Time-space dependency of utility-scale solar photovoltaic power generation in the National Electricity Market

Jing Huang [ID]

CSIRO Oceans and Atmosphere, Black Mountain, ACT, Australia
Email: jing.duke@gmail.com

Abstract: Driven by declining costs and the Large-scale Renewable Energy Target (LRET) in particular in Australia, the installation of utility-scale solar photovoltaic power is rapidly increasing in the recent few years. Consequently, accurate solar power forecasts are urgently needed to efficiently and economically integrate variable solar power generation into the inter-connected power system and to ensure national energy security. In addition to a central point power forecast, the associated uncertainty information is more and more being regarded useful to inform critical decision making such as planning for contingency events. As such, probabilistic forecasting of solar irradiance or power is receiving more and more attention from solar energy researchers.

Most probabilistic methods forecast solar irradiance or power at specific time and location independently without explicitly treating time-space dependency of scattered solar power generators. This is fine as far as individual solar power generators and independent time periods are concerned. However, under scenarios where the aggregated power in time and space is most concerned such as supply-demand balancing for inter-connected power systems, the time-space correlation of solar power generation will likely play a role and thus an explicit treatment of its time-space dependency becomes important. In this study, we investigate the power output of four adjacent utility-scale solar farms which are connected to the National Electricity Market using gaussian copula. We have quantitatively demonstrated the benefit of retaining the time-space dependency in probabilistic forecasting of solar power.

Keywords: *Time-space dependency, solar forecasting, probabilistic forecasting, gaussian copula, quantile regression*

1. INTRODUCTION

The conventional way to forecast solar energy generation is to provide a single value for a specific time and a specific location. However, deterministic forecasts will never be perfect because of the complexity of solar irradiance and power modelling. Therefore, the solar research community and the solar industry are becoming increasingly aware of the importance of the uncertainty information of solar forecasts for critical decision making, in particular given the trend of increasing solar energy penetration in the energy mix of most countries over the world. Probabilistic energy forecasting is also important to power grid operators. In Australia, the Australian Energy Market Operator (AEMO) has adopted probabilistic energy forecasting in the form of probability of exceedance (POE) (e.g. 10% POE and 90% POE) and assesses its forecasting models on an annual basis (AEMO, 2018a). In addition, the AEMO needs the information of the probability distribution of the supply-demand balance in the National Electricity Market (NEM) to quantify its reliability (AEMO, 2018b).

Probabilistic solar forecast are normally supplied by technologies such as quantile regression (e.g. Lauret et al., 2017) or analogy to historical scenarios (e.g. Huang and Perry, 2016). Depending on forecasting horizon, explanatory information could include historical observed values, relevant meteorological forecasts from numerical weather prediction (NWP) models (e.g. surface solar irradiance and ambient air temperature) and other useful inputs from measurement instruments such as sky imaging cameras or satellites. In a recent notice sent to its registered users, the Bureau of Meteorology (BOM) has announced that their new ensemble NWP models will be able to run operationally in 2020, providing probabilistic weather forecasts over Australia. Given the strong reliance of energy demand and renewable energy generation on weather conditions, this will provide rich information for probabilistic energy forecasting in Australia. However, it is still relatively rare for solar forecasters to explicitly treat and examine time-space dependency of solar power generation of spatially distributed solar generators.

Nevertheless, there are a few recent works which used copula to investigate the time-space correlation structure of solar irradiance or power (Golestaneh et al., 2016; Golestaneh and Gooi, 2017; Munkhammar and Widén, 2017; Munkhammar et al., 2017). In this study, we examine time-space dependency of solar power generation time series from four adjacent NEM-connected solar farms and have quantitatively shown the benefit by explicitly incorporating them in probabilistic forecasting models.

2. BACKGROUND

2.1. Selected solar farms

Table 1. Details of selected solar farms

Name	AEMO ID	Capacity (MWac)	Technology	Operation since	Gap filling (%)
Moree Solar Farm	MOREESF1	56	Single-axis tracking	04/2016	14
Nyngan Solar Plant	NYNGAN1	102	Fixed	06/2015	10
Parkes Solar Farm	PARSF1	50.5	Single-axis tracking	05/2018	13
Royalla Solar Farm	ROYALLA1	20	Fixed	09/2014	4

The majority of the solar farms which are currently operational and connected to the NEM were commissioned in 2018 or 2019. Many of them are still undergoing system testing. And consequently, they are running only at their partial capacity and/or their output power is not stable at monthly and longer scale. For the purpose of this study, we wish to select solar farms which are able to generate power stably and are in geographic proximity. Finally, four solar farms in the NSW and the ACT are identified (see Table 1 and Figure 1). One year (08/2018-07/2019) of 5-min power data are obtained for all four solar farms from the AEMO's online archive (AEMO, 2019), which are then upscaled to hourly to match the temporal resolution of forecasts by the ACCESS regional model.

Figure 1. Location of the four selected solar farms in the map of southeastern coasts. They spread approximately 6 degrees in latitude and 3 degrees in longitude.

2.2. ACCESS regional model

It has been shown that the ACCESS model exhibits excellent performance in solar irradiance forecasting, particularly over southeastern Australia (Huang et al., 2018). Hence, in this study we use forecasts of eight surface variables produced by the ACCESS regional (ACCESS-R) model to predict solar power generation (Bureau of Meteorology, 2016; Huang et al., 2018). The eight variables are hourly averages of global

horizontal irradiance (GHI, *av_swsfcdown*), direct horizontal irradiance (DHI, *av_sfc_sw_dir*), specific humidity (*av_qsair_scrn*), mean sea level pressure (*av_mslp*), air temperature (*av_temp_scrn*), total cloud cover (*av_ttld*), and zonal and meridional 10m wind speed (*av_uwnd10m* and *av_vwnd10m*), respectively. Note that all the timestamps are at the end of the hourly averaging period and all variables are extracted from the gird points nearest to the corresponding solar farms. We use forecasts initiated at UTC Hour 18 (i.e., Australian Eastern Standard Time or AEST Hour 4 next day) each day out to 24h ahead to compose continuous time series for one year (08/2018-07/2019).

2.3. Quantile regression

Historically, quantile regression has been widely used in statistics and econometrics (see Koenker, 2005). In recent years, it has been proliferating in addressing energy problems and become almost a standard approach for probabilistic solar irradiance and power forecasting (e.g. Bacher et al., 2009; Lauret et al., 2017). In this study, we use linear quantile regression to produce probabilistic solar power forecasts for the four selected solar farms from the eight surface meteorological variables by ACCESS-R. A separate linear quantile regression model is built for each solar farm and each hour from 7am to 5pm in AEST, totaling 40 quantile regression models. To minimize seasonal effects, every other day beginning with 01/Aug/2018 is used for training and the rest (i.e. beginning with 02/Aug/2018) for testing. Thus, 183 days are used for training and 182 days for testing.

2.4. Quantile score

We use the pinball loss function as the measure of quantile forecasts (Hong et al., 2016; Huang and Perry, 2016), which is comparable to the mean absolute error (MAE) for deterministic forecasting. For a quantile forecast q_α with $\alpha \in (0,1)$ as the target quantile, the corresponding meausre L for a quantile forecast q_α defined by the pinball loss function is:

$$L(q_\alpha, y) = \begin{cases} (1-\alpha)(q_\alpha - y), & \text{if } y < q_\alpha \\ \alpha(y - q_\alpha), & \text{if } y \geq q_\alpha , \end{cases}$$

where *y* is the observation truth. In this study, we mainly use quantile score (QS) in an integrated and normalized form, i.e. QS $= \sum_{\alpha=0.05,0.10,...,0.95} L(q_\alpha, y)/A$, where *A* is the nameplate capacity (MWac) of a solar farm or the aggregated nameplate capacity of multiple solar farms depending on the target of forecasts. By definition, QS is nonnegative and a zero QS implies a perfect forecast. Similar to relative MAE, QS increases with deteriorating performance.

3. GAUSSIAN COPULA

Even subject to space limit, we feel that it is necessary to briefly present some background about copula here since its use in solar irradiance and power forecasting is not very popular to date. The readers can refer to, for example, Golestaneh et al. (2016) and Munkhammar et al. (2017), for more details.

Let us consider a solar PV power generation $P_{h,z}$ for a given hour of day *h* and a given location *z*. Its cumulative distribution function (CDF) can be empirically represented by its quantile forecasts q_α as:

$$F_{h,z}(q_\alpha) = \alpha.$$

Classical probabilistic theory can be used to obtain $F_{h,z}$ independently for *h* and *z* from selected explanatory information (e.g. modelling NWP forecasts via quantile regression here in this study). To further model the time-space dependency contained in dispersed solar power generation connected in the NEM, gaussian copula needs to be introduced. According to Sklar's theorem (Sklar, 1959), the multivariate CDF of solar power generation in time and space **F** can be linked to their marginal distribution $F_{h,z}$ via a gaussian copula **C**, i.e.,

$$\mathbf{F}\big(P_{8,Mr}, P_{9,Mr}, ..., P_{16,Ryl}, P_{17,Ryl}\big) = \mathbf{C}(F_{8,Mr}, F_{9,Mr}, ..., F_{16,Ryl}, F_{17,Ryl}).$$

Here *Mr* is short for Moree and *Ryl* for Royalla. Note that the multivariate distribution can be composed similarly in time or space only, as will be seen in Section 4.3. However, the formulas written here are for both time and space. The gaussian copula is a multivariate normal distribution with a predetermined correlation structure, i.e.,

$$\mathbf{C}\big(U_{8,Mr}, U_{9,Mr}, ..., U_{16,Ryl}, U_{17,Ryl}\big) = \mathbf{F}_\Sigma(N^{-1}(U_{8,Mr}), N^{-1}(U_{8,Mr}), ..., N^{-1}(U_{16,Ryl}), N^{-1}(U_{17,Ryl})),$$

where $U_{8,Mr}, U_{9,Mr}, ..., U_{16,Ryl}, U_{17,Ryl}$ are uniformly distributed random variables transformed from the CDF of the original solar PV power generation, N^{-1} is the inverse CDF of a standard univariate normal

distribution and $\mathbf{F_\Sigma}$ is the joint CDF of multivariate normal distribution with zero mean values and a feed-in correlation matrix $\mathbf{\Sigma}$.

As can be seen above, the input information needed for gaussian copula is the correlation matrix $\mathbf{\Sigma}$ and all the marginal distribution $F_{8,Mr}, F_{9,Mr}, \ldots, F_{16,Ryl}, F_{17,Ryl}$. $\mathbf{\Sigma}$ can be calculated from the observed solar power values after they are transformed to the normally distributed domain. In this study the marginal CDF is empirically obtained through forecasts by quantile regression. Also note that all the input information is produced by data in the 183 training days only.

Finally, the transformation between an arbitrary marginal distribution and a uniform distribution is given by:

$$U_{h,z} = F_{h,z}(P_{h,z}) \text{ or } P_{h,z} = F^{-1}{}_{h,z}(U_{h,z}).$$
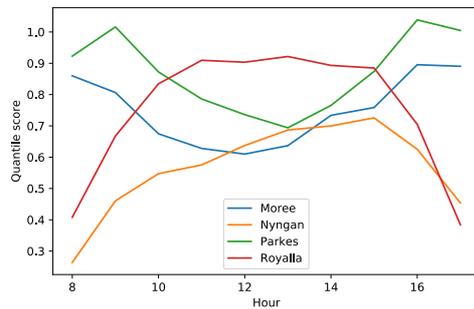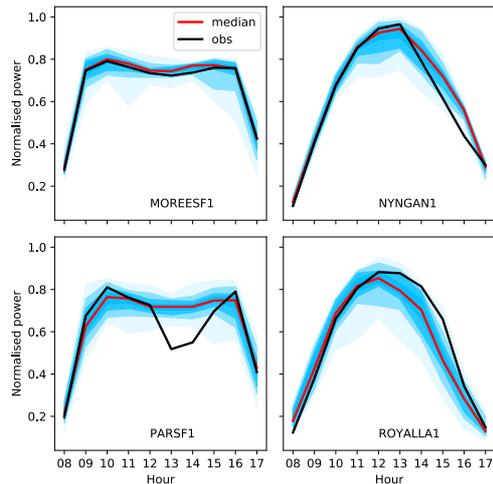
4. RESULTS

4.1. Pre-processing solar power data
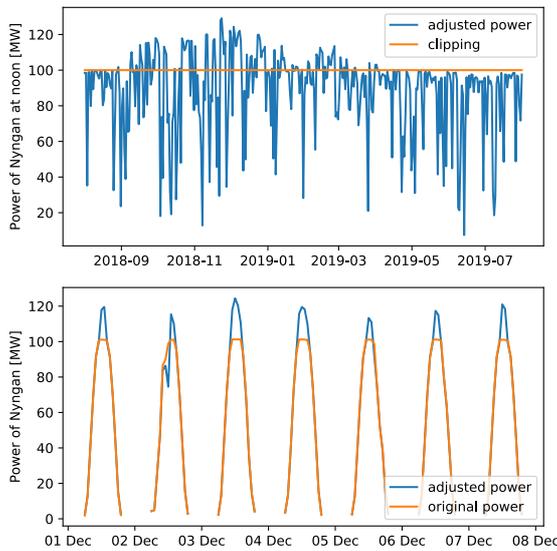
Figure 2. Illustration of preprocessing of Nyngan power data: (top) power generation at noon for a year with all power above the clipping line (100 MWac) gap-filled; (bottom) original power vs. adjusted power for seven summer days.

The AEMO-published 5 min power generation data needs to be pre-processed before being used for probabilistic modelling. Three factors are considered during this step: (1) the power time series are averaged to an hourly resolution to match the resolution of the weather forecasts by ACCESS-R; (2) small fraction of missing data can be filled with meaningful values; (3) all the

four solar farms studied here clip their power generation to some extent, which could be due to capacity limit of their AC inverters or other practical reasons. Firstly, we exclude the power values exceeding or being close to their corresponding AC capacity by using thresholds slightly smaller than their capacity. By visually examining the probability density functions of power values, the thresholds are set at 53, 100, 50, 19 MW, respectively. Then, power time series are averaged to an hourly resolution. Quantile regression is used for each solar farm to construct a prediction model of median. All meaningful data between 8am and 5pm are used for training and then missing and excluded data are gap filled by the median model. Throughout this process, the percentages of gap filling are 14% for Moree, 10% for Nyngan, 13% for Parkes and 4% for Royalla, respectively. This process is illustrated in Figure 2 using Nyngan solar farm as an example.

Figure 3. (top 4) Probabilistic forecasts of four solar farms on 28/07/2019 by quantile regression. The bands of prediction intervals are bound sequentially by the decile forecasts (i.e. q_{10}, q_{20} ,…, q_{90} upward). (bottom) averaged intra-day patterns of quantile scores of four solar farms.

4.2. Independent quantile regression

With the missing and clipped power data filled, we are able to demonstrate the performance of quantile regression in Figure 3. There is a clear difference in the intra-day patterns between solar farms deploying fixed PV panels (i.e. Nyngan and Royalla) and deploying single-axis tracking PV panels (i.e. Moree and Parkes). The power generation pattern of solar farms using fixed PV panels resembles that of global horizontal irradiance whilst solar farms using single-axis tracking PV panels have a long plateau of power generation approximately during 10am to 4pm. As shown in Figure 3, quantile regression is able to forecast power generation reasonably well from ACCESS-R forecasts with its median forecast close to the actual generation most of the time. However, large deviations exist such as Parkes solar farm during 1pm-2pm, which may be due to forecasting error of NWP or internal contingency events of the solar farms. Day-averaged quantile score is then calculated for each solar farm and each hour of day (Figure 3). The difference between the two technologies is reflected in the quantile score as well. Solar farms fixing PV panels have higher forecasting errors around noon whilst solar farms tracking the sun from east to west actually forecast the noon period more accurately than early morning and late afternoon. Nyngan solar farm has the lowest QS overall because its inland location favours a clear-sky weather which is generally easier for ACCESS-R to forecast (see Huang et al., 2018). In terms of reading QS, we note that it is difficult to associate the value of QS with the quality or certainty of probabilistic forecasts mainly because of a lack of reported QS values for similar studies in literature. However, it makes sense to use QS of the forecasts by quantile regression as a benchmark since it's becoming increasingly popular in probabilistic solar forecasting.
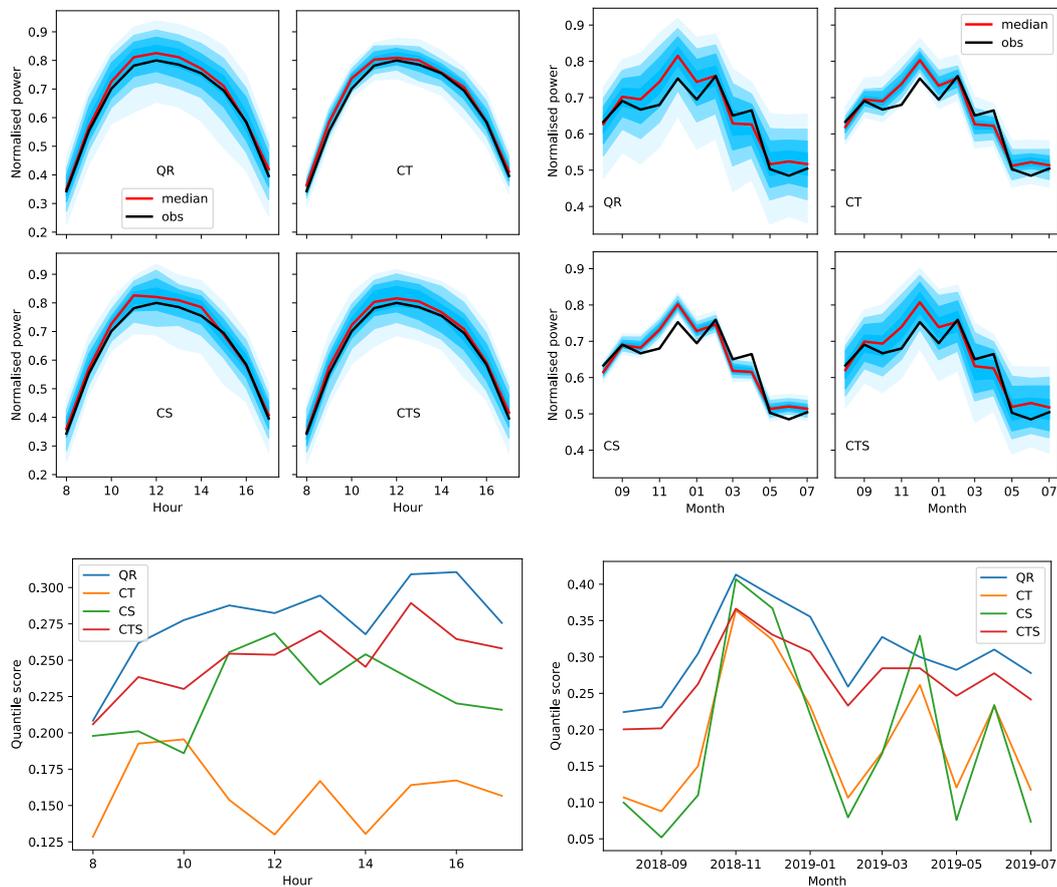
Figure 4. (top left 4) Quantile/decile forecasts (i.e. q_{10}, q_{20},…, q_{90} similarly as in Figure 3) of solar power generation for each hour of day, which are aggregated over 4 solar farms and 182 testing days and then normalized by the aggregated capacity (228.5 MWac). They are produced by independent quantile regression (QR), gaussian copula in time (CT), gaussian copula in space (CS) and gaussian copula in time and space (CTS), respectively. (bottom left) The intra-day patterns of the corresponding quantile scores of the four approaches calculated for testing period only. (top right 4) Quantile/decile forecasts of solar power generation for each month, which are aggregated over 4 solar farms and 182 testing days in each month and then normalized by the aggregated capacity. Again, they are produced by QR, CT, CS and CTS, respectively. (bottom right) The seasonal patterns of the corresponding quantile scores of the four approaches calculated for testing period only.

4.3. Time-space dependency

We now employ the gaussian copula introduced in Section 3 to generate scenarios of probabilistic solar power forecasts. More specifically, the procedure of scenario generation can be divided into three steps: (1) all the marginal distributions $F_{h,z}$ are obtained from independent quantile regression (cf. Section 4.2); (2) the correlation matrix Σ is derived from observed solar power in the training period, thus formulating the gaussian copula $\mathbf{C}$ as $\mathbf{F_\Sigma}$; (3) scenarios are generated by $\mathbf{C}$, which are then calibrated to conform to $F_{h,z}$ for each time h and location z. As mentioned, it is possible to construct a gaussian copula in time or in space only or in both time and space. Here we generate 100 scenarios for each time and location incorporating correlation structure in time only (CT), in space only (CS) and in both time and space (CTS), respectively. The time-space structure of solar power generation is best revealed by aggregation of probabilistic solar power forecasts in time and space, which are calculated from the 100 scenarios. This is because that copula uses the marginal probability distribution functions (PDFs) produced by quantile regression as input so that the forecasts given by quantile regression and by copula at a specific time and a specific location would be similar in a probabilistic sense.

In Figure 4, we examine forecasts of such aggregations against hour of day and against month, respectively. It is shown that copulas are generally able to forecast the aggregated power more accurately than quantile regression evidenced by narrower prediction intervals which still encompasses the actual power generation. This can be further quantified by comparing their corresponding quantile scores. It is interesting to observe that the quantile scores of CT and CS are often lower than those of CTS, indicating better performance, whilst all copula forecasts outperform the quantile regression forecasts. However, it appears that the performance of CTS is generally more robust than CT and CS alone (e.g. 11/2018 and 04/2019 for the forecasts of monthly power aggregation). This may be because that less constraint on the correlation of power forecasts in time-space may improve the performance of power forecasts in general but does not do well in terms of forecasting anomalies.

5. DISCUSSION AND CONCLUSION

In this study, we managed to explicitly incorporate the time-space correlation structure of solar power generation into probabilistic forecasts. We firstly used classic quantile regression techniques to obtain these independent probabilistic forecasts from intra-day weather forecasts by the BOM's ACCESS-R model. The difference between fixed PV panel technology and single-axis tracking technology was revealed in terms of intra-day pattern of probabilistic solar forecasting errors. And then, we employ gaussian copula to generate scenarios of solar forecasts which conform to both the probabilistic structures of the independent forecasts as well as the time-space correlation structure of actual power generation. We note some limits and potential improvements of this work as follows.

- More physical approaches can be employed to improve the irradiance-to-power modelling. In particular, the power forecasts of the single-axis trackers have significant errors during early morning and late afternoon, which can definitely be improved by a realistic model for plane-of-array irradiance. However, this does not affect the main conclusion of this work on the role of time-space dependency.
- Given that the 5 min resolution is the most important time frame for the participants of the NEM and the availability of the 5 min power generation time series, it is possible to extend this work at hourly resolution to 5 min.
- It is possible to improve the pre-processing step of solar farm data by more realistically identifying the reasons of anomalies (e.g. by communicating with the solar farms investigated) and modelling the behaviours of their AC inverters. Longer (e.g. multi-years) power data for training and testing is always desired.
- The time-space correlation structure may be influenced by seasonal patterns to some extent. Hence, by removing these patterns from the observed power prior to calculating and passing the correlation matrix to copula may achieve higher accuracy than demonstrated in this work.
- Although gaussian copula is among the most popular ones, it may worth trying and comparing with other types of copulas for probabilistic solar forecasting, such as R-Vine copula investigated by Golestaneh and Gooi (2017).

To conclude, we quantified the performance of probabilistic solar forecasting by the so-called pinball loss function and are able to show the superiority of explicitly retaining this time-space dependency relative to the approach of producing probabilistic forecasts independently. In addition to central power grid operators such as the AEMO, this work is potentially useful to any party who is interested in aggregated power forecasting in time and space such as local distribution network service providers (DNSPs) and companies who own multiple distributed renewable generators.

ACKNOWLEDGMENTS

REFERENCES

AEMO (2018a) *Forecast Accuracy Report 2018*. Available at: https://www.aemo.com.au/-/media/Files/Electricity/NEM/Planning_and_Forecasting/Accuracy-Report/Forecast-Accuracy-report-2018.pdf.

AEMO (2018b) *The NEM Reliability Framework*. Available at: https://www.aemc.gov.au/sites/default/files/2018-11/Additional information from AEMO to support its Enhanced RERT rule change proposal.pdf.

AEMO (2019) *Public Dispatch SCADA*. Available at: http://nemweb.com.au/Reports/Archive/Dispatch_SCADA/.

Bacher, P., Madsen, H. and Nielsen, H. A. (2009) 'Online short-term solar power forecasting', *Solar Energy*, 83(10), pp. 1772–1783. doi: 10.1016/j.solener.2009.05.016.

Bureau of Meteorology (2016) *BNOC Operations Bulletin Number 107 - APS2 Upgrade to the ACCESS-R Numerical Weather Prediction System*. Available at: http://www.bom.gov.au/australia/charts/bulletins/apob107-external.pdf (Accessed: 10 April 2018).

Golestaneh, F. and Gooi, H. B. (2017) 'Multivariate prediction intervals for photovoltaic power generation', in *2017 IEEE Innovative Smart Grid Technologies - Asia (ISGT-Asia)*. IEEE, pp. 1–5. doi: 10.1109/ISGT-Asia.2017.8378380.

Golestaneh, F., Gooi, H. B. and Pinson, P. (2016) 'Generation and evaluation of space-time trajectories of photovoltaic power', *Applied Energy*. doi: 10.1016/j.apenergy.2016.05.025.

Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A. and Hyndman, R. J. (2016) 'Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond', *International Journal of Forecasting*, pp. 896–913. doi: 10.1016/j.ijforecast.2016.02.001.

Huang, J. and Perry, M. (2016) 'A semi-empirical approach using gradient boosting and k-nearest neighbors regression for GEFCom2014 probabilistic solar power forecasting', *International Journal of Forecasting*, 32(3), pp. 1081–1086. doi: 10.1016/j.ijforecast.2015.11.002.

Huang, J., Rikus, L. J., Qin, Y. and Katzfey, J. (2018) 'Assessing model performance of daily solar irradiance forecasts over Australia', *Solar Energy*, 176, pp. 615–626. doi: 10.1016/j.solener.2018.10.080.

Koenker, R. (2005) *Quantile Regression*. London: Cambridge University Press.

Lauret, P., David, M. and Pedro, H. T. C. (2017) 'Probabilistic solar forecasting using quantile regression models', *Energies*, 10(10), p. 1591. doi: 10.3390/en10101591.

Munkhammar, J. and Widén, J. (2017) 'An autocorrelation-based copula model for generating realistic clear-sky index time-series', *Solar Energy*. doi: 10.1016/j.solener.2017.09.028.

Munkhammar, J., Widén, J. and Hinkelman, L. M. (2017) 'A copula method for simulating correlated instantaneous solar irradiance in spatial networks', *Solar Energy*. doi: 10.1016/j.solener.2016.12.022.

Sklar, A. (1959) 'Fonctions de répartition à n dimensions et leurs marges', *Publ. Inst. Statist. Univ. Paris*, 8, pp. 229–231.