

Research Article

A Short-Term Power Output Forecasting Model Based on Correlation Analysis and ELM-LSTM for Distributed PV System

Deng Yongsheng,¹ Jiao Fengshun,¹ Zhang Jie,¹ and Li Zhikeng² 

¹Shenzhen Power Supply Co., Ltd., Shenzhen 518033, China

²China Energy Engineering Group Guangdong Electric Power Design Institute Co. Ltd., Guangzhou, China

Correspondence should be addressed to Li Zhikeng; lizhikeng@gedi.com.cn

Received 27 October 2019; Revised 31 January 2020; Accepted 13 March 2020; Published 29 June 2020

Academic Editor: Nicola Sorrentino

Copyright © 2020 Deng Yongsheng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Accurate short-term power output forecasting results are conducive to reducing the scheduling difficulty of grid-connected operation of distributed photovoltaic (PV) systems, thus improving the safety and stability of power grid operation. In this paper, a one-day-ahead short-term power output forecasting model based on correlation analysis and combination algorithms for distributed PV system is proposed to solve the problems within the current methods. Firstly, the basic information of distributed PV system is introduced, and the main influence factors affecting the power output of distributed PV system are determined. Secondly, the influence factors with higher correlation with PV output are selected by Spearman rank-order correlation coefficient (SROCC) analysis in multiple timescales. Then, based on the multimodel univariate extreme learning machine (ELM) submodel and the single-model multivariate long short-term memory (LSTM) submodel, the ELM-LSTM model is established. The case study analysis based on the actual data indicates that the ELM-LSTM forecasting model proposed in this paper has higher forecasting accuracy than the traditional forecasting methods.

1. Introduction

With the continuous consumption of global fossil energy sources, the development of new energy sources and the improvement of the utilization efficiency of renewable energy have become an important means to solve problems such as energy shortage and environmental pollution. Among many new energy sources, solar photovoltaic (PV) power generation has been widely used due to the excellent performances such as cleanliness and high efficiency, and it has been rapidly developed in recent years. According to the latest data of the International Renewable Energy Agency (IRENA), the global installed capacity of grid-connected PV is 94.3 GW in 2018, accounting for more than half of the installed capacity of renewable energy [1]. The construction of large-capacity and distributed PV systems has become a significant solution to fully absorb solar energy resources and alleviate the peaking pressure of power grids. Due to the volatility and intermittent characteristics of PV power generation, the grid-connected operation of distributed PV systems increases the difficulty of power system scheduling

[2]. Therefore, accurate output forecasting results of the distributed PV system will be beneficial to the dispatching department to reasonably plan the coordination and co-operation between conventional energy power plants and distributed PV systems, which can improve the safety and stability of power grid operation.

Generally speaking, there are two main methods for the short-term power output forecasting of PV system: indirect forecasting method and direct forecasting method. The indirect forecasting method firstly forecasts the solar radiation intensity and then the short-term power output is given based on the physical model of the PV power plant [3–5]. The direct forecasting method is to use the historical output data of PV power plant to forecast directly. In recent years, with the development of numerical weather prediction (NWP) technology, accurate meteorological values can be added as important factors to the forecasting model and, accordingly, the direct forecasting method will be more efficient and convenient, avoiding complex physical modeling of PV power plants [6–8]. Abuella and Chowdhury [9] used weather prediction and artificial neural network (ANN) models to

forecast PV output and [10] compared the PV output results by using ANN, ANN-Particle Swarm Optimization (ANN-PSO), and ANN-Firefly Algorithm (ANN-FA). Jie et al. [11] propose a forecasting method based on weather classification and support vector machine (SVM), and select four types of weather conditions for testing; SVM shows excellent forecasting performance. A hybrid forecasting method combining SVM, PSO, and wavelet transform (WT) has been given in [12], which has better result and shorter calculation time than traditional methods.

Two main problems exist in the mentioned methods. (1) The input variables of most PV forecasting models are selected as historical data. For example, the PV output of the forecasting day is given through the relevant information of the past few days before the forecasting day. In fact, the PV output varies remarkably with meteorological factors [13]; therefore, historical data often cannot accurately contain the real information of the forecasting day, resulting in a decrease in accuracy. (2) Though there are many influence factors affecting PV output, most of the forecasting methods lack the analyses of the correlations between the factors and the PV output. Reference [14] proposes a forecasting method based on Principal Component Analysis (PCA) and SVM. PCA is a statistical analysis method that transforms multiple influence factors into a few unrelated comprehensive variables but cannot select the most effective factors properly. Excessive selection of factors may lead to the increase of complexity in the forecasting model, while too few choices may result in the lack of effective information, causing greater risk of error.

Aiming at the above problems, this paper proposes a one-day-ahead short-term power output forecasting model based on correlation analysis and combination algorithms for distributed PV system, which adopts the direct forecasting method. Firstly, by analyzing the factors affecting the output of distributed PV system, the multitime-scale Spearman rank-order correlation coefficient analysis method is used to select the most effective factors related to PV output. Secondly, based upon the 24 h ahead weather report, the next day's meteorological information can be obtained in advance in the forecasting process. Then, the multimodel univariate extreme learning machine (ELM) submodel and the single-model multivariate long short-term memory (LSTM) submodel are established, respectively. The excellent generalization performance and fast forecasting ability of ELM facilitate the preliminary establishment of the correlation between the input and output of the forecasting model, and the forecasting results are provided to LSTM submodel for deep learning together with the relevant influence factors of the forecasting day, which reduces the dependence on historical data. In the whole ELM-LSTM model, LSTM submodel modifies the forecasting results of ELM submodel and achieves higher accuracy.

2. Distributed PV System and Correlation Analysis of the Influence Factors

2.1. Distributed PV System. Distributed PV system is a power generation system which uses PV modules to directly convert solar energy into electrical energy [15]. Generally, distributed

PV system is arranged near the users, which is often directly connected to the distribution network, and together with which provides electricity to the surrounding users. Compared with centralized PV power plants, distributed PV system has smaller installed capacity and output power, geographically dispersed [16]. At present, the most widely used distributed PV system is the PV power generation project built on the roof of urban buildings. Figure 1(a) shows the distributed PV system on the roof of an urban building and Figure 1(b) shows the centralized PV power plant.

In engineering, the output power of distributed PV system is related to solar radiation intensity, PV area, temperature, conversion efficiency, and so forth, and actually, there are other factors in addition to the parameters above [17]. In summary, the factors can be summarized into two categories: (1) internal influence factors: geographical location of PV system, installation angle of PV panel, conversion efficiency of PV array, and so forth; and (2) external influence factors: mainly meteorological factors, such as solar radiation intensity, atmospheric temperature, humidity, wind speed, cloud cover, and atmospheric pressure [18]. In the short-term output forecasting of PV system, many influence factors are often considered to improve the accuracy. As mentioned, too many selection factors may increase the complexity of the forecasting model, and too few may lose important information. Therefore, selecting the appropriate influence factors will help further improving the forecasting accuracy.

For a given distributed PV system, the internal influence factors such as the conversion efficiency and installation angle of the PV array have been already included in the historical output power. Therefore, the input of the forecasting model can be reduced without considering the internal influence factors. Based on this, this paper will only consider external factors that affect the output of distributed PV system.

2.2. Correlation Analysis of the Influence Factors. The data of an actual distributed PV system is selected for analysis, and the distributed PV system is arranged on the roof of a factory in a city in southern China. The data are composed of historical PV power output (kW), temperature ($^{\circ}\text{C}$), solar radiation (W/m^2), and meteorological information, including wind speed (m/s), wind direction ($^{\circ}$), atmospheric pressure (hPa), and humidity (%). The above data are obtained through the power sampling equipment (power meters), thermometers, radiometers equipped in distributed PV system, and external weather stations. The sampling time is from 1st April 2016 to 1st December 2016, for a total of eight months. The daily sampling frequency is 30 minutes, that is, 48 sampling points within 24 hours. The output power of the PV system in the sampling time range is shown in Figure 2, and the temperature, solar radiation, and meteorological information (wind speed, wind direction, atmospheric pressure, and humidity) are shown in Figure 3.

The randomness and volatility of meteorological conditions increase the difficulty of PV output forecasting. In order to select the influence factors with higher correlation with PV output to reduce the complexity of the forecasting model, Spearman rank-order correlation coefficient (SROCC) analysis method is used



FIGURE 1: The PV power systems: (a) distributed PV system and (b) centralized PV power plant.

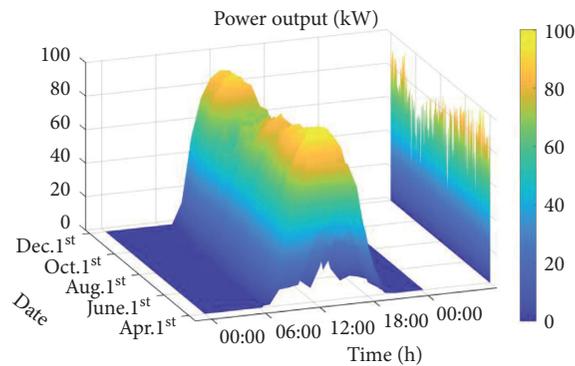


FIGURE 2: The output power (kW) of the PV system in the sampling time range.

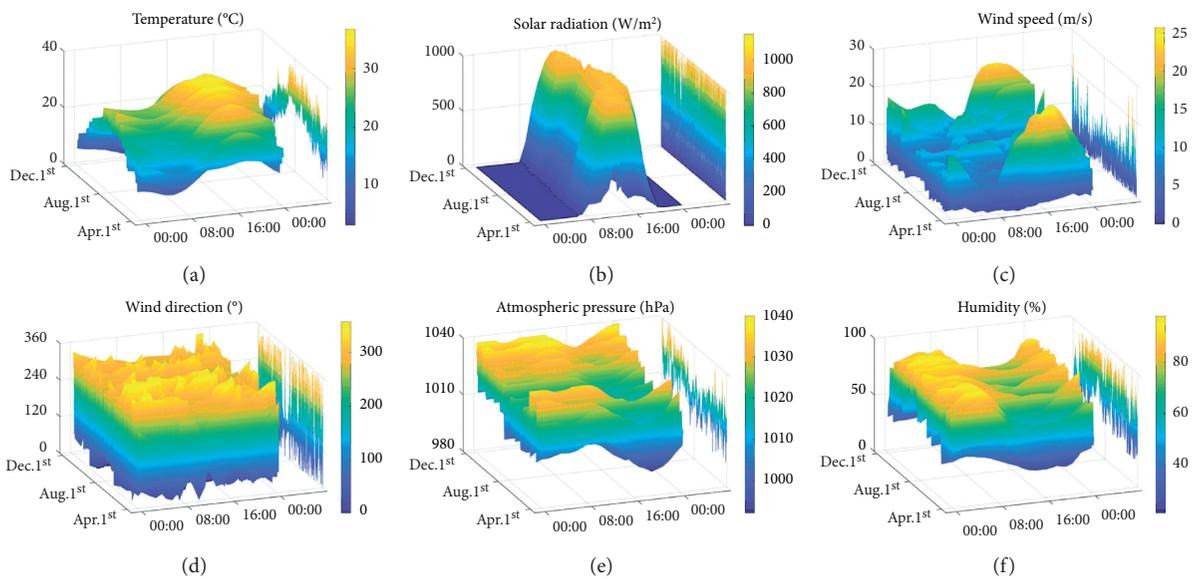


FIGURE 3: The data of temperature, solar radiation, and meteorological information: (a) temperature, (b) solar radiation, (c) wind speed, (d) wind direction, (e) atmospheric pressure, and (f) humidity.

to analyze various influence factors in multiple timescales. SROCC is a statistical index used to determine the correlation degree between two variables (called independent variable and dependent variable). SROCC can be used for research regardless of the overall distribution pattern and sample size of the two

variables. Consequently, SROCC is an effective method to determine the nonlinear correlation, usually in the $[-1, 1]$ interval [19]. When the independent variable is positively correlated with the dependent variable, SROCC is greater than zero, and when the two are completely monotonically positively

correlated, SROCC is equal to 1. Conversely, when the independent variable is negatively correlated with the dependent variable, SROCC is less than zero, and when the two are completely monotonically negatively correlated, SROCC is equal to -1 . Generally, the degree of correlation between variables can be determined by SROCC statistically, and when SROCC is in the range $[0, \pm 0.2)$, it is called weakly related; when SROCC is in the range $[\pm 0.2, \pm 0.5)$, it is called partially related; when SROCC is in the range $[\pm 0.5, \pm 0.8)$, it is called significantly related, and when SROCC is in the range $[\pm 0.8, \pm 1]$, it is called highly related [20].

In the multitimescale SROCC analysis, the PV outputs at all sampling points in a day are selected as the independent variable, and all other influence factors at the same sampling points in a day are selected as the dependent variables, respectively, to calculate SROCC in a day. In three timescales of one week, one month, and six months, the distribution rule of SROCC between each type of factors and PV output has been analyzed. Box-whisker plot is used to plot SROCC of each type of factors in three different timescales, which are selected as the first week of April, the whole month of April, and six months from April to October, and the correlation intervals are divided into different colors according to the correlation degree mentioned above, as shown in Figure 4. Box-whisker plot intuitively demonstrates the upper and lower edges, upper and lower quartiles, median, and outliers of a set of data. The shorter the box-whisker plot is, the more concentrated the data distribution is. Under different timescales, the distribution rule of SROCC tends to change.

As shown in Figure 4, SROCC of solar radiation (SR) is in the highly related interval in both long and short timescales; SROCC of temperature (TEM) and humidity (HUM) are distributed in positive and negative correlation intervals, respectively, and the interiors of the upper and lower quartiles (box region) are all located in partially related and significantly related intervals; furthermore, the distribution rules are basically consistent in different timescales. Therefore, there is a relatively obvious correlation between temperature and humidity with PV output; SROCC of atmospheric pressure (AP) is located in weakly related and partially related interval at different timescales; SROCC of wind speed (WS) shows a strong correlation in the short term, but with the increase of timescale, the distribution range of SROCC becomes larger, indicating that it has the correlation uncertainty of positive and negative; similarly,

SROCC of wind direction (WD) also presents a larger distribution range and higher volatility. According to the above analysis, the solar radiation, temperature, and humidity have a higher correlation with PV output, which can be selected as the main and effective factors affecting PV output to be added to the forecasting model.

3. The Forecasting Model Based on ELM-LSTM

3.1. Extreme Learning Machine (ELM). Extreme learning machine (ELM) is a single-hidden layer feedforward neural network (SLFN). Compared with the traditional single-hidden layer feedforward neural networks, ELM is faster and more accurate, has stronger learning and generalization ability, and can greatly shorten the forecasting period of PV output [21]. A typical ELM has a 3-layer structure: the input, hidden, and output layer, as shown in Figure 5.

For a SLFN, suppose there are N samples (t_j, X_j) , where $X_j = [x_{j1}, x_{j2}, \dots, x_{jm}]^T \in R^n$, $t_j = [t_{j1}, t_{j2}, \dots, t_{jm}]^T \in R^m$. For a SLFN with L hidden layer nodes, it can be expressed as

$$\sum_{i=1}^L \beta_i g(W_i \cdot X_j + b_i) = o_j, \quad j = 1, \dots, N \quad (1)$$

where $g(\cdot)$ is the activation function, $W_i = [w_{i1}, w_{i2}, \dots, w_{im}]^T$ is the input weight, β_i is the output weight, b_i is the bias of the i^{th} hidden layer unit, and $W_i \cdot X_j$ is the inner product of W_i and X_j .

The goal of SLFN learning is to minimize the output error, that is,

$$\sum_{j=1}^N \|o_j - t_j\| = 0, \quad (2)$$

that is, there are β_i , W_i , and b_i so that

$$\sum_{i=1}^L \beta_i g(W_i \cdot X_j + b_i) = t_j, \quad j = 1, \dots, N, \quad (3)$$

and it can be expressed in the form of a matrix:

$$H\beta = T, \quad (4)$$

where H is the output of the hidden layer nodes, β is the output weight, and T is the expected output, which can be expressed as

$$H(W_1, \dots, W_L, b_1, \dots, b_L, X_1, \dots, X_N) = \begin{bmatrix} g(W_1 \cdot X_1 + b_1) & \cdots & g(W_L \cdot X_1 + b_L) \\ \vdots & \cdots & \vdots \\ g(W_1 \cdot X_N + b_1) & \cdots & g(W_L \cdot X_N + b_L) \end{bmatrix}_{N \times L},$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m},$$

$$T = \begin{bmatrix} T_1^T \\ \vdots \\ T_N^T \end{bmatrix}_{N \times m}. \quad (5)$$

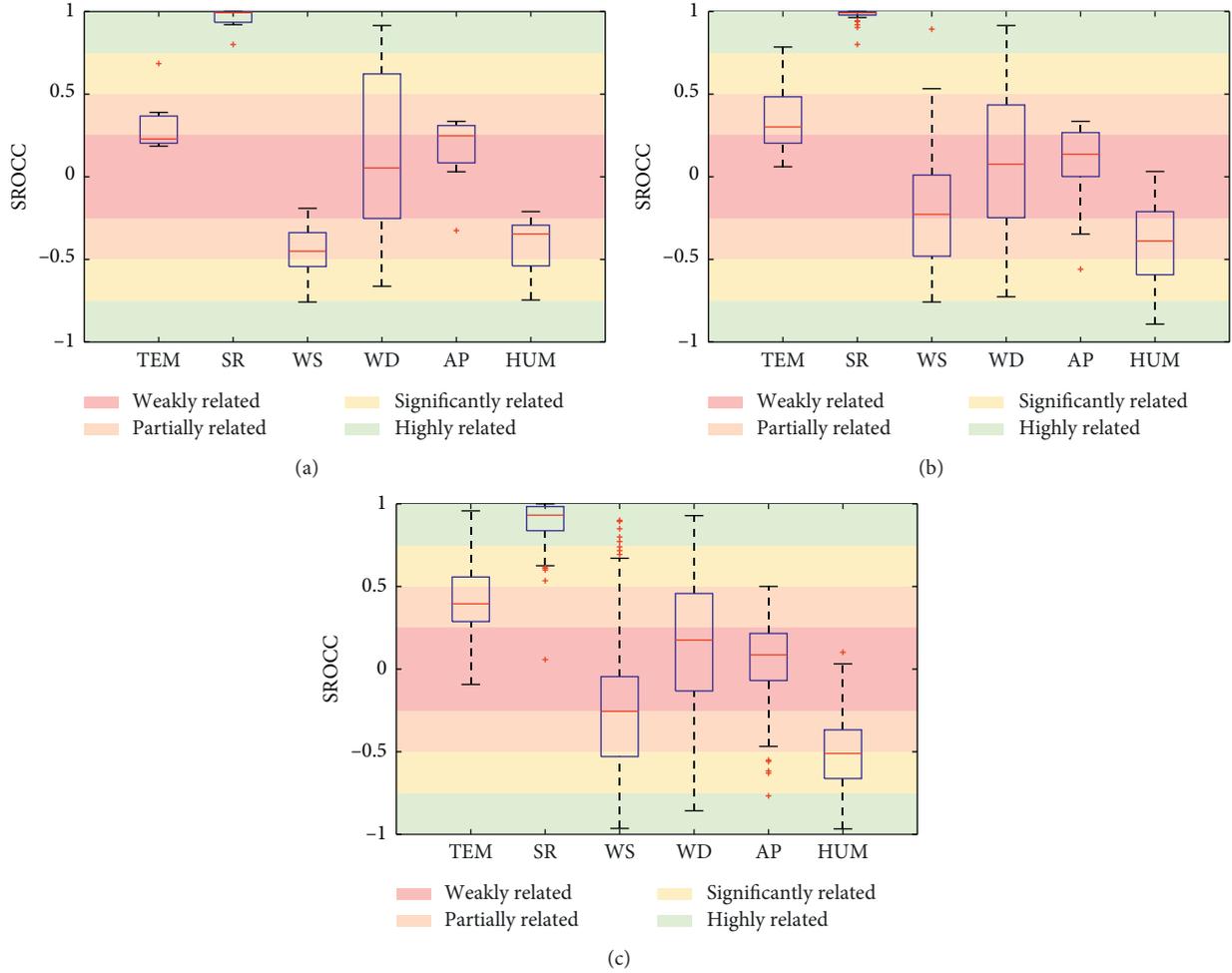


FIGURE 4: Box-whisker plot of SROCC of each type of factors in three different timescales: (a) a week, (b) a month, and (c) 6 months. TEM refers to temperature; SR refers to solar radiation; WS refers to wind speed; WD refers to wind direction; AP refers to atmospheric pressure; and HUM refers to humidity.

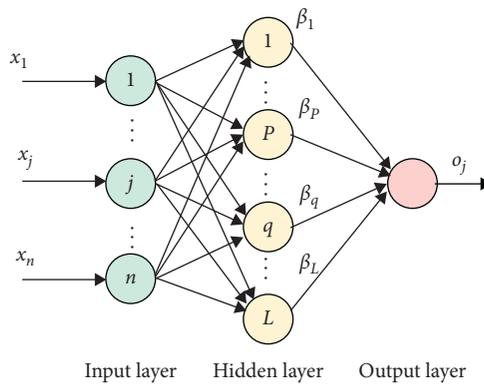


FIGURE 5: The neural network structure of extreme learning machine (ELM).

In order to be able to train SLFN, it is desirable to have \widehat{W}_i , $\widehat{\beta}_i$, and \widehat{b}_i such that

$$\|H(\widehat{W}_i, \widehat{b}_i)\widehat{\beta}_i - T\| = \min_{W, b, \beta} \|H(W_i, b_i)\beta_i - T\|, \quad i = 1, \dots, L, \quad (6)$$

and it is equivalent to minimizing the loss function:

$$E = \sum_{j=1}^N \left(\sum_{i=1}^L \beta_i g(W_i \times X_j + b_i) - t_j \right)^2. \quad (7)$$

Traditional gradient descent algorithms can be used to solve such problems, but basic gradient-based learning algorithms require all parameters to be adjusted during the iteration. In ELM, once the input weight W_i and the hidden layer offset b_i are randomly determined, the output matrix H of the hidden layer is randomly determined. The training of SLFN can be converted to solving the solution of a linear system $H\beta = T$, and the output weight $\hat{\beta}$ can be determined:

$$\hat{\beta} = H^+T, \quad (8)$$

where H^+ is the Moore-Penrose generalized inverse matrix of matrix H , and it can be proved that the norm of the solution $\hat{\beta}$ obtained is the smallest and unique.

3.2. Long Short-Term Memory (LSTM). Long short-term memory is a recurrent neural network (RNN) algorithm for deep learning which has a better effect in processing time series data. LSTM inherits the recursive property of RNN; while fully utilizing the time series data, it makes up for the shortcomings of RNN, such as gradient disappearance and gradient explosion, as well as the insufficient long-term memory ability [22]. The structure of the LSTM unit is shown in Figure 6.

The basic structure mainly includes the input gate, the forget gate, the output gate, and the memory unit, wherein x_t is the input and y_t is the output, h_t and c_t can be regarded as the short-term state and the long-term state, and g_t is a candidate value. The input gate i_t accepts the input information and updates the state of the memory unit according to different conditions. The forget gate f_t determines the discarded information according to specific condition, and the output gate o_t determines the output content according to the input information and the memory unit. The working process of the LSTM unit is as follows: at each moment, the forget gate receives the current state x_t and the hidden layer state h_{t-1} at the previous moment, and the output value of the forget gate is $[0, 1]$ through the activation function σ . When the output of the forget gate is 0, it indicates that information at the previous state is completely discarded, and when the output is 1, the information at the previous state is all retained. After the input of the input gate is transformed by the nonlinear function, it is superimposed with the output of the forget gate to get the updated memory unit c_t . Finally, the output gate can dynamically control the output h_t (y_t) of LSTM according to c_t after the operation of the nonlinear function. The calculation formulas between the variables are as follows:

$$\begin{aligned} i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i), \\ f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f), \\ o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o), \\ g_t &= \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c), \\ c(t) &= f_t c_{t-1} + i_t g_t, \\ h_t &= y_t = o_t \tanh(c_t), \end{aligned} \quad (9)$$

where W_{xf} , W_{hf} , W_{cf} , W_{xi} , W_{hi} , W_{ci} , W_{xc} , W_{hc} , W_{xo} , W_{ho} , W_{co} are weight matrices; b_i , b_c , b_f , b_o are offset parameters; and σ is an activation function, which is usually a relu or sigmoid function.

3.3. The ELM-LSTM Forecasting Model. ELM has a fast convergence speed and excellent generalization performance. However, since the bias between the input layer and the hidden layer is generated randomly, when a single ELM is used for PV output forecasting, the result may be unstable thus affecting the final forecasting accuracy. Therefore, ELM is suitable for extracting the relationship between PV input and output to make preliminary forecasting of the distributed PV system output. As a deep learning algorithm, LSTM can reflect the overall dynamic characteristics, with strong learning ability and nonlinear fitting ability. LSTM combines historical data to make adaptive adjustment to the forecasting results of ELM, so as to correct the forecasting error of ELM and further improve the forecasting accuracy.

The overall forecasting model consists of two submodels. The first submodel is the ELM forecasting model using multimodel univariate method. The multimodel univariate method is to forecast only one PV output value at each time of one model and combines forecasting values at all times by using multiple models to obtain the forecasting results of a whole day. The multimodel univariate method divides the continuous time series into multiple discrete values, and each of the models is smaller in scale hence faster in convergence. Meanwhile, considering the correlation analysis of the aforementioned influence factors affecting the PV output, the solar radiation, temperature, and humidity with higher correlation with the PV output are taken as the input nodes of the ELM model. As shown in Figure 7, assuming that the PV output value $P_{ELM}(i)$ at time i on day n is forecasted, the input nodes are selected as the solar radiation value, the temperature value, and the humidity value at times $i-1$, i , and $i+1$ of day n . Note that the meteorological data used in the forecasting model are the data on the forecast day, which is based on the currently mature weather forecasting technology, avoiding the uncertainty of using historical data to forecast future information.

The second submodel is the LSTM forecasting model which adopts the single-model multivariable method. The single-model multivariable method uses one model to forecast all the PV output values of a day in one time. Therefore, LSTM is well suited for this method due to its good forecasting performance for processing continuous time series data. Figure 8 shows the LSTM forecasting model. To forecast the PV output values at all N times of the day, the output values of N ELM models are selected as the inputs. At the same time, the influence factors with higher correlation with PV output are still added to the LSTM model to further improve the forecasting accuracy. The final output results of multiple ELM models and single LSTM model are the PV output results at all times of the forecasting day.

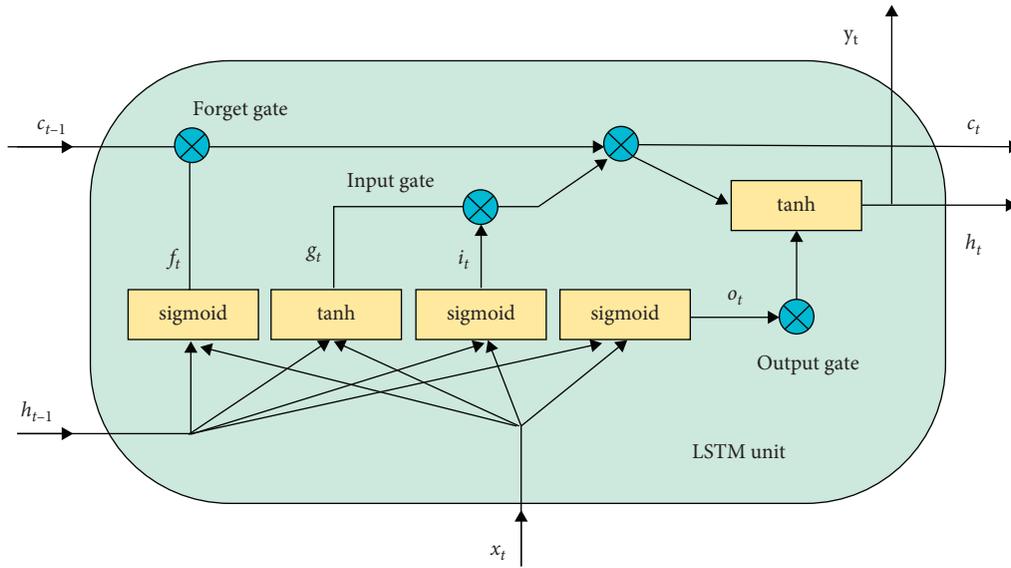


FIGURE 6: The structure of the LSTM unit.

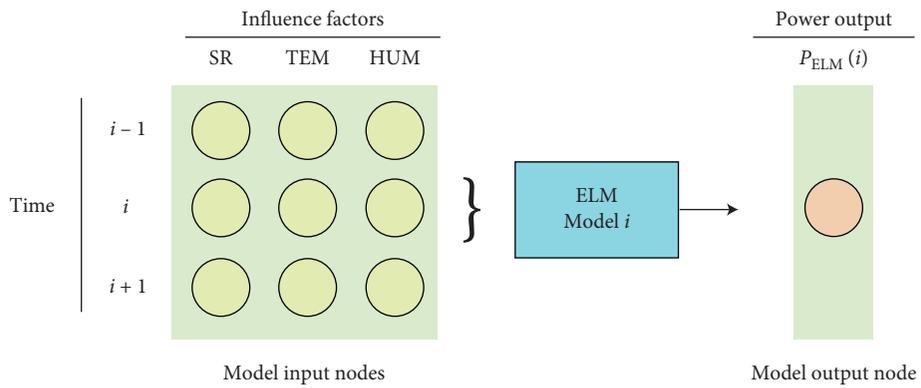


FIGURE 7: The multimodel univariate ELM forecasting model at time i of the forecasting day.

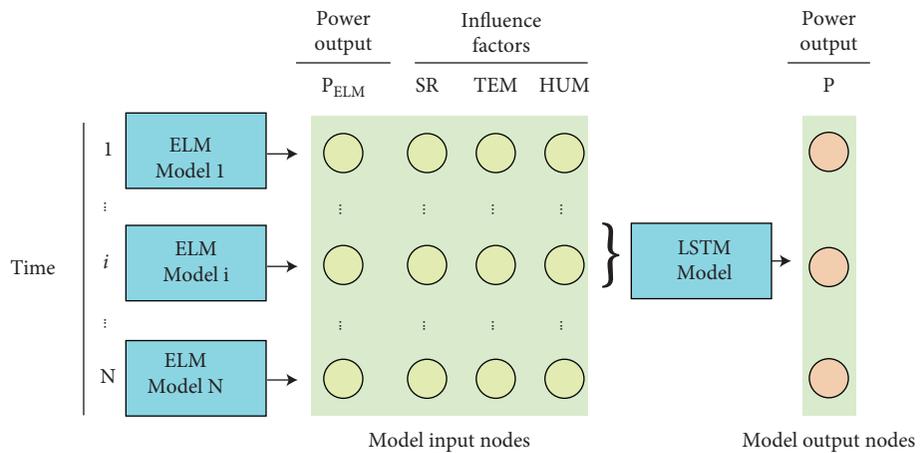


FIGURE 8: The single-model multivariable LSTM forecasting model of the forecasting day.

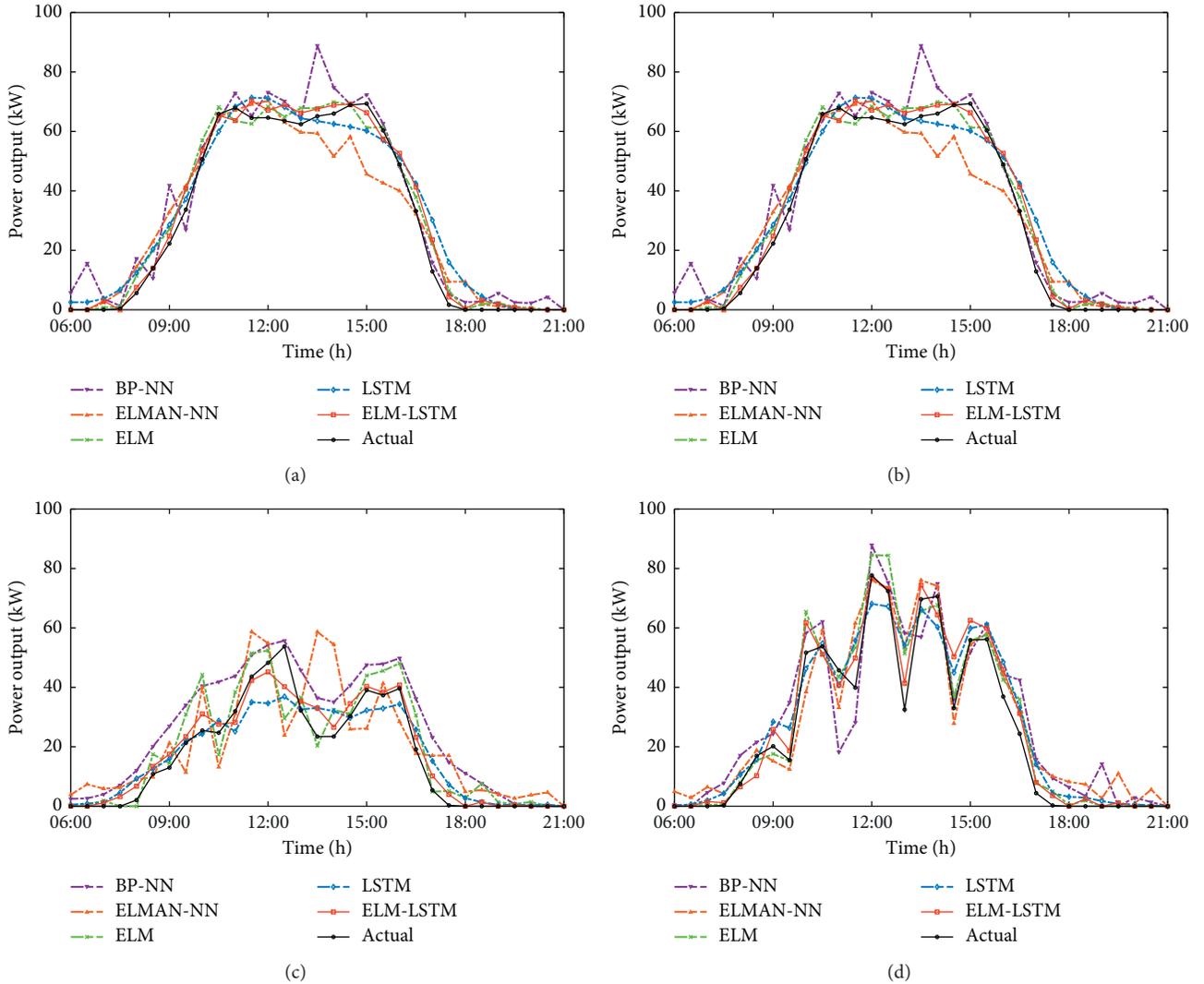


FIGURE 9: PV power output results of different forecasting models in different cases: (a) case 1, sunny day, (b) case 2, cloudy day, (c) case 3, cloudy and rainy day, and (d) case 4, day with large weather changes.

4. Case Study

4.1. Data Preprocessing and Forecasting Evaluation Criteria.

The actual output data sets of distributed PV system mentioned in Section II are divided into training sets, verification sets, and test sets according to the ratio of 0.8 : 0.1 : 0.1. The training set is used to train the ELM-LSTM forecasting model, the verification set is used to adjust the parameters of the forecasting model, and the test set is used to test the forecasting performance of the model.

Since the dimensions of each type of data in the data set are not uniform, the normalization method is needed to preprocess the data. Normalization converts all data into the range of [0, 1], which is conducive to better generalization of the ELM-LSTM forecasting model and reduction of the calculation amount of the model. The normalization formula is as follows:

$$x'_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \quad (10)$$

since the results given by the forecasting model are still normalized values, the results need to be denormalized to compare with the actual values. The denormalization formula is as follows:

$$\hat{x}_i = x''_i \times (x_{\max} - x_{\min}) + x_{\min}, \quad (11)$$

where x_i is the original data at time i ; x'_i is the data after normalization; x_{\max} and x_{\min} are the maximum and minimum values of the input samples, respectively; x''_i is the normalized forecasting value; and \hat{x}_i is the forecasting value.

Since the PV output often contains zero values, the evaluation criteria of calculating relative error, such as mean absolute percentage error (MAPE) are not applicable. Accordingly, in order to estimate the performance of the ELM-LSTM forecasting model, this paper adopts evaluation criteria of mean relative error (MRE) and root mean square error (RMSE) to evaluate the forecasting results. MRE is the evaluating index for the precision of a method for constructing fitted values, especially in trend estimation, and

TABLE 1: Statistics of forecasting results errors of MRE and RMSE.

Model	Case 1		Case 2		Case 3		Case 4	
	MRE (%)	RMSE (kW)						
BP-NN	5.435	7.071	4.908	7.702	8.272	10.434	7.522	10.746
ELMAN-NN	5.484	7.099	5.230	8.138	8.666	12.825	5.986	8.322
ELM	3.637	5.691	2.784	4.094	4.898	7.662	3.728	6.287
LSTM	3.572	5.909	4.529	6.355	4.143	6.209	5.232	7.647
ELM-LSTM	2.327	3.678	2.532	3.837	2.623	4.083	3.537	5.817

RMSE measures the average magnitude of the errors. The smaller the MRE and RMSE values are, the better the model forecasting effect and the higher the accuracy will be. The calculation formulas are as follows:

$$\text{MRE} = \frac{1}{N} \sum_{i=1}^N \frac{|P_i - \hat{P}_i|}{P_{\text{total}}} \times 100\%, \quad (12)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - \hat{P}_i)^2},$$

where P_i and \hat{P}_i are actual data and forecasting data at time i , respectively; P_{total} is the PV power capacity, in this paper the value is 106 kW; N is the number of PV output moments in one day, in this paper, the forecasting period is 6 a.m. to 9 p.m. with the interval of 30 minutes; thus $N = 31$.

4.2. Forecasting Results and Analysis. In order to evaluate the performance of the ELM-LSTM forecasting model proposed in this paper, the ELM-LSTM forecasting model is compared with the single ELM forecasting model, the single LSTM forecasting model, and two traditional neural network forecasting models: backpropagation neural network (BP-NN) and the ELMAN neural network (ELMAN-NN). In the test data set, PV outputs at all moments of sunny day (case 1), cloudy day (case 2), cloudy and rainy day (case 3), and day with large weather changes (case 4) are forecasted, and the comparison results are shown in Figure 9. The MRE and RMSE statistics of various forecasting models in four cases are given in Table 1.

It can be seen from the forecasting results that ELM-LSTM forecasting model proposed in this paper, ELM model, and LSTM model are significantly more accurate than the traditional BP-NN and ELMAN-NN models. In the case that the PV output fluctuates little (case 1 and case 2), the forecasting accuracies of ELM-LSTM model, ELM model, and LSTM model are not significantly different, but the result of ELM-LSTM model is closer to the true value. When the weather is bad or changes greatly (case 3 and case 4), the ELM-LSTM model can still maintain the best forecasting performance and have higher forecasting accuracy compared with the ELM model and LSTM model. As a one-day-ahead forecasting model and compared with [11] and [23], this paper achieves the average MRE of 2.75% rather than 8.64% and 3.29%. In addition, the forecasting results of the ELM-LSTM model can meet the industry requirements in four cases, that is, the RMSE of short-term PV output forecasting should be less than 20%

[24]. In conclusion, the ELM-LSTM forecasting model proposed in this paper, based on the single ELM and LSTM forecasting models, fully utilizes the fast learning ability and excellent generalization performance of ELM for preliminary forecasting, then learns historical data according to the deep learning algorithm of LSTM, makes self-adaptive adjustment, and finally improves the forecasting accuracy compared with the two independent models.

5. Conclusion

This paper proposes a one-day-ahead short-term power output forecasting model based on correlation analysis and ELM-LSTM for distributed PV system.

- (1) Based on the actual data of the distributed PV system, the multitime-scale Spearman rank-order correlation coefficient analysis method is used to select the external influence factors with higher correlation with PV output and provides them to the forecasting model, which reduces the complexity of the forecasting model.
- (2) Based on the multimodel univariate ELM submodel and the single-model multivariate LSTM submodel, the composite ELM-LSTM forecasting model is established, which utilizes the ELM model for information generalization and preliminary forecasting and the LSTM model for adaptively adjusting to improve the whole forecasting accuracy.

Compared with different forecasting models through the actual data, the ELM-LSTM forecasting model has the highest accuracy under various weather conditions.

Data Availability

The PV output power and meteorological information data used to support the findings of this study have been deposited in the Figshare repository ([DOI: 10.6084/m9.figshare.10119314]).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by Science and Technology Research Program of Shenzhen Power Supply Co., Ltd. (no. 0002200000042674).

References

- [1] IRENA, *Renewable Power Generation Costs in 2018*, International Renewable Energy Agency (IRENA), Abu Dhabi, UK, 2019, <https://www.polity.org.za/article/renewable-power-generation-costs-in-2018-2019-05-31>.
- [2] M. A. Eltawil, Z. Zhao, and S. E. reviews, "Grid-connected photovoltaic power systems: technical and potential problems-A review," *Renewable and Sustainable Energy Reviews*, vol. 14, no. 1, pp. 112–129, 2010.
- [3] A. Yona, T. Senjyu, T. Funabashi, P. Mandal, and C.-H. Kim, "Decision technique of solar radiation prediction applying recurrent neural network for short-term ahead power output of photovoltaic system," *Smart Grid and Renewable Energy*, vol. 4, no. 6, pp. 32–38, 2013.
- [4] B. Amrouche and X. Le Pivert, "Artificial neural network based daily local forecasting for global solar radiation," *Applied Energy*, vol. 130, no. 5, pp. 333–341, 2014.
- [5] E. Akarslan and F. O. Hocaoglu, "A novel adaptive approach for hourly solar radiation forecasting," *Renewable Energy*, vol. 87, pp. 628–633, 2016.
- [6] W. F. Holmgren, A. T. Lorenzo, M. Leuthold, C. K. Kim, A. D. Cronin, and E. A. Betterton, "An operational, real-time forecasting system for 250 MW of PV power using NWP, satellite, and DG production data," in *Proceedings of the 2014 IEEE 40th Photovoltaic Specialist Conference (PVSC)*, pp. 0080–0084, Denver, CO, USA, June 2014.
- [7] S. Tiwari, R. Sabzehgar, and M. Rasouli, "Short term solar irradiance forecast using numerical weather prediction (NWP) with gradient boost regression," in *Proceedings of the 2018 9th IEEE International Symposium on Power Electronics for Distributed Generation Systems (PEDG)*, pp. 1–8, Charlotte, NC, USA, June 2018.
- [8] C. K. Kim, H. Kim, Y. Kang, and C. Yun, "Toward improved solar irradiance forecasts: evaluation of operational numerical weather prediction model for solar irradiance over the Korean Peninsula," in *Proceedings of the 2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC) (A Joint Conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC)*, pp. 2317–2319, Waikoloa Village, HI, USA, June 2018.
- [9] M. Abuella and B. Chowdhury, "Solar power forecasting using artificial neural networks," in *Proceedings of the North American Power Symposium*, Charlotte, NC, USA, October 2015.
- [10] T. Demirdelen, I. O. Aksu, B. Esenboga, K. Aygul, F. Ekinci, and M. Bilgili, "A new method for generating short-term power forecasting based on artificial neural networks and optimization methods for solar photovoltaic power plants," in *Solar Photovoltaic Power Plants*, pp. 165–189, Springer, Berlin, Germany, 2019.
- [11] S. Jie, W. J. Lee, Y. Liu, Y. Yang, and P. Wang, "Forecasting power output of photovoltaic systems based on weather classification and support vector machines," *IEEE Transactions on Industry Applications*, vol. 48, no. 3, pp. 1064–1069, 2012.
- [12] A. T. Eseye, J. Zhang, and D. Zheng, "Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and meteorological information," *Renewable Energy*, vol. 118, pp. 357–367, 2018.
- [13] J. V. Paatero and P. D. Lund, "Effects of large-scale photovoltaic power integration on electricity distribution networks," *Renewable Energy*, vol. 32, no. 2, pp. 216–234, 2007.
- [14] S. Qijun, L. Fen, Q. Jialin, Z. Jinbin, and C. Zhenghong, "Photovoltaic power prediction based on principal component analysis and Support Vector Machine," in *Proceedings of the 2016 IEEE Innovative Smart Grid Technologies—Asia (ISGT-Asia)*, pp. 815–820, Melbourne, VIC, Australia, December 2016.
- [15] R. A. Messenger and A. Abtahi, *Photovoltaic Systems Engineering*, CRC Press, Boca Raton, FL, USA, 2010.
- [16] A. Sanz, I. Vidaurrazaga, A. Pereda, R. Alonso, E. Roman, and V. Martinez, "Centralized vs distributed (power optimizer) PV system architecture field test results under mismatched operating conditions," in *Proceedings of the 2011 37th IEEE Photovoltaic Specialists Conference*, pp. 002435–002440, IEEE, Seattle, WA, USA, June 2011.
- [17] A. Yona, T. Senjyu, and T. Funabashi, "Application of recurrent neural network to short-term-ahead generating power forecasting for photovoltaic system," in *Proceedings of the 2007 IEEE Power Engineering Society General Meeting*, Tampa, FL, USA, June 2007.
- [18] S. N. Zhang and S. Y. Tian, "The institution of the hourly solar radiation model," *Acta and MVS Energy Solutions*, vol. 18, pp. 273–277, 1997.
- [19] P. H. Ramsey, "Critical values for spearman's rank order correlation," *Journal of Educational Statistics*, vol. 14, no. 3, pp. 245–253, 1989.
- [20] J. L. Myers, A. D. Well, and R. F. Lorch Jr., *Research Design and Statistical Analysis*, Routledge, Abingdon, UK, 2013.
- [21] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *Proceedings of the 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541)*, vol. 2, pp. 985–990, Budapest, Hungary, July 2004.
- [22] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [23] H.-T. Yang, C.-M. Huang, Y.-C. Huang, and Y.-S. Pai, "A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 3, pp. 917–926, 2014.
- [24] T. Daye, *Managing Intermittency: Standards and Recommended Practices in Solar Power Forecasting*, ERCOT, Austin, TX, USA, 2011.