Solar PV Maps for Estimation and Forecasting of Distributed Solar Generation

Julian de Hoog [1] **Maneesha Perera** [1] **Kasun Bandara** [2] **Damith Senanayake** [1] **Saman Halgamuge** [1]

Abstract

Rapid uptake of distributed solar PV is starting to make the operation of grids and energy markets more challenging, and better methods are needed for measuring and forecasting distributed solar PV generation across entire regions. We propose a method for converting time series data from a number of point sources (power measurements at individual sites) into 2-dimensional maps that estimate total solar PV generation across large areas. These maps may be used on their own, or in conjunction with additional data sources (such as satellite imagery) in a deep learning framework that enables improved regional solar PV estimation and forecasting. We provide some early validation and results, discuss anticipated benefits of this approach, and argue that this method has the potential to further enable significant uptake of solar PV, assisting a shift away from fossil fuel-based generation.

1. Introduction

Global installation of solar photovoltaic (PV) generation continues to accelerate, and solar PV is now the fastest growing form of electricity generation worldwide [5]. However, while this increasing uptake of solar is a welcome development, high levels of solar PV can lead to complex issues for our existing electrical grids.

An early example for this is the state of Western Australia, which has one of the highest rates of solar uptake in the world, with some neighbourhoods having solar on more than 50% of homes [1]. The national energy market operator recently reported extensively on the issues that all this distributed rooftop solar PV is starting to cause [17]:

Grid instability: many distribution feeders are now feeding back into the grid for extended periods of the day, leading to problems with voltage control and frequency stability.
System volatility: when cloud bands pass over neighbourhoods having high levels of solar PV, there can be significant fluctuations in generation that must be handled by short-term reserves and ancillary services.
New contingency risks: during disturbances, inverters may trip for short periods, which at scale can be equivalent to a major generator temporarily going offline.
System restart: after blackouts, a certain level of load is required to re-energise the system, and this can be difficult to achieve when many neighbourhoods are in fact behaving like generators.
Market volatility: high levels of distributed solar PV (which affect net demand) can be difficult for market operators to forecast, and have made markets more volatile.

The above list is not exhaustive and many other issues exist [3, 9, 10]. These issues are already starting to have a major impact on system and market operation in Australia and other markets where uptake is high. Given the rates of uptake across many other geographies, it is likely that these problems will start to emerge in many parts of the world in the coming 5-10 years. If they are not addressed, there is a risk that uptake of solar PV will be slowed, generation levels of existing systems may be curtailed, and the full potential for reducing greenhouse gas emissions may not be realised.

Fortunately, many of these issues can be addressed by emerging technologies and tools, such as energy storage and improved solar PV estimation and forecasting. There has been extensive work on forecasting solar output of individual systems, and forecasting approaches can range from numerical weather prediction [8], to on-site sky imaging systems [7], to the use of satellite imagery [6, 13], to standard statistical forecasting tools [14], to (more recently) the use of machine learning to develop probabilistic forecasts [19].

However, what is urgently required now is a better ability to forecast distributed solar PV generation for entire regions, such as a whole distribution feeder, or an entire market zone. To date this has been difficult to do: there is no region in the world that has direct access to metering and sensing of all distributed solar PV systems; in many locations, the actual number and generating capacities of all systems may not even be known; and there is no ground truth against which competing approaches may be evaluated. However, an ability to estimate and forecast solar PV generation across entire regions is becoming increasingly essential, since this would allow system operators to better handle and plan for many of the previously mentioned issues that distributed solar PV is starting to introduce.

In this paper we present a proposal for a new approach to estimation and forecasting of solar PV generation across entire regions. We describe a method for generating solar PV "maps" for large regions using only a small set of metered

[1]Department of Mechanical Engineering, School of Electrical, Mechanical and Infrastructure Engineering, The University of Melbourne [2]School of Computing and Information Systems, Melbourne Centre for Data Science, The University of Melbourne.

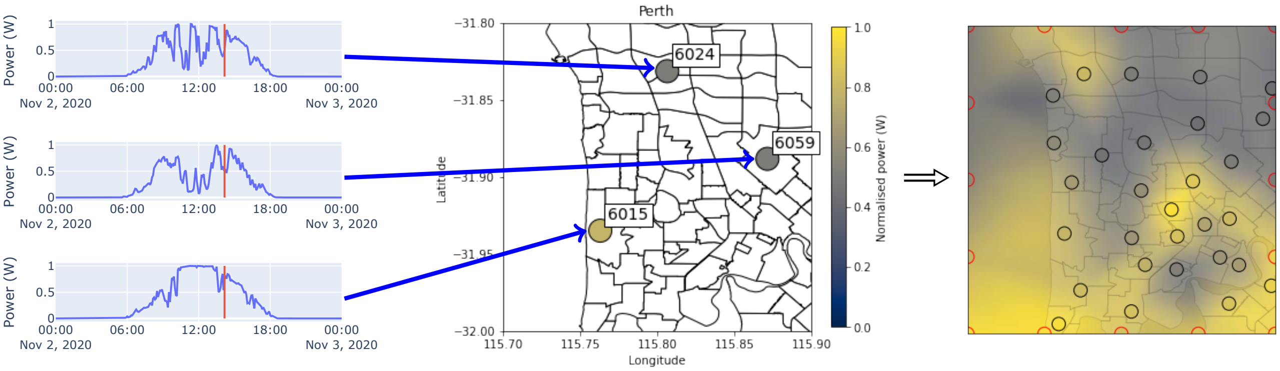

Figure 1. One day's generation data for three example point sources (solar PV systems) is shown on the left. In any interval – here represented by vertical red lines – power generation values can be represented geo-spatially (middle), in this case using markers at the centers of the postcodes that these three systems are located in. A 2D solar PV map can be generated (right) by fitting a surface to the full set of point sources (markers having black edges) and estimated boundary conditions (markers having red edges). Although this map appears to resemble cloud cover, it was generated purely from a small set of individual point sources on the ground.

sites. These maps can then be used in conjunction with other data sources (such as weather data or satellite imagery) for more accurate forecasting of the total distributed solar PV generation of an entire region. We further discuss a number of additional potential use cases for this approach and propose several further research directions.

2. Method

The method is guided by the fact that in most regions there is availability of metered data from some subset of existing solar PV systems – either through the network operator's metering infrastructure, or through third parties such as inverter manufacturers or services providers. Even when data for only a small number of PV systems is available, we can use this data to generate 2-dimensional solar PV "maps" that allow us to estimate and forecast solar PV generation across the wider region that these systems are located in.

We use a dataset of solar PV generation values for homes and businesses in Western Australia. The dataset contains values for instantaneous solar power output (in W) of 740 systems across the state, collected at 5-minute intervals over the period March 2020 to March 2021. For each solar PV system we know the postcode that it is located in, but not its exact location (due to privacy preservation). As a result, we can estimate the location of each system to within a few kilometres of its true location, using postcode centres.

The process for converting the time series data of individual solar PV systems into 2-dimensional maps is outlined in Fig. 1. The left figure shows normalised data for one day (2-Nov-2020) for three example postcodes (6024, 6059, and 6015). The middle figure shows a map of Perth, Australia, where these postcodes are located (polygons in the map represent postcodes). As can be seen, despite these postcodes being fairly close to one another (approximately 10km apart), generation profiles can vary significantly throughout the day due to the passing of clouds.

For one interval (14:10), the instantaneous power generation values can be represented as geo-spatially located values (see markers in map). Such geo-spatially located values can subsequently be converted into a full 2-dimensional "map" by fitting a surface to them. There are many ways this can be done; the resulting map shown in the right of Fig. 1 was obtained by fitting a piecewise cubic, continuously differentiable, and approximately curvature-minimizing polynomial surface (available within `scipy` [15]).

When fitting such surfaces, care must be taken to consider boundary conditions: cubic surfaces may rapidly reach large positive or negative values when extrapolating beyond the convex hull of the points used to fit the surface. Our existing approach to handle this is to introduce a set of artificial points along the corners and boundary of the area of interest (marker with red edges in right of Fig. 1), and populate them with values of nearest known points before fitting the surface. There may be better ways to do this and we intend to evaluate and explore this in future work.

These solar maps form just one component in the process for forecasting either individual systems or the total distributed solar PV system of entire regions, as shown in Fig. 2. However, they may be a helpful way to integrate individual point sources with geospatial data that is often used in solar forecasting (such as satellite imagery or weather data).

Some preliminary validation results are presented in Appendix A, and a first attempt at training a CNN with a set of sequential solar maps is presented in Appendix B. A side-by-side comparison of solar PV maps generated in this manner and satellite imagery obtained from the Advanced Himawari Imager is presented in Appendix C.

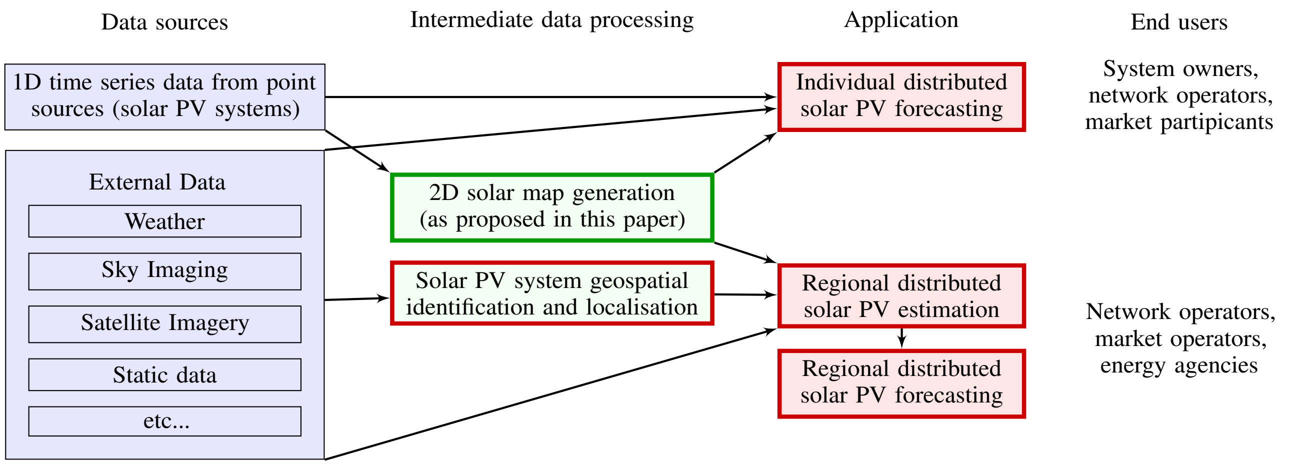

Figure 2. Overview of solar PV estimation and forecasting components and processes. Forecasting of individual systems is in general a well-studied domain, but solar PV maps may help to further improve forecast accuracy. Forecasting of distributed solar PV across entire regions remains an open area of research, and solar PV maps may provide the missing link between sparse time series data and geospatial external data sources. Applications with solid red borders are likely to benefit from deep learning-based approaches.

3. Potential Value and Future Work

This approach may be valuable for multiple reasons:

Leveraging deep learning-based approaches:
There is evident predictability in the movement of clouds across a given region, and recently it has been shown that deep learning based approaches on satellite imagery provide valuable predictive power and outperform many existing models commonly used today, for example using CNNs [12], U-Net networks [2] or gated recurrent unit - recurrent convolutional networks [16]. Given that these solar PV maps are generated using fitted surfaces, they can be considered continuous and can be converted to any desirable resolution. This enables detection of features at multiple levels of complexity, and certainly at higher resolution than the currently best available satellite imagery (which has a resolution of 1km).

Regional solar PV estimation and forecasting:
There is no region in the world for which a complete set of measurements of all distributed solar PV resources is available. Many systems are not metered at all, or may be net-metered, in which case it can be difficult to extract the solar PV signal from net demand. It is therefore not possible to assume that precise measurement of all distributed solar PV will be possible. These solar maps, however, make it possible to approximate the solar output of any given system in the area covered by the solar map. It is realistic to assume that the locations and sizes of all solar PV systems are attainable – either through accurate record keeping or via identification of solar PV systems from overhead imagery [4, 18]. As shown in Fig. 2, by combining the locations of all distributed solar PV resources with estimates of their respective generation using these solar maps, it may become possible to generate reasonably accurate solar PV estimates and forecasts of distributed solar PV for entire regions.

Accuracy of realised power measurements:
While there are many models and forecasting packages available to estimate solar production at any given set of coordinates, these are typically based on estimates of solar irradiation. However, many factors can affect the conversion of irradiation into actual solar power generation, such as system tilt and orientation, local shading, losses, and impacts due to the network. These solar PV maps are generated using actual power measurements, meaning that many of these additional factors are inherently already taken into account. For example, for some neighbourhoods, solar power generation may be impacted by geographical features such as hills or escarpments. Such impacts would likely not be taken into account in solar irradiation-based models, but are directly visible in the power output signal of individual solar PV systems in that neighbourhood.

Forecasting individual (possibly un-metered) locations:
Any pixel in the map can be converted back to a normalised value that can be used to forecast an individual system, even if it is not being actively measured on the ground. Recent advances in deep learning that outperform well established statistical approaches may be used for this purpose [11]. At the same time, an individual system will be impacted by regional dynamics of cloud movement, and therefore integrating a solar PV map into the forecasting process is likely to lead to gains in accuracy (and again, deep learning based approaches are expected to perform best).

Correlation with satellite imagery:
Finally, we see some additional potential benefits to this approach that may be of value beyond regional solar estimation and forecasting. It may be possible to use this ground-based geospatial knowledge to super-resolve satellite imagery in reverse – leading to higher resolution satellite imagery that is of value to many applications beyond solar power (*e.g.* agriculture). To date, we have only used the

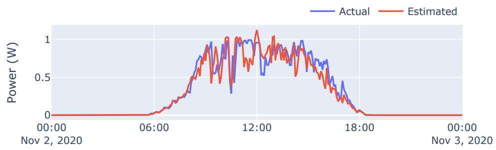

Figure 3. Validation of estimated solar PV generation for point source in postcode 6016 (in middle of area of study)

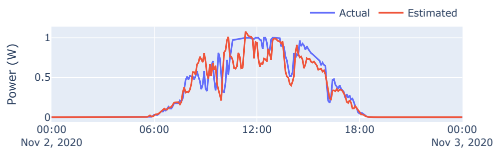

Figure 4. Validation of estimated solar PV generation for point source in postcode 6023 (near outer edge of area of study)

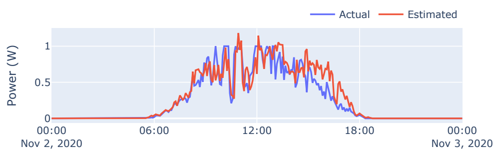

Figure 5. Validation of estimated solar PV generation for point source in postcode 6007 (when it is one of five adjacent postcodes that are withheld)

visible bands to compare satellite imagery to our solar maps (see Appendix C), but most satellites provide multi-spectral imaging across a wide range of wavelengths. The Advanced Himawari Imager, for example, can sense wavelengths from $0.43\mu m - 14.4\mu m$. A reverse analysis of the correlation between generated solar power and individual bands at different wavelengths may provide some valuable insights.

A. Preliminary Validation

We conducted preliminary validation of whether the method presented in Section 2 produces reasonable estimates of solar PV generation at non-metered locations (*i.e.*, locations for which we don't have time series data). To conduct the validation we withheld one or more point sources from the dataset, generated the 2D solar maps, and then extracted the resulting values from the 2D map at the locations of the withheld source(s), comparing these to the actual withheld measured values.

We chose the following sets of point sources to withhold as part of this validation process:

1. Postcode 6016, which is in the centre of our region of study and surrounded by other point sources.
2. Postcode 6023, which is located near the outer edge of our region of study.
3. A cluster of postcodes located close to one another (6007, 6008, 6014, 6016, 6017) – to explore whether a "hole" in the point sources used to generate the 2D map would make an impact.
4. Full k-fold cross validation.

Figures 3 and 4 show the comparison over a full day of withholding individual point sources in the center or at the outer edge of the region of study. In general there appears to be strong correlation between the estimated solar output at a single location and the true measured values, with these two point sources having a MAPE of 5.4% and 6.3%, respectively.

Figure 5 shows the comparison for a single postcode (6007), when it is one of several adjacent postcodes being withheld. Again, the estimated data matches true data well, with an average MAPE of 6.5% across all postcodes that were withheld.

Finally, we conducted a full k-fold cross validation. For k=5 – in other words, iteratively withholding groups of 20% of the point sources – the average MAPE across all point sources in all folds was 5.8%. For k=3, average MAPE was also 5.8%. This suggests that only a small number of point sources can be used to generate reasonably accurate estimates of solar power generation across larger regions.

B. Preliminary Results

In a first attempt at using these solar maps in a deep learning framework, a vanilla CNN-based network was trained on sets of five sequential solar maps to output the subsequent solar map – in other words, to forecast solar PV generation for the whole region in the next interval. We used a single convolution layer having a 3X3 kernel size with a rectified linear unit activation. Adam optimizer was used for training to minimize Mean Squared Error (MSE). Some example results are shown in Fig. 6. Each row represents a sample having five consecutive solar maps as inputs, and the ensuing solar map as a target.

To compare the network's prediction with the actual target solar map, we use MSE, calculated on a pixel by pixel basis. For this small study, a MSE of 0.035 (averaged across all samples) was achieved.

Although this is very early work, we consider these results encouraging and look forward to further exploring the use of solar maps in deep-learning based distributed solar PV forecasting.

C. Side-by-side Comparison: Satellite Images and Solar PV Maps

An overview of the process for obtaining and comparing satellite images with solar PV maps is presented in Fig. 7. The region chosen for this study (Perth, Australia) is located within the Southwest Interconnected System that covers part of Western Australia. This region has high levels of solar PV uptake which is introducing multiple challenges for the network and market operators [17].

A side-by-side comparison of satellite imagery with solar maps is presented in Fig. 8.

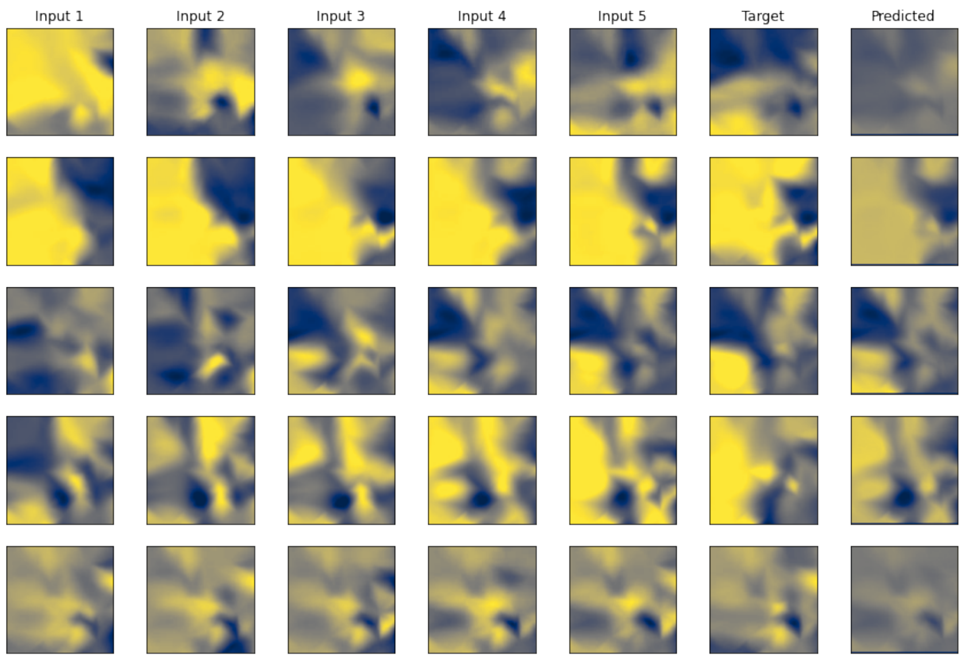

Figure 6. An example of how solar maps can be used in a deep learning framework. Each row represents a sample, where five consecutive solar maps form the input and the subsequent solar map is the target. Even with the use of a simple network, good predictions can be obtained.

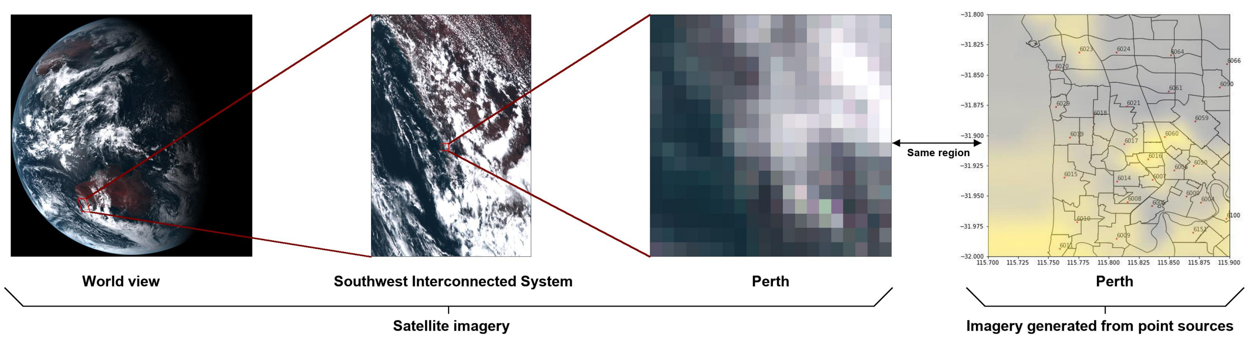

Figure 7. Overview of region under study. Images shown are for Perth, Australia, on 2 November 2020. Satellite images were obtained from the Japanese Advanced Himawari Imager and were generated using visible spectrum bands 1-3, which have a resolution of 1km.

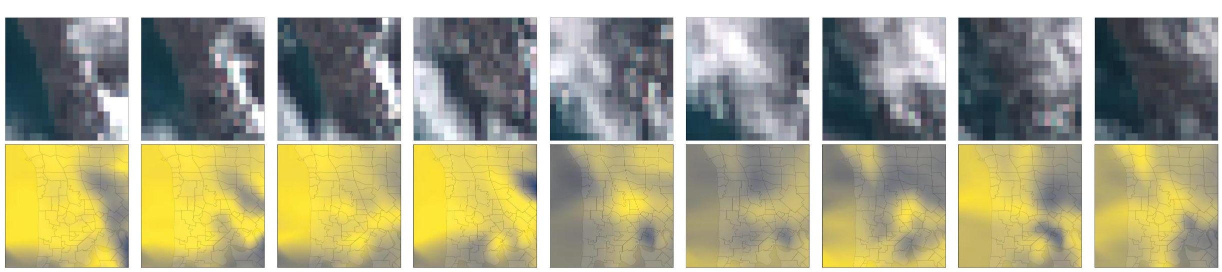

Figure 8. Comparison of satellite imagery with solar PV maps generated in ten minute intervals on 2-Nov-2020 from 13:00 – 14:20. Cloud cover in satellite images (white) appears to show a certain level of correlation with low levels of generation in the solar maps (blue/grey).

Acknowledgements

The authors are grateful to Solar Analytics for providing anonymised solar power generation data, and to NCI Australia for access to historical satellite imagery.

References

[1] Australian PV Institute. Mapping Australian Photovoltaic installations, 2021. Available online: https://pv-map.apvi.org.au/historical, accessed 30 May 2021.

[2] Léa Berthomier, Bruno Pradel, and Lior Perez. Cloud cover nowcasting with deep learning. *CoRR*, abs/2009.11577, 2020.

[3] California ISO. What the duck curve tells us about managing a green grid, 2016. Available at https://www.caiso.com/Documents/FlexibleResourcesHelpRenewables_FastFacts.pdf. Accessed 24 April 2020.

[4] Julian de Hoog, Stefan Maetschke, Peter Ilfrich, and Ramachandra Rao Kolluri. Using Satellite and Aerial Imagery for Identification of Solar PV: State of the Art and Research Opportunities. In *Proceedings of the Eleventh ACM International Conference on Future Energy Systems*, e-Energy '20, page 308–313, New York, NY, USA, 2020. Association for Computing Machinery.

[5] International Energy Agency. Renewables 2019, 2019. Available at https://www.iea.org/reports/renewables-2019. Accessed 6 February 2020.

[6] Elke Lorenz, Annette Hammer, and Detlev Heinemann. Short term forecasting of solar radiation based on satellite data. *EUROSUN2004 (ISES Europe Solar Congress)*, 01 2004.

[7] Ricardo Marquez and Carlos F.M. Coimbra. Intra-hour dni forecasting based on cloud tracking image analysis. *Solar Energy*, 91(Supplement C):327 – 336, 2013.

[8] Patrick Mathiesen and Jan Kleissl. Evaluation of numerical weather prediction for intra-day solar forecasting in the continental united states. *Solar Energy*, 85(5):967 – 977, 2011.

[9] Alejandro Navarro-Espinosa and Luis(Nando) Ochoa. Probabilistic impact assessment of low carbon technologies in lv distribution systems. *IEEE Transactions on Power Systems*, 31:2192–2203, 05 2016.

[10] Australian Energy Market Operator. Maintaining power system security with high penetrations of wind and solar generation, 2019. Available at https://www.aemo.com.au/-/media/Files/Electricity/NEM/Security_and_Reliability/Future-Energy-Systems/2019/AEMO-RIS-International-Review-Oct-19.pdf. Accessed 24 April 2020.

[11] Boris N. Oreshkin, Dmitri Carpov, Nicolas Chapados, and Yoshua Bengio. N-BEATS: neural basis expansion analysis for interpretable time series forecasting. *CoRR*, abs/1905.10437, 2019.

[12] Emilio Pérez, Javier Pérez, Jorge Segarra-Tamarit, and Hector Beltran. A deep learning model for intra-day forecasting of solar irradiance using satellite-based estimations in the vicinity of a PV power plant. *Solar Energy*, 218:652–660, April 2021.

[13] Richard Perez, Kathleen Moore, Steve Wilcox, David Renné, and Antoine Zelenka. Forecasting solar radiation – preliminary evaluation of an approach based upon the national forecast database. *Solar Energy*, 81(6):809 – 812, 2007.

[14] Gordon Reikard. Predicting solar radiation at high resolutions: A comparison of time series forecasts. *Solar Energy*, 83(3):342 – 349, 2009.

[15] scipy. Python-based ecosystem of open-source software for mathematics, science, and engineering. https://www.scipy.org/.

[16] Xinyue Su, Tiejian Li, Chenge An, and Guangqian Wang. Prediction of short-time cloud motion using a deep-learning model. *Atmosphere*, 11(11), 2020.

[17] Julius Susanto. Operational impacts of der integration in the southwest interconnected system, April 2021. Available at https://www.energy-transition-hub.org/content/operational-impacts-der-integration-swis. Accessed 10 May 2021.

[18] Jiafan Yu, Zhecheng Wang, Arun Majumdar, and Ram Rajagopal. DeepSolar: A Machine Learning Framework to Efficiently Construct a Solar Deployment Database in the United States. *Joule*, 2(12):2605 – 2617, 2018.

[19] Eric Zelikman, Sharon Zhou, Jeremy Irvin, Cooper Raterink, Hao Sheng, Jack Kelly, Ram Rajagopal, Andrew Y. Ng, and David Gagne. Short-term solar irradiance forecasting using calibrated probabilistic models. *CoRR*, abs/2010.04715, 2020.