

Article

ForecastNet Wind Power Prediction Based on Spatio-Temporal Distribution

Shurong Peng ^{1,2}, Lijuan Guo ², Haoyu Huang ², Xiaoxu Liu ¹ and Jiayi Peng ^{3,*}

¹ Sino-German College of Intelligent Manufacturing, Shenzhen Technology University, Shenzhen 518118, China; peng_sr@126.com (S.P.); liuxiaoxu@sztu.edu.cn (X.L.)

² School of Electrical and Information Engineering, Changsha University of Science and Technology, Changsha 410114, China; guolijuan@stu.csust.edu.cn (L.G.); 23205061121@stu.csust.edu.cn (H.H.)

³ State Grid Zhuzhou Power Supply Company, State Grid Hunan Power Company, Zhuzhou 412011, China

* Correspondence: zzd_w_pengjy@163.com

Abstract: The integration of large-scale wind power into the power grid threatens the stable operation of the power system. Traditional wind power prediction is based on time series without considering the variability between wind turbines in different locations. This paper proposes a wind power probability density prediction method based on a time-variant deep feed-forward neural network (ForecastNet) considering a spatio-temporal distribution. First, the outliers in the wind turbine data are detected based on the isolated forest algorithm and repaired through Lagrange interpolation. Then, based on the graph attention mechanism, the features of the proximity node information of the individual wind turbines in the wind farm are extracted and the input feature matrix is constructed. Finally, the wind power probability density prediction results are obtained using the ForecastNet model based on three different hidden layer variants. The experimental results show that the ForecastNet model with a hidden layer as a dense network based on the attention mechanism (ADFN) predicts better. The average width of the prediction intervals at achieved confidence levels for all interval coverage is reduced by 34.19%, 35.41%, and 35.17%, respectively, when compared to the model with the hidden layer as a multilayer perceptron. For different categories of wind turbines, ADFN also achieves relatively narrow interval average widths of 368.37 kW, 315.87 kW, and 299.13 kW, respectively.

Keywords: time-variant deep feed-forward neural network; probability density prediction; spatio-temporal distribution



Citation: Peng, S.; Guo, L.; Huang, H.; Liu, X.; Peng, J. ForecastNet Wind Power Prediction Based on Spatio-Temporal Distribution. *Appl. Sci.* **2024**, *14*, 937. <https://doi.org/10.3390/app14020937>

Academic Editor: Maria Vicidomini

Received: 14 November 2023

Revised: 18 January 2024

Accepted: 19 January 2024

Published: 22 January 2024



Copyright: © 2024 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

With energy and environmental issues becoming more and more prominent [1], global energy is moving towards low-carbon, clean, safe, and efficient development [2]. In recent years, the wind energy industry has been developing rapidly around the world. According to statistics, the global newly installed capacity of wind power reached 93.6 GW in 2021. However, wind power generation is random and fluctuating due to climate conditions. Therefore, when large-scale wind power is integrated into the grid, it can jeopardize the safe and stable operation of the grid [3,4]. Wind power prediction helps to optimize the grid integration of wind power, reduce operating costs, and promote the development of energy systems.

Existing wind power prediction methods are categorized into ultra-short-term prediction [5,6], short-term prediction [7,8] and mid-to-long-term prediction [9,10], with multiple prediction time dimensions. Commonly used wind power prediction models include physical [11–14] and statistical [15–18] methods. Physical methods utilize weather forecast data and related geographic information, and the models are generally more complex. Moreover, the weather forecast data will affect the prediction accuracy to a certain extent. They are often used for the mid-to-long-term prediction of new wind farms or

wind farms with incomplete data. Statistical methods use learning algorithms to analyze historical data and establish the intrinsic connections between historical data. They are commonly used in short-term and ultra-short-term prediction.

Of the many forecasting methods, deep learning [19] algorithms are the most widely used, mainly including time series methods [20] and artificial intelligence algorithms [21]. Common artificial intelligence algorithms include random forests [22], support vector machines [23], extreme learning machines [24], and artificial neural networks [25]. Long- and short-term memory neural networks [26,27] are widely used in wind power prediction due to their excellent properties in analyzing time-series data. Convolutional neural networks (CNN) [28], which have achieved good performance in image processing, have also been applied to wind power prediction. Zhou et al. [29] proposed a combination of RANSAC (random sample consensus) noise screening and the Seq2Seq-Attention-BiGRU model to enhance prediction accuracy. Zhang et al. [30] proposed a novel hybrid prediction method involving individual prediction model training, model ensemble, and error correction considering temporal correlation. Chen et al. [31] proposed a variational mode decomposition-gate recurrent unit network prediction mode to enhance the accuracy of ultra-short wind power forecasting. Although these studies have achieved more accurate wind power prediction results, they lacked research on the spatial characteristics of the wind power data.

Most wind power prediction studies predict the sum of the power of all turbines in the region, taking into account the temporal characteristics. However, the location and contextual information of wind turbines are neglected, and consideration of the variability of turbine power output at different spatial locations is lacking. The studies show that considering both temporal and spatial characteristics of wind turbine data can help improve the accuracy of wind power prediction. Yu et al. [32] proposed a spatiotemporal wind power forecasting model, which utilized a GAT (graph attention network), GRU (gated recursion unit), and GAT-TCN (temporal convolutional network) as the main prediction methods. Zhang et al. [33] proposed a wind power prediction method based on spatiotemporal correlations considering the influences of wind speed, wind direction, and temperature. Wang et al. [34] proposed a short-term wind power forecasting method based on feature clustering and correlation analysis, improving forecasting accuracy through data feature clustering, variable correlation analysis, and building forecasting models.

Comprehensively analyzing existing studies, this paper analyzes the spatial distribution and dynamic contextual information of wind turbines. In order to facilitate the study of the spatial characteristics of wind power data and better analyze the location information, this paper predicts wind power from the perspective of wind turbines rather than wind farms. The dataset used in this paper includes data such as the angle of the wind received by each turbine, the environmental temperatures around the different wind turbines, the internal temperatures of the turbine nacelles, and the orientation of each turbine nacelle. These data are used to model the spatial correlation between the wind turbines. Predictions for each wind turbine are obtained considering the spatial and temporal distribution of the turbines, and then the predictions for the different turbines are summed to obtain the final prediction. Specifically, this paper makes the following contributions:

- We detect outliers based on the isolated forest algorithm and repair them using the Lagrange interpolation method.
- We extract the information about the neighboring nodes of individual turbines in a wind farm based on the graph attention mechanism and construct the input feature matrix.
- We use a time-variant deep feed-forward neural network (ForecastNet) model to obtain the wind power probability density prediction results based on three different hidden layer variants.

The rest of the paper is structured as follows: Section 2 presents the theoretical knowledge of the method proposed in this paper. Section 3 presents the experiments

using the method proposed in this paper and analyzes the experimental results. Section 4 provides the conclusions of this paper.

2. Materials and Methods

2.1. Analysis and Processing of Wind Power Data

2.1.1. Wind Power Anomaly Data Processing

Wind farm data mainly includes the spatial distribution of wind turbines, as well as dynamic background factors such as the temperature, weather, and internal state of turbines. Wind power data will inevitably have anomalies and missing information in the process of collection, transmission, processing, and storage. The existence of these anomalous data can reduce the efficiency of the model training and the accuracy of the prediction [35]. The anomalous data need to be handled without interfering too much with the original data.

The isolated forest algorithm quickly separates the outliers and vacancies in the data from the normal data. Based on the length of the data path to score the data to determine the degree of abnormality, the higher the score, the higher the degree of abnormality of the data. By setting a certain proportion of abnormal data, the algorithm can quickly distinguish between abnormal and normal data, and the abnormal score of the isolated forest algorithm is calculated as shown in Equation (1).

$$S(x) = 2^{-\frac{E(h(x))}{c(x)}}, \tag{1}$$

where $S(x)$ is the outlier score of samples x , with a value range of [0,1]. The larger the value, the more likely the sample is to be labeled as an outlier. $h(x)$ is the path length of the sample in the tree, $h(x) = \ln(x) + \zeta$, and ζ is Euler's constant. $E(h(x))$ is the mean value of the path length of the sample x in the tree. $c(x)$ is the average search path length of the binary tree constituted by a dataset containing x samples.

The Lagrange interpolation method is used to repair the abnormal data detected by the isolated forest, as shown in Equation (2).

$$L(x) = \sum_{i=0}^n y_i \prod_{j=0, j \neq i}^n \frac{x - x_j}{x_i - x_j}, \tag{2}$$

where x_j corresponds to wind speed and y_j corresponds to wind power. Figure 1 shows the wind speed–power plot for the detection of anomalous data. Figure 1a shows the results before detection and repair, and Figure 1b shows the results after detection and repair. The orange dots are the anomalous and repaired data, respectively, and the blue dots are the normal data.

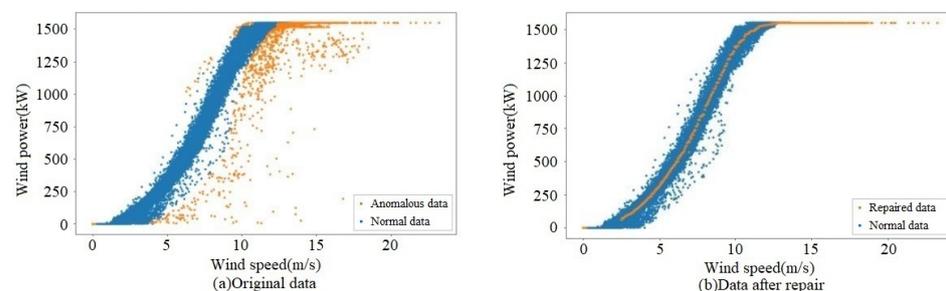


Figure 1. Detection and repair of wind power outliers.

2.1.2. Feature Clustering of Wind Power Data

Wind power output is closely related to factors in the weather environment. Generally, the feature with the strongest correlation with wind power is wind speed. Figure 2 shows the Pearson correlation heat map of wind power data features. From Figure 2, it can be seen that the features that have the greatest degree of influence on wind turbine power

output are wind speed (Wspd), the angle between the wind direction and the position of the turbine nacelle (Wdir), and the pitch angle of blades 1–3 (Pab1–Pab3).

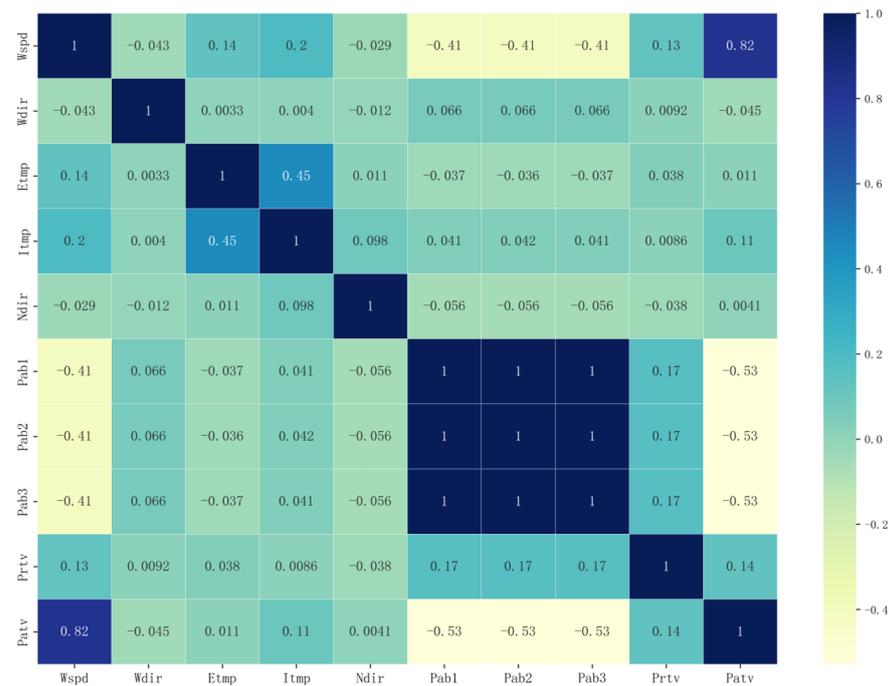


Figure 2. Characteristic correlation heat map.

In order to improve the intrinsic connection between wind power data and better explore the spatial and temporal distribution state of wind power, the K-sums algorithm is used to segmentally cluster wind turbines at different locations [36]. Based on the features related to wind speed, wind turbines distributed in different spatial locations are differentiated. In this paper, the wind direction factor is added on the basis of wind speed. The wind speed and direction of the natural wind in the duration of time is constantly changing, and the atmosphere has a certain inertia and incompressible mobility in a small-time scale range, which leads to periodic changes in wind speed and direction within a certain range. This is manifested by the direction of the wind force and the angle between the wind turbine oscillating back and forth within a certain range [37].

The wind turbines are clustered into two classes based on wind speed characteristics, defined as the class that receives higher wind speeds and the class that receives lower wind speeds. In fact, due to geographic location and environmental factors, low output units still exist in the high wind speed range. For this reason, the high wind speed category will be clustered for the second time using wind direction characteristics such as Wdir and Pab as input variables, and finally, three types of wind turbines will be obtained. Type 0 is for high wind speeds and high output units, type 1 is for high wind speeds and low output units, and type 2 is for low wind speeds and low output units. The results of the K-sums algorithm clustering for visualizing the location of the wind farm are shown in Figure 3.

The spatial categorization of the three types of turbines is more concentrated, and the existence of some outliers is due to the different topography and ground roughness at different turbine locations. The gap between the turbines in the upper right corner of Figure 3 is exactly a railroad.

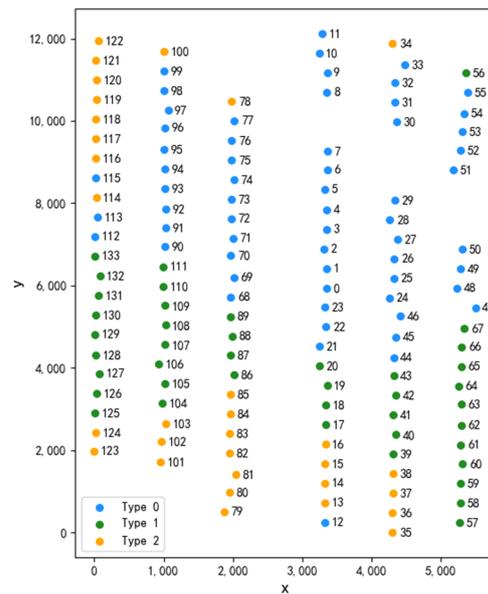


Figure 3. Results of K-sums clustering.

2.2. Wind Power Prediction Model

2.2.1. Euclidean Distance and Differential Distance

In order to obtain the spatio-temporal distribution characteristics of wind turbines, the spatial distance perception function G_d is constructed based on the Euclidean distance, and the differential distance perception function G_s is constructed based on the differential similarity. G_d can reflect the explicit neighboring relationship between wind turbines, and G_s can reflect the invisible neighboring relationship. We calculate the Euclidean distance between two nodes, take the K nodes with the closest distance as neighboring nodes, and form the set $N_d(i)$ to obtain the spatial matrix $A(i, j)$. The expression of $A(i, j)$ is shown in Equation (3). We calculate the differential similarity $Sim(i, j)$ between two nodes, which is calculated as in Equation (4), and the closest K nodes are taken as the differential proximity nodes to form the set $N_s(i)$.

$$A(i, j) = \begin{cases} 1, & j \in N_d(i) \\ 0, & j \notin N_d(i) \end{cases} \quad (3)$$

$$Sim(i, j) = \sum_{t=1}^T ((x_{i,w}^t - x_{i,w}^{t-1}) \cdot (x_{j,w}^t - x_{j,w}^{t-1})), \quad (4)$$

where $x_{i,w} \in R^{T \times 1}$ denotes the wind speed sequence of the i -th wind turbine.

For wind turbines, both $N_d(i)$ and $N_s(i)$ are combined to aggregate wind speed information from two neighboring turbines and merged as input features to improve the performance of the wind power prediction.

2.2.2. Graph Attention

In order to establish spatial and temporal correlations between wind turbines in different geographic locations, an attention-based spatio-temporal graph network [38] is introduced into the time series prediction model. In order to improve the training efficiency and prediction accuracy of the model, the feature information needs to be filtered, and the information with low importance is ignored [39].

Graph attention networks process information by calculating the weights of the information and weighting the information according to certain weights to aggregate the information. Specifically, the attention score of the information is calculated using query

vector (Query), key-value vector (Key), and value vector (Value) [40]. The formula is shown in Equation (5):

$$Attention(Q, K, V) = \sum_i^L \langle Q, K_i \rangle V_i, \quad (5)$$

where Q is the feature vector of the current node, K is the feature vector of the neighboring node, and V is the feature vector of the neighboring node after mapping the weight value W . An attention score is obtained by performing an inner product calculation on the feature vector of a node itself and the feature vectors of the neighboring nodes, and the score is weighted with the node's feature vectors after a normalization operation. Specifically:

1. The central target node and neighbor node attention scores are calculated.

The formula for calculating the value of attention on node i based on the features of node j is shown in Equation (6), where e_{ij} is the attention score between node i and node j , a represents the correlation calculation function between nodes, h_j is the output vector of node i , and W is the weight.

$$e_{ij} = a(Wh_i, Wh_j). \quad (6)$$

2. The weight scores are activated using the activation function.

$$e_{i,j} = LeakyReLU(a^T[Wh_i || Wh_j]). \quad (7)$$

The calculation formula is shown in Equation (7), where $||$ is the vector vertical splicing operation, which splices mapped column vectors; $LeakyReLU$ is the activation function; a is the vector to be learnt; and a^T is the transpose of vector a .

3. Weight normalization

The sum of all the weights should be 1; therefore, the attention score is normalized. The weights are normalized for all neighboring node attention values of node i using the softmax function. The calculation formula is shown in Equation (8).

$$a_{ij} = \frac{\exp(LeakyReLU(e_{ij}))}{\sum_{k \in N_i} \exp(LeakyReLU(e_{ik}))}. \quad (8)$$

4. Information aggregation

The graph attention layer adds the node's feature information and neighbor node's feature information according to a certain weight coefficient, performs feature extraction to form a new node to represent the feature, and outputs the new node feature as a result. The calculation formula is shown in Equation (9):

$$h_i' = \sigma\left(\sum_{j \in N_i} a_{ij} Wh_j\right), \quad (9)$$

where h_i' is the new node features, $\sigma(\cdot)$ is the activation function, and a_{ij} is the effect of node i features on the node j attention score.

The features of the new node obtained according to the parameter W will be different in dimension from the features of the original node, and the parameter W will map the information of the original node to the new space.

5. Multi-head attention mechanisms

The multi-head attention mechanism introduces multiple attention mechanisms to aggregate information, and each attention mechanism is able to focus on different features to enhance the expression ability of the attention layer [41]. The multi-head attention

mechanism splices the outputs of multiple nodes into column vectors to obtain the final new node features, and the new node features are calculated using the Equation (10):

$$h_i' = ||_{k=1}^K \sigma(\sum_{j=N_i} a_{ij}^k W^k h_j), \tag{10}$$

where $||$ denotes the vector splicing operation, a_{ij}^k represents the normalized value of the attention score calculated by the k -th attention mechanism, and W^k is the weight matrix of the linear transformation.

Using the graph attention mechanism to deal with the pair of spatial feature information, combined with the multi-head attention mechanism, different attention score weights are calculated for the target node and its neighboring nodes [42]. It is beneficial to improve the model’s representation of spatial dimensions and reduce the risk of overfitting.

2.2.3. ForecastNet Model

ForecastNet [43] is a multi-layer feed-forward neural network model that is commonly used to perform multi-step time series forecasting. Its network structure is shown in Figure 4, where each neuron can receive signals from the previous layer of neurons, resulting in a multilayer structure. This neural network structure is commonly used for processing time series data because it captures trends and periodicities in time series, as well as other complex dynamic features.

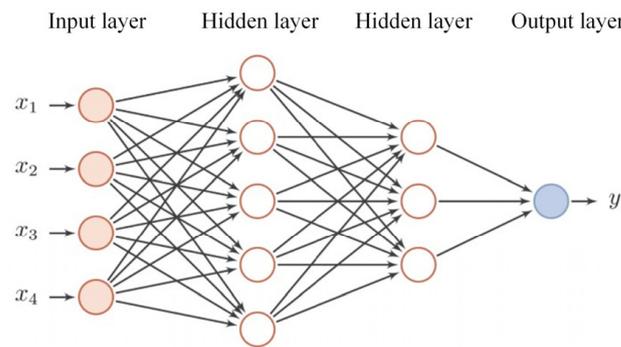


Figure 4. Multilayer feedforward neural network structure.

ForecastNet has time-varying and interleaved output properties among model neurons. The former improves the problem of gradient vanishing during the training process of recurrent neural network (RNN) and CNN models. The latter improves the problem of gradient explosion and gradient vanishing that occurs during neural network training. The root cause of gradient explosion and gradient vanishing is the reuse of chaining laws in neural network gradient computation. The interleaved output properties of ForecastNet are shown in Figure 5, where the upper row represents the hidden layer network, the lower row represents the output layer, $a^{[l]}$ is the output vector of the l -th layer, $z^{[l]} = W^{[l]T} a^{[l-1]} b^{[l]}$, $W^{[l]}$ is the weight matrix of the l -th layer, and $b^{[l]}$ is the bias parameter matrix. It breaks down the chain multiplication chain in the chain rule into the sum of multiple terms. The factor accumulation process is more stable than the multiplication process, which can significantly reduce the depth of the network while alleviating the gradient explosion and gradient vanishing problems.

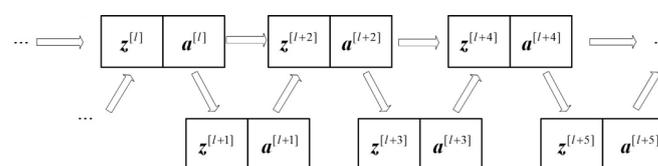


Figure 5. The interleaved output properties of ForecastNet.

The input layer of ForecastNet is univariate or a set of multivariate inputs; the hidden layers are different forms of feed-forward neural networks, such as common back propagation (BP) networks, radial basis function (RBF) networks, etc. The architecture of each hidden layer can be heterogeneous or identical, and the parameters of each hidden layer are independent of each other, which are used to simulate the dynamic characteristics of the time series. Different variants of the ForecastNet model can be obtained by using different feed-forward networks in the hidden layer:

6. ForecastNet model with a multilayer perceptron (MLP) as the hidden layer (MLPFN)

A multilayer perceptron is a special form of a fully connected neural network. Its main difference from a fully connected network is its hidden layer. The hidden layer can improve the MLP's expressive ability for the network, thus improving its ability to solve complex prediction or classification problems. Figure 6 shows a schematic diagram of the ForecastNet hidden layer using an MLP structure, where dense is the fully connected layer and h represents the number of neuron nodes in each hidden layer. The hidden layer and the output layer of the MLP network are fully connected layers. Each layer has 24 ReLU neuron units, where the neuron nodes are fully connected.

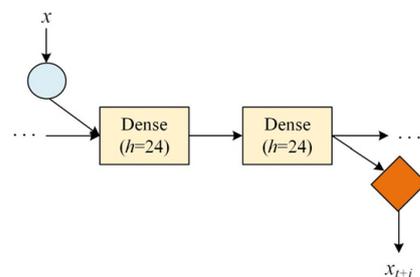


Figure 6. Hidden layer structure of MLPFN.

7. ForecastNet model with CNN as the hidden layer (CNNFN)

CNN is a kind of artificial neural network. Its structure is mainly composed of three parts: the convolution layer, pooling layer, and dense layer, as shown in Figure 7. The main role of the convolution layer is to extract the features, where f is the number of convolution kernels and k is the size of the convolution kernel. The pooling layer is used for down sampling, where s is the filling of the pooling layer and p is the step size of the pooling layer. The dense layer is mainly used for feature classification, where h is the number of hidden layers of the dense layer, which is composed of 24 ReLU neurons.

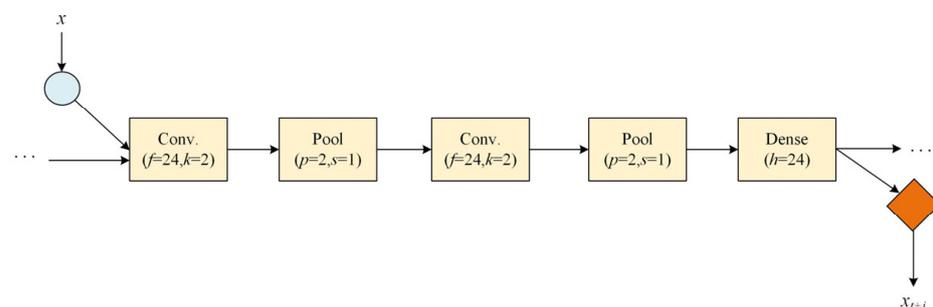


Figure 7. Hidden layer structure of CNNFN.

8. ForecastNet model with the hidden layer as a dense network based on the attention mechanism (ADFN)

The model is shown in Figure 8.

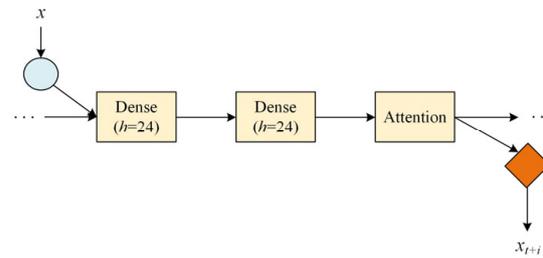


Figure 8. Hidden layer structure of ADFN.

In this paper, the ForecastNet model is used as the basis for wind power spatio-temporal distribution prediction research by using a MLP, CNN, and the dense network based on the attention mechanism in its hidden layer. Each output in the output layer is a prediction result, and a linear output can be used to obtain a deterministic prediction result, or a Gaussian mixture density can be used to establish a probability density output.

2.2.4. ForecastNet Methodology

There is a certain difference in the output power of each wind turbine, and it is not only a large amount of work to build a model for each wind turbine, but also a tedious and complicated model training process. Therefore, the prediction is based on the average wind power. Specifically, the higher the active power (Patv) data of the wind turbines, the closer the output of the turbines to the theoretical output, which means that it is more appropriate to select the turbines with high wind speeds in the k-sums clustering. The data from the N wind turbines with the highest output and high spatio-temporal correlation are selected as the historical data for training the model. Predictions were made using these data as the output of other turbines, and the average of the predictions was used as the prediction for the other turbines in the wind plant.

The new node data obtained from the data processed by the graph attention network mentioned in Section 2.2.2 contains not only the wind turbine feature data and historical wind power data of its own node but also the feature data of the neighboring nodes. These new node data are used as inputs to the ForecastNet model to build a prediction model and perform multi-step predictions. Each step predicts the data for the next 12 time points, and the prediction results obtained from the previous step are merged with the historical features before the next moment prediction. Together, they are used as inputs to the next model. This allows for more accurate learning of wind power trends over time while also fully considering the correlation of wind power data between inputs and outputs and between outputs at different moments. Following the multi-step prediction steps shown in Figure 9, a multi-step prediction of wind power is performed to obtain the prediction results at multiple future points in time. Adding a linear output model to the ForecastNet output layer allows for obtaining deterministic prediction results, and adding a Gaussian mixture density module [44] allows for obtaining probability density prediction results. The specific process framework diagram is shown in Figure 10.

2.2.5. Evaluation Metrics

Common error evaluation metrics such as root mean squared error (RMSE), mean absolute error (MAE), etc., are only applicable to deterministic prediction results and cannot be applied to probabilistic prediction results. In interval probabilistic prediction, the prediction interval coverage is a major metric for evaluating uncertain predictions. The formula is shown in Equation (11):

$$R_{cover} = \frac{c}{v} \times 100\%, \quad (11)$$

where R_{cover} is the interval coverage, c is the number of true values in the test set that fall within the prediction interval, and v is the total number of true values in the test set.

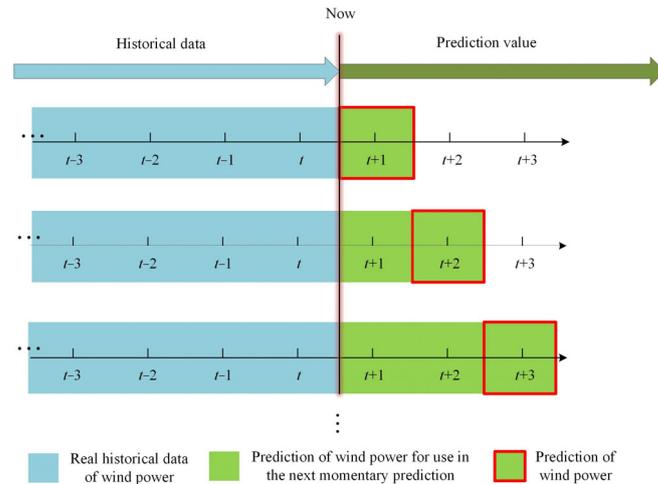


Figure 9. Schematic of multi-step prediction.

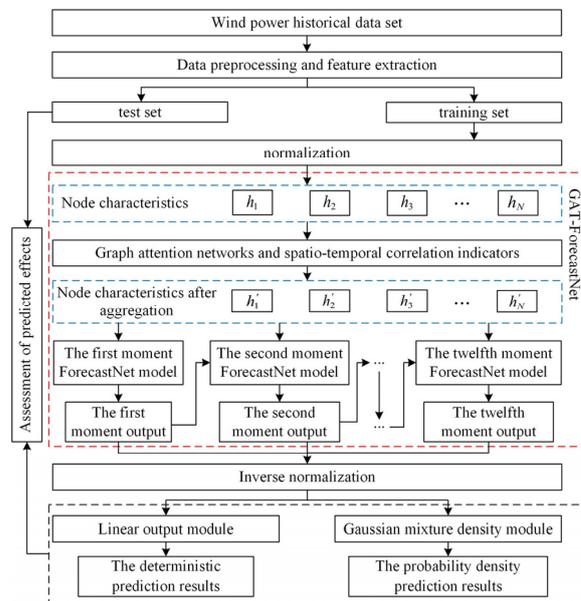


Figure 10. A wind power prediction framework based on spatio-temporally distributed ForecastNet.

Interval coverage indicates the proportion of real values included in the prediction interval. A higher coverage indicates a more accurate prediction. Under a given confidence level, the coverage of the prediction interval cannot be lower than the requirement of the confidence level; otherwise, the prediction result is unreliable. In wind power prediction, interval coverage is one of the important metrics used to assess the prediction accuracy. However, relying solely on the interval coverage is insufficient to evaluate quality of the interval predictions; the interval width metric is also essential. The interval average width metric describes the average width of the prediction interval, which is usually used to assess the accuracy and effectiveness of interval prediction. When the interval coverage is higher, the average width of the interval should be smaller, indicating that the model can provide a more accurate range of intervals in the prediction process. Relatively speaking, the model prediction results of the interval coverage are high-quality, but the average width of the interval is too large may have certain defects. The calculation of the average width of the interval metric is shown in Equation (12):

$$\delta_{mean} = \frac{1}{v} \sum_{t=1}^v \frac{(U_t - L_t)}{(Y_{max} - Y_{min})}, \tag{12}$$

where Y_{\max} is the maximum value of wind power in the test set at the prediction time t , Y_{\min} is the minimum value of wind power in the test set at the prediction time t , U_t is the upper boundary of the prediction interval, and L_t is the lower boundary of the prediction interval.

The larger the average interval width indicator, the larger the power interval obtained from the prediction, and the more valid the information it can provide. Combining the two metrics, interval average width and interval coverage, to jointly judge the uncertain prediction results is of practical significance.

3. Results and Discussion

In this paper, we utilize the data from the open-source dataset 2022 Long yuan Power Group Co., Beijing, China wind farm to verify the effectiveness of the proposed method.

The proximity nodes are identified based on the spatial perception function and differential distance perception function. The data of the neighboring turbine nodes are aggregated based on the graph attention model to improve the model effect. Predictions are made using the ForecastNet model with different implicit layers, predicting data at 12 time points in the future at a time. Rolling predictions are used to determine the power generation from wind turbines in the next two days. The deterministic prediction results are obtained using linear output at the output layer, and the interval prediction results of wind power are obtained using a Gaussian mixture density network for the probability density output.

Wind energy fluctuates over time, and there are differences in the variability at different locations in the same wind farm. A spatio-temporal graph is constructed based on the physical spatial location of individual wind turbines, and the data of the model are reconstructed based on the established relationships between nodes and edges. The strong temporal correlation of the time series model is taken into account while also fully considering the characteristics of the wind turbines in terms of their physical location in space.

3.1. Analysis of Deterministic Prediction Results

Deterministic prediction values can be obtained by using linear output in the output layer of ForecastNet. The RMSE and MAE metric values are calculated separately for the ForecastNet model and the other three prediction models: support vector regression (SVR), K-nearest neighbor (KNN), and light gradient-boosting machine (LightGBM). Tables 1 and 2 show the RMSE and MAE metrics for selecting the top 1–5 proximate wind turbine prediction results with the highest power outputs.

Table 1. RMSE metrics for wind turbine prediction results.

Models	Top = 1	Top = 2	Top = 3	Top = 4	Top = 5
ForecastNet	55.930	59.595	62.336	53.447	65.332
SVR	114.660	109.169	120.044	100.730	106.931
KNN	121.807	120.008	137.426	106.321	121.102
LightGBM	82.583	95.353	106.340	92.396	96.301

Table 2. MAE metrics for wind turbine prediction results.

Models	Wind Turbine 1	Wind Turbine 2	Wind Turbine 3	Wind Turbine 4	Wind Turbine 5
ForecastNet	48.828	43.145	53.664	44.771	62.602
SVR	96.914	93.318	103.126	92.814	110.663
KNN	105.121	100.702	112.003	113.630	129.361
LightGBM	69.743	73.088	80.355	75.230	90.505

Figure 11 demonstrates the trend in RMSE evaluation metrics for selecting different numbers of wind turbines with the highest power output (THPO) in the multi-step prediction process.

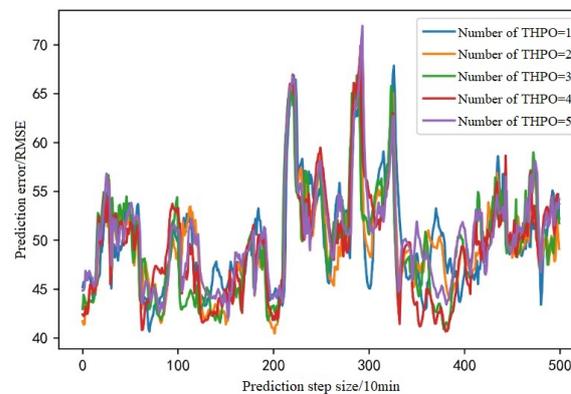


Figure 11. Multi-step prediction error curves for different numbers of THPO.

From Figure 11, it can be concluded that ForecastNet obtains similar error distributions when predicting proximity wind turbines. It indicates that the ForecastNet model achieves relative feature aggregation in dealing with the proximity of wind turbines, obtaining similar error outputs. The RMSE within the short-term prediction is controlled to be less than 60, and its prediction trend is relatively good and stable. However, after 200 predictions, a large error fluctuation occurs, and then the error fluctuation tends to be more stable. The reason for this may be that the wind speed and other characteristics have fluctuated more during this prediction. Overall, most of the prediction steps have good error performance, and from the data in the Figure 11 and Tables 1 and 2, it can be seen that better prediction results can be achieved by selecting the four wind turbines with the highest power outputs for prediction.

3.2. Analysis of Probability Density Prediction Results

In order to verify the prediction effect of ForecastNet on the uncertainty of the spatio-temporal distribution of wind power, the hidden layer of ForecastNet is improved. Different hidden layer connection methods such as MLP, CNN, and attention mechanism fully connected network are chosen to analyze their effects on the prediction effect. The rolling forecast is used to predict the wind power, and each time the data of the next 12 time points are predicted, the multi-step prediction is performed to predict the distribution of wind power in the next 2 days. The prediction results of different prediction steps in the multi-step prediction process were randomly selected, and the prediction intervals at different confidence levels (80%, 85%, and 90%) are shown in Figure 12.

As can be seen from the Figure 12, the prediction intervals of each probabilistic prediction model cover most of the real values more accurately, and the trends in the upper and lower limits of the prediction intervals are also consistent with the trend in the wind power. The wind power prediction intervals obtained at different confidence levels are different, and the larger the confidence level, the wider the given prediction interval. The specific prediction evaluation indexes are shown in Table 3.

Different variants can be obtained by changing the hidden layer structure of ForecastNet. As can be seen from Table 3, the highest coverage of prediction intervals at different confidence levels is achieved by the CNNFN model, with interval coverage of reliability metrics at 90%, 85%, and 80%, and confidence levels of 100%, 98.95%, and 95.32%, respectively. However, interval coverage that is too high may cause the average width of the intervals to rise. The average width of the intervals of the CNNFN model is also the highest among the three models, which is 72.11 kW, 67.75 kW, and 63.31 kW at 90%, 85%, and 80% confidence levels, respectively. The wider the average width of the intervals in the case of certainty of the interval coverage, the less practical significance of the interval prediction.

Both the ADFN model and the MLPFN model obtained relatively good prediction results, and the predicted interval coverages obtained are up to the pre-specified confidence level. At confidence levels of 90%, 85%, and 80%, the interval coverage of the MLPFN model is higher than that of the ADFN model by 4.11%, 4.28%, and 3.59%, respectively. However, the average width of the intervals of the ADFN is narrower, which is reduced by 34.19%, 35.41%, and 35.17% compared with that of the MLPFN model, respectively, achieving better prediction results. Combining the above analyses, it can be concluded that the prediction effects of the ADFN and MLPFN models are better than that of the CNNFN model. The ADFN model achieves a narrower interval prediction width based on the attention mechanism at the confidence level of interval coverage, and it is able to better fit the trend in wind power. The CNNFN model, on the other hand, obtained 100% interval coverage at a higher confidence level, making the prediction intervals less meaningful in practice.

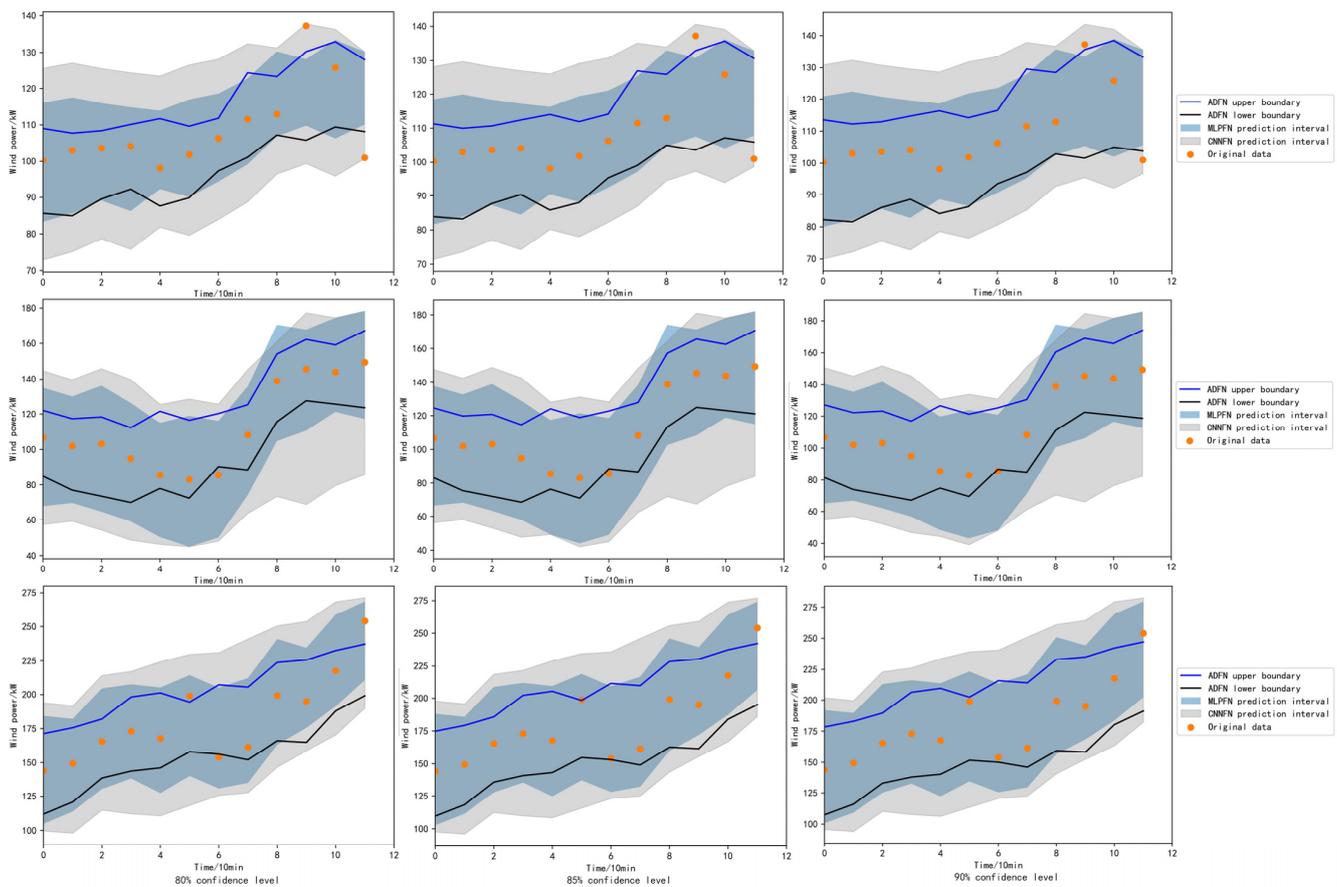


Figure 12. Comparison of wind power prediction intervals for different ForecastNet models.

Table 3. Comparison of evaluation indexes for the different models.

Confidence Levels %	Models	Interval Coverage/%	Average Width of the Interval/kW
80	CNNFN	95.32	63.31
	MLPFN	87.26	50.46
	ADFN	83.67	32.71
85	CNNFN	98.95	67.75
	MLPFN	90.89	54.36
	ADFN	86.61	35.11
90	CNNFN	100.0	72.11
	MLPFN	94.54	59.95
	ADFN	90.43	39.45

In addition, predictions are made for the three different classes of wind turbines obtained from the previous clustering. Figures 13–15 demonstrate the probability density predictions of the ADFN model over the next two days. The more common probabilistic prediction models selected for comparison at the 90% confidence level are LSTM (combination of long short-term memory network and Gaussian mixture density model) and the QRGBM model (gradient booster using quantile output). Table 4 shows the corresponding prediction results.

As can be seen from Table 4, the turbines in the high wind speed and high output units and the turbines in the high wind speed and low output units achieve the given confidence level when uncertainty prediction is made for turbines in different clusters at the 90% confidence level. The highest interval coverage is obtained by QRGBM in the prediction results, with 99.50%, 88.13%, and 97.01%, respectively. However, the interval width of QRGBM is wider than the other models, reaching 452.58 kW, 358.46 kW, and 329.57 kW, respectively. The overly wide interval width will lead to less valid information available in the prediction results. The LSTM and ADFN models have similar interval coverages. For type 0 (high wind speed and high output units), the ADFN model obtains a slightly narrower average width of intervals, but it is wider for other types, and its average width of intervals increases by 5.23% and 4.34% compared to LSTM model. ADFN, based on the attention mechanism, with a small increase in interval coverage, also increases the average width of intervals by a certain amount, with better prediction results. For type 2 (low wind speed and low output units), the interval coverages of all three models do not reach the confidence level, and the prediction effects are not good.

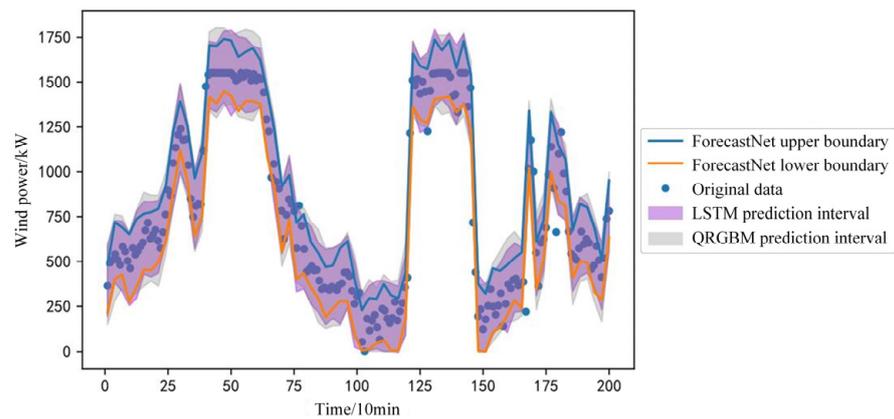


Figure 13. Comparison of interval prediction results for different models for turbine #1.

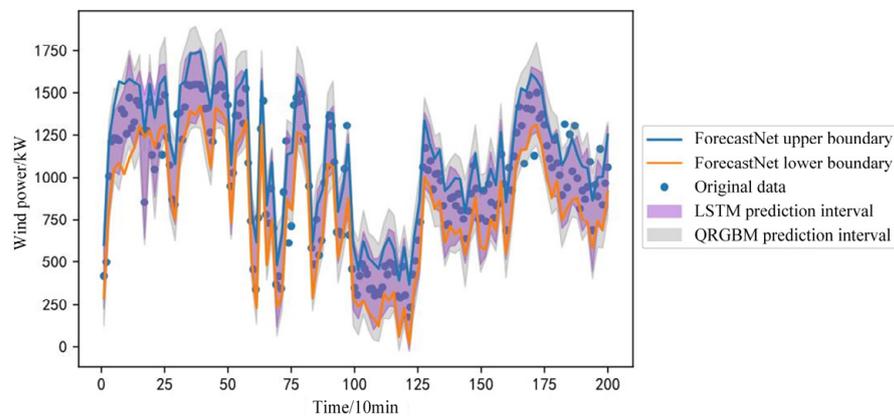


Figure 14. Comparison of interval prediction results for different models for turbine #96.

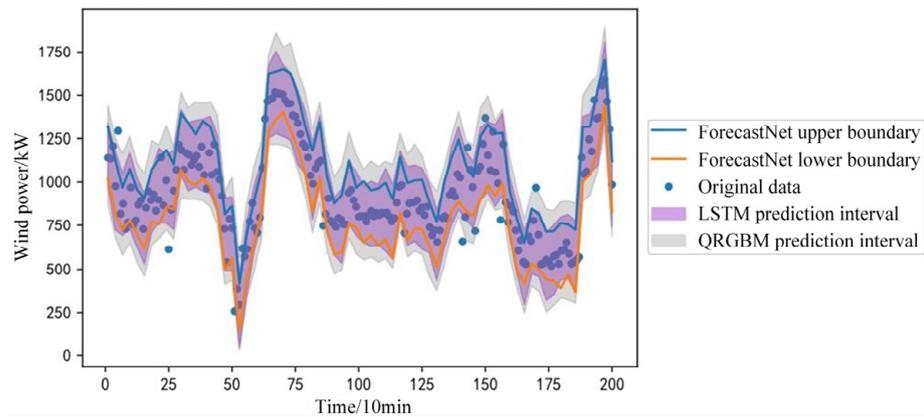


Figure 15. Comparison of interval prediction results for different models for turbine #128.

Table 4. Comparison of forecast indicators for different wind turbines.

Wind Turbine Serial Number	Models	Interval Coverage/%	Average Width of the Interval/kW
#1 (Type 0)	LSTM	94.52	373.43
	ADFN	95.36	368.37
	QRGBM	99.50	452.58
#96 (Type 2)	LSTM	83.92	300.16
	ADFN	84.13	315.87
	QRGBM	88.13	358.46
#128 (Type 1)	LSTM	94.43	286.68
	ADFN	95.06	299.13
	QRGBM	97.01	329.57

Furthermore, this paper randomly selected the probability density function of the ADFN prediction model at different wind turbines and different time prediction points from the multi-step prediction, as shown in Figure 16.

The straight line perpendicular to the x-coordinate axis in Figure 16 is the true value at that prediction moment, and the curve is the probability density distribution of the corresponding model. From Figure 16, it can be seen that the probability density curve is complete and smooth, and there is no missing, very high, or very low values. The curve is also not too broad or too narrow, indicating that the prediction effect of the algorithm is appropriate. Most of the real values in the prediction results of the ADFN model fall near the highest probability points of the probability density curve, such as when $t = 167$, $t = 110$, $t = 100$, and $t = 0$. This indicates that the algorithm predicts with a high accuracy. When the actual wind power value is in the vicinity of the crest of the probability density curve, it indicates that the real value of wind power falls in the high probability interval given by the interval prediction, and the prediction error of these moments is small. When $t = 20$, the real value of wind power deviates from the center of the probability curve, and when $t = 30$, the deviation is even further away. This indicates that the prediction error at these moments is large, and the prediction interval may not even cover the real value of wind power. In the probability density curve results obtained from the test set, the prediction results are reliable if most of the actual values are close to the center range of the curve and only a small number of actual values deviate farther. If most of the actual values are deviated or even far away from the position of the peak of the probability density curve, the obtained probability density prediction results are unreliable. In summary, the ADFN prediction model is reliable in predicting the probability density of different wind turbines.

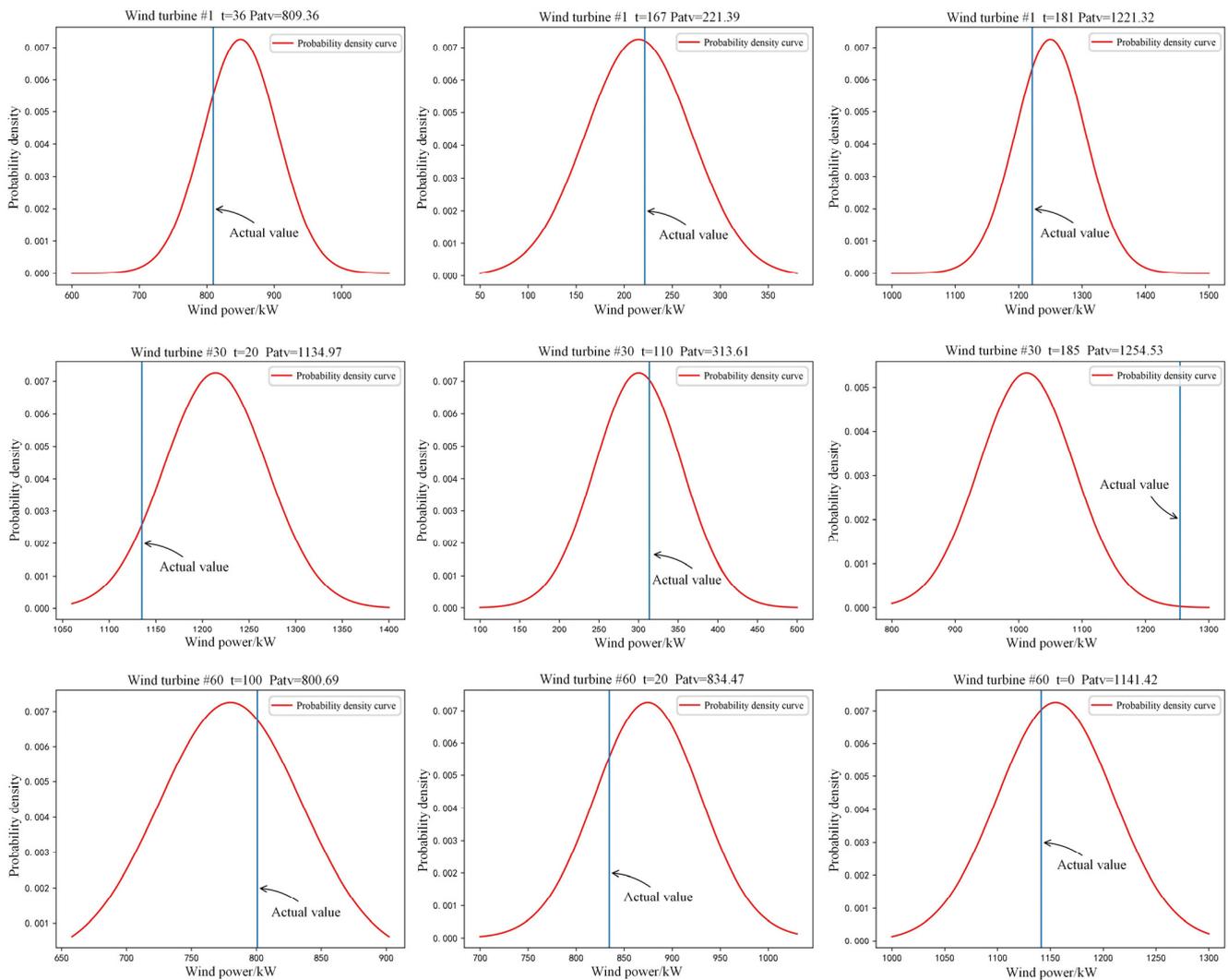


Figure 16. Probability density function of wind power at different moments.

4. Conclusions

This paper focuses on predicting the spatio-temporal distribution of wind power from different turbines in wind farms and proposes a ForecastNet prediction model based on the spatio-temporal distribution. In order to simplify the prediction, the data of N wind turbines with the highest output power in the wind farm are selected to train the model, where the determination of the N value is based on the deterministic prediction results. Then, the uncertainty prediction is performed on the wind power data. The prediction results of ForecastNet variant models based on different hidden layers are compared with those of common prediction models. The results show that the ADFN model has better prediction results when considering the reliability index and acuity index. In addition, the prediction model in this paper has the following features:

- The model uses Euclidean distance and difference distance combined with a graph attention network to aggregate spatio-temporal information on input features.
- In order to avoid the gradient problem when the neural network model predicts wind power in the short term, the time-varying characteristics and interleaved output characteristics of ForecastNet are used to improve the prediction effect.
- The model can predict the probability density curve of wind power at future moments, which can provide more effective information for grid decision makers.

Author Contributions: Conceptualization, S.P., X.L. and J.P.; methodology, S.P. and J.P.; software, L.G. and J.P.; validation, L.G., H.H. and J.P.; writing—original draft preparation, L.G. and J.P.; writing—review and editing, L.G., H.H. and J.P.; supervision, S.P. and X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by National Natural Science Foundation of China (grant number: No. 62003218) and in part by the Stable Support Projects for Shenzhen Higher Education Institutions (grant number: No. 20220717223051001).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the data used in this paper can be obtained by contacting the authors of this study.

Conflicts of Interest: Author Jiayi Peng was employed by the company State Grid Zhuzhou Power Supply Company. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

ForecastNet	Deep feed-forward neural network
Wspd	Wind speed
Wdir	The angle between the wind direction and the position of turbine nacelle
Etmp	Environmental temperature
Itmp	Turbine nacelle internal temperature
Ndir	Nacelle direction
Pab 1	Pitch angle of blade 1
Pab 2	Pitch angle of blade 2
Pab 3	Pitch angle of blade 3
Prtv	Reactive power
Patv	Active power
CNN	Convolutional neural network
MLP	Multilayer perceptron
MLPFN	ForecastNet model with multilayer perceptron as hidden layer
CNNFN	ForecastNet model with CNN as hidden layer
ADFN	ForecastNet model with hidden layer as a dense network based on the attention mechanism
RMSE	Root mean squared error
MAE	Mean absolute error
SVR	Support vector regression
KNN	K nearest neighbor
LightGBM	Light gradient boosting machine
THPO	Wind turbines with the highest power output

References

- Li, J.; Liu, K.; Yang, Z.X.; Qu, Y. Evolution and impacting factors of global renewable energy products trade network: An empirical investigation based on ERGM model. *Sustainability* **2023**, *15*, 8701. [[CrossRef](#)]
- Lu, Y.; Liu, X.; Li, H.J.; Wang, H.R.; Kong, J.J.; Zhong, C.; Cui, M.L.; Li, Y.; Sun, X.Q.; Xuan, J.D.; et al. What is the impact of the renewable energy power absorption guarantee mechanism on China's green electricity market? *Energies* **2023**, *16*, 7434. [[CrossRef](#)]
- Hu, S.; Xiang, Y.; Zhang, H.C.; Xie, S.Y.; Li, J.H.; Gu, C.H.; Sun, W.; Liu, J.H. Hybrid forecasting method for wind power integrating spatial correlation and corrected numerical weather prediction. *Appl. Energy* **2021**, *293*, 116951. [[CrossRef](#)]
- Chen, W.H.; Zhou, H.T.; Cheng, L.S.; Xia, M. Prediction of regional wind power generation using a multi-objective optimized deep learning model with temporal pattern attention. *Energy* **2023**, *278*, 127942. [[CrossRef](#)]
- Yang, M.; Guo, Y.F.; Huang, Y.T. Wind power ultra-short-term prediction method based on NWP wind speed correction and double clustering division of transitional weather process. *Energy* **2023**, *282*, 128947. [[CrossRef](#)]
- Peng, C.; Zhang, Y.Q.; Zhang, B.W.; Song, D.; Lyu, Y.; Tsoi, A. A novel ultra-short-term wind power prediction method based on XA mechanism. *Appl. Energy* **2023**, *351*, 121905. [[CrossRef](#)]
- Ying, H.M.; Deng, C.H.; Xu, Z.H.; Huang, H.X.; Deng, W.S.; Yang, Q.L. Short-term prediction of wind power based on phase space reconstruction and BiLSTM. *Energy Rep.* **2023**, *9*, 474–482. [[CrossRef](#)]

8. Xu, P.H.; Zhang, M.Y.; Chen, Z.H.; Wang, B.Q.; Cheng, C.; Liu, R.F. A deep learning framework for day ahead wind power short-term prediction. *Appl. Sci.* **2023**, *13*, 4042. [[CrossRef](#)]
9. Cai, Z.; Dai, S.; Ding, Q.; Zhang, J.L.; Xu, D.; Li, Y.X. Gray wolf optimization-based wind power load mid-long term forecasting algorithm. *Comput. Electr. Eng.* **2023**, *109*, 108769. [[CrossRef](#)]
10. Han, S.; Qiao, Y.H.; Yan, J.; Liu, Y.Q.; Li, L.; Wang, Z. Mid-to-long term wind and photovoltaic power generation prediction based on copula function and long short term memory network. *Appl. Energy* **2019**, *239*, 181–191. [[CrossRef](#)]
11. Guo, N.Z.; Shi, K.Z.; Li, B.; Qi, L.W.; Wu, H.H.; Zhang, Z.L.; Xu, J.Z. A physics-inspired neural network model for short-term wind power prediction considering wake effects. *Energy* **2022**, *261*, 125208. [[CrossRef](#)]
12. Yakoub, G.; Mathew, S.; Leal, J. Intelligent estimation of wind farm performance with direct and indirect ‘point’ forecasting approaches integrating several NWP models. *Energy* **2023**, *263*, 125893. [[CrossRef](#)]
13. Zhang, Y.; Li, Y.T.; Zhang, G.Y. Short-term wind power forecasting approach based on Seq2Seq model using NWP data. *Energy* **2020**, *213*, 118371. [[CrossRef](#)]
14. Liu, C.Y.; Zhang, X.M.; Mei, S.W.; Zhou, Q.Y.; Fan, H. Series-wise attention network for wind power forecasting considering temporal lag of numerical weather prediction. *Appl. Energy* **2023**, *336*, 120815. [[CrossRef](#)]
15. Wang, J.N.; Zhu, H.Q.; Cheng, F.; Zhou, C.; Zhang, Y.J.; Xu, H.L.; Liu, M.H. A novel wind power prediction model improved with feature enhancement and autoregressive error compensation. *J. Clean. Prod.* **2023**, *420*, 138386. [[CrossRef](#)]
16. Li, D.Q.; Yu, X.D.; Liu, S.L.; Dong, X.; Zang, H.Z.; Xu, R. Wind power prediction based on PSO-Kalman. *Energy Rep.* **2022**, *8*, 958–968. [[CrossRef](#)]
17. Zhang, W.Q.; Lin, Z.; Liu, X.L. Short-term offshore wind power forecasting—A hybrid model based on Discrete Wavelet Transform (DWT), Seasonal Autoregressive Integrated Moving Average (SARIMA), and deep-learning-based Long Short-Term Memory (LSTM). *Renew. Energy* **2022**, *185*, 611–628. [[CrossRef](#)]
18. Le Goff Latimier, R.; Le Bouedec, E.; Monbet, V. Markov switching autoregressive modeling of wind power forecast errors. *Electr. Power Syst. Res.* **2020**, *189*, 106641. [[CrossRef](#)]
19. Zhen, Z.; Liu, J.M.; Zhang, Z.Y.; Wang, F.; Chai, H.; Yu, Y.L.; Lu, X.X.; Wang, T.Q.; Lin, Y.Z. Deep learning based surface irradiance mapping model for solar PV power forecasting using sky image. *IEEE Trans. Ind. Appl.* **2020**, *56*, 3385–3396. [[CrossRef](#)]
20. Yang, M.; Chen, X.X.; Du, J.; Cui, Y. Ultra-short-term multistep wind power prediction based on improved EMD and reconstruction method using run-length analysis. *IEEE Access* **2018**, *6*, 31908–31917. [[CrossRef](#)]
21. Lu, H.F.; Ma, X.; Huang, K.; Azimi, M. Prediction of offshore wind farm power using a novel two-stage model combining kernel-based nonlinear extension of the Arps decline model with a multi-objective grey wolf optimizer. *Renew. Sustain. Energy Rev.* **2020**, *127*, 109856. [[CrossRef](#)]
22. Lahouar, A.; Ben Hadj Slama, J. Hour-ahead wind power forest based on random forest. *Renew. Energy* **2017**, *109*, 529–541. [[CrossRef](#)]
23. Zhao, Y.Y.; Song, H.Y.; Guo, Y.J.; Zhao, L.J.; Sun, H.X. Super short term combined power prediction for wind power hydrogen production. *Energy Rep.* **2022**, *8*, 1387–1395. [[CrossRef](#)]
24. Wang, X.Y.; Li, J.; Shao, L.; Liu, H.L.; Ren, L.; Zhu, L.H. Short-term wind power prediction by an extreme learning machine based on an improved hunter–prey optimization algorithm. *Sustainability* **2023**, *15*, 991. [[CrossRef](#)]
25. Ogliari, E.; Guilizzoni, M.; Giglio, A.; Pretto, S. Wind power 24-h ahead forecast by an artificial neural network and an hybrid model: Comparison of the predictive performance. *Renew. Energy* **2021**, *178*, 1466–1474. [[CrossRef](#)]
26. Wei, J.Q.; Wu, X.J.; Yang, T.M.; Jiao, R.H. Ultra-short-term forecasting of wind power based on multi-task learning and LSTM. *Int. J. Electr. Power Energy Syst.* **2023**, *149*, 109073. [[CrossRef](#)]
27. Xiao, Z.X.; Tang, F.; Wang, M.Y. Wind power short-term forecasting method based on LSTM and multiple error correction. *Sustainability* **2023**, *15*, 3798. [[CrossRef](#)]
28. Liu, X.; Yang, L.X.; Zhang, Z.J. 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](#)]
29. Zhou, G.Y.; Hu, G.F.; Zhang, D.X.; Zhang, Y. A novel algorithm system for wind power prediction based on RANSAC data screening and Seq2Seq-Attention-BiGRU model. *Energy* **2023**, *283*, 128986. [[CrossRef](#)]
30. Zhang, S.Y.; Liu, M.B.; Liu, M.D.; Lei, Z.X.; Zeng, G.H.; Chen, Z.R. Day-ahead wind power prediction using an ensemble model considering multiple indicators combined with error correction. *Appl. Soft. Comput.* **2023**, *148*, 110873. [[CrossRef](#)]
31. Chen, H.P.; Wu, H.; Kan, T.Y.; Zhang, J.H.; Li, H.L. Low-carbon economic dispatch of integrated energy system containing electric hydrogen production based on VMD-GRU short-term wind power prediction. *Int. J. Electr. Power Energy Syst.* **2023**, *154*, 109420. [[CrossRef](#)]
32. Yu, C.Q.; Yan, G.X.; Yu, C.M.; Zhang, Y.; Mi, X.W. A multi-factor driven spatiotemporal wind power prediction model based on ensemble deep graph attention reinforcement learning networks. *Energy* **2023**, *263*, 126034. [[CrossRef](#)]
33. Zhang, J.A.; Liu, D.; Li, Z.J.; Han, X.; Liu, H.; Dong, C.; Wang, J.Y.; Liu, C.Y.; Xia, Y.P. Power prediction of a wind farm cluster based on spatiotemporal correlations. *Appl. Energy* **2021**, *302*, 117568. [[CrossRef](#)]
34. Wang, Y.J.; Wang, J.D.; Cao, M.; Li, W.X.; Yuan, L.; Wang, N. Prediction method of wind farm power generation capacity based on feature clustering and correlation analysis. *Electr. Power Syst. Res.* **2022**, *212*, 108634. [[CrossRef](#)]
35. Li, Y.; Xiao, Z.Q.; Nie, S.S.; Cao, J.W.; Hua, H.C. Review of research on generative adversarial network and its application in new energy data quality. *South. Power Syst. Technol.* **2020**, *14*, 25–33. [[CrossRef](#)]

36. Xu, W.J.; Wu, S.L.; Wang, F.Y.; Lin, L.; Li, G.J.; Zhang, Z. Large-scale post-disaster user distributed coverage optimization based on multi-agent reinforcement learning. *J. Commun.* **2022**, *43*, 1–16.
37. Xu, B.K. Fan yaw control optimization strategy based on clustering algorithm and wind direction prediction. *J. Shandong Electr. Power Coll.* **2017**, *20*, 50–54.
38. Ni, Q.J.; Wang, Y.H.; Yuan, J.Y. Adaptive scalable spatio-temporal graph convolutional network for PM2.5 prediction. *Eng. Appl. Artif. Intell.* **2023**, *126*, 107080. [[CrossRef](#)]
39. Yin, L.S.; Wu, Y.Y. Traffic flow combination prediction model based on improved VMD-GAT-GRU. *J. Electron. Meas. Instrum.* **2022**, *36*, 62–72. [[CrossRef](#)]
40. Feng, B.; Zhang, Y.W.; Tang, X.; Guo, C.X.; Wang, J.J.; Yang, Q.; Wang, H.F. Power equipment defect record text mining based on BiLSTM-attention neural network. *Proc. CSEE* **2020**, *40*, 1–10. [[CrossRef](#)]
41. Wei, J.; Zhao, H.T.; Liu, D.N.; Jia, H.P.; Wang, X.Y.; Zhang, H.; Liu, Z. Short-term power load forecasting method by attention-based CNN-LSTM. *J. North China Electr. Power Univ. (Nat. Sci. Ed.)* **2021**, *48*, 42–47.
42. Li, J.Q.; Hu, X.D.; Qin, J.R.; Zhang, C.Z. Ultra-short term wind power prediction based on multi-head attention and CNN model. *Electr. Power Sci. Eng.* **2022**, *38*, 34–40.
43. Dabrowski, J.J.; Zhang, Y.; Rahman, A. ForecastNet: A Time-Variant Deep Feed-Forward Neural Network Architecture for Multi-step-Ahead Time-Series Forecasting. *Int. Conf. Neural Inf. Process.* **2020**, *12534*, 579–591. [[CrossRef](#)]
44. Zheng, W.Q.; Peng, X.G.; Lu, D.; Zhang, D.; Liu, Y.; Lin, Z.H.; Lin, L.X. Composite quantile regression extreme learning machine with feature selection for short-term wind speed forecasting: A new approach. *Energy Convers. Manag.* **2017**, *151*, 737–752. [[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.