



# Article Multi-Step Short-Term Wind Speed Prediction Models Based on Adaptive Robust Decomposition Coupled with Deep Gated Recurrent Unit

Kui Yang <sup>1</sup>, Bofu Wang <sup>2,3,\*</sup>, Xiang Qiu <sup>1</sup>, Jiahua Li <sup>4</sup>, Yuze Wang <sup>5</sup> and Yulu Liu <sup>1,2</sup>

- <sup>1</sup> School of Science, Shanghai Institute of Technology, Shanghai 201418, China; yangcdk@163.com (K.Y.); qiux@sit.edu.cn (X.Q.); ylliu@sit.edu.cn (Y.L.)
- <sup>2</sup> Shanghai Key Laboratory of Mechanics in Energy Engineering, Shanghai Institute of Applied Mathematics and Mechanics, School of Mechanics and Engineering Science, Shanghai Frontiers Science Base for Mechanoinfomatics, Shanghai University, Shanghai 200072, China
- <sup>3</sup> Guangdong Provincial Key Laboratory of Turbulence Research and Applications, Southern University of Science and Technology, Shenzhen 518055, China
- <sup>4</sup> School of Urban Construction and Safety Engineering, Shanghai Institute of Technology, Shanghai 201418, China; lijiahua@sit.edu.cn
- <sup>5</sup> School of Mechanical Engineering, Shanghai Institute of Technology, Shanghai 201418, China; wang18321226730@163.com
- \* Correspondence: bofuwang@shu.edu.cn

Abstract: Accurate wind speed prediction is a premise that guarantees the reliable operation of the power grid. This study presents a combined prediction model that integrates data preprocessing, cascade optimization, and deep learning prediction to improve prediction performance. In data preprocessing, the wavelet soft threshold denoising (WSTD) is employed to filter the blurring noise of the original data. Then, the robust empirical mode decomposition (REMD) and adaptive variational mode decomposition (AVMD) are adopted to carry out a two-stage adaptive decomposition. Spearman correlation is used to quantify the mode that need to be decomposed for the second time. In the cascade optimization, the hybrid grey wolf algorithm (HGWO) is employed to optimize the parameters of the VMD and the gated recurrent unit (GRU), which overcomes the problem of empirical parameter adjustment. The HGWO is also adopted in the prediction strategy to optimize the GRU model to predict the grouped intrinsic mode functions (IMFs). Lastly, the final wind speed prediction result is obtained by superimposing the values of all the predicted models. The proposed model was validated with the measured wind speed data of the four quarters in the Bay area of China and was compared with 20 models of the classic method to further evaluate the effectiveness of the model. The results show that the whole process of the proposed model is adaptive, the final multi-step prediction performance is good, and high prediction accuracy can be attained.

**Keywords:** wind speed prediction; wavelet soft threshold denoising; robust empirical mode decomposition; cascade optimization strategy; deep gated recurrent unit; adaptive model

# 1. Introduction

Wind energy is a highly efficient and renewable energy. Its development and utilization have been widely recognized [1]. The intermittent and randomness of wind speed bring stern challenges for the stable operation of power systems [2,3]. Accurate prediction of wind speed will benefit the exploitation of wind energy. Wind speed prediction is a hot research topic, and a number of prediction methods have been proposed over the past decades which can be classified into physical models, statistical models, machine learning models, and combination models [4].

As wind is a multi-scale physical phenomenon, numerical weather predictions (NWP) provide straightforwad wind speed predictions with physical advantages [5]. The physical



Citation: Yang, K.; Wang, B.; Qiu, X.; Li, J.; Wang, Y.; Liu, Y. Multi-Step Short-Term Wind Speed Prediction Models Based on Adaptive Robust Decomposition Coupled with Deep Gated Recurrent Unit. *Energies* **2022**, *15*, 4221. https: //doi.org/10.3390/ en15124221

Academic Editor: Galih Bangga

Received: 20 April 2022 Accepted: 4 June 2022 Published: 8 June 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). model based on NWP applies historical meteorological factors such as temperature and barometric pressure to predict the wind speed over a long period [6]. Additionally, wind speed prediction is applied using physical approximation and spatial correlation fluid dynamics models [7,8]. However, the physical model is complex in modeling, poor in prediction accuracy, and its scope of application is limited. It is not suitable for short-term wind speed prediction. Recently, a data-driven method based on stacked bidirectional long short-term memory (BiLSTM) for wind turbine wake prediction was proposed by Geibel and Bangga [9], and satisfactory prediction accuracy was achieved, indicating that a data-driven approach can offer an alternative to conventional prediction methods.

Statistical models use the linear mapping relationship between historical weather conditions and wind speed and utilize series data to predict the future wind speed series through this relationship [10]. Traditional statistical modes such as auto-regression (AR) [11], auto-regressive moving average (ARMA) [12], auto-regressive integrated moving average (ARIMA), and auto-regressive conditional heteroskedasticity (ARCH) are widely used in the field of time series prediction [13]. For instance, the ARCH model was used to predict the revenues of the financial stock market [14]. Radziukynas adopted a classical ARIMA model to predict short-term wind speed series [15]. Tian and Wang et al. proposed a prediction method based on ARMA and echo state network (ESN) compensation to deal with the statistical characteristics of wind speed series, which can obtain accurate prediction results [16]. As mentioned above, the statistical models often uses the historical wind speed data to establish linear wind speed prediction models, but due to the strong nonlinearity of data, the model parameters are difficult to determine, and the prediction accuracy is not satisfied.

To improve the accuracy of the wind speed prediction, statistical models based on artificial intelligence were proposed. Compared with statistical models, artificial intelligence prediction models have outstanding capabilities in processing nonlinear wind speed time series. With the rapid development of artificial intelligence, great effort has been made to improve the accuracy and universality of wind speed prediction. Nair and Jisma presented ANN and ARIMA to predict the wind speed at three different locations in India in different time periods, which reduces the nonlinear characteristics of the wind speed series [17]. In another work, Shukur and Lee proposed the Kalman filter (KF) and an ANN hybrid model based on ARIMA to deal with the nonlinearity and uncertainty of wind speed [18]. Machine learning models based on neural network have became popular in short-term wind speed prediction [19]. For example, the prediction models based on back-propagation neural networks (BPNN) [20], least square support vector machines (LSSVM) [21], support vector regressions (SVR) [22], extreme learning machines (ELM) [23], Elman neural networks (ENN) [24], adaptive wavelet neural networks (AWNN) [25], and recurrent neural networks (RNN) [26] perform well when dealing with time series with nonlinear characteristics. However, these techniques are all built with single neural network models, which may result in local optimization or over-fitting problems in short-term wind speed prediction.

To improve the prediction performance, researchers introduced combination models to integrate the advantages of every single model. The combination prediction model combines data preprocessing, an optimization algorithm, a and predictor, which shows outstanding performance in wind speed prediction. The data preprocessing strategy composed of outlier detection and data decomposition can greatly ameliorate the prediction performance of the whole model [27]. The missing original data and outlier problems caused by human, weather, and other factors are the primary problems to be addressed in the field of wind speed prediction. After outlier processing, taking into account the nonlinearity and noise characteristics of wind speed [28], the data decomposition method can drop the instability of wind speed correlation series and abandon redundant information and combine with a machine learning model to enhance the predictability of wind speed [29]. For instance, Mi and Zhao [30] applied the singular spectrum analysis (SSA) model to denoise the original wind speed data to capture the complex dynamic characteristics of

wind speed, which combines the adaptive structure learning of neural networks with longand short-term memory networks (LSTM) to predict wind speed in three wind farms in Xinjiang, China. The proposed model has good prediction performance. A hybrid method of empirical mode decomposition (EMD) and ARIMA-ANN is proposed to improve the prediction accuracy of time series [31]. Zhan and Tian et al. [32,33] studied the two-stage decomposition model of the complementary ensemble empirical mode decomposition (CEEMD) and local mean decomposition (LMD) to achieve intrinsic mode functions (IMFs) of different regularity degrees, which was applied to (support vector machine) SVM and T-S fuzzy neural network (FNN) prediction. Duan et al. [34] proposed variational mode decomposition (VMD) to extract the local characteristics of the original wind speed series and constructed an integrated prediction model using deep belief network (DBN) optimized by particle swarm optimization (PSO), which overcomes the shortcomings of linear weighted combination, and the performance of the prediction model is better than many traditional models. Meanwhile, the wavelet transform (WT) [35] method shows advantages in extracting and studying the characteristics of wind speed in the time domain and frequency domains and solves the randomness and complexity of wind speed signals. Liu et al. [36] designed a hybrid model combining wavelet decomposition (WD) and LSTM to predict China's wind power generation in the next two years. The experiment showed that this model effectively improves the accuracy of the prediction.

In addition to the above models, the decomposition-based method may contribute to large differences in prediction performance. For the decomposed models, in terms of generalization performance, sometimes they cannot capture the characteristics of wind speed. As such, the prediction accuracy and training speed of the model are affected because the advantage of parameter optimization is not considered. To avoid these problems, an adaptive short-term wind speed predictor is formed based on the model of optimization algorithm [37], which reduces the prediction error and achieves good prediction results. As an example, Bai et al. [38] proposed a dynamic integrated wind speed prediction model, a hybrid model composed of VMD and a genetic algorithm (GA)-optimized double-layer staged training echo state network (DESN), which used the DESN to process nonlinear series and capture time information of different time scales, and the model has better timevarying and robustness through nonlinear weighted combination mechanism. Wu and Wang et al. [39], considering the accuracy and stability of the model, applied multi-objective grey wolf optimization (MOGWO) to optimize ELM to form a new integrated global wind speed prediction method. Neshat [40] combined effective hierarchical decomposition technology and deep learning optimization methods to develop a combined model with deep feature selection and optimal intrinsic mode functions to predict the forward time step of wind speed data from Baltic offshore wind farms. Tian et al. [41] utilized EMD to decompose the original wind speed data into IMFs with different frequencies, and then embedded multiple IMFs into the enhanced network of improved sparrow search algorithm (ISSA) optimized LSTM for prediction, which solved the problem of slow convergence speed of previous models and being easy to fall into local optimum. The results indicate that the model of EMD and ISSA optimization LSTM has good predictive ability. The literature review shows that using the diversity of optimization algorithms to predict the wind speed prediction model with the best parameters, which can improve the prediction accuracy and stability of the model to a certain extent.

Short-term prediction is to predict wind speed from 10 min to 30 min in advance. It is conducive to the timely and reasonable dispatching of the power grid, maintenance of power quality, and the stable operation of the power system. Because of the uncertainty of wind speed, short-term wind speed prediction has great practical significance and application value. In the field of short-term wind speed prediction, wind speed is an important indicator that affects wind power generation. However, due to many uncertain factors, the integrity of the original wind speed data has been destroyed. The rapid change in wind speed leads to the nonlinear and nonstationary characteristics of wind speed series. EMD is an adaptive data decomposition method [42] suitable for processing nonlinear

and unstable time series, but its decomposition has the problem of model-mixing and the uncertainty of the maximum number of iterations in the sifting process, which affects the performance of the wind speed data decomposition method. Hence, VMD has a better decomposition effect and is more robust when dealing with complex wind speed time series [43]. What is undesirable is that VMD requires artificial parameters and lacks rigor. Therefore, in consideration of the completeness and predictability of wind speed data, data preprocessing strategy plays an important role in the study of short-term wind speed prediction.

Reviewing the above-mentioned methods, the combined method based on the decomposition denoising method and the optimization parameter method are discussed, and the research contributions of these methods and their shortcomings are summarized. To this end, this study considers the generalization ability and stability of the prediction model and develops a short-term multi-step wind speed prediction method with adaptive robust decomposition characteristics that combines a data-processing strategy, a cascade optimization strategy, and a prediction strategy [44]. The method includes a data preprocessing strategy based on wavelet soft threshold denoising (WSTD), robust empirical mode decomposition (REMD) and variational mode decomposition (VMD), cascade optimization based on the hybrid grey wolf optimization algorithm (HGWO) strategy, and a prediction strategy based on deep gated recurrent unit (DGRU), which is achieved satisfactory results in the field of short-term wind speed prediction. The primary innovations and contributions of this research are as follows:

- (1) Data preprocessing strategy: A novel and efficient two-stage data preprocessing technology is proposed. WSTD filters out the redundant noise of the original wind speed series. One-stage REMD decomposes to obtain a series of IMFs to eliminate random fluctuations. To reduce the error, Spearman correlation analysis is used to analyze the correlation between each IMF and the original wind speed time series, group reconstruction, reduce the accumulation of errors, and prepare high-quality data for prediction purposes.
- (2) Cascade optimization strategy: The cascading optimization strategy based on HGWO, which is used for the first time, and the optimized VMD is used to decompose the IMFs with strong correlation in the wind speed correlation series in the second stage to further explore the potential characteristic information of the wind speed. On this basis, it is more robust to deal with time series of complex characteristics.
- (3) Prediction strategy: The strategy of cascading optimization is adopted to dynamically analyze the optimal input parameters and optimal network structure of the GRU deep learning model, and the reorganized wind speed correlation sub-series are predicted and superimposed in the future time step to complete deeper wind speed characteristic extraction and learning, which greatly enhance the stability and generalization of the model.
- (4) The combined multi-step wind speed prediction method of WSTD, REMD, and HGWO-VMD-GRU is proposed, which integrates the advantages of each single model. The wind speed datasets of different seasons in the Shanghai Bay area are selected to verify the validity of the model, and the final conclusion is reached by testing and analyzing three different benchmark models with the classic single models, the decomposition optimization models, and other combined models.

The paper is organized as follows. In Section 2, we introduce the basic theory of the relevant methods and the framework of the proposed prediction system. Section 3 gives a comprehensive discussion on the experimental results from various prediction models. In Section 4, we draw the conclusion.

## 2. Related Methodology

An adaptive decomposition integrated combined wind speed prediction system has been developed in this study. The basic framework of the proposed model is shown in Figure 1. First, a wind speed data preprocessing strategy based on WSTD and REMD is proposed. Meanwhile, the Spearman correlation coefficient is used to quantitatively analyze the wind speed component after pretreatment to reduce the accumulation of errors. Secondly, the cascade optimization strategy of VMD and GRU is optimized based on HGWO. Finally, the prediction strategy based on GRU is proposed. It can be observed that each integrated model has adaptive characteristics, which makes the whole prediction system more robust and accurate. The strategies adopted for data preprocessing, cascade optimization and building prediction modeling are introduced in detail in the following.



Figure 1. Overall block diagram of the proposed combined model.

## 2.1. Data Preprocessing

#### 2.1.1. Data Collection

The wind speed series used in this study are measured from the Shanghai Bay area in China in 2019 and were provided by Fengxian Meteorological Bureau. Four different short-term wind speed samples in the four seasons are selected as datasets. The four datasets are termed as dataset A, B, C, and D. The data sampling interval is 10 min, and there are 2000 data in each dataset, as shown in Figure 2. The average, standard deviation, median, variance, and box plot are shown in the figure as well. The difference among the four datasets are reflected from the statistics.



Figure 2. Basic information of the four datasets used in this study.

## 2.1.2. Data Denoising Based on WSTD

WSTD is a time-frequency domain signal processing method based on wavelet transform (WT) [45]. WT is usually used for denoising or filtering of strongly nonstationary signals. There are a variety of criteria for selecting wavelet bases and thresholds in WT, and an inappropriate selection will disrupt the performance. Suppose the measured noisy wind speed data are f(t). They can be decomposed as:

$$f(t) = x(t) + \delta e(t), \tag{1}$$

where *t* is the time interval, x(t) is the real wind signal, e(t) is the Gaussian white noise,  $\delta$  represents the correlation coefficient of the noise.

WSTD generally includes three steps, i.e., WT, threshold function, and wavelet reconstruction. In WT, the denoising effect is quantified by the signal-to-noise ratio (SNR) and the RMSE. The larger the SNR or the smaller the RMSE is, the better the noise reduction. The denoising effect of WT was tested on dataset C. Table 1 demonstrates the SNR and RMSE obtained with different wavelet basis functions. The results suggest that WT with db4 wavelet basis function shows the best noise reduction.

Table 1. Comparison of WT with various wavelet basis functions.

Wavelet Basis Function	SNR/dB	RMSE
db4	12.8588	0.5677
haar	11.3318	0.6768
db3	12.5165	0.5905
sym2	12.3492	0.6020

Threshold function is used to quantify the high-frequency coefficients at different scales obtained by WT. The selection of the threshold function has a great influence on the denoising results. The two well-known threshold functions are the hard threshold function and the soft threshold function. The hard threshold function is not continuous, which leads to additional oscillations in the signal reconstruction. On the contrary, the soft threshold function maintains good continuity by flexibly managing the discontinuity of the hard threshold in various threshold estimation methods and enhances the effect of denoising in the wind speed prediction models. The soft threshold function is:

$$\hat{W_X} = \begin{cases} sign(W_X)|W_X|, & |W_X| \ge \lambda \\ 0, & |W_X| < \lambda \end{cases}$$
(2)

where  $W_X$  is the WT coefficient,  $\hat{W}_X$  is the calculated WT coefficient,  $\lambda$  indicates the threshold, *sign* denotes the symbolic function.

Finally, the denoised wind speed time series is obtained by wavelet reconstruction, which reconstructs the processed high- and low-frequency wavelet coefficients.

Overall, the denoising effect of WSTD is better than the traditional denoising methods. The denoised datasets by WSTD are shown in Figure 3, and it is observed that the wind speed time series are smoother than the original ones.



**Figure 3.** The four datasets denoised by WSTD. The vertical dashed lines denote the separation of training data and test data.

## 2.1.3. One-Stage Decomposition Based on REMD

The nonlinearity and nonstationarity of wind speed seriously affect its forecast accuracy. The proper decomposition of the original wind speed data can effectively alleviate the complex characteristics of the wind speed and improve the prediction accuracy. The usually used decomposition methods such as EMD and EEMD require manually preset thresholds, which are not adaptive, resulting in mode-mixing problems after decomposition and affecting the prediction results.

The implementation of EMD is mainly based on three parameters: envelope estimation, boundary condition and sifting stop criterion (SSC). The classical cubic spline interpolation method and mirror extension method are adopted for envelope estimation and boundary condition. Among them, the SSC parameter has the greatest impact on the accuracy and efficiency of EMD. This parameter directly affects the sifting iterations number and then controls the decomposition component of EMD, but it opts to cause the 'undersifting' or 'over-sifting' phenomenon. Therefore, the study of SSC is essential to deal with the mode-mixing problem in EMD. The definition of SSC parameter is as follows: (1) the number of extreme points and zeros must be equal or no more than one, i.e.,  $|N_{zores} - N_{extreme}| \leq 1$ ; (2) if the envelope mean signal  $m_{ik}[n]$  of EMD is 0, the sifting process stops, that is,  $\lim_{k\to\infty} m_{ik}[n] = 0$ . The objective function is designed to describe the envelope mean signal, and the RMS and EK are taken into account, so that all sample points of the signal uniformly tend to zero. Based on this, Liu et al. [46] exploited SSC for adaptive control of the sifting process, which is used to ease the mode-mixing phenomenon and improve the decomposition results, so as to improve the accuracy of wind speed prediction. The improved EMD performs robust decomposition of wind speed data owing to its adaptability. The original wind speed series is written as:

$$x(t) = \sum_{k=1}^{i} IMF_{K}(t) + r_{i}(t)$$
(3)

The detailed processes of robust empirical mode decomposition (REMD) are as follows:

- Step 1: Initialize parameters k and i, set the maximum number of sifting iterations  $I_{max}$ ;
- Step 2: Find the maximum and minimum values of the wind speed signal  $h_{ik-1}[n]$ . The upper and lower envelopes are obtained by cubic spline interpolation. Then, calculate the average value of the upper and lower envelopes  $m_{ik}[n]$ :

$$h_{ik}[n] = h_{ik-1}[n] - m_{ik}[n]$$
(4)

where  $h_{ik}[n]$  indicates the signal after the *i*th IMF sifting *k* times;

Step 3: Apply the objective function of SSC to calculate the objective value  $f_{ik}$ . The objective function is defined as follows:

$$f_{ik} = RMS(m_{ik}[n]) + |EK(m_{ik}[n])|$$
(5)

$$RMS_{ik} = \sqrt{\frac{1}{N_s} \sum_{n=1}^{N_s} (m_{ik}[n])^2}$$
(6)

$$EK_{ik} = \frac{\frac{1}{N_s} \sum_{n=1}^{N_s} (m_{ik}[n] - \bar{m})^4}{\left[\frac{1}{N_s} \sum_{n=1}^{N_s} (m_{ik}[n] - \bar{m})^2\right]^2} - 3$$
(7)

where  $\bar{m}$  stands for the arithmetic mean of  $m_{ik}[n]$ ;

Step 4: Execute SSC to determine the sifting stop process. If it is satisfied at the same time, stop and output; otherwise, return to step 2 and continue to iterate until the maximum number of sifting iterations  $I_{max}$  is received, and output the k-2nd  $h_{ik-2}[n]$  as  $IMF_i$ . The two criteria are expressed as follows:

$$\begin{cases} f_{ik-2} < f_{ik-1}, f_{ik-1} < f_{ik} \\ abs(N_{zores} - N_{extreme}) \le 1 \end{cases}$$

$$\tag{8}$$

## 2.2. Cascade Optimization

## 2.2.1. The Hybridizing Grey Wolf Optimization Algorithm

The HGWO uses the differential evolution (DE) to carry out population mutation in order to maintain the diversity of the population and then takes it as the initial population of the GWO to find the optimal individual [47]. The crossover and selection operation of DE are used to update the positions of other grey wolf individuals, and the cycle process is used to obtain the optimal solution. Step 4 in Figure 1 shows the flow chart of hybrid grey wolf algorithm. The process of DE and GWO are presented below.

DE is a random selection model that simulates biological evolution, which is used to solve global optimization problems. The implementation of DE includes initialize population, mutation, crossover, and selection operations as explained below.

Initialize population: Randomly generate population individuals.

Mutation: The operation of the mutated individual is implemented as follows:

$$h_{ij}^{t+1} = X_{p1}^t + F(X_{p2}^t - X_{p3}^t)$$
(9)

where *F* is the scaling factor; *t* is the current iteration number;  $p_1$ ,  $p_2$ , and  $p_3$  are a random unequal integers in *N* that are not equal to *t*; and *N* is the population size.

Crossover: Crossover operation can increase the diversity of the population, the operation is applied as below:

$$U_{ij}^{t+1} = \begin{cases} h_{ij}^{t+1}, & rand(j) \le CR \text{ or } j = rand(1,n) \\ X_{ij}^t, & rand(j) > CR \text{ or } j \ne rand(1,n) \end{cases}$$
(10)

where *CR* represents the crossover probability between 0 and 1, *j* is a random integer between 1 and *D*, and *D* is the dimension of the solution.

Selection: Utilize a greedy strategy to select the offspring of an individual, as defined below:

$$X_{i}^{t+1} = \begin{cases} U_{i}^{t+1}, & f(U_{i}^{t+1}) < f(X_{i}^{t}) \\ X_{i}^{t}, & f(U_{i}^{t+1}) \ge f(X_{i}^{t}) \end{cases}$$
(11)

where  $X_i^{t+1}$  is the *i*th current population individual in the *t*+1 generation population, and  $U_i^{t+1}$  is the *i*th crossover individual in the *t*+1 generation population.

According to the DE algrithm, the superiority of the progeny population can be guaranteed, and the average performance of the population can be improved.

GWO is an intelligent optimization algorithm inspired by the predation behavior of grey wolves in nature. With its advantages of simple mechanism, strong global search ability, and few adjustable parameters, it achieves parameter optimization based on the mechanism of wolf group cooperation. Wolves have a very strict social hierarchy. As shown in Figure 4, the wolf pack is divided into four levels to simulate leadership levels. The social hierarchy of grey wolves plays an important role in pack hunting. The predation process is completed under the leadership of Wolf  $\alpha$ , Wolf  $\beta$ , and Wolf  $\zeta$ , and the existence of Wolf  $\omega$  maintains the stability of the pack hierarchy. The hunting of grey wolves includes the following three main components:



Figure 4. The hierarchy of the grey wolf pack.

Pursuing prey: The behavior of grey wolves chasing prey is defined as follows. The distance D between the two is expressed as:

$$\vec{D} = |\vec{C} \cdot \vec{X}_p(t) - \vec{X}(t)| \tag{12}$$

$$\vec{C} = 2\vec{r}_1 \tag{13}$$

where *t* denotes the current iteration,  $\vec{A}$  and  $\vec{C}$  are coefficient vectors,  $\vec{X}(t)$  represents the position of the grey wolf after iteration *t*,  $\vec{X}_p(t)$  stands for the position of prey after iteration *t*, and  $\vec{C}$  is the swing factor.

The grey wolves' position update formula is described as follows:

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A}\vec{D} \tag{14}$$

$$\vec{A} = 2\vec{a}\vec{r}_2 - \vec{a} \tag{15}$$

where  $\vec{r}_1$  and  $\vec{r}_2$  are random vectors in [0, 1],  $\vec{A}$  is the convergence factor, and  $\vec{a}$  decreases linearly from 2 to 0 as the number of iterations increases.

Rounding up prey: Grey wolves identify the location of their prey and encircle their prey. This behavior is depicted as follows:

$$\vec{D}_{\alpha} = |\vec{C}_1 \vec{X}_{\alpha}(t) - \vec{X}(t)| \tag{16}$$

$$\vec{D}_{\beta} = |\vec{C}_2 \vec{X}_{\beta}(t) - \vec{X}(t)|$$
(17)

$$\vec{D}_{\zeta} = |\vec{C}_3 \vec{X}_{\zeta}(t) - \vec{X}(t)| \tag{18}$$

Attack prey: Prey stops moving, and the grey wolf attacks to complete the hunt:

$$\vec{X}_1 = \vec{X}_\alpha(t) - \vec{A}_1 \vec{D}_\alpha \tag{19}$$

$$\vec{X}_2 = \vec{X}_\beta(t) - \vec{A}_2 \vec{D}_\beta$$
 (20)

$$\vec{X}_3 = \vec{X}_{\zeta}(t) - \vec{A}_3 \vec{D}_{\zeta} \tag{21}$$

$$\vec{X}_p(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3}$$
(22)

where  $\bar{X}_{\alpha}$ ,  $\bar{X}_{\beta}$ , and  $\bar{X}_{\zeta}$  indicate the position vectors of  $\alpha$ ,  $\beta$ , and  $\zeta$  wolves, respectively, and  $\vec{D}_{\alpha}$ ,  $\vec{D}_{\beta}$ ,  $\vec{D}_{\zeta}$  are the distances from the current wolf to the three optimal solutions, respectively.

To avoid the defects of local optimization and poor stability of DE and GWO, the hybrid of the two algorithms, namely, DE-GWO, improves the global search ability and shows the adaptive and fast convergence ability in wind speed prediction.

## 2.2.2. Two-Stage Decomposition Based on VMD

VMD is a signal decomposition method for nonlinear and nonstationary characteristics proposed by Dragomiretski et al. [48]. It has the characteristics of being adaptive, quasiorthogonal, and completely nonrecursive, the essence of which is to construct and solve the variational problem. The VMD method is as follows:

To obtain a unilateral spectrum, the means of the Hilbert transform are adopted to calculate the associated analytic signal for each mode:

$$(\delta(t) + \frac{j}{\pi t}) * u_k(t) \tag{23}$$

The center frequency of each mode is assessed:

$$[(\delta(t) + \frac{j}{\pi t}) * u_k(t)]e^{-j\omega_k t}$$
(24)

The L2 norm of the gradient for the demodulated signal is employed to calculate the bandwidth of each mode. The variational constraint problem is constructed as follows:

$$\min_{u_k,\omega_2} \sum_k \|\partial_t [(\delta(t) + \frac{j}{\pi t}) \cdot u_k(t)] e^{-j\omega_k t}\|_2^2$$
(25)

$$\sum_{k} u_k(t) = f(t) \tag{26}$$

where f(t) represents the original signal,  $u_k$  is each modal component,  $\omega_k$  indicates the central frequency of each modal,  $\delta(t)$  denotes the unit impulse function, and *K* is the total number of modal components.

Solving variational problems: Introduces the Lagrange multipliers  $\lambda$  and quadratic penalty factors  $\alpha$ , so that the constructed constrained variational solution problem becomes an unconstrained variational decomposition problem. The augmented Lagrange function of the original constrained variational solution problem can be expressed as:

$$L(u_{k}, \omega_{k}, \lambda) = \alpha \sum_{k=1}^{K} \|\partial_{t} [(\delta(t) + \frac{j}{\pi t}) \cdot u_{k}(t)] e^{-j\omega_{k}t} \|_{2}^{2} + \|f(t) - \sum_{k=1}^{K} u_{k}(t)\|_{2}^{2} + \langle \lambda(t), f(t) - \sum_{k=1}^{K} u_{k}(t) \rangle$$
(27)

The Alternating Direction Multiplier Method (ADMM) is used to search the saddle point of the augmented Lagrange function to solve the variational problem, where each mode  $u_k$ , center frequency  $\omega_k$ , and Lagrange operator  $\lambda$  are iteratively updated as follows:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega)\sum_{ik}\hat{u}_k^n(\omega) + \frac{\lambda(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2}$$
(28)

$$\omega_k^{n=1} = \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k^{n+1}(\omega)|^2 d\omega}$$
(29)

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau(\hat{\omega} - \sum_k \hat{u}_k^{n+1}(\omega))$$
(30)

where  $\tau$  is the update parameter.

If the following discriminant formula is met, stop the iteration; otherwise, repeat the above operation:

$$\sum_{k} \frac{\|\hat{u}_{k}^{n+1}(\omega) - \hat{u}_{k}^{n}(\omega)\|_{2}^{2}}{\|\hat{u}_{k}^{n}(\omega)\|_{2}^{2}}$$
(31)

VMD exhibits some advantages over traditional recursive mode decomposition in signal decomposition problems, and the decomposition result is determined by the number of modal components and the secondary penalty factor. The remaining parameters have weak effect on the decomposition result and are generally set with default values. However, the choice of the two key parameters is empirical. Usually, the optimal combination of the parameters is not adopted. The adaptive decomposition of VMD optimized by HGWO overcomes this deficiency, which provides the best decomposition as well as sufficient feature mining of the data.

In this study, the decomposition performance of VMD is measured with the minimum envelope entropy (MEE). Taking the MEE as the fitness function of HGWO algorithm, under the influence of a set of modal decomposition parameters and the combination of the secondary penalty factor parameters, the one with the smallest entropy value is selected as the MEE, which is recorded as the best fitness function MEE. Therefore, the parameter combination corresponding to the global optimal component is obtained, namely, *K* and  $\alpha$ . The steps to solve the *MEE* are as follows:

$$b_j = \frac{a_j}{\sum_{j=1}^M b_j}, j = 1, 2, \cdots, M$$
 (32)

$$IMF_{EE}(k) = -\sum_{j=1}^{M} b_j \log_2(b_j)$$
 (33)

$$MEE = \min(IMF_{EE}(1), \cdots, IMF_{EE}(k))$$
(34)

where *M* is the length of the modal component of the wind speed time series by VMD,  $b_j$  is the normalized envelope of the modal components,  $a_j$  is the envelope amplitude of the *j*th point of the modal component after VMD, and  $IMF_{EE}(k)$  is the envelope entropy of a modal component.

The parameter determination algorithm based on the MEE is used to find the optimal combination of VMD parameters and the fitness function set by HGWO, which solves the problem of the under-decomposition and over-decomposition of the modal caused by the improper selection of VMD parameters, and forms an adaptive wind speed decomposition method.

VMD is commonly used to perform data preprocessing in the field of wind power prediction, which can extract abundant characteristic information from wind speed signals. In this study, the local characteristics of the wind speed time series are extracted through the secondary decomposition of VMD with the purpose of tracking wind speed change information in real time.

## 2.3. Prediction Model

## Deep Gated Recurrent Unit

GRU is a network model for processing time series. It is a variant of RNN [49], which simplifies the network structure while ensuring the original accuracy of LSTM, has strong learning ability and efficiency, and can selectively forget or retain the input historical information through the update gate and reset gate, so as to realize the prediction of future data.

The GRU model can control the degree of the state information of the previous moment by updating and resetting the gate. The GRU network structure is shown in Figure 5. The parameters that the entire model needs to be trained are  $W_z$ ,  $W_r$ , and  $W_h$ , and the forward training process is:

$$z_t = \sigma(W_{zh}h_{t-1} + W_{zx}x_t) \tag{35}$$

$$r_t = \sigma(W_{rh}h_{t-1} + W_{rx}x_t) \tag{36}$$

$$\tilde{h}_t = tanh(W_{hh}(r_t \odot h_{t-1}) + W_{hx}x_t)$$
(37)

$$h_t = (1 - z_t) \bigodot h_{t-1} + \tilde{h}_t \bigodot z_t \tag{38}$$

where  $W_{zh}$ ,  $W_{zx}$  is the weight matrix of the updated gate,  $W_{rh}$ ,  $W_{rx}$  is the weight matrix of the reset gate,  $W_{hh}$ ,  $W_{hx}$  is the weight matrix when calculating the output candidate value  $h_t$ , and  $\odot$  is the product of the elements of the matrix.



Figure 5. GRU structure.

The concrete procedures of HGWO-GRU (DGRU) are described as follows:

- Step 1: Define the parameters of HGWO, such as population size N, maximum number of iterations  $t_{max}$ , and crossover probability parameters CR.
- Step 2: Initialize the parameters *a*, *A*, *C*, implement DE mutation and competitive selection on the population individuals according to the formula, and generate the initial population.
- Step 3: Apply formula (12)–(21), calculate the objective function value of each grey wolf individual in the population, and select the positions of the three individual grey wolves  $\vec{X}_{\alpha}$ ,  $\vec{X}_{\beta}$ , and  $\vec{X}_{\zeta}$  with the optimal value. Then, calculate the distance between other grey wolves in the population and the optimal individual position, and update the current position.
- Step 4: According to the formula (9)–(11), cross the positions of individuals and screen out new individuals.
- Step 5: Perform formula (22), calculate the target fitness value of all grey wolf individuals, and update the grey wolf individuals in the three optimal positions of  $\vec{X}_{\alpha}$ ,  $\vec{X}_{\beta}$ , and  $\vec{X}_{\zeta}$ .
- Step 6: Cycle process, judge whether the maximum number of iterations is reached; if so, save the global optimal solution and exit. Otherwise, return to step 3 to continue the iterative update.

Step 7: Output the optimal position, that is,  $\vec{X}_{\alpha}$ . The GRU network determines the optimal combination parameters (GRU-size, Learning-rate).

For the construction of prediction model, the rolling prediction mechanism is used in training set and test set. Each input of 3 measured data corresponds to 1 output as the predicted value. Accordingly, 2000 data are formed into 1997 datasets, of which the first 1800 sets of data are used as training sets. The remaining 197 sets of data are used as the test set. In addition, before entering the model, the wind speed time series are normalized, and, finally, the future wind speed data are predicted by one, two, and three step ahead. The expressions of the one-step and multi-step predictions are as follows. Assuming the wind speed series is  $(x_1, x_2, ..., x_n)$ , the length of Moving Window is *h*. In m-step prediction, the moving step length of the Moving Window is *m*. Note that the length h = 3 is adopted in our following predictions based on multiple trials.

The prediction formula at time *t* in one-step prediction is:

$$(x_1, x_2, \cdots, x_h) \Rightarrow (y_{h+1}), t = 1$$
(39)

$$(x_2, x_3, \cdots, x_{h+1}) \Rightarrow (y_{h+2}), t = 2$$
 (40)

$$(x_{1+k-1}, x_{1+k}, \cdots, x_{h+k-1}) \Rightarrow (y_{h+k}), t = k$$
 (41)

The prediction formula at time *t* in m-step prediction is:

$$(x_1, x_2, \cdots, x_h) \Rightarrow (y_{h+1}, \cdots, y_{h+m}), t = 1$$

$$(42)$$

$$(x_{1+m}, x_{2+m}, \cdots, x_{h+m}) \Rightarrow (y_{h+m+1}, \cdots, y_{h+2m}), t = 2$$
 (43)

$$(x_{1+(k-1)m}, x_{2+(k-1)m}, \cdots, x_{h+(k-1)m}) \Rightarrow (y_{h+(k-1)m+1}, \cdots, y_{h+kn}), t = k$$
 (44)

## 3. Experimental Results and Discussion

This section introduces the Evaluation index, Denoising verification, Model parameters selection, and Experimental results.

#### 3.1. Evaluation Index

To evaluate the performance of the prediction model, five evaluation indicators are used in this paper, which are the Mean absolute error, Root mean square error, R-squared, Theil inequality coefficient, and Square sum error. The indicators are listed in Table 2, where  $y_i$  is the *i*th actual value of short-time wind speed,  $\hat{y}_i$  is the corresponding predicted value from the prediction model for performance estimation, and  $\hat{y}$  is the mean value of the short-time wind speed.

Table 2. The performance evaluation index.

Index	Meaning	Equation
SNR	Signal-to-noise ratio	$SNR = 101g(\frac{\sum_{i=1}^{N} y_i^2}{\sum_{i=1}^{N} (y_i - \hat{y_i})^2})$
MAE	Mean absolute error	$MAE = rac{1}{N} \sum_{i=1}^{N}  y_i - \hat{y}_i $
RMSE	Root mean square error	$RMSE = \sqrt{rac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y_i})^2}$
$R^2$	Coefficient of determination	$R^2 = 1 - rac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \hat{y})^2}$
TIC	Theil inequality coefficient	$TIC = \frac{\sqrt{\frac{1}{N}\sum_{t=1}^{N}(y_i - \hat{y}_i)^2}}{\sqrt{\frac{1}{N}\sum_{t=1}^{N}y_i^2} + \sqrt{\frac{1}{N}\sum_{t=1}^{N}\hat{y}_i^2}}$
SSE	Square sum error	$SSE = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$

### 3.2. Denoising Verification

In order to verify the improvement in wind speed prediction by denoising, the predicted results by BPNN from the original data and the WSTD denoised data are compared. BPNN is a classical neural network with adaptive learning and an error feedback mechanism. The four datasets are trained and tested by BPNN, and the prediction results are shown in Figure 6. It can be seen from the figure that the data quality after WSTD processing is significantly improved, the data become smoother, and the redundant information is filtered out. Moreover, the noise inhibited in the datasets reduces the prediction accuracy, whereas the the prediction performance is significantly improved after denoising. This experiment suggests that WSTD is essential to the hybrid wind speed prediction model.



**Figure 6.** Comparison of BPNN prediction results of four wind speed series with and without denoising: (a) without the WSTD denoising; (b) with the WSTD denoising.

## 3.3. Experimental Results

In this section, three simulation experiments based on a decomposition optimization model, a decomposition model, and a classical model are presented. The four wind speed datasets measured in the bay area of Shanghai, China, are used to train and test the developed model. The one-step, two-step, and three-step predictions are performed. More specifically, the REMD, VMD, CEEMD, and WD models are used as benchmark models for comparison. Furthermore, the results of one-stage decomposition of REMD in Section 2.1.3 and two-stage decomposition of HGWO-VMD in Section 2.2 are analyzed in more detail. The dataset C is taken as an example for demonstration and discussion.

#### 3.3.1. Validation of REMD Method

Proper decomposition can effectively reduce the random fluctuation in the original wind speed series and improve the performance of wind speed prediction. By taking the advantages of the REMD method with SCC, the mode-mixing problem is avoided, and the adaptive robust decomposition is utilized. Then, REMD is used to decompose the original wind speed series into IMFs of different frequencies for subsequent model analysis and prediction. Figure 7 shows the decomposition results of REMD. With the data preprocessing of the REMD method, each IMF is separated to maintain different degrees of non-stationary. Using these separated IMFs can reduce the impact of strong data volatility on the predicting process. Meanwhile, it can be further shown from the figure that from IMF1 to IMF7, the complexity of each modal decreases, which is more conducive to the analysis and processing of wind speed data. Moreover, in Section 3.3.3, the improvement in prediction accuracy with REMD will be focused on.



**Figure 7.** REMD decomposition results of the four wind speed series. (**A**–**D**) are REMD decomposition results of four datasets in spring, summer, autumn and winter respectively.

## 3.3.2. Rationality of Adaptive VMD Method

500

С

0

The Spearman correlation coefficient is generally used to establish the correlation between each time series, analyzing the correlation level with the original wind speed series, and perform a secondary decomposition of the modal components according the correlation. At the same time, to reduce the calculation time and error accumulation and to improve the real-time correlation of wind speed characteristics, the components with high correlation are selected for the secondary decomposition to carry out reconstruction, which is convenient for feature mining.

Figure 8 shows the heat map of Spearman correlation coefficients between the wind speed series from dataset C and each IMF from the REMD. It can be recorded intuitively from the figure that the correlation index between IMF1 and the dataset C is 0.6438 with the positive correlation, which reflects that IMF1 has more detailed intrinsic wind characteristic information, indicating incomplete first decomposition. The correlation coefficient between the dataset C and IMF5 is -0.06734, showing a negative correlation, and the absolute value is close to 0, indicating that IMF5 is completely decomposed and few wind series characteristics are retained. According to the Spearman correlation analysis, the AVMD mode that has the highest correlation when the wind speed series is chosen to perform the secondary decomposition. The resulting parameters [K, $\alpha$ ] are shown in Figure 9, where K is 2, 4, 10 and 4 and  $\alpha$  is 200,684,336 and 452.

The characteristic of wind speed determines that it is not suitable to be decomposed into too many levels. The adaptive VMD method does not require a predetermined number of decompositions, based on the parameter optimization ability of the metaheuristic algorithm, to achieve the purpose of adaptive decomposition. When the number of decomposition levels increases, the characteristic information of the wind speed is more detailed, and the high- and low-frequency characteristics of wind speed are clearly identified, which can reflect the change in wind speed series from a long period to a short period. The IMFs left over from the first decomposition of REMD still retain some information of the original data. Through the secondary decomposition of the IMFs, the

1000

500

0 0

D

Spearman Correlation Coefficient 0.2654 0.3797 Autumn 0.6438 0.239 0.1681 0.1925 1 0.8 IMF1 0.6438 1 0.6 IMF2 0.239 1 IMF3 0.2654 1 0.4IMF4 0.3797 1 0.1599 0.2 IMF5 1 0 IMF6 0.1681 0.1599 1 -0.2 IMF7 0.1925 1 IMF4 IMF6 IMET Autumn IMEI IME2 IMF3 IMF5

local characteristics of the original data are highlighted, which is conducive to exploring the internal variation features of the original wind speed data.

**Figure 8.** Heat map of Spearman correlation coefficients between the original wind speed series from dataset C and each IMF from the REMD.



**Figure 9.** Combination of *K* and *α* parameters after AVMD. A, B, C and D are the combined parameters of AVMD decomposition of the four datasets in spring, summer, autumn and winter respectively.

#### 3.3.3. Prediction Results

In this section, the combined model WSTD-REMD-AVMD-DGRU proposed in this study is trained and verified on the four datasets. Figure 10 shows the visualization of the single-step prediction of the four wind speed datasets, including a prediction curve, an error distribution, and an error box plot, and each dataset is depicted with a different color. The prediction curve of the WSTD-REMD-AVMD-DGRU model on the test set fits well with the edge of the actual value in each dataset. The prediction error is distributed near the '0' baseline, illustrating that the error of the prediction system is controllable, especially on the datasets A and D. Furthermore, the effectiveness of the model is verified by various performance indices. The specific values are shown in the Tables 3–6. The predicted results from various other models are also given in the tables for comparison, including prediction in one step, two steps, and three steps ahead. The comparison models include single models such as GRU, LSTM, and RF; the decomposition optimization models such as WD-HGWO-SVR, WSTD-REMD-GRU, and VMD-HGWO-SVR; and the combined models of REMD-HGWO-GRU, WD-GWO-SVR and CEEMD-HGWO-GRU based on WSTD. The evaluation indicators MAE, RMSE, R2, TIC, and SSE are used to measure the prediction results and comprehensively test the effectiveness of the proposed model.



Figure 10. Single-step prediction results of the four wind speed series.

Table 3. The prediction performance from various models on dataset A.

Model		MAE			RMSE			$R^2$			TIC			SSE	
	1-Step	2-Step	3-Step	1-Step	2-Step	3-Step	1-Step	2-Step	3-Step	1-Step	2-Step	3-Step	1-Step	2-Step	3-Step
The proposed	0.0265	0.0531	0.0818	0.0338	0.0782	0.1273	0.9986	0.9957	0.9875	0.0080	0.0283	0.0296	0.0098	0.0184	0.0330
WSTD-VMD-HGWO-SVR	0.0873	0.0997	0.1105	0.0936	0.1203	0.1346	0.9942	0.9937	0.9902	0.0246	0.0287	0.0396	0.5648	1.6243	2.6591
WSTD-REMD-HGWO-GRU	0.1082	0.1394	0.1772	0.1257	0.1678	0.2177	0.9914	0.9821	0.9701	0.0296	0.0397	0.0515	2.3157	4.9175	7.8733
WSTD-CEEMD-HGWO-GRU	0.2381	0.3242	0.4317	0.3112	0.4662	0.5826	0.9459	0.8741	0.8050	0.0735	0.1106	0.1374	16.2923	39.2686	57.5463
WSTD-REMD-GWO-GRU	0.2029	0.2151	0.2266	0.2459	0.2569	0.2745	0.9740	0.9662	0.9595	0.0566	0.0597	0.0647	6.2862	9.9638	9.2274
WSTD-REMD-HGWO-SVR	0.1482	0.1554	0.4801	0.2364	0.2888	0.5893	0.9752	0.9461	0.7757	0.0542	0.0678	0.1417	7.3970	16.1000	59.1824
WSTD-WD-GWO-SVR	0.1631	0.2624	0.3381	0.2062	0.3430	0.4412	0.9758	0.9311	0.8829	0.0493	0.0809	0.1046	7.7958	20.7852	34.2191
WSTD-VMD-GRU	0.0289	0.0968	0.1045	0.0414	0.0916	0.1389	0.9970	0.9925	0.9866	0.0108	0.0219	0.0342	0.2666	0.9548	3.5586
WSTD-CEEMD-GRU	0.2256	0.2760	0.3255	0.3190	0.3605	0.3817	0.9366	0.9203	0.9132	0.0748	0.0850	0.0888	19.7238	25.5663	27.3503
WSTD-WD-GRU	0.2537	0.3061	0.4751	0.3380	0.4051	0.5911	0.9342	0.9058	0.8011	0.0826	0.0959	0.1314	22.5032	32.1648	68.1359
VMD-HGWO-SVR	0.1351	0.1786	0.1776	0.1634	0.2157	0.2160	0.9870	0.9774	0.9724	0.0383	0.0502	0.0509	3.3464	6.6253	7.0425
WD-HGWO-SVR	0.1634	0.2616	0.3354	0.2063	0.3396	0.4351	0.9867	0.9338	0.8917	0.0489	0.0799	0.1028	8.3818	22.6097	36.9103
WSTD-REMD-GRU	0.0510	0.1557	0.1725	0.0672	0.1883	0.2406	0.9970	0.9863	0.9654	0.0159	0.04433	0.0580	0.8734	4.0478	9.9962
WSTD-GWO-SVR	0.2002	0.2138	0.2258	0.2427	0.2560	0.2703	0.9732	0.9668	0.9577	0.0560	0.0595	0.0636	6.5164	7.8110	10.0248
GRU	0.5100	0.6724	0.7834	0.6597	0.8546	0.9884	0.7673	0.6078	0.4853	0.1541	0.1989	0.2262	55.3926	70.7938	69.7912
LSTM	0.5022	0.6801	0.7733	0.6515	0.8735	0.9701	0.7737	0.5942	0.4943	0.1526	0.2024	0.2241	53.0837	68.5102	79.3900
ARIMA	0.5056	0.6550	0.7655	0.6450	0.8205	0.9535	0.7722	0.6336	0.5084	0.1525	0.1930	0.2231	59.5121	74.0400	76.1862
BP	0.5153	0.6376	0.7578	0.6597	0.8394	0.9440	0.7632	0.6156	0.5191	0.1557	0.2015	0.2213	58.4474	82.8701	74.1387
LSSVM	0.5119	0.6616	0.7596	0.6499	0.8292	0.9538	0.7724	0.6317	0.5202	0.1537	0.1951	0.2233	53.9873	65.1392	61.5543
RF	0.5714	0.7492	0.8794	0.7190	0.9324	1.0929	0.7178	0.5325	0.3642	0.1700	0.2176	0.2562	79.6895	109.9353	119.3138

|--|

Model		MAE			RMSE			<i>R</i> <sup>2</sup>			TIC			SSE	
	1-Step	2-Step	3-Step	1-Step	2-Step	3-Step	1-Step	2-Step	3-Step	1-Step	2-Step	3-Step	1-Step	2-Step	3-Step
The proposed	0.0534	0.0748	0.0856	0.0758	0.1036	0.1246	0.9987	0.9977	0.9969	0.0115	0.0157	0.0173	0.7567	2.1019	2.5629
WSTD-VMD-HGWO-SVR	0.1390	0.1518	0.1626	0.1788	0.1861	0.2010	0.9901	0.9891	0,9843	0.0270	0.0279	0.0301	6.0389	4.9523	5.7395
WSTD-REMD-HGWO-GRU	0.1952	0.2454	0.3140	0.2387	0.3039	0.4096	0.9858	0.9730	0.9479	0.0360	0.0456	0.0611	7.8560	15.2756	30.4725
WSTD-CEEMD-HGWO-GRU	0.2891	0.3531	0.4867	0.3697	0.4552	0.6076	0.9613	0.9443	0.8993	0.0555	0.0688	0.0913	24.8642	32.8804	54.8275
WSTD-REMD-GWO-GRU	0.1972	0.2117	0.2952	0.2332	0.3947	0.4950	0.9895	0.9538	0.9232	0.0350	0.0603	0.0745	5.5570	27.4678	47.3507
WSTD-REMD-HGWO-SVR	0.2649	0.3818	0.4564	0.3185	0.4751	0.5583	0.9715	0.9354	0.9108	0.0474	0.0707	0.0831	18.7953	39.9718	53.7150
WSTD-WD-GWO-SVR	0.4010	0.4073	0.4079	0.4735	0.4783	0.4850	0.9639	0.9620	0.9618	0.0723	0.0735	0.0746	14.9314	15.3618	15.3963
WSTD-VMD-GRU	0.1387	0.2000	0.2849	0.1810	0.2160	0.5525	0.9934	0.9849	0.9078	0.0278	0.0325	0.0826	3.0832	8.8005	59.2937
WSTD-CEEMD-GRU	0.3969	0.5127	0.6638	0.6430	0.9566	1.0510	0.9234	0.8393	0.8288	0.1013	0.1513	0.1728	59.4176	156.9631	170.2382
WSTD-WD-GRU	0.2893	0.3905	0.5251	0.3994	0.5134	0.6400	0.9544	0.9250	0.8841	0.0597	0.0765	0.0990	31.4280	51.6582	79.8677
VMD-HGWO-SVR	0.2329	0.3986	0.4242	0.3178	0.4100	0.5365	0.9720	0.9689	0.9312	0.0472	0.0514	0.0763	6.9342	23.7923	36.5286
WD-HGWO-SVR	0.2691	0.3872	0.4514	0.3249	0.4809	0.5544	0.9701	0.9342	0.9130	0.0484	0.0715	0.0824	20.8039	45.3356	59.9384
WSTD-REMD-GRU	0.1197	0.2083	0.3410	0.1600	0.3150	0.4440	0.9945	0.9760	0.8813	0.0243	0.0464	0.1052	3.4960	16.6538	34.7916
WSTD-GWO-SVR	0.3983	0.4048	0.4108	0.4751	0.4802	0.4858	0.9654	0.9633	0.9636	0.0726	0.0733	0.0736	14.1472	14.7856	14.4973
GRU	0.5302	0.7081	0.8722	0.6819	0.9143	1.1275	0.8722	0.7722	0.6566	0.1032	0.1390	0.1731	75.5531	114.9054	160.5826
LSTM	0.5649	0.8180	0.9325	0.7485	1.0120	1.1752	0.8453	0.7275	0.6305	0.1127	0.1560	0.1766	99.0382	127.8319	213.4905
ARIMA	0.5125	0.7678	0.9572	0.7233	0.9730	1.1723	0.8542	0.7396	0.6236	0.1087	0.1462	0.1776	88.6792	151.3385	184.1856
BP	0.6498	0.7737	0.9624	0.8109	0.9660	1.1717	0.8435	0.7497	0.6239	0.1251	0.1478	0.1780	69.9745	117.8563	175.6531
LSSVM	0.5950	0.7953	0.9820	0.7468	0.9849	1.1870	0.8480	0.7362	0.6179	0.1130	0.1498	0.1817	82.4907	123.8675	152.3672
RF	0.6588	0.8508	1.0154	0.8397	1.0670	1.2544	0.8069	0.6898	0.5753	0.1277	0.1628	0.1929	106.3476	161.5313	188.5835

Table 5. The prediction performance from various models on dataset C.

Model		MAE			RMSE			$R^2$			TIC			SSE	
	1-Step	2-Step	3-Step	1-Step	2-Step	3-Step									
The proposed	0.0440	0.0459	0.0462	0.0582	0.0615	0.0626	0.9983	0.9967	0.9921	0.0138	0.0146	0.0149	0.5298	0.7387	0.7675
WSTD-VMD-HGWO-SVR	0.0865	0.0901	0.1166	0.1047	0.1138	0.1458	0.9895	0.9887	0.9821	0.0250	0.0270	0.0343	1.9670	2.1397	3.2653
WSTD-REMD-HGWO-GRU	0.0991	0.1640	0.2186	0.1243	0.2013	0.2724	0.9879	0.9647	0.9313	0.0294	0.0475	0.0642	2.5469	7.4293	13.6527
WSTD-CEEMD-HGWO-GRU	0.1971	0.4087	0.4118	0.2605	0.5005	0.5159	0.9372	0.8014	0.7835	0.0627	0.1179	0.1230	14.6358	41.9232	48.0972
WSTD-REMD-GWO-GRU	0.1450	0.1610	0.1615	0.1745	0.1931	0.1971	0.9806	0.9751	0.9709	0.0406	0.0450	0.0462	3.7592	4.7490	5.4628
WSTD-REMD-HGWO-SVR	0.1307	0.2025	0.2095	0.1808	0.2626	0.2607	0.9777	0.9482	0.9303	0.0424	0.0613	0.0618	4.8231	11.6438	12.9346
WSTD-WD-GWO-SVR	0.1611	0.2712	0.3614	0.2100	0.3456	0.4450	0.9667	0.9041	0.8389	0.0481	0.0816	0.1043	7.5073	19.4635	32.5792
WSTD-VMD-GRU	0.0226	0.0678	0.0952	0.0370	0.0628	0.1785	0.9988	0.9967	0.9691	0.0089	0.0149	0.0424	0.2404	0.6309	6.1653
WSTD-CEEMD-GRU	0.0776	0.1957	0.2532	0.1272	0.3812	0.4456	0.9848	0.8644	0.8288	0.0304	0.0916	0.1064	3.1937	31.4186	38.9903
WSTD-WD-GRU	0.2212	0.2427	0.5399	0.3211	0.3155	0.7330	0.9148	0.9179	0.5567	0.0785	0.0773	0.1738	20.3076	19.5159	104.7591
VMD-HGWO-SVR	0.1149	0.1456	0.1680	0.1420	0.1765	0.2053	0.9894	0.9781	0.9640	0.0348	0.0431	0.0496	2.1390	4.2846	6.6573
WD-HGWO-SVR	0.1663	0.2757	0.3611	0.2111	0.3563	0.4448	0.9632	0.9271	0.8367	0.0504	0.0839	0.1042	8.7813	24.8823	38.5802
WSTD-REMD-GRU	0.0870	0.1813	0.2141	0.1224	0.2266	0.3163	0.9883	0.9653	0.9087	0.0289	0.0526	0.0761	2.4138	7.1025	19.0146
WSTD-GWO-SVR	0.1612	0.1599	0.1504	0.1968	0.1921	0.1819	0.9713	0.9758	0.9801	0.0461	0.0447	0.0422	5.3970	4.5678	3.8964
GRU	0.4571	0.5368	0.6011	0.5844	0.6870	0.7726	0.7350	0.6345	0.5386	0.1405	0.1637	0.1859	49.5267	55.2589	63.1174
LSTM	0.4524	0.5246	0.5916	0.5788	0.6620	0.7439	0.7402	0.6615	0.5716	0.1384	0.1577	0.1760	46.5672	61.7483	60.8625
ARIMA	0.4785	0.5895	0.6983	0.6028	0.7429	0.8610	0.7180	0.5722	0.4257	0.1445	0.1782	0.2060	52.9226	60.7234	59.1549
BP	0.4778	0.5902	0.6930	0.6106	0.7447	0.8580	0.7142	0.5699	0.4321	0.1480	0.1782	0.2067	56.9368	63.3287	55.3472
LSSVM	0.4818	0.5973	0.7148	0.6030	0.7487	0.8709	0.7178	0.5641	0.4122	0.1443	0.1789	0.2072	50.0967	61.7453	63.9388
RF	0.5319	0.6997	0.8122	0.6806	0.8910	1.0285	0.6471	0.4123	0.2503	0.1646	0.2163	0.2496	68.8850	94.6278	107.4784

It is clearly shown in Tables 3–6 that the performance of single models are all ranking behind. The decomposition optimization model is improved on the basis of single models, and it is recognized that proper decomposition can help improve the prediction performance. The R2 and TIC indicators in the table explain the correlation between the predicted data and the actual data. The larger the R2 value and the smaller the TIC value, the higher correlation, which can effectively evaluate the model. For the dataset A as shown in Table 3, the predictive performance of WSTD-REMD-GRU is much better than that of WSTD-CEEMD-GRU. The average values of R2 are 0.9829 and 0.9234, and the average values of TIC are 4.9725 and 24.2135, respectively. The prediction results suggest that the data preprocessing with REMD can effectively improve the prediction accuracy.

Table 6. The prediction performance from various models on d	lataset D.
--	------------

Model		MAE			RMSE			<i>R</i> <sup>2</sup>			TIC			SSE	
	1-Step	2-Step	3-Step	1-Step	2-Step	3-Step	1-Step	2-Step	3-Step	1-Step	2-Step	3-Step	1-Step	2-Step	3-Step
The proposed	0.0309	0.0374	0.0444	0.0384	0.0467	0.0562	0.9973	0.9945	0.9873	0.0101	0.0124	0.0150	0.2902	0.4274	0.6168
WSTD-VMD-HGWO-SVR	0.0433	0.0663	0.0891	0.0522	0.0824	0.1101	0.9916	0.9808	0.9672	0.0139	0.0219	0.0291	0.4850	1.0614	1.7848
WSTD-REMD-HGWO-GRU	0.0436	0.0716	0.0905	0.0540	0.0904	0.1213	0.9886	0.9689	0.9453	0.0145	0.0243	0.0326	0.5424	1.3934	2.1183
WSTD-CEEMD-HGWO-GRU	0.2609	0.2667	0.3327	0.3548	0.3585	0.4742	0.7073	0.7018	0.5317	0.0957	0.0941	0.1248	18.6286	20.3429	35.7302
WSTD-REMD-GWO-GRU	0.0659	0.0860	0.1120	0.0831	0.1057	0.1368	0.9755	0.9566	0.9372	0.0223	0.0284	0.0371	1.0129	1.9540	2.7816
WSTD-REMD-HGWO-SVR	0.1647	0.1284	0.1945	0.2472	0.3176	0.3504	0.8809	0.6958	0.6593	0.0633	0.0845	0.0910	6.7752	19.6783	21.3897
WSTD-WD-GWO-SVR	0.1442	0.2172	0.2682	0.1870	0.2965	0.3525	0.9191	0.7941	0.7111	0.0502	0.0793	0.0946	5.8147	13.1518	18.5134
WSTD-VMD-GRU	0.0853	0.0952	0.2703	0.2217	0.2425	0.5223	0.8105	0.7838	0.4470	0.0595	0.0651	0.1397	9.5462	11.5238	53.1585
WSTD-CEEMD-GRU	0.1467	0.2385	0.2718	0.4163	0.5655	0.6323	0.5813	0.3350	0.2327	0.1103	0.1518	0.1715	34.6628	90.1783	137.5041
WSTD-WD-GRU	0.2199	0.2785	0.3247	0.3290	0.3685	0.4063	0.7460	0.6801	0.6134	0.0877	0.0989	0.1093	26.7563	21.2180	32.1939
VMD-HGWO-SVR	0.0610	0.0888	0.1199	0.0760	0.1131	0.1543	0.9876	0.9542	0.9206	0.0204	0.0385	0.0407	0.5707	1.8336	2.8740
WD-HGWO-SVR	0.1441	0.2173	0.2617	0.1860	0.2966	0.3405	0.9186	0.7936	0.7189	0.0498	0.0792	0.0928	6.8137	17.2469	23.4062
WSTD-REMD-GRU	0.0491	0.1258	0.1437	0.1049	0.2151	0.2820	0.9568	0.8595	0.7316	0.0281	0.0561	0.0748	2.1447	7.4762	15.1173
WSTD-GWO-SVR	0.0670	0.0843	0.1101	0.0840	0.1045	0.1358	0.9755	0.9573	0.9380	0.0225	0.0281	0.0368	1.0138	1.9295	2.7120
GRU	0.4448	0.5470	0.6011	0.5883	0.6957	0.7395	0.6546	0.5567	0.4347	0.1558	0.1824	0.1908	35.8524	37.3562	40.5836
LSTM	0.4490	0.5502	0.5809	0.5929	0.6935	0.7261	0.6436	0.5571	0.4912	0.1570	0.1800	0.1904	36.9078	36.6845	35.8163
ARIMA	0.4511	0.5359	0.5850	0.6012	0.6955	0.7390	0.5479	0.4617	0.3526	0.1596	0.1834	0.1921	43.9867	42.9321	37.8196
BP	0.4532	0.5419	0.5760	0.6007	0.6946	0.7172	0.5454	0.4739	0.3015	0.1600	0.1812	0.1914	42.9672	43.8373	36.1058
LSSVM	0.4423	0.5255	0.5745	0.5861	0.6798	0.7194	0.4601	0.3772	0.2928	0.1555	0.1793	0.1886	37.8521	39.0385	32.8654
RF	0.5283	0.5863	0.6179	0.6826	0.7481	0.7985	0.5758	0.4455	0.3592	0.1806	0.1968	0.2082	63.2678	65.9756	66.6372

For the dataset B shown in Table 4, it can be seen from Figure 10 that the wind speed trend is more volatile, and the comparison from Tables 3–6 also reflects that the prediction accuracy on dataset B is generally lower than the other three datasets. Compared with the classic decomposition models CEEMD, VMD, and WD, REMD can better reduce the nonstationarity of the wind speed. When these decomposition algorithms are combined with GRU and SVR, the prediction results of the GRU method are more accurate than SVR, especially in RMSE. Table 4 gives the evaluation indicators of different models. Taking the REMD-HGWO-SVR model as an example, the MAE value is improved by 35.71% in one-step prediction, and the accuracy of the two-steps- and three-steps-ahead predictions is improved by 55.53% and 45.35%, respectively. The large improvements indicate the superiority of the REMD, HGWO, and GRU integrated model. Therefore, this approach is considered as part of the proposed model.

For the dataset C shown in Table 5, the six single models of GRU, LSTM, ARIMA, BP, LSSVM, and RF are predicted one step ahead, and the accuracy of the MAE is maintained at about 50%. However, as the number of prediction steps increases, the performance of the model decreases, making it difficult to capture wind speed information. The RF model shows the worst performance. By comparing with single models, it reflects the advantages of the proposed model. Among all models, the baseline model prediction performance of REMD and VMD is better than the rest of the models. In terms of MAE and RMSE, the multi-step prediction accuracy of the WSTD-REMD-AVMD-DGRU model is improved by more than 50% compared with other models, and the value of SSE is basically close to 0, indicating that the multi-step prediction error of the combined model is small. In addition, the average value of each index predicted by the model in multiple steps is also optimal.

The prediction results on the dataset C is shown in Figure 11. The best performance from the WSTD-REMD-AVMD-DGRU model can be observed. The performance shown by histogram and spider graph is more straightforward. When the wind speed changes rapidly, the model has excellent adaptability and good extrapolation performance of multistep prediction. The prediction performance of the decomposition optimization model is not as good as that of the combination model, but the decomposition optimization model fits the wind speed trend well in one-step-ahead prediction, and there are multiple spikes. However, after the wind speed series is processed by WSTD and VMD, the wind speed prediction accuracy is significantly improved, and the error between the actual value and the predicted value is also reduced. For the dataset D shown in Table 6, the preprocessing of WSTD improves the quality of the dataset. In VMD-HGWO-SVR and WSTD-VMD-HGWO-SVR, the noise filtering of WSTD has a slight advantage for the construction of prediction models. The combination of HGWO and GRU is compared with the combination of single GRU, GWO and GRU. The MAE, RMSE, R2, TIC, and SSE values are improved in different degrees when the present method is used to predict multiple steps ahead. Meanwhile, the improved GWO algorithm also avoids the problem of parameters falling into local optimality. In addition, based on the four experiments, the nonlinear processing capability of the VMD algorithm has always performed well in the field of wind speed prediction with high accuracy and practicability, and its robustness on different datasets is also good.

Generally speaking, the performance of the combined models is better than that of the single models. The WSTD-REMD-AVMD-DGRU model proposed in this paper has excellent robustness and adaptability. The model fits well in predicting the trend of wind speed, and all indicators rank the first, of which SSE is significantly smaller than the other models. The model has been decomposed twice, and the superiority of the model is generally better than that of one decomposition and no decomposition, which is proven by experiments on four datasets. Finally, the wind speed data decomposed by REMD and AVMD reduce the difficulty of constructing the HGWO-GRU prediction model and further improve the prediction accuracy. In summary, the present model has superior prediction ability and adaptive ability.



Figure 11. The prediction results on the dataset C.

#### 3.3.4. Error Analysis

The establishment of wind speed prediction model should take the stability and universality of the model into consideration. Adopting the integration strategy to form a suitable prediction model will improve the performance of the model as much as possible. Due to the complex characteristics of wind speed, errors are inevitable. Therefore, it is necessary to avoid the generation of errors as much as possible, and effectively use the errors to fully understand the potential characteristic information of the wind speed series and feed it back to the prediction model to improve the accuracy of wind speed prediction. As can be seen from Figure 12, when the combined model developed in this study performs one-step-, twos-step-, and threes-step-ahead prediction on the wind speed datasets in spring, summer, autumn, and winter, the prediction error probability plot conforms to the rule of Gaussian distribution on the whole. However, there is an outlier phenomenon in the tail of the prediction error distribution in summer B# and autumn C#, indicating that the fitting effect of the tail of the two datasets is not good, but the overall error of the model is controllable. The error distribution of one-step-ahead prediction is basically concentrated in [-0.1 m/s, 0.1 m/s]. The error distribution of two-steps-ahead prediction is concentrated in [-0.2 m/s, 0.2 m/s], and the error distribution of three-steps-ahead prediction is basically concentrated in [-0.3 m/s, 0.3 m/s]. In particular, the WSTD-REMD-AVMD-DGRU model has a good prediction effect on the performance of the dataset D. The error distribution of the multi-step ahead prediction is concentrated in [-0.15 m/s, 0.15 m/s], and the error distribution is more concentrated. The error analysis shows that the proposed model has good generalization and accuracy, and the wind speed prediction error is controllable. The prediction effect is the best among the compared models.



**Figure 12.** Error probability plot of multi-step prediction for the WSTD-REMD-AVMD-DGRU model. (**A**–**D**) are the prediction error probability plot of four data sets in spring, summer, autumn and winter respectively.

## 4. Conclusions

Accurate wind speed prediction can effectively improve wind energy utilization. This study proposes an adaptive two-stage decomposition integrated system for short-term wind speed prediction. The system is based on a data preprocessing strategy, a cascade optimization strategy, and a deep learning prediction strategy.

First, the wind speed is decomposed using WSTD and REMD into a series of components that change smoothly and have obvious changing regularity, thus greatly reducing the interference and coupling between different features and improving the quality of subsequent data. The VMD is employed for the secondary decomposition, and it is integrated with the Spearman criterion to revise the accumulation of errors of the model during the first decomposition, reduce the model complexity, and obtain the long-term, fluctuating, trends of the wind power signal. The potential characteristics of wind speed series are acquired, and their dynamic characteristics are captured by in-depth analysis.

Then, HGWO is adopted to optimize VMD and GRU, which effectively avoids the limitation of empirically set parameters, and make up for the defect that the parameters fall into the local optimum. Using the most advanced deep learning model, an adaptive parameter selection process is advanced.

Finally, the improved GRU method strengthens the characteristic information and inline relationships of wind speed data and can comprehensively mine the characteristics of wind power, perform feature mining, and ensure the prediction accuracy and stability of the model. Four datasets in different seasons are selected for multi-step ahead prediction of the wind speed. The results show that the proposed WSTD-REMD-AVMD-DGRU model developed in this study has robust and accurate prediction performance, which can provide the best forecast results for wind series.

The present model realizes a whole adaptive process of wind speed prediction and overcomes the problems of experience adjustment, incomplete wind speed information mining, and inaccurate single-model prediction that inhibited traditional models. Moreover, the present model is robust and generalizable, which can be easily extended to time series prediction in meteorology, mechanical engineering, finance, biology, etc.

**Author Contributions:** Conceptualization, K.Y., B.W. and Y.L.; methodology, K.Y.; software, K.Y. and Y.W.; validation, X.Q.; formal analysis, K.Y. and Y.L.; investigation, B.W.; resources, X.Q.; data curation, J.L.; writing—original draft preparation, K.Y.; writing—review and editing, B.W.; visualization, K.Y.; supervision, J.L.; project administration, Y.L.; funding acquisition, Y.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (Grant Nos. 91952102, 12032016, 11972220). It was also funded by Program supported by the Shanghai Education Development Foundation and the Shanghai Municipal Education Commission in China via Project (Grant No. 18SG53), as well as the Guangdong Provincial Key Laboratory under Grant No. 2019B121203001.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

#### References

- 1. Wang, Y.; Zou, R.; Liu, F.; Zhang, L.; Liu, Q. A review of wind speed and wind power forecasting with deep neural networks. *Appl. Energy* **2021**, *304*, 117766. [CrossRef]
- Liu, L.; Wang, Z.; Wang, Y.; Wang, J.; Chang, R.; He, G.; Tang, W. Optimizing wind/solar combinations at finer scales to mitigate renewable energy variability in China. *Renew. Sustain. Energy Rev.* 2020, 156, 321–330. [CrossRef]
- Zuluaga, D.; Alvarez, A.; Giraldo, T. Short-term wind speed prediction based on robust Kalman filtering: An experimental comparison. *Appl. Energy* 2015, 156, 321–330. [CrossRef]
- 4. Deng, Y.; Wang, B.; Lu, Z. A hybrid model based on data preprocessing strategy and error correction system for wind speed forecasting. *Energy Convers. Manag.* **2020**, *212*, 112779. [CrossRef]
- Brabec, M.; Craciun, A.; Dumitrescu, A. Hybrid numerical models for wind speed forecasting. J. Atmos. Sol. Terr. Phys. 2021, 220, 105669. [CrossRef]
- 6. Bouzgou, H.; Benoudjit, N. Multiple architecture system for wind speed prediction. Appl. Energy 2011, 88, 2463–2471. [CrossRef]
- 7. Wang, H.; Han, S.; Liu, Y.; Yan, J.; Li, L. Sequence transfer correction algorithm for numerical weather prediction wind speed and its application in a wind power forecasting system. *Appl. Energy* **2019**, 237, 1–10. [CrossRef]
- Ye, L.; Zhao, Y.; Zeng, C.; Zhang, C. Short-term wind power prediction based on spatial model. *Renew. Energy* 2017, 101, 1067–1074. [CrossRef]
- 9. Geibel, M.; Bangga, G. Data Reduction and Reconstruction of Wind Turbine Wake Employing Data Driven Approaches. *Energies* **2022**, *15*, 3773. [CrossRef]

- Liang, T.; Zhao, Q.; Lv, Q.; Sun, H. A novel wind speed prediction strategy based on Bi-LSTM, MOOFADA and transfer learning for centralized control centers. *Energy* 2021, 230, 120904. [CrossRef]
- Lydia, M.; Kumar, S.; Selvakumar, A.; Kumar, G. Linear and non-linear autoregressive models for short-term wind speed forecasting. *Energy Convers. Manag.* 2016, 112, 115–124. [CrossRef]
- 12. Erdem, E.; Shi, J. ARMA based approaches for forecasting the tuple of wind speed and direction. *Appl. Energy* **2011**, *88*, 1405–1414. [CrossRef]
- 13. Valipour, M.; Banihabib, M.; Behbahani, S. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. *J. Hydrol.* **2013**, *476*, 433–441. [CrossRef]
- Corba, B.; Egrioglu, E.; Dalar, A. AR–ARCH type artificial neural network for forecasting. *Neural Process. Lett.* 2020, 51, 819–836. [CrossRef]
- Radziukynas, V.; Klementavicius, A. Short-term wind speed forecasting with ARIMA model. In Proceedings of the 2014 55th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON), Riga, Latvia, 14 October 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 145–149.
- 16. Tian, Z.; Wang, G.; Ren, Y. Short-term wind speed forecasting based on autoregressive moving average with echo state network compensation. *Wind. Eng.* **2020**, *44*, 152–167. [CrossRef]
- Nair, A.; Krishnaveny, R.; Vanitha, B.; Jisma, C. Forecasting of wind speed using ANN, ARIMA and hybrid models. In Proceedings of the 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Kannur, India, 6–7 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 170–175.
- 18. Shukur, O.; Lee, M. Daily wind speed forecasting through hybrid KF-ANN model based on ARIMA. *Renew. Energy* 2015, 76, 637–647. [CrossRef]
- 19. Chen, Y.; Dong, Z.; Wang, Y.; Su, J.; Han, Z.; Zhou, D. Short-term wind speed predicting framework based on EEMD-GA-LSTM method under large scaled wind history. *Energy Convers. Manag.* **2021**, 227, 113559. [CrossRef]
- 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]
- 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]
- 22. Dai, F.; Nie, G.; Chen, Y. The municipal solid waste generation distribution prediction system based on FIG–GA-SVR model. *J. Mater. Cycles Waste Manag.* 2020, *22*, 1352–1369. [CrossRef]
- Li, C.; Xiao, Z.; Xia, X.; Zou, W.; Zhang, C. A hybrid model based on synchronous optimisation for multi-step short-term wind speed forecasting. *Appl. Energy* 2018, 215, 131–144. [CrossRef]
- 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. Santhosh, M.; Venkaiah, C.; Kumar, D. Ensemble empirical mode decomposition based adaptive wavelet neural network method for wind speed prediction. *Energy Convers. Manag.* **2018**, *168*, 482–493. [CrossRef]
- 26. Mora, E.; Cifuentes, J.; Marulanda, G. Short-Term Forecasting of Wind Energy: A Comparison of Deep Learning Frameworks. *Energies* **2021**, *14*, 7943. [CrossRef]
- Liu, H.; Chen, C. Data processing strategies in wind energy forecasting models and applications: A comprehensive review. *Appl. Energy* 2019, 249, 392–408. [CrossRef]
- 28. Zhou, Q.; Wang, C.; Zhang, G. Hybrid forecasting system based on an optimal model selection strategy for different wind speed forecasting problems. *Appl. Energy* **2019**, 250, 1559–1580. [CrossRef]
- Qian, Z.; Pei, Y.; Zareipour, H.; Chen, N. A review and discussion of decomposition-based hybrid models for wind energy forecasting applications. *Appl. Energy* 2019, 235, 939–953. [CrossRef]
- Mi, X.; Zhao, S. Wind speed prediction based on singular spectrum analysis and neural network structural learning. *Energy* Convers. Manag. 2020, 216, 112956. [CrossRef]
- 31. Büyükşahin, Ü.Ç.; Ertekin, Ş. Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition. *Neurocomputing* **2019**, *361*, 151–163. [CrossRef]
- 32. Tian, Z.; Chen, H. Multi-step short-term wind speed prediction based on integrated multi-model fusion. *Appl. Energy* **2021**, 298, 117248. [CrossRef]
- Zhang, Y.; Han, J.; Pan, G.; Xu, Y.; Wang, F. A multi-stage predicting methodology based on data decomposition and error correction for ultra-short-term wind energy prediction. *J. Clean. Prod.* 2021, 292, 125981. [CrossRef]
- 34. Duan, J.; Wang, P.; Ma, W.; Fang, S.; Hou, Z. A novel hybrid model based on nonlinear weighted combination for short-term wind power forecasting. *Int. J. Electr. Power Energy Syst.* **2022**, *134*, 107452. [CrossRef]
- 35. Liao, X.; Liu, Z.; Deng, W. Short-term wind speed multistep combined forecasting model based on two-stage decomposition and LSTM. *Wind. Energy* **2021**, *24*, 991–1012. [CrossRef]
- Liu, B.; Zhao, S.; Yu, X.; Zhang, L.; Wang, Q. A Novel Deep Learning Approach for Wind Power Forecasting Based on WD-LSTM Model. *Energies* 2020, 13, 4964. [CrossRef]
- Lu, P.; Ye, L.; Zhao, Y.; Dai, B.; Pei, M.; Tang, Y. Review of meta-heuristic algorithms for wind power prediction: Methodologies, applications and challenges. *Appl. Energy* 2021, 301, 117446. [CrossRef]

- Bai, Y.; Liu, M.; Ding, L.; Ma, Y. Double-layer staged training echo-state networks for wind speed prediction using variational mode decomposition. *Appl. Energy* 2021, 301, 117461. [CrossRef]
- 39. Wu, C.; Wang, J.; Chen, X.; Du, P.; Yang, W. A novel hybrid system based on multi-objective optimization for wind speed forecasting. *Renew. Energy* **2020**, *146*, 149–165. [CrossRef]
- Neshat, M.; Nezhad, M.; Abbasnejad, E.; Mirjalili, S.; Tjernberg, L.; Garcia, D.; Alexander, B.; Wagner, M. A deep learning-based evolutionary model for short-term wind speed forecasting: A case study of the Lillgrund offshore wind farm. *Energy Convers. Manag.* 2021, 236, 114002. [CrossRef]
- Tian, Z.; Chen, H. A novel decomposition-ensemble prediction model for ultra-short-term wind speed. *Energy Convers. Manag.* 2021, 248, 114775. [CrossRef]
- 42. Huang, N.; Shen, Z.; Long, S.; Wu, M.; Shih, H.; Zheng, Q.; Yen, N.; Tung, C.; Liu, H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. Ser. Math. Phys. Eng. Sci.* **1998**, 454, 903–995. [CrossRef]
- 43. Tian, Z.; Li, S.; Wang, Y. A prediction approach using ensemble empirical mode decomposition-permutation entropy and regularized extreme learning machine for short-term wind speed. *Wind. Energy* **2020**, *23*, 177–206. [CrossRef]
- Wu, Z.; Xia, X.; Xiao, L.; Liu, Y. Combined model with secondary decomposition-model selection and sample selection for multi-step wind power forecasting. *Appl. Energy* 2020, 261, 114345. [CrossRef]
- Peng, Z.; Peng, S.; Fu, L.; Lu, B.; Tang, J.; Wang, K.; Li, W. A novel deep learning ensemble model with data denoising for short-term wind speed forecasting. *Energy Convers. Manag.* 2020, 207, 112524. [CrossRef]
- 46. Liu, Z.; Peng, D.; Zuo, M.; Xia, J.; Qin, Y. Improved Hilbert–Huang transform with soft sifting stopping criterion and its application to fault diagnosis of wheelset bearings. *ISA Trans.* **2021**, *21*, 0019–0578. [CrossRef] [PubMed]
- 47. Zhu, A.; Xu, C.; Li, Z.; Wu, J.; Liu, Z. Hybridizing grey wolf optimization with differential evolution for global optimization and test scheduling for 3D stacked SoC. *J. Syst. Eng. Electron.* **2015**, *26*, 317–328. [CrossRef]
- 48. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. IEEE Trans. Signal Process. 2013, 62, 531–544. [CrossRef]
- 49. Shivam, K.; Tzou, J.; Wu, S. Multi-Step Short-Term Wind Speed Prediction Using a Residual Dilated Causal Convolutional Network with Nonlinear Attention. *Energies* **2020**, *13*, 1772. [CrossRef]