Article

Research on the Short-Term Prediction of Offshore Wind Power Based on Unit Classification

Jinhua Zhang [1], Xin Liu [1,*] and Jie Yan [2]

[1] School of Electrical Engineering, North China University of Water Resources and Electric Power, Zhengzhou 450045, China; zhangjh@ncwu.edu.cn
[2] School of New Energy, North China Electric Power University, Beijing 100096, China; yanjie@ncepu.edu.cn
* Correspondence: 17638567502@163.com

Abstract: The traditional power prediction methods cannot fully take into account the differences and similarities between units. In the face of the complex and changeable sea climate, the strong coupling effect of atmospheric circulation, ocean current movement, and wave fluctuation, the characteristics of wind processes under different incoming currents and different weather are very different, and the spatio-temporal correlation law of offshore wind processes is highly complex, which leads to traditional power prediction not being able to accurately predict the short-term power of offshore wind farms. Therefore, aiming at the characteristics and complexity of offshore wind power, this paper proposes an innovative short-term power prediction method for offshore wind farms based on a Gaussian mixture model (GMM). This method considers the correlation between units according to the characteristics of the measured data of units, and it divides units with high correlation into a category. The Bayesian information criterion (BIC) and contour coefficient method (SC) were used to obtain the optimal number of groups. The average intra-group correlation coefficient (AICC) was used to evaluate the reliability of measurements for the same quantized feature to select the representative units for each classification. Practical examples show that the short-term power prediction accuracy of the model after unit classification is 2.12% and 1.1% higher than that without group processing, and the mean square error and average absolute error of the short-term power prediction accuracy are reduced, respectively, which provides a basis for the optimization of prediction accuracy and economic operation of offshore wind farms.

Keywords: offshore wind farms; Gaussian mixture model; unit classification; short-term power prediction

Citation: Zhang, J.; Liu, X.; Yan, J. Research on the Short-Term Prediction of Offshore Wind Power Based on Unit Classification. *Electronics* **2024**, *13*, 2293. https://doi.org/10.3390/electronics13122293

Academic Editors: Mohamed Benbouzid, Sara Deilami, Jahangir Hossain, Antonio J. Marques Cardoso and Seyedfoad F. Taghizadeh

Received: 26 April 2024
Revised: 7 June 2024
Accepted: 10 June 2024
Published: 12 June 2024

1. Introduction

1.1. Background

China's coastal areas are rich in wind energy resources, which has created excellent conditions for the rapid development of the offshore wind power industry. As a key field of renewable energy expansion, offshore wind power is becoming an important direction of wind energy utilization [1]. In the past few years, China's offshore wind power industry has developed rapidly from the initial stage and made remarkable achievements. According to the statistics of the China Wind Energy Association, by the end of 2023, the cumulative installed capacity of China's offshore wind farms has reached 34.7 million kilowatts, accounting for 46.2% of the global total installed capacity. Faced with the challenges of limited development potential and resource shortage in offshore areas during the 14th Five-Year Plan period, China's offshore wind power, relying on the development experience of Europe, is expanding to the offshore area and entering a new stage of development. Between 2017 and 2023, China's total installed capacity of offshore wind power increased at different rates, but the overall trend was upward.

1.2. Related Works

Offshore wind power, an environmentally friendly and renewable energy solution, plays a vital role in the sustainable development of global energy. However, the research of offshore wind power generation in China is still in the initial stage, and there are few studies on offshore wind power. Domestic researchers have carried out a series of studies on the power prediction of offshore wind power and made certain progress. Most of the studies are based on the model of the power prediction of onshore wind power. Relevant studies show that this approach is feasible [2,3]. At present, different techniques and methods are used for each type of power prediction, including physical methods [4] and statistical methods [5], etc., to meet the prediction challenges on different time scales. With the advancement of technology and the enhancement of data acquisition capabilities, the accuracy and reliability of offshore wind power prediction are constantly improving, providing support for the efficient use of wind energy and the stable operation of the grid.

Physical methods represent one of the classical techniques in the field of wind power prediction, which mainly uses historical wind speed and power data to predict future generation performance. Physical methods rely on factors such as numerical weather forecasting, terrain, and elevation to build prediction models. Statistical methods are increasingly used in wind power prediction, especially when dealing with complex nonlinear relationships and large datasets. Commonly used machine-learning algorithms are mainly concentrated in auto-regression and moving average model (ARMA), artificial neural network (ANN), support vector machines (SVM), and so on. These algorithms can learn from historical data the influence of various factors such as wind speed, wind direction, temperature, and air pressure on the wind power output, thereby improving the accuracy of prediction. Huang and Qin [6] proposes a short-term offshore wind power prediction method that considers dynamic time-delay effects to intuitively capture power prediction information. Based on the nonlinear coupling relationship, dynamic sliding windows matching different mean periods are introduced. Then, the dynamic delay time is calculated and multiple delay relationships between variables are defined. Finally, the Elman network is used to predict the short-term offshore wind power. Aiming at the problems of strong randomness and time correlation in offshore wind power prediction, Wang et al. [7] proposed a principal component analysis (PCA), sparrow algorithm (SSA), variational mode decomposition (VMD), and bidirectional long short-term memory neural network (Bi LSTM), and finally verified the results by simulation experiments. The results showed that the proposed model effectively improved the prediction accuracy. The validity of the prediction model is verified. In reference [8], An et al. carried out work based on the spatio-temporal correlation features between wind turbine outputs in which a diffused convolutional neural network (DCNN) is embedded into a gated recurrent unit (GRU) for feature extraction of the spatio-temporal correlation of wind turbine outputs. Combined with graph structure learning, a sequence-to-sequence model for the ultra-short-term power prediction of large offshore wind farms is proposed. The actual case simulation shows that the model has a good forecasting performance in the ultra-short-term power prediction of large offshore wind farms. Sun et al. [9] proposed a CNN-LSTM-AM network for predicting the power of offshore wind turbines using signals from multiple sensors. A variable control comparison was performed to complete the sensitivity analysis of the sensors to determine the most suitable sensor group for power prediction. Compared with existing deep-learning algorithms, the model achieved a maximum 13.77% improvement in power prediction. Zhang et al. [10] proposed a GAT-LSTM short-term wind power prediction model, which adopts a random sampling algorithm to optimize hyperparameters and improve the learning rate and performance of the model. The results show that the proposed model has a higher prediction accuracy than other traditional models and is reasonably interpretable in terms of time and space.

In the operation and management of offshore wind farms, effectively categorizing wind turbine units is a crucial task. It helps us to better understand the operating characteristics, maintenance requirements, and power output behavior of each unit. The cluster method is a widely used data grouping technique. In short, the goal of clustering is to divide the dataset into multiple categories according to some criteria (such as the closest distance between elements, the farthest distance, or the average distance), so that the characteristics of data points within the same category are as consistent as possible, while the data points between different categories show greater differences. For example, K-means [11], density clustering [12–14], hierarchical clustering [15,16], spectral clustering [17–19], and incremental clustering [20–22] can effectively classify wind turbines to optimize operation and maintenance strategies and improve energy output efficiency. ST-TRACLUS was proposed in reference [23], which is a novel spatio-temporal clustering algorithm, which enhances the DBSCAN framework through spatial and temporal analysis to identify similarities in trajectory data. They showed a better performance than traditional methods such as TRACLUS and ST-OPTICS. Pu et al. [24] proposed KDE-AHIAC, an improved HIAC clustering algorithm based on kernel density estimation (KDE), to solve the problem of small datasets and improve the accuracy of threshold determination. By automatically adjusting bandwidth and smoothing density curves, the clustering accuracy is significantly improved and the performance is excellent on a variety of datasets. Wang et al. [25] proposed hybrid sand cat swarm optimization and improved fuzzy C-means clustering algorithm to determine power deviation and other feature data related to icing detection. Real sensor data from the monitoring and data acquisition system were used to validate the proposed icing risk assessment method (considering WPP). Hou et al. [26] uses density-based noise applied spatial clustering (DBSCAN) and the enhanced prey optimization algorithm (ENHPO) to design a new hybrid power prediction model for wind turbine clusters. The simulation results show that compared with other clustering methods such as fuzzy C-means, balanced iterative reduction, hierarchical clustering, K-means clustering, and density peak clustering, the prediction accuracy and efficiency of this method are improved.

To sum up, significant progress has been made in the field of offshore wind power prediction. By drawing on land-based wind power models, applying physical and statistical methods, and developing new algorithms, researchers have achieved important results in improving forecast accuracy and reliability. These contributions have laid a solid foundation for the future research and practical application of offshore wind power, and promoted the technological progress and industrial development of offshore wind power. To sum up, however, most of these studies on offshore wind power use different methods to decompose time series or make power predictions based on important features in time series, and rarely consider the correlation and difference between units in wind farms. For large wind farms, wind turbines are widely distributed, and climatic factors such as wind speed, output power and wind direction are different in different locations. However, traditional methods fail to accurately capture these individual characteristics, resulting in relatively inaccurate prediction results. Therefore, in the power prediction of wind turbines, the correlation and difference between wind turbines are considered, and the wind turbines with a high correlation between wind speed and output power are grouped into a class by the Gaussian mixture model clustering algorithm. The Bayesian information criterion and contour coefficient were used to judge the optimal number, and the CNN-LSTM neural network was used to establish power prediction sub-models for each category. The example verified that the clustering algorithm had a good effect on improving the prediction accuracy of offshore wind power. It also played a key supporting role in the operation control, safety and stability guarantee, and market strategy formulation of the power system. The block diagram of the research content of this paper is shown in Figure 1.

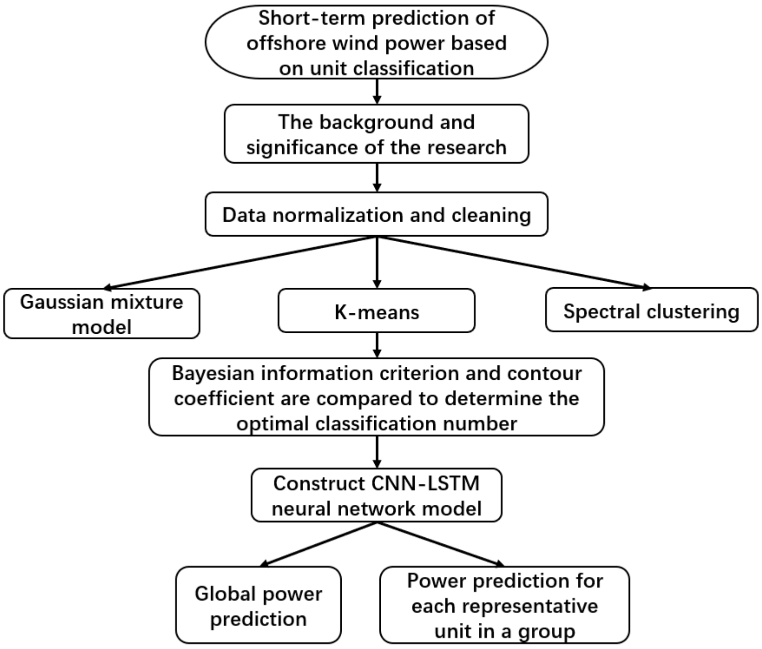

Short-term prediction of offshore wind power based on unit classification
The background and significance of the research
Data normalization and cleaning
Gaussian mixture model
K-means
Spectral clustering
Bayesian information criterion and contour coefficient are compared to determine the optimal classification number
Construct CNN-LSTM neural network model
Global power prediction
Power prediction for each representative unit in a group

Figure 1. Block diagram of main research contents.

2. Materials and Methods

2.1. Wind Turbine Grouping Model

2.1.1. Gaussian Mixture Model

The Gaussian mixture model (GMM) is a probabilistic machine learning technique [27], which is composed of K individual Gaussian models. When there are multiple Gaussian distributions and these distributions obey the same population distribution, they can be grouped into the same category. After classification, the model uses the expectation maximization (EM) algorithm to estimate the parameters. The EM algorithm evaluates the matching degree between the probability of model prediction and the probability of observation data, and it brings the prediction probability of model closer to the actual observation probability by adjusting the model parameters. This adjustment and evaluation process will be repeated several times until the probability predicted by the model is close enough to the probability actually observed, at which point the algorithm stops iterating and the model training is completed. The total Gaussian distribution can be expressed as (1):

$$\sum_{k=1}^{k} \alpha_k N\left(\mu_k, \sigma_k^2\right) = P(y|\theta) \tag{1}$$

In the framework of the Gaussian mixture model, each independent Gaussian component is defined with its own mean μ and standard deviation σ parameters, as well as a specific weight parameter. These weight parameters are all positive numbers, and their sum must be ensured to be 1, in order to ensure that the probability density value of the model as a whole is kept within a reasonable range. In other words, the integral sum of the probability density function of each independent Gaussian component of the model over the entire input space should be equal to 1. In this context, y represents the observed data point, while θ represents all the parameters of the model, including the relevant parameters of all the Gaussian distributions. Each Gaussian distribution can be expressed as follows:

$$N\left(\mu_k, \sigma_k^2\right) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(y-\mu_k)^2}{2\sigma_k^2}} \tag{2}$$

In the process of wind turbine grouping, the half-year measured wind speed mean $V_{i,mean}$ and standard deviation $V_{i,std}$ and the measured power mean $P_{i,mean}$ and standard deviation $P_{i,std}$ of 134 units are used together as the representation of a single unit.

The specific steps are as follows:

1. Initialize the k multivariate Gaussian distribution of parameters μ_j and σ_j, $(j = 1, 2, \cdots, k)$; since the input data are a 4-dimensional array of length m, each element has its own corresponding matrix.
2. After initializing the parameters, calculate the probability density γ_{ij} of each sample point $y_i (i = 1, 2, 3, \cdots, m)$ belonging to the j Gaussian distribution. The formula is as follows:
$$\gamma_{ij} = P(y_i|z_i = j) = \frac{1}{(2\pi)^{\frac{d}{2}} |\sigma_j|^{\frac{1}{2}}} e^{-\frac{1}{2}(y_i - \mu_j)^T \sigma_j{}^{-1} (y_i - \mu_j)} \tag{3}$$
where P is the probability function, z_i is the class to which y_i belongs, and d is the dimension of y_i.
3. Obtain the updated values μ'_j and σ'_j of parameters μ_j and σ_j of each Gaussian distribution according to Formulas (4) and (5):
$$\mu'_j = \frac{\sum\limits_{i=1}^{m} \gamma_{ij} y_i}{\sum\limits_{i=1}^{m} \gamma_{ij}} \tag{4}$$
$$\sigma'_j = \frac{\sum\limits_{i=1}^{m} \gamma_{ij} \left(y_i - \mu'_j \right) \left(y_i - \mu_j \right)^T}{\sum\limits_{i=1}^{m} \gamma_{ij}} \tag{5}$$
4. Repeat the preceding steps until the parameters of each Gaussian component stabilize and converge.
5. After the parameters converge, traverse each sample point and divide it into the category with the greatest probability.

2.1.2. The K-Means Model

K-means is a simple and efficient unsupervised learning algorithm [11] known for its simple structure and easy operation. It is widely used in clustering analysis, and is favored by researchers and data scientists because of its ability to quickly process large datasets.

In the process of wind turbine grouping, the half-year measured wind speed mean $V_{i,mean}$ and standard deviation $V_{i,std}$ and the measured power mean $P_{i,mean}$ and standard deviation $P_{i,std}$ of 134 units are used together as the representation of a single unit. Given dataset $X(t) = (x_1, x_2 \cdots x_n)$ containing n samples, where each sample is a 4-dimensional data vector, the objective function of K-means algorithm modeling can be obtained as follows:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left\| X_i^{(j)} - c_j \right\|^2 \tag{6}$$

In the formula, $\left\| X_i^{(j)} - c_j \right\|^2$ represents the distance measurement from cluster sample point x_i^j to cluster center c_j, which is the similarity measurement process of the original data. For the data characteristics of wind farms, Euclidean distance is selected as the similarity measurement method. The modeling steps of the K-means algorithm are as follows:

1. Enter the number of group categories k and to be clustered n, $(k \leq n)$.
2. Randomly select samples from the dataset in numbers equivalent to the number of clusters as the initial clustering centers.

3. Calculate the distance between each additional sample object and the selected cluster centers of each class, and assign it to the nearest class.
4. Calculate the average value of each class of data objects obtained from the previous step, and use it as the new clustering center.
5. Repeat Steps 3 and 4 until the centers of the clusters no longer change. With the number of classifications set at C, this process will identify C cluster centers. Following these steps results in C cluster classifications, thereby completing the optimal clustering classification for the K-means algorithm.

2.1.3. Spectral Clustering Model

Spectral clustering is a clustering method based on graph theory [17–19], which clusters data points by analyzing the spectrum (eigenvalues) of the graph formed by the data points. Compared to traditional clustering methods such as K-means, spectral clustering is more adept at uncovering the global structure of data and can handle the clustering of non-convex sets and data with irregular boundaries. It is insensitive to the size and shape of the dataset, making it particularly suitable for discovering datasets with complex structures. The fundamental idea is to utilize the spectrum of the similarity matrix to capture the nonlinear low-dimensional manifold structure of the data, thereby achieving the clustering of data points.

The modeling process for the spectral clustering algorithm is outlined as follows: In the process of wind turbine grouping, the half-year measured wind speed mean $V_{i,mean}$ and standard deviation $V_{i,std}$ and the measured power mean $P_{i,mean}$ and standard deviation $P_{i,std}$ of 134 units are used together as the representation of a single unit. For a given dataset $X(t) = (x_1, x_2 \cdots x_n)$ containing n samples, where each sample is a 4-dimensional data vector, these vectors are clustered into k classes. The specific steps are as follows:

1. Selecting Euclidean distance as the similarity measure, construct a similarity matrix $S \in R^{n\times n}$ based on the similarity between data points. This matrix is symmetric, where S_{ij} represents the similarity between the i-th and j-th data points.
2. Construct the Laplacian matrix L, $L = D - S$, which is formed by the degree matrix D and the similarity matrix S. The degree matrix is a diagonal matrix, where the elements on the diagonal are equal to the sum of the corresponding rows in the similarity matrix.
3. Carry out the eigenvalue decomposition of the Laplacian matrix in Step 2 to obtain the eigenvalues and corresponding eigenvectors. Select the eigenvectors corresponding to k eigenvalues with relatively small eigenvalues as the input of clustering, where k is the number of predefined clusters.
4. Arrange the selected k eigenvectors column-wise to form a new matrix $U \in R^{n\times k}$, where each row represents the coordinates of the original data in the new low-dimensional space. Normalize each row of matrix U so that the length of each point's feature vector is 1.
5. Use the selected eigenvectors as the new data representation and apply the K-means clustering algorithm to cluster them.

2.1.4. Bayesian Information Criteria

We utilize the Bayesian information criterion (BIC) [28] model selection theory to probabilistically estimate the number of classifications for the units. This theory obtains the optimal number of clusters through an approximation method, defined by the following equation.

$$C_{BIC} = n_p ln(m) - 2ln(L) \tag{7}$$

In the formula, C_{BIC} represents the BIC value; n_p is the number of hyperparameters; L is the maximum value of the likelihood function of the estimated model.

Assuming the model's errors or perturbations follow a normal distribution, then the Bayesian information criterion (BIC) can be expressed as follows:

$$C_{BIC} = mln\left(\frac{S_{RSS}}{m}\right) + n_p ln(m) \tag{8}$$

In the formula, S_{RSS} represents the sum of squared residuals of the estimated model. C_{BIC} is an increasing function of S_{RSS} and n_p, meaning that the introduction of residuals and unknown parameters will cause C_{BIC} to increase. Therefore, when determining the optimal number of groupings for wind turbine units, models with lower BIC values are preferred.

2.1.5. Silhouette Coefficient

The silhouette coefficient (SC) is an indicator used to measure the cohesion within clusters and the separation between clusters, and it is widely applied in the assessment of clustering validity [29]. The silhouette value and the definition of SC are given by the following formulas.

$$s(x_i) = \frac{b(x_i) - a(x_i)}{max(a(x_i), b(x_i))} \tag{9}$$

$$C_{SC} = \frac{1}{m}\sum_{i=1}^{m} s(x_i) \tag{10}$$

In the formula, $a(x_i)$ is the average distance between the sample point x_i and all other points within the same cluster; $b(x_i)$ is the minimum of the average distances between the sample point x_i and the sample points in all other clusters; C_{SC} represents the SC value. The silhouette value $s(x_i)$ ranges between -1 and 1. Specifically, a silhouette coefficient close to -1 indicates a clustering result that is least satisfactory; conversely, a silhouette coefficient approaching 1 signifies an excellent clustering effect. Therefore, when assessing the optimal number of groupings for wind turbine units, models with higher silhouette coefficient values are considered more preferable.

2.2. Short-Term Power Prediction Mode

2.2.1. Convolutional Neural Network

Convolutional neural network (CNN) is a feedforward neural network with a convolutional structure, composed of an input layer, convolutional layers, pooling layers, and fully connected layers. It has widespread applications in fields such as image recognition, natural language processing, and remote sensing [30,31]. Compared to traditional multilayer neural networks, CNNs introduce convolutional layers and pooling layers before the fully connected layers, which allows for more effective feature extraction and learning. The formula for feature extraction in time-series data using one-dimensional convolution is as follows:

$$Y = \sigma(W * T + b) \tag{11}$$

where Y is the extracted feature; σ is the sigmoid activation function; W is the weight matrix; T is the time series; b is the bias vector.

2.2.2. Long Short-Term Memory Neural Network

The long short-term memory neural network (LSTM) is an efficient recurrent neural network (RNN) architecture, which overcomes the problems of gradient disappearance and gradient explosion when RNN networks deal with long-term dependence problems [32,33]. The core concept of LSTM is the cell state and "gate" structure, and each LSTM unit consists of a cell state, a forgetting gate, an input gate, and an output gate. The LSTM structure is shown in Figure 2.

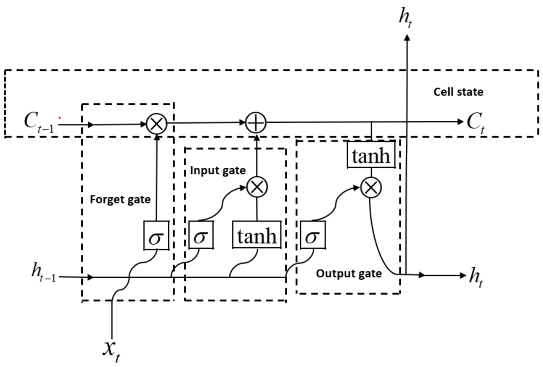


Figure 2. Chain structure of the LSTM hidden layer.

The input time series is set as $X = (X_1, X_2, \cdots X_n)$, and the two output series after LSTM mapping are $h = (h_1, h_2, \cdots h_n)$ and $y = (y_1, y_2, \cdots y_n)$, respectively. The forgotten gate in the LSTM unit determines what information should be discarded or retained, and its formula is as follows:

$$f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f) \tag{12}$$

where σ represents the sigmoid function, and W and b are the parameters of the training network. By reading the previous output h_{t-1} and the current output x_t, and then processing by the sigmoid function, the output f_t is obtained. The output value is between 0 and 1, and it is deleted when it is close to 0 and retained when it is close to 1.

The input gate determines what new information is put into the cell state and consists of two steps; its formula is as follows:

$$i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i) \tag{13}$$

$$\tilde{C}_t = tanh(W_C * [h_{t-1}, x_t] + b_C) \tag{14}$$

where $\tilde{C}_t$ represents the new vector created by the layer tanh. The input gate yields data processed separately by the sigmoid and tanh functions, which are combined into the cell state.

The cell state is to update C_{t-1} to C_t, and the formula is as follows:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{15}$$

The output gate determines what value needs to be output in the end, and the formula is as follows:

$$o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o) \tag{16}$$

$$h_t = o_t * tanh(C_t) \tag{17}$$

It can be seen from the formula that the input data processed by sigmoid function are multiplied by the cell state data processed by the tanh function, and the final data obtained are the output part.

The CNN-LSTM short-term wind power prediction model consists of two parts: The first part uses a convolutional neural network to extract data features from the original time series and form the data feature information sequence; the second part predicts the feature information sequence extracted from the CNN by LSTM. The training process of the whole model is divided into two stages: forward propagation and back propagation. In the forward propagation stage, the error of the target loss function is mainly calculated, while in the back propagation stage, the adaptive moment estimation (ADAM) algorithm is used to optimize the network parameters.

$$L = \frac{1}{F}\sum_{T=1}^{F}\left(y_T - \hat{y}_T\right) \tag{18}$$

In the above formula, y_T is the real value of wind power at moment T , and $\hat{y}_T$ is the predicted value of wind power at moment T , where F is the number of samples in the training set sample set.

3. Case Analysis

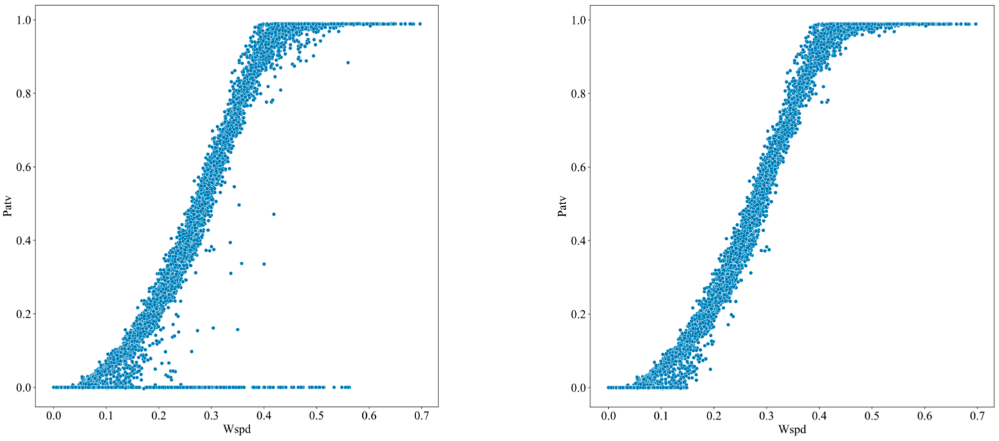
3.1. Wind Farm Introduction

Since this forecast is a short-term wind power forecast, the forecast model usually has a strong adaptability and can quickly adjust the parameters and structure according to recent data. On the contrary, too much historical data may make the model overfit some historical patterns that are no longer applicable, affecting the prediction accuracy. Therefore, the experimental data are used in the actual wind speed and power generation records of an offshore wind farm in China in 2021, so the time resolution is 5 min. The coverage period is from 0:00 1 January to 24:00 30 June. The total number of wind turbines involved is 134.

3.2. Cleaning of Measured Data of Wind Turbine

This paper focuses on data cleaning of actual measurement data in the wind power sector, emphasizing that the raw data recorded by SCADA systems often contain abnormal data due to power limitations, unit defects, problems with recording instruments, or communication failures.

In actual measurements based on individual wind turbines, scatter plots of wind speed and power often contain anomalous data such as noise points and zero power accumulation points (i.e., data points above the starting wind speed but with zero power). Using DB-SCAN [34] to process these anomalies can effectively identify and remove them, resulting in a cleaned up, more accurate scatter map of the wind speed vs. power relationship. Figure 3 below shows the results before and after processing.



(**a**) (**b**)

Figure 3. (**a**) Wind power scatter diagram (before processing); (**b**) Wind power scatter diagram (after processing).

3.3. Wind Turbine Grouping Scheme

According to the constructed wind turbine grouping model, these 134 units were grouped and studied. Among the many possible factors, the output power is the most important performance index of the wind turbine, which directly reflects the generation efficiency and ability of the wind turbine. Wind speed is one of the main environmental factors affecting the output power, and it is an important parameter to evaluate the performance and efficiency of the unit. In addition, the wind speed and output are the conventional monitoring record data of the wind farm, which have a high frequency and high accuracy, and are easy to obtain and use. Therefore, the wind speed and output are selected as the influencing factors of the evaluation. Therefore, the two most direct parameters, wind speed and generation power, are selected as the criteria to evaluate the performance of a single unit. Specifically, the performance of Unit $i(i = 1, 2, 3, \cdots, 134)$ over six months is described using four parameters: the mean wind speed $V_{i,mean}$, its standard deviation $V_{i,std}$, the mean power output $P_{i,mean}$, and its standard deviation $P_{i,std}$.

Therefore, the input data for the wind turbine grouping model consist of a 134 × 4-dimensional array. When the number of classifications is set between 3 and 7, the BIC and SC indices for different group numbers are calculated using the GMM clustering method, K-means, and spectral clustering methods. The evaluation of the rationality of different models as the number of groups changes is illustrated in Figures 4 and 5.

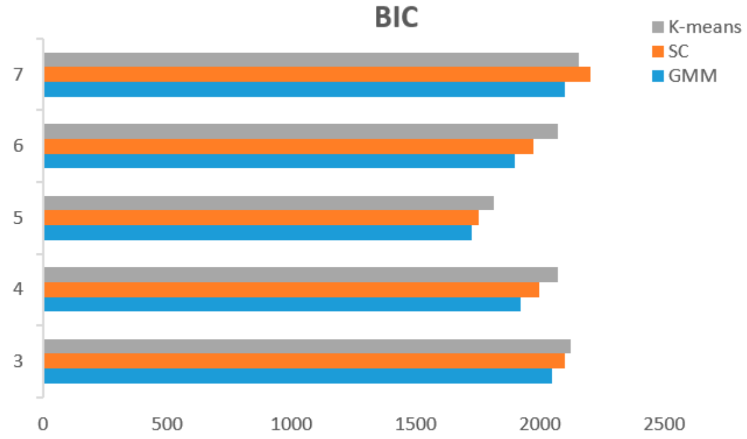


Figure 4. BIC values of each clustering model.

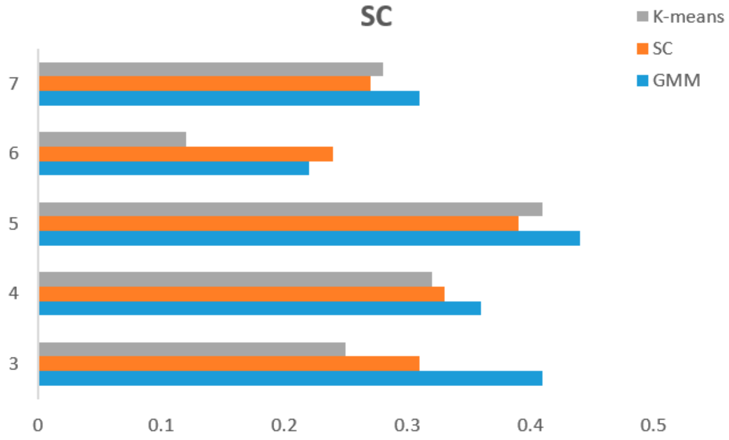


Figure 5. SC values of each clustering model.

As can be seen from Figures 4 and 5, with the increase in the number of groups, the BIC value decreases first and then increases, and the lowest point appears when the number of groups is five. When judging the optimal number of groups of wind turbines, the model with a low BIC value is the best. In general, the maximum SC value is reached when the number of packets is five, which is closer to one. According to the criteria that the lower the BIC value, the better, and the closer the SC value to one, the higher the better, when the number of groups is five, the GMM clustering algorithm has the lowest BIC value and the highest SC value. Therefore, it is concluded that the wind farm unit classification is best when divided into five categories. Compared with K-means and spectral clustering, GMM clustering has more advantages.

3.4. Selection of Representative Units

The average intra-cluster correlation coefficient (AICC) is a statistical index used to measure the similarity among members in a group [35]. The AICC ranges from −1 (perfect negative correlation) to +1 (perfect positive correlation). A value close to +1 means that members within the group are very similar in measured attributes, while a value close to −1 indicates large internal differences. If the AICC value is close to 0, it may mean that there is no significant correlation between the members of the group on that attribute. When performing classification clustering, those units with the highest AICC values are considered representative and can be used to build power prediction models for each subgroup.

$$I_{AICC,P} = \frac{1}{n_l} \sum_{q \in C_l} \frac{Cov(X_P, X_q)}{\sqrt{Var(X_P) * Var(X_q)}} \tag{19}$$

In the formula, C_l and n_l, respectively, represent the set of units in group l and the number of units in the group, where $l = 1, 2, \cdots, g$; p and q represent any two units in Group l, $(p, q \in C_l)$; X_p and X_q are the measured power time series of Units p and q, respectively. $Cov(X_p, X_q)$ represents the covariance of X_p and X_q; $Var(X_p)$ and $Var(X_q)$ are the variances of X_p and X_q, respectively.

According to Section 3.3, the clustering effect is the best when the number of clusters is five. The unit with the highest AICC value in each class is taken as the representative unit. According to Table 1, in the GMM cluster, Units 1, 8, 40, 83, and 113 have the highest AICC value in each category and the average AICC value is 0.774. The average AICC value of the units with the highest AICC in each category of K-means clustering is 0.724. The average AICC value of the units with the highest AICC in each category of spectral clustering is 0.716. It can be seen from the data display that the average AICC value of GMM clustering is higher than that of K-means clustering and spectral clustering. Therefore, using the results from GMM clustering, Units 1, 8, 40, 83, and 113 are selected as representative units to construct a short-term power prediction sub-model for offshore wind farms. For the overall power prediction model, Unit 88, which has the highest AICC value of 0.71, is chosen as the representative unit.

Table 1. AICC values of each representative unit.

Number	GMM		K-Means		SC	
	Representative Unit	AICC	Representative Unit	AICC	Representative Unit	AICC
1	1	0.79	5	0.68	5	0.72
2	8	0.83	38	0.73	27	0.78
3	40	0.75	61	0.64	56	0.69
4	83	0.64	96	0.77	91	0.64
5	113	0.86	115	0.80	117	0.75
Average		0.774		0.724		0.716

3.5. Short-Term Forecast Evaluation Index and Analysis of Examples

In the field of error analysis, common evaluation metrics include the root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R-squared). To accurately assess the estimation performance of the model, this paper will employ RMSE and MAE as the primary tools. The formulas for calculating RMSE and MAE are as follows:

$$MAE = \frac{1}{NP} \sum_{i=1}^{N} \left| P_i - \hat{P}_i \right| \tag{20}$$

$$RMSE = \frac{1}{P}\sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - \hat{p}_i)^2} \tag{21}$$

In the above formula, N is the number of samples; P is the single-unit capacity of the wind farm; P_i is the actual value of wind power; $\hat{P}_i$ is the predicted value of wind power.

To validate the effectiveness of the CNN-LSTM short-term wind power forecasting model, experimental data from a certain offshore wind farm in China in 2021 were selected for analysis. The wind farm consists of 134 units, each with a capacity of 2000 kW, totaling 268 MW. The sampling frequency of the wind farm is 5 min, resulting in 288 data points sampled per day, spanning from 1 April 2021, to 15 April 2021, totaling 4320 points (with 4176 datasets used as training samples and 144 sets used for prediction verification). The data include information on wind turbine units for short-term power prediction 12 h in advance. Clustering results were obtained according to Table 1, and for unit classification prediction, representative Units 1, 8, 40, 83, and 113 with the highest AICC values were selected for each category. For the overall prediction, Unit 88 (0.71) with the highest overall AICC value was selected as the representative unit to establish the prediction model.

In the CNN-LSTM model constructed this time, CNN is responsible for extracting original data features, and the LSTM network is responsible for wind power prediction. The CNN layer model uses two layers of convolution kernel and one layer of pooling to carry out the feature extraction of data series. The first layer has 64 convolution nuclei, the size of which is 1 × 4; the second layer has 32 convolution nuclei, the size is 1 × 3, and the moving step is 2. The size of the pooling layer is 1 × 2. We set the LSTM batch size to 256, learning rate to 0.001, and epoch to 300. In order to show the prediction effect more directly, a total of 4320 units of Unit 88 with the highest AICC value in the whole wind farm were selected for 15 days in April (in which the sampling frequency was 5 min, 4176 sets of data were used as training samples, and 144 sets of data were used as prediction verification samples) for the overall wind power prediction analysis, in order to clearly show the performance of the prediction algorithm. The forecast results are shown in Figure 6.

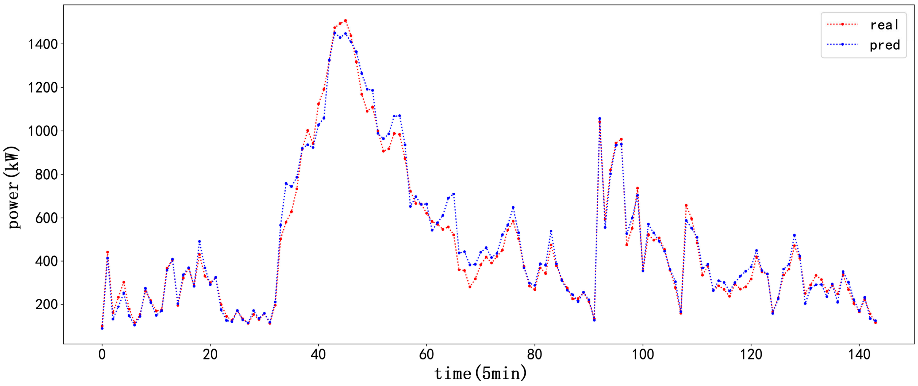


Figure 6. Short-term variation lines of measured and modeled power of wind farms.

As shown in Figure 6, real represents the true value and pred represents the predicted value; there are minor deviations between the predicted values and the actual values at the peaks and troughs, but the overall prediction trend is consistent. The overall root mean square error is 10.86%, and the mean absolute error is 6.69%. Following the clustering of groups, power predictions were conducted for Units 1, 8, 40, 83, and 113, selecting data from April for CNN-LSTM forecasting. Of these, 4176 datasets were used as training samples, and 144 datasets were used as prediction control samples. The power predictions are illustrated in Figures 7–11.

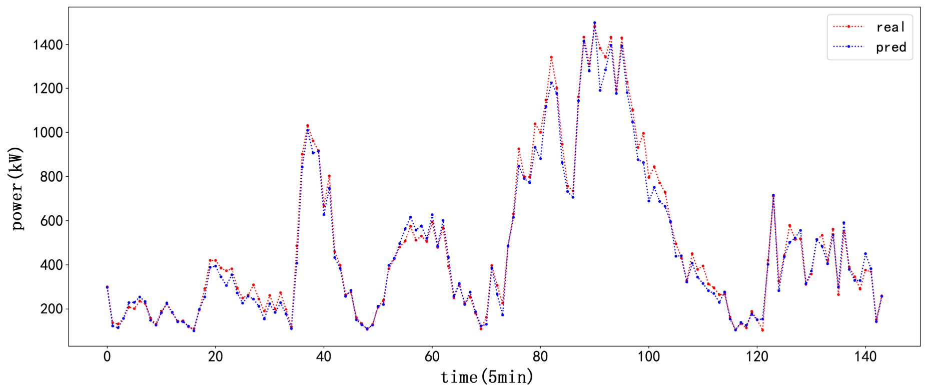


Figure 7. The short-term variation in the measured and predicted power of Unit 1.

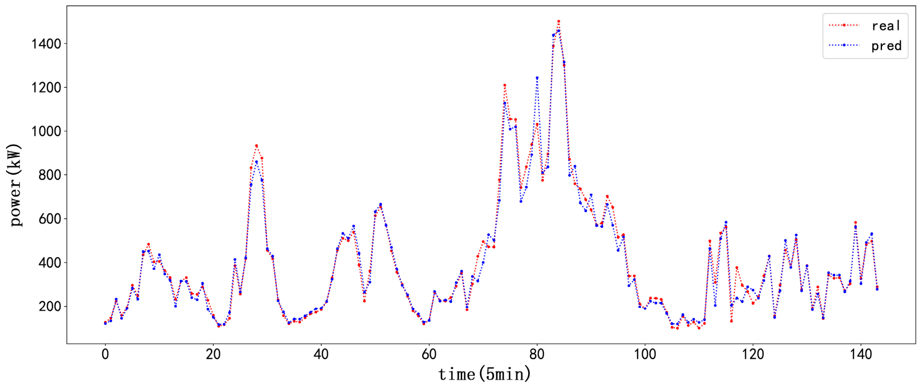


Figure 8. The short-term variation in the measured and predicted power of Unit 8.

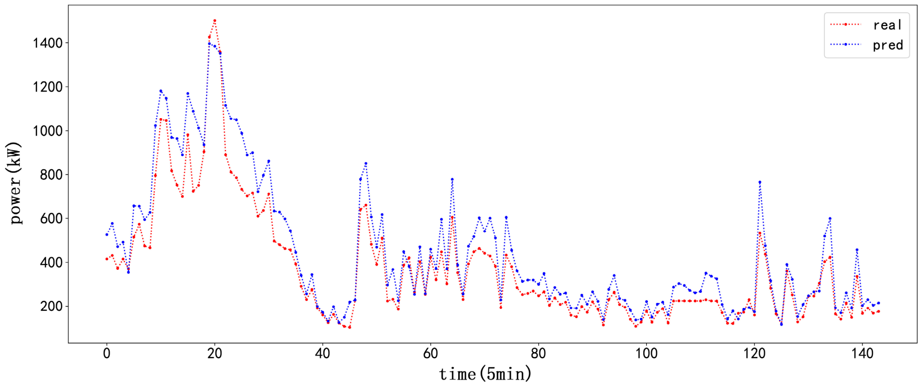


Figure 9. The short-term variation in the measured and predicted power of Unit 40.

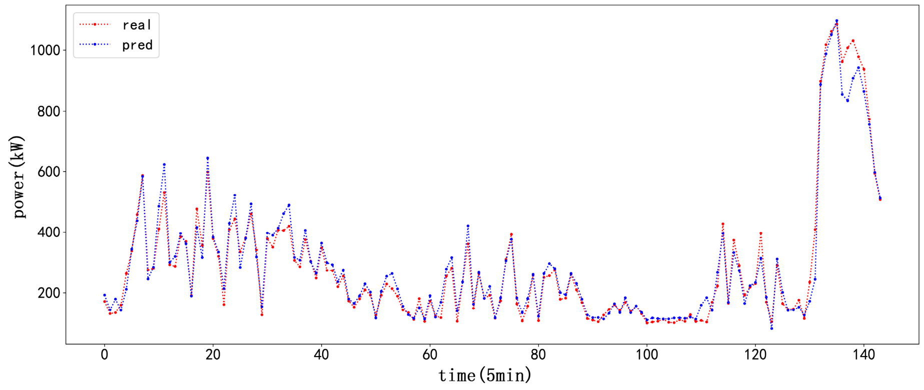


Figure 10. The short-term variation in the measured and predicted power of Unit 83.

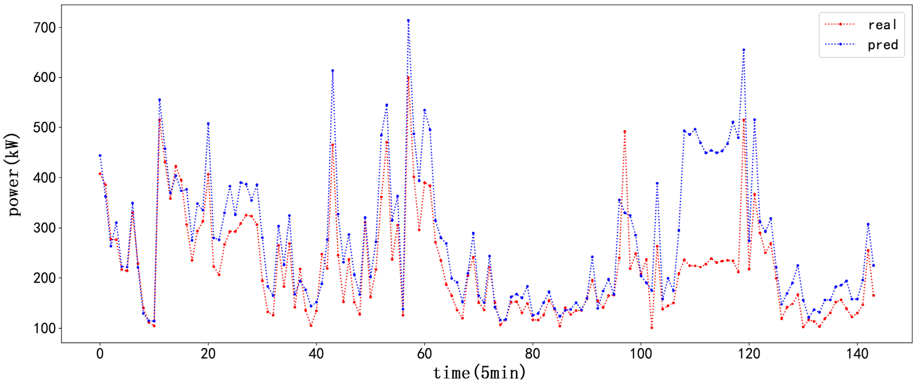


Figure 11. The short-term variation in the measured and predicted power of Unit 113.

Power prediction sub-models were established for the above five units, respectively, to obtain the short-time change curves of the measured units and the model power prediction corresponding to the five units, respectively. It can be seen from the line charts that the trend change in the model prediction sequence diagram for Units 1, 8, and 83 closely follows the true value, and the root mean square error (RMSE) is 4.58% and 7.87%, respectively. The mean absolute error (MAE) is 2.64%, 4.77%, and 4.59%, respectively. For Unit 113 and Unit 40, the prediction effect is slightly weaker than the other three units, the root mean square error (RMSE) is 13.68% and 11.27%, and the mean absolute error (MAE) is 9.85% and 8.12%, respectively. The root mean square error and average absolute error of most representative units are lower than the root mean square error and average absolute error of the overall prediction.

According to the weights of each representative unit, the weighted power forecast diagram of each classified representative unit and the comparison diagram of the overall forecast and the actual value are obtained through calculation.

As can be seen from Figure 12, the changes in the power prediction series curve and overall prediction series curve after grouping offshore wind farms closely follow the true value, and the trend and coincidence degree of classification prediction are higher than the overall prediction; the root mean square error (RMSE) and average absolute error (MAE) are shown in Table 2 below.

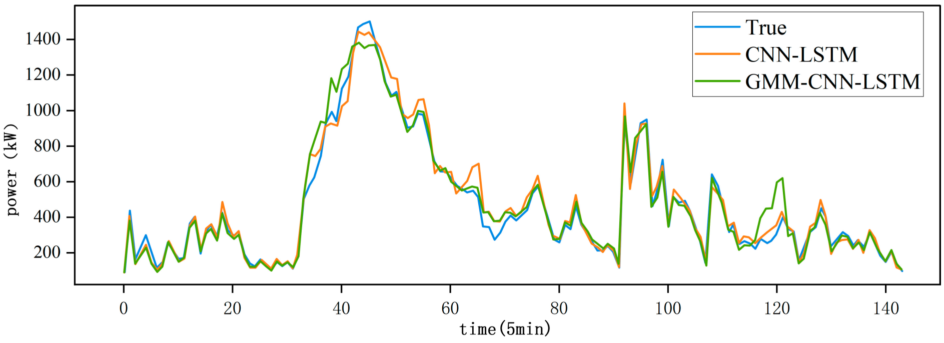


Figure 12. Breakdown lines of short-term variation between the true value of wind farm power and the predicted value of the model.

Table 2. RMSE and MAE values are predicted for different types of power.

Prediction Type	RMSE	MAE
Global forecast	10.86%	6.69%
Classification prediction	8.74%	5.59%

Table 2 above shows the prediction results of the overall prediction and classification prediction. It can be seen from the table that the error of classification prediction is smaller than that of the overall prediction regardless of the RMSE value or MAE value. This verifies that in the face of complex climate change at sea, CNN-LSTM classification prediction refined by a Gaussian mixture model (GMM) space can adapt to such rapid climate change at sea better than that predicted by a single model. The application of this method not only improves the accuracy of prediction, but also provides a new perspective and technical path for offshore wind farm power prediction research.

4. Conclusions and Future Work

The complexity and variability of maritime climate, coupled with atmospheric circulation, ocean current movements, and wave fluctuations, make the accurate prediction of wind power generation crucial. This is significant for ensuring the stability of the electrical grid and enhancing the efficiency of wind energy utilization. Unlike models that predict power output for a single wind farm, this paper considers the interrelations among units, grouping highly correlated units together. On this basis, different wind power prediction models are constructed. This predictive method, which accounts for the similarities among units, better captures the factors affecting power output and adapts more effectively to rapid environmental changes and variations in unit performance at sea. The case study confirms that the proposed method improves the accuracy of short-term power predictions for offshore wind farms, as indicated by the results:

1. With the change in cluster number, the BIC value and SC value of Gaussian mixture model clustering, K-means clustering, and spectral clustering have the best clustering effect when the cluster number is five.
2. The AICC index is used as the basis for the selection of representative points in the whole field and each sub-wind turbine cluster. The AICC results of the GMM grouping model are significantly higher than the K-means and spectral clustering grouping models, the power correlation of the units in the group is higher, and the representation of the prediction model is stronger.
3. According to the power prediction model based on the CNN-LSTM neural network, a better prediction accuracy can be obtained by considering the group of wind turbines. Compared with the overall prediction error, the root mean square error and average absolute error of classification prediction are reduced by 2.12% and 1.1%, respectively.

In unit clustering and power prediction, this paper only considers the effects of wind speed, output power, and wind direction, but it does not consider the effects of offshore fan torque, blade angle, and temperature. In the future work, these factors can be added and a more refined prediction model can be adopted to conduct simulation experiments.

Author Contributions: Conceptualization, J.Z.; methodology, X.L.; software, X.L.; validation, J.Z. and J.Y.; formal analysis, X.L.; investigation, J.Y.; resources, J.Z.; writing—original draft, X.L.; visualization, J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Support Plan for Scientific and Technological Innovation Team of Colleges and Universities in Henan Province (No. 22IRTSTHN011) and the Outstanding Youth Fund of Henan Provincial Natural Science Foundation (No. 242300421054).

Data Availability Statement: The data are available and explained in this article, and the data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

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

References

1. Song, D.; Fan, T.; Li, Q.; Joo, Y.H. Advances in Offshore Wind. *J. Mar. Sci. Eng.* **2024**, *12*, 359. [CrossRef]
2. Yu, M.; Zhang, Z.; Li, X.; Yu, J.; Gao, J.; Liu, Z.; You, B.; Zheng, X.; Yu, R. Superposition Graph Neural Network for offshore wind power prediction. *Future Gener. Comput. Syst.* **2020**, *113*, 145–157. [CrossRef]

3. Li, S.; Huang, L.-L.; Liu, Y.; Zhang, M.-Y. Modeling of Ultra-Short Term Offshore Wind Power Prediction Based on Condition-Assessment of Wind Turbines. *Energies* **2021**, *14*, 891. [CrossRef]
4. Wang, Y.; Liu, Y.; Li, L.; Infield, D.; Han, S. Short-Term Wind Power Forecasting Based on Clustering Pre-Calculated CFD Method. *Energies* **2018**, *11*, 854. [CrossRef]
5. Fang, S.; Chiang, H.-D. A High-Accuracy Wind Power Forecasting Model. *IEEE Trans. Power Syst.* **2017**, *32*, 1589–1590. [CrossRef]
6. Huang, J.; Qin, R. Elman neural network considering dynamic time delay estimation for short-term forecasting of offshore wind power. *Appl. Energy* **2024**, *358*, 122671. [CrossRef]
7. Wang, Z.; Ying, Y.; Kou, L.; Ke, W.; Wan, J.; Yu, Z.; Liu, H.; Zhang, F. Ultra-Short-Term Offshore Wind Power Prediction Based on PCA-SSA-VMD and Bi LSTM. *Sensors* **2024**, *24*, 444. [CrossRef]
8. An, Y.; Zhang, Y.; Lin, J.; Yi, Y.; Fan, W.; Cai, Z. Ultra-Short-Term Power Prediction of Large Offshore Wind Farms Based on Spatiotemporal Adaptation of Wind Turbines. *Processes* **2024**, *12*, 696. [CrossRef]
9. Sun, Y.; Zhou, Q.; Sun, L.; Sun, L.; Kang, J.; Li, H. CNN–LSTM–AM: A power prediction model for offshore wind turbines. *Ocean Eng.* **2024**, *301*, 117598. [CrossRef]
10. Zhang, J.; Li, H.; Cheng, P.; Yan, J. Interpretable Wind Power Short-Term Power Prediction Model Using Deep Graph Attention Network. *Energies* **2024**, *17*, 384. [CrossRef]
11. Liu, W.; Sun, Y.; Yu, B.; Wang, H.; Peng, Q.; Hou, M.; Guo, H.; Wang, H.; Liu, C. Automatic Text Summarization Method Based on Improved Text Rank Algorithm and K-Means Clustering. *Knowl. Based Syst.* **2024**, *287*, 111447. [CrossRef]
12. Cheng, D.; Xu, R.; Zhang, B.; Jin, R. Fast density estimation for density-based clustering methods. *Neurocomputing* **2023**, *532*, 170–182. [CrossRef]
13. Bhattacharjee, P.; Mitra, P. A survey of density based clustering algorithms. *Front. Comput. Sci.* **2020**, *15*, 151308. [CrossRef]
14. Wang, M.; Zhang, Y.-Y.; Min, F.; Deng, L.-P.; Gao, L. A two-stage density clustering algorithm. *Soft Comput.* **2020**, *24*, 17797–17819. [CrossRef]
15. Shang, S.; Wang, X.; Liu, A. ABAC policy mining method based on hierarchical clustering and relationship extraction. *Comput. Secur.* **2024**, *139*, 103717. [CrossRef]
16. Jafarzadegan, M.; Safi-Esfahani, F.; Beheshti, Z. An Agglomerative Hierarchical Clustering Framework for Improving the Ensemble Clustering Process. *Cybern. Syst.* **2022**, *53*, 679–701. [CrossRef]
17. Ding, L.; Li, C.; Jin, D.; Ding, S. Survey of spectral clustering based on graph theory. *Pattern Recognit.* **2024**, *151*, 110366. [CrossRef]
18. Gao, C.; Chen, W.; Nie, F.; Yu, W.; Wang, Z. Spectral clustering with linear embedding: A discrete clustering method for large-scale data. *Pattern Recognit.* **2024**, *151*, 110396. [CrossRef]
19. Wang, C.; Gu, Z.; Wei, J.-M. Spectral clustering and embedding with inter-class topology-preserving. *Knowl. Based Syst.* **2024**, *284*, 111278. [CrossRef]
20. Laohakiat, S.; Sa-Ing, V. An incremental density-based clustering framework using fuzzy local clustering. *Inf. Sci.* **2021**, *547*, 404–426. [CrossRef]
21. Gorrab, S.; Ben, R.F.; Nouira, K. Split incremental clustering algorithm of mixed data stream. *Prog. Artif. Intell.* **2024**, *13*, 51–64. [CrossRef]
22. Zhang, Y.; Li, X.; Jiang, S.; Tseng, M.L.; Wang, L.; Fan, S. Dynamic conditional score model-based weighted incremental fuzzy clustering of consumer power load data. *Appl. Soft Comput.* **2023**, *143*, 110395. [CrossRef]
23. Ansari, M.Y.; Ahmad, A.; Bhushan, G. Spatiotemporal trajectory clustering: A clustering algorithm for spatiotemporal data. *Expert Syst. Appl.* **2021**, *178*, 115048. [CrossRef]
24. Pu, Y.; Yao, W.; Li, X.; Alhudhaif, A. An adaptive highly improving the accuracy of clustering algorithm based on kernel density estimation. *Inf. Sci.* **2024**, *663*, 120187. [CrossRef]
25. Wang, L.; He, Y.; He, Y.; Zhou, Y.; Zhao, Q. Wind turbine blade icing risk assessment considering power output predictions based on SCSO-IFCM clustering algorithm. *Renew. Energy* **2024**, *223*, 119969. [CrossRef]
26. Hou, G.; Wang, J.; Fan, Y. Wind power forecasting method of large-scale wind turbine clusters based on DBSCAN clustering and an enhanced hunter-prey optimization algorithm. *Energy Convers. Manag.* **2024**, *307*, 118341. [CrossRef]
27. Na, L.; Jige, Q.S. Identifying urban form typologies in Seoul using a new Gaussian mixture model-based clustering framework. *Environ. Plan. B Urban Anal. City Sci.* **2023**, *50*, 2342–2358.
28. Lorah, J.; Womack, A. Value of sample size for computation of the Bayesian information criterion (BIC) in multilevel modeling. *Behav. Res. Methods* **2019**, *51*, 440–450. [CrossRef] [PubMed]
29. Lin, Z.; Wen, F.; Ding, Y.; Xue, Y. Data-Driven Coherency Identification for Generators Based on Spectral Clustering. *IEEE Trans. Ind. Inform.* **2018**, *14*, 1275–1285. [CrossRef]
30. Liu, M.; Jiao, L.; Liu, X.; Liu, F.; Yang, S. C-CNN: Contourlet Convolutional Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 2636–2649. [CrossRef]
31. Roy, D.; Panda, P.; Roy, K. Tree-CNN: A hierarchical Deep Convolutional Neural Network for incremental learning. *Neural Netw.* **2020**, *121*, 148–160. [CrossRef] [PubMed]
32. Han, L.; Jing, H.; Zhang, R.; Gao, Z. Wind power forecast based on improved Long Short Term Memory network. *Energy* **2019**, *189*, 116300. [CrossRef]
33. Wang, J.; Zhu, H.; Zhang, Y.; Cheng, F.; Zhou, C. A novel prediction model for wind power based on improved long short-term memory neural network. *Energy* **2023**, *265*, 126283. [CrossRef]

34. Gholizadeh, N.; Saadatfar, H.; Hanafi, N. K-DBSCAN: An improved DBSCAN algorithm for big data. *J. Supercomput.* **2021**, *77*, 6214–6235. [CrossRef]
35. Shieh, G. Choosing the best index for the average score intraclass correlation coefficient. *Behav. Res. Methods* **2016**, *48*, 994–1003. [CrossRef]

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.