Solar Power Forecasting in Smart Grids Using Distributed Information

R.J. Bessa, A. Trindade, A. Monteiro, V. Miranda
INESC TEC - INESC Technology and Science and FEUP - Faculty of Engineering, University of Porto
Porto, Portugal
rbessa@inescporto.pt, ee08089@fe.up.pt,
andre.b.monteiro@inescporto.pt, vmiranda@inescporto.pt

Catia S. P. Silva
NeuroEngineering Laboratory, University of Florida
Gainesville, FL USA
catiaspsilva@ufl.edu

***Abstract*—The growing penetration of solar power technology at low voltage (LV) level introduces new challenges in the distribution grid operation. Across the world, Distribution System Operators (DSO) are implementing the Smart Grid concept and one key function, in this new paradigm, is solar power forecasting. This paper presents a new forecasting framework, based on vector autoregression theory, that combines spatial-temporal data collected by smart meters and distribution transformer controllers to produce six-hour-ahead forecasts at the residential solar photovoltaic (PV) and secondary substation (i.e., MV/LV substation) levels. This framework has been tested for 44 micro-generation units and 10 secondary substations from the Smart Grid pilot in Évora, Portugal (one demonstration site of the EU Project SuSTAINABLE). A comparison was made with the well-known Autoregressive forecasting Model (AR – univariate model) leading to an improvement between 8% and 12% for the first 3 lead-times.**

Keywords—Solar power; forecasting; smart grid; spatial-temporal; smart metering

I. INTRODUCTION

Several countries are achieving remarkable levels of installed solar power. By the end of 2012, the installed solar power in Germany and Italy was around 32.6 GW and 16.7 GW respectively. In terms of cost, solar photovoltaic (PV) is reaching grid parity in many countries, meaning that it can generate electricity at a levelized cost less than or equal to the electricity retailing tariffs [1]. In this context, the deployment of solar PV will likely continue even if subsidies are withdrawn (e.g., feed-in tariffs).

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 concept [2], strongly supported by Information and Communication Technologies (ICT), provides additional capabilities for monitoring and controlling the distribution grid and its assets. This creates conditions to develop new management tools (or improve existing ones) that maximize the integration of distributed generation at the MV and LV levels [3], such as voltage control [4], state-estimation [5] and congestion management [6].

Furthermore, a massive deployment of local storage (such as batteries) at the residential level might occur if governments create incentives for such goal [7]. For instance, Germany has created financial incentives for owners of solar systems with batteries [8].

The abovementioned management tools and the joint coordination of PV panels and local storage require the use of solar power forecasts for several hours ahead. The time-horizon of interest for power system operations can be divided into two classes [9]: (a) very short-term (up to six hours ahead); (b) short-term (up to three days ahead).

In recent publications, several works combine statistical algorithms with numerical weather predictions (NWP) to produce solar power forecasts for the short-term horizon. Bacher *et al.* [10] 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 two hours ahead the most important inputs are the past observations, while for a horizon up to 36 hours ahead the NWP prevail. Fernandez-Jimenez *et al.* [11] also used NWP as input in several machine learning algorithms (i.e., Auto-Regressive Integrated Moving Average – ARIMA, k-nearest neighbors - kNNs, neural networks – NN, and adaptive neuro-fuzzy models) to produce solar power forecasts for the next 39 hours. The best performance was obtained with a Neural Network (NN).

For the very short term horizon, two different classes of models can be found. The first class is based on satellite images. Hammer *et al.* [12] describe an algorithm based on cloud-index images that are predicted with motion vector fields derived from two consecutive images.

This work was made in the framework of the BEST CASE project ("NORTE-07-0124-FEDER-000056") financed 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), and by national funds, through Fundação para a Ciência e a Tecnologia (FCT). It was also co-financed by the 7th RTD Framework Programme within the SuSTAINABLE project (contract no. 308755), and by the COMPETE Programme and the FCT within projects «SMAGIS – PTDC/SEN-ENR/113094/2009» and «DYMONDS – CMU-PT/SIA/0043/2009».

The second class consists of univariate time-series models. Pedro and Coimbra [13] compared the performance of different machine learning algorithms (i.e., ARIMA, kNNs, 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.

An extended overview for the solar power forecast literature can be found in [14].

The forecasting framework presented in this paper addresses the very short-term horizon and is included in the second class of models. The main difference is that the new proposed method is spatial-temporal, since it combines the past observations of time series distributed in space. To our knowledge, only two works combined information from neighboring sites to improve solar power forecast. Berdugo *et al.* [15] described a method based on searching similar local and global current states (e.g. kNNs), considering neighbor sites, whose goal is not to produce the "optimal" forecast (i.e., with minimum error); instead, it handles distributed data streams and maintains power measurements' privacy. Yang *et al.* [16] proposed an ARX model for each solar site where the exogenous variables are measurements from neighbor sites.

This paper proposes three main contributions:

a) new forecasting method, constructed on the top of a Smart Grid infrastructure, based on vector auto-regression theory (VAR), combining information from the distributed PV panels;

b) improved accuracy by introducing exogenous variables to the model, i.e. observations from micro-generation smart meters;

c) an online fitting method, based on recursive least squares, for the VAR model.

This forecasting framework will be applied to produce six-hour-ahead forecasts for each residential PV and secondary substation (i.e., MV/LV). 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 intra-day electricity markets.

The paper is organized as follows: section II presents the Smart Grid infrastructure in Portugal; section III describes the solar power forecasting algorithms; the test case results are presented in section IV; section V presents the conclusions and future work.

II. SMART GRID INFRASTRUCTURE

During the recent years, several DSO have conducted different Smart Grid pilot tests and a deployment of these technological solutions is expected for the following years. An example is the InovGrid Project in Portugal [17], promoted by EDP Distribution, aiming to develop new ICT technology and computational tools for automating network management and to create a complete smart distribution grid. This project resulted in a large-scale demonstration pilot in the city of Évora in Portugal, named InovCity [18], which is also one demonstration site of the EU Project SuSTAINABLE.

The main components of this infrastructure are the EDP Box (EB) and the Distribution Transformer Controller (DTC). The EB is a smart meter with load and generation management functions, located at each delivery point. Load and generation are metered separately. It can interact with other devices through a home area network. The DTC is located at the secondary substation level, comprising modules for measurement, remote control and communication actions. It collects data from the EB and the secondary substation. Both the EB and the DTC are part of a hierarchical control and communication architecture. Each EB has a bi-directional communication with the corresponding DTC through GPRS (General Packet Radio Service) or PLC (Power Line Communications), and the DTC communicates with the SCADA/DMS through a wide area network based on GPRS [18]. Other communication technologies, such as radio frequency mesh modules, can be explored in this framework.

At the primary substation level (HV/MV), a Smart Substation Controller (SSC) is installed. The SSC is responsible for aggregating and managing the operational data from EB and DTC, and for applying demand-side management, self-healing and generation management strategies. Therefore, the MV grid is managed by the SSC. On the top of these technologies, there are services capable of handling large volumes of data and, at the same time, providing an overview of all existing devices.

Section III presents the forecasting framework that follows the Smart Grid infrastructure of the InovCity pilot. The forecasting system is installed at the central management level (i.e., in the DMS). Although installed in the DMS, the forecasting system can be virtually distributed by HV/MV substation, but using information from DTC connected to different primary substations.

The outputs are forecasts for each DTC and EB. For the forecast at the DTC level, the EB measurements can be used as distributed sensors to better capture the effect of clouds in solar generation and consequently improve the forecasts, which in turn increase the amount of transmitted data. Section IV evaluates the improvement that is obtained with this additional information.

Finally, it is important to stress that the forecasting system is also valid for a non-hierarchical architecture, such as one where each device communicates directly with the DSO control center via TCP/IP [19]. One has to pose a centralized data flow topology as a requirement. A forecasting approach for peer-to-peer data flow is presented in [14].

III. FORECASTING FRAMEWORK

A. Seasonal De-Trending 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, which results in a trend component during the day. 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) [14].

One of these statistical approaches is the clear-sky model described in [10] 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. Note that most classical models, such as AR and VAR, assume stationarity of the time-series.

Following the results in [10] and [20], the predictors of the clear-sky model are the time of the day (h) and day of the year (doy). The clear-sky generation ($\hat{p}_t^{cs}$) is estimated as a local constant model and the weighted quantile regression for quantile τ can be expressed as:

$$\hat{p}_t^{cs} = \arg\min_{\hat{p}_t^{cs}} \sum_{i=1}^{N} K(h_t, doy_t, h_i, doy_i) \cdot \rho(\tau, e_i) \quad (1)$$

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

$$K(h_t, doy_t, h_i, doy_i) = \frac{K(h_t, h_i, \sigma_h) \cdot K(doy_t, doy_i, \sigma_{doy})}{\sum_{i=1}^{N} [K(h_t, h_i, \sigma_h) \cdot K(doy_t, doy_i, \sigma_{doy})]} \quad (2)$$

is the kernel product of the two predictors which locally weights each observation, and

$$\rho(\tau, e_i) = \begin{cases} \tau \cdot e_i & , e_i \geq 0 \\ (1-\tau) \cdot e_i & , e_i < 0 \end{cases} \quad (3)$$

is the loss function of the quantile regression problem [21]. Since both variables are circular, a circular kernel is used:

$$K(x_t, x_i, \sigma) = e^{\frac{1}{\sigma} \cdot \cos\left[2\pi \cdot \frac{(x_t - x_i)}{d}\right]} \quad (4)$$

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 output of the model from Eq. 1 is used to normalize the measured solar power (P_t) as follows:

$$p_t^{norm} = \frac{p_t}{\hat{p}_t^{cs}} \quad (5)$$

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, inducing a value equal to 1 for the variable p_t^{norm} during clear-sky days.

B. Vector Autoregressive (VAR)Model

The ARIMA process is a well-known class for univariate time-series models [22]. A special model of this class is the autoregressive (AR), in which the value of the response variable for time interval t can be interpreted as a regression on past observations (or lags) of the time-series. For one hour-ahead forecast, the AR model is:

$$\hat{p}_{t+1|t} = \alpha + \beta_1 \cdot p_t + \beta_2 \cdot p_{t-1} + \cdots + \beta_l \cdot p_{t-l} + e_{t+1|t} \quad (6)$$

where β are the model's coefficients, α a constant term, l the order of the AR model and $e_{t+l|t}$ is a white noise process 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 dependent variable. This model can be extended with exogenous variables (such as NWP), forming an ARX model. Nevertheless, as mentioned in the literature [10] and [14], NWP can only improve the forecast error for time horizons greater than four hours-ahead, while for shorter time horizons the relevant information consists in time-series observations.

In order to improve the forecasts for the very short-term horizon, a VAR model [23] is used to combine past observations from the solar power in each site with past values from neighbor sites. This consists in a multi-output (or spatial-temporal) 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:

$$\hat{P}_{t+1|t} = \alpha + \mathrm{B} \cdot P_{t-l} + \mathrm{E}_{t+1|t} \quad (7)$$

where $\hat{P}_{t+1|t}$ is the response matrix with dimension $N.q$, B the coefficient matrix with dimension $d.q$, P_{t-l} the predictor matrix with dimension $N.d$, α is a vector with q intercept (or constant) terms, $E_{t+l|t}$ is a matrix with dimension $N.q$ containing i.i.d. residuals with zero mean and constant covariance Σ_e.

Eq. 7, for an example with two response variables and two lagged terms, becomes:

$$\begin{aligned} \hat{p}_{t+1|t,1} &= \alpha_1 + \beta_{11} \cdot p_{t,1} + \beta_{12} \cdot p_{t-1,1} + \beta_{13} \cdot p_{t,2} + \beta_{14} \cdot p_{t-1,2} + e_{t+1|t,1} \\ \hat{p}_{t+1|t,2} &= \alpha_2 + \beta_{21} \cdot p_{t,1} + \beta_{22} \cdot p_{t-1,1} + \beta_{23} \cdot p_{t,2} + \beta_{24} \cdot p_{t-1,2} + e_{t+1|t,2} \end{aligned} \quad (8)$$

As shown in Eq. 8, the VAR consists of linear univariate 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-l}) 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, Eq. 7 and 8 have the form of a Seemingly Unrelated Regression (SUR) model [23].

Estimating the coefficients of the VAR model is straightforward, that is, the Ordinary Least Squares (OLS) can be applied if the same predictors appear in every equation. In such a case, OLS gives the same solution of Generalized Least Squares (GLS) and can be applied independently to each regression equation.

C. Recursive Least Squares

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 recursive least squares (RLS) method with a forgetting factor [24]. This method overcomes the problem of handling "big data" since it is not necessary to store historical data for fitting (or re-fitting) the model. Furthermore, the RLS method, with a *forgetting factor* λ, 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 Eq. 7) is performed with the RLS method as follows for time step t:

$$B_t = B_{t-1} + K_t \cdot \left[P_t - \left(\alpha + B_{t-1} \cdot P_{t-l}\right)\right] \tag{9}$$

where K_t is given by,

$$K_t = Q_t \cdot P_{t-l} \tag{10}$$

and Q_t by

$$Q_t = \frac{1}{\lambda} \cdot \left[Q_{t-1} - \frac{Q_{t-1} \cdot P_{t-l} \cdot P_{t-l}^T \cdot Q_{t-1}}{\lambda + P_{t-l}^T \cdot Q_{t-1} \cdot P_{t-l}} \right] \tag{11}$$

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 observation P_t).

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

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 is 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. VAR Model for Solar Power Forecast

In order to apply the forecasting techniques from Eq. 6 and 7, in a first phase, it is necessary to normalize the solar power time-series with the clear-sky model from section III.A. In [20], 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 night time.

In this paper, a different approach that enables the use of the VAR model was followed: the normalized solar power values outside the period between 7h00 and 19h00 (i.e., the average period with almost no solar generation during the whole year in Portugal) are removed from all sites.

In a second phase, the normalized solar power values are used to fit the AR and VAR models with RLS. Both models are 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: first, the autocorrelation plot of the normalized time series is analyzed to make a coarse estimation of the necessary lags; then, the autocorrelation plot of the residuals is analyzed to check if the residuals are i.i.d. (i.e., white noise). 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 [25].

For instance, Fig. 1 depicts the autocorrelation plot of the residuals obtained with an AR model (that includes lags *t-1*, *t-2* and *t-24*) for one DTC. As depicted, the residuals are almost uncorrelated (i.e., in the sense that the autocorrelation values are inside the 95% confidence interval), which validates the choice of the lagged terms. This is the reason why the second lag was included in the model.

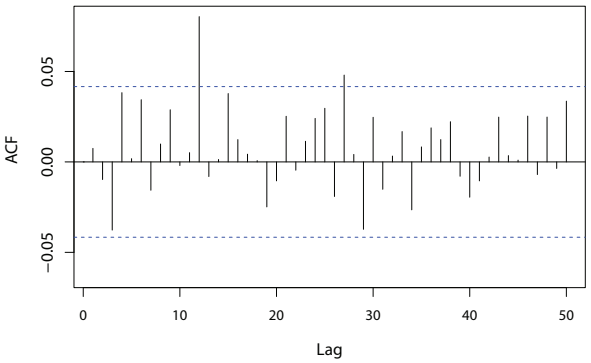

Fig. 1. Autocorrelation plot of the AR model's residuals.

Since the goal is to produce six hours-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:

$$\hat{P}_{t+2|t} = \alpha_2 + \mathrm{B}_1 \cdot P_t + \mathrm{B}_2 \cdot P_{t-1} + \mathrm{B}_3 \cdot P_{t-22} + \mathrm{E}_{t+2} \tag{12}$$

$$\hat{P}_{t+6|t} = \alpha_6 + \mathrm{B}_1 \cdot P_t + \mathrm{B}_2 \cdot P_{t-1} + \mathrm{B}_3 \cdot P_{t-18} + \mathrm{E}_{t+6} \tag{13}$$

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

The RLS algorithm (Eq. 9-11) is used to update the coefficients of each lead-time (e.g., coefficients B of Eq. 12 and 13).

Finally, in addition to the AR and VAR models described in the previous sections, a VAR with exogenous variables (VARX) is also proposed and tested. The model consists in adding exogenous variables to Eq. 7: 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. 2. 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.

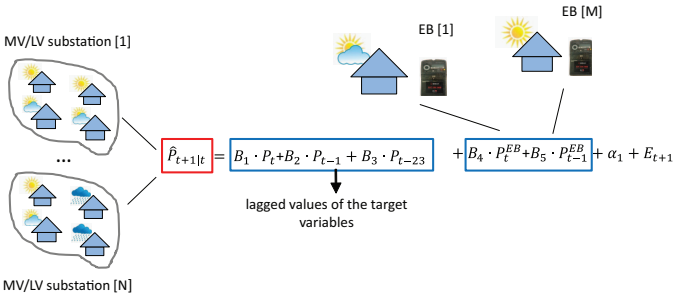

Fig. 2. VARX model for lead-time t+1 with EB measurements.

IV. Test Case Results

A. Description

The solar power dataset used as test case is from the city of Évora, with about 54,000 residents and an area of 1307 km^2. The distribution network in Évora municipality is supplied by two 60 kV substations with 15 kV and 30 kV feeders. 25 of these feeders supply a total of 655 secondary substations.

In what concerns distributed generation, in August 2013, there were 218 micro-generation producers, mainly solar PV, with an installed capacity of 761 kW. During 2 years, more than 30,000 EB and 300 DTC were installed in Évora, including all customers and substations, in order to have the entire municipality covered.

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 kWp and 3.7 kWp. These EB measurements were related to 10 different DTC, and the total values of each DTC are also forecasted.

The "optimal" parameters of the clear-sky model are σ_h=0.01, σ_{doy}=0.01, τ=85%. From trial-error tests, the *forgetting factor* λ for both AR and VAR was found to be 0.999.

The original data was sampled in 15 minutes, but it was resampled to hourly values (i.e., the same length of the electricity market). The period between 1 February 2011 and 31 January 2012 was used to fit the models, and the period between 1 February 2012 and 6 March 2013 was used to calculate the forecast errors.

The forecasting results are evaluated with the root mean square error (RMSE) calculated for the k^{th} horizon [10]:

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

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:

$$\mathrm{Imp}_k = \frac{RMSE_{k,AR} - RMSE_{k,VAR}}{RMSE_{k,AR}} \cdot 100\% \qquad (15)$$

B. Results

The improvement of the VAR and VARX over the AR model for each lead-time is plotted in Fig. 3 for two DTC and for the RMSE_k calculated with the full dataset of DTC forecast errors.

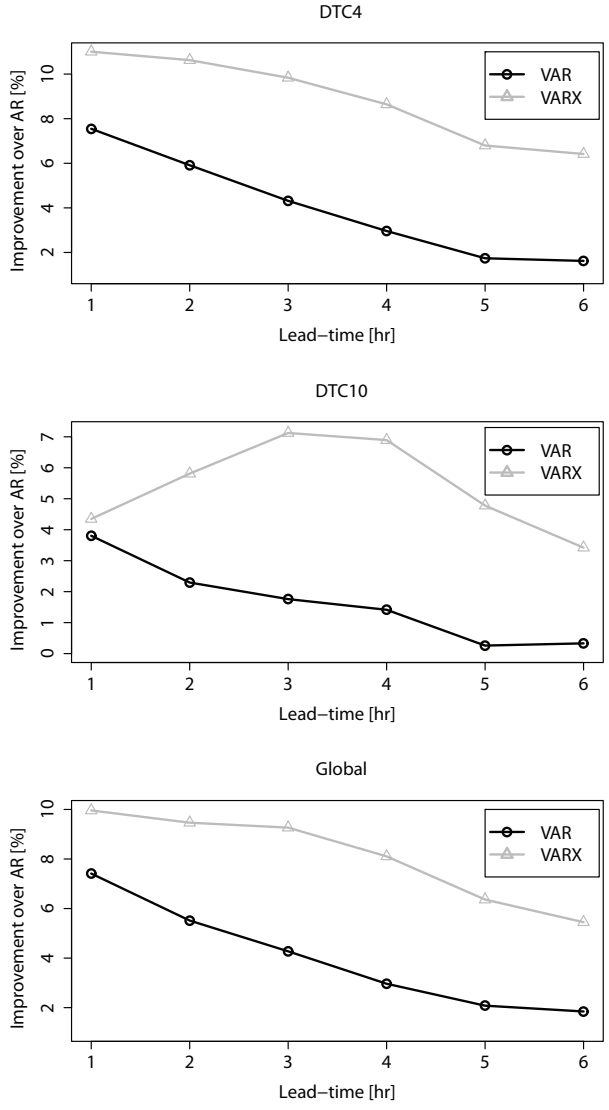

Fig. 3. Imp_k of the VAR and VARX models for two DTC and for the RMSE calculated using the entire set of forecast errors.

These three plots clearly show that the VARX 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 11% for the first lead-time and around 6% for the sixth lead-time. Number 10 is the one with the lowest improvement, ranging between 4% and 3%.

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 (or MV) level.

The global improvement of the VARX varies between 10% and 5.5%, which shows the benefit from using spatial-temporal models for very short-term solar power forecasting. Another interesting conclusion is that the improvement decays with the lead-time, meaning that the spatial-temporal information is more relevant for the first three hours. 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 is included in the model, a higher improvement for lead-times between 4 and 6 is expected.

Fig. 4 shows the improvement obtained with the VAR model for two EB and for the RMSE calculated with the full dataset of EB forecast errors.

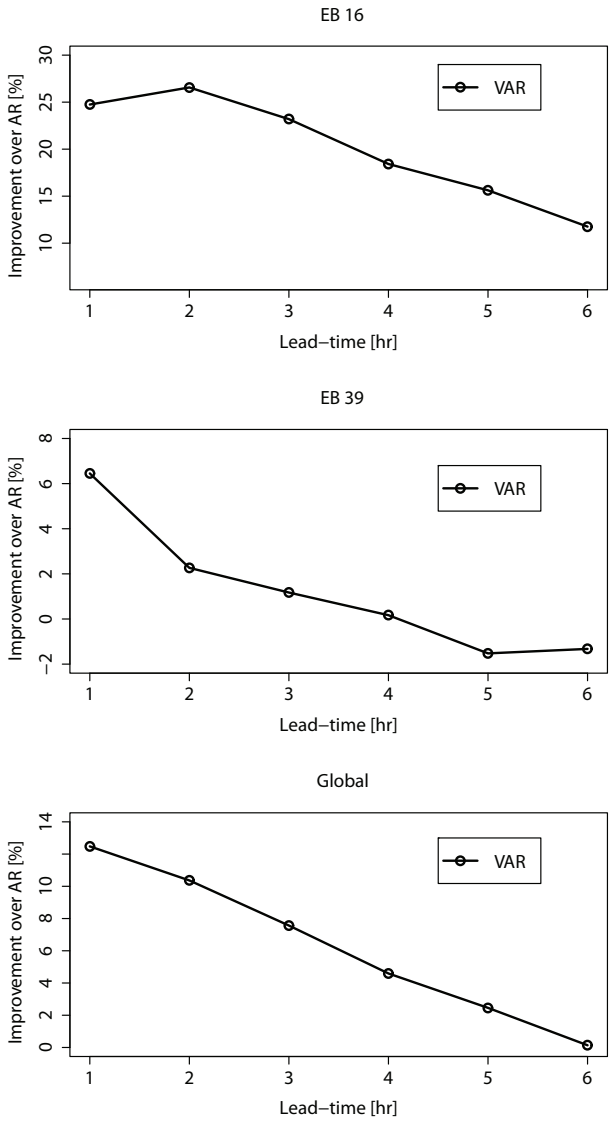

Fig. 4. Imp_k of the VAR model for two EB and for the RMSE calculated using the entire set of forecast errors.

The VAR model attained the highest improvement for EB number 16, with a value around 25% for lead-time 1 and around 11% for lead-time 6. The lowest improvement was attained for EB number 39, with 6.5% for lead-time 1 and a negative value of around -1% for lead-time 6. The global improvement for the EB dataset varies between 12.5% and 0.1%. Compared to the DTC results, the improvement obtained for the EB dataset is higher for the first two lead-times.

Fig. 5 depicts the average, minimum and maximum values of the normalized RMSE for the EB and DTC datasets, calculated from the individual $RMSE_k$ values of each EB and DTC. For the DTC, the forecast errors are from the VARX model, while for the EB are from the VAR model. The RMSE magnitude is consistent with the state of the art [10][11].

The average $RMSE_k$ of the EB and DTC is similar for the first two lead-times, but the difference increases for the other lead-times. The main difference is in the minimum and maximum values, with a higher amplitude for the EB. For instance, there are EB with an RMSE below 10% (even for lead-time 6), but also EB with an RMSE close to 20%. This is an expected result, since the variability in solar power due to the clouds is smoothed by the aggregation of EB in one DTC. Nevertheless, this also shows that the forecasts for some EB are significantly improved with the spatial-temporal information (e.g., EB number 16 in Fig. 4).

Fig. 6 depicts the improvement of the VAR over the AR model, calculated using the entire set of EB forecast errors and for three seasons (winter, spring and summer). The highest improvement for the first three lead-times occurs in the winter season, which frequently exhibits overcast days ("stable" clouds). The summer season frequently has clear-sky days, thus the improvement of the VAR model is lower. Nevertheless, for the last three lead-times, the results are contrasting and the improvement is higher in the summer season. This might be explained by the presence of small and rapid clouds ("unstable" clouds) during summer, which are captured by the six hours time-window. However, additional weather data (e.g., from weather stations) that would enable a more detailed analysis are needed to explain these differences. Finally, it is important to emphasize that the VAR model only showed negative improvement in lead-time 6 of the spring season.

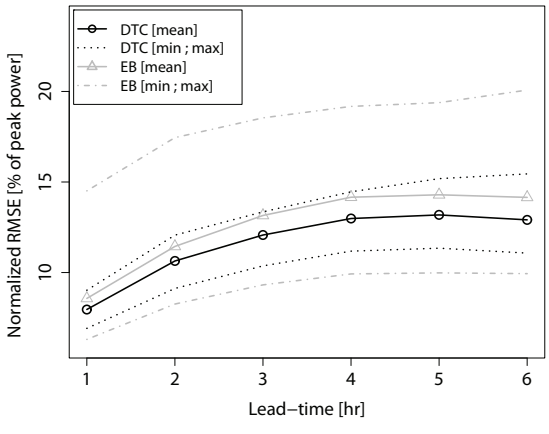

Fig. 5. Average, minimum and maximum normalized $RMSE_k$ calculated with the individual $RMSE_k$ obtained for each EB and DTC.

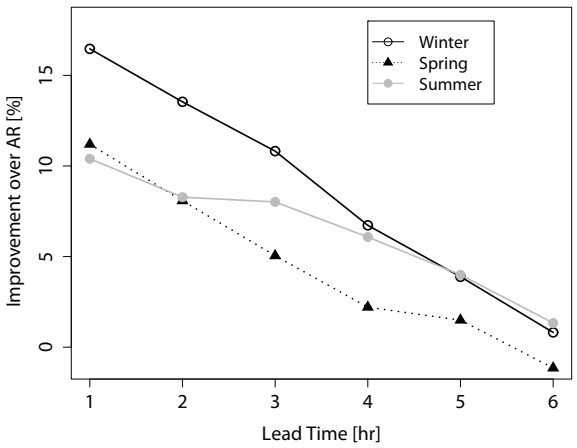

Fig. 6. Improvement of the VAR over AR for different seasons of the year, calculated with the full set of forecast errors for each EB.

V. CONCLUSIONS

A new forecasting approach for very short-term solar power forecast, based on a vector autoregressive model fitted with recursive least squares, is proposed in this paper. It takes advantage of a Smart Grid infrastructure with smart meters and advanced control functions installed at the MV/LV substation level.

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, between 8% and 12% on average for the first 3 lead-times. Furthermore, the inclusion of the EB observations in the forecast at the DTC level also decreases the error. Therefore, the adoption of multivariate models such as VAR with an online tracking of its coefficients is recommended to improve the solar power forecasting in a Smart Grid environment. This forecasting framework can be explored either by the DSO or by a solar PV aggregator.

These results open new lines for future research and improvement, such as: development of a feature selection algorithm that reduces the amount of information used and consequently the communication requirements; inclusion of data from weather stations and numerical weather prediction models; increasing the spatial coverage of the PV data.

REFERENCES

[1] P. Hummel, P. Lekander, A. Gandolfi, S. Hunt and I. Cossio, "The unsubsidised solar revolution," Technical Report, UBS Investment Research, January 2013.

[2] H. Farhangi, "The path of the Smart Grid," *IEEE Power and Energy Magazine*, vol. 8, no. 1, pp. 18-28, 2010.

[3] A. Madureira, L. Seca, J. Peças Lopes, P. Matos, N. Silva, "Maximizing the integration of distributed generation in smart grids distribution systems," in *Proceedings of the CIGRE Symposium*, Lisbon, Portugal, April 2013.

[4] A.G. Madureira, J.A. Peças Lopes, "Ancillary services market framework for voltage control in distribution networks with microgrids," *Electric Power Systems Research,* vol.86, pp.1-7, May 2012.

[5] J. Krstulovic, V. Miranda, A. Simões Costa, J. Correia Pereira, "Towards an auto-associative topology state estimator," *IEEE Transactions on Power Systems*, vol.28, no.3, pp.3311-3318, August 2013.

[6] J. Soares, J.A. Peças Lopes, "Controlling electric vehicles in quasi-real-time," in *Proceedings of the IEEE PowerTech 2013*, Grenoble, France, 16-20 June 2013.

[7] European Photovoltaic Industry Association (EPIA), "EPIA Response to DG ENER working paper - The future role and challenges of energy storage," Position Paper, February 2013.

[8] Website (accessed on November 2013): http://www.renewableenergyworld.com/rea/news/article/2013/03/solar-storage-market-set-for-rapid-growth?cmpid=rss

[9] C. Monteiro, R. Bessa, V. Miranda, A. Botterud, J. Wang, G. Conzelmann, "Wind power forecasting: state-of-the-art 2009," Report ANL/DIS-10-1, Ar-gonne National Laboratory, November 2009.

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

[11] L.A. Fernandez-Jimenez, A. Muñoz-Jimenez, A. Falces, M. Mendoza-Villena, E. Garcia-Garrido, P.M. Lara-Santillan, E. Zorzano-Alba, P.J. Zorzano-Santamaria, "Short-term power forecasting system for photovoltaic plants," *Renewable Energy*, vol. 44, pp. 311-317, August 2012.

[12] A. Hammer, D. Heinemann, C. Hoyer, R. Kuhlemann, E. Lorenz, R. Müller, H.G. Beyer, "Solar energy assessment using remote sensing technologies," *Remote Sensing of Environment*, vol. 86, no. 3, pp. 423-432, August 2003.

[13] H. Pedro, C. Coimbra, "Assessment of Forecasting Techniques for Solar Power Production with no Exogenous inputs," *Solar Energy*, vol. 86, no. 7, pp. 2017-2028, July 2012.

[14] H.M. Diagne, M. David, P. Lauret, J. Boland, "Solar irradiation forecasting: state-of-the-art and proposition for future developments for small-scale insular grids," *World Renewable Energy Forum (WREF 2012)*, Colorado, USA, 13-17 May 2012.

[15] V. Berdugo, C. Chaussin, L. Dubus, G. Hebrail, V. Leboucher, "Analog method for collaborative very-short-term forecasting of power generation from photovoltaic systems," *Next Generation Data Mining Summit (NGDM '11)*, Athens, Greece, 4 September 2011.

[16] C. Yang, L. Xie, "A novel ARX-based multi-scale spatiotemporal solar power forecast model," in *Processings of the North American Power Symposium (NAPS)*, USA, September 2012.

[17] P. Godinho Matos, P. Daniel, A. Veiga, A. Messias, M. Oliveira, P. Monteiro, "InovGrid, a smart vision for a next generation distribution system," in *Proceedings of the 22nd International Conference on Electricity Distribution (CIRED)*, Stockholm, Sweden, 10-13 June 2013.

[18] P. Lúcio, P. Paulo, H. Craveiro, "InovCity - Building smart grids in Portugal," in *Proceedings of the 21nd International Conference on Electricity Distribution (CIRED)*, Frankfurt, Germany, 6-9 June 2011.

[19] X. Lu, W. Wang, J. Ma, "An empirical study of communication infrastructures towards the smart grid: design, implementation, and evaluation," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 170-183, March 2013.

[20] P. Bacher, "Short-term solar power forecasting," MSc Thesis, Technical University of Denmark, 2008.

[21] R. Koenker, G. Bassett, "Regression quantiles", *Econometrica*, vol. 46, pp. 33-50, 1978.

[22] H. Madsen, *Time Series Analysis*, London: Chapman and Hall/CRC, 2006.

[23] R. Davidson, J.G. MacKinnon, *Econometric Theory and Methods*, New York: Oxford University Press, 2003.

[24] L. Ljung, T. Soderstrom, T*heory and Practice of Recursive Identification*, Cambridge: The MIT Press, 1983.

[25] J.M. Wooldridge, *Introductory Econometrics: A Modern Approach*, 2nd Edition, South-Western College Pub, 2002.