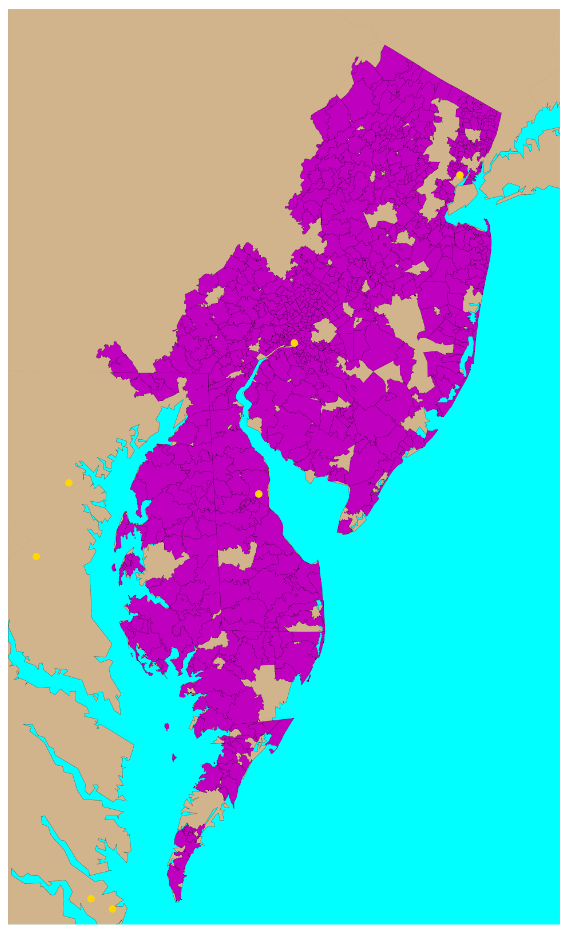
Final Capstone Proposal

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 have been sourced 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.

Please describe your data sources and acess in detail and how exactly are you accessing them. What is the size of the data? How many rows and columns exist? What are some important features for this analysis etc

Multiple sources of data will be used in this project. The first is a CSV file posted to Kaggle of hourly electricity consumption in megawatts from the PJM regional transmission organization between 1998 and 2018 and has 178262 rows. The second dataset is a CSV file obtained from the Weather.com's API that contains hourly weather data from airports in the region for the same dates as above and includes 46 different variables such as temperature pressure humidity.

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, 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. I am also trying to better understand how to clean and handle missing data, especially for the 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.

Project Milestones

- Data Cleaning
    - Convert weather data from Unix epoch time to date time
    - Resample weather data hourly
    - Drop irrelevant variables such as wave height
    - Impute values for variables such as temperature pressure and humidity using interpolation
    - Fill empty values with zero for variables such as snow and precipitation
    - Merge weather data with power consumption data using Date Time
- Data Exploration
    - Explore the seasonality of weather and electrical demand with autocorrelation functions and Fast Fourier Transformations.
    - Use unsupervised learning to extract behavioral information such as how holidays affect electrical consumption, and as a method to detect anomalies such as power losses due to storms.

- Explore the relation between temperature and electrical consumption. Through plotting and polynomial fitting
  - Conduct dimensionality reduction on weather data to consolidate variables
- Model Fitting
  - Create dummy variables from unsupervised and categorical data
  - Split data into training and test sections use 2018 as a test data
  - Detrend and deseasonalize data in preparation for autoregression and vector autoregression
  - Fit ARIMA and VAR models to Dataset
  - Fit LSTM Model
  - Test and score models
- Draw conclusions and explore ideas for further experimentation
  - Discuss the benefits and caveats of each model and how they might be further improved
  - Discuss how additional data from renewables might be incorporated.
  - Discuss how various climate models might be incorporated for the future.

By finding an accurate and robust model that can forecast power demand with weather data, power plant and grid operators would be able to implement this model in order to plan for things such as ordering fuel, coordinating with renewable power generation, and planning for storms that might affect electrical grid integrity. This is important because electricity production must match demand to prevent power surges and brownouts. The analysis of these models or the models themselves could be used by companies such as PJM to improve grid operation, not just the Eastern Mid Atlantic Area, but for other parts of their grid as well.