25th International Congress on Modelling and Simulation, Darwin, NT, Australia, 9 to 14 July 2023
mssanz.org.au/modsim2023

Probabilistic forecasting for solar energy

John Boland

UniSA STEM, University of South Australia, Adelaide
Email: john.boland@unisa.edu.au

Abstract: 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 $\{0.005, 0.025, 0.05, 0.1, 0.2, \ldots, 0.8, 0.9, 0.95, 0.975, 0.995\}$. The forecast horizons for both variables are one step ahead (for time $t + 1$ time interval performed at time t). 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 initial 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, the goal is to

$$\min_{\beta_0(\tau),\beta_1(\tau),\ldots,\beta_p(\tau)} \sum_{i=1}^{n} \rho_\tau(y_i - \beta_0(\tau) - \sum_{j=1}^{p} z_{ij}\beta_j(\tau))^2 \tag{1}$$

$$\rho = \tau \max(r, 0) + (1 - \tau) \max(-r, 0) \tag{2}$$

is the check function. If the error in the regression in a single period, r, is positive, then the check function multiplies the error by τ and by $1 - \tau$ if negative. In the study performed on both the Seattle GHI data and the Broken Hill output data, the predictor variables are the previous 5 lagged values of the noise.

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. 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%, The interval widths for both that case and 90% using the normal assumption were much greater than the quantile regression usage.

In conclusion, using quantile regression to form the prediction intervals proved to be a valid approach, using the evaluation metrics we adopted. Further metrics will be tested such as Interval Score, and Cumulative Rank Probability Score (CRPS). The study will be extended to wind speed and wind farm output as well.

Keywords: Global horizontal irradiation, solar farms, forecasting, prediction intervals, quantile regression