Solar Energy 166 (2018) 116–122

Contents lists available at ScienceDirect

Solar Energy

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

Data-driven estimation of expected photovoltaic generation

R. Amaro e Silva*, J. Monteiro Baptista, M. C. Brito

Instituto Dom Luiz, Faculdade de Ciências, Universidade de Lisboa, Portugal

ARTICLE INFO

Keywords:
PV estimation
PV performance assessment
Data-driven approach

ABSTRACT

Photovoltaic (PV) systems' monitoring and performance assessment are relevant tools to ensure its correct operation. The standard methodology consists in determining irradiance in the plane of the array and convert it into PV power, considering system efficiency. This work proposes an effective data-driven approach based on past PV generation of the system being assessed and local measurements of solar radiation. Using local global horizontal irradiance (GHI) measurements, model performance degrades with the difference in tilt between the PV module and the pyranometer; when radiation is measured in the plane of the array (in this case for a horizontal module) the data driven method is more accurate than the standard approach. The method was also tested for a small-scale residential PV system context, replacing the local irradiation measurements by satellite GHI estimates. Although errors increase significantly, the data-driven method outperforms the standard approach for all tilts.

1. Introduction

Photovoltaic (PV) technology is becoming cost-competitive in many locations. Large scale deployment of PV systems heightens the need for accurate assessment of their performance, for regular monitoring and early detection of malfunctions, hence maximizing its power output.

The performance of a PV system is affected by several factors including conversion losses at the module, cables and inverter, module temperature, shading, soiling, or inverter malfunction. The most common criteria for the assessment of PV systems' performance are the final yield (Y_f) and the performance ratio (PR) parameters (Khalid et al., 2016; Marion et al., 2005). The final yield gives an estimate of the number of hours a system is expected to work at its rated power Eq. (1). However, as it does not integrate the actual incident irradiance, this value can vary greatly, depending on the location and the system orientation/tilt.

$$Y_f = \frac{PV_{effective}}{PV_{rated\ power}}\left[\frac{\text{kWh}}{\text{kWp}}\right] \quad (1)$$

The performance ratio (PR) is the ratio between the effective and expected generation, and corresponds to how much of the expected generation the system generated Eq. (2).

$$PR = \frac{Y_f}{Y_r} = \frac{PV_{effective}}{PV_{expected}} = \frac{PV_{effective}}{PV\ \ rated\ power \times \frac{I_{POA}}{1000\text{W m}^{-2}}} \quad [1] \quad (2)$$

Properly managed PV plants achieve a PR up to 80–90% (Nordmann et al., 2007; SMA, 2010) as there are unavoidable losses (e.g. thermal and connection losses), with lower PR values indicating an under-performing plant.

To determine the PR using Eq. (2), the effective generation can be simply monitored using a data logger. The standard procedure to estimate the expected generation is the use of a physical model of the PV system, converting measured irradiance to PV power using conversion efficiency and area of the modules, their orientation and tilt.

Most medium or large size PV plants install irradiance sensors with the same tilt as the modules (Chouder et al., 2013; degli Uberti et al., 2010; Marion et al., 2005; Sugiura et al., 2003) but this is economically unfeasible for smaller-scale systems. Weather stations and satellites, the most common sources for irradiance data usually yield horizontal irradiance. When only horizontal data is available, a transposition model has to be used to estimate irradiance at the module's plane of array (Taylor et al., 2015), often using software tools such as PVsyst (Burgess et al., 2011).

When transposing global horizontal irradiance (GHI) to a specific plane of array, its direct (or beam, BHI) and diffuse (DHI) components need to be transposed and an additional reflected component needs to be considered Eq. (3).

$$G_{POA} = f_1(BHI) + f_2(DHI) + f_3(GHI) \quad [\text{kW/m}^2] \quad (3)$$

where G_{POA} is the global irradiance on the plane of array, f_1 and f_2 are the transposition functions for the direct component and diffuse components, respectively, and f_3 the additional reflected component. While f_1 is a matter of trigonometry, the other two depend on the (an)isotropic

* Corresponding author.
E-mail address: rasilva@fc.ul.pt (R. Amaro e Silva).

https://doi.org/10.1016/j.solener.2018.03.039
Received 18 January 2018; Received in revised form 3 March 2018; Accepted 14 March 2018

nature of the diffuse and reflected irradiance and the reflectivity of the local environment.

The earlier diffuse radiation transposition models assumed that this component was isotropic (Badescu, 2002; Koronakis, 1986; Liu and Jordan, 1960), only depending on the variation in sky-view factor due to the change in tilt (β). However, this simplistic approach tends to underestimate global irradiance, in particular for equator-oriented surfaces in clear-sky days. More refined approaches consider the anisotropy of diffuse radiation (Gueymard, 1987; Hay, 1979; Perez et al., 1990; Reindl et al., 1990). The Perez model is the most commonly used due to its simplicity, good performance and its applicability to different time scales. It estimates the diffuse irradiance at the plane of array (D_{POA}) as

$$D_{POA} = DHI\left[(1-F_1)\left(\frac{1+\cos\beta}{2}\right) + F_1\frac{a}{b} + F_2\sin\beta\right] \quad [\mathrm{kW/m^2}] \tag{4}$$

hence describing the circumsolar and horizon enhancements by means of empirically defined coefficients (F_1, F_2, a and b). The reflected component is typically assumed to be the GHI multiplied by a constant ground albedo, standardized according to the type of the ground.

These requirements for local solar radiation measurements are expensive, making it unfeasible for small-scale systems. Thus, several methodologies deal with this issue by replacing irradiance data by power records from an ensemble of PV systems.

Golnas et al. use measurements from neighbouring systems normalized by their nameplate power and weighted by their distance and correlation degree (Golnas et al., 2011). Engerer and Mills estimate the PV clear-sky profile of an individual system (Engerer and Mills, 2014) and assume that, in normal working conditions, its PV clear-sky index (K_{PV}) is similar to its neighbouring systems, independent of their rated power and tilt/orientation. Thus, the expected generation for a specific system is calculated by multiplying the clear-sky expectation by the neighbour's K_{PV}. Killinger et al. propose projecting power between differently-oriented PV systems (Killinger et al., 2016): power measurements from a different system is first converted into the corresponding G_{POA}; GHI, BHI and DHI are estimated by means of an iterative process, inverting a transposition model while coupled with a decomposition model; irradiance is then transposed to the assessed system's plane of orientation and reconverted to power. Marion and Smith use the same method but propose a different approach to convert PV power into G_{POA} (Marion and Smith, 2017). All these approaches require knowing the modules' technical details (e.g. rated power and temperature coefficient), their tilt and orientation and temperature measurements. They also depend on irradiance transposition models which, in general, underperform in cloudy days, particularly for vertical surfaces, and are significantly worse when a decomposition model is used (Gueymard, 2009; Notton et al., 2006).

Lonij et al. proposed a data-driven performance assessment methodology (Lonij et al., 2012) for residential PV systems, which relies only on historical generation records from an ensemble of neighbouring systems. Performance parameters are calculated based on both effective and clear-sky working conditions from such an ensemble allowing the detection of losses from shadowing, outages and cloud. However, this depends on the modules being fairly similar in terms of tilt and orientation. Additionally, the use of data from different PV systems raises issues concerning data ownership rights (Berdugo et al., 2011).

The method proposed in this work seeks to circumvent both the need for the PV system's technical characteristics and fetching data from different neighbouring systems. The expected generation of a PV system is estimated from historical PV generation and GHI measurements, either from local ground sensor or from satellite data.

The manuscript is organized as follows. Section 2 introduces the proposed method, along with a baseline approach for comparison; the details regarding the used data set; and the error metrics considered. Results are shown in Section 3; sensitivity analyses are presented, which motivate the inclusion of a bias-correction step; the impact of replacing local measurements by satellite estimations is also assessed. In Section Section 4 the main conclusions of this work are discussed.

2. Method

2.1. Proposed and baseline models

For solar forecasting purposes it is standard to convert either irradiance or PV power into its clear-sky index (K_G or K_{PV}, respectively), indicating the sky attenuation when compared to clear-sky conditions (CS_G or CS_{PV}) as shown in Eqs. (5) and (6) (Amaro e Silva and Brito, 2018). This is of relevant since it removes the easy to determine seasonal variation and singles out weather-induced variability.

$$K_G = \frac{G}{CS_G} \quad [1] \tag{5}$$

$$K_{PV} = \frac{PV}{CS_{PV}} \quad [1] \tag{6}$$

The method proposed herein is based on the hypothesis that, if measured at the same site, the clear-sky index estimated from either a pyranometer or a PV system should be similar. If so, a pyranometer would be a suitable sky condition estimator as it would be unbiased by potential malfunctions of a PV system. Thus, the expected PV generation ($PV_{expected}$) of a certain system in normal operation conditions could be estimated using Eq. (7). The fact that this approach is based on the relationship between K_G and K_{PV}, so forth it will be designated as *K*2 *method*.

$$PV_{expected} = CS_{PV} \times K_G \quad [\mathrm{W}] \tag{7}$$

In this work, CS_G and CS_{PV} are estimated using Lonij's model (Lonij et al., 2012). This data-driven approach assumes that the clear-sky expectation for every specific timestamp is the 85th percentile of the values measured at that same hour over the previous 15 days, as shown in Eqs. (8) and (9).

$$CS_G = perc(G, 0.85) \quad [\mathrm{W/m^2}] \tag{8}$$

$$CS_{PV} = perc(PV, 0.85) \quad [\mathrm{W}] \tag{9}$$

Since the estimation of both CS_{PV} and K_{GHI} avoid the need for any technical data (PV system power, configuration, temperature coefficient, etc.), the proposed method bypasses many of the limitations identified in the previous section: the need for a considerable set of parameters and the data ownership rights that would be implied in an approach based on generation from neighbouring PV systems.

For high temporal resolutions data, the clear-sky model may feature high frequency noise. This noise was mitigated by applying a locally weighted 2nd degree polynomial regression at each data point. Similar to a moving average approach, weights are attributed to the neighbouring data points based on how distant, time-wise, they are. The weights are defined by a tricube function and then a 2nd degree polynomial is applied based on these same weights and data points. Fig. 1 shows that a filter considering only the 10% nearest points to each data point fares better at the beginning/end of the day, while considering 50% nearest points is more appropriate for the mid-day period. Thus, the first range is applied for solar elevation angles below 15 ° and the second for higher solar elevations. Since this filter smooths out over-irradiance and/or over-generation events, the 100th percentile may now be considered in the clear-sky model.

To assess the relative performance of the model, the standard physical approach was implemented as baseline. PV generation at Standard Test Conditions ($PV_{STC,phys}$) is estimated based on GHI data (transposed, when needed, into G_{POA} using the Perez model (Perez et al., 1990) with DHI data and assuming a 0.2 ground albedo, characteristic of grassy fields (Stanhill, 1970)), and the modules' rated power: