Probabilistic Forecasting of Solar Energy

John Boland[1]
[1]Industrial AI Research Centre, University of South Australia
John.boland@unisa.edu.au

This paper describes the forecasting of 15 minute solar irradiation on a horizontal plane (GHI) for Seattle, USA, as well as 15 minute solar farm output for Broken Hill, Australia. The goal is to set error bounds on the forecast, specifically estimating 15 quantiles, from essentially minimum to maximum. In practice, the quantiles calculated are $0.005, 0.025, 0.05, 0.1, 0.2, \dots, 0.8, 0.9, 0.95, 0.975, 0.995$. The forecast horizons for the solar farm output is one step ahead (for time $t + 1$ time interval performed at time t). The forecast for GHI is part of a benchmarking project under Task 16 of the International Energy Agency. The Task focuses on Solar resource for High Penetration and Large Scale Applications. The procedure entails first calculating point forecasts, and then using quantile regression techniques to form the quantiles of the resulting noise terms. The modelling process is performed on a year's data for 2017 for both locations, and then tested on data from 2018. In the standard modelling manner, the models developed for both the point forecasts and quantiles on the 2017 data are applied to the 2018 data, whereupon the quantiles are added to the point forecasts for verification of the efficacy of the procedure.

The point forecast contains a model for the seasonality using Fourier series for the significant cycles. For GHI, they are once a year, once and twice a day, plus beat frequencies to modulate the daily cycle to suit the time of year. Since the solar farm has an oversized field, thus capping the output, the only necessary cycles are once and twice a day. Once the seasonality model is subtracted from the original series, the residuals are represented by an $ARMA(p, q)$ forecast model. The combination of the models forms the point forecast. The noise terms from this process are modelled using quantile regression.

For quantile level τ of the response variable,

$$\min_{\beta_0(\tau), \beta_1(\tau), \dots, \beta_p(\tau)} \sum_{i=1}^{n} \rho_\tau \left(z_i - \beta_0(\tau) - \sum_{j=1}^{p} z_{i-j} \beta_j(\tau) \right)^2 .$$

In this ρ is called the check function and is equal to

$$\rho = \tau \max(r, 0) + (1 - \tau) \max(-r, 0)$$

r is the error betwen the noise z_i at time i and the model, and if the error in the regression in a single period is positive, then the check function multiplies the error by τ and by $1 - \tau$ if negative. The regression is for the noise z_i at time i as a function of the noise at the previous 5 time steps.

The quantile regression approach for obtaining the prediction intervals was used because of the skewness of the noise distributions in each case. To evaluate the worth of this method, the results were compared to assuming the noise terms are independent and identically distributed (iid) normal variates. Two metrics are used for the comparison - Coverage and Mean Width of the intervals. Coverage means that if one is designing a 95% prediction interval, approximately 95% of the observations should fall within the interval. As well as coverage, a smaller mean width of intervals is sharper and better. The comparison was performed for three probabilities - 80%, 90% and 95%. Interestingly, the assumption of iid normal was slightly better for the 95% case as both approaches had good coverage, but the normal intervals had a smaller mean width. For the other two cases, the quantile regression approach was significantly better. In fact, the coverage of the normal assumption case for the 80% case was of the order of 88%, significantly different from that desired. The interval widths for both that case and 90% using the normal assumption were much greater than the quantile regression usage.

Figure 1 gives a representation of the resulting prediction intervals for Seattle. Note that the intervals, as well as the forecast, are zeroed at night.

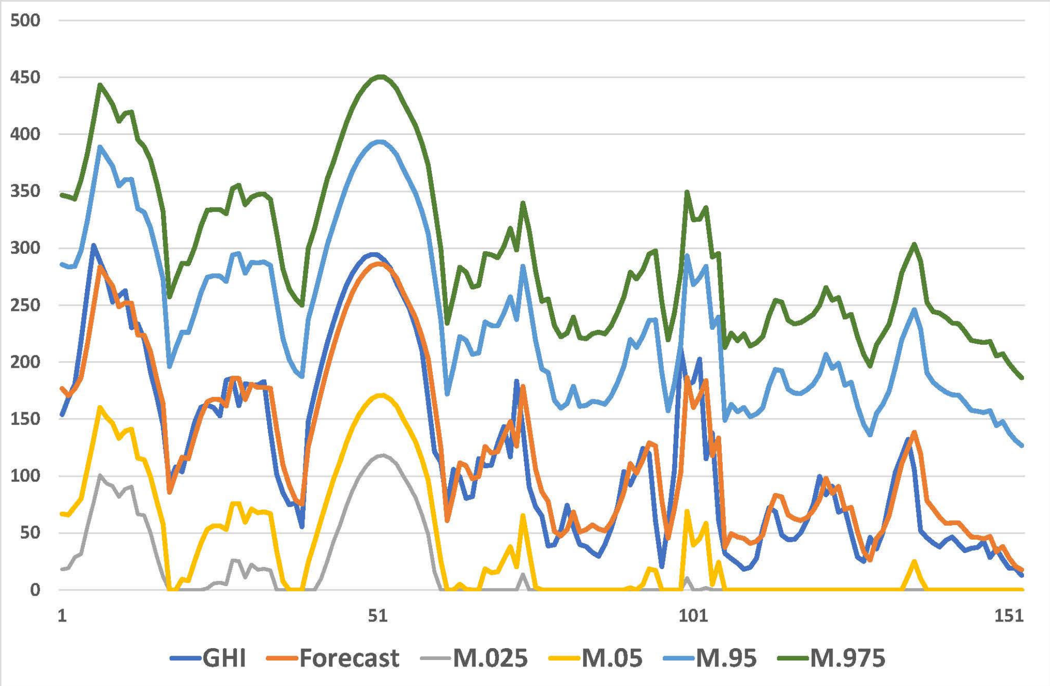

Figure 1. Seattle GHI, one step forecast with 90 and 95% prediction intervals

Table 1 gives the quantitative metrics for in sample (2017) and out of sample (2018) results and comparison with the benchmark model. As is the custom in the literature all evaluations are done for the solar altitude equal to or greater than 10^{o}.

Table 1 Seattle coverage and interval widths

	Percentage	Present	Benchmark
Coverage			
	80	80.4	88.7
	90	90.1	92.1
	95	95.0	94.1
Width			
	80	122.0	184.4
	90	211.0	233.0
	95	308.7	273.8

When analysing the forecasting for solar farms, it is important to note that the Broken Hill installation, like many in Australia, is capped at a specific level of output. Since the panel field is oversized, it also means that on a clear day, the output will be constant for a number of hours, even in winter – see Figure 2 as an example.

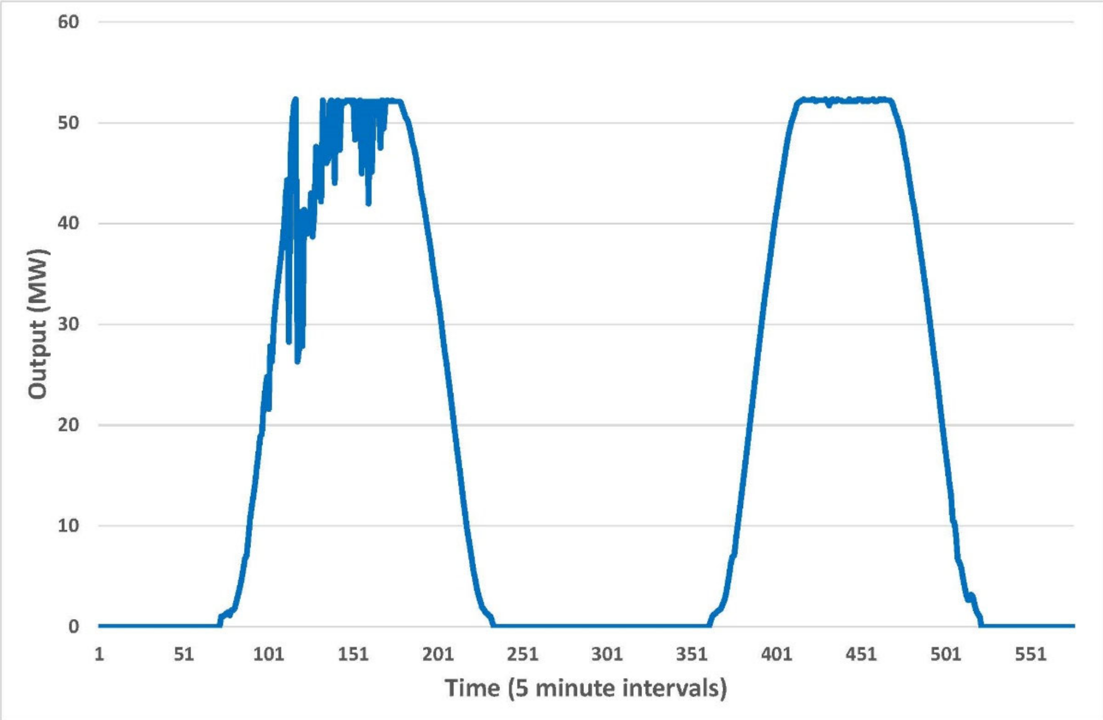

Figure 2. Two days of output from the Broken Hill solar farm

As a result, when constructing the prediction intervals, they are constrained to not exceed the cap, as well as being set to zero at night, as was the case with the Seattle GHI.

Figure 3 shows the output, forecast and 90 and 95% prediction intervals for Broken Hill, and Table 2 gives the associated error metrics.

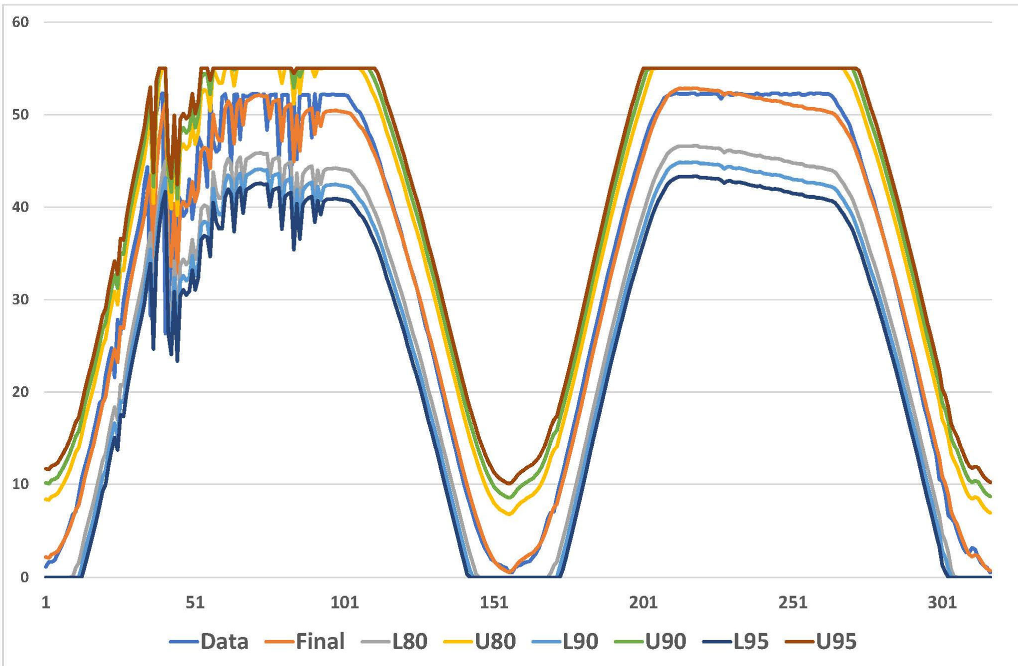

Figure 3. Output, one step ahead forecast and prediction intervals for Broken Hill

Table 2: Broken Hill coverage and interval widths

	Percentage	Present	Benchmark
Coverage			
	80	83.6	91.0
	90	92.2	93.3
	95	96.2	94.9
Width			
	80	6.10	11.24
	90	12.54	14.01
	95	18.91	16.37

Conclusion

The use of quantile regression has been demonstrated for construction of prediction intervals for both GHI and solar farm output. The GHI analysis is part of an IEA Task 16 benchmarking project. The project includes forecasting with prediction intervals for various horizons. Since the statistical techniques shown here are only sensible for short term forecasting, the extension is only being undertaken for steps $t + 15$, $t + 30$, $t + 45$ and $t + 60$ minute horizons. The other extension of the work being undertaken is to compare the results with the more complicated procedure outlined in Boland and Grantham (2018). In that paper the errors underwent a normalising transformation, and an exponential smoothing average forecast was used on the squared errors. Then prediction intervals were constructed and back transformed and inserted around the point forecasts.

References

Boland, J. and Grantham , A., 2018, Nonparametric Conditional Heteroscedastic Hourly Probabilistic Forecasting of Solar Radiation, *J Multidisciplinary Journal*, pp. 174–191, doi:10.3390/j1010016.