Solar Irradiance and Power Forecasting in Piscataway, NJ Using Data Science

Rutgers Aresty Research Center for Undergraduates

Irfan Peer, Dr. Ahmed Aziz Ezzat

Industrial and Systems Engineering, Rutgers University, Piscataway, New Jersey 08854

Abstract

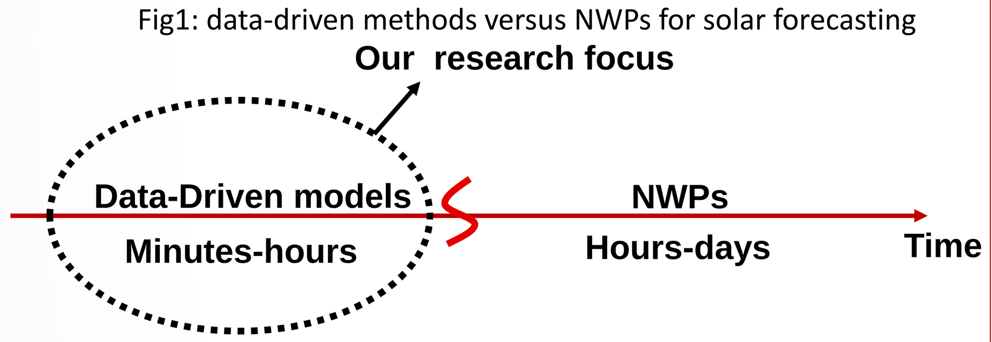
In the last decade, the US. Solar power capacity increased from just 1.2GW up to 70GW, supplying ~2.5% of the U.S. electricity demand [1]. Despite the promise, solar energy still faces several technical challenges due to its intermittency and limited predictability. To ensure a reliable integration into the grid, power system operators need to accurately anticipate, ahead of time, how much sunlight will be there on a given timing on a given day, in turn helping them forecast how much solar energy can be produced. Motivated by that, this research is focused on developing an accurate, data-driven method to produce accurate short-term forecasts of solar irradiance and power at fine spatial and temporal resolutions (6-hours ahead, 10-min resolution, in Piscataway, NJ). Various statistical forecasting models have been formulated and tested using real-world irradiance data collected at the Energy Lab at Rutgers University. Our numerical experiments suggest that an autoregressive-based model, L-ARIMA, can achieve up to 97% and 94% improvement, in terms of forecast accuracy, relative to Neural nets and persistence forecasts, respectively. This work contributes to accurately characterizing the variability of solar energy, one of the main challenges to boost its economic outlook.

Background, Data Collection & Analysis

Due to the intermittency of solar energy, forecasting is critical to running operations as grid operators need to know, within a certain confidence, the energy that can be produced on a given timing on a given day. This allows operators to set appropriate prices for buyers/sellers, deal with under- or over- energy production, and plan how to distribute electricity. Accurate forecasts can inform those decisions and hence, save billions of dollars a year in a lucrative market on the rise.

There are two approaches, each with pros and cons, to forecasting renewable energy (See Fig1):
1. Physics-based Numerical Weather Prediction (NWP) models
2. Data-driven statistical models

Fig1: data-driven methods versus NWPs for solar forecasting



Statistical methods rely on large amounts of data to learn the historical trends and correlations and make accurate predictions of the future. Data for this research was collected using a CS320 Digital thermopile pyranometer sensor installed at the Energy Lab at Rutgers University in Piscataway, NJ. The data comprises 10-min irradiance measurements recorded between April-June 2020 in 10-min resolution, corresponding to 8856 data points. Fig. 2a shows the data collection and processing procedures.

Data Collection **Statistical Analysis**

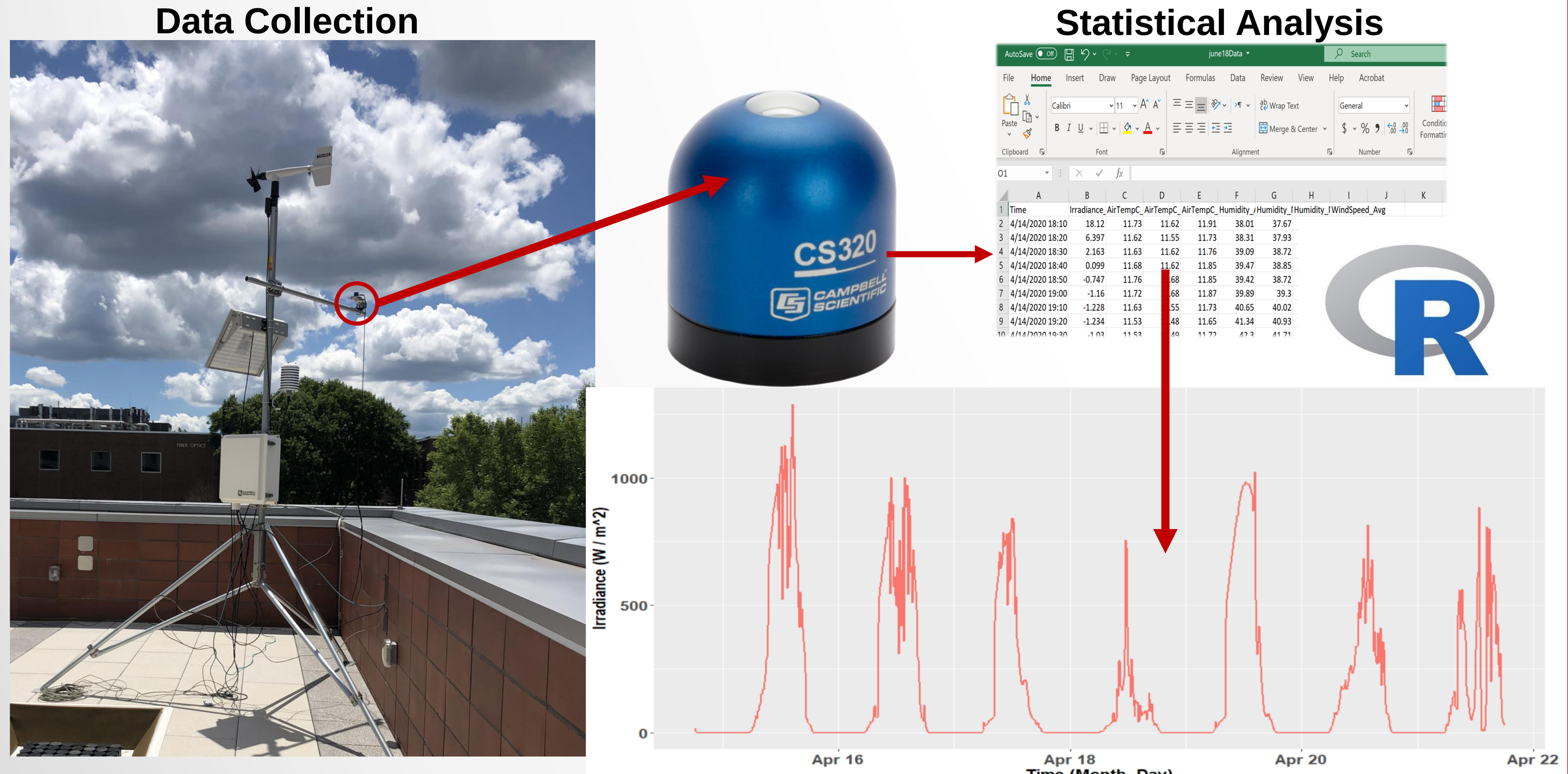


Fig 2. Plot of a subset (April 15 to April 22) of solar irradiance data collected from the energy lab in Richard Weeks Hall. Notice the seasonal trend in the data, a periodic rising and falling motion that mirrors the trajectory of the Sun throughout the day. Some peaks are lower than others, potentially due to clouds or other atmospheric disturbances encountered.

Acknowledgements

[1] Solar Energy Industries Association. http://www.seia.org/
[2]Renewables' and Industrial Analytics (RIA) Group.
- https://sites.rutgers.edu/azizezzat/research/

Irfan Peer - ip233@scarletmail.rutgers.edu

Materials and Methods

Five models were formulated and tested to determine which performed the best and resulted in the highest accuracy improvements compared to its compatriots:
1. *ARIMA*: Leverages temporal correlations by regressing future values on a combination of past values and lagged noise terms.
2. *L-ARIMA*: Trained on data parsed through ln function to ensure positivity, reduce error and unwanted variance
3. *Daylight ARIMA*: Trained on nonzero data values to account for only daytime measurements
4. *Artificial Neural Network*: Uses hidden layers, neurons, backpropagation, nodes to correct itself during forecasting
5. *Persistence Forecast*: Assumes status-quo "persists" in the forecast horizon. Used as benchmark to compare final results

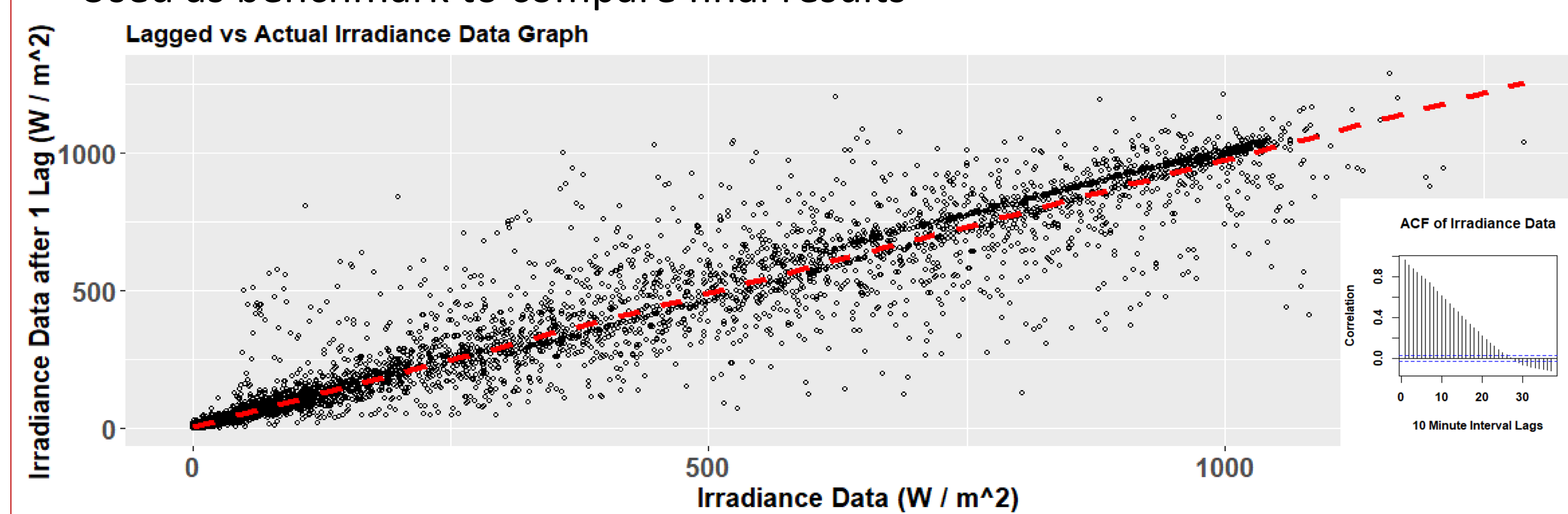


Fig. 3. The positive trend between the lagged values and actual data show the applicability of an autoregressive model. The ACF plot furthers this conclusion by showing the high correlation in lagged values in the time series.

ARMA (p,q) model

X_t is predicted value; X_{t-i} are lagged values; ϵ_{t-i} are lagged white noise values

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \phi_i \epsilon_{t-i} + \epsilon_t$$

c is an intercept; φ_i is coefficient of autoregression, weight dependent on value of p; ϕ_i is weight determining correlation between white noise value and dependent variable

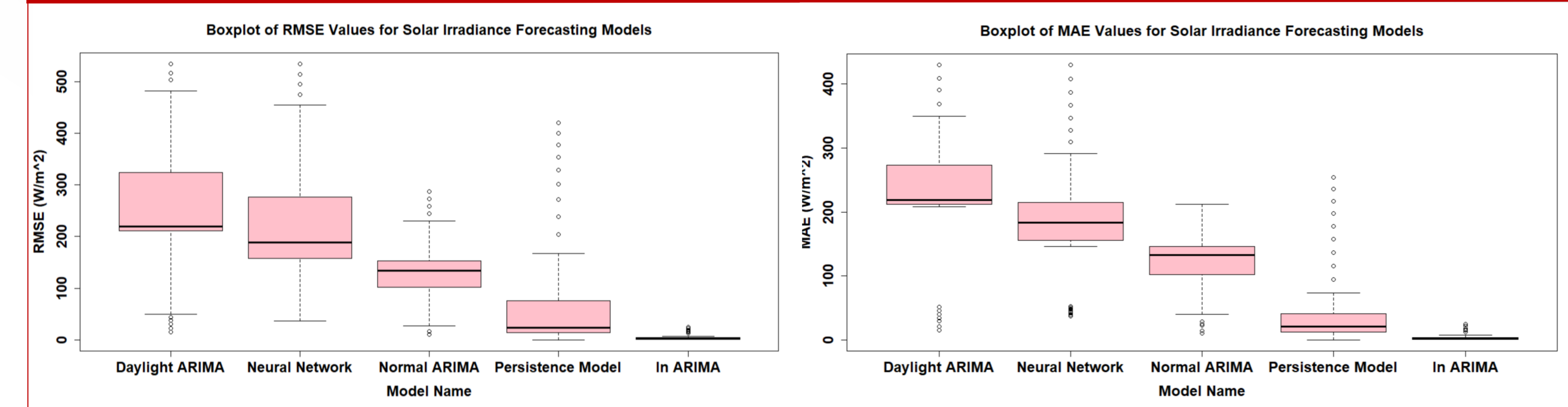
Each model constructed requires a slightly different approach to processing the data. However, there were some commonalities;
- Time stamp from data converted into time object to be usable within R programming language
- All negative irradiance values, if any, in data converted to zero

The following details the ***uniqueness*** in some of the approaches:
1. *L-ARIMA*: All data points not equal to zero stored in separate column after taking natural log of their value.
   - Negative and zero irradiance values essentially irrelevant
   - Converted back to irradiance forecast through mathematical process.
   - Forecasting error can be significantly reduced.
2. *Daylight ARIMA*: All nonzero data points stored separately for use.
   - Method focused on using only daytime irradiance measurements, assuming nighttime values not necessary for forecasting irradiance.
   - Time stamp details of these daytime-only values stored separately as well.

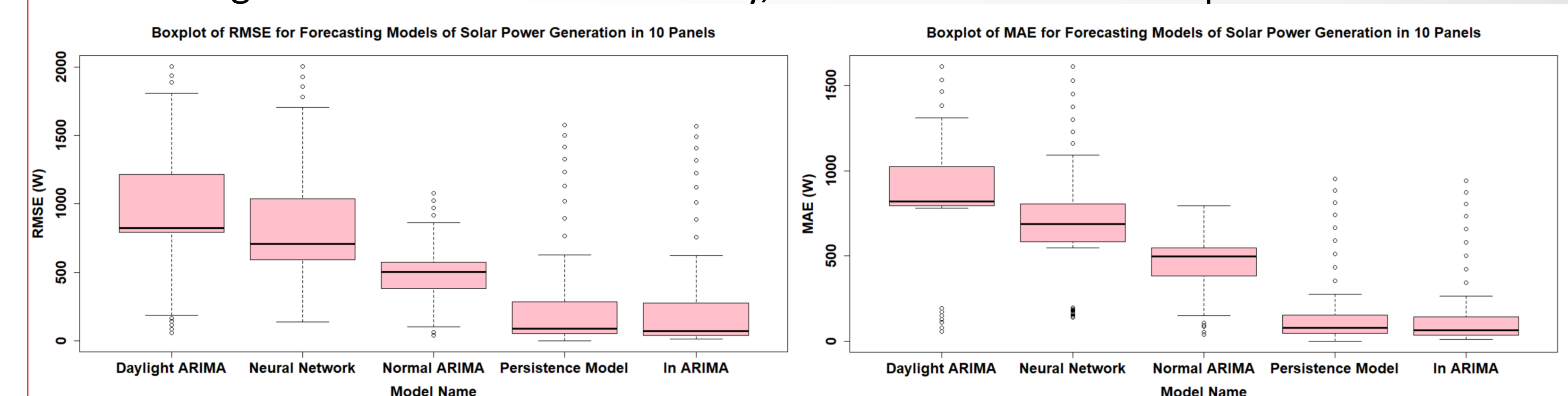
All five models were trained on two months of data and were tasked with producing forecasts of 6 hours into the future (in 10-min resolution). Training and testing data was repeated over 72 forecasting rolls, each shifted by 10 minutes.

Irradiance-to-power conversion: An array of ten solar panels. The power values used above were not measured, but were calculated by multiplying the irradiance data with assumed details of 25% efficiency, a 1 by 1.5 square meter area, and a total of ten in total.

Results



Boxplots of RMSE and MAE values respectively for all 5 methods across 72 rolls of forecasting solar irradiance in Piscataway, NJ. The ln ARIMA model performs best.



Boxplots of RMSE and MAE values respectively for all 5 methods across 72 rolls of forecasting solar power in Piscataway, NJ. The ln ARIMA model performs best

Improvements	in % for Irradiance Forecasting		in % for Power Generation	
ln ARIMA to....	Aggregate (Mean) RMSE Improvement (%)	Aggregate (Mean) MAE Improvement (%)	Aggregate (Mean) RMSE Improvement (%)	Aggregate (Mean) MAE Improvement (%)
Normal ARIMA	96.488	96.428	36.784	59.826
Daylight ARIMA	98.174	98.125	67.138	78.913
ELM Neural Network	97.909	97.751	62.358	74.708
Persistence Forecast	94.645	91.647	3.609	6.046

The table above shows the improvements made by the ln ARIMA model when compared with the other forecasting models. For example, if looking at the "Normal ARIMA" row in the "Improvements in Percents for Irradiance Forecasting" table, the ln ARIMA shows a 96.488% improvement in RMSE when forecasting irradiance in Piscataway, NJ when compared with the Normal ARIMA model.

For irradiance forecasting, the ln ARIMA model showed improvements of over 90% in all statistical metrics against all four of the other forecasting models tested.

For power forecasting, the ln ARIMA model still performed very well with 3.61% and 6.05% mustered by the ln ARIMA against the persistence forecast . The ln ARIMA also drastically outperforms the other forecasting models and the neural network, cementing its position at the top.

Conclusion

The forecasting trials make clear that the ln ARIMA model performs significantly well compared with its compatriots when predicting irradiance and power generation in Piscataway, NJ. This finding can be optimally used by grid operators in diverting their resources and planning ahead on pricing and storage matters. The ln ARIMA was rigorously tested against four other models, outperforming them all, showing its robustness in handling forecasting in Piscataway. Based on the findings, the ln ARIMA is reliable and can be trusted to accurately forecast in the short-term when necessary.

Future Direction

The robustness of the ln ARIMA can be further verified by performing more exhaustive forecasting experiment at different times of the year. Furthermore, integrating different types of data can be attempted, perhaps trying to forecast using irradiance indicators other than GHI, or even attempting to better train machine learning models using exogeneous variables, many of which are available online through satellite-based providers such as Solcast. The future is bright for solar energy, as the Sun has a long way to burning out, unlike fossil fuels.