

Article

A Novel and Robust Wind Speed Prediction Method Based on Spatial Features of Wind Farm Cluster

Mumin Zhang ¹, Yuzhi Wang ¹, Haochen Zhang ², Zhiyun Peng ³ and Junjie Tang ^{4,*} ¹ University of Cincinnati Joint Co-op Institute, Chongqing University, Chongqing 400044, China² Department of Electrical and Computer Engineering, University of California, Los Angeles, CA 90095, USA³ School of Computing Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada⁴ State Key Laboratory of Power Transmission Equipment & System Security and New Technology, Chongqing University, Chongqing 400044, China

* Correspondence: tangjunjie@cqu.edu.cn

Abstract: Wind energy has been widely used in recent decades to achieve green and sustainable development. However, wind speed prediction in wind farm clusters remains one of the less studied areas. Spatial features of cluster data of wind speed are not fully exploited in existing work. In addition, missing data, which dramatically deteriorate the forecasting performance, have not been addressed thoroughly. To tackle these tough issues, a new method, termed input set based on wind farm cluster data–deep extreme learning machine (IWC-DELM), is developed herein. This model builds an input set based on IWC, which takes advantage of the historical data of relevant wind farms to utilize the spatial characteristics of wind speed sequences within such wind farm clusters. Finally, wind speed prediction is obtained after the training of DELM, which results in a better performance in forecasting accuracy and training speed. The structure IWC, complete with the multidimensional average method (MDAM), is also beneficial to make up the missing data, thus enhancing data robustness in comparison to the traditional method of the moving average approach (MAA). Experiments are conducted with some real-world data, and the results of gate recurrent unit (GRU), long- and short-term memory (LSTM) and sliced recurrent neural networks (SRNNs) are also taken for comparison. These comparative tests clearly verify the superiority of IWC-DELM, whose accuracy and efficiency both rank at the top among the four candidates.

Keywords: wind speed forecasting; wind farm cluster; input set based on wind farm cluster data; robustness analysis; deep extreme learning machine; multidimensional average method

MSC: 68T07

Citation: Zhang, M.; Wang, Y.; Zhang, H.; Peng, Z.; Tang, J. A Novel and Robust Wind Speed Prediction Method Based on Spatial Features of Wind Farm Cluster. *Mathematics* **2023**, *11*, 499. <https://doi.org/10.3390/math11030499>

Academic Editors: Zhuojia Fu, Yiqian He and Hui Zheng

Received: 18 December 2022

Revised: 8 January 2023

Accepted: 12 January 2023

Published: 17 January 2023



Copyright: © 2023 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/>).

1. Introduction

Clean energy is largely needed to achieve peak carbon emission and carbon neutrality [1]. Wind power, as a renewable and widely distributed energy source, has received increasing attention in the past two decades [2]. Large-scale wind energy integration brings challenges for grid security due to the intermittent and random nature of wind speed [2–4]. Therefore, accurate forecasting of wind speed among wind farms in the cluster has gradually taken on a key role in operating strategies, capacity planning and power balance [5].

Existing work on wind speed and wind power forecasting can be classified as single-wind-farm prediction and wind farm cluster prediction according to the scale of research objects. Single-wind-farm forecasting methods mainly include four categories: physical models, traditional statistical models, artificial-intelligence-based models and hybrid models [6]. Deep learning methods have emerged as a powerful tool in wind speed and wind power prediction due to their ability to realize nonlinear fitting [7,8]. Gate recurrent unit (GRU) and sliced recurrent neural networks (SRNNs) are used in wind speed forecasting

in [9,10]. References [11,12] adopt deep extreme learning machine (DELM) to predict wind power. However, the prediction of a single farm only focuses on its own data analysis without considering surrounding environmental factors including humidity, temperature, latitude and orography, leading to insufficient prediction accuracy [13,14]. These environmental factors can be reflected by the historical data of adjacent wind farms [15]. Additionally, the data capacity of a wind farm cluster is several times that of a single wind farm, which indicates that single-wind-farm forecasting methods may be not suitable for wind farm cluster forecasting.

Wind farm cluster prediction imposes a significant influence on the generation schedule and reserve capacity of the power system compared with a single wind farm [16,17]. Taking spatial-temporal correlation into consideration, a wind farm cluster fully utilizes information of the surrounding environment. Existing works related to multi-wind-farm prediction mainly focus on wind power prediction (WPP). In [18], three coefficients representing the characteristics of a wind farm in a wind cluster are weighted by the Shapley value method. The characteristics of the wind cluster are extracted by a convolutional neural network (CNN), and then such characteristics are fed into a long- and short-term memory (LSTM) neural network to establish the relationship between key characteristics and power generation. Peng et al. [19] proposed a regional WPP method called multifeature similarity matching (MFSM) on the basis of the single feature similarity matching (SFSM) method. The four key parameters in MFSM are proposed while the impact of each parameter on forecasting error and the method applicability in varying regional scales are analyzed.

There are few studies concerning wind speed forecasting (WSF) of wind farm clusters. However, WSF has a wider range of applications, including meteorological uses and energy uses. Additionally, unlike wind power data, historical wind speed sequences can directly represent the relevance between different wind farms. In [20], a prediction method based on collaborative filtering against a virtual edge expansion graph structure is proposed in order to tackle the problem of underutilization of wind speed sequences. This method ensures that the spatial correlation can be fully learned by extending the scale of the dataset. It connects the wind turbines in different wind farms through virtual edges and takes LSTM as the main body for wind speed prediction. In [15], the CNN and LSTM are combined to build a deep architecture termed predictive spatiotemporal network (PSTN). CNNs at the bottom of the prediction model are used to extract spatial features from the spatial wind speed matrices, and LSTM captures the temporal dependencies amongst the spatial characteristics. This model is trained by a loss function in an end-to-end manner to learn the temporal correlations along with spatial correlations. Reference [21] proposed a predictive deep convolutional neural network (PDCNN), which is an integration of CNNs and a multilayer perceptron (MLP). Spatial characteristics are extracted by CNNs, and MLP is intended to construct a relationship between temporal and spatial features. However, the structure of [15,21] cannot be directly applied to WSF of wind farm clusters, since these two studies focus on wind turbines that are neatly arranged, and the CNN is intended to solve graphic issues [22].

It can be found that most existing works apply hybrid models, which bring about framework complications and calculation costs, thus reducing efficiency. They tend to have a longer training time period.

Additionally, few of them consider the robustness training of the input set [23]. The prediction accuracy cannot maintain a high degree of accuracy when the input data are continuously missing in a time interval, for instance, due to damage to the measuring devices or data transmission failure. Interpolation methods are always used to solve this problem, typically from two perspectives, spatial characteristics and temporal characteristics [24]. From a spatial perspective, the “1/7 power law”, “revised power law” and “ANFIS” are typical interpolation methods for dealing with wind speed prediction at different heights. Recently, a new method, the vertically correlated echelon model (VCEM), which utilizes vertical correlation of wind speeds, is proposed with a significant improvement in the

prediction accuracy [25]. From a temporal perspective, there are a few interpolation methods based on a time sequence. The most commonly used method is the moving average approach (MAA), which entirely neglects the spatiotemporal features of wind speed data. Therefore, the multidimensional average method (MDAM) is first proposed in this paper to utilize spatial characteristics so as to enhance data robustness.

To address the issues of insufficient utilization of the spatiotemporal features and inefficiency in large-volume data processing, as well as to improve the input data robustness, this paper proposes a new model termed input set based on wind farm cluster data–deep extreme learning machine (IWC-DELM). This model enlarges the input dataset by utilizing the historical data of adjacent wind farms with full consideration of their data correlation. This model constructs DELM as the main body for achieving high prediction efficiency. By means of adopting the multidimensional input set, this model can also enhance the wind speed prediction robustness. It is demonstrated in [26,27] that the RNN has a better performance than the CNN in time series data prediction. Therefore, some variants of CNNs, for instance, GRU, LSTM and SRNN, are selected to prove the validity of IWC-DELM.

The main contributions of this paper can be summarized as follows:

1. A new input configuration of the wind speed prediction model, i.e., an input set based on wind farm cluster data (IWC), is built. The capacity of the input set has been expanded by utilizing historical data of adjacent wind farms, thus fully considering the spatial features of wind speed sequences.
2. A new machine learning architecture, IWC-DELM, is proposed for the WSF within wind farm clusters. This model contributes to more accurate and efficient prediction compared to some promising deep learning methods. Three algorithms, GRU, LSTM and SRNN, are selected to verify the superiority of the proposed method.
3. Robustness analysis on the input set is performed. The forecasting accuracy is required to maintain a high level even if some input data are missing in a time interval. The MDAM, which completes the temporal features of data with the spatial features of the wind farm cluster, is first proposed for this purpose.

The rest of this paper is organized as follows: Section 2 introduces the main methods in data preprocessing, and Section 3 illustrates the model structure of the proposed method. A case study located in the USA is discussed in Section 4. Section 5 provides the conclusion.

2. Data Preprocessing Theory

2.1. Weighted Mean Filtering

Weighted mean filtering (WMF) is employed as a denoise method to replace the traditional methods in order to overcome the boundary effects issue [28]. Its transfer function can be described as:

$$\frac{D_{(d)}(Z)}{D(Z)} = \frac{\sum_{i=0}^L \alpha_{\tau-i} \cdot Z^{-i}}{\sum_{i=0}^L \alpha_{\tau-i}} \quad (1)$$

The denoised wind speed data and the corresponding raw data are represented by $D_{(d)}(Z)$ and $D(Z)$, respectively. (Z) denotes its Z-transform. $\alpha_{\tau-i}$ denotes the weight for each timestamp, and $L - 1$ represents the window size. Accordingly, the output of WMF in the time domain can be defined as:

$$D_{(d)}[\tau] = \frac{\sum_{i=0}^L \alpha_{\tau-i} \cdot D[\tau]}{\sum_{i=0}^L \alpha_{\tau-i}} \quad (2)$$

where $D[\tau]$ is the original wind speed data and $D_{(d)}[\tau]$ is the denoised data at a particular time instant τ .

2.2. Multidimensional Average Method

Traditional interpolation methods, for example, the MAA, make the missing data be determined as the average of the preceding data number with a defined autoregressive order [24]. In this paper, it is modified to apply in the cluster, which is defined as:

$$x_j = \frac{1}{\omega} \sum_{i=j-\omega}^{j-1} x_i \quad (3)$$

where x_j is the first missing data and ω is the autoregressive order. However, this method only uses the temporal characteristics of wind speed data.

To better utilize the spatiotemporal characteristics of wind speed, the MDAM is first proposed in this paper. As shown in Figure 1, the relevant data sequences from other wind farms within the same wind farm cluster are adopted in the MDAM to make up for the missing data. Assuming there are k relevant sequences with the same length, $x_{1,j}$, which stands for the first missing data of wind farm speed series x_1 , can be represented as:

$$x_{1,j} = \frac{x_{2,j} + x_{3,j} + \dots + x_{k,j}}{k-1} \quad (4)$$

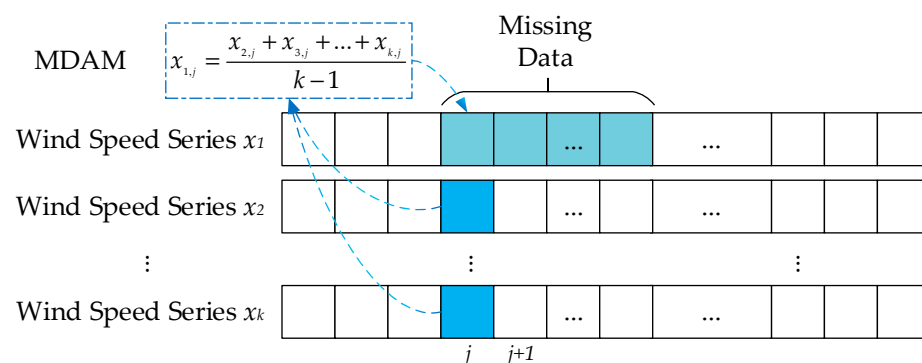


Figure 1. Schematic diagram of MDAM.

3. The Ensembled Model IWC-DELM

3.1. Deep Extreme Learning Machine

Extreme learning machine (ELM) is a popular feed-forward neural network for classification or regression uses, which was first proposed by Huang et al. in [29]. ELM has a good generalization performance along with a comparatively fast speed. Its trainable parameters connecting the input layer and hidden layer are randomly assigned instead of backpropagation [30]. Its output weights are obtained by calculating the generalized inverse operation of the hidden matrix [31]. Assuming there are l training samples, the output of ELM with L hidden neurons can be represented as:

$$y = \sum_{i=1}^L \eta_i h_i(x) = H\eta, \text{ for } i = 1, 2, \dots, l \quad (5)$$

$$H = g(wx + b) \quad (6)$$

where y represents the output vector and η_i represents output weight connecting the i th hidden layer and output neuron. H is the hidden layer matrix, and $g(\cdot)$ is the activation function. w and b denote input weight and bias, respectively. We can also have

$$H\eta = T \quad (7)$$

where T is the matrix of targets. η is determined by reaching the smallest training error between the output y and the target T .

$$\eta = \min \|T - H\eta\|_2^2 = H^+ T \quad (8)$$

where H^+ is the generalized inverse matrix of H .

Due to the shallow architecture of ELM, it is incapable of capturing the complex characteristics of input data [32]. To tackle this issue, deep extreme learning machine (DELM) was proposed in [33], whose configuration is shown in Figure 2. This model, utilizing a multilayer extreme learning machine (MLELM) and based on an extreme learning machine autoencoder (ELM-AE), takes advantage of both deep learning and ELM. The output of ELM-AE is the same as (6) and (9) is used to ensure the orthogonality of w and b .

$$w^T w = 1, b^T b = 1 \quad (9)$$

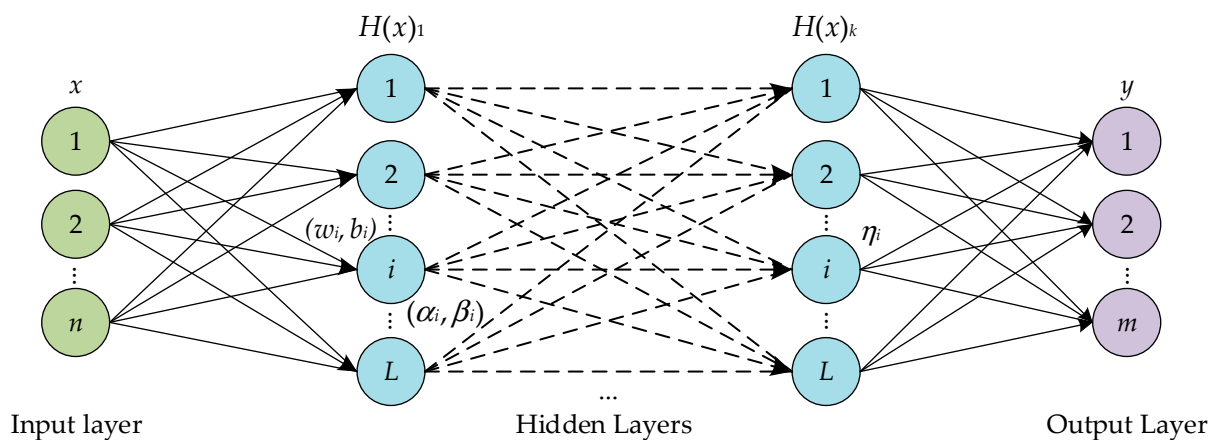


Figure 2. DELM configuration with n inputs and L -length hidden layer.

The relationship between the adjacent hidden layers can be expressed as:

$$H_j = g(\alpha_i H_{j-1} + \beta_i), \text{ for } i = 1, 2, \dots, L; j = 1, 2, \dots, k; \quad (10)$$

where α_i and β_i denote the weight and bias of the i th hidden neuron.

Unlike traditional machine learning methods using a gradient-based method, which include many iterations and deep learning models, which contain a memory unit leading to a slow procession, DELM determines the output weight by calculation of a hidden matrix [34]. Therefore, DELM shows great efficiency in processing big-capacity data.

3.2. Input Set Based on Wind Farm Cluster Data

Figure 3 illustrates a newly proposed approach, IWC, for input set construction. It is defined as follows:

$$X = \begin{bmatrix} x_{a,1} & x_{a,2} & \cdots & x_{a,m} \\ x_{b,1} & x_{b,2} & \cdots & x_{b,m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,m} \end{bmatrix} \quad (11)$$

where X is the constructed n -dimensional input set, $x_a, x_b, \dots$, and x_n is wind speed series from wind farm $a, b, \dots$, and n .

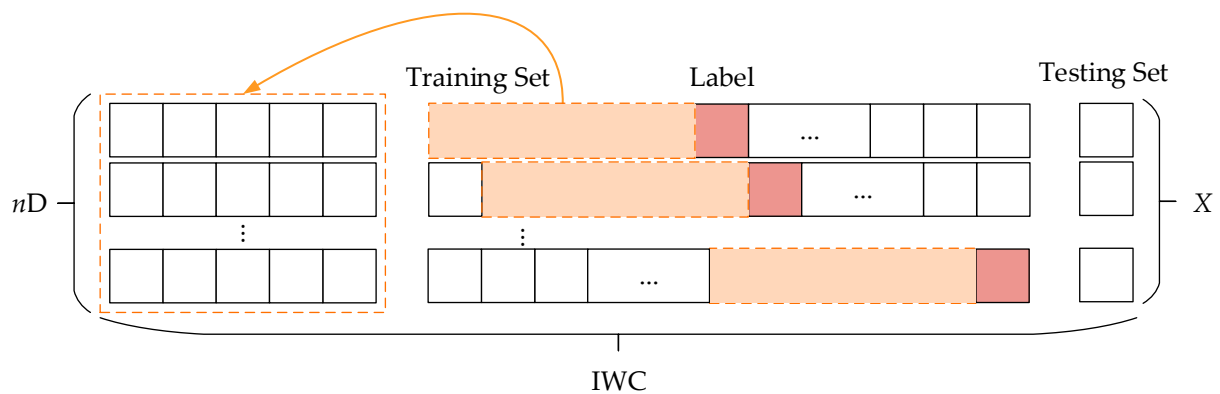


Figure 3. Schematic diagram of IWC configuration.

In Figure 3, an n -dimensional training set is constructed based on wind farm cluster data containing n wind farms. X contains both the training set and the testing set. The overall process is named the IWC approach.

3.3. The Proposed IWC-DELM

Figure 4 illustrates a newly proposed model for IWC-DELM that consists of n -dimensional input sets and three major steps.

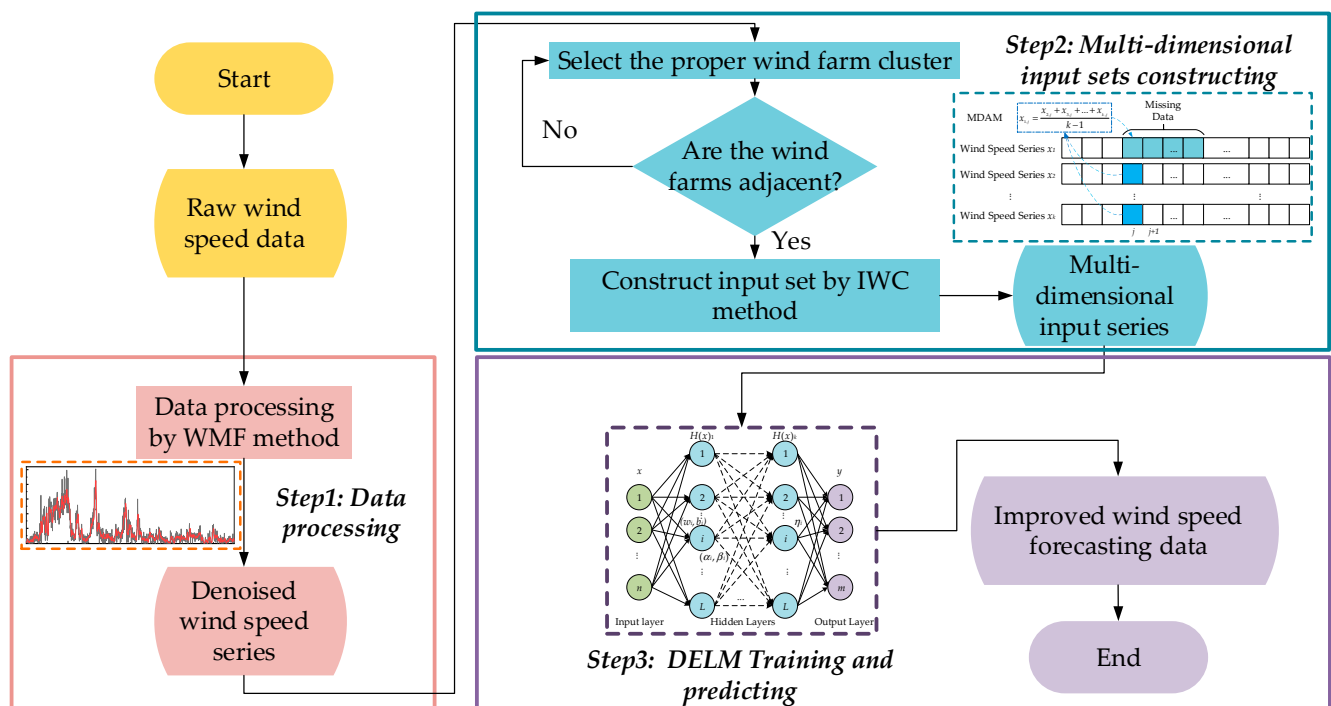


Figure 4. Flowchart of IWC-DELM.

Step 1: During the data cleaning preprocess, the original wind speed data is filtered by WMF to suppress the white noise in the original time series. WMF can maintain the causality of the whole system and reduce the noise at the same time [28].

Step 2: Once the denoised wind speed series is obtained, a proper wind farm cluster should be selected to construct a multidimensional input set of DELM using the IWC method. Moreover, whenever the wind speed of any wind farm in a cluster is unavailable, the model will make up for the continuous missing data via the MDAM.

Step 3: As described in Section 3.2, the multidimensional input set is used to train the prediction model of DELM. Then, it is applied to future prediction to obtain improved wind speed forecasting data.

Parameters of the IWC-DELM model can be found in Table 1.

Table 1. Configuration of the proposed ensemble model IWC-DELM.

Type	Configuration	
WMF	Batch extent	5
	Weights	[0.80, 0.64, 0.51, 0.41, 0.33]
IWC	$X = \begin{bmatrix} x_{a,1} & x_{a,2} & \cdots & x_{a,m} \\ x_{b,1} & x_{b,2} & \cdots & x_{b,m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,m} \end{bmatrix} \begin{cases} X_a \text{ data series of wind farm } a \\ X_b \text{ data series of wind farm } b \\ \cdots \\ X_n \text{ data series of wind farm } n \end{cases}$	
	Layers	Hyperparameters
DELM	Hidden layer 1	Input: 5×3 Nodes: 30
	Hidden layer 2	Input: 10×1 Nodes: 15

4. Case Study

4.1. Datasets and Evaluation Indices

The datasets used in this paper are shown in Figure 5 with their longitudes and latitudes. Further information, both data and maps, can be found in the data availability statement. Site 1 to Site 5 are adjoining to one other, and Site 6 and Site 7 are apart from these five wind farms. Figure 6 illustrates the data preprocessing of seven selected wind farms by WMF.

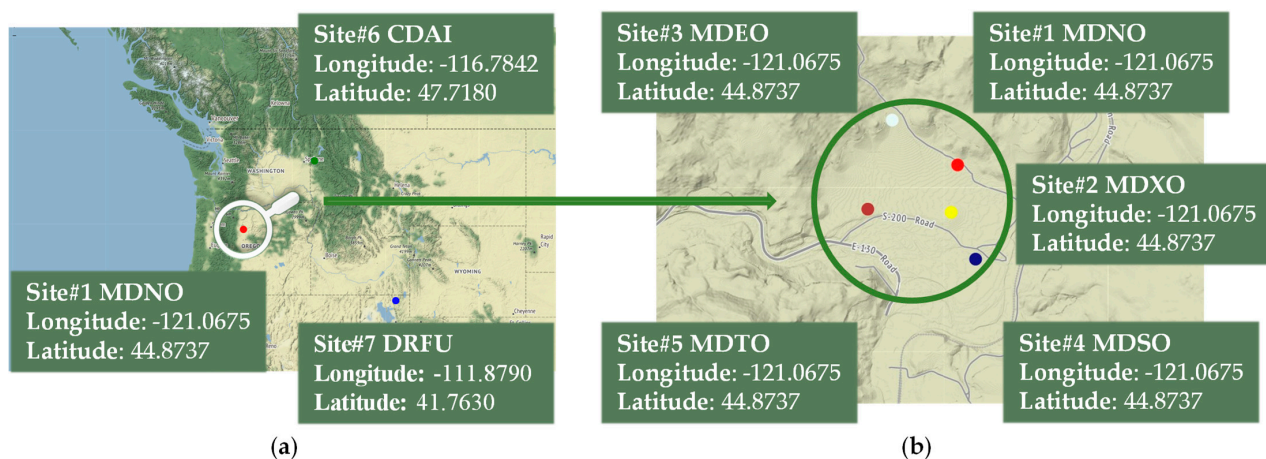


Figure 5. Seven selected wind farms with their longitudes and latitudes: (a) nonadjacent wind farms from Site 1 to Site 6 to Site 7; (b) adjacent wind farms from Site 1 to Site 5.

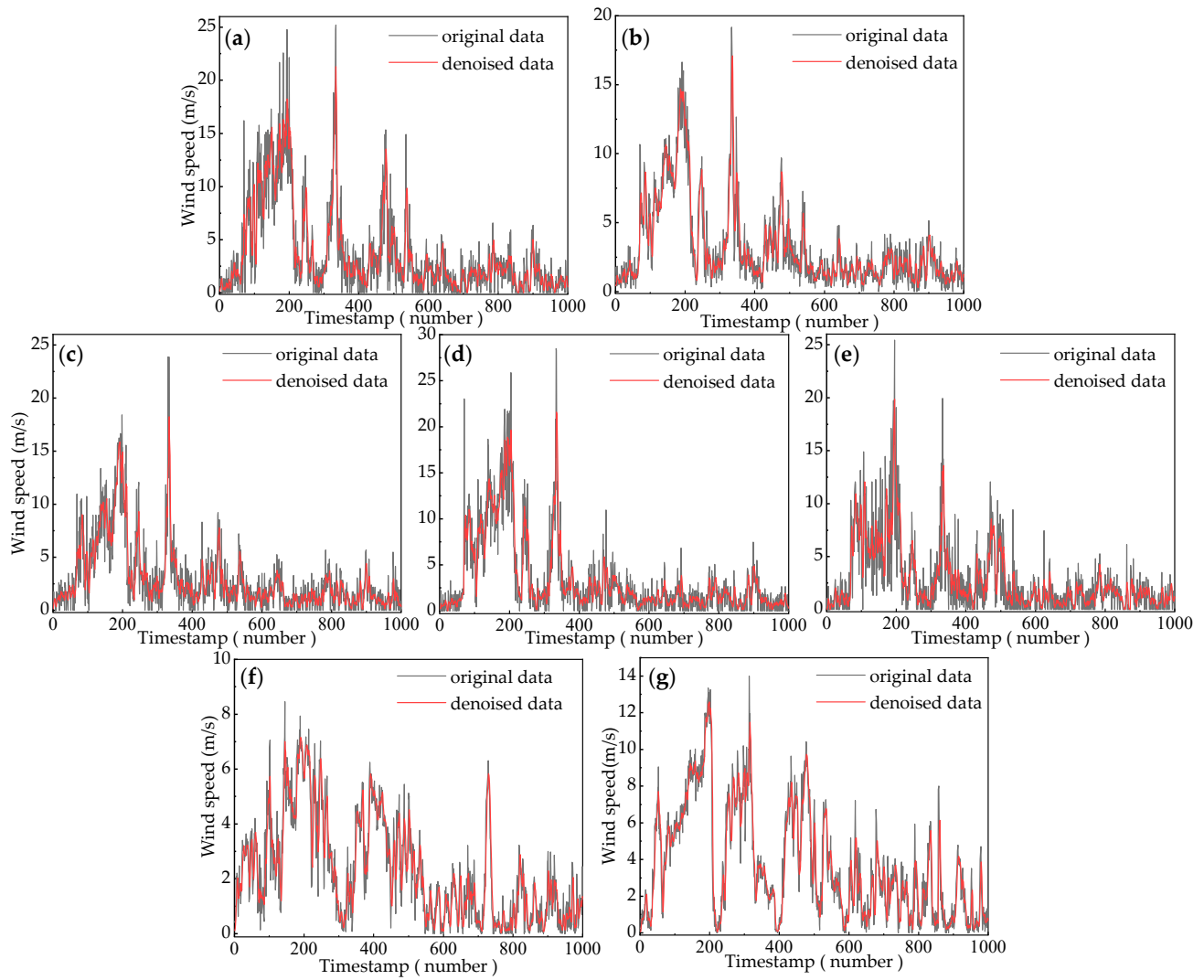


Figure 6. Data preprocessing on the historical wind speed sequences of seven selected wind farms: (a) Site#1, (b) Site#2, (c) Site#3, (d) Site#4, (e) Site#5, (f) Site#6, (g) Site#7.

Four indices, the root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and R-squared (R^2), are used for evaluation [28,35]. Their definitions are available below:

$$\text{RMSE} = \sqrt{\frac{1}{\mu} \sum_{i=1}^{\mu} (y_i^p - y_i)^2} \quad (12)$$

$$\text{MAE} = \frac{1}{\mu} \sum_{i=1}^{\mu} |y_i^p - y_i| \quad (13)$$

$$\text{MAPE} = \frac{1}{\mu} \sum_{i=1}^{\mu} \left| \frac{y_i^p - y_i}{y_i} \right| \quad (14)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{\mu} (y_i^p - y_i)^2}{\sum_{i=1}^{\mu} (\bar{y} - y_i)^2} \quad (15)$$

where y^p denotes the wind speed prediction, y is the real data and $\bar{y}$ represents the mean value of real data. μ represents the wind speed sequence length.

To measure the improvement of these four indices, η_I is defined as:

$$\eta_I = -\frac{I' - I}{I} \times 100\% \quad (16)$$

where the I represents any of the four indices and I' is the improved one. If η_I is positive, it indicates lower error and better performance, and vice versa.

4.2. Tests under Various Influential Factors

4.2.1. Accuracy Analysis of Different Methods Operated on Different Input Sets

Historical data of the adjacent wind farms are involved in the input dataset in order to utilize the spatial relevance of wind sequences in wind farm clusters. The number of adopted wind farms changes from one to five, which means the input dataset can be from one-dimensional to five-dimensional (i.e., 1D to 5D). Three models, GRU, LSTM and SRNN, are adopted as candidate algorithms for comparison.

Table 2 and Figure 7 illustrate the performance of three promising algorithms and the selected algorithm DELM. SRNN and DELM obtain a similar performance, which is better than those of GRU and LSTM. DELM shows more competitive capability in terms of all error indices at 3D input, with an RMSE of 0.288, MAE of 0.224, MAPE of 0.140 and R^2 of 0.898. Four indices of GRU and LSTM become larger along with an increase in input dimension, indicating a drop in prediction accuracy. For instance, the RMSE of GRU increases from 0.316 to 0.628. Their lack of capabilities of addressing large-volume data and overfitting issues may result in this phenomenon, which indicates that GRU and LSTM are not suitable for the prediction of wind farm clusters herein. In contrast, the forecasting accuracies of SRNN and DELM are improved when enlarging the input dimension. The RMSE of DELM declines from 0.325 to 0.287. The fitting lines shown in Figure 8 clearly represent their capacities for prediction. Prediction of DELM is always the closest to the real wind speed data. These experimental results indicate that utilizing relevant spatial data plays a significant role in prediction accuracy improvement.

Table 2. Performance of different models with diverse input dimensions from 1 to 5 *.

Model	Indices	1D	2D	3D	4D	5D
GRU	RMSE	0.316	0.452	0.564	0.658	0.628
	MAE	0.267	0.353	0.442	0.507	0.489
	MAPE	0.529	0.674	1.412	1.978	1.979
	R^2	0.898	0.831	0.766	0.765	0.723
LSTM	RMSE	0.319	0.390	0.452	0.722	0.725
	MAE	0.257	0.301	0.344	0.558	0.585
	MAPE	0.438	0.364	0.648	1.911	2.458
	R^2	0.896	0.858	0.874	0.763	0.728
SRNN	RMSE	0.352	0.302	0.289	0.267	0.298
	MAE	0.280	0.241	0.235	0.216	0.234
	MAPE	0.235	0.137	0.266	0.226	0.237
	R^2	0.841	0.872	0.891	0.903	0.872
DELM (selected algorithm)	RMSE	0.325	0.301	0.288	0.286	0.287
	MAE	0.262	0.234	0.224	0.228	0.236
	MAPE	0.354	0.168	0.140	0.145	0.297
	R^2	0.777	0.809	0.898	0.886	0.875

* The minimum error of 4 indices among candidate algorithms with different input dimensions are in bold.

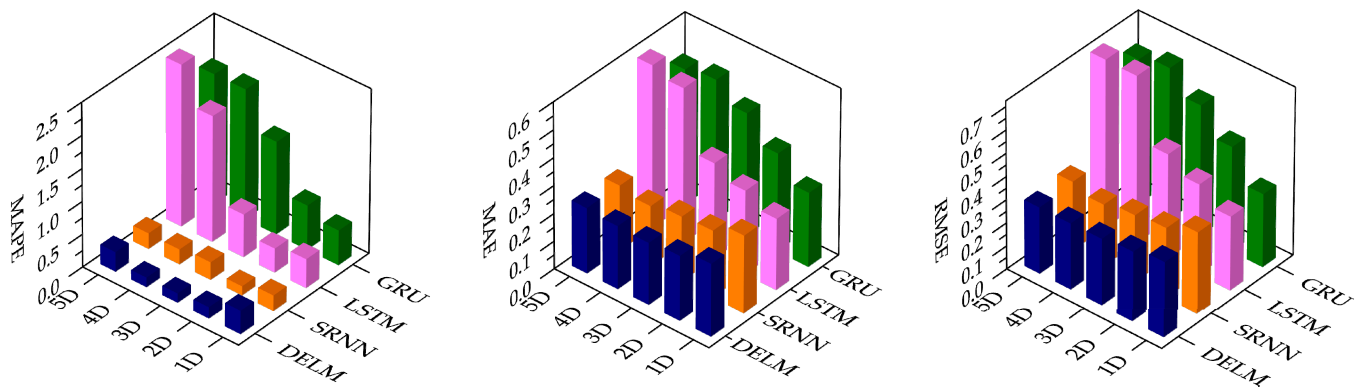


Figure 7. Error indices of different models.

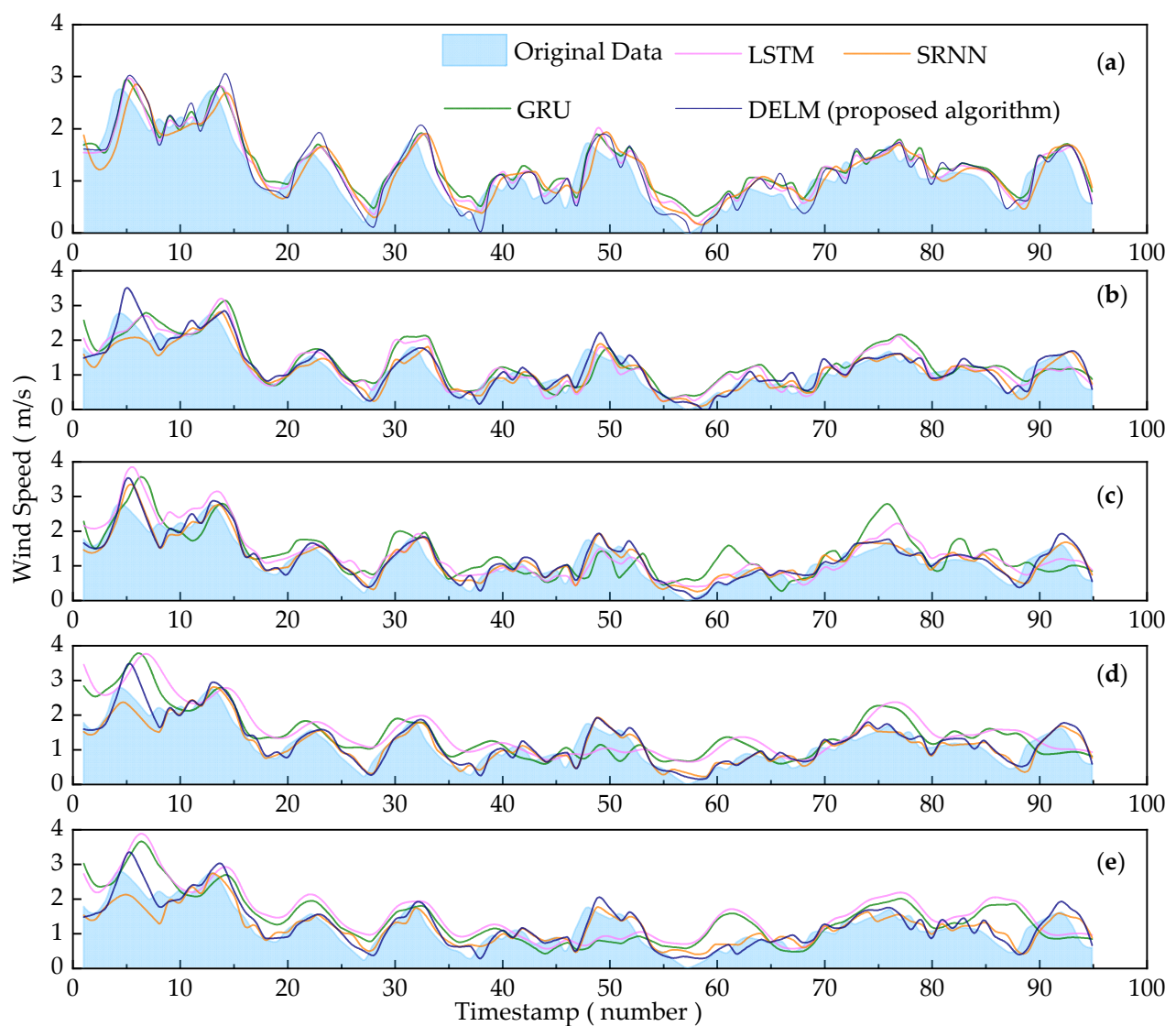


Figure 8. Fitting lines of different models with input dimensions from 1 to 5: (a) 1D input set; (b) 2D input set; (c) 3D input set; (d) 4D input set; (e) 5D input set.

4.2.2. Efficiency Analysis of Different Methods Operated on Different Input Sets

To measure the efficiency of different prediction methods, their training periods are displayed in Table 3. Three comparative methods and DELM are conducted with 1D to

5D input set conditions. Figure 9 shows the efficiency improvement percentage of three candidates compared to DELM.

Table 3. CPU time for both training and testing sets of different models (s) *.

Model	1D	2D	3D	4D	5D
GRU	105.225	165.201	195.250	227.745	306.538
LSTM	140.558	240.385	342.736	410.983	501.438
SRNN	51.800	55.793	57.255	58.372	56.018
DELM (selected algorithm)	0.228	0.268	0.279	0.343	0.462

* The minimum training time with different input dimensions is in bold.

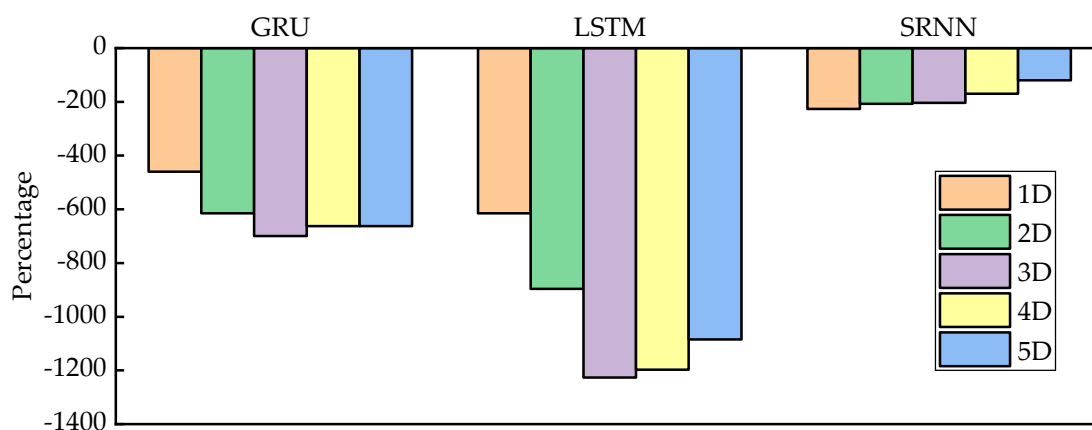


Figure 9. The efficiency improvement percentage of three candidates compared to DELM.

According to Table 3, when the input dimension ranges from one to five, the training times of GRU and LSTM increase sharply from 105.225 s to 306.538 s and from 140.558 s to 501.438 s, respectively. The time cost of SRNN and DELM shows a slight rise. The SRNN period increases by 4.218 s, and the DELM period with five dimensions is only 0.234 s longer than with one dimension. Despite the fact that the training time of DELM increases by 102.63% compared to the 1D input, its absolute CPU time is comparatively shorter than others. These results mainly stem from different model configurations. GRU and LSTM cannot be computed in parallel owing to their recurrent structure [10]. Every current input is connected to its previous step, so the larger the input database is, the longer it will take for computation. However, SRNN, as an improvement of this recurrent structure, slices input data into subsequences so that each subsequence can be operated simultaneously, leading to a markable reduction in training time [36]. The larger the input dataset is, the more significant the speed advantage SRNN achieves. When the input dimension reaches five, SRNN only needs 56.018 s, while GRU and LSTM take 306.538 s and 501.438 s, respectively.

It is mentioned that the parameters of the three comparison models are consistent with those in [1,9,10] as shown in Table 4. Among the mentioned four models, the selected DELM markedly surpasses the other models in training time. As shown in Figure 9 and Table 4, its speed is hundreds or even thousands of times faster than the other three candidate methods as it has the least parameters. The learning period of DELM is extremely fast, which can be completed within one second [36,37].

Table 4. Number of parameters of different methods.

Model	GRU	LSTM	SRNN	DELM (Selected Algorithm)
Number of parameters	845,601	1,849,441	4137	3600

4.2.3. Comparison of Datasets with Different Correlation Degrees

To evaluate the validity of IWC-DELM model, we chose the prediction results of a single wind farm as a blank control group and compared the results of wind farm clusters with high relevance and weak relevance. Adjacent wind farms and nonadjacent wind farms represent strong and weak correlations, respectively, since the distance between wind farms is positively related to correlation.

The performance of three different input datasets is displayed in Table 5. When the input set includes adjacent wind farm data, the RMSE declines by 11.38%, MAE drops by 14.50%, MAPE dramatically falls by 60.56% and R^2 increases by 5.40% compared to the single-wind-farm prediction. Inversely, nonadjacent farm data input leads to the RMSE, MAE and MAPE experiencing a rise of 2.15%, 5.34% and 1.98%, respectively. Figure 10 intuitively shows the fitting lines of the three conditions mentioned above. It can be obviously observed that forecasting of adjacent wind farms is most close to the original data.

Table 5. Performance of different input datasets *.

	Single Wind Farm	Wind Farm Cluster			
		Nonadjacent Farms	Improvement	Adjacent Farms	Improvement
RMSE	0.325	0.332	−2.15%	0.288	11.38%
MAE	0.262	0.276	−5.34%	0.224	14.50%
MAPE	0.354	0.361	−1.98%	0.140	60.45%
R^2	0.777	0.748	−3.73%	0.819	5.40%

* The minimum error and maximum improvement are in bold.

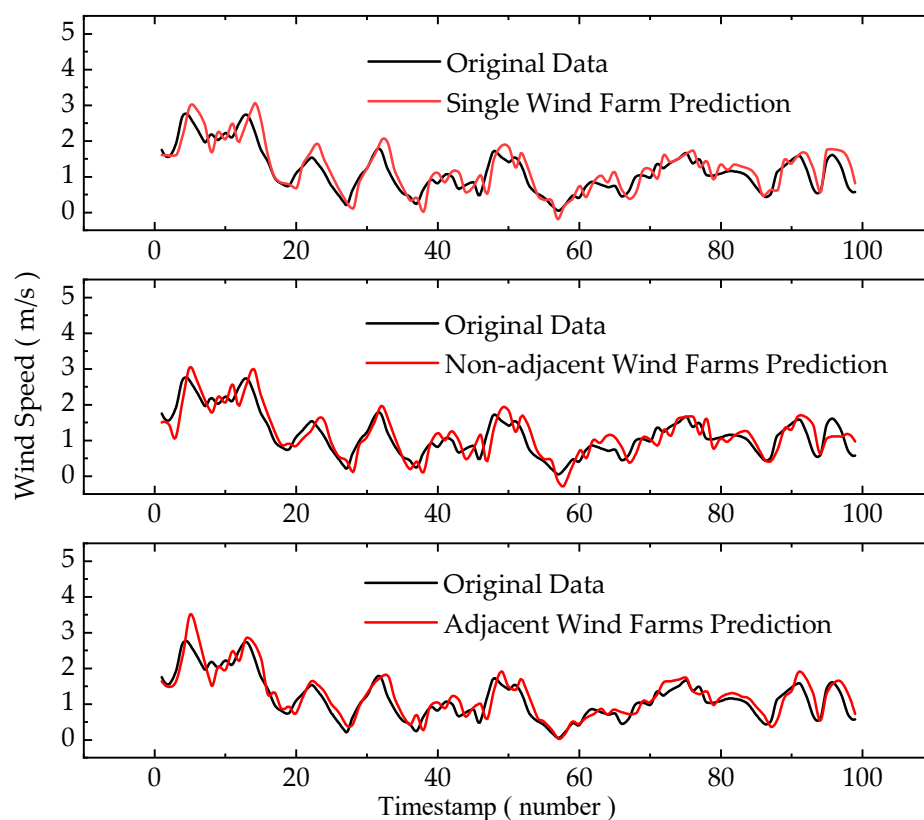


Figure 10. Fitting lines of different data inputs.

Therefore, a conclusion can be drawn that the forecasting accuracy is improved with a highly correlated historical data input, which verifies the effectiveness of the proposed model utilizing spatial corrections of wind speed sequences in wind farm clusters.

4.2.4. Discussion

According to the experiment results shown above, the prediction accuracies of SRNN and DELM are higher than those of GRU and LSTM. In addition, the operation efficiency of DELM is outstanding. Therefore, DELM is selected as the main body of the model, and the 3D input is designed due to the balance between training cost and prediction accuracy. The input dataset is constructed by wind farm data with a strong correlation. This proposed model, termed IWC-DELM and proved validly, can achieve a better wind speed forecasting result.

The main advantages of the proposed model can be summarized as follows:

1. DELM has boasted its operational efficiency as it determines the output weight by simple computations of the hidden matrix. It also employs the multi-hidden-layer structure to capture complex nonlinear characteristics. Hence, using DELM as the main body of the model is critical for forecasting accuracy and training speed improvement.
2. Considering the spatial correlation, the multidimension input can improve the forecasting capacity. The proposed model is an integration of DELM and IWC, which outperforms the other candidate algorithms with a smaller prediction fluctuance, better adaptiveness and greatly enhanced efficiency.

4.3. Input Dataset Robustness Analysis

Robustness analysis is conducted to prevent the prediction capability from a significant drop in extreme cases. Figure 11 shows the selected two periods that experience 5, 10 and 15 missing points, respectively. Those missing points in single-farm prediction are made up by the MAA [25], while those in wind farm cluster prediction are made up by the MDAM. Error indices of prediction results of the single wind farm and wind farm cluster are displayed in Table 6. Their improvement compared to the original forecasting is shown in Table 7.

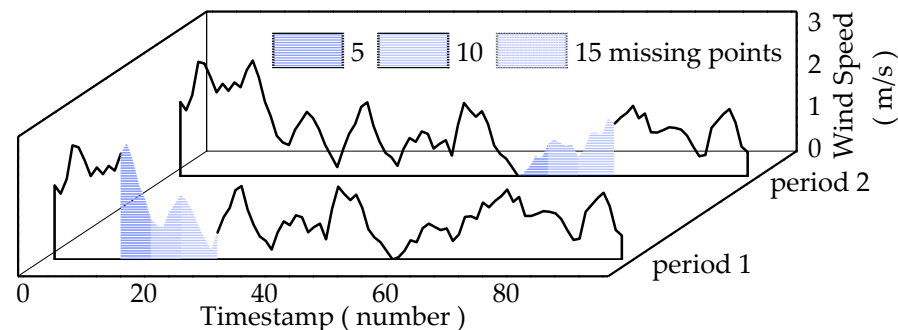


Figure 11. Selected two periods where data are missing.

Table 6. Comparison of single-farm and wind farm cluster performance in different conditions of missing data *.

Missing Period	Indices	Original Prediction	Single Farm			Wind Farm Cluster		
			MAA			MDAM (Proposed Method)		
			5 Missing	10 Missing	15 Missing	5 Missing	10 Missing	15 Missing
Period 1	RMSE	0.288	0.340	0.386	0.448	0.296	0.312	0.331
	MAE	0.224	0.267	0.303	0.342	0.234	0.245	0.257
	MAPE	0.140	0.373	0.439	1.078	0.153	0.168	0.308
	R ²	0.898	0.724	0.548	0.238	0.858	0.734	0.702
Period 2	RMSE	0.288	0.328	0.445	0.454	0.291	0.288	0.292
	MAE	0.224	0.263	0.352	0.365	0.231	0.226	0.230
	MAPE	0.140	0.386	0.281	0.306	0.145	0.144	0.145
	R ²	0.898	0.827	0.582	0.292	0.796	0.726	0.705

* The minimum error in the different missing conditions is in bold.

Table 7. Improvement of single-farm and wind farm cluster prediction in different conditions of missing data *.

Missing Period	Indices	Single Farm			Wind Farm Cluster		
		MAA			MDAM (Proposed Method)		
		5 Missing	10 Missing	15 Missing	5 Missing	10 Missing	15 Missing
Period 1	η_{RMSE}	−18.06%	−34.03%	−55.56%	−2.78%	−8.33%	−14.93%
	η_{MAE}	−19.20%	−35.27%	−52.68%	−4.46%	−9.38%	−14.73%
	η_{MAPE}	−166.43%	−213.57%	−670.00%	−9.29%	−20.00%	−120.00%
	η_{R2}	−19.38%	−38.98%	−73.50%	−4.45%	−18.26%	−21.83%
Period 2	η_{RMSE}	−13.89%	−54.51%	−57.64%	−1.04%	−0.32%	−1.39%
	η_{MAE}	−17.41%	−57.14%	−62.95%	−3.13%	−0.89%	−2.68%
	η_{MAPE}	−175.71%	−100.71%	−118.57%	−3.57%	−2.86%	−3.57%
	η_{R2}	−7.91%	−35.19%	−67.48%	−11.36%	−19.15%	−21.49%

* The maximum improvement ratios in the different missing conditions are in bold.

According to Table 6, as a result of missing data, the negative η_I indicates that the performance of both the single farm and wind farm cluster are worse than the original one without missing data. As the numbers of lost data increase, the four indices decline in both prediction forms and in both periods. Table 7 illustrates that a wind farm cluster can achieve a better prediction than a single wind farm. Especially in the 15-point-missing condition in period 2, the improvement of the RMSE in the wind farm cluster only falls by 1.39% compared to 57.64% in the single wind farm. This can be attributed to different interpolation methods. Our MDAM adopts the historical wind speed sequences of two adjacent wind farms, which characterize the original wind speed data.

Additionally, the accuracy reduction in period 2 is not as serious as in period 1, which is mainly arisen from the sharp variation in wind speed in period 1. For instance, in the 15-point-missing condition, the RMSE in period 2 drops by 1.39%, while that in period 1 declines by 14.93%.

Using a paired t -test to study the differences in experimental data [38], it can be seen from the Table 8 that there are six groups of paired data in total, amongst which five groups of paired data show differences ($p < 0.05$). With the increase in missing data, the p value becomes smaller ($p < 0.01$), and this indicates a more significant difference between the MAA and MDAM.

Table 8. Results of paired t -test analysis of missing data in both MAA and MDAM methods *.

Missing Period	Paired Name	Pair (Mean $\pm$ Standard Deviation)		Difference (Pair 1–Pair 2)	t	p
		Pair 1	Pair 2			
Period 1	MAA(5) vs. MDAM(5)	1.17 $\pm$ 0.48	1.15 $\pm$ 0.48	0.03	0.664	0.517
	MAA(10) vs. MDAM(10)	1.55 $\pm$ 0.62	1.28 $\pm$ 0.47	0.27	4.144	0.001 ***
	MAA(15) vs. MDAM(15)	1.57 $\pm$ 0.00	1.37 $\pm$ 0.29	0.21	2.756	0.015 **
Period 2	MAA(5) vs. MDAM(5)	0.71 $\pm$ 0.29	1.01 $\pm$ 0.32	−0.30	−2.611	0.021 **
	MAA(10) vs. MDAM(10)	0.53 $\pm$ 0.37	0.98 $\pm$ 0.35	−0.45	−3.822	0.002 ***
	MAA(15) vs. MDAM(15)	0.62 $\pm$ null	1.02 $\pm$ 0.34	−0.40	−4.578	0.000 ***

* MAA(5), MAA(10) and MAA(15) mean 5, 10 and 15 missing values in MAA method, while the same rule holds for MDAM as well. ** $p < 0.05$; *** $p < 0.01$

From these experimental results, it can be concluded that the proposed MDAM can enhance the data robustness, thus leading to a reduction in accuracy loss in some special cases.

5. Conclusions

Existing studies focus on WSF in wind farm clusters without consideration of spatial correlation. Additionally, there are few effective solutions for dealing with missing data conditions. Therefore, IWC-DELM is proposed in this paper based on a combination of IWC and MDAM. IWC takes advantage of spatial correlations within wind farm clusters by employing the historical data of adjacent wind farms. DELM is selected as the main body of this proposed model for its extremely fast speed in processing. Robustness analysis is conducted using the MDAM to prevent prediction accuracy from a dramatic drop in the missing data conditions. Therefore, IWC-DELM outperforms some promising deep learning algorithms, for instance, GRU, LSTM and SRNN, in both accuracy and efficiency, especially when processing large volumes of wind farm cluster data. It is mentioned above that the parameters of the three comparison models are consistent with those in [1,9,10]. Four indices, RMSE, MAE, MAPE and R^2 , in the 3D input condition using DELM are improved by 48.94%, 49.32%, 90.08% and 17.23%, respectively, compared to GRU and are improved by 36.28%, 34.88%, 78.40% and 2.75%, respectively, compared to LSTM. As for different input datasets, IWC-DELM used in adjacent wind farms achieves an improvement in RMSE, MAE, MAPE and R^2 of 11.38%, 14.50%, 60.45% and 5.40%, respectively, compared to nonadjacent wind farms. When dealing with missing data, the proposed method, MDAM, also performs better than the MAA. For instance, the MAE is improved by 9.38~25.89% with 5, 10 and 15 missing data. Such an accurate prediction can be applied in broad fields, such as the making of grid operation strategies, and is available for meteorological usage.

Meanwhile, there is still room for improvement. Firstly, the selected DELM could be not the best option with the advent of other state-of-the-art forecasting methods. Moreover, hybrid state-of-the-art approaches in both wind prediction studies and other AI-assisted contributions can be applied in terms of handling missing data. Some promising statistics and probability methods should be taken into consideration as well. Numerical weather prediction (NWP) can be also adopted to further expand the input dataset since the current dataset only includes historical data on wind speed.

Author Contributions: Methodology, M.Z. and J.T.; software, M.Z. and H.Z.; writing—original draft preparation, Y.W.; writing—review and editing, M.Z., Y.W., H.Z., Z.P. and J.T.; supervision, J.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China [52177071] and Chongqing Municipal Training Program of Innovation and Entrepreneurship for Undergraduates [S 202110611442]. And the APC was funded by [52177071].

Data Availability Statement: The datasets used in this paper are available at the website: <https://www.usbr.gov/pn/agrimet/webaghrread.html>, accessed on 16 January 2023. The main prediction method code used in this article can be accessed at the following website for reference. In addition, the codes written by the authors, which are not available below, will be provided by the corresponding author via email once they are required by the readers. GRU: <https://github.com/keras-team/keras/blob/v2.11.0/keras/layers/rnn/gru.py>, accessed on 16 January 2023. LSTM: <https://github.com/keras-team/keras/blob/v2.11.0/keras/layers/rnn/lstm.py#L382-L893>, accessed on 16 January 2023. SRNN: <https://github.com/zepingyu0512/srnn>, accessed on 16 January 2023. DELM: <https://github.com/atefeta/Deep-Extreme-Learning-Machine>, accessed on 16 January 2023.

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

References

1. Zhou, B.; Ma, X.; Luo, Y.; Yang, D. Wind Power Prediction Based on LSTM Networks and Nonparametric Kernel Density Estimation. *IEEE Access* **2019**, *7*, 165279–165292. [\[CrossRef\]](#)
2. Qin, Y.; Li, K.; Liang, Z.; Lee, B.; Zhang, F.; Gu, Y.; Zhang, L.; Wu, F.; Rodriguez, D. Hybrid forecasting model based on long short term memory network and deep learning neural network for wind signal. *Appl. Energy* **2019**, *236*, 262–272. [\[CrossRef\]](#)
3. Liu, X.; Yang, L.; Zhang, Z. Short-Term Multi-Step Ahead Wind Power Predictions Based on A Novel Deep Convolutional Recurrent Network Method. *IEEE Trans. Sustain. Energy* **2021**, *12*, 1820–1833. [\[CrossRef\]](#)
4. Alanis, A.Y.; Sanchez, O.D.; Alvarez, J.G. Time Series Forecasting for Wind Energy Systems Based on High Order Neural Networks. *Mathematics* **2021**, *9*, 1075. [\[CrossRef\]](#)
5. Wang, H.Z.; Wang, G.B.; Li, G.Q.; Peng, J.C.; Liu, Y.T. Deep belief network based deterministic and probabilistic wind speed forecasting approach. *Appl. Energy* **2016**, *182*, 80–93. [\[CrossRef\]](#)
6. 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. [\[CrossRef\]](#)
7. Zhang, H.; Liu, Y.; Yan, J.; Han, S.; Li, L.; Long, Q. Improved Deep Mixture Density Network for Regional Wind Power Probabilistic Forecasting. *IEEE Trans. Power Syst.* **2020**, *35*, 2549–2560. [\[CrossRef\]](#)
8. Khan, N.; Ullah, F.U.M.; Haq, I.U.; Khan, S.U.; Lee, M.Y.; Baik, S.W. AB-Net: A Novel Deep Learning Assisted Framework for Renewable Energy Generation Forecasting. *Mathematics* **2021**, *9*, 2456. [\[CrossRef\]](#)
9. 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. [\[CrossRef\]](#)
10. Yu, Z.; Liu, G. Sliced recurrent neural networks. *arXiv* **2018**, arXiv:1807.02291.
11. Meng, A.; Zhu, Z.; Deng, W.; Ou, Z.; Lin, S.; Wang, C.; Xu, X.; Wang, X.; Yin, H.; Luo, J. A novel wind power prediction approach using multivariate variational mode decomposition and multi-objective crisscross optimization based deep extreme learning machine. *Energy* **2022**, *260*, 124957. [\[CrossRef\]](#)
12. An, G.; Jiang, Z.; Chen, L.; Cao, X.; Li, Z.; Zhao, Y.; Sun, H. Ultra Short-Term Wind Power Forecasting Based on Sparrow Search Algorithm Optimization Deep Extreme Learning Machine. *Sustainability* **2021**, *13*, 10453. [\[CrossRef\]](#)
13. Chen, N.; Qian, Z.; Nabney, I.T.; Meng, X. Wind Power Forecasts Using Gaussian Processes and Numerical Weather Prediction. *IEEE Trans. Power Syst.* **2014**, *29*, 656–665. [\[CrossRef\]](#)
14. Pinhão, M.; Fonseca, M.; Covas, R. Electricity Spot Price Forecast by Modelling Supply and Demand Curve. *Mathematics* **2022**, *10*, 2012. [\[CrossRef\]](#)
15. Zhu, Q.; Chen, J.; Shi, D.; Zhu, L.; Bai, X.; Duan, X.; Liu, Y. Learning Temporal and Spatial Correlations Jointly: A Unified Framework for Wind Speed Prediction. *IEEE Trans. Sustain. Energy* **2020**, *11*, 509–523. [\[CrossRef\]](#)
16. Wang, Q.; Martinez-Anido, C.B.; Wu, H.; Florita, A.R.; Hodge, B.-M. Quantifying the Economic and Grid Reliability Impacts of Improved Wind Power Forecasting. *IEEE Trans. Sustain. Energy* **2016**, *7*, 1525–1537. [\[CrossRef\]](#)
17. Hu, Q.; Zhang, S.; Yu, M.; Xie, Z. Short-Term Wind Speed or Power Forecasting with Heteroscedastic Support Vector Regression. *IEEE Trans. Sustain. Energy* **2016**, *7*, 241–249. [\[CrossRef\]](#)
18. Zhang, J.; Liu, D.; Li, Z.; Han, X.; Liu, H.; Dong, C.; Wang, J.; Liu, C.; Xia, Y. Power prediction of a wind farm cluster based on spatiotemporal correlations. *Appl. Energy* **2021**, *302*, 117568. [\[CrossRef\]](#)
19. Peng, X.; Chen, Y.; Cheng, K.; Wang, H.; Zhao, Y.; Wang, B.; Che, J.; Liu, C.; Wen, J.; Lu, C.; et al. Wind Power Prediction for Wind Farm Clusters Based on the Multifeature Similarity Matching Method. *IEEE Trans. Ind. Appl.* **2020**, *56*, 4679–4688. [\[CrossRef\]](#)
20. Ying, X.; Zhao, K.; Liu, Z.; Gao, J.; He, D.; Li, X.; Xiong, W. Wind Speed Prediction via Collaborative Filtering on Virtual Edge Expanding Graphs. *Mathematics* **2022**, *10*, 1943. [\[CrossRef\]](#)
21. Zhu, Q.; Chen, J.; Zhu, L.; Duan, X.; Liu, Y. Wind Speed Prediction with Spatio-Temporal Correlation: A Deep Learning Approach. *Energies* **2018**, *11*, 705. [\[CrossRef\]](#)
22. Ali, M.; Prasad, R.; Xiang, Y.; Sankaran, A.; Deo, R.C.; Xiao, F.; Zhu, S. Advanced extreme learning machines vs. deep learning models for peak wave energy period forecasting: A case study in Queensland, Australia. *Renew. Energy* **2021**, *177*, 1031–1044. [\[CrossRef\]](#)
23. Wan, J.; Huang, J.; Liao, Z.; Li, C.; Liu, P.X. A Multi-View Ensemble Width-Depth Neural Network for Short-Term Wind Power Forecasting. *Mathematics* **2022**, *10*, 1824. [\[CrossRef\]](#)
24. Lotfi, B.; Mourad, M.; Najiba, M.B.; Mohamed, E. Treatment methodology of erroneous and missing data in wind farm dataset. In Proceedings of the Eighth International Multi-Conference on Systems, Signals & Devices, Sousse, Tunisia, 22–25 March 2011; pp. 1–6.
25. Lin, Q.; Wang, J. Vertically Correlated Echelon Model for the Interpolation of Missing Wind Speed Data. *IEEE Trans. Sustain. Energy* **2014**, *5*, 804–812. [\[CrossRef\]](#)
26. O’Leary, C.; Lynch, C.; Bain, R.; Smith, G.; Grimes, D. A Comparison of Deep Learning vs Traditional Machine Learning for Electricity Price Forecasting. In Proceedings of the 2021 4th International Conference on Information and Computer Technologies (ICICT), Kahului, HI, USA, 11–14 March 2021; pp. 6–12.
27. Peña-Gallardo, R.; Medina-Rios, A. A comparison of deep learning methods for wind speed forecasting. In Proceedings of the 2020 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC), Ixtapa, Mexico, 4–6 November 2020; pp. 1–6.

28. Zhang, H.; Peng, Z.; Tang, J.; Dong, M.; Wang, K.; Li, W. A multi-layer extreme learning machine refined by sparrow search algorithm and weighted mean filter for short-term multi-step wind speed forecasting. *Sustain. Energy Technol. Assess.* **2022**, *50*, 101698. [[CrossRef](#)]
29. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme learning machine: A new learning scheme of feedforward neural networks. In Proceedings of the 2004 IEEE International Joint Conference on Neural Networks, Budapest, Hungary, 25–29 July 2004; Volume 1–4, pp. 985–990.
30. Zhao, Y.; Ye, L.; Li, Z.; Song, X.; Lang, Y.; Su, J. A novel bidirectional mechanism based on time series model for wind power forecasting. *Appl. Energy* **2016**, *177*, 793–803. [[CrossRef](#)]
31. Fu, W.; Wang, K.; Li, C.; Tan, J. Multi-step short-term wind speed forecasting approach based on multi-scale dominant ingredient chaotic analysis, improved hybrid GWO-SCA optimization and ELM. *Energy Convers. Manag.* **2019**, *187*, 356–377. [[CrossRef](#)]
32. Wei, J.; Liu, H.; Yan, G.; Sun, F. Robotic grasping recognition using multi-modal deep extreme learning machine. *Multidimens. Syst. Signal Process.* **2016**, *28*, 817–833. [[CrossRef](#)]
33. Ding, S.; Zhang, N.; Xu, X.; Guo, L.; Zhang, J. Deep Extreme Learning Machine and Its Application in EEG Classification. *Math. Probl. Eng.* **2015**, *2015*, 129021. [[CrossRef](#)]
34. Wan, C.; Xu, Z.; Pinson, P.; Dong, Z.Y.; Wong, K.P. Probabilistic Forecasting of Wind Power Generation Using Extreme Learning Machine. *IEEE Trans. Power Syst.* **2014**, *29*, 1033–1044. [[CrossRef](#)]
35. Farsi, B.; Amayri, M.; Bouguila, N.; Eicker, U. On Short-Term Load Forecasting Using Machine Learning Techniques and a Novel Parallel Deep LSTM-CNN Approach. *IEEE Access* **2021**, *9*, 31191–31212. [[CrossRef](#)]
36. Huang, G.-B.; Zhu, Q.-Y.; Siew, C.-K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [[CrossRef](#)]
37. Chorowski, J.; Wang, J.; Zurada, J.M. Review and performance comparison of SVM- and ELM-based classifiers. *Neurocomputing* **2014**, *128*, 507–516. [[CrossRef](#)]
38. Menke, J.; Martinez, T.R. In Using permutations instead of student's t distribution for p-values in paired-difference algorithm comparisons. In Proceedings of the 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541), Budapest, Hungary, 25–29 July 2004; Volume 2, pp. 1331–1335.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.