26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

Dynamic short-term solar forecasting using historical data and machine learning under varying weather conditions

L. Bai [a]

[a] *Industrial AI Research Centre, STEM, University of South Australia, Mawson Lakes, 5095, Australia*
Email: baily012@mymail.unisa.edu.au

Abstract: This study investigates the forecasting potential of LSTM, XGBoost, and Transformer-based models in comparison to Smart Persistence and a simple yet effective benchmark model for univariate time series forecasting. Designed for practical deployment in real-world grid operation settings, all models use only 5-minute historical power output data from large-scale, grid-connected solar farms. The aim is to evaluate the performance of different forecasting techniques under varying weather conditions and to develop a dynamic system that selects the most appropriate model for predicting the next 5-minute solar power output.

The benchmark model, adapted from Farah and Boland (2021), combines a Fourier series to capture seasonal patterns with an autoregressive moving average (ARMA) component to model residuals. Persistence detection is also included to enhance performance on clear days. Despite its simplicity, the benchmark model performs well, particularly in terms of normalised root mean squared error (NRMSE), but it exhibits under-forecasting during capped periods and shows room for improvement in normalised mean absolute error (NMAE).

Clear-sky index (CSI)-based models are commonly used to classify weather conditions, serving either as baseline comparators or input features. However, sky conditions can vary considerably within 5-minute intervals and remain unknown for future time steps. This study reviews several approaches in the literature to identify an optimal weather classification method based on historical power outputs. To evaluate the performance of the benchmark model (Farah & Boland 2021), clear-sky power output for each timestamp is defined as the output on the day with the highest total energy generation over the past 365 days. Boland, Farah and Bai (2022) propose using the maximum value for each time step across the last 30 days. Huang et al. (2024) introduce a Datum Day Array (DDA) model using solar radiation from the highest-radiation day of each month. Poddar, Kay and Boland (2024) present a clustering-based sky classification approach that uses daily CSI and aggregate ramp rates to define five weather types: clear, overcast, low intermittent, high intermittent and high variability.

Initial analysis of 5-minute data from the Broken Hill Solar Farm (1 September 2016 - 31 August 2018) shows that XGBoost performs particularly well on sunny days, reducing both RMSE and MAE while effectively capturing capped output periods. The LSTM model shows slight improvements under overcast conditions, whereas the Transformer-based model offers limited value relative to its computational cost.

The study will continue to investigate robust weather classification techniques for intermittent periods and identify the most effective forecasting models accordingly. Ultimately, a generalised classification framework will be developed to support a dynamic and adaptive short-term forecasting system using only historical power output data.

REFERENCES

Boland, J, Farah, S & Bai, L 2022, 'Forecasting of Wind and Solar Farm Output in the Australian National Electricity Market: A Review', Energies, vol. 15, no. 1, p. 370.
Farah, S & Boland, J 2021, 'Time series model for real-time forecasting of Australian photovoltaic solar farms power output', Journal of Renewable and Sustainable Energy, vol. 13, no. 4, p. 046102.
Huang, J, Yuan, C, Boland, J, Guo, S & Liu, W 2024, 'One-step ahead short-term hourly global solar radiation forecasting with a dynamical system based on classification of days', Renewable Energy, vol. 237, p. 121639.
Poddar, S, Kay, M & Boland, J 2024, 'A Standardized Sky Condition Classification Method for Multiple Timescales and Its Applications in the Solar Industry', Energies, vol. 17, Multidisciplinary Digital Publishing Institute, no. 18, pp. 4616–4616.

Keywords: Solar forecasting, machine learning, time series prediction, weather classification