

Article

Very-Short-Term Power Prediction for PV Power Plants Using a Simple and Effective RCC-LSTM Model Based on Short Term Multivariate Historical Datasets

Biaowei Chen ^{1,2}, Peijie Lin ^{1,2,*}, Yunfeng Lai ^{1,2}, Shuying Cheng ^{1,2}, Zhicong Chen ^{1,2} and Lijun Wu ^{1,2}

¹ School of Physics and Information Engineering, and Institute of Micro-Nano Devices and Solar Cells, Fuzhou University, Fuzhou 350108, China; chen.b.wei@163.com (B.C.); Yunfeng.Lai@fzu.edu.cn (Y.L.); N171127055@fzu.edu.cn (S.C.); yongcheng.wang9527@163.com (Z.C.); luxiaoyang_fzu@sina.com (L.W.)

² Jiangsu Collaborative Innovation Center of Photovoltaic Science and Engineering, Changzhou 21316, China

* Correspondence: linpeijie@fzu.edu.cn

Received: 30 December 2019; Accepted: 3 February 2020; Published: 8 February 2020



Abstract: Improving the accuracy of very-short-term (VST) photovoltaic (PV) power generation prediction can effectively enhance the quality of operational scheduling of PV power plants, and provide a reference for PV maintenance and emergency response. In this paper, the effects of different meteorological factors on PV power generation as well as the degree of impact at different time periods are analyzed. Secondly, according to the characteristics of radiation coordinate, a simple radiation classification coordinate (RCC) method is proposed to classify and select similar time periods. Based on the characteristics of PV power time-series, the selected similar time period dataset (include power output and multivariate meteorological factors data) is reconstructed as the training dataset. Then, the long short-term memory (LSTM) recurrent neural network is applied as the learning network of the proposed model. The proposed model is tested on two independent PV systems from the Desert Knowledge Australia Solar Centre (DKASC) PV data. The proposed model achieving mean absolute percentage error of 2.74–7.25%, and according to four error metrics, the results show that the robustness and accuracy of the RCC-LSTM model are better than the other four comparison models.

Keywords: photovoltaic power generation; long short-term memory; very short-term Power prediction; similarity time period

1. Introduction

Developing renewable energy can effectively reduce dependence on fossil energy and other burning energy sources, thereby improving the world's energy and economic security [1,2]. Thus, the importance of developing new energy sources is increasingly prominent [3,4]. Due to its clean, safe, and sustainable characteristics, photovoltaic (PV) power generation is still receiving continuous attention worldwide. PV technology has been improved in material [5] and maintenance strategy [6,7] in recent years. According to the latest data [8], the global installed capacity of new PV has reached 100 GW in 2018, accumulated to 505 GW. Among them, the newly installed capacity of PV in China, US, Japan and Germany reached 45 GW, 10.6 GW, 6.5 GW, and 3.0 GW, respectively. However, due to the power output of the PV power generation system is largely affected by environmental factors, the economic benefits of the PV plant depend on the flexibility of PV power systems [9]. In order to improve the flexibility of the demand side and supply side in the PV market, increasing the resolution and accuracy of PV Power generation predictions becomes critical and urgent [10].

When there is a high penetration level of PV power energy, the fluctuations caused by PV power energy and flexible loads degrade the frequency performance of modern power systems [11]. While the predicting technology related to PV energy can not only provide a reference for the price of electricity for a PV plant [12], as well as cooperate with power grid dispatching and power load distribution [13,14], it could also help for the scientific planning and operation of the entire power system. Nottrott et al. [15] developed a linear programming routine to optimize the energy storage dispatch schedule for a grid-connected, which is based on PV power and load forecasting. In addition, Shibani Ghosh et al. propose a local voltage regulation technique that adaptively changes real/reactive injection/absorption from the PV inverter. This method is based on 15-s PV generation forecasting to prevent distribution overvoltage [16].

Over the past decade, various technologies for PV generation forecasting have been presented, which can be roughly separated into physical modeling approaches and data-driven models [17]. The traditional physical modeling approaches focus on studying equivalent circuits of PV cells. And then deriving the power output based on numerical weather prediction input parameters (includes temperature, humidity, and global radiation among others) [18]. This kind of method can be applied to different time scales PV generation forecasting. However, the physical method involves complex calculations and requires numerous circuit parameters (includes series, shunt resistors, various temperature coefficients and diode influence factors). Therefore, it's hard to ensure robustness due to the diversity of parameters among different power stations, which obstructs the development of the physical PV forecasting method. On the other hand, the data-driven models have greater feasibility than the physical method own to the advantages given by artificial intelligence methods, they are now widely used in different forecasting applications.

According to the raw data used, the data-driven models can be divided into three types: time-series, sky imaging, and numerical weather prediction (NWP). The time-series models are dependent on long-term historical data, which are usually applied in long-term or medium-term PV power forecasting. The most commonly used model is linear models based on the auto-regressive (AR) method which is simple but inflexible. Li and Su employ the auto-regressive integrated moving average (ARIMA) model to estimate monthly solar radiation based on the measured radiation and temperatures dataset [19]. In addition, in [20], a partially functional linear regression model (PFLRM) for predicting the daily production of a PV system is proposed. Furthermore, there are also some nonlinear methods based on time-series. For example, in [21], based on the historical time-series data-set, the traditional ANN and support vector regression (SVR) are employed to improves prediction accuracy at a 24-h scale. Furthermore, in [22], an adaptive learning hybrid model (ALHM) model combine meteorological data is applied to improve the accuracy of long-term prediction of solar intensity. Compared to the linear methods, the nonlinear approaches can notably improve the accuracy of forecasting since the ability of adaptability and self-update.

Moreover, owing to the development of image technology, both the sky imaging methods and satellite monitoring techniques are applied to the field of PV power prediction [23]. This kind of model is more suitable for short-term and VST PV power forecasting [24]. Caldas et al. used sky imaging to monitor cloud motion, then forecasted 1–10 min real-time irradiance value according to the empirical formula [25]. In addition, Crisosto et al. [26] established a method using the artificial neural network (ANN) and sky imaging to predict solar irradiance at an hour scale and the result shows the outstanding prediction performance of the proposed method. However, the image technology-based model requires more calculation and higher equipment cost and may also encounter camera occlusion problems, which makes it difficult to further promote.

In addition, PV prediction based on NWP represents one of the most promising research directions of PV power forecasting [27–29]. The NWP based PV forecasting method usually classify the weather types [30] and then input the NWP data into the trained model for prediction. A prediction model based upon the grey relational analysis (GRA) algorithm and support vector machine (SVM) is proposed in [31]. This model applies the PV power of the similar day sample and the meteorological factor of

the prediction day as inputs dataset. The medium-range weather forecasting and ANN are employed to forecast 72 h PV power output [32]. However, the precise of these prediction models rely on the accuracy of weather prediction results [33], if the weather prediction result is not accurate or the weather prediction information is difficult to obtain, the applicability and accuracy of such models may be affected.

Generally, the choice of prediction model often depends on the application scenario. However, due to the PV power generation being almost entirely determined by environmental factors [34], the PV power prediction method based on meteorological data has become the main orientation of current research. The presented models in many recent literatures try to search the nonlinear relationship between power generation and its influencing factors based on learning networks (LNs) [35,36]. Typical LNs structures include back propagation neural network (BPNN), radial basis function neural network (RBFNN), and Elman neural network (Elman) among others. Among them, BPNN is a basic but widely used neural network [37]. And in [38], the model based on historical power datasets. The GRA algorithm is applied to classify meteorological types. According to the experiment threshold value to select similar day samples as the training dataset and the Elman is used as a learning network of the prediction model, the results verify the reliability of Elman in PV power prediction. RBFNN is also a network structure with a good learning ability for nonlinear problems [17]. Its excellent performance in the forecast has been proven in many studies, as the model in [39], it using solar radiation and panel module temperature combined with two RBFNN networks to predict the daily PV array power generation in 26 different cities. At the same time, the deep learning network has the powerful ability to learn and discover potential relationships [40], Mei Fei et al. propose a model based on phase space reconstruction and deep learning neural network (DLNN) to predict ultra-short-term net load [41]. The prediction of short-term power fluctuations by recurrent neural network LSTM is presented in recent research [42], which proves the application potential of this model in short-term prediction tasks. It is evident that the predictive model based on the learning network is the main direction of the future development of PV power forecasting technology.

With regard to the time horizon, the forecast scales are usually defined as the following types [43]: (1) Long-term and Medium-term prediction is adopted as a reference for the planning of PV power plants and the setting of some parameters, the horizon ranged from a week to a year; (2) Short-term horizon ranged from a few hours to a week: this prediction scale is focused on the PV power generation grid-connected scheduling and economic dispatching; (3) Very-short-term (VST) PV power forecasting: the horizon is regarded as several minutes to several hours [44]. This forecasting scale can be employed as a reference for the power supply market. It also used for the real-time emergency response and health monitoring of PV systems [45]. With the improvement of computing power and the rise of 5G communication technology in recent years, the rhythm of human social activities has accelerated. Both the energy application and the supply side hope to obtain updated and accurate real-time data. This demand drives the development of VST PV real-time prediction technology [46,47], which could ensure more stable and reliable operation of the PV system.

As previously mentioned, obviously, the learning network models have outstanding performance on various forecasting tasks [48]. Unlike the physical method, the LNs model does not require a complete understanding of the photoelectric conversion relationship inside the PV system in advance. However, the prediction accuracy of the LNs model usually relies on the calculation of higher dimensions to ensure the effect, which increases the calculation and slows down the prediction speed. On the other hand, for a real-time VST forecasting model, reasonable training time is required, and there are two methods that can be applied to solve this contradiction: reducing training sample size and improving the quality of samples. Thus, the influence of different meteorological parameters on PV power output in a short period of time is firstly analyzed. Then, in order to reduce the training time and improve the accuracy of predictions, the dataset of similar time periods is selected according to the relationship between short-term radiation coordinates and power output. Considering the advantage and flexibility of LSTM, this model applies LSTM as the learning network. Furthermore, the proposed

RCC-LSTM model does not require very long-term historical data, which allows the system to forecast VST PV power generation after a short-period of self-running, and the sliding window enables the predictive model to be self-updated in real-time and adapt to the natural attenuation of PV systems.

The main achievements of this work can be summarized as follows:

(1) We present a new method for VST PV power forecasting that combines similar time period collection using the RCC algorithm with neural network learning prediction algorithms. The models use only the previous PV power data and meteorological data, i.e., solar radiation, temperature, and humidity. A notable advantage of our method is that it uses only variables that are easily obtainable (previous PV power and simple weather data). In comparison to other methods, it does not use future weather predictions that are not always available for all PV plant. or sky images that require special equipment to be processed and recorded and

(2) On the five minutes time resolution, the correlation between different meteorological data and power output at different time periods was explored. The specified time point radiation coordinates which had the highest correlation with power output are further proposed. The similar time period collected by the RCC method is used as training sample for the prediction model. This method reduces the calculation cost of the model and enhances the prediction accuracy.

(3) Based on the dataset from two independent PV systems, a comprehensive comparative study is conducted comparing the proposed method with mainstream data-driven methods, including RCC-BPNN, RCC-Elman, RCC-RBFNN models and LSTM model on all four seasons. The experimental results show that the proposed RCC-LSTM model has an obvious advantage in forecasting accuracy.

The rest of the work is organized as follows. Section 2 presents the materials and methods in this paper. Section 3 describes the proposed methodology in this paper. Section 4 shows the experimental results and provides an analysis and comparison of the test results. Finally, the conclusions and future work are presented in Section 5.

2. Materials and Methods

2.1. The Description of The Experimental Data

This paper uses the measured data from the YULARA PV system in Alice Springs, Australia at a latitude of $22^{\circ}79'$ S and a longitude of $130^{\circ}16'$ E. In order to verify the scope and robustness of the proposed model, two separate systems (i.e., 3A and 4) with different PV technologies and panel ratings are selected. Figure 1 shows the map of the systems and their power generation ranges are 22.56 KW and 327.6 KW, respectively. The detailed information of these two systems is shown in Table 1.

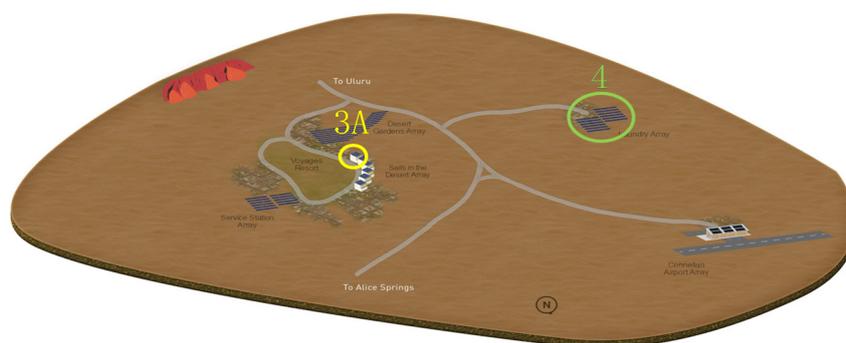


Figure 1. The map of the two systems.

Table 1. The specific information of two systems (3A and 4).

Parameter	SITE 3A	SITE 4
PV Technology	mono-Si	poly-Si
Panel rating	327 W	315 W
Number of panels	69	1040
Array rating	22.56 kW	327.6 kW
Panel type	SunPower SPR-327NE	Jinko Solar JKM315PP
Array Structure	Fixed: Roof Mount	Fixed: Ground Mount
Inverter size/type	SMA STP 25000TL-30 & 20000TL-30	SMA STP 25000TL-30
Installation completed	Sat, 19 Mar 2016	Thu, 3 Mar 2016
Array Tilt/Azimuth	Tilt = 10, Azi = 0 (Solar North)	Tilt = 20, Azi = 0 (Solar North)

The data of the two consecutive years (2017 and 2018) are chosen for this experiment and can be download from [49]. The resolution of the historical dataset is 5-min, and the data mainly includes active power (KW), temperature (°C), relative humidity (%), global horizontal radiation ($w/m^2 \times sr$), and diffuse horizontal radiation ($w/m^2 \times sr$).

2.2. General Structure of the Proposed Model

The detailed overall structure of the proposed method is described in Figure 2. To further understand the details of the method, an additional description of each part is given in this section.

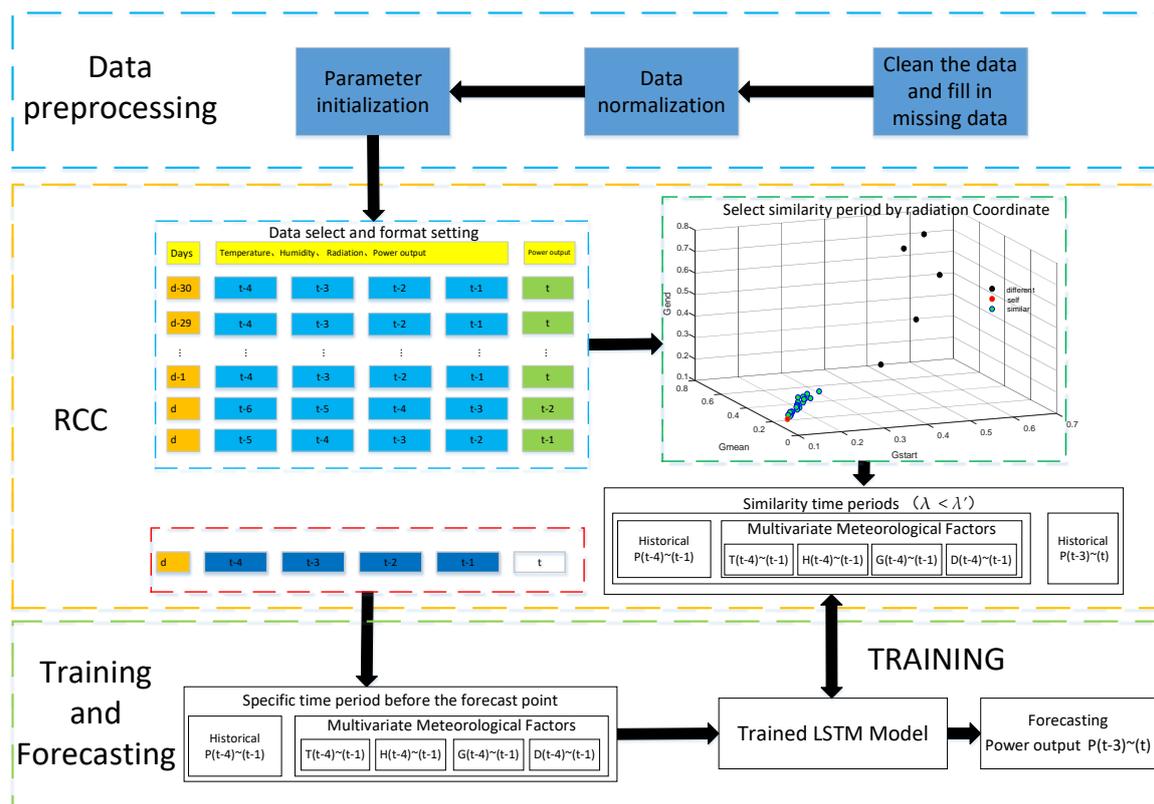


Figure 2. General structure of the RCC-LSTM forecasting model.

2.2.1. Data Preprocessing

During the training process of the deep learning network, the quality of the training data will affect the accuracy of the prediction model. Therefore, the training-data should be preprocessed before it is transmitted to the network, which includes the cleaning of the abnormal data (such as PV panel anomalies) and filling in missing data (such as system and equipment failure). After that, in order to

meet the data requirements of the training network and to avoid the unbalanced data distribution caused by different unit ranges of the different feature vectors, these types of data are normalized into the same unit of measurement. Finally, the training dataset is rearranged according to the PV output sequence and the structure of the neural network.

2.2.2. Radiation Coordinate Classification Method

PV power output is related to many factors [50], including some meteorological factors, type of PV module, the installation structure, and the working characteristics of the PV module among others and it is almost impossible to include all the influencing factors. However, it is easy to understand that the weight of these factors on the PV power output is not constant. They act differently under different time periods and different weather conditions. Moreover, the natural attenuation of PV systems has a certain effect on their degree of impact.

This paper selects three random days for analysis in four quarters (includes January 15, February 22, December 20, summer in Australia, April 30, May 20, June 1, Autumn in Australia, July 7, July 18, August 6, winter in Australia, and September 1, September 22, spring in Australia). Time range from 8:30–17:30, 109-time points in total, four representative and easily available meteorological factors are selected for comparison. The correlations between different features with power output at different time period are calculated by $\rho_{d,p}$, and the average of three days was randomly selected.

The $\rho_{d,p}$ is defined as follows:

$$\rho_{X,Y} = \frac{N \sum XY - \sum X \sum Y}{\sqrt{N \sum X^2 - (\sum X)^2} \sqrt{N \sum Y^2 - (\sum Y)^2}}, \tag{1}$$

where X represents the different meteorological influential factors, Y represents the power generation, and N represents the number of time points.

The correlation coefficient ranges from 0.8 to 1.0, the representation has a strong correlation, the strong correlation at 0.6–0.8, the medium correlation between 0.4–0.6, the weak correlation at 0.2–0.4. And the very weak correlation at 0–0.2. Figure 3 represents the correlation coefficient of four different meteorological factors with PV power generation at different time periods in different seasons. In general, it can be seen that the correlation between different impact factors and PV power output also changes in different seasons. Their respective influences on PV power output are also constantly changing, which further explains why the linear models are difficult to solve.

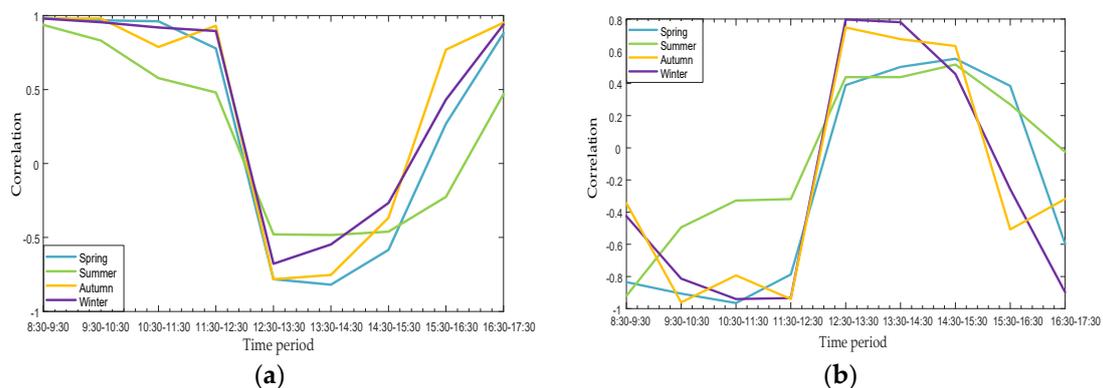


Figure 3. Cont.

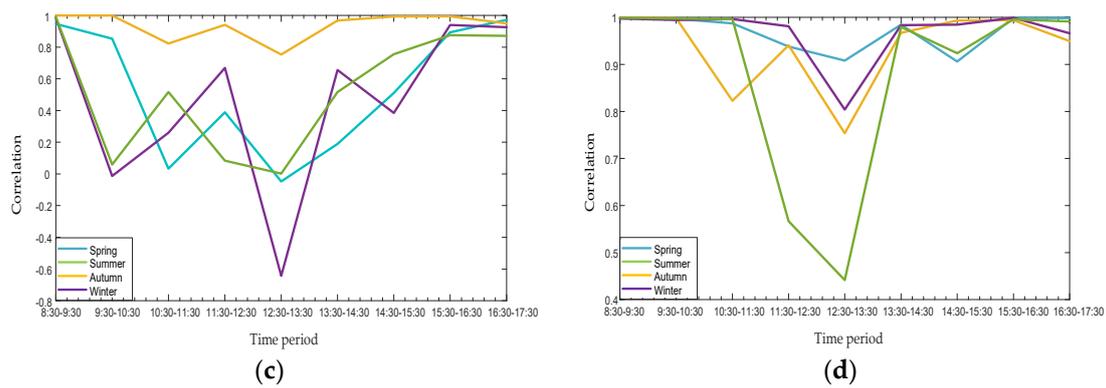


Figure 3. (a) Correlation between PV power and air temperature; (b) Correlation between PV power and relative humidity; (c) Correlation between PV power and diffuse horizontal radiation; (d) Correlation between PV power and global horizontal radiation.

Figure 3a illustrates the correlation between temperature and PV power generation over different hours of the day. It can be seen that there is a certain degree of similarity between the four seasons, but there are still minor differences, and the correlation varies from -0.75 to 1 during a single day, this fluctuation is large. At the same time, in Figure 3b. The fluctuation range of correlation between humidity and PV power generation over the different time period is between -1 and 0.75 , there are also subtle differences in four seasons. Moreover, the Figure 3c represents the correlation between diffuse horizontal radiation and PV power generation. The correlation is in the range from -0.6 to 1 and there is a kind of regulation within one day, but the degree of correlation in different seasons is quite different. However, as shown in Figure 3d. The correlation between global horizontal radiation and PV power output always keeps a high value, most are distributed between 0.8 and 1.0 , a few are between 0.4 and 0.7 . Thus, it is more reasonable to use global horizontal radiation to collect similar time periods.

Therefore, in order to ensure that the forecasting model can be adapted to different seasons while learning the slight gaps in different time periods. The method separates the different time periods and uses the sliding window for VST power prediction.

The range of climate parameters in a short period of time is small, and climate change is relatively stable in the northwestern region of the PV generation. It may work to speculate on the climate of the predicted point by analyzing the climate situation before the predicted point. On this basis, set the origin $(0,0,0)$ as the reference point, and select the global horizontal radiation which has the highest correlation with power output. By setting the global radiation in the time period for different time points to the coordinates of the start, mean, and end values, Euclidean distance can be calculated between these coordinates and the origin as follows:

$$d_i = \sqrt{(G_{start_i})^2 + (G_{mean_i})^2 + (G_{end_i})^2}, \tag{2}$$

where G_{start_i} , G_{mean_i} and G_{end_i} are the start, average, and end values of the global horizontal radiation for the time period $(i-n, i-1)$ before the predicted time point, respectively. n is the number of time point in the selected time period. Five conditions with n equal to 2, 3, 4, 5, and 6 are selected for verification. i represents the number of times during the day of the test.

Calculate the correlation between d and p .

$$\rho_{d,p} = \frac{\sum_{i=1}^N d_i p_i - \sum_{i=1}^N d_i \sum_{i=1}^N p_i}{\sqrt{\sum_{i=1}^N d_i^2 - \left(\sum_{i=1}^N d_i\right)^2} \sqrt{\sum_{i=1}^N p_i^2 - \left(\sum_{i=1}^N p_i\right)^2}}, \tag{3}$$

where p is the PV power output value at time-step i .

As shown in Table 2. The correlation between the d of different scale radiation coordinates and the PV power output p at the next moment is obtained. In Figure 4, it can be observed that they have high correlation value, more than that, the $\rho_{d,p}$ mean value and standard deviation of different time-steps are shown in Figure 4a,b, and when the time point is 4, the correlation is higher and the stability is better. Thus, the 4-time points before the prediction point are selected as the analysis time period.

Table 2. The correlation between the d and p under different seasons and different time scales.

Time-steps(n)	Seasons			
	Spring	Summer	Autumn	Winter
2	0.9936	0.8569	0.9896	0.9950
3	0.9937	0.8729	0.9914	0.9958
4	0.9928	0.8842	0.9924	0.9954
5	0.9901	0.8821	0.9919	0.9937
6	0.9881	0.8784	0.9905	0.9909

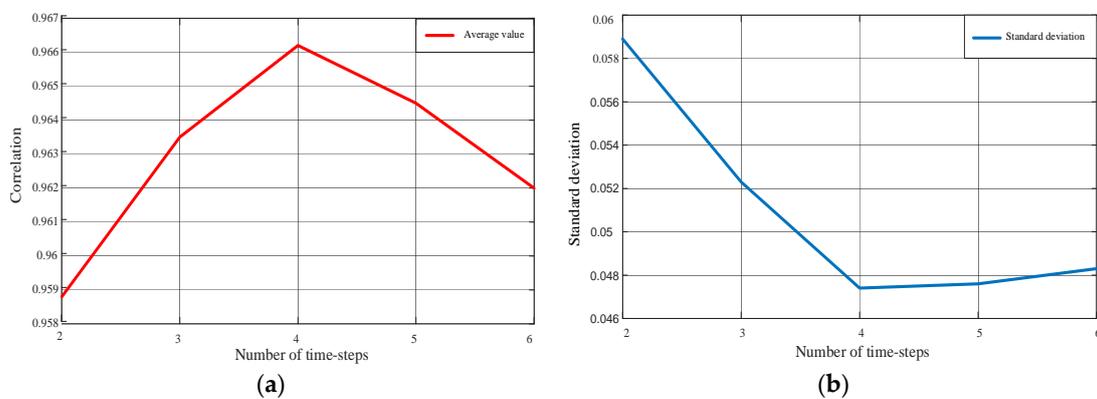


Figure 4. (a) The mean value of $\rho_{d,p}$ in different time-steps; (b) Standard deviation of $\rho_{d,p}$ in different time-steps.

According to the above-mentioned characteristics, the radiation coordinate classification (RCC) method is proposed as the classification method for selecting similar time periods. The corresponding data of a similar time obtained is reconstructed into a training data set. After training, input the corresponding data of this period to predict power output. The specific process is as follows:

Firstly, collect the data in the same time period of the last 30 days before the day, and the first two time periods before the target time point. Reconstitute these data into a feature array A which is composed of $(P_t, T_t, H_t, G_t, D_t, P_{t+1})$. The structure of array A is shown below:

$$A = \begin{pmatrix} A_{d-30} \\ A_{d-29} \\ \vdots \\ A_{d-1} \\ A_d \end{pmatrix}, A_d = \begin{pmatrix} P_{t-4} & T_{t-4} & H_{t-4} & G_{t-4} & D_{t-4} & P_{t-3} \\ P_{t-3} & T_{t-3} & H_{t-3} & G_{t-3} & D_{t-3} & P_{t-2} \\ P_{t-2} & T_{t-2} & H_{t-2} & G_{t-2} & D_{t-2} & P_{t-1} \\ P_{t-1} & T_{t-1} & H_{t-1} & G_{t-1} & D_{t-1} & P_t \end{pmatrix} \quad (4)$$

Then, the power and meteorological parameters in the feature array are normalized. The normalization formula is defined as:

$$A_{kl_new} = \frac{A_{kl} - A_{kmin}}{A_{kmax} - A_{kmin}}, \quad (5)$$

where A_{kl_new} represents the data obtained after normalization, A_{kl} is the specific value of the power and meteorological data, k represents the star, mean or end value listed, l indicates the number of hours

in the unit time period, A_{kmin} and A_{kmax} are the minimum and maximum values of the meteorological data in the feature column.

Secondly, by defining the radiation classification feature coordinates as $(G_{start}, G_{mean}, G_{end})$, the definition of each parameter is the same as above.

Combine the 32 time periods into corresponding three-dimensional vectors. Calculate the Euclidean distance λ between these 32 feature coordinates and the target period feature coordinates. The formula is defined as:

$$\lambda = \sqrt{(G'_{start} - G_{start})^2 + (G'_{mean} - G_{mean})^2 + (G'_{end} - G_{end})^2}, \tag{6}$$

where G'_{start} , G'_{mean} , and G'_{end} represent the start, mean, and end values of global horizontal radiation for a specified time period before the predicted time, respectively.

Compare them with the value of λ , set the experience threshold λ' and select the time period which λ is less than λ' . The meteorological and power data corresponding to these time periods are used as the dataset for training network.

The data include the temperature (T_t), relative humidity (H_t), global horizontal radiation (G_t), diffusion horizontal radiation (D_t), PV power output (P_t), and PV power output at the next time point P_{t+1} of each step. These data have been normalized, and they are arranged from far to near as the training set of the prediction network according to the distance from the target time period.

2.2.3. LSTM Recurrent Neural Network

As one of the most advanced recurrent neural networks, the Long Short-term Memory (LSTM) recurrent neural network has shown remarkable results in numerous time-series learning tasks [51,52]. Unlike the neurons of traditional recurrent neural networks, the LSTM has memory blocks connected by successive layers, and it enables the network to selectively memorize the input training data through a unique three-gate structure. These structures ensure that the network structure can learn multivariate influences of nonlinear tasks. In addition, the cascade structure of the LSTM makes it has an excellent performance in time series problems. For example, there are some good examples of forecasting work based on LSTM. Wang et al. [53] establish a hybrid day-ahead PV power forecasting model based on CNN and LSTM. This model uses CNN first extracts local features of data and applies LSTM to extracts the overall timing features of data, and the prediction performance is outstanding.

Fundamentally, there are three logic gate structures in every single cell, including forgetting gate, input gate and the output gate. And each operation process mainly includes four sub-operations, as shown in Figure 5.

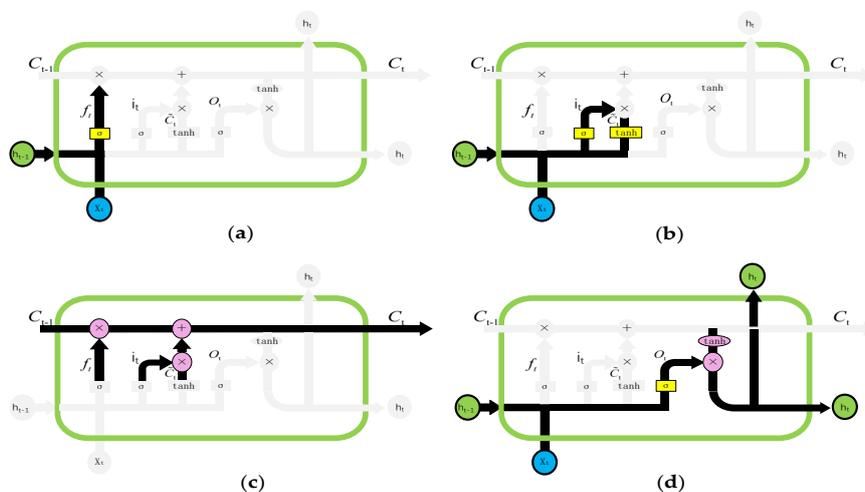


Figure 5. (a) Forget gate; (b) Input gate; (c) Merge process; (d) Output gate.

The formula corresponding to each part of the operation is as follows [54]:

Forget gate:

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

Input gate:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \tag{8}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \tag{9}$$

Merge process:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t, \tag{10}$$

Output gate:

$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o), \tag{11}$$

$$h_t = o_t * \tanh(C_t), \tag{12}$$

The cell cascade structure is shown in Figure 6 [51]:

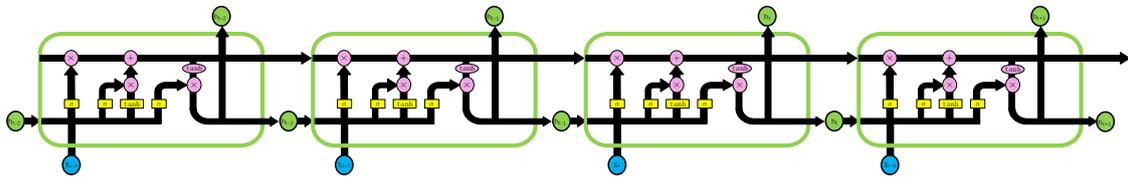


Figure 6. Cascade structure of LSTM.

where, h_t is P_{t+1} , represents the power output at the next moment, X_t is an eigenvector composed of $(T_t, H_t, G_t, D_t, P_t)$, and T_t, H_t, G_t, D_t present four multivariate meteorological factors at time point t , while P_t presents power output at time point t . The training process of LSTM is shown in Figure 7.

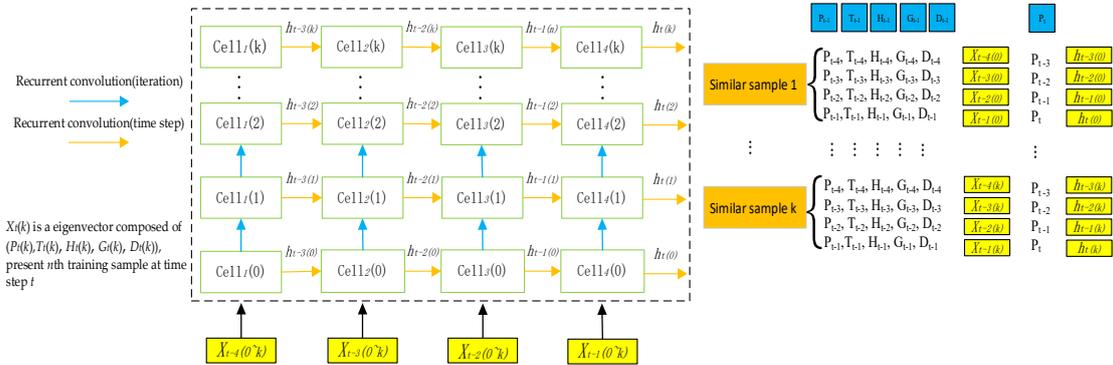


Figure 7. The training process of LSTM.

2.3. Error (Evaluation) Metrics

To prove the stability of the present RCC-LSTM based PV output prediction method, the coefficient of determination (R^2), the mean absolute percentage error (MAPE) and the root mean square error (RMSE) metrics are calculated, respectively. The definitions of these error metrics are shown below [1].

The R^2 is defined as:

$$R^2 = \frac{\left(N \sum_{i=1}^N P_{f,i} P_{a,i} - \sum_{i=1}^N P_{f,i} \sum_{i=1}^N P_{a,i} \right)^2}{\left(N \sum_{i=1}^N P_{f,i}^2 - \left(\sum_{i=1}^N P_{f,i} \right)^2 \right) \left(N \sum_{i=1}^N P_{a,i}^2 - \left(\sum_{i=1}^N P_{a,i} \right)^2 \right)}, \tag{13}$$

The MAPE is defined as:

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left(\left| \frac{P_{f,i} - P_{a,i}}{\bar{p}_{a,i}} \right| \right) \times 100, \tag{14}$$

$$\bar{p}_{a,i} = \frac{1}{N} \sum_{i=1}^N p_{a,i}, \tag{15}$$

The MAE is defined as:

$$MAE = \frac{1}{N} \sum_{i=1}^N |p_{f,i} - P_{a,i}|, \tag{16}$$

The RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_{f,i} - P_{a,i})^2}, \tag{17}$$

In this study, forecast results of the model run for a whole day are evaluated. $P_{f,i}$ and $P_{a,i}$ represent the predicted and real PV output at i time-point, respectively. $\bar{p}_{a,i}$ is the average value of actual PV output, and N is the prediction sample point numbers. N equals to 109 in this study.

3. Proposed Methodology

In order to describe the method more intuitively, the implementation procedure of the prediction method is shown in Figure 8. The detailed steps of the RCC-LSTM prediction model are shown below:

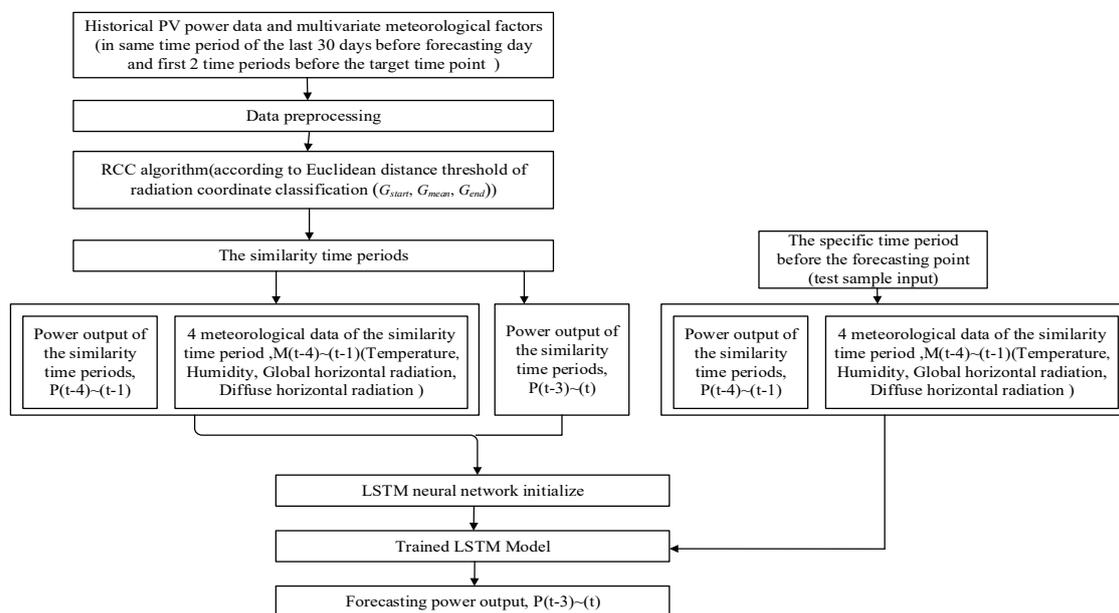


Figure 8. The implementation procedure of the proposed model.

Step 1: Collect historical PV power output and multivariate meteorological factors datasets. The meteorological factors include air temperature, relative humidity, global horizontal radiation and diffuse horizontal radiation.

Step 2: Preprocess the data, including abnormal data and normalization.

Step 3: According to the meteorological characteristic values of the time period before the forecasting point. RCC algorithm is used to determine the similarity time periods of the forecasting time period in the sample sets. By setting the threshold λ' in RCC, the data of the first 10 time periods

with smaller λ values are selected as the training dataset of the neural network if there are fewer than 10 samples.

Step 4: Determine the cell numbers of the LSTM, and initialize the threshold values and weights of LSTM RNN, respectively.

Step 5: The LSTM neural network is trained by using similarity time period samples, and then the prediction model is obtained.

Step 6: Input the power output and the values of the meteorological factors of the specific time period before forecasting time points into the prediction model to forecast the power output value.

4. Results and Discussion

To verify the validity of the proposed RCC-LSTM model, several typical networks, including RCC-BPNN, RCC-RBFNN [47], RCC-Elman, and LSTM-RNN [55] are chosen to make comparison, and the test are conducted in four seasons and two different PV systems. In addition, four different evaluation metrics (RMSE, MAPE, MAE, and R2) are applied to verify the prediction accuracy of the RCC-LSTM model. Figures 9 and 10 represent the forecasting result curves obtained by running different prediction models on two random days, respectively.

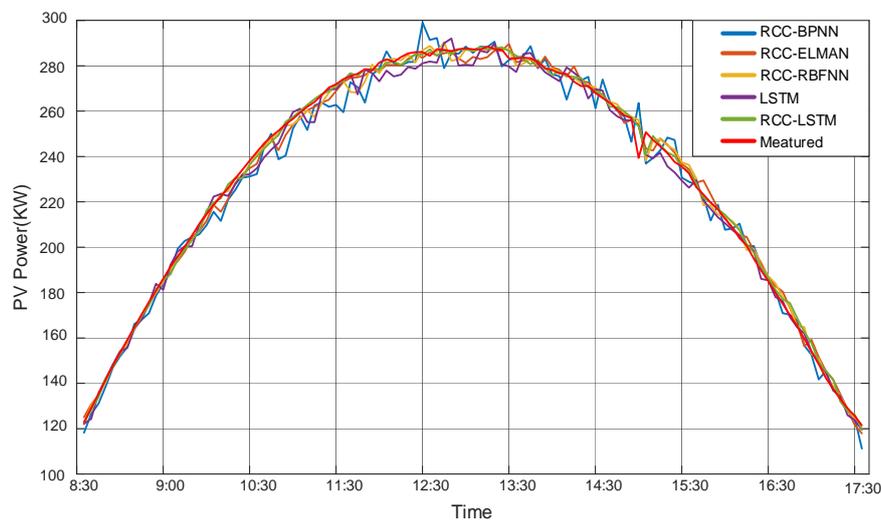


Figure 9. Forecasting result curve by different models for random days in SITE 4.

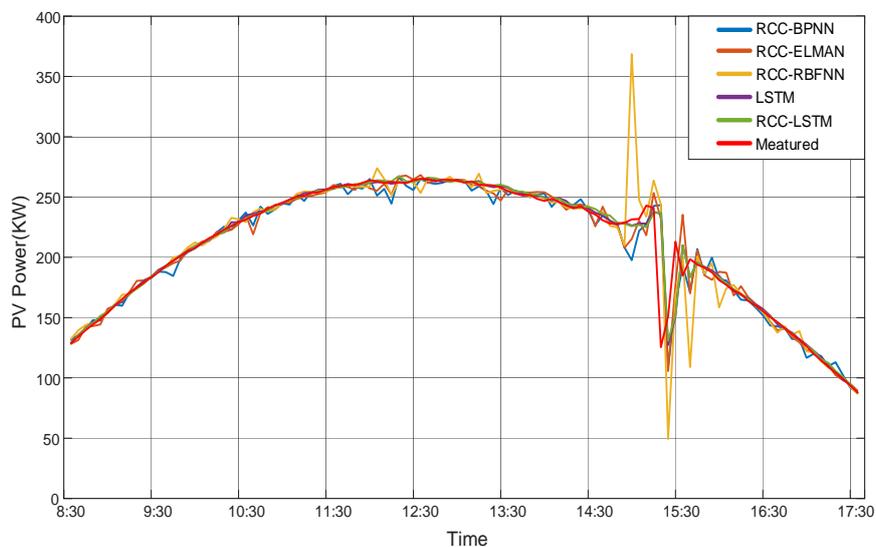


Figure 10. Forecasting result curve by different models for random days in SITE 4.

To further test the performance of the proposed RCC-LSTM model in different seasons, several days in different seasons are chosen to expand the validation sample set, three random days in each season. The detail information about different evaluation metrics is shown below.

The metrics results of the RMSE in SITE 3A and SITE 4 are shown in Figures 11 and 12, respectively. The RCC-LSTM model has the best prediction accuracy: the mean value of RMSE is 0.94 KW (in SITE 3A), which is the minimum for all models, compare with other models, the enhancement is 24.79%, 23.25%, 45.83, and 8.23%, respectively. In SITE 4, the RMSE has 12.58 KW mean value, and the enhancement is 38.38%, 28.18%, 55.33%, and 16.08%, respectively. However, due to the limit of data set, in RMSE, the standard deviation performance of LSTM is slightly lower than RCC-LSTM.

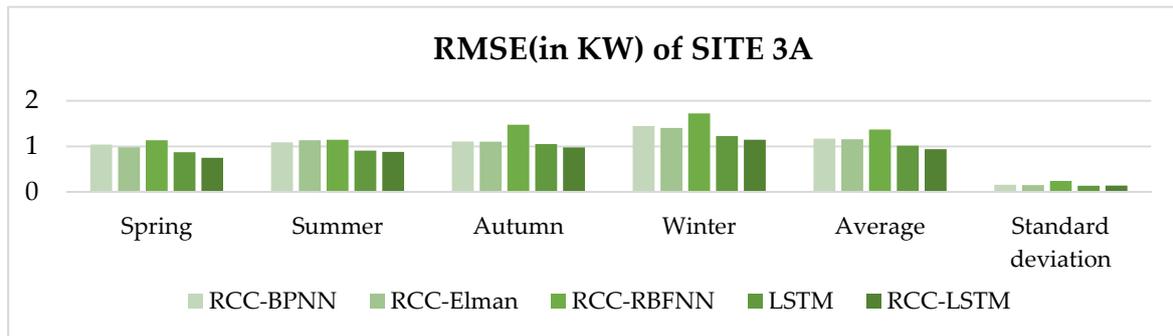


Figure 11. Comparative RMSE (in KW) results of SITE 3A.

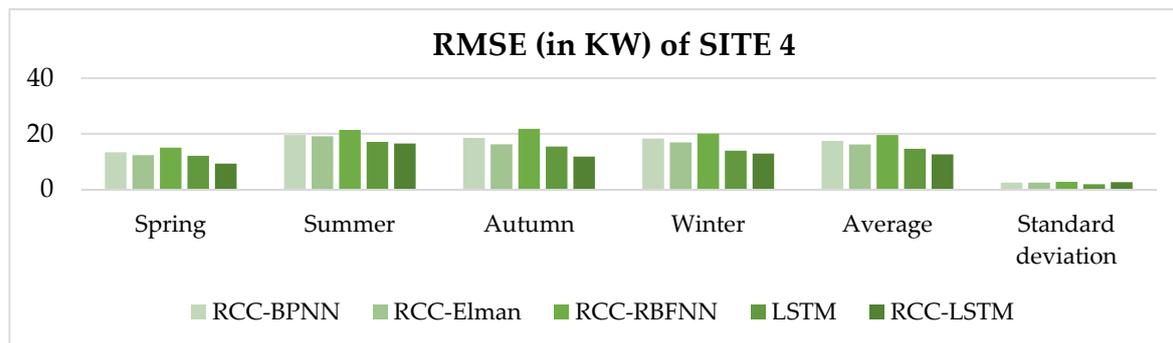


Figure 12. Comparative RMSE (in KW) results of SITE 4.

As seen in Figures 13 and 14, and compared with four other models, in SITE 3A, the average MAPE of the presented model reduced by 28.70%, 23.30%, 43.40%, and 9.67%, respectively. In SITE 4, the presented model’s average MAPE improvement relative to the compared four models is 44.04%, 31.31%, 53.11%, and 18.40%, respectively.

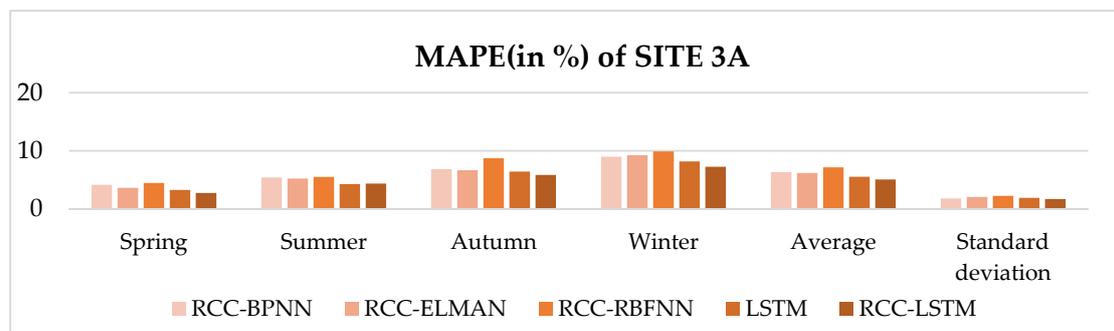


Figure 13. Comparative MAPE (in %) results of SITE 3A.

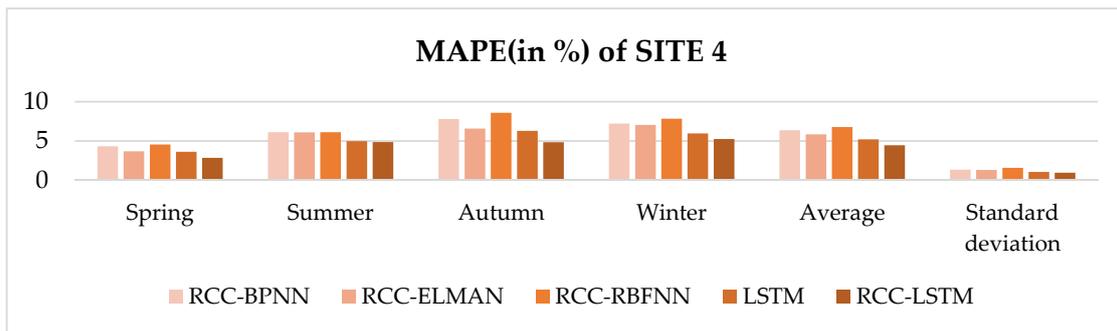


Figure 14. Comparative MAPE (in %) results of SITE 4.

Figures 15 and 16 show the MAE results of two PV systems, in SITE 3A, the presented model's average MAE improvement relative to the compared four models is 27.30%, 23.15%, 41.93%, and 9.34%, respectively. In SITE 4, the presented model's average MAE improvement relative to the compared four models is 45.29%, 32.10%, 53.83%, and 19.06%, respectively.

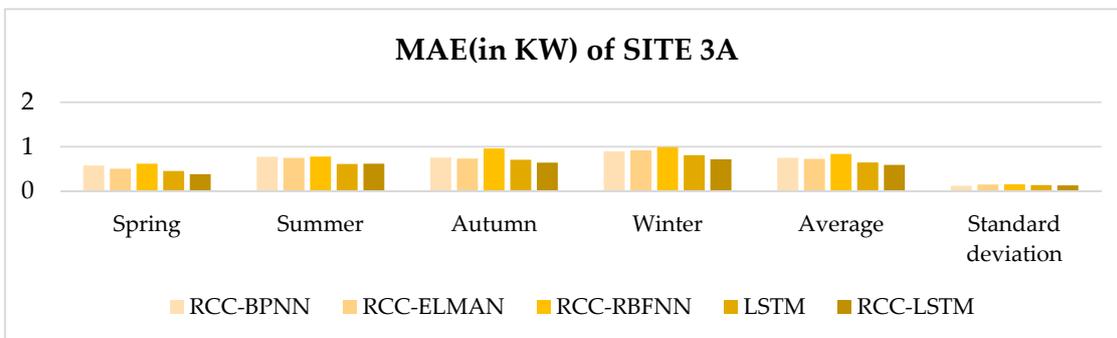


Figure 15. Comparative MAE (in KW) results of SITE 3A.

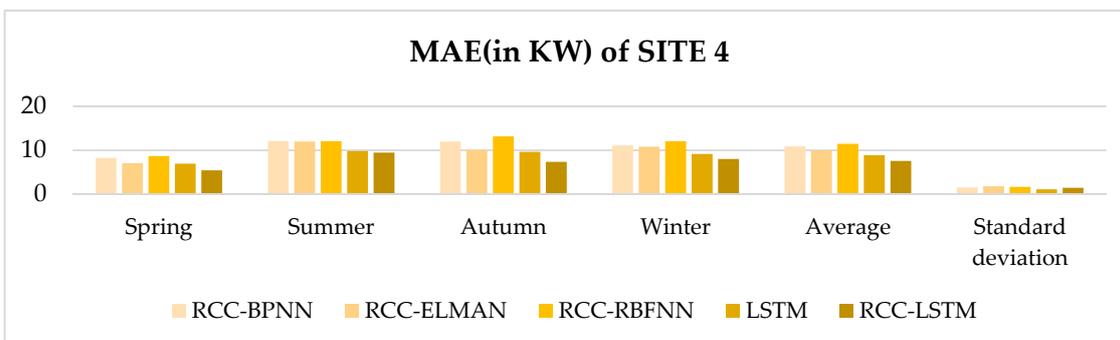


Figure 16. Comparative MAE (in KW) results of SITE 4.

In addition to the RMSE, MAPE, and MAE, R^2 is also a meaningful parameter to evaluate prediction models, the average value of R^2 (in %) and the standard deviation of different forecasting models are presented in Figure 17. Owing to the self-update time window, these models have good correlation performance, while the results of the proposed model are still better than others. In SITE 4, the mean value and the standard deviation of R^2 are 0.9747 and 0.0176, respectively. They are both the best among all models. Further, the situation in SITE 3A is the same as SITE 4.

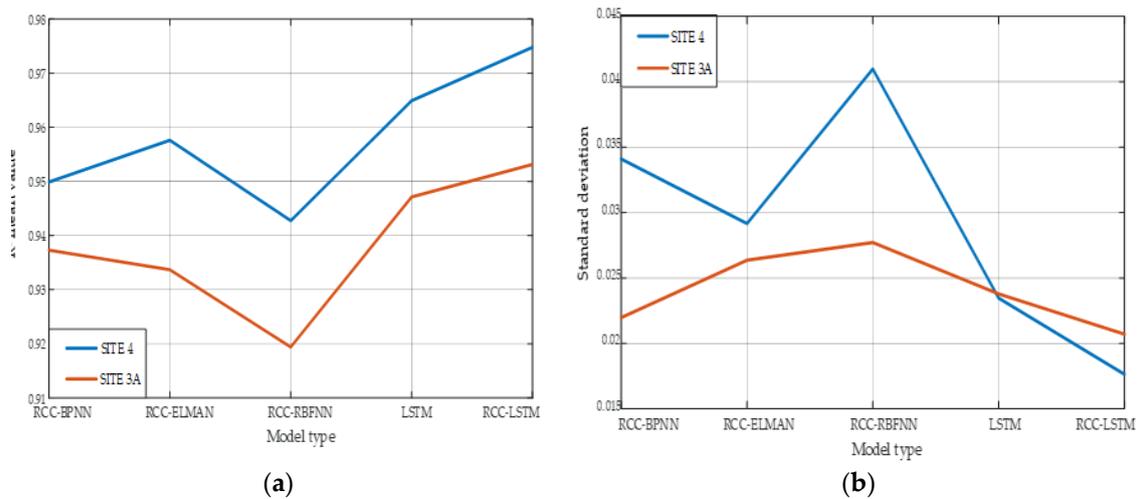


Figure 17. (a) The R^2 of different models in two PV systems; (b) The standard deviation of R^2 in two PV systems.

Therefore, the RCC-LSTM model has outstanding performance in the VST prediction of PV generation, especially in desert areas where the weather changes are more moderate. As the data set accumulates, the forecasting result of cloudy weather will also be improved.

The environment framework of this experiment is TensorFlow, which is implemented based on Python3.6 and a 64-bit operating system personal computer with Intel (R) Core (7M) i5-7300HQ CPU@2.50GHZ 2.50GHZ and 8.00GB of RAM. As shown in Figure 18, the average training time-cost at the different time points and the average runtime of every time point are shown in Figure 18a,b, respectively. Owing to the same time periods collection strategy of the training set, the dataset of training is small. Thus, every real-time predict step only requires a few seconds, which is acceptable in practical applications. Furthermore, due to the training dataset of RCC-LSTM that selected by RCC, compared with LSTM model, the RCC-LSTM not only improves the accuracy of the prediction but also reduces the training time cost of the prediction model. The average training time-cost of RCC-LSTM is 28.84% lower than that LSTM, and the runtime will be much lower by improving the hardware environments or optimizing code.

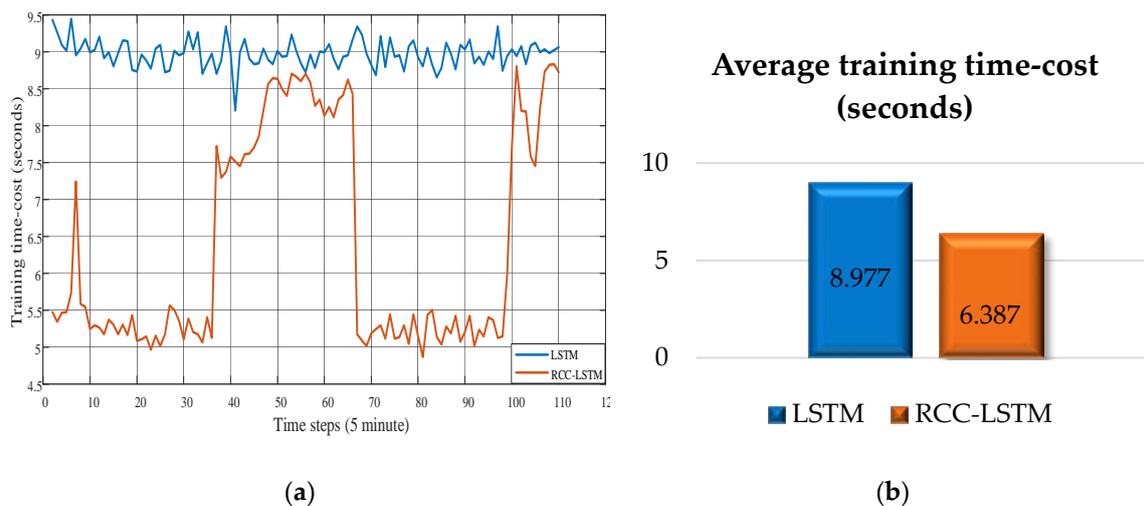


Figure 18. (a) The average training time-cost at different time point; (b) The average training time-cost of different models.

5. Conclusions and Future Work

A simple and effective RCC-LSTM model for VST PV power forecasting is proposed in this paper. The proposed method applies the RCC method as a tool for collecting similar time periods and employs the LSTM to extract features from the time series photovoltaic power data and to learn long-term information in sequence. Based on the dataset from two independent PV systems located in Central Australia. In five-minute forecasting horizons, a comprehensive comparative study is conducted to compare the proposed RCC-LSTM method with available four data-driven methods, including RCC-BPNN, RCC-Elman, RCC-RBFNN, and LSTM. Then, four error metrics are calculated and compared. The average daily RMSE, MAPE, and MAE of the RCC-LSTM model in site 3A are 0.940 kW, 5.053% and 0.587 KW, respectively, and those of site 4 are 12.58 kW, 4.449% and 7.590 KW, respectively. Compared with other methods, the average enhancement of RMSE, MAPE, MAE and R^2 is 30.01%, 31.49%, 31.50%, and 2.152%, respectively, which illustrates the superior performance of the proposed method. In addition, the average prediction time-cost of the RCC-LSTM is 28.84% lower than the basic LSTM. Therefore, it is proven that the proposed model can be used to predict VST PV power generation for PV power plants.

In future work, the proposed model can be improved in terms of its architecture and training data, and a more flexible selection of the threshold values should be implemented. Also, the cell number can be adjusted to be applied to different weather conditions.

Author Contributions: Methodology model built, B.C.; writing–original draft preparation, B.C.; conceptualization, P.L.; writing–review and editing, P.L.; supervision, Y.L.; project administration, S.C.; funding acquisition, S.C.; formal analysis Z.C.; data curation, L.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 61574038, 61601127, and 51508105), by the Foundation of Fujian Provincial Department of Science and Technology of China (Grant Nos. 2018J01774, 2018J01645, and 2019H0006), by the Fujian Provincial Economic and Information Technology Commission of China (Grant Nos. 82318075). The performance of the proposed forecasting model was tested by using the database from the Desert Knowledge Australia Solar Center (DKASC) website.

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

References

1. Antonanzas, J.; Osorio, N.; Escobar, R.; Urraca, R. Review of photovoltaic power forecasting. *Sol. Energy* **2016**, *136*, 78–111. [[CrossRef](#)]
2. Ye, B.; Zhang, K.; Jiang, J.; Miao, L. Towards a 90% renewable energy future: A case study of an island in the South China Sea. *Energy Convers. Manag.* **2017**, *142*, 28–41. [[CrossRef](#)]
3. Hua, H.; Xieb, N.; Fangb, D.; Zhang, X. The role of renewable energy consumption and commercial services trade in carbon dioxide reduction: Evidence from 25 developing countries. *Appl. Energy* **2018**, *211*, 1229–1244. [[CrossRef](#)]
4. Wang, B.; Wang, Q.; Wei, Y.-M.; Lie, Z.-P. Role of renewable energy in China's energy security and climate change mitigation: An index decomposition analysis. *Renew. Sustain. Energy Rev.* **2018**, *90*, 187–194. [[CrossRef](#)]
5. Yang, K.-J.; Kim, S.; Kim, S.-Y.; Ahn, K. Flexible $\text{Cu}_2\text{ZnSn}(\text{S},\text{Se})_4$ solar cells with over 10% efficiency and methods of enlarging the cell area. *Nat. Commun.* **2019**, *10*, 1–10. [[CrossRef](#)] [[PubMed](#)]
6. Andenæs, E.; Jelle, B.P.; Ramlo, K.; Kolås, T. The influence of snow and ice coverage on the energy generation from photovoltaic solar cells. *Sol. Energy* **2018**, 318–328. [[CrossRef](#)]
7. Krauter, S. Simple and effective methods to match photovoltaic power generation to the grid load profile for a PV based energy system. *Sol. Energy* **2018**, 768–776. [[CrossRef](#)]
8. REN21. Renewables 2019: Global status report. *REN21 Secretariat*, 2019.
9. García-Olivares, A.; Solé, J.; Osychenko, O. Transportation in a 100% renewable energy system. *Energy Convers. Manag.* **2018**, *158*, 266–285. [[CrossRef](#)]

10. Salpakari, J.; Mikkola, J.; Lund, P.D. Improved flexibility with large-scale variable renewable power in cities through optimal demand side management and power-to-heat conversion. *Energy Convers. Manag.* **2016**, *126*, 649–661. [[CrossRef](#)]
11. Eltawil, M.A.; Zhao, Z. Grid-connected photovoltaic power systems: Technical and potential problems—A review. *Renew. Sustain. Energy Rev.* **2010**, *14*, 112–129. [[CrossRef](#)]
12. Uniejewski, B.; Marcjasz, G.; Weron, R. Understanding intraday electricity markets: Variable selection and very short-term price forecasting using LASSO. *Int. J. Forecast.* **2019**, *35*, 1533–1547. [[CrossRef](#)]
13. Kraas, B.; Schroedter-Homscheidt, M.; Madlener, R. Economic merits of a state-of-the-art concentrating solar power forecasting system for participation in the Spanish electricity market. *Sol. Energy* **2013**, *93*, 244–255. [[CrossRef](#)]
14. Nottona, G.; Niveta, M.-L.; Voyanta, C.; Paolib, C. Intermittent and stochastic character of renewable energy sources: Consequences, cost of intermittence and benefit of forecasting. *Renew. Sustain. Energy Rev.* **2018**, *87*, 96–105. [[CrossRef](#)]
15. Nottrott, A.; Kleissl, J.; Washom, B. Energy dispatch schedule optimization and cost benefit analysis for grid-connected, photovoltaic-battery storage systems. *Renew. Energy* **2013**, *55*, 230–240. [[CrossRef](#)]
16. Ghosh, S.; Rahman, S.; Pipattanasomporn, M. Distribution Voltage Regulation Through Active Power Curtailment With PV Inverters and Solar Generation Forecasts. *Ieee Trans. Sustain. Energy* **2017**, *8*, 13–22. [[CrossRef](#)]
17. Sobri, S.; Koochi-Kamali, S.; Rahim, N.A. Solar photovoltaic generation forecasting methods: A review. *Energy Convers. Manag.* **2018**, *156*, 459–497. [[CrossRef](#)]
18. Ramsami, P.; Oree, V. A hybrid method for forecasting the energy output of photovoltaic systems. *Energy Convers. Manag.* **2015**, *95*, 406–413. [[CrossRef](#)]
19. Li, Y.; Su, Y.; Shu, L. An ARMAX model for forecasting the power output of a grid connected photovoltaic system. *Renew. Energy* **2015**, *66*, 78–89. [[CrossRef](#)]
20. Wang, G.; Su, Y.; Shu, L. One-day-ahead daily power forecasting of photovoltaic systems based on partial functional linear regression models. *Renew. Energy* **2016**, *96*, 469–478. [[CrossRef](#)]
21. Das, U.K.; Tey, K.S.; Seyedmahmoudian, M.; Idris, M.Y.I. SVR-Based Model to Forecast PV Power Generation under Different Weather Conditions. *Energies* **2017**, *10*, 876. [[CrossRef](#)]
22. Wang, Y.; Shen, Y.; Mao, S.; Cao, G. Adaptive Learning Hybrid Model for Solar Intensity Forecasting. *IEEE Trans. Ind. Inform.* **2018**, *14*, 1635–1645. [[CrossRef](#)]
23. Wang, F.; Zhen, Z.; Liu, C.; Mi, Z. Image phase shift invariance based cloud motion displacement vector calculation method for ultra-short-term solar PV power forecasting. *Energy Convers. Manag.* **2018**, *157*, 123–135. [[CrossRef](#)]
24. Alonso-Montesinos, J.; Battles, F.J.; Portillo, C. Solar irradiance forecasting at one-minute intervals for different sky conditions using sky camera images. *Energy Convers. Manag.* **2015**, *105*, 1166–1177. [[CrossRef](#)]
25. Caldas, M.; Alonso-Suarez, R. Very short-term solar irradiance forecast using all-sky imaging and real-time irradiance measurements. *Renew. Energy* **2019**, *143*, 1643–1658. [[CrossRef](#)]
26. Crisosto, C.; Hofmann, M.; Mubarak, R.; Seckmeyer, G. One-Hour Prediction of the Global Solar Irradiance from All-Sky Images Using Artificial Neural Networks. *Energies* **2018**, *11*, 2906. [[CrossRef](#)]
27. Eseye, A.T.; Zhang, J.; Zheng, D. Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and Meteorological information. *Renew. Energy* **2018**, *118*, 357–367. [[CrossRef](#)]
28. Koster, D.; Minette, F.; Braun, C.; O’Nagy, O. Short-term and regionalized photovoltaic power forecasting, enhanced by reference systems, on the example of Luxembourg. *Renew. Energy* **2019**, *132*, 455–470. [[CrossRef](#)]
29. Wang, J.; Li, P.; Ran, R.; Che, Y. A Short-Term Photovoltaic Power Prediction Model Based on the Gradient Boost Decision Tree. *Appl. Sci.* **2018**, *8*, 689. [[CrossRef](#)]
30. Wang, F.; Zhang, Z.; Liu, C.; Yu, Y. Generative adversarial networks and convolutional neural networks based weather classification model for day ahead short-term photovoltaic power forecasting. *Energy Convers. Manag.* **2019**, *181*, 443–462. [[CrossRef](#)]
31. Wang, J.; Ran, R.; Song, Z.; Sun, J. Short-Term Photovoltaic Power Generation Forecasting Based on Environmental Factors and GA-SVM. *J. Electr. Eng. Technol.* **2017**, *12*, 64–71. [[CrossRef](#)]
32. Sperati, S.; Alessandrini, S.; Monache, L.D. An application of the ECMWF Ensemble Prediction System for short-term solar power forecasting. *Sol. Energy* **2016**, *133*, 437–450. [[CrossRef](#)]

33. Pierro, M.; De Felice, M.; Maggioni, E.; Moser, D.; Perotto, A.; Spada, F.; Cornaro, C. Data-driven upscaling methods for regional photovoltaic power estimation and forecast using satellite and numerical weather prediction data. *Sol. Energy* **2017**, *158*, 1026–1038. [[CrossRef](#)]
34. Achouby, H.E.; Zaimi, M.; Ibral, A.; Assaid, E.M. New analytical approach for modelling effects of temperature and irradiance on physical parameters of photovoltaic solar module. *Energy Convers. Manag.* **2018**, *177*, 258–271. [[CrossRef](#)]
35. Azimi, R.; Ghayekhloo, M.; Ghofrani, M. A hybrid method based on a new clustering technique and multilayer perceptron neural networks for hourly solar radiation forecasting. *Energy Convers. Manag.* **2016**, *118*, 331–344. [[CrossRef](#)]
36. Ogliari, E.; Dolara, A.; Manzolini, G.; Leva, S. Physical and hybrid methods comparison for the day ahead PV output power forecast. *Renew. Energy* **2017**, *113*, 11–21. [[CrossRef](#)]
37. Das, U.K.; Tey, K.S.; Seyedmahmoudian, M.; Mekhilef, S. Forecasting of photovoltaic power generation and model optimization: A review. *Renew. Sustain. Energy Rev.* **2018**, *81*, 912–928. [[CrossRef](#)]
38. Lin, P.; Peng, Z.; Lai, Y.; Cheng, S. Short-term power prediction for photovoltaic power plants using a hybrid improved Kmeans-GRA-Elman model based on multivariate meteorological factors and historical power datasets. *Energy Convers. Manag.* **2018**, *77*, 704–717. [[CrossRef](#)]
39. Yadava, A.K.; Sharmab, V.; Malick, H.; Chandela, S.S. Daily array yield prediction of grid-interactive photovoltaic plant using relief attribute evaluator based Radial Basis Function Neural Network. *Renew. Sustain. Energy Rev.* **2018**, *81*, 2115–2127. [[CrossRef](#)]
40. Wang, H.; Lei, Z.; Zhang, X.; Zhou, B. A review of deep learning for renewable energy forecasting. *Energy Convers. Manag.* **2019**, *198*, 111799–111814. [[CrossRef](#)]
41. Mei, F.; Wu, Q.; Shi, T.; Lu, J. An Ultrashort-Term Net Load Forecasting Model Based on Phase Space Reconstruction and Deep Neural Network. *Appl. Sci.* **2019**, *9*, 1487. [[CrossRef](#)]
42. Wen, S.; Wang, Y.; Tang, Y.; Xu, Y. Real-Time Identification of Power Fluctuations based on LSTM Recurrent Neural Network: A Case Study on Singapore Power System. *IEEE Trans. Ind. Inform.* **2019**, *15*, 5266–5275. [[CrossRef](#)]
43. Barbieri, F.; Rajakaruna, S.; Ghosh, A. Very short-term photovoltaic power forecasting with cloud modeling: A review. *Renew. Sustain. Energy Rev.* **2017**, *75*, 242–263. [[CrossRef](#)]
44. Aryaputera, A.W.; Yang, D.; Zhao, L.; Walsh, W.M. Very short-term irradiance forecasting at unobserved locations using spatio-temporal kriging. *Sol. Energy* **2015**, *122*, 1266–1278. [[CrossRef](#)]
45. Zheng, L.; Liu, Z.; Shen, J.; Wu, C. Very short-term maximum Lyapunov exponent forecasting tool for distributed photovoltaic output. *Appl. Energy* **2018**, *229*, 1128–1139. [[CrossRef](#)]
46. Pedro, H.T.C.; Lim, E.; Coimbra, C.F.M. A database infrastructure to implement real-time solar and wind power generation intra-hour forecasts. *Renew. Energy* **2018**, *123*, 513–525. [[CrossRef](#)]
47. Rana, M.; Koprinska, I.; Agelidis, V.G. Univariate and multivariate methods for very short-term solar photovoltaic power forecasting. *Energy Convers. Manag.* **2016**, *121*, 380–390. [[CrossRef](#)]
48. Hwangbo, S.; Nam, K.; Heo, S.; Yoo, C. Hydrogen-based self-sustaining integrated renewable electricity network (HySIREN) using a supply-demand forecasting model and deep-learning algorithms. *Energy Convers. Manag.* **2019**, *185*, 353–367. [[CrossRef](#)]
49. The Desert Knowledge Australia Solar Centre. Available online: <http://dkasolarcentre.com.au/locations/yulara> (accessed on 30 April 2019).
50. Rahman, M.M.; Hasanuzzaman, M.; Rahim, N.A. Effects of various parameters on PV-module power and efficiency. *Energy Convers. Manag.* **2015**, *103*, 348–358. [[CrossRef](#)]
51. Qing, X.; Niu, Y. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy* **2018**, *148*, 461–468. [[CrossRef](#)]
52. Zhang, J.; Yan, J.; Infield, D.; Liu, Y.; Lien, F.S. Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model. *Appl. Energy* **2019**, *241*, 229–244. [[CrossRef](#)]
53. Wang, K.; Qi, X.; Liu, H.; Wu, C. A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network. *Appl. Energy* **2019**, *251*, 113315–113328. [[CrossRef](#)]

54. Wang, F.; Yu, Y.; Zhang, Z.; Li, J. Wavelet Decomposition and Convolutional LSTM Networks Based Improved Deep Learning Model for Solar Irradiance Forecasting. *Appl. Sci.* **2018**, *8*, 1286. [[CrossRef](#)]
55. Zhou, H.; Zhang, Y.; Yang, L.; Liu, Q. Short-Term Photovoltaic Power Forecasting Based on Long Short Term Memory Neural Network and Attention Mechanism. *IEEE Access* **2019**, *7*, 78063–78074. [[CrossRef](#)]



© 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/>).