



Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief Paper

Estimation of photovoltaic generation forecasting models using limited information[☆]

Gianni Bianchini^{*}, Daniele Pepe, Antonio Vicino

Dipartimento di Ingegneria dell'Informazione e Scienze Matematiche, Università di Siena, Italy

ARTICLE INFO

Article history:

Received 21 January 2019

Received in revised form 15 October 2019

Accepted 21 October 2019

Available online xxxx

Keywords:

Energy systems

Model fitting

Forecasting

Photovoltaic generation

ABSTRACT

This work deals with the problem of estimating a photovoltaic generation forecasting model in scenarios where measurements of meteorological variables (i.e., solar irradiance and temperature) at the plant site are not available. A novel algorithm for the estimation of the parameters of the well-known PVUSA model of a photovoltaic plant is proposed. Such a method is characterized by a low computational complexity, and efficiently exploits only power generation measurements, a theoretical clear-sky irradiance model, and temperature forecasts provided by a meteorological service. The proposed method is validated on real data.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

A major challenge in the integration of renewable energy sources into the grid (Schiffer, Zonetti, Ortega, Stankovic, Sezi, & Raisch, 2016) is that power generation is intermittent, difficult to control, and strongly dependent on the variation of weather conditions. For this reason, forecasting of renewable distributed generation has become a fundamental requirement in order to reliably manage conventional power plant operation, grid balancing, real-time unit dispatching (Kim, Oh, Moore, & Ahn, 2016), demand constraints (Ishizaki et al., 2016), and energy market requirements. In this respect, renewable generation forecasts on different time horizons are of special interest to various players that operate in the active grid, in particular to Distribution System Operators (DSO) and Transmission System Operators (TSO) (see Albuyeh (2009), Denholm and Margolis (2007a, 2007b) and references therein).

Concerning photovoltaic (PV) power generation (Coimbra, Kleissl, & Marquez, 2013), most contributions, focus on the problem of solar irradiance prediction (Inman, Pedro, & Coimbra, 2013; Kang & Tam, 2015; Perez, Kivalov, Schlemmer, Hemker, Renné, & Hoff, 2010). To tackle this problem, several approaches based on Artificial Neural Networks (ANNs) (Capizzi, Napoli, & Bonanno, 2012; Wu & Chan, 2011) or Support Vector Machines (Ragnacci, Pastorelli, Valigi, & Ricci, 2012) can be found

in the literature. Alternatively, classical linear time series prediction methods are used in Bacher, Madsena, and Nielsen (2009), Reikard (2009), where the considered time series is typically the global horizontal irradiance (GHI) (Wong & Chow, 2001). GHI forecasts are typically used along with temperature forecasts in a simulation model of the PV plant (Patel, 2006) in order to calculate generated power predictions. In all cases, computing reliable forecasts from predicted meteorological variables hinges upon the availability of an accurate model of the plant, be it physical or estimated from data.

Unfortunately, in many common scenarios, neither a plant model, nor direct on-site measurements of solar irradiance and other meteorological variables (e.g., temperature) are available. This is always the case with a DSO dealing with hundreds or thousands of heterogeneous, independently owned and operated PV plants; in this case, the only available data consist of generated power measurements provided by smart meters, and of irradiance and temperature forecasts provided by a meteorological service. The problem of forecasting power generation in this case is addressed in Tao, Shanxu, and Changsong (2012) by means of a neural network and in Pepe, Bianchini, and Vicino (2016, 2017) using a parametric model. In these approaches, however, further information on the cloud cover index at the plant site is assumed to be available. In Bianchini, Paoletti, Vicino, Corti, and Nebiacolombo (2013a, 2013b), a heuristic method for the estimation of the parameters of the well-known PVUSA model (Dows & Gough, 1995) based on theoretical clear-sky irradiance is presented, while in Pepe, Bianchini, and Vicino (2018), a recursive procedure based on the clear-sky criteria proposed in Reno and Hansen (2016) is devised. However, the former approach does not allow for capturing possible parameter variations or seasonal drifts, and moreover both approaches require

[☆] The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Angelo Alessandri under the direction of Editor Thomas Parisini.

^{*} Corresponding author.

E-mail addresses: giannibi@diism.unisi.it (G. Bianchini), pepe@diism.unisi.it (D. Pepe), vicino@diism.unisi.it (A. Vicino).

<https://doi.org/10.1016/j.automatica.2019.108688>

0005-1098/© 2019 Elsevier Ltd. All rights reserved.

trial-and-error procedures in order to manually tune a number of algorithm parameters whose values may vary significantly according to the climate zone.

In this paper, a novel approach to the problem of estimating the parameters of the PVUSA model in the partial information case is presented. The only historical data used by the method consist of generated power, and temperature forecasts. Our approach is based on three tests to be performed on generated power data in order to detect portions of such data that were generated under clear-sky conditions. The information contained in such portions is then exploited in a recursive parameter estimation algorithm in combination with theoretical clear-sky irradiance data provided by a suitable model. The method proposed in this paper improves over (Bianchini et al., 2013a, 2013b; Pepe et al., 2018), since it is able to adapt to parameter variations and requires the tuning of a single threshold coefficient whose physical role is well defined.

The paper is structured as follows: in Section 2 the modeling tools are introduced; in Section 3 the proposed clear-sky detection tests are developed; the model estimation procedure is presented in Section 4. Experimental validation results are reported in Section 5, and conclusions are drawn in Section 6.

2. Preliminaries

A PV plant can be efficiently modeled using the PVUSA model (Dows & Gough, 1995), which expresses the instantaneous generated power as a function of irradiance and air temperature according to the equation:

$$P = \mu_1 I + \mu_2 I^2 + \mu_3 I T, \quad (1)$$

where P , I , and T are the generated power (kW), irradiance (W/m^2), and air temperature ($^{\circ}\text{C}$), respectively, and $\mu = [\mu_1 \ \mu_2 \ \mu_3]'$ is the model parameter vector. It is important to notice that model (1) is linear in the parameters. For the purpose of this work, it is useful to express (1) in the form

$$P = \mu_1 \cdot \alpha(I, T) \cdot I, \quad (2)$$

where $\alpha(I, T)$ can be written as a function of the ratios

$$\eta_2 = \mu_2/\mu_1, \quad \eta_3 = \mu_3/\mu_1 \quad (3)$$

as

$$\alpha(I, T) = 1 + \eta_2 I + \eta_3 T. \quad (4)$$

From (2), it is apparent that μ_1 represents the main power/irradiance gain of the plant, while $\alpha(I, T)$ in (4) can be seen as correction term. In this respect, it is worth noticing that the ratios η_2 and η_3 in (3) are characterized by well-established variability ranges among different PV technologies (see Dows and Gough (1995)). Such ranges are given by:

$$\begin{aligned} \eta_2 \in [\underline{\eta}_2, \bar{\eta}_2] &= [-2.5 \times 10^{-4}, -1.9 \times 10^{-5}], \\ \eta_3 \in [\underline{\eta}_3, \bar{\eta}_3] &= [-4.8 \times 10^{-3}, -1.7 \times 10^{-3}]. \end{aligned} \quad (5)$$

This property will be used in the proposed estimation procedure. The PVUSA model can be fruitfully exploited for the purpose of computing forecasts of generated power on the basis of predicted meteorological variables. Indeed, once a correct estimate $\hat{\mu}$ of the parameter vector is available, a reliable power generation forecast \hat{P} can be obtained by substituting predicted irradiance \hat{I} and temperature \hat{T} , provided by a meteorological service, into the model equation (1). Similarly, a clear-sky generation forecast \hat{P}^{cs} can be obtained by using the theoretical irradiance I^{cs} at the plant location, as provided by a suitable model, along with a temperature forecast.

Despite its simplicity, very good accuracy is obtained from the PVUSA model when μ is estimated using measured irradiance and temperature data via, e.g., standard least squares fitting (see, e.g., Bianchini et al. (2013b)).

A DSO that manages a high number of independent generation facilities may not have access to time series of irradiance and temperature measured on the premises of each plant, while power generation data are always available through meters. In order to estimate model parameters, replacing the measured values of I and T with forecasts \hat{I} and \hat{T} provided by a meteorological service is not a viable solution, due to the fact that forecasting errors on the irradiance are in general too large. On the contrary, temperature forecasts are quite reliable and can be used in place of actual measurements (see Bianchini et al. (2013a, 2013b) for details).

In this paper, a theoretical model for the global clear-sky irradiance I^{cs} on a given surface is also needed. Any of the several different models in the literature (Ineichen, 2006) is suitable for the proposed technique. In the experimental part of this work, the Heliodon simulator model (Meinel & Meinel, 1976) is used. We refer the reader to Section 2.2 of Bianchini, Pepe, and Vicino (2019) for details.

3. Clear-sky data detection

In this paper, the following key idea is exploited for the purpose of estimating the parameters of the PVUSA model (1) of a PV plant without resorting to on-site irradiance measurements. Given a time series composed of generated power measurements and temperature forecasts (or measurements, if available), we propose three tests to be performed on the data in order to detect portions of the power curve which have been generated under a clear-sky condition; this allows for fitting the parameters of the PVUSA model to such data by using the theoretical clear-sky irradiance in place of the actual measured irradiance. This section deals with the derivation of such tests.

In view of (5), suitable bounds can be derived on $\alpha(I, T)$ and P in the PVUSA model (2)–(4). Indeed, from (4) and (5), it is easily checked that

$$\underline{\alpha}(I, T) \leq \alpha(I, T) \leq \bar{\alpha}(I, T), \quad (6)$$

where

$$\underline{\alpha}(I, T) = \begin{cases} 1 + \underline{\eta}_2 I + \underline{\eta}_3 T, & \text{if } T \geq 0 \\ 1 + \underline{\eta}_2 I + \bar{\eta}_3 T, & \text{if } T < 0 \end{cases} \quad (7)$$

$$\bar{\alpha}(I, T) = \begin{cases} 1 + \bar{\eta}_2 I + \bar{\eta}_3 T, & \text{if } T \geq 0 \\ 1 + \bar{\eta}_2 I + \underline{\eta}_3 T, & \text{if } T < 0. \end{cases} \quad (8)$$

Moreover, for realistic values of I and T , it always holds that $\underline{\alpha}(I, T) > 0$ and $\bar{\alpha}(I, T) < 1$. From (6) and (2), the following bound on P is obtained:

$$\mu_1 \cdot I \cdot \underline{\alpha}(I, T) \leq P \leq \mu_1 \cdot I \cdot \bar{\alpha}(I, T). \quad (9)$$

Let us now consider a time series $\{P(j), I(j), T(j)\}$ of the variables in (1), where j represents a discrete time index. The increment of $P(j)$ can be expressed as

$$\begin{aligned} \Delta P(j) &= P(j) - P(j-1) \\ &= \mu_1 [I(j-1)\Delta\alpha(j) + \Delta I(j)\alpha(I(j), T(j))], \end{aligned} \quad (10)$$

where $\Delta I(j) = I(j) - I(j-1)$ and $\Delta\alpha(j) = \alpha(I(j), T(j)) - \alpha(I(j-1), T(j-1))$.

Let $\Delta T(j) = T(j) - T(j-1)$. Taking into account (7)–(8), it is easily checked that the following bounds on $\Delta\alpha(j)$ hold:

$$\underline{\Delta\alpha}(j) \leq \Delta\alpha(j) \leq \bar{\Delta\alpha}(j), \quad (11)$$

where

$$\begin{bmatrix} \frac{\Delta\alpha(j)}{\Delta\alpha(j)} \end{bmatrix} = Q(j) \begin{bmatrix} \Delta I(j) \\ \Delta T(j) \end{bmatrix} \quad (12)$$

and the matrix $Q(j)$ depends on the signs of $\Delta I(j)$ and $\Delta T(j)$ according to the following table:

$Q(j)$	$\Delta T(j) \geq 0$	$\Delta T(j) < 0$
$\Delta I(j) \geq 0$	$\begin{bmatrix} \eta_2 & \eta_3 \\ \bar{\eta}_2 & \bar{\eta}_3 \end{bmatrix}$	$\begin{bmatrix} \eta_2 & \bar{\eta}_3 \\ \bar{\eta}_2 & \eta_3 \end{bmatrix}$
$\Delta I(j) < 0$	$\begin{bmatrix} \bar{\eta}_2 & \bar{\eta}_3 \\ \eta_2 & \eta_3 \end{bmatrix}$	$\begin{bmatrix} \bar{\eta}_2 & \eta_3 \\ \eta_2 & \bar{\eta}_3 \end{bmatrix}$

In view of (10), this allows to derive the following bounds on $\Delta P(j)$:

$$\mu_1 \delta_P(j) \leq \Delta P(j) \leq \mu_1 \bar{\delta}_P(j), \quad (13)$$

where

$$\begin{bmatrix} \delta_P(j) \\ \bar{\delta}_P(j) \end{bmatrix} = R(j) \begin{bmatrix} I(j) - 1 \\ \Delta I(j) \end{bmatrix} \quad (14)$$

and the matrix $R(j)$, depending on the sign of $\Delta I(j)$, is given by

$R(j)$	
$\Delta I(j) \geq 0$	$\begin{bmatrix} \frac{\Delta\alpha(j)}{\Delta\alpha(j)} & \alpha(I(j), T(j)) \\ \frac{\Delta\alpha(j)}{\Delta\alpha(j)} & \bar{\alpha}(I(j), T(j)) \end{bmatrix}$
$\Delta I(j) < 0$	$\begin{bmatrix} \frac{\Delta\alpha(j)}{\Delta\alpha(j)} & \bar{\alpha}(I(j), T(j)) \\ \frac{\Delta\alpha(j)}{\Delta\alpha(j)} & \alpha(I(j), T(j)) \end{bmatrix}$

The bounds (9) and (13) allow to devise the sought tests. Let us consider a time interval \mathcal{J} , and the following associated time series

$$\mathcal{P}_{\mathcal{J}} = \{ \{P^m(j), T(j), P^{cs}(j)\}, j \in \mathcal{J} \}, \quad (15)$$

where, for each time instant j , $P^m(j)$ represents the measured plant power reported by meters, $T(j)$ is a temperature forecast (or measurement), and $P^{cs}(j)$ is the clear-sky generated power predicted by a PVUSA model characterized by given values of the parameters μ_1, μ_2 , and μ_3 , i.e.,

$$P^{cs}(j) = \mu_1 I^{cs}(j) \alpha(I^{cs}(j), T(j)). \quad (16)$$

Clearly, by (9),

$$\begin{aligned} \mu_1 \cdot I^{cs}(j) \cdot \alpha(I^{cs}(j), T(j)) &\leq P^{cs}(j) \\ &\leq \mu_1 \cdot I^{cs}(j) \cdot \bar{\alpha}(I^{cs}(j), T(j)). \end{aligned} \quad (17)$$

Let

$$j_{max} = \arg \max_{j \in \mathcal{J}} \{I^{cs}(j)\}, \quad (18)$$

$$I_{max}^{cs} = I^{cs}(j_{max}), \quad (19)$$

$$P_{max}^{cs} = P^{cs}(j_{max}) = \mu_1 I_{max}^{cs} \alpha(I_{max}^{cs}, T(j_{max})). \quad (20)$$

The quantities I_{max}^{cs} , j_{max} , and P_{max}^{cs} define, respectively, the maximum clear-sky irradiance within the given time window \mathcal{J} , the time index for which this maximum value occurs, and the predicted clear-sky generated power at j_{max} . Note that, since the sensitivity of power with respect to temperature is quite small in general, it can be assumed that the actual power peak within a given time window always corresponds to the irradiance peak.

Normalizing (17) with respect to P_{max}^{cs} and taking (6) into account yields the following bounds on the ratio $\frac{P^{cs}(j)}{P_{max}^{cs}}$:

$$\underline{\gamma}_1(j) \leq \frac{P^{cs}(j)}{P_{max}^{cs}} \leq \bar{\gamma}_1(j), \quad (21)$$

where

$$\begin{aligned} \underline{\gamma}_1(j) &= \frac{\alpha(I^{cs}(j), T(j))}{\bar{\alpha}(I_{max}^{cs}, T(j_{max}))} \cdot \frac{I^{cs}(j)}{I_{max}^{cs}}, \\ \bar{\gamma}_1(j) &= \frac{\bar{\alpha}(I^{cs}(j), T(j))}{\alpha(I_{max}^{cs}, T(j_{max}))} \cdot \frac{I^{cs}(j)}{I_{max}^{cs}}. \end{aligned} \quad (22)$$

It is important to observe that (21)–(22) define a condition to be satisfied by the clear-sky power time series provided that the reference model is given by (1), and that such bounds are independent of the model parameters. Condition (21) can therefore be exploited as a first test in order to single out candidate time windows \mathcal{J} in which power data have been generated under clear-sky conditions. To this aim, given the time series $\{P^m(j), T(j)\}$, $j \in \mathcal{J}$, the ratio $\frac{P^m(j)}{P^m(j_{max})}$ is compared against the bounds (22), yielding the following test:

$$\boxed{\text{T1}} \quad \underline{\gamma}_1(j) \leq \frac{P^m(j)}{P^m(j_{max})} \leq \bar{\gamma}_1(j), \quad \forall j \in \mathcal{J}. \quad (23)$$

The satisfaction of T1 is in general not sufficient to classify power data within \mathcal{J} as having been generated under a clear-sky condition. Specifically, if the sky is partially cloudy during the time interval \mathcal{J} , the measured power may heavily oscillate, but could remain quite close to the clear-sky power at the maximum (Reno & Hansen, 2016), thus satisfying (23). To overcome this issue, a further condition on the normalized increment of the power time series is derived. Let $\delta_P^{cs}(j)$ and $\bar{\delta}_P^{cs}(j)$ be defined by (14) evaluated for $I(j) = I^{cs}(j)$ and $\Delta I(j) = \Delta I^{cs}(j) = I^{cs}(j) - I^{cs}(j-1)$. The increment $\Delta P^{cs}(j) = P^{cs}(j) - P^{cs}(j-1)$ satisfies

$$\mu_1 \delta_P^{cs}(j) \leq \Delta P^{cs}(j) \leq \mu_1 \bar{\delta}_P^{cs}(j) \quad (24)$$

by (13). Normalizing (24) with respect to P_{max}^{cs} , the following bounds on the normalized increment $\frac{\Delta P^{cs}(j)}{P_{max}^{cs}}$ are obtained:

$$\underline{\gamma}_2(j) \leq \frac{\Delta P^{cs}(j)}{P_{max}^{cs}} \leq \bar{\gamma}_2(j), \quad (25)$$

where

$$\begin{aligned} \underline{\gamma}_2(j) &= \frac{\delta_P^{cs}(j)}{\alpha(j_{max})} \cdot \frac{1}{I_{max}^{cs}}, \\ \bar{\gamma}_2(j) &= \frac{\bar{\delta}_P^{cs}(j)}{\alpha(j_{max})} \cdot \frac{1}{I_{max}^{cs}}. \end{aligned} \quad (26)$$

Note that the bounds (25)–(26), similar to (21)–(22), do not depend on the model parameters. Condition (25) provides a second criterion for classifying a time window \mathcal{J} of measured power data points as clear-sky. The following test is therefore introduced:

$$\boxed{\text{T2}} \quad \underline{\gamma}_2(j) \leq \frac{\Delta P^m(j)}{P^m(j_{max})} \leq \bar{\gamma}_2(j), \quad \forall j \in \mathcal{J}, \quad (27)$$

where $\Delta P^m(j)$ is the increment of the measured power, i.e., $\Delta P^m(j) = P^m(j) - P^m(j-1)$.

Tests T1 and T2 detect deviations in the shape of the normalized power curve from the clear-sky condition caused by cloudiness in different scenarios. However, due to normalization, such conditions may turn out to be fulfilled on a given time window \mathcal{J} when the corresponding data are generated under perfectly uniform cloudiness, i.e., when the actual irradiance satisfies $I(j) = \beta I^{cs}(j) \forall j \in \mathcal{J}$, where $0 < \beta < 1$ is a constant that represents a uniform cloud cover factor (see Kimura and Stephenson (1969)). If the data collected within such a time window are used to perform a model parameter adaptation step in a recursive estimation procedure, the algorithm may tend to underestimate the power/irradiance gain of the plant at such step. This fact may be detrimental when a long series of data collected under uniform cloudiness is processed. To mitigate this effect, a further

test is introduced. Suppose that a current estimate $\hat{\mu}$ of the model parameters is available, and let $\hat{P}^{cs}(j)$ be the current estimate of the generated power under clear-sky conditions, computed via (16) using $\hat{\mu}$. Let \hat{P}_{max}^{cs} be the peak value of $\hat{P}^{cs}(j)$ in \mathcal{J} , i.e., $\hat{P}_{max}^{cs} = \hat{P}^{cs}(j_{max})$. Provided that T1 and T2 are passed by the data in time window \mathcal{J} , the following further condition is introduced, which involves a comparison of the maximum currently predicted clear-sky power \hat{P}_{max}^{cs} with the corresponding generated power $P^m(j_{max})$ as follows:

$$\boxed{\text{T3}} \quad \frac{P^m(j_{max})}{\hat{P}_{max}^{cs}} \geq 1 - \epsilon, \quad (28)$$

where $0 < \epsilon < 1$ is a parameter chosen by the designer, typically a number slightly higher than 0. T3 has the specific role of detecting, under the condition that T1 and T2 are satisfied, whether the peak value of measured power within the considered time window lies above a given fraction of the clear-sky power currently estimated by the model. Condition (28) is satisfied when (a) the model is currently underestimating clear-sky power, or (b) the current model is overestimating the generated power by a small amount, or (c) uniform cloudiness is present within the given time window, so that generation is marginally lower than the clear-sky power currently predicted by the model. With the exception of the latter case, the simultaneous satisfaction of T1, T2, and T3 requires that the model parameters be adapted in order to fit the measured power series with the predicted one within \mathcal{J} .

Remark 1. The parameter ϵ plays a key role in detecting whether the clear-sky curve provided by the model matches or underestimates power data satisfying T1 and T2, which are related to the shape of the normalized power curve. Setting this value very close to zero allows for good adaptation when the model is underestimating the clear-sky power (for this reason it is advisable to choose an underestimate of μ_1 as the initial guess in the estimation procedure, as detailed in the next section). Higher values, on the other hand, allow for adjusting the model when it is overestimating; the latter case is very important for capturing possible slow parameter drifts as well as seasonal variations in the accuracy of the theoretical clear-sky model. However, increasing ϵ may cause adaptation to long series of data generated under uniform cloudiness. Actually, it can be shown (see Remark 1 in Bianchini et al. (2019)) that under uniform cloud cover factor β , T3 is passed when β approximately satisfies

$$\beta \gtrsim \hat{\mu}_1 \cdot \frac{1000}{P_{nom}} \cdot (1 - \epsilon), \quad (29)$$

where P_{nom} is the nominal plant power and $\hat{\mu}_1$ is the current estimate of μ_1 . The relationship (29) provides an interpretation of the parameter ϵ and represents a possible guideline for tuning such parameter on the basis of the minimum value of the cloud cover factor for which the designer allows data generated under uniform cloudiness to be considered for parameter adaptation.

4. Model estimation

According to the observations in the previous sections, we now introduce the proposed PVUSA plant model estimation method, which yields an on-line update of the parameter vector estimate $\hat{\mu}$ by relying only on the information contained on a time series composed by measured power P^m and forecast (or measured) temperature T . The model estimation procedure is recursive, and combines tests T1, T2, T3 with a standard Recursive Least-Squares (RLS) algorithm using a dynamical time window. The following definitions are instrumental for building up the procedure:

- k : present time index;
- d : present day;
- $\mathcal{I}_d = [k_d, \bar{k}_d]$: time interval corresponding to light hours in day d , i.e., $I^{cs}(k) > 0$ for all $k_d \leq k \leq \bar{k}_d$;
- $\mathcal{J}_{k,l}$: set of time indices corresponding to a time window of given length l ending at k , i.e., $\mathcal{J}_{k,l} = \{k-l+1, \dots, k\}$;
- $\hat{\mu}(k)$: estimate of the parameter vector at time k , being $\hat{\mu}(0)$ the initial guess;
- $I^{cs}(j)$: theoretical clear-sky solar irradiance at time step j , computed according to a suitable model for the plant site;
- $T(j)$: temperature forecast (or measurement, if available) at time j at the plant site, provided by a meteorological service;
- $P^m(j)$: measured generated power at time j ;
- $D(j) = \{P^m(j), T(j), I^{cs}(j)\}$: data sample at time j ;
- $\mathcal{D}(\mathcal{J}) = \{D(j), j \in \mathcal{J}\}$: data set pertaining to time window \mathcal{J} ;
- ϵ : fixed threshold value ($0 < \epsilon < 1$);
- l_{min} : minimum time window length.

The estimation algorithm is constructed as follows. The procedure is reset on each day d at time $k = k_d$. The current parameter estimate $\hat{\mu}(k_d)$ is initialized with the last estimate obtained on day $d-1$. An initial data set $\mathcal{D}(\mathcal{J}_{k,l_{min}})$ is constructed at time $k = k_d + l_{min} - 1$ corresponding to an initial time window $\mathcal{J}_{k,l_{min}}$ of length l_{min} . If $\mathcal{D}(\mathcal{J}_{k,l_{min}})$ does not pass T1, T2 and T3, then the procedure is reset at time $k = k_d + 1$. Otherwise (i.e., if $\mathcal{D}(\mathcal{J}_{k,l_{min}})$ is recognized as generated under clear-sky), a new data sample $D(k)$ is acquired at each following step k and added to the current data set $\mathcal{D}(\mathcal{J}_{k,l})$, incrementing the length of the time window $\mathcal{J}_{k,l}$ by one. Then, T1, T2, and T3 are performed on $\mathcal{D}(\mathcal{J}_{k,l})$. If tests are passed, then further data samples are added to the data set until one of the tests fails (or the end of the day is reached) at some time k' . When this occurs, the data set $\mathcal{D}(\mathcal{J}_{k'-1,l-1})$ is deemed to be generated under clear-sky conditions and an RLS adaptation step is performed using such data in order to obtain an updated parameter estimate $\hat{\mu}(k')$. Then, the algorithm is reset at time $k = k'$ and repeated. A detailed description of the procedure is reported in Algorithm 1. Concerning the selection of the initial parameter guess $\hat{\mu}(0)$, the following observations are in order.

Algorithm 1 Parameter estimation

```

1: On each day  $d$ 
2:  $k' \leftarrow k_d$ 
3: while  $k' + l_{min} - 1 \leq k_d$  do
4:   for  $k = k' : k' + l_{min} - 1$  do   ▷ Get the initial data set  $\mathcal{D}(\mathcal{J}_{k,l_{min}})$ 
5:     Acquire  $D(k)$ 
6:   end for
7:   if  $\mathcal{D}(\mathcal{J}_{k,l_{min}})$  does not satisfy T1, T2, T3 then
8:      $k' \leftarrow k' + 1$    ▷  $D(k')$  is rejected and the algorithm is reset at
time  $k' + 1$ 
9:   else
10:     $l \leftarrow l_{min}$ 
11:    do   ▷ Try to increase the window length by one
12:       $k \leftarrow k + 1$ 
13:       $l \leftarrow l + 1$ 
14:      Acquire  $D(k)$ 
15:    while  $\mathcal{D}(\mathcal{J}_{k,l})$  satisfies T1, T2, T3 and  $k \leq \bar{k}_d$ 
16:      Compute updated parameter estimate  $\hat{\mu}(k)$  via RLS using
 $\mathcal{D}(\mathcal{J}_{k-1,l-1})$ 
17:       $k' \leftarrow k + 1$ 
18:    end if
19: end while

```

- As previously stated, a good guess for the main power/irradiance gain μ_1 is represented by $\hat{\mu}_1(0) = P_{nom}/1000$, where P_{nom} denotes the nominal plant power (Bianchini et al., 2013a; Pepe et al., 2017). As pointed out in Remark 1,

it is appropriate to start with an underestimate of this value, e.g., 75%, to ensure faster parameter adaptation.

- As for the initial values $\hat{\mu}_2(0)$ and $\hat{\mu}_3(0)$, it is convenient to choose them so that $\mu_2(0)/\mu_1(0)$ and $\mu_3(0)/\mu_1(0)$ are equal to the central values of the intervals \mathcal{S}_2 and \mathcal{S}_3 in (5), respectively (Pepe et al., 2017).

5. Experimental results

Two experiments have been run to evaluate the performance of the proposed estimation algorithm. In the first one, both model estimation and validation have been conducted using measured data (power and temperature for estimation, irradiance and temperature for forecasting) in order to assess the performance of the estimation procedure net of errors due to inaccuracies of weather forecasts. In the second, meteorological predictions have been used both for model parameter fitting and generation forecasting. The latter scenario corresponds to a typical DSO use case. To this purpose, the forecasting performance of the PVUSA model estimated using the proposed procedure has been evaluated on the widely used Day-Ahead (DA) and Hour-Ahead (HA) forecasts (International Energy Agency, 2013). The details on how such forecasts can be computed from the estimated model are omitted here due to space limitations and the reader is referred to Section 5 of Pepe et al. (2017). The following standard error measures are employed: Root Mean Square Error (RMSE), Mean Bias Error (MBE), Mean Absolute Percentage Error (MAPE), Normalized RMSE (NRMSE), $R^2 = 1 - \text{NRMSE}^2$. Details on how such indices are computed in the specific contexts of DA or HA forecasting are provided in Pepe et al. (2017). Two further indices, i.e., RMSE_{NP} and MAPE_{NP} , are the normalized values of RMSE and MAPE with respect to the nominal plant power P_{nom} and are of practical interest for network operation. In particular, values lower than 10% are considered acceptable for network operation (Coimbra et al., 2013; Widiss & Porter, 2014).

The performance indices achieved using the proposed approach have been compared to those obtained using both the One-Day-ahead Naive Predictor (ODNP) and a PVUSA model estimated via a standard RLS algorithm in the complete information case, i.e., using actual measurements of generated power, irradiance and temperature (SRLS).

For the two experiments performed, the following data sets have been used, respectively:

(D1) data from a PV plant P1 with nominal power $P_{nom} = 960$ kWp located in the campus of the University of Salento, in Monteroni di Lecce, Puglia, Italy (see Malvoni, De Giorgi, and Congedo (2016) for details). Data, ranging from March 5th, 2012 to December 31st, 2013, consist of hourly samples of averaged measured power $P^m(k)$, air temperature $T^m(k)$, and normal irradiance (the latter used only for comparison in the SRLS benchmark);

(D2) data from a PV plant P2 with nominal power $P_{nom} = 920$ kWp located in Sardinia. Data, ranging from February 2nd, 2012, to May 1st, 2012, consist of hourly samples of averaged measured power $P^m(k)$, one day-ahead forecasts of air temperature $\hat{T}(k)$, and one day-ahead forecasts of normal irradiance. Information about the quality of such forecasts is reported in Table 1 of Bianchini et al. (2019).

The sampling time has been chosen equal to $\tau_s = 1$ h, and $I^s(k)$ has been generated using the Heliodon model. Clearly, only time indices k corresponding to hours of light have been considered. The minimum window length has been set to $l_{min} = 3$. Taking higher values has proven to make the algorithm unnecessarily selective for the chosen sampling time.

The initial parameter vector has been chosen according to the criteria in Section 4, i.e., $\hat{\mu}_1(0) = 0.75 P_{nom}/1000$, $\hat{\mu}_2(0) = -1.34 \times 10^{-4} \cdot \hat{\mu}_1(0)$, and $\hat{\mu}_3(0) = -3.25 \times 10^{-3} \cdot \hat{\mu}_1(0)$. Further

Table 1

Performance comparison of CSD, SRLS and ODNP computed starting from day 28 (D1).

Performance indices	CSD	SRLS	ODNP
RMSE (kW)	31.0	23.1	143.2
MAPE	31%	26%	109%
MBE (kW)	-7.01	-6.73	3.00
DA Forecast			
R^2	0.98	0.99	0.65
NRMSE	0.13	0.10	0.59
RMSE_{NP}	0.032	0.024	0.15
MAPE_{NP}	2.2%	1.5%	8.4%

details on the setup can be found in Section 7.1 of Bianchini et al. (2019). Concerning the choice of ϵ , it is worth recalling (see Remark 1) that in order for CS test 3 to reject uniformly cloudy data with a CCF $\beta \leq \beta_0$, ϵ can be chosen approximately as

$$\epsilon = 1 - \frac{P_{nom}}{1000} \cdot \frac{1}{\hat{\mu}_1} \cdot \beta_0, \quad (30)$$

where $\hat{\mu}_1$ represents the currently available estimate of μ_1 . Therefore, we find it convenient to fix the CCF bound β_0 and adjust ϵ dynamically via (30) as soon as a new estimate $\hat{\mu}_1$ is computed. In this respect, we observe that the range of variability of the CCF depends on the climate of the macro-area where the plant is located, which is usually available. For the Italian case, typical values of the CCF range from 0.5 to 1 (Spena, D'Angiolini, & Strati, 2010). In the experiments of this section, we choose $\beta_0 = 0.9$. However, higher/smaller values of β_0 within the typical variability range make the CS detection algorithm more/less selective. The evaluation of this effect for the proposed experiments is not presented here due to space limitations; an extensive discussion can be found in Bianchini et al. (2019).

5.1. Validation on measured data (D1)

The proposed method (denoted as CSD) has been evaluated with reference to day-ahead (DA) forecasts (Pepe et al., 2017) by taking actual measurements of meteorological variables as the respective forecasts. The performance is compared with that of both the ODNP and the SRLS.

The time evolution of the parameters estimated using CSD and SRLS algorithms are shown in Fig. 1. Since the two algorithms use different data, namely theoretical irradiance for CSD and measured irradiance for SRLS, it is not surprising that parameters tend to slightly different values.

As far as the forecasting performance is concerned, all error measures on DA predictions were computed over the period starting from day 28, in order to guarantee at least a rough adaptation of the model parameters. In Table 1 the performance indices achieved by the proposed CSD approach are compared with SRLS and ODNP. Errors computed on CSD and SRLS are comparable and clearly show better performance with respect to the ODNP.

A visual representation of the algorithm behavior with special attention to clear-sky detection is shown in Fig. 2. In those graphs, sequences of red markers denote time windows in which the measured power is detected as being generated under a clear-sky condition. The plots show three, non consecutive days: days 32 and 419 are completely clear-sky; day 91 is a partially clear-sky day, in which about a half of the samples is rejected by the algorithm.

Finally, in Fig. 3, DA forecasts provided by CSD and SRLS during three different days and under three different weather conditions are compared with the measures of generated power.

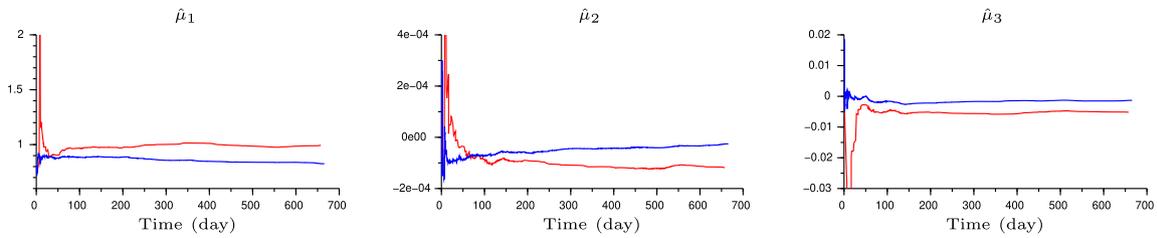


Fig. 1. PVUSA parameters estimation using the CSD algorithm (red line) and a SRLS algorithm (blue line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

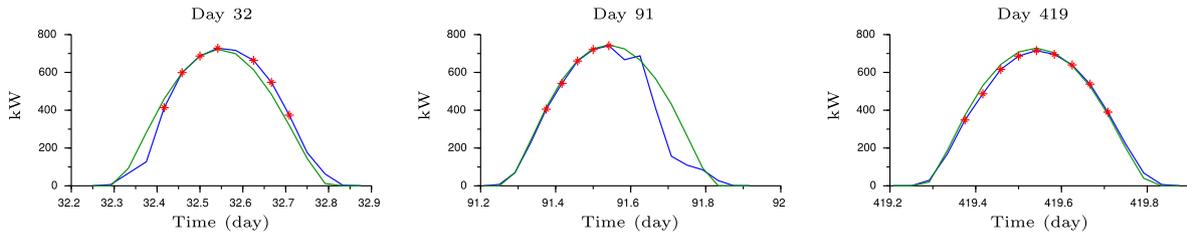


Fig. 2. Visual representation of an algorithm run (D1). Measured power is in blue, current predicted clear-sky power is in green, red markers denote detected clear-sky windows. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

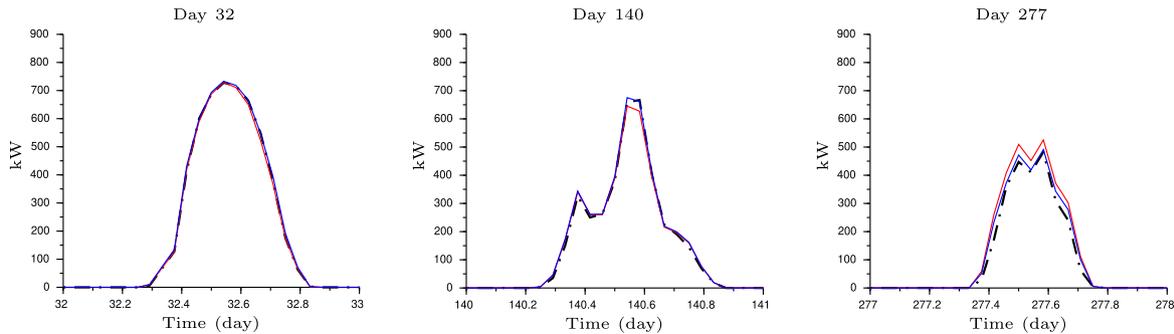


Fig. 3. (D1): Comparison between the measured power (dash dot line), DA CSD forecast (red line) and DA SRLS forecast (blue line). From right to left, a clear-sky day, an overcast day and a partially clear-sky day. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5.2. Validation on predicted data (D2)

In this section a typical DSO scenario is reproduced, in which it is assumed that measurements of weather variables are not available at the plant site. Therefore, measured power and temperature forecasts are used to estimate the plant parameters, while irradiance and temperature forecasts are used to cast predictions of generated power. In this scenario, the performance of the proposed method has been evaluated with reference to both day-ahead (DA) and hour-ahead (HA) forecasts, and compared with the performance achieved by SRLS and ODNP. Forecasting error measures are reported in Table 2. While the ODNP still has the worst performance indices, CSD performs even better than SRLS. However, it should be observed that forecasting errors in this case are to a large extent due to the quality of weather reports (see Table 1 in Bianchini et al. (2019)).

Three examples of DA forecasts computed using the CSD approach and SRLS during different weather conditions are shown in Fig. 4.

5.3. Further remarks

With reference to Tables 1 and 2, it is important to observe that the normalized errors (MAPE_{NP}) computed on DA forecasts

Table 2

Performance comparison of CSD, SRLS and ODNP computed starting from day 28 (D2).

Performance indices		CSD	SRLS	ODNP
DA Forecast	RMSE (kW)	117.9	118.5	193.3
	MAPE	58.8%	55.2%	85.6%
	MBE (kW)	-7.69	35.6	-5.6
	R ²	0.799	0.797	0.458
	NRMSE	0.448	0.451	0.736
	RMSE _{NP}	0.128	0.129	0.201
	MAPE _{NP}	8.3%	9.8%	12.4%
HA Forecast	RMSE (kW)	138.2	136.2	-
	MAPE	52.1%	46.0%	-
	MBE (kW)	-25.8	33.0	-
	R ²	0.655	0.665	-
	NRMSE	0.588	0.579	-
	RMSE _{NP}	0.150	0.148	-
	MAPE _{NP}	10.0%	11.9%	-

are below 10%, which demonstrates viability for network operation. Furthermore, the performance indices achieved by CSD are very close to those obtained by SRLS, i.e., via a PVUSA model estimated using measured irradiance. It is also worth pointing out that CSD significantly improves over Bianchini et al. (2013a, 2013b) as far as all error measures are concerned. In particular,

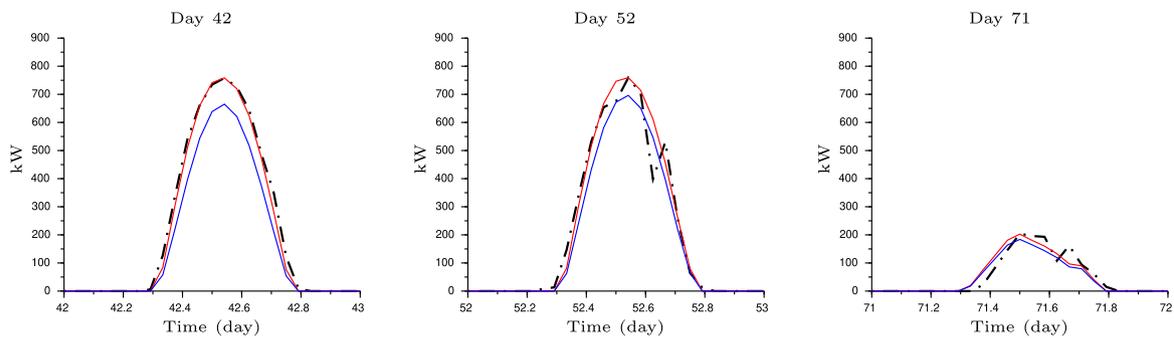


Fig. 4. (D2): Comparison between measured power (dash dot line), DA CSD forecast (red line) and DA SRLS forecast (blue line). From right to left, a clear-sky day, an overcast day and a uniformly overcast day. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

on data set (D2) it bears a 45% reduction in ($MAPE_{NP}$) and a 39% reduction in ($RMSE_{NP}$) for DA forecasts.

The proposed algorithm has been implemented in Scilab. Each iteration took on average less than one second on an i7 2.6 Ghz processor, thus demonstrating that the approach carries an extremely low computational burden.

6. Conclusions

In this paper, an efficient technique for estimating a forecasting model of photovoltaic power generation from limited information has been proposed. The approach is based on a set of tests performed on power data combined with a recursive estimation framework. It only exploits the time series of generated power and forecasts of temperature, the latter obtained from a meteorological service. The procedure especially fits the typical scenario where the network operator has no access to on-site measurements of irradiance and temperature, due to the large number of plants connected to the grid.

The algorithm has been extensively validated on two plants located in Italy, both on measured data and on forecasts of weather variables. The latter case reproduces a typical DSO scenario. Experiments worked out show very good forecasting performance, with limited computational burden.

References

- Albuyeh, F. (2009). Grid of the future. *IEEE Power and Energy Magazine*, 7(2), 52–62.
- Bacher, P., Madsena, H., & Nielsen, H. A. (2009). Online short-term solar power forecasting. *Solar Energy*, 83, 1772–1783.
- Bianchini, Gianni, Paoletti, Simone, Vicino, Antonio, Corti, Franco, & Nebiacolombo, Federico (2013a). Estimation of a simple model of solar power generation using partial information. In *52nd IEEE conference on decision and control* (pp. 996–1000).
- Bianchini, Gianni, Paoletti, Simone, Vicino, Antonio, Corti, Franco, & Nebiacolombo, Federico (2013b). Model estimation of photovoltaic power generation using partial information. In *Innovative smart grid technologies Europe (ISGT EUROPE), 2013 4th IEEE/PES* (pp. 1–5). IEEE.
- Bianchini, Gianni, Pepe, Daniele, & Vicino, Antonio (2019). Estimation of photovoltaic generation forecasting models using limited information. arXiv, <https://arxiv.org/abs/1903.04827>.
- Capizzi, G., Napoli, C., & Bonanno, F. (2012). Innovative second-generation wavelets construction with recurrent neural networks for solar radiation forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, 23(11), 1805–1815.
- Coimbra, C. F., Kleissl, Jan, & Marquez, Ricardo (2013). Overview of solar-forecasting methods and a metric for accuracy evaluation. In Jan Kleissl (Ed.), *Solar energy forecasting and resource assessment*. Waltham, Massachusetts, USA: Elsevier Inc.
- Denholm, P., & Margolis, Robert M. (2007a). Evaluating the limits of solar photovoltaics (PV) in electric power systems utilizing energy storage and other enabling technologies. *Energy Policy*, 35(9), 4424–4433.
- Denholm, P., & Margolis, Robert M. (2007b). Evaluating the limits of solar photovoltaics (PV) in traditional electric power systems. *Energy Policy*, 35(5), 2852–2861.
- Dows, R., & Gough, E. J. (1995). *PVUSA procurement, acceptance, and rating practices for photovoltaic power plants: Technical report*, San Ramon, CA: Pacific Gas and Electric Company.
- Ineichen, P. (2006). Comparison of eight clear sky broadband models against 16 independent data banks. *Solar Energy*, 80(4), 468–478.
- Inman, R. H., Pedro, Hugo T. C., & Coimbra, Carlos F. M. (2013). Solar forecasting methods for renewable energy integration. *Progress in Energy and Combustion Science*, 39(6), 535–576.
- International Energy Agency (2013). Photovoltaic and solar forecasting: State of the art. Photovoltaic Power Systems Programme Report IEA PVPS T14 01, 2013.
- Ishizaki, T., Koike, Masakazu, Ramdani, Nacim, Ueda, Yuzuru, Masuta, Taisuke, Oozeki, Takashi, et al. (2016). Interval quadratic programming for day-ahead dispatch of uncertain predicted demand. *Automatica*, 64, 163–173.
- Kang, B. O., & Tam, Kwa-Sur (2015). New and improved methods to estimate day-ahead quantity and quality of solar irradiance. *Applied Energy*, 137, 240–249.
- Kim, B.-Y., Oh, Kwang-Kyo, Moore, Kevin L., & Ahn, Hyo-Sung (2016). Distributed coordination and control of multiple photovoltaic generators for power distribution in a microgrid. *Automatica*, 73, 193–199.
- Kimura, K., & Stephenson, D. G. (1969). Solar radiation on cloudy days. *ASHRAE Transactions*, 75(1), 1–8.
- Malvoni, M., De Giorgi, M. G., & Congedo, P. M. (2016). Data on photovoltaic power forecasting models for mediterranean climate. *Data in brief*, 7, 1639–1642.
- Meinel, A. B., & Meinel, Marjorie P. (1976). *Applied solar energy. An introduction*. Reading, MA: Addison-Wesley Publishing Co.
- Patel, M. (2006). *Wind and solar power systems: Design, analysis and operation*. Boca Raton, FL: Taylor and Francis.
- Pepe, D., Bianchini, Gianni, & Vicino, Antonio (2016). Model estimation of photovoltaic power generation using partial information. In *Proc. IEEE EnergyCon 2016*. IEEE.
- Pepe, D., Bianchini, Gianni, & Vicino, Antonio (2017). Model estimation for solar generation forecasting using cloud cover data. *Solar Energy*, 157, 1032–1046.
- Pepe, D., Bianchini, Gianni, & Vicino, Antonio (2018). Estimating PV forecasting model from power data. In *Proc. IEEE EnergyCon 2018*. IEEE.
- Perez, R, Kivalov, Sergey, Schlemmer, James, Hemker, Karl, Renné, David, & Hoff, Thomas E (2010). Validation of short and medium term operational solar radiation forecasts in the US. *Solar Energy*, 84(12), 2161–2172.
- Ragnacci, A., Pastorelli, M., Valigi, P., & Ricci, E. (2012). Exploiting dimensionality reduction techniques for photovoltaic power forecasting. In *Proc. 2nd IEEE EnergyCon*.
- Reikard, G. (2009). Predicting solar radiation at high resolutions: A comparison of time series forecasts. *Solar Energy*, 83, 342–349.
- Reno, M. J., & Hansen, Clifford W. (2016). Identification of periods of clear sky irradiance in time series of GHI measurements. *Renewable Energy*, 90, 520–531.

- Schiffer, J., Zonetti, Daniele, Ortega, Romeo, Stankovic, Aleksandar M., Sezi, Tevfik, & Raisch, Jörg (2016). A survey on modeling of microgrids - From fundamental physics to phasors and voltage sources. *Automatica*, 74, 135–150.
- Spena, A., D'Angiolini, Giulia, & Strati, Cecilia (2010). First correlations for solar radiation on cloudy days in Italy. In *ASME-ATI-UIT 2010 conference on thermal and environmental issues in energy systems*.
- Tao, C., Shanxu, D., & Changsong, C. (2012). 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* (pp. 773–777).
- Widiss, R., & Porter, K. (2014). A review of variable generation forecasting in the west. National Renewable Energy Laboratory Report.
- Wong, L. T., & Chow, W. K. (2001). Solar radiation model. *Applied Energy*, 69, 191–224.
- Wu, J., & Chan, C. K. (2011). Prediction of hourly solar radiation using a novel hybrid model. *Solar Energy*, 85, 808–817.



Daniele Pepe received the BS degree cum laude in Automation Engineering in 2012, the MS degree cum laude in Computer and Automation Engineering in 2015, and the PhD in Information Engineering and Science in 2019, all from the University of Siena. He currently works in industry. His research interests include renewable energy forecasting and optimal energy management in buildings.



Antonio Vicino is Professor Control Systems the Dept. of Information Engineering and Mathematical Sciences, University of Siena (Italy), where he covered the position of Head of the Department from 1996 to 1999 and Dean of the Engineering Faculty from 1999 to 2005. In 2000 he founded the Center for Complex Systems Studies of the University of Siena, where he covered the position of Director from 1999 to 2012. Since 2016 he is the Director of the PhD Program in Information Engineering and Science at the same university. Since May 2019 he is President of the Italian University

Council. He is the author of 350 technical publications; co-author of the book *Homogeneous Polynomial Forms for Robustness Analysis of Uncertain Systems* (Springer-Verlag, 2009); co-editor of several books and guest editor of several special issues on robustness in identification and control. He has worked on stability analysis of nonlinear systems, time series analysis and prediction, robust control of uncertain systems. His present interests include identification of nonlinear systems, energy systems and smart grids, systems biology and applied system modeling.

He has served as Associate Editor or Associate Editor at Large for several journals like *IEEE Transactions on Automatic Control*, *IEEE Transactions on Circuits and Systems II* and *Automatica*. He is Fellow of the IEEE and of the IFAC.



Gianni Bianchini received the MS degree in electronic engineering from the University of Firenze, Italy, in 1997, and the PhD degree in control systems engineering from the University of Bologna, Italy, in 2001. In 2000, he was a visiting student at the Center for Control Engineering and Computation, University of California at Santa Barbara. In 2001–2002, he was a research associate at the University of Siena, and since 2003 he is an assistant professor of control systems at the same University. He is interested in energy systems, robust and nonlinear control, analysis and control of

hybrid systems, aerospace applications, and haptics. He is associate editor of the *European Journal of Control* and past associate editor of *IEEE Transactions on Circuits and Systems-II*.