

# Estimation of a Simple Model of Solar Power Generation using Partial Information

Gianni Bianchini<sup>1</sup>, Simone Paoletti<sup>1</sup>, Antonio Vicino<sup>1</sup>, Franco Corti<sup>2</sup>, and Federico Nebiacolombo<sup>3</sup>

**Abstract**—This paper presents a heuristic method for the estimation of a model of a photovoltaic plant to be used for power generation forecasting purposes. The problem is addressed in a scenario where measurements of meteorological variables (i.e., irradiance and temperature) at the plant site are not available. The proposed approach efficiently exploits only power generation measurements and theoretical clear-sky irradiance, and is characterized by very low computational effort. A real-world application is presented to validate the procedure, which is also currently in pre-deployment phase at a Distribution System Operator control center in Northern Italy.

## I. INTRODUCTION

The electrical grid can be no longer considered a uni-directional means of distributing energy from conventional plants to the final users, but a Smart Grid, where strong interaction between producers and users takes place [1]. In this respect, the forecasting of demand and the balancing of production and consumption of electric energy, taking into account the distributed generation located at the lower levels of the network, are fundamental issues. In particular, a major challenge in the integration of renewable energy sources, such as solar and eolic, into the grid is that power generation is intermittent, difficult to control, and strongly depending on the variation of weather conditions. Therefore, it is of utmost importance that accurate forecasts of future generation be available to the various players. In active grid operation, generation forecasts of renewable plants (and aggregates thereof) are required by Distribution System Operators (DSOs) and Transmission System Operators (TSOs) in order to fulfil grid constraints, dispatch generators and satisfy demand [2].

Concerning solar power generation, a lot of attention has been paid to the problem of obtaining accurate day-ahead or hourly forecasts of solar irradiance and/or generated power at a given site. A widely adopted approach is the use of neural networks [3], [4] and support vector machines [5] with different types of input. Alternatively, classical linear time series forecasting methods are used [6]–[8], where the time series considered is typically the global irradiance normalized with a clear-sky model [9], [10] in order to obtain an almost stationary behaviour. Global radiation forecasts are then fed along with temperature forecasts to a simulation model of the plant [11] to compute the prediction of power generation. In any case, computing good generation forecasts

from predicted meteorological variables hinges upon on the availability of a reliable model of the plant, be it physical or estimated from data.

There are, however, some relevant scenarios in which neither an accurate model of the plant, nor direct on-site measurements of solar radiation and other meteorological variables such as temperature are available as data. This is typically the case of a DSO dealing with several heterogeneous and independently-operated solar plants connected to the network, of which only generated power measurements and meteorological forecasts of radiation and temperature are available. The problem of forecasting power generation in similar cases is addressed in [12] by means of a neural network, in which the inputs are the measured power series and the clear-sky radiation generated by a theoretical model that uses a-priori information on plant-specific configurations such as the orientation of the solar panels.

In this paper, we present a heuristic approach to the solution of the power forecasting problem in the partial information case. The approach consists in the estimation of a simple plant model using only historical data of generated power, theoretical clear-sky global irradiance data, and temperature forecasts. The main idea of the proposed algorithm is to perform a suitable least-squares adaptation of the gain parameter of a PVUSA model [13], [14] of the plant. Such adaptation is obtained from the information contained in portions of the generated power data which turn out to be meaningful if combined with the corresponding theoretical clear-sky irradiance data over the same period.

The paper is organized as follows: in Section II we recall the properties of the PVUSA model which are relevant to the approach presented here, in Section III we illustrate the proposed parameter estimation algorithm, an application using real data is presented in Section IV, and conclusions are drawn in Section V.

## II. PVUSA MODEL

The PVUSA method for rating photovoltaic power plants [14] expresses the generated power as a function of irradiance and ambient temperature according to the equation:

$$P = aI + bI^2 + cIT, \quad (1)$$

where  $P$ ,  $I$  and  $T$  are respectively the generated power, irradiance and ambient temperature, and  $a > 0$ ,  $b < 0$  and  $c < 0$  are the model parameters.

Model (1) shows several attractive features:

- 1) The model is simple and parsimonious in terms of number of parameters. This implies very little memory

<sup>1</sup>Dipartimento di Ingegneria dell'Informazione e Scienze Matematiche, Università di Siena, Italy, {giannibi,paoletti,vicino}@diism.unisi.it

<sup>2</sup>Siemens Italia, Italy, franco.corti@siemens.com

<sup>3</sup>Softco Sismat, Italy, federico.nebiacolombo@softco.it

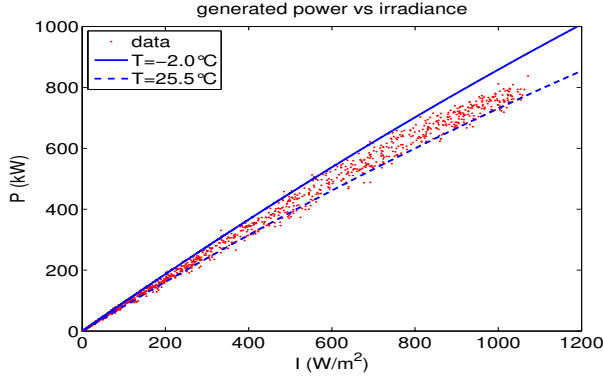


Fig. 1. Generated power  $P$  versus irradiance  $I$ . The two lines represent the estimated characteristic curves (1) for  $T = -2.0^\circ\text{C}$  (solid) and  $T = 25.5^\circ\text{C}$  (dashed), which are the minimum and maximum temperature in the data set.

occupation (which is a critical issue in some grid applications), and makes it easier to figure out the role of each parameter. Moreover, the model is linear-in-the-parameters, which makes parameter estimation accomplishable very efficiently via classical least-squares once power, irradiance and temperature measurements are available. To this aim, model (1) can be rewritten in regression form as follows:

$$P_k = \theta^T \varphi_k, \quad (2)$$

where  $\theta = [a \ b \ c]^T$  and  $\varphi_k = [I_k \ I_k^2 \ (I_k T_k)]^T$ , being  $k$  the discrete-time index, and  $P_k, I_k, T_k$  the average values of power, irradiance and temperature, respectively, in the  $k$ -th time interval.

- 2) It expresses the generated power as a function of measurable meteorological variables for which forecasts provided by meteorological services are also available, thus making generation forecasts straightforward. Other models of solar generation exist, such as [15], but they often involve also electrical, thermal and optical characteristics of photovoltaic modules which are too detailed for most grid applications.
- 3) Good accuracy is obtained when a model of the type (1) is fitted to available real measured data. Fig. 1 shows a data set obtained from a photovoltaic plant of 920 kWp located in Italy. It also shows the estimated characteristic curves (1) for  $T = -2.0^\circ\text{C}$  (solid) and  $T = 25.5^\circ\text{C}$  (dashed), which are the minimum and maximum temperature in the data set. It can be seen that only a very small portion of the data remains outside the two curves. Fig. 2 shows the generated power (solid), the prediction computed via model (1) using measured meteorological variables (dashed), and the corresponding prediction error (dash-dotted). It can be seen that the prediction error is relatively small.

The PVUSA model is very useful to DSOs, since it allows for computing forecasts of generated power on the basis of predicted meteorological variables. Indeed, once the parameters have been estimated, the resulting model can be

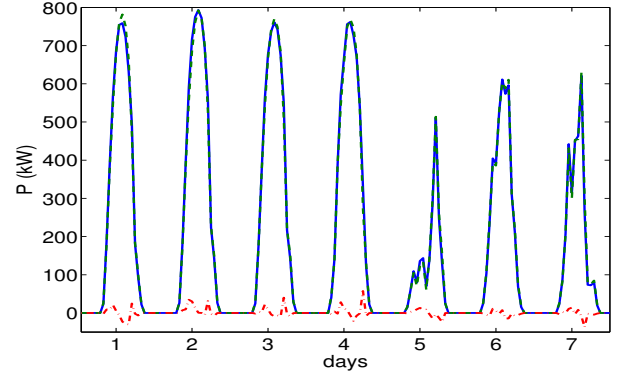


Fig. 2. Generated power (solid), prediction computed via model (1) using measured meteorological variables (dashed), and corresponding prediction error (dash-dotted).

used to compute power generation forecasts by substituting predicted irradiance  $\hat{I}_k$  and temperature  $\hat{T}_k$  provided by a meteorological service into the model equation. Similarly, generation forecasts under clear-sky conditions can be obtained by using the theoretical irradiance  $I_k^{cs}$  and temperature forecasts. Notice that clear-sky generation forecasts represent an upper bound on the power that can be generated by a plant, and can be used by the DSO, for instance, when scheduling the maintenance of the portion of the grid where the plant is located.

#### A. Parameter estimation with partial information

Unfortunately, DSOs do not typically have all the information needed to estimate the parameters of model (1) directly, since they often have to deal with a high number of independently-operated plants. In particular, it is quite frequent that DSOs hold a time series of the measurements of  $P_k$ , but not of  $I_k$  and  $T_k$  for a given plant. This means that the straightforward least squares approach described above cannot be applied, in practice, to estimate the parameter vector  $\theta$ . A possible alternative could be to replace the measured values of  $I_k$  and  $T_k$  with forecasts  $\hat{I}_k$  and  $\hat{T}_k$  provided by a meteorological service. However, as can be seen in Fig. 3, this idea is typically not viable, mainly due to large forecasting errors on the irradiance, while the errors on the predicted temperature are usually acceptable.

Motivated by this observation, we propose a heuristic algorithm to estimate a PVUSA model (1) of a plant using the measured generated power, theoretical clear-sky irradiance (which can be computed analytically through well-known models [9]) and the predicted temperature. To this purpose, some simplifications are in order. Let us rewrite (1) as

$$P = aI(1 + \beta I + \gamma T), \quad (3)$$

where  $\beta = b/a$ ,  $\gamma = c/a$ . At the present stage, we limit ourselves to the estimation of the main power/irradiance gain  $a$ , while considering  $\beta$  and  $\gamma$  equal to given constant values  $\bar{\beta}$  and  $\bar{\gamma}$ . This assumption can be partially justified by the fact that the sensitivity of  $P$  w.r.t.  $\beta$  and  $\gamma$  is about one order of magnitude lower than the sensitivity w.r.t.  $a$  for typical values

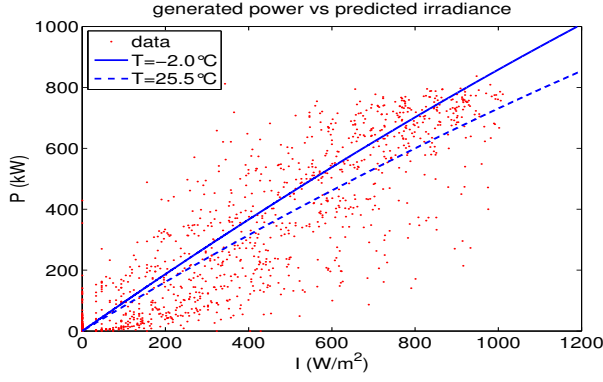


Fig. 3. Generated power  $P_k$  versus predicted irradiance  $\hat{I}_k$  (dots). For comparison purposes, we show also the same two characteristic curves of Fig. 1.

of  $I$  and  $T$ . Moreover, the ratios  $\beta$  and  $\gamma$  typically have little variability among plants using the same technology. For the devices rated in [14],  $\beta$  and  $\gamma$  lie in the following ranges:

$$\begin{aligned} \beta &\in [-2.5 \cdot 10^{-4} \div -1.9 \cdot 10^{-5}], \\ \gamma &\in [-4.8 \cdot 10^{-3} \div -1.7 \cdot 10^{-3}]. \end{aligned} \quad (4)$$

We recognize that this is a current limitation of the approach, which nevertheless shows quite good performance in practice, even when little a-priori information is available on  $\beta$  and  $\gamma$ , as illustrated in the application section.

### III. ALGORITHM

Our approach is based on updating the estimate of the gain parameter  $a$  on-line. As soon as a new data set  $\mathcal{S}^i$ ,  $i = 1, 2, \dots$  (measured power and temperature forecast) becomes available, the algorithm is run and a current estimate  $a_i$  of  $a$  is obtained. To the purpose of illustrating the procedure, let us define the following quantities:

- $\mathcal{K}^i$ : set of time indices  $k$  corresponding to data set  $\mathcal{S}^i$ ,
- $a_i$ : estimate of the gain parameter  $a$  after processing data set  $\mathcal{S}^i$ , being  $a_0$  the initial guess,
- $I_k^{cs}$ : the clear-sky solar irradiance at time  $k$ , computed according to a given theoretical model,
- $\hat{T}_k$ : temperature forecast for time interval  $k$  at the plant site, provided by a meteorological service,
- $P_k$ : measured generated power at time  $k$ ,
- $P_k^{cs}$ : estimate of the generated power under clear-sky conditions using the current parameter value at time  $k$ ,
- $\delta$ : a global binary variable, initially set to 1 and used to modify the behaviour of the algorithm as detailed below,
- $J_{max} > 0$ ,  $\alpha_{max} > 1$  and  $0 < \alpha_{min} < 1$ : user-defined threshold values.

The algorithm is recursive. When the data set  $\mathcal{S}^i$ , indexed by  $\mathcal{K}^i$ , is ready (for instance,  $\mathcal{K}^i$  may span 6 hours), we compute the estimated power under clear-sky conditions  $P_k^{cs}$ , to be used as a reference, for all  $k \in \mathcal{K}^i$  using the current value of  $a = a_{i-1}$ . Then, the following simple least squares problem

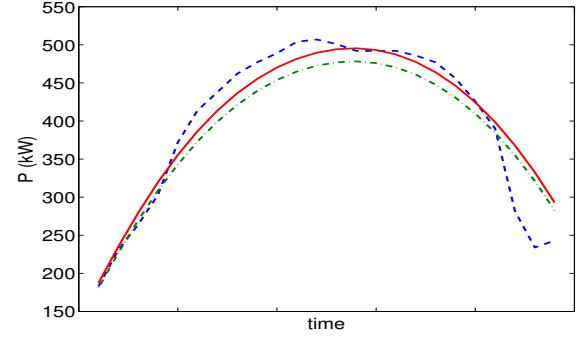


Fig. 4. Case  $\alpha > 1$  after solving problem (5). The dashed curve is  $P_k$ , the dash-dotted curve is  $P_k^{cs}$ , and the solid curve is  $\alpha P_k^{cs}$ .

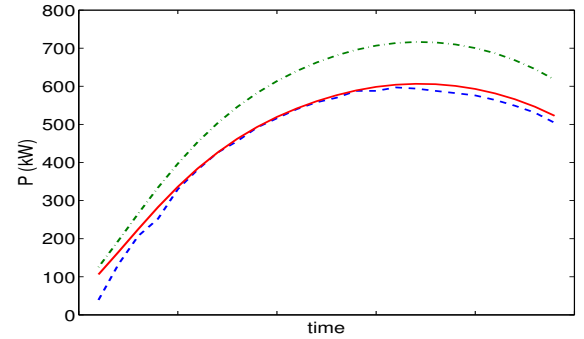


Fig. 5. Situation after solving problem (6). The dashed curve is  $P_k$ , the dash-dotted curve is  $P_k^{cs}$ , and the solid curve is  $\alpha P_k^{cs}$ .

is solved:

$$\alpha = \arg \min_{\alpha} \sum_{k \in \mathcal{K}^i} (P_k - \alpha P_k^{cs})^2. \quad (5)$$

Now, if  $\alpha > 1$ , we are in a situation similar to that represented in Fig. 4. The measured power  $P_k$  is on average above the clear-sky power curve  $P_k^{cs}$ . Since  $P_k^{cs}$  should be always above  $P_k$  for a fixed temperature, we understand from this that the true gain  $a$  is currently underestimated by the value  $a_{i-1}$ , and therefore we increase its estimate by a factor  $\alpha$ , i.e. we set  $a_i = \alpha a_{i-1}$ . We also capture the fact that this situation has occurred by setting  $\delta$  to 0. On the contrary, if  $\alpha < 1$ , we have that  $P_k$  lies on average below  $P_k^{cs}$ , but we cannot directly conclude that  $a_{i-1}$  overestimates  $a$ . For instance,  $\alpha$  turns out to be significantly less than 1 in cloudy days. Therefore, we compute a new scaling factor  $\alpha$  as follows:

$$\alpha = \max_{k \in \mathcal{K}^i} \frac{P_k}{P_k^{cs}}. \quad (6)$$

The situation is now that represented in Fig. 5. The scaled clear-sky power curve  $\alpha P_k^{cs}$  is by construction always above the curve  $P_k$ , and touches the curve  $P_k$  in the point(s) corresponding to the maximum value of the ratio  $\frac{P_k}{P_k^{cs}}$ . The following quantity:

$$J = \frac{|\sum_{k \in \mathcal{K}^i} (P_k - \alpha P_k^{cs})|}{\sum_{k \in \mathcal{K}^i} P_k} \quad (7)$$

TABLE I  
ALGORITHM FOR PARAMETER UPDATE

---



---

GIVEN: $\{I_k^{cs}, \hat{T}_k, P_k\}_{k \in \mathcal{K}^i}, a_{i-1}, \delta, \bar{\beta}, \bar{\gamma}$
Compute $P_k^{cs} = a_{i-1} I_k^{cs} (1 + \bar{\beta} I_k^{cs} + \bar{\gamma} \hat{T}_k) \forall k \in \mathcal{K}^i$
IF $(I_k^{cs} \neq 0 \forall k \in \mathcal{K}^i)$ AND $(\sum_{k \in \mathcal{K}^i} P_k \neq 0)$
$\alpha = \arg \min_{\alpha} \sum_{k \in \mathcal{K}^i} (P_k - \alpha P_k^{cs})^2$
IF $\alpha < 1$
$\alpha = \max_{k \in \mathcal{K}^i} \frac{P_k}{P_k^{cs}}$
$J = \frac{ \sum_{k \in \mathcal{K}^i} (P_k - \alpha P_k^{cs}) }{\sum_{k \in \mathcal{K}^i} P_k}$
IF $J > J_{max}$
$\alpha = 1$
END IF
END IF
ELSE
$\alpha = 1$
END IF
IF $\alpha > 1$
$\delta = 0$
ELSE IF $(\alpha < 1)$ AND $(\delta = 0)$
$\alpha = 1$
END IF
$\alpha = \min(\alpha_{max}, \max(\alpha_{min}, \alpha))$
$a_i = \alpha a_{i-1}$

---



---

represents a measure of how much the two curves  $P_k$  and  $\alpha P_k^{cs}$  are tight to each other. If  $J$  is sufficiently small, i.e., below a threshold value  $J_{max}$  fixed by the user, then it is assumed that the curve  $P_k$  was generated under (almost) clear-sky conditions, and therefore it may be acceptable to reduce the updated estimate  $a_i$  w.r.t.  $a_{i-1}$  by the scaling factor  $\alpha$ . This update is actually performed only if  $\delta = 1$ , i.e., when  $a_{i-1}$  has never been recognized as an underestimate of  $a$  (and therefore increased) in the past history. Basically, this means that once an update from below has occurred (i.e.,  $\delta$  is set to 0), only updates from below are allowed in further algorithm runs. This is done in order to avoid oscillations of the estimate. At initialization, i.e., before the first run,  $\delta$  is set to 1 since we do not know yet whether the initial estimate  $a_0$  over- or under-estimates the true  $a$ . Clearly, it is a good practice to reset  $\delta$  to 1 periodically, in order to allow the algorithm to capture possible gain decreases due to, e.g., the loss of efficiency of the plant with time. The two threshold values  $\alpha_{max} > 1$  and  $\alpha_{min} < 1$  may be used to limit the update of  $a_i$  in both directions. In particular,  $\alpha_{max}$  may be used to avoid, for instance, adaptation to outliers, while  $\alpha_{min}$  may be used to avoid adaptation to data sets generated under uniformly cloudy sky. As far as the initial guess  $a_0$  is concerned, a reasonable value can be the ratio between the nominal power, if known, and the irradiance of 1000 W/m<sup>2</sup>. The proposed algorithm is detailed in Table I.

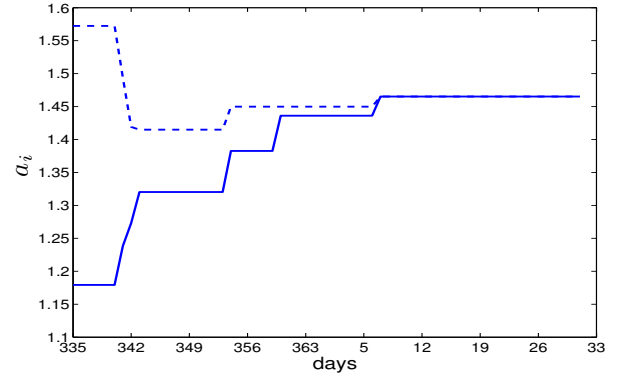


Fig. 6. Evolution of the parameter estimate  $a_i$  with time starting from under-estimated (solid) and over-estimated (dashed) initial values.

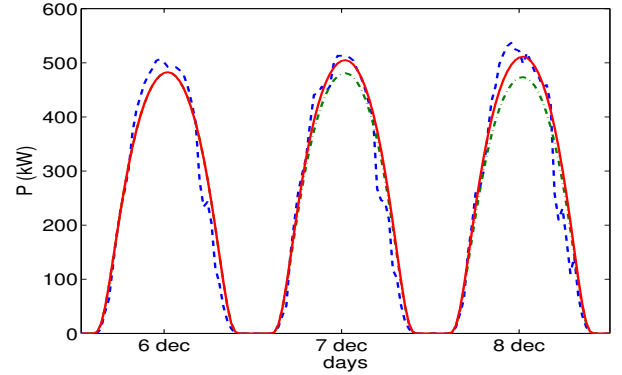


Fig. 7. Comparison of  $P_k$  (dashed line) and of  $P_k^{cs}$  computed without parameter updating (dash-dotted line) and with parameter updating (solid line). The three days are at the beginning of the data set.

#### IV. APPLICATION

In this section, we validate the proposed method using the data collected from a generation plant of 1075 kWp located in Northern Italy during the period ranging from December 1, 2011 to January 31, 2012. Further details on the plant, such as exact latitude and longitude, cannot be provided at the moment due to confidentiality constraints. Sampling time is 15 minutes. Data sets  $\mathcal{S}^i$  are constructed by considering a sliding window of 6 hours. Reference values of parameters  $\beta$  and  $\gamma$  are chosen as  $\bar{\beta} = -1.1 \cdot 10^{-4}$  and  $\bar{\gamma} = -3.3 \cdot 10^{-3}$ , which represent the central values of the ranges in (4). The following values are used for the thresholds:  $J_{max} = 0.1$ ,  $\alpha_{max} = 1.2$ ,  $\alpha_{min} = 0.95$ . The theoretical clear-sky irradiance is computed according to the celebrated ASHRAE model [9].

The solid line in Fig. 6 shows the evolution with time of the plant gain estimate  $a_i$  starting from under-estimated initial condition  $a_0$ . Fig. 7 compares the measured power  $P_k$  (dashed line) and the clear-sky curves  $P_k^{cs}$  computed without parameter updating (dash-dotted line) and with parameter updating (solid line). December 6 is the first day when a parameter update occurs (in particular, with  $\alpha > 1$ ), and therefore the dash-dotted and solid lines still overlap. The parameter updated on December 6 is used to compute  $P_k^{cs}$ .

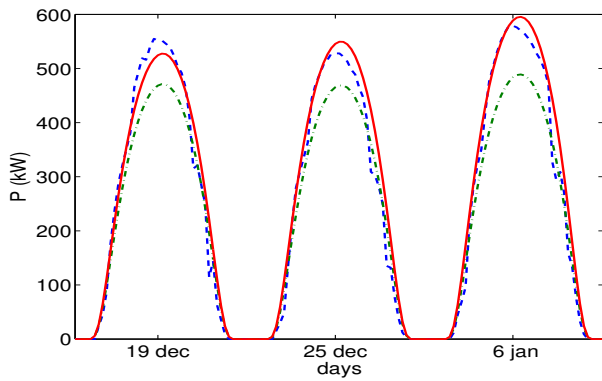


Fig. 8. Comparison of  $P_k$  (dashed line) and of  $P_k^{cs}$  computed without parameter updating (dash-dotted line) and with parameter updating (solid line). The three days are in the central part of the data set.

on December 7, and it can be seen that  $P_k^{cs}$  computed with parameter update is now above  $P_k^{cs}$  computed with the initial conditions, but it is still below  $P_k$  in some parts. It turns out that  $\alpha > 1$ , and then  $a_i$  is increased once more. The same happens on December 8. Fig. 8 shows the same comparison on December 19, December 25 and January 6. It can be seen that  $a_i$  has progressively increased, and finally the solid clear-sky curve accurately encapsulates the generated power curve on January 6, which is under clear-sky conditions at least in the first half of the day, thus showing that the model has now correctly adapted. Notice that very similar performance is obtained if the model estimation procedure is repeated, using the same data, for all the values of  $\beta$  and  $\gamma$  relative to the plants rated in [14]. The corresponding clear-sky power curves  $P_k^{cs}$  for January 6 are shown in Fig. 9 (blue lines) together with the analogous clear-sky power curve shown in Fig. 8 (red line).

To complete the analysis, Fig. 6 also shows the adaptation process of  $a_i$  starting from an over-estimated initial condition  $a_0$  (dashed line). Quite nicely, the estimate converges to the same value as in the case of an under-estimated initial condition  $a_0$  (solid line).

## V. CONCLUSION

A heuristic method for estimating a forecasting model of generation from photovoltaic plants has been proposed. It can be applied in scenarios where the historical time series of generated power and meteorological forecasts (irradiance and temperature) are the only available data. The approach relies on the estimation of a simple plant model exploiting theoretical clear-sky irradiance. Due to its simplicity, it has modest memory and computational requirements. The proposed procedure is currently implemented in a new grid management software package and it is in the pre-deployment phase at a DSO control center located in Northern Italy. This control centre supervises around 650 MV producers and about 6500 MV/LV transformers, connecting at least one LV generator each. The algorithm is currently limited to cover a one-parameter family of models but nonetheless it has performed quite well in the tests conducted. We are currently

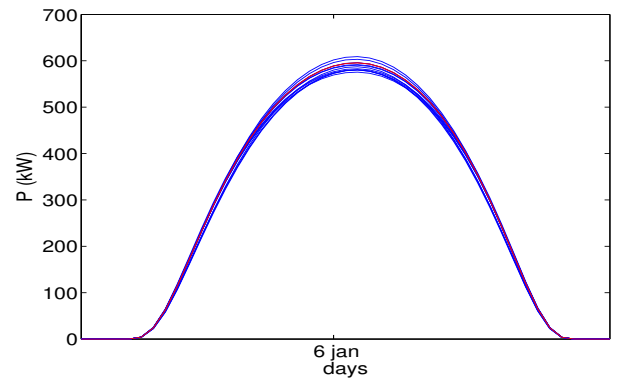


Fig. 9. January 6 clear-sky power curves  $P_k^{cs}$  for different values of  $\beta$  and  $\gamma$ .

investigating a refinement for independently updating the three parameters of the PVUSA model. This will require the development of more sophisticated techniques to detect (portions of) clear-sky days from power data.

## REFERENCES

- [1] F. Albuyeh, "Grid of the future," *IEEE Power and Energy Magazine*, vol. 7, no. 2, pp. 52–62, 2009.
- [2] C. Baldi, F. Corti, G. Di Lembo, and F. Nebiacolombo, "Monitoring and control of active distribution grid," in *Proc. CIRED Workshop 2012*, Lisbon, 2012.
- [3] G. Capizzi, C. Napoli, and F. Bonanno, "Innovative Second-Generation Wavelets Construction With Recurrent Neural Networks for Solar Radiation Forecasting," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 11, pp. 1805–1815, 2012.
- [4] J. Wu and C. Chan, "Prediction of hourly solar radiation using a novel hybrid model," *Solar Energy*, vol. 85, pp. 808–817, 2011.
- [5] A. Ragnacci, M. Pastorelli, P. Valigi, and E. Ricci, "Exploiting dimensionality reduction techniques for photovoltaic power forecasting," in *Proc. 2nd IEEE EnergyCon*, Florence, Italy, 2012.
- [6] G. Reikard, "Predicting solar radiation at high resolutions: A comparison of time series forecasts," *Solar Energy*, vol. 83, pp. 342–349, 2009.
- [7] 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, 2009.
- [8] P. Bacher, H. Madsena, and H. Nielsen, "Online short-term solar power forecasting," *Solar Energy*, vol. 83, pp. 1772–1783, 2009.
- [9] L. Wong and W. Chow, "Solar Radiation Model," *Applied Energy*, vol. 69, pp. 191–224, 2001.
- [10] C. A. Gueymard and D. Thevenard, "Monthly average clear-sky broadband irradiance database for worldwide solar heat gain and building cooling load calculations," *Solar Energy*, vol. 83, pp. 1998–2018, 2009.
- [11] M. Patel, *Wind and Solar Power Systems: Design, Analysis and Operation*. Boca Raton, FL: Taylor and Francis, 2006.
- [12] D. Tao, C. Shanxu and C. Changsong, "Forecasting power output for grid-connected photovoltaic power system without using solar radiation measurement," in *Proc. 2nd IEEE International Symposium on Power Electronics for Distributed Generation Systems (PEDG)*, 2012, pp. 773–777.
- [13] C. Whitaker, T. Townsend, J. Newmiller, D. King, E. Boyson, J. Kratochvil, D. Collier, and D. Osborn, "Application and validation of a new PV performance characterization method," in *Proc. 26th PVSC*, Anaheim, CA, 1997.
- [14] R. Dows and E. Gough, "PVUSA procurement, acceptance, and rating practices for photovoltaic power plants," Pacific Gas and Electric Company, San Ramon, CA, Tech. Rep., 1995.
- [15] D. King, W. Boyson, and J. Kratochvil, "Photovoltaic array performance model," Sandia National Laboratories, Albuquerque, NM, Tech. Rep., 2004.