

PV Power Forecast Using a Nonparametric PV Model

M. Pinho Almeida^{a,*}, O. Perpiñán Lamigueiro^{b,c}, L. Narvarte^c

^a*Instituto de Energia e Ambiente, Universidade de São Paulo, São Paulo, Brazil*

^b*Electrical Engineering Department, ETSIDI-UPM, Ronda de Valencia 3, 28012 Madrid, Spain.*

^c*Instituto de Energía Solar, Ciudad Universitaria s/n, Madrid, Spain*

Abstract

Forecasting the AC power output of a PV plant accurately is important both for plant owners and electric system operators. Two main categories of PV modeling are available: the parametric and the nonparametric. In this paper, a methodology using a nonparametric PV model is proposed, using as inputs several forecasts of meteorological variables from a Numerical Weather Forecast model, and actual AC power measurements of PV plants. The methodology was built upon the R environment and uses Quantile Regression Forests as machine learning tool to forecast AC power with a confidence interval. Real data from five PV plants was used to validate the methodology, and results show that daily production is predicted with an absolute *cv*MBE lower than 1.3%.

Keywords:

PV plant, Numerical Weather Prediction, Weather Research and Forecasting, PV power forecast, Random Forest, Quantile Regression

1. Introduction

Plant owners and electric system operators can both benefit from an accurate AC power output forecast of PV plants. The first group needs to plan their energy commercialization in advance and the latter has to program the dispatch of the electric system.

A PV plant can be seen as a box with several inputs (irradiation, temperature and wind speed, for example) and one output, the AC power injected into the electrical grid. Two main types of modeling can be employed to estimate the AC power output given the required inputs:

- The parametric model, that conceives the PV system as a white box where each subsystem can be modeled using a collection of parameters.
- The nonparametric model, that conceives the PV system as a black box. This approach does not presume any knowledge of internal characteristics and processes of the system. Instead, it is a data-driven model that estimates the behavior of the system from a historical time series of inputs and outputs.

Both approaches may share the same inputs (irradiation, temperature, etc.) and, of course, the same output, AC power in the context of this paper.

The parametric approach relies on a set of sub-models to compute AC power injected into the electrical grid, namely:

*Corresponding author

Email addresses: marcelopa@iee.usp.br (M. Pinho Almeida), oscar.perpinan@upm.es (O. Perpiñán Lamigueiro)

- Decomposition model that estimates diffuse and beam components with the global irradiance on the horizontal plane as the input.
- Transposition and shading models that estimate the effective irradiance on the generator plane with the diffuse and beam irradiances on the horizontal plane as the input.
- PV generator model that estimates DC power with the effective irradiance on the generator plane and the ambient temperature as the inputs.
- Inverter model that estimates AC power with the DC power as the input.
- Wiring and electrical equipment models that estimate Joule and conversion losses in the way from the PV generator and inverter to the electrical grid.

This approach requires detailed information about the characteristics and the behavior of each relevant component of the PV plant. Because this information is not always available some simplifications and assumptions are needed, with the subsequent uncertainty in the output. Consequently, the accuracy and precision of the estimations of a parametric model is driven by the performance of its component models and the accuracy of the measurements or estimations of parameters. However, its main advantage is the possibility to compute the AC power output prior the construction of a PV plant, during the project and planning stages.

This scheme has been adopted in recent research to forecast PV power. (Lorenz et al., 2011) derives solar irradiance forecasts based on forecasts of the global model of the European Centre for Medium-Range Forecasts (ECMWF) with a post-processing procedure, and PV power is derived with a set of physical modeling steps. (Pelland et al., 2013) uses photovoltaic simulation models with PV forecasts based on 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.

On the other hand, nonparametric models circumvent the need for simplifying assumptions and accurate internal parameters with the use of historical time series of meteorological variables and AC power measurements. Therefore, the accuracy of a nonparametric model depends mainly on the quality of the data. To illustrate how this feature could be useful, let's suppose an electric system operator needs estimations of future generation of a PV plant, but he does not know anything about the plant, not even its nominal peak power. As system operators normally have access to the records of power output of generation plants, this data could be used to solve this problem. However, this characteristic also leads to its main disadvantage: the PV plant must exist and be operational for some time.

One interesting advantage of a nonparametric model is the potential to compensate systematic errors associated to the inputs. For example, if irradiance data has a consistent error, the model will learn to associate the incorrect irradiance with the correct AC power output value during the training process. When supplied with new data from the same source, the output will not be compromised if the same error persists.

The nonparametric approach has been implemented in several recent researches. (Bacher et al., 2009) forecasts hourly values of AC power of PV systems 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. (Mandal et al., 2012) forecasts one-hour-ahead power output of a PV system using a combination of wavelet transform and neural network techniques by incorporating the interactions of PV system with solar radiation and temperature data. (Pedro and Coimbra, 2012) predicts 1 and 2 h-ahead solar power of a PV system comparing several forecasting techniques without exogenous inputs such as Auto-Regressive Integrated Moving Average, k-Nearest-Neighbors,

Artificial Neural Networks, and Neural Networks optimized by Genetic Algorithms. (Zamo et al., 2014) analyzes a mix of eight statistical methods to forecast PV power one day ahead in an hourly basis, and the Random Forests method presents the best results.

This paper proposes a methodology to derive AC power forecasts one day ahead with hourly resolution using a nonparametric PV model based on Quantile Regression Forests. It improves the current state of art with several contributions. First, both a single-valued forecast and a probabilistic forecast are produced, providing statistical information about the uncertainty of the output. Moreover, several variability indexes derived from the original variables are proposed, and a systematic and exhaustive variable importance analysis is carried out with different scenarios. Finally, both the length of the time series used to learn from data, and the method for selecting the days included in this training time series are analyzed regarding the model performance. Its methodology is as follows:

- Previous AC power measurements from a PV plant are collected. (Section 2)
- 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. (Section 3)
- 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. (Section 3)
- 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). (Section 4)
- A machine learning tool (Quantile Regression Forests) is trained with the train time series. (Section 4)
- Predictions of the median (quantile 0.5) and the confidence interval (quantiles 0.1 and 0.9) for the AC power are generated with the test time series. (Section 4)

The methodology has been validated by comparing the predictions with measured AC power from several PV plants, as described in Section 5. The validation results are presented in Section 6.

2. PV Plants Database

The database of AC power time series used in this paper is the result of a measurements campaign carried out in five PV plants situated in northern Spain (latitude 42.2°), with a 5-s resolution, previously analyzed in (Marcos et al., 2011).

The monitoring systems of the PV plants record operational and meteorological data including AC power, in-plane irradiance, ambient temperature and wind speed. Measurements are synchronized with a GPS. Data recording started on April 17th, 2008, but this paper is restricted to the period comprised between January 1st, 2009 and December 29th, 2010. Moreover, in order to reduce file sizes and to filter noise, the raw data has been aggregated to produce 1-min records. This recorded data was then aggregated into 1-hour values to be used in the present study.

Table 1 summarizes the main characteristics of these plants. Their installed power ranges from 2.64 MWp to 958 kWp, with areas ranging from 11.8 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. These plants use vertical axis trackers.

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

Table 1: PV plants characteristics.

3. Weather Research and Forecasting variables

Weather forecasting is the science of predicting the state of the atmosphere in the future at some location. Numerical Weather Prediction, NWP, uses mathematical models of the atmosphere and oceans to predict the weather based on current conditions. The domain of a model is either global, covering the entire Earth, or regional, covering only part of the Earth. Regional models use finer grid spacing than global models because computational resources are focused on a specific area. Thus, regional models are able to resolve smaller-scale meteorological phenomena.

The methods proposed in this paper use a regional mesoscale¹ NWP model called Weather Research and Forecasting, WRF (Skamarock et al., 2005).

Meteogalicia, a meteorological institute of the Xunta de Galicia (Spain), publishes regularly results from a WRF model freely available at its Thredds server². The WRF model runs twice a day, initialized at 00UTC (forecast for the next 96 hours) and 12UTC (forecast for the next 84 hours). The spatial resolution is 12 km \times 12 km, in an area comprised between 21.58°W to 6.36°E and 33.64°N to 49.57°N, and the time resolution is hourly. Meteogalicia maintains an historical archive of past forecasts that is available on-line.

The model output includes a wide collection of variables, such as solar radiation, temperature or cloud cover³. 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 2 presents the name and description of the WRF variables considered in the predictor set of this paper.

For each WRF variable, both its value at the location of interest and a spatially interpolated value are used. The interpolated value is computed using Inverse Distance Weighting for a central cell (the location of interest) from its 24 surrounding cells in a 5 by 5 grid. The value of a certain variable at the location of interest will be denoted with the name of the variable modified with the suffix `.point`, while the interpolated value will use the suffix `.idw`.

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 ($Bo0$). These three variables are computed only for the location of interest.

On the other hand, there is additional information related with the spatial and time evolution of the raster data. An attempt to take advantage of this information is a set of spatial and

¹Mesoscale 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>

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

Label	Description
swflx	surface downwelling shortwave flux
temp	temperature at 2 m
cfh	cloud cover at high levels
cfl	cloud cover at low levels
cfm	cloud cover at mid levels
cft	cloud cover at low and mid levels
u	longitude-wind at 10 m
v	latitude-wind at 10 m
mod	wind speed module at 10 m
dir	wind direction at 10 m
rh	relative humidity at 2 m
mslp	mean sea level pressure
visibility	visibility in air

Table 2: WRF-NWP variables used to forecast AC power.

time variability indexes described in next section.

3.1. Variability indexes

The chaotic nature of the atmosphere imposes an uncertainty to the forecasts produced by a NWP model. Figure 1 displays WRF-NWP forecasts of global horizontal irradiance for several nearby locations around the PV plant P1. Figure 2 displays the forecast of solar irradiance produced by several consecutive model runs for the location of the PV plant P1. During clear sky or completely cloudy days, forecasts of different model runs and at different nearby locations are similar and coherent among them. However, during partially cloudy days, forecasts vary both spatially (different locations) and temporally (different model runs).

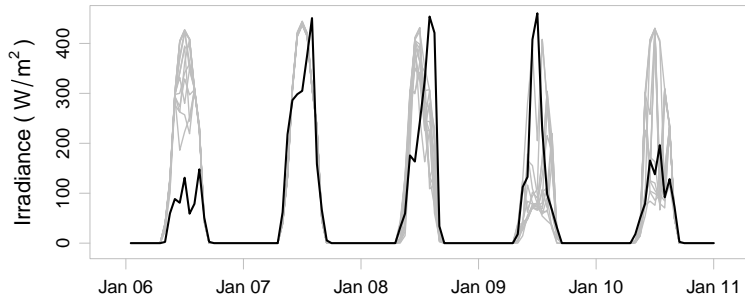


Figure 1: Hourly global horizontal irradiance WRF-NWP forecasts for several nearby locations around the PV plant P1 for the year 2009, and comparison with on-ground measurements (dark line).

Our proposal to deal with this inherent uncertainty consists in taking into account the forecasts for nearby locations and consecutive model runs for the point of interest through additional spatial and time variability indexes derived from the original WRF outputs. These indexes are calculated for each WRF variable.

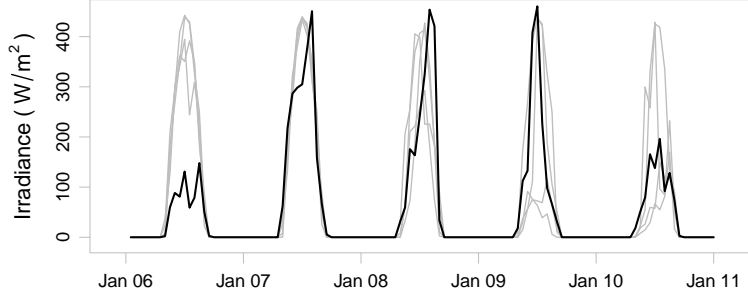


Figure 2: Hourly global horizontal irradiance WRF-NWP forecasts produced by several consecutive model runs for the location of the PV plant P1 for the year 2009, and comparison with on-ground measurements (dark line).

To quantify the spatial variability of raster data, three indexes have been selected: *TRI*, *TPI* and *roughness* (Wilson et al., 2007). These indexes refer to a 3 by 3 grid on the terrain, where 8 cells surround a central cell. In this paper, each cell is a 12 km × 12 km square from the WRF output. In the following equations, the central cell, which corresponds to the location of a PV plant, is denoted with c_0 , and the surrounding cells are denoted with c_i .

- *TRI*: Terrain Ruggedness Index, is defined as the mean of the absolute differences between a central cell and its surrounding 8 cells in a 3 by 3 grid.

$$TRI = \frac{1}{8} \sum_{i=1}^8 |c_i - c_0| \quad (1)$$

- *TPI*: Topographic Position Index, is defined as the difference between a central cell and the mean of its surrounding 8 cells in a 3 by 3 grid.

$$TPI = c_0 - \frac{1}{8} \sum_{i=1}^8 c_i \quad (2)$$

- *Roughness*: is the largest inter-cell difference of a central cell and its surrounding 8 cells in a 3 by 3 grid.

$$rough = \max \{c_i : i = 1 \dots 8\} - \min \{c_i : i = 1 \dots 8\} \quad (3)$$

With regard to time variability, we take advantage on the fact that any WRF variable, on a specified hour, will have 7 (before 12UTC) or 8 (after 12 UTC) past forecasts. For example, considering an irradiance forecast for today at 16UTC, this hour was included in the forecasts produced at 00UTC and 12UTC three days ago, two days ago, yesterday, and today. An initial hypothesis that can be made is that forecasts closer to the hour of interest would be more reliable. In order to investigate this behavior, the errors between measured and predicted global horizontal irradiance for each model run of Meteogalicia forecasts were calculated at the locations of interest. However, at least for the Meteogalicia NWP-WRF model, it was not possible to find a consistent relation between forecasts errors and time distance between forecast and actual event, as can be seen in Figure 3. This figure shows the solar irradiance measurements for three

consecutive days of the dataset in the upper panel, and the error between these measurements and the NWP-WRF forecast in the lower panel. It is clear that errors are higher during cloudy days, but there is not a consistent relation between the model run and the error magnitude. Thus, considering that every model run for a specified hour has the same level of reliability, the standard deviation of this collection of forecasts (*sdr*) can be used as an index to quantify the time variability of a WRF variable.

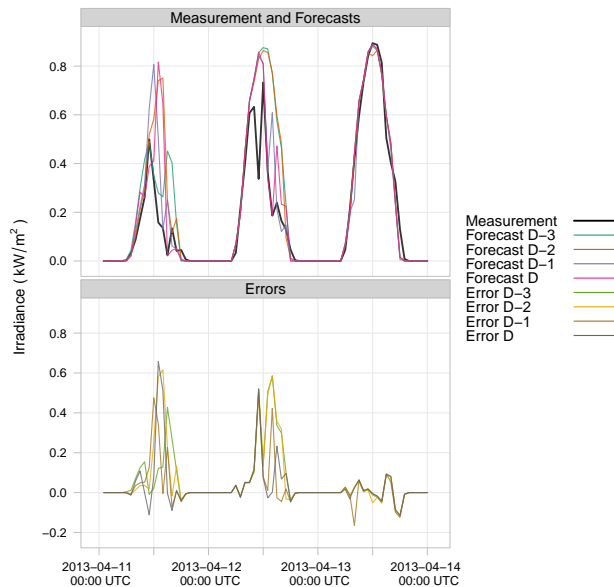


Figure 3: Errors between measured and forecasted global horizontal irradiance. The upper panel displays the solar irradiance measurements and the NWP-WRF forecasts. The lower panel displays the errors between these measurements and the respective NWP-WRF forecasts. Each line of forecast and error corresponds to a different model run. For example, D-3 is the forecast or the error related to the forecast published 3 days ago (at 00 UTC), and so on.

Finally, Figure 4 displays the complete set of predictors extracted from the WRF variables for one day (sun geometry variables, and WRF variables with their variability indexes).

4. Implementation of the Proposed Method

The AC power measurements time series, and the collection of predictors, are merged together to build a dataset, that is a multivariate hourly time series. This dataset is divided into two time series. Machine learning nomenclature denotes these two data sets as training and test sets. The training time series comprises several days of past values of AC power measurements and predictors, and it is used to train the machine learning tool. The test time series includes the hourly values of the predictors for the day ahead.

The training time series, or training set, may have different lengths, what eventually leads to different results. For practical purposes, the length of the training set is defined here in days. Therefore, the training set can be composed by N days, selected from a larger database. Three selecting methods were used in our model:

- *Previous*

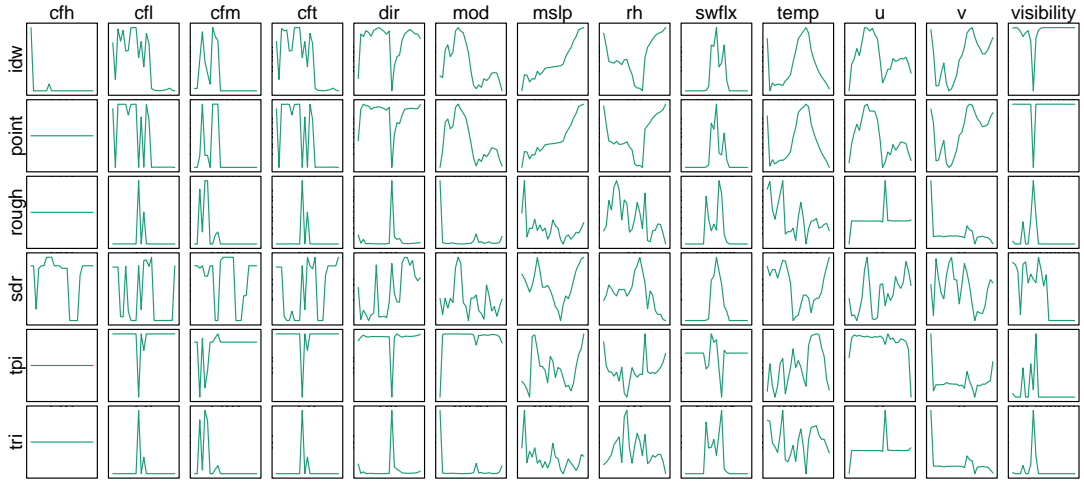


Figure 4: Set of predictors comprising one day for PV plant P1. Each column represents a different WRF variable, and each row corresponds to the point and interpolated time series, or to each variability index. See Table 2 for information about the denomination of each WRF variable.

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 daily clearness index of the day to be predicted and the daily 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. Therefore, the database does not need to be complete up to the day prior the prediction. For example, the method could work with a dataset built some time ago provided that the performance of the PV plant remains stable.

- *KS*

This method selects N days according to the similarity between the empirical distribution function of the intradaily irradiance forecast for the day to be predicted and the empirical distribution function of the intradaily 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. As in the method *KT*, the database does not need to be complete up to the day prior the prediction.

After training, the machine learning tool returns the predictions of the quantile $Q_{.5}$ (median) and a confidence interval within the quantiles $Q_{.1}$ and $Q_{.9}$, i.e., a 80% prediction interval, using the test time series.

The machine learning tool used in the proposed methodology is based on Random Forests (Breiman, 2001), which consists of a collection, or ensemble, of a multitude of decision trees, each one built from a sample drawn with replacement (a bootstrap sample) from a training set (X_i, Y_i) , $i = 1, 2, \dots$, where X is the group of inputs and Y is the group of outputs. In addition,

when splitting a node during the construction of a tree, only a random subset of variables is used. As a consequence, the final nodes, or leafs, may contain one or several observations from Y . For regression problems, each tree is capable of producing a response y when presented with a set of predictors x , being y the conditional mean (approximated by a weighted mean) of the observations from Y present on the resulting leaf. As a result of the random construction of the trees, the bias of the forest usually slightly increases (with respect to the bias of a single non-random tree) but, due to the averaging, its variance also decreases, usually more than compensating for the increase in bias, hence yielding an overall better model. Finally, the responses of all trees are also averaged to obtain a single response variable for the model, and again weighted mean is used.

However, the group of observations of Y in a result leaf can also be used to produce a good estimation of the full conditional distribution, which is known as Quantile Regression Forests. Standard regression analysis tries to estimate the conditional mean of a response variable, which focuses on only one aspect of its conditional distribution, but neglects all other features. On the other hand, the quantiles give more information about the distribution of Y as a function of the predictors than the conditional mean alone. While Random Forests keeps only an average of the observations of Y that fall into each leaf of each tree and neglects all other information, Quantile Regression Forests keeps the values of all observations of Y in every leaf, not just their weighted mean, and assesses the conditional distribution based on this information, enabling the construction of prediction intervals (Meinshausen, 2006). For example, a 80% prediction interval for y is given by $[Q_{.1}(x), Q_{.9}(x)]$.

The simplified algorithm of Quantile Regression Forests is as follows:

- Grow trees as in Random Forests.
- For a given x , drop it down all trees. However, for every leaf of every tree, register all observations from Y in this leaf, not just their average.
- Compute the weights as in Random Forests.
- Compute the estimate of the distribution function for all responses with the weights from previous step.

It is important to state that, due to the restriction imposed by the WRF variables from Meteorological, this paper uses an hourly time basis, both for predictors and predictions. Figure 5 shows an example of simulation using the proposed methodology with $N = 30$ days and selecting method *KS*.

4.1. Software

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).
- *solr* for sun geometry calculation (Perpiñán, 2012).

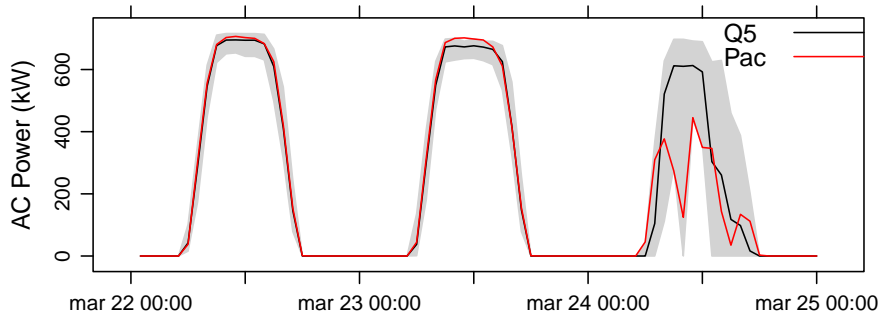


Figure 5: Example of simulation with $N = 30$ days and selecting method KS .

- `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).

5. Validation procedures

A model performance is commonly evaluated by quantifying the discrepancy between forecasts and actual observations through the use of different statistics such as the Mean Bias Error or the Root Mean Square Error (Gueymard, 2014). Because each performance statistic characterizes a certain aspect of the overall model performance, a complete evaluation needs the combined use of a collection of statistics tools. Section 5.2 describes the performance statistics used in this paper. On the other hand, the proposed methodology produces both the forecast of the median and the forecast of the quantiles $Q_{.1}$ and $Q_{.9}$, which must also be assessed. Section 5.3 describes the performance statistics used to evaluate the confidence interval.

The performance of the proposed model has been measured using a leave-one-out cross-validation procedure:

- One day is extracted from the database to be the test set.
- The training set is constructed with N days extracted from the remaining days of the data set, according to the selecting method (KS , KT , or *Previous*). This training set is used to train the QRF.
- The trained QRF is used to predict AC power, with hourly quantiles $Q_{.1}$, $Q_{.5}$ and $Q_{.9}$, for the test set.
- The error between these predictions and AC power measurements for the test day is characterized with the model performance statistics described in sections 5.2 and 5.3.

This procedure is repeated for every day in the dataset (over 600 days) resulting in a large collection of performance statistics. For ease of understanding, the results of each performance statistic have been aggregated with the quantiles 0.25, 0.5 and 0.75, hereafter denominated $QS_{.25}$, $QS_{.5}$ and $QS_{.75}$, respectively, to distinguish them from the quantiles of the predictions.

Moreover, the model performance has been compared with a persistence method commonly used as reference in forecast problems related with PV generation. This comparison has been measured with the skill score defined in section 5.4.

5.1. Scenarios

The full picture of the strengths and weakness of a complex model is only grasped when its performance is evaluated under different conditions. The proposed methodology has been analyzed varying the predictors included in the training and test time series.

A variable importance analysis observes the output to determine which variables statistically significantly contribute to the model. The standard approach to measure the importance of each predictor with Random Forests (Breiman, 2001) consists on random permutations of a certain predictor among the training data. The error is again computed on this perturbed data set, and the importance score for that predictor is computed by averaging the difference in error before and after the permutation over all trees. Predictors which produce large values for this score are ranked as more important than predictors which produce small values.

We have followed a different approach with the objective of a better understanding of the model and the interactions between variables. We have not tested blindly the full space of predictors combinations. Instead, 17 scenarios have been defined to analyze the model performance, where a scenario is defined as a collection of certain predictors conveniently chosen to test their relative importance in the model performance. These scenarios have been divided into four groups, described as follows:

- Group 1
Scenarios include all WRF variables from Meteogalicia listed on Table 2 and all the spatial and time indexes.
- Group 2
Scenarios do not include the time index and the spatial indexes, but keep the forecast for the location and the interpolated value. Besides, these scenarios do not include the whole set of WRF variables listed on Table 2, but all contain irradiance forecast.
- Group 3
Scenarios do not include the time index and the spatial indexes, but keep the forecast for the location and the interpolated value. Besides, these scenarios do not include the whole set of WRF variables listed on Table 2, and none contain irradiance forecast.
- Group 4
Scenarios were selected from group 2 to include calculated hourly clearness index (kt).

Each scenario has been subjected to 15 different cross-validations changing the number of days included in the training set ($N = 7, 15, 30, 45$ or 60 days) and the selecting method (*Previous*, *KS* and *KT*). Consequently, a total of 255 cross-validations per each PV plant described in Section 2 have been performed. Table 3 summarizes the characteristics of all the scenarios.

5.2. Performance Statistics for the Quantile $Q_{.5}$

Forecast performance measures the differences, $\mathbf{D} = \{d_i : i = 1 \dots n\}$, between forecasts, $\mathbf{F} = \{f_i : i = 1 \dots n\}$, and observations, $\mathbf{O} = \{o_i : i = 1 \dots n\}$. Thus,

Group		G1				G2					G3					G4			
Scenario		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
WRF variables (<i>point</i> and <i>IDW</i>)	<i>swflx</i>	×	×	×	×	×	×	×	×	×	×						×	×	
	<i>temp</i>	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	
	<i>cft</i>	×	×	×	×	×	×					×	×		×				
	<i>cfl</i>	×	×	×	×							×	×						
	<i>cfm</i>	×	×	×	×							×	×						
	<i>cfh</i>	×	×	×	×							×	×						
	<i>u</i>	×	×	×	×	×	×												
	<i>v</i>	×	×	×	×	×	×												
	<i>mod</i>	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×		×	×
	<i>dir</i>	×	×	×	×	×	×	×	×										
	<i>rh</i>	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×		×	×
	<i>visibility</i>	×	×	×	×	×	×	×	×										
	<i>mslp</i>	×	×	×	×	×	×	×	×										
	Sun-Earth Geometry	<i>ALS</i>	×		×		×		×		×		×		×	×	×	×	×
<i>AzS</i>		×		×		×		×		×		×		×	×	×	×	×	
<i>Bo0</i>		×		×		×		×		×		×		×	×	×	×	×	
Indexes (for WRF variables)	<i>TRI</i>	×	×																
	<i>TPI</i>	×	×																
	<i>rough</i>	×	×																
	<i>sdr</i>	×	×																
Clearness Index	<i>kt</i>																×	×	

Table 3: A total of 17 scenarios are defined to analyze the model performance. They differ on the variables and indexes used as input.

$$\mathbf{D} = \mathbf{F} - \mathbf{O} \quad (4)$$

$$= \{d_i = f_i - o_i : i = 1 \dots n\} \quad (5)$$

where d_i , f_i and o_i are the elements of the vectors \mathbf{D} , \mathbf{F} and \mathbf{O} , respectively.

There is a variety of forecast performance metrics, each of them providing a projection of the forecast errors and, therefore, highlighting a certain aspect of the error characteristics. Therefore, a combination of metrics is often required to assess the forecast performance. The Mean Bias Error (*MBE*, Equation 6) and the Root Mean Square Error (*RMSE*, Equation 7) are the most widely reported.

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

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

where the upper line denotes average. It is important to note that negative and zero values of observations are not considered in these metrics. Therefore, n is the number of hours when either the observations or the model output are strictly positive.

The *RMSE* aggregates information both of 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 measure, such as the *MBE*, that allows discerning between average and variance. This property will be exploited in section 6.1 with the Target Diagram, a graphical tool to compare the performance of several models.

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

$$RMSE^2 = \sigma_{\mathbf{D}}^2 + \bar{\mathbf{D}}^2 = \sigma_{\mathbf{D}}^2 + MBE^2 \quad (9)$$

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).

$$MAE = |\bar{\mathbf{D}}| = \frac{1}{n} \sum_{i=1}^n |d_i| = \frac{1}{n} \sum_{i=1}^n |f_i - o_i| \quad (10)$$

As previously stated, the *RMSE* is more widely reported, although the *MAE* is useful to unveil the occurrence of outliers. Thus, the validation results included in the section 6 use both of them to quantify the forecast performance regarding the median.

5.3. Performance Statistics for the Confidence Interval

The proposed methodology outputs a confidence interval comprised between the quantiles $Q_{.1}$ and $Q_{.9}$. Therefore, the accuracy of this confidence interval, as well as its amplitude, must be quantified in order to complement the model performance assessment. Three performance statistics are proposed for this purpose, one concerning its amplitude and two regarding its accuracy, i.e., the goodness of containing all observations within.

To assess the amplitude of the confidence interval, its area, normalized to the area of the observations, $Q1Q9_{Sum}$, is calculated with Equation 11. This statistic gives information on how wide the interval is, as well as how many times the area (or energy) inside the interval is larger than the area comprised under the observed power curve, so greater values of $Q1Q9_{Sum}$ means more uncertainty related to the quantile $Q_{.5}$.

$$Q1Q9_{Sum} = \frac{\sum_{i=1}^n (Q_{.9i} - Q_{.1i})}{\sum_{i=1}^n o_i} \quad (11)$$

Once again, n is the number of hours when either the observations or the model output are strictly positive.

The performance statistic for the accuracy of quantile $Q_{.1}$ considers only the instants (or hours, specifically for this paper) when the observed value is smaller than the quantile. First, these instants are identified using Equation 12, resulting in the vector $Q1u$ (the numbers 0 and 1 were arbitrarily chosen, but simplify the next step). The sum of all elements of $Q1u$ results in the number of events when the observed value is below $Q_{.1}$ in the period considered, resulting in $Q1_{Num}$ (Equation 13).

$$Q1u_i = \begin{cases} 1 & Q_{.1i} - o_i > 0 \\ 0 & Q_{.1i} - o_i \leq 0 \end{cases} \quad \text{where } i = 1, 2, \dots, n \quad (12)$$

$$Q1_{Num} = \sum_{i=1}^n Q1u_i \quad (13)$$

A similar approach is used to compute the statistic for the accuracy of quantile $Q_{.9}$, but now considering the instants when the observations are higher than the quantile, as described in Equations 14 and 15.

$$Q9u_i = \begin{cases} 1 & o_i - Q_{.9i} > 0 \\ 0 & o_i - Q_{.9i} \leq 0 \end{cases} \quad \text{where } i = 1, 2, \dots, n \quad (14)$$

$$Q9_{Num} = \sum_{i=1}^n Q9u_i \quad (15)$$

5.4. Skill Score

Skill scores 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 reference. A commonly used reference is the persistence forecast, the most used method of forecasting AC power in photovoltaics, which 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 Skill Score is defined as (Murphy, 1988):

$$SS = 1 - \frac{RMSE_f}{RMSE_p} \quad (16)$$

where $RMSE_f$ and $RMSE_p$ are the Root Mean Square Error of the forecast method and of the persistence, respectively. $SS > 0$ ($SS < 0$) 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 skill score 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 skill scores up to 0.36 for the forecast of AC power for next day horizons. It must be underlined that this result corresponds to the average power of a set of 21 different PV systems in a region.

5.5. Summary of Performance Statistics

To make comparison between simulations easier, MBE , $RMSE$ and MAE have been normalized in order to fall in a more restricted range of values. In statistic studies, it is common to normalize these statistics to the range, $\max(\mathbf{O}) - \min(\mathbf{O})$, or the mean, $\bar{\mathbf{O}}$, of the observations. The first option was chosen to ensure most of the values fall in a range between 0 and 1. Table 4 summarizes all the performance statistics used in this paper.

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.

6. Validation results

The validation has been performed by comparing the AC power forecast with actual measurements from the PV plants summarized in Section 2, following the procedures detailed in Section 5. The possible combinations considering different training set length (number of days,

	Statistic	Modification
Quantile $Q_{.5}$	$nMBE$	MBE normalized respect to the daily range of the observations
	$nRMSE$	$RMSE$ normalized respect to the daily range of the observations
	$nMAE$	MAE normalized respect to the daily range of the observations
	SS	As it is (calculated with normalized $RMSE$)
Confidence interval	$Q1Q9_{Sum}$	As it is
	$Q1_{Num}$	As it is
	$Q9_{Num}$	As it is

Table 4: Performance statistics used to validate the proposed methodology and the modifications made to facilitate the comparison between different simulations.

N), selecting method, scenarios, and PV plants, result in a total of 1.275 cross-validations. For practical purposes, only the detailed results for PV plant P1 are presented in this paper. We have checked that these results represent adequately the behavior of the whole set of PV plants.

As detailed in section 5, the different variants of the model have been analyzed using an adapted leave-one-out cross-validation method. It provides the model performance statistics described in section 5.2 for every day of the dataset, and, consequently, it is possible to compute a daily value of the skill score. Figure 6 displays the median of these daily skill scores for each scenario considering the possible combinations of clearness index classes, length of the training set, and selection method. Scenarios are ordered according to their Skill Score performance.

Several comments are deserved here:

- KS behaves consistently as the best selecting method, although the relative improvement over the other two methods depends on the rest of parameters.
- Forecasts using 7 days in the training set are significantly worse than those with longer training sets. On the other hand, the results produced with 30, 45, and 60 days are frequently overlapped. Therefore, the use of training sets with 30 days seems adequate.
- The best average results are achieved with the scenario 1, which includes the whole set of predictors. Nevertheless, some scenarios can beat it under specific combinations of clearness index, selecting method, and length of the training set.
- The best six methods according to the skill score include both forecasted irradiance ($swflx$) and the sun geometry variables. Moreover, the sun geometry variables are included in all the top 10 scenarios. The step between the worst scenario using sun geometry (15) and the best scenario without it (17) is clearly visible in the KS and KT panels with high clearness indexes. It must be stressed the importance of the sun geometry variables. Scenario 2, which is a reduced copy of scenario 1 without sun geometry, is one of the worst scenarios. Moreover, scenario 12, without both $swflx$ and $Bo0$, presents the worst performance, with SS between -0.07 and -2.04.
- Provided that both irradiance and sun geometry information are included in the predictors set, it is remarkable that a large collection of WRF variables is not mandatory. Focusing on the KS method and $N \geq 30$, the reduction in the model performance, in terms of the skill score, between the scenario 1 and other more parsimonious models is acceptable. A similar result was obtained by (Zamo et al., 2014). Thus, if the NWP provider publishes a reduced set of WRF variables the performance of the forecasts will not be compromised. The scenario 9 is particularly interesting because with only four WRF variables and sun

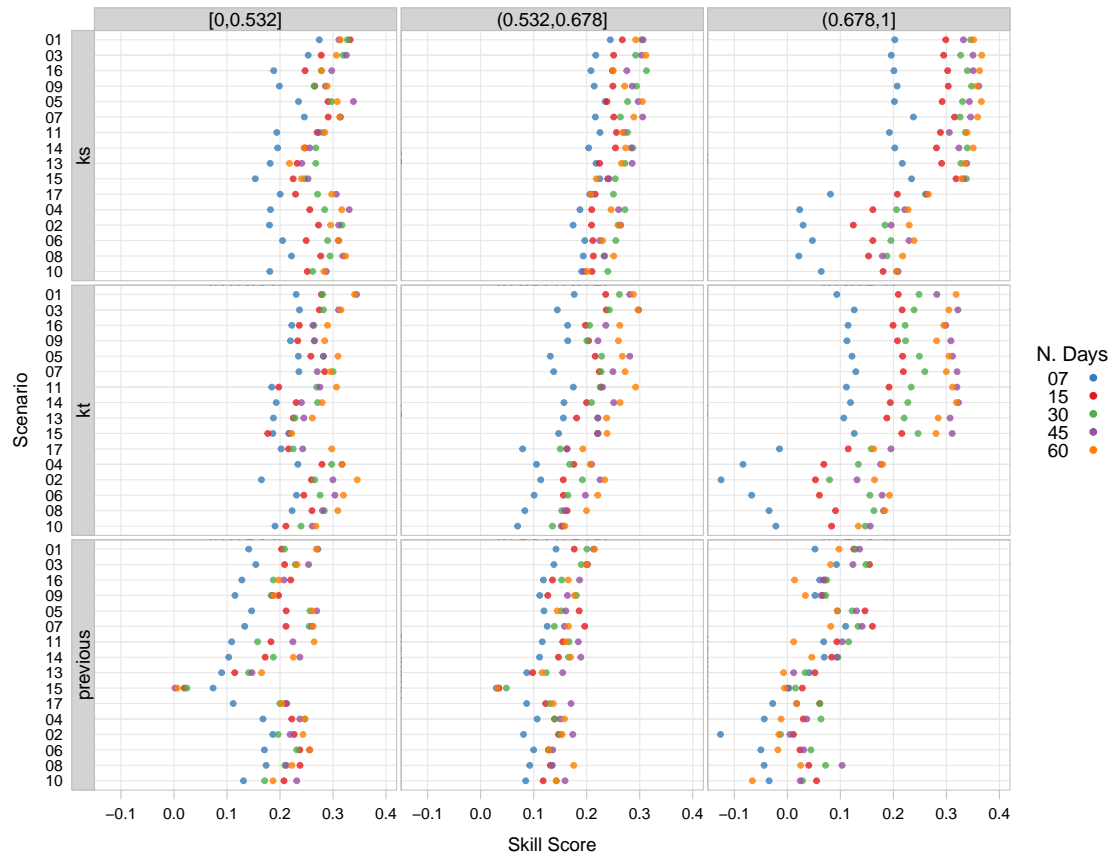


Figure 6: Median of the daily skill scores for each scenario considering the possible combinations of clearness index classes, length of the training set, and selection method. Scenarios are ordered according to their Skill Score performance. Results for scenario 12 are not presented due to its lower performance (SS between -0.07 and -2.04).

geometry information, it is able to produce results comparable to other more complex scenarios: scenario 1 has a SS of 0.325 ± 0.021 , while scenario 9 has a SS of 0.306 ± 0.042 .

- The use of variability indexes increases the skill score, mainly with low clearness indexes, although the improvement is not very high. This is illustrated with the comparison of scenarios 1 and 3, almost identical except by the fact that the last one do not include variability indexes.

Consequently, because the selecting method KS returns the best performance, and because using $N > 30$ days does not increase performance significantly, from now on, and for clarity reasons, only the results corresponding to KS and $N = 30$ will be presented. On the other hand, only the results for the scenarios 1, and 9, as representatives of complexity and parsimony, will be detailed.

Considering all five PV plants described in table 1 the range of SS for scenarios 1 and 9 are 0.336 to 0.361 and 0.324 to 0.350, respectively. These results compare satisfactorily with those reported in (Bacher et al., 2009) with a set of forecast methods of AC power for next day horizons. These authors published skill scores up to 0.36 for aggregated forecasts corresponding to the average power of a set of 21 different PV systems in a region. In contrast, our proposal is focused on the forecast of different PV plants as individual entities.

The quantiles $QS_{.25}$, $QS_{.5}$ and $QS_{.75}$ of the performance statistics for each daily clearness index class are presented in Tables 5 and 6 for these scenarios. The statistics are computed for a period of one day, so 24 individual errors are resumed into one single value. The $nMBE$ indicates the daily mean accumulated error, while the $nRMSE$ and $nMAE$ give some insight on the individual errors.

Statistic	$0 \leq KTd < 0.532$			$0.532 \leq KTd < 0.678$			$0.678 \leq KTd \leq 1$		
	$QS_{.25}$	$QS_{.5}$	$QS_{.75}$	$QS_{.25}$	$QS_{.5}$	$QS_{.75}$	$QS_{.25}$	$QS_{.5}$	$QS_{.75}$
$nMBE$	9.88%	-4.02%	-18.49%	8.11%	-0.92%	-7.26%	3.74%	-0.75%	-3.28%
$nRMSE$	26.21%	31.65%	42.95%	13.74%	20.90%	27.68%	3.71%	7.71%	15.49%
$nMAE$	19.38%	24.19%	34.64%	10.61%	16.25%	22.05%	2.80%	5.48%	11.13%
$Q1Q9_{Sum}$	1.35	1.99	3.60	0.73	0.99	1.32	0.27	0.36	0.51
$Q1_{Num}$	0	0	1	0	0	2	0	0	1
$Q9_{Num}$	0	1	2	0	1	2	0	1	2

Table 5: Quantiles $QS_{.25}$, $QS_{.5}$ and $QS_{.75}$ of the performance statistics for each daily clearness index class, with $N = 30$ days, selecting method KS and scenario 1.

Statistic	$0 \leq KTd < 0.532$			$0.532 \leq KTd < 0.678$			$0.678 \leq KTd \leq 1$		
	$QS_{.25}$	$QS_{.5}$	$QS_{.75}$	$QS_{.25}$	$QS_{.5}$	$QS_{.75}$	$QS_{.25}$	$QS_{.5}$	$QS_{.75}$
$nMBE$	17.83%	-1.65%	-19.46%	9.70%	1.98%	-6.35%	3.70%	-0.66%	-3.03%
$nRMSE$	27.03%	35.61%	47.82%	14.13%	21.03%	29.27%	3.29%	7.62%	15.24%
$nMAE$	19.80%	26.53%	39.07%	10.52%	16.13%	22.69%	2.66%	5.36%	10.80%
$Q1Q9_{Sum}$	1.24	1.98	3.62	0.70	0.95	1.22	0.22	0.34	0.49
$Q1_{Num}$	0	0	2	0	0	2	0	0	2
$Q9_{Num}$	0	1	3	0	1	2	0	1	2

Table 6: Quantiles $QS_{.25}$, $QS_{.5}$ and $QS_{.75}$ of the performance statistics for each daily clearness index class, with $N = 30$ days, selecting method KS and scenario 9.

The median ($QS_{.5}$) $nMBE$ is small for both scenarios and all KTd classes, with a maximum absolute value around 4%. This is expected from a statistical method based on Random Forests, which has the tendency to give unbiased results. Individual errors are somewhat bigger, as can be observed from the higher values of $nRMSE$ and the $nMAE$. Nevertheless, the medians of these two statistics for clear days, which concentrate most of the electricity generation (almost 50%), are excellent. Results for scenarios 1 and 9 are again very similar (for all the quantiles presented), reinforcing the fact that as long as irradiance data is present in the predictors, a large collection of WRF variables is not mandatory.

The uncertainty related to the quantile $Q_{.5}$ is relatively low for clear and partially clouded days. For cloudy days, $Q1Q9Sum$ indicates a higher level of uncertainty, but this is strictly related to the variability of the solar resource due to unstable cloud cover and the small amount of energy generated during cloudy days, which is closer to the magnitude of the generation uncertainty.

Statistical methods based on Random Forests, due to their inherent averaging, tend to avoid minimums and maximums. Therefore, extrapolations regarding the quantile $Q_{.9}$ are more likely to happen. As the confidence interval is limited by the quantiles $Q_{.1}$ and $Q_{.9}$ and it is calculated for a day (24 hours), no more than 2.4 extrapolations are expected under or over the confidence interval. Median $Q1Num$ and $Q9Num$ are consistent with what was expected. Only with cloudy days and scenario 9 $Q9Num$ presented a number of extrapolations bigger than 2.4 in the range between $QS_{.25}$ and $QS_{.75}$.

Next section details the daily performance of the scenarios 1 and 9 focusing on the daily results of $nRMSE$ and $nMBE$.

6.1. Daily results compared between scenarios

The previous section has analyzed the aggregated results of the model using the skill score metric and the quantiles of the error statistics proposed in section 5.5. This section offers an insight into the raw results, that is, all the daily outputs of the forecasts.

With this approach, scenarios cannot be compared through tables or summary graphics due to the large amount of numbers linked to each scenario and metric. A suitable approach to make a more detailed comparison of the performance of different scenarios is merging the results of $nMBE$ and $nRMSE$ with a graphical tool such as the Target Diagram (Jolliff et al., 2009), derived from Equation 9. Moreover, this daily analysis is restricted to the scenarios 1 and 9, using the selecting method KS and training length $N = 30$, because this is the combination that has been highlighted in the previous aggregated comparative.

The Target Diagram uses a Cartesian coordinate system, where the x-axis represents the unbiased $nRMSE$ (variance of the error, σ_D^2), and the y-axis represents the $nMBE$. Therefore, the distance between any point to the origin is equal to the $nRMSE$. Because the unbiased $nRMSE$ is always positive, the left area of the coordinate system is empty with this scheme. With additional information this region may be also used: the unbiased $nRMSE$ is multiplied by the sign of the difference between the standard deviations of forecast and observations, $\text{sign}_\sigma = \text{sign}(\sigma_F - \sigma_O)$.

The resulting diagram provides three different measures: whether the forecast overestimates or underestimates (positive or negative values of the $nMBE$ on the y-axis, respectively), whether the forecast standard deviation is larger or smaller than the standard deviation of the measurements (positive or negative values on the x-axis, respectively), and the error performance as quantified by the $nRMSE$ represented as the distance to the coordinates origin.

Figure 7 displays the target diagrams corresponding to the daily results of the scenarios 1 and 9, where each dot is related to a certain day of the database. Each column corresponds to a different class of clearness index. Both scenarios are displayed with different colors.

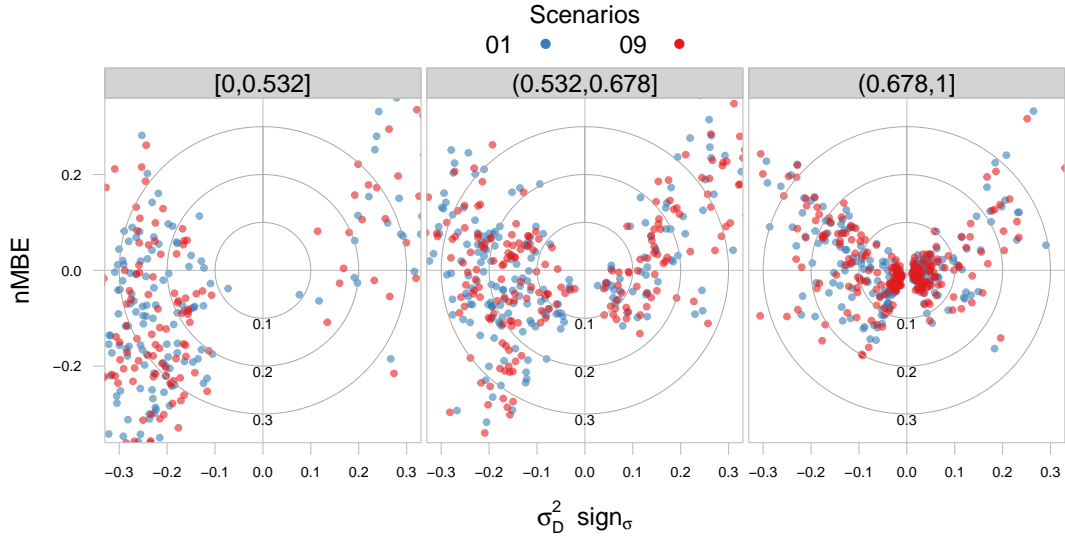


Figure 7: Target diagrams comparing scenarios 1 and 9, with $N = 30$ days and selecting method *KS*. Each diagram comprises the daily results for a different clearness index class.

From Figure 7 it can be easily observed that the two clouds of points corresponding to scenarios 1 and 9 overlap. Therefore, even though scenario 9 is more parsimonious than scenario 1, their performances are almost the same.

During cloudy days forecasts have higher $RMSE$ values, because of both higher MBE and higher σ_D^2 . Moreover, MBE tends to be negative, indicating that observations are consistently higher than forecasts. On the other hand, $sign_\sigma$ tends to be negative, what suggests that the standard deviation of the forecast is lower than the standard deviation of the observations. Most points are located between the circles with radius 0.1 and 0.3.

During clear days points are closer to the center of the diagram, inside a $nRMSE$ radius of 10%. Although MBE values tend to zero, they are slightly positive when they are apart from the center (observations are lower than forecasts). There is not a clear tendency in the values of $sign_\sigma$.

7. Impacts on Daily Energy Production Forecast

Previous sections have evaluated the model performance under a statistical framework using tools and metrics commonly found in this discipline. However, PV power forecasting is mainly used for trading energy in electricity power markets. Therefore, a different analysis is necessary due to the nature of the final application of the method: energy assessment.

This section discusses the performance of the model under this framework, taking into account the economical benefits and penalties stated in the market regulations. There is a variety of market practices and regulations that provokes that a certain forecasting model can perform better or worse due to the different impact of the success and failures in each market. Therefore, the metrics used to evaluate the model performance, in terms of the quantile $Q_{.5}$ of the forecast, must be adequate to the market configuration. 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.

Besides, in this context these metrics are more useful if presented as an energy ratio, and thus they must be normalized to the daily measured energy. Equations 17 and 18 shows the resulting metrics, denoted as the coefficient of variation of the *MBE* and *MAE* respectively.

$$cvMBE = \frac{MBE}{\bar{O}} \quad (17)$$

$$cvMAE = \frac{MAE}{\bar{O}} \quad (18)$$

The *cvMAE* measures the goodness of the predictions for applications requiring hourly predictions during a period of a day, whereas the *cvMBE* is an index of the goodness of the total daily energy production. Both have been computed for every day included in the database. The median of the results has been calculated for each daily clearness index class. Besides, the median was weighted with the energy generated by the PV plant under the corresponding clearness index class.

The statistics have been computed with $N = 30$ days, selecting method *KS* and scenarios 1 and 9, as presented in Tables 7 and 8 for PV plants P1 and P5, respectively.

Statistic	Scenario	$0 \leq KTd < 0.532$	$0.532 \leq KTd < 0.678$	$0.678 \leq KTd < 1$
<i>cvMBE</i>	1	-1.27%	-0.47%	-0.54%
	9	-0.51%	1.22%	-0.49%
<i>cvMAE</i>	1	8.63%	9.14%	4.13%
	9	9.49%	8.76%	4.22%

Table 7: Weighted errors of energy forecast for PV plant P1 according to the daily clearness index class, with $N = 30$ days, selecting method *KS* and scenarios 1 and 9.

Statistic	Scenario	$0 \leq KTd < 0.532$	$0.532 \leq KTd < 0.678$	$0.678 \leq KTd < 1$
<i>cvMBE</i>	1	-1.16%	0.07%	-0.72%
	9	-0.28%	0.58%	-0.49%
<i>cvMAE</i>	1	9.09%	9.32%	3.83%
	9	9.45%	9.33%	3.73%

Table 8: Weighted errors of energy forecast for PV plant P5 according to the daily clearness index class, with $N = 30$ days, selecting method *KS* and scenarios 1 and 9.

Values of *cvMBE* are small, what was expected due to the machine learning tool used. Total daily energy is forecasted with an absolute *cvMBE* of less than 1.3% for all *KTd* classes. In terms of hourly prediction, the overall *cvMAE* is less than 9.5%. Both results are very good and appear to be independent of the size of the PV plant.

8. Conclusion

A methodology to forecast one day ahead hourly AC power produced by a PV plant has been proposed. This approach conceives the PV system as a black box (nonparametric PV model), and it does not presume any knowledge of internal characteristics and processes of

the system. It is a data-driven model that estimates the behavior of the system from a historical time series of inputs and output.

The methodology uses forecasts of several meteorological variables (produced by a Numerical Weather Prediction model), and spatial and temporal indexes (estimated from the forecasted variables) as inputs to predict the hourly AC power of the PV plant. The PV model is constructed upon a machine learning tool, the Quantile Regression Forests, which must be trained with past series of both NWP forecasts and AC power measurements. This tool is able to produce both a central forecast (median) and a confidence interval, which is useful to assess the reliability of the forecast.

Training set has been constructed with three different selection methods: choosing N days immediately before the day to be predicted; including 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; or choosing 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.

The validation procedure has analyzed the performance of the methodology according to the daily clearness index, the training set length (number of days composing the historical NWP forecast and AC power measurements), the NWP variables and indexes used, and the training set selecting method. The main conclusions from this analysis are:

- The presence of irradiance data, predicted ($swflx$) and calculated ($B00$), leads to better results.
- Increasing the number of WRF variables do not necessarily increase the accuracy of the forecast.
- Training set selecting methods based on similarity measures (KT and KS) behave better than choosing recent days ($Previous$). Method KS achieves the best results.
- The training set length has no significant impact on the model performance with time series longer than 15 days. A value of $N = 30$ days was used with good results.
- The confidence interval ability to contain all observations within is very good, especially for the quantile Q_1 .
- Total daily energy is forecast with an absolute $cvMBE$ of less than 1.3% for all KTd classes.
- In terms of hourly prediction, the overall $cvMAE$ is less than 9.5%.

The model performance has also been evaluated using the Skill Score a measure of the relative improvement of the forecast method over the persistence forecast. The results range from 0.33 to 0.36 that compare satisfactorily with the set of forecast methods reported in (Bacher et al., 2009). These authors published skill scores up to 0.36 for aggregated forecasts corresponding to the average power of a set of 21 different PV systems in a region. In contrast, our proposal is focused on the forecast of different PV plants as individual entities.

The full code implementing the procedure is freely available from the repository <https://github.com/iesiee/PVF>.

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. 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>
- 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.
- 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.
- Jolliff, J., Kindle, J. C., Shulman, I., Penta, B., Friedrichs, M. A. M., Helber, R., Arnone, R. A., 2009. Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. *Journal of Marine Systems* 76, 64–82.
- 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.
- 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
- 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>

- 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.
- 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 post-processing of the global environmental multiscale numerical weather prediction model. *Progress in Photovoltaics: Research and Applications* 21 (3), 284–296.
- 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
- 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/>
- Wilson, M. F., O’Connell, B., Brown, C., Guinan, J. C., Grehan, A. J., 2007. Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. *Marine Geodesy* 30 (1-2), 3–35.
URL <http://trac.osgeo.org/postgis/raw-attachment/ticket/2164/Wilson-MarGeo-2007.pdf>
- 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/>