Comparative Study of PV Power Forecast Using Parametric and Nonparametric PV Models

Marcelo Pinho Almeida[a,*], Mikel Muñoz[b], Iñigo de la Parra[b], Oscar Perpiñán[c,d]

[a] *Instituto de Energia e Ambiente, Universidade de São Paulo, São Paulo, Brazil*
[b] *Departamento de Ingeniería Eléctrica y Electrónica, Universidad Pública de Navarra, Pamplona, Spain*
[c] *Electrical Engineering Department, ETSIDI-UPM, Madrid, Spain*
[d] *Instituto de Energía Solar, Universidad Politécnica de Madrid, Madrid, Spain*

Abstract

Forecast procedures for large ground mounted PV plants or smaller BIPV or BAPV systems may use a parametric or a nonparametric model of the PV system. In this paper, both approaches are used independently to calculate the energy delivered to the grid on an hourly basis in forecast procedures that use meteorological variables from a Numerical Weather Prediction model as inputs, and their performances against real generation data from six PV plants are analyzed. The parametric approach relies on mathematical models with several parameters that describe the PV systems and it was implemented in MATLAB®, whereas the nonparametric approach is based on Quantile Regression Forests with training and forecast stages and its code was built in `R`. The parametric approach presented more significant bias on its results, mostly due to the input data and the transposition model of irradiance from a horizontal surface to the plane of the PV array.

Keywords:
PV plant, PV power forecast, Quantile Regression Forests, PV system modeling

1. Introduction

Forecasting the AC power delivered to the grid by a PV system is important for both plant owners and electric system operators in order to minimize technical risks to the grid and expenses related to the uncertainty of generation, to program the dispatch of conventional power generators and to maximize profits with the sale of electricity.

In this context, a PV plant can be treated as a box with several inputs (irradiation, temperature and wind speed, for example) and the AC power injected into the electrical grid as the output. Two major kinds of modeling can be employed to represent this box and estimate the AC power output:

*Corresponding author

Email addresses: `marcelopa@iee.usp.br` (Marcelo Pinho Almeida), `mikel.munoz@unavarra.es` (Mikel Muñoz), `inigo.delaparra@unavarra.es` (Iñigo de la Parra), `oscar.perpinan@upm.es` (Oscar Perpiñán)

- The parametric modeling, which conceives the PV system as a white box where each subsystem, or component, can be modeled using a collection of parameters and physical equations. Some of these equations are used to model the components' performance and others to obtain the operation conditions of the PV generator.
- The nonparametric modeling, which conceives the PV system as a black box, presuming no knowledge of internal characteristics and processes of the PV system. It estimates the behavior of the PV system from a historical time series of inputs and outputs.

The parametric approach is a multi-step model that requires detailed characterization of the performance of each significant component of the PV plant, and to obtain the operation conditions of the PV generator, it is necessary to know the correlation model that fits correctly to the site. Because all this information is not always available, some simplifications and preliminary assumptions are usually made, what leads to a subsequent uncertainty in the output. Consequently, the accuracy and precision of the estimations of a parametric model is driven not only by the performance of its sub-models, but also by the accuracy of the parameters. A parametric methodology can also be used to compute the AC power output prior to the construction of a PV plant, during the project and planning stages, by using, for example, the nameplate characteristics of the PV plant components and generic decomposition/translation models.

Some recent researches have been using the parametric approach in PV power forecast procedures. For example, (Lorenz et al., 2011) derives PV power with a set of physical modeling steps using as input post-processed solar irradiance forecasts based on forecasts of the global model of the European Centre for Medium-Range Forecasts (ECMWF) and (Pelland et al., 2013) uses PV simulation models with the spatially averaged solar irradiance forecasts derived from post-processing of a global numerical weather prediction model, namely Environment Canada's Global Environmental Multiscale (GEM) model.

The nonparametric approach avoids the need for simplifying assumptions and accurate parameters by using historical time series of meteorological variables (inputs) and AC power measurements (output). Therefore, the accuracy of a nonparametric model depends mainly on the quality of these data. However, this characteristic also leads to its main disadvantage: the PV plant must exist and be operational for some time, so relevant input/output information is available. One interesting advantage of a nonparametric model is the potential to compensate systematic errors (bias) associated with the inputs.

Recent investigations have been employing the nonparametric approach on PV power forecast procedures. (Zamo et al., 2014) analyzes a mix of eight statistical methods to forecast PV power one day ahead on an hourly basis. The Random Forests method presents the best results. (Mandal et al., 2012) forecasts one-hour-ahead power output using a combination of wavelet transform and neural network techniques by incorporating the interactions of PV system with solar radiation and temperature data. (Bacher et al., 2009) forecasts hourly values of AC power for horizons of up to 36 hours using adaptive linear time series models, namely autoregressive without and with exogenous input models. The autoregressive with exogenous model uses numerical weather predictions as input. (Pedro and Coimbra, 2012) predicts 1 and 2 hour-ahead solar power comparing several forecast techniques without exogenous inputs such as Auto-Regressive Integrated Moving Average, k-Nearest-Neighbors, Artificial Neural Networks, and Neural Networks optimized by Genetic Algorithms.

This paper presents the results of a comparative study between the performances of two PV power output forecast procedures, each employing one of the modeling approaches presented above, using meteorological forecast (see Section 2) as input. It is not the objective of this paper to select the "best" PV modeling approach to be used with PV power forecast procedure, but to discuss their pros and cons. The predictions using both approaches have been compared with measured AC power from six PV plants (see Section 5) as described in Section 6, and the results and discussions are presented in Sections 7, 8 and 9.

2. Meteorological Forecast

The meteorological forecast was downloaded from Meteogalicia, a Spanish meteorological institute located in Xunta de Galicia. It regularly publishes results from a regional mesoscale[1] NWP model, namely Weather Research and Forecasting (WRF) (Skamarock et al., 2005), freely at its THREDDS server[2]. The WRF model runs twice a day, at 00UTC (forecast for the next 96 hours) and 12UTC (forecast for the next 84 hours), with a spatial resolution of 12 km × 12 km in an area comprised between 21.58°W to 6.36°E and 33.64°N to 49.57°N and a hourly time resolution. Meteogalicia also maintains an historical archive of past forecasts available on-line.

The WRF outputs includes a wide collection of variables, such as solar radiation, temperature, wind speed or cloud cover[3]. These outputs are provided as raster data, that is, a matrix (or layers of matrices) of cells organized into rows and columns. Each raster file corresponds to a certain WRF variable, with each cell containing the value for a geographical location defined by a spatial grid. The Meteogalicia raster files comprise several layers, with a different layer for each hourly forecast. Table 1 presents the name and description of the WRF variables considered in this paper for both forecast procedures.

Label	Description	Forecast procedure with
$swflx$	surface downwelling shortwave flux	Parametric / Non-parametric PV model
$temp$	temperature at 2 m	Parametric / Non-parametric PV model
mod	wind speed at 2 m	Parametric PV model

Table 1: WRF-NWP variables used to calculate AC power.

3. Forecast Procedure with Parametric PV Model

Figure 1 displays a diagram for a general configuration of a grid-connected PV system, which is composed by a PV generator, an inverter (MPPT + DC/AC converter), a low voltage/medium voltage (LV/MV) transformer and a medium voltage/high voltage (MV/HV) transformer.

A parametric model estimates the power produced by this system using as input variables the ambient temperature (T_a), the wind speed (W_S) and the global horizontal irradiance (G_0). In this case, these variables are all forecasts by the Meteogalicia service (WRF variables $temp$, mod and $swflx$, respectively). The forecast procedure is as follows:

1. The effective irradiance on the plane of the PV array, G_{eff}, is estimated from the irradiance on the horizontal plane using the following procedure:
   (a) Sun and trackers geometry with the set of equations as provided in (Lorenzo, 2011).

[1]Mesoscale meteorology studies weather systems with horizontal dimensions ranging from around 5 kilometers to several hundred kilometers.

[2]http://www.meteogalicia.es/web/modelos/threddsIndex.action

[3]http://mandeo.meteogalicia.es/thredds/catalogos/WRF_2D/catalog.html

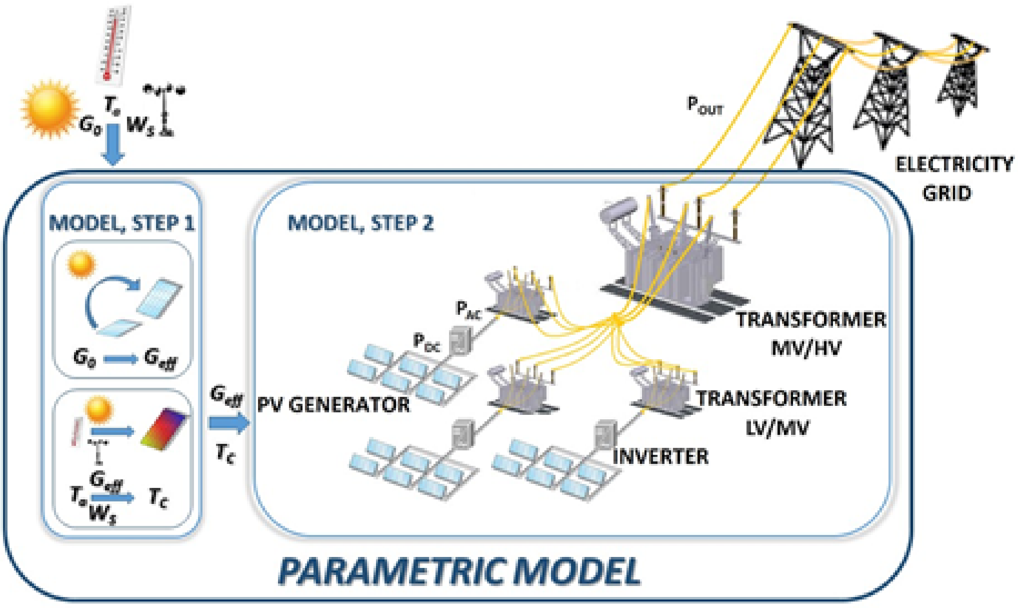

Figure 1: Diagram for a general configuration of a grid-connected PV system.

(b) Decomposition of hourly global horizontal irradiation using correlations between the diffuse fraction of horizontal radiation and the clearness index. . Different correlations for different locations are available in the literature (Engerer, 2015; Erbs et al., 1982). Each correlation best suits for a specific location and it is not possible to determine in advance which is the best for the chosen location. In this paper, due to its generality and simplicity, the Erbs' correlation has been chosen. Other correlations were also tested but the results (not shown here) were slightly worse. This decomposition model represents the first source of inaccuracy of the parametric model (Moretón, 2016).

(c) Estimation of irradiance on the inclined plane: the direct irradiance is calculated with geometrical equations, while the estimation of the diffuse component uses an anisotropic model. The most popular models present in the literature were proposed by Hay and Mckay (Hay and Mckay, 1960) and Perez (Perez et al., 1992). The latter has been used for the results which are shown here. It is worth noticing that, although Perez's model is more complicated than Hay and Mckay's, differences between the statistics used for measuring the performance of the predictions with both models are lower than 1%.

(d) Effects of dirt and angle of incidence with the equations proposed in (Martin and Ruiz, 2001).

(e) Shading losses (Martínez, 2012) and spectral corrections (Fuentes et al., 2007).

2. The cell temperature, T_c, is estimated from T_a, W_S and G_{eff} (equation 1) (Faiman, 2008):

$$T_C = T_a + \frac{G_{eff}}{U_0 + U_1 W_S} \tag{1}$$

where U_0 is the constant heat transfer component (W/m^{2}K) and U_1 is the convective heat transfer component (W/m^{2}K). U_0 varies from 23.5 W/m^{2}K to 26.5 W/m^{2}K and U_1 varies from 6.25 W/m^{2}K to 7.68 W/m^{2}K. The values used in this paper were U_0 = 25 W/m^{2}K and U_1 = 6.84 W/m^{2}K.

3. Finally, the power injected into the grid, P_{out}, is estimated based on G_{eff}, T_c, and the technical characteristics of the PV system. As a result of the characteristics of Meteogalicia's forecast, the PV output power is forecasted with hourly resolution.
The PV generator performance has been modeled using the simple, yet precise, expression given by Equation 2 proposed by Evans (Evans, 1981) and Osterwald (Osterwald, 1986).

$$P_{DC} = P_{STC}\frac{G_{eff}}{G_{STC}}(1-\gamma(T_C - T_{STC})) \tag{2}$$

where P_{DC} is the PV DC power (in W), P_{STC}, G_{STC} and T_{STC} are the generator power (in W), the irradiance (in W/m^2) and the temperature (in °C) under the Standard Test Conditions, respectively, and γ is the temperature coefficient of power (in °C).
The inverter is characterized by its nominal output power (P_I) and three experimental parameters (k_0, k_1 and k_2) used to calculate its efficiency, η_I (equation 3), according to the method proposed in (Jantsch et al., 1992):

$$\eta_I = \frac{P_{AC}}{P_{DC}} = \frac{p_o}{p_o + k_0^o + k_1^o p_o + k_2^o p_o^2} \tag{3}$$

where $p_o = P_{AC}/P_I$ is the normalized output power of the inverter.
The efficiency of the transformers, η_T, can be expressed as a function of the AC power delivered by the inverter, P_{AC}, using Equation 4.

$$\eta_T = \frac{P_{out}}{P_{AC}} = \frac{P_{AC} - P_{core} - P_{Cu}}{P_{AC}} \tag{4}$$

where P_{AC} and P_{out} are the AC power in the primary and the secondary coils of the transformer, respectively, P_{core} are the core losses of the transformer and P_{Cu} are the copper losses of the transformer, which can be calculated with Equation 5.

$$P_{Cu} = P_{Cu,nom}\frac{P_{out}}{P_T}^2 \cong P_{Cu,nom}\frac{P_{AC}}{P_T}^2 \tag{5}$$

where $P_{Cu,nom}$ is the copper losses of the transformer when it operates at its nominal power, P_T. Power losses in DC and AC wiring are calculated using equations analogous to Equation 5.
The parameters needed in each of the PV plant components' models can be obtained from the standard information provided by manufacturers or promoters, which may be adjusted experimentally by on-site quality control testing procedures. In the case of the PV generator's nominal power, the use of standard information provided by datasheets generally leads to systematic errors in predictions. For example, using the power value provided by the datasheet of the module to calculate the generator's nominal power may cause errors within the predictions of up to 5% (Garcia et al., 2009; Martínez-Moreno et al., 2011), therefore its use is not recommended. In this paper, the PV generator nominal power is the one obtained from field measurements carried out at the PV plants.

4. Forecast Procedure with Nonparametric PV Model

The forecast procedure with nonparametric PV model used in this paper was extensively detailed and validated in (Almeida et al., 2015). PV output power is forecasted one day ahead with hourly resolution using Quantile Regression Forests, allowing statistical information about the quantiles of the hourly prediction. Its methodology is as follows:

- Previous AC power measurements from a PV plant are collected.
- Forecasts of a set of Weather Research and Forecasting (WRF) variables (solar radiation, cloud cover, temperature, wind speed, etc.) from a Numerical Weather Prediction (NWP) model run by a meteorological institute are downloaded.
- Each WRF variable is processed to extract information about the value at the location of interest and its relation with the surrounding locations and previous forecasts. In addition, three calculated variables describing the Sun-Earth geometry are included in the predictor set: azimuth angle (AzS), altitude angle (AlS), and extra-terrestrial irradiance on the horizontal plane ($Bo0$).
- The time series of processed WRF variables and AC power measurements is divided into two time series: train and test. The train time series comprises past values of both WRF variables and AC power, whereas the test time series contains only present WRF variables from the NWP model (forecasts).

  The training time series, or training set, may have different sizes, what eventually leads to distinct results. For practical purposes, the size of the training set is defined in days. Therefore, the training set can be composed by N days, selected from a bigger database.

  Three selection methods were analyzed:

  - *Previous*

    This method selects those N days immediately before the day to be predicted. As a consequence, the database must be complete up to the day prior the prediction.
  - *KT*

    This method selects N days according to the absolute difference between the clearness index of the day to be predicted and the clearness index of each day included in the database. Both clearness indexes are computed with the irradiance forecast retrieved from the NWP model. The N days with the lowest absolute difference are chosen to conform the training set.
  - *KS*

    This method selects N days according to the similarity between the empirical distribution function of the irradiance forecast for the day to be predicted and the empirical distribution function of the irradiance forecast for each day included in the database. Here the Kolmogorov-Smirnov statistic is used to compute the distance between the distributions. The N days with the lowest Kolmogorov-Smirnov distance are chosen to conform the training set.

  Both selecting methods KT and KS do not need the database to be completed up to the day prior the prediction, and could also be composed by older information.
- A machine learning tool based on Random Forests (Breiman, 2001; Meinshausen, 2006) (namely Quantile Regression Forests, QRF) is trained with the train time series.
- Predictions of the median (quantile 0.5) and a confidence interval (quantiles 0.1 and 0.9) for the AC power are generated with the test time series.

The study in (Almeida et al., 2015) also contributes with an analysis on how additional variability indexes, daily clearness index, training set length, training set selection method and different configurations of predictors influence on the final results.

The methodology was built upon the `R` environment (R Development Core Team, 2014) using a list of contributed packages:

- `rgdal` and `raster` for raster data manipulation (Bivand et al., 2013; Hijmans, 2013).
- `zoo`, `xts`, and `data.table` for time series analysis (Dowle et al., 2014; Ryan and Ulrich, 2013; Zeileis and Grothendieck, 2005).
- `gstat` for spatial interpolation (Pebesma, 2004).
- `meteoForecast` to import NWP-WRF forecasts (Perpiñán and Almeida, 2015).
- `solaR` for sun geometry calculation (Perpiñán, 2012).
- `quantregforest` for Quantile Regression Forests (Meinshausen, 2006).

The full code is freely available from the repository `https://github.com/iesiee/PVF`, which itself is a `R` package named `PVF` (Almeida and Perpiñán, 2014). An online toolbox that implements the nonparametric methodology is freely available at `http://vps156.cesvima.upm.es:3838/predictPac`.

5. PV Plants Database

As mentioned before, data from six PV plants were used in this paper. Five are located in northern Spain (latitude between 42° and 42.5°), with 5-s resolution measurements synchronized via GPS, previously analyzed in (Marcos et al., 2011), and the other is located in southern Portugal ((latitude around 38.2°), also with 5-s resolution measurements.

Their monitoring systems record operational and meteorological data, including AC power, global horizontal irradiance, on-plane irradiance[4], ambient temperature and wind speed. The database used in this study is restricted to the period comprised between January 1st, 2010 and December 29th, 2010. The database was reduced to 1-h resolution due to the restrictions of the forecast data used in this study.

All PV plants have an azimuthal one-axis tracker, with a receiving surface tilted 45°. Table 2 summarizes the main individual characteristics of each PV plant. Their installed power ranges from 45.6 MWp to 958 kWp, with areas ranging from 250 ha to 4.1 ha. The inverter sizing factor, $c_{inv} = P_{PV}^{STC}/P_{inv}$, ranges from 1.24 to 1.32, where P_{PV}^{STC} is installed PV peak power and P_{inv} is inverter rated power.

6. Procedures to Assess the Performance

The performance of the forecast procedure with nonparametric PV model has been assessed using a leave-one-out cross-validation procedure:

- The considered day is extracted from the database to be the test set.

[4] On-plane irradiance is the irradiance on the same plane of the PV modules that compose the PV generator.

Label	Peak power (kWp)	Rated power (kW)	Area (Ha)	Location
P1	958	775	4.1	Spain
P2	990	780	4.2	Spain
P3	1438	1155	6.4	Spain
P4	1780	1400	8.7	Spain
P5	2640	2000	11.8	Spain
P6	45600	38500	250	Portugal

Table 2: PV plants characteristics.

- The training set, used to train the QRF, is constructed with $N = 30$ days extracted from the remaining days of the data set, according to the selecting method KS. These configurations were selected due to the good performance presented in (Almeida et al., 2015).
- The trained QRF is used to predict hourly AC power (quantile $Q_{.5}$), for the test set.
- The hourly error between these predictions and AC power measurements for the test day is characterized with the model performance statistics described in section 7.

On the other hand, the performance of the forecast procedure with parametric PV model has been assessed with the following procedure:

- Hourly AC power for the test day is calculated with the corresponding weather forecast.
- The hourly error between predictions and AC power measurements is characterized with the model performance statistics described in section 7.

7. Statistical Comparison

Different statistics are commonly used to quantify the discrepancy, or error, between forecast and actual observation in order to evaluate the performance of a model. The Mean Bias Error (MBE, Equation 6) and the Root Mean Square Error ($RMSE$, Equation 7) are the most widely reported statistics (Gueymard, 2014).

$$MBE = \overline{\mathbf{D}} = \overline{\mathbf{F}} - \overline{\mathbf{O}} = \frac{1}{n}\sum_{i=1}^{n}(f_i - o_i) \tag{6}$$

$$RMSE = \left(\overline{\mathbf{D}^2}\right)^{1/2} = \left(\frac{1}{n}\sum_{i=1}^{n} d_i^2\right)^{1/2} = \left(\frac{1}{n}\sum_{i=1}^{n}(f_i - o_i)^2\right)^{1/2} \tag{7}$$

where the upper line denotes average. It is important to note that negative and zero values of observations are not considered in this paper. Therefore, n is the number of hours when both observations and predicted values are strictly positive.

The $RMSE$ aggregates information for both the average and the variance (or unbiased $RMSE$, Equation 8) of the error, as described in Equation 9. Therefore, the $RMSE$ must be reported together with another measurement, such as the MBE, that allows discerning between average and variance.

$$\sigma_{\mathbf{D}}^2 = \frac{1}{n}\sum_{i=1}^{n}(d_i - \overline{\mathbf{D}})^2 \tag{8}$$

$$RMSE^2 = \sigma_{\mathbf{D}}^2 + \overline{\mathbf{D}}^2 = \sigma_{\mathbf{D}}^2 + MBE^2 \tag{9}$$

It must be noted that in the $RMSE$ each error influences the total in proportion to its square, rather than its magnitude. Thus, large errors have a relatively greater influence on the total square error than do the smaller errors. In consequence, the $RMSE$ may be distorted if the total error is concentrated within a small number of outliers (Willmott and Matsuura, 2005).

The Mean Absolute Error, MAE, (Equation 10), is less influenced by outliers, although the relative influence depends on the number of samples and the error distribution (Chai and Draxler, 2014).

$$MAE = \overline{|\mathbf{D}|} = \frac{1}{n}\sum_{i=1}^{n}|d_i| = \frac{1}{n}\sum_{i=1}^{n}|f_i - o_i| \tag{10}$$

As previously stated, the $RMSE$ is more widely reported, although the MAE is useful to unveil the occurrence of outliers.

As the PV plants used in this study have very different nominal powers, absolute values of MBE, $RMSE$ and MAE may blur the comparison between simulations, so they have to be normalized. In statistical studies, it is common to normalize these statistics to the range, $\max(\mathbf{O}) - \min(\mathbf{O})$, or the mean, $\overline{\mathbf{O}}$, of the observations. For a statistical comparison, the first option was chosen to ensure most of the values fall in a range between 0 and 1. Table 3 summarizes the performance statistics as they are used in this paper.

Statistic	Description
$nMBE$	MBE normalized by the daily range of the observations
$nRMSE$	$RMSE$ normalized by the daily range of the observations
$nMAE$	MAE normalized by the daily range of the observations

Table 3: Performance statistics used to compare both the performances of AC power forecast with parametric and nonparametric PV models.

Tables 4 and 5 show the results for the parametric and the nonparametric model, respectively. As a large number of days were simulated, the values presented in Tables 4 and 5 are the quantiles 25% ($QS_{.25}$), 50% ($QS_{.5}$) and 75% ($QS_{.75}$) obtained from the pool of results of each performance statistic. Also, the results are grouped according to the daily clearness index (KTd) into three classes: cloudy days ($0 \leq KTd < 0.532$), partially cloudy days ($0.532 \leq KTd < 0.678$) and clear days ($0.678 \leq KTd < 1$). The ranges of KTd were selected so that each class comprises one third of the total number of days present in the database. Both Tables are displayed together in Figure 2.

It is easy to observe that the $nRMSE$ of the nonparametric approach is due almost entirely to the $nMAE$, showing that the overall daily prediction has small energy errors ($|nMBE| < 3\%$ for quantile $QS_{.5}$ for all six PV plants), but individual hours may present significant deviations from real measurements ($4\% < |nMAE| < 30\%$ for quantile $QS_{.5}$ for all six PV plants).

The performance of the parametric approach is lower than the performance of the nonparametric approach, especially during cloudy days. During clear and partially clouded days, when most of the energy is generated, the difference of $nRMSE$ (quantile $QS_{.5}$) is bellow 23% for all

PV plant	Statistic	$0 \leq KTd < 0.532$			$0.532 \leq KTd < 0.678$			$0.678 \leq KTd \leq 1$		
		$QS_{.25}$	$QS_{.5}$	$QS_{.75}$	$QS_{.25}$	$QS_{.5}$	$QS_{.75}$	$QS_{.25}$	$QS_{.5}$	$QS_{.75}$
P1	$nMBE$	1.46%	25.91%	74.57%	12.16%	21.30%	39.18%	10.88%	15.2%	23.62%
	$nRMSE$	39.97%	72.39%	139.44%	27.92%	37.80%	55.97%	17.51%	22.46%	31.62%
	$nMAE$	34.13%	60.73%	103.74%	22.25%	30.49%	47.09%	16.16%	20.10%	28.16%
P2	$nMBE$	-4.23%	21.83%	65.17%	4.27%	15.64%	40.20%	8.20%	12.68%	23.93%
	$nRMSE$	40.87%	61.03%	105.03%	23.15%	34.65%	61.58%	15.45%	20.85%	29.99%
	$nMAE$	32.38%	52.31%	88.18%	18.16%	28.52%	48.72%	13.95%	17.51%	25.59%
P3	$nMBE$	-4.21%	21.02%	72.90%	5.56%	15.21%	35.97%	7.19%	12.81%	21.46%
	$nRMSE$	43.54%	63.19%	106.47%	20.26%	33.78%	52.58%	14.02%	20.84%	34.49%
	$nMAE$	35.55%	56.59%	85.74%	16.54%	28.05%	43.09%	12.99%	17.95%	27.06%
P4	$nMBE$	-4.82%	21.91%	64.02%	1.74%	14.21%	35.20%	9.55%	14.88%	25.28%
	$nRMSE$	37.61%	60.47%	100.05%	19.75%	29.40%	48.50%	14.74%	21.77%	33.44%
	$nMAE$	29.92%	50.62%	81.24%	15.73%	24.12%	40.22%	13.11%	18.59%	26.57%
P5	$nMBE$	-4.85%	18.08%	50.17%	5.61%	16.53%	35.50%	8.18%	12.56%	23.24%
	$nRMSE$	39.96%	59.18%	96.19%	22.96%	34.56%	51.81%	14.81%	21.98%	26.65%
	$nMAE$	31.40%	47.76%	79.11%	17.53%	28.62%	44.48%	13.64%	17.90%	26.65%
P6	$nMBE$	-28.98%	-3.92%	21.32%	-11.77%	1.71%	19.14%	-2.21%	0.27%	6.38%
	$nRMSE$	42.68%	54.52%	77.24%	22.85%	29.32%	38.91%	11.02%	14.51%	19.16%
	$nMAE$	32.28%	42.88%	65.89%	17.05%	24.46%	32.78%	7.53%	9.75%	12.12%

Table 4: Performance statistics for the forecast procedure with parametric PV model for all the six PV plants.

PV plant	Statistic	$0 \leq KTd < 0.532$			$0.532 \leq KTd < 0.678$			$0.678 \leq KTd \leq 1$		
		$QS_{.25}$	$QS_{.5}$	$QS_{.75}$	$QS_{.25}$	$QS_{.5}$	$QS_{.75}$	$QS_{.25}$	$QS_{.5}$	$QS_{.75}$
P1	$nMBE$	-20.55%	2.47%	14.24%	-8.56%	-1.60%	5.64%	-5.03%	0.29%	2.65%
	$nRMSE$	25.29%	32.43%	41.31%	16.52%	20.42%	27.74%	3.70%	7.77%	17.42%
	$nMAE$	18.25%	24.97%	32.29%	11.99%	15.92%	21.21%	2.89%	5.81%	12.76%
P2	$nMBE$	-15.22%	2.20%	13.24%	-9.69%	0.46%	8.16%	-6.28%	-0.46%	2.79%
	$nRMSE$	24.52%	30.96%	41.34%	15.03%	21.61%	28.65%	4.31%	9.73%	20.93%
	$nMAE$	18.35%	21.98%	33.85%	11.33%	16.15%	22.97%	3.15%	6.43%	14.54%
P3	$nMBE$	-9.78%	2.76%	14.33%	-12.97%	-0.09%	8.00%	-4.03%	0.31%	2.16%
	$nRMSE$	24.53%	32.00%	40.28%	15.59%	20.84%	28.55%	3.39%	7.40%	18.44%
	$nMAE$	17.87%	23.79%	31.41%	11.93%	17.16%	21.88%	2.81%	5.69%	13.71%
P4	$nMBE$	-13.85%	0.45%	17.1%	-6.69%	-1.69%	7.74%	-2.53%	0.40%	6.28%
	$nRMSE$	25.04%	33.05%	44.63%	11.90%	20.17%	27.28%	3.98%	9.68%	17.62%
	$nMAE$	18.67%	25.06%	34.92%	8.96%	15.82%	21.25%	2.99%	6.57%	12.72%
P5	$nMBE$	-14.76%	0.02%	15.14%	-4.63%	0.86%	9.80%	-3.25%	-0.69%	3.94%
	$nRMSE$	25.00%	31.69%	39.76%	13.11%	18.87%	27.48%	3.81%	8.29%	17.44%
	$nMAE$	18.09%	23.56%	30.70%	9.49%	14.07%	21.40%	2.82%	6.20%	13.50%
P6	$nMBE$	-15.37%	-1.57%	13.94%	-8.02%	-0.22%	7.40%	-2.10%	-0.25%	4.29%
	$nRMSE$	26.28%	35.24%	44.69%	13.51%	19.86%	26.24%	3.48%	5.91%	10.97%
	$nMAE$	20.06%	28.94%	36.43%	10.17%	14.86%	20.34%	2.76%	4.35%	7.52%

Table 5: Performance statistics for the forecast procedure with nonparametric PV model for all the six PV plants.

PV plants. Another noticeable difference is the confidence interval (colored areas in Figure 2 between quantiles $QS_{.25}$ and $QS_{.75}$) of the nonparametric approach, which is significantly narrower than the confidence interval of the parametric approach.

Finally, the two approaches were compared regarding their Skill Scores (SS), which are widely used in evaluating the performance of meteorological forecasting methods. Skill Scores are defined as a measure of the relative improvement of a forecast method over a commonly

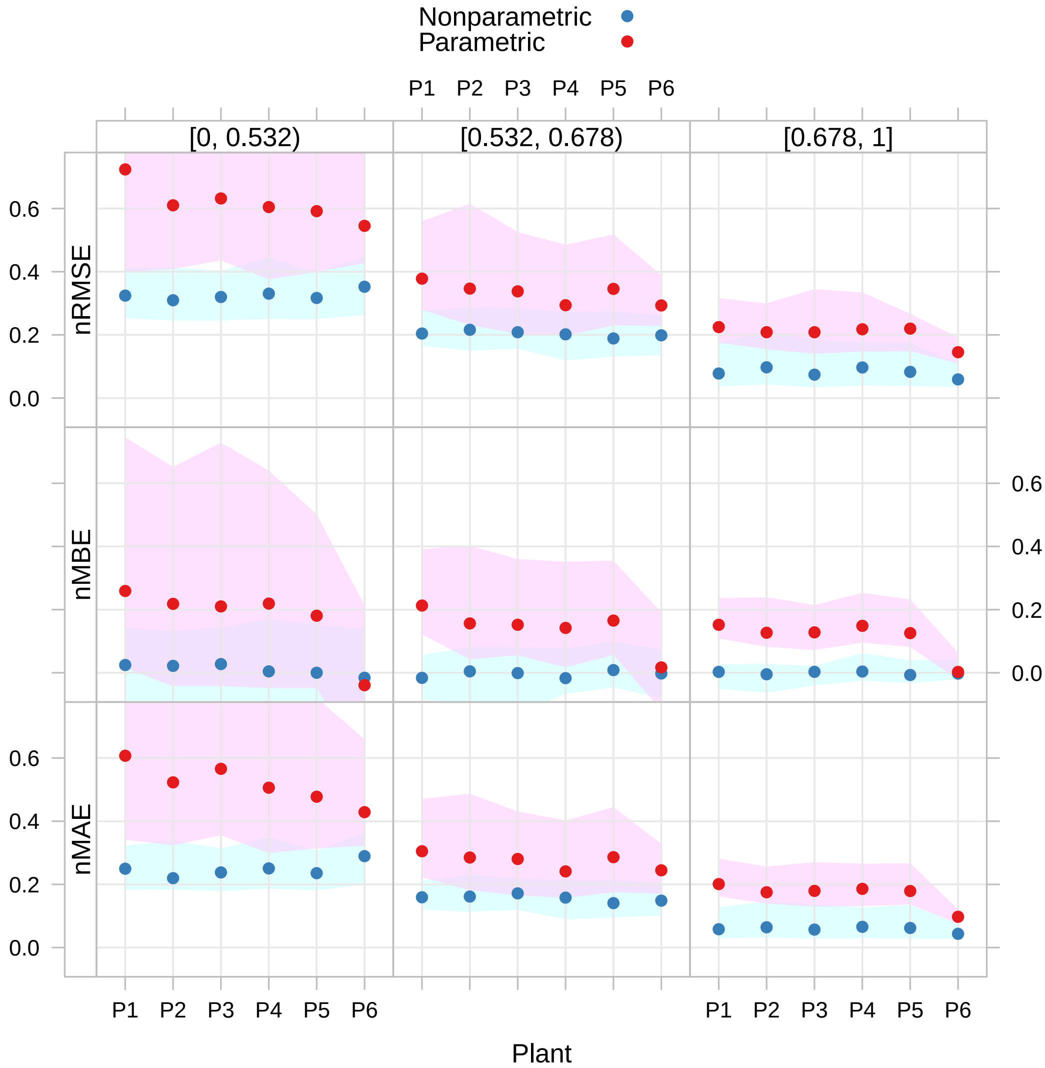

Figure 2: Performance statistics for the forecast procedures with parametric and nonparametric PV models including the six PV plants. The lines represent the value of the quantile $QS_{.5}$, and the bands are comprised between the quantiles $QS_{.25}$ and $QS_{.75}$. Each column corresponds to an interval of KTd.

used reference, which is, in this case, the persistence forecast, the most common choice in forecasting AC power from photovoltaics systems and atmospheric sciences (Wilks, 2011). This reference forecast predicts that AC power during day D will be the same as during day $D-1$. Using the $RMSE$ as the measure of accuracy, the SS is defined as (Murphy, 1988):

$$SS = 1 - \frac{RMSE_f}{RMSE_p} \tag{11}$$

where $RMSE_f$ and $RMSE_p$ are the Root Mean Square Error of the forecast method and of the persistence, respectively.

A $SS > 0$ ($SS < 0$) implies that the forecast method is better (worse) than persistence. $SS = 1$ indicates a perfect forecast and $SS = 0$ implies that the forecast method equals the persistence.

(Inman et al., 2013) provides a comprehensive review of solar forecasting methods with several tables comprising the SS for the various methods. They report values ranging from 0 to 0.42 for a wide variety of objective variables (hourly global or direct irradiation, or AC power), and forecast time horizon (from 15 minutes to 6 days ahead). In the context of this paper, it is relevant the contribution of (Bacher et al., 2009) that reports SS up to 0.36 for the forecast of AC power for next day horizon. It must be underlined that this result corresponds to the average power of a set of 21 different PV systems in a region. Figure 3 shows the SS for every day from the database divided regarding their KTd classes and PV plant.

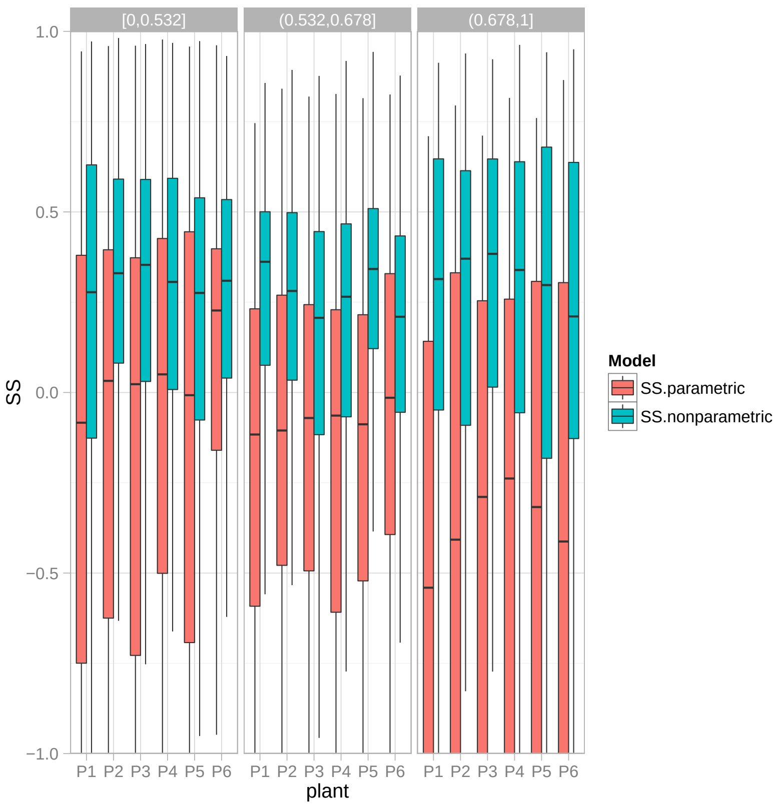

Figure 3: Boxplot of the SS for the forecast procedures with parametric and nonparametric PV models. Each column corresponds to an interval of KTd.

The mean median SS for all KTd classes of the nonparametric approach are greater than the corresponding SS of the parametric approach, and they are also very close to the value of 0.36 reported by (Bacher et al., 2009). In fact, a significant number of values of SS for the nonparametric approach are even bigger than 0.36, what is not true for the parametric approach. In the range of higher KTd, a great number of SS values for this approach are smaller than 0, indicating that a simple method as the persistence could give better results when clear sky conditions remain during several consecutive days.

8. Impacts on Daily Energy Production Forecast

PV power forecast can be used for trading energy in electricity power markets, so an analysis other than a purely statistical one is necessary due to the nature of this final application. This section discusses the performances of the two forecast approaches under the framework of energy assessment, taking into account the general economical benefits and penalties stated in the market regulations. Two important scenarios are accounted here: on the one hand, markets that penalize the daily energy error for which the MBE is appropriate; on the other hand, markets that penalize the hourly energy error, for which the MAE is preferred.

The impacts on daily energy forecast can be estimated considering either the accumulated error at the end of the day or the total hourly energy error during the day. The first estimation will be quantified by the daily MBE normalized by the mean of all the observations from the same day, given by Equation 12, and the second by the daily MAE, also normalized by the mean of observations from the same day, given by Equation 13. As both MBE (or MAE) and the mean of the observations are computed for the same time interval, the ratio between them can be understood as an energy ratio, with the total energy generated during the day in the denominator.

$$cvMBE = \frac{MBE}{\overline{\mathbf{O}}} \tag{12}$$

$$cvMAE = \frac{MAE}{\overline{\mathbf{O}}} \tag{13}$$

The quantiles $QS_{.50}$ for all daily $cvMBE$ and $cvMAE$ for each PV plant were weighted with the fraction of the energy generated by the PV plant under the corresponding clearness index class to give a more meaningful value in terms of energetic impact during the period analyzed. Tables 6 and 7 show the results for the parametric and the nonparametric models, respectively. Both Tables are displayed together in Figure 4.

PV plant	Statistic	$0 \leq KTd < 0.532$	$0.532 \leq KTd < 0.678$	$0.678 \leq KTd < 1$
P1	$cvMBE$	5.36%	9.03%	9.91%
	$cvMAE$	8.68%	12.05%	13.00%
P2	$cvMBE$	6.63%	8.85%	5.55%
	$cvMAE$	13.12%	14.82%	7.80%
P3	$cvMBE$	5.69%	7.06%	6.58%
	$cvMAE$	12.96%	12.00%	8.83%
P4	$cvMBE$	5.39%	6.71%	7.94%
	$cvMAE$	10.15%	10.89%	9.88%
P5	$cvMBE$	4.71%	8.00%	6.86%
	$cvMAE$	9.63%	12.51%	9.36%
P6	$cvMBE$	-1.01%	0.49%	0.14%
	$cvMAE$	8.93%	7.84%	4.83%

Table 6: Weighted errors of energy forecast for the procedure with parametric PV model.

Values of $cvMBE$ for the forecast procedure with a nonparametric PV model are smaller, but this is expected due to the machine learning tool used, which gives low biased results.

PV plant	Statistic	$0 \leq KTd < 0.532$	$0.532 \leq KTd < 0.678$	$0.678 \leq KTd < 1$
P1	$cvMBE$	0.73%	-0.78%	0.22%
	$cvMAE$	7.27%	9.27%	4.59%
P2	$cvMBE$	1.12%	0.30%	-0.27%
	$cvMAE$	11.31%	12.57%	3.82%
P3	$cvMBE$	1.03%	-0.05%	0.22%
	$cvMAE$	9.81%	10.13%	4.05%
P4	$cvMBE$	0.17%	-1.20%	0.27%
	$cvMAE$	9.16%	9.80%	4.49%
P5	$cvMBE$	0.01%	0.59%	-0.47%
	$cvMAE$	9.58%	8.86%	4.08%
P6	$cvMBE$	-0.66%	-0.05%	-0.13%
	$cvMAE$	9.83%	7.45%	2.38%

Table 7: Weighted errors of energy forecast for the procedure with nonparametric PV model.

In terms of hourly prediction, the performances of both forecast procedures are similar, especially for cloudy or partially clouded days. Most of the difference between their performances is due to the bias presented by the parametric method.

9. Sources of Error in the Forecast Procedure with Parametric PV Model

If a PV plant is available with the necessary information, then it is possible to determine the error made at each step of the forecast procedure with parametric PV model. Such is the case of plant P6, equipped with measurements of ambient temperature, horizontal irradiance, wind speed, in-plane irradiance and cell temperature. These measurements are termed $T_{a,m}$, $G_{0,m}$, $W_{S,m}$, $G_{eff,m}$ and $T_{C,m}$, respectively. The hourly energy produced by the PV plant is termed $E_{out,m}$. Table 8 shows the impact of the errors made at each step of the parametric approach. For this purpose, we used daily quantiles $QS_{.50}$ of $nMBE$, $nMAE$ and $nRMSE$. The performance statistics for T_a and T_C are weighted with the coefficient for the power output variation in relation to the module temperature. This better reflects the impact that the errors related to T_a and T_C have on the power output. The error in the wind speed predicted by Meteogalicia was not included in Table 8, given the fact that it has a slight impact on energy production.

The last line of Table 8 corresponds to the errors in the whole parametric approach for plant P6 (the same errors appearing in Table 4). For this particular plant, it can be checked that the bias no longer clearly appears and the cumulative values are very much like those of non-parametric model. However, $nRMSE$ values are significantly different between both methods. This makes high values of SS for the parametric model.

Table 8 shows the analysis for each of the stages of the parametric model. The highest $nRMSE$ value for clear days (higher production) corresponds to the G_{eff} calculation from G_0. Different models have been tested to do this calculation ((Erbs et al., 1982)) ((Engerer, 2015)) and the Erbs' model was the one that led to better results. In the other five PV plants (P1-P5), other decomposition models led to smaller $nRMSE$ of the whole forecast procedure. These results confirm the site-dependence of transposition models and put them as one of the main sources of error of a parametric approach.

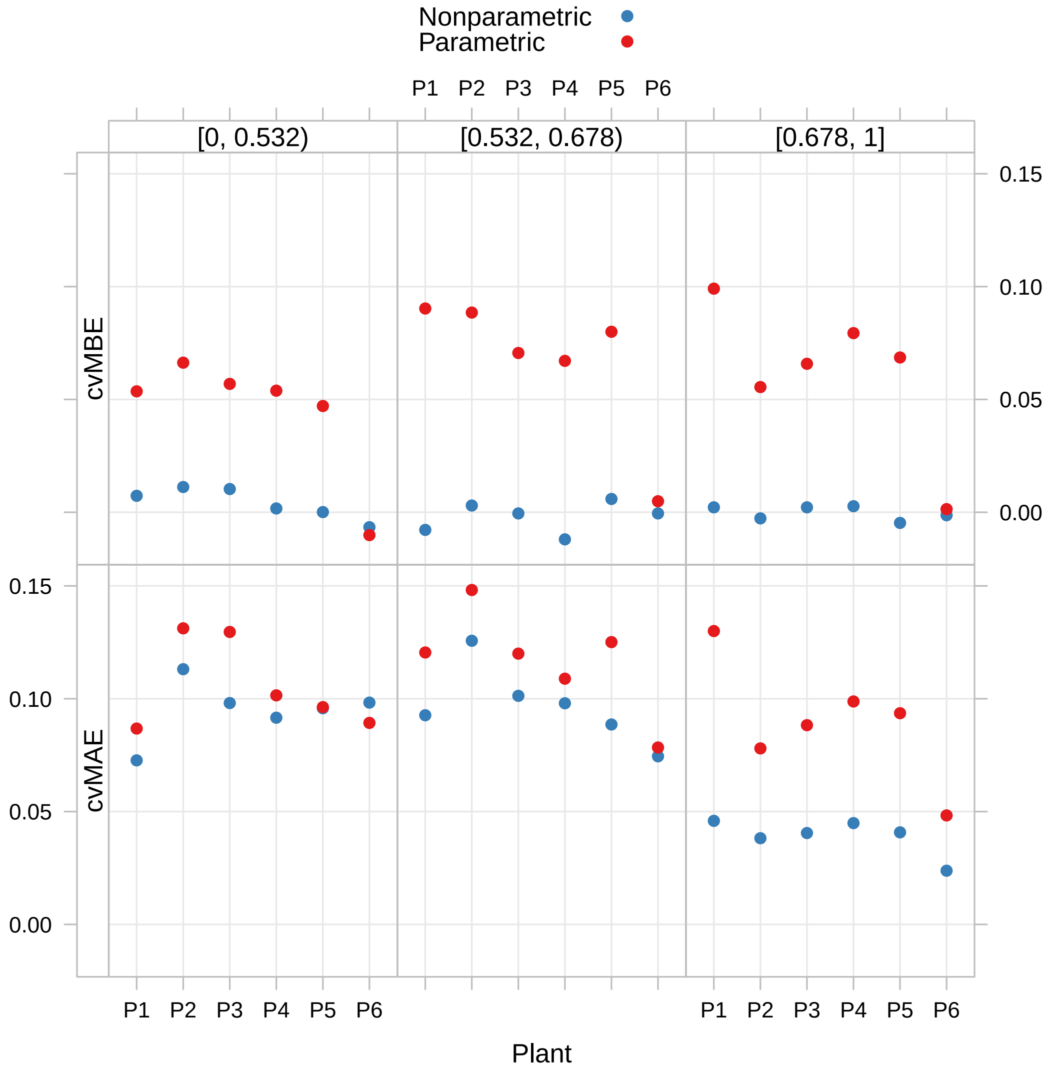

Figure 4: Weighted errors of energy forecast for the procedures with parametric and nonparametric PV models including the six PV plants. Each column corresponds to an interval of KTd.

Finally, if the prediction is analyzed, in particular G_0 from Meteogalicia's WRF model, the $nRMSE$ and $nMAE$ observed (Table 8) suggests that the WRF, and NWP models in general, functions reasonably well when determining whether a day will be sunny or cloudy, yet there is considerable error when predicting the specific times at which these clouds are to appear. In this way, the errors in the actual prediction of the meteorological variables based on NWP models are an important part of the hourly prediction errors of a parametric approach. This is sure to affect, in a similar way, the hourly errors in the nonparametric approach.

Model Step	Variable [Error description]	Statistic	KTd [0 - 0.532)	[0.532 - 0.678)	[0.678 - 1]
Forecast	T_a [T_a - $T_{a,m}$]	$nMBE$	-2.56%	-4.16%	-5.75%
		$nMAE$	6.60%	5.83%	6.01%
		$nRMSE$	8.05%	7.10%	6.76%
	G_0 [G_0 - $G_{0,m}$]	$nMBE$	-9.79%	0.05%	0.27%
		$nMAE$	22.64%	9.92%	2.82%
		$nRMSE$	30.22%	12.49%	3.53%
Operational conditions	G_{eff} [${G_{eff}}^a$ - $G_{eff,m}$]	$nMBE$	1.12%	1.34%	1.54%
		$nMAE$	5.76%	5.83%	7.99%
		$nRMSE$	7.74%	7.98%	12.27%
	T_C [${T_C}^b$ - $T_{C,m}$]	$nMBE$	-4.46%	5.19%	7.41%
		$nMAE$	11.80%	8.25%	7.67%
		$nRMSE$	13.48%	9.54%	8.88%
Electrical models	$E_{out,model}$ [${E_{out,model}}^c$ - $E_{out,m}$]	$nMBE$	0.46%	0.48%	-1.57%
		$nMAE$	3.80%	3.89%	4.63%
		$nRMSE$	5.39%	5.93%	7.63%
Whole Model	$E_{out,forecast}$ [$E_{out,forecast}$ - $E_{out,m}$]	$nMBE$	-3.92%	1.71%	0.27%
		$nMAE$	42.88%	24.46%	9.75%
		$nRMSE$	54.52%	29.32%	14.51%

a Modeled from $G_{0,m}$.

b Modeled from $G_{0,m}$, $T_{a,m}$ and $W_{S,m}$.

c Modeled from $G_{eff,m}$ and $T_{C,m}$.

Table 8: Impact of the errors made at each step of the parametric approach.

10. Post-processing of input data applied to the parametric approach

To improve site-specific forecast accuracy, some authors have proposed statistical methods which correct irradiance forecasts derived from NWP models with actual ground data (Lorenz et al., 2009; Mathiesen and Kleissl, 2011). Those methods are based on MBE corrections as a function of the clearness index and the solar zenith angle. Since post-processing requires actual irradiance data, those methods cannot be used during the initial months of operation. However, the application of statistical post-processing to the NWP irradiance output can improve the performance of PV power forecasts once the PV plant has already been in operation for some time.

In some PV installations there are no or very few valid irradiance measurements available. Such is the case of the first five PV plants analyzed in this paper, which irradiance measurements presented a considerable amount of interruptions. In these cases, the post-processing can be made by means of PV power output. Table 9 shows the weighted errors (quantile $QS_{.5}$) of the parametric approach after the MOS-correction proposed by Lorenz et al (Lorenz et al., 2009) is applied to the power output forecast. Thirty days were used for training data. Data from Tables 7 and 9, for nonparametric and parametric with MOS-correction approaches, respectively, are displayed in Figure 5.

Comparing the results in Tables 6 and 9, it is evident that the addition of MOS-correction into the parametric approach was successful in minimizing $cvMBE$ and reducing $cvMAE$ of daily output forecasts, but the non-parametric approach still performs better than the parametric one.

PV plant	Statistic	$0 \leq KTd < 0.532$	$0.532 \leq KTd < 0.678$	$0.678 \leq KTd < 1$
P1	$cvMBE$	3.79%	0.78%	-4.21%
	$cvMAE$	15.64%	10.94%	7.33%
P2	$cvMBE$	7.53%	-3.64%	-3.00%
	$cvMAE$	21.55%	14.64%	4.75%
P3	$cvMBE$	4.69%	2.26%	-0.11%
	$cvMAE$	19.53%	12.42%	4.54%
P4	$cvMBE$	2.54%	-5.88%	-4.49%
	$cvMAE$	16.39%	13.51%	6.35%
P5	$cvMBE$	0.99%	-2.61%	-3.49%
	$cvMAE$	14.78%	13.65%	5.11%
P6	$cvMBE$	1.52%	-0.65%	-1.20%
	$cvMAE$	11.70%	9.90%	5.19%

Table 9: Weighted errors of energy forecast (quantile $QS_{.5}$) for the procedure with parametric PV model and MOS-correction.

Figure 6 shows the SS of the parametric approach forecasts after the MOS-correction. As expected, the SS improves considerably regarding the previous results without MOS-correction (Figure 3).

11. Conclusions

Both approaches discussed in this paper offer a number of pros and cons, and choosing one depends on the context. The forecast procedure with parametric PV model requires prior information on the system itself which, as a first approach, can be obtained from the data sheets for the various components. More accurate information can be obtained from measurements taken at the plant itself, for example, during commissioning. Given the fact that plant performance can be predicted without the need for production records, the advantage of this approach is that it can be used before the plant is actually constructed or during the initial months of operation. However, given the fact that detailed information on the plant components is not generally available to the system operators, it is of little benefit to them.

Although the nonparametric approach requires no prior information on the plant components, it does need the forecasting and production historical data. For this reason, it cannot be used before the construction of a PV plant or during its initial months of operation. However, once sufficient production records have been generated, these records are generally available to owners and system operators alike.

As far as accuracy is concerned, the nonparametric approach offers better results, particularly in the range of the high and medium clearness indexes, for hourly errors ($cvMAE$) and the total daily error ($cvMBE$) alike. For medium-low KTd (greater cloudiness) the hourly errors for both approaches are similar, although the daily error continues to be greater for the parametric approach. In any case, they both show reasonable accuracy, with hourly errors that are normally under 8% or even less than 5% in the case of the nonparametric approach on clear days.

Most of the performance differences for both forecast procedures are due to the bias shown by the parametric approach and which is generally related to the presence of systematic errors

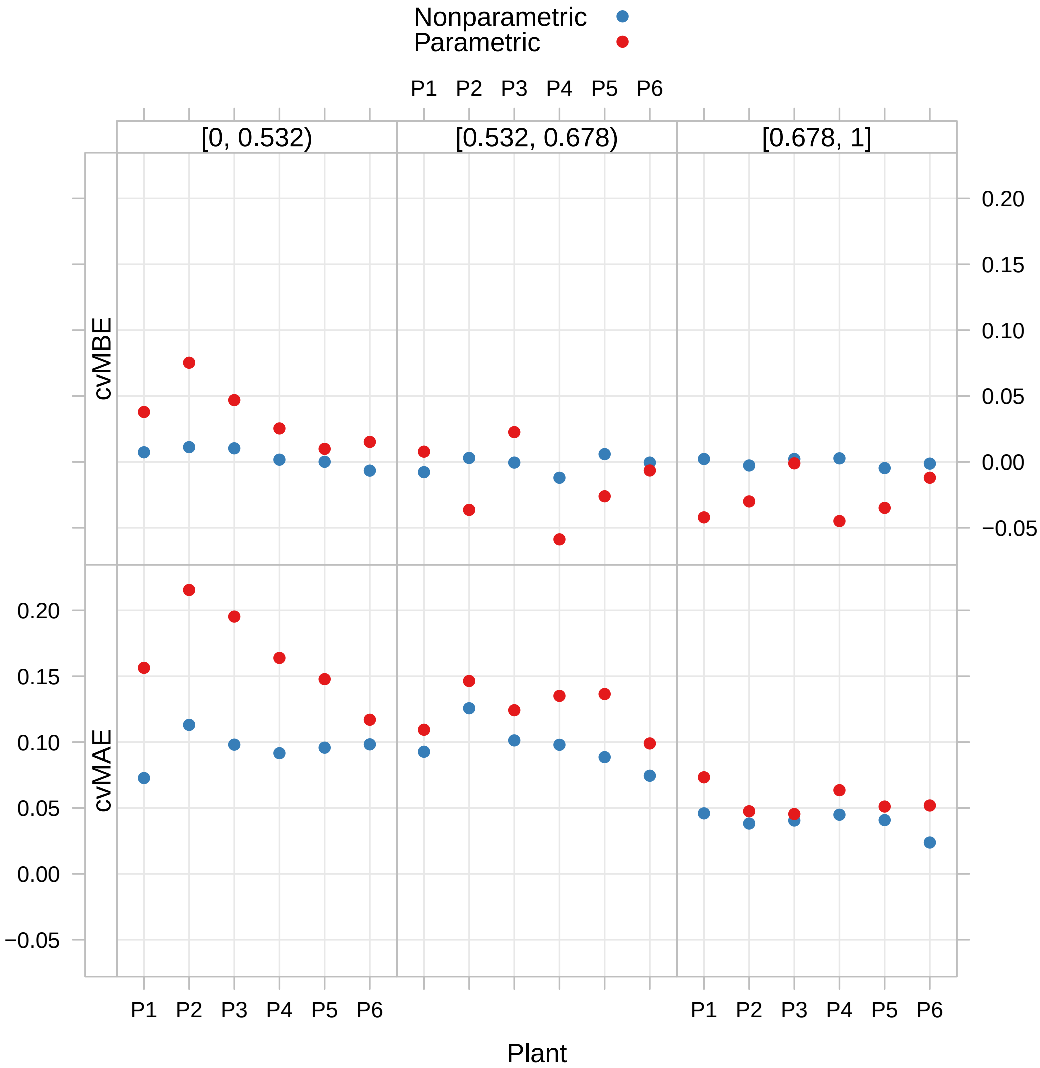

Figure 5: Weighted errors of energy forecast for the procedures with parametric and nonparametric PV models including the six PV plants. Each column corresponds to an interval of KTd and for the parametric approach MOS-correction was used.

(bias) in some of its steps, especially the calculation of G_{eff} from G_0 through empirical correlations FD-KT. Systematic errors are also possible with the electrical model of the plant. For example, if the parameters are not correctly set or if phenomena that could affect the energy performance (auxiliary consumptions, shading, etc.) are not taken into account in the model. However, errors in these parts of the model can be gradually reduced, as more information becomes available on the plant itself and its location. On the other hand, the actual prediction of the meteorological variables used as inputs is generally another source of systematic error. The machine-learning techniques used by the non-parametric approach helps compensate for these input-associated errors.

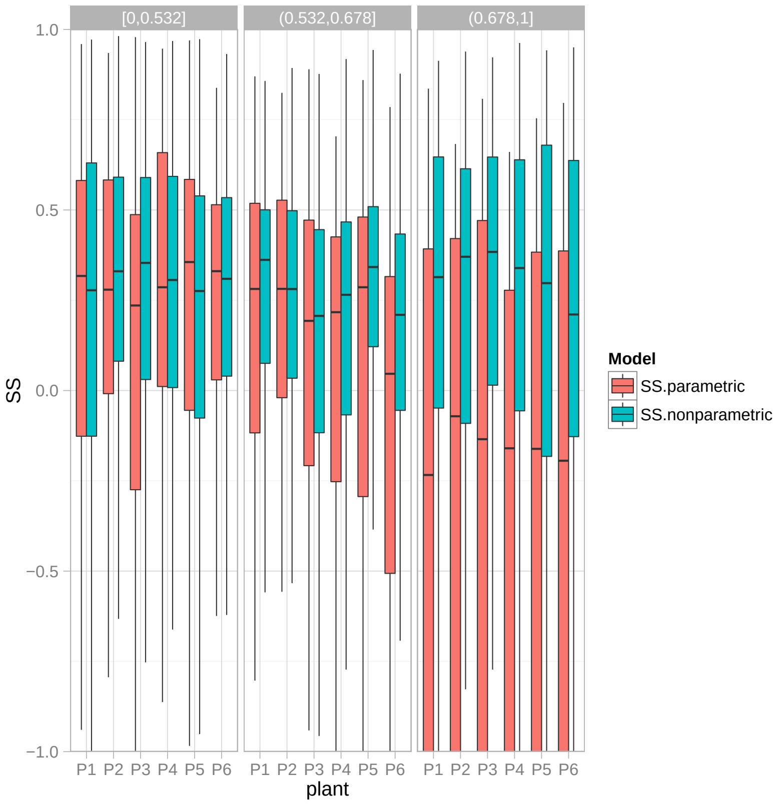

Figure 6: Boxplot of the *SS* for the forecast procedures with parametric and nonparametric PV approaches. Each column corresponds to an interval of *KTd*. Post-processing of input data using MOS-correction was applied to the parametric approach.

Statistical post-processing techniques can be used in order to improve the parametric forecast accuracy. Those methods are based on *MBE* corrections as a function of the clearness index and the solar zenith angle and they require irradiance or production records. The application of these post-processing techniques succeeds in minimizing *cvMBE* and reducing *cvMAE*, but it appears to be less effective than the machine-learning methods used in the non-parametric models.

Acknowledgements

This work has been partially financed by the Seventh Framework Programme of the European Commission with the Project Photovoltaic Cost Reduction, Reliability, Operational Performance, Prediction and Simulation (PVCROPS—Grant Agreement No. 308468).

References

Almeida, M., Perpiñán, O., Narvarte, L., 2015. Pv power forecast using a nonparametric pv model. Solar Energy (115), 354 – 368.

Almeida, M. P., Perpiñán, O., Dec 2014. PVF 0.20.
URL http://dx.doi.org/10.5281/zenodo.13348

Bacher, P., Madsen, H., Nielsen, H. A., 2009. Online short-term solar power forecasting. Solar Energy 83 (10), 1772 – 1783.

Bivand, R., Keitt, T., Rowlingson, B., 2013. rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 0.8-11.
URL http://CRAN.R-project.org/package=rgdal

Breiman, L., Oct. 2001. Random Forests. Machine Learning 45 (1), 5–32.
URL http://oz.berkeley.edu/~breiman/randomforest2001.pdf

Chai, T., Draxler, R. R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)? – arguments against avoiding RMSE in the literature. Geoscientific Model Development 7 (3), 1247–1250.
URL http://www.geosci-model-dev.net/7/1247/2014/

Dowle, M., Short, T., Lianoglou, S., Srinivasan, A., 2014. data.table: Extension of data.frame. R package version 1.9.2.
URL http://CRAN.R-project.org/package=data.table

Engerer, N. A., 2015. Sciencedirect minute resolution estimates of the diffuse fraction of global irradiance for southeastern australia. Solar Energy 116, 215–237.

Erbs, D. G., Klein, S. A., Duffie, J. A., 1982. Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation. Solar Energy 28 (4), 293–302.

Evans, D. L., 1981. Simplified method for predicting photovoltaic array output. Solar Energy 27 (6), 555–560.

Faiman, D., 2008. Assessing the outdoor operating temperature of photovoltaic modules. Progress in Photovoltaics: Research and Applications 16, 307–315.

Fuentes, M., Nofuentes, G., Aguilera, J., Talavera, D. L., Castro, M., 2007. Application and validation of algebraic methods to predict the behavior of crystalline silicon pv modules in mediterranean climates. Solar Energy 81 (11), 1396–1408.

Garcia, M., Vera, J. A., Marroyo, M., Lorenzo, E., Pérez, M., 2009. Solar-tracking pv plants in navarra: A 10 mw assessment. Progress in Photovoltaics: Research and Applications 17, 337–346.

Gueymard, C., 2014. A review of validation methodologies and statistical performance indicators for modeled solar radiation data: Towards a better bankability of solar projects. Renewable and Sustainable Energy Reviews (39), 1024 – 1034.

Hay, J. E., Mckay, D. C., 1960. Estimating solar irradiance on inclined surfaces: A review and assessment of methodologies. International Journal of Solar Energy 4 (3), 203–240.

Hijmans, R. J., 2013. `raster`: Geographic Analysis and Modeling with Raster Data. R package version 2.1-66.
URL http://CRAN.R-project.org/package=raster

Inman, R. H., Pedro, H. T., Coimbra, C. F., Dec. 2013. Solar forecasting methods for renewable energy integration. Progress in Energy and Combustion Science 39 (6), 535–576.

Jantsch, M., Schmidt, H., Schmid, J., 1992. Results on the concerted action on power conditioning and control. In: 11th European Photovoltaic Solar Energy Conference. pp. 1589–1592.

Lorenz, E., Hurka, J., Heinemann, D., Beyer, H. G., 2009. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 2, 2–10.

Lorenz, E., Scheidsteger, T., Hurka, J., Heinemann, D., Kurz, C., Nov. 2011. Regional PV power prediction for improved grid integration. Progress in Photovoltaics: Research and Applications 19 (7), 757–771.

Lorenzo, E., 2011. Energy collected and delivered by PV modules. Ch. 22, pp. 984–1042.

Mandal, P., Madhira, S. T. S., haque, A. U., Meng, J., Pineda, R. L., 2012. Forecasting power output of solar photovoltaic system using wavelet transform and artificial intelligence techniques. Procedia Computer Science 12 (0), 332 – 337, complex Adaptive Systems 2012.

Marcos, J., Marroyo, L., Lorenzo, E., Alvira, D., Izco, E., 2011. Power output fluctuations in large scale PV plants: One year observations with 1 second resolution and a derived analytic model. Progress in Photovoltaics: Research and Applications 19, 218–227.
URL http://138.4.46.62:8080/ies/ficheros/2_52_ref14.pdf

Martin, N., Ruiz, J. M., 2001. Calculation of the pv modules angular losses under field conditions by means of an analytical model. Solar Energy Materials and Solar Cells (70), 25–38.

Martínez, F., 2012. Caracterización y modelado de grandes centrales fotovoltaicas. Ph.D. thesis, Universidad Politécnica de Madrid.

Martínez-Moreno, F., Lorenzo, E., Muñoz, J., Moretón, R., 2011. On the testing of large pv arrays. Progress in Photovoltaics: Research and Applications 20, 100–105.

Mathiesen, P., Kleissl, J., 2011. Evaluation of numerical weather prediction for intra-day solar forecasting in the continental united states. Solar Energy 85, 967–977.

Meinshausen, N., 2006. Quantile regression forests. The Journal of Machine Learning Research 7, 983–999.
URL http://www.stats.ox.ac.uk/~meinshau/quantregforests.pdf

Moretón, R., 2016. Contributions to uncertainty reduction in the estimation of pv plants performance. Ph.D. thesis, Universidad Politécnica de Madrid.

Murphy, A. H., 1988. Skill scores based on the mean square error and their relationships to the correlation coefficient. Monthly weather review 116 (12), 2417–2424.

Osterwald, C. R., 1986. Translation of device performance measurements to reference conditions. Solar Cells 3 (18), 269–279.

Pebesma, E. J., 2004. Multivariable geostatistics in S: The `gstat` package. Computers and Geosciences 30, 683–691.

Pedro, H. T., Coimbra, C. F., 2012. Assessment of forecasting techniques for solar power production with no exogenous inputs. Solar Energy 86 (7), 2017–2028.

Pelland, S., Galanis, G., Kallos, G., 2013. Solar and photovoltaic forecasting through postprocessing of the global environmental multiscale numerical weather prediction model. Progress in Photovoltaics: Research and Applications 21 (3), 284–296.

Perez, R. R., Ineiche, P., Maxwell, E. L., Seals, R. D., A., Z., 1992. Dynamic global-to-direct irradiance conversion models. ASHRAE Transactions. 98 (1), 354–369.

Perpiñán, O., 2012. solaR: Solar radiation and photovoltaic systems with R. Journal of Statistical Software 50 (9), 1–32.
URL http://www.jstatsoft.org/v50/i09/

Perpiñán, O., Almeida, M. P., Jan 2015. meteoForecast 0.44.
URL http://dx.doi.org/10.5281/zenodo.13882

R Development Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
URL http://www.R-project.org

Ryan, J. A., Ulrich, J. M., 2013. xts: eXtensible Time Series. R package version 0.9-5.
URL http://CRAN.R-project.org/package=xts

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W., Powers, J. G., 2005. A description of the advanced research wrf version 2. Tech. rep., National Center for Atmospheric Research.
URL http://www2.mmm.ucar.edu/wrf/users/docs/arw_v2.pdf

Wilks, D., 2011. Statistical Methods in the Atmospheric Sciences. Academic Press.

Willmott, C. J., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research 30, 79–82.
URL http://www.int-res.com/abstracts/cr/v30/n1/p79-82/

Zamo, M., Mestre, O., Arbogast, P., Pannekoucke, O., 2014. A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production, part I: Deterministic forecast of hourly production. Solar Energy 105, 792–803.

Zeileis, A., Grothendieck, G., 2005. zoo: S3 infrastructure for regular and irregular time series. Journal of Statistical Software 14 (6), 1–27.
URL http://www.jstatsoft.org/v14/i06/