Project Proposal

Team members: Kanika Agarwal (ka2522), Chris Chen (cc7969), Weber Meng (qm351)

Question: Can we understand the dynamics of electricity market data (demand, supply and imbalance) and forecast into the future?

Datasets and Algorithms: We plan on using 3 electricity market datasets and apply HMM, Kalman and extensions of ARIMA to them.
3 datasets: Hourly load (demand) (SARIMA, Kalman), 5-min generation fuel mix (supply) (Kalman, HMM), 10-minute Area Control Error (ARIMA, SARIMA/other ARIMA extensions), covering 1 or 2 years.

Evaluation: Keep validation sets for selecting models and test sets for measuring forecasting error with RMSE, MSE, measuring variance. Evaluate how far into the future we can forecast with what accuracy/variance.

Who does what:
Chris is the subject matter expert, and will also work on preprocessing and the Generation Fuel Mix dataset (Kalman filtering and HMM).
Weber will be working general data loading preparation and will focus on modeling hourly data.
Kanika would be working upon the ACE dataset by applying ARIMA and SARIMA models to it.

Dataset Background:
Electricity Market Datasets: Open data is published by independent system operators (ISO), such as Southwest Power Pool (SPP), which cover multi-state regions of the United States. SPP covers states from Oklahoma to North Dakota.
Load:
Electrical system load is the amount of electricity used or consumed within an area, and represents demand in the electricity markets. These loads represent a multi-dimensional time series. There are strong seasonal effects at the daily, weekly, and yearly level. The main factor in load variability is temperature, where extreme temperatures lead to higher loads. The temperature dependency causes correlation between nearby loads. Load may be analyzable with Kalman filtering/smoothing as well as with seasonal ARIMA (SARIMA).
Generation fuel mix:
Generation fuel mix data shows generation amounts grouped by fuel type, such as natural gas, nuclear plants, wind, solar farms, coal, or hydropower. Factors that affect generation fuel mix include price of the underlying fuel, the amount of wind and solar power generated and outages of generators for refueling or repair. Thus, generation fuel mix may be modeled by Kalman filtering or Hidden Markov Models, where the underlying latents may represent fuel price, renewables, and electrical load.
Area Control Error (ACE):
ACE effectively measures the error in load/generation balance, the amount of excess load or supply. ACE is an example of a stationary time series centered around 0, so may be tractable with ARIMA models or other ARIMA extensions like SARIMA. We may also consider the ACE of multiple neighboring areas to form a multi-dimensional data set.

Data Citations:

Spp integrated marketplace. (n.d.). Retrieved October 10, 2022, from https://marketplace.spp.org/

Ace chart. (n.d.). Retrieved October 10, 2022, from https://marketplace.spp.org/pages/ace-chart

Hourly load. (n.d.). Retrieved October 10, 2022, from https://marketplace.spp.org/pages/hourly-load#

Generation mix rolling 365. (n.d.). Retrieved October 10, 2022, from https://marketplace.spp.org/pages/generation-mix-rolling-365