

## Article

# Quantitative Analysis of the Impact of Meteorological Environment on Photovoltaic System Feasibility

Dengchang Ma <sup>1,2</sup> , Guobing Pan <sup>1,\*</sup> , Fang Xu <sup>1</sup> and Hongfei Sun <sup>1</sup>

<sup>1</sup> The Institute of Distributed Energy and Microgrid, Zhejiang University of Technology, Hangzhou 310013, China; madc@zjweu.edu.cn (D.M.); fangx@zjut.edu.cn (F.X.); shongfei@126.com (H.S.)

<sup>2</sup> The College of Electrical Engineering, Zhejiang University of Water Resources and Electric Power, Hangzhou 310018, China

\* Correspondence: gbpan@zjut.edu.cn; Tel.: +86-1598-886-9763

**Abstract:** The meteorological environment is a determining factor in photovoltaic (PV) system feasibility (PVSF). To evaluate this impact more accurately, a quantitative analysis model based on multimeteorological factors and the Random Forest Regression model is proposed in this work. Firstly, an evaluation system is established to assess the impact. Then, to predict the indicators of the evaluation system, a parameter, i.e., performance ratio in sampling period, is defined. Secondly, a set of essential influences on the performance ratio in the sampling period is established through analyzing and reducing the discovered influences on the PV system performance. Finally, data from the Desert Knowledge Australia Solar Centre (DKASC) website are used to conduct the experiment. During the experiment, the sample set is cleaned using the model based on the cosine of the zenith angle. The functional relationship between the performance ratio in the sampling period and its essential influences is established through training a Random Forest Regression model with the data of the modeling system. The data of the test system are used to verify the forecast performance of the proposed model. Compared with the reference model, which is based on the traditional physical experiment, the results of the proposed model accord better with the measured values.

**Keywords:** PV system; feasibility study; meteorological environment; quantitative analysis; Random Forest Regression



**Citation:** Ma, D.; Pan, G.; Xu, F.; Sun, H. Quantitative Analysis of the Impact of Meteorological Environment on Photovoltaic System Feasibility. *Energies* **2021**, *14*, 2893. <https://doi.org/10.3390/en14102893>

Academic Editor: Marco Pasetti

Received: 9 April 2021  
Accepted: 7 May 2021  
Published: 17 May 2021

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



**Copyright:** © 2021 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 (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

It is expected that by 2088, the world will run out of fossil fuels [1]. To overcome the energy crisis, solar energy, along with other renewable energy sources, has gathered significant attention due to its abundance, universality and cleanliness [2]. By the end of 2020, the global cumulative installed capacity of PV power plants amounted 714 GW, with an additional installed capacity of 127 GW (+22%) [3]. Such a scale of growth puts forward higher requirements for the study of the PVSF. Improving the accuracy of PVSF assessment can effectively promote scientific decisions concerning project approval by government departments, and reduce the investment risk of power operators. In recent years, the rapid development of big data storage and data mining technology has promoted technological innovation in all walks of life. To establish a big data system by integrating the data extracted from the PVSF study report and other useful data, and to rapidly and quantitatively evaluate the PVSF using data-driven methods have attracted the interest of the PV project evaluation department. The impact of the meteorological environment plays an important role for the PVSF, because it has a great influence on the generation potential of PV systems.

In the traditional method, the impact is analyzed by using solar resources and the estimated output energy. Solar energy resources are easy to calculate based on the dataset of horizontal irradiance measured at the site of the PV system to be evaluated. PV system output energy is generally predicted by using the model based on the system physical

structure (SPS). The first step of this model is to determine the ideal energy conversion efficiency of the system according to its SPS. The second step is to create a data set of the incident solar irradiance on the module or array plane, which is denoted as plane of array irradiance (POAI). The third step is to establish a data set of the ideal output power (IOP) of the system based on the data set of the POAI and the ideal energy conversion efficiency of the system. The last step is to correct the IOPs and predict the system output energy.

In the second step of the model based on the SPS, there are a number of models available to estimate the solar irradiance on the inclined surface (the POAI in this article) from the corresponding horizontal data [4], including isotropic models (e.g., Liu-Jordan [5], Korokanis [6], Badescu [7]) and anisotropic models (e.g., Bugler [8], Ma-Iqbal [9], Hay [10], Skartveit-Olseth [11], Muneer [12], Willmot [13], Gueymard [14], Reindel et al. [15], Temps [16], Klucher [17], Perez et al. [18]). Researchers have compared and improved these models (e.g., [19–21]).

The power output of PV system is not only dependent on the POAI, but is also affected by the system performance. The fluctuation of system performance is mainly caused by the change of PV module efficiency. The operating temperature of a PV module has a dramatic effect on its electrical efficiency [22], which is well documented [23]. Radziemska [24] experimentally investigated the effect of the operating temperature and found a decrease of about 0.65% electrical efficiency of the PV module for every 1 K increase in cell temperature. Teo et al. [25] observed that the operating temperature of PV modules attained a value as high as 68 °C, and the electrical efficiency dropped significantly to 8.6%. In the last step of the model based on the SPS, traditionally, the ideal power was corrected by using the module power temperature coefficient provided by the manufacturer and the operating temperature, calculated from the ambient temperature (AT). In this correction model, only the influence of AT on the operating temperature of PV modules is taken into account.

Many scholars have studied the PV module temperature and found that, in addition to the AT, the operating temperature of PV modules is also affected by other meteorological factors. By comparison, Ayompe, Duffy et al. [26] found that high average wind speed (WS) and low AT can improve PV module efficiency. According to some tests, Griffith et al. [27] found that the cell temperature is extremely sensitive to WS, and moderately sensitive to wind direction. Teo et al. [25] experimentally found that the PV module increases by 1.4 °C for every 100W/m<sup>2</sup> increment of solar irradiance with active cooling, and increases more, i.e., by 1.8 °C, without active cooling. There are many correlations expressing the PV cell temperature as a function of weather variables such as AT, local WS, and solar irradiance, with material and system-dependent properties as parameters, e.g., glazing-cover transmittance, plate absorptance, etc. [23,28,29].

The PV module performance is affected by the operating temperature in the short term; however, in the long term, the degradation in the performance of a PV module due to its aging is also an influencing factor. Jordan and Kurtz [30] reviewed the degradation rates of flat-plate terrestrial modules and systems reported in the published literature from field testing from 1974 to 2013. Nearly 2000 degradation rates measured on individual modules or entire systems were assembled from the literature, showing a median value of 0.5% each year. Ndiaye et al. [31] reviewed the most frequent PV module degradation types of silicon PV modules according to the literature.

Some other factors also affect the PV module performance, such as front surface soiling [32–34], mismatch losses [35], dust accumulation [36–38], incident solar spectrum [39–41] et al.

The impact of solar radiation and AT on the PVSF can be analyzed through the model based on the SPS, but it is difficult to analyze the comprehensive impact of more meteorological factors through this model. Ignoring other meteorological factors which have some impact on the PV system performance not only has a negative impact on the accuracy of prediction results, but also leads to a waste of data resources that can be extracted from the feasibility study report.

To analyze the impact of meteorological environment on the PVSF, the main step is to predict the PV output energy. In addition to the PV output energy prediction method of the model based on the SPS, a variety of more advanced PV output energy prediction models have been proposed, as reviewed and compared in [42–47]. These models can be classified as statistical techniques, linear models or time series models, artificial intelligence techniques and hybrid models [42]. Most of these models are designed to eliminate the negative impact created by the inherent variability of PV output power on the electrical performance of the grid. The purpose of these models is to predict the output energy of the built PV system in the short-term. The PV system performance can be affected by the degradation in the performance of PV modules in the long term; this factor is generally ignored in short-term forecasting. However, in the PVSF study, the output energy of the unbuilt system in the ultralong term in the future needs to be predicted, so the impact of the degradation in the performance of PV modules cannot be ignored. Each proposed prediction model has its specific applicable conditions, and few PV output energy prediction models have been designed based on the data available in the PVSF study report, so when using these models, it is difficult to make full use of the data resources in feasibility reports.

Therefore, in this study, we aimed to develop a model for analyzing the impact of the meteorological environment on the PVSF in data-driven way. Firstly, we sought to establish a more scientific indicator system to quantify the impact by optimizing the traditional indicators and introducing a new indicator. Then, to improve the indicator prediction accuracy by making the most of the data resources and increasing the no extra data measurement cost, we analyzed and reduced the influences on the system performance based on the data available in the PVSF study report and picked out the essential influences which corresponded to the inputs of the indicator prediction model. Finally, we used the Random Forest Regression (RFR) [48] model to develop an indicator prediction model and took not only the two parameters (POAI and AT) which are the inputs of the model based on the SPS, but also some other meteorological parameters which also correlate with the PV system performance as the inputs of the model to quantify the comprehensive impact of multiple meteorological factors on the PVSF.

## 2. Development of the Quantitative Analysis Model

### 2.1. Development of the Indicator System

The indicators for analyzing the impact of meteorological environment on the PVSF in this article include the final PV system yield ( $Y_f$ ), the reference yield ( $Y_r$ ), and the performance ratio (PR), which are defined in the IEC standard 61,724 [49] as follows:

$$Y_f = \frac{E}{P_0}, \quad (1)$$

$$Y_r = \frac{H}{G}, \quad (2)$$

$$PR = \frac{Y_f}{Y_r}, \quad (3)$$

where  $E$  is the output energy,  $P_0$  is the system capacity,  $H$  is the total in-plane irradiation per square meter, and  $G$  is the reference irradiance at which  $P_0$  is determined.

The indicator  $Y_f$  is used to quantify the impact of meteorological environment on the system generation potential, because if the feasibilities of several PV systems with different capacities are compared,  $Y_f$  is more applicative than  $E$  which is one of the traditional indicators. The indicator  $Y_r$  is used as one of the indicators, because compared with the horizontal irradiation per square meter, which is the other traditional indicator,  $Y_r$  is more appropriate to quantify the impact of meteorological environment on the solar resources in PVSF study. We also take PR into the indicator system, because PR can be used to quantify the impact of meteorological environment on the system performance. Thus, we establish

an indicator system by optimizing the two traditional indicators and introducing a new one. More scientific and reasonable analysis of the impact of the meteorological environment on the PVSF can be realized based on the proposed indicator system.

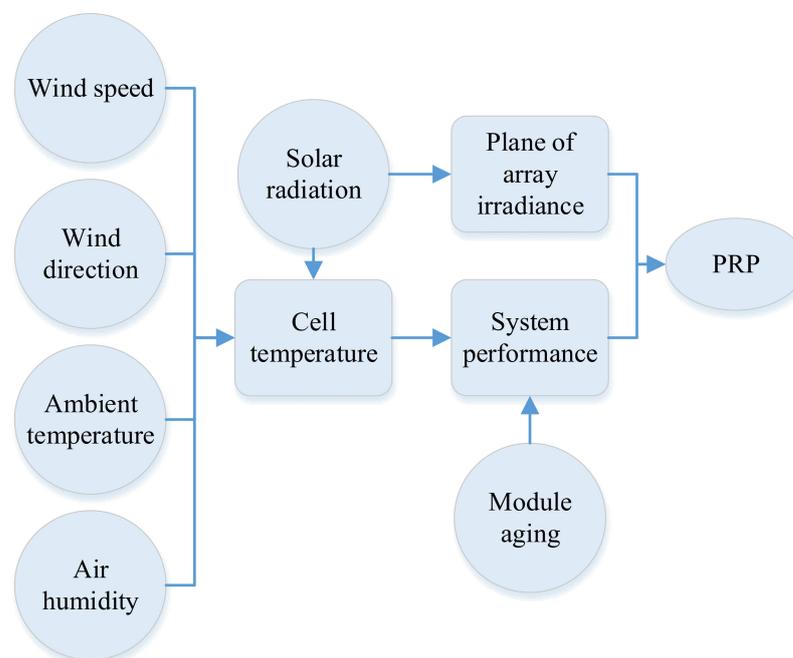
## 2.2. Essential Influences on the PV System Performance

To improve the indicator prediction accuracy by making the most of the data resources and increasing the no extra data measurement cost, the influences on the system performance were analyzed and reduced based on the data available in the PVSF study report and several influences were taken as the essential ones, which correspond to the inputs of the indicator prediction model. In this study, the performance ratio in the sampling period (PRP) expressed by the Equation (4) is taken as the parameter of the PV system performance:

$$\text{PRP} = \frac{p}{P_0}, \quad (4)$$

where  $p$  is the system power in sampling period.

The solar radiation, WS, wind direction, AT, air humidity (AH), and degradation in the performance of the PV module due to aging are taken as the essential influences (Figure 1), because, according to the definition of the PRP, the determinants are POAI and system performance. Solar radiation, WS, wind direction, and AT can affect cell temperature [23,25,27–29], and indirectly affect the PRP. Solar radiation also affects the PRP by affecting the POAI. The reason the AH is taken as one of the essential influences is that this factor has some effect on the heat dissipation of PV modules and on the PRP. Degradation in the performance of PV modules is also taken as an essential influence, because in the long run, the effect of this factor on the PRP is too significant to be ignored.



**Figure 1.** The essential influences on the PRP.

The data of the parameters which correspond to the essential influences is available in PVSF study report, but the available data resources are limited in the PVSF study stage. To increase the no extra data measurement cost, some other influences on the PRP, such as front surface soiling, mismatch losses, dust accumulation, incident solar spectrum, etc. were reduced, although this may introduce some errors into the prediction results. However, this error is minimized with the proposed indicator prediction model.

### 2.3. Proposed Indicator Prediction Model

The steps of the proposed indicator prediction model are:

- (1) Utilize the RFR model to learn the functional relationship between the PRP and the parameters of the essential influences in Section 2.2 with the data of the modelling system.
- (2) Predict the PRP of the system to be evaluated using the data measured at the site of the system and the functional relationship established in the previous step.
- (3) Predict the indicators through statistics and further calculation with the predicted PRP dataset.

#### 2.3.1. RFR Model

The RFR model is an improved bagging regression tree (RT) model [50–53], and one of the most effective machine learning models for forecasting [54]. Thus, the RFR model is used to forecast the indicators to analyze the impact of the meteorological environment on the PVSF in this paper. The process can be divided into the following five essential steps as shown in Figure 2.

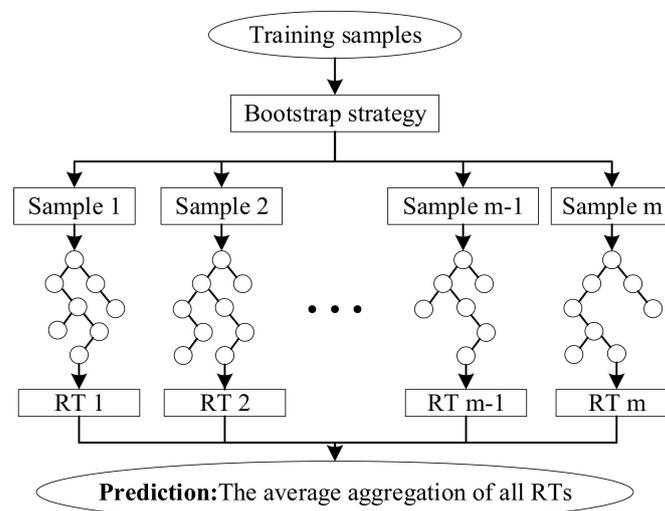


Figure 2. RFR model.

- (1) A sample set is obtained via bootstrap strategy. The bootstrap strategy aims to generate a dataset by resampling existing training samples. One-third of the whole training samples are left out. These samples form the out-of-bag (OOB) data. The other samples are called in-bag data which are used to create a RT of the RFR.
- (2) A RT is trained based on the in-bag data by a binary number of splits. More specifically, the training phase are: (a) If the number of all features is  $a$ , randomly select  $k$  ( $k < a$ ) features from the all features as a feature subset of the current node; (b) Pick an optimal feature from the  $k$  features to split the current node into two daughter nodes. (c) Repeat (a) and (b) until ending up at a leaf node.
- (3) Repeat (1) and (2)  $m$  times to generate  $m$  RTs which make up the trained RFR model. Breiman [48] suggests to set the number of trees as 500 (default).
- (4) The RFR model provides a significant measure, variable-importance (VI) measure, which aims to calculate the contribution of each input to the output. The calculating process are: (1) Calculate the prediction error of the result which is predicted by using a RT and its relative OOB data. (2) Randomly alter the sequence of the element of an input vector in the OOB data. Then calculate the increased prediction error. (3) The definition of the VI measure is the mean value of the increased prediction errors before and after altering the input vector over all RTs in the forest.
- (5) After obtaining the required number of RTs, the final predicted response of the developed RFR model is the average of the RT predictions in the test data set.

### 2.3.2. Model Training

The system with the same type of modules as the system to be evaluated was taken as the modelling system for two reasons. The first is that under the same conditions, the degradation rates of different types of modules may be different. Bayandelger et al. [55] found that the annual degradation rates of multicrystalline silicon (mc-Si) and single-crystalline silicon (sc-Si) are  $-1.28\%$  and  $-0.86\%$ , respectively. The other reason is that the relationships between the electrical efficiency and the operating temperature of different types of modules are also different. Polysilicon PV modules is the most sensitive to module temperature variations, compared with amorphous silicon (a-Si) PV modules and stacked a-Si PV modules [39], and a-Si PV modules are more affected by temperature than monocrystalline PV modules [36]. Under different climatic conditions, the degradation rates of PV modules also vary for the same technology of solar cells forming part of the PV module, which has a negative impact on the prediction performance of the proposed model. However, this impact can be minimized through selecting the PV system with similar climatic conditions to the system to be evaluated as the modelling system.

To obtain the essential influence set on the PRP, some influences were reduced for which data is not available in the PVSF study report. To minimize the error caused by this reduction, the system with similar geomorphic conditions and the closest distance to the system to be evaluated was selected as the modeling system, because we believe that the external environment of those reduced factors is very similar in the locations of the two systems when the modeling system is chosen in this way. Thus, the impact of the reduced factors on the functional relationship between the parameters which correspond to the essential influences and the PRP of the modeling system is similar to the system to be evaluated.

In training the RFR model, the process of selecting the hyper-parameters is as follows:

- (1) Preset the hyper-parameters which include the number of regression trees ( $N_{\text{tree}}$ ), the number of random variables ( $N_{\text{var}}$ ), and the number of observations in leaves ( $N_{\text{obs}}$ ). Then train the model and calculate the OOB error (OOBE).
- (2) Increase the value of  $N_{\text{tree}}$  from 1 until the OOBE is no longer increased, then select the value at this point for  $N_{\text{tree}}$ .
- (3) Increase the value of  $N_{\text{var}}$  from 1 to the input number of the RFR model which is 9 in this article, then select the value with the minimum OOBE for  $N_{\text{var}}$ .
- (4) Increase the value of  $N_{\text{obs}}$  from 1 until the OOBE begins to increase significantly, then select the value with the minimum OOBE for  $N_{\text{obs}}$ .

### 2.3.3. Inputs of the RFR Model

The PRP was taken as the output of the RFR model, and the inputs of the model were the WS, wind incidence angle on the module (WIAM), AT, AH, global horizontal irradiance (GHI), diffuse horizontal irradiance (DHI), direct normal irradiance (DNI), POAI, and the system operated days (SOD). The data of the WS, AT, AH, GHI, DHI, DNI and SOD were derived from the measured historical dataset. The POAI data was obtained through the Hay model [10], and the WIAM data were acquired according to formula (5):

$$\theta_{\text{ws}} = \arccos(\sin \beta \cdot \cos \theta_{\text{wg}}), \quad (5)$$

where  $\theta_{\text{ws}}$  is the WIAM,  $\beta$  is the inclined angle of the module,  $\theta_{\text{wg}}$  is the angle between the wind direction and the azimuth of the module which was obtained through formula (6):

$$\begin{cases} \theta_{\text{wg}} = 180^\circ - |\gamma_{\text{m}} - \theta_{\text{w}}|, & \gamma_{\text{m}} \geq 0, \\ \theta_{\text{wg}} = 180^\circ - |\gamma_{\text{m}} + 360^\circ - \theta_{\text{w}}|, & \gamma_{\text{m}} < 0, \end{cases} \quad (6)$$

where  $\gamma_{\text{m}}$  is the azimuth of the module,  $\theta_{\text{w}}$  is the wind direction.

The inputs of the RFR model correspond to the essential influences on the PRP. WS, AT, and AH correspond to the eponymous influences; GHI, DHI, DNI, and POAI correspond to the solar radiation; SOD corresponds to the module aging; WIAM corresponds to the

wind direction. The WIAM was taken as one of the predictors instead of wind direction, because the effective windward area of a PV module, which has some effect on the heat dissipation and thus affects the performance of the module, depends not only on the wind direction, but also on the module azimuth.

#### 2.3.4. Data and Data Preprocessing

Both the modelling dataset [56] and the testing dataset [57] were derived from the Desert Knowledge Australia Solar Centre (Desert Knowledge Australia, the Australian Government, the Northern Territory Government and the project managers; Ekistica do not endorse, and accept no legal liability whatsoever arising from, or connected to, the outcomes and conclusions associated with the use of data from the Desert Knowledge Australia Solar Centre).

In the experiment, two systems (see Table 1) were selected as the modelling system and the system to be evaluated. In the selected period, the sample missing rates of the sample sets were so small (see Table 2) that the influence of missing samples on the experiment was ignored.

**Table 1.** System Parameters.

System	NO.	Technology	Array Rating (kW)	Installed Year
The modelling system	site 30	Q CELLS poly-Si	5.61	2013
The system to be evaluated	site 35	Elkem poly-Si	5.5	2014

**Table 2.** Sample missing rate.

System	Dates	Sample Size	Sample Missing Rate
The modelling system	1 December 2013 to 30 November 2014	105,089	0.03%
The system to be evaluated	9 December 2014 to 8 December 2015	105,083	0.04%

The meteorological data measured in the data acquisition year was taken as the representative of the annual meteorological data in the lifecycle (25 years) of the system to be evaluated, since it is almost impossible to accurately predict the meteorological information in the next 25 years, and the interannual variation of meteorological information can be ignored in analyzing the impact of the meteorological environment on the PVSF.

Cleaning redundant data not only promotes the efficiency of the prediction model, but also avoids the interference of redundant data to the prediction. In this paper, if the system power is almost zero, the relevant sample is considered as a redundant sample. For the modeling system, the system power value can be used to determine whether the sample is redundant or not. However, for the system to be evaluated, there is no system power data, so the redundant samples cannot be cleaned according to the power value. When the cosine of zenith angle is less than or equal to zero, the solar radiation is almost zero, so in this paper, the data was cleaned according the cosine of zenith angle.

#### 2.3.5. Statistics of Indicators

Due to the difference in the rated conversion efficiency of the modules between the modelling system and the system to be evaluated, the prediction results of the RFR model were modified as follows:

$$\text{PRP}_f = \text{PRP}'_f \cdot \frac{\eta_f}{\eta_m}, \quad (7)$$

where  $\text{PRP}_f$  is the PRP of the system to be evaluated,  $\text{PRP}'_f$  is the prediction result of the RFR model,  $\eta_f$  is the rated conversion efficiency of the modules of the system to be evaluated, and  $\eta_m$  is the rated conversion efficiency of the modules of the modelling system.

Based on the  $PRP_f$  dataset, the values of each indicator in the first year of the lifecycle were obtained. The annual output energy values were obtained by Formula (8), and the annual values of each indicator were got according to the definitions (Formulas (1)–(3)). The predicted annual average values of  $Y_f$ , PR and  $Y_r$  in the lifecycle of the system are taken as the quantitative results of the indicators of the evaluation system.

$$E_k = E \cdot (1 - \Delta\eta)^{k-1}, \quad 1 \leq k \leq 25 \quad (8)$$

where  $E_k$  is the output energy in the  $k$  year,  $E$  is the output energy in the first year, and  $\Delta\eta$  is the guaranteed annual PV module degradation rate.

According to the definition of the indicators, both  $Y_f$  and PR are linearly correlated with  $E$ , so the prediction errors of  $E$ ,  $Y_f$  and PR are the same. The errors of the results were analyzed through the Mean Absolute Percentage Error (MAPE) and Root Mean-Square Percentage Error (RMSPE) of the predicted monthly values of  $E$  in the first year. There are two sources of error in the proposed model:

- (1) The intrinsic error existing in RFR model;
- (2) The error caused by reducing the nonessential influences on the PRP in this paper.

The errors of the proposed model and the reference model were obtained based on the predicted results and the measurements of the system to be evaluated. The intrinsic error existing in RFR model was got based on the OOB prediction results and the measurements of the modelling system.

### 2.3.6. Reference Indicator Prediction Model

The main difference between the reference model used in this paper and the proposed model is that the models to obtain the PRP are different. In the proposed model, PRP is predicted based on the RFR model. The WS, WIAM, AT, AH, GHI, DHI, DNI, POAI, and SOD were taken as the inputs of the model. In the reference model, PRP is predicted based on the SPS. The GHI, POAI and AT were taken as the inputs of the model. The proposed model is based on the RFR model. So before predicting the indicators of the system to be evaluated, the RFR model need to be trained with the data of the modeling system. The reference model is based on the SPS. So, the indicators of the system to be evaluated can be predicted directly with the data of this system.

The main steps of the reference model to predict the PRP are:

- (1) Obtain the IOP of the system by Formula (9).

$$IOP = P_0 \cdot \frac{POAI}{I_R} \quad (9)$$

where the POAI is obtained by using the Hay model [10], and the  $I_R$  is the Standard Test Irradiance ( $1000 \text{ W/m}^2$ ).

- (2) Estimate the operating temperature of PV modules by Formula (10).

$$T_O = GHI \frac{NOCT - 20}{800} + AT \quad (10)$$

where  $T_O$  is the operating temperature and NOCT is the Normal Operating Cell Temperature.

- (3) Correct the IOP by Formula (11).

$$p_C = IOP \left[ 1 + (T_O - 25) \frac{C_T}{100} \right] \quad (11)$$

where  $p_C$  is the corrected power output and the  $C_T$  is the Cell Temperature Coefficient.

- (4) Predict the indicators according to the corresponding definitions.

### 3. Results and Discussion

#### 3.1. Data Preprocessing Results

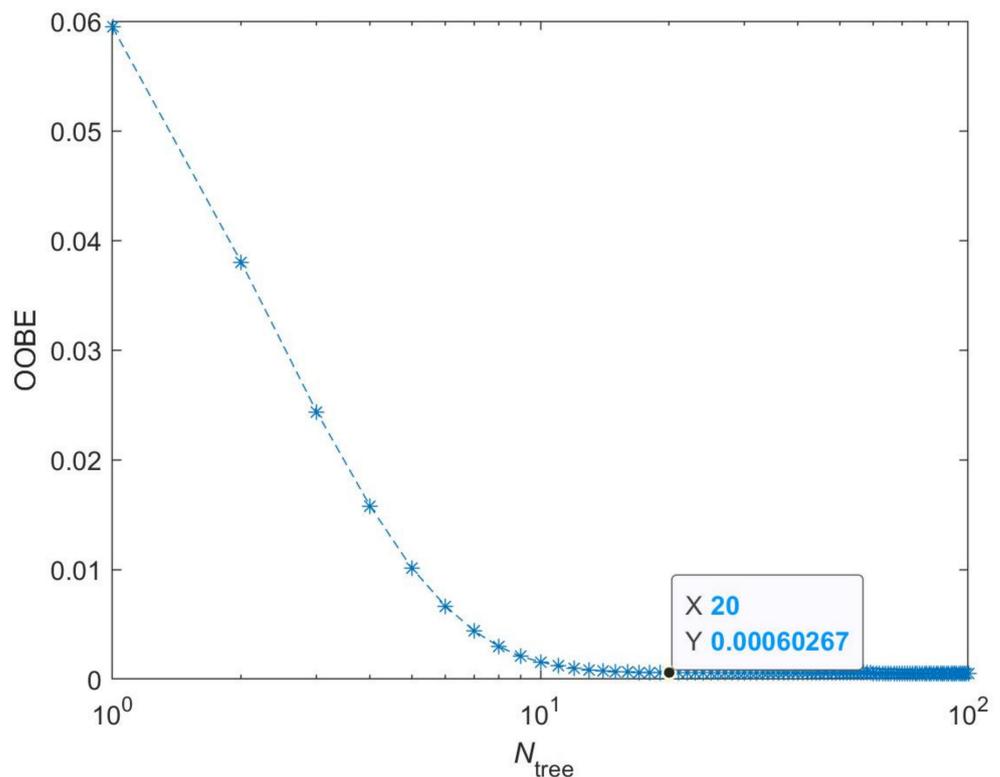
We found that, by using the proposed sample cleaning method, the cleaning rates of the two systems were so high (Table 3) that a large number of redundant samples were cleaned, which improved the prediction efficiency and avoided the interference of redundant samples in the modelling and prediction process.

**Table 3.** Sample cleaning results.

System Role	Original Sample Size	Cleaned Sample Size	Sample Cleaning Rate
The modelling system	105,089	51,686	49.18%
The system to be evaluated	105,083	50,545	48.10%

#### 3.2. Hyper-Parameters Selection Results

In selecting the hyperparameters of the RFR model, the initial value of  $N_{\text{tree}}$  was set at 100,  $N_{\text{var}}$  was 3, and  $N_{\text{obs}}$  was 5. The OOB E was no longer decreased when  $N_{\text{tree}}$  was greater than 20 (Figure 3). With the increase of  $N_{\text{var}}$  from 1 to 9, the OOB E decreased at first and then increased, and the value of  $N_{\text{var}}$  corresponding to the minimum OOB E was 4 (Figure 4). The OOB E decreased at the early stage and then increased with the increase of  $N_{\text{obs}}$  from 1, and the value of  $N_{\text{obs}}$  corresponding to the minimum OOB E was 5 (Figure 5).



**Figure 3.** Variation of the OOB E with the  $N_{\text{tree}}$  increasing from 1 to 100.

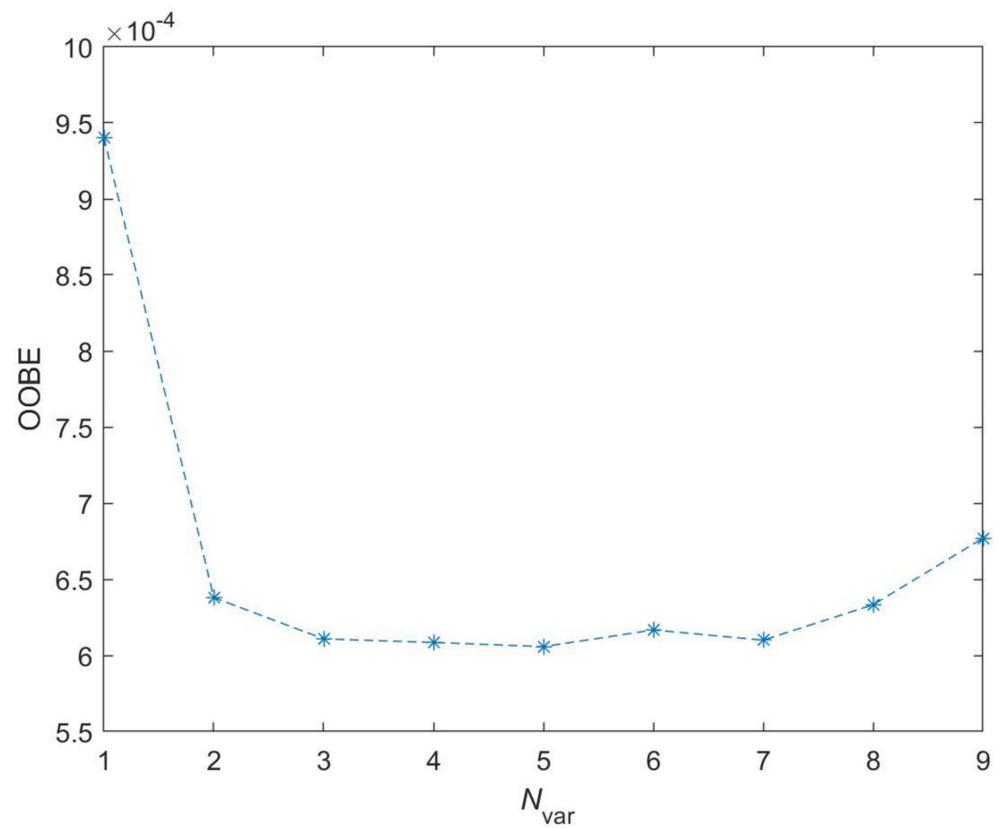


Figure 4. Variation of the OOBE with the  $N_{var}$  increasing from 1 to 9.

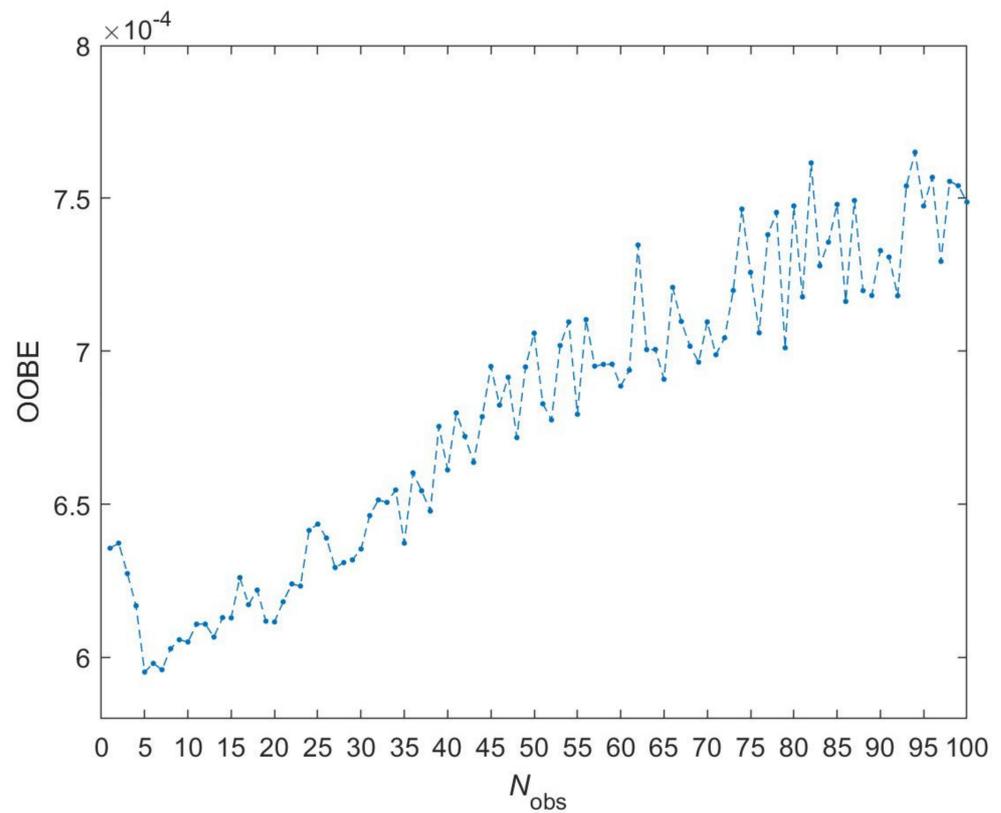


Figure 5. Variation of the OOBE with the  $N_{obs}$  increasing from 1 to 100.

### 3.3. VIs Calculation Results

We found that the VIs of the inputs were positive and on the same order of magnitude (Figure 6), which means the correlations between the influences corresponding to the inputs and the PRP are approximate. Compared with the reference model, which can be used to analyze the impacts of solar radiation and AT, more essential influences on the PVSF can be quantitatively analyzed by using the proposed model.

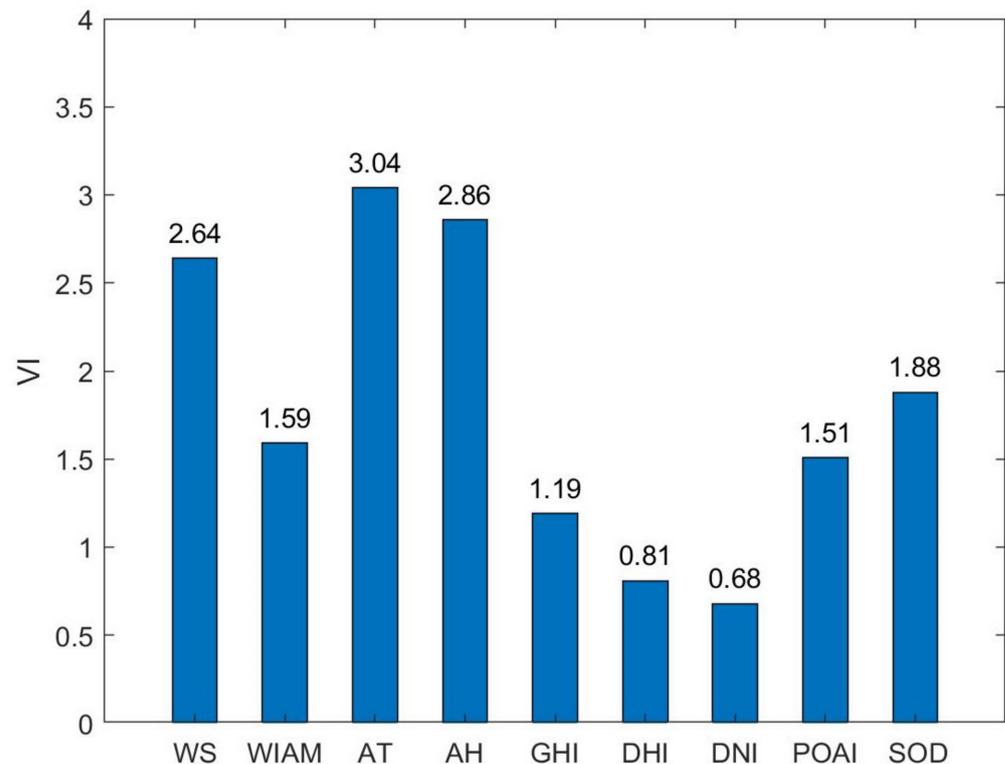


Figure 6. VI of each input.

### 3.4. Indicators Predation Results

The most important input is AT, followed by AH, WS, etc., but it was quite unexpected that the inputs GHI, DHI, DNI, and POAI, which correspond to the influencing factor, solar radiation, would be of lower importance. We believe that the main reason for this is that the four variables were strongly correlated, and the importance of each variable was weakened.

The monthly values of  $Y_T$  were statistically obtained based on the measured dataset, so the predicted values are equal to the observed ones (Figure 7). The monthly  $E$ ,  $Y_f$ , and PR predicted by using the proposed model are generally closer to the observed values than the predictions of the reference model (Figures 8–10), which indicates that the stability of the proposed model is higher than the reference model. The monthly observed results are slightly lower than those predicted by us, which, we believe, was most probably because there was some inevitable lost yield due to the unexpected shutdown of the system when the system runs in practice. However, the predictions of the reference model are generally lower than the observed values, which may be because the empirical value of the transmission efficiency of PV system commonly used in the reference model is lower than the real value.

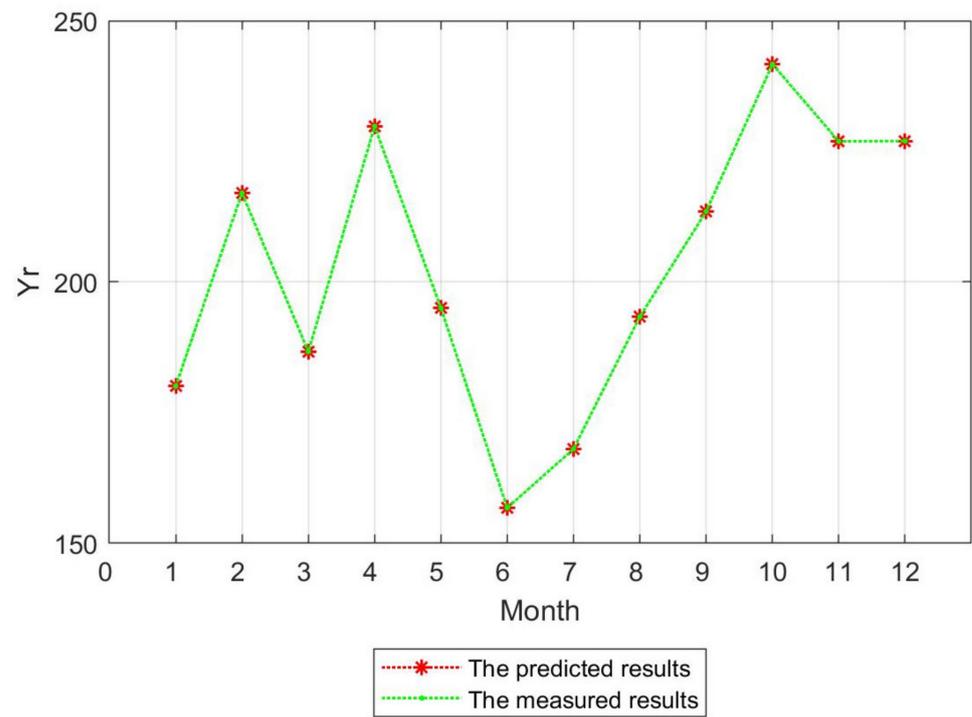


Figure 7. Monthly values of  $Y_r$ .

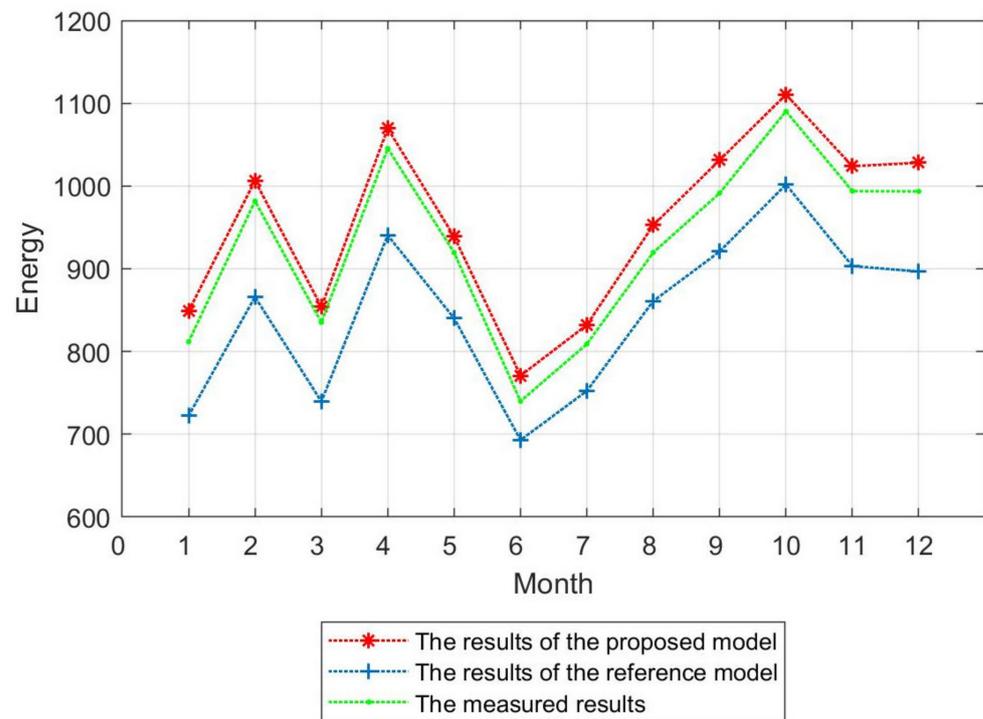


Figure 8. Monthly values of  $E$ .

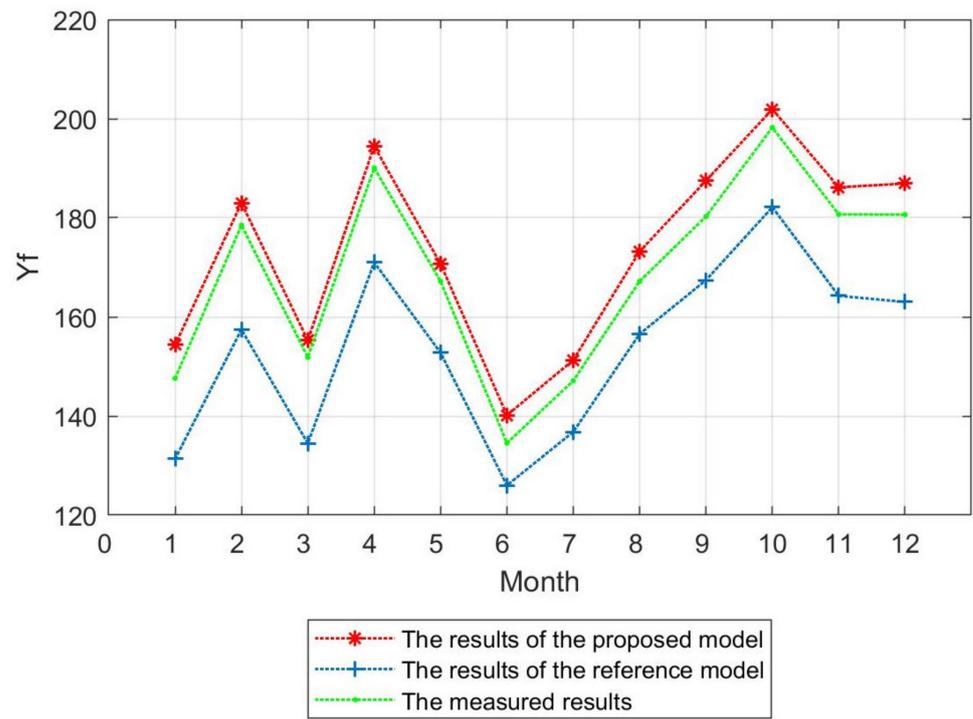


Figure 9. Monthly values of  $Y_f$ .

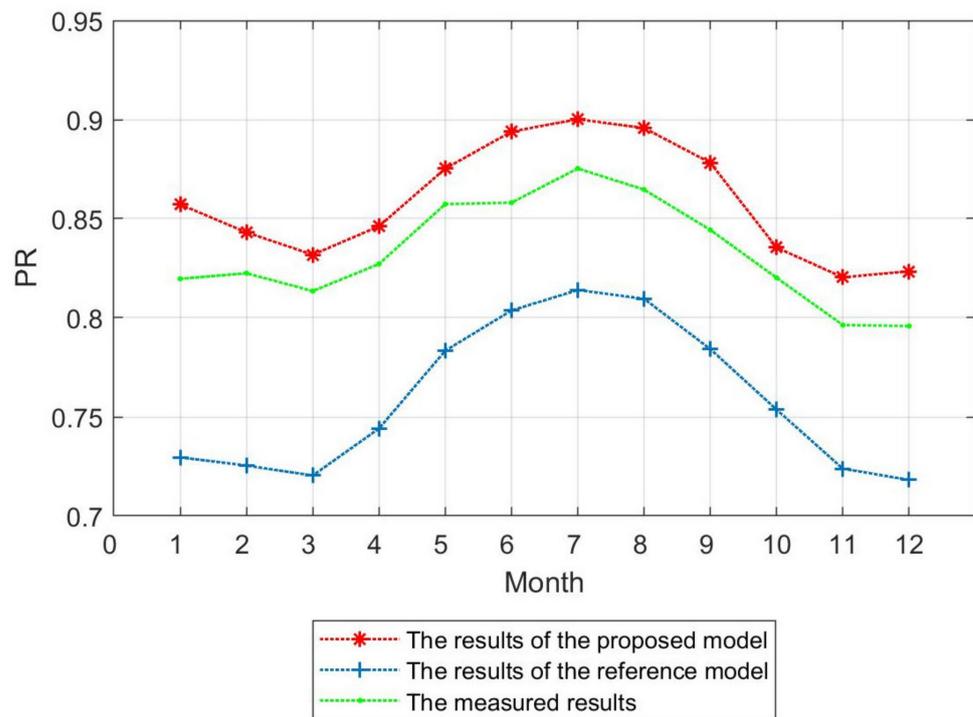


Figure 10. Monthly values of PR.

The annual average values of  $Y_f$ , PR and  $Y_r$  (Table 4) in the lifecycle of the system, which were taken as the results of the indicators to analyze the impact of meteorological environment on the PVSF in this paper, were obtained based on the predicted results of the first year.

**Table 4.** Predicted values of the indicators.

Indicators	Predicted Value
$Y_f$ (h)	1868
$Y_r$ (h)	2436
PR	0.77

The MAPE, due to the intrinsic error of the RFR model, is 0.26%, and the RMSPE is 0.21%. The MAPE of the result of the proposed prediction model is 3.06%, and the RMSPE is 0.9%. The MAPE of the result predicted by using the reference model is 8.89%, and the RMSPE is 1.95%. The intrinsic error of the RFR model is smaller than the error caused by reducing the nonessential influences on the PRP, which is the other part of the error resources of the proposed model. Although reducing the influences on the PRP has some negative impact on the accuracy and stability in predicting the indicators, compared with the reference model, the proposed model still has greater advantages.

#### 4. Conclusions

In this study, a model to analyze the impact of the meteorological environment on the PVSF was proposed. To quantify this impact, a new indicator system was proposed by optimizing the two traditional indicators and introducing a new one. Compared with traditional indicators, the proposed indicator system is more comprehensive and conducive to comparing the feasibilities of several PV systems with different capacities. It is hard to improve the indicators prediction accuracy by using the reference model. Since the input sets of the proposed short-term energy prediction models do not match with the data available in the PVSF study report, the indicators cannot be predicted based on these models.

To predict the indicators, a prediction model based on multimeteorological factors and RFR model was proposed. In this model, to improve the indicator prediction accuracy by making the most of the data resources and increasing the no extra data measurement cost, the influences on the system performance were analyzed and reduced based on the data available in the PVSF study report, and several influences were taken as the essential ones, which corresponded to the inputs of the indicator prediction model. According to the calculated VIs of the inputs, the set of essential influences and the corresponding inputs we created were found to adequately support the prediction of the indicators. To eliminate redundant data of the system to be evaluated, a data cleaning model applicable to PVSF study stage was proposed. In the experiment, the original dataset was preprocessed by using the proposed data cleaning model, and a large number of redundant samples were cleaned. The error analysis demonstrates that the precision of the proposed prediction method is higher than that of the reference model.

In this paper, the indicators were predicted based on the data resources available in the PVSF study report. The accuracy of the prediction results depends not only on the prediction model, but also on the available data. Therefore, if the data resources in the PVSF study report are more abundant in future, the accuracy of the prediction results can be further improved.

**Author Contributions:** All of the authors have contributed to this research. F.X. and G.P. provided the research ideas and guided the revision of the draft; D.M. and H.S. performed the case studies and analysis and wrote and revised the paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Key Research and Development Program of China [Grant No. 2017YFA0700300], Zhejiang Province Key Research and Development Program [Grant No. 2021C01112], and Zhejiang Provincial Natural Science Foundation of China (LGG21E060001).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data supporting reported results can be found at <http://dkasolarcentre.com.au/download?location=alice-springs>, accessed on 17 May 2021.

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

## References

1. Ahmed, S.M.M.; Al-Amin, M.R.; Ahammed, S.; Ahmed, F.; Saleque, A.M.; Rahman, M.A. Design, construction and testing of parabolic solar cooker for rural households and refugee camp. *Sol. Energy* **2020**, *205*, 230–240. [CrossRef]
2. Wang, D.J.; Qi, T.; Liu, Y.F.; Wang, Y.Y.; Fan, J.H.; Wang, Y.; Du, H. A method for evaluating both shading and power generation effects of rooftop solar PV panels for different climate zones of China. *Sol. Energy* **2020**, *205*, 432–445. [CrossRef]
3. IRENA. Available online: [https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2021/Apr/IRENA\\_-RE\\_Capacity\\_Highlights\\_2021.pdf?la=en&hash=1E133689564BC40C2392E85026F71A0D7A9C0B91](https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2021/Apr/IRENA_-RE_Capacity_Highlights_2021.pdf?la=en&hash=1E133689564BC40C2392E85026F71A0D7A9C0B91) (accessed on 21 April 2021).
4. Khalil, S.A.; Shaffie, A. M Evaluation of transposition models of solar irradi: Iance over Egypt. *Renew. Sustain. Energy Rev.* **2016**, *66*, 105–119. [CrossRef]
5. Liu, B.Y.H.; Jordan, R.C. The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. *Sol. Energy* **1960**, *4*, 1–19. [CrossRef]
6. Koronakis, P.S. On the choice of the angle of tilt for south facing solar collectors in the Athens basin area. *Sol. Energy* **1986**, *36*, 217–225. [CrossRef]
7. Badescu, V. 3D isotropic approximation for solar diffuse irradiance on tilted surfaces. *Renew. Energy* **2002**, *26*, 221–233. [CrossRef]
8. Bugler, J.W. The determination of hourly insolation on an inclined plane using a diffuse irradiance model based on hourly measured global horizontal insolation. *Sol. Energy* **1977**, *19*, 477–491. [CrossRef]
9. Ma, C.C.Y.; Iqbal, M. Statistical comparison of models for estimating solar radiation on inclined surfaces. *Sol. Energy* **1983**, *31*, 313–317. [CrossRef]
10. Hay, J.E.; Davies, J.A. Calculation of the solar radiation incident on an inclined surface. In Proceedings of the First Canadian Solar Radiation Data Workshop, Toronto, ON, Canada, 1 January 1980.
11. Skartveit, A.; Olseth, J.A. Modelling slope irradiance at high latitudes. *Sol. Energy* **1986**, *36*, 333–344. [CrossRef]
12. Muneer, T.; Kambezidis, T. *Solar Radiation and Daylight Models for the Energy Efficient Design of Buildings*; Architectural Press: Boston, MA, USA, 1997.
13. Willmott, C.J. On the climatic optimization of the tilt and azimuth of flat-plate solar collectors. *Sol. Energy* **1982**, *28*, 205–216. [CrossRef]
14. Gueymard, C. An anisotropic solar irradiance model for tilted surfaces and its comparison with selected engineering algorithms. *Sol. Energy* **1987**, *38*, 367–386. [CrossRef]
15. Reindl, D.T.; Beckman, W.A.; Duffie, J.A. Evaluation of hourly tilted surface radiation models. *Sol. Energy* **1990**, *45*, 9–17. [CrossRef]
16. Temps, R.C.; Coulson, K.L. Solar radiation incident upon slopes of different orientations. *Sol. Energy* **1977**, *19*, 179–184. [CrossRef]
17. Klucher, T.M. Evaluation of models to predict insolation on tilted surfaces. *Sol. Energy* **1979**, *23*, 111–114. [CrossRef]
18. Perez, R.; Seals, R.; Ineichen, P.; Stewart, R.; Menicucci, D. A new simplified version of the Perez diffuse irradiance model for tilted surfaces. *Sol. Energy* **1987**, *39*, 221–231. [CrossRef]
19. Wattan, R.; Janjai, S. An investigation of the performance of 14 models for estimating hourly diffuse irradiation on inclined surfaces at tropical sites. *Renew. Energy* **2016**, *93*, 667–674. [CrossRef]
20. Demain, C.; Journée, M.; Bertrand, C. Evaluation of different models to estimate the global solar radiation on inclined surfaces. *Renew. Energy* **2013**, *50*, 710–721. [CrossRef]
21. Khalil, S.A.; Shaffie, A.M. A comparative study of total, direct and diffuse solar irradiance by using different models on horizontal and inclined surfaces for Cairo, Egypt. *Renew. Sustain. Energy Rev.* **2013**, *27*, 853–863. [CrossRef]
22. Shan, F.; Tang, F.; Cao, L.; Fang, G. Comparative simulation analyses on dynamic performances of photovoltaic–thermal solar collectors with different configurations. *Energy Convers. Manag.* **2014**, *87*, 778–786. [CrossRef]
23. Skoplaki, E.; Palyvos, J.A. On the temperature dependence of photovoltaic module electrical performance: A review of efficiency/power correlations. *Sol. Energy* **2009**, *83*, 614–624. [CrossRef]
24. Radziemska, E. The effect of temperature on the power drop in crystalline silicon solar cells. *Renew. Energy* **2003**, *28*, 1–12. [CrossRef]
25. Teo, H.G.; Lee, P.S.; Hawlader, M.N.A. An active cooling system for photovoltaic modules. *Appl. Energy* **2012**, *90*, 309–315. [CrossRef]
26. Ayompe, L.M.; Duffy, A.; McCormack, S.J.; Conlon, M. Measured performance of a 1.72 kW rooftop grid connected photovoltaic system in Ireland. *Energy Convers. Manag.* **2011**, *52*, 816–825. [CrossRef]
27. Griffith, J.S.; Rathod, M.S.; Paskalski, J. Some tests of flat plate photovoltaic module cell temperatures in simulated field conditions. In Proceedings of the Fifteenth IEEE Photovoltaic Specialists Conference, Pasadena, CA, USA, 12–15 May 1981.
28. Skoplaki, E.; Palyvos, J.A. Operating temperature of photovoltaic modules: A survey of pertinent correlations. *Renew. Energy* **2009**, *34*, 23–29. [CrossRef]

29. Dubey, S.; Sarvaiya, J.N.; Seshadri, B. Temperature dependent photovoltaic (PV) efficiency and its effect on PV production in the world—A review. In Proceedings of the PV Asia Pacific Conference, Singapore, 23–25 October 2012.
30. Jordan, D.C.; Kurtz, S.R. Photovoltaic degradation Rates—An analytical review. *Prog. Photovolt. Res. Appl.* **2013**, *21*, 12–29. [[CrossRef](#)]
31. Ndiaye, A.; Charki, A.; Kobi, A.; Kebe, C.M.F.; Ndiaye, P.A.; Sambou, V. Degradations of silicon photovoltaic modules: A literature review. *Sol. Energy* **2013**, *96*, 140–151. [[CrossRef](#)]
32. Kalogirou, S.A.; Agathokleous, R.; Panayiotou, G. On-site PV characterization and the effect of soiling on their performance. *Energy* **2013**, *51*, 439–446. [[CrossRef](#)]
33. Mejia, F.A.; Kleissl, J. Soiling losses for solar photovoltaic systems in California. *Sol. Energy* **2013**, *95*, 357–363. [[CrossRef](#)]
34. Ullah, A.; Imran, H.; Maqsood, Z.; Butt, N.Z. Investigation of optimal tilt angles and effects of soiling on PV energy production in Pakistan. *Renew. Energy* **2019**, *139*, 830–843. [[CrossRef](#)]
35. Lorente, D.G.; Pedrazzi, S.; Zini, G.; Dalla Rosa, A.; Tartarini, P. Mismatch losses in PV power plants. *Sol. Energy* **2014**, *100*, 42–49. [[CrossRef](#)]
36. Touati, F.A.; Al-Hitmi, M.A.; Bouchech, H.J. Study of the effects of dust, relative humidity, and temperature on solar PV Performance in Doha: Comparison between Monocrystalline and Amorphous PVS. *Int. J. Green Energy* **2013**, *10*, 680–689. [[CrossRef](#)]
37. Beattie, N.S.; Moir, R.S.; Chacko, C.; Buffoni, G.; Roberts, S.H.; Pearsall, N.M. Understanding the effects of sand and dust accumulation on photovoltaic modules. *Renew. Energy* **2012**, *48*, 448–452. [[CrossRef](#)]
38. Said, S.A.M.; Walwil, H.M. Fundamental studies on dust fouling effects on PV module performance. *Sol. Energy* **2014**, *107*, 328–337. [[CrossRef](#)]
39. Nagae, S.; Toda, M.; Minemoto, T.; Takakura, H.; Hamakawa, Y. Evaluation of the impact of solar spectrum and temperature variations on output power of silicon-based photovoltaic modules. *Sol. Energy Mater. Sol. Cells* **2006**, *90*, 3568–3575. [[CrossRef](#)]
40. Behrendt, T.; Kuehnert, J.; Hammer, A.; Lorenz, E.; Betcke, J.; Heinemann, D. Solar spectral irradiance derived from satellite data: A tool to improve thin film PV performance estimations? *Sol. Energy* **2013**, *98*, 100–110. [[CrossRef](#)]
41. Leitao, D.; Torres, J.P.N.; Fernandes, J.F.P. Spectral irradiance influence on solar cells efficiency. *Energies* **2020**, *13*, 5017. [[CrossRef](#)]
42. Mellit, A.; Pavan, A.M.; Oglari, E.; Leva, S.; Lughi, V. Advanced methods for photovoltaic output power forecasting: A review. *Appl. Sci.* **2020**, *10*, 487. [[CrossRef](#)]
43. Das, U.K.; Tey, K.S.; Seyedmahmoudian, M.; Mekhilef, S.; Idris, M.Y.I.; Van Deventer, W.; Horan, B.; Stojcevski, A. Forecasting of photovoltaic power generation and model optimization: A review. *Renew. Sustain. Energy Rev.* **2018**, *81*, 912–928. [[CrossRef](#)]
44. Raza, M.Q.; Nadarajah, M.; Ekanayake, C. On recent advances in PV output power forecast. *Sol. Energy* **2016**, *136*, 125–144. [[CrossRef](#)]
45. Antonanzas, J.; Osorio, N.; Escobar, R.; Urraca, R.; Martinez-de-Pison, F.J.; Antonanzas-Torres, F. Review of photovoltaic power forecasting. *Sol. Energy* **2016**, *136*, 78–111. [[CrossRef](#)]
46. Sobri, S.; Koochi-Kamali, S.; Abd Rahim, N. *Sol. photovoltaic generation forecasting methods: A review. Energy Convers. Manag.* **2018**, *156*, 459–497. [[CrossRef](#)]
47. Wang, H.Z.; Liu, Y.Y.; Zhou, B.; Li, C.B.; Cao, G.Z.; Voropai, N.; Barakhtenko, E. Taxonomy research of artificial intelligence for deterministic solar power forecasting. *Energy Convers. Manag.* **2020**, *214*, 112909. [[CrossRef](#)]
48. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
49. IEC. *Photovoltaic System Performance—Part 3: Energy Evaluation Method*, 1.0 ed.; IEC TS 61724-3; IEC: Geneva, Switzerland, 2016.
50. Voyant, C.; Notton, G.; Kalogirou, S.; Nivet, M.L.; Paoli, C.; Motte, F.; Fouilloy, A. Machine learning methods for solar radiation forecasting: A review. *Renew. Energy* **2017**, *105*, 569–582. [[CrossRef](#)]
51. Prasad, A.M.; Iverson, L.R.; Liaw, A. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems* **2006**, *9*, 181–199. [[CrossRef](#)]
52. 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](#)]
53. Yu, P.S.; Yang, T.C.; Chen, S.Y.; Kuo, C.M.; Tseng, H.W. Comparison of random forests and support vector machine for real-time radar-derived rainfall forecasting. *J. Hydrol.* **2017**, *552*, 92–104. [[CrossRef](#)]
54. Benali, L.; Notton, G.; Fouilloy, A.; Voyant, C.; Dizene, R. Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components. *Renew. Energy* **2019**, *132*, 871–884. [[CrossRef](#)]
55. Bayandelger, B.E.; Ueda, Y.; Batbayar, B.; Adiyabat, A.; Enebish, N.; Otani, K. Long-term performance and degradation rate analyses of photovoltaic modules exposed in the Gobi Desert, Mongolia. *Jpn. J. Appl. Phys.* **2018**, *57*, 6. [[CrossRef](#)]
56. Desert Knowledge Australia Centre. 1 December 2013 to 30 November 2014. Download Data: Array 30 Q CELLS. Alice Springs. Available online: <http://dkasolarcentre.com.au/download?location=alice-springs> (accessed on 27 July 2020).
57. Desert Knowledge Australia Centre. 9 December 2014 to 8 December 2015. Download Data: Array 35 Elkem. Alice Springs. Available online: <http://dkasolarcentre.com.au/download?location=alice-springs> (accessed on 27 July 2020).