# Using LSTM Model to predict Short-term wind power

Zhiqi Wan[1], Jiali Wang[1], Jialin Tang[1], Bo Zhang[1] and Feilin Zhu[2,*]

[1] *DAYU College, Hohai University, Nanjing, China.*
[2] *College of Hydrology and Water Resources, Hohai University, Nanjing, China*

**Abstract**

Wind energy is an essential renewable energy. For Carbon Neutrality, wind power has received extensive attention around the world. However, short-term wind power time series are difficult to predict with complex characteristics such as non-stationary and nonlinear. Therefore, this paper proposes a short-term prediction method of wind power using the Long Short-term Memory (LSTM) model. In view of insufficient memory ability, gradient disappearance and explosion in traditional prediction methods, the strategy of "Data Processing - Autocorrelation Analysis - Model Prediction" is proposed. Firstly, test the outliers by Z-Score method, and linearly interpolate the missing values and outliers. Secondly, determine the model input length through autocorrelation and partial autocorrelation coefficients. Finally, predict each subsequence with the LSTM model. On the testing set, the root mean square error is 58.55 (MW), mean absolute error is 79.60 (MW), and coefficient of determination is 0.86. In brief, using LSTM prediction model can obtain a higher accuracy of short-term wind power prediction.

**Keywords**

Long Short-term Memory, Short-term Wind Power Prediction, Data Processing

## 1. Introduction

Energy is an essential foundation for human beings. With global climate change, environmental pollutions and other issues becoming more and more prominent, the transformation of the energy structure is imminent. Recently, many countries have been vigorously developing new energy. Wind energy, as an important renewable energy source, has the advantages of a long history of research, many technological innovations and broad development prospects, and is crucial for promoting energy transformation. However, wind energy is characterized by intermittency, volatility, and randomness. The grid integration of wind power intensifies the pressure of peak regulation. In order to ensure national energy security, accurate short-term wind power prediction is an important challenge.

*ICCIC 2024: International Conference on Computer and Intelligent Control, June 29–30, 2024, Kuala Lumpur, Malaysia*

* Corresponding author.
2119010405@hhu.edu.cn (Z. Wan); 1093347909@qq.com (J. Wang); tangjialin2003@163.com (J. Tang); 1970649539@qq.com@163.com (B. Zhang); zhufeilin163@163.com (F. Zhu)
0009-0001-6180-0031 (Z. Wan); 0009-00003566-1931 (J. Wang); 0009-0007-7897-1674 (J. Tang); 0009-0008-0174-5000 (B. Zhang); 0000-0002-9780-9361 (F. Zhu)

© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR-WS.org/Vol-3734/paper8.pdf

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

Many scholars conducted in-depth research on the wind power prediction and proposed many models and methods. According to the principal classification, they can be divided into physical methods based on numerical weather predictions and statistical methods with historical data [1]. The physical prediction method mainly predicts wind power by screening related physical quantities and establishing space-time physical equations. This method does not require historical wind power data, but is limited by the accuracy of measurement data, modeling errors and economic conditions of complex wind farms. It is mostly applied to new wind farms lacking historical data [2-4]. Statistical prediction methods mainly include Time-Series Analysis, Kalman filter, Artificial Neural Network, etc. [5-12], which is based on the learning of historical data and does not need to consider the complex physical calculation process. It has the characteristics of self-adaptation, self-adjustment, and self-learning. The model structure is simple and suitable for wind farms with historical data. Many studies have shown that statistical prediction methods have higher applicability and accuracy in short-term wind power prediction [13].

As deep learning theories develop, the accuracy of wind power prediction has been continuously improved [14-16]. Long Short-Term Memory neural network is an improvement of traditional RNN. It retains important features through the gate function, and effectively alleviates the problems of insufficient long sequence memory capacity, gradient disappearance and explosion. For nonlinearly varying wind power time series, the accuracy of LSTM model is better.

Summarily, according to the idea of "Data Processing - Autocorrelation Analysis - Model Prediction", this paper proposes a short-term prediction method of wind power based on LSTM model. Taking the annual hourly data of a power station in Qinghai Province as an example, the autocorrelation of wind power historical data is proved. The applicability and accuracy of the model are also analyzed and tested, and the overall idea is shown in Figure 1.

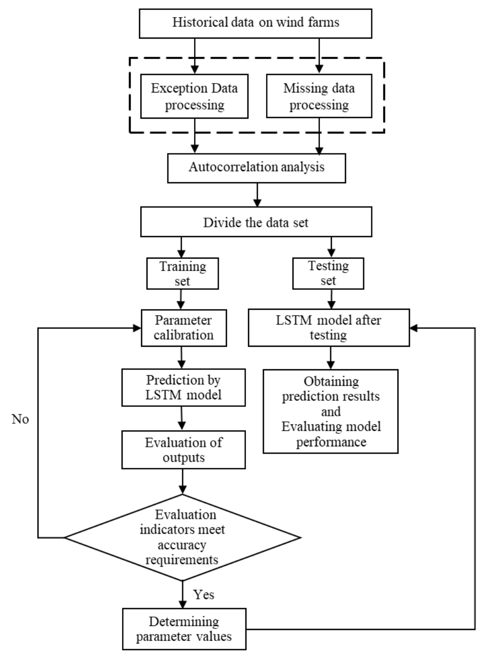



**Figure 1:** Framework of wind power prediction using LSTM model.

# 2. Research Methods

## 2.1 Data Processing

For data-driven short-term prediction models, missing values and outliers in the historical data can seriously affect the accuracy. Therefore, data processing is required before the model forecasting.

For missing values, this paper adopts the method of linear interpolation. It is assumed that the missing values can be represented by the data of its previous and next moments through a linear relationship, given by the formula:

$$x_i = x_{i-1} + \frac{(x_{i+1} - x_{i-1})}{t_{i+1} - t_{i-1}} \cdot (t_i - t_{i-1}) \tag{1}$$

In the equation, $x_i$ is the missing value, $x_{i-1}$,$x_{i+1}$ are the previous data and next data. $t_i$ is the moment corresponding to missing value. $t_{i-1}$ is the previous period and $t_{i+1}$ is the next period.

For outliers testing, this paper uses the Z-Score method, and the formula as:

$$Z_i = \frac{(x_i - \overline{x})}{\sigma} \tag{2}$$

In the equation, $Z_i$ is the Z-Score value. $x_i$ is the $i$ th data in the time series. $\overline{x}$ is the mean value of the time series. $\sigma$ is the standard deviation of the time series. The threshold of Z-Score is taken as 3. When $Z_i > 3$, the data value is considered to have a large difference with other values and is regarded as an outlier. For outliers, the treatment is the same as for missing values, with the linear interpolation method.

## 2.2 Correlation Analysis

The core of the predicted model with historical data is to find the patterns hidden in the time series and then predict the data of next period based on the discovered patterns. Therefore, before using such prediction models, it is necessary to ensure that the time series exhibit autocorrelation. To test the autocorrelation of historical wind power data, autocorrelation coefficient is introduced, like:

$$\rho_h = \frac{\mathrm{Cov}(x_t, x_{t-h})}{\sqrt{\mathrm{Var}(x_t) \cdot \mathrm{Var}(x_{t-h})}} \tag{3}$$

In the equation, $\rho_h$ is the $h$ th autocorrelation coefficient, representing the correlation between data points separated by $h$ time units; $\mathrm{Cov}$ is covariance; $\mathrm{Var}$ is variance; $h$ is the lag order; $x_t$ and $x_{t-h}$ are the observed values at time moments $t$ and $t-h$. By using autocorrelation coefficients, a linear relationship between the perdiction point and the observed point is constructed as follows:

$$x_t = \rho_h x_{t-h} + \varepsilon_t \tag{4}$$

In the equation, $\varepsilon_t$ represents the white noise error term. When $\rho_h = 1$, it indicates complete correlation between the prediction point and the observed point, meaning data at one time point can be entirely predicted by data at another time point; when $\rho_h = 0$, it indicates no correlation between $x_t$ and $x_{t-h}$; when $\rho_h = -1$, it indicates complete negative correlation between $x_t$ and $x_{t-h}$, implying data at one time point can be entirely reversely predicted by data at another time point.

Since time series are correlated, autocorrelation coefficients cannot represent the correlation between $x_t$ and $x_{t-h}$ without the influence of other time series. To eliminate interference from $x_{t-1}, x_{t-2}, \cdots, x_{t-h+1}$ on the correlation between two points, partial autocorrelation coefficient is introduced as follows:

$$\phi_h = \rho_h - \sum_{j=1}^{h-1} \phi_{h-1,j} \cdot \rho_{h-j} \tag{5}$$

Where $\phi_h$ is the $h$ th partial autocorrelation coefficient, representing the correlation between data points separated by $h$ time while considering the situation of the previous $h$ time points; $\phi_{h-1,j}$ is the $j$ th term in the $h-1$ th order partial autocorrelation coefficient; $j$ is the summation index; and $h$ is the order of autocorrelation and partial autocorrelation functions, representing the time interval to be calculated. Based on the definition of partial autocorrelation, a linear relationship between the $x_t$ and all observed points within the previous $h$ time units is constructed as follows:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_h x_{t-h} + \varepsilon_t \tag{6}$$

### 2.3 LSTM Model

Long Short-Term Memory (LSTM) is an improvement of recurrent neural network (RNN). Compared to traditional neural networks, LSTM neural networks have more effective memory and forgetting patterns for long time series. By introducing gate mechanisms, LSTM can better capture long-term dependencies, effectively addressing issues like gradient disappearance and explosion. Additionally, for nonlinear system time series prediction, LSTM neural networks also have significant advantages[17].

The key components of LSTM neural networks include forget gate, input gate, cell state, and output gate, which together form the unique four-layer structure of LSTM, as shown in Figure 2.

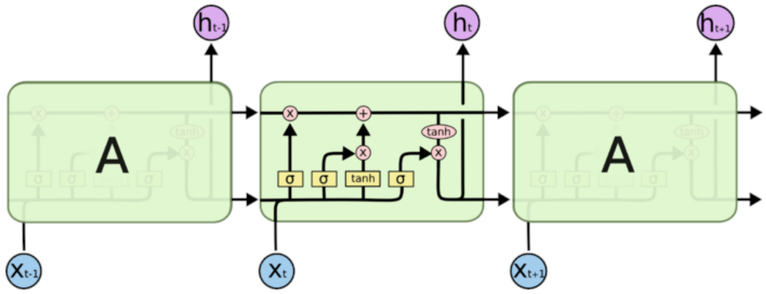



**Figure 2:** Structure of LSTM .

According to the flow of information within neurons, the four-layer structure is explained as follows:

1. Forget Gate: Distinguish and forget minor information.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{7}$$

2. Input Gate: Determine and filter new information stored in the cell state.

$$\begin{aligned} i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \end{aligned} \tag{8}$$

3. Cell State: Update the cell state based on the previous cell state and input gate information.

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{9}$$

4. Output Gate: Select the information to be input to the next neuron.

$$\begin{aligned} o_t &= \sigma(W_o [h_{t-1}, x_t] + b_o) \\ h_t &= o_t * \tanh(C_t) \end{aligned} \tag{10}$$

Where $w_f$, $w_i$ , $w_c$ , $w_o$ are weight matrices; $b_f$, $b_i$ , $b_c$ , $b_o$ are bias vectors; $h_t$, $h_{t-1}$ represent the input and output of previous neuron and current neuron; $C_t$, $C_{t-1}$represent the cell states of the previous neuron and the current neuron; $\tilde{C}_t$ is the cell state of the input gate;$\sigma(\cdot)$ is the sigmoid function.

### 2.4 Performance Evaluation

To quantitatively evaluate the prediction effectiveness, three evaluation metrics are selected, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination ($R^2$), defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{11}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{12}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{13}$$

Where $n$ is the number of samples in the testing set; $y_i$ is the actual wind power; $\widehat{y}_i$ is the prediction value; and $\overline{y}$ is the mean of the wind power historical sequence.

# 3. Research Example

This paper collects hourly data of wind power in Qinghai Province throughout the year. Through the detection of missing values and outliers, there are 5 missing values and 7 outliers. With linear interpolation processing, 8785 normal data can be obtained.

Due to the limitations of data and algorithms, the prediction model often has problems such as underfitting and overfitting. To improve accuracy, the data set needs to be divided into a training set and a testing set. The division of the data set and the amount of data are shown in Table 1.

**Table 1**
Experimental data sets

| Location of the wind farm | Data type | Time span | Number of samples (strip) |
|---|---|---|---|
| Qinghai Province | All data | Jan. 1 to Dec. 31 | 8785 |
| | Training set | Jan. 1 to Nov. 26 | 7902 |
| | Testing set | Nov. 26 to Dec. 31 | 883 |

# 4. Results and Discussion

## 4.1 Correlation Analysis

Bring the 8,785 processed data sets into Eq. (1), and set the significance level $\alpha = 5\%$. Draw the autocorrelation analysis diagram of historical data, as shown in Figure 3. The diagram shows that most of the observation points are outside the significance band. It proves that the historical data of wind power has strong autocorrelation and can be used as model inputs. With the increase of the time, the autocorrelation coefficient first decreases and then oscillates around the value of 0, indicating that the closer date to the prediction point are more likely to affect the prediction results, and wind power has a certain periodic pattern on the time scale.

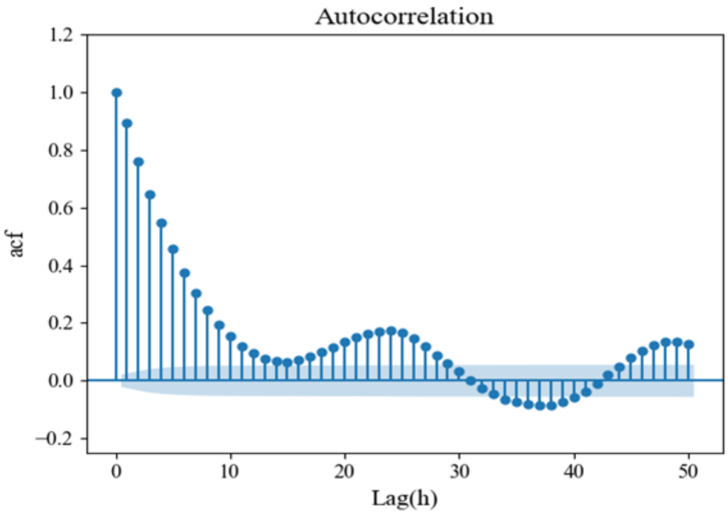



**Figure 3:** Autocorrelation analysis of wind power historical data.

Bring the historical data into Eq (3), set the significance level $\alpha = 5\%$, and draw the partial autocorrelation analysis diagram of historical data, as shown in Figure 4. The diagram shows that with the increase of time, the partial autocorrelation coefficient and autocorrelation coefficient change in a similar trend, but the partial autocorrelation coefficient decreases faster and the oscillation peak is smaller. This indicates that if only consider the partial autocorrelation coefficient between the two points, the prediction point is more correlated with the observation points in the first 2 time intervals.

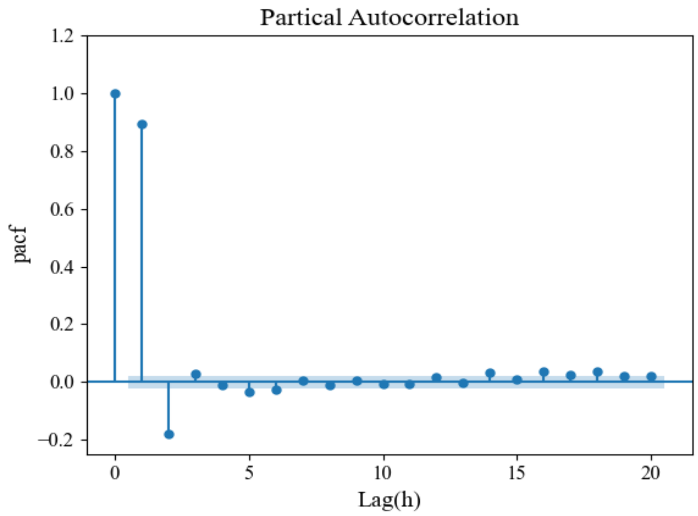
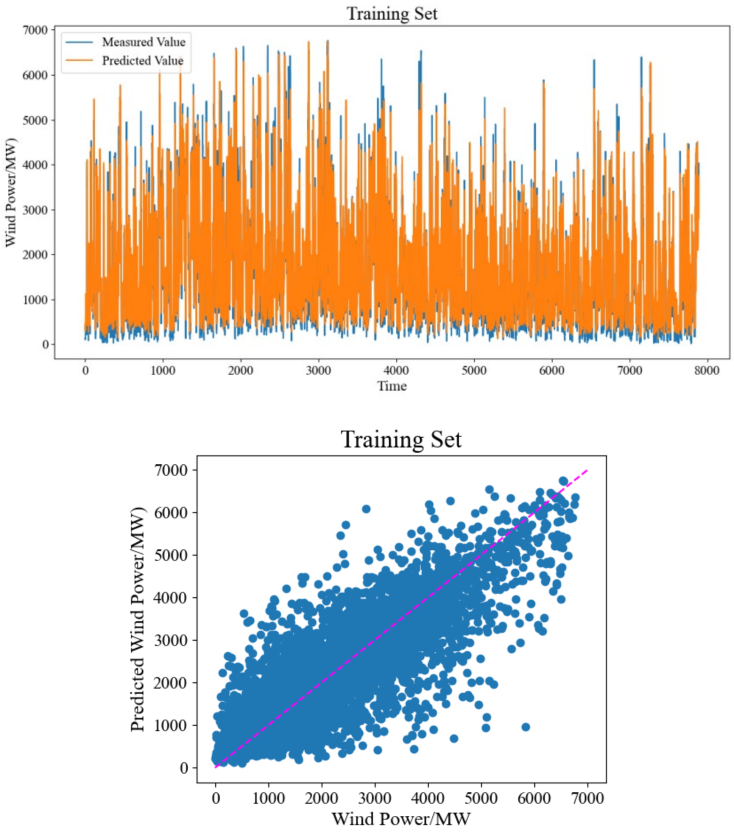



**Figure 4:** Partial autocorrelation analysis of wind power historical data.

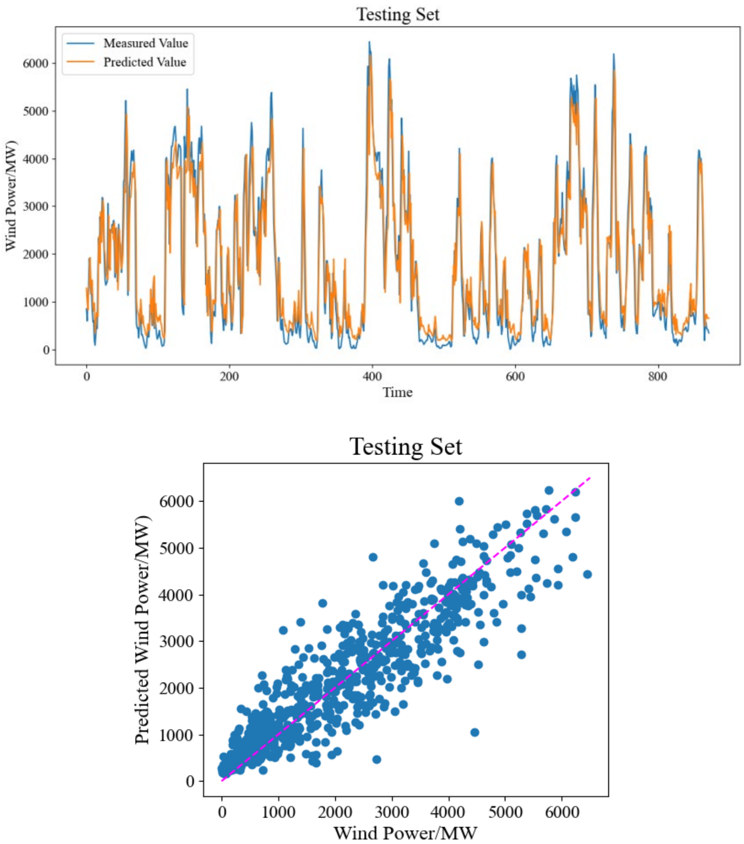
Considering the high accuracy and timeliness requirements of short-term prediction model, and the model inputs have a major influence on the short-term prediction accuracy, selecting the appropriate length of input data is an important prerequisite. Since the historical data input to the model is correlated with each other, the autocorrelation coefficient is mainly considered when selecting the length of input data. However, if the input data is too long, it will lead to error superposition. Therefore, considering the autocorrelation and partial autocorrelation coefficients comprehensively, model input length should not exceed 15 time series.

### 4.2 Model Training and Prediction

Taking the hourly wind power of Qinghai Province as an example, the predicted step size is 1, and brought it into LSTM neural network for manual parameter calibration. According to the above three evaluation indexes, the activation function is set as the "Relu" function, the time step is 10, the dimension of LSTM layer is 128, epoch=12, batch_size=32. The model prediction result under this parameter is shown in Figure 4, and the analysis of indexes is shown in Table 2.




(a) Training Set




(b) Testing Set

**Figure 5:** Comparison of LSTM model prediction results for training set and testing set

**Table 2**
Calculation results of evaluation indexes

| | $RMSE$ ($MW$) | $MAPE$ ($MW$) | $R^2$ |
|---|---|---|---|
| Training Sets | 65.72 | 55.15 | 0.79 |
| Testing Sets | 58.55 | 79.60 | 0.86 |

From Fig. 5 and Table 2, the LSTM prediction model with historical data has a better prediction accuracy of wind power. Therefore, when other information is missing and only historical time series are available, the LSTM prediction model has significant advantages of high accuracy and convenient calculation.

For the power grid system implementing regulation, the higher the accuracy and the longer the forecast period, the greater the benefit. For better regulation of the "Wind - Light - Water" system, the prediction accuracy is still a work in progress. Based on the research in this paper, in-depth research, such as input data feature extraction, prediction model coupling, parameter optimization, can be promoted subsequently.

## 5. Conclusion

Accurate short-term wind power prediction is a prerequisite for implementing “Wind - Light - Water” complementary optimal scheduling. According to the idea of "Data Processing - Autocorrelation Analysis - Model Prediction", this paper constructs a LSTM prediction model based on historical data. Taking the annual hourly data of a power station in Qinghai Province as an example, this paper analyzes and tests the applicability and accuracy, and the following conclusions are obtained:

(1) The historical data of wind power has autocorrelation and periodicity, making it suitable for use as input data in prediction models..

(2) If the interval between the predicted point and the observed point is smaller, the autocorrelation coefficient is larger and the autocorrelation is stronger. Therefore, several observation points close to the front of the prediction point should be selected for prediction. For each prediction point, the number of input observation points should not exceed 15.

(3) When the available data is limited, the LSTM prediction model can obtain the single-value prediction of wind power with high precision in a short time.

## Acknowledgment

This research was funded by the College Student Innovation and Entrepreneurship Training Program of Hohai University: Ultra-short-term probabilistic prediction method of wind and photovoltaic power in watersheds under the background of "dual-carbon" strategy Nexus (No. 202310294022Z).

## References

[1] J. Yan, Y. Liu, S. Han, et al., “Reviews on uncertainty analysis of wind power forecasting,” Renewable and Sustainable Energy Reviews, pp. 521322-1330, 2015.

[2] G. Giebel, J. Badger, I. MartiP, “Short-term forecasting using advanced physical modeling-the results of the anemos project,” European Wind Energy Conference & Exhibition, Athens, pp. 236-264, 2006.

[3] L. L. Li, X. Zhao, M. L. Tseng, et al., “Shot-term wind power forecasting based on support vector machine with improved dragonfly algorithm,” Journal of Cleaner Production, vol. 242 pp. 38-47, 2020.

[4] L. Ye, Y. Zhao, C. Zeng, et al., “Short-term wind power prediction based on spatial model,” Renew Energy, vol. 101 pp. 1067-74, 2017.

[5] E. Erden, J. Shi, “ARMA based approaches for forecasting the tuple of wind speed and direction,” Applied Energy, vol. 88, no. 4, pp. 1405-1414, 2011.

[6] D. A. Bechrakis, P. D. Sparis, “Wind speed prediction using artificial neural networks,” Wind Engineering, vol. 22, no. 6, pp. 287-295, 1998.

[7] E. A. Bossanyi, “Short-term wind prediction using Kalman filters,” Wind Engineering, vol. 9, no. 1, pp. 1-8, 1985.

[8] E. T. Renani, M. F. M. Elias, N. A. Rahim, "Using data-driven approach for wind power prediction: A comparative study," Energy Conversion & Management, vol. 118, pp. 193-203, 2016.

[9] A. E. Saleh, M. S. Moustafa, K. M. Abo-Al-Ez, et al., "A hybrid neuro-fuzzy power prediction system for wind energy generation," International Journal of Electrical Power & Energy Systems, vol. 74, pp. 384-395, 2016.

[10] S. Wang, M. Li, L. Zhao, et al., "Short-term wind power prediction based on improved small-world neural network," Neural Computing & Applications, vol. 31, pp. 3173-3185, 2019.

[11] Y. Hao, L. Dong, X. Liao, et al., "A novel clustering algorithm based on mathematical morphology for wind power generation prediction," Renewable Energy, vol. 136, pp. 572-585, 2019.

[12] C. D. Zuluaga, M. A. Álvarez, E. Giraldo, "Short-term wind speed prediction based on robust Kalman filtering: An experimental comparison," Applied Energy, vol. 156, 2015.

[13] J. Duan, P. Wang, W. Ma, et al., "A novel hybrid model based on nonlinear weighted combination for short-term wind power forecasting," International Journal of Electrical Power and Energy Systems, vol. 134, 2022.

[14] M. Gong, C. Yan, W. Xu, et al., "Short-term wind power forecasting model based on temporal convolutional network and Informer," Energy, vol. 283, 2023.

[15] S. Wang, J. Shi, W. Yang, et al., "High and low frequency wind power prediction based on Transformer and BiGRU-Attention," Energy, vol. 288, pp. 129753, 2024.

[16] L. Xiang, X. Fu, Q. Yao, et al., "A novel model for ultra-short term wind power prediction based on Vision Transformer," Energy, vol. 294, pp. 294130854, 2024.

[17] S. Hochreiter, J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735-1780, 1997. DOI:10.1162/ neco.1997. 9. 8. 1735.