Analysis and Forecasting of Electricity Market Data

Kanika Agarwal
New York University
New York, NY 10012
ka2522@nyu.edu

Chris Chen
New York university
New York, NY 10012
cc7969@nyu.edu

Weber Meng
New York university
New York, NY 10012
qm351@nyu.edu

1 Introduction

Electricity markets allow economically-efficient buying of selling of electricity. They are based on the physical electric power grid, a complex interconnected network of generators, transmission lines, and consumers.

Our goals are to: 1) understand the dynamics of electricity market data (demand, supply and imbalance) and 2) forecast their future values at different forecasting scales (short-term/long-term). Electrical system load is the amount of electricity used or consumed within an area, and represents demand in the electricity markets. Generation fuel mix data shows the amount of generation attributed to different fuel types. Area Control Error (ACE) measures the error in the balance between load and generation, or in other words, the amount of excess load or supply.

Understanding and forecasting the large-scale dynamics of an electricity market's demand (load), supply (generation), and imbalance (Area Control Error) are important goals for market participants and for grid operators. For a market participant, such as a financial trader trying to buy and sell power for a profit, load forecasts can be used as inputs to forecasts for power prices. Understanding the changes in generation fuel mix would help understand which generators turn on or off and when, and which fuel prices are determining the price of electricity. A short-term financial participant may be interested in forecasts for the next 24 to 48 hours, while a longer-term financial participant may be interested in load forecasts for the next month to year.

For an operator of the electric grid such as a electric utility or a Balancing Authority (BA), the end goal is that there is enough generation to match the demand for electricity, and that it is generated in an economically efficient way. Good short-term load forecasts would help indicate whether the operators need to increase or decrease generation, as would short-term generation forecasts for the self-scheduled fuel types the operator cannot control, such as wind generation. Balancing Authorities are mandated to manage Area Control Error (ACE) within certain limits, so understanding the persistence of ACE could help evaluate the control performance of the Balancing Authority or indicate possible improvements in operations.

We model nuclear generation data with a Hidden Markov Model HMM. We analyze hourly load data by applying ARIMA and an LSTM model. We also used walk-forward forecasting with ARIMA to explore different forecasting lengths. For ACE, we use an ARIMA model. With these time series methods, we are able to predict nuclear generation, and load and understand ACE.

2 Related work

While we did not base our analysis here on any previous study, we note that time series methods have been used in the past on electricity market data. Juberias et al. applied ARIMA (autoregressive integrated moving average) to load forecasting for an electrical grid in Spain, using meteorology as an explanatory variable. Similarly, Contreras et al. use ARIMA models to predict next-day electricity prices for Spain and California. Yu and Sheble model electricity market states as a Hidden Markov Model, working with data from New York. Tokgöz and Ünal use Recurrent Neural Networks (RNN), Long-Short Term Memory (LSTM), and Gated Recurrent Units (GRU) to forecast Turkish electricity load.

3 Problem definition and algorithms

3.1 Task

We want to understand the dynamics of electricity demand and supply, namely generation fuel mix data and hourly load data. To understand the nature of these data, we train various models including HMM, ARIMA, LSTM and variations of these and try to fit the data. Both these data seem seasonal in nature and we except to these models to be able to forecast well. For the imbalance data we have in our Area Control Error data, we want to find out if the imbalance has any structure to it. If so, how can we model it. [1]

3.2 Algorithm

3.2.1 AR

It is an **Auto-Regressive (AR)** model. Data regresses on its own preceding values. AR(p) is an AR model where p is the number of autoregressive terms.

3.2.2 ARIMA

It is an autoregressive integrated moving average model. It has an extended version of the AR model and has two addition components apart from the **Auto-Regressive** component. **Integrated (I)** component enables the time series to become stationary by separating the raw observations. **Moving average (MA)** component forecasts errors which are linear functions of prior errors. ARIMA(p,d,q) is an ARIMA model where p is the number of autoregressive terms, d is the differencing degree and q is the count of forecasting errors.

3.2.3 Walk-forward ARIMA

ARIMAs do not give accurate results in a long run: the further the ARIMAs try to predict, the smaller their amplitude is. The speed of amplitude decay is very quick. To resolve this problem and force seasonality, we have adopted a walk-forward approach to ARIMA (5). Here is our strategy:

1. Train on first x data points to make the next predictions for the next y steps.
2. Record those y steps predictions in 1.
3. Train on the first $x + y$ data points as the training data set, and then make predictions for the next y steps. We retrain all ARIMA parameters using the *auto_arima* function, so the parameters of ARIMA parameters are likely to be different if the step size y is large.
4. Record those y steps predictions in 3.
5. Continue including more y steps into the training data and predict the next y steps until the data exhausts.

[1] All code files for the current project can be found here: `https://github.com/QingweiMeng1234/Project-1018`

3.2.4 Hidden Markov Model

A Hidden Markov Model (HMM) is a latent-space model where the latent dynamics are represented as a Markov Chain. At each time step, the latent space can take on one of a discrete set of k states, and the probabilities of transitioning between two states after a time step is given by a transition matrix A. Each of these latent states is associated with a different observation or emission distribution. The expectation maximization (EM) algorithm is used to train the HMM. The Viterbi algorithm is used to decode the most likely sequence of latent states, given a set of observations and known model parameters:

$$\hat{z}_{1:t} = \arg\max_{z_{1:t}} \{P(z_{1:t}|x_{1:t})\}.$$

In this case, we use a Gaussian Hidden Markov Model, where each state emits observations according to a Gaussian distribution with mean and covariance parameters. We use HMM on the Nuclear Generation data.

3.2.5 Long short-term memory

LSTM is a type of Recurrent Neural Network. RNNs are used for identifying trends in data series. The standard RNN might not be able to capture long term dependencies which are taken care of by LSTMS by providing a short-term memory that could persist for long. A typical LSTM cell comprises of a forget gate, input gate and an output gate.

4 Experimental evaluation

4.1 Data

The data is published by the Southwest Power Pool (SPP), which is a regional transmission organization (RTO) that manages the operation of the electricity system across portions of 14 states from Oklahoma to North Dakota (4).

Hourly Load: Electrical system load is electrical usage or electrical demand in an area. There are 17 different Balancing Areas in the dataset, forming a multi-dimensional time series, but we focus on analyzing load for one area, "CSWS", which on average has the most load of the areas. 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. As the dataset name implies, Hourly Load is reported as an hourly average, and the data ranges from dates 2019-06-21 to 2022-10-16, for 29112 rows (3).

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. In this analysis, we focus on analyzing nuclear generation. Generation fuel mix is reported every 5 minutes as an average for that interval, and we have data from the dates 2018-01-01 to 2022-10-16, for 502144 total rows (2).

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. ACE is reported every one minute, and we have data for 2014-03-01 to 2022-10-16, for 3805578 total rows (1).

For the three datasets above, the units are reported in megawatts (MW). The data was downloaded from the original sources. As it generally was published with multiple reports for different time periods (days/months/years), the relevant reports were concatenated to create the multi-year time series. For the purposes of this analysis, missing or null data was simply removed, and analysis is performed assuming the time series has discrete, regular time steps.

4.2 Methodology

4.2.1 Nuclear generation

Hidden markov model: For nuclear generation, we split the dataset to train on the earlier half of the time period, from January 1, 2018 to May 24, 2020. We used the *hmmlearn* Python package and the *GaussianHMM* function. Because the result of fitting an HMM model with the expectation-maximization algorithm (EM) depends on the random initialization, we generated multiple (11) candidate HMM Models which also varied in parameters, trying different numbers of latent states (7 to 9), different number of EM iterations (13 to 40) and different random state (random seed) values. The model with the highest log likelihood score was kept as the final model.

After training the HMM, the model was used to infer the most likely latent states corresponding to the observations in the training data with the Viterbi algorithm. This sequence of latent states was associated with their corresponding mean and standard deviations to plot against the observed sequence and to calculate a training RMSE.

4.2.2 Hourly load data

ARIMA: The ARIMA model is trained on data from April 9, 2022 6:00 pm to October 9, 2022 5:00 am and tested on data from October 9, 2022 6:00 am to October 16, 2022 5:00 am, that is exactly a week's data. The ARIMA model used for predictions is ARIMA(2,1,3).

LSTM: Hourly load data is split into train and test data. Training data has data from June 21, 2019 6:00 am for about 172 weeks until October 9, 2022 5:00 am. Model is tested on the upcoming week right after that the train data. Input size is 24 and output size is 5. Our LSTM model uses *tanh* as the activation function. The model is trained for 3 epochs since it started to overfit beyond that.

Walk forward ARIMA: In the code, we trained the first 96 data points (4 cycles), then for each step, since our cycles have around 24 hours, we have chosen $1, 6, 12, 24, 36, 48, 60, 72, 84, 96, 168$ as step sizes.

4.2.3 Area Control Error

We ran the Augmented Dickey–Fuller test to see if the data is stationary or not. We plotted ACF and PACF for the data to understand the correlation amongst various observations within the data.

AR: AR model is trained on ACE data with observations every minute from March 1, 2014 6:00 am to March 8, 2014 4:41 am and tested for short-term forecast on 10 observations from March 8, 2014 4:42 am. We used the *api.tsa.ARIMA* function from *statsmodels* Package in Python to build the AR model and chose the p value to be 2 based on the PACF plot of the data shown in Figure 1.

ARIMA: ARIMA model is trained on ACE data with observations every minute from March 1, 2014 6:00 am to March 8, 2014 4:41 am and tested for short-term forecast on 10 observations from March 8, 2014 4:42 am. We used the *auto_arima* function from the *pmdarima* library in Python .

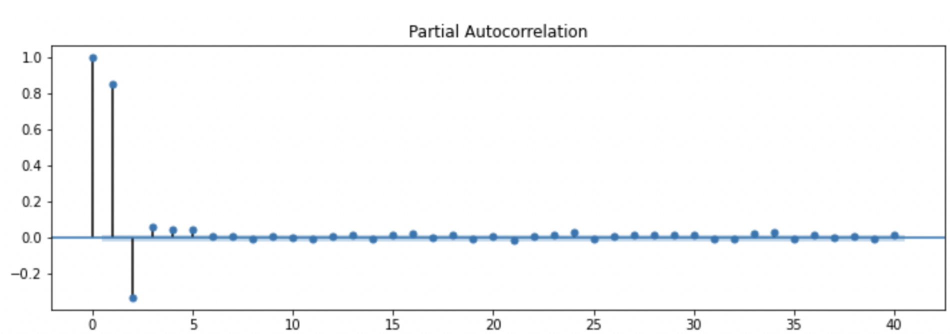


Figure 1: Partial Autocorrelation plot of ACE Data

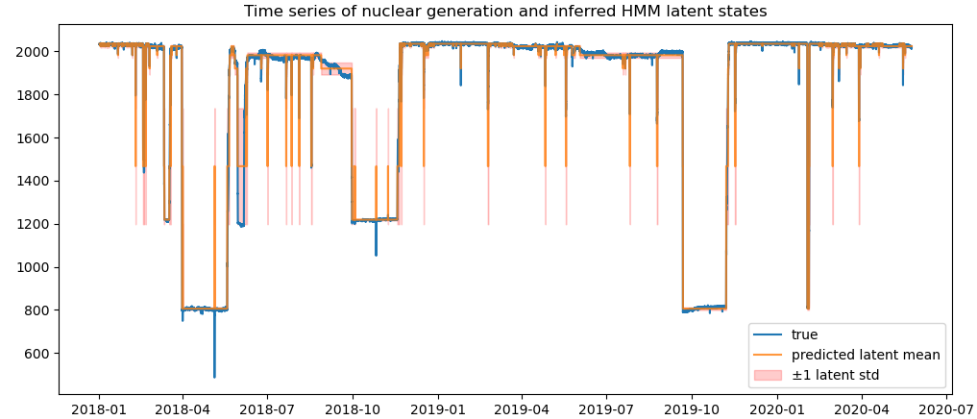
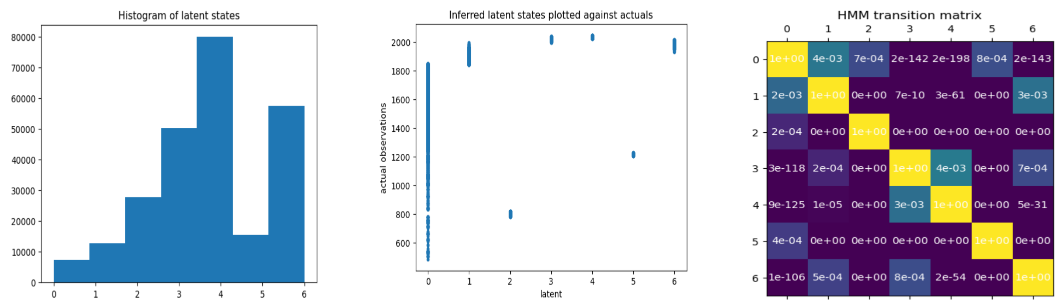


Figure 2: Plot of nuclear generation data and inferred HMM latent states, showing confidence bounds of 1 standard deviation of the emissions distribution.



(a) Histogram of the latent states of HMM

(b) Plot of latent states vs observations

(c) HMM transition matrix, where color is plotted on a log scale.

Figure 3: Examining the HMM latent-state representation of the nuclear generation dataset. From these plots, we can see state 4 is the most common state and represents full capacity at around 2000 MW. States 1, 3, and 5 also represent generation near 1900 to 2000. State 2 represents the 800 MW state and state 5 represents 1200 MW. State 0, the least common state represents temporary drops to other values, or a transition state between other states and has the highest variance/standard deviation.

4.3 Results

4.3.1 Nuclear Generation Hidden Markov Model

Qualitatively, the HMM captures the dynamics and states of the Nuclear Generation dataset well. There are multiple states that represent nuclear generation output near the maximum capacity of 2000 MW, as well as separate states for 800 MW and 1200 MW. There is also a transition state that captures temporary dips in generation to other values or transitions between the other states. See Figures 2 and 3.

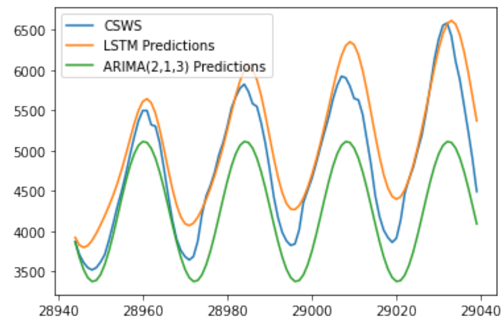
The RMSE between the actual training data observations and the inferred latent means is 46.98. The log likelihood of the final chosen model is -812389.

4.3.2 Hourly load data

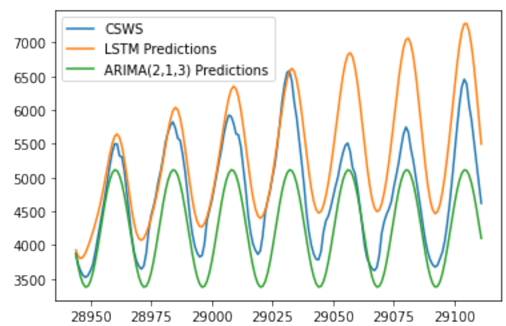
We used RMSE as metrics to compare the different models. ARIMA model gives us RMSE value of 642.77, LSTM gives RMSE value of 749.83 for 1 week forecasting. Onthe other hand, for 4-day forecasting, ARIMA model gives us RMSE value of 681.79 and LSTM gives us RMSE of 370.43. These results are displayed in Table 1. Refer to Figures 4a and 4b for the ARIMA and LSTM model predictions on the test data. Please refer to Figures 5 and 6 for model predictions using Walk-Forward ARIMA and RMSE values for various step sizes for the same.

Model	4 day forecasting	7 day forecasting
ARIMA(2,1,3)	681.79	**642.77**
LSTM	**370.43**	749.83

Table 1: RMSE values for ARIMA(2,1,3) and LSTM for Hourly Load data for forecasting within a week



, height=4cm, height=4cm (a) forecasting for 4 days , height=4cm,



height=4cm (b) forecasting for 7 days

Figure 4: ARIMA(2,1,3) and LSTM model predictions on Hourly Load Data

4.3.3 Area Control Error Data

Refer to Figure 7 for predictions of AR and ARIMA models on ACE test dataset.

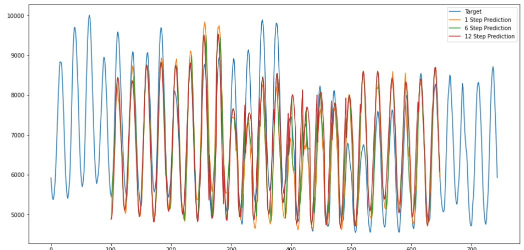
4.4 Discussion

4.4.1 Nuclear Generation Hidden Markov Model

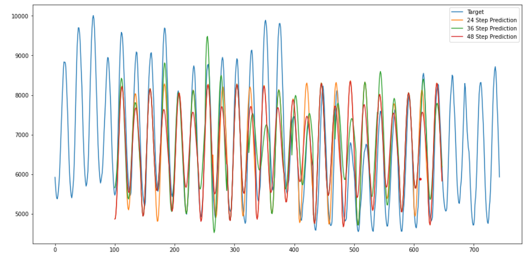
A Hidden Markov Model can represent SPP Nuclear Generation data well. The Nuclear Generation data has a set of discrete, distinguishable states in it that can be identified by the HMM training procedure.

It is helpful to understand the fundamental factors behind why nuclear generation in SPP specifically appears to have these states. There are two nuclear plants in SPP, each with one generating unit: 1. Wolf Creek Generating Station in Kansas, with capacity 1200 MW. 2. Cooper Nuclear Station in Nebraska, with capacity 835 MW. In general, it is economic to always run nuclear generators at their full capacity, unless they need to be turned off for repairs or refueling. That is why we see three distinguishable states at 800, 1200 and 2000.

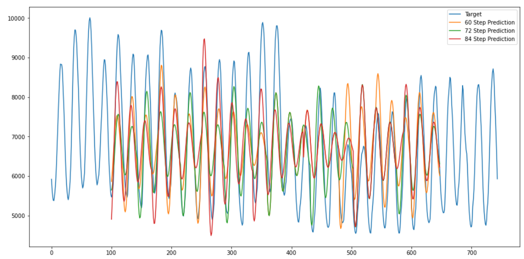
Modeling nuclear generation data gives further insights into the dynamics of the system. For example, from the transition matrix, we can estimate the probability of remaining at a state or of transitioning to another state, which may be useful if an operator or financial participant want to quantify a risk that one of the nuclear generators will shut down or turn on in the next interval. We can further investigate the states on the ground, and ask what are the fundamental factors that cause some of the less obvious



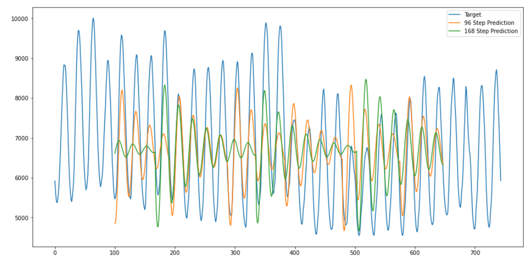
(a)



(b)



(c)



(d)

Figure 5: Walk Forward Model Predictions on Hourly Load Data for next 27 days

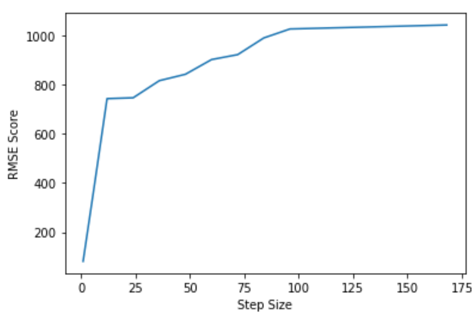


Figure 6: RMSE for Walk-Forward ARIMA on Hourly Load data

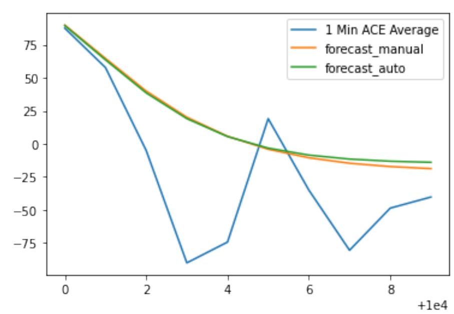


Figure 7: ARIMA model predictions on ACE data

states, like the transition state or the states near 1900 MW. As a final example, if we see a new dip in nuclear generation, we can infer whether this is likely to be a long term or short term outage using the Viterbi algorithm or the forward-backward algorithm to predict latent probabilities.

4.4.2 Hourly load data

Even though ARIMA model seems to give better results for a week, LSTM seems to give better results for the initial 3-4 days as seen in Figure 4 which shows forecasting by ARIMA(2,1,3) and LSTM for 4 days of data. We see that ARIMA overall performs better for longer-term forecasting whereas LSTM performs better for shorter-term forecasting.

To help with ARIMA's "long-term" forecasting, we are motivated by a walk-forward approach. During the walk-forward approach on ARIMA, we discover that the RMSE scores increase sharply when we increase the step size at the lower end, while the scores increase very slowly for the higher end.

We also allow the ARIMA parameters to vary according to the newly incorporated training data. Each time we walk forward, the ARIMA model used for each step might vary, but most of the time it switches between ARIMA(2,0,2),ARIMA(2.0,3),ARIMA(2,1,3) and ARIMA(5,0,0). The reason why the model parameters change might be because of the structure of the data. As we can see, the second half of the data has less amplitude than the first half. The smaller amplitude might be the reason why the ARIMA shifts from $(2, 0, 2)$ or $(2, 0, 3)$ to $(2, 1, 3)$ or $(5, 0, 0)$.

Overall, this method allows forecasting to adapt the ARIMA model to recent changes in the dynamics, and may be suitable when there is a belief that newer data has different dynamics.

4.4.3 Area control data

ARIMA model doesn't forecast well on ACE dataset and revert back to the mean after two timesteps. This is expected due to the nature of the ACE dataset as a control error in load and generation balance. The errors are not persistent. It seems to be very close to being white noise except from having very slight autocorrelation. Although ACE data is stationary with mean at 0, not much can be forecast about the load-generation imbalance.

5 Conclusion and future work

Nuclear Generation Hidden Markov Model The Hidden Markov Model is an appropriate representation for SPP Nuclear Generation output. The trained HMM reports seven distinct states, each with corresponding Gaussian distribution mean and variance, which correspond to different configurations of the two nuclear generators in SPP. The HMM can assist operators and financial participants in understanding the dynamics of nuclear generation.

One shortcoming of our method is that it is possible that better models could be found, as we only tried a small number of manually specified models. Ideally, we would test more candidate models and systematically search over the hyperparameter space with grid search or another search algorithm.

There are some limitations inherent to the HMM representation: the model necessarily has a fixed number of discrete latent states. Therefore, predictions can only assume the form of the means associated to the latent states (or if using probabilities, a weighted average of those means). The HMM is better at representing "jumps" in the data to different states rather than gradual continuous decreases that can be seen in the nuclear data.

Future work includes doing predictions with the HMM, but this would probably perform worse (or similarly) to a naive forecast: this HMM, with transition matrix diagonal entries near one, would predict future latent states as the same as the current most-likely latent state, and thus would predict the same latent mean.

Hourly Load Forecasting By leveraging the regular periodic structure of the hourly load data, ARIMA and LSTM methods both are viable methods for forecasting future periods. They capture different aspects of the data and have different long term forecasting behavior in forecasting a trending mean.

For ARIMA, experimenting with different forecasting lengths shows that shorter timeframes such as one-step-ahead forecasting results in the most accuracy, and forecasting RMSE drops significantly when forecasting beyond 24 hours.

For our approach of dynamically retraining all ARIMA parameters during the walk-forward forecasting, future work would involve further comparisons against a more traditional method where the ARIMA model hyperparameters or parameters are trained once on the training data and remain fixed when incorporating new data.

One limitation of both ARIMA and LSTM methods is that they both predict future observations based on current and past observations, and do not take into account any exogenous variables such as temperature that may disrupt the current trends. Future work could incorporate such variables with different methods like ARIMAX or as inputs into the same algorithms.

Area Control Error While we can create an ARIMA model for ACE, similar to the model for load, forecasts for ACE diverge much more rapidly from the realized actual values. ARIMA forecasts for ACE are only viable for one or two time steps before they diverge. Thus, we conclude that Area Control Error has a minimal amount of time series structure and autocorrelation that can be used to perform limited forecasting.

For future work, we think it is unlikely that more sophisticated methods would be able to improve on ACE forecasting, and we have already experimented with using RNN but did not achieve significant forecasting performance. We may be able to look at changes in variance in the ACE data and try to employ a model like GARCH.

6 Student contributions

Chris Chen worked on data sourcing and preprocessing, the generation data (Nuclear HMM) and contributed to the ACE analysis.
Kanika Agarwal worked on training the ARIMA model and the LSTM model for Hourly Load Data and also worked upon the ACE analysis.
Weber Meng worked on the Hourly Load ARIMA and walk-forward analysis.
Everybody worked upon writing the report and presentation.

We performed several exploratory analyses that we ultimately chose not to report (Gen Fuel Mix Kalman filter, ACE LSTM, Load PCA).

References

[1] Ace chart. (n.d.), October 10, 2022. URL `https://marketplace.spp.org/pages/ace-chart`.

[2] Generation fuel mix rolling 365 (n.d.), October 10, 2022. URL `https://marketplace.spp.org/pages/generation-mix-rolling-365252`.

[3] Hourly load data (n.d.), October 10, 2022. URL `https://marketplace.spp.org/pages/hourly-load`.

[4] Spp integrated marketplace. (n.d.), October 10, 2022. URL `https://marketplace.spp.org/`.

[5] Stein R. M. Bohn, J. R. *Active credit portfolio management in practice*. Essay, John Wiley & Sons, 2009.

[6] J. Contreras, Rodrigo Espinola, Francisco Nogales, and Antonio Conejo. Arima models to predict next-day electricity prices. *Power Systems, IEEE Transactions on*, 18:1014 – 1020, 09 2003. doi: 10.1109/TPWRS.2002.804943.

[7] G. Juberias, R. Yunta, J. Garcia Moreno, and Cloe Ortiz de Mendívil. A new arima model for hourly load forecasting. *1999 IEEE Transmission and Distribution Conference (Cat. No. 99CH36333)*, 1:314–319 vol.1, 1999.

[8] Alper Tokgöz and Gözde B. Ünal. A rnn based time series approach for forecasting turkish electricity load. *2018 26th Signal Processing and Communications Applications Conference (SIU)*, pages 1–4, 2018.

[9] Wang Yu and Gerald Sheble. Modeling electricity markets with hidden markov model. *Electric Power Systems Research*, 76:445–451, 04 2006. doi: 10.1016/j.epsr.2005.09.013.