**Project Proposal**

**Team members:** Kanika Agarwal, Chris Chen, Weber Meng

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

**Datasets and Algorithms**: 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 published by independent system operators (ISO), such as Southwest Power Pool (SPP).

So far, we have considered analyzing three types of electricity datasets, load, generation mix, and Area Control Error (ACE).

**Load**:

Electrical system load is the amount of electricity used or consumed within an area, and represents demand in the electricity markets. Load may be reported at a regional level and at a “balancing area” (BA) level, approximately on a scale similar to a state. For example, SPP publishes hourly loads for 17 balancing areas. These loads represent a multi-dimensional time series. There are strong seasonal effects, showing a regular daily cycle (high loads in the day, low load at night), as well as weekly seasonality for day of the week and yearly seasonality. The main factor in load variability is weather, especially temperature, where extreme temperatures lead to higher loads. The temperature dependency also leads to a correlation between nearby loads.

Load may be analyzable with Kalman filtering/smoothing, perhaps after some de-seasoning with ARIMA techniques.

[SARIMA, other extensions to arima]

**Generation fuel mix:**

Generation fuel mix data shows how much electricity is being generated by categories of generators grouped by fuel type, such as natural gas generators, nuclear plants, wind farms, solar farms, coal, hydropower. Because electricity needs to be balanced, the sum of the generation generally matches the load at that time. Electricity markets are balanced economically, with the cheapest marginal generators being used for economic efficiency. Factors that affect generation fuel mix include price of the underlying fuel (natural gas, coal, oil), the amount of wind and solar power generated (based on weather conditions) and outages (shutdowns) 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.

If generation fuel mix is reported as a proportion of total generation instead of absolute amount (MW) of power, this may be an appropriate dataset for Hidden Markov Models (HMM).

**Area Control Error** (ACE):

A system operator (or balancing authority area’s) goal is to match electricity demand (load) with supply (generation), with adjustments for scheduled imports and exports of electricity from neighboring areas. ACE effectively measures the error in this balance, how much excess load or supply at any one time for an area. If there is error, the amount of power flowing into or out of the area is different than expected/scheduled, and measurements over these boundaries are used to calculate ACE.

ACE is an example of a stationary time series centered around 0, so may be tractable with ARIMA models. The ACE of one area, such as SPP, is a one-dimensional data set, but we can also consider the ACE of multiple neighboring areas to form a multi-dimensional data set. Such neighboring ACEs would then have a relationship in that excess flow out of one area corresponds to excess flow into the neighboring areas, so they have common underlying factors/summands.

Definitions

Area control error. (n.d.). NERCipedia. Retrieved October 10, 2022, from <https://nercipedia.com/glossary/area-control-error/>

Area control error (Ace) · Energy KnowledgeBase. (n.d.). Retrieved October 10, 2022, from <https://energyknowledgebase.com/topics/area-control-error-ace.asp>

* Good description to read!

Paper using deep learning:

Abdeltawab, H., & Radwan, A. (2022). Area control error forecasting using deep learning for an interconnected power system. 2022 IEEE Power and Energy Conference at Illinois (PECI), 1–5. https://doi.org/10.1109/PECI54197.2022.9744044

<https://ieeexplore.ieee.org/document/9744044>

SPP data sources:

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>