energies

MDPI

Communication

Exploring the PV Power Forecasting at Building Façades Using Gradient Boosting Methods

Jesús Polo *, Nuria Martín-Chivelet, Miguel Alonso-Abella, Carlos Sanz-Saiz, José Cuenca and Marina de la Cruz

Photovoltaic Solar Energy Unit, Renewable Energy Division, CIEMAT, Avda. Complutense 40, 28040 Madrid, Spain

* Correspondence: jesus.polo@ciemat.es

Abstract: Solar power forecasting is of high interest in managing any power system based on solar energy. In the case of photovoltaic (PV) systems, and building integrated PV (BIPV) in particular, it may help to better operate the power grid and to manage the power load and storage. Power forecasting directly based on PV time series has some advantages over solar irradiance forecasting first and PV power modeling afterwards. In this paper, the power forecasting for BIPV systems in a vertical façade is studied using machine learning algorithms based on decision trees. The forecasting scheme employs the skforecast library from the Python environment, which facilitates the implementation of different schemes for both deterministic and probabilistic forecasting applications. Firstly, deterministic forecasting of hourly BIPV power was performed with XGBoost and Random Forest algorithms for different cases, showing an improvement in forecasting accuracy when some exogenous variables were used. Secondly, probabilistic forecasting was performed with XGBoost combined with the Bootstrap method. The results of this paper show the capabilities of Random Forest and gradient boosting algorithms, such as XGBoost, to work as regressors in time series forecasting of BIPV power. Mean absolute error in the deterministic forecast, using the most influencing exogenous variables, were around 40% and close below 30% for the south and east array, respectively.

Keywords: BIPV; PV power forecasting; machine learning; gradient boosting algorithms

Citation: Polo, J.; Martín-Chivelet, N.; Alonso-Abella, M.; Sanz-Saiz, C.; Cuenca, J.; de la Cruz, M. Exploring the PV Power Forecasting at Building Façades Using Gradient Boosting Methods. *Energies* **2023**, *16*, 1495. https://doi.org/10.3390/en16031495

Academic Editors: Fouzi Harrou, Ying Sun, Bilal Taghezouit and Dairi Abdelkader

Received: 22 December 2022
Revised: 16 January 2023
Accepted: 20 January 2023
Published: 2 February 2023

1. Introduction

The large expansion and growth of photovoltaic (PV) systems and the foreseen future increase are driving a strong impulse for distributed power generation. PV operation in urban environments and building integrated photovoltaics (BIPV) are good examples of topics that are significantly drawing the attention of the PV community and industry [1]. BIPV comprises those modules and systems that can serve as conventional building components and produce energy at the same time. Architectural, power and aesthetic aspects are all important in designing BIPV systems, and thus a wide variety of designs are emerging in the industry [2–4]. The specific features of BIPV systems have an impact on the thermal and electrical performance so adapted or new testing procedures, standards and modeling methodologies are being developed [5,6]. Modeling BIPV systems is challenging since the modules work under frequent partial shading conditions and also their thermal boundary conditions can be different from conventional PV plants [7–9]. Parametric 3D tools and models are being used to better approach the complex geometries and shading patterns produced by neighboring buildings [10]. In addition, the usual PV cell temperature models may be less accurate in the case of PV in façades affecting, consequently, the power prediction accuracy [11]. Challenges in modeling PV façades performance have been remarked on in several previous works of the authors at Ciemat using different approaches [12,13].

Solar forecasting is an important part of the management of solar power generation. In particular, PV solar power forecasting is helpful in electric grid management, load management and battery storage operation. A comprehensive overview of solar forecasting,

including basic concepts, models and selected examples can be found in the reports delivered by IEA-PVPS Task 16 [14]. Most developments in solar forecasting have been made in the prediction of solar irradiance as a previous step to solar power forecasting, and many important contributions have been reported in recent years [15,16]. Thus, PV power forecast is performed by forecasting solar irradiance as a first step and deriving afterwards the PV power with a simulation model [17]. An alternative way consists of forecasting directly the AC PV power from historical time series, which requires no detailed knowledge of the PV system and the involved meteorological variables (module temperature and irradiance forecasts) and relying on the quality of the historical data [18,19].

Time series PV power forecasting is usually performed by autoregressive models or by machine learning methods. The latter are being recently used preferably due to the availability of advanced and powerful machine learning algorithms [20–23]. Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) are two of the most popular methods for forecasting time series of solar irradiance [24,25]. However, gradient boosting decision trees (GBDT) methods are gaining popularity in the most recent years due to their accuracy and computation speed proven in many Kaggle data science competitions. XGBoost (extreme gradient boosting) and CatBoost (categorical boosting) algorithms are two of the most successful GBDT methods for regression and classification [26,27].

Despite there being many works in the literature dealing with solar irradiance forecasting, PV power forecasting based on time series is much less studied and, in particular, BIPV power forecasting is practically not covered by the solar forecasting community. The present work is thus aimed at contributing to filling the gap of studies focused on BIPV power forecasting. BIPV power forecast applying the usual developments in solar irradiance forecasting involves deriving the Plane of Array (POA) irradiance from the forecasted global irradiance using a transposition model and then modeling the power with a PV model, which also requires the forecasting of the module temperature. Therefore, the forecasting uncertainty of the strategy based on solar irradiance forecasting has to include the combined uncertainty of the intermediate steps (transposition, power modeling, temperature and so on). In comparison, BIPV power forecasting using time series and machine learning only requires the measuring values of the power in the past, and it might be a more suitable approach. There are many works dealing with machine learning in solar forecasting, and most of them are based on solar irradiance forecasting. Support vector machines, neural networks and Long Short-Term Memory (LSTM) networks are very popular [28–30]. However, for large datasets, parameter tuning can be complicated [31]. Decision trees are versatile, fast and very efficient machine learning algorithms due to their simplicity. In this work, we are exploring the performance of the XGBoost and Random Forest models in forecasting hourly time series of BIPV power in a building at Ciemat headquarters using different exogenous variables. In addition to deterministic forecasting with machine learning, this work explores the capability of the XGBoost algorithm in probabilistic forecasting. Both Random Forest and XGBoost can be implemented in a straightforward way into forecasting schemes; in particular, they were implemented in the *skforecast* scheme (a very powerful library for addressing time series forecasting problems in python). In this work, the forecasting of BIPV power for two small arrays (south- and east-oriented) has been analyzed with these two algorithms. General good performance is observed in both methods for the case of deterministic forecasting, where the mean absolute error was placed at around 30% in the case of the east-oriented array and around 40% in the case of the south-oriented.

2. Materials and Methods

2.1. Test Facility and Experimental Data

In this work, we are using two years of monitored data corresponding to two PV arrays integrated into the south and east façades of Building 42 at Ciemat headquarters in Madrid (Spain). The geographic coordinates of the site are 40.45° N, −3.73° E, and the altitude is 695 m; the climatology according to the Koeppen climate classification is Csa

(i.e., dry-summer subtropical, often referred to as Mediterranean climate). The PV modules are situated at the top of each façade (Figure 1). The arrays consist of monocrystalline silicon modules with seven series by two parallel (7sx2p) configurations for the east array and 7sx4p configuration for the south array. The south array, with a power of 8 kW, contains SunPower E18-325 modules (305 W each one) and a Fronius IG Plus 100 V-3 inverter. The east array, with a nominal power of 4 kW, consists of SunPower E20-327 modules (327 W each one) and a Fronius IG Plus 50 V-1 inverter.

Figure 1. View of the south and east arrays in the Ciemat building.

The experimental database contains hourly mean values of power, current and voltage at maximum power for each array, as well as hourly values of the module temperature and ambient temperature. Global horizontal irradiance and POA irradiance at south and east azimuths are measured with calibrated PV cells and are also available on an hourly basis.

2.2. Methodology

Gradient boosting decision trees (GBDT) are a machine learning technique that combines the prediction from multiple models (weak learners) to obtain a better predictive performance in an iterative and sequential way, where each predictor fixes its predecessor's error. XGBoost is a powerful and very popular open-source algorithm due to its speed and accuracy. XGBoost is a parallelized and optimized version of the gradient boosting algorithm [32]. It has been successfully used in many different disciplines involving big data science during the last recent years [26]. On the other hand, Random Forest is another popular effective ensemble machine learning algorithm, also based on decision trees, which has been used recently in solar irradiance forecasting [33,34]. Both algorithms can be implemented in a time series forecasting scheme. In this work, we have used the *skforecast* library in python, which contains the necessary classes and functions to adapt any Scikit-learn regression model, such as XGBoost and Random Forest, to forecasting problems [35].

The data used in this work consist of two years (2017–2018) of monitored PV power on an hourly basis of two BIPV arrays corresponding to the façade south and east of a building. The data are divided into three sets, namely the training, validation and test datasets. The model is first trained with the training set, and secondly, it is evaluated with the validation set using back-testing in order to select the combination of hyper-parameters and lags leading to the lowest error. Finally, the model is trained again with the best combination using both training and validation data. Figure 2 shows the different data sets

selected for this work. Using the *skforecast* scheme, we have trained two different models to forecast hourly values of the power of each array using 72 lags. Explorative work, using back-testing, has been addressed by trying a different number of lags and has shown that 72 lags yield better results than larger or shorter ones. The hyper-parameters are certain values or weights that control the learning process. In this work, the hyper-parameters selected were: maximum depth per tree which took 3, 5 and 10 as possible values; the learning rate (0.01 and 0.1) for controlling the step size at each iteration; the number of trees in the ensemble, which was selected in 100 and 500.

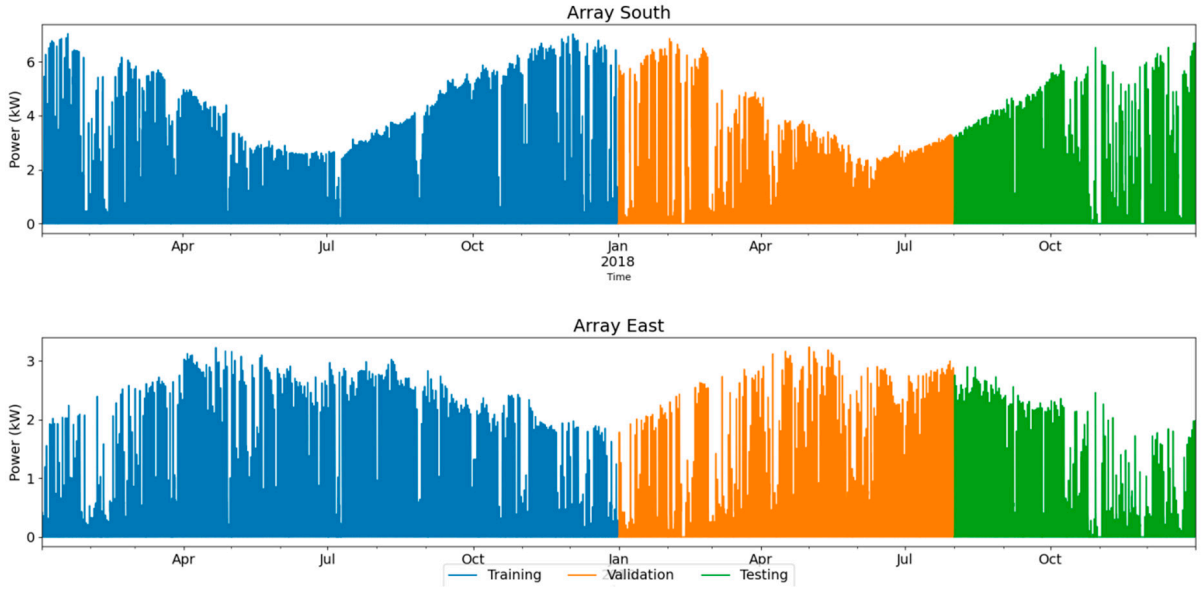

Figure 2. BIPV power datasets for south and east arrays.

In addition to the predictors (i.e., the array power), it is possible to add exogenous variables that may explain the regression model. Exogenous variables are those parameters that can be known or derived in the future. Clear sky irradiance, solar geometry-related variables and shadowing fraction are plausible examples of exogenous variables. In this work, we have explored the role of several exogenous variables: POA irradiance under clear sky (POA_cs), cosine of the incidence angle (cosAOI), shaded fraction of array surface (FS), sun azimuth (SunAz) and month number (m). The solar geometry and the POA irradiance under clear sky are calculated with the pvlib library [36], which contains the most updated information and models for solar resource computation. POA_cs is computed by calculating first the three solar irradiance components (global, direct and diffuse irradiance) for a horizontal surface using Ineichen's model [37]; then, Perez's model is used to transpose the horizontal components to the vertically tilted surfaces [38]. The shaded fraction of array surface was calculated in previous works where LIDAR data were used for computing a digital surface model (DSM) of the building and surroundings that allows shadow computation throughout the year [13]. In order to explore which exogenous variables work better in the forecasting scheme, the case matrix in Table 1 has been tested for each array and each algorithm.

Table 1. Case matrix for exploring the role of exogenous variables.

Case	Predictor	Exogenous Variables
Base Case	Power	None
Case 1	Power	POA_cs
Case 2	Power	POA_cs, cosAOI
Case 3	Power	POA_cs, cosAOI, FS
Case 4	Power	POA_cs, cosAOI, FS, SunAz
Case 5	Power	POA_cs, cosAOI, FS, SunAz, m

The correlation coefficients of each exogenous variable with the array power calculated with the Pearson linear correlation (R_{xy}) were, in descending order, 0.75, 0.63, 0.45, 0.09 and −0.08 for POA_cs, cosAOI, FS, m and SunAz, respectively.

$$R_{xy} = \frac{N \sum xy - \sum x \sum y}{\sqrt{N \sum x^2 - (\sum x)^2} \sqrt{N \sum y^2 - (\sum y)^2}} \quad (1)$$

where x represents the power and y refers to any exogenous variable.

Figure 3 shows the autocorrelation of the power time series for the two arrays under study.

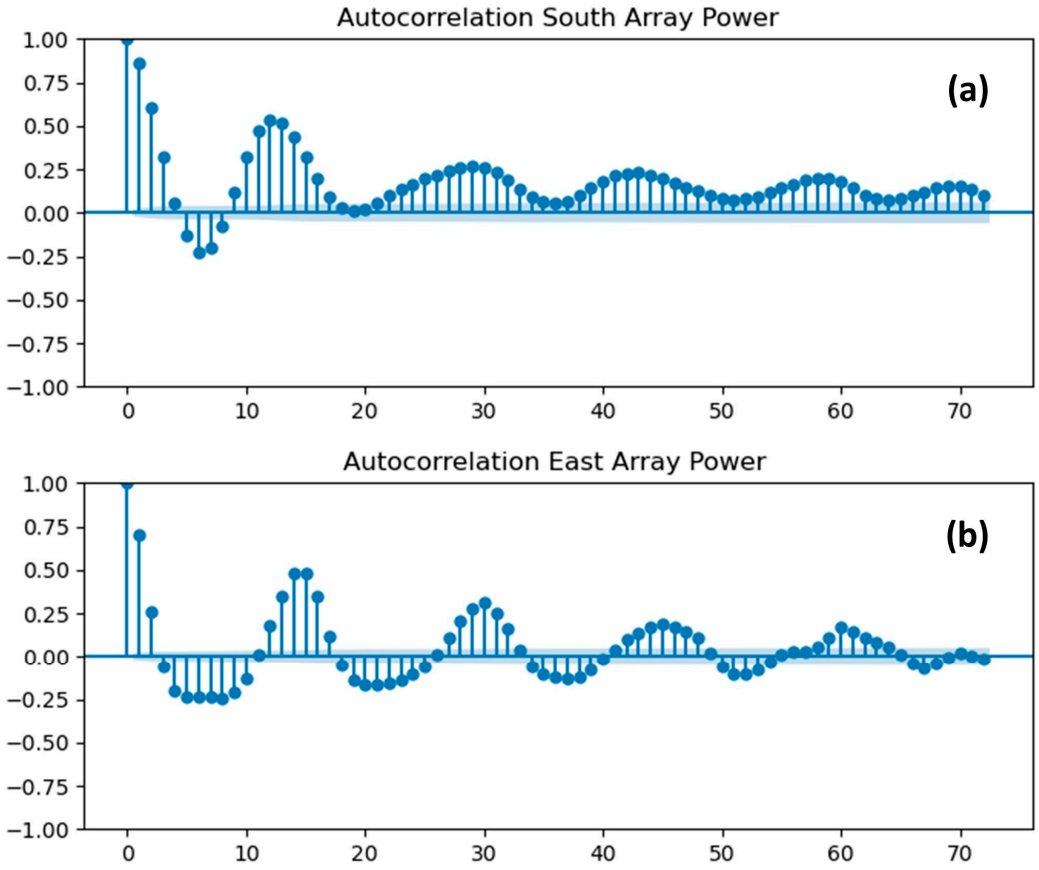

Figure 3. Autocorrelation plots for the array power: (**a**) South array. (**b**) East array.

3. Results

3.1. Deterministic Forecasting Using XGBoost and Random Forest

Deterministic forecasting refers to the prediction of single values (hourly values in the case of this work) in the future. Hence, hourly values of BIPV array power are forecasted for each array covering the different cases listed in Table 1 using XGBoost and Random Forest as regression algorithms in the skforecast scheme. The results are validated with the test dataset. Figures 4 and 5 show the scatter plot of the hourly power forecasted values for the south and east arrays, respectively.

Random Forest as a regressor in the forecasting model resulted in slightly more accurate forecasting than XGBoost. Furthermore, in both cases, the use of exogenous variables in the forecasting scheme improves the results.

The metrics used in the evaluation of forecasting the hourly power are mean bias error (MBE), normalized mean absolute error (nMAE) and root mean squared error (RMSE) [39].

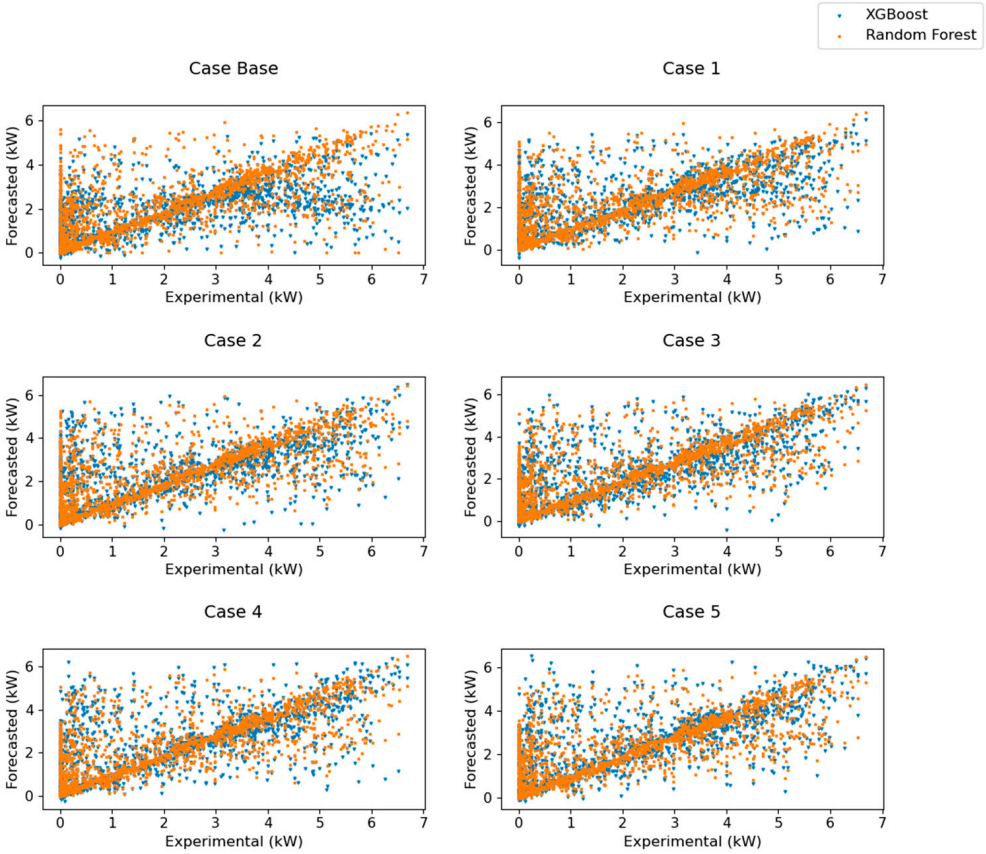

Figure 4. Scatter plot of forecasted power for south array.

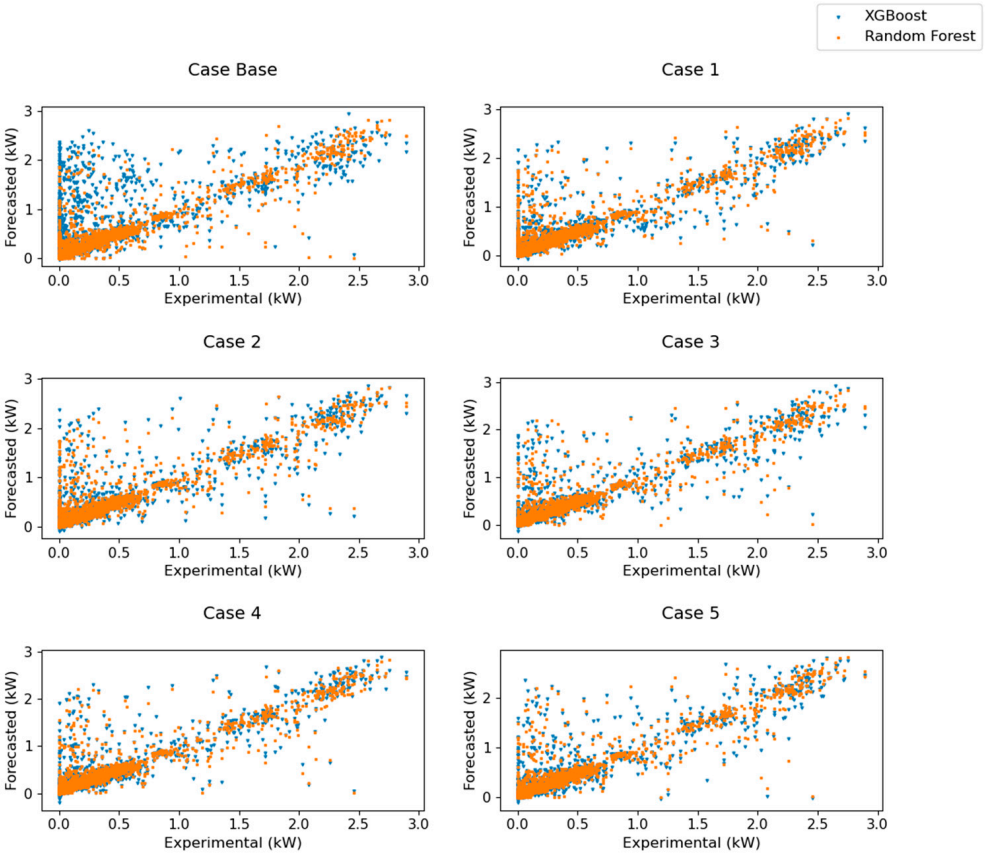

Figure 5. Scatter plot of forecasted power for east array.

$$MBE = \frac{1}{n} \sum \left(P_{measured} - P_{forecasted} \right) \quad (2)$$

$$nMAE = \frac{1}{\langle P_{measured} \rangle} \frac{1}{n} \sum abs \left(P_{measured} - P_{forecasted} \right) \quad (3)$$

$$RMSE = \sqrt{\frac{\sum \left(P_{measured} - P_{forecasted} \right)^2}{n}} \quad (4)$$

where $P_{measured}$ is the experimental power of the array, $P_{forecasted}$ refers to the forecasted power values and $\langle P_{measured} \rangle$ is the arithmetic mean value of $P_{measured}$.

Figure 6 compares the MBE and RMSE metrics for XGBoost and Random Forest in both arrays. In the case of array south, the nMAE ranges from 54% to 42% with XGBoost, and from 43% to 40% with the Random Forest algorithm used in the *skforecast* scheme. On the other hand, in the case of array east, the nMAE is very high for the case base with XGBoost; the nMAE of all the studied cases varies from 72% to 28%. For Random Forest, the nMAE is in the range of 29–23%. The determination coefficient R^2 varies in the range of 0.60–0.69 in both methods (Random Forest and XGBoost) for the south array, and it reaches 0.80–0.85 in the case of the east array. Therefore, in all cases, the use of exogenous variables as additional explicative variables improves the performance of the forecasting. Moreover, the forecasting scheme performs better for the array east than for the south. For the south array, the use of clear sky in-plane irradiance (POA_cs) as an exogenous variable produces power forecasting with the minimum bias compared to the other cases. In the case of array east, the inclusion of shadowing information (FS) in addition to POA_cs as exogenous variables resulted in a more accurate forecast since this façade is much more influenced by shadows than the south-oriented one. Finally, Figure 7 illustrates the hourly power forecasted in several days for south and east arrays with the best performance cases, which are case 1 and case 3 (Table 1). The forecasting results improve significantly with the use of those exogenous variables with higher correlation coefficients, which correspond to case 3. Cases 4 and 5 do not add significant improvement to case 3. In terms of RMSE, the forecasting using Random Forest performs slightly better than XGBoost for most of the cases. However, it should be remarked that in the south array, case 3 (the best case of exogenous variables) resulted in much lower bias. Lowering the bias is important in forecasting when the dispersion of the forecasted data (RMSE) are very similar with both machine learning methods.

3.2. Probabilistic and Intervals Forecasting

The *skforecast* library allows the estimation of forecasting intervals based on bootstrapped residuals. Detailed information on the bootstrap method for interval prediction in solar power can be found in recent literature [40]. For forecasting intervals, case 3 of Table 1 has been selected since the inclusion of such exogenous variables improves the accuracy of the forecast. Moreover, preliminary simulations found that XGBoost was more suitable than Rain Forest for this type of forecast. Thus, the XGBoost regressor has been used with the bootstrap method to produce forecasting intervals. The results for forecasting 10–90% intervals with the bootstrap method for several days are illustrated for both arrays in Figure 8.

In order to evaluate the probabilistic prediction results, several metrics are frequently used. The Prediction Interval Coverage Probability (PICP) is defined as the proportion of the data covered by the forecasted interval [31].

$$\text{PICP} = \frac{1}{n} \sum c_i \, , \; c_i = \begin{cases} 1, & lower_i \leq c_i \leq upper_i \\ 0, & otherwise \end{cases} \quad (5)$$

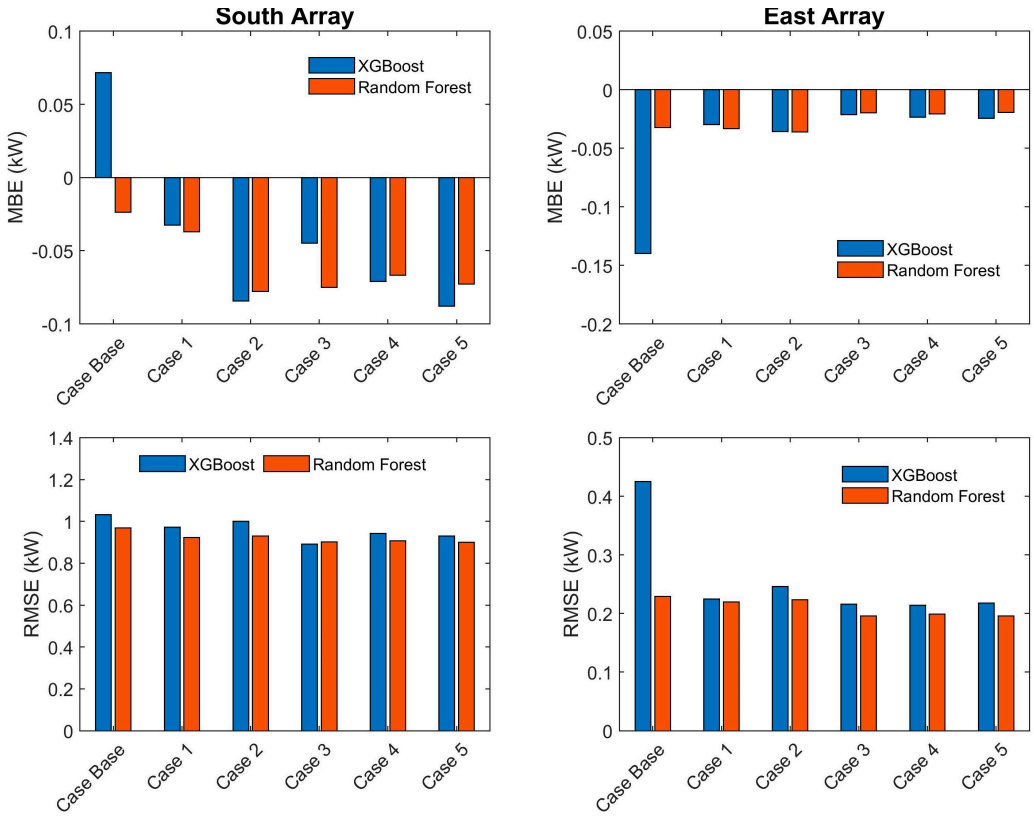

Figure 6. MBE and RMSE values for both arrays and all the cases.

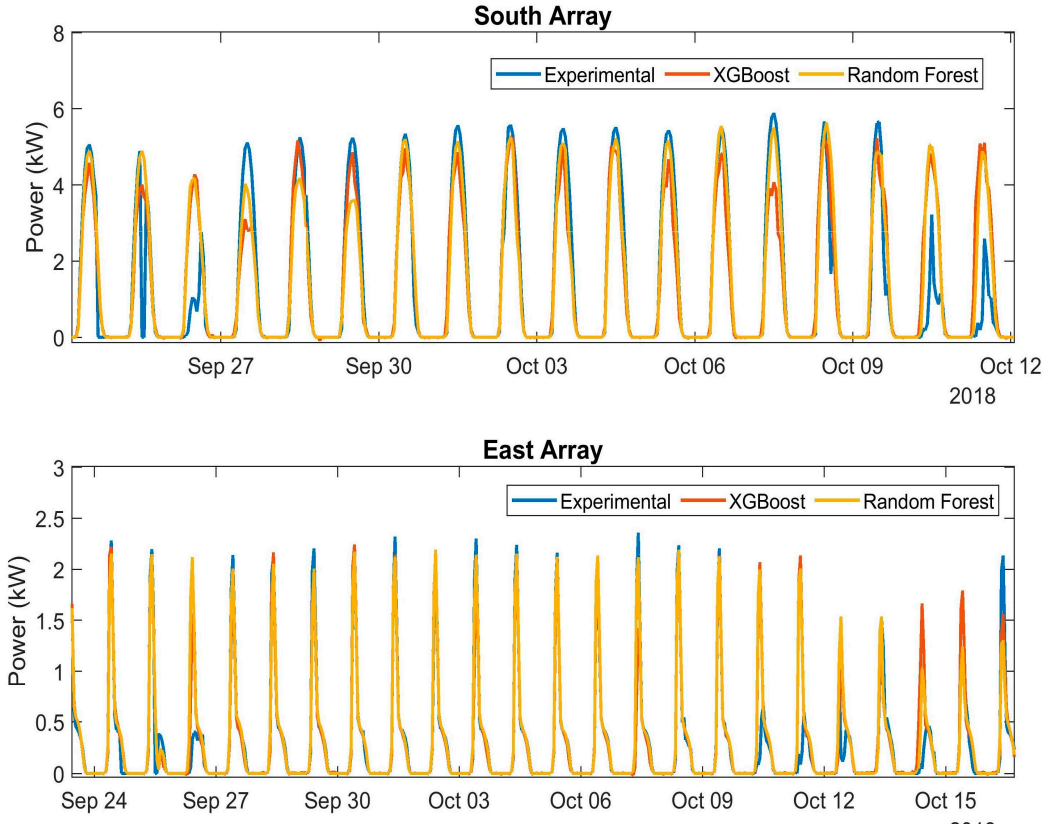

Figure 7. Hourly power forecasted for south and west arrays for the best performance cases.

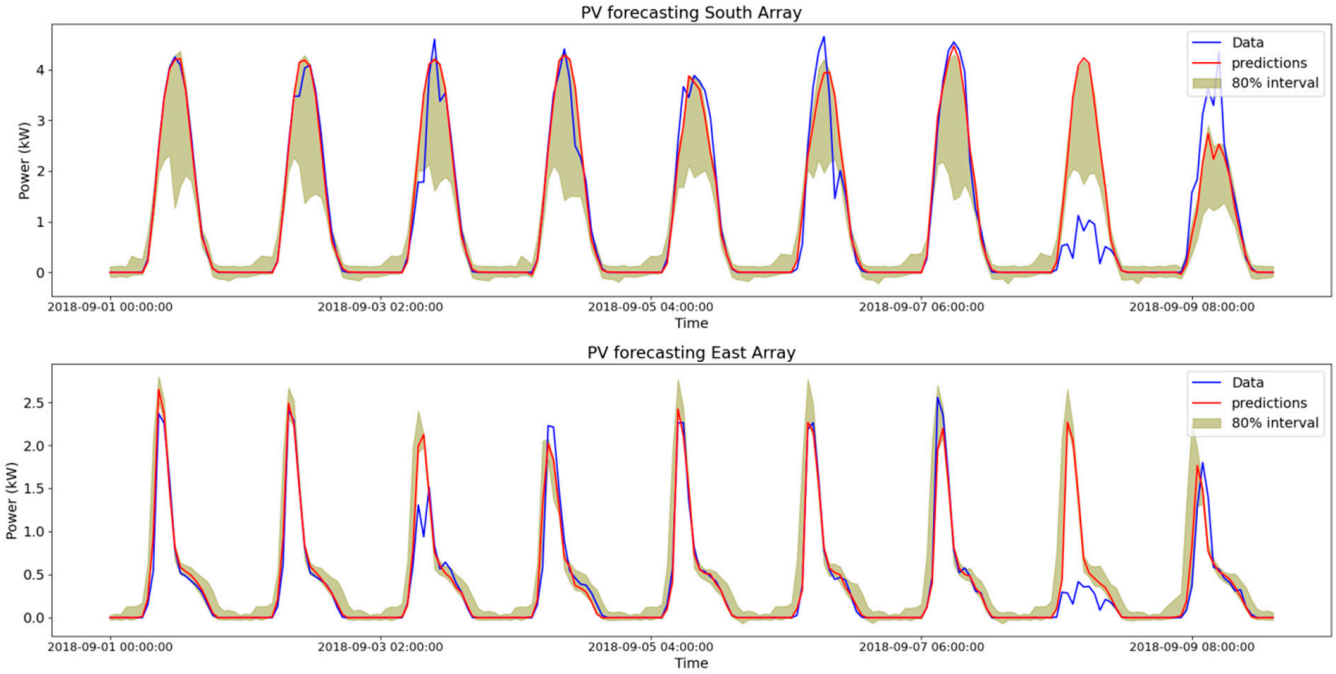

Figure 8. Probabilistic forecasting of the BIPV power for south and east arrays.

The Mean Prediction Interval Width (MPIW) is defined as,

$$\text{MPIW} = \frac{1}{R\,n} \sum(upper_i - lower_i) \tag{6}$$

where R is the range of the target values (i.e., the difference between maximum and minimum power of the array). The higher the PICP value, the higher the reliability of the forecast, and the lower the MPIW value, the narrower the width of the forecasted interval. Table 2 shows the metrics for the south and east arrays. The coverage of both arrays is below 80% in both cases, but higher reliability is achieved in forecasting intervals for the east array. The intervals of probabilistic forecasting of the east array are narrower than in the case of the south array, and the coverage (PICP) is also greater (nearly 80%).

Table 2. Metrics for evaluating interval forecasting.

Array	PICP (%)	MPIW (%)
South	69.52	11.34
East	78.01	7.05

4. Discussion

Forecasting PV power directly offers some advantages over forecasting solar irradiance and modeling afterwards the PV power from the forecasted irradiance. In the case of BIPV, and particularly for PV modules in façade, the intermediate steps between solar irradiance forecasting and PV power forecasting include the use of transposition models, temperature forecasting and PV modeling. The alternative strategy is to directly forecast the PV power based on historical values (i.e., time series forecasting). Machine learning algorithms have evolved and grown considerably in the recent few years, and nowadays, there are available tools that integrate machine learning algorithms with forecasting schemes. The library *skforecast* is a good example of a tool that makes it easier to use scikit-learn machine learning regressors as forecasters.

Decision tree regressors such as XGBoost and Random Forest have been explored in this work as forecasters for BIPV power hourly forecasting. Both deterministic and probabilistic forecasting schemes have been applied to two BIPV arrays (south and east

façades) monitored in a building in Madrid (Spain). Furthermore, XGBoost and Random Forest can be effectively integrated into a deterministic forecasting scheme. In both cases, the use of exogenous variables (i.e., those variables that can be securely known in the future), such as clear sky irradiance, solar geometry-related variables and shading patterns in the building, resulted in more accurate forecasted hourly values. The minimum mean absolute errors achieved were around 40% for the south array and close to 30% in the case of the East array. In addition, probabilistic forecasting has been explored using the XGBoost regressor with bootstrapped residuals. Better results were found for the East array, where the coverage of the forecasted interval was 78%. This work has shown the potential of these new forecasting tools for both deterministic and probabilistic forecasting of BIPV power on an hourly basis. The combination of different machine learning algorithms and different forecasting strategies (e.g., recursive multi-steps, hyper-parameter tuning, back-testing) could conduct more accurate results. Future work is expected in this regard to explore new forecasting schemes.

5. Conclusions

PV power forecasting is essential in managing the electrical grid and storage systems. In the particular case of BIPV, with the increasing growth of its deployment, it can be even more important. This paper has explored two machine learning algorithms based of decision trees in forecasting BIPV power from time series of power. Random Forest and XGBoost are able to forecast directly PV power with much better performance when highly correlated variables are used as exogenous variables. Random Forest achieved, in general, slightly better RMSE than XGBoost for most cases in deterministic forecasting. However, the best accuracy was observed in case 3, with XGBoost showing an RMSE of 0.89 kW and 0.21 kW for south and east arrays, respectively; moreover, very similar RMSE results of 0.9 kW and 0.20 kW for south and east arrays were found in the case of Random Forest. In probabilistic forecasting, XGBoost was used for south and east arrays with good results within the 10–90% interval.

XGBoost algorithm presents the advantage of a very easy and fast implementation method in forecasting schemes (as the *skforecast* library), having fast execution and making possible the use of both deterministic and probabilistic forecasting strategies based on time series forecasting. This versatility makes it a very interesting option in machine learning forecasting techniques for BIPV applications.

Author Contributions: Conceptualization, J.P. and N.M.-C.; methodology, J.P.; software, J.P.; validation, J.P., N.M.-C. and M.A.-A.; formal analysis, J.P., N.M.-C. and C.S.-S.; investigation, J.P., N.M.-C. and C.S.-S.; resources, M.A.-A., J.C. and M.d.l.C.; data curation, M.A.-A., J.C. and M.d.l.C.; writing—original draft preparation, J.P.; writing—review and editing, N.M.-C. and C.S.-S.; funding acquisition, J.P. and N.M.-C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Spanish Ministry of Science and Innovation, grant number PID2021-124910OB-C31.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank the RINGS-BIPV (Advanced Modeling and Prediction of BIPV) Project (PID2021-124910OB-C31), which is funded by the Ministerio de Ciencia e Innovación (Spain). The authors would also like to recognize the efforts, research and contributions of the experts groups of IEA PVPS Program, in particular those corresponding to Task 15 (BIPV) and Task 16 (Solar Resource), where the authors have an active collaboration.

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

References

1. Jakica, N.; Ynag, R.J.; Eisenlohr, J. *BIPV Design and Performance Modelling: Tools and Methods*; International Energy Agency: Paris, France, 2019; ISBN 9783906042862. Available online: https://www.iea.org/ (accessed on 21 December 2022).
2. Eder, G.; Peharz, G.G.; Trattni, R.; Bonomo, P.; Saretta, E.; Frontini, F.; Lopez, C.S.P.; Wilson, H.R.; Jakica, N.; Eisenlohr, J.; et al. *Report IEA-PVPS T15-07: 2019—Coloured BIPV Market, Research and Development*; International Energy Agency: Paris, France, 2019; Available online: https://www.iea.org/ (accessed on 21 December 2022).
3. Pelle, M.; Lucchi, E.; Maturi, L.; Astigarraga, A.; Causone, F. Coloured BIPV Technologies: Methodological and Experimental Assessment for Architecturally Sensitive Areas. *Energies* **2020**, *13*, 4506. [CrossRef]
4. Kuhn, T.E.; Erban, C.; Heinrich, M.; Eisenlohr, J.; Ensslen, F.; Neuhaus, D.H. Review of technological design options for building integrated photovoltaics (BIPV). *Energy Build.* **2020**, *231*, 110381. [CrossRef]
5. Martín-Chivelet, N.; Kapsis, K.; Wilson, H.R.; Delisle, V.; Yang, R.; Olivieri, L.; Polo, J.; Eisenlohr, J.; Roy, B.; Maturi, L.; et al. Building-Integrated Photovoltaic (BIPV) products and systems: A review of energy-related behavior. *Energy Build.* **2022**, *262*, 111998. [CrossRef]
6. Berger, K.; Boddaert, S.; Del Buono, M.; Fedorova, A.; Frontini, F.; Inoue, S.; Ishii, H.; Kapsis, K.; Kim, J.-T.; Kovacs, P.; et al. Analysis of requirements, specifications and regulation of BIPV. In *Report IEA-PVPS T15-08: 2019*; Inoue, S., Wilson, H.R., Eds.; International Energy Agency—PVPS: Paris, France, 2019; Available online: https://www.iea.org/ (accessed on 21 December 2022).
7. Al-Janahi, S.A.; Ellabban, O.; Al-Ghamdi, S.G. A Novel BIPV Reconfiguration Algorithm for Maximum Power Generation under Partial Shading. *Energies* **2020**, *13*, 4470. [CrossRef]
8. Zomer, C.; Custódio, I.; Antoniolli, A.; Rüther, R. Performance assessment of partially shaded building-integrated photovoltaic (BIPV) systems in a positive-energy solar energy laboratory building: Architecture perspectives. *Sol. Energy* **2020**, *211*, 879–896. [CrossRef]
9. Yadav, S.; Panda, S.; Hachem-Vermette, C. Optimum azimuth and inclination angle of BIPV panel owing to different factors influencing the shadow of adjacent building. *Renew. Energy* **2020**, *162*, 381–396. [CrossRef]
10. Walker, L.; Hofer, J.; Schlueter, A. High-resolution, parametric BIPV and electrical systems modeling and design. *Appl. Energy* **2019**, *238*, 164–179. [CrossRef]
11. Martín-Chivelet, N.; Polo, J.; Sanz-Saiz, C.; Tamara, L.; Benítez, N.; Alonso-Abella, M.; Cuenca, J. Assessment of PV Module Temperature Models for Building-Integrated Photovoltaics (BIPV). *Sustainability* **2022**, *14*, 1500. [CrossRef]
12. Polo, J.; Martín-Chivelet, N.; Sanz-Saiz, C. BIPV Modeling with Artificial Neural Networks: Towards a BIPV Digital Twin. *Energies* **2022**, *15*, 4173. [CrossRef]
13. Polo, J.; Martín-Chivelet, N.; Alonso-Abella, M.; Alonso-García, C. Photovoltaic generation on vertical façades in urban context from open satellite-derived solar resource data. *Sol. Energy* **2021**, *224*, 1396–1405. [CrossRef]
14. Sengupta, M.; Habte, A.; Gueymard, C.; Wilbert, S.; Renne, D. *Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications*, 2nd ed.; National Renewable Energy Lab. (NREL): Golden, CO, USA, 2017; Available online: https://www.iea.org/ (accessed on 21 December 2022). [CrossRef]
15. Yang, D.; Wang, W.; Gueymard, C.A.; Hong, T.; Kleissl, J.; Huang, J.; Perez, M.J.; Perez, R.; Bright, J.M.; Xia, X.; et al. A review of solar forecasting, its dependence on atmospheric sciences and implications for grid integration: Towards carbon neutrality. *Renew. Sustain. Energy Rev.* **2022**, *161*, 112348. [CrossRef]
16. Yang, D.; Kleissl, J.; Gueymard, C.A.; Pedro, H.T.; Coimbra, C.F. History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. *Sol. Energy* **2018**, *168*, 60–101. [CrossRef]
17. Lorenz, E.; Heinemann, D. 1.13—Prediction of Solar Irradiance and Photovoltaic Power. In *Comprehensive Renewable Energy*; Sayigh, A., Ed.; Elsevier: Oxford, UK, 2012; pp. 239–292. ISBN 978-0-08-087873-7.
18. 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]
19. Almeida, M.P.; Perpiñán, O.; Narvarte, L. PV power forecast using a nonparametric PV model. *Sol. Energy* **2015**, *115*, 354–368. [CrossRef]
20. Sharadga, H.; Hajimirza, S.; Balog, R.S. Time series forecasting of solar power generation for large-scale photovoltaic plants. *Renew. Energy* **2020**, *150*, 797–807. [CrossRef]
21. Theocharides, S.; Makrides, G.; Georghiou, G.E.; Kyprianou, A. Machine learning algorithms for photovoltaic system power output prediction. In Proceedings of the 2018 IEEE International Energy Conference (ENERGYCON), Limassol, Cyprus, 3–7 June 2018; pp. 1–6. [CrossRef]
22. Mittal, A.K.; Mathur, K.; Mittal, S. A Review on forecasting the photovoltaic power Using Machine Learning. *J. Physics Conf. Ser.* **2022**, *2286*, 012010. [CrossRef]
23. Bae, D.J.; Kwon, B.S.; Song, K. Bin XGboost-based day-ahead load forecasting algorithm considering behind-the-meter solar PV generation. *Energies* **2022**, *15*, 128. [CrossRef]
24. 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]
25. Narvaez, G.; Giraldo, L.F.; Bressan, M.; Pantoja, A. Machine learning for site-adaptation and solar radiation forecasting. *Renew. Energy* **2021**, *167*, 333–342. [CrossRef]

26. Hancock, J.T.; Khoshgoftaar, T.M. CatBoost for big data: An interdisciplinary review. *J. Big Data* **2020**, *7*, 1–45. [CrossRef]
27. Divina, F.; Torres, M.G.; Vela, F.A.G.; Noguera, J.L.V. A Comparative Study of Time Series Forecasting Methods for Short Term Electric Energy Consumption Prediction in Smart Buildings. *Energies* **2019**, *12*, 1934. [CrossRef]
28. Qing, X.; Niu, Y. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy* **2018**, *148*, 461–468. [CrossRef]
29. Kazem, H.A.; Yousif, J.H. Comparison of prediction methods of photovoltaic power system production using a measured dataset. *Energy Convers. Manag.* **2017**, *148*, 1070–1081. [CrossRef]
30. Zhang, J.; Verschae, R.; Nobuhara, S.; Lalonde, J.-F. Deep photovoltaic nowcasting. *Sol. Energy* **2018**, *176*, 267–276. [CrossRef]
31. Li, X.; Ma, L.; Chen, P.; Xu, H.; Xing, Q.; Yan, J.; Lu, S.; Fan, H.; Yang, L.; Cheng, Y. Probabilistic solar irradiance forecasting based on XGBoost. *Energy Rep.* **2022**, *8*, 1087–1095. [CrossRef]
32. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; ACM: San Francisco, CA, USA, 2016; pp. 785–794.
33. Ali, M.; Prasad, R.; Xiang, Y.; Khan, M.; Farooque, A.A.; Zong, T.; Yaseen, Z.M. Variational mode decomposition based random forest model for solar radiation forecasting: New emerging machine learning technology. *Energy Rep.* **2021**, *7*, 6700–6717. [CrossRef]
34. Liu, D.; Sun, K. Random forest solar power forecast based on classification optimization. *Energy* **2019**, *187*, 115940. [CrossRef]
35. Amat Rodrigo, J.; Escobar Ortiz, J. skforecast. Available online: https://www.cienciadedatos.net/py27-forecasting-series-temporales-python-scikitlearn.html (accessed on 26 September 2022).
36. Holmgren, W.F.; Hansen, C.W.; Mikofski, M.A. pvlib python: A python package for modeling solar energy systems. *J. Open Source Softw.* **2018**, *3*, 884. [CrossRef]
37. Ineichen, P.; Perez, R. A new airmass independent formulation for the Linke turbidity coefficient. *Sol. Energy* **2002**, *73*, 151–157. [CrossRef]
38. Perez, R.; Ineichen, P.; Seals, R.; Michalsky, J.; Stewart, R. Modeling daylight availability and irradiance components from direct and global irradiance. *Sol. Energy* **1990**, *44*, 271–289. [CrossRef]
39. Gueymard, C.A. A review of validation methodologies and statistical performance indicators for modeled solar radiation data: Towards a better bankability of solar projects. *Renew. Sustain. Energy Rev.* **2014**, *39*, 1024–1034. [CrossRef]
40. Li, K.; Wang, R.; Lei, H.; Zhang, T.; Liu, Y.; Zheng, X. Interval prediction of solar power using an Improved Bootstrap method. *Sol. Energy* **2018**, *159*, 97–112. [CrossRef]