

Article

Wind Power Group Prediction Model Based on Multi-Task Learning

Da Wang, Mao Yang * and Wei Zhang

Key Laboratory of Modern Power System Simulation and Control & Renewable Energy Technology,
Ministry of Education (Northeast Electric Power University), Jilin 132012, China

* Correspondence: 20102339@neepu.edu.cn; Tel.: +86-15844221802

Abstract: Large-scale wind power grid connection increases the uncertainty of the power system, which reduces the economy and security of power system operations. Wind power prediction technology provides the wind power sequence for a period of time in the future, which provides key technical support for the reasonable development of the power generation plan and the arrangement of spare capacity. For large-scale wind farm groups, we propose a cluster model of wind power prediction based on multi-task learning, which can directly output the power prediction results of multiple wind farms. Firstly, the spatial and temporal feature matrix is constructed based on the meteorological forecast data provided by eight wind farms, and the dimensionality of each attribute is reduced by the principal component analysis algorithm to form the spatial fusion feature set. Then, a network structure with bidirectional gated cycle units is constructed, and a multi-output network structure is designed based on the Multi-gate Mixture-of-Experts (MMoE) framework to design the wind power group prediction model. Finally, the data provided by eight wind farms in Jilin, China, was used for experimental analysis, and the predicted average normalized root mean square error is 0.1754, meaning the prediction precision meets the scheduling requirement, which verifies the validity of the wind power prediction model.

Keywords: multi-task learning; wind farms; wind power cluster; MMoE; wind power prediction



Citation: Wang, D.; Yang, M.; Zhang, W. Wind Power Group Prediction Model Based on Multi-Task Learning. *Electronics* **2023**, *12*, 3683. <https://doi.org/10.3390/electronics12173683>

Academic Editor: Shoji Nishikata

Received: 1 August 2023

Revised: 28 August 2023

Accepted: 29 August 2023

Published: 31 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

According to the Global Wind Energy Council (GWEC), which released the “Global Wind Energy Report 2023”, the renewable capacity was expected to further increase by over 8% in 2022, reaching almost 320 GW. Among them, the wind power (WP) development scale of China is the first in the world [1]. Wind power is a kind of uncertain power supply, which has certain intermittently, fluctuation, and randomness. The grid connection of large-scale WP brings serious challenges to the safe and stable operation of the power system. Wind power prediction (WPP) technology provides a power prediction sequence for a certain period in the future, which provides technical support for the formulation of a power generation plan and reserve of spare capacity [2,3]. The improvement in the WPP accuracy is of great significance for promoting the consumption of WP, reducing the cost of thermal power generation, and enhancing the competitive advantage of WP in the power market.

According to the prediction time scale, WPP technology can be divided into ultra-short-term prediction, short-term prediction, and medium- and long-term prediction. According to a spatial scale, WPP technology can be divided into single wind turbine prediction, wind farm prediction, and wind power cluster prediction [4,5].

The prediction time scale of the short-term wind power prediction (STWPP) model is 1–3 days in the future, that is, it provides the WPP sequence of 0–72 h the next day, and the number of executions of the model is determined by the refresh frequency of the numerical weather prediction (NWP) [6]. So, the STWPP technology is of great significance for the

formulation of the power generation plan, reserve of spare capacity, and adjustment of unit combination plan [7]. According to the modeling method, the STWPP can be divided into physical-based methods and statistical-based methods. Based on the actual geographic information of the wind farm (WF), the physics-based prediction model can predict the wind speed (WS) and wind direction (WD) in the range of the WF by solving high-dimensional nonlinear equations. Then, according to the wind speed–power conversion model, the predicted output of the WP is obtained. This prediction model can reflect the operating characteristics of wind farms and has good interpretability. Since the complex topography and geomorphology of wind farms are difficult to describe mathematically, the accuracy of such prediction methods is often limited [8].

By mining the autocorrelation of the time series and establishing a nonlinear mapping relationship between history and the future, a statistical-based model transforms the static time series modeling problem into a dynamic time series modeling task [9,10]. Such methods include time series predictive modeling algorithms and traditional machine learning algorithms. The time series prediction algorithms include the continuous method, moving average method, auto-regressive moving average method, and so on. Machine learning algorithms include support vector machines, artificial neural networks, and so on. The prediction method based on statistics avoids the physical mechanism to some extent, and the modeling is more efficient. However, the ability to mine the temporal characteristics of time series is limited, which is often suitable for ultra-short-term power prediction tasks [11,12]. The WP series is a dynamic time series, and the WPP results are not only related to the current state but also to the past state. Therefore, traditional machine learning algorithms have certain limitations for STWPP tasks.

With the development of deep learning (DL) technology, its algorithm has also made obvious breakthroughs in time series prediction. Therefore, the WPP algorithm also adopts DL technology in a wide range. On the one hand, DL algorithms can improve the ability of the nonlinear feature extraction of data through the stacking of network layers. On the other hand, for the development of WPP systems, DL models are easy to maintain and deploy [13,14]. The authors in [15] further improved the complementary ensemble empirical mode decomposition (CEEMD) algorithm based on those in [16] to reduce data noise in the mode decomposition process, and then the moth–flame algorithm was applied to optimize the wavelet neural network parameters to enhance the reliability of the modeling. Duan et al. adopted the maximum entropy iteration algorithm to automatically search for the optimal number of modes, improved the loss function of the regression task, and significantly improved the modeling accuracy [17]. Due to the large number of sub-models, the efficiency of modeling is relatively low for the technical route of decomposing power data and then modeling each component separately. Zhang et al. used the K-means algorithm to divide the data into several scenes, and then Sequence-to-Sequence modeling was carried out for each scene, which improved the learning ability of the model for each scene with a high degree of refinement [18]; this is the most common modeling method in the current research. In [19], Ding et al. proposed a wind speed correction algorithm to further construct statistical characteristics based on the WS correction, and then the bidirectional gate recurrent unit (Bi_GRU) network was constructed to establish the WPP model. As the NWP is the main input of WPP modeling, numerical weather prediction error is the main cause of power prediction error, and the correction of WS can significantly improve the prediction accuracy of WP. In addition, new network structures, such as deep residual networks and graph neural networks, can also improve the accuracy of WPP [20,21].

The DL algorithm has been widely used in the field of WF power prediction, but for the dispatching side, the regional wind power output has a greater value for power and electricity balance. Although the dispatching side can collect the forecast results of each WF and then obtain the regional forecast power by superimposing the predicted power of all the WFs, it is difficult to ensure the simultaneity of the data reported by each WF. To solve this problem, a wind power group prediction model is proposed in this paper, which

can ensure the simultaneity of the predicted power of each WF. The framework utilizes the advantages of recurrent neural networks in a time series prediction and is combined with a multi-task learning framework to develop a synchronous multi-output, the specific research contents of which are as follows:

- (1) The dimensionality of WS, WD, temperature, humidity, and pressure data of each WF is reduced based on the principal component analysis algorithm, and the input feature set of the WPP model is formed together with the original meteorological attributes.
- (2) A Bi_GRU network is built as the base learner and a multi-task learning mechanism is designed based on the MMoE algorithm to train the power of multiple WFs in space at the same time to improve modeling efficiency.
- (3) Simulation experiments were conducted on the data provided by eight WFs in Jilin Province, China, and RMSE and MAE indexes were used to evaluate the prediction performance.

The rest of the manuscript is structured as follows: Section 2 is a summary of the problems and a description of the methods; Section 3 is the technical route; Section 4 is the experimental analysis; and Section 5 is the conclusion of this paper.

2. Materials and Methods

2.1. Power Group Prediction of Wind Power Cluster

Wind power group prediction requires the power prediction results of all wind WFs in the output space. Because there is a certain correlation between the output of WFs with close spatial distribution, the utilization of this correlation can effectively improve the training accuracy of the network. Due to the confidentiality of WFs to data, the data of each wind farm are not transmitted to each other but rather uploaded to the dispatch center, so only the dispatch center has the multi-source data uploaded by each WF [22,23]. Based on this consideration, a centralized wind power group prediction model is developed in this paper. Using the data uploaded by each WF as the input, the multi-task learning model is used as a predictor to directly input the power prediction results of all WFs in the space. The output power of each WF is equivalent to a prediction task. The MMoe multi-task learning framework is adopted to directly output the power prediction results of 8 wind farms to reduce the modeling complexity.

2.2. The MMoe Multi-Task Learning Framework

Multiple prediction tasks are generally carried out in a separate modeling way. However, when there is some potential correlation between multiple tasks, if the idea of independent modeling is still adopted, the correlation of multiple problems will be ignored. In this case, the multi-task learning strategy can be used to improve the overall effect. Multi-task learning framework only needs to build a deep learning model, regard multiple related problems as sub-tasks of the model, and control the features in the model, so that multiple tasks can share information, to improve the performance of each task.

The traditional multi-task learning strategy generally adopts the parameter hard-sharing mechanism. The hard-sharing mechanism divides the model into a parameter-sharing layer and subtask-learning layers. The parameter-sharing layer obtains the input features of all tasks and extracts the features [24]. The sub-task learning layer obtains the features related to the task itself from the parameter-sharing layer, then trains the sub-task and outputs the predicted results. Taking a model with 3 outputs as an example, the network model is shown in Figure 1. For each task, all the parameters and structure of the model are the same, only the goal of the final mapping is different. The interaction between all tasks is fixed in the training process and has certain limitations.

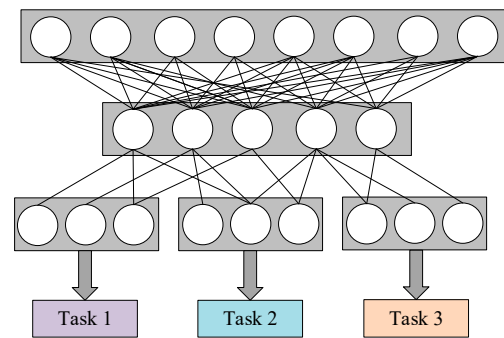


Figure 1. The parameter hard sharing mechanism.

In the hard sharing mechanism, the stronger the correlation of subtasks, the better the model training effect; if some tasks are weakly correlated, the performance of the model may degrade because other tasks are misleading [25]. To this end, Google Inc proposed the MMoE multi-task learning framework, the basic principle of which is shown in Figure 2.

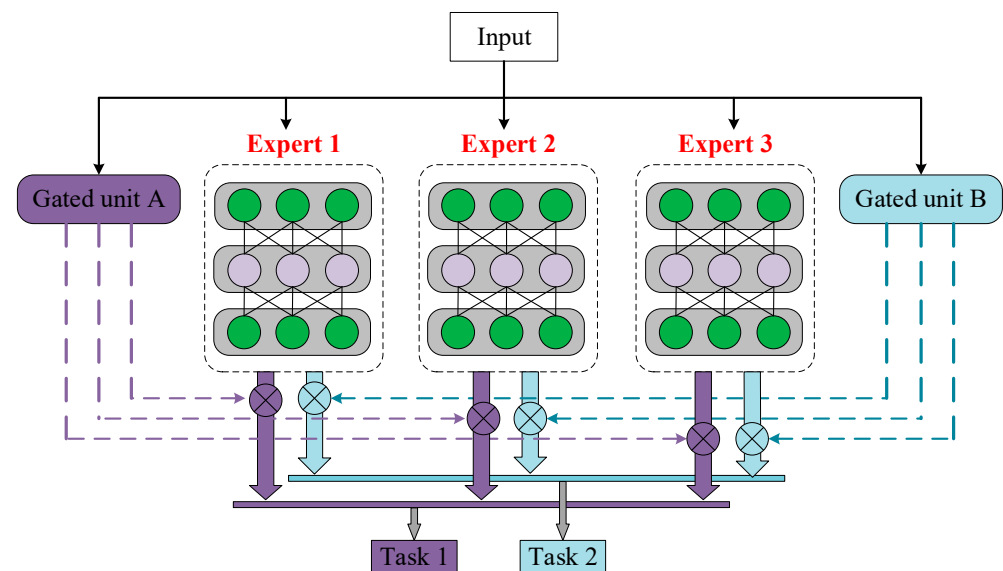


Figure 2. The MMoE multi-task learning framework.

The MMoE algorithm divides the parameter-sharing layer into several expert subnets. Each expert subnetwork is a multi-layer perceptron, which is responsible for independently learning the coupling relationship between multiple tasks. Different expert subnets do not share parameters. At the same time, the MMoE algorithm sets a gate unit for each subtask, which is responsible for calculating the weight of each expert subnet in the subtask so that different tasks can choose expert subnets more flexibly and avoid the mutual interference between weakly related tasks. The output y_k of the k th subtask in the MMoE multi-task learning model can be expressed as Formulas (1) and (2):

$$y_k = \sum_{i=1}^n g_i^k(x) f_i(x) \quad (1)$$

$$g_i^k(x) = \text{softmax}(W_g^{k,i} x) \quad (2)$$

where n is the number of expert subnets, k indicates the task number, x is the input feature of the model, $g_i^k(x)$ is the weight of expert subnet i in the task k , $f_i(x)$ is the output of expert subnet i , $W_g^{k,i}$ is the linear transformation matrix of the i th expert subnet corresponding to the k th gated unit, and $\text{softmax}()$ stands for activation function.

The gate control unit maps the input features to the dimension through linear transformation and obtains the weight coefficient of each expert subnet through the softmax activation function, which realizes the flexible control of expert subnet output in the task. Due to the weak correlation between multi-time scale power prediction, the MMoE multi-task learning model can separately train the weight coefficient of the expert subnet for the prediction of 8 WFs subtasks on the premise of sharing multi-time scale power information to ensure that each subtask can learn the most effective information during network training. Therefore, the MMoE algorithm is more suitable for modeling.

2.3. Bidirectional Gated Recurrent Unit

The structure of the GRU is shown in Figure 3. The GRU structure contains two gates. The reset gate r determines how to combine the new input information with the previous memory, and the update gate z controls how much information from the past is passed to the future; the network parameters of GRU are shown in Formula (3):

$$\begin{cases} r_t = \sigma(i_t W_{xr} + h_{t-1} W_{hr} + b_r) \\ z_t = \sigma(i_t W_{xz} + h_{t-1} W_{hz} + b_z) \\ \tilde{h}_t = \tanh(i_t W_{xh} + r_t \odot h_{t-1} W_{hh} + b_n) \\ h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h}_t \end{cases} \quad (3)$$

where i_t represents the input of the current state, and $\tilde{h}_t$ and y_t represent the output of the current state and the input of the next neuron, respectively. W_{xr} , W_{xz} , and W_{hh} represent the three parameter matrices of the previous hidden layer states h_{t-1} through r_t , z_t , and h_t , respectively.

The structure of the recurrent neural network with GRU as neurons is shown in Figure 4. The backpropagation network is used to train and calculate the parameters of RNN. It is described in Formula (4):

$$o_t = Vf(Ux_t + Wf(Ux_{t-1} + Wf(Ux_{t-2} + Wf(Ux_{t-3} + \dots)))) \quad (4)$$

where o_t represents the output of a single neuron, x_t represents the input at the time t , and U , V , and W represent the weight matrix of the x , h , and output layers, respectively.

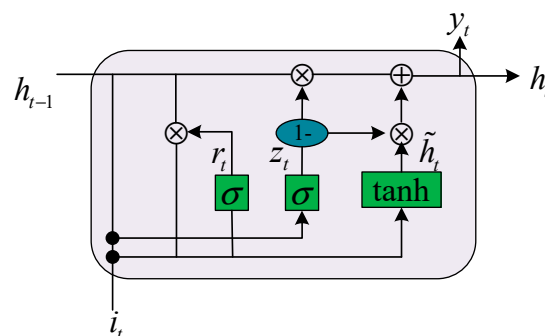


Figure 3. The internal structure of the GRU.

For a GRU-based RNN structure, as shown in Figure 4, if independent GRU neural networks are connected, the structure is developed into a Bi-GRU. This structure enables the GRU neural network to process sequence inputs in two directions, both forward and backward with two separate hidden layers. Each hidden layer can capture both past (forward) and future (backward) data information. This bidirectional structure increases the capacity and flexibility of the model and improves the feature extraction capability of the network for time series [26,27].

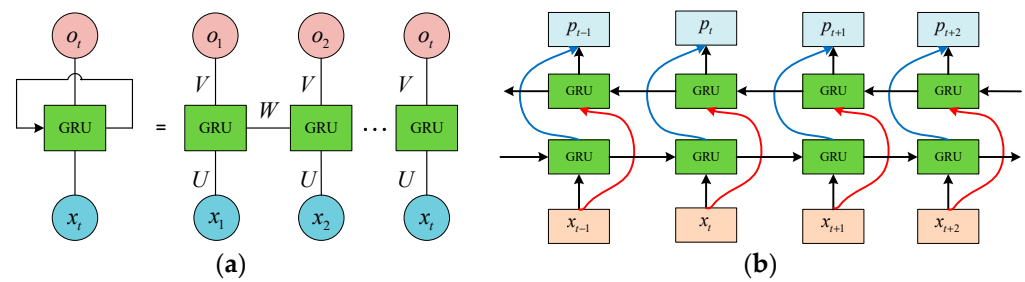


Figure 4. Structure of GRU and Bi_GRU. (a) Architecture of GRU-based RNN; (b) structure of bidirectional GRU.

The research objective of this paper proposed a short-term wind power somatic prediction model for all wind farms in the wind power cluster and output the prediction results of all WFs simultaneously. The model is realized by the combination of a multi-task learning framework and bidirectional GRU network. The Bi_GRU model is the base learner, and the correlation between tasks in the MMoE framework adjusts the learning parameters.

3. Technical Route

Based on multi-task learning technology, this paper designs an STWPP model for WFs with different output spaces at the same time. The technology roadmap is shown in Figure 5:

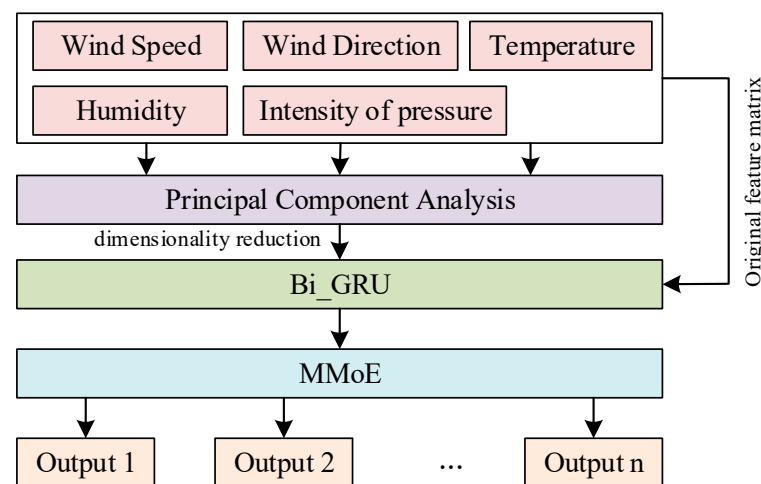


Figure 5. Technology roadmap.

The detailed steps are as follows:

- (1) The WS, WD, temperature, humidity, pressure, and other attributes of different WFs are collected to form a spatial feature matrix, which is used as the original input of the model. Based on the above features, the principal component analysis algorithm is used to reduce the dimensionality of each attribute, and the combination of the original feature matrix and the space matrix after dimensionality reduction is taken as the input.
- (2) A multi-output model based on the MMoE framework is designed, in which Bi_GRU is used as the base learner.
- (3) The dataset is divided into the training set, verification set, and test set. The wind power group prediction model is trained on the training set, the network parameters are fine-tuned by the validation set, and the performance of the model is tested on the test set.

4. Experimental Analysis

4.1. Dataset and Network Parameters

The data from eight wind farms in Jilin, China, were used to conduct simulation experiments. The data included historical power data of actual measurement of wind farms and numerical weather forecast data, with a data resolution of 15min and a time span from 1 January 2018 to 31 December 2018. The NWP includes six attributes of WS, WD, temperature, humidity, pressure, and momentum flux. Where WS represents the predicted wind speed of 100 m high, WD represents the predicted wind direction of 100 m high, and temperature, humidity, pressure, and momentum flux represent the predicted amount of temperature, humidity, pressure, and momentum flux in NWP, respectively. The WD is normalized by the cosine trigonometric function. The installed capacity of the single typhoon generator unit is 1.5 MW, the inlet WS is 3 m/s, and the hub height is 85 m. The data from January to August are the training set, the data from September to October are the verification set, and the data from November to December are the test set. The input data are processed by a sliding window; the length of the time window is 16. The experiments in this paper are completed in python3.8, and the deep learning framework is the Tensorflow(CPU). The computer hardware parameters are as follows {CPU: Intel(R) Core(TM) i5-7300HQ CPU @ 2.50 GHz 2.50 GHz; RAM: 16.0 GB}.

The network built in this paper contains eight outputs, which can realize the synchronous output of eight WFs. The network structure parameters are shown in Table 1; the network contains a total of eight tasks, that is, the network simultaneously outputs the short-term power prediction results of eight WFs. In the multi-task learning layer, the expert subnet parameters of the MMoE multi-task learning model are hyperparameters. The network consists of four expert subnets with 32 neurons. Each expert subnet is used to learn the specific relationship between input features and eight prediction tasks. The subtask layer includes three layers of Bi_GRU layer, the number of neurons is 16 and 8, respectively, and the activation function is “Relu”, which is used to capture the long-distance dependence of the sequence. Dropout layers are added between the three Bi_GRU layers to force the random deactivation of neurons with a random deactivation ratio of 0.2, which is used to prevent the overfitting of the network. The end of the Bi_GRU layer is the fully connected layer, and the load prediction results are the output combined with the linear activation function. During network training, the weights of the eight tasks in the loss function are equal, the optimizer is Adam, the loss function is the mse loss, and the number of iterations is 500. The early stop strategy is adopted in the training process, that is, the training process is stopped if the loss does not decrease in the course of several successive iterations.

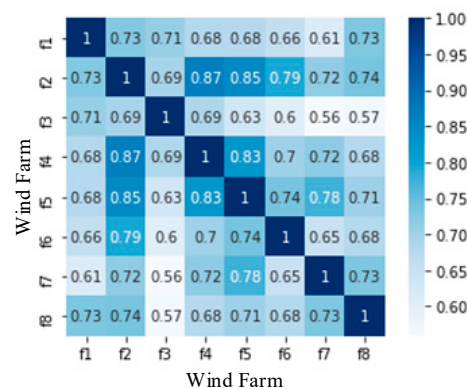
Table 1. Network parameters.

Name of Parameter			Parameter Values	
Network structure	Multi-task layer	MMoE	Number of neurons	32
			Number of experts	4
			Number of tasks	8
	Subtask layer	Bi_GRU	Number of neurons	16
		Dropout	Ratio	0.2
		Bi_GRU	Number of neurons	32
		Dropout	Ratio	0.2
		Bi_GRU	Number of neurons	16
		Dense	Number of neurons	1

Table 1. *Cont.*

Name of Parameter		Parameter Values
Network parameter	Activation function	Relu
	Last layer activation function	Linear
	Multi-task loss function weights	Average
	optimizer	Adam
	Number of iterations	5000
	Loss function	mse

The correlation coefficient of the power of each WF is calculated on the training set, and the results are shown in Figure 6. Wind farms and their numbers are simply “f” + “number”. The correlation coefficient between the power of f4 and f2 is the largest, reaching 0.87; the correlation coefficient between f3 and f7 is the smallest, reaching 0.56. The correlation coefficient represents the degree of linear correlation between the two sequences, and the greater the correlation coefficient, the more similar the two sequences are. Table 2 provides the quantitative index of correlation level; there is a significant correlation between the power of the eight WFs. The strong correlation of each wind farm power reflects the coupling between each task, and joint training helps to improve the prediction accuracy.

**Figure 6.** Power correlation coefficients among wind farms.**Table 2.** The corresponding relationship between the correlation coefficient and correlation degree.

I	Degree of Correlation
<0.3	Weak correlation
[0.3, 0.5)	Low correlation
[0.5, 0.8)	Significant correlation
[0.8, 1]	Strong correlation

4.2. Error Evaluation Index

The normalized root mean square error (RMSE), normalized mean absolute error (MAE), and normalized mean error percentage (MEP) were used to evaluate the performance of the proposed wind power prediction model. The root mean square error is calculated as shown in Formula (5):

$$R_{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{nCap^2}} \quad (5)$$

where y_i represents the real power at time i , $\hat{y}_i$ represents the predicted power, n represents the length of the test set, and Cap represents the actual start-up capacity of the WF. Since

the actual start-up capacity of the WF is difficult to obtain, the installed capacity is used instead in this paper.

The mean absolute error R_{MAE} is calculated as shown in Formula (6):

$$R_{MAE} = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i| \times 100\%}{Cap} \quad (6)$$

The mean error percentage R_{MEP} is as shown in Formula (7):

$$R_{MEP} = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \times 100\% \quad (7)$$

To maintain the same dimension and improve computing efficiency, the maximum-minimum normalization method is adopted to normalize the input and output data. The normalization principle is shown in Formula (8):

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (8)$$

where x' represents the normalized eigenvector, and $x_{\max}$ and $x_{\min}$ represent the maximum and minimum values, respectively. In the prediction stage, the prediction results are restored to the original power interval according to the inverse normalization formula. The inverse normalization principle is shown in Formula (9):

$$x = x'(x_{\max} - x_{\min}) + x_{\min} \quad (9)$$

4.3. Experimental Results

The NWP provided by each WF includes six attributes of wind speed, wind direction, temperature, humidity, pressure, and momentum flux, and eight WFs have a total of 48 attributes. To reduce the dimensionality of the data, principal component analysis is used to reduce the dimensionality of the data. The dimensionality reduction strategy is not to reduce the dimensionality of 48 attributes at the same time but to reduce the dimensionality of the attributes of each NWP, respectively, and then the data after dimensionality reduction are merged.

Since a total of eight wind farms are included, each meteorological attribute in the data contains eight features. The principal component analysis algorithm is used to extract the features of the data, and the number of principal components is set to eight. The principal component contribution rate is sorted by dimensionality reduction, and the sorting result is shown in Figure 7:

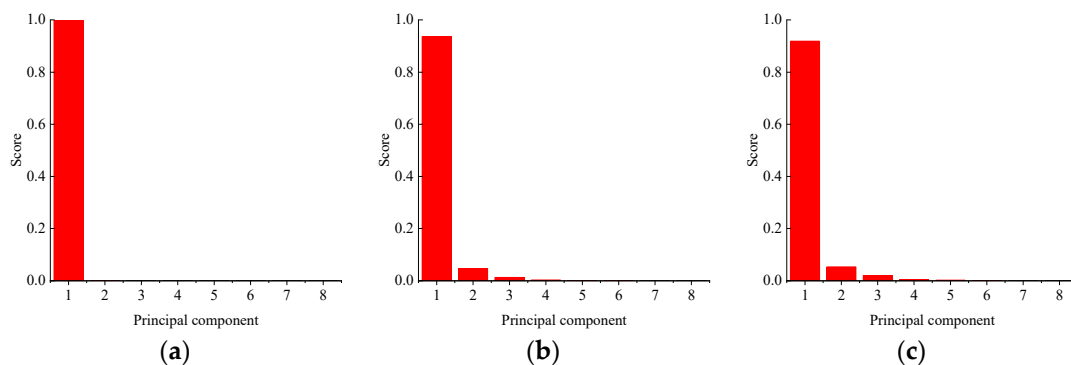


Figure 7. Cont.

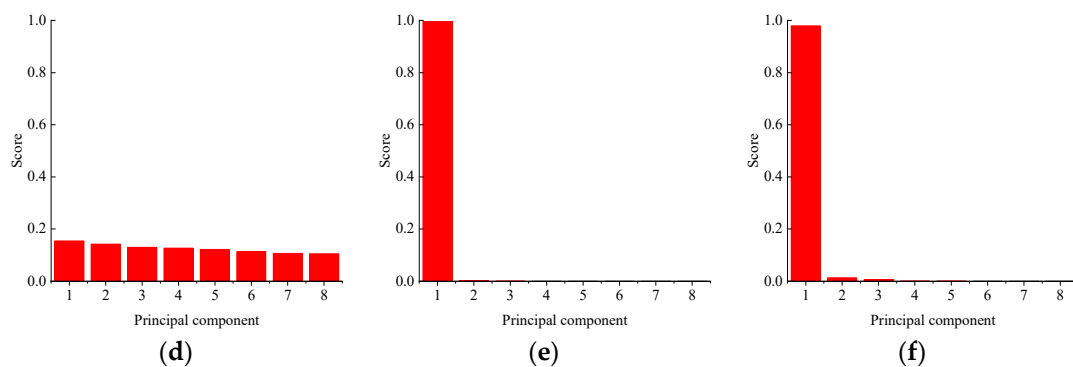


Figure 7. Principal component contribution of each attribute. (a) Temperature; (b) momentum flux; (c) WS; (d) WD; (e) atmospheric pressure; (f) humidity.

A value of 95% is taken as the threshold of principal component retention, in which temperature retains the first principal component, momentum flux retains the first two principal components, WS retains the first two principal components, air pressure takes the first principal component, and humidity takes the first principal component. Because the principal component of WD is too divergent, it is considered that the principal component analysis algorithm does not affect the wind direction, so the WD after dimensionality reduction is not considered. The features after dimensionality reduction and the original features are used as the input of the MMoE algorithm.

The prediction model proposed in this paper is named MMoE-PCA-Bi-GRU. The predicted power curves of each wind farm are shown in Figure 8. To show clearly, the predicted power curves of 1 November 2018–7 November 2018 are shown. Combined with the analysis of Figure 8a–i, the wind power group prediction framework proposed in this paper can well predict the wind power trend, and the prediction curve can follow the actual power curve. By adding the power prediction results of all WFs, the cluster wind power prediction results are obtained, as shown in Figure 8i. The peaks and troughs can be well predicted.

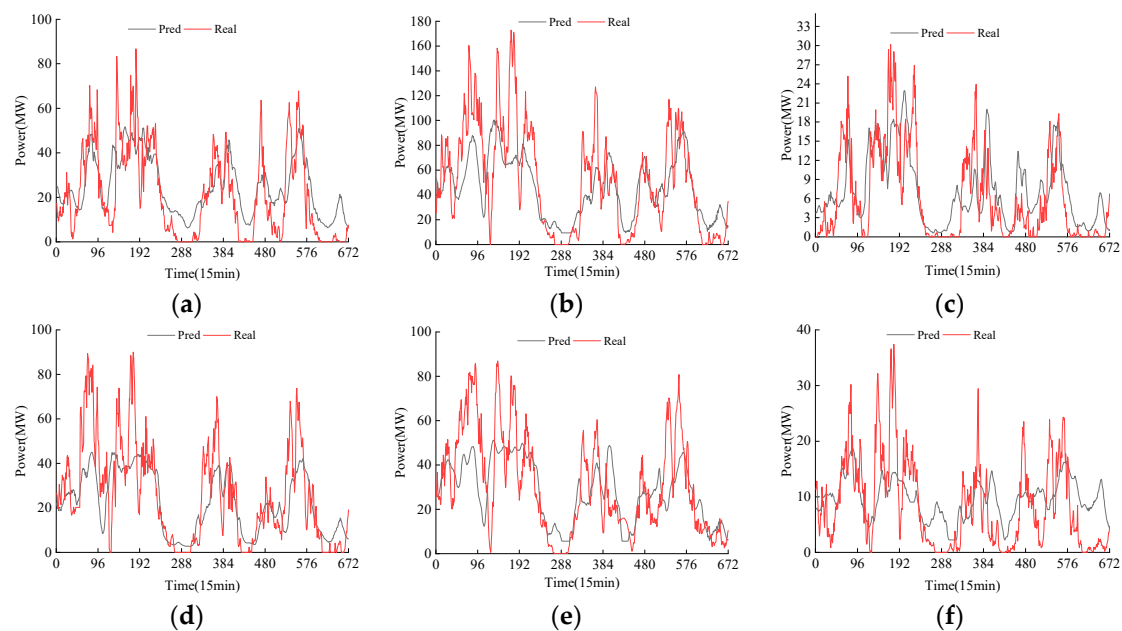


Figure 8. Cont.

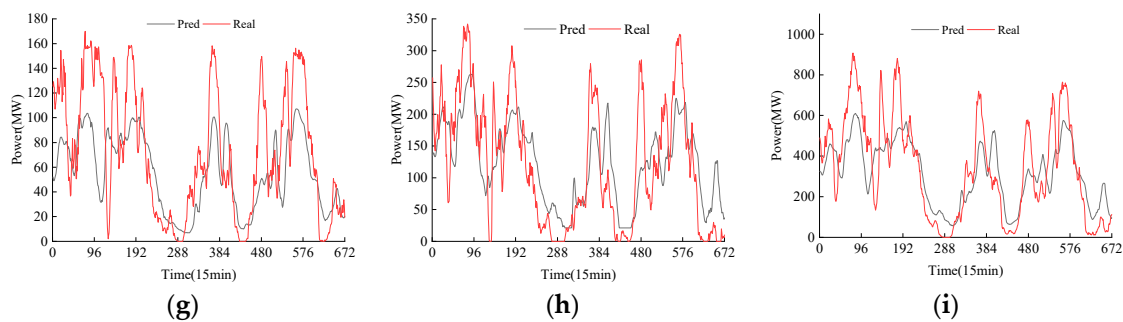


Figure 8. Prediction power curve. (a) Farm 1; (b) Farm 2; (c) Farm 3; (d) Farm; (e) Farm 5; (f) Farm 6; (g) Farm 7; (h) Farm 8; (i) cluster.

The statistical error indicators of eight wind farms and wind power clusters in November and December are shown in Table 3. Among them, the prediction accuracy of the wind farm on 5 November was low, but the prediction error in December significantly decreased. The 7 December wind farm had the highest forecast error, but the November forecast was more accurate than the December 1, 2, and 5 WFs. On average, the eight wind farms had an average RMSE of 0.1760 and an average MAE of 0.1738 in November and an average RMSE of 0.1367 and an average MAE of 0.1325 in December, all below 20% of the installed capacity. From the analysis of the wind power cluster composed of eight WFs, the predicted RMSE and MAE of the whole wind power cluster are lower than 15% of the installed capacity and 12% of the installed capacity, indicating that the MMoE-PCA-Bi_GRU model is valid. The MEP indicator reflects the predicted error as a percentage of the actual power, the average error deviation in November is higher, reaching 20.32%, and the maximum MEP of a single WF is 27.71%. The average MEP in December reached 18.65%, and the maximum MEP for a single WF is 23.32.

Table 3. Prediction indexes of MMoE-PCA-Bi_GRU model.

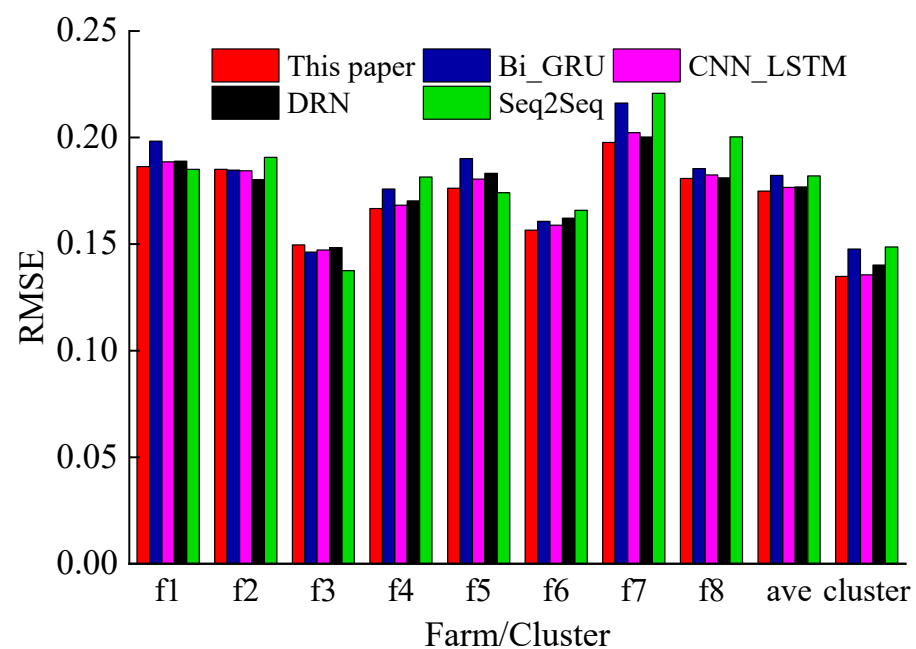
Farm/Cluster	RMSE		MAE		MEP(%)	
	November	December	November	December	November	December
Farm 1	0.1769	0.1959	0.1404	0.1476	21.22	23.32
Farm 2	0.1763	0.1939	0.1361	0.1441	20.83	22.18
Farm 3	0.1575	0.1417	0.1165	0.1021	18.69	16.84
Farm 4	0.1872	0.1461	0.1399	0.1146	19.17	16.03
Farm 5	0.1970	0.1554	0.1553	0.1211	27.71	16.98
Farm 6	0.1595	0.1536	0.1274	0.1159	18.46	16.34
Farm 7	0.1739	0.2214	0.1338	0.1731	18.04	18.56
Farm 8	0.1793	0.1822	0.1445	0.1415	18.41	18.97
Average	0.1760	0.1738	0.1367	0.1325	20.32	18.65
Cluster	0.1394	0.1302	0.1114	0.1023	16.43	16.02

Take traditional Bi_GRU as the prediction model, and the parameters were set the same as those of MMoE-PCA-Bi_GRU. Each WF established its prediction model, and the prediction results are shown in Table 4. Among them, the MMoE-PCA-Bi_GRU model is superior to the traditional Bi_GRU model for the predicted RMSE and MAE of all wind farms. However, in some cases, the performance of the MMoE-PCA-Bi_GRU model is not as good as the traditional Bi_GRU model, such as the RMSE of the MMoE-PCA-Bi_GRU model in December of the No. 1 wind farm, but from the global performance, the performance of the MMoE-PCA-Bi_GRU model is better. The single wind farm MEP of the Bi_GRU model in November was 22.56%, the average MEP was 19.8%, and the maximum MEP of the single WF in December was 29.38%, among which the maximum MEP of the single WF in December was larger. For the whole wind power cluster, the MEP of the Bi_GRU model on the test set is higher than that of the model proposed in this paper.

Table 4. Prediction indexes of traditional Bi_GRU model.

Farm/Cluster	RMSE		MAE		MEP	
	November	December	November	December	November	December
Farm 1	0.2077	0.1889	0.1655	0.1442	22.56	20.88
Farm 2	0.1818	0.1875	0.1384	0.1375	18.59	19.03
Farm 3	0.1549	0.1377	0.1238	0.1068	16.54	15.58
Farm 4	0.2008	0.1509	0.1542	0.1156	20.35	16.87
Farm 5	0.2184	0.1618	0.1690	0.1256	20.79	17.89
Farm 6	0.1672	0.1542	0.1308	0.1128	17.03	17.25
Farm 7	0.1927	0.2396	0.1497	0.1868	21.22	29.38
Farm 8	0.1837	0.1871	0.1461	0.1424	21.32	20.16
Average	0.1837	0.1760	0.1472	0.1340	19.80	19.63
Cluster	0.1547	0.1406	0.1220	0.1078	0.1673	0.1639

With RMSE as the evaluation index, the performance of the proposed model is compared with other deep learning models, including the CNN–LSTM model, Seq2Seq model, deep residual network (DRN), and Bi_GRU model. The matching results are shown in Figure 9:

**Figure 9.** RMSE comparison of different models.

When taking a single wind farm as the target, the prediction error of the model proposed in this paper is higher for the No. 3 wind farm, but for other wind farms, the model proposed in this paper is lower. The MMoE–PCA–Bi_GRU model has the lowest average error and cluster error for eight WFs, which verifies the validity of the proposed model.

5. Conclusions

A wind power group prediction model is proposed to directly output the STWPP results of all WFs in the region, which improves the modeling efficiency. The conclusions are as follows.

- The principal component analysis algorithm is used to extract features from meteorological data of multiple wind farms, and the dimensionality of the data is reduced

from 48 to 8 dimensions by screening the principal component components, which reduces the complexity of the model.

- The STWPP of the wind power cluster is designed based on multi-task learning, and the power prediction sequence of all wind farms in the output region is synchronized, which simplifies the modeling complexity.
- The average RMSE of the MMoE–PCA–Bi_GRU model for eight wind farms is 0.1754; compared with the model predicted by each wind farm separately, the prediction precision has been significantly improved.

The model proposed in this paper is more suitable for small- and medium-sized wind power clusters. When the wind power cluster covers a wider area and more target tasks are predicted, each task is difficult to converge at the same time, and the model proposed in this paper is difficult to apply. In future studies, large-scale graph computation and distributed training algorithms will be introduced to improve the ability of the model for spatio-temporal feature extraction, and it is suitable for large-scale wind power cluster power prediction tasks.

Author Contributions: M.Y. collected data and built the model, D.W. analyzed the experimental results, and W.Z. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by National Key R&D Program of China (Multi-timescale Forecast Technology for Large-scale Wind/Photovoltaic Power Supply Capability) (2022YFB2403000).

Data Availability Statement: The authors are not authorized to share these data.

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

References

1. International Energy Agency. Renewables 2019. Available online: <https://www.iea.org/reports/renewables-2019> (accessed on 15 July 2023).
2. Yang, M.; Shi, C.; Liu, H. Day-ahead wind power forecasting based on the clustering of equivalent power curves. *Energy* **2021**, *218*, 119515. [CrossRef]
3. Lu, P.; Ye, L.; Zhao, Y. Review of meta-heuristic algorithms for wind power prediction: Methodologies, applications and challenges. *Appl. Energy* **2021**, *301*, 117446. [CrossRef]
4. Ma, Z.; Mei, G. A hybrid attention-based deep learning approach for wind power prediction. *Appl. Energy* **2022**, *323*, 119608. [CrossRef]
5. Alkesaiberi, A.; Harrou, F.; Sun, Y. Efficient wind power prediction using machine learning methods: A comparative study. *Energies* **2022**, *15*, 2327. [CrossRef]
6. Potter, C.W.; Negnevitsky, M. Very short-term wind forecasting for Tasmanian power generation. *IEEE Trans. Power Syst.* **2006**, *21*, 965–972. [CrossRef]
7. Tawn, R.; Browell, J. A review of very short-term wind and solar power forecasting. *Renew. Sustain. Energy Rev.* **2022**, *153*, 111758. [CrossRef]
8. Xue, Y.; Lei, X.; Xue, F. A Review on Impacts of Wind Power Uncertainties on Power Systems. *Proc. CSEE* **2014**, *34*, 5029–5040.
9. Rajagopalan, S.; Santoso, S. Wind power forecasting and error analysis using the autoregressive moving average modeling. In Proceedings of the 2009 IEEE Power & Energy Society General Meeting, Calgary, AB, Canada, 26–30 July 2009; pp. 1–6.
10. Billinton, R.; Chen, H.; Ghajar, R. A sequential simulation technique for adequacy evaluation of generating systems including wind energy. *IEEE Trans. Energy Convers.* **1996**, *11*, 728–734. [CrossRef]
11. Foley, A.M.; Leahy, P.G.; Marvuglia, A. Current methods and advances in forecasting of wind power generation. *Renew. Energy* **2012**, *37*, 1–8. [CrossRef]
12. Hanifi, S.; Liu, X.; Lin, Z. A critical review of wind power forecasting methods—Past, present and future. *Energies* **2020**, *13*, 3764. [CrossRef]
13. Wang, Y.; Zou, R.; Liu, F. A review of wind speed and wind power forecasting with deep neural networks. *Appl. Energy* **2021**, *304*, 117766. [CrossRef]
14. Rayi, V.K.; Mishra, S.P.; Naik, J. Adaptive VMD based optimized deep learning mixed kernel ELM autoencoder for single and multistep wind power forecasting. *Energy* **2022**, *244*, 122585. [CrossRef]
15. Xing, Z.; Qu, B.; Liu, Y. Comparative study of reformed neural network based short-term wind power forecasting models. *IET Renew. Power Gener.* **2022**, *16*, 885–899. [CrossRef]
16. Du, P.; Wang, J.; Yang, W. A novel hybrid model for short-term wind power forecasting. *Appl. Soft Comput.* **2019**, *80*, 93–106. [CrossRef]

17. Duan, J.; Wang, P.; Ma, W. Short-term wind power forecasting using the hybrid model of improved variational mode decomposition and Correntropy Long Short-term memory neural network. *Energy* **2021**, *214*, 118980. [[CrossRef](#)]
18. Zhang, Y.; Li, Y.; Zhang, G. Short-term wind power forecasting approach based on Seq2Seq model using NWP data. *Energy* **2020**, *213*, 118371. [[CrossRef](#)]
19. Ding, M.; Zhou, H.; Xie, H. A gated recurrent unit neural networks based wind speed error correction model for short-term wind power forecasting. *Neurocomputing* **2019**, *365*, 54–61. [[CrossRef](#)]
20. Ko, M.; Lee, K.; Kim, J.K. Deep Concatenated Residual Network with Bidirectional LSTM for One-Hour-Ahead Wind Power Forecasting. *IEEE Trans. Sustain. Energy* **2020**, *12*, 1321–1335. [[CrossRef](#)]
21. Mei, Y.; Zhuo, Z.; Xlac, D. Superposition Graph Neural Network for offshore wind power prediction. *Future Gener. Comput. Syst.* **2020**, *113*, 145–157.
22. Yang, M.; Wang, D.; Zhang, W. A short-term wind power prediction method based on dynamic and static feature fusion mining. *Energy* **2023**, *280*, 128226. [[CrossRef](#)]
23. Yang, M.; Wang, D.; Xu, C. Power transfer characteristics in fluctuation partition algorithm for wind speed and its application to wind power forecasting. *Renew. Energy* **2023**, *211*, 582–594. [[CrossRef](#)]
24. Kumari, R.; Ashok, N.; Ghosal, T. Misinformation detection using multitask learning with mutual learning for novelty detection and emotion recognition. *Inf. Process. Manag.* **2021**, *58*, 102631. [[CrossRef](#)]
25. Ma, J.; Zhao, Z.; Yi, X. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 1930–1939.
26. Liu, H.; Han, H.; Sun, Y. Short-term wind power interval prediction method using VMD-RFG and Att-GRU. *Energy* **2022**, *251*, 123807. [[CrossRef](#)]
27. Liu, X.; Yang, L.; Zhang, Z. Short-term multi-step ahead wind power predictions based on a novel deep convolutional recurrent network method. *IEEE Trans. Sustain. Energy* **2021**, *12*, 1820–1833. [[CrossRef](#)]

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.