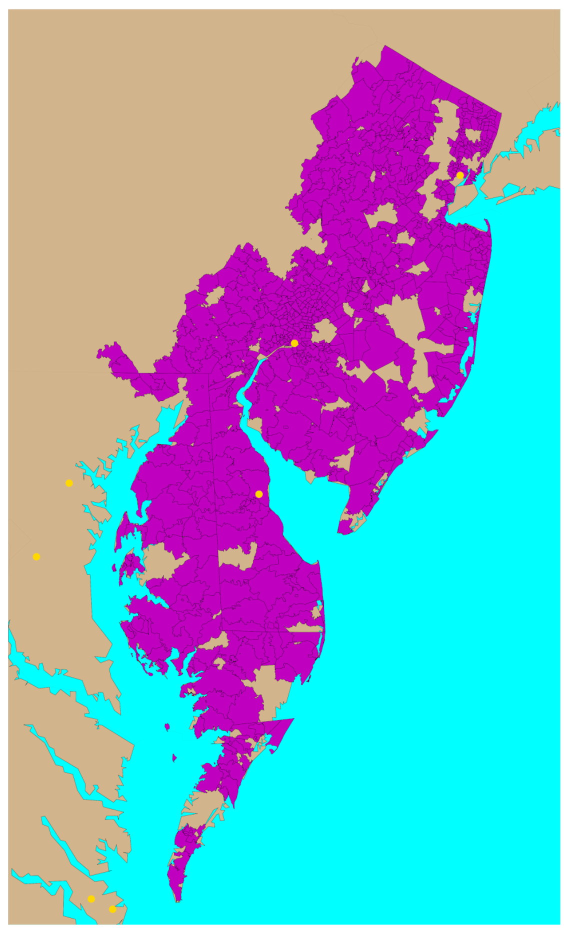
Final Capstone Proposal
Mitch Shiles

Effects of Weather on Power Consumption and Production

With projections for climate change becoming more dire year over year, the need for the world's economies to rapidly switch over to carbon-free sources of energy is more necessary than ever before. Renewable energy sources such as wind and solar are becoming cheaper and more cost-effective to install. However, one problem that continues to arise with these sources of energy, is intermittency, which has resulted in a common critique that the fossil fuel industry voices against renewables is "What happens when the wind stops blowing and the sun goes down?". Renewable energy is directly tied to weather conditions, for example: wind speeds affect turbines, cloud conditions affect solar, and annual precipitation can affect hydroelectric generators.

In addition to electric production by renewables, electrical power consumption is inextricably connected to the weather. Air conditioning and space heating are some of the most energy-intensive processes that rely on the electric grid and are both affected by temperature and humidity, while seasonal daylight fluctuations affect lighting usage. Extreme weather events also affect the functionality of the grid, intense storms can snap limbs, uproot trees, and topple power lines resulting in outages. The ability to accurately and robustly forecast both energy production and consumption with weather data will be critical to coordinating energy production between renewable and nonrenewable energy sources. In doing so this would allow grid operators to extend the capacity for a renewable electric generation, and reduce the near term demand for additional grid-level electricity storage.



This project will attempt to discern how meteorological phenomena affect the Eastern Mid Atlantic Area Grid as seen in the image to the right, which includes New Jersey, Delaware, Eastern Pennsylvania, along with the Eastern Shore of Maryland and Virginia. The map to the right has been created by cross-referencing two lists, one of zip codes and one of electrical nodes. In addition, fourteen years of hourly weather data has been scraped from three major airports and one minor airport in the region. Hourly power consumption with aggregate weather data for the region will be modeled with vector autoregression and multi-variate long short term memory (LSTM). Models will be tested for both accuracy and robustness.

The data sets will be explored with unsupervised clustering to search for outliers and patterns in human behavior. It will also be investigated by plotting weather data against power consumption data while searching for cross-correlations. Each data set will be explored for seasonality and trends to be removed for the vector autoregression model. I anticipate using mapping skills, dimensionality reduction, clustering, time series interpolation, data cleaning, web scraping, and date-time functions to complete the project.

The areas where I expect to encounter the most difficulty is in manipulating the time series data, building the LSTM model, and removing seasonality. With other time-series data, I found working with the data quite cumbersome in that I was switching between date-time objects, time stamps, and Unix Epoch time. It often became difficult to align multiple time-series data from different sources. I am also trying to better understand how to clean and handle missing data, especially for the scraped weather data. I expect the seasonality removal to be difficult because the power consumption data appears to have hourly, weekly and annual seasonality (the seasonality appears to be affected by both weather and human activity). The human activity aspect will likely present a problem because during daylight savings time humans often shift their schedules while the weather conditions do not. In addition, I expect to run into difficulty implementing the models for LSTM, as I have not yet used that model yet. Overall I see this model as a stepping stone to a large problem. In the future, I would like to implement hypothetical models using roof space data, solar panel installation data, and real-time PV production data to predict photovoltaic output.