Article

Daily Photovoltaic Power Generation Forecasting Model Based on Random Forest Algorithm for North China in Winter

Ming Meng and Chenge Song *

Department of Economics and Management, North China Electric Power University, Baoding 071003, China; 51851341@ncepu.edu.cn

* Correspondence: songchenge001@163.com

Received: 17 February 2020; Accepted: 12 March 2020; Published: 13 March 2020

Abstract: North China is one of the country's most important socio-economic centers, but its severe air pollution is a huge concern. In this region, precisely forecasting the daily photovoltaic power generation in winter is essential to improve equipment utilization rate and mitigate effects of power system on the environment. Considering the climatic characteristics of North China, the winter days are divided into three classifications. A forecasting model based on random forest algorithm is then designed for each classification. To evaluate its performance, the proposed model and three other methods are separately used to forecast the daily power generation at the Zhonghe PV station, which is located in the center of North China. Empirical results show that, because of its ability to reduce the risk of overfitting by balancing decision trees, the proposed model obtains mean absolute percentage errors as low as 2.83% and 3.89% for clear and cloudy days, respectively. For days in which weather conditions are unusual, forecasting errors are relatively large. On these days, enlarging training samples, performing subdivision, and imposing manual intervention can improve the forecasting precision. Generally, the proposed model is better than the other three methods for nearly all error evaluation indicators in each classification.

Keywords: photovoltaic power generation; forecasting; North China; random forest; weather classification

1. Introduction

North China, the geographical boundary of the area shown in Figure 2 [1], is one of the country's most important socio-economic centers. In recent years, this region has accumulated over one quarter of China's population and has produced a similar share of the country's GDP [2].

In this region, vast amounts of electric power are consumed to meet the requirements of socio-economic development. In 2017, electricity consumption had reached 1500.5 billion kWh, an increase by 70 billion kWh from 2016 [3]. To ensure the electric power supply, North China has built large-scale thermal power plants in the past several decades [4]. Given its fragile environment, the operation of thermal power plants further deteriorates the prevailing air pollution. Among China's 169 large- and medium-sized cities, the top 20 with the worst air conditions are all located in North China, and 24%–38% of different pollutants in this region are emitted from thermal power plants [5,6]. During winter, poor diffusion conditions of air pollutants, high power load, and the lack of other clean power sources significantly increase the effects of thermal power plants on air quality [7]. It is both necessary to promote the development of photovoltaic (PV) power generation and improve the utilization efficiency of power that are generated by solar PV in North China.

North China has abundant solar resources. Its annual solar horizontal radiation is higher than 1050 kWh/m² [8]. With the advancement of technology and the decrease in electricity generation cost,

this region has recently constructed many photovoltaic power plants to mitigate the air pollution. Following the development plan released by the National Energy Bureau of China, the installed capacity of PV power generation in this region is expected to reach 49 million kilowatts in 2020, which is four times that of 2015 [9].

The power generation of PV stations is considerably unstable, and specifically, determined by weather conditions to a great extent [10]. For the dispatch department of the grid company, the next day power generation of PV stations must be learned to plan the operation of the power system. Inaccurate prediction of daily PV power generation leads to low utilization rate of equipment. China has published a target of maintaining the utilization rate of PV equipment above 95% in the 2020s [11]. Achieving this target requires the support of an advanced forecasting method that is suitable for North China. Due to the worse air pollution in winter, when the PV power is more significant in this period than during other days, this research focuses on the daily generation forecasting for North China in this season.

At present, several methods of PV electricity generation forecasting have been proposed. According to time span, previous studies can be divided into: (1) very short-term—seconds, minutes, or hours ahead [12], (2) short-term—one or several days ahead [13], and (3) mid- and long-term—one or several weeks, months, or years ahead forecasting [14]. The present research belongs to the short-term forecasting. In addition, most studies follow one of two ideas, namely, indirect and direct predictions. The former establishes solar radiation forecast model and then converts the radiation data to PV power values, whereas the latter directly forecasts the PV electricity generation using historical and weather data. Solar radiation is closely related to the running state of the PV system, while the daily electricity generation is affected by various other factors; as such, the indirect prediction is usually more suitable than the direct method for forecasting the power output of PV panels [15,16]. The present research concentrates on the electricity generation forecasting and then does not adopt the indirect prediction.

When forecasting the daily PV power generation using the direct prediction, trend extrapolation methods, e.g., time series methods [17] and hybrid trend model [18], we can forecast according to the development trend of historical data. However, without considering weather conditions, these methods have limited accuracy, though the forecasting process is relatively simple. Artificial intelligence methods, e.g., artificial neural network [19] and genetic algorithm [20], can fit the relationship between weather conditions and PV power generation, but need large samples for model training. Otherwise, the generalization capacity degrades due to overfitting. In North China, the climate is changeable in winter. As a result, the sample number in each weather type is commonly limited to meet the requirements of the above intelligent methods. In recent years, tree-based methods began to be used for solar PV prediction. Random forest (RF) is a newly developed collection and forecasting method that can reduce the risk of overfitting by balancing decision trees [21,22]. Compared with other machine learning methods, RF presents a pronounced advantage in handling small samples. Relevant researches have proved that RF has effective capability in both solar radiation forecasting and PV power forecasting. Huang et al. [23] predicted the daily solar radiation of four sites in Australia, and show that the RF is the most accurate model. Another example of RF application in solar radiation prediction is detailed in the research of Liu et al. [24]. According to Ahmad et al. [25], two tree-based ensemble methods, including extra trees and RF, were performed marginally better than SVR in hourly PV output prediction. The authors of Ref. [26] indicate that RF and gradient boosting regression give the best results in forecasting solar power in GEF2014 competition. Zamo et al. [27] undertake a comparison of several forecasts based on statistical techniques to forecast hourly PV electricity production at some power plants in mainland France. The results have shown that RF has a superior performance. In a recent article by Kim et al. [28], a two-step solar power generation prediction model was proposed. On comparing the different machine learning methods, RF was also proved to be the most accurate model.

However, few of these previous studies made in-depth analysis on the selection of influence factors, especially in the complex climate background. On the basis of climatic characteristics in North

China, the present research classifies the winter days in this region to different weather types and then constructs the RF model to forecast the corresponding daily PV power generation.

The contributions of this paper are as follows:

1. **Topic:** In this paper, a daily PV power generation forecasting model for North China in winter is proposed. The proposed forecasting model is based on the random forest algorithm and can obtain satisfactory forecasting results using small samples. The results of this study provide a reference to the sustainable development of PV generation in this area.
2. **Influence factor selection:** The unique winter climatic characteristics of North China were considered. In consideration of the serious air pollution in winter, “PM2.5” is especially selected as one of the influence factors.
3. **Weather classification:** To ensure the operation of the model, the weather classification analysis is used to fix the varied weather. By combining weather types with similar characteristics, the problem of balance between category and sample number was solved.
4. **Methodology:** The application of RF for solar PV systems, as most of the previous researches are focused on trend extrapolation methods, artificial intelligence methods or support vector machines.

The remainder of this paper is organized as follows. Section 2 introduces the factors affecting PV power generation and weather classification. The principles of the method used for forecasting are described in Section 3. A case study of Zhonghe PV power station and result discussions are provided in Section 4. Conclusions and perspectives are summarized in the last section.

2. Influence Factors and Weather Classification

2.1. Influence Factor Selection

As mentioned in Section 1, weather conditions affect the electricity power generated by PV stations to a large extent. Involving these key factors into the forecasting model is essential to improve the prediction precision. When selecting key factors, the technical requirements of PV stations and the unique climatic characteristics of North China in winter require consideration.

In this research, six factors are involved in the forecasting model. (1) Total solar radiation. Solar radiation is the sum of direct and diffuse radiation and is the base of PV power generation. On this basis, this research takes total solar radiation as a selected factor. (2) Mean atmospheric pressure and (3) Wind speed. North China is located in the southeast end of Eurasia; winter in this area is usually from November to January, weather conditions are significantly affected by the monsoon [29]. The winter monsoon, which is caused by the gradient between the Siberian high and Aleutian low pressures, is the most important atmospheric circulation in the Northern Hemisphere [30]. This research uses the daily mean atmospheric pressure and wind speed to describe the influence of the winter monsoon. (4) Mean temperature. The winter monsoon brings strong cold air from Siberia to North China, and thus, the winter temperature in this region is generally lower and changeable than that in other areas of the same latitude [1]. Hence, the daily mean temperature is included into the forecasting model. (5) Relative humidity. Most areas of North China belong to temperate and semi-humid zones. During winter, different levels of fog commonly appear due to vapor in the air. In this research, the daily mean of relative humidity is considered in the forecasting model. (6) PM2.5 concentration. As mentioned in the preceding sections, air pollution further deteriorates in North China in the winter. Air pollutants can affect the generation of PV power stations on the ground. PM 2.5 concentration has considerable influence on the loss of power generation. This relationship has a coincidence level of over 90% [31]. In North China, PM2.5 is monitored once an hour in winter. This research excludes the nighttime monitored results and only considers the daytime mean values of PM2.5 concentration as a variable in the forecasting model.

2.2. Weather Classification

Dividing the observations into several classes according to weather characteristics and then building the prediction model for each classification is essential to improving the forecasting precision. However, a large classification number does not necessarily indicate an improvement, but rather means a small amount of information in training samples, thereby shortening the training time. However, small samples inevitably weaken the generalization ability of the forecasting model. As a result, observation classification should balance the classification and the sample numbers.

Ways to describe the weather state are varied. The China Meteorological Administration classifies the weather state into more than 30 types [32]. According to historical observation data of major cities in North China, seven main types of weather occur in winter: sunny, cloudy, overcast, rainy, snowy, fog and haze. Considering the sunshine conditions and the number of samples of these different weather types, we combined the cloudy, overcast, fog and haze into one class and treated rainy or snowy days as another. The main weather type, the one that lasts the longest in a day, is used as the classification criteria. In this manner, all observations are divided into three categories: clear, cloudy, and rainy or snowy days.

3. RF Forecasting Model

3.1. RF Algorithm

RF is a machine learning algorithm proposed by Breiman in 2001 [33]. As an ensemble method, RF improves learning performance with a voting system given a set number of decision trees. RF exhibits the characteristics of (1) random feature selection, (2) bootstrap sampling, (3) out-of-bag error estimation, and (4) full depth decision tree growing [34]. These features make random forest suitable for PV power generation prediction. PV power generation is easily affected by environmental factors, the data series usually contain a lot of noise. These noises may reduce the generalization ability of the model. After inputting data samples, RF model will first extract some of the samples by bootstrap sampling, and then randomly select the features of these samples. These two steps of random sampling make RF more tolerant to outliers and noise, and reduce the possibility of over fitting. Besides, due to the special winter climate conditions in North China, there are few samples of rainy and snowy days. However, RF can be applied to data sets of various sizes.

In this research, the RF regression algorithm is used when establishing the forecast model. The main steps to set up a RF algorithm are as follows:

1. Select samples by Bootstrap method [35] and regard them as a training set.
2. Grow an initial tree in the set.
3. Calculate the best node split of the initial tree according to its features.
4. Split the nodes until the samples belong to the same class.
5. Aggregate all trees into a forest, and then consider the mean value of the results given by each tree as the final prediction of the forest.

In addition, model building inevitably involves the determination of model parameters. The RF regression model has several main parameters, namely, number of estimators, criterion index, and max features. The functions of each parameter are as follows:

1. Number of estimators or the number of trees in the forest.
2. Criterion index measures the quality of the split. Alternatives include the mean absolute error (MAE) or mean square error (MSE) criteria.
3. Max features. A function is chosen to select the best number of features when searching for the best node. Three options are available: (1) original value, corresponding to the function auto, (2) square root of original value, corresponding to sqrt, and (3) the logarithm of original value, corresponding to log2.

This research selects the optimal values of the above parameters by comparing the performance of different parameter values. In addition, this research adopts the default values in the Sklearn module, which is written using the Python programming language [36].

3.2. Forecasting Process

Figure 1 shows the process of forecasting the daily PV power generation.

Figure 1. Forecasting the daily PV power generation.

To build a forecast model, the collected raw data require initial cleaning and filtering. In other words, fragmentary data—mainly referring to days for technical breakdown, and evidently outliers—mainly referring to statistic errors, must be amended. For preprocessed observations, the classification should be performed in accordance with their weather conditions. In this research, for each classification, a RF model was used for PV power generation. After training, the forecasting results are obtained by inputting the weather conditions of the target days into the model. When evaluating the performance of the RF model, the forecasting results for the three classifications are considered comprehensively.

3.3. Performance Evaluation Indicators

To understand the performance of the forecasting model, error analysis indicators are necessary. This research uses MAE, mean absolute percentage error (MAPE), root mean square error (RMSE), and explained variance (EV).

MAE represents the mean of the absolute errors and is used to reflect the real situation of the forecast error. MAE is calculated as:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|, \quad (1)$$

where $\hat{y}_i$ is the prediction and y_i is the real value. The range of values for MAE is 0 to $+\infty$. Evidently, as the result of MAE increases, the size of error likewise increases.

MAPE is a straightforward measure of the prediction accuracy of a forecasting method, and is thus usually considered as the fairest indicator [37]. On this basis, MAPE is usually considered as the

most important precision evaluation indicator. Dividing each error by the real value to provide an average result in percentage, MAPE is defined by:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \cdot 100\%, \quad (2)$$

RMSE is another common indicator used in the evaluation of the accuracy of the forecasting model. RMSE is the square root of MSE and is more sensitive to large errors than MAPE. The formula is:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}, \quad (3)$$

EV measures the part of the variation in a given data set that can be explained by a model. EV is calculated as:

$$\text{EV} = 1 - \frac{\text{Var}(y_i - \hat{y}_i)}{\text{Var}(y_i)}, \quad (4)$$

where Var is the variance of a sequence of values.

The largest value of EV is 1, which represents the best prediction.

4. Model Application and Evaluation

4.1. Model Application

In this research, the power generation from Zhonghe PV station is used to test the above forecasting method. The central position of the station is 114.32°E, 37.45°N, in Xingtai city, Hebei province, the center of North China, see Figure 2.

Figure 2. Geographical boundary of North China and the position of Zhonghe PV station in Xingtai city, Hebei province.

In addition to the daily power generation, meteorological data are also needed in the forecasting model. These data are selected from the Xingtai Meteorological Bureau in Hebei province.

Following the process shown in Figure 1, before the raw data being classified and introducing into the forecasting model, these data are preprocessed by cleaning and filtering to mend fragmentary information and eliminate statistical errors. In particular, data on PM2.5 and total solar radiation are preprocessed. In the sequence of PM2.5, approximately 10% of data was missing due to observation equipment failure and other reasons. To fill the missing values, linear interpolation method is adopted in this research. Regarding the total solar radiation, despite the possibility of automatic solar radiation

recording, erroneous data remain because of the measuring instruments [38]. To replace problematic data, alternate values are obtained through different measures: (1) analysis of data in similar days; (2) calculation of mean values of observations in adjacent areas; and (3) linear interpolation.

Finally, data from November to December 2016 and from January to November in 2017 are set as the training samples. The testing samples are composed of data from 1 November to 31 December 2018. Value distribution ranges of different indicators in the training and testing samples are shown in Table 1.

Table 1. Value distribution ranges of different indicators in the training and testing samples.

Data Set	Indicator	Range
Training samples	Temperature (°C)	−4.1–14.7
	Atmospheric pressure (kPa)	99.91–102.35
	Relative humidity (%)	18–99
	Wind speed (m/s)	0.7–4.6
	PM2.5 (µg/m ³)	16–409.15
	Total solar radiation (0.01 MJ/m ²) ¹	54–1376
Testing samples	Temperature (°C)	−8.4–12.4
	Atmospheric pressure (kPa)	100.42–102.98
	Relative humidity (%)	39–94
	Wind speed (m/s)	1.1–3.1
	PM2.5 (µg/m ³)	31.67–353.33
	Total solar radiation (0.01 MJ/m ²) ¹	191–1195

¹ Total solar radiation is a cumulative value and the rest are daily mean values.

As shown in Table 1, the training and testing samples have similar distribution range for most indicators. Thus, the RF model obtained by the training samples can be used for the forecasting analysis of the testing samples.

The next step is to classify the samples in both training set and test set. According to Section 2.2, both of the two data sets are divided into three categories: clear, cloudy, rainy or snowy days. The data in the same category of training set and test set are corresponding to each other. For instance, the data under the “clear days” classification in the test set will be input into the RF model trained by the data under the same classification in training set.

After classification, RF models will be established for each weather category. To ensure the good performance of the proposed model, ascertaining the optimal parameters is necessary after inputting the training samples into the corresponding model. As mentioned in Section 3.1, the proposed model can automatically search for the optimal parameters when providing a rough range of the number of trees. On the basis of the sample size, this research limited the tree number to the range of 100–1000. The final results of optimal parameters under different weather classification are shown in Table 2. After the above steps are completed, the influencing factors in the training set were input into the corresponding RF model, and then the prediction results can be obtained.

Table 2. Optimal parameters of RF regression models for three weather classifications.

Weather Classification	N Estimators	Criterion	Max Features
Clear days	900	MAE	auto
Cloudy days	800	MAE	auto
Rainy or snowy days	500	MSE	auto

4.2. Forecasting Results Analysis

As mentioned in Section 4.1, the RF models are used to forecast the daily power generation of the Zhonghe PV station from 1 November to 31 December 2018. This period comprises 61 days. However,

11 days are abandoned due to the serious loss of weather records and equipment maintenance, leaving only 50 days. To facilitate the following analysis, each day was assigned a code. These results are listed in Table 3.

Table 3. Code, date, classification, real value, and forecasting result of power generation of the Zhonghe PV station.

Code	1	2	3	4	5	6	7	8	9
Date	Nov.1	Nov.2	Nov.3	Nov.4	Nov.5	Nov.6	Nov.7	Nov.8	Nov.9
Classification ¹	Clear	Clear	Clear	R or S	R or S	R or S	Cloudy	Clear	Clear
Real value ²	162.03	160.79	189.85	61.23	40.19	57.88	149.91	190.32	217.15
Forecasting result ²	170.98	167.27	191.68	57.86	57.91	56.21	144.53	190.85	215.87
Code	10	11	12	13	14	15	16	17	18
Date	Nov.11	Nov.12	Nov.13	Nov.14	Nov.15	Nov.16	Nov.17	Nov.18	Nov.19
Classification ¹	Clear	Clear	Cloudy	Cloudy	R or S	Cloudy	Cloudy	Clear	Clear
Real value ²	194.27	183.15	127.50	100.38	28.82	152.87	157.07	175.82	220.41
Forecasting result ²	202.29	197.10	132.65	105.59	18.53	149.32	151.11	169.59	218.26
Code	19	20	21	22	23	24	25	26	27
Date	Nov.20	Nov.21	Nov.22	Nov.23	Nov.24	Nov.25	Nov.26	Nov.27	Nov.28
Classification ¹	Cloudy	Clear	Clear	Cloudy	Clear	Clear	Cloudy	Cloudy	Cloudy
Real value ²	157.33	231.79	233.79	156.58	168.88	191.59	126.30	115.87	134.24
Forecasting result ²	149.50	233.37	233.30	149.43	175.86	197.06	136.88	109.43	144.84
Code	28	29	30	31	32	33	34	35	36
Date	Nov.29	Nov.30	Dec.1	Dec.2	Dec.3	Dec.4	Dec.6	Dec.7	Dec.8
Classification ¹	Cloudy	Cloudy	R or S	R or S	R or S	Clear	Cloudy	Clear	Clear
Real value ²	155.80	106.35	61.33	28.82	60.99	194.41	115.10	175.14	185.19
Forecasting result ²	150.18	104.67	57.06	29.50	57.86	202.94	113.08	182.08	185.39
Code	37	38	39	40	41	42	43	44	45
Date	Dec.9	Dec.10	Dec.13	Dec.18	Dec.20	Dec.21	Dec.22	Dec.24	Dec.25
Classification ¹	Cloudy	Cloudy	Cloudy	Clear	R or S	Cloudy	R or S	Clear	Cloudy
Real value ²	123.02	117.17	158.11	177.99	56.21	117.28	57.68	168.05	103.45
Forecasting result ²	115.28	115.71	148.22	175.68	56.99	115.22	56.21	175.99	101.67
Code	46	47	48	49	50				
Date	Dec.27	Dec.28	Dec.29	Dec.30	Dec.31				
Classification ¹	R or S	Cloudy	Clear	Clear	Cloudy				
Real value ²	12.78	141.91	173.72	182.20	132.52				
Forecasting result ²	17.42	139.03	181.70	185.84	134.14				

¹ R and S refers to rainy or snowy days; ² The unit of power generation is MWh.

To understand the forecasting results, a line chart is used to demonstrate the forecasting performance, and a histogram is used to show the forecasting percentage errors. The line chart and histogram are shown in Figure 3a,b, respectively.

Figure 3a shows that affected by weather conditions, the real power generation of the PV station extremely changes. Generally, forecasting results can always be close to the real values. Hence, the proposed model can be applied to this problem.

Figure 3b shows three evidently largest forecasting errors, coded as 5, 14, and 46. Figure 3a shows that these days are all extremely rainy or snowy with the least power generation. As these weather conditions seldom appear, such training samples are limited and thereby produce the largest forecasting errors. In fact, rainy or snowy days are rarer than other weather types in North China, and observations in this classification vary greatly. As a result, the forecasting error of rainy or snowy days

is evidently larger than that of other weather classifications. Furthermore, enlarging training samples and performing subdivision can improve the forecasting precision of this classification.

Figure 3. Forecasting performance and errors.

Figure 3b shows that in addition to the rainy or snowy days, there are still some other days which have large forecasting errors. The forecasting errors of the 11th, 25th, 30th, and several other days are larger than their neighbors. In accordance with Table 3, these days are all classification conversion days, e.g., the first day of continuous cloudy days. Given that several weather indicators are close to its neighbor classification during these days, the trained RF model in this research cannot perfectly handle such a situation and therefore, cause large forecasting errors. To improve forecasting precision of these days, manual intervention is necessary. Specifically, the forecasting results of these days should be adjusted in accordance with its neighbor classifications.

According to statistics, errors of photovoltaic power generation prediction using random forest method are within the range of 8.5% to 12.3% [39]. Except for the three days that had been explained, the prediction errors of the random forest model established in this paper are almost all less than this range.

4.3. Performance Evaluation

To evaluate the performance of the proposed model, the support vector regression (SVR) with Linear kernel, elastic net (EN), and gradient boosting decision tree (GBDT) models are also used to forecast the power generation of the Zhonghe PV station from 1 November to 31 December 2018. The training samples and their classifications are the same as those of the proposed model. The forecasting

results of these three methods are listed in the Appendix A. Figure 4 shows the comparison of the different methods for each classification.

Figure 4. Forecasting result comparison of different methods.

Figure 4 shows that for clear and cloudy days, which have large samples, the forecasting results of RF, SVR, EN, and GBDT are relatively close. However, for rainy or snowy days with small samples, the forecasting performance of SVR and EN presents higher instability than other models. Several results, e.g., 41st, 43rd, and 46th days, have large errors. To evaluate the performance of the above models, the MAE, MAPE, RMSE, and EV are calculated in each classification and listed in Table 4.

Table 4. MAE, MAPE, RMSE, and EV of each forecasting model.

Weather	Algorithm	MAE (MWh)	MAPE (%)	RMSE(MWh)	EV
Clear	RF	5.07	2.83	6.23	0.95
	SVR	6.69	3.69	7.91	0.95
	EN	6.08	3.36	7.91	0.90
	GBDT	6.02	3.30	6.73	0.91
Cloudy	RF	5.23	3.89	6.02	0.91
	SVR	5.52	4.26	6.39	0.89
	EN	7.21	5.39	8.22	0.88
	GBDT	6.46	4.87	7.35	0.87
Rainy or snowy	RF	4.80	14.29	6.98	0.83
	SVR	9.70	24.84	13.11	0.72
	EN	11.03	33.76	16.10	0.30
	GBDT	5.29	16.19	7.17	0.82

By calculating error indicators, as shown in Table 4, except EV slightly smaller than SVR in the clear days, the RF-based method proposed in this research performs the best for all indicators in all classifications. Especially for the MAPE in the clear and cloudy days, the superiority of the proposed model is especially distinct.

5. Conclusion

Better forecasting the daily PV power generation in winter is essential to improving the equipment utilization rate and mitigate effects of power system on the environment in North China. Given that RF algorithms can lower the risk of overfitting by balancing decision trees in small samples, the winter days are divided into three classifications and an RF-based daily PV power generation forecasting model is built for each classification.

The proposed model and three other methods are separately used to forecast the daily power generation from 1 November to 31 December 2018 at the Zhonghe PV station, which is located in the center of North China. Empirical results show the following conclusions: (1) The values of MAPE, which is usually considered as the most important precision evaluation indicator, of the proposed model for clear and cloudy days are as low as 2.83% and 3.89%, respectively. (2) During the rainy or snowy days, which rarely appear in this region, the forecasting errors of the proposed model are relatively larger than those during other weather conditions. In those days, the forecasting precision can be improved by enlarging training samples and performing subdivision. (3) For classification conversion days, manual intervention based on neighbor classifications can further improve the forecasting precision. (4) The proposed model is better than the other three methods for nearly all error evaluation indicators in each classification.

Author Contributions: The authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 71471061, and the Fundamental Research Funds for the Central Universities, grant number 2017MS171.

Conflicts of Interest: There are no conflicting interests needing to be clarified.

Appendix A

Table A1. Training samples.

Date	Classification ¹	Temperature (°C)	Atmospheric Pressure (kPa)	Relative Humidity (%)	Wind Speed (m/s)	PM2.5 (µg/m ³)	Radiant Exposure (0.01 MJ/m ²)	Daily Power Generation (MWh)
Nov.2.2016	Clear	6.1	101.34	85	1.2	110	1251	233.72
Nov.3.2016	Clear	7.4	100.53	90	1.8	129	1120	197.05
Nov.5.2016	Clear	10.2	99.92	90	2.4	127	1164	208.24
Nov.7.2016	Cloudy	7.3	101.56	91	1.3	109	424	76.66
Nov.8.2016	Clear	6.1	102.11	77	1.4	81	1376	260.91
Nov.9.2016	Cloudy	6.1	101.52	81	3	138	577	108.90
Nov.10.2016	Clear	8.1	100.25	62	2.1	51	1185	231.03
Nov.11.2016	Clear	8.2	99.91	65	1.4	112	1118	210.87
Nov.12.2016	Clear	8	100.6	77	2	115	934	154.30
Nov.14.2016	Cloudy	7	100.75	85	1.3	67	583	108.29
Nov.15.2016	Cloudy	6	101.15	73	1.5	161	993	151.00
Nov.16.2016	Cloudy	7	100.66	69	1.3	141	861	148.59
Nov.17.2016	Cloudy	6.7	100.93	79	1.4	132	526	90.14
Nov.19.2016	Cloudy	9	100.44	84	1.1	157	429	83.41
Nov.20.2016	R or S	6.5	100.89	90	2	74	230	18.55
Nov.21.2016	R or S	0.9	101.47	95	4.6	43	54	5.98
Nov.23.2016	Clear	−1.7	102.22	71	1.3	112	910	191.26
Nov.24.2016	Clear	−0.7	101.58	76	2	166	974	201.23
Nov.25.2016	Clear	1.1	101.14	81	1.8	242	848	161.25

Table A1. Cont.

Date	Classification ¹	Temperature (°C)	Atmospheric Pressure (kPa)	Relative Humidity (%)	Wind Speed (m/s)	PM2.5 (µg/m ³)	Radiant Exposure (0.01 MJ/m ²)	Daily Power Generation (MWh)
Nov.26.2016	Clear	6.6	100.78	48	2.2	160	1048	210.21
Nov.27.2016	Clear	4.6	101.28	58	2.2	108	998	199.54
Nov.28.2016	Cloudy	2.7	101.85	62	1.4	143	796	147.18
Nov.29.2016	Cloudy	3.5	101.9	70	1.4	220	313	50.09
Nov.30.2016	Clear	4.8	101.18	72	1.4	222	1026	196.39
Dec.1.2016	Clear	4.3	101.83	59	1.8	162	906	185.11
Dec.2.2016	Clear	3.3	101.58	70	1.1	194	885	180.53
Dec.3.2016	Cloudy	4.4	100.71	81	1.4	276	621	116.64
Dec.4.2016	Cloudy	3.5	100.29	80	1.4	260	712	122.56
Dec.5.2016	Cloudy	4.4	101.31	72	2.9	99	563	89.90
Dec.6.2016	Clear	2.6	100.95	75	1.5	133	947	199.79
Dec.7.2016	Clear	5.7	100.97	57	2	74	894	183.91
Dec.8.2016	Cloudy	5.2	100.35	53	2.3	52	791	146.21
Dec.10.2016	Cloudy	−0.5	101.53	72	1.1	79	666	112.29
Dec.12.2016	R or S	3.8	100.96	86	1.9	154	372	58.66
Dec.15.2016	Clear	−0.2	101.87	75	2	86	860	179.06
Dec.16.2016	Clear	0	101.37	77	1.2	136	1091	199.70
Dec.17.2016	Cloudy	0	100.78	81	1.1	235	677	120.06
Dec.18.2016	Cloudy	0.1	101.04	83	0.8	225	635	105.17
Dec.19.2016	Cloudy	0	101.14	87	1.4	212	645	114.64
Dec.20.2016	R or S	−1.7	101.45	95	1.2	185	376	53.67
Dec.21.2016	Cloudy	2.8	101.22	96	1.1	217	212	30.85
Dec.22.2016	Clear	5.7	100.97	61	2.6	45	1145	229.62
Dec.23.2016	Cloudy	0.1	101.6	67	2	83	608	120.08
Dec.24.2016	Cloudy	−0.3	101.58	80	1.4	82	382	59.82
Dec.25.2016	R or S	0.9	101.47	90	1.5	67	217	17.42
Dec.27.2016	Clear	−1.7	102.07	89	2.3	182	902	179.27
Dec.28.2016	Cloudy	−2.1	101.87	89	1.8	259	684	114.12
Dec.29.2016	Clear	−2.4	102.14	77	2.2	89	899	179.25
Dec.30.2016	Cloudy	−2	101.54	85	1.1	122	646	118.11
Dec.31.2016	Cloudy	−1.8	101.3	94	1.3	108	771	131.67
Jan.1.2017	R or S	−2.1	101.1	99	2.4	92	175	25.66
Jan.2.2017	Clear	−0.6	101.25	91	1.8	128	891	176.30
Jan.3.2017	Cloudy	−1.7	100.94	99	1.1	109	370	61.80
Jan.4.2017	Cloudy	−1.5	101.14	99	1.6	144	329	34.55
Jan.5.2017	R or S	0.1	101.53	96	2.9	247	86	8.57
Jan.6.2017	R or S	2.3	101.33	92	1.2	250	285	30.22
Jan.7.2017	R or S	1.9	100.97	95	1.9	140	75	7.01
Jan.8.2017	Cloudy	0.1	100.94	98	2.3	98	455	95.71
Jan.9.2017	Cloudy	0.2	101.26	83	1.5	143	824	153.38
Jan.10.2017	Cloudy	−0.6	101.52	68	1.7	127	786	154.87
Jan.11.2017	Clear	0.7	101.2	73	2.1	241	854	185.42
Jan.12.2017	Clear	3.6	100.73	40	2.2	91	968	223.92
Jan.13.2017	Clear	0.4	101.07	51	1.7	114	867	175.80
Jan.14.2017	Cloudy	−2.3	101.75	65	1.4	138	720	122.26
Jan.15.2017	Cloudy	−2.5	101.81	64	1.5	179	531	90.27
Jan.16.2017	Cloudy	−2.3	101.66	74	2.3	211	572	69.70
Jan.17.2017	Cloudy	−2.5	101.64	77	1.3	275	575	114.43
Jan.18.2017	R or S	−2.2	101.89	82	1.1	356	126	18.41
Jan.19.2017	Clear	−1.2	101.62	53	3.7	215	1186	229.56
Jan.20.2017	Clear	−4	102	38	2.8	116	1013	213.61
Jan.21.2017	Clear	−2.7	101.57	45	1.4	190	975	194.99
Jan.22.2017	Clear	−3.9	102.08	41	1.4	109	900	180.35
Jan.23.2017	Cloudy	−3.7	101.99	55	1.7	264	776	140.14
Jan.24.2017	Cloudy	−2.2	102.21	45	2	251	679	118.29
Jan.25.2017	Cloudy	−2.5	101.78	62	1.4	324	682	116.38
Jan.26.2017	Cloudy	0.3	100.92	64	1.5	409	565	86.72
Jan.28.2017	Cloudy	0.7	100.66	47	1.1	349	413	64.13
Jan.29.2017	Cloudy	−0.1	101.36	41	3	196	720	96.10
Jan.30.2017	Clear	−4.1	102.35	37	1.6	64	1080	226.64
Jan.31.2017	Cloudy	0.2	101.53	31	1.6	139	798	146.80

Table A1. Cont.

Date	Classification ¹	Temperature (°C)	Atmospheric Pressure (kPa)	Relative Humidity (%)	Wind Speed (m/s)	PM2.5 (µg/m ³)	Radiant Exposure (0.01 MJ/m ²)	Daily Power Generation (MWh)
Nov.2.2017	Clear	13.2	100.31	68	1.9	119	1187	227.75
Nov.5.2017	Clear	9.4	100.92	75	1.7	61	1119	202.59
Nov.6.2017	Clear	10.2	100.44	80	0.7	151	1023	173.67
Nov.7.2017	Clear	14.7	100.59	57	2.8	65	1248	245.59
Nov.8.2017	Clear	10.9	101	56	2.5	94	918	157.07
Nov.10.2017	Clear	12.3	101.09	26	2.7	20	1209	254.84
Nov.11.2017	Clear	5.6	101.22	50	1.9	44	1185	239.73
Nov.12.2017	Cloudy	5.8	100.36	66	1.2	100	605	73.47
Nov.14.2017	Clear	4.6	101.07	38	1.4	38	1213	245.78
Nov.15.2017	Clear	3.1	101.08	50	1.6	48	1168	224.78
Nov.16.2017	Clear	3.3	100.75	52	1.2	95	844	176.55
Nov.17.2017	Cloudy	4	101.16	48	2.3	71	832	151.65
Nov.18.2017	Clear	−0.5	102	37	1.9	57	1179	243.87
Nov.19.2017	Cloudy	−0.3	101.43	52	1.3	128	436	66.00
Nov.20.2017	Clear	3.3	101.4	49	2.4	143	1113	213.70
Nov.21.2017	Clear	5.6	100.78	57	2.6	104	1085	199.50
Nov.22.2017	Clear	8.1	101.24	18	4.1	16	1194	232.54
Nov.23.2017	Clear	3.1	101.17	39	1.3	60	1119	217.04
Nov.24.2017	Clear	2.2	101.24	39	2.6	40	1078	216.00
Nov.25.2017	Clear	5	100.55	42	3.1	68	1063	204.74
Nov.26.2017	Clear	3	101.52	44	2.3	64	995	187.32
Nov.27.2017	Cloudy	1.2	100.7	60	2	122	655	119.84
Nov.28.2017	Clear	3.1	100.94	55	2.1	133	843	165.24
Nov.29.2017	Cloudy	−0.4	102.07	31	1.5	31	467	73.71

¹ R and S refers to rainy or snowy days.

Table A2. Test samples.

Date	Classification ¹	Temperature (°C)	Atmospheric Pressure (kPa)	Relative Humidity (%)	Wind Speed (m/s)	PM2.5 (µg/m ³)	Radiant Exposure (0.01 MJ/m ²)	Daily Power Generation (MWh)
Nov.1.2018	Clear	10.4	101.36	64	1.6	120	834	162.03
Nov.2.2018	Clear	11.1	101.15	69	1.9	131	822	160.79
Nov.3.2018	Clear	12.4	100.79	65	2.9	76	995	189.85
Nov.4.2018	R or S	11.9	100.92	75	3.1	164	471	61.23
Nov.5.2018	R or S	7.9	101.69	78	2.3	35	354	40.19
Nov.6.2018	R or S	6.5	101.83	72	1.1	43	425	57.88
Nov.7.2018	Cloudy	7.8	101.44	58	1.6	51	781	149.91
Nov.8.2018	Clear	9.4	100.44	66	2.9	63	993	190.32
Nov.9.2018	Clear	8.5	100.54	67	1.9	88	1135	217.15
Nov.11.2018	Clear	8.4	101.11	76	1.3	75	1045	194.27
Nov.12.2018	Clear	7.4	100.9	82	1.6	149	1037	183.15
Nov.13.2018	Cloudy	6.1	100.79	93	1.4	193	760	127.50
Nov.14.2018	Cloudy	8.7	100.89	93	1.6	167	581	100.38
Nov.15.2018	R or S	8.2	101.22	85	2.2	210	227	28.82
Nov.16.2018	Cloudy	6.1	101.75	52	1.9	32	789	152.87
Nov.17.2018	Cloudy	6.5	101.22	69	2.2	78	828	157.07
Nov.18.2018	Clear	8.5	100.93	45	1.6	41	921	175.82
Nov.19.2018	Clear	6	100.78	54	2	65	1154	220.41
Nov.20.2018	Cloudy	6.1	100.87	63	1.9	141	859	157.33
Nov.21.2018	Clear	6.6	101.4	51	2.6	88	1187	231.79
Nov.22.2018	Clear	3.8	101.31	54	1.7	51	1195	233.79
Nov.23.2018	Cloudy	4	100.71	57	2.5	150	828	156.58
Nov.24.2018	Clear	3.7	101.09	68	1.9	153	854	168.88
Nov.25.2018	Clear	5	100.78	67	1.7	179	970	191.59
Nov.26.2018	Cloudy	4.4	100.61	80	2	353	769	126.30

Table A2. Cont.

Date	Classification ¹	Temperature (°C)	Atmospheric Pressure (kPa)	Relative Humidity (%)	Wind Speed (m/s)	PM2.5 (µg/m ³)	Radiant Exposure (0.01 MJ/m ²)	Daily Power Generation (MWh)
Nov.27.2018	Cloudy	9.4	101.14	39	2.1	118	722	115.87
Nov.28.2018	Cloudy	5.1	101	45	2	127	779	134.24
Nov.29.2018	Cloudy	3.5	101.49	62	1.6	106	867	155.80
Nov.30.2018	Cloudy	2.9	101.21	80	1.1	133	581	106.35
Dec.1.2018	R or S	3.2	101.17	85	1.5	167	454	61.33
Dec.2.2018	R or S	5.2	100.42	94	1.2	194	236	28.82
Dec.3.2018	R or S	5.2	100.95	73	1.7	151	483	60.99
Dec.4.2018	Clear	2	101.64	42	2.1	55	1001	194.41
Dec.6.2018	Cloudy	−1.4	101.98	82	2.4	94	610	115.10
Dec.7.2018	Clear	−5.4	102.61	54	2	44	891	175.14
Dec.8.2018	Clear	−6.7	102.7	41	1.5	46	928	185.19
Dec.9.2018	Cloudy	−4.2	102.17	44	1.5	68	734	123.02
Dec.10.2018	Cloudy	−4.4	102.05	56	1.1	129	629	117.17
Dec.13.2018	Cloudy	−1.7	102.16	44	1.8	95	858	158.11
Dec.18.2018	Clear	2.7	100.53	46	1.4	98	879	177.99
Dec.20.2018	R or S	0.9	100.85	56	1.6	126	565	56.21
Dec.21.2018	Cloudy	4.3	100.98	45	1.6	113	629	117.28
Dec.22.2018	R or S	4.4	101.37	43	1.4	146	585	57.68
Dec.24.2018	Clear	−3.7	101.24	56	1.3	84	848	168.05
Dec.25.2018	Cloudy	−0.2	101.06	45	1.5	159	600	103.45
Dec.27.2018	R or S	−6.7	102.63	76	2.6	49	191	12.78
Dec.28.2018	Cloudy	−8.4	102.98	58	1.9	93	784	141.91
Dec.29.2018	Clear	−7.7	102.91	52	1.7	33	865	173.72
Dec.30.2018	Clear	−7.9	102.91	50	1.9	65	912	182.20
Dec.31.2018	Cloudy	−5.9	102.75	58	2	106	694	132.52

¹ R and S refers to rainy or snowy days.

Table A3. Forecasting results of three other methods.

Code	Date	Real Value	Forecasting Result ¹		
			SVR	EN	GBDT
1	Nov.01	162.03	171.01	160.33	159.45
2	Nov.02	160.79	168.14	155.81	154.38
3	Nov.03	189.85	198.04	187.03	185.39
4	Nov.04	61.23	67.74	60.73	58.08
5	Nov.05	40.19	47.84	41.92	57.86
6	Nov.06	57.88	58.83	53.37	56.02
7	Nov.07	149.91	142.84	137.60	142.75
8	Nov.08	190.32	197.99	188.93	184.76
9	Nov.09	217.15	220.35	214.86	218.87
10	Nov.11	194.27	204.90	196.73	200.91
11	Nov.12	183.15	201.07	193.50	178.80
12	Nov.13	127.50	135.87	134.44	131.76
13	Nov.14	100.38	102.53	101.12	106.45
14	Nov.15	28.82	31.30	22.52	18.52
15	Nov.16	152.87	144.70	138.88	147.24
16	Nov.17	157.07	151.33	146.71	151.88
17	Nov.18	175.82	189.52	183.71	165.61
18	Nov.19	220.41	224.09	222.86	225.36
19	Nov.20	157.33	155.34	150.72	148.71
20	Nov.21	231.79	230.87	230.98	236.91
21	Nov.22	233.79	232.78	234.45	235.77
22	Nov.23	156.58	148.98	143.68	149.35
23	Nov.24	168.88	173.14	168.10	176.83

Table A3. Cont.

Code	Date	Real Value	Forecasting Result ¹		
			SVR	EN	GBDT
24	Nov.25	191.59	191.44	187.63	199.46
25	Nov.26	126.30	132.78	131.55	136.04
26	Nov.27	115.87	129.29	122.90	101.47
27	Nov.28	134.24	139.97	134.02	144.52
28	Nov.29	155.80	157.74	153.28	150.61
29	Nov.30	106.35	102.99	100.75	103.71
30	Dec.01	61.33	65.26	62.93	56.51
31	Dec.02	28.82	32.39	24.32	24.50
32	Dec.03	60.99	69.44	67.71	58.08
33	Dec.04	194.41	202.75	204.88	200.24
34	Dec.06	115.10	109.56	107.01	119.16
35	Dec.07	175.14	183.88	188.88	185.28
36	Dec.08	185.19	191.53	200.60	188.96
37	Dec.09	123.02	132.78	127.39	122.34
38	Dec.10	117.17	111.43	107.87	109.55
39	Dec.13	158.11	155.71	150.20	148.44
40	Dec.18	177.99	181.35	180.21	177.09
41	Dec.20	56.21	81.67	88.52	57.07
42	Dec.21	117.28	111.78	106.13	118.41
43	Dec.22	57.68	85.06	89.13	56.02
44	Dec.24	168.05	175.61	177.95	179.12
45	Dec.25	103.45	104.88	99.89	105.10
46	Dec.27	12.78	23.40	33.43	18.12
47	Dec.28	141.91	141.93	138.20	148.29
48	Dec.29	173.72	180.26	187.17	185.04
49	Dec.30	182.20	187.36	196.01	189.67
50	Dec.31	132.52	124.47	120.80	120.92

¹ The unit of forecasting result is MWh.

References

1. Zhao, J. *The Physical Geography of China (in Chinese)*, 3rd ed.; Higher Education Press: Beijing, China, 1995; p. 203.
2. National Bureau of Statistics. *China Statistical Yearbook 2019 (in Chinese)*; China Statistics Press: Beijing, China, 2019; pp. 34–35, 69–71.
3. China Electricity Council. *Annual Development Report of China's Power Industry 2018 (in Chinese)*; China Electricity Council: Beijing, China, 2018.
4. Pei, Z.; Wang, C.; He, Q.; Wang, Y.; Fan, G. Analysis and suggestions on renewable energy integration problems in China (in Chinese). *Electr. Power* **2016**, *11*, 1–7. [\[CrossRef\]](#)
5. Ministry of Ecology and Environment. Bulletin of China's Ecological Environment (in Chinese). Available online: <http://www.mee.gov.cn/hjzl/zghjzkgb/lnzghjzkgb/201905/P020190619587632630618.pdf> (accessed on 10 January 2020).
6. Wang, L.; Li, P.; Yu, S.; Mehmood, K.; Li, Z. Predicted impact of thermal power generation emission control measures in the Beijing-Tianjin-Hebei region on air pollution over Beijing, China. *Sci. Rep.* **2018**, *8*, 934. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Chen, Q.; Sun, F.; Xu, Y. Does winter heating cause smog? Evidence from a city panel in North China (in Chinese). *Nankai Econ. Stud.* **2017**, *4*, 25–40. [\[CrossRef\]](#)
8. China Meteorological Administration. *Annual Bulletin of China's Wind and Solar Energy Resources in 2018 (in Chinese)*; China Meteorological Administration: Beijing, China, 2019.

9. National Energy Bureau of China. *The 13th Five Year Plan for Solar Energy Development (in Chinese)*; National Energy Bureau of China: Beijing, China. Available online: http://zfxgk.nea.gov.cn/auto87/201612/t20161216_2358.htm (accessed on 10 January 2020).
10. Jin, D.; Olama, M.M.; Kuruganti, T.; Melin, A.M.; Djouadi, S.M.; Zhang, Y.; Xue, Y. Novel stochastic methods to predict short-term solar radiation and photovoltaic power. *Renew. Energy* **2020**, *145*, 333–346. [CrossRef]
11. National Development and Reform Commission of the People's Republic of China. National Energy Bureau of China. Clean Energy Consumption Plan 2018–2020 (in Chinese). Available online: <http://zjb.nea.gov.cn/article/zygg/d1/201812/3022.htm> (accessed on 10 January 2020).
12. Kushwaha, V.; Pindoriya, N.M. A SARIMA-RVFL hybrid model assisted by wavelet decomposition for very short-term solar PV power generation forecast. *Renew. Energy* **2019**, *140*, 124–139. [CrossRef]
13. Yang, M.; Meng, L. Short-term photovoltaic power dynamic weighted combination forecasting based on least squares method. *Ieej Trans. Electr. Electron. Eng.* **2019**, *14*, 1739–1746. [CrossRef]
14. Alanazi, M.; Alanazi, A.; Khodaei, A. Long-term solar generation forecasting. In Proceedings of the IEEE/PES Transmission and Distribution Conference and Exposition (T&D), Dallas, TX, USA, 3–5 May 2016. [CrossRef]
15. Ostrometzky, J.; Bernstein, A.; Zussman, G. Irradiance field reconstruction from partial observability of solar radiation. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1698–1702. [CrossRef]
16. Larson, D.P.; Nonnenmacher, L.; Coimbra, C.F.M. Day-ahead forecasting of solar power output from photovoltaic plants in the American Southwest. *Renew. Energy* **2016**, *91*, 11–20. [CrossRef]
17. Yang, Z.; Zhu, F.; Zhang, C.; Ge, L.; Yuan, X. Photovoltaic power generation short-term power forecasting based on adaptive fuzzy time sequence method (in Chinese). *J. Nanjing Inst. Technol. (Nat. Sci. Ed.)* **2014**, *1*, 6–13. [CrossRef]
18. Meng, M.; Niu, D.; Shang, W. A small-sample hybrid model for forecasting energy-related CO₂ emissions. *Energy* **2014**, *64*, 673–677. [CrossRef]
19. Chaouachi, A.; Kamel, R.M.; Nagasaka, K. Neural network ensemble-based solar power generation short-term forecasting. *J. Adv. Comput. Intell. Inform.* **2010**, *14*, 69–75. [CrossRef]
20. Yang, Z.; Cao, Y.; Xiu, J. Power generation forecasting model for photovoltaic array based on generic algorithm and BP neural network. In Proceedings of the 2014 IEEE 3rd International Conference on Cloud Computing and Intelligence Systems, Shenzhen, China, 27–29 November 2014. [CrossRef]
21. Behrens, C.; Pierdzioch, C.; Risse, M. Testing the optimality of inflation forecasts under flexible loss with random forests. *Econ. Model.* **2018**, *72*, 270–277. [CrossRef]
22. Lei, C.; Deng, J.; Cao, K.; Ma, L.; Xiao, Y.; Ren, L. A random forest approach for predicting coal spontaneous combustion. *Fuel* **2018**, *223*, 63–73. [CrossRef]
23. Huang, J.; Troccoli, A.; Coppin, P. An analytical comparison of four approaches to modelling the daily variability of solar irradiance using meteorological records. *Renew. Energy* **2014**, *72*, 195–202. [CrossRef]
24. Liu, J.; Cao, M.; Gao, Z.; Xu, K. A solar radiation prediction model based on random forest (in Chinese). *Control. Eng. China* **2017**, *24*, 2472–2477. [CrossRef]
25. Ahmad, M.W.; Mourshed, M.; Rezugui, Y. Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression. *Energy* **2018**, *164*, 465–474. [CrossRef]
26. Mohammed, A.A.; Yaqub, W.; Aung, Z. *Probabilistic Forecasting of Solar Power: An Ensemble Learning Approach. Smart Innovation, Systems and Technologies*; Neves-Silva, R., Jain, L., Howlett, R., Eds.; Springer: Cham, Switzerland, 2015; Volume 39, pp. 449–458. ISBN 978-3-319-19856-9.
27. Zamo, M.; Mestre, O.; Arbogast, P.; Pannekoucke, O. A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production, part I: Deterministic forecast of hourly production. *Sol. Energy* **2014**, *105*, 792–803. [CrossRef]
28. Kim, S.G.; Jung, J.Y.; Sim, M.K. A two-step approach to solar power generation prediction based on weather data using machine learning. *Sustainability* **2019**, *11*, 1501. [CrossRef]
29. Chernokulsky, A.; Mokhov, I.I.; Nikitina, N. Winter cloudiness variability over Northern Eurasia related to the Siberian High during 1966–2010. *Environ. Res. Lett.* **2013**, *8*, 045012. [CrossRef]
30. Chang, C.; Lu, M. Intraseasonal predictability of Siberian High and East Asian Winter Monsoon and Its Interdecadal Variability (in Chinese). *J. Clim.* **2012**, *25*, 1773–1778. [CrossRef]
31. Ren, J.; Chen, T.; Xu, Z.; Wu, C.; Zhao, C. Mechanism analysis and experimental study on the influence of haze on photovoltaic power generation (in Chinese). *Res. Explor. Lab.* **2019**, *38*, 42–46. [CrossRef]

32. China Meteorological Administration. *Public Meteorological Service—Weather Graphic Symbols (in Chinese)*; China Meteorological Administration: Beijing, China, 2017.
33. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
34. Jiang, R.; Tang, W.; Wu, X.; Fu, W. A random forest approach to the detection of epistatic interactions in case-control studies. *BMC Bioinform.* **2009**, *10*, S65. [[CrossRef](#)]
35. Feng, W.; Dauphin, G.; Huang, W.; Quan, Y.; Liao, W. New margin-based subsampling iterative technique in modified random forests for classification. *Knowl. Based Syst.* **2019**, *182*, 104845. [[CrossRef](#)]
36. Scikit-learn, Random Forest Regressor. Available online: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html> (accessed on 2 December 2019).
37. Paulescu, M.; Paulescu, E. Short-term forecasting of solar irradiance. *Renew. Energy* **2019**, *143*, 985–994. [[CrossRef](#)]
38. Zarzo, M.; Martí, P. Modeling the variability of solar radiation data among weather stations by means of principal components analysis. *Appl. Energy* **2011**, *88*, 2775–2784. [[CrossRef](#)]
39. Cui, Y.; Sun, Y.; Chang, Z. A review of short-term solar photovoltaic power generation prediction methods (in Chinese). *Resour. Sci.* **2013**, *35*, 1474–1481.

© 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/>).