High-frequency validation of predictions for energy production from large solar PV plants

Stefania Castelletto\textsuperscript{1} and Alberto Boretti\textsuperscript{2,*}

\textsuperscript{1}Mechanical and Automotive Engineering, School of Engineering, RMIT University, 3083 Bundoora, Victoria, Australia
\textsuperscript{2}Prince Mohammad Bin Fahd University, P.O. Box 1664, Al Khobar 31952, Kingdom of Saudi Arabia

\textsuperscript{*}Email: a.a.boretti@gmail.com

Abstract. The solar energy source is characterized by variability, intermittency, and unpredictability. The annual average direct normal irradiance and global horizontal irradiance only gives a rough estimation of the amount of energy that can be harvested by a solar photovoltaic plant. It is however also relevant to quantify the variability of the energy production, which is not only the result of alternating daylight and nighttime but also to accept variability from one day to another due to weather and seasonality. A design that permits a more stable controllable output is preferable because it reduces the need for external energy storage by batteries. Here we quantify the coefficient of variability of existing photovoltaic plants, and compare it with the coefficient of variability of the resource, to validate solar PV plant models.

1. Introduction

The estimation of power production yield by large solar photovoltaic (PV) plants is currently crucial for investment decisions and policymakers relying on technical and economic feasibility design analyses. The main reason is to tackle the well-known nature of renewable resources unpredictability together with the learning curve achieved by current installations performance. The aim is to properly estimate the power production to suitably balancing the generation and demand at the grid level.

Many methods \cite{1} to \cite{4} have been applied to forecast renewables energy production. These methods are based on physical or phenomenological, statistical, and machine learning approaches. They are based in some cases on sparse empirical analysis and are applied to small renewable energy systems. Many software tools based on the above methods are available to provide prediction and forecast of energy production and related Levelized cost of electricity (LCOE). Estimations deliver an apparent high level of details, as predictions are provided to hourly based power production. Most of the work so far has focussed on the resource’s accuracy and its forecasting, as the main cause of possible departure from measured values.

Given the available information of measured energy production of power plants in some countries, such for example the U.S. and Australia \cite{5} to \cite{7}, it is here proposed to use this electricity production data to validate the predictive models of large scale solar PV plants level up to the higher sampling frequency. So far similar approaches were applied to very small-scale systems \cite{8}, \cite{9}, providing limited confidence in their extension to large-scale systems.

Here we compare the measured energy production of solar PV energy facilities of over 50 MW capacity in Australia with the values computed by the (US) National Renewable Energy Laboratory (NREL) System Advisor Model (SAM) physical model \cite{10}.

The contribution highlights the issue of the validation of renewable energy tools, in need of proper high-frequency (hour level or less) simultaneous weather and plant data. Only these operating data at
the system and individual components level coupled to the weather permits the validation. Only validated tools can be used for reliable performance estimation and forecasting of energy production and related LCOE.

2. **PV Solar Energy**

Power generation from solar PV is estimated to have increased by more than 30% in 2018, to over 580 TWh and 720 TWh in 2019. With this increase, the solar PV share in global electricity generation exceeded 2% for the first time. It remains the fourth-largest renewable electricity technology in terms of generation, after hydropower, onshore wind, and bioenergy [11]. PV installation in 2019 was 204.7 GW in China, 131.7 GW in the European Union, 75.9 GW in the U.S. [12].

Within Australia, solar generation is growing significantly particularly regarding PV power plants and rooftop residential installations. The total PV installation accounts for 13.9 GW, including 7.6 GW residential installations under 10 kW, 1.9 GW for medium 10-100 kW installation, and 4.4 GW for installation over 100 kW [13]. There is no large solar thermal grid-connected power plant in Australia, only PV.

Solar energy facilities connected to the Australian National Electricity Market (NEM) grid, include the solar energy facilities of New South Wales (registered power 1,226 MW), Queensland (registered power 1,798 MW), South Australia (registered power 378 MW), and Victoria (registered power 430 MW), for a total registered power of 3,832 MW [5], [6]. Tasmania has no solar power generation facility.

The NEM grid stretches for about 5,000 km from Port Lincoln, South Australia, to Port Douglas, Queensland, also serving Tasmania through the Bass Strait. The NEM grid, managed by the Australian Energy Market Operator (AEMO) is the world’s largest interconnected power system.

The correlation between the performances of the different solar facilities is significant, only complicated by the changes in the weather patterns, quite different moving from the southern state of Victoria and South Australia to New South Wales, and then to Queensland in the North.

It is worth mentioning that the single concentrated solar power energy facility connected to the NEM grid is a small 10 MW unit, with the rest being solar PV, flat panel, or tracking flat panel. Large-scale solar generated 1,875 GWh of electricity in 2018, accounting for 0.8% of overall electricity generation [15] in Australia. The installed capacity of 2018 was 1,824 MW.

The solar energy production correlates to the variable resource [15], [16], [17]. However, limited information is available about the actual electricity production of solar energy facilities sampled with high frequency. The capacity factor of a solar energy facility oscillates from zero to close to 100% every day, following a sequence of daylight and night. Additionally, the production is affected by the weather conditions of the specific day, the seasonality, and inter-annual and multi-decadal variability. This variability is not appropriately investigated in the literature.

Although electricity demand and solar generation have some similarities [16], [17], more often the two are completely uncorrelated. Thus, huge energy storage is needed to compensate for the fact that solar energy supply by PV is completely unavailable on average 12 hours a day, and in the remaining 12 hours, it still suffers from significant variability.

The statistic of power vs time of solar energy facilities is often limited to relatively large sampling intervals, [18] for example, monthly average values of power. These averages over a long period say nothing about the integration of the specific facility into a stable grid where demand and supply must be balanced at any time, and not over a month or a year.

A high-frequency sampling, every minute and even less, is needed to properly understand the issue of variable, unpredictable, and intermittent renewable energy integration into a grid [14]. These data are hard to be sourced.

Only the Australian Energy Market Operator (AEMO) makes public the energy production data of the different energy facilities connected to the grid with a frequency of 3 minutes. This is simply because the AEMO is an open market, and the contracts to buy and sell electricity are public domain. In every other part of the world, electricity production from renewables is almost always covered by high secrecy.
3. PVout Model

Simplified models are often adopted to evaluate the annual electricity production for a specific location. One of these models is PVout. For a selected location, the possible electricity production from a solar PV system can be calculated by following a few simple steps, with results made available from online facilities such as Ref.[19].

The global irradiation descending on a tilted PV solar panel is calculated from both the Global Horizontal Irradiance (GHI) and the Direct Normal Irradiance (DNI). The GHI is the total amount of shortwave radiation received from above by horizontal surfaces. This value is of particular interest to PV and includes both DNI and Diffuse Horizontal Irradiance (DIF or DHI). DNI is the radiation coming straight from the sun. DIF is the radiation scattered in the atmosphere and coming from all other directions.

The albedo of the terrain and the position of the sun affect the irradiation received by the solar panel. The model of [19] accounts for tilted surfaces diffuse irradiance and reflection losses. Terrain shading is computed from elevation data. Shading from buildings, vegetation, and other obstacles is ignored.

The solar panel performance is approximated through a single-diode equivalent circuit by using the De Soto five-parameter model [12]. A typical crystalline-silicon solar panel is considered. The temperature of the solar panel is included in the model, computed from the module temperature calculated from air temperature, irradiance, and typical panel thermal characteristic. Inter-row shading of solar panels in arrays is simply approximated. Higher losses are experienced during seasons with low sun elevation angles and considered negligible otherwise. The Sandia Inverter Model is used to calculate DC to AC conversion losses [12]. The other existing losses due to many other factors, including environmental agents and cleaning of the surface of the panels, mismatch or power transportation losses, are also included simply.

Fig. 1 presents the DNI, GHI, and the PV power potential across Australia (from [19]). This is a very rough estimation of the annual average solar energy production. This estimation provides an idea of the amount of electricity production, but not of the quality of this electricity. Electricity supply is more valuable when matching the grid demand, and the more it is stable.
4. Other Models

Many other methods have been developed to forecast renewables energy production. For example, in [1], a phenomenological model (deterministic) proposed by Sandia National Laboratories and two statistical learning models are applied to a small renewable energy system (1 kW) connected to a microgrid. Previously a deterministic approach was used to model electrical, thermal, and optical characteristics of PV modules, by using the hourly solar resource and typical meteorological year (TMY) data, to validate different module types or small systems [20].

These models are usually implemented in commercial or freely accessible software such for example PVSyst [21], or SAM [10]. These software tools based on the above methods are available to provide predictions of energy production and related LCOE at an apparent high-frequency time series to hourly based power production estimates. SAM has been validated for commercial rooftop solar PV systems and small-size PV utility plants up to 10 MW [8].

Other solar simulations have been proposed for concentrated solar power (CSP) and not PV. CSP only uses the DNI and not the GHI, as the DIF is lost, and it suffers much more than PV from cloud coverage and transients. Despite CSP are much less widespread than PV, there are in the literature more comparisons of experiments with simulations for CSP rather than PV power plants.

Other validations based on electricity production using SAM simulation modeling of a CSP plant of 50 kW was proposed in [2] with an agreement on few hours in specific days in the range from 3.6 to 13%. SAM energy production predictions have been compared to actual production data for CSP large facilities such as Genesis, Mojave, and Solana, and the 110 MW solar tower facility of Crescent Dunes, all in the United States South-West [3]. These comparisons showed a relatively good agreement albeit the not optimal sun radiation resources referred to the typical year and not the specific year.

Most of the work so far has been focused on the resource’s accuracy and its forecasting, as the main cause of possible departure from measured production values. However, from the correct resource, not necessarily the model returns the correct energy production.

Simply interpolated time series irradiance data, such as the ones associated with typical meteorological years, can lead to significant errors in studies examining renewable energy systems production.

In [4] a database of measured PV panel outputs is used to validate the power output from PV simulations using reanalysis and satellite datasets. However, even more, correct irradiance measurements may be insufficient to accurately model the output from real plants. This is because models do not cover everything, and do not perfectly represent every phenomenon. Other effects such as temperature derating and panel shading, as an example, can influence the production. The exact
configuration and locations of the PV arrays in a plant are often not known. Validation against the measured production data is of paramount importance before making any use of the predictions.

5. Present Simulations
Here, we compare the energy production of large PV facilities in Australia over a full year sampled every 5 minutes. Simulated energy production is obtained by using SAM PV [10] detailed single owner utility model. Comprehensive information about the method is provided in [10], plus [8], [9], [18], [20], [22] to [24].

The measured hourly capacity factors variation on a typical day for the 12 different months of the year, obtained by 5 minutes sampling data, are compared to the simulated hourly capacity factors for a typical day of each month, obtained from hourly simulated data. Values of measured and simulated capacity factors are also compared as daily, weekly, and monthly average.

This work evidences the need for modeling energy production using high-frequency data from the solar resources to better validate the energy forecast. The common practice to compare annual measured energy production and modeling results, or at the best monthly values, is not a satisfactory validation exercise. This approach cannot provide accurate modeling for real-time design of energy storage for grid balancing due to the extremely large variability of the capacity factors during a single day.

The validation exercise of the NREL SAM software and the supporting resource database is carried out for three large PV solar energy facilities in Australia, namely Broken Hill, Moree, and Nyngan Solar Farms (Table 1).

| Table 1. PV Solar facilities |
|-----------------------------|
| **Broken Hills**             |
| Latitude | 31.983 S |
| Longitude | 141.467 E |
| Year Completed | 2014 |
| Technology | ~650,000 flat-panel PV thin-film Cd Te-Fixed at 25-degree tilt |
| Nameplate capacity [MW] | 54.5 |
| Site area [km²] | 1.4 |
| Cost [m$] | 150 |
| Specific cost [$/kW] | 2752 |
| **Nyngan** |
| Latitude | 31.566° S |
| Longitude | 147.072° E |
| Year Completed | 2014 |
| Technology | ~1,350,000 flat-panel PV thin-film Cd Te, fixed tilt |
| Nameplate capacity [MW] | 102 |
| Site area [km²] | 2.5 |
| Cost [m$] | 440 |
| Specific cost [$/kW] | 4314 |
| **Moree** |
| Latitude | 29.467 S |
| Longitude | 149.85 E |
| Year Completed | 2015 |
| Technology | ~222,880 polycrystalline modules & a single-axis horizontal tracking |
| Nameplate capacity [MW] | 57 |
| Site area [km²] | |
| Cost [m$] | 201 |
| Specific cost [$/kW] | 3526 |
The first facility is Broken Hill Solar Plant in New South Wales, a flat panel facility with a registered capacity of 54 MW. The construction works began in 2013 and operations commenced in 2015. It has ~650,000 flat-panel PV thin-film Cd Te fixed at a 25-degree tilt.

The second facility is Nyngan Solar Plant also in New South Wales, a PV flat panel, of registered capacity 102 MW. It has ~1,350,000 flat-panel PV thin-film Cd Te, fixed tilt.

The third facility is Moree Solar Farm in New South Wales, a single-axis horizontal tracking flat-panel facility, with a registered capacity of 57 MW and completed in 2016. Construction began in 2015 and operations commenced in 2016. The facility uses solar PV modules and a single-axis horizontal tracking system. It has ~222,880 polycrystalline modules & single-axis horizontal tracking.

In Fig. 2 the three locations solar resources used for the simulations are shown.

![Graphs showing solar radiation data for Broken Hill, Nyngan, and Moree](image)

**Figure 2.** continued. Input GHI and DNI irradiance for the three locations based on weather data.

Solar data resources are hourly values of the DNI, GHI, and DHI/DIF of the typical year. This is the average between the measurements performed between 2004 and 2018 from [25]. Other relevant weather parameters include wind speed and direction, and air temperature and pressure.

Data for generating power versus time and registered capacity for the different solar PV energy facilities are obtained from [5]. The sampling frequency is 5 minutes, for a total of 105,120 samples in 1 year (2018).
6. Results

Simulations are performed by using SAM [10] with the number and type of modules specifications provided based on Table 1. Tilt conditions and tracking is defined based on the information available. The simulated energy produced is determined hourly and the capacity factor (CF) is averaged hourly, over a typical day of each month, then daily and monthly.

Table 2 presents the mean and standard deviation (hourly frequency) of the measured and computed capacity factors for the three solar PV facilities.

| Facility     | Measure | Mean CF | Std Dev | CV  |
|--------------|---------|---------|---------|-----|
| Broken Hill  | Exp     | 0.27    | 0.36    | 1.34|
|              | Theor   | 0.29    | 0.38    | 1.30|
| Nyngann      | Exp     | 0.28    | 0.37    | 1.33|
|              | Theor   | 0.29    | 0.39    | 1.32|
| Moree        | Exp     | 0.31    | 0.37    | 1.23|
|              | Theor   | 0.32    | 0.39    | 1.23|

The data are averaged for hourly, weekly, and monthly capacity factors as shown in Fig. 2 to 7. Fig. 3 presents, from top to bottom, the hourly measured CFs compared to the simulated ones for the three facilities of Broken Hills, Nyngan, and Moree.

While the typical year resource is about the same in the three locations, the operation of the three facilities is slightly different. The typical year resource is not specific, thus the differences between computations and measurements can be attributed to a mismatch between the resource of the typical and actual year, or limitations of the model.

Fig. 4 presents the scatter plot of the hourly CF computed and measured for Nyngan. While $R^2$ is relatively large ($R^2=0.9123$), there are however significant differences. This is mostly due to the typical year not representing the specific year, as the weather is not the same on the same days of different years.
This is more evident in Fig. 5 that presents from the top to bottom the daily measured CFs of the three facilities of Broken Hills, Nyngan, and Moree compared to the computed ones.

**Figure 5.** Top to bottom daily measured CFs of the three facilities of Broken Hills, Nyngan, and Moree compared to the simulated ones.

Fig. 6 presents from the top to bottom the weekly measured CFs of the three facilities of Broken Hills, Nyngan, and Moree compared to the computed ones. The weakly average suffers much less of variability. Averaging over a month, Fig. 7, differences further reduce.

**Figure 6.** Continues. Top to bottom Weekly measured CFs of the three facilities of Broken Hills, Nyngan, and Moree compared to the simulated ones.
Fig. 7. Top to bottom, left to right, monthly measured energy production (blue) of the three facilities of Broken Hills, Nyngan, and Moree compared to the simulated ones.

Fig. 8, 9, and 10 finally presents for Broken Hills, Nyngan, and Moree, the computed operation over the typical day of every month of the typical year, and the averaged day of every month of 2018. Differences are relatively small, but not negligible, especially in the winter and spring months. They are different from one facility from the other.

Apart from June, October, and November, differences are quite small in Broken Hills. Differences are also larger in October and November in Nyngan. Larger differences are generally found in the same and other months in Moree. Moree is the best operating facility, with a 10% larger mean capacity factor (0.31 vs. 0.27-0.28).

Fig. 8. Daily variation of the measured CF (blue) for Broken Hill compared to the simulated ones.
Figure 9. Daily variation of the measured CF (blue) for Nyngan compared to the simulated ones.

Figure 10. Daily variation of the measured CF (blue) for Moree compared to the simulated ones.

7. Conclusions
The present simulations of large solar PV facilities by using software tools such as SAM suffers from two uncertainties.

First of all, the solar irradiance and other weather conditions for the typical year are not an accurate representation of the solar irradiance and weather conditions of the specific year. There is significant variability between one day and another in a given year, and interannual and multidecadal variability is also important.

After model tuning, more than model validation, both the mean and the standard deviation with the hourly frequency of the capacity factors are well predicted. However, the specific output at a specific time cannot be predicted with great accuracy.

The modeling of a specific plant operation is everything but perfect.
Only having a simultaneous measurement of the solar resource, weather conditions, and detailed plant operating parameters, at the system as well as the component level, it would be possible to make the models better and properly validate them.

References

[1] Graditi G, Ferlito S and Adinolfi G 2016 Renewable Energy 90 513-9
[2] Ezeanya EK, Massiha GH, Simon WE, Raush JR and Chambers TL 2018 Cogent Engineering 5(1) 1524051
[3] Boretti A, Nayfeh J and Al-Kouz W 2020 Energies 13(8) p.1949
[4] Pfenninger S and Staffell 1 2016 Energy 114 1251-1265
[5] Australian Energy Market Energy data from the Australian Energy Market Operator anero.id/energy/ (accessed December 31, 2020)
[6] Solar Energy Solar power in the Australian Energy Market. anero.id/energy/solar-energy (accessed December 31, 2020)
[7] U.S. Energy Information Administration, Electricity data browser. www.eia.gov/electricity/data/browser/ (accessed December 31, 2020)
[8] Freeman J, Whitmore J, Kaffine L, Blair N and Dobos AP 2013 System advisor model: Flat plate photovoltaic performance modeling validation report (NREL, Golden, CO)
[9] Gilman P 2015 SAM Photovoltaic Model Technical Reference (NREL, Golden, CO)
[10] NREL System Advisor Model sam.nrel.gov/ (accessed December 31, 2020)
[11] Solar PV power generation in the Sustainable Development Scenario, 2000-2030. www.iea.org/data-and-statistics/charts/solar-pv-power-generation-in-the-sustainable-development-scenario-2000-2030. (accessed December 31, 2020)
[12] Renewable Capacity Statistics 2020. irena.org/publications/2020/Mar/Renewable-Capacity-Statistics-2020 (accessed December 31, 2020)
[13] Australian PV institute, Solar PV Maps and Tools pv-map.apvi.org.au/ (accessed December 31, 2020)
[14] Boretti A 2020 Energy Storage 2(1) p.e92
[15] BP Statistical Review of World Energy 2019 | 68th edition. www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2019-full-report.pdf (accessed December 31, 2020)
[16] Boretti A 2018 Renewable Energy Focus 26 93-99
[17] Boretti A 2020 Energy Storage doi.org/10.1002/est2.206
[18] King DL, Gonzalez S, Galbraith GM and Boyson WE 2007 Performance model for grid-connected photovoltaic inverters. (Sandia National Laboratories, Albuquerque, NM).
[19] SolarGIS, Global solar atlas globalsolaratlas.info/ (accessed December 31, 2020)
[20] King DL, Kratochvil JA and Boyson WE 2003 Photovoltaic array performance model. (Sandia National Laboratories, Albuquerque, NM).
[21] PVsyst, Photovoltaic software www.pvsyst.com/ (accessed December 31, 2020)
[22] Gilman P, et al. 2018 SAM Photovoltaic Model Technical Reference Update. (NREL, Golden, CO).
[23] Marion W and Dobos A 2013 Rotation Angle for the Optimum Tracking of One-Axis Trackers. (NREL, Golden, CO).
[24] MacAlpine S and Deline C 2015 Simplified Method for Modeling the Impact of Arbitrary Partial Shading Conditions on PV Array Performance. (NREL, Golden, CO).
[25] Climate.OneBuilding.Org, Repository of free climate data for building performance simulation. climate.onebuilding.org/ (accessed December 31, 2020)