

# Combined Optimization Prediction Model of Regional Wind Power Based on Convolution Neural Network and Similar Days

Yalong Li \*, Fan Yang, Wenting Zha and Licheng Yan

School of Mechanical Electronic and Information Engineering, China University of Mining and Technology-Beijing, Beijing 100083, China; yangedu0420@163.com (F.Y.); wtzha@seu.edu.cn (W.Z.); wzdsylc@163.com (L.Y.)

\* Correspondence: lylwyyx@163.com; Tel.: +86-10-62331257

Received: 10 October 2020; Accepted: 18 November 2020; Published: 20 November 2020



**Abstract:** With the continuous optimization of energy structures, wind power generation has become the dominant new energy source. The strong random fluctuation of natural wind will bring challenges to power system dispatching, so it is necessary to predict wind power. In order to improve the short-term prediction accuracy of regional wind power, this paper proposes a new combination prediction model based on convolutional neural network (CNN) and similar days analysis. Firstly, the least square fitting and batch normalization (BN) are used to preprocess the data, and then the recent historical wind power data set for CNN is established. Secondly, the Pearson correlation coefficient and cosine similarity combination method are utilized to find similar days in the long-term data set, and the prediction model based on similar days is constructed by the weighting method. Finally, based on the particle swarm optimization (PSO) method, a combined forecasting model is established. The results show that the combined model can accurately predict the future short-term wind power curve, and the prediction accuracy is improved to different extents compared to a single method.

**Keywords:** wind power prediction; combination model; convolutional neural network; similar days; optimization algorithm

## 1. Introduction

In recent years, the new energy input, dominated by wind power generation, is increasing rapidly. The strong volatility and randomness of natural wind make it difficult to accurately predict wind power, which brings new problems for the safe and stable operation of power grids [1,2]. More and more researchers have been committed to investigating the best model to predict the wind power. Generally speaking, model complexity is an index to measure the quality of the model and the simple one is preferred. Traditional methods [3,4], such as the autoregressive moving average model (ARMA) and grey model, cannot dig deeper into the law, which leads to low prediction accuracy. Thus, the common way for high accuracy is to increase the complexity of the model. With the continuous development of artificial intelligence technology, wind power prediction based on machine learning [5–8] and deep learning [9–13] has become a hot topic, compared to the traditional physical and statistical methods, and the prediction accuracy has been greatly improved. Due to the complexity of the structures, the models have the ability to process a large number of data and to better reflect the non-linear and non-stationary nature of wind power. The work in Reference [5] used the Markov model divided residual error into several different states, the residual values were further forecasted through calculating the probability distribution of each state, the residual errors were used to correct

the predicted values to improve the accuracy of wind speeds, and the results showed that the Markov model can improve the short-term forecasting accuracy of wind power effectively. The work in Reference [6] used the K-nearest neighbor (KNN) algorithm to select historical wind speed points and built an SVR model to obtain the fusion wind speed, which improves the accuracy of the wind speed prediction. A prediction model based on the Long short-term memory network (LSTM) was proposed in Reference [9], which used NWP prediction data for prediction. The prediction accuracy has some advantages over the traditional network, but depends on the accuracy of numerical weather prediction (NWP) itself. Two hidden layers in the Gated Recurrent Unit network were used in Reference [10] to forecast the wind power, and the results show that the proposed model has higher prediction accuracy than the Autoregressive moving average (ARMA) model and the LSTM model. Even though the above research can meet the basic goal of wind power prediction, it is still easy to fall into the limitations of the model itself, which can be supplemented by the optimization algorithm or other models to further improve the prediction effect.

Therefore, the method of multi model combination was introduced to overcome the limitations of the single model [14–21]. The work in Reference [14] proposed a combination model of the support vector regression model and grey model, and the result showed that the combination model can improve the prediction accuracy. The work in Reference [15] used CNN to extract information from input data and integrated the Light Gradient Boosting Machine (LightGBM) algorithm, and the combination model has better performance in terms of the accuracy and efficiency of the single model. A prediction method of combining empirical mode decomposition (EMD) and D-LSTM was introduced in Reference [16], which used EMD to decompose wind power curve and weighted combination after component prediction of D-LSTM; the results show that the combined model has better prediction ability. The work in Reference [17] verified that the ensemble learning method had better prediction performance than the traditional independent method; ensemble learning reduced the overall prediction error and has the capacity to merge various models. The above research shows that the multi model combination better than the single one due to the complementary advantages. However, there are still some shortcomings of combination weight subjectivity and multi model learning time dimension overlap.

In addition, according to the actual situation of power grid operation, the wind power prediction of a single wind farm cannot meet the dispatching needs, and the importance of regional wind power prediction is gradually reflected. Regional wind power prediction refers to the prediction of the total wind power of each wind farm in the same region as a whole. The regional prediction method emphasizes the regularity of the total wind power in the prediction process, which effectively reduces the amount of information needed and the interference of local random factors of individual wind farms. It is of great practical significance for the power sector to optimize the power system dispatching, and reduces the rotating reserve capacity and the operating cost of the power system. Physical methods are no longer applicable because it is difficult to obtain all physical parameters in the whole region. Moreover, the statistical method and the traditional artificial intelligence method cannot dig out the rule of historical wind power, so the prediction effect is not good. Therefore, the problem of the regional wind power prediction has not been effectively solved yet.

This paper further studies the influence of combined weights on multi model prediction methods, and proposes a combined optimization prediction model of regional wind power based on CNN and similar days. First, the recent data after preprocessing is used as the input of CNN model, which is able to discover the short-term rule existing in the historical data. Second, the Pearson correlation coefficient and cosine similarity are used to select the most similar days and search for the long-term rule in all historical data. Finally, the optimal combination weight is selected by the PSO algorithm to obtain the predicted results, which are analyzed by examples.

## 2. Short-Term Law Prediction Based on CNN

The prediction goal is to achieve the prediction of  $x_{t+1}$  to  $x_{t+\sigma}$  in advance at the time of  $t$ . The time length of the forecast interval is called the forecast period and the prediction period of short-term prediction is generally in hours. In this paper, the sampling interval of regional wind power data is 15 min, and the prediction period is 4 h, which means that  $\sigma$  is taken as 16. The recent data used for CNN model prediction is defined as  $[x_{t-a^2+1}, x_{t-a^2+2}, \dots, x_t]$ , where  $a$  is the side length of the input matrix.

### 2.1. Data Process

Interference points exist in the wind power curve because the natural wind has the characteristics of fluctuation and intermittence, and the total output of multiple wind farms will introduce the phenomenon of offsetting or aggravating the fluctuation. In order to eliminate abnormal fluctuation points, the  $K$ -order least square curve fitting method is used to smooth the wind power curve and eliminate certain noise in recent data sets. The fitting objective function can be defined as follows:

$$\min \sum_{i=1}^m |p(x_i) - y_i|$$

$$K = \deg[p(x)] \quad (1)$$

where  $m$  is the number of fitting continuous time points,  $p(\cdot)$  is the fitting polynomial function,  $y_i$  is actual wind power curve, and  $K$  is the degree of  $p(\cdot)$ .

On this basis, the fitting data are normalized and made into matrices to meet the input data structure of the CNN model. Normalization can unify the wind power value into (0, 1) for calculation and also help to find the global optimal solution through gradient descent when the error propagates in the back direction. The common normalization methods are linear normalization, Z-score normalization and sigmoid normalization in order to avoid constraining the distribution of data and weakening the expression ability of the network. This paper utilizes the batch normalization (BN) [22] algorithm proposed by Google in 2015. The input data is transformed into new data, satisfying normal distribution, and the parameters of translation and scaling that can be learned are added. The conduction process formula is defined as follows:

$$\mu_B \leftarrow \frac{1}{n} \sum_{i=1}^n x_i$$

$$\sigma_B^2 \leftarrow \frac{1}{n} \sum_{i=1}^n (x_i - \mu_B)^2$$

$$\hat{x}_i \leftarrow \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}}$$

$$y_i \leftarrow \gamma \hat{x}_i + \beta \triangleq BN_{\gamma, \beta}(x_i) \quad (2)$$

where  $B = \{x_1, x_2, \dots, x_n\}$ ,  $\mu_B$  and  $\sigma_B$  are average value and standard deviation of set  $B$ , which contains  $n$  elements,  $\hat{x}_i$  is the normalized value (mean value is 0, variance is 1),  $\varepsilon$  is a small positive number,  $\gamma$  and  $\beta$  are scaling parameters and translation parameters, respectively, and  $y_i$  is the final normalized result.

### 2.2. CNN Model Structure and Applicability Analysis

CNN is a powerful feedforward neural network, which is good at extracting features from matrix data and has strong data mining ability. The structure diagram is shown in Figure 1, which includes input layer, convolution layer, pooling layer, fully connection layer and output layer. The convolution layer can be regarded as the process of extracting features from data sources, which generally requires multiple convolution structures to work together. Due to data after the convolution layer having larger dimensions, we use the pool layer to segment the feature, so as to reduce the size of the feature graph. Finally, the full connection layer combines the local features into the global features.

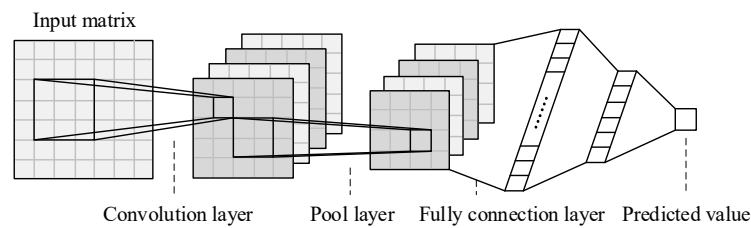


Figure 1. CNN structure diagram.

The operation of the model is divided into training process, verification process and test process, which correspond to the training set, verification set and test set, respectively. The total data set consisted of wind power data in the last two years. The training process is divided into two stages: the forward propagation of data and the backward propagation error. By calculating the error between the output value and the real one, the model parameters will be continuously modified based on the gradient descent method and the training process ends after certain iterations. The verification process is accompanied by the training process. After each iteration, the model calculates the accuracy of the verification set according to the current network parameters, and saves the best ones as the optimal model. In order to guarantee the validity of the experiment, the intersection between the test set and the first two sets is not allowed. In the prediction process, the test set data is imported into the optimal model and its output is the prediction value of the CNN model.

In this paper, the CNN model is a single output structure, using the recursive prediction method to achieve 16 points of continuous prediction. It means that the predicted value is inserted into the last part of the input matrix, and the data points in the front of the input matrix are removed, and the prediction  $x_{n+1}$  to  $x_{n+16}$  can be achieved. After a prediction period is completed, real value  $[x_{n+1}, x_{n+2} \cdots x_{n+\tau}]$  can be introduced after the rolling step  $\tau$ , and the input data of the model can be updated. The method of intraday prediction by introducing the real value through the rolling step is called rolling prediction.

In addition, model hyperparameter optimization is also an important step to establish the network. An effective hyperparameter adjustment algorithm can not only improve the prediction accuracy, but also save the operation cost. In all, the input matrix data source has a great influence on the accuracy of the algorithm, as well as the combination of the convolution layer and pooling layer. Therefore, this paper focuses on data preprocessing and network structure determination when CNN is applied.

### 3. Long-Term Law Prediction Based on Similar Days

All data sets in CNN are wind power matrices arranged in chronological order and the time span of input data should not be too large, which implies that the long-term law cannot be captured. In this paper, the generalized similar day long-term forecast is proposed to make up for this shortcoming, and all the historical data are deepened, so as to effectively capture the law of historical similar data. Divide all data set  $\{x_1, x_2 \cdots x_t\}$  into  $\{D_1, D_2 \cdots D_n\}$ ,  $D_n = [x_{t-nd+1}, x_{t-nd+2} \cdots x_{t-(n-1)d}]$ , where  $d = 96$  is the total number of sampling points in a day. Then, the grouped and divided data are defined as the long-term data set, which are used to search for the generalized similar days. The rolling prediction method is also used in multi-step prediction.

#### 3.1. Similar Day Search

The goal of looking for similar days is to find out the data closest to the trend of wind power on the day (base day) before the forecast day in all historical wind power data, and to use the trend of the next day to revise the forecast results, so as to improve the forecast accuracy. The similarity criteria are as follows:

Pearson correlation coefficient:

$$\rho_{(X,Y)} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad (3)$$

Cosine similarity:

$$\cos(X,Y) = \frac{\sum_{i=1}^n X_i \times Y_i}{\sqrt{\sum_{i=1}^n (X_i)^2} \times \sqrt{\sum_{i=1}^n (Y_i)^2}} \quad (4)$$

Combination similarity:

$$S_{(X,Y)} = \rho_{(X,Y)} + \cos(X,Y) \quad (5)$$

where  $X$  and  $Y$  are the comparison day data and base day data, respectively.

On the basis of the single method, considering that the Pearson correlation coefficient (PCCs) and cosine similarity value interval are the same, we put these two methods together as the combination similarity day selection method. In the actual case, it is necessary to carry out multiple groups of comparative experiments on large-scale data, and the best calculation method can be determined to predict the test set to avoid data leakage.

### 3.2. Applicability Analysis of Similar Days

Similar days mainly take meteorological similarity into account, which means that there is still a certain correlation between the follow-up trends of similar data in the two historical periods. The trend comparison of similar days selected by Formula (5) on a base day is shown in Figure 2.

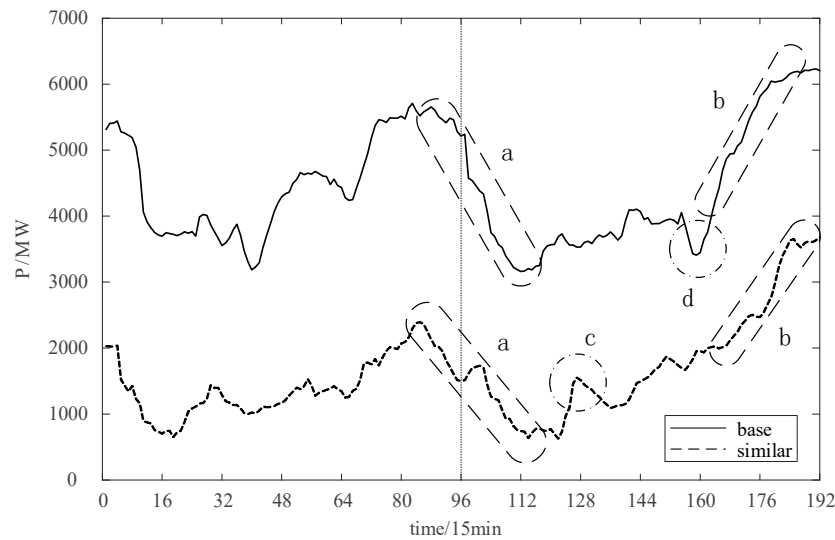


Figure 2. Trend comparison of similar days.

The solid line and dotted line on the left side of Figure 2 are the base day (2019-1-1) and similar day (2019-2-3), respectively, and the trends of these two in the next day are on the right side of the figure. It can be seen that the wind power of the next day in similar days tends to converge in the time interval of downhill (a), climbing (b) and stable period, and the trends have a high degree of correlation. While there are differences in the small peaks (c) and troughs (d), it can still be an effective aid for prediction.

#### 4. Long Short-Term Combined Optimization Prediction

The operation principle of the short-term and long-term prediction model is described above, and the feasibility is analyzed. In some cases, the CNN model can predict well, but in other cases, the prediction accuracy of similar days is higher. In the actual prediction, we require the scientific and rigorous combination method to make the total statistical index optimal in a long time, since we cannot know the best combination weight in advance.

##### 4.1. Select Combination Weighting Method

The combination weighting method can be divided into the subjective weighting method and objective weighting one, where the subjective weighting method is represented by the analytic hierarchy process (AHP). The weights are reasonably evaluated according to the decision-making problems and their own experience. However, the decision-making results are subjective and arbitrary, which means that it is difficult to provide effective and scientific weighting criteria in the wind power prediction problem. Due to these great limitations, this paper focuses on the objective weighting method.

The objective weighting methods for wind power prediction include the equal weight average method, time-varying weight method, entropy weight method, genetic algorithm and particle swarm optimization (PSO). Among them, particle swarm optimization is a random search algorithm based on group cooperation, which has strong global optimization search ability, fast convergence and high accuracy, and can effectively avoid falling into local optimization. The optimization process simulates the behavior of birds' predation, where each particle is a possible solution of search space, and the particle dynamically adjusts the speed and direction of flight according to its historical optimal position and population optimal position. The optimal solution of a single particle search is called the individual extremum, and the optimal one in the particle swarm is called the global extremum. Through continuous iteration, the optimal solution satisfying the termination condition can be obtained.

In addition, the neural network is designed in several works for weight selection, which considerably increases the model complexity and time cost. Considered synthetically, particle swarm optimization is chosen in the experiment as the objective combination weighting method to avoid the subjectivity of artificial weighting or AHP.

##### 4.2. Total Prediction Structure

The general structure of the model is shown in Figure 3. On the left is the short-term law prediction model based on CNN. By using the verification set data, the optimal network structure can be achieved and then the base day data are input into the optimal CNN model, and the CNN predicted results can be obtained through recursive prediction. On the right is the long-term law prediction model of similar days. Similar to the CNN model, the optimal similar day selection algorithm is selected according to the verification set, and the candidate similar days are determined by calculating the similarity. Then, the PSO algorithm is used to determine the combination weight of the verification set, which is used to get the final predicted results.

In the CNN prediction process, we use the recursive prediction method above to forecast the data in the future period without introducing the true value, which belongs to multi-step prediction. After the prediction, the evaluation indexes of root mean square error (*RMSE*), percent error (*PE*) and qualified rate (*QR*) are used. The *PE* indicator is similar to *RMSE*, which evaluates the error between the predicted value and the true value. The *QR* indicator is closely related to power quality, because even if the average error is very low, a certain percentage of large errors will still threaten the operation of the power grid. Parameters *a* and *b* in *QR* are determined according to the specific requirements of the power grid. The difference among them is that *RMSE* reflects the absolute size of the prediction error, *PE* reflects the relative value of the prediction error, and *QR* reflects the compliance

degree of the prediction results. Therefore, the predicted results can be evaluated comprehensively. The formulas for calculating the indicators are as follows:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (\hat{y}_i - y_i)^2} \quad (6)$$

$$PE = \frac{1}{m} \sum_{i=1}^m \left| \frac{\hat{y}_i - y_i}{Z} \right| \quad (7)$$

$$QR = \frac{1}{m} \text{count} \begin{cases} |\hat{y}_i - y_i| \leq ay + b & \text{True} \\ \text{else} & \text{False} \end{cases} \quad (8)$$

where  $m$  is the number of sampling points,  $\hat{y}$  and  $y$  are the predicted value and the real value,  $Z$  is the max power per day,  $a$  and  $b$  are adjustable parameters, determined by the actual dispatching situation of the power grid, and the proportion of *True* is the qualification rate.

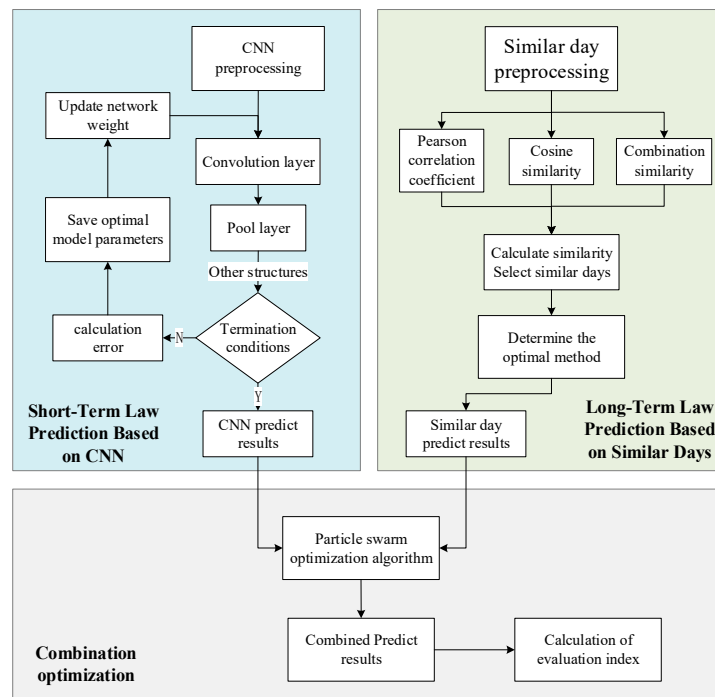


Figure 3. Model structure diagram.

## 5. Case Analysis

Take the wind power data of a regional wind farm as an example. Specifically, there is a prediction system in the Beijing–Tianjin–Tangshan region of China where the maximum wind power is 7446 MW. Through this prediction system, only the real data and predicted data of historical wind power in the last decade are sampled at a sampling frequency of 15 min. Ten years of data from 2010 to 2019 are used in similar day models for long-term law prediction, and the data of the last two years are used in the CNN model for short-term law prediction, so as to achieve the prediction goal of the next 4 h. The feasibility of this method is verified by simulations, and the single CNN method, the single similar day method and the existing level (the prediction accuracy of this prediction system, which has been running for more than three years) are compared using three evaluation indexes to verify the superiority of the combined prediction method.



### 5.1. Select Combination Weighting Method

This paper selects the Tensorflow deep learning framework in Python to build the model. First, determine the  $K$  value of the least square fitting. By testing the  $K$  value in References [10,20], the optimal  $K = 15$  can be achieved. Then, normalize and matrix the fitted data with BN. Take the winter data of the year 2019 as an example, the CNN data set is divided as follows:

In the data set divided in Figure 4, considering the seasonal characteristics of wind power, the validation set is selected as the same quarter of the year before the test set. In the verification set, the method of grid optimization is used to optimize the super parameters of CNN. In the CNN training, a simple grid search method is used to traverse the value interval set by hyperparameters and selects the optimal hyperparameters without excessively increasing the complexity of the model. The structure and training parameters of the optimal CNN network are shown in Tables 1 and 2, respectively.

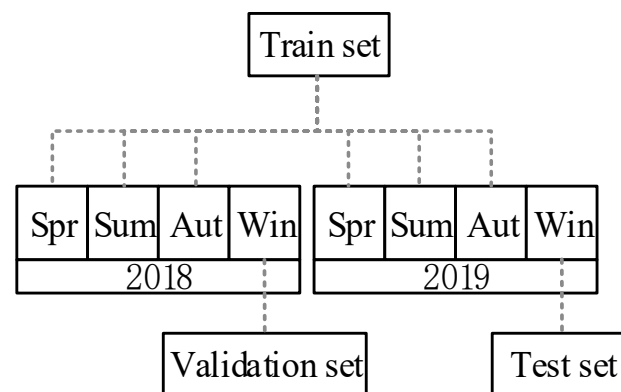


Figure 4. CNN data set division diagram.

Table 1. CNN networks structure parameters.

Structure	Input Size	Conv Kernel/Filter
Conv1	$8 \times 8$	4, $(2 \times 2)$
Average Pool1	$6 \times 6$	$2 \times 2$
Conv2	$3 \times 3$	16, $(2 \times 2)$
Dense	128	
Dense	64	
Dense	1	

Table 2. CNN training parameters.

Parameter	Value
Loss	Mean squared error
Optimizer	Adam
Epoch	90
Batch size	128

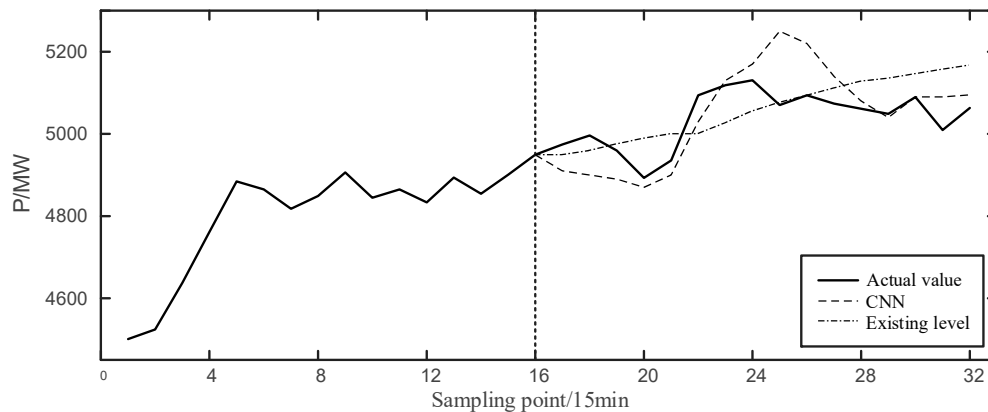
In addition, each layer of the CNN uses the relu activation function, and the steps of convolution kernel and pooling kernel are 1 and 2, respectively. In the end, the recursive prediction method is adopted to achieve 64 to 16 prediction structures.

According to the prediction process of CNN in Figure 3, take the time interval 0:00–4:00 on 6 November 2019 as an example, and the predictive effect is as follows:

In Figure 5, the left side of the vertical dotted line is the actual curve at 20:00–24:00 on the previous day, and the right side is the effect comparison of the forecast curve.  $RMSE$  and  $PE$  indexes of CNN are 266.85 MW and 4.63%, respectively. It can be seen that the existing level can track the actual value



on the general trend, but there is no response at the fluctuation, so it can only meet the basic forecast demand. CNN predicted results successfully predicted the time of the first trough, and followed the climbing process well. However, after the 24th point, it continued to rise and deviated from the actual curve until the 29th point. Yet the climbing response time is too long.



**Figure 5.** CNN prediction result at 0:00–4:00 on 6 November 2019.

Generally speaking, CNN is more sensitive to the short-term data and has a good tracking effect. The farther it is from the prediction starting point, the more difficult it is to achieve accurate prediction. Thus, similar days need to be introduced to assist prediction, so as to improve the prediction accuracy.

### 5.2. Similar Day Forecast

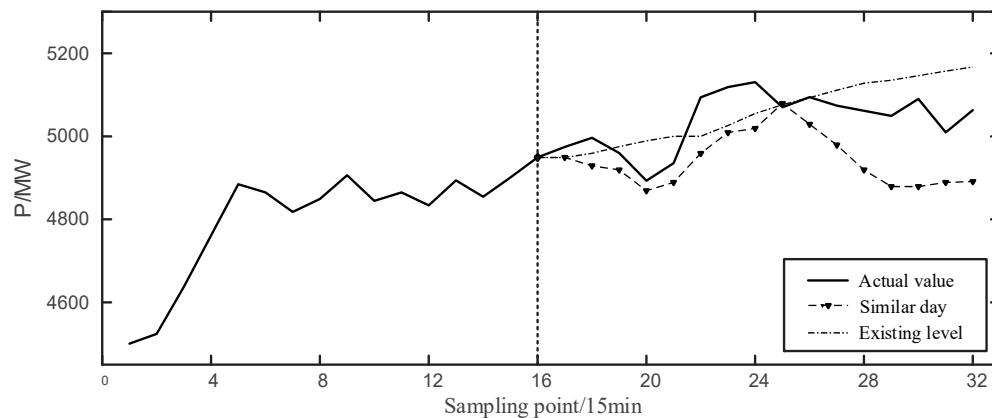
In order to select the most suitable similar day, the similar day method in Section 2.1 is used as the validation set for prediction, and the prediction effect is shown in Table 3. It is shown that by using the combination of the Pearson correlation coefficient and cosine similarity, the prediction effect is the best. Therefore, this combination is selected to predict the test set data.

**Table 3.** Statistical error of different similar days selection methods.

Method	PE
Euclidean distance	6.44%
PCCs	6.39%
Cosine similarity	6.36%
PCCs + Cosine similarity	6.31%

We also take the time interval 0:00–4:00 on 6 November 2019 as an example to conduct the prediction based on similar day, which is convenient for comparison between different models. The predicted results are calculated as follows:

In Figure 6, the prediction values of similar days are lower than the actual ones, but some inflexion points appear similar to the actual value. Moreover, the response of the similar days prediction at the wave crest is slow, and the downward trend in the second half is greater than the real situation, which can neutralize the long response of the CNN prediction.



**Figure 6.** Similar day result at 0:00–4:00 on 6 November 2019.

### 5.3. PSO Combination Prediction and Error Analysis

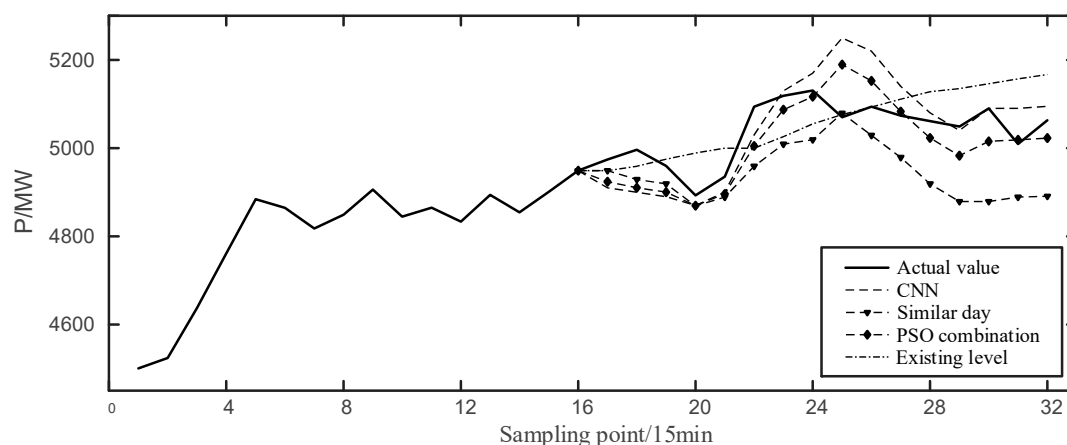
The subsection tests the combination method mentioned above in the verification set, and the comparison results are shown in Table 4.

**Table 4.** Statistical effect of combination weighting.

Method	CNN Weight	RMSE (MW)
Average weight	0.5	608.51
Time varying weight	change	620.91
Entropy weight method	0.474	607.43
Genetic algorithm	0.486	607.30
PSO	0.487	607.29

It shows that the RMSE of the combination weight determined by PSO is the smallest, and the predicted results of multiple models are shown as follows:

It can be seen from Figure 7 that the prediction curve using the PSO combination tracks the first trough well, corrects the response at the peak, fits the real value more closely, and the prediction accuracy is better than other individual predicted results.



**Figure 7.** Predicted results of multiple models at 0:00–4:00 on 6 November 2019.

To better verify this conclusion, statistical analysis is carried out on the whole year 2019, and the parameters  $a$  and  $b$  of the qualified rate in the evaluation index are taken 15% and 350 respectively. The annual prediction indexes are shown in the table below.

From Table 5, we know that, compared to the existing level, *RMSE* and *PE* indexes of the combination model are reduced by 73.2 MW and 1.69%, respectively, and the corresponding *QR* index is increased by 10.1%. Moreover, even if the CNN method and the similar days method can achieve higher prediction accuracy than the existing level, the prediction errors based on PSO weighting combination are the lowest, and the qualification rate is the highest at the same time. This is consistent with the theoretical analysis, that is, the combination model synthesizes the short-term laws (by CNN) and the long-term ones (by the similar days) to give a more accurate prediction. In a word, this method has a great application prospect to accurately predict the short-term wind power.

**Table 5.** Forecast indicators for the whole year of 2019.

Evaluation Indexes	CNN	Similar Days	Combination Model	Existing Level
RMSE (MW)	621.2	627.9	592.9	666.1
PE	6.07%	6.16%	5.79%	7.48%
QR	80.3%	78.6%	81.1%	71.0%

## 6. Conclusions

In this paper, based on the traditional CNN model, a new multi model combination forecasting method is implemented. The prediction accuracy is greatly improved compared to the existing level. The main contributions are as follows:

- On the basis of CNN's wind power prediction, we add the prediction method of similar days, discuss its feasibility, make up for the shortcomings of the same time dimension of multi model combination prediction, and form the characteristics of CNN's short-term law learning and similar days' long-term law searching.
- In this paper, the PSO method is used to avoid the subjectivity in choosing combination weight. The result shows that the combined prediction algorithm of CNN and similar days with PSO has higher prediction accuracy and longer prediction period, which is an effective method.
- Regional wind power prediction is different from single base station prediction or single wind farm prediction. Regional wind power prediction is difficult to rely on real-time data or single regional meteorological data, and needs a longer time margin. Therefore, it requires fully mining the law in historical data to meet the prediction demand.

**Author Contributions:** Conceptualization, Y.L. and W.Z.; Methodology, Y.L. and W.Z.; Software, F.Y.; Validation, F.Y. and L.Y.; Writing—original draft, F.Y. and Y.L.; Writing—review & editing, Y.L. and W.Z.; Project Administration, Y.L. and W.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China grant number 61703405 and the Science and Technology Project of State Grid grant number SGGJY00PSJS1800138.

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

## References

1. Qian, Z.; Pei, Y.; Zareipour, H.; Chen, N. A review and discussion of decomposition-based hybrid models for wind energy forecasting applications. *Appl. Energy* **2019**, *235*, 939–953. [\[CrossRef\]](#)
2. Huang, N.; Wu, Y.; Lu, G.; Wang, W.; Cao, X. Combined probability prediction of wind power considering the conflict of evaluation indicators. *IEEE Access* **2019**, *7*, 174709–174724. [\[CrossRef\]](#)
3. Wang, J.; Zhou, Q.; Zhang, X. Wind power forecasting based on time series ARMA model. *IOP Conf. Ser. Earth Environ. Sci.* **2018**, *199*, 022015. [\[CrossRef\]](#)
4. Zhang, Y.; Sun, H.; Guo, Y.; Lu, Z. Wind turbine power prediction research based on improved grey system model. *Renew. Energy Resour.* **2018**, *36*, 1865–1870. (In Chinese)
5. Chen, S.; Ye, L.; Zhang, G.; Zeng, C.; Dong, S.; Dai, C. Short-term wind power prediction based on combined grey-Markov model. In Proceedings of the 2011 International Conference on Advanced Power System Automation and Protection, Beijing, China, 16–20 October 2011; Volume 3, pp. 1705–1711.

6. An, J.; Chen, Z.; Wu, M.; She, J.; Ding, M. A multi-source wind speed fusion method for wind power prediction based on kNN-SVR. In Proceedings of the IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society, Washington, DC, USA, 21–23 October 2018; pp. 4245–4250.
7. Tian, S.; Fu, Y.; Ling, P.; Wei, S.; Liu, S.; Li, K. Wind power forecasting based on arima-lgarch model. In Proceedings of the 2018 International Conference on Power System Technology (POWERCON), Guangzhou, China, 6–8 November 2018; pp. 1285–1289.
8. Zhu, L.; Shi, H.; Ding, M. A chaotic BP neural network used to wind power prediction. In Proceedings of the 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Xi'an, China, 25–27 May 2018; pp. 1169–1172.
9. Qu, X.; Kang, X.; Zhang, C.; Jiang, S.; Ma, X. Short-term prediction of wind power based on deep long short-term memory. In Proceedings of the 2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), Xi'an, China, 25–28 October 2016; pp. 1148–1152.
10. Zhang, X.; Wang, Y.; Chen, Y.; Cheng, X.; Gao, L. Short-term wind speed prediction based on GRU. In Proceedings of the 2019 IEEE Sustainable Power and Energy Conference (iSPEC), Beijing, China, 21–23 November 2019; pp. 882–887.
11. Li, C.; Tang, G.; Xue, X.; Saeed, A.; Hu, X. Short-term wind speed interval prediction based on ensemble GRU model. *IEEE Trans. Sustain. Energy* **2019**, *11*, 1370–1380. [[CrossRef](#)]
12. Fu, Y.; Hu, W.; Tang, M.; Yu, R.; Liu, B. Multi-step ahead wind power forecasting based on recurrent neural networks. In Proceedings of the 2018 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), Kota Kinabalu, Malaysia, 7–10 October 2018; pp. 217–222.
13. Dong, D.; Sheng, Z.; Yang, T. Wind power prediction based on recurrent neural network with long short-term memory units. In Proceedings of the 2018 International Conference on Renewable Energy and Power Engineering (REPE), Toronto, ON, Canada, 24–26 November 2018; pp. 34–38.
14. Zhang, Y.; Sun, H.; Guo, Y. Wind power prediction based on PSO-SVR and grey combination model. *IEEE Access* **2019**, *7*, 136254–136267. [[CrossRef](#)]
15. Ju, Y.; Sun, G.; Chen, Q.; Zhang, M.; Zhu, H.; Rehman, M.U. A model combining convolutional neural network and LightGBM algorithm for ultra-short-term wind power forecasting. *IEEE Access* **2019**, *7*, 28309–28318. [[CrossRef](#)]
16. Zhou, B.; Sun, B.; Gong, X.; Liu, C. Ultra-short-term prediction of wind power based on EMD and DLSTM. In Proceedings of the 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA), Xi'an, China, 19–21 June 2019; pp. 1909–1913.
17. Lee, J.; Wang, W.; Harrou, F.; Sun, Y. Wind power prediction using ensemble learning-based models. *IEEE Access* **2020**, *8*, 61517–61527. [[CrossRef](#)]
18. Wang, Y.; Wang, D.; Tang, Y. Clustered hybrid wind power prediction model based on ARMA, PSO-SVM, and clustering methods. *IEEE Access* **2020**, *8*, 17071–17079. [[CrossRef](#)]
19. Lu, P.; Ye, L.; Sun, B.; Zhang, C.; Zhao, Y.; Teng, J. A new hybrid prediction method of ultra-short-term wind power forecasting based on EEMD-PE and LSSVM optimized by the GSA. *Energies* **2018**, *11*, 697. [[CrossRef](#)]
20. Zhang, C.; Ding, M.; Wang, W.; Bi, R.; Miao, L.; Yu, H.; Liu, L. An improved ELM model based on CEEMD-LZC and manifold learning for short-term wind power prediction. *IEEE Access* **2019**, *7*, 121472–121481. [[CrossRef](#)]
21. López, E.; Valle, C.; Allende, H.; Gil, E.; Madsen, H. Wind power forecasting based on echo state networks and long short-term memory. *Energies* **2018**, *11*, 526. [[CrossRef](#)]
22. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv* **2015**, arXiv:1502.03167v3.

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



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).