Available online at www.sciencedirect.com

ScienceDirect

Energy Reports 9 (2023) 495–502

www.elsevier.com/locate/egyr

2022 3rd International Conference on Power, Energy and Electrical Engineering (PEEE 2022)
18–20 November 2022

Short-term probabilistic forecasting models using Beta distributions for photovoltaic plants

L. Alfredo Fernandez-Jimenez[a,*], Claudio Monteiro[b], Ignacio J. Ramirez-Rosado[c]

[a] *Electrical Engineering Department, University of La Rioja, Logroño 26004, Spain*
[b] *FEUP, Faculdade Engenharia Universidade do Porto, Porto 4200-465, Portugal*
[c] *Electrical Engineering Department, University of Zaragoza, Zaragoza 50018, Spain*

Received 13 December 2022; accepted 16 January 2023
Available online xxxx

Abstract

This article presents original probabilistic forecasting models for day-ahead hourly energy generation forecasts for a photovoltaic (PV) plant, based on a semi-parametric approach using three deterministic forecasts. Input information of these new models consists of data of hourly weather forecasts obtained from a Numerical Weather Prediction model and variables related to the sun position for future instants. The proposed models were satisfactorily applied to the case study of a real-life PV plant in Portugal. Probabilistic benchmark models were also applied to the same case study and their forecasting results compared with the ones of the proposed models. The computer results obtained with these proposed models achieve better point and probabilistic forecasting evaluation indexes values than the ones obtained with the benchmark models.

Peer-review under responsibility of the scientific committee of the 3rd International Conference on Power, Energy and Electrical Engineering, PEEE, 2022.

Keywords: Short-term forecasting; PV power generation; Probabilistic forecasting; Photovoltaic

1. Introduction

The need to integrate power generating plants based on renewable sources into electric power system has boosted the installation of photovoltaic (PV) plants. The installed power capacity of PV plants worldwide has exceeded 843 GW by the end of 2021 [1], and a large expansion is foreseen until 2030, reaching a global energy generation between 2.4 and 6.1 thousand TWh, depending on the growth scenario considered [2]. However, energy production from PV plants is affected by sunlight, temperature, cloud cover, etc., which makes its power production highly variable over time. Thus, as the level of power integration from PV plants into the power system grows along the coming years, the need for accurate and reliable short-term forecasting models (STFMs) of their power production will increase. The forecast of power generation in PV plants is necessary for the correct operation of the power system, helping to reduce costs and uncertainties [3].

[*] Corresponding author.
E-mail address: luisalfredo.fernandez@unirioja.es (L.A. Fernandez-Jimenez).

https://doi.org/10.1016/j.egyr.2023.01.059

Peer-review under responsibility of the scientific committee of the 3rd International Conference on Power, Energy and Electrical Engineering, PEEE, 2022.

STFMs (horizons up to 72 h) have had the greatest development in literature, as they enable interesting applications for PV plants managers (preparing bids to electricity markets) and for Transmission or Distribution System Operators (scheduling of power system operations). STFMs for PV plants can be approached by two different perspectives. The first one, more traditional, corresponds to deterministic or point-forecast models. The forecast with these models is focused on the expected mean/average value of the hourly PV energy generation at the desired future instant. Deterministic forecasting models are not able to quantify the uncertainties associated with the predicted value. In contrast, the second type of models are capable of such quantification. These are the probabilistic forecasting models (PFMs), which are gaining interest as they provide predictions for the future with their associated probability, allowing the risk of a decision relying on such predictions to be quantified [4].

PFMs are intended to obtain a probability density function (PDF) of the predicted variable or the range of values where it will be according to a confidence level [5]. In the development of PFMs a statistical distribution for the PDF of the predicted variable can be assumed beforehand. This is the case of the parametric approach; by contrast, the nonparametric approach does not assume any distribution. Parametric PFMs fit the forecasting errors of a deterministic model to a supposed density function. In [6] a parametric PFM for solar irradiance prediction is presented where a normal distribution is assumed to calculate prediction intervals with the idea of transferring them to PV power intervals. Another work [7], describing a parametric PFM for solar irradiance, assumes a normal distribution for the error of the deterministic forecast provided by the combination of two linear models.

Van der Meer et al. [5] present a description of the methods (parametric and nonparametric) used for PFMs of PV plants production, with their most relevant characteristics, and show that the quantile regression has been the most widely used method. The inclusion, as explanatory variables, of forecasts of weather variables obtained with a Numerical Weather Prediction (NWP) model and variables related to the solar position at the future instant can considerably reduce the forecasting error [8]. A review of the latest advances in the development of forecasting models for PV plants can be found in [9,10]. The new forecasting models are becoming more and more complex, focusing the development of probabilistic models on the nonparametric approach. Despite this, we consider that there is still room for models based on the parametric approach that can be competitive (similar or better forecasting results) against those of the nonparametric approach.

This paper presents two PFMs (called PFMB1 and PFMB2 models) for the hourly energy generation of a PV plant for the next day. These PFMB models are semi-parametric models, in the sense that the Beta distribution is used for the PDF of the hourly PV energy generation variable, although the parameters of such Beta distribution are obtained by two deterministic forecasting models. The first model, PFMB1, only uses explanatory variables related to the sun position in the forecasting horizon, while the second model, PFMB2, uses also forecasts of weather variables for that future instant. Their forecasting results, for a real PV plant, are compared with those obtained with two benchmark probabilistic models, being superior those obtained by models PFMB1 and PFMB2. The best of the two proposed models, the PFMB2 model, by providing as forecasting result the probability density function (PDF) of the hourly power generated in the PV plant at the future instant, can be useful both for plant managers, who can evaluate his/her risks when offering the PV energy produced in the electricity market, and for distribution system operators.

This article is structured as follows: Section 2 presents the proposed PFMB models, with the methodology used to determine the parameters of the selected PDF distribution (Beta distribution). Section 3 presents probabilistic benchmark models and the evaluation indexes used for comparisons of PV power forecasting results. Section 4 presents a case study and the application of the proposed and benchmark models to a real-life PV plant with the comparisons of their results. Section 5 presents the conclusions of this article.

2. The PFMB models

The PFMB models seek to obtain the PDF for the PV energy generated in each of the hours of the following day. A Beta distribution has been chosen for its flexibility in modelling the variable to be predicted. The PDF of a Beta distribution for the hour h is given by (1), where $\alpha\,(h)$ represents the value of the first parameter for hour h ($\alpha\,(h) > 0$), $\beta\,(h)$ the value of the second parameter ($\beta\,(h) > 0$), and $\boldsymbol{B}(\alpha(h), \beta(h))$, represents the Beta function for hour h, defined by (2).

$$f\,(x; \alpha\,(h)\,, \beta(h)) = \frac{x^{\alpha(h)-1}\,(1-x)^{\alpha(h)-1}}{\boldsymbol{B}(\alpha(h), \beta(h))} \qquad 0 < x < 1 \tag{1}$$

$$\boldsymbol{B}(\alpha(h), \beta(h)) = \int_0^1 z^{\alpha(h)-1} (1-z)^{\beta(h)-1}\, dz \tag{2}$$

The methodology is structured in two stages: a first stage of normalization of the variable to be predicted and a second stage of calculation of the parameters of the Beta distribution for the prediction horizon. Three point forecasting models are used, one in the first stage and two in the second stage, as it is described in the next paragraphs.

Since the Beta distribution is only defined in the range [0,1], it is necessary to normalize the hourly energy produced in the PV plant by (3),

$$E_n(h) = \frac{E(h)}{E_{MAX}(h)} \tag{3}$$

where $E_n(t)$ represents the normalized hourly PV energy generation, in per unit, at hour h; $E(h)$ represents the hourly PV energy generation, in MWh, at hour h; and $E_{MAX}(h)$ represents the maximum value of hourly PV energy generation, in MWh, at hour h, which corresponds to the value that would be generated in the PV plant under a clear-sky condition. In this work we have used an ensemble of artificial neural networks to calculate the $E_{MAX}(h)$ value.

In the second stage, two point forecasting models are used to provide the value of the point forecast of the normalized hourly PV energy generation, $\hat{E}_n(h)$, and the point forecast value of the square of the normalized hourly PV energy generation, $\widehat{E_n^2}(h)$, both for hour h. Both models were also implemented with ensembles of neural networks. The values of the two parameters of the Beta distribution for hour h are obtained by the matching moments method [11], as expressed in (4), where both parameters are calculated from the expected value of the normalized hourly PV energy generated for hour h and the expected value of the square of the normalized hourly PV energy for hour h. These expected values are the forecasted ones for hour h obtained from the previous point forecasting models. Notice that the denominators of (4) correspond to the variance of the hourly PV energy generation normalized variable.

$$\alpha(h) = \frac{\hat{E}_n(h) \times \left(\hat{E}_n(h) - \widehat{E_n^2}(h)\right)}{\widehat{E_n^2}(h) - \left(\hat{E}_n(h)\right)^2} \qquad \beta(h) = \frac{\left(1 - \hat{E}_n(h)\right) \times \left(\hat{E}_n(h) - \widehat{E_n^2}(h)\right)}{\widehat{E_n^2}(h) - \left(\hat{E}_n(h)\right)^2} \tag{4}$$

The values of both parameters must be greater than zero, which is a condition imposed by the Beta distribution. Therefore, (4) must meet some conditions: the denominator and the numerators must be positive, what leads to the constraints given on the left side of (5). These restrictions are fulfilled by applying the corrections expressed on the right-hand side of (5),

$$\left\{\begin{array}{c} \hat{E}_n(h) > 0 \\ \hat{E}_n(h) < 1 \\ \widehat{E_n^2}(h) < \hat{E}_n(h) \\ \widehat{E_n^2}(h) > \left(\hat{E}_n(h)\right)^2 \end{array}\right. \Rightarrow \left\{\begin{array}{ll} if\ \left(\hat{E}_n(h)\right) \leq 0 & then\ \tilde{\hat{E}}_n(h) = \epsilon \\ if\ \left(\hat{E}_n(h)\right) \geq 1 & then\ \tilde{\hat{E}}_n(h) = (1-\epsilon) \\ if\ \left(\hat{E}_n(h)\right) \geq 1 & then\ \widetilde{\widehat{E_n^2}}(h) = (1-\epsilon) \times \left(\tilde{\hat{E}}_n(h)\right) \\ if\ \left(\widehat{E_n^2}(h)\right) \leq \left(\hat{E}_n(h)\right)^2 & then\ \widetilde{\widehat{E_n^2}}(h) = (1+\epsilon) \times \left(\tilde{\hat{E}}_n(h)\right)^2 \end{array}\right. \tag{5}$$

where $\tilde{\hat{E}}_n(h)$ represents the corrected point forecast value of the normalized PV energy generation for hour h; $\widetilde{\widehat{E_n^2}}(h)$ represents the corrected point forecast value of square of the normalized hourly PV energy generation for hour h; and ε is a small marginal value. Thus, the final corrected parameters, $\alpha_c(h)$ and $\beta_c(h)$, of the Beta distribution are calculated by substituting in (4) the values of $\hat{E}_n(t)$ and $\widehat{E_n^2}(h)$ by their corrected values, when the correction is necessary.

The expected value of the hourly PV energy generation for the hour h, i.e., its point forecast with values expressed in MWh, in the interval [0; $E_{MAX}(h)$], can be obtained by using the mean value of $f(x; \alpha_c(h), \beta_c(h))$ multiplied by the corresponding maximum possible value for the hour h, that is, $E_{MAX}(h)$.

3. Evaluation indexes and benchmark probabilistic forecasting models

In order to check the goodness of the forecasts obtained by the PFMB1 and PFMB2 models, we defined three indexes for models' performance evaluations and two benchmark probabilistic forecasting models (linear quantile regression models, LQRM1 and LQRM2).

The first two indexes were used to evaluate deterministic forecasts, and the third one to evaluate probabilistic forecasts. The first index was the Mean Absolute Error (*MAE*) and the second one was the Root Mean Square Error (*RMSE*) which are defined by (6) and (7), respectively,

$$MAE = \frac{1}{N}\sum_{h=h_1}^{h_2}\left|\hat{E}\left(h\right) - E(h)\right| \tag{6}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{h=h_1}^{h_2}\left(\hat{E}\left(h\right) - E(h)\right)^2} \tag{7}$$

where $\hat{E}\left(h\right)$ is the point forecast for the hourly PV energy, in MWh, generated in the PV plant for hour h; $E(h)$ is the real hourly PV energy generation, in MWh, for hour h; h_1 and h_2 are the first and the last, respectively, sunlight hours of the evaluation time period; and N is the total number of sunlight hours in such period.

The third evaluation index was the Continuous Ranked Probability Score (*CRPS*), which has widely used in the literature to evaluate probabilistic forecasts. A probabilistic forecast is characterized by its reliability or calibration, sharpness and resolution [12]. *CRPS* evaluates jointly the three characteristics [13]. The instantaneous *CRPS*, that is, the value corresponding to a forecast that is subsequently materialized into an actual value, is defined by (8), assuming that the probabilistic forecast is expressed by the cumulative distribution function CF; $\mathbb{1}$ corresponds to the indicator function (Heaviside step function); and y is the actual value. *CRPS* is negatively oriented (that is, a lower value of *CRPS* indicates a better forecasting performance) and its value is expressed in the same unit as the forecasted variable.

$$CRPS\left(CF, y\right) = \int_{\mathbb{R}}\left[CF\left(x\right) - \mathbb{1}\left(x \geq y\right)\right]^2 dx \tag{8}$$

The quantile regression consists of estimating the parameters of the functions that link the quantiles of the dependent variable with their explanatory variables. If the functions are linear then the model is known as a Linear Quantile Regression (LQR) model. In [14] the authors propose a method to fit the linear quantile regression model, so that the model estimates the θth quantile, $0 < \theta < 1$, of the forecasted variable on the basis of a linear relationship, as expressed in (9), where $\hat{E}^{\theta}\left(t\right)_{LQR}$ is the θth conditional quantile of forecasted variable (hourly PV energy generation) for the hour t; β_θ is the vector of coefficients; $\boldsymbol{x}\left(h\right)$ is the vector of explanatory variables for such moment; and δ is the error term.

$$\hat{E}^{\theta}\left(h\right)_{LQR} = \beta_\theta \boldsymbol{x}\left(h\right) + \delta \tag{9}$$

The vector of coefficients β_θ can be estimated by minimizing the loss function $\rho_\theta\left(z\right)$ (known as the pinball loss function) defined by (10) so that the expected values of the vector of coefficients for the θth quantile, $\hat{\beta}_\theta$, are obtained by (11), where i represents any sunlight hour in the fitting period and N_h is the number of data (hours) used to fit the model. Then, the expected value of the θth conditional quantile of forecasted variable (hourly PV energy generation) for the hour h is obtained by (12).

$$\rho_\theta\left(z\right) = \begin{cases} \theta z & if\ z \geq 0 \\ \left(1-\theta\right)z & if\ z < 0 \end{cases} \tag{10}$$

$$\hat{\beta}_\theta = \underset{\beta}{\arg\min}\left[\sum_{i=1}^{N_h}\rho_\theta\left(E\left(i\right) - \beta\boldsymbol{x}\left(i\right)\right)\right] \tag{11}$$

$$\hat{\beta}_\theta \boldsymbol{x}\left(h\right) \tag{12}$$

Notice that any conditional quantile has its own vector of coefficients, $\hat{\beta}_\theta$. If the vectors of coefficients of different quantiles are computed separately, crossing quantiles can appear, i.e., $\hat{E}^{\theta_1}\left(h\right)_{LQR} > \hat{E}^{\theta_2}\left(h\right)_{LQR}$ when $\theta_1 < \theta_2$. In order to avoid crossing quantiles, the rearrangement method described in [15] was used.

4. Case study and computer results

The PFMB models and the benchmark models were used to forecast the hourly energy production of a PV plant placed in Portugal. The forecast is carry out at 9:00 each day, obtaining the predictions for the following day's sunlight hours. The PV plant has a rated power capacity of 10 MW and it is composed of single-axis tracking PV modules. The time series of hourly electric energy production of that PV plant for 41 months (from January 2017 to May 2020) was available. In addition to that time series, the energy production data were completed with weather forecasts data and data related to the solar position for each of the registers of the hourly energy production of the PV plant.

The weather forecasts data were obtained from Meteogalicia, the Galician weather service [16], which provide daily weather forecasts with high resolution for the region of Galicia (north-west of the Iberian Peninsula) and with medium resolution for the entire peninsula. The forecasts provided by Meteogalicia for the region where the PV plant is located include the hourly values of the main weather variables on the earth's surface for the next 72 h, with a spatial resolution of about 12 km, that is, it provides the forecasts of these variables for a set of geographical points that form a grid with a spatial resolution of 12 km.

The forecasts for the four grid points corresponding to the geographical positions closest to the PV plant were downloaded from the Meteogalicia server for the whole study period. Subsequently, the forecasted value of each weather variable for the PV plant position was calculated as the weighted average of the values for these four closest points. The weighting factor used was the square of the Euclidean distance between the PV plant position and the position corresponding to each of these four points. Initially the number of weather variables was very large, but it was reduced to 9 after a correlation study between each weather variable and the hourly energy produced in the PV plant. The variables thus selected are those listed in Table 1 within the group of "Forecasted weather variables".

Table 1. Selected explanatory input variables for the probabilistic forecasting models.

Group	Denomination	Name of variable	Meaning
Solar variables	V1	*Declination*	Solar declination (°)
	V2	H_0	Extra-terrestrial solar irradiance (W/m^2)
	V3	*Altitude*	Solar altitude (°)
	V4	*Azimuth*	Solar azimuth (°)
Forecasted weather variables	V5	*temp*	Temperature at 2 m (K)
	V6	*swflux*	Surface downwelling shortwave flux (W/m^2)
	V7	*mslp*	Mean sea level pressure
	V8	*mod*	Wind module at 10 m (m/s)
	V9	*rh*	Relative humidity at 2 m (0 to 1)
	V10	*cft*	Cloud cover at low and mid levels (0 to 1)
	V11	*cfl*	Cloud cover at low levels (0 to 1)
	V12	*cfm*	Cloud cover at mid levels (0 to 1)
	V13	*cfh*	Cloud cover at high levels (0 to 1)

The complete set of explanatory information includes four additional variables related to the solar position for each one of the hours in the period under study, which are included in the group "Solar variables" of Table 1, and they correspond to the angle of solar declination, the extra-terrestrial solar irradiance, the angle of solar altitude, and the angle of solar azimuth.

In order to collect only the information relevant to PV energy generation, only data corresponding to hours of sunlight were included in the database. The complete dataset was divided in training or fitting dataset and testing dataset. The training dataset comprised the period from January 2017 to May 2019 (29 months), and the testing dataset comprised the period from June 2019 to May 2020 (12 months).

The selection of the distribution used in the models was carried out with a statistical analysis of the variable to be predicted. The normalized hourly PV energy generation of the training dataset was tested with the Skewness–Kurtosis plot proposed by Cullen and Frey [17], which suggested that the Beta distribution was an appropriate choice.

Two models, PFMB1 and PFMB2, were developed with the proposed methodology. The PFMB1 model used as explanatory variables only those from the "Solar variables" group of Table 1. The PFMB2 model used the two sets of explanatory variables, that is, the 13 variables in Table 1. The comparison of the results of both models allows us to determine the predictive value of the variables corresponding to the group of "Forecasted weather variables".

In the first stage, a point forecasting model was created to predict the hourly energy that would be generated in each hour in the case of a clear-sky day condition. This model was developed in the following steps:

1. The records corresponding to clear-sky days were selected from the database. This selection was made by taking those records in which the value of the variable V11 was null (absence of clouds at low and medium levels). This selection was then filtered by visual inspection of the hourly energy generation curve.
2. A Bayesian regularized neural network (BRNN) [18] was trained using as input variables only the explanatory weather variables (V1 to V4) of this reduced dataset. This type of single hidden layer neural network was chosen because of its better generalization characteristics and lower tendency to overfitting than back-propagation neural networks. The number of neurons in the hidden layer was determined using a 5-folds cross-validation procedure.
3. To improve the performance of the hourly PV energy generated forecasting model for clear-sky hours, we used an ensemble of 100 BRNNs as the described in the previous step. Each one of the BRNNs in the ensemble corresponded to a training process of the same BRNN with different initial weight values. Once the ensemble was trained, the hourly PV energy generation values, assuming a clear-sky condition for all the records in the training and testing dataset, were obtained as the mean value of those values provided by each one of the members (BRNNs) of the ensemble.
4. The values obtained in the previous step were used to normalize the hourly PV energy generation of the training and testing dataset according to (3).

In the second stage, two point forecasting models were developed using a similar solution to the one described in the previous stage, that is, two ensembles of BRNNs neural networks. These models were trained to achieved deterministic forecasts of the normalized hourly PV energy generation and of the square of the normalized hourly PV energy generation. The selection of the optimal number of neurons for the members of each ensemble was carry out with a 5-folds procedure with the training dataset. The mean value of the outputs of each of the two ensembles constituted the two deterministic forecasts needed for the determination of the parameters of the Beta distributions. Once the ensembles were trained, they were used to obtain the forecasts for the testing dataset. These forecasts were used to determine the values of the two parameters of the Beta distribution for each hour in the testing dataset. Their values were corrected (when it was needed) by applying (5) with an ε value of 0.001. The expected value of the hourly PV generated energy for each hour was calculated as the mean value of the corresponding Beta distribution multiplied by the energy generated during each hour in a clear-sky day. The expected hourly PV energy values were used to compute the *RMSE* and *MAE* indexes values for the complete testing dataset.

The probabilistic forecasts of the two PFMB models were evaluated by means of *CRPS*. Instantaneous *CRPS* was calculated for each hour of the testing dataset using the cumulative probability function of each hourly Beta distribution. Finally, the mean *CRPS* was calculated for the entire testing dataset. The values of the *CRPS*, *RMSE* and *MAE* indexes obtained for the two PFMB models are shown in Table 2.

Table 2. Forecasting results for the testing dataset.

Model	CRPS (MW)	MAE (MW)	RMSE (MW)
PFMB1	0.702	1.037	1.632
PFMB2	0.556	0.757	1.233
LQRM1	0.851	1.222	1.632
LQRM2	0.692	0.994	1.319

The two linear quantile regression models, LQRM1 and LQRM2 models, were developed by fitting their coefficients using the training dataset for 99 quantiles, from quantile 0.01 to 0.99. The LQRM1 model used as explanatory variables exclusively the group of "Solar variables", while the LQRM2 model used the total set of explanatory variables shown in Table 1 (variables V1 to V13). Once the coefficients were determined, they were used to calculate the outputs corresponding to the testing dataset. The output corresponding to the 0.5 quantile was taken as the expected value, and the corresponding values of the *RMSE* and *MAE* indexes were determined. For the calculation of the instantaneous *CRPS* for each hour, the empirical distribution formed by the 99 values of the quantiles was used. The results for the three evaluation indexes (*CRPS, MAE* and *RMSE*) of the linear quantile regression models LQRM1 and LQRM2 for the testing dataset are shown in Table 2.

By comparing the prediction results shown in Table 2, various conclusions can be drawn:

- The use of weather forecast as explanatory variables improves the forecasting results of the models. Thus, the PFMB2 model obtains better values in all the evaluation indexes than the PFMB1 model. The LQRM2 model also performs better than the LQRM1 model.
- If we compare models using the same set of explanatory variables, PFMB models are superior to LQRM models. Thus, the PFMB1 model presents better values of *MAE* and *CRPS* than the LQRM1 model, and a similar value for the *RMSE* index. Both models use the "Solar variables" group (Table 1) as explanatory variables. The PFMB2 model presents better values in all three evaluation indexes than the LQRM2 model. These two models use all the explanatory variables listed in Table 1 as explanatory variables.
- From the four developed models, the PFMB2 model is the one that shows the best forecasting results from the viewpoints of probabilistic forecast and point forecast.

Fig. 1 represents the forecasts and actual hourly PV energy generation values for the sunlight hours of seven consecutive days in the testing period (1 Oct 2019 to 7 Oct 2019) obtained with the PFMB2 model. The expected value of hourly PV energy generation (point forecast) is plotted in red colour; the real value is plotted in blue colour; and the prediction intervals for 50% of probability (PI 50%, corresponding to values between quantiles 0.25 and 0.75) and for 90% of probability (PI 90%, corresponding to the values between quantiles 0.05 and 0.95).

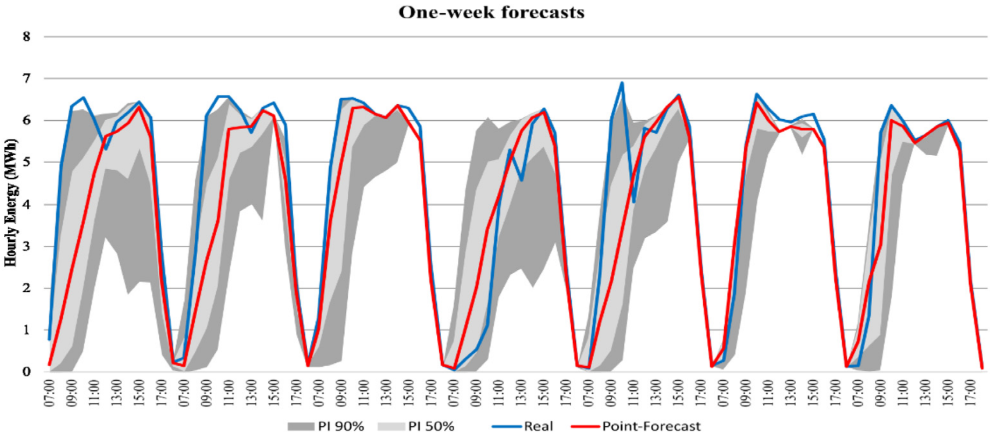

Fig. 1. Actual and forecasted hourly PV energy generation values for sunlight hours in seven consecutive days in the testing period. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5. Conclusions

This paper presents two novel probabilistic forecasting models, called PFMB1 and PFMB2, designed to provide the probability density functions of hourly energy generation in a PV plant for all the hours of the following day. These models are based on a semi-parametric approach by which the two parameters of a Beta distribution, corresponding to the PDF of the hourly PV energy production, are determined by means of forecasts obtained from deterministic forecasting models. Such deterministic forecasting models provide normalized values of hourly PV energy generation and of its square value. The data are previously normalized in the range 0 to 1 using the hourly PV energy generation in each hour assuming a clear-sky day.

The PFMB1 model uses four explanatory variables corresponding to the position of the sun with respect to the location of the PV plant in each of the hours. The PFMB2 model uses nine additional explanatory variables corresponding to weather forecasts. This model yields better computer forecasting results than the PFMB1 model, illustrating the predictive value of using weather forecasts as explanatory variables.

The computer results of the two PFMB models developed with the proposed methodology, PFMB1 and PFMB2, were compared with those provided by benchmark probabilistic forecasting models. These benchmark models used the same explanatory variables that the PFMB models. The results of the new PFMB models were better than those of the benchmark models, from the viewpoints of probabilistic forecast and point forecast.

Further research is undergoing to improve the forecasting results of the proposed models. The new research lines on which we are working include the extension of the explanatory variables set including new variables related to the temporal and spatial variation of the forecasts of weather variables more directly related to the PV power production.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

This research is a part of the project Afianza 2021/03 funded by Government of La Rioja, Spain, and the grant RED2018-102618-T funded by MCIN/AEI/10.13039/501100011033, Spain.

References

[1] Renewable capacity statistics 2022. International Renewable Energy Agency; 2022, https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2022/Apr/IRENA_RE_Capacity_Statistics_2022.pdf, (accessed 2nd Jun 2022).

[2] World energy outlook 2021. International Energy Agency; 2021, https://www.iea.org/reports/world-energy-outlook-2021, (accessed 2nd Jun 2022).

[3] Antonanzas J, Osorio N, Escobar R, Urraca R, Martinez-de Pison FJ, Antonanzas-Torres F. Review of photovoltaic power forecasting. Sol Energy 2016;136:78–111. http://dx.doi.org/10.1016/j.solener.2016.06.069.

[4] Lucas Segarra E, Ramos Ruiz G, Fernández Bandera C. Probabilistic load forecasting for building energy models. Sensors 2020;20:6525. http://dx.doi.org/10.3390/s20226525.

[5] van der Meer DW, Widén J, Munkhammar J. Review on probabilistic forecasting of photovoltaic power production and electricity consumption. Renew Sustain Energy Rev 2018;81:1484–512. http://dx.doi.org/10.1016/j.rser.2017.05.212.

[6] Lorenz E, Hurka J, Heinemann D, Beyer HG. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. IEEE J Sel Top Appl Earth Obs Remote Sens 2009;2:2–10. http://dx.doi.org/10.1109/JSTARS.2009.2020300.

[7] David M, Ramahatana F, Trombe PJ, Lauret P. Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models. Sol Energy 2016;133:55–72. http://dx.doi.org/10.1016/j.solener.2016.03.064.

[8] Markovics D, Mayer MJ. Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction. Renew Sustain Energy Rev 2022;161:112364. http://dx.doi.org/10.1016/j.rser.2022.112364.

[9] Tina GM, Ventura C, Ferlito S, De Vito S. A state-of-art-review on machine-learning based methods for PV. Appl Sci 2021;11:7550. http://dx.doi.org/10.3390/app11167550.

[10] Mellit A, Massi Pavan A, Ogliari E, Leva S, Lughi V. Advanced methods for photovoltaic output power forecasting: A review. Appl Sci 2020;10:487. http://dx.doi.org/10.3390/app10020487.

[11] Forbes C, Evans M, Hastings N, Peacock B. Beta distribution. In: Forbes C, Evans M, Hastings N, Peacock B, editors. Statistical distributions. 4th ed.. Hoboken, NJ, USA: Wiley; 2010, p. 55–61. http://dx.doi.org/10.1002/9780470627242.ch8.

[12] Pinson P, Nielsen HA, Møller JK, Madsen H, Kariniotakis GN. Non-parametric probabilistic forecasts of wind power: required properties and evaluation. Wind Energy 2007;10:497–516. http://dx.doi.org/10.1002/we.230.

[13] Gneiting T, Katzfuss M. Probabilistic forecasting. Annu Rev Stat Appl 2014;1:125–51. http://dx.doi.org/10.1146/annurev-statistics-062713-085831.

[14] Koenker R, Bassett G. Regression quantiles. Econometrica 1978;46:33–50.

[15] Chernozhukov V, Fernández-Val I, Galichon A. Quantile and probability curves without crossing. Econometrica 2010;78:1093–125. http://dx.doi.org/10.3982/ECTA7880.

[16] Meteogalicia. Conselleria de Medio Ambiente, Territorio e Vivienda. Xunta de Galicia; 2022, https://www.meteogalicia.gal (accessed 10th Jan 2022).

[17] Cullen AC, Frey HC HC. Probabilistic techniques in exposure assessment. 1st ed.. NY: Plenum Publishing Co.; 1999.

[18] Burden F, Winkler D. Bayesian regularization of neural networks. In: Livingstone DJ, editor. Artificial neural networks. Methods in molecular biology, Vol. 458. Humana Press; 2008, http://dx.doi.org/10.1007/978-1-60327-101-1_3.