# Improving Renewable Energy Forecasting with a Grid of Numerical Weather Predictions

José R. Andrade and Ricardo J. Bessa

*Abstract*—In the last two decades, renewable energy forecasting progressed towards the development of advanced physical and statistical algorithms aiming at improving point and probabilistic forecast skill. This paper describes a forecasting framework to explore information from a grid of numerical weather predictions (NWP) applied to both wind and solar energy. The methodology combines the gradient boosting trees algorithm with feature engineering techniques that extract the maximum information from the NWP grid. Compared to a model that only considers one NWP point for a specific location, the results show an average point forecast improvement (in terms of mean absolute error) of 16.09% and 12.85% for solar and wind power respectively. The probabilistic forecast improvement, in terms of continuous ranking probabilistic score, was 13.11% and 12.06% respectively.

*Index Terms*—Solar energy, wind energy, forecasting, probabilistic, feature engineering, spatial, temporal, weather predictions

## I. INTRODUCTION

RENEWABLE energy sources (RES) are increasing steadily in several countries, mainly driven by the cost decrease of photovoltaic (PV) panels and wind turbines. Presently, onshore wind energy provides electricity competitively compared to fossil-fuel; the levelized cost of energy (LCOE) of solar PV decreased 58% between 2010-15; concentrated solar power and offshore wind energy showed a LCOE decrease of 43% and 35% respectively in the last five years [1]. In this context, RES can participate directly in the electricity market, which is the case in Denmark, Spain and Germany, and accurate forecasts are a key requirement.

RES variability and uncertainty introduce challenges in power system management, which requires high-quality probabilistic forecasts. At the transmission system operator (TSO) level, the aggregation of individual RES installations smooths variability and uncertainty, e.g. mean absolute error (MAE) around 4% of nominal power for 10 GW of aggregated PV capacity [2], in contrast to around 18% for 10 MW PV power plants; for wind energy the smoothing effect is similar [3].

For a distribution system operator (DSO), the local variability and uncertainty may create technical problems, such as over/under-voltage and branches over-current. Therefore, accurate forecasts at the power plant level is a key requirement for this type of end-user. This information when included in

a predictive grid management algorithm helps to [4]: (i) solve voltage problems and minimize the active power losses; (ii) maximize RES hosting capacity.

Bacher *et al.* proposed one of the first non-parametric models for probabilistic PV forecasting, based on weighted quantile regression conditioned to a clearness index [5]. In the same year, Lorenz *et al.* proposed a parametric method where the Gaussian distribution parameters were conditioned by a combination of a clear-sky index and solar zenith angle [6]. Similar models were developed for wind power probabilistic forecasts, like the local quantile regression [7] or the conditional kernel density estimation (KDE) [8]. A recent development applied to both technologies is an analog search method that used an Euclidean distance metric to rank historical forecasts' similarity to the current NWP forecast; the probability density function was estimated using a set of "best" analogs [9], [10]. A similar concept that combines k-nearest neighbors algorithm and KDE was proposed in [11]. A detailed literature review about solar and wind power forecasting methods can be found in [12] and [13] correspondingly.

All the aforementioned forecasting methodologies do not explore information from a spatial grid of Numerical Weather Predictions (NWP) and use a single point for the RES power plant location.

Recent works showed the added value of spatial-temporal time series information to improve RES forecasting skill. In this scope, the most adopted model was the vector autoregression (VAR) that used past observations from the power in each site, combined with past values from neighboring sites. This framework was applied to solar ([14], [15]) and wind ([16], [17]) power time series. More advanced models include conditional VAR models where the coefficients are allowed to vary with external variables (e.g., average wind direction) [18] or online forecasting with Markov chain using real-time measurements of the wind turbines [19]. However, all these methods only explored spatially distributed measured time series to improve the forecast skill in the very short-term horizon (e.g., up to six hours-ahead).

The models that ranked first and second in the Global Energy Forecasting Competition 2014 (GEFCom2014) combined feature engineering techniques with gradient boosting trees (GBT) and quantile regression forests [20]. The random forest (point forecast)/GBT (quantile forecast) stack with onsite information from each power plant achieved the second place in both wind and solar tracks [21]. Landry *et al.* won the wind power forecast track by exploring offsite information with a multi-layer approach: in the first layer, a separated GBT is fitted for each wind zone; then, its out-of-sample forecasts are used as inputs to a second layer [22]. The

winner of the solar power forecast track also included offsite information in the GBT, i.e. a single input dataset with all weather variables was used for each solar power plant [23]. The analysis of the GEFCom2014 results emphasized that it was not clear if offsite information was a deciding factor (e.g., methods with offsite information, such as [24], were outperformed by [21]) and highlighted that feature engineering with onsite data was probably the key factor [20]. Also using this data, a Gaussian process was proposed for point forecast in which the covariance matrix incorporates joint effects between neighboring wind power plants [25]. Nevertheless, none of the revised methods explored information from a complete NWP grid.

A NWP grid was explored by Almeida *et al.* to construct indexes that quantify spatial variability of raster NWP data applied to PV forecasting, inspired by terrain analysis indexes, such as topographic position and roughness [26]. Davò *et al.* explored principal component analysis (PCA) to extract information from a grid of wind speed and downward shortwave flux to forecast the total PV and wind power in one region [27]. Additional variables were not derived from the NWP grid and only forecasts at the regional level were explored. In fact, only the benefits of dimension reduction were demonstrated in that work.

The present paper aims at exploring information from a grid of NWP to improve RES point and probabilistic forecasting skill for a short-term time horizon (i.e., up to three days ahead). Compared to the state of the art, this is the first work to propose a framework to extract features from a NWP grid by using domain knowledge and to prove that this information can improve the forecast skill of state of the art forecasting systems. The proposed methodology constructs new variables from the raw NWP data that are used as inputs in the GBT algorithm. This results in significant improvement in point and probabilistic RES forecast. In particular, compared to reference [26] this work explores different variability features that result in higher accuracy and generalizes the framework to both wind and solar energy. Compared to reference [27], it applies a different framework (i.e., more features and a PCA by layer of variables) at the power plant level.

The remaining of this paper is organized as follows: section II describes the data used from a wind power plant located in Spain (Galicia region) and a PV site located in Portugal (Porto); section III describes the forecasting framework developed to explore temporal and spatial data; the test results are discussed in section IV; conclusions are presented in section V.

## II. DATA DESCRIPTION

### A. PV Power Plant

The solar power dataset consists of hourly power measurements from a 16.32 kW peak PV power plant located on the North of Portugal (Porto city), divided in the following strings: 3x2160W, 2x3360 W and 1x3120 W.

The available historical period spans from March $28^{th}$, 2013 to June $28^{th}$, 2016. The missing values in each string were interpolated using the available measured data from the other
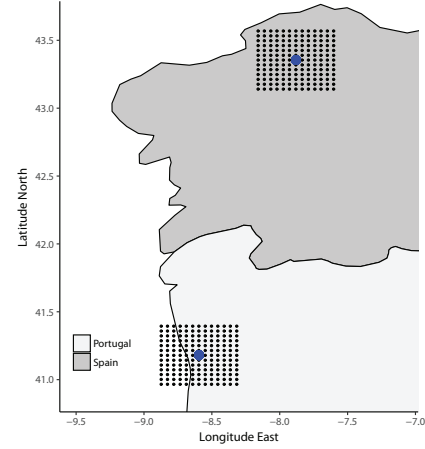


Fig. 1. NWP grids for both Porto PV site and Sotavento's wind power plant.

strings, thus the number of missing values was very small (only 3.82%).

### B. Wind Power Plant

The wind power dataset is from the Sotavento Experimental wind power plant, located in Galicia region in the North-Western part of Spain. The site is located 600-700m above sea level, placed in a moderately complex orography. The power plant is composed by 24 wind turbines from five different technologies, with rated capacity ranging from 600 kW to 1320 kW, which makes a total capacity of 17.56 MW.

The available historical data consists of wind power, speed and direction measurements, covering a time period from January $1^{st}$, 2014 to September $22^{nd}$, 2016. The percentage of missing values was only 1.16%.

### C. Spatial Grid of Numerical Weather Predictions

The access to NWP data is crucial to improve the forecast accuracy in the short-term horizon and a grid of NWP points can be generated for a specific region. Mesoscale models, such as the Weather Research and Forecasting (WRF) model, generate high resolution NWP grids suitable for RES applications [28].

The NWP-WRF data was retrieved from MeteoGalicia THREDDS server, which is a publicly available service that provides historical and daily forecasts of several weather variables. Even though the service has available data for three different spatial domains, specifically 36 km, 12 km and 4 km, for this work the latter was considered. The temporal resolution is one hour, each run is initialized at 00h UTC and the time horizon is 96 hours-ahead.

Figure 1 depicts the NWP grid for the solar and wind power plants and a different set of weather variables was selected for each point accordingly to the RES technology.

The defined geographical grid for both case studies is made up by 169 equally distributed points in a 13 x 13 spatial configuration, covering an approximated area of 2400 km$^2$.

Table I presents the list of variables considered for each technology. The raw NWP dataset consisted in 2704 variables for the wind power plant and 1014 variables for the PV site.

TABLE I
WEATHER INFORMATION FOR EACH CASE STUDY

| Case Study | Variables |
|---|---|
| PV | swflx [W/m$^2$] - surface downwelling shortwave flux |
| | temp [K] - ambient temperature at 2 meters |
| | cfl [0,1] - cloud cover at low levels |
| | cfm [0,1] - cloud cover at mid levels |
| | cfh [0,1] - cloud cover at high levels |
| | cft [0,1] - cloud cover at low and mid levels |
| Wind Power | Considering four model levels for each variable: |
| | u [m/s] - azimuthal wind speed |
| | v [m/s] - meridional wind speed |
| | mod [m/s] - wind speed module |
| | dir [0°,360°] - wind direction |

## III. FORECAST FRAMEWORK

### A. Feature Engineering

One of the main obstacles regarding PV and wind power forecasting is the variability of the respective endogenous resource. Even though it is common sense that PV generation is well characterized by diurnal high variability periods in partially cloudy days and lower during sunny or total overcast days, the WRF information provided for a specific site (or coordinates) is not always precise enough to correctly model these time periods, which increases the overall error of the models for cloudy days while maintaining a significant and counterintuitive degree of uncertainty for clear days forecast.

In the same way, the variability of wind energy derives not only from sudden changes in the local wind dynamics but it is also influenced by larger scale wind-flow patterns, which might not be correctly explained by the series of WRF variables available for a specific location.

Therefore, the use of a equally distributed grid of NWP geographical points aims to capture each site's surrounding information in order to complement the current local information, increasing the precision and robustness of the forecasting model and decreasing both uncertainty and point forecast errors.

In the following sub-sections a set of features used to quantify both spatial and temporal variability are described.

*1) Temporal Information:* Regarding the PV case study, in a pre-processing phase a weighted quantile regression model was applied to the power and shortwave flux time series to generate a set of clear-sky estimations, used to normalize the original data (e.g. ratio between observed and clear sky solar solar power) and remove its periodic component [5] .

For each day, a set of four forecasts, i.e. generated in the current day at 00 UTC plus three generated on the previous days, are available for the central points of each grid represented in Figure 1. Three different variables are created to explore the temporal information in this set of forecasts.

From the present day forecasts, two features are extracted:

- *Temporal Variance* $\left(\sigma^2_{time}\right)$: considering a $N_h \in \{3, 7, 11\}$ hours moving window centered in a specific lead-time of interest $(t+k)$, three temporal variance

indexes are computed along each WRF variable time series $x(t+k)$:

$$\sigma^2_{time}(t+k) = \frac{\sum_{i=t+k-N_h}^{t+k+N_h}(x(i) - \bar{x})^2}{N_h - 1} \qquad (1)$$

In Figure 2 a comparison between both power time series (first panel) and respective main WRF variables for the central location (*swflx* and *mod$_{level3}$*, second panel), shows that high power variability periods are accompanied by high temporal variance (depicted in the third panel). In contrast, in time periods with low fluctuations (e.g. seventh in PV case and fourth day in wind power case) the hourly difference between forecasts is low and the temporal variance index exhibits a low value.

- *Lags and leads* $(t+k \pm z)$: The temporal dependency of a NWP variable in regression models can be modeled by including the preceding (lags, $t+k-z$, $z = 1, 2, \ldots$) and following (leads, $t+k+z$, $z = 1, 2, \ldots$) with respect to a specific lead-time. In previous works (c.f., [21], [24]), the inclusion of these variables resulted in accuracy improvement. In this work, the importance of this conventional approach is verified and used as benchmark in order to demonstrate the value of the new features.

For the majority of days in the PV case study, a comparison between the four forecast runs (depicted in the fourth panel of Figure 2) shows high similarity regarding clear days forecasts and an increase in the hourly discrepancy between all the values for high variability days. For complete overcast days, this pattern is not completely clear; however, there is a significant difference between high and low/mid clearness days.

After testing different methods to model this information, a smoothing approach based on the weighted mean $(\bar{x}_{past}(t+k))$ of all the available values was considered in order to reduce the number of input information while slightly improving the overall results:

$$\bar{x}_{past}(t+k) = \frac{\sum_{d=0}^{3} \omega_d x_d(t+k)}{\sum_{d=0}^{3} \omega_d} \qquad (2)$$

Where $d \in \{0, 1, 2, 3\}$ represents the four model runs initialized at 00 UTC for the same lead-time and $\omega_i$ are weights associated to each forecast. The weights were estimated with a constrained global optimization solver[1] that is built upon Bayesian techniques and Gaussian processes [29]. The objective function is to maximize the correlation between the smoothed series and the respective power time series.

Regarding the wind energy case study, the correlation between the discrepancy of all the forecasts and different variability periods was not so prominent. Still, the performance of the model was slightly improved with the introduction of the four NWP runs as individual inputs.

*2) Spatial Information:* Three different approaches were defined in order to extract the most relevant information of the WRF grid data.

---

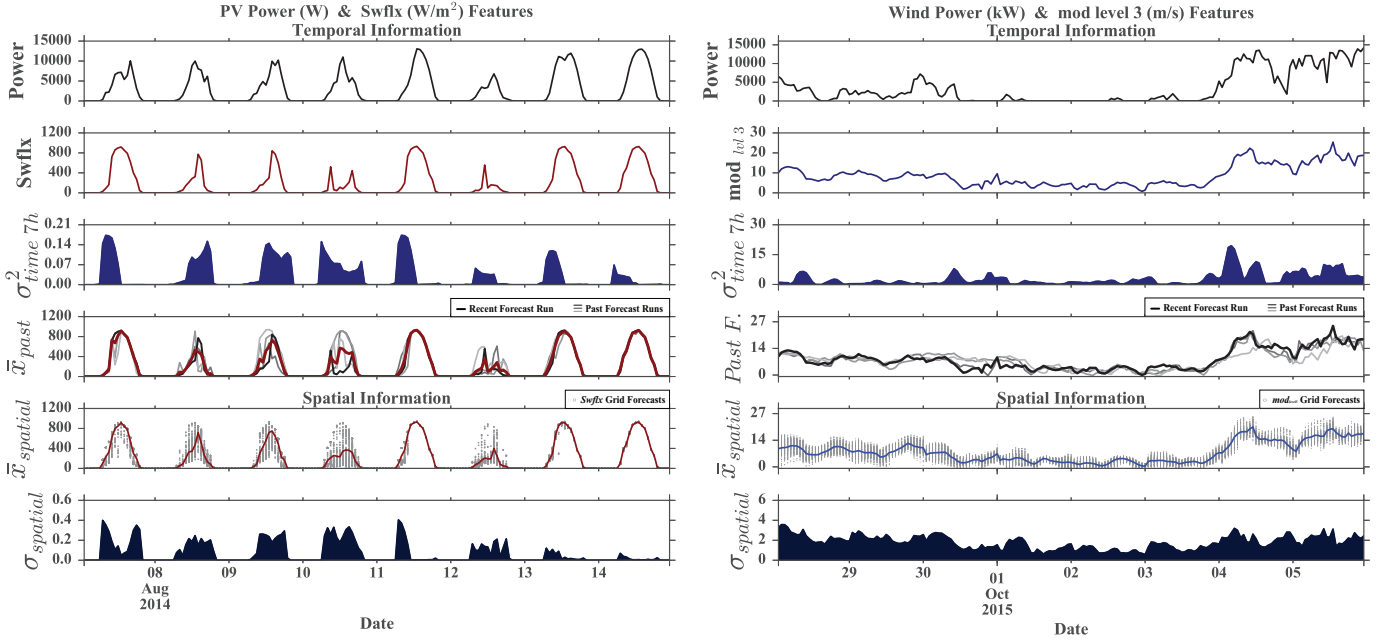[1]Bayesian Optimization Package: github.com/fmfn/BayesianOptimization

Fig. 2. Power time series comparison with new temporal and spatial features.

- *Principal Component Analysis (PCA)*: dimension reduction technique that uncovers the most important patterns and dynamics of the original data. The output features result from a linear transformation process in which the initial data is projected into a new set of orthogonal axis (also called principal components) accordingly to the direction that maximizes the original data information represented by each component. The PCA implementation from the Python `scikit-learn` library was used in order to transform each of the normalized set of 169 WRF variables (grouped by type) in the same number of principal components, which represents 100% of the initial data variance. To reduce the set of components to a lower dimension one may arbitrarily choose a variance threshold, removing some noise and redundant information.

  Concerning the PV case study, the overall best results were achieved by applying the PCA individually to every *swflx*, *cfl*, *cfm* and *cft* grid data, considering a 90% variance threshold. For the wind power case, the best model included the PCA applied to the model levels 1,2,3 *mod* and *dir*, and to the *u* and *v* in model level 3 with a 95% variance threshold.

- *Spatial Standard Deviation ($\sigma_{spatial}$)*: An initial comparison between the raw datasets of each technology most relevant WRF variables (*swflx* and $mod_{level3}$) along all the geographical points, and the observed power in the two power plants, revealed a significant correlation between the dispersion of the NWP grid values and different power variability time periods. This evidence is illustrated in Figure 2 fifth panel where it is clear a discrepancy between NWP grid forecasts (represented as grey dots) for periods with low or high variability, e.g. the contrast between first and last two days for the PV

case, and the fourth and sixth day for the wind power case study.

Thus, in order to quantify this information, a hourly standard deviation index was computed individually for every type of available WRF variables ($x_i$ being $i$ the index of each geographical point):

$$\sigma_{spatial}(t+k) = \sqrt{\frac{\sum_{i=1}^{N_p}(x(t+k)_i - \bar{x})^2}{N_p - 1}} \quad (3)$$

where $t + k$ is a specific lead-time of the time horizon and $N_p$ is the number of geographical WRF points considered. This feature, similarly to the temporal variance, is able to characterize different variability periods for both PV and wind energy, providing additional information extracted from the NWP grid forecasts and, therefore, balancing some of the NWP errors. This feature is able to characterize different variability periods for both PV and wind energy, even though the relation is more evident for the first.

- *Spatial Smoothing*: Even though the spatial standard deviation index tends to accentuate and differentiate high and low statistical dispersion periods along all the grid data, sometimes, due to poor forecast quality this dispersion may not characterize well enough the magnitude of power fluctuations, which may lead to some over/under estimation periods. To complement this index, a new variable is derived to smooth the NWP of the main variables for the central location based on surrounding forecasts for the entire domain.

  For the wind power case study, a simple hourly mean computed individually for the WRF grid values of *mod*, *u* and *v* at model levels 1, 2, 3, provided the best results.
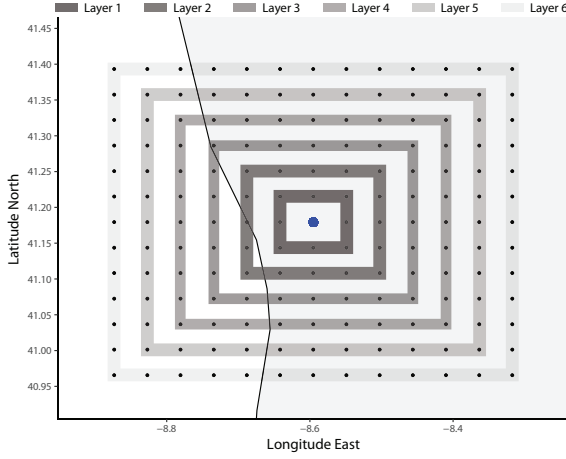
Fig. 3. Scheme of the spatial grid division by layers.

TABLE II
CORRELATION COEFFICIENTS BETWEEN EACH LAYER *swflx* SMOOTHED MEAN AND THE OBSERVED PV GENERATION. ALL DOMAINS ARE IN ASCENDING ORDER (1-6) ACCORDING TO THE DISTANCE TO SITE.

| Spatial Grid Domains | | | | | |
|---|---|---|---|---|---|
| Layer 1 | Layer 2 | Layer 3 | Layer 4 | Layer 5 | Layer 6 |
| 0.8287 | 0.8347 | 0.8382 | 0.8406 | 0.8427 | 0.8452 |

However, for the PV case, a decomposition of all the shortwave flux ($swflx$) grid information in smaller layers (each domain illustrated by a different color gradient in Figure 3) showed that the farther the domain considered for the smoothing is from the site, the higher is the new time series correlation with the observed PV generation. These results are presented in Table II, demonstrating the correlation increase between the central location *swflx* time series smoothed with the most inner spatial domain (*Layer 1*) and the same values smoothed with the most further information (*Layer 6*). Considering this empirical result, a new feature is constructed in two steps:

1) Smooth the central *swflx* time series (point in the center of the grid represented in Figure 3) by calculating the hourly mean $\bar{s}(t + k)$ considering every point in each defined layer. Six smoothed time series will be obtained, retaining every domain different information.
2) Compute the hourly weighted mean ($\bar{x}_{spatial}(t+k)$) considering all the smoothed time series:

$$\bar{x}_{spatial}(t + k) = \frac{\sum_{i=1}^{6} \beta_i \bar{s}_i(t + k)}{\sum_{i=1}^{6} \beta_i} \quad (4)$$

where $i \in [1,6]$ represents the six smoothed series and $\beta$ the respective weight coefficient estimated with Bayesian optimization to maximize the correlation between the $\bar{x}_{spatial}(t + k)$ and the observed power for each train dataset. It is important to stress that the number of layers depend on the specific case study and NWP spatial resolution.

## B. Weather-to-Power Model

*1) Gradient Boosting Trees (GBT):* Gradient boosting is an ensemble machine learning algorithm that can be applied to both classification and regression problems. It conducts numerical optimization using steepest-descent minimization in function space by combining base learners recurrently on modified data that is the output from the previous iterations [30], [31]. GBT is a particular case of the gradient boosting algorithm where the base learners are regression trees. The additive training process can be summarized as follows:

Start with an initial model, which can be the average of the target variable $F(\mathbf{x}) = \sum_{i=1}^{N} y_i \big/ N$

Iterate until converge:

1) Calculate negative gradients for each training sample $i$: $-g(\mathbf{x}_i) = -\partial L(y_i, F(\mathbf{x}_i))/\partial F(\mathbf{x}_i) = y_i - F(\mathbf{x}_i)$, were $L$ is the loss function
2) fit a regression tree $h(\mathbf{x}; a)$ using the negative gradients $-g(\mathbf{x}_i)$ as target variable; $a$ is the set of hyperparameters of the regression tree
3) $F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \tau \cdot h(\mathbf{x}; a)$, where $\tau$ is a learning parameter. This last step means that function $F$ is updated based on the negative gradient fitted by the base learner.

This algorithm has three major advantages for the RES forecasting problem: (i) non-parametric; (ii) can consider different loss functions, e.g. generate point forecasts by using the absolute or square loss and probabilistic forecasts with the quantile loss function; (iii) scalable for a high number of explanatory variables (suitable for an industrial integration). In this work, the GBT implementation from the Python open-source `scikit-learn` library [32] was used.

*2) Hyperparameters tuning:* The challenge in the application of GBT algorithm is to tune the different hyperparameters, which are divided into two groups: (i) regression tree; (ii) boosting process.

The tree-specific hyperparameters are:

- *Maximum depth*: limits the number of nodes in the tree in order to control the overfitting. Higher depth will give more flexibility to learn specific relations of a particular dataset. This value was set between 5 and 9 since the base learners should have a low complexity.
- *Minimum number of samples required to split an internal node*: higher values prevent overfitting since avoids that a particular sample is selected for a tree. This value varies between 150 and 350.
- *Minimum number of samples required to be at a leaf node*: used to control overfitting. Value set between 20 and 80 and tuned accordingly to the maximum depth and minimum samples split.
- *Number of features to consider when looking for the best split*: choosing a number of features below the full set leads to a reduction of variance and an increase in bias. As a thumb-rule, square root of the total number of features is a good option, which was followed and confirmed to perform well.

The boosting-specific hyperparameters are:

TABLE III
TIME PERIOD OF EACH TEST DATA FOLD

| Fold | Test Set Range | |
|------|----------------|----------------|
| | PV | Wind Power |
| 1 | 01/05/2014 - 30/09/2014 | 01/01/2015 - 31/05/2015 |
| 2 | 01/10/2014 - 28/02/2015 | 01/06/2015 - 31/10/2015 |
| 3 | 01/03/2015 - 31/07/2015 | 01/11/2015 - 31/03/2016 |
| 4 | 01/08/2015 - 31/12/2015 | 01/04/2016 - 22/09/2016 |
| 5 | 01/01/2016 - 28/06/2016 | |

- *Learning rate*: controls the contribution of each regression tree in the additive training process. A low value shrinks the contribution of each tree and increases the generalization of the model. A thumb-rule is that a small value (e.g, between 0.01 and 0.05) does not influence significantly the results. However, low values require a higher number of iterations to model all the relations, which increases the computational time.
- *Number of boosting iterations*: defines the number of base learners in the final model. The trade-off with the learning rate should be considered to avoid overfitting. Its value was set between 500 and 800, which proven to be the ideal range for the particular range of learning rates;
- *Fraction of samples to be used for fitting the individual base learners*: a value slightly below 100% makes the model robust by reducing the variance. The typical value is around 80%, which was fixed during the tuning process.

The hyperparameters of the GBT model were estimated with the Bayesian Optimization algorithm. A 12-fold cross validation was employed and since all training datasets have a yearly time span, this number of folds guarantees 12 different monthly validation scenarios, each with a respective error. For the final evaluation, the average of monthly errors is considered for each training set in the optimization process.

## IV. EVALUATION OF THE FORECASTING SKILL

### A. Validation Scenarios

The extent of the historical databases available for this work allowed to test each model with a sliding window approach divided in three steps:

1) Select a starting point that guarantees at least one year of historical data (train dataset). In section IV-D2, the model's performance was also evaluated for distinct train dataset sizes, i.e. 12, 9, 6 and 3 months;
2) Forecast the next five months (test dataset);
3) Slide the train and test window five months forward and repeat the procedure until the dataset limit is reached.

Five and four distinct test data folds were created for the PV/Wind Power case studies respectively, as presented in Table III.

This test methodology ensures that every model is tested under different conditions, providing relevant information on which features combinations provide a substantial contribution to the forecasting skill.

### B. Evaluation Metrics

The Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) metrics were considered to evaluate the point forecast quality.

In this work, each probabilistic forecast is represented by a set of quantiles between 5% and 95%, with a 5% increment. The quality of this forecast is then measured by the Continuous Ranked Probability Score (CRPS), which is a state of art metric to compare probability forecasts in the form of cumulative distribution functions.

For each fold presented in Table III, the available set of forecasts was evaluated by the monthly MAE, RMSE and CRPS values, normalized by the rated power. The average of all individual monthly scores was considered for model comparisons.

For each forecasting model ($m$), the relative improvement over a benchmark model is calculated. Assuming $\epsilon$ as the average of all individual monthly scores, and *base* as a defined benchmark model, the improvement of each model is given by $\left(1 - \frac{\epsilon_m}{\epsilon_{base}}\right) \times 100$ (%).

### C. Forecasting Models, Temporal and Spatial Features

To provide a comprehensive study of the importance of every temporal and spatial features discussed in section III-A, the feature analysis is divided in two phases.

Firstly, a base model with the most relevant NWP variables for each grid central point is considered with the following information:

- Solar energy: *month*, *hour*, *swflx*, *temp*, *cfl*, *cfm*, *cfh*, *cft*;
- Wind energy: *hour*, *mod* and *dir* at the four model levels.

For solar energy forecasting, besides the short-wave-flux and cloud cover information, ambient temperature forecast is also considered since PV cells efficiency has some dependency on local temperature. By adding this information to the statistical model, an improvement of 0.42%, 1.07% and 1.20% is verified, concerning MAE, RMSE and CRPS.

Regarding wind energy, in addition to the wind speed variable, the inclusion of wind direction helps to model features of the prevailing wind that justify, at some degree, wind turbines power output variability. The introduction of wind direction revealed a significant improvement of 11.23%, 10.28% and 9.93% of MAE, RMSE and CRPS over a model exclusively based on wind speed information.

This base model is initially compared to different sets of temporal and spatial features, described in Table IV, in order to demonstrate the inherent value of each type of information. Then the two types of feature domains (temporal and spatial) are merged to create a final model.

It is important to mention that every model in Table IV was defined by a careful and extensive variable selection process, removing any redundant information while maintaining the most relevant variables.

Finally, for the PV case study, the temporal and spatial features used by Almeida *et al.* in [26] were computed for the *swflx* and cloud cover (at all heights) in order to construct a benchmark model from the state of the art.

TABLE IV
FEATURES CONSIDERED IN EVERY FORECASTING MODEL

| Domain | ID | Features |
|---|---|---|
| Temporal | T1 | Lags and leads |
| | T2 | $\sigma^2_{time}$ and different NWP runs |
| | T | Combination of models T1 and T2 inputs |
| Spatial | S1 | $\sigma_{spatial}$ and $\bar{x}_{spatial}$ |
| | S2 | Principal components |
| | S | Combination of models S1 and S2 inputs |
| Temporal & Spatial | F | Combination of both domain features |

### D. Analysis of the Results

*1) Forecasting skill for 24 hours ahead:* Figure 4 depicts the improvements of every model in Table IV over the base model considering 12 months historical data for model training. The night periods were removed for the PV generation.

A close look to models T1, T2 and T results reveals that the use of NWP variables lags and leads (T1) is highly relevant, providing the most significant individual improvement of all the tested input combinations.

A comparison between model T2 for both case studies reveals that the use of the moving window temporal variance and past forecast runs provide better results for the PV case study. Nonetheless, there is some relevant improvement for the wind power case study. Interestingly, models T1 and T2 variables complement each other in both case studies, increasing the amount of information one can retrieve from the central point NWP variables.

Regarding the inclusion of NWP spatial features in the PV case study, the PCA approach (S2) provided similar overall results compared to the spatial variables approach (S1). However, the combination of these two models presented better results, as demonstrated by the improvement of model S over S1 and S2. A note to the comparison between model S, that comprises all spatial information, and the approach from Almeida *et al.* that, even though demonstrated a significant improvement over the base model, was outperformed by the combination of features presented in this work.

For the wind power case study, the PCA approach (S2) surpassed the remaining features, retrieving more information from the NWP grid, which led to better accuracy as it may be seen with a comparison between S1 and S2.

It is important to underline that for the NWP *swflx* spatial grid only one principal component was able to explain more than 90% of the original input data variance. On the other hand, for the wind speed spatial grid, with a 95% variance threshold, the number of principal components varies between 3 and 5. This result means that there is more information in the wind speed grid, which partially explains the discrepancy verified in Figure 4 between models S1 and S2 on solar and wind case studies. Another justification for this result was already discussed in section III-A2, where the relation between $\sigma_{spatial}$ and the power observations was not as prominent as for the PV case. Withal, when merging the two approaches (model S), the use of $\sigma_{spatial}$ and $\bar{x}_{spatial}$ features improved the overall score provided by the PCA-based approach.

Figure 5 depicts the monthly relative improvement of model F over T1. For the PV case, it is noticeable that the final model does not outperform the benchmark in every month. Nonetheless the improvement is clearly in favor of the first, denoting an average MAE, RMSE, CRPS improvement of 4.28%, 3.69% and 3.01% respectively, with three months showing average MAE improvements above 10%.

A visual inspection of the probabilistic forecasts, depicted in Figure 6, shows a global reduction in the forecast sharpness (i.e., size of the 95%-5% interval), more acute in clear or total overcast days (see days 7 Jul, 2 Aug, 30 Sep, 2014 and 22 Jan, 4 Feb, 1 Mar, 2016). Interestingly, some over/under estimation situations inherent to T1 point forecasts appear to be smoothed with the introduction of the new information (see days 2 Aug, 2014 and 22 Jan, 9 Jun, 2016), which, along with the better overall point and probabilistic clear-sky days forecast, leads to the improvement verified in January and February.

For wind energy, even though December denotes a negative improvement on MAE and RMSE scores, the overall improvement is 4.47%, 4.83% and 4.66% on MAE, RMSE and CRPS, with a standout 10.41% and 12.39% RMSE improvements on May and August 2015. Similarly to the PV case study, the introduction of spatial information reduced the global sharpness of probabilistic forecasts, while maintaining similar calibration values, which led to the overall improvement in the CRPS values. A visual inspection of Figure 8 shows the reduction in sharpness of the probabilistic forecast, more prominent in some specific days (e.g., 20th, 23rd March and 18th, 19th April 2016 where the shape of probabilistic forecasts is more adjusted to the observed values). Although this sharpness reduction leads to an overall improvement on the forecasting skill, there are some power peaks that might be not correctly modeled (e.g. peak observed in the first hours of 18th March 2016) and even periods where, due to misleading spatial information, model T1 outperforms model F (e.g. 15th and 17th April 2016 where both point and probabilistic forecasts of model T1 surpass model F).

*2) Impact of train dataset size:* Four different tests were conducted to compare the performance of models T1 and F under different train dataset sizes. For both case studies, the best result of this model was achieved with 12 month of historical data. Interestingly, as exposed in Tables V and VI, even though there was a decay in model performance when shorter historical datasets were considered (e.g. for solar energy, MAE increases from 5.85% to 6.47% with a change from 12 to 3 months), the improvement of F over T1 is higher when a range of historical data lower than 12 months is used to train each model on the sliding window approach (e.g., improvement increases from 4.28% to 5.87%), revealing a significant dependency on the extra spatial and temporal information when a shorter range of historical data is used.

*3) Forecasting skill for longer time horizons:* The number of past forecast runs available at the time of forecast narrows when longer forecast horizons are pretended, therefore, for horizons longer than 24h, the past NWP runs were used as individual inputs instead of the weighted average approach. For both case studies, model F overall improvement over model T1 decays as the forecast horizon increases, as depicted
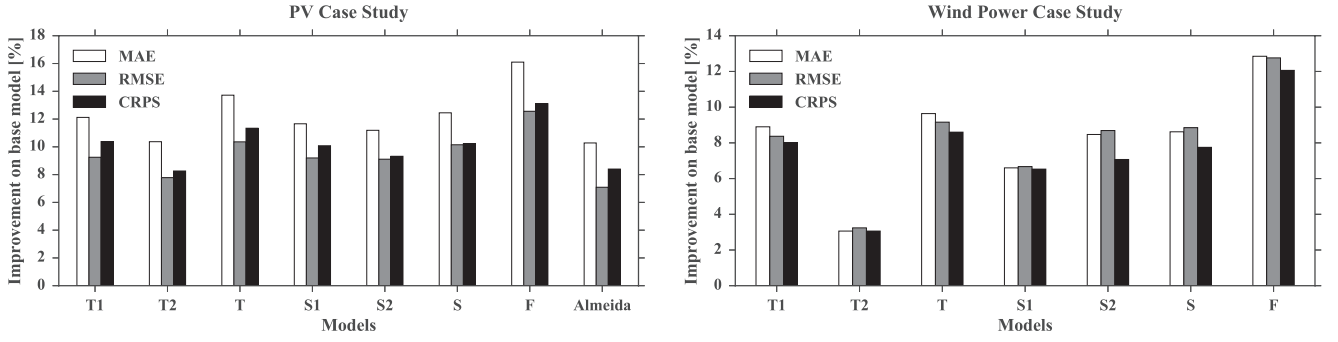
Fig. 4. Relative forecast improvement over the base model for each case study.
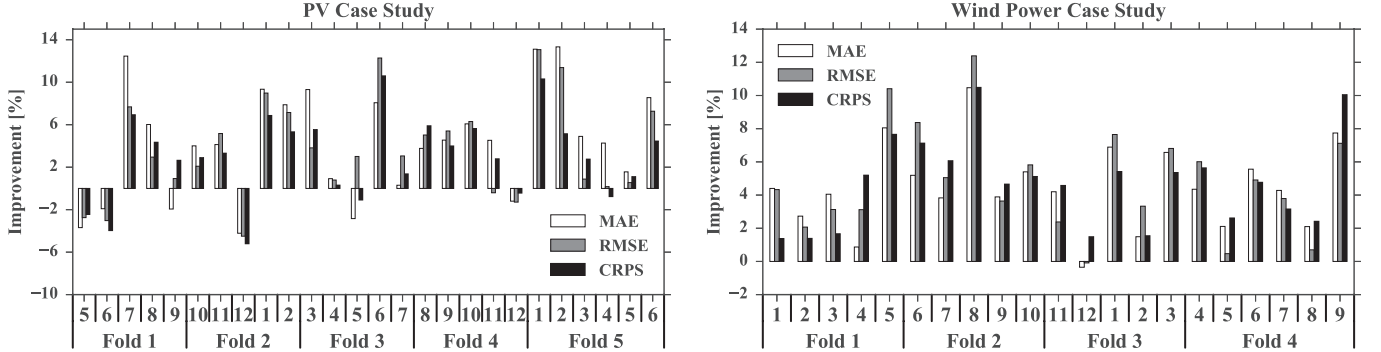


Fig. 5. Monthly relative improvement of model F over T1.
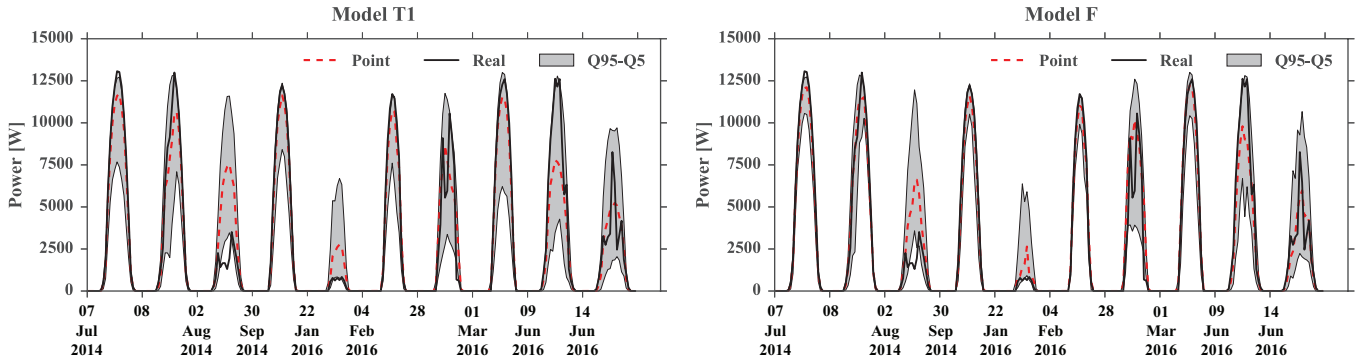


Fig. 6. Subset of days representative of the visual overall reduction of sharpness in the PV power forecasts.

in Figure 7. Regarding solar energy, a close analysis of this figure shows a decay of the MAE, RMSE and CRPS monthly average improvements from 4.28%, 3.69% and 3.01% (1 to 24 hours horizon) to 1.54%, 1.32%, 1.69% (48 to 72 hours horizon). Regarding wind energy, the MAE, RMSE and CRPS improvements also decrease from 4.47%, 4.83% and 4.66% to 3.53%, 2.90% and 2.89% respectively.

One hypothesis for model F performance decrease for longer horizons is the inherent dependency on 169 NWP forecasts along the spatial domain, which quality may decrease for longer lead times, as well as the reduced number of available past forecast runs at the time of forecast generation. All in all, the model F denotes positive improvements for all the aforementioned time horizons.

*4) Computational time:* When dealing with a large number of input features, it is important to assess the total computational time of each model run. Furthermore, the GBT requires a separated training for each quantile. The following times[2], considering the 19 quantiles of model F, were obtained:

- Solar energy (35 input features): higher computational time recorded during the model's fitting phase with 90, 60, 45 and 30 seconds for training datasets of size 12, 9, 6 and 3 months respectively.
- Wind energy (100 input features): considering equal training dataset sizes as the former, all required calculations for model's fitting are performed under 240, 180, 120 and 60 seconds respectively.

The computational time of the daily operational forecast (i.e., feature creation plus forecast calculation) is lower than 2 seconds for both cases.

---

[2]The tests were performed in a Intel(R) Core(TM) i7-2600 CPU, with 8 GB RAM, running a 64-bit Windows 7 operating system.
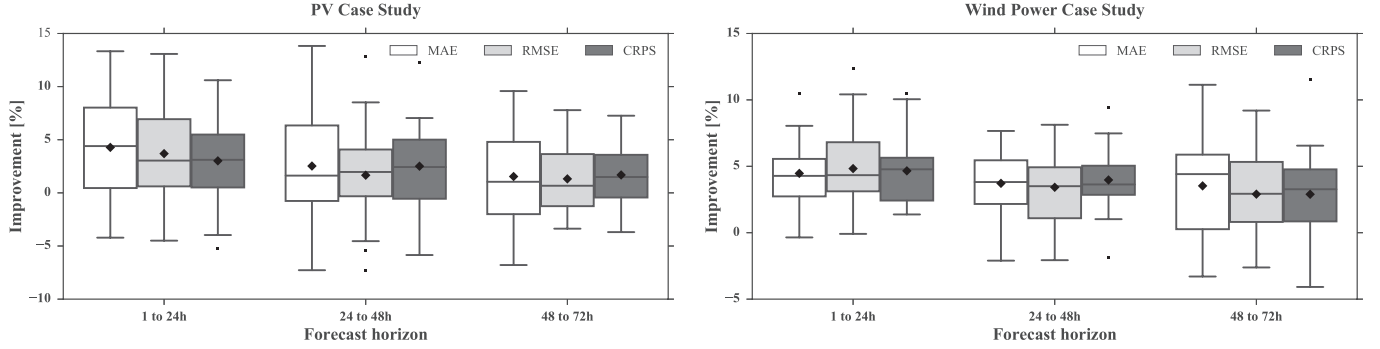
Fig. 7. Representation of model F monthly average improvement over model T1 for distinct forecast horizons. The line and marker in the middle of each box represent the median and mean of all the individual monthly improvements. Small squared markers outside the box represent improvement outliers.
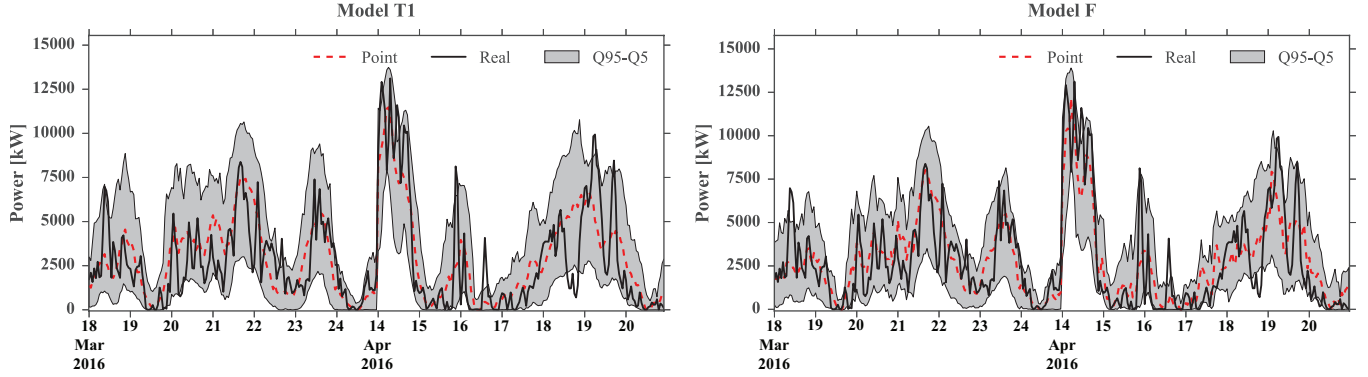


Fig. 8. Subset of days representative of the visual overall improvements in the wind power forecasts.

TABLE V
SOLAR ENERGY: MODEL F ERROR METRICS AND IMPROVEMENT OVER MODEL T1 FOR DIFFERENT TRAIN DATASET SIZES

| Metric | Train dataset size | | | |
|---|---|---|---|---|
| | 3 months | 6 months | 9 months | 12 months |
| | Error (Improvement) | Error (Improvement1) | Error (Improvement) | Error (Improvement) |
| MAE | 6.47% (5.87%) | 6.74% (5.07%) | 5.92% (4.53%) | 5.85% (4.28%) |
| RMSE | 10.16% (3.90%) | 10.35% (4.86%) | 9.42% (4.12%) | 9.37% (3.69%) |
| CRPS | 4.37% (3.41%) | 4.46% (3.97%) | 4.06% (3.83%) | 4.01% (3.01%) |

TABLE VI
WIND ENERGY: MODEL F ERROR METRICS AND IMPROVEMENT OVER MODEL T1 FOR DIFFERENT TRAIN DATASET SIZES

| Metric | Train dataset size | | | |
|---|---|---|---|---|
| | 3 months | 6 months | 9 months | 12 months |
| | Error (Improvement) | Error (Improvement) | Error (Improvement) | Error (Improvement) |
| MAE | 6.43% (6.44%) | 6.18% (4.48%) | 6.04% (3.74%) | 5.94% (4.47%) |
| RMSE | 9.09% (6.29%) | 8.80% (5.16%) | 8.62% (4.46%) | 8.51% (4.83%) |
| CRPS | 4.51% (5.69%) | 4.34% (4.97%) | 4.23% (4.70%) | 4.18% (4.66%) |

## V. CONCLUSIONS

This paper proposed a feature engineering methodology to extract more information from a NWP grid for both wind and solar energy forecast. For solar energy, considering all the NWP grid forecasts, a combination of temporal and spatial indexes computed for the shortwave flux and PCA applied to the cloud cover at different heights provided the best results. This led to a reduction in the overall sharpness of probabilistic forecasts, more substantial for clear-sky days, removing some of the abnormal high uncertainty levels verified in those cases.

Concerning wind energy, the PCA approach leads to better results than the use of spatial and temporal indexes separately. The coupling of this different information led to a significant improvement on deterministic and probabilistic forecast.

This work is a step towards big data approaches for RES forecasting that seek to use the maximum number of variables (and information) to improve the forecast skill. The results presented in this paper showed that the adequate extraction of features from the raw NWP data could improve the forecasting skill and the renewable energy forecaster should invest time

in this data mining phase, in addition to the choice of the statistical learning algorithm. Therefore, the future trend is twofold, feature extraction with advanced techniques and machine learning algorithms that can handle the maximum number of features. This will pave the way for the application of deep learning techniques to this problem.

## REFERENCES

[1] M. Taylor, P. Ralon, and A. Ilas, "The power to change: solar and wind cost reduction potential to 2025," International Renewable Energy Agency (IRENA), Tech. Rep., June 2016.

[2] A. Tuohy, J. Zack, S. E. Haupt, J. Sharp, M. Ahlstrom, S. Dise, E. Grimit, C. Mohrlen, M. Lange, M. G. Casado, J. Black, M. Marquis, and C. Collier, "Solar forecasting: Methods, challenges, and performance," *IEEE Power and Energy Magazine*, vol. 13, no. 6, pp. 50–59, November/December 2015.

[3] M. G. Lobo and I. Sánchez, "Regional wind power forecasting based on smoothing techniques, with application to the spanish peninsular system," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 1990–1997, Nov. 2012.

[4] Z. Ziadi, M. Oshiro, T. Senjyu, A. Yona, N. Urasaki, T. Funabashi, and C.-H. Kim, "Optimal voltage control using inverters interfaced with PV systems considering forecast error in a distribution system." *IEEE Transactions on Sustainable Energy*, vol. 5, no. 2, pp. 682–690, April 2014.

[5] P. Bacher, H. Madsen, and H. A. Nielsen, "Online short-term solar power forecasting," *Solar Energy*, vol. 83, no. 10, pp. 1772–1783, October 2009.

[6] E. Lorenz, J. Hurka, D. Heinemann, and H. G. Beyer, "Irradiance forecasting for the power prediction of grid-connected photovoltaic systems," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 2, no. 1, pp. 2–10, March 2009.

[7] J. B. Bremnes, "Probabilistic wind power forecasts using local quantile regression," *Wind Energy*, vol. 7, no. 1, pp. 47–54, Jan.-Mar. 2004.

[8] R. Bessa, V. Miranda, A. Botterud, J. Wang, and E. M. Constantinescu, "Time adaptive conditional kernel density estimation for wind power forecasting," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 4, pp. 660–669, Oct. 2012.

[9] S. Alessandrini, L. D. Monache, S. Sperati, and J. Nissen, "A novel application of an analog ensemble for short-term wind power forecasting," *Renewable Energy*, vol. 76, no. 768-781, April 2015.

[10] S. Alessandrini, L. D. Monache, S. Sperati, and G. Cervone, "An analog ensemble for short-term probabilistic solar power forecast," *Applied Energy*, vol. 157, pp. 95–110, November 2015.

[11] Y. Zhang and J. Wang, "K-nearest neighbors and a kernel density estimator for GEFCom2014 probabilistic wind power forecasting," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1074–1080, Jul.-Sept. 2016.

[12] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. M. de Pison, and F. Antonanzas-Torres, "Review of photovoltaic power forecasting," *Solar Energy*, vol. 136, pp. 78–111, October 2016.

[13] J. Jung and R. P. Broadwater, "Current status and future advances for wind speed and power forecasting," *Renewable and Sustainable Energy Reviews*, vol. 32, pp. 762–777, March 2014.

[14] R. Bessa, A. Trindade, C. Silva, and V. Miranda, "Probabilistic solar power forecasting in smart grids using distributed information," *International Journal of Electrical Power & Energy Systems*, vol. 72, pp. 16–23, November 2015.

[15] A. Tascikaraoglu, B. M. Sanandaji, G. Chicco, V. Cocina, F. Spertino, O. Erdinc, N. G. Paterakis, and J. P. Catalo, "Compressive spatio-temporal forecasting of meteorological quantities and photovoltaic power," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 3, pp. 1295–1305, July 2016.

[16] L. Cavalcante, R. J. Bessa, M. Reis, and J. Browell, "LASSO vector autoregression structures for very short-term wind power forecasting," *Wind Energy*, In Press 2016.

[17] J. Dowell and P. Pinson, "Very-short-term probabilistic wind power forecasts by sparse vector autoregression," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 763–770, March.

[18] J. Tastu, P. Pinson, and H. Madsen, "Multivariate conditional parametric models for a spatio-temporal analysis of short-term wind power forecast errors," in *Proceedings of the European Wind Energy Conference (EWEC 2010)*, Warsaw, Poland, April 2010.

[19] M. He, L. Yang, J. Zhang, and V. Vittal, "A spatio-temporal analysis approach for short-term forecast of wind farm generation," *IEEE Transactions on Power Systems*, vol. 29, no. 4, pp. 1611–1622, July 2014.

[20] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond," *International Journal of Forecasting*, vol. 32, no. 3, pp. 896–913, Jul.-Sept. 2016.

[21] G. I. Nagy, G. Barta, S. Kazi, G. Borbély, and G. Simon, "GEF-Com2014: Probabilistic solar and wind power forecasting using a generalized additive tree ensemble approach," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1087–1093, July?September 2016.

[22] M. Landry, T. P. Erlinger, D. Patschke, and C. Varrichio, "Probabilistic gradient boosting machines for GEFCom2014 wind forecasting," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1061–1066, Jul.-Sept. 2016.

[23] J. Huang and M. Perry, "A semi-empirical approach using gradient boosting and k-nearest neighbors regression for GEFCom2014 probabilistic solar power forecasting," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1081–1086, July-September 2016.

[24] R. Juban, H. Ohlsson, M. Maasoumy, L. Poirier, and J. Z. Kolter, "A multiple quantile regression approach to the wind, solar, and price tracks of GEFCom2014," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1094–1102, July-September 2016.

[25] S. Fang and H.-D. Chiang, "A high-accuracy wind power forecasting model," *IEEE Transactions on Power Systems*, In Press, 2016.

[26] M. P. Almeida, O. Perpinán, and L. Narvarte, "PV power forecast using a nonparametric PV model," *Solar Energy*, vol. 115, pp. 354–368, May 2015.

[27] F. Davò, S. Alessandrini, S. Sperati, L. D. Monache, D. Airoldi, and M. T. Vespucci, "Post-processing techniques and principal component analysis for regional wind power and solar irradiance forecasting," *Solar Energy*, vol. 134, pp. 327–338, September 2016.

[28] C. Monteiro, R. Bessa, V. Miranda, A. Botterud, J. Wang, and G. Conzelmann, "Wind power forecasting: state-of-the-art 2009," Argonne National Laboratory, Tech. Rep. ANL/DIS-10-1, 2009.

[29] J. Snoek and H. Larochelle, "Practical bayesian optimization of machine learning algorithms," in *26th Annual Conference on Neural Information Processing Systems (NIPS)*, Nevada, USA, 3-6 Dec. 2012, pp. 2951–2959.

[30] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.

[31] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Berlin: Springer, Jan. 2013.

[32] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

**Jose R. Andrade** was born in 1992 in Sta. Maria da Feira, Portugal, received his M.S. degree from the Faculty of Engineering of the University of Porto, Portugal (FEUP) in 2016 in Electrical and Computer Engineering. Currently, he is a Researcher at INESC TEC in its Center for Power and Energy Systems. His research interests include renewable energy and electricity prices forecasting, as well as offshore generation technologies.

**Ricardo J. Bessa** was born in 1983 in Viseu, Portugal, received his Licenciado (five-year) degree from the Faculty of Engineering of the University of Porto, Portugal (FEUP) in 2006 in Electrical and Computer Engineering. In 2008, he received the M.Sc. degree in Data Analysis and Decision Support Systems on the Faculty of Economics of the University of Porto (FEP). He obtained his Ph.D. degree in the Doctoral Program in Sustainable Energy Systems (MIT Portugal) at FEUP in 2013. Currently, he is a Senior Researcher and Area Manager at INESC TEC in its Center for Power and Energy Systems.

His research interests include renewable energy, energy analytics, smart power systems and electricity markets. He worked in several international projects, such as the European Projects FP6 ANEMOS.plus, FP7 SuSTAINABLE, FP7 evolvDSO, Horizon 2020 UPGRID, and an international collaboration with Argonne National Laboratory for the U.S. Department of Energy.