Accurate Solar Energy Generation Predictions Using Weather Forecasts

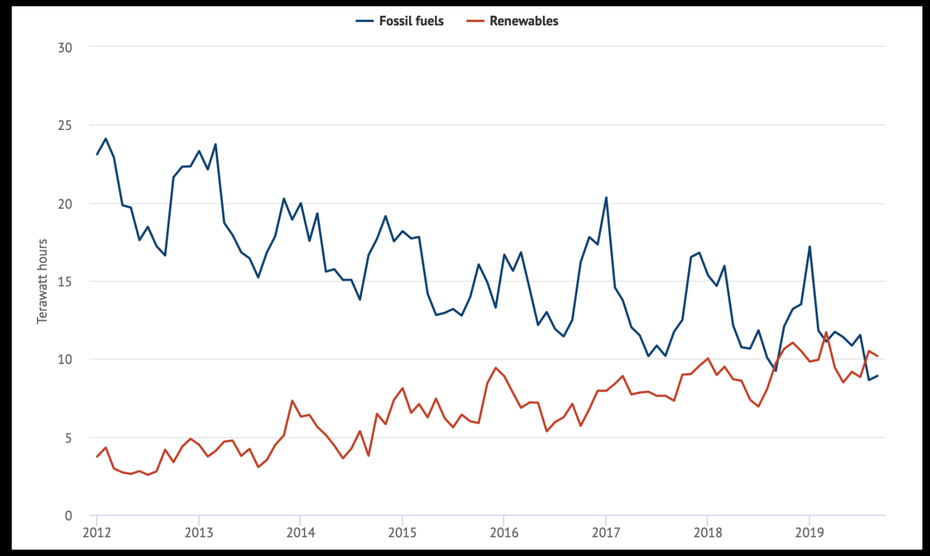
DATA 698 - CUNY SPS

Jose A. Mawyin - December 16, 2020

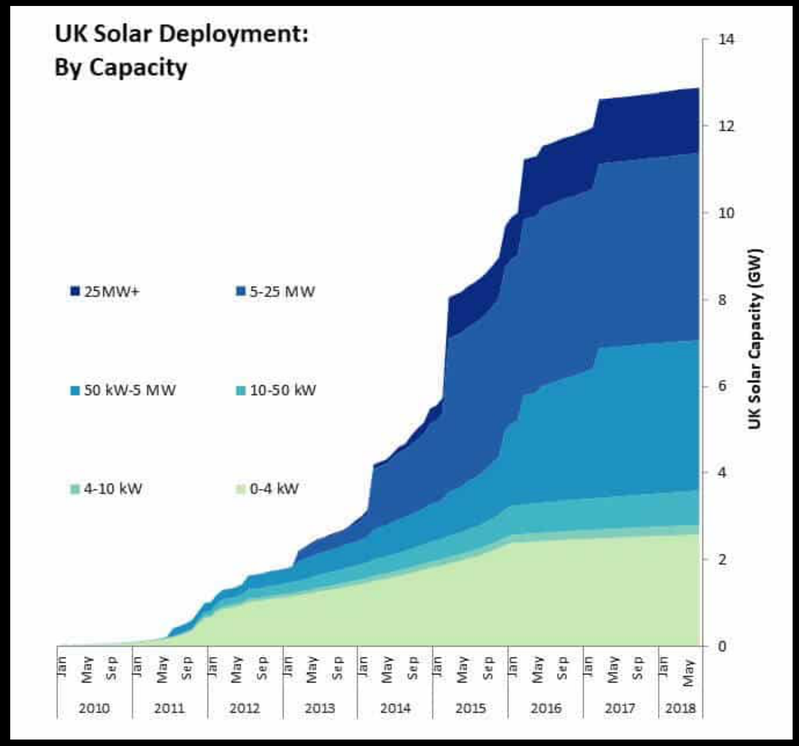
Problem Statement

How can the country make better use of a variable resource?

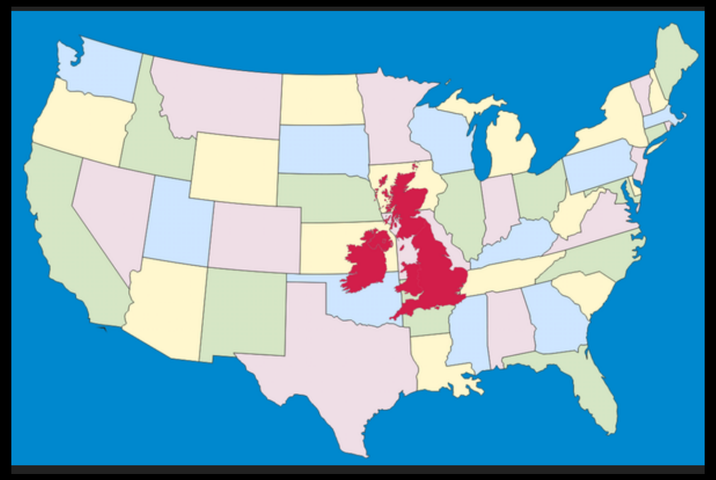
- In the UK, there is enough solar panels deployed to generate up to 6% of the total electricity demand of the country.
- Solar generation ranges from 0% to 30% depending of the time of day and from day to day.



https://www.carbonbrief.org/analysis-uk-renewables-generate-more-electricity-than-fossil-fuels-for-first-time



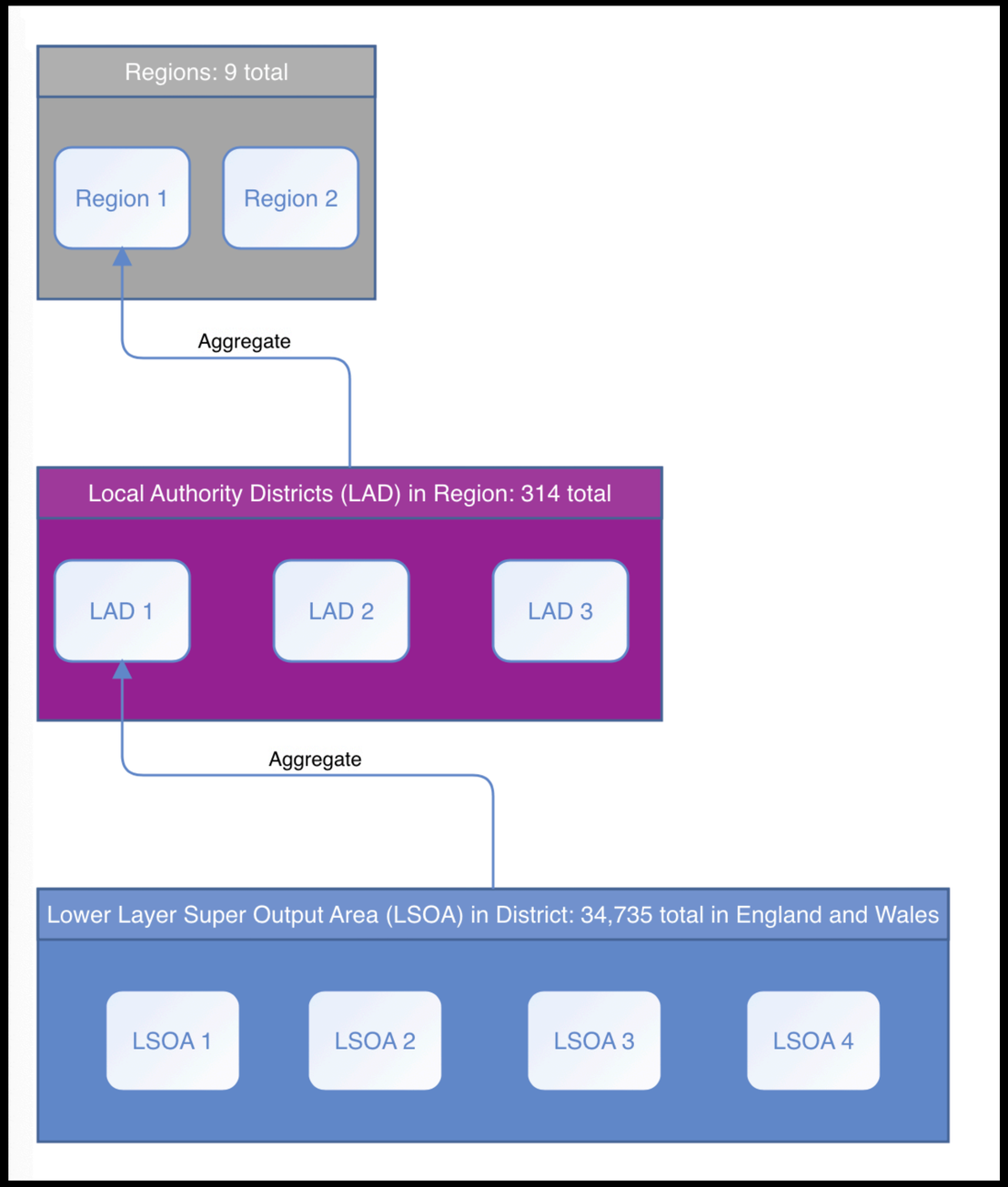
https://ukbusinessenergy.co.uk/uk-solar-pv-market-report/



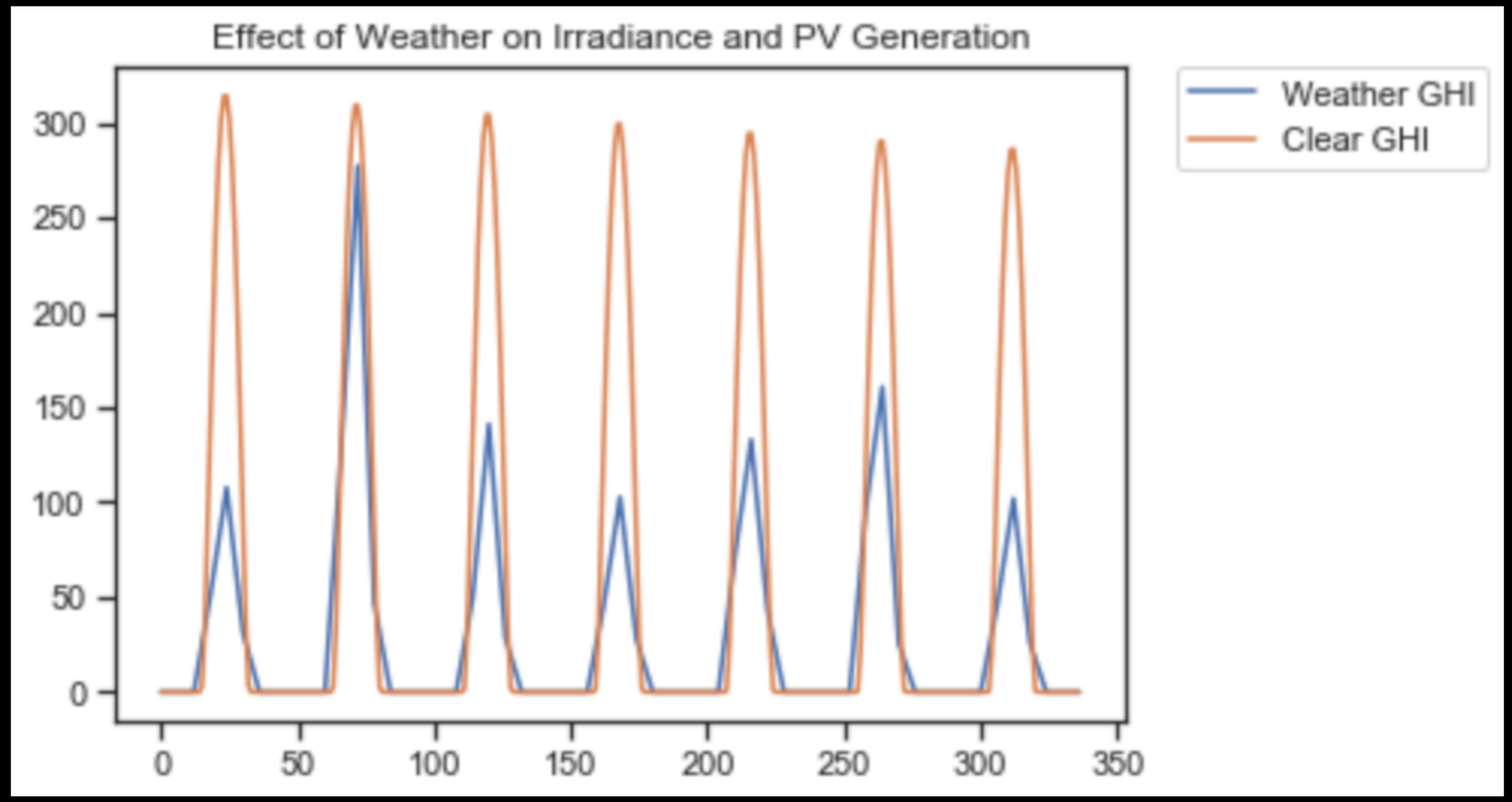
Geographic Levels of Details

Based on human statistical geography:

Output Area —> Local Area District —> Region



Effect of Weather on Irradiance and PV Generation



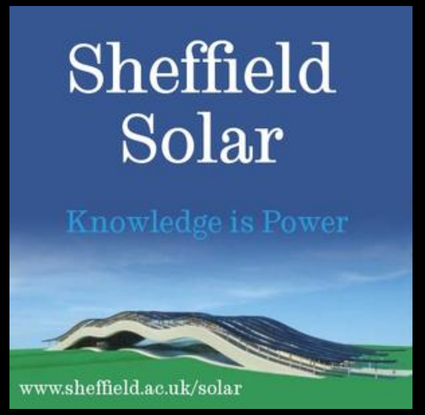
Data

Office for National Statistics: Population Density, LSOA boundaries, electricity demand, # of meters, District boundaries and centroid.

Ofgem: Number of deployed photovoltaic systems and DC capacity.

NOAA: Hourly weather conditions at District centroid.

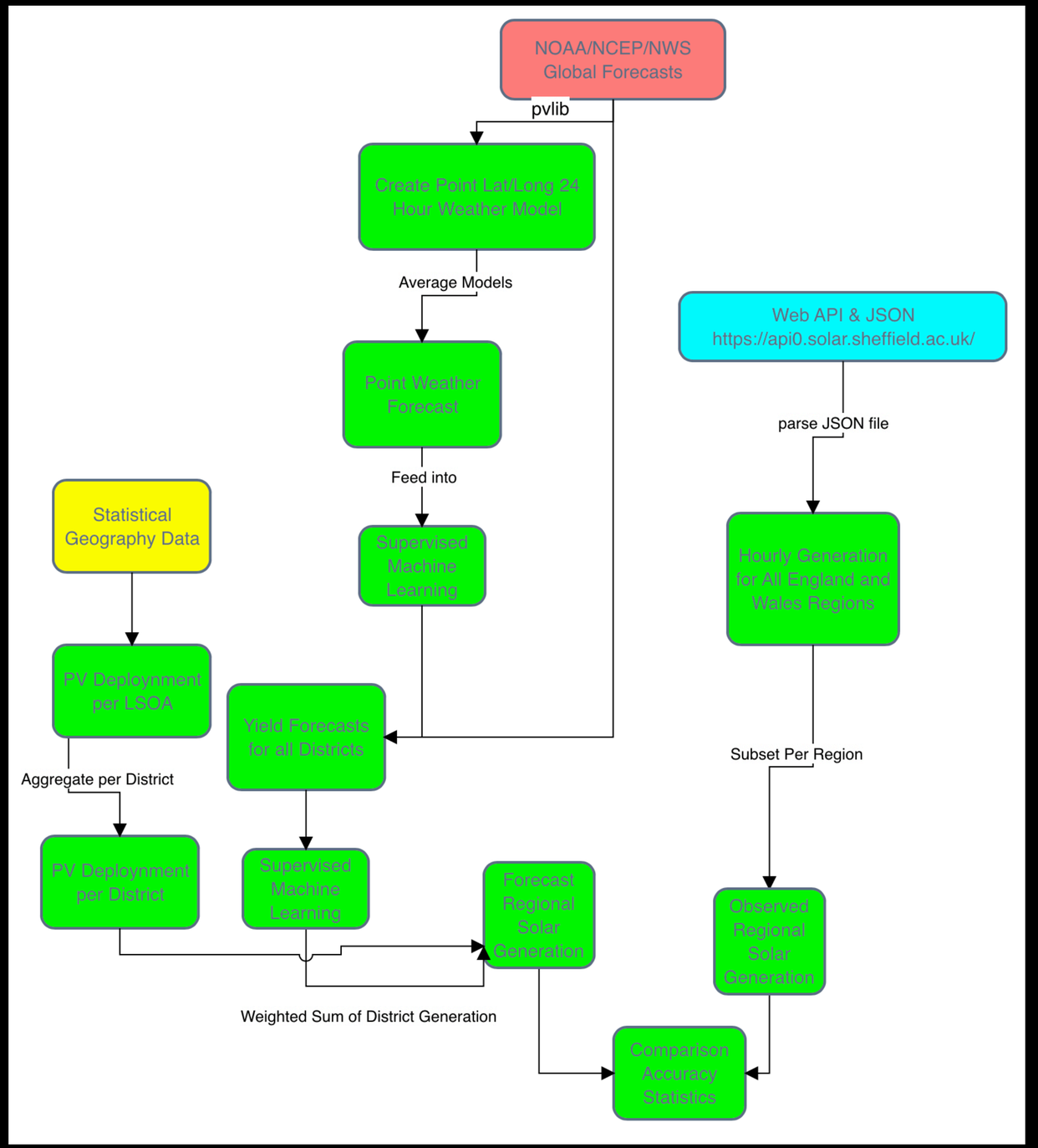
Sheffield Solar Farm: Regional hourly generation and yield.



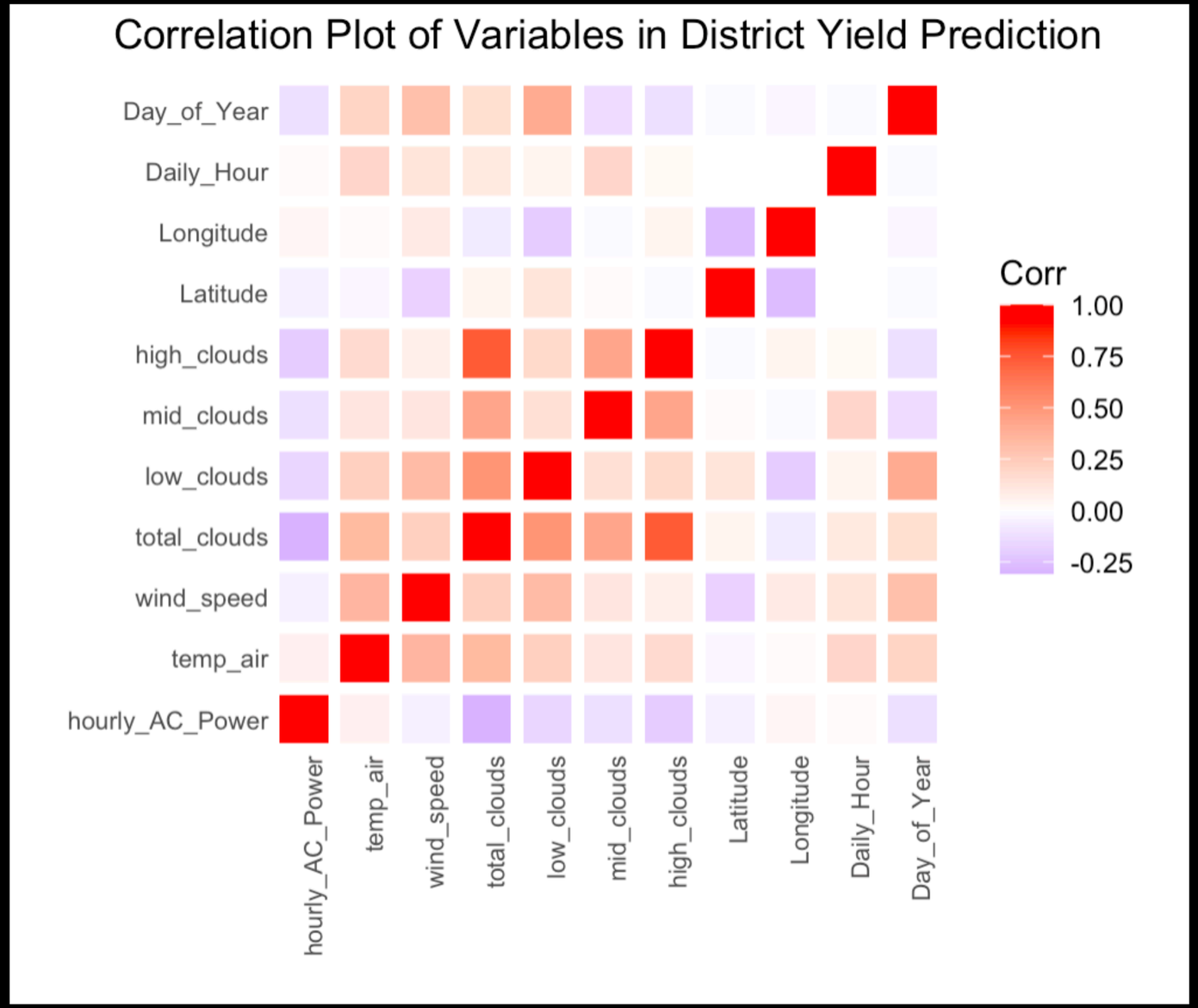
Machine Learning for Electricity Forecasting

- Observations:
  - From 331 geographical Areas
  - 7 days of observations with a 30 minute temporal resolution
- The variables used in the model are:
  - From location: Latitude, Longitude
  - From calculations: Global Horizontal Irradiance - Extraterrestrial
  - From weather report: Hour of the Day, Day of the Year, Air Temperature, Wind Speed, Cloud Coverage at Different Heights and Total Cloud Coverage
- The following linear, non-linear and tree-based models were trained:
  - Linear Regression:
  - k - Nearest Neighbors:
  - Random Forest:

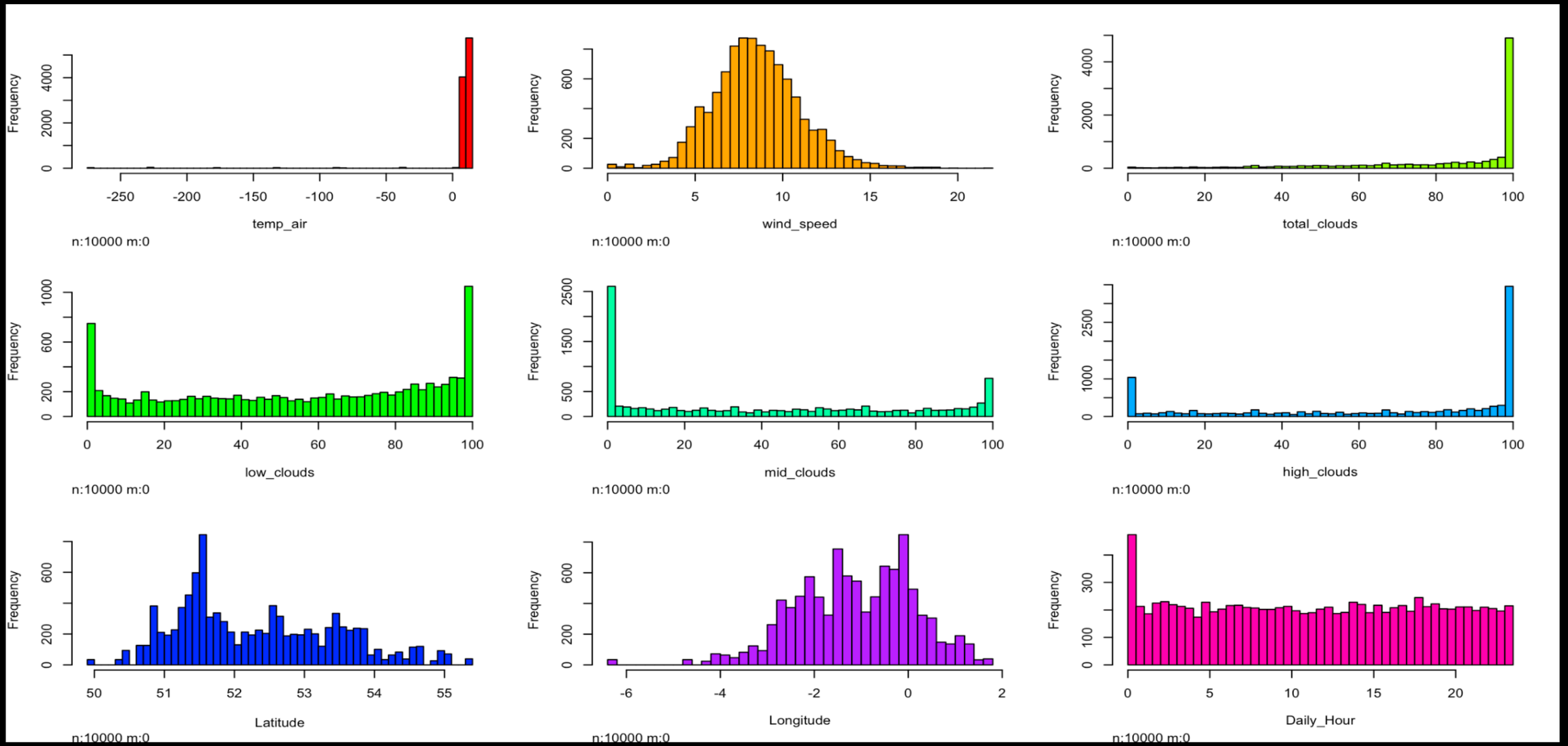
Workflow to Process Information



Data Exploration



Data Exploration

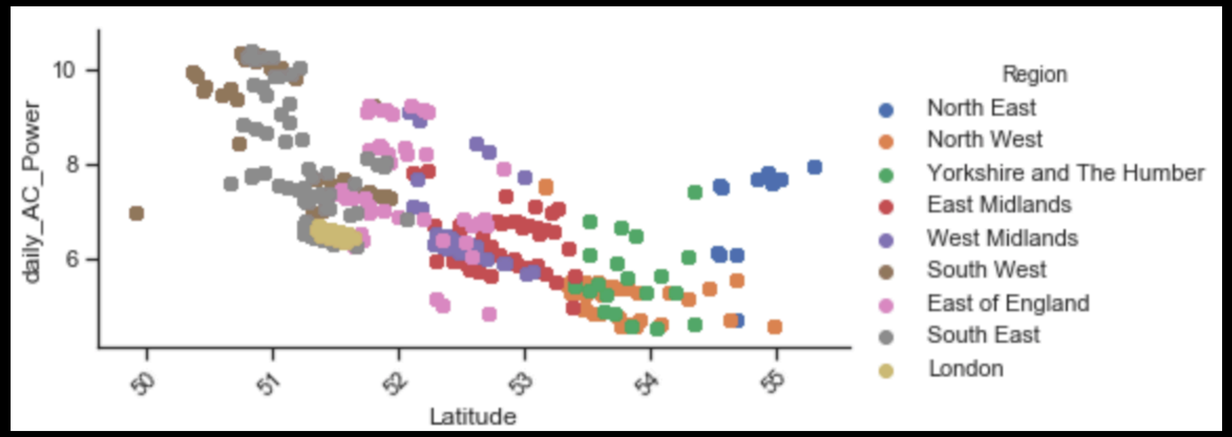


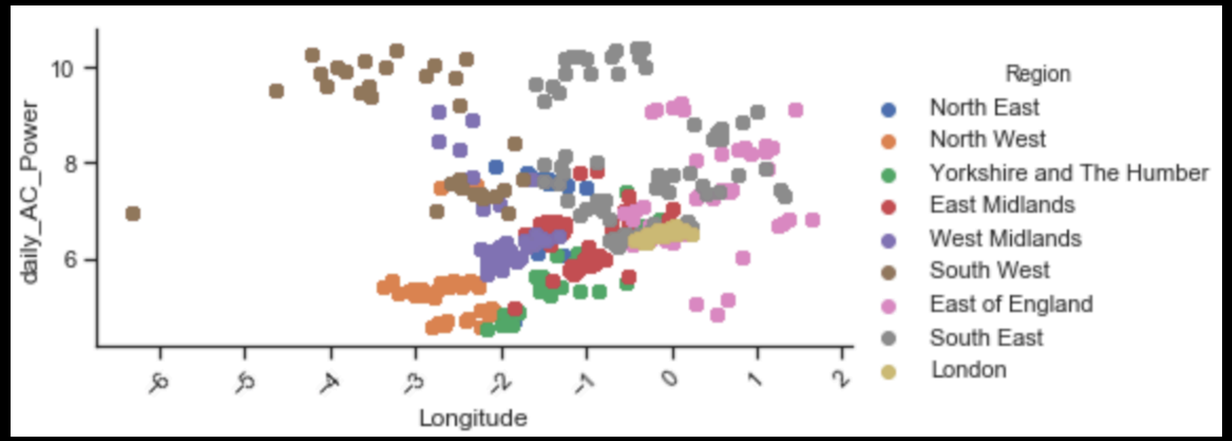
Data Exploration

Yield dependence on geographical location

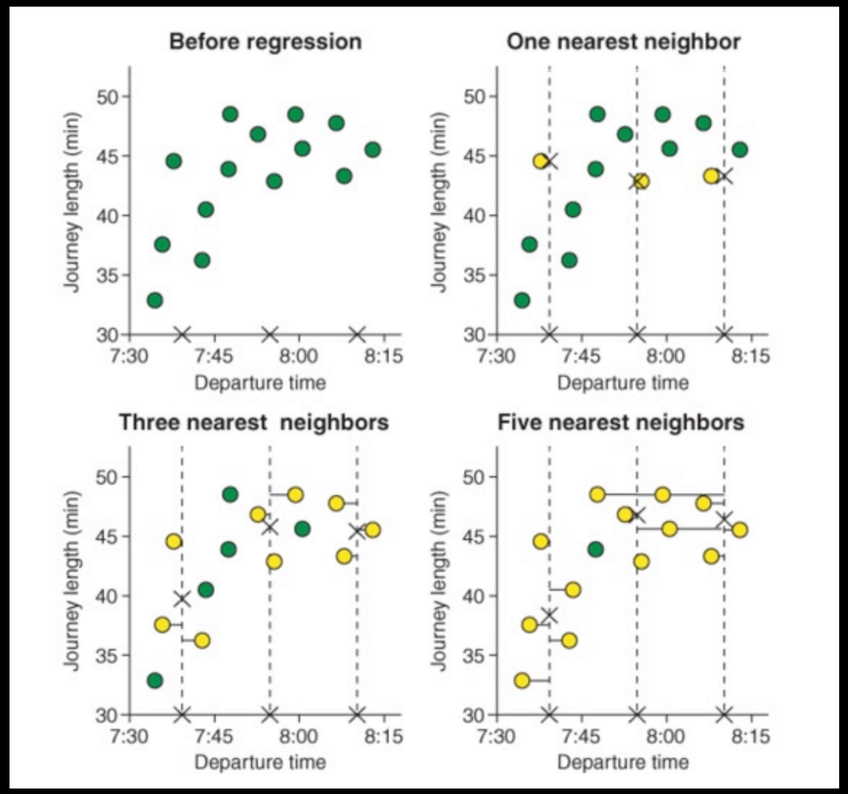
Latitude: South —> North

Longitude: West —> East

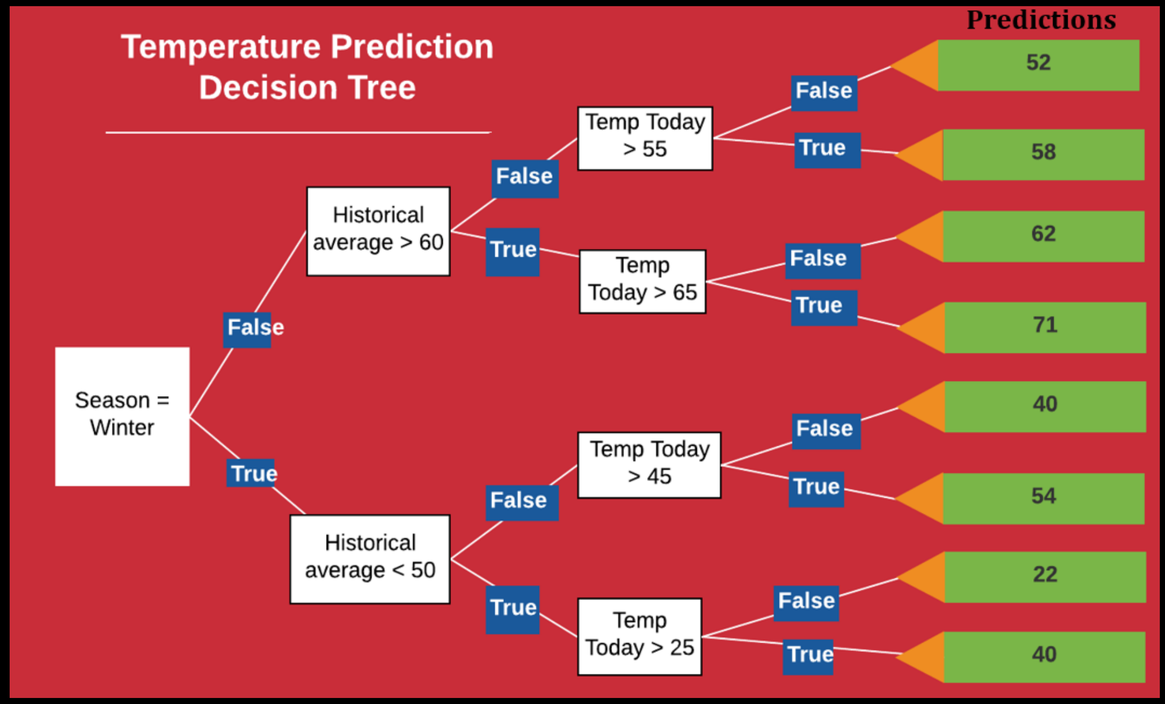




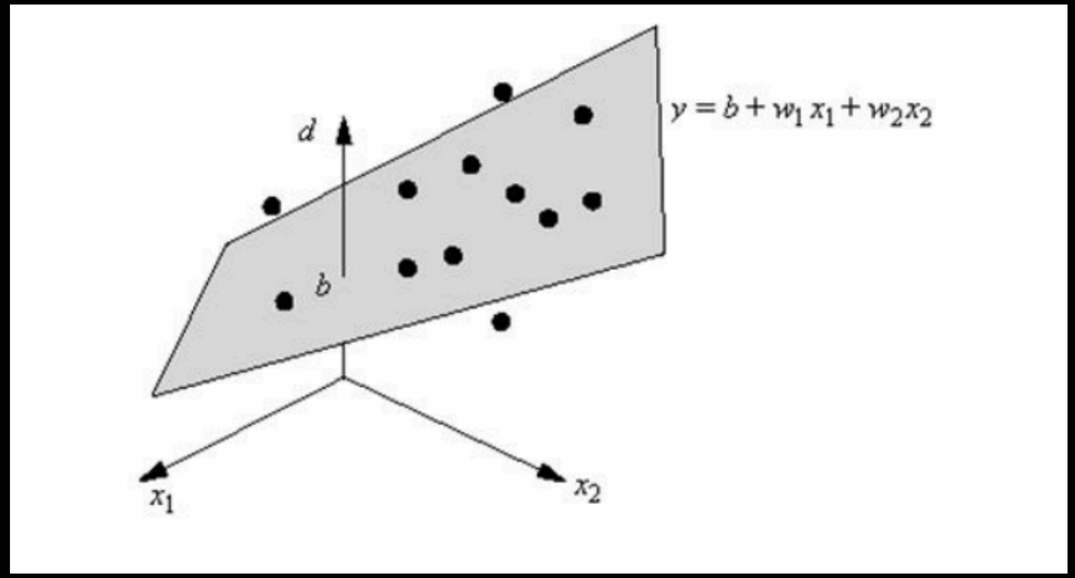
Supervised Machine Learning Algorithms



k Nearest Neighbor



Random Forest



Linear Regression

Model Performance

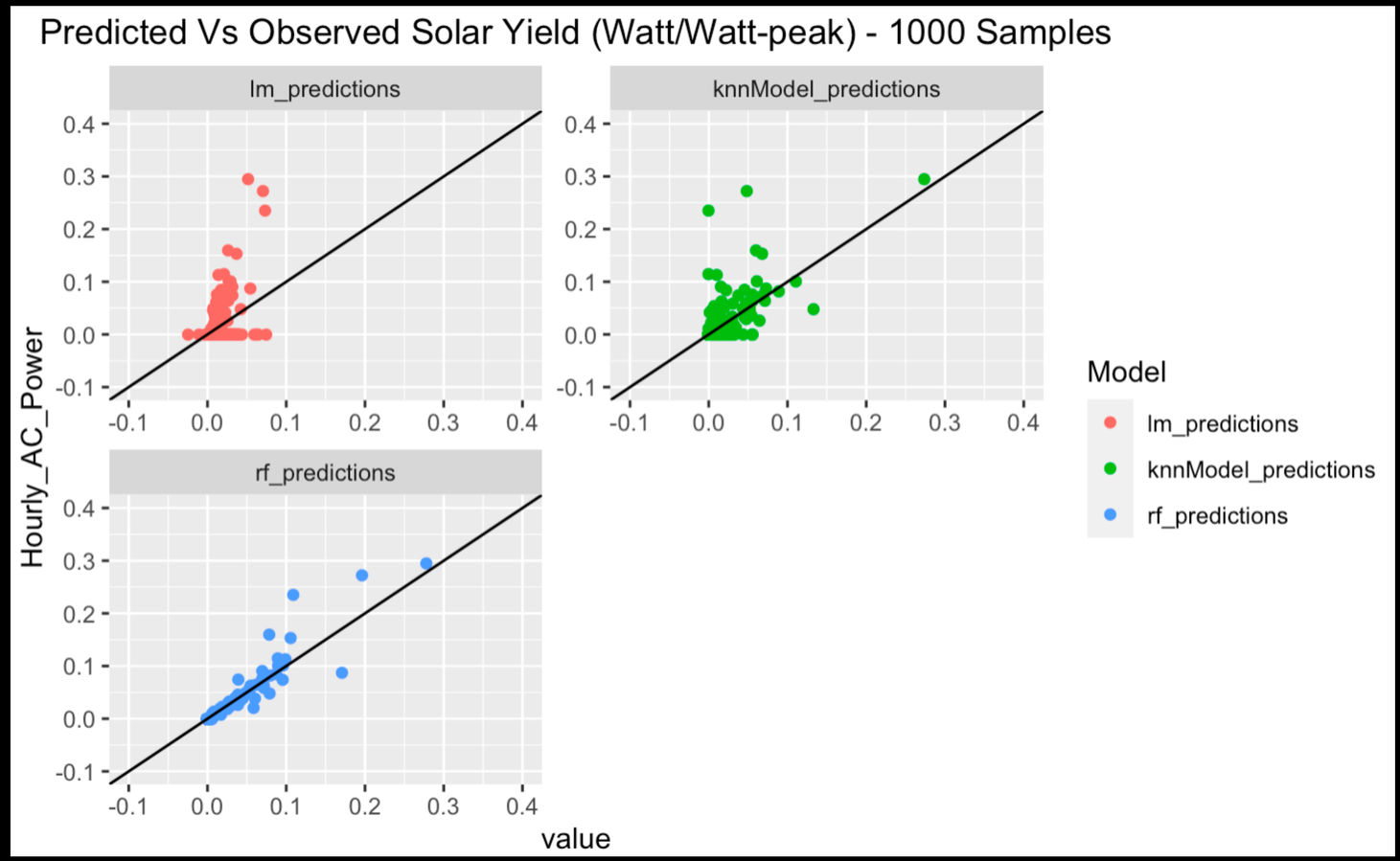
1000 Samples

```
Linear Regression training time:
0.667 sec elapsed

knn training time:
1.488 sec elapsed

Random Forest training time:
227.694 sec elapsed
```

Model <chr>	RMSE <dbl>	Rsquare <dbl>
Linear Regression	0.03955913	0.1399861
k-Nearest Neighbors	0.03229901	0.4289623
Random Forest Grid	0.01504436	0.8834463



3.8 minutes for Random Forest training @ 1000 samples

Model Performance

10000 Samples

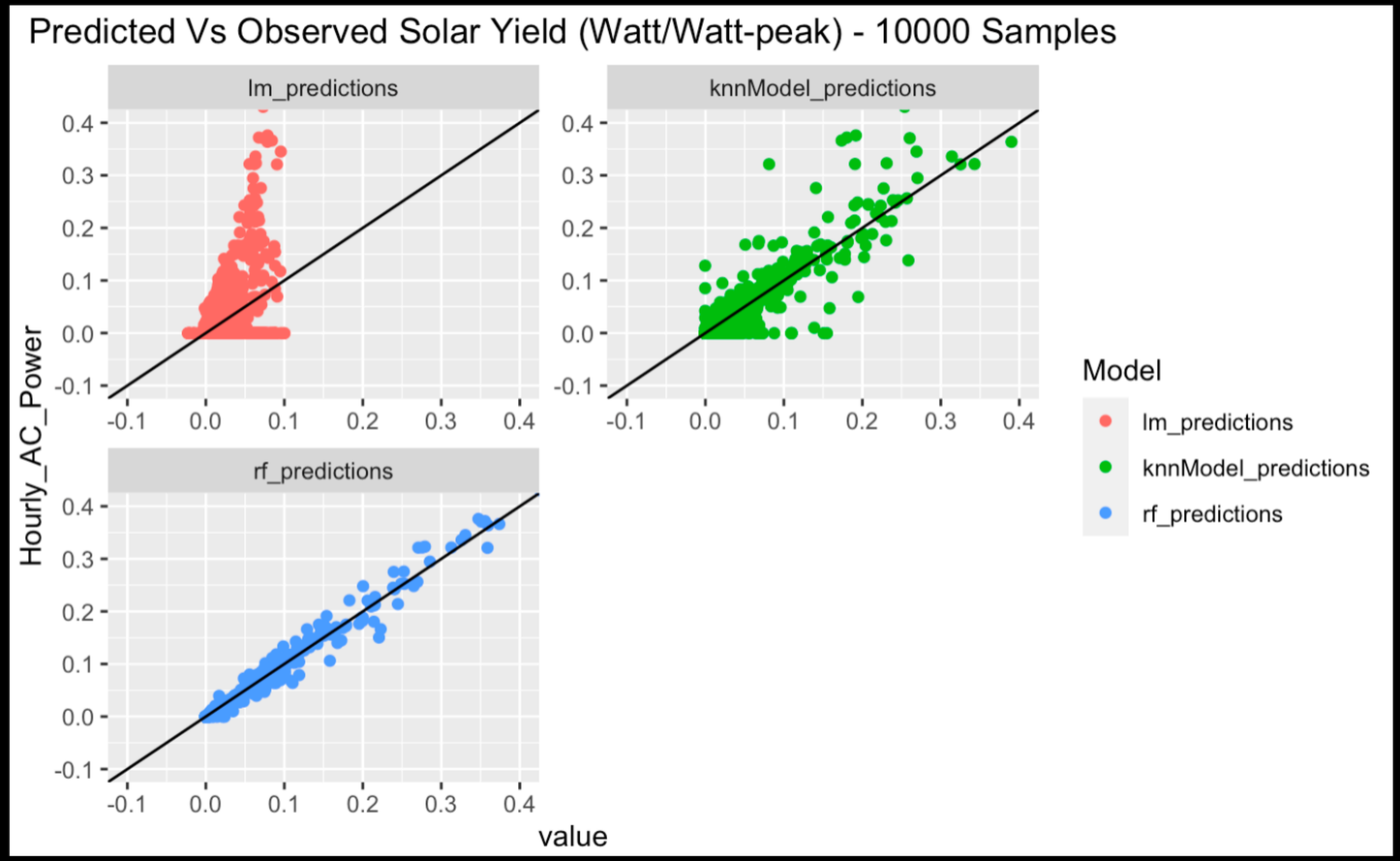
```
Linear Regression training time:
1.17 sec elapsed

knn training time:
24.579 sec elapsed

Random Forest training time:
4027.059 sec elapsed
```

**Using 4 core parallel processing in a desktop computer*

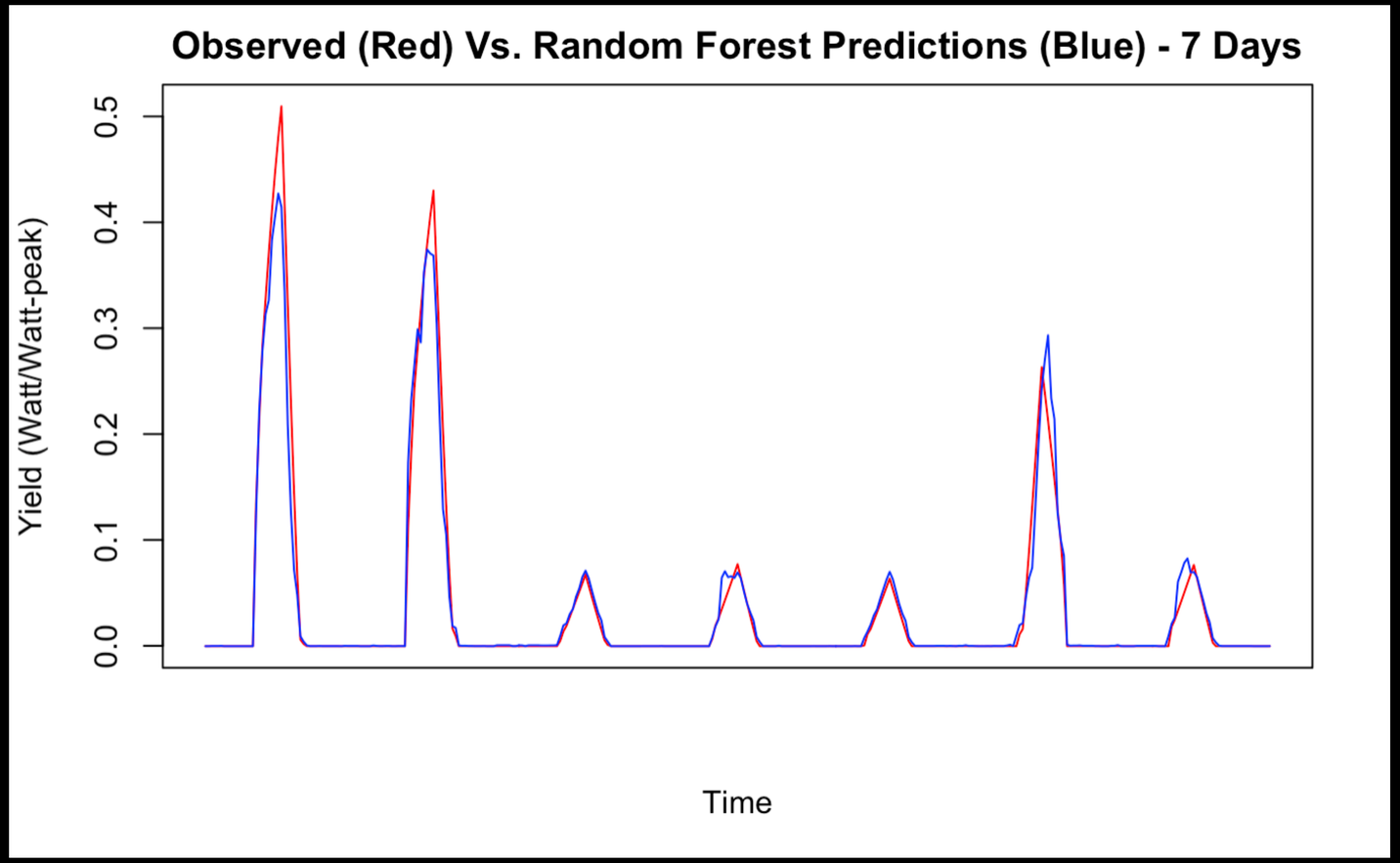
Model <chr>	RMSE <dbl>	Rsquare <dbl>
Linear Regression	0.049338757	0.1460002
k-Nearest Neighbors	0.020040717	0.8617034
Random Forest Grid	0.006896239	0.9843703

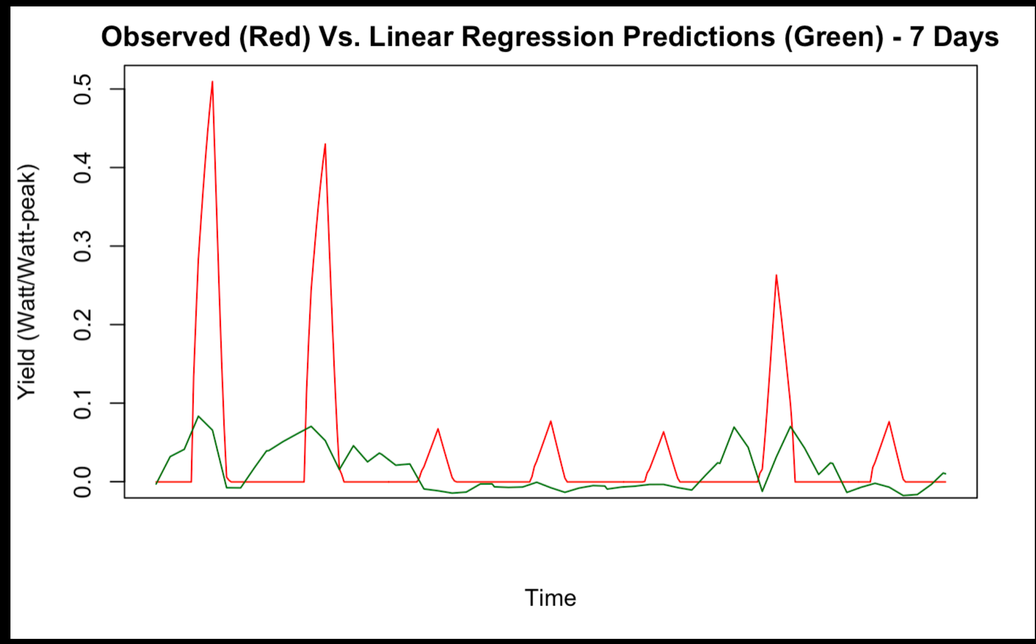


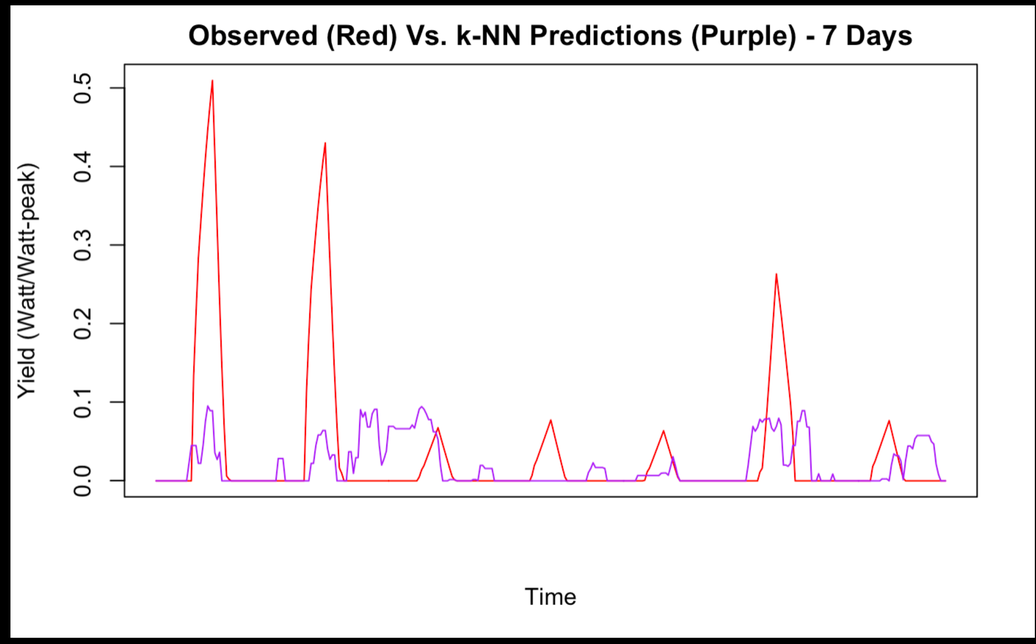
167 minutes for Random Forest training @ 1000 samples
10 times sample size - 43 times the training time increase!

Model Performance

10000 Samples





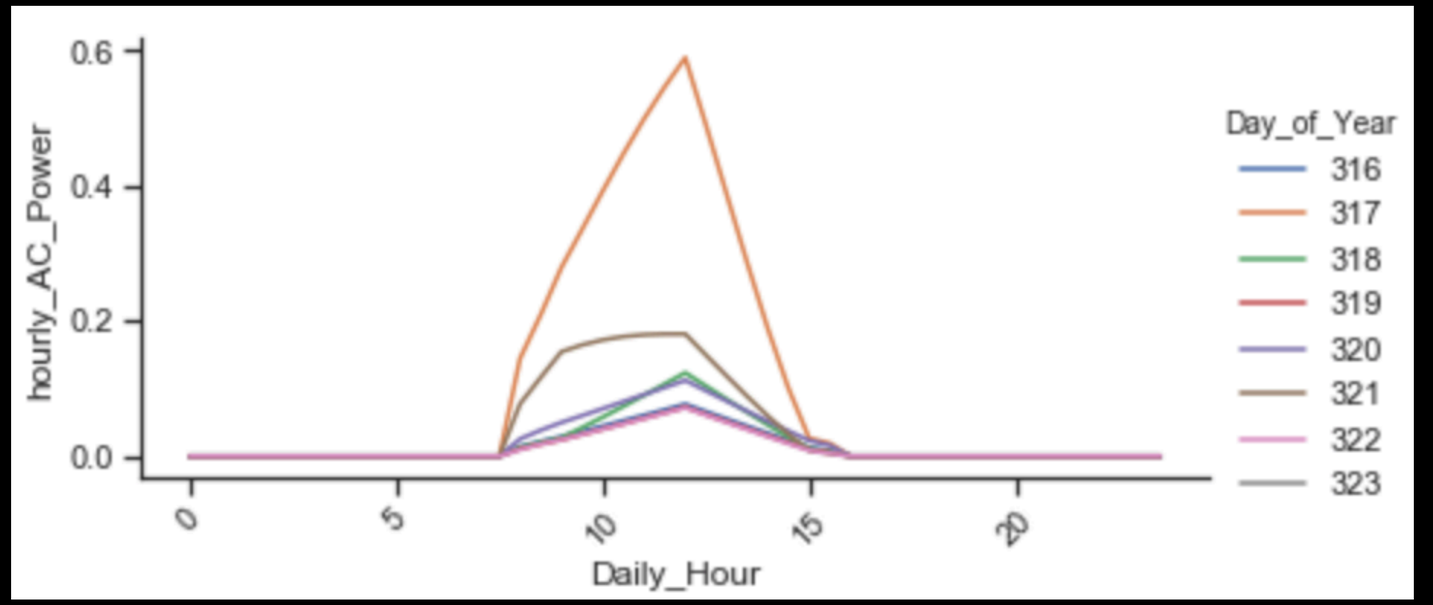


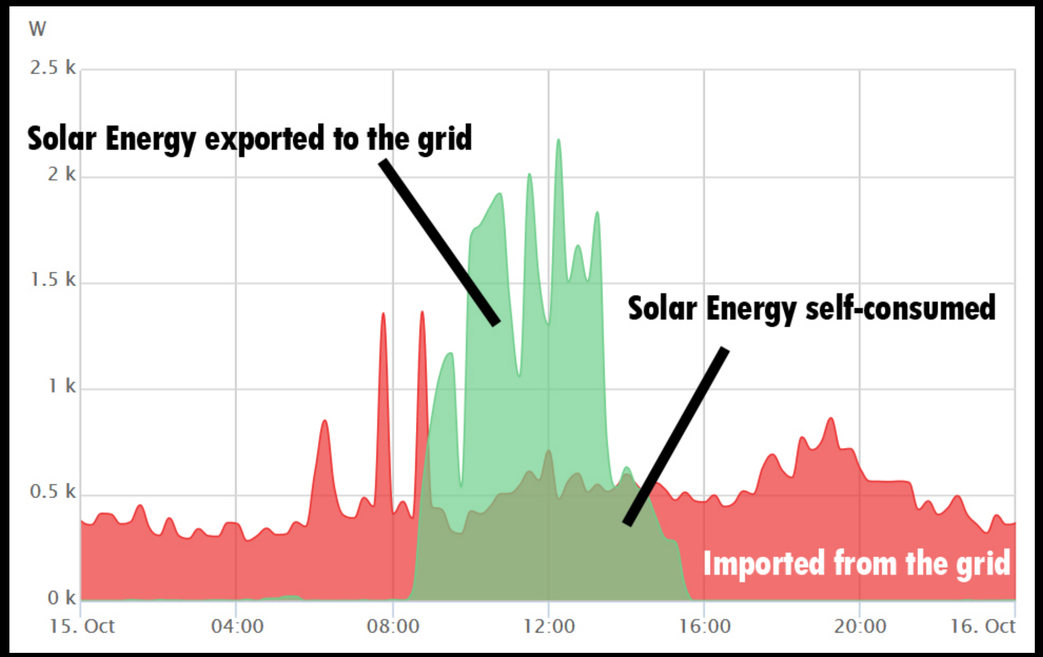
Forecasting benefit for the home owner

- How much energy is my roof top array going to produce tomorrow and the day after?
- When should I schedule energy intensive uses?
- When is the best time to sell or buy electricity?

When and how much energy?

Variable generation but dependable with accurate forecasting





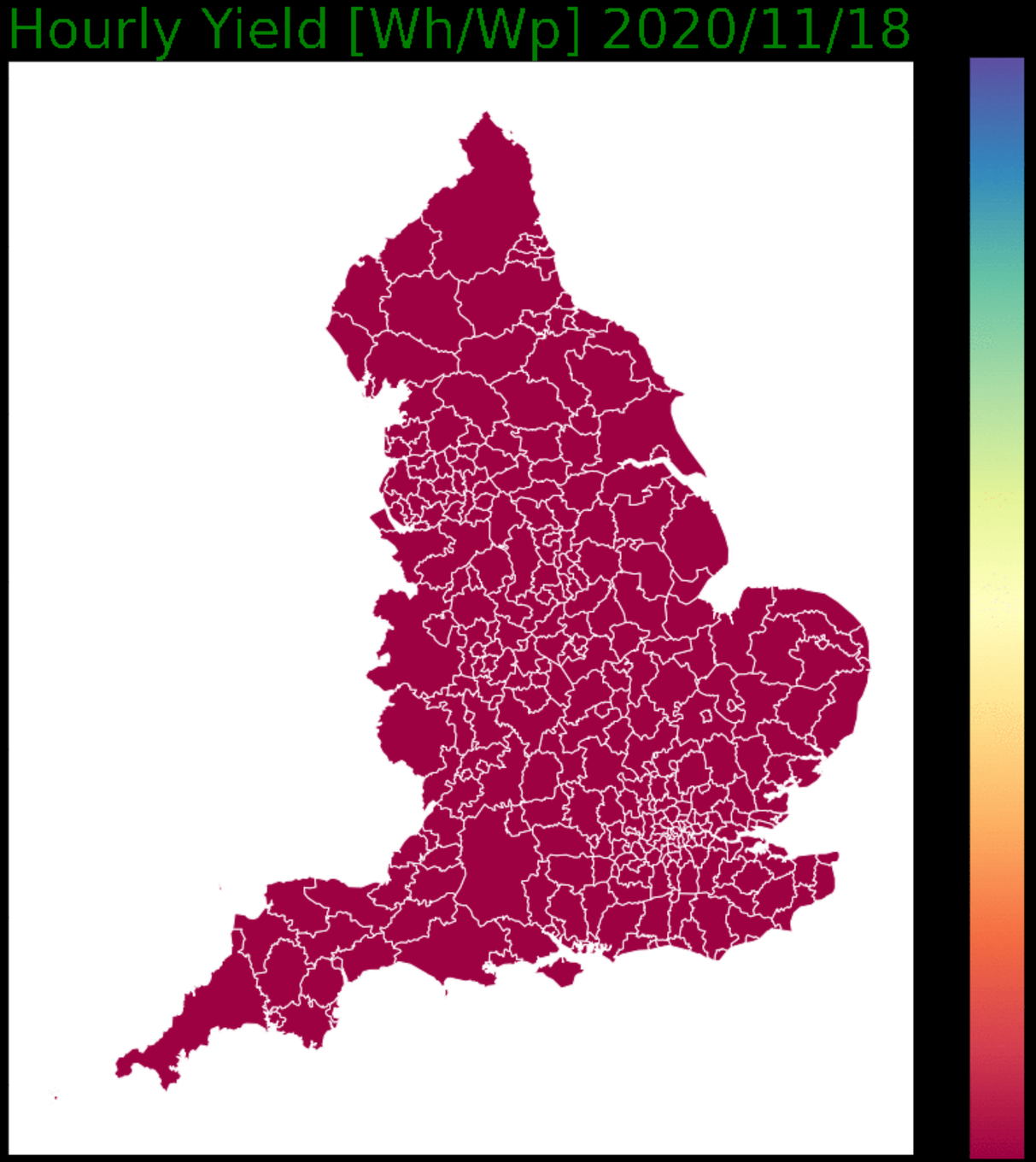
https://www.exeoenergy.co.uk/solar-panels/

Forecasting benefit for the energy provider

- Given limited resources, which location will produce the highest return on investment?
- Which location offers the most stable generation in a given time period?
- How to balance renewable and non-renewable generation?

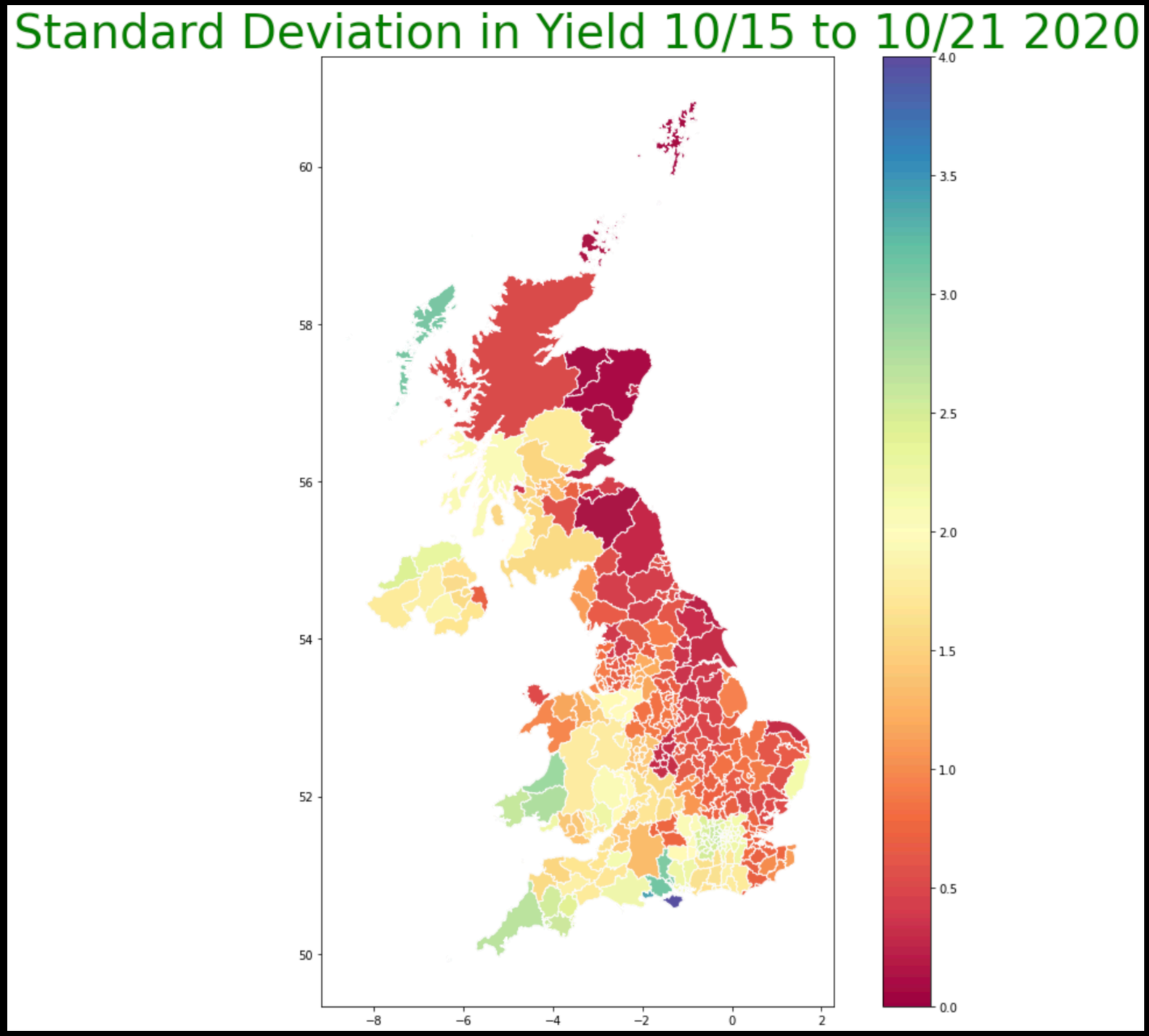
District Forecasting

Variability in Generation in Time and Space



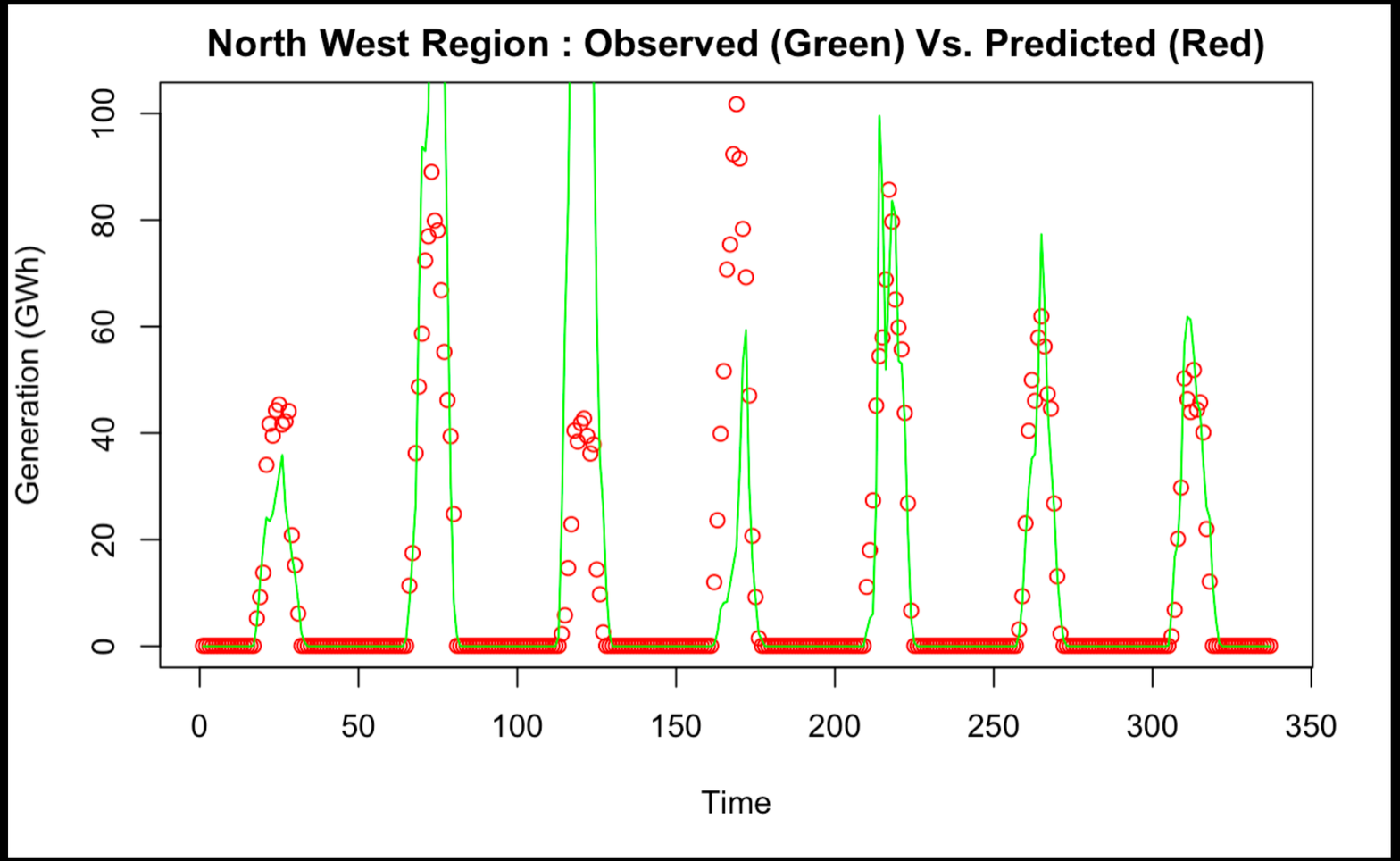
District Forecasting

Where to deploy for stable generation



Regional Forecasting

Random Forest model trained using Regional Generation (MWh) and District Yield (Wh/Wp)



Final Thoughts

- Shown predictive power of supervised machine learning
- Results useful to homeowners, local and national level planners.
- Extend model to whole year:
  - District yield accurate
  - Regional generation needs improvement
- Work beyond initial scope: Open up results to public through App - Tableau

Things that I learned during this project:

- Working with R and Python
- Data wrangling: filtering, reshaping and joining
- API’s and JSON
- Piping data from Python libraries
- Automating machine learning models
- Visualizations
- Patience