Solar Energy 173 (2018) 1087–1106

Contents lists available at ScienceDirect

Solar Energy

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

On the search for representative characteristics of PV systems: Data collection and analysis of PV system azimuth, tilt, capacity, yield and shading

Sven Killinger[a,b], David Lingfors[c], Yves-Marie Saint-Drenan[d], Panagiotis Moraitis[e], Wilfried van Sark[e], Jamie Taylor[f], Nicholas A. Engerer[a,1], Jamie M. Bright[a,*]

[a] Fenner School of Environment and Society, The Australian National University, 2601 Canberra, Australia
[b] Fraunhofer Institute for Solar Energy Systems ISE, 79100 Freiburg, Germany
[c] Department of Engineering Sciences, Uppsala University, Lgerhyddsvgen 1, 752 37 Uppsala, Sweden
[d] MINES ParisTech, PSL Research University, O.I.E. Centre Observation, Impacts, Energy, 06904 Sophia Antipolis, France
[e] Copernicus Institute of Sustainable Development, Utrecht University, 3508 TC Utrecht, The Netherlands
[f] Sheffield Solar, University of Sheffield, Hicks Building, Hounsfield Road, Sheffield S3 7RH, UK

ARTICLE INFO

Keywords:
PV system characteristics
Metadata
Shading
Data analysis

ABSTRACT

Knowledge of PV system characteristics is needed in different regional PV modelling approaches. It is the aim of this paper to provide that knowledge by a twofold method that focuses on (1) metadata (tilt and azimuth of modules, installed capacity and specific annual yield) as well as (2) the impact of shading.

Metadata from 2,802,797 PV systems located in Europe, USA, Japan and Australia, representing a total capacity of 59 GWp (14.8% of installed capacity worldwide), is analysed. Visually striking interdependencies of the installed capacity and the geographic location to the other parameters tilt, azimuth and specific annual yield motivated a clustering on a country level and between systems sizes. For an eased future utilisation of the analysed metadata, each parameter in a cluster was approximated by a distribution function. Results show strong characteristics unique to each cluster, however, there are some commonalities across all clusters. Mean tilt values were reported in a range between 16.1° (Australia) and 35.6° (Belgium), average specific annual yield values occur between 786 kWh/kWp (Denmark) and 1426 kWh/kWp (USA South). The region with smallest median capacity was the UK (2.94 kWp) and the largest was Germany (8.96 kWp). Almost all countries had a mean azimuth angle facing the equator.

PV system shading was considered by deriving viewsheds for ≈48,000 buildings in Uppsala, Sweden (all ranges of solar angles were explored). From these viewsheds, two empirical equations were derived related to irradiance losses on roofs due to shading. The first expresses the loss of beam irradiance as a function of the solar elevation angle. The second determines the view factor as a function of the roof tilt including the impact from shading and can be used to estimate the losses of diffuse and reflected irradiance.

1. Introduction

With 402.5 GW of installed photovoltaic (PV) capacity globally (IEA, 2018), the integration of the large amounts of energy generated by the numerous distributed solar power systems into the electricity supply system is an issue ever gaining in importance. Modelling of the power generated by those decentralised solar systems is of utmost importance for several issues ranging from energy trading to network flow control. The estimation and forecast of PV power is made difficult by the fact that only a minority of systems continuously report their generation and are publicly accessible.

Different strategies have been proposed to overcome the lack of reporting (e.g. upscaling approaches or power simulations based on satellite derived irradiance); an extensive literature overview is provided in Bright et al. (2017b). Within this paper, the estimation of the aggregated power generated in a given region by a fleet of unknown PV systems is referred to as regional PV power modelling. Knowledge of PV system characteristics is required in the different regional PV modelling

* Corresponding author.
E-mail addresses: nicholas.engerer@anu.edu.au (N.A. Engerer), jamie.bright@anu.edu.au (J.M. Bright).
[1] Co-corresponding author.

https://doi.org/10.1016/j.solener.2018.08.051
Received 24 April 2018; Received in revised form 8 August 2018; Accepted 19 August 2018

approaches to reconstruct the missing power measurements (Lorenz et al., 2011; Saint-Drenan et al., 2016). Some studies assign simplified assumptions of the PV system characteristics. This can result in over-exaggerated grid impacts (Bright et al., 2017a). Unfortunately in most cases, characteristics from PV systems are either unknown or only accessible for a small number of stakeholders (inverter manufacturers, monitoring solutions providers, etc.). As a result, progress in the area of regional PV power estimation or forecasting can be considered sub-optimal as potential contributors like universities or small companies are partially excluded from access to larger datasets of measurements or metadata. This is still the case despite grid integration of solar energy being considered a strategic societal issue. Therefore, it is the aim of this paper to offer any stakeholders the possibility to develop activities on this research field by collecting, analysing and disseminating metadata on millions of PV systems installed worldwide. To begin, we must establish which metadata are the most important.

Saint-Drenan (2015) carried out a sensitivity analysis and found that the four most influential characteristics impacting PV output generation are: (1) tilt angle and (2) azimuth angle of PV modules, (3) installed capacity and (4) total efficiency (represented herein as the specific annual yield). Furthermore, (5) shading is of crucial influence on the PV power generation but is not accessible from PV system metadata. The impact of shading can only be accessed with considerable effort, e.g. simulations that consider digital elevation models (DEM) including buildings, trees and other obstacles, by analysing PV power profiles or even weekly performance ratios (see Paulescu et al. (2012), Freitas et al. (2015), Lingfors et al. (2018), Tsafarakis et al. (2017) for further reading). Due to its significant influence, a shading analysis complements the focus of this study.

These five identified characteristics are the central focus of this paper because of their general importance for regional PV modelling approaches. The overall aim of this paper is to achieve a full reproducibility of the five characteristics so that they can be used in regional PV power modelling applications such as nowcasting or forecasting, but also in power simulations that are used for energy system analysis, studying the grid impact, defining the PV power potential, etc.

1.1. Related work

The relevant literature for this research has three prominent categories: (1) metadata analysis with intention to improve regional PV power simulations, (2) PV performance due to specific yield, and (3) models that consider shading analysis.

Category 1: Examples of literature using metadata to improve regional PV power simulations. Schubert (2012) provides a useful guidebook for the simulation of PV power that sketches important parts of the simulation chain and delivering assumptions for characteristics. An overview of different characteristics of tilt, azimuth, the module and installation type are given together with suggested weights. However, these weights seem to be assumptions with no datasets being cited as an empirical basis and so using these weights in PV simulations raise questions of trust.

Datasets are used by Lorenz et al. (2011), who evaluated the representativeness of a set of reference PV systems to predict regional PV power by analysing the orientation and module types of ≈ 8000 systems in Germany. The authors note that their dataset seem to have a disproportionate share of large PV systems and so do not fully represent a larger portfolio.

The problem of poor representativeness was bypassed in Saint-Drenan (2015) and Saint-Drenan et al. (2017) by feeding a PV model with metadata statistics from a larger sample of PV systems as opposed to a smaller and unrepresentative subset. They derived joint probabilities of azimuth and tilt from 35,000 systems and clustered them by their system size and geographic location. These empiric distributions where then used to estimate the characteristics of all 1,500,000 PV systems installed in Germany at that time. Saint-Drenan et al. (2018) complemented their earlier research by reproducing it for more European countries using statistical distributions from 35,000 PV systems in Germany and 20,000 in France. This demonstrates the significant potential of generating representative statistical distributions with intended use in regional PV power simulations.

Kühnert (2016, pp. 80-85) followed a similar approach and derived statistical distributions for tilt and azimuth from ≈1300 PV systems in Germany. Based on this portfolio, the author evaluated the representativeness should PV systems be clustered into different geographic regions and system sizes. The authors quantitatively derived recommendations between the two extremes of (1) a portfolio covering all PV systems and (2) a high number of subclasses with a very small number of PV systems. From this, we observe that there must be a well considered clustering approach in order to derive representative subclasses.

Killinger et al. (2017c) detailed a regional PV power upscaling approach which estimated the power of ≈2000 target PV systems based on 45 continuously measured PV systems in Freiburg, Germany. Whereas the azimuth and tilt of the 45 measured systems were known in their case, both parameters were derived through a geographic information system (GIS) based approach for the target PV systems.

Furthermore, Pfenninger and Staffell (2016) use PV power measurements and incorporate metadata from 1029 systems in 25 European countries to derive empirical correction factors for PV power simulations. A comparison between the analysed tilt and latitude showed an trend towards steeper angles at higher latitudes, indicating that metadata might vary with the geographic location.

PV system metadata is thus used to successfully improve regional PV power upscaling across Europe in Pfenninger and Staffell (2016), Killinger et al. (2017c), Saint-Drenan (2015), Saint-Drenan et al. (2017, 2018) and Kühnert (2016). These works applied information of azimuth, tilt, installed capacity and the geographic location from PV systems to estimate the power output of a larger PV fleet for similar geographies and different countries. They stand as an powerful and excellent example for how representative metadata distribution statistics can be employed. It is these examples that guide the first usage of our vast dataset towards deriving representative metadata distributions.

Category 2: Excerpts of literature that analyse the performance of PV systems. Performance is more complex than just tilt and azimuth as it is inherently influenced by other components, such as soiling and meteorology.

Nordmann et al. (2014) found a positive correlation between specific annual yield and incoming irradiance, as well as an observed negative correlation between system performance and ambient temperature. Their data was obtained via web-scraping of Solar-Log (2914 systems in the Netherlands, Germany, Belgium, France and Italy) and collected by participants of the IEA task (>60,000 systems in the USA).

Moraitis et al. (2015) observed an increasing yield with decreasing latitude from ≈20,000 systems in Netherlands, Germany, Belgium, France and Italy, also achieved using web-scraping techniques. We therefore expect to observe geographical differences due to latitude and climate.

Taylor (2015) explored the generation of 4369 distributed systems in the UK to derive the performance ratio and degradation rate. To allow reproducibility, the analysis of the performance ratio was enriched by approximating it with distribution functions. We intend to extend this style of analysis to PV system metadata.

Leloux et al. (2012a) examined data from residential PV systems in Belgium; Leloux et al. (2012b) focused on France. In Belgium, specific annual yield was analysed for 158 systems in 2009 and normalised by a factor which compared the incoming irradiance in this year to a 10 year average. The mean value was 836 kWh/kWp. The same approach led to a mean value of 1163 kWh/kWp for 1635 systems in 2010 in France. Weibull distributions were used throughout both papers to approximate the specific yield and performance indicators; Weibull distributions