

Spatial–Temporal Solar Power Forecasting for Smart Grids

Ricardo J. Bessa, Artur Trindade, and Vladimiro Miranda, *Fellow, IEEE*

Abstract—The solar power penetration in distribution grids is growing fast during the last years, particularly at the low-voltage (LV) level, which introduces new challenges when operating distribution grids. Across the world, distribution system operators (DSO) are developing the smart grid concept, and one key tool for this new paradigm is solar power forecasting. This paper presents a new spatial–temporal forecasting method based on the vector autoregression framework, which combines observations of solar generation collected by smart meters and distribution transformer controllers. The scope is 6-h-ahead forecasts at the residential solar photovoltaic and medium-voltage (MV)/LV substation levels. This framework has been tested in the smart grid pilot of Évora, Portugal, and using data from 44 microgeneration units and 10 MV/LV substations. A benchmark comparison was made with the autoregressive forecasting model (AR—univariate model) leading to an improvement on average between 8% and 10%.

Index Terms—Distribution network, forecasting, smart grid, smart metering, solar power, spatial–temporal.

I. INTRODUCTION

THE PENETRATION of solar-based generation is reaching nonmarginal levels in numerous power systems. For example, by the end of 2012, the installed solar power in Germany, Italy and Japan was around 32.6, 16.7, and 6.9 GW, respectively. In fact, the global solar photovoltaic (PV) capacity has grown from around 1.4 GW in 2000 to over 102 GW in 2012, with 78 GW installed during the 2010–2012 period [1]. In terms of cost, PV is reaching grid parity in many countries, meaning that it can generate electricity at a leveled cost less than or equal than the electricity retailing tariffs [2]. In this

Manuscript received January 23, 2014; revised May 21, 2014, July 28, 2014, and September 15, 2014; accepted October 23, 2014. Date of publication October 29, 2014; date of current version February 02, 2015. This work was made in the framework of the BEST CASE project (“NORTE-07-0124-FEDER-000056”); supported in part by the North Portugal Regional Operational Programme (ON.2—O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF); in part by the national funds, through Fundação para a Ciência e a Tecnologia (FCT); in part by the 7th Research and Technological Development (RTD) Framework Programme within the SuSTAINABLE project under Contract 308755; and in part by the COMPETE Programme and the FCT within projects SMAGIS—PTDC/SEN-ENR/113094/2009 and DYMONDS—CMU-PT/SIA/0043/2009. Paper no. TII-14-0090.

R. J. Bessa is with the INESC TEC—Instituto de Engenharia de Sistemas e Computadores Tecnologia e Ciência, 4200-465 Porto, Portugal (e-mail: ricardo.j.bessa@inesctec.pt).

A. Trindade and V. Miranda are with the INESC TEC and Faculty of Engineering of the University of Porto (FEUP), 4200-465 Porto, Portugal (e-mail: ee08089@fe.up.pt; vladimiro.miranda@inesctec.pt).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TII.2014.2365703

context, the deployment of solar PV will likely continue even if financial subsidies are withdrawn.

The majority of this installed capacity is connected to the medium- and low-voltage (MV and LV) distribution grids. The roll-out of the Smart Grid infrastructure provides additional capabilities for monitoring and controlling assets at the distribution grid level and fosters demand-side management and renewable energy integration [3]. This creates conditions to develop a new generation of management tools that maximize the integration of distributed generation at the MV and LV levels [4], such as boosted voltage control [5], state-estimation [6] algorithms, and energy management systems (including an optimal power flow and machine learning algorithms) [7].

Furthermore, a massive deployment of small-scale storage at the residential level (e.g., thermal storage and batteries) might occur if governments create incentives for such goal [8]. For instance, Germany has created financial incentives for owners of solar systems with batteries. At the building and microgrid level, solar PV can be combined with storage (e.g., supercapacitors) and nonrenewable energy microturbines (e.g., gas) using centralized and local energy management functions [9].

The new management tools, and the joint coordination of PV generation and storage at the building and micro-grid levels, require the use of solar and load power forecasts for several hours ahead. The time-horizon of interest for power system operations and electricity markets can be divided into two classes [10]: 1) very short-term; (up to 6 h ahead); and 2) short-term (up to 3 days ahead).

For short-term load forecast in a smart grid environment, Borges *et al.* [11] propose three different methods: (top-down) adding up single-observed consumptions and perform a global forecast; (bottom-up) adding up the sum of individual forecasts for each load in order to create a global forecast; (regression method) regression of the individual loads recorded by the meters. The main goal was to evaluate combined forecasting.

In recent publications about solar PV, several works combine statistical/machine learning algorithms with numerical weather predictions (NWP) to produce solar power forecasts for the short-term horizon. Bacher *et al.* [12] describe a two-stage forecasting approach. First, a clear sky model, based on weighted quantile regression, is proposed to remove the diurnal component of solar generation and global irradiance. Then, an autoregressive model with exogenous inputs (ARX) is used to combine past observations of solar power with NWP. According to the authors, up to 2 h ahead, the most important inputs are the past observations, while for a horizon up to 36 h ahead, the NWP prevail. Fernandez-Jimenez *et al.* [13] also used NWP as input in several machine-learning algorithms

(i.e., autoregressive-integrated moving average (ARIMA), k -nearest neighbors (k NNs), neural networks (NN), and adaptive neurofuzzy models) to produce solar power forecasts for the next 39 h. The best performance was obtained with a NN.

For the very short-term horizon, which is addressed in this paper, two different classes of models can be found. The first class is based on satellite images. Hammer *et al.* [14] describe an algorithm based on cloud-index images that are predicted with motion vector fields derived from two consecutive images.

The second class consists of univariate time-series models. Pedro and Coimbra [15] compared the performance of different machine-learning algorithms (i.e., ARIMA, k NNs, NN, and NN optimized by genetic algorithms), which only use past observations of the time-series as inputs. The NN, combined with genetic algorithms, obtained the best performance. Huang *et al.* [16] combine an AR model with a dynamical system model (i.e., resonating model introduced by Lucheroni) for 1-h-ahead forecast of global solar radiation. The proposed method outperformed other methods such as NN and time-delay NN. An extended overview for the solar power forecast literature can be found in [17].

The forecasting framework presented in this paper addresses the very short-term horizon and is included in the second class of models. Compared to [15] and [16], the main advantage is that the new proposed method is spatial-temporal, since it combines the past observations of time-series distributed in space (i.e., collected by smart meters) in order to capture the effect of cloud movement. On the other hand, compared to models based on NWP (such as [13]), the accuracy for time horizons greater than 4 h is lower. Thus, it is only suitable for very short-term horizons. The use of satellite images (like in [14]) can provide information about cloud movement in the local area for very short-term horizons, which also improves the forecasting accuracy during the first few hours. However, the main constraint is that satellite images must be available in almost real time and it might be more complex (and expensive) to operationalize such forecasting service. It should be stressed that information from spatially distributed smart meters (or sensors) is readily available in a smart grid and it can be combined with satellite information.

To our knowledge, only two works combined information from neighboring sites to improve solar power forecast. Berdugo *et al.* [18] described a method based on searching similar local and global current states, considering neighbor sites. Compared to our proposed method, the main disadvantage is that the goal is not to produce the “optimal” forecast (i.e., with minimum error); instead, it is to handle distributed data streams and maintains power measurements’ privacy. In this sense, our approach also handles data streaming but seeks, at the same time, the “best” forecast accuracy. Yang and Xie [19] proposed an ARX model for each solar site where the exogenous variables are measurements from neighbor sites. The main disadvantage, compared to our method, is that the relevant exogenous variables are not selected and the vector autoregression (VAR) framework is not explored. Therefore, this paper proposes three original contributions:

- 1) a new forecasting method, constructed on the top of a smart grid infrastructure, that combines VAR and gradient

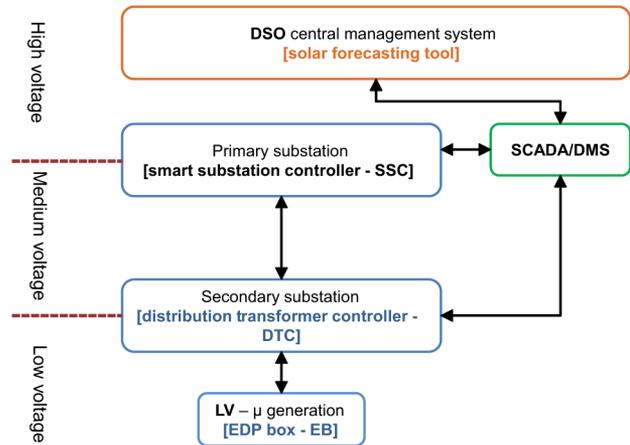


Fig. 1. Smart grid architecture in Portugal.

boosting (GB) frameworks to explore information from the distributed PV panels;

- 2) improved accuracy, at the MV/LV level, by introducing exogenous variables to the model, i.e., observations from microgeneration smart meters;
- 3) an online fitting method, based on recursive least squares (RLS), for the VAR framework. A sparse model, in terms of model’s coefficients, is also explored with the GB approach.

This forecasting framework will be applied to produce 6-h-ahead forecasts for each residential PV and secondary substation (i.e., MV/LV substation). The proposed method can operate as a centralized forecasting system to be used by a distribution system operator (DSO) for managing distributed energy resources or by a solar power aggregator for participating in intraday electricity markets.

The paper is organized as follows. Section II presents the smart grid infrastructure in Portugal. Section III describes the solar power forecasting framework. The test case results are presented in Section IV. Section V presents the conclusion and future work.

II. SMART GRID ARCHITECTURE IN PORTUGAL

As a followup of Smart Grid pilots conducted by several DSO, a generalized deployment of these technological solutions is expected for the following years. An example is the InovGrid Project in Portugal [20], promoted by Energias de Portugal (EDP) Distribution, aiming to develop new Information and Communications Technology (ICT) and computational tools for automating network management and to create a full smart distribution grid. This project resulted in a large-scale demonstration pilot in the city of Évora in Portugal, named InovCity [21]. This city is also one demonstration site of the EU Project SuSustainable [4].

The main components of this infrastructure, depicted in Fig. 1, are the EDP Box (EB) and the distribution transformer controller (DTC). The EB is a smart meter device installed at the consumer/producer premises that include a measurement module, control module, and communications module. It can

interact with other devices through a home area network. The DTC is a local control equipment located at the secondary substation level, comprising components for measurement, remote control, and communication actions. Its main functions are collecting data from EB and MV/LV substation, data analysis functions, grid monitoring, and interface with commercial and technical central systems.

Both the EB and the DTC are part of a hierarchical control and communication architecture. Each EB has a bidirectional communication with the corresponding DTC through General Packet Radio Service (GPRS) or Power Line Communications, and the DTC communicates with the System Control and Data Acquisition/Distribution Management System (SCADA/DMS) through a wide-area network based on GPRS [21]. Further communication technologies, such as radio frequency mesh modules, can be explored in this architecture.

At the primary substation level (i.e., HV/MV substation), a Smart Substation Controller (SSC) is installed, which is responsible for aggregating and managing the operational data from EB and DTC, and for applying demand/generation management and self-healing strategies. In this hierarchical architecture, the SSC is responsible for managing the MV grid and communicates with the DSO central system via the SCADA/DMS system.

On the top of these technologies, there are data management services capable of handling large volumes of data and, at the same time, providing an overview of all existing devices.

In this architecture, the forecasting system is installed at the central management level and explores time-series information from DTC and EB connected to different primary/secondary substations.

The outputs are forecasts for each DTC (groups of consumers with PV generation) and EB point (individual consumer with PV panels). For the forecast at the DTC level, the EB measurements can be used as distributed sensors to capture the spatial-temporal effect of clouds in solar generation and consequently improve the forecasts, which in turn increase the amount of transmitted data.

Finally, it is important to stress that the forecasting system is also valid for a nonhierarchical architecture, such as one where each device communicates directly with the DSO control center via Internet Protocol using an internet connection [22]. However, one has to pose a centralized data flow topology as a requirement. Moreover, the proposed forecasting framework can also take advantage of information collected by wireless sensor networks [23] associated to pyranometers that measure global and direct solar irradiance.

III. SPATIAL-TEMPORAL FORECASTING MODEL

The forecasting framework described in this section is constructed on the top of the Smart Grid infrastructure of the InovCity pilot that was described Section II.

A. Seasonal Detrending of the Time Series

The solar power time series present a seasonal pattern dependent on the time of the day and day of the year, and most

classical models, such as AR and VAR, assume stationary time series.

In the literature, different physical and statistical approaches are proposed to estimate the deterministic variation of the solar irradiance (i.e., excluding the influence of clouds and other factors) [17]. One of these statistical approaches is the clear-sky model described in [12] that, based on weighted quantile regression, is directly applied to solar power time series. The method is described as a statistical normalization, capable of generating a stationary time series with normalized solar power. Details of the clear-sky model are given in the Appendix, and following the results in [12], the selected predictors are the time of the day (h) and day of the year (doy).

The output of the clear-sky model is clear-sky solar generation (\hat{p}_t^{cs}), which is used to normalize the measured solar power (p_t) as follows:

$$p_t^{\text{norm}} = \frac{p_t}{\hat{p}_t^{cs}}. \quad (1)$$

The model's parameters are the kernel bandwidths σ_h and σ_{doy} as well as the quantile τ . These parameters are determined by trial-error experiences [12], inducing a value equal to one for the variable p_t^{norm} during clear-sky days.

B. Vector Autoregressive Framework

The ARIMA process is a well-known class for univariate time-series models [24]. A special model of this class is the AR, in which the predictors for time interval t are past observations (or lags) of the univariate time series. For 1-h-ahead forecast, the AR model is

$$\hat{p}_{t+1|t} = \alpha + \beta_1 \cdot p_t + \beta_2 \cdot p_{t-1} + \dots + \beta_l \cdot p_{t-l} + e_{t+1|t} \quad (2)$$

where β is the model's coefficient, α is a constant (or intercept) term, l is the order of the AR model, and $e_{t+1|t}$ is a contemporaneous white noise (or residuals) with zero mean and constant variance σ_e^2 .

The main limitation of this model is that it only uses, as predictors, the past observations from the response variable. This model can be extended with exogenous variables (such as NWP), forming an ARX model. Nevertheless, as mentioned in the literature [12] and [17], NWP can only improve the forecast error for time horizons greater than 3 or 4 h ahead, while for shorter time horizons, the relevant information consists of time-series observations or satellite frames.

In order to improve the forecasts for the very short-term horizon, a VAR framework [25] is used to combine past observations from the solar power in each site with past observations from neighbor sites. This consists of a spatial-temporal (or multioutput) linear regression model with N observations, q -dimensional response, and d -dimensional predictors.

In matrix form, for one step-ahead forecast, it is given by

$$\hat{P}_{t+1|t} = \alpha + B \cdot P_{t-l} + E_{t+1|t} \quad (3)$$

where $\hat{P}_{t+1|t}$ is the response matrix with dimension $N \cdot q$, B is the coefficient matrix with dimension $d \cdot q$, P_{t-l} is the predictor (lagged terms) matrix with dimension $N \cdot d$, α is a vector

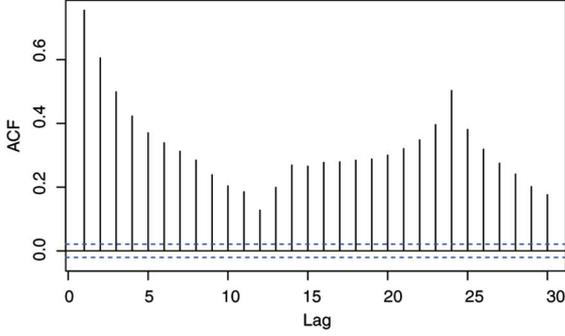


Fig. 2. Autocorrelation function (ACF) plot of the normalized solar power for one DTC time series with hourly resolution.

with q constant terms, E_{t+1} is a matrix with dimension $N \cdot q$ containing i.i.d. residuals with zero mean and constant covariance Σ_e .

Equation (3), for an example with two response variables (i.e., sites 1 and 2) and two lagged terms, becomes

$$\begin{aligned} \hat{p}_{t+1|t,1} &= \alpha_1 + \beta_{1,1} \cdot p_{t,1} + \beta_{1,2} \cdot p_{t-1,1} + \beta_{1,3} \cdot p_{t,2} \\ &\quad + \beta_{1,4} \cdot p_{t-1,2} + e_{t+1|t,1} \\ \hat{p}_{t+1|t,2} &= \alpha_2 + \beta_{2,1} \cdot p_{t,1} + \beta_{2,2} \cdot p_{t-1,1} + \beta_{2,3} \cdot p_{t,2} \\ &\quad + \beta_{2,4} \cdot p_{t-1,2} + e_{t+1|t,2}. \end{aligned} \quad (4)$$

Note that in (4), the forecasted power at site 1 for the lead time $t + 1$ is expressed as a function of past observations at sites 1 and 2 (i.e., a neighbor site). The same is valid for the forecasted power for site 2 (the past values of site 1 are included in the model).

In order to apply the forecasting techniques from (3) and (4), in a first phase, it is necessary to normalize the solar power time series with the clear-sky model from Section III-A. In [12], it is recommended to remove the small \hat{p}_t^{CS} values, since for these values, p_t^{norm} increases considerably and reaches infinity in the nighttime. In the work reported in this paper, the normalized solar power values outside the period 7 h:00–19 h:00 (i.e., the average period with almost no solar generation during the whole year in Portugal) are removed for all sites.

In the second phase, the normalized solar power values are used to fit the AR and VAR frameworks (i.e., estimate the coefficients β) with one of the methods that are described in Sections III-C and III-D.

The VAR model is applied to forecast the solar power for each DTC and EB. Furthermore, when specifying an AR or VAR model, it is important to determine how many lagged terms need to be included. This can be achieved with the following method.

- 1) First, the autocorrelation plot (depicted in Fig. 2 for one DTC with an hourly time resolution) of the normalized time series is analyzed to make a coarse estimation of the necessary lags.
- 2) Then, the autocorrelation plot of the residuals is analyzed to check if the residuals are i.i.d. (i.e., white noise).

Fig. 2 was created with the full time series of normalized power, but excluding the hours with zero clear-sky generation and where both clear sky and observed power are zero.

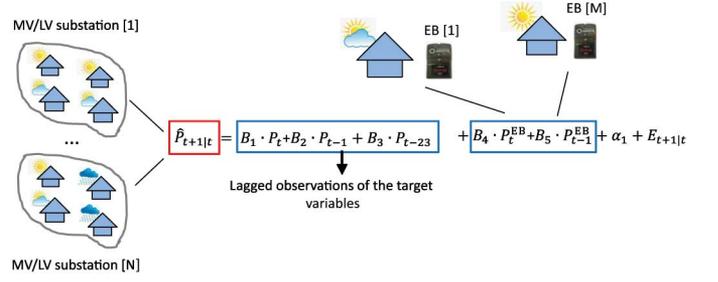


Fig. 3. VARX model for lead time $t + 1$ with EB observations.

In Fig. 2, the exponential decay of the first few lags indicates that at least the first lag should be included in the model, while the peak at lag 24 with an exponential decay suggests the inclusion of a diurnal effect. Note that, by increasing the order of the model (i.e., including more lagged terms), it is possible to remove the serial dependency of the residuals [26]. This is the reason why the second lag is included in the model. From these analyses, the lags 1 and 2, as well as a diurnal effect (i.e., seasonal daily variation), were selected, which is consistent with the literature [12]. However, the square error in the training dataset is calculated using a cross-validation approach in order to check if the diurnal effect should be included or not for each lead time (more details are given for the test case in Section IV).

Since the goal is to produce 6-h-ahead forecasts, a different AR and VAR model is fitted for each lead time. For instance, for lead times 2 and 6, the VAR model has the following form:

$$\hat{P}_{t+2|t} = \alpha_2 + B_1 \cdot P_t + B_2 \cdot P_{t-1} + B_3 \cdot P_{t-22} + E_{t+2|t} \quad (5)$$

$$\hat{P}_{t+6|t} = \alpha_6 + B_1 \cdot P_t + B_2 \cdot P_{t-1} + B_3 \cdot P_{t-18} + E_{t+6|t} \quad (6)$$

where the terms P_t and P_{t-1} (lags 1 and 2) remain the same, and the seasonal effect associated to the previous day changes with the lead time.

Finally, in addition to the AR and VAR frameworks, a VAR with exogenous variables (VARX) is also proposed and tested. The model consists of adding exogenous variables to (3): the solar power values observed in each EB.

The goal is to assess if the EB measurements (P_t^{EB}) can improve the solar power forecast at the DTC level. The VARX for lead time $t + 1$ has the structure depicted in Fig. 3. Note that, for the EB observations, only the past observations t and $t - 1$ are included in the model since the goal is to use the EB as distributed sensors that characterize the current atmospheric conditions (in terms of solar power) across the region.

C. Ordinary Least-Squares Fitting

As shown in (4), the VAR consists of linear regression models, in which the $\hat{p}_{t+k|t}$ of each site depends on a constant term and lagged terms of the q response variables. Note that each regression equation takes the same matrix of predictors (P_{t-1}) and a vector of coefficients (a row from matrix B) is estimated from data. However, this model is capable of modeling the dynamic relation among spatially distributed time series.

In fact, (3) and (4) have the form of a seemingly unrelated regression (SUR) model.

Estimating the coefficients of the VAR model is straightforward, i.e., the ordinary least squares (OLS) can be applied if the same predictors appear in every equation [25]. In such a case, OLS gives the same solution of generalized least squares and can be applied independently to each regression equation.

The ICT infrastructure of a Smart Grid generates a continuous stream of data that must be handled in quasi real time and with low data storage requirements. The AR and VAR models described in the previous section can be fitted in quasi real time using the RLS method with a *forgetting factor* λ [27]; λ is a weighting factor; it intends to ensure that the observations in the distant past are forgotten in order to enable following the statistical variations of the observable data in a nonstationary environment. This overcomes the problem of handling “big data” since it is not necessary to store historical data for fitting (or refitting) the model and tracks changes in the dynamics of the data generating structure, such as loss of performance due to dust in PV panels or changes in the surrounding environment (e.g., shadows).

Since both VAR and AR can be fitted with OLS, the RLS method can also be applied to this model and it is of great importance since the spatial–temporal relation between PV sites is very dynamic and requires time-varying coefficients.

The update of the parameters of the VAR model [using notation from (3)] is performed with the RLS method as follows for time step t :

$$B_t = B_{t-1} + K_t \cdot [P_t - (\alpha + B_{t-1} \cdot P_{t-1})] \quad (7)$$

where K_t is given by

$$K_t = Q_t \cdot P_{t-1} \quad (8)$$

and Q_t by

$$Q_t = \frac{1}{\lambda} \cdot \left[Q_{t-1} - \frac{Q_{t-1} \cdot P_{t-1} \cdot P_{t-1}^T \cdot Q_{t-1}}{\lambda + P_{t-1}^T \cdot Q_{t-1} \cdot P_{t-1}} \right]. \quad (9)$$

B_{t-1} is the coefficient matrix from time step t and B_t is the matrix with updated coefficients (i.e., after receiving the observed solar power during time step t). K_t can be interpreted as a gain vector, meaning that the parameter estimates decrease along a line in parameter space, determined by the gain vector; Q_t can be interpreted as a covariance matrix [28]. The demonstration of (7)–(9) and corresponding properties can be found in [27] and [28].

A λ equal to 1 leads to a recursive estimation of the coefficients, while a smaller value discounts old data with an exponential decay.

The RLS algorithm is used to update the coefficients of each lead time [e.g., coefficients B of (5) and (6)]. This fitting method perfectly copes with streaming data since at time step t only B_t , K_t , and Q_t have to be stored in memory, and the remaining data are dropped.

This algorithm requires some initial values for B_0 and Q_0 . A simple and robust approach is to initialize B_0 with zeros and Q_0 as a diagonal matrix with a large constant value.

D. Gradient Boosting Fitting

Boosting is an ensemble machine-learning algorithm for classification and regression, which combines base (or *weak*) learners. Friedman proposed, in [29], one variant of boosting, named Gradient Boosting (GB). GB conducts numerical optimization, via steepest-descent, in function space by using a user-defined base learner recurrently on modified data that are the output from the previous iterations. Following the optimization phase, the final solution $\hat{F}(x)$ is a linear combination of the base learners, as follows:

$$\hat{F}(x) = \hat{f}_0(x) + \sum_{m=1}^M \hat{f}_m(x) \quad (10)$$

where $\hat{f}_0(x)$ is an initial guess, $\hat{f}_m(x)$ are the base learners, and M is the maximum number of “boosts” (which is a model’s parameter). In (3), the function $\hat{F}(x)$ corresponds to $\alpha + b \cdot P_{t-1}$, where b is a column of matrix B .

Several authors noted that successive boosting steps reduce the bias (at a cost of a slightly increase in variance), particularly if the base learner has low variance and high bias, which is the case of a linear model [30]. Therefore, by setting a value for M , the bias–variance tradeoff is being controlled.

Bühlmann in [30] propose the component-wise linear least squares, which is a GB algorithm where the base learner selects only one predictor among all the d -predictors. In the VAR framework, the component-wise GB can be applied to each equation [e.g., (4)] of the VAR model. It is an alternative to the OLS method and performs automatic variable selection and coefficients shrinkage. For the solar power forecast problem, the empirical risk function is the quadratic loss, and the base learner is a linear effect of a continuous predictor [like in (3)].

The component-wise GB applied to each q -response variable of the VAR framework works as follows.

- 1) Initialize $\hat{f}_0(P_{t-1})$ with the mean value of the $p_{t,q}$ response variable.
- 2) For each m , given $\hat{F}_{m-1}(P_{t-1})$, calculate the residuals $u[i] = p_{t,q}[i] - \hat{F}_{m-1}(P_{t-1}[i])$, where i takes values between 1 and N (number of training samples).
- 3) Using the residuals calculated in step 2) as the response variable, produce an estimate of the coefficients associated to each candidate base learner (i.e., $\beta_j \cdot P_{t-1}^{(j)}$)

$$\beta_j = \frac{\sum_{i=1}^N u[i] \cdot P_{t-1}^{(j)}[i]}{\sum_{i=1}^N \left(P_{t-1}^{(j)}[i] \right)^2}. \quad (11)$$

- 4) Determine the s th predictor or base learner (from a set with d candidates) that minimizes the quadratic loss function

$$s = \arg \min_{1 \leq j \leq d} \sum_{i=1}^N \left(u[i] - \beta_j \cdot P_{t-1}^{(j)}[i] \right)^2 \quad (12)$$

which gives the selected base learner: $\hat{f}_m(P_{t-1}^{(s)}) = \beta_s \cdot P_{t-1}^{(s)}$.

- 5) Update function $\hat{F}_m(P_{t-1})$ as follows:

$$\hat{F}_m(P_{t-1}) = \hat{F}_{m-1}(P_{t-1}) + \nu \cdot \hat{f}_m(P_{t-1}^{(s)}) \quad (13)$$

where ν is a shrinkage parameter.

6) Stop when $m = M$.

The final estimator is obtained in step 6). The GB method has two parameters that need to be set: 1) maximum number of boosting iterations (M) and 2) shrinkage parameter (ν).

The value of M is estimated through fivefold cross validation, where the m value with the lowest square error is selected for each lead time and response variable. According to [30], the value of ν does not influence significantly the results if set to be a low value (e.g., 0.1).

In order to cope with structural changes in the time series, a sliding-window approach is used to fit the VAR and VARX frameworks using the GB algorithm. In terms of data storage, this approach is not efficient as the RLS method. However, the component-wise GB has two important advantages over RLS, which are relevant for the problem addressed in this paper.

First, it is particularly suitable for high-dimensional problems since the variable selection capability [step 4) of the algorithm] can lead to a sparse B matrix of coefficients. This sparse matrix can reduce the data flow volume from EB; for instance, in the VARX model of Fig. 3, component-wise GB can reduce the set of EB measurements that are needed to improve the forecasts at the DTC level. Second, it provides a framework where different loss functions can be included, which enables probabilistic forecasting through quantile regression in the framework described in this paper.

IV. TEST CASE RESULTS

A. Description

The solar power dataset used as test case is from the municipality of Évora, with 1307 km² of area. During 2 years, more than 30,000 EB and 300 DTC were installed, including all customers and substations, in order to have the entire municipality covered. In what concerns distributed generation, in January 2014, there were 220 microgeneration producers, mainly solar PV, with a total installed power of 768 kW.

In order to test the proposed forecasting framework, time series from 44 EB were used. These were the time series with better quality, or in other words, the ones with the lowest number of missing values and hours with zero generation due to maintenance operation or communication problems. The EB data comprise domestic PV, with installed capacity ranging between 1.1 and 3.7 kWp. These EB measurements were related to 10 different DTC, and the total values of each DTC are also forecasted.

The parameters of the clear-sky model, estimated by trial-error in order to guarantee a normalized power equal to one during clear-sky days (see [12]), are $\sigma_h = 0.01$, $\sigma_{\text{day}} = 0.01$, and $\tau = 85\%$. The *forgetting factor* λ , estimated by fivefold cross validation for both AR and VAR, is equal to 0.999. A ν equal to 0.2 is used in the GB algorithm since it leads to the best results.

The original data were sampled in 15 min, but it was resampled to hourly values (i.e., the same length of the electricity market). The period between February 1, 2011 and 31 January 31, 2012 was used to fit the models and it is also the size of the sliding window. The period between February 1, 2012 and

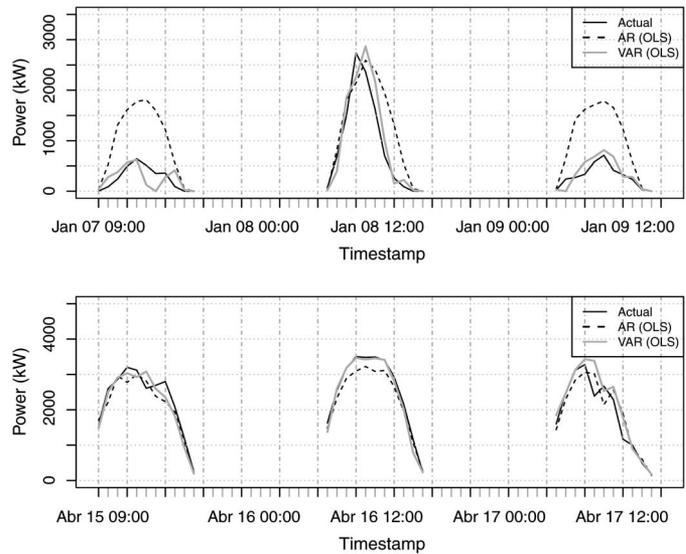


Fig. 4. One step-ahead solar power forecast obtained with the AR and VAR models trained with RLS for one EB.

March 6, 2013 was used to calculate the forecast errors (i.e., test period).

The forecasting results are evaluated with the root-mean-square error (RMSE) calculated for the k th lead time [12]

$$\text{RMSE}_k = \sqrt{\frac{1}{N} \sum_{t=1}^N \left(\hat{P}_{t+k|t} - P_{t+k} \right)^2}. \quad (14)$$

The RMSE is normalized with the solar peak power.

The RMSE is calculated separately for each EB or DTC, but it is also calculated using the full dataset of errors as a summary performance metric for all DTC or EB. The performance of two models (AR and VAR) is compared by computing the improvement in terms of RMSE

$$\text{Imp}_k = \frac{\text{RMSE}_{k,AR} - \text{RMSE}_{k,VAR}}{\text{RMSE}_{k,AR}} \cdot 100\%. \quad (15)$$

B. Forecasting Results

For the DTC dataset, and by fivefold cross validation on the fitting period, it was determined that a second-order AR model (i.e., without the diurnal term) achieves the lowest square error for lead times 1 and 2. For the EB dataset, this only occurs for lead time 1. The VAR model includes the diurnal term in all lead times.

A one step-ahead solar power forecast obtained with the AR and VAR models fitted with the RLS method (described in Section III-C) is depicted in Fig. 4 for one EB (with 3.5 kWp) and six different days. For the 3 days in January, the forecast produced by the VAR model fits better the actual observation under cloudy and overcast conditions. In this case, the AR fails because it does not include information from neighbor sites. In the VAR framework, since all the other EB show a low solar-generation level in the past observations, the solar power level for this specific EB is forecasted to be lower than one provided

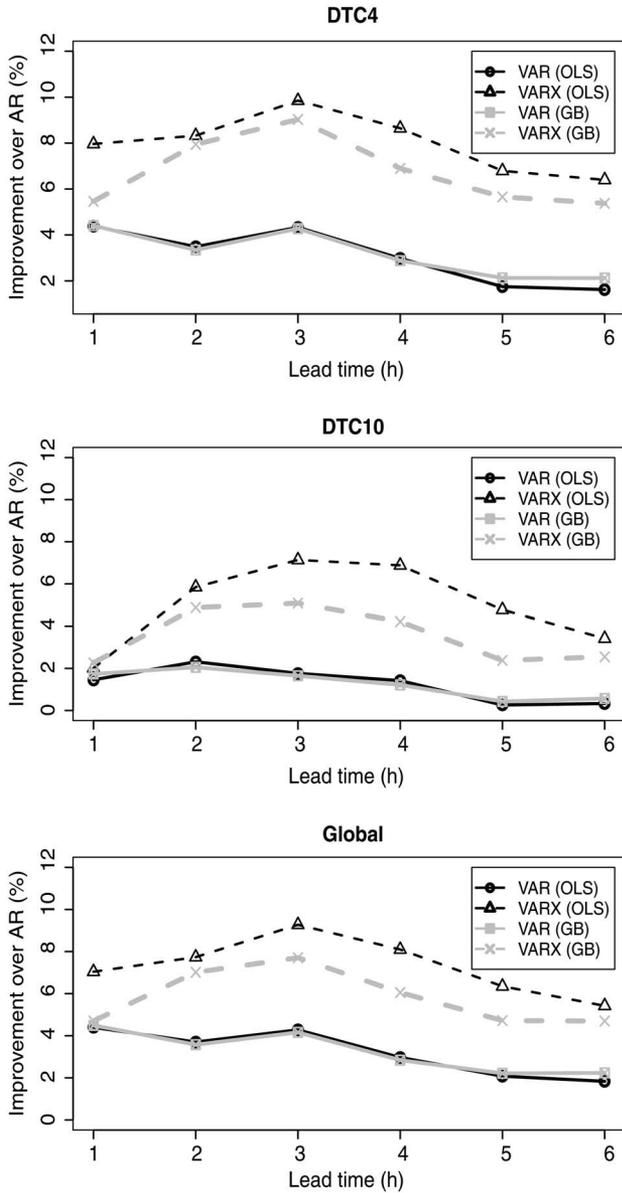


Fig. 5. Imp_k of the VAR and VARX models fitted with OLS and GB for two DTC and for the RMSE_k calculated using the entire set of forecast errors.

by the AR model (which only includes observations from the same EB). For the 3 days in April, the difference between the AR and VAR forecast is not very substantial, however, April 16 was a clear-sky day and the AR, compared to VAR, also shows worst forecast accuracy.

The improvement of the VAR and VARX (estimated with OLS and GB) over the AR model for each lead time is plotted in Fig. 5 for two DTC and for the RMSE_k calculated with the full dataset of DTC forecast errors.

These three plots clearly show that the VARX [OLS] model achieves the highest improvement. From the full set of DTC, number 4 is the one with the highest overall improvement, reaching a value around 8% for the first lead time and around 6% for the sixth lead time. Number 10 is the one with the lowest improvement, ranging between 2% and 7%. The VARX [GB] also presents a higher improvement compared to both

TABLE I

p -VALUES OF THE DIEBOLD–MARIANO STATISTICAL TEST FOR FORECAST ACCURACY PERFORMANCE AT THE DTC LEVEL

Lead time	1	2	3	4	5	6
AR vs. VARX [OLS]	0.00*	0.00*	0.00*	0.00*	0.00*	0.00*
VARX [OLS] vs. VARX [GB]	0.00*	0.02	0.00*	0.00*	0.00*	0.02

*Identifies statistically different accuracies.

VAR [OLS] and VAR [GB], nevertheless lower than the VARX [OLS] model.

The VAR model also achieves a positive improvement in all lead times, but lower than VARX. This means that the EB measurements, used as distributed sensors, can improve the forecast at the DTC level. Note that the performance of VAR [OLS] and VAR [GB] is rather similar.

The global improvement of the VARX [OLS] varies between 9% and 5.7%, which shows the benefit from using spatial-temporal models for very short-term solar power forecasting. It is important to stress that the coefficients matrix of the VARX [GB] is very sparse compared to VARX [OLS]. For instance, in lead time 1, only the coefficients associated to past observations of 3 EB are nonzero, while only 11 coefficients (or EB observations) have nonzero values for lead time 6. This represents a decrease in terms of volume of transmitted data (i.e., only a few reference EB are used) at the cost of a slightly increase in forecast error.

Table I shows the p -value results obtained with the Diebold–Mariano (DB) test [31], which is used to assess the statistical significance of the forecast error improvement. The null hypothesis is “no difference in the accuracy of two competing forecasts,” and if the p -value is less than a significance level (i.e., 0.01 in this paper), then the observed result would be highly unlikely under the null hypothesis. As shown in Table I, the difference in accuracy between the competing forecasts is significant in all lead times. The only exception is between VARX [GB]- and VARX [OLS]-based forecasts for lead times 2 and 6.

Fig. 6 shows the improvement obtained with the VAR model fitted with OLS and GB for two EB and for the RMSE_k calculated with the full dataset of EB forecast errors.

For lead times 1–4, the VAR [OLS] model attained the highest improvement for EB number 16, with a value around 27% for lead time 1 and around 18% for lead time 4. The VAR [GB] attains the same improvement as VAR [OLS] for lead time 5 and a higher one for lead time 6 (i.e., around 14%).

The lowest improvement was attained for EB number 39, with 3.5% obtained with VAR [OLS] for lead time 1 and a negative value of around -1% for lead time 5. The improvement obtained with VAR [GB] is around 1% for all lead times. The global improvement for the EB dataset varies between 10% and 0.4% for VAR [OLS] and between 8% and 5% for VAR [GB]. Table II presents the DB test results. All the differences are statistically significant with the exception of lead time 6 for the difference between AR and VAR [OLS] and lead time 3 for the difference between VAR [OLS] and VAR [GB].

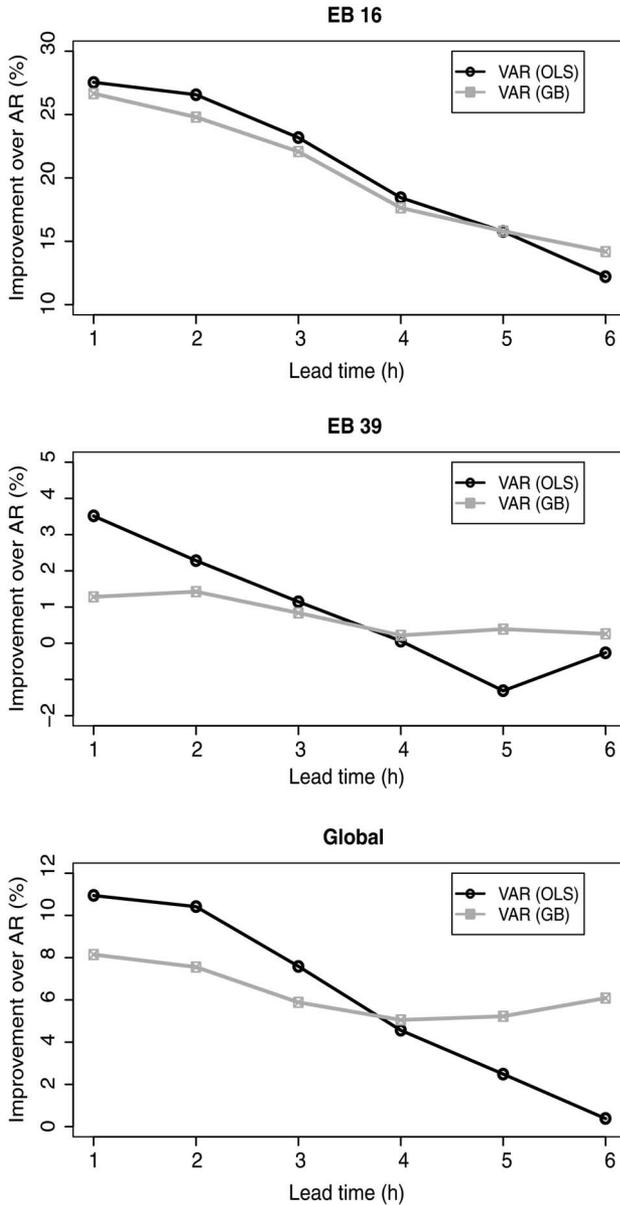


Fig. 6. Imp_k of the VAR model fitted with OLS and GB for two EB and for the $RMSE_k$ calculated using the entire set of forecast errors.

TABLE II
p-VALUES OF THE DIEBOLD-MARIANO STATISTICAL TEST FOR FORECAST ACCURACY PERFORMANCE AT THE EB LEVEL

Lead time	1	2	3	4	5	6
AR vs. VARX [OLS]	0.00*	0.00*	0.00*	0.00*	0.00*	0.26
AR vs. VAR [GB]	0.00*	0.00*	0.00*	0.00*	0.00*	0.00*
VARX [OLS] vs. VARX [GB]	0.00*	0.00*	0.02	0.00*	0.00*	0.00*

*Identifies statistically different accuracies.

As mentioned before, the benefit from using the GB is to get a sparse matrix of coefficients. For the EB dataset, 78% of the coefficients have a null value for lead time 1 and 80% for lead time 6. The GB fitting, compared to OLS, also attains a higher and statistically significant improvement for lead times 4–6.

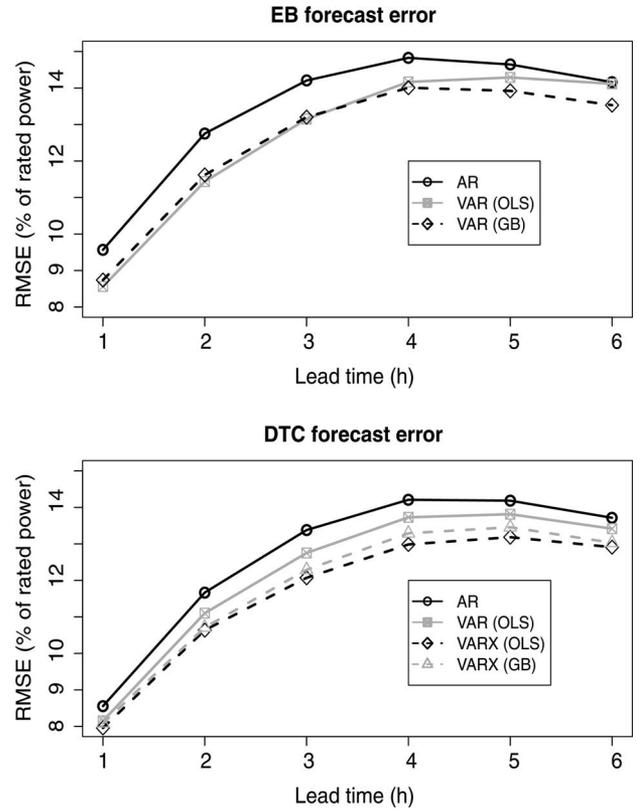


Fig. 7. Average $RMSE_k$ for each lead time calculated with the individual $RMSE_k$ obtained for each EB and DTC.

Compared to the DTC results (Fig. 5), the improvement obtained for the EB dataset is higher for the first two lead times. Another interesting conclusion is that, in most cases, the improvement decays with the lead time, meaning that the spatial-temporal information is more relevant for the first 3 h. This makes sense since the forecasting model in this test case only includes information from a small municipality. If solar power data from neighboring municipalities and regions are included in the model, a higher improvement for lead times between 4 and 6 is expected.

Fig. 7 depicts the average values of the normalized $RMSE_k$ for the EB and DTC datasets, calculated from the individual normalized $RMSE_k$ values (i.e., divided by the rated power of each PV site) of each EB and DTC. For the EB, the forecast errors are from the AR (the benchmark model), VAR [OLS], and VAR [GB] model, while for the DTC are from the VAR [OLS], VARX [OLS], and VARX [GB] models.

This plot shows a RMSE, for the best model, ranging between 8.5% and 14.1% for the EB, and between 7.9% and 12.9% for the DTC. The AR model exhibits an error between 9.5% and 14.1% for the EB, and between 8.6% and 13.7% for EB. The $RMSE_k$ magnitude is consistent with the state-of-the-art [12], [13]. As shown in Tables I and II, the forecast error difference is significant for each lead time.

The forecasting algorithms were programmed in R environment. In terms of computational effort, for an Intel Core i7-2600 CPU @ 3.40 GHz processor and 8 GB of RAM and programmed in R, the average fitting times for 6 h ahead are:

(VAR OLS) 20.2 s for 44 EB and 2.7 s for 10 DTC; (VAR GB) 736.4 s for 44 EB; and 46.7 s for 10 DTC, including fivefold cross validation. After fitting the model's coefficients, a forecast can be produced almost instantaneously. The average fitting time of the AR model for one time series is 1.6 s.

V. CONCLUSION

A new forecasting approach for very short-term solar power forecast is proposed in this paper. It relies on a spatial-temporal model, based on a vector autoregressive (VAR) framework fitted with two alternative methods (Recursive Least Squares and Gradient Boosting). It takes advantage of the territorially distributed nature of a smart grid infrastructure with smart meters and the advanced control functions installed at the MV/LV substation level. In terms of communication requirements, this type of tool needs that solar time series are transmitted on an hourly basis to the DSO central management system, which requires technological solutions that ensure reduced data latency (such as General Packet Radio Service).

The results for data from a smart grid pilot, in the city of Évora, Portugal, indicate that information from distributed PV generation can improve the forecast error, compared to an autoregressive model, in a reduction of between 8% and 12% on average for the first three lead times (which are the most important ones).

The RLS fitting method provides a lower error compared to GB and it is capable of tracking time-varying coefficients. Contrariwise, the GB generates a sparse coefficient's matrix, which decreases the volume of transmitted data, particularly from the LV grid. Finally, the inclusion of the EB observations in the forecast at the MV/LV substation level also decreases the forecast error.

These results open new lines for future research, such as: 1) inclusion of data from weather stations and NWP models; 2) increasing the spatial coverage of the PV data; 3) probabilistic forecasting; and 4) combination of metered and satellite information.

APPENDIX CLEAR-SKY MODEL

The clear-sky generation (\hat{p}_t^{CS}) is estimated as a local constant model and the weighted quantile regression [12] for quantile τ can be expressed as

$$\hat{p}_t^{\text{CS}} = \arg \min_{\hat{p}_t^{\text{CS}}} \sum_{i=1}^N K(h_t, \text{doy}_t, h_i, \text{doy}_i) \cdot \rho(\tau, e_i) \quad (16)$$

with $e_i = p_t - \hat{p}_t^{\text{CS}}$, where

$$K(h_t, \text{doy}_t, h_i, \text{doy}_i) = \frac{K(h_t, h_i, \sigma_h) \cdot K(\text{doy}_t, \text{doy}_i, \sigma_{\text{doy}})}{\sum_{i=1}^N [K(h_t, h_i, \sigma_h) \cdot K(\text{doy}_t, \text{doy}_i, \sigma_{\text{doy}})]} \quad (17)$$

is the kernel product of the two predictors [i.e., time of the day (h) and day of the year (doy)] which locally weights each observation, and

$$\rho(\tau, e_i) = \begin{cases} \tau \cdot e_i & , e_i \geq 0 \\ (1 - \tau) \cdot e_i & , e_i < 0 \end{cases} \quad (18)$$

is the loss function of the quantile regression problem [32]. Since both variables are circular, a circular kernel is used

$$K(x_t, x_i, \sigma) = e^{-\frac{1}{\sigma} \cdot \cos \left[2\pi \cdot \frac{(x_t - x_i)}{d} \right]} \quad (19)$$

where σ is the smoothing parameter and d is the period of variable x (e.g., equal to 24 in the time of the day).

The model's parameters are the kernel bandwidths σ_h and σ_{doy} , as well as the quantile τ .

REFERENCES

- [1] G. Masson, M. Latour, M. Reking, I. Theologitis, and M. Papoutsis, "Global market outlook for photovoltaics: 2013–2017," European Photovoltaic Industry Association, Tech. Rep., 2013 [online] Available: http://www.epia.org/fileadmin/user_upload/Publications/GMO_2013_-_Final_PDF.pdf
- [2] P. Hummel, P. Lekander, A. Gandolfi, S. Hunt, and I. Cossio, "The unsubsidised solar revolution," UBS Investment Research, Tech. Rep., Jan. 2013 [online] Available: <http://www.qualenergia.it/sites/default/files/articolo-doc/UBS.pdf>
- [3] V. C. Gungor *et al.*, "A survey on smart grid potential applications and communication requirements," *IEEE Trans. Ind. Informat.*, vol. 9, no. 1, pp. 28–42, Feb. 2013.
- [4] A. Madureira, L. Seca, J. Peças Lopes, P. Matos, and N. Silva, "Maximizing the integration of distributed generation in smart grids distribution systems," in *Proc. CIGRE Symp.*, Lisbon, Portugal, 2013, pp. 1–8.
- [5] A. G. Madureira and J. A. Peças Lopes, "Ancillary services market framework for voltage control in distribution networks with microgrids," *Elect. Power Syst. Res.*, vol. 86, pp. 1–7, May 2012.
- [6] J. Krstulovic, V. Miranda, A. Simões Costa, and J. Pereira, "Towards an auto-associative topology state estimator," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3311–3318, Aug. 2013.
- [7] P. Siano, C. Cecati, H. Yu, and J. Kolbusz, "Real time operation of smart grids via FCN networks and optimal power flow," *IEEE Trans. Ind. Informat.*, vol. 8, no. 4, pp. 944–952, Nov. 2012.
- [8] European Photovoltaic Industry Association (EPIA), "EPIA response to DG ENER working paper—The future role and challenges of energy storage," Position Paper, Feb. 2013.
- [9] H. Kanchev, D. Lu, F. Colas, V. Lazarov, and B. Francois, "Energy management and operational planning of a microgrid with a PV-based active generator for smart grid applications," *IEEE Trans. Ind. Electron.*, vol. 58, no. 10, pp. 4583–4592, Oct. 2011.
- [10] R. J. Bessa, C. L. Moreira, B. Silva, and M. A. Matos, "Handling renewable energy variability and uncertainty in power systems operation," *Wiley Interdiscip. Rev.: Energy Environ.*, vol. 3, no. 2, pp. 156–178, Mar./Apr. 2014.
- [11] C. E. Borges, Y. K. Peña, and I. Fernández, "Evaluating combined load forecasting in large power systems and smart grids," *IEEE Trans. Ind. Informat.*, vol. 9, no. 3, pp. 1570–1577, Aug. 2013.
- [12] P. Bacher, H. Madsen, and H. A. Nielsen, "Online short-term solar power forecasting," *Solar Energy*, vol. 83, no. 10, pp. 1772–1783, 2009.
- [13] L. A. Fernandez-Jimenez *et al.*, "Short-term power forecasting system for photovoltaic plants," *Renew. Energy*, vol. 44, pp. 311–317, Aug. 2012.
- [14] A. Hammer *et al.*, "Solar energy assessment using remote sensing technologies," *Remote Sens. Environ.*, vol. 86, no. 3, pp. 423–432, 2003.
- [15] H. Pedro and C. Coimbra, "Assessment of forecasting techniques for solar power production with no exogenous inputs," *Solar Energy*, vol. 86, no. 7, pp. 2017–2028, Jul. 2012.

- [16] J. Huang, M. Korolkiewicz, M. Agrawal, and J. Boland, "Forecasting solar radiation on an hourly time scale using a Coupled AutoRegressive and Dynamical System (CARDS) model," *Solar Energy*, vol. 87, pp. 136–149, Jan. 2013.
- [17] H. M. Diagne, M. David, P. Lauret, and J. Boland, "Solar irradiation forecasting: State-of-the-art and proposition for future developments for small-scale insular grids," in *Proc. World Renew. Energy Forum*, Denver, CO, USA, May 2012, pp. 1–8.
- [18] V. Berdugo, C. Chaussin, L. Dubus, G. Hebrail, and V. Leboucher, "Analog method for collaborative very-short-term forecasting of power generation from photovoltaic systems," in *Proc. Next Gener. Data Min. Summit*, Greece, Sep. 2011, pp. 1–5.
- [19] C. Yang and L. Xie, "A novel ARX-based multi-scale spatiotemporal solar power forecast model," in *Proc. North Amer. Power Symp.*, Champaign, IL, USA, Sep. 2012, pp. 1–6.
- [20] P. Godinho Matos *et al.*, "InovGrid, a smart vision for a next generation distribution system," in *Proc. 22nd Int. Conf. Elect. Distrib.*, Stockholm, Sweden, Jun. 10–13, 2013, pp. 1–4.
- [21] P. Lúcio, P. Paulo, and H. Craveiro, "InovCity—Building smart grids in Portugal," in *Proc. 21st Int. Conf. Elect. Distrib.*, Frankfurt, Germany, Jun. 6–9, 2011, pp. 1–5.
- [22] V. C. Güngör *et al.*, "Smart grid technologies: Communication technologies and standards," *IEEE Trans. Ind. Informat.*, vol. 7, no. 5, pp. 529–539, Nov. 2011.
- [23] V. C. Güngör, B. Lu, and G. P. Hancke, "Opportunities and challenges of wireless sensor networks in smart grid," *IEEE Trans. Ind. Electron.*, vol. 57, no. 10, pp. 3557–3564, Oct. 2010.
- [24] H. Madsen, *Time Series Analysis*. London, U.K.: Chapman & Hall, 2006.
- [25] R. Davidson and J. G. MacKinnon, *Econometric Theory and Methods*. New York, NY, USA: Oxford Univ. Press, 2003.
- [26] J. M. Wooldridge, *Introductory Econometrics: A Modern Approach*, 2nd ed. Chula Vista, CA, USA: South-Western College, 2002.
- [27] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification*. Cambridge, MA, USA: MIT Press, 1983.
- [28] H. Madsen, J. Holst, and E. Lindstrom, *Modelling Non-Linear and Non-Stationary Time Series*, Lingby, Denmark: Dept. Mathematical Modelling, Technical Univ. Denmark, 2006 [Online]. Available: <http://www.control.aau.dk/~tk/undervisning/PhDAdvSI/Litterature/MadsenAndHolst2006.pdf>
- [29] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annu. Stat.*, vol. 29, no. 5, pp. 1189–1232, 2001.
- [30] P. Bühlmann, "Boosting for high-dimensional linear models," *Annu. Stat.*, vol. 34, no. 2, pp. 559–583, 2006.
- [31] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *J. Bus. Econ. Stat.*, vol. 13, no. 3, pp. 253–263, 1995.
- [32] R. Koenker and G. Bassett, "Regression quantiles," *Econometrica*, vol. 46, pp. 33–50, 1978.

Ricardo J. Bessa received the Licenciado degree in electrical and computer engineering from the Faculty of Engineering, University of Porto (FEUP), Porto, Portugal, in 2006; the M.S. degree in data analysis and decision support systems from the Faculty of Economics, University of Porto (FEP), Porto, in 2008; and the Ph.D. degree in sustainable energy systems from FEUP, in 2013.

Currently, he is a Senior Researcher with the Center for Power and Energy Systems, INESC TEC, Porto. His research interests include renewable energy forecasting, smart grids, electricity markets, and decision-aid methods.

Artur Trindade received the Licenciado degree in physics and chemistry from the Faculty of Sciences, University of Porto, Porto, Portugal, in 2005, and the M.S. degree in electrical and computer engineering from the Faculty of Engineering, University of Porto, in 2014.

Currently, he is a Researcher with Elergone Energia Lda, Leça da Palmeira, Portugal. His research interests include forecasting area.

Vladimiro Miranda (M'90–SM'04–F'06) received the Ph.D. degree in electrical engineering from the Faculty of Engineering, University of Porto (FEUP), Porto, Portugal, in 1982.

In 1981, he joined FEUP and currently holds the position of Full Professor. He has been a Researcher with INESC, Porto, since 1985, and is currently the Director with INESC, the leading institution of INESC TEC—INESC Technology and Science, an advanced research network in Portugal. He is also the President of INESC P&D Brasil, Santos - São Paulo, Brasil. He has authored many papers and has been responsible for many projects in areas related with the application of computational intelligence to power systems.

Dr. Miranda was the 2013 recipient of the IEEE Power and Energy Society Ramakumar Family Renewable Energy Excellence Award.