Institute for Artificial Intelligence
Franklin College of Arts and Sciences
UNIVERSITY OF GEORGIA

SOLAR IRRADIANCE PREDICTION USING DISTRIBUTED MACHINE LEARNING TECHNIQUES

Aashish Yadavally, Anirudh K. M. Kakarlapudi, Jayant Parashar, Frederick Maier, Khaled Rasheed, Shannon Quinn

INTRODUCTION

Solar power is one of the fastest-growing and most promising sources of renewable energy. The output of a solar farm is heavily dependent on the angle and intensity of the sunlight that strikes the panel, which is in turn affected by existing environmental conditions. Because of this external dependency, it is difficult to forecast solar energy production without an accurate predictor of solar radiation.

In this work, with the help of the weather prediction models such as North America Mesoscale (NAM) Forecast System and the data from the solar farm in University of Georgia, we compare several machine learning models which attempt to predict observed solar radiation in a 1-24 target offset hour future window across varying grid sizes such as (1, 1), (3, 3) and (5, 5), which represents the grid of 12 sq.km cells around the target location.

DATASET

The dataset has two components:

1. North America Mesoscale (NAM) Forecast Data

- One of the major weather models run by National Centers for Environmental Prediction (NCEP)
- NAM-CONUS data has a spatial resolution of approximately 12 km^2, and is released every six hours, at a one-hour temporal resolution.
- Available in two variants:
  (a) 2D Surface Fields
  (b) **3D Pressure Fields**

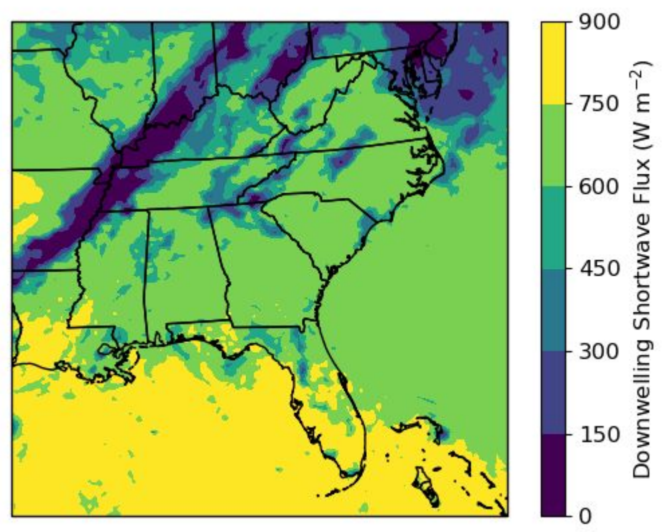

Figure 2: Geographic region encompassed by our NAM dataset

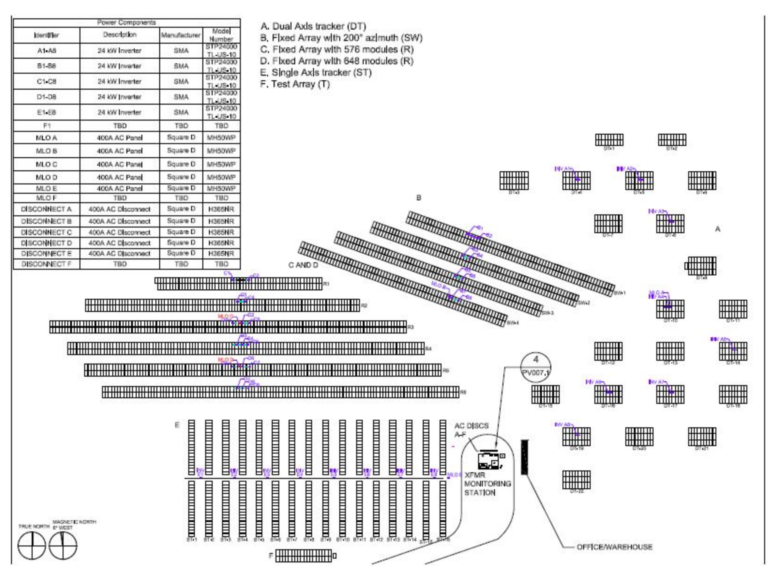

Figure 1: UGA Solar Farm Site

2. UGA Solar Farm Data

- UGA partnered with Georgia Power to set up a 1-MW solar array using a variety of single-axis, dual-axis, fixed and tracking pyranometers.

NAM forecasts between January 1st, 2017 and December 31st, 2017 occupying about 2.3TB of disk space were used for this work.

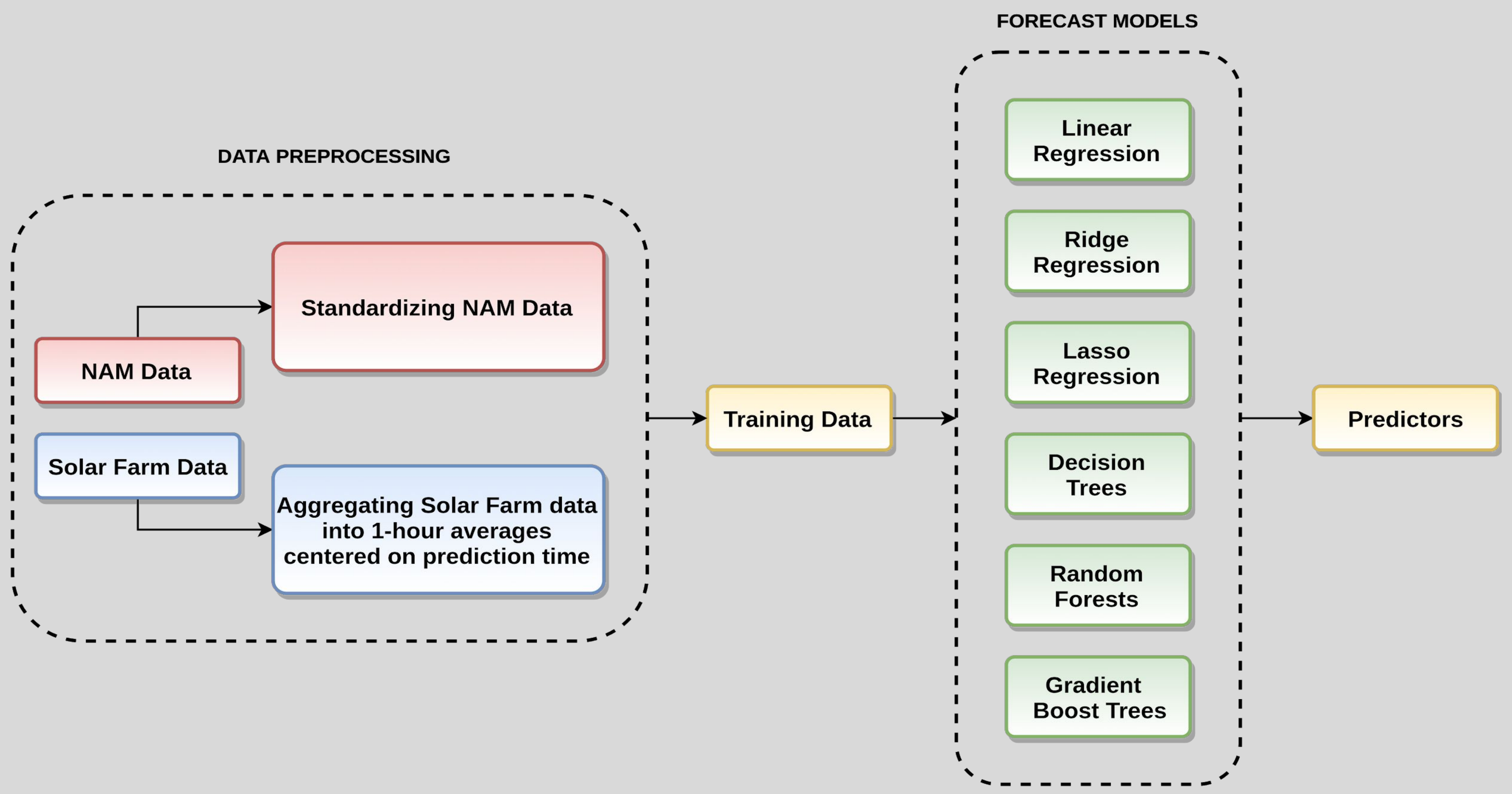

Figure 3: Prediction Generation Pipeline

METHODOLOGY

Models were trained with features from NAM forecasts and ground observations from UGA Solar Farm's fixed array (Array B) pyranometer, with a 80:20 train-test split.

Machine learning provides various powerful techniques for developing sophisticated regression models.

Models used for training in a distributed fashion

- Linear Regression
- Ridge Regression
- Lasso Regression
- Decision Tree
- Random Forests
- Gradient Boost Trees

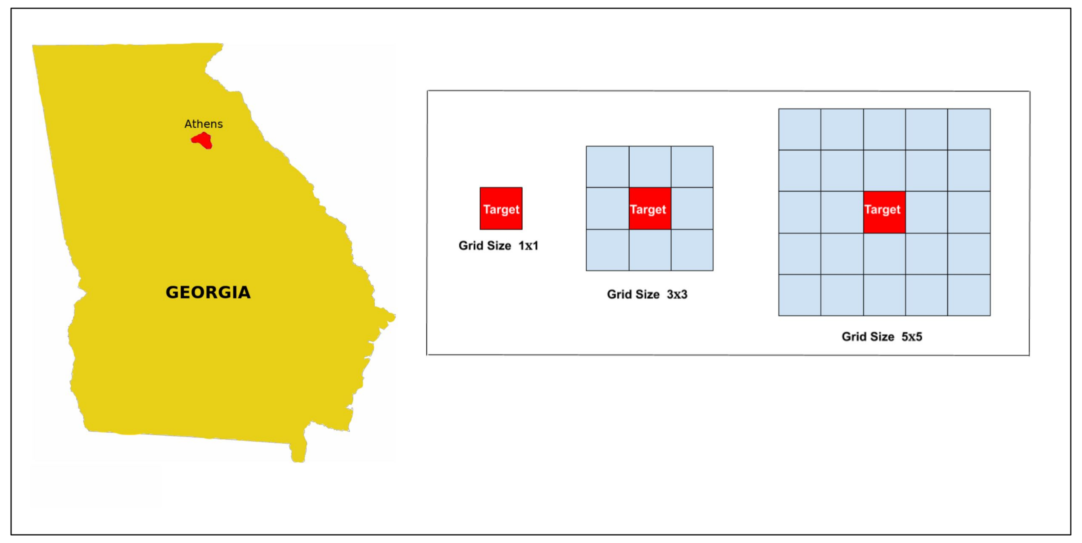

Figure 4: Grid of cells around Athens

NAM data from the grid containing Athens was used in three capacities, with a 1x1 grid size, 3x3 grid size and 5x5 grid size.

TECHNOLOGIES USED

Google Cloud Platform, python, pvlib, Spark MLlib

RESULTS

Evaluation metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R^2 were used for tracking the performance of the models across different grid sizes.

Random Forests generally performed better across different grid sizes, for majority of the target offset hours.

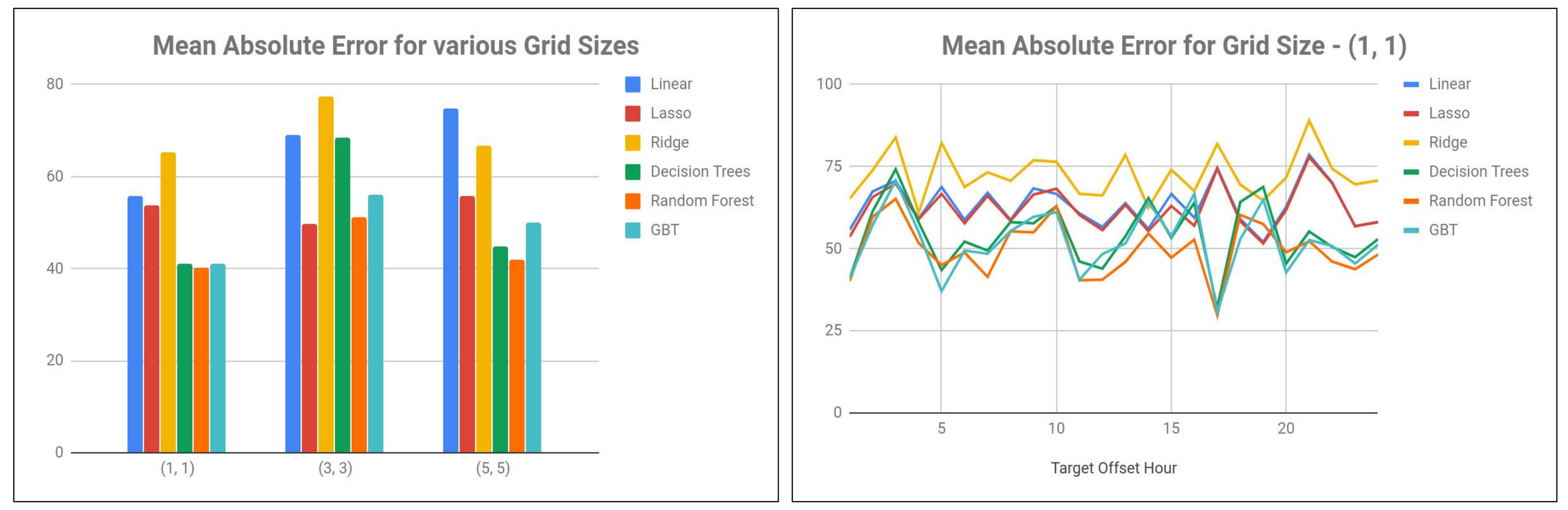

Figure 5: (left) MAE Performance of models for various grid sizes
(right) MAE Performance of models for various grid sizes across 1 - 24 target hours

For the 1-hour ahead models, random forests perform best for (1,1), (3,3) and (5,5) grid sizes, with MAEs of 40.18, 49.81 and 41.93 respectively. The results generally align with the trends found by Jones in [2], for whom kNN and Random Forests performed best, while he trained on 2017 data and tested on 2018 data.

FUTURE WORK

Mathiesen et.al [3] found that NAM models systematically overforecast GHI during clear sky conditions by up to 40%. PVLIB-Python [4] contains implementations to convert the data from NAM models to irradiance metrics using the physical models, providing scope for better model development.

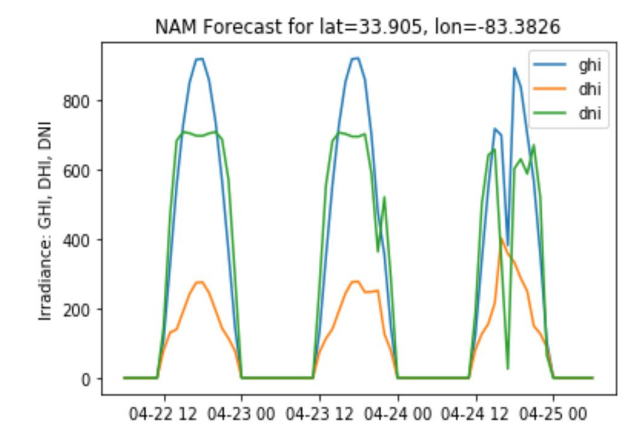

Figure X: PVLIB Irradiance Forecast

REFERENCES

1. Sam Sanders et al. "Solar Radiation Prediction Improvement Using Weather Forecasts". In: Machine Learning and Applications (ICMLA), 2017 16th IEEE International Conference on. IEEE. 2017, pp. 499–504.
2. Jones Z. (2018). "An Analysis of Machine Learning for Solar Irradiance Forecasting" (Unpublished Master's Thesis). Institute for Artificial Intelligence, University of Georgia, Athens, GA.
3. Mathiesen, Patrick, Craig Collier, and Jan Kleissl. "A high-resolution, cloud-assimilating numerical weather prediction model for solar irradiance forecasting." *Solar Energy* 92 (2013): 47-61.
4. https://github.com/pvlib/pvlib-python