Article

Photovoltaic Power Forecasting: Assessment of the Impact of Multiple Sources of Spatio-Temporal Data on Forecast Accuracy

Xwégnon Ghislain Agoua ^{*}, Robin Girard and Georges Kariniotakis

Centre for Processes, Renewable Energies and Energy Systems (PERSEE), MINES ParisTech, PSL University, CS 10207, 1 rue Claude Daunesse, 06904 Sophia Antipolis, France; robin.girard@mines-paristech.fr (R.G.); georges.kariniotakis@mines-paristech.fr (G.K.)

* Correspondence: ghislain.agoua@gmail.com

Abstract: The efficient integration of photovoltaic (PV) production in energy systems is conditioned by the capacity to anticipate its variability, that is, the capacity to provide accurate forecasts. From the classical forecasting methods in the state of the art dealing with a single power plant, the focus has moved in recent years to spatio-temporal approaches, where geographically dispersed data are used as input to improve forecasts of a site for the horizons up to 6 h ahead. These spatio-temporal approaches provide different performances according to the data sources available but the question of the impact of each source on the actual forecasting performance is still not evaluated. In this paper, we propose a flexible spatio-temporal model to generate PV production forecasts for horizons up to 6 h ahead and we use this model to evaluate the effect of different spatial and temporal data sources on the accuracy of the forecasts. The sources considered are measurements from neighboring PV plants, local meteorological stations, Numerical Weather Predictions, and satellite images. The evaluation of the performance is carried out using a real-world test case featuring a high number of 136 PV plants. The forecasting error has been evaluated for each data source using the Mean Absolute Error and Root Mean Square Error. The results show that neighboring PV plants help to achieve around 10% reduction in forecasting error for the first three hours, followed by satellite images which help to gain an additional 3% all over the horizons up to 6 h ahead. The NWP data show no improvement for horizons up to 6 h but is essential for greater horizons.

Keywords: Lasso; forecasts; photovoltaic generation; spatio-temporal; satellite images; Numerical Weather Predictions; weather stations

Citation: Agoua, X.G.; Girard, R.; Kariniotakis, G. Photovoltaic Power Forecasting: Assessment of the Impact of Multiple Sources of Spatio-Temporal Data on Forecast Accuracy. *Energies* **2021**, *14*, 1432. <https://doi.org/10.3390/en14051432>

Academic Editor: Venizelos Efthymiou

Received: 30 January 2021

Accepted: 27 February 2021

Published: 5 March 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

The urge of response to climate change and the necessity to reduce the global carbon footprint have put renewable energy in the spotlight. Photovoltaic (PV) energy generation has grown in many countries with the reduction of its costs. However, PV power generation is not controllable as it depends on the meteorological conditions. Increasing the PV penetration in the grid then require a better control of the production variability. The ability to accurately forecast the future production of the PV power plants is then decisive for both power producers and network operators.

The literature features several methods to forecast PV production. Detailed reviews of the state of the art are provided in [1–3]. They can be classified according to the forecast horizon, the available data, and the type of approach, which may be based on statistics, physics or a hybrid combination [2]. Although early methods were deterministic, probabilistic approaches are increasingly popular since they provide additional information about the distribution of future production and thus about uncertainty in the forecasts. Some of these probabilistic approaches are based on Numerical Weather Predictions (NWP) issued by meteorological models or sky imaging, and provide ensemble forecasts of the future PV generation [4–6]. Analog ensembles [7], regression trees [8,9] and k-nearest

neighbors (kNN) [10,11] are also found in the related literature on probabilistic PV forecasting. A wide range of models based on Artificial Neural Networks (ANN) also exist for short-term PV power production [12,13]. These models have evolved from simple neural networks to radial neural networks (more suitable for time series prediction) and more recently to deep learning methods [14]. Geostationary satellite imagery can be used to estimate ground irradiation. The literature mentions different methods to make this estimate. The main difference between these methods is the characterization of interactions between solar radiation and the atmosphere. Ref. [15,16] provide a review of the first methods used, classified according to whether they are physical or statistical. The various evolutions in the characterization of atmospheric phenomena and technological advances in the field of satellite imagery have led to increasingly efficient methods for deriving irradiation data from satellite images [17,18]. Satellite data can be coupled with ground irradiation measurements to improve the quality of the estimates provided; the site-adaptation method allows estimates to be compared with actual on-site measurements [19].

Ground-based irradiation data, produced either by satellite imagery alone or by coupling to ground measurements, is used to provide irradiation forecasts for horizons ranging from 0 (nowcasting) to 6 h, based on physical, statistical or hybrid prediction methods presented in literature reviews [2,20]. Among other things, cloud motion vectors (CMV) determine the speed and direction of clouds by analyzing satellite images [21–23] to provide better forecasts of irradiation. Artificial neural networks [24,25], the SVM [26,27] and the Bayesian estimation [28] are also used in the framework of the irradiation forecasting from satellite images. Approaches also include spatio-temporal methods [29,30] and methods that combine both satellite images, NWP forecasts and ground measurements [31,32].

Considering the state of the art, the key contributions of this paper can be resumed as follows: (1) we propose a spatio-temporal model which can extract and use both spatial and temporal data from the different available sources of data to improve the forecasts accuracy. The model follows a data-driven approach, where the available data are directly fed as input without other advanced pre-treatment than normalization (i.e., to produce information like cloud motion vectors); (2) we show that the large dimensionality of the model can be efficiently addressed by a Lasso approach that permits to select the most relevant input; (3) we provide a thorough quantitative comparison of the impact that the multiple heterogeneous sources of spatio-temporal data have on the forecasting performance. This data include measurements from neighboring PV plants, local meteorological stations, NWP forecasts and satellite images. Each addition of a new data source is done in relation to the forecasting horizon, making it possible to indicate which data are beneficial for which horizon. A method to define the radius of useful pixels around a PV plant for preselecting the information used as input to the model is proposed. (4) Finally, an exhaustive validation of the proposed approach is made with a real world case study comprising 136 PV installations in France. These contributions will help to build more efficient forecasting models, incite data sharing, contribute to cost-benefit analysis for new measuring infrastructures.

The paper is structured as follows: the PV data and other data sources are presented in Section 2; the proposed incremental spatio-temporal model is presented in Section 3, while Section 4 presents an evaluation and analysis of the performance of the forecasts. Finally, the conclusions of the study are discussed in Section 5.

2. Experimental Framework for Spatio-Temporal Forecasting

2.1. PV Power Data and Weather Forecasts

The data set, denoted d is a set of 136 different PV power plants in mid-west France. Each power plant is an aggregation of power inverters with peak power ranging from 3.2 kWp to 58 kWp. The distance between the power plants varies from 1 km to 230 km and the available data cover November 2014 to March 2016 with a 15 min temporal resolution. The locations of the power plants are represented in Figure 1. In the following, the power plants are labeled $P_{i,1 \leq i \leq 136}$. The production data have been normalized employing the

same procedure as that proposed in [33]. This permits to avoid that the effect of the daily course of the sun dominates in the correlations that are estimated among two sites. The NWP prediction come from the European Centre for Medium-Range Weather Forecasts (ECMWF) applying its HRES solution (<https://www.ecmwf.int/en/forecasts/datasets/> accessed on April 2017). The local meteorological measurements are obtained from the closest meteorological station (of Meteo France network).

Figure 1. The 136 PV power plants in the test case *d*. The figure on the right is a zoom of the one on the left.

2.2. Satellite Images

The satellite images used in this paper are extracted from the Helioclim database [34,35]. This database was created using MFG EUMETSAT (European Organization for the Exploitation of Meteorological Satellites) satellite observations. The Helioclim3 version is one of the most efficient versions of the data-base, featuring improved spatial (3 km at nadir) and temporal (15 min) resolutions. The images are treated in nearly real time : there is no analysis time before the reception of the images like there is for NWP forecasts (with 2 h runtime for the fastest NWP models); the small delay is due to internet speed and is in the range of millisecond. The pixels for low solar elevation are interpolated. An example of satellite data providing GHI over an area that covers the power plants of the test case is presented in Figure 2. The figure shows GHI values for two instants in January and July, representing respectively a spatial “screenshot” of the GHI intensity in winter and summer. The lower figure of July presents a higher level of GHI than the upper figure of January. Moreover, in the lower figure, the majority of power plants fall in a region of high GHI with low variability among pixels compared to the upper figure, except for the power plants around 1 degree longitude. More details about the characteristics of the satellite images database can be found in the above-mentioned references. It is noted that for the purpose of this study we obtained the data in the form of data files resulting from the translation of the information in the pixels into numerical information. The pre-processing to obtain the numerical values is done from the service that delivers operationally the satellite images.

Here we consider that the satellite image information employed consists of the time-series that can be generated from a sequence of images. Each pixel location corresponds to a time series. The resulting data are highly correlated. It is thus necessary to select the number of time series, and thus pixels, that provide informative input for the forecasting model. A methodology to achieve this is proposed in Section 3.1. Note that to derive Cloud Motion Vectors, we consider the basic GHI information derived from the images and not from pre-processing [22]. This is set as a requirement for the data-driven approach of the proposed forecasting model. In other words, we expect that the consideration of spatially distributed GHI time series resulting from a series of past images up to the most recent one

is informative about the evolution of the clouds in time, and this can be captured implicitly by the data-driven forecasting model.

Figure 2. GHI from satellite data covering the power plants of data set *d*. The black dots represent the power plants.

3. Proposed Model

We present here the spatio-temporal model proposed to integrate the information from power measurements to satellite images in an incremental way in order to make possible the assessment of the impact of each data source. We have used satellite images in the form of maps that span the geographic area of the data set *d*. The first step in considering this type of data as input is to define the map points which are of interest for forecasting the power output of a specific PV plant and also the appropriate treatment to apply to these points. We then present the statistical forecast model that integrates the satellite data. Finally, we detail the results of the evaluation of the performances of the forecasts resulting from this model and its comparison with models of the state of the art.

3.1. Identifying the Pixels of Interest

Identifying for each PV plant, the points of the satellite image that are of interest in the context of spatio-temporal forecasting has a double objective: the first is to determine

the sub-part of the image, the pixels of which (thus the series of irradiation) are the most related to the production of the site and to quantify this link. It is evident that neighbor pixel carry very similar information that can be redundant and increase the dimensionality of the model. The second objective is to evaluate the interest itself of using satellite images. For each of the $s = 1, \dots, n$ PV installations, the identification of the points of the image that are interesting for the forecast is done in several steps. The first step is to choose the pixels of interest around the site of interest. A correlation analysis between the production and the series of irradiation for some pixels located at 10, 20, ... 100 km have been conducted. The results are presented in Figure 3 as boxplots of the correlation values between each production series and the pixels located from 10 km to 100 km to the power plants. The Figure show that there is no interest going further than 50 km as the correlations values for distance greater to 50 km are lower than 0.5 in mean and those values tend to zero for higher distances. We can see in Figure 4 for three PV power plants, the 50 km area retained for 1 January 2015 at 12:00 UTC. The picture on the upper left size is a bit truncated as the power plants is close to the border of the provided satellite image (which was truncated over the area covering all power plants).

Figure 3. Distribution of correlation values between production and satellite data according to the distance of the pixels.

We chose a fixed block size independent from the forecast horizon, although some methods choose scalable block sizes depending on the horizon, especially for motion detection applications of one-dimensional structures [36]. The second step involves transforming the GHI irradiation series into production series assuming that the relation between the irradiation and the production is an efficiency factor.

Then, we evaluate the link between the measurements of production on the PV site and the data derived from the satellite image. For this, we use a bi-varied criterion of spatial association proposed by Wartenberg [37] which is a transformation of the Moran index.

Let $X_{i,j}(t)$ be the estimate of the output provided by the satellite map at the point (i, j) for the moment t , $Y_s(t)$ the measure of production on the site s at the moment t and τ a time delay. The coefficient of spatial association of Wartenberg is written:

$$I_{(i,j),s}(\tau) = \frac{\sum_t (X_{i,j}(t) - \bar{X})(Y_s(t - \tau) - \bar{Y}_s)}{\sqrt{\sum_t (X_{i,j}(t) - \bar{X})^2} \sqrt{\sum_t (Y_s(t - \tau) - \bar{Y}_s)^2}}. \quad (1)$$

The coefficient of Wartenberg allows the estimation of the links both spatial and temporal. For a zero time offset of the measurement series ($\tau = 0$), the association coefficient makes it possible to evaluate the correlation between the grid points retained and the measurement. Figure 5 presents the correlation values obtained between the production and the estimates for the pixels of the satellite image for three PV plants. We note that for each grid point, we have a time series of GHI estimation and that the correlations were

calculated between the GHI and the on-site measurement. The most important correlations are observed for the points closest to the power stations with correlation values that remain high over the entire area of interest.

Figure 4. Area of interest around 3 PV power plants of the dataset *d*. The time is 1 January 2015 at 12:00 UTC. The black dots represent the power plants.

Figure 5. Correlation between on-site measurements and times series resulting from the satellite images for three PV plants. Correlations are calculated for the month of January 2015; the series are at the time step of 15 min. PV plants are represented by black dots.

The calculation of the association coefficient with non-zero time delay values ($\tau > 0$) makes it possible to evaluate the interest of using the satellite images for the horizons envisaged. In our case, the time delays considered are related to the forecast horizons envisaged, that is to say 6 h. We applied time offsets from 1 h to 6 h to PV production series. The association coefficients between these series and the production estimates for the pixels of the satellite image are calculated. They make it possible to determine the areas of interest of the satellite images for the forecasts for horizons corresponding to the offset applied. Figure 6 shows for a power plant in the West of the region covered by the data set, the values of the association coefficient for different time offsets. Note that for small time offsets (or horizons), the area of interest that corresponds to the highest values of the association coefficient remains close to the center of interest. This zone moves away progressively as time offset values increase. This translation can be explained by the advection of clouds. In addition, the association coefficient values decrease with the time offset and the area of interest shifts to the northeast as the offset increases. As mentioned in Section 3.1, the area of interest in the short-term forecasting frame is the 50 km around the power plant. This area represents the pixels which provide information to help improving the forecasts. The area is not yet associated to a specific pixels selection. It is based on the coefficient of spatial association; it contains all the pixels for which the coefficient value is significant. It is also a first step in order to reduce the dimensionality of the problem. It is noted that the initial image has more than 2000 pixels while by going down to a part of the image based on 50 km radius we limit to around 400 pixels. A further selection among the pixels will be done later in Section 3.2.

3.2. The Forecasting Model

The deterministic spatio-temporal model with Lasso variable selection proposed in [33] was used as basis and extended to integrate satellite image data. Let's recall that this model is defined by

$$\begin{aligned} P_{t+h|t}^x &= \beta_h^0 + \sum_{l=0}^{Ls} \sum_{y \in \mathcal{X}} \beta_h^{l,y} P_{t-l}^y \\ \text{s.c. } \operatorname{argmin}_{\beta, \gamma} &\left\{ \frac{1}{2} \text{RSS}(\beta, \gamma) + \lambda \|\beta\|_1 \right\} \end{aligned} \quad (2)$$

where $\mathcal{X}$ represents the set of all the neighboring plants.

The method we propose for integrating satellite image data into this model is to add this information as exogenous variables in the model. In order to do so it is necessary to select which pixels are the most informative to integrate into the model for the PV production forecast. Indeed, Figure 4 represents the pixels of interest around some central dataset. The 50 km zone defined around these plants represents a significant number of pixels that could pose a problem of dimension for the model. We therefore propose to further select the most pertinent pixels by applying the Lasso's variable selection approach. This choice makes it possible to avoid loss of information that could occur in the case of arbitrary choice of pixels and helps reducing the dimension of the problem. The final model obtained is as follows:

$$\begin{aligned} P_{t+h|t}^x &= \beta_h^0 + \sum_{l=0}^{Ls} \sum_{y \in \mathcal{X}} \beta_h^{l,y} P_{t-l}^y + \sum_{k=1}^p \sum_{l=0}^{Ls'} \gamma_l P_{sat,t-l}^k \\ \text{s.c. } \operatorname{argmin}_{\beta, \gamma} &\left\{ \frac{1}{2} \text{RSS}(\beta, \gamma) + \lambda_1 \|\beta\|_1 + \lambda_2 \|\gamma\|_1 \right\}. \end{aligned} \quad (3)$$

with $P_{sat,t}$ the satellite data, Ls' the maximal lag applied to the pixels. The penalties λ_1 and λ_2 are respectively associated to production data and data from satellite images.

Figure 6. Values of the association coefficient between on-site measurement with time offsets and satellite image estimates for a power plant in the West of the covered region. The PV plant is represented by the black dot, τ represents the time offset.

3.3. Comparison of the Models

With the previously defined spatio-temporal model that integrates exogenous variables from satellite images, we propose here an incremental evaluation approach that aims to quantify the contribution of each source of data in terms of forecast performance.

1. The reference model is the autoregressive model AR which is a model exploiting only the temporal dependencies of the production data only from the site of interest :

$$P_t = f(P_{t-1}, \dots, P_{t-k})$$

- The first model we evaluate is the spatio-temporal model which exploits both temporal dependencies in the measurements but also spatial correlation between measurements of different power plants :

$$ST = f(\text{Production data for the site of interest,} \\ \text{Production data of neighboring sites,} \\ \text{lags of all these production data})$$

In this model, the Lasso variable selection procedure proposed in [33,38] is integrated, thus ensuring the processing of problems of parsimony and dimension.

- The second model investigated is an enhancement of the “ST” model with the integration of local meteorological data. This new model is called a spatio-temporal model with conditioning ST(Z) as the parameters are estimated according to the value of the local meteorological variable Z used.
- The last models investigated are the spatio-temporal model which exploits satellite images, NWP forecasts or both.

$$\begin{aligned} ST + SAT &= ST + \sum(\text{Satellites images data}) \\ ST + NWP &= ST + \sum(\text{NWP data}) \\ ST_{global} &= ST + SAT + NWP \end{aligned}$$

A visual synthesis of all these models is presented in Figure 7.

Figure 7. The different models implemented for the comparison of data source.

4. Evaluation of the Forecasts

4.1. Variable Selection and Reduction of Dimension

The optimal AR model has been obtained by using the production data of the site of interest and try different lags configurations. The optimal lag has been obtained by minimization of the AIC criteria. For most of the power plants of the data set, the optimal

maximum lag is around 1 h (4 time steps). The only variables in this AR(4) model are then the respective lags of the production.

The area of interest of the satellite image around each PV plant is 50 km (see Section 3.1). This area contains approximately 400 pixels. All these 400 pixels are initially integrated into the spatio-temporal forecasting model. The initial number of input variables in the spatio-temporal model with satellite images for a given plant is therefore 2015; which corresponds to the pixels with their respective delays ($400 * 3$ (3 h)) and to the production series of neighboring plants with their respective delays ($136 * 6$ (number of lags) $- 1$).

Table 1 shows for a power plant the number of variables selected according to the horizon. The numbers of pixels and different PV units (without the delayed series) selected are also shown in the table. The small number of variables selected shows that the variable selection procedure is effective for reducing the size of the problem. In addition, we note that the variables selected are mainly variables related to the production of neighboring sites, followed by the pixels of satellite images. The number of selected pixels increases slowly with the forecast horizon. There are 4 selected pixels for 15 min when there are 7 for 3 h. One may expect a more important increase but it should be noted that adjacent neighboring sites concentrate most of the spatio-temporal information and only pixels which provided more information are selected to keep the dimensionality of the model low.

Table 1. Number of pixels and PV plants selected by horizon for a power plant of interest.

Horizons	Initial Number of Variables = 1351		
	Number of Pixels Selected	Number of PV Plants Selected	Total Number of Variables Selected (Lags and Pixels Included)
15 min	4	28	67
1 h	4	22	64
3 h	7	25	56

4.2. Forecasting Performances

The performances of the forecasting models previously presented are then evaluated using state-of-the-art criteria like RMSE, MAE [2] normalized by the maximum power observed for the power plant.

The Tables 2 and 3 present respectively for a selected power plant (P10) the MAE and RMSE for the each of the five models for different horizon.

Table 2. MAE of the models for one power plant (P10) of interest.

Metric	Models				
MAE (% Pmax)	AR	ST	ST(Z)	ST + SAT	ST + NWP
15 min	2.75	2.69	2.62	2.53	2.69
1 h	5.69	5.33	5.23	4.82	5.59
3 h	11.81	10.21	10.21	8.66	11.61
6 h	15.15	12.59	12.59	10.14	14.76

Table 3. RMSE of the models for a power plant of interest.

Metric	Models				
RMSE (% Pmax)	AR	ST	ST(Z)	ST + SAT	ST + NWP
15 min	4.32	3.12	3.00	2.90	3.34
1 h	8.34	6.72	6.5	6.30	6.90
3 h	10.46	8.84	8.84	8.41	9.12
6 h	16.67	9.68	9.68	9.31	10.53

The analysis of these tables show :

- for 15 min horizon, all the models show similar performances;
- for longer horizons the ST model outperforms the AR model;
- the integration of the local meteorological information reduce the MAE compared to when this info is not used;
- the model resulting from the combination of spatio-temporal and satellite data is the best model;
- the use of satellite data in combination with ST measured data results to more efficient forecasts for the short-term forecasting than the combination of ST and NWP;
- the level of the observed errors is similar to the lowest observed in the literature.

These conclusions can be extended to the entire test case according to the Table A1 in Appendix A presenting the evaluation results (RMSE, MAE) for 9 other power plants for 6-h horizon. An additional evaluation process of the performance of each model compared to the reference AR model over all the power plants (in Table A1) is presented in Figure 8. The figure is produced as follows :

- M_0 is the reference AR model
- For each power plants $P_i, i = 1, \dots, 9$ in Table A1 and for each model $M_i, i = 1, \dots, 4$ between the one presented in Section 3.2 (ST, ST(Z), ST+SAT and ST+NWP)
 - For each hour $h, h = 1, \dots, 6$ of the forecasting horizon, compute on the testing set the RMSE improvement of model M_i over reference model M_0 :

$$Improvement(P_i, M_i, M_0, h) = 100 * \frac{RMSE(P_i, M_i, h) - RMSE(P_i, M_0, h)}{RMSE(P_i, M_0, h)}$$

- Each line on Figure 8 represents the average improvement at each horizon over all the 9 power plants of a model M_i (over M_0).

The figure show that the spatio-temporal model allows an average improvement of RMSE of 10% for 3 h. This improvement can reach 20% depending on the plants. This result is consistent with those presented in [38]. Using local wind speed measurements with weather stations near power stations improves forecasting performance by an average of 2% for the first two hours of forecasting. Beyond 3 h, these measurements do not contribute to further improving the prediction performance compared to the basic spatio-temporal model. The spatio-temporal model that integrates the NWP forecasts shows no significant improvement over the initial spatio-temporal model over the 6 h of forecast. It should be noted, however, a slight improvement in the performance of this model for horizon values from 5 h. Integration of satellite images further reduces forecast errors. Indeed, we see in the figure an improvement of the RMSE of the order of 3% on average of the model with integrated satellite images compared to the simple spatio-temporal model. The hierarchy in term of global performances of these models is the ST + STAT model first, followed by the ST(Z) model, the ST model, the ST + NWP model and the AR model.

Figure 8. Comparison of the average (over all power plants) forecasting performances of the models. The time step is 15 min.

5. Conclusions

In this paper, we have proposed a spatio-temporal model which exploits not only the spatio-temporal information of the production measurements of the neighboring sites, but also the satellite images and NWP predictions. Since the latter are characterized by finer resolutions and faster update rates than NWP forecasts, they are a very interesting source of data for short-term PV prediction. We have presented a pixels selection procedure around the plants for which the PV production forecast is being considered. This procedure makes it possible to go from images covering all the power stations considered to a finer image that focuses around the power plant. The relationship between the points of interest of the area around the power plants and the production of the site showed that the pixels closest to the images are the most correlated to the production. We have quantified the contribution of each of the different sources of information namely satellite images, measurements of neighboring power plants, NWP forecasts and local meteorological measurements on forecast performances in comparison with an exclusively temporal reference model. The forecast horizons envisaged are 6 h. The biggest source of improvement comes from the use of power plant measurements. Satellite images can further reduce forecast errors when they are associated with spatio-temporal patterns. The effect of the NWP forecasts is very small on the early horizons in opposition to that of the local meteorological measurements. The NWP predictions, however, correct the poor performance of the spatio-temporal model for horizons greater than 12 h, thus confirming the importance of meteorology for these forecast horizons. It is important to mention that the results obtained in this paper show that the use of geographically distributed data motivates data sharing (as open data or monetised through data markets) as a good practice for the future.

Author Contributions: Conceptualization, X.G.A., R.G. and G.K.; methodology, X.G.A., R.G. and G.K.; software, X.G.A.; validation, X.G.A., R.G. and G.K.; formal analysis, X.G.A.; investigation, X.G.A.; resources, X.G.A.; data curation, X.G.A.; writing—original draft preparation, G.A.; writing—review and editing, X.G.A., R.G. and G.K.; visualization, X.G.A.; supervision, R.G. and G.K.; project administration, R.G.; funding acquisition, X.G.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work was carried out within the research project entitled “Improvement of PV power forecasting and predictive management including storage solutions”, funded by the company Coruscant SA in the frame of its participation to a tender of the French Energy Regulator CRE for the development of PV plants above 250 kWc.

Data Availability Statement: ECMWF HRES data available at <https://www.ecmwf.int/en/forecasts/datasets/set-i> accessed on April 2017.

Acknowledgments: The authors would like to thank the French industrial Hespul for providing the PV data in the frame of the PhD thesis of the first author, as well as the European Center for Medium-Range Weather Forecasts for providing the NWP data. We would like also to thank the partners of the European project Smart4RES (European Union’s Horizon 2020, No. 864337) for their useful comments that helped improving the paper.

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

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
TLA	Three letter acronym
LD	Linear dichroism

Appendix A. Detailed Evaluation Results

Table A1. Evaluation results for 9 power plants for the horizon of 6 h ahead (in % of Nominal Power).

Power Plant	Criterion	AR	ST	ST(Z)	ST_SAT	ST_NWP
P1	RMSE	16.67	9.68	9.68	9.31	10.53
	MAE	15.15	12.59	12.59	10.14	14.76
P2	RMSE	17.02	9.53	9.55	9.32	10.12
	MAE	14.73	12.56	12.56	11.23	13.48
P3	RMSE	16.82	9.72	9.72	9.22	10.01
	MAE	14.33	12.84	12.84	10.89	12.33
P4	RMSE	18.21	10.02	10.02	9.88	10.35
	MAE	16.13	12.94	12.94	12.72	13.08
P5	RMSE	16.34	9.88	9.88	9.33	10.54
	MAE	14.92	12.33	12.33	10.14	14.83
P6	RMSE	17.12	9.66	9.66	9.29	10.21
	MAE	17.23	12.10	12.10	10.48	14.55
P7	RMSE	15.88	9.55	9.55	9.22	10.12
	MAE	14.67	12.43	12.43	10.08	13.97
P8	RMSE	16.42	9.77	9.77	9.4	10.02
	MAE	14.87	12.65	12.65	10.42	14.88
P9	RMSE	17.01	9.82	9.82	9.62	10.33
	MAE	15.64	12.71	12.71	11.01	14.66

References

- Kariniotakis, G. *Renewable Energy Forecasting: From Models to Applications*; Woodhead Publishing: Cambridge, UK, 2017.
- 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. [\[CrossRef\]](#)
- Sobri, S.; Koochi-Kamali, S.; Rahim, N.A. Solar photovoltaic generation forecasting methods: A review. *Energy Convers. Manag.* **2018**, *156*, 459–497. [\[CrossRef\]](#)
- Sperati, S.; Alessandrini, S.; Delle Monache, L. An application of the ECMWF Ensemble Prediction System for short-term solar power forecasting. *Sol. Energy* **2016**, *133*, 437–450. [\[CrossRef\]](#)
- Pierro, M.; Bucci, F.; De Felice, M.; Maggioni, E.; Moser, D.; Perotto, A.; Spada, F.; Cornaro, C. Multi-Model Ensemble for day ahead prediction of photovoltaic power generation. *Sol. Energy* **2016**, *134*, 132–146. [\[CrossRef\]](#)
- Liu, Y.; Shimada, S.; Yoshino, J.; Kobayashi, T.; Miwa, Y.; Furuta, K. Ensemble forecasting of solar irradiance by applying a mesoscale meteorological model. *Sol. Energy* **2016**, *136*, 597–605. [\[CrossRef\]](#)
- Alessandrini, S.; Delle Monache, L.; Sperati, S.; Cervone, G. An analog ensemble for short-term probabilistic solar power forecast. *Appl. Energy* **2015**, *157*, 95–110. [\[CrossRef\]](#)
- Zamo, M.; Mestre, O.; Arbogast, P.; Pannekoucke, O. A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production. Part II: Probabilistic forecast of daily production. *Sol. Energy* **2014**, *105*, 804–816. [\[CrossRef\]](#)
- Nagy, G.I.; Barta, G.; Kazi, S.; Borbely, G.; Simon, G. GEFCom2014: Probabilistic solar and wind power forecasting using a generalized additive tree ensemble approach. *Int. J. Forecast.* **2016**, *32*, 1087–1093. [\[CrossRef\]](#)
- Huang, J.; Perry, M. A semi-empirical approach using gradient boosting and -nearest neighbors regression for GEFCom2014 probabilistic solar power forecasting. *Int. J. Forecast.* **2016**, *32*, 1081–1086. [\[CrossRef\]](#)
- Chu, Y.; Coimbra, C.F.M. Short-term probabilistic forecasts for Direct Normal Irradiance. *Renew. Energy* **2017**, *101*, 526–536. [\[CrossRef\]](#)
- Vagropoulos, S.I.; Kardakos, E.G.; Simoglou, C.K.; Bakirtzis, A.G.; Catalão, J.P.S. ANN-based scenario generation methodology for stochastic variables of electric power systems. *Electr. Power Syst. Res.* **2016**, *134*, 9–18. [\[CrossRef\]](#)
- Dolara, A.; Grimaccia, F.; Leva, S.; Mussetta, M.; Ogliari, E. A Physical Hybrid Artificial Neural Network for Short Term Forecasting of PV Plant Power Output. *Energies* **2015**, *8*, 1138. [\[CrossRef\]](#)
- Khodayar, M.; Kaynak, O.; Khodayar, M.E. Rough Deep Neural Architecture for Short-Term Wind Speed Forecasting. *IEEE Trans. Ind. Inform.* **2017**, *13*, 2770–2779. [\[CrossRef\]](#)
- Noia, M.; Ratto, C.F.; Festa, R. Solar irradiance estimation from geostationary satellite data: I. Statistical models. *Sol. Energy* **1993**, *51*, 449–456. [\[CrossRef\]](#)
- Noia, M.; Ratto, C.F.; Festa, R. Solar irradiance estimation from geostationary satellite data: II. Physical models. *Sol. Energy* **1993**, *51*, 457–465. [\[CrossRef\]](#)
- Ineichen, P. Comparison of eight clear sky broadband models against 16 independent data banks. *Sol. Energy* **2006**, *80*, 468–478. [\[CrossRef\]](#)
- Perez, R.; Moore, K.; Wilcox, S.; Renné, D.; Zelenka, A. Forecasting solar radiation—Preliminary evaluation of an approach based upon the national forecast database. *Sol. Energy* **2007**, *81*, 809–812. [\[CrossRef\]](#)
- Polo, J.; Wilbert, S.; Ruiz-Arias, J.A.; Meyer, R.; Gueymard, C.; Sári, M.; Martín, L.; Mieslinger, T.; Blanc, P.; Grant, I.; et al. Preliminary survey on site-adaptation techniques for satellite-derived and reanalysis solar radiation datasets. *Sol. Energy* **2016**, *132*, 25–37. [\[CrossRef\]](#)
- Blanc, P.; Remund, J.; Vallance, L. Short-term solar power forecasting based on satellite images. In *Renewable Energy Forecasting*; Kariniotakis, G., Ed.; Woodhead Publishing Series in Energy; Woodhead Publishing: Cambridge, UK, 2017; pp. 179–198. [\[CrossRef\]](#)
- Peng, Z.; Yu, D.; Huang, D.; Heiser, J.; Kalb, P. A hybrid approach to estimate the complex motions of clouds in sky images. *Sol. Energy* **2016**, *138*, 10–25. [\[CrossRef\]](#)
- Dong, Z.; Yang, D.; Reindl, T.; Walsh, W.M. Satellite image analysis and a hybrid ESSS/ANN model to forecast solar irradiance in the tropics. *Energy Convers. Manag.* **2014**, *79*, 66–73. [\[CrossRef\]](#)
- Bosch, J.L.; Kleissl, J. Cloud motion vectors from a network of ground sensors in a solar power plant. *Sol. Energy* **2013**, *95*, 13–20. [\[CrossRef\]](#)
- Rosiek, S.; Alonso-Montesinos, J.; Batlles, F. Online 3-h forecasting of the power output from a BIPV system using satellite observations and ANN. *Electr. Power Energy Syst.* **2018**, *99*, 261–272. [\[CrossRef\]](#)
- Linares-Rodriguez, A.; Quesada-Ruiz, S.; Pozo-Vazquez, D.; Tovar-Pescador, J. An evolutionary artificial neural network ensemble model for estimating hourly direct normal irradiances from meteosat imagery. *Energy* **2015**, *91*, 264–273. [\[CrossRef\]](#)
- Jang, H.S.; Bae, K.Y.; Park, H.S.; Sung, D.K. Solar Power Prediction Based on Satellite Images and Support Vector Machine. *IEEE Trans. Sustain. Energy* **2016**, *7*, 1255–1263. [\[CrossRef\]](#)
- Wolff, B.; Kühnert, J.; Lorenz, E.; Kramer, O.; Heinemann, D. Comparing support vector regression for PV power forecasting to a physical modeling approach using measurement, numerical weather prediction, and cloud motion data. *Sol. Energy* **2016**, *135*, 197–208. [\[CrossRef\]](#)
- Mazorra Aguiar, L.; Pereira, B.; David, M.; Díaz, F.; Lauret, P. Use of satellite data to improve solar radiation forecasting with Bayesian Artificial Neural Networks. *Sol. Energy* **2015**, *122*, 1309–1324. [\[CrossRef\]](#)

-
29. Messner, J.W.; Pinson, P. Online adaptive lasso estimation in vector autoregressive models for high dimensional wind power forecasting. *Int. J. Forecast.* **2018**, *35*, 1485–1498.
 30. Dambreville, R.; Blanc, P.; Chanussot, J.; Boldo, D. Very short term forecasting of the Global Horizontal Irradiance using a spatio-temporal autoregressive model. *Renew. Energy* **2014**, *72*, 291–300. [[CrossRef](#)]
 31. Koster, D.; Minette, F.; Braun, C.; O’Nagy, O. Short-term and regionalized photovoltaic power forecasting, enhanced by reference systems, on the example of Luxembourg. *Renew. Energy* **2019**, *132*, 455–470. [[CrossRef](#)]
 32. Lorenz, E.; Kühnert, J.; Heinemann, D.; Nielsen, K.P.; Remund, J.; Müller, S.C. Comparison of global horizontal irradiance forecasts based on numerical weather prediction models with different spatio-temporal resolutions. *Prog. Photovolt. Res. Appl.* **2016**, *24*, 1626–1640. [[CrossRef](#)]
 33. Agoua, X.G.; Girard, R.; Kariniotakis, G. Short-Term Spatio-Temporal Forecasting of Photovoltaic Power Production. *IEEE Trans. Sustain. Energy* **2018**, *9*, 538–546. [[CrossRef](#)]
 34. Vernay, C.; Blanc, P.; Pitaval, S. Characterizing measurements campaigns for an innovative calibration approach of the global horizontal irradiation estimated by HelioClim-3. *Renew. Energy* **2013**, *57*, 339–347. [[CrossRef](#)]
 35. Blanc, P.; Gschwind, B.; Lefèvre, M.; Wald, L. The HelioClim Project: Surface Solar Irradiance Data for Climate Applications. *Remote Sens.* **2011**, *3*, 343–361. [[CrossRef](#)]
 36. Dambreville, R. Nowcasting and Very Short Term Forecasting of the Global Horizontal Irradiance at Ground Level: Application to Photovoltaic Output Forecasting. Ph.D. Thesis, Université de Grenoble, Saint-Martin-d’Hères, France, 2014.
 37. Wartenberg, D. Multivariate Spatial Correlation: A Method for Exploratory Geographical Analysis. *Geogr. Anal.* **1985**, *17*, 263–283. [[CrossRef](#)]
 38. Agoua, X.G.; Girard, R.; Kariniotakis, G. Probabilistic Model for Spatio-Temporal Photovoltaic Power Forecasting. *IEEE Trans. Sustain. Energy* **2019**, *10*, 780–789. [[CrossRef](#)]