Forecasting New England Energy Consumption

Jonathan Clifford, @takao8, jcliff@udel.edu, Data/Literature Manager

Desiderio Pilla, @DesiPilla, dmpilla@udel.edu, Methodology Manager

Ramiz Qudsi, @ahmadryan, ahmadr@udel.edu, Communication/Visualization Manager

Abstract

Energy consumption forecasting is vitally important for both energy companies and government agencies to properly hedge their assets and to ensure sustainability and access to the resource. Using historical weather data for the city of Boston (2000 to 2019), we performed a statistical analysis of energy forecasting for the New England region. Two different models using Facebook's Prophet package and feedforward neural networks were developed to forecast load upto 90 days in advance. Results from the two methods show differing statistical behaviours with a best performance from Prophet of 7.45% hourly mean average percentage error over a 3 month period.

Introduction

In the United States, transmission and wholesale of electricity is regulated by the Federal Energy Regulatory Commission (FERC). Regions of the country that are deregulated are split into separate entities created by FERC, known as independent system operators (ISOs), which coordinate the electric grid of that region. For each of these regions, load (the amount of energy that is consumed at a single point, expressed in Megawatt-hours) forecasting is vitally important for the smooth and sustainable functioning of the area. Since produced electricity must be consumed immediately [1,2] (approximately true due to limited implementation of battery technology except for California [17]), an accurate forecasting of load will avoid wastage of energy in cases of over production, risk of blackouts, or in case of under production the necessity to buy energy to meet that demand from real-time energy markets, which tend to be expensive. In this paper, we present our analyses and results from forecasting energy demand in ISO New England (commonly called NEPOOL, the New England Power Pool), which encompasses Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont.

Load forecasting is a well-documented problem and ample literature is available for various methods. ARMA [3,4], ARIMA, linear regression [5], expert systems, neural networks [6,7], Fuzzy logic [4], support vector machine [8,9], and ant colony optimization [9] are some of the major methods used for electrical load forecasting. [10] describes several of these modern methods in detail and [11] presents an extensive review. One of the first in-depth analyses of neural networks in load forecasting is explored by [11], reporting a mean absolute percent error of 2% on day-ahead forecasts.

For our study, we used two models: Facebook's Prophet (FP), which is based on an autoregressive method, and a feed forward neural network. We also explore an attempt to implement recurrent neural networks.

Data

The primary dataset used in this project is the net energy load by source for NEPOOL. This dataset was acquired from the ISO New England website and covers the time interval from January 1st, 2000 to July 29th, 2019, with a cadence of 1 hour for each observation. Figure 1 shows a typical energy demand for the year 2015. The other relevant feature in this dataset is the Heat Index (HI) [12], which combines historical temperature and humidity data at Boston Logan International Airport, measured in degrees celsius. Figure 2 shows the correlation between load and HI in Boston for the period between 2000 and 2019. The graph shows two populations during the winter months, which merge into one as the HI increases. As expected, load increases as the HI increases (decreases) from the global minima point as people turn on their cooling (heating) systems.

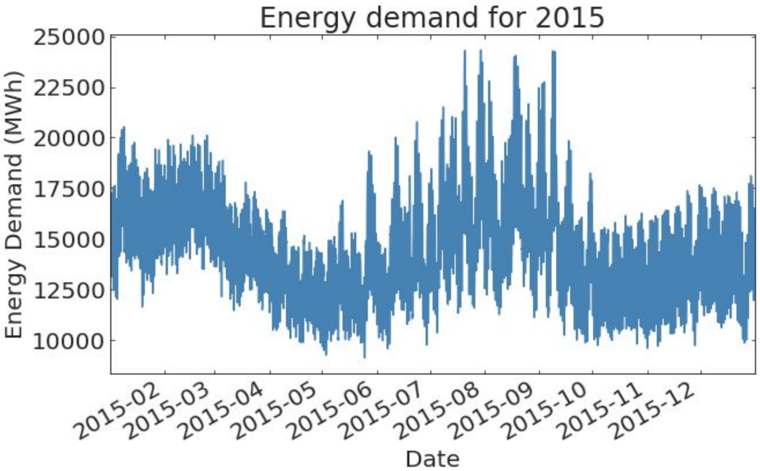
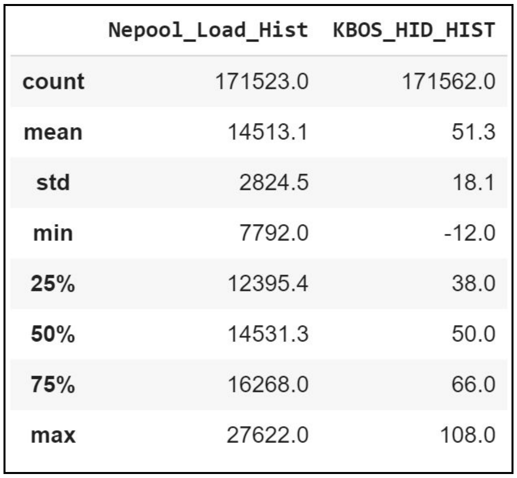


Figure 1 : A time series sample of energy demand for 2015 for the New England area.



	Nepool_Load_Hist	KBOS_HID_HIST
count	171523.0	171562.0
mean	14513.1	51.3
std	2824.5	18.1
min	7792.0	-12.0
25%	12395.4	38.0
50%	14531.3	50.0
75%	16268.0	66.0
max	27622.0	108.0

Table 1: Snapshot of `nepool.describe()`

The second dataset includes historical weather data for the Boston area. This data originates from Climate Data Online (CDO), presented by NOAA. The data was acquired using the following criteria:

- Weather Observation Type/Dataset: Normals Hourly
- Date Range:
    - 2000-01-01 00:00 to 2009-12-31 23:59
    - 2010-01-01 00:00 to 2019-07-29 23:59
- Search for: Stations
- Search Term: WBAN:14739

Due to limits on the size of datasets that can be ordered at a time, two separate datasets were requested and then concatenated together. The resulting dataset contained several attributes relating to the hourly weather conditions over the time period covered by the NEPOOL dataset. Table 2 shows which weather features were gathered, whereas Table 3 gives a high level description of the dataset used.

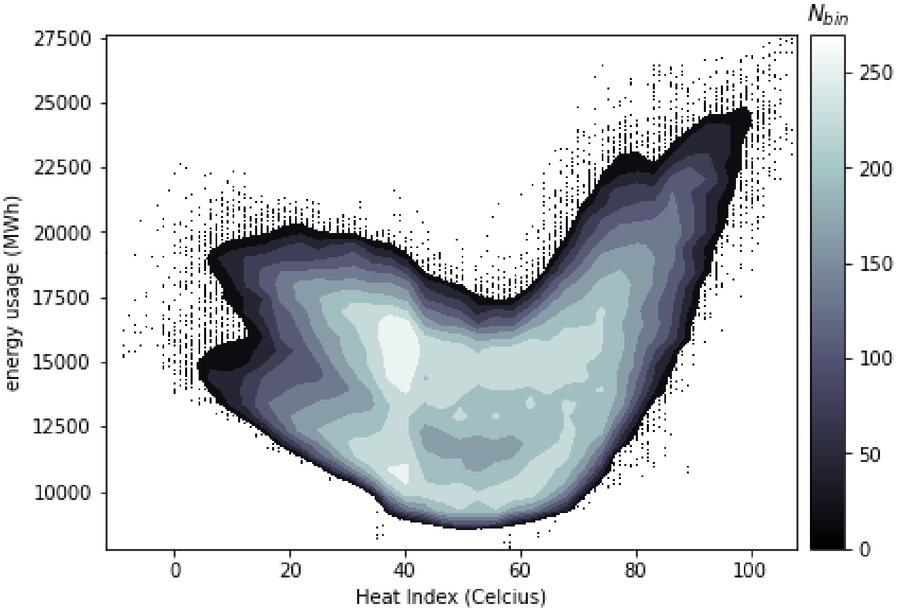
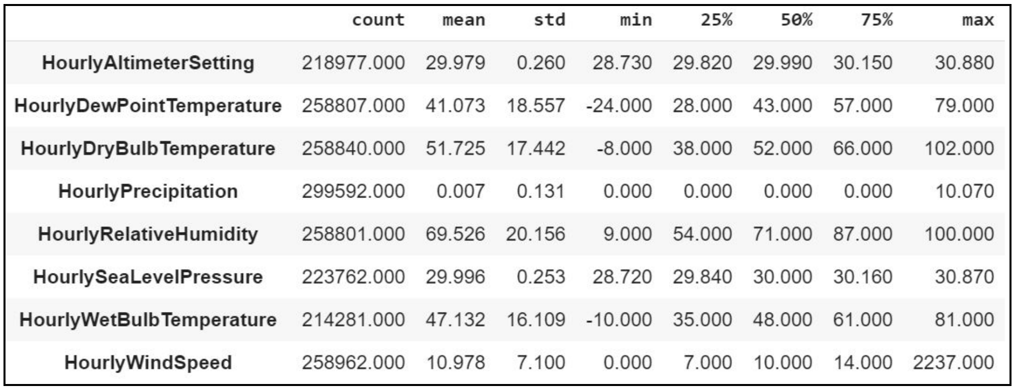


Figure 2: Contour plot between energy consumption and Heat Index of Boston for 2000-2019. Although showing evidence of two populations in the winter, they merge into one in the summer, dissuading the possibility of disjoint populations.



	count	mean	std	min	25%	50%	75%	max
HourlyAltimeterSetting	218977.000	29.979	0.260	28.730	29.820	29.990	30.150	30.880
HourlyDewPointTemperature	258807.000	41.073	18.557	-24.000	28.000	43.000	57.000	79.000
HourlyDryBulbTemperature	258840.000	51.725	17.442	-8.000	38.000	52.000	66.000	102.000
HourlyPrecipitation	299592.000	0.007	0.131	0.000	0.000	0.000	0.000	10.070
HourlyRelativeHumidity	258801.000	69.526	20.156	9.000	54.000	71.000	87.000	100.000
HourlySeaLevelPressure	223762.000	29.996	0.253	28.720	29.840	30.000	30.160	30.870
HourlyWetBulbTemperature	214281.000	47.132	16.109	-10.000	35.000	48.000	61.000	81.000
HourlyWindSpeed	258962.000	10.978	7.100	0.000	7.000	10.000	14.000	2237.000

Table 2: A description of NOAA weather data. Count refers to the number of data points and other columns have standard meaning

Dataset name	URL	Number of rows	Number of valid rows (not NaN on relevant columns)	Number of columns	Number of relevant columns	Data type for each relevant column
nepool	https://www.iso-ne.com/isoexpress/web/reports/load-and-demand/-/tree/net-ener-peak-load	171,562	171,562	18	2	All floats
weather_noaa	https://www.ncdc.noaa.gov/cdo-web/	299,952	299,952	124	9	All floats

Table 3: High level description of datasets.

Methodology

Facebook Prophet

The Prophet package by Facebook is one of the two tools utilized to complete this project. This model optimizes the parameters using a direct optimization approach.

An initial exploration of the NEPOOL dataset suggests the presence of three significant seasonalities:

1. **Yearly:** Because of various seasons over the period of a year, we observe a periodic rise and fall in the consumption of energy. The demand is higher in summer and winter, and lower in spring and autumn.
2. **Weekly:** The typical work week is five days, and the weekend two days. There is a considerable difference in power usage during the off-days compared to the work days, with energy consumption tending to be greater during business days than on the weekend.
3. **Daily:** Initial analysis shows an increased demand in the morning and during the day. The energy load drops off at night time.

These seasonalities have been considered in the model by implementing the specified seasonalities in the Prophet implementation.

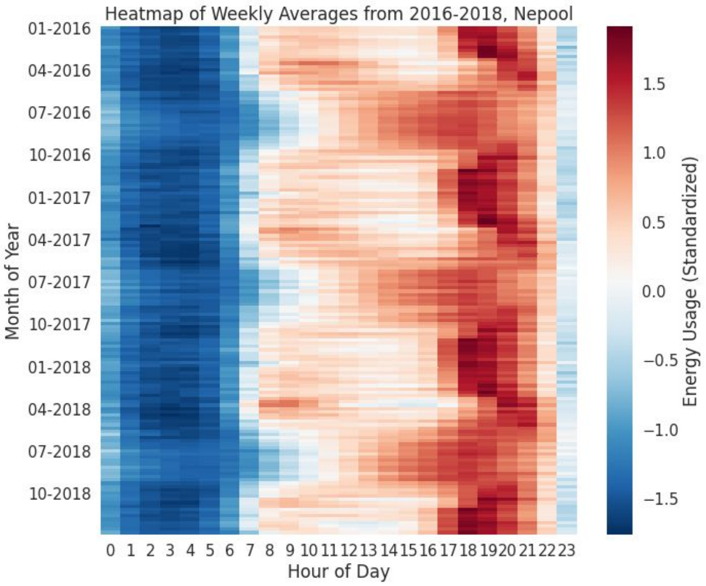


Figure 3: Heatmap showing weekly averages of daily energy load from 2016-2018. In this figure, the summer and winter seasonalities are clearly shown. Note the smooth transition of daily seasonalities between summer and winter.

Energy demands follow different patterns in the summer as in the winter. In the summer months, energy load rises throughout the day, peaks in the evening, and then sharply drops at night. However, during the winter months the data exhibits a double peak. The energy load rises sharply in the morning, retracts during the afternoon, and then reaches a second, higher peak in the evening before dropping at night. The spring and autumn months follow daily motions somewhere in between these two distributions. The Prophet model has been trained to learn different daily seasonalities for days in different seasons. (See Appendix A for more details)

Holidays must also be considered. On a holiday, the energy demand more closely follows that of a weekend. The Prophet package has the built-in capability to recognize holidays in its model, and has been implemented in the model as well.

We used the NOAA dataset to analyze the effect of variation in weather on the energy demand. Rather than use all available weather variables, a principal component analysis (PCA) was conducted to decrease the number of features in the dataset (see **Results** for more details). The most relevant components were added as regressors to the Prophet implementation.

To measure the performance of the model, we compute the mean absolute percent error (MAPE) of its predictions:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{F_t - A_t}{A_t}\right|$$

where A_t is true load, F_t is forecasted load, and n is the size of our test set. The objective is to minimize MAPE on the test set.

Neural Networks

We implemented two different neural networks; a feed-forward network and a recurrent neural network.

For both the neural networks, we discarded the PCA and fit on the standardized features. The reason for discarding is from the neural network's ability to learn patterns from correlated data. The training set available was large enough to optimize the parameters without fear of overfitting from lack of data. Additionally, due to inclusions of dummy variables (described below), its mixture with PCA components produced noticeably worse results that were difficult to resolve. As this model doesn't require uncorrelated features, we chose to drop the PCA components.

Both the networks are supplied with dummy variables to help them learn the seasonalities of the data. These dummy variables include hourly, monthly, holiday, and weekend indicator variables. Overall we retain 45 features to feed into the neural nets.

In constructing the feedforward network, we created 8 dense layers, each decreasing in size. 4 dropout layers are interwoven to prevent overfitting. We used the mean squared error for our loss function as it is a natural loss function (including any Minkowski metric) for regression problems. Additionally, each layer uses a 'tanh' activation function, chosen due to its common usage in regression tasks. In our model, we split our training, validation, and testing as the following: after choosing a point in our data set to test forward from, we select the previous year from that point to validate, the previous 2 years from the validation to train, and the upcoming 3 months to test. The points we chose are the first of February, May, August, and November, to test on Spring, Summer, Fall, and Winter, respectively. We cross-validate for an accurate test error by training/validating and testing through multiple years, and obtain an overall MAPE.

Due to computational issues with the cross-validation, only a limited range could be cross-validated at one time. In this paper we obtained cross-validation results from 2011-2012. Due to a memory leak in the Tensorflow code, training of further than 1,000 epochs resulted in a likelihood of a computer crash. For the purposes of tuning this model, these computational limits are debilitating and require to be resolved for proper testing.

A recurrent neural network was also implemented, with properties described in the **Results** section.

Results

Principal Component Analysis

The principal component analysis revealed that only the first two components were needed to explain nearly 98% of the variance in the weather dataset (Figure B.1). This is very beneficial, as it allowed us to reduce the dimensionality of the dataset from 9 features to 2. This saved hours in the model training / cross-validation step of the analysis.

It is also noteworthy which weather features were most represented in the PCA components. The two components were almost entirely composed of the Dew Point Temperature, Wet Bulb Temperature, Dry Bulb Temperature, and Relative Humidity (Figure 4). This is important, as it tells us precipitation is essentially excluded from our model.

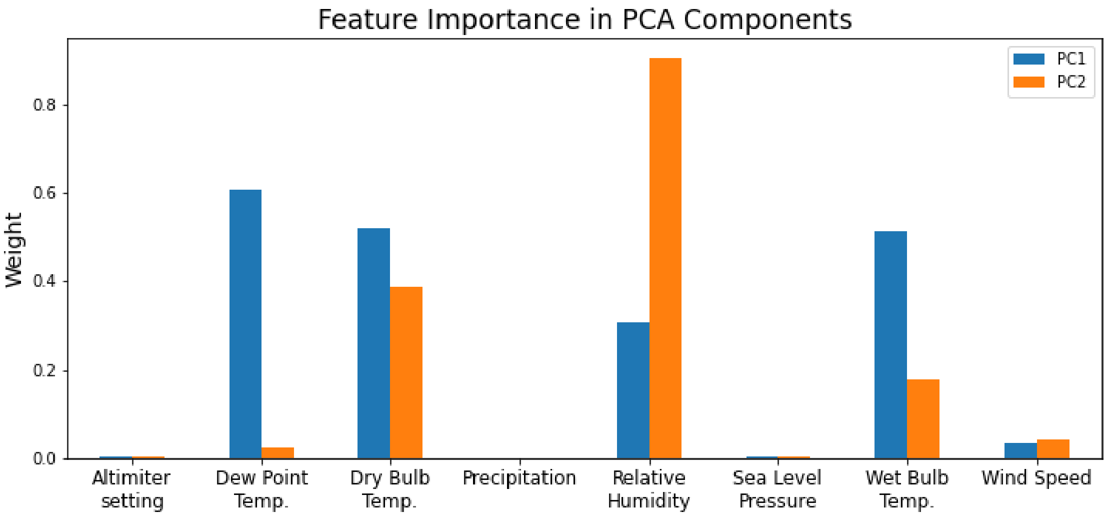


Figure 4: The bar chart shows the weight attributed to each weather feature for the first two principal components. These components assign most of the weight to the three temperature features and relative humidity.

Additive Regression with Facebook Prophet

A few different models were created before to tune the parameters of the Prophet implementation. An initial glance at predictions shows that the model was able to capture enough of the seasonality to make hourly forecasts without the confidence interval blowing up (Figure C.1).

The mean absolute percent error (MAPE) of the model was acceptable for this model. Depending on the time period, the model was able to produce very accurate predictions. On the hourly level (the scale at which predictions are made), the model had a MAPE of 7.45%. However, across the 16 cross-validation periods, each of which forecasted 90 days of energy demand, the 3-month MAPE was much lower, at roughly 2.35%.

Hourly	Daily	Monthly	3 Months
7.45%	6.22%	3.97%	2.35%

Table 3: Results of the Additive regression model based on the granularity of the evaluation window.

The reason for this can be seen by examining the residuals (Figure C.4). The model's errors are well behaved; they follow a mostly normal distribution that is centered very close to 0. The median error was 23.88 MWh, which is less than 0.2% of the average energy load. With the error being mostly balanced, any under-estimated predictions were mostly offset by over-estimates.

One concern with this model is that it relies on weather data to make predictions. However, when forecasting future energy consumption, the hourly weather data is unknown. This becomes more unclear when considering this model uses principal components as opposed to distinct weather features. In these forecasts, future weather values were assumed to be the average value–over a three hour rolling window for that specific hour, day, and month–of the principal component over the nearly 20 years of training data. While this adds a level of uncertainty to the model's predictions, we claim that this will not affect the predictions very much. Figure C.5 shows that the prediction residuals were nearly identical for this model and a model that excluded weather regressors altogether. In the sans-weather model, the hourly MAPE only rose to 7.78%, and even had a 3-month MAPE of 2.32%, slightly *better* than our final model.

Another good sign from this model was that its prediction error stayed relatively constant throughout the forecast range (Figure 5). This means that the model can be used with the same confidence for predicting energy consumption 10 days out as it can be predicting 90 days out. One limitation, however, is that the model tends to forecast conservatively when there are large spikes in energy load. Figure C.5 shows that the density of high energy load predictions was smaller than the true frequency of high-load observations.

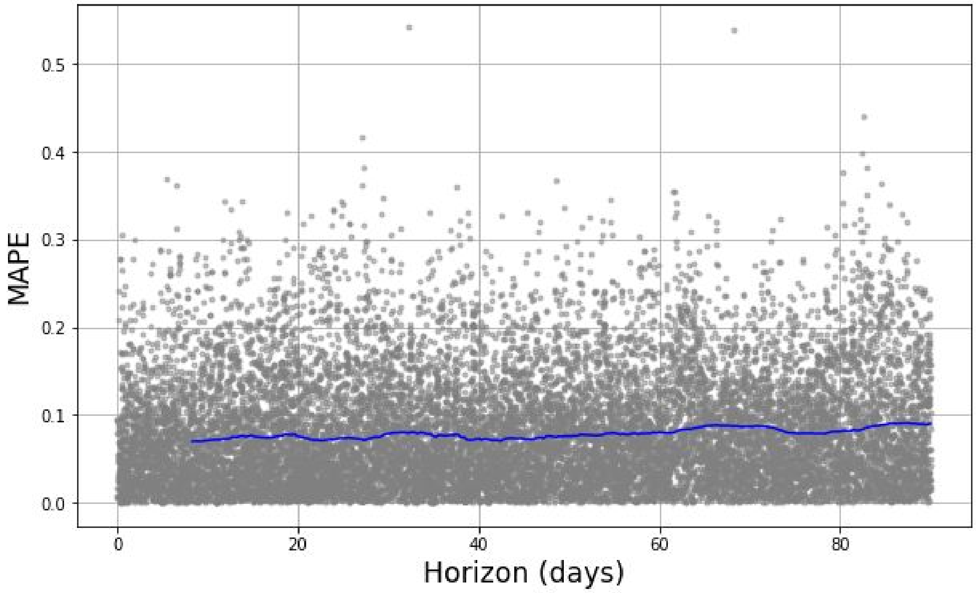


Figure 5: The absolute percent error of each observation in the 16 cross validation periods, plotted as a function of their time into the future. Observations on the left are predictions closer to the time of forecast, while observations plotted on the right are predictions farther away from the time of estimation. The blue line is the rolling average of the absolute percent errors.

Feedforward Neural Network

An analysis of the feedforward neural network's output shows promise but fundamental errors in its predictive power. Viewing the residual plot in Figure 6 shows the nature of the issue–although producing a favorable MAPE, an instability emerges in the MAPE rolling average. This portrays an inconsistency in the accuracy. Overall, the model returned an hourly MAPE of 9.12% over 3 months.

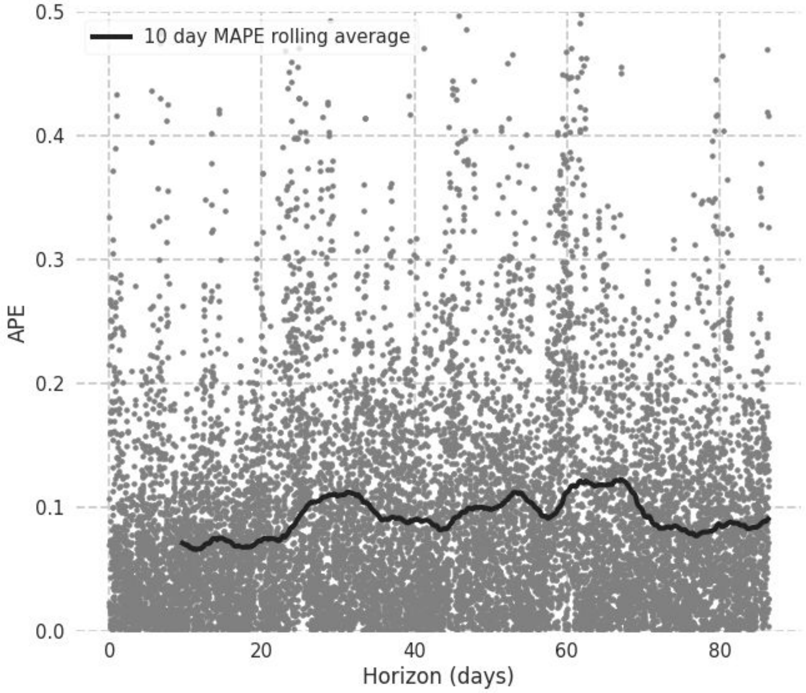


Figure 6: The absolute percent error of each observation over a timespan of 2 years with the feedforward neural network model, similar to Figure 5.

Although stated with uncertainty, through analysis of the output of the model, the nature of the neural network's inaccuracy stems from a tendency to overpredict the true values (see Appendix D.2). Analysis of the unnormalized residuals shows a median of +200.03 MWh, further confirming its tendency to overpredict. Previous versions of this model overpredicted due to incorrectly incorporating weekend behavior; this model doesn't exhibit such behavior, yet overpredicts values on a seemingly arbitrary basis (see Figures D.3, D.4), which is likely explained by poor fitting.

Due to limited cross-validation (described in **Methodology**), difficulty was had in properly tuning the feedforward net. This likely produced the unstable MAPE shown in Figure 6. Not only could the model's hyperparameters not be tuned, training was difficult after 1000 epochs, even though the model could potentially train further (see Figure D.1). Further tuning and effective cross-validation will need to be performed before appropriate testing and evaluation can be done on the neural network's forecasting power for energy demand.

Recurrent Neural Network

We also made an attempt to use a recurrent neural network [14]. The algorithm employed 3 LSTM and 2 Dense layers. Each layer, except for the last, was followed by a 20% dropout rate and all LSTM layers had batch normalization. The model was trained on 45 parameters, as discussed in the previous subsection. We observed that although the training loss stabilized and was below the validation loss (Figure E.1), the trained model failed to adequately forecast future energy demand (Figure E.2). Forecasting performed in cross-validation yielded a slightly better result (Figure E.3). However, the improvement was not enough to warrant an in depth statistical analysis of the model and its results.

Conclusions

Overall, we found that the Additive Regression model was able to predict the future energy load with a higher accuracy than the Neural Networks. Our target was to predict energy demand within a 5% error. On the hourly level, our best model was only able to achieve a 7.45% MAPE. However, when predicting cumulative energy demand over a 3 month period, the best model was able to achieve a 2.35% MAPE. While there is still room for improvements, this model is successfully able to forecast up to 3 months in advance.

While statistically not the most successful model created in this analysis, the feed-forward neural networks would benefit from numerous improvements. These include but are not limited to tuning the

- number of layers
- activation function
- density of the net
- size of the validation data.

This paper is not a statement of the failure of the neural net, but of the room for its improvement.

Future Work

Further work needs to be done to further verify the power of these models. In particular, we desire two such advancements:

1. Expand the weather features by testing on forecasted weather or incorporating average weather deviations from their yearly means. This is desired since we will not have real time weather data when forecasting on future energy. Overall this will return a more realistic test error with error bands. [15] shows weather measurement error, and [16] shows annual weather deviations for selected stations.
2. Prove transferability by testing our models on load series of different spatial scales. This is our ultimate goal, since the spatial scale can change extensively between different loads. Showing the models work just as well on the city scale as on the multi-state scale will be vital in proving the models are powerful for all load forecasting.

Link to GitHub repository:

https://github.com/takao8/MLTSAProject

Bibliography

[1] B. E. Psiloglou, C. Giannakopoulos, S. Majithia, “Comparison of energy load demand and thermal comfort levels in Athens, Greece and London, UK”, 2nd PALENC Conference and 28th AIVC Conference on Building Low Energy Cooling and Advanced Ventilation Technologies in the 21st Century, September 2007.
[2] Eisa Almeshaiei a, Hassan Soltan, “A methodology for Electric Power Load Forecasting”, Alexandria Engineering Journal vol. 50, pp. 137–144, July 2011.
[3] Y. Ohtsuka, T. Oga, K. Kakamu, “Forecasting electricity demand in Japan: a bayesian spatial autoregressive ARMA approach,” Computational Statistics and Data Analysis, vol. 54, pp. 2721–2735, 2010.
[4] O. Valenzuela, I. Rojas, F. Rojas, H. Pomares, L.J. Herrera, A.Guillen, L. Marquez, M. Pasadas, “Hybridization of intelligent techniques and ARIMA models for time series prediction,” Fuzzy Sets and Systems, vol. 159, pp. 821–845, 2008.

[5] L.F. Amaral, R.C. Souza, M. Stevenson, “A smooth transition periodic autoregressive (STPAR) model for short-term load forecasting,” International Journal of Forecasting, vol. 24, pp. 603–615, 2008.
[6] J.V. Ringwood, D. Bofelli, F.T. Murray, “Forecasting electricity demand on short, medium and long time scales using neural networks,” Journal of Intelligent and Robotic Systems, vol. 31, pp. 129–147, 2001.
[7] H. Wang, B.-S. Li, X.-Y. Han, D.-L. Wang, H. Jin, “Study of neural networks for electric power load forecasting”, in: J. Wang et al. (Eds.), ISNN 2006, LNCS 3972, Springer-Verlag, pp. 1277–1283, 2006.
[8] D.-X. Niu, Q. Wang, J.-C. Li, “Short term load forecasting model based on support vector machine,” in: D.S. Yeung et al. (Eds.), ICMLC 2005, LNAI 3930, Springer-Verlag, pp. 880–888, 2006.
[9] D. Niu, Y. Wang, D.D. Wu, “Power load forecasting using support vector machine and ant colony optimization,” Expert Systems with Applications, vol. 37, pp. 2531–2539, 2010.
[10] B. Yildiz, J.I. Bilbao, A.B. Sproul, "A review and analysis of regression and machine learning models on commercial building electricity load forecasting," in IEEE Transactions on Power Systems, vol. 73, pp. 1104-1122, June 2017.
[11] R. Weron, “Modeling and forecasting electricity loads and prices: a statistical approach,” John Wiley & Sons Ltd., England, 2006.
[12] R. G. Steadman, “The Assessment of Sultriness. Par I: A Temperature-Humidity Index Based on Human Physiology and Clothing Science”, Journal of Applied Meteorology, vol. 18, pp. 861-873., 1979.
[13] D. C. Park, M. A. El-Sharkawi, R. J. Marks, L. E. Atlas and M. J. Damborg, "Electric load forecasting using an artificial neural network," in IEEE Transactions on Power Systems, vol. 6, no. 2, pp. 442-449, May 1991.
[14] Dupond, Samuel, "A thorough review on the current advance of neural network structures". Annual Reviews in Control, vol. 14, pp. 200–230, 2019.
[15] Frank, Patrick. (2010). Refereed Papers: Uncertainty in the Global Average Surface Air Temperature Index: A Representative Lower Limit. Energy & Environment. 21. 10.1260/0958-305X.21.8.969.
[16] Bailey, H. (1966). The Mean Annual Range and Standard Deviation as Measures of Dispersion of Temperature around the Annual Mean. *Geografiska Annaler. Series A, Physical Geography, 48*(4), 183-194. doi:10.2307/520501
[17] https://en.wikipedia.org/wiki/Moss_Landing_Power_Plant

Appendices

Appendix A: Data Exploration Figures

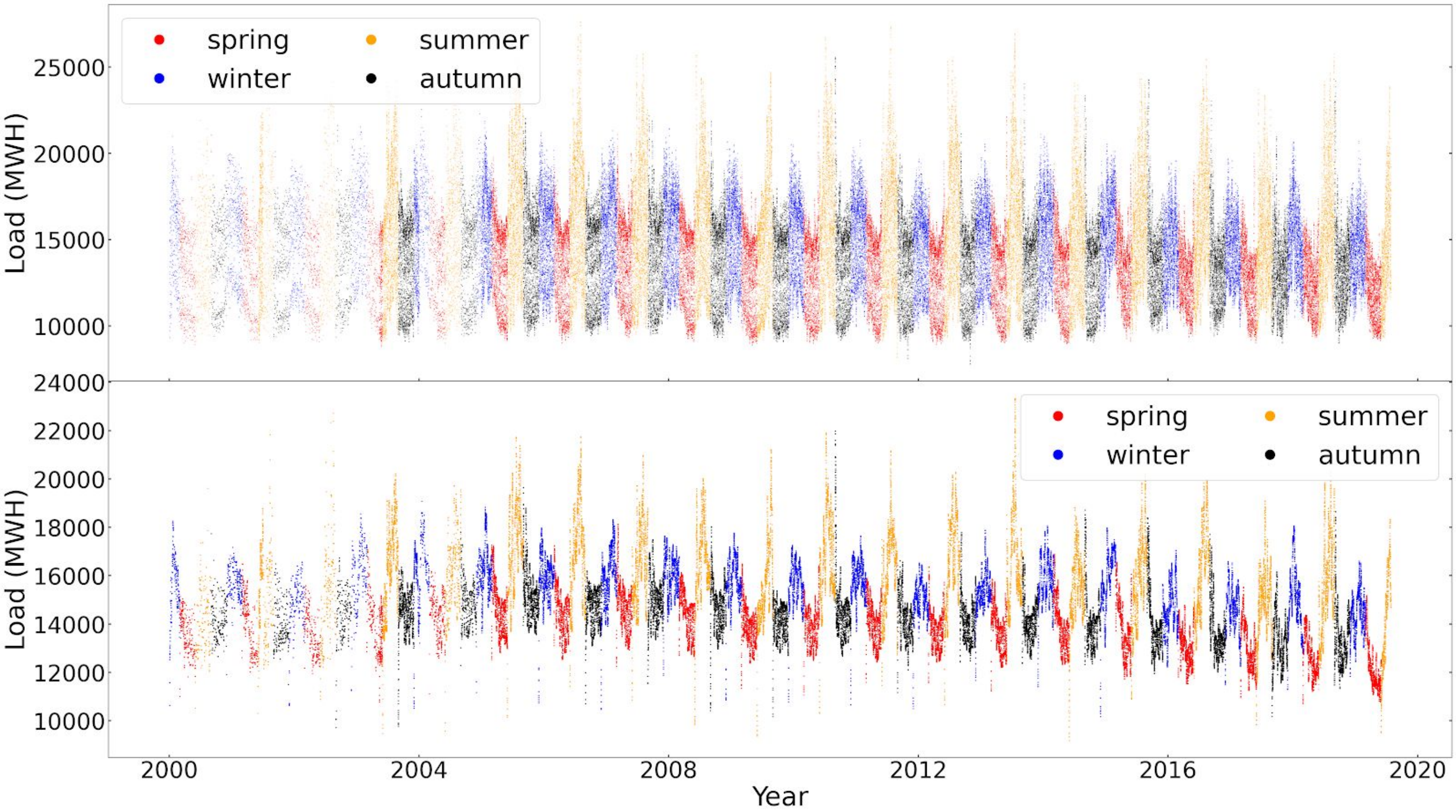


Figure A.1: An overview of load corresponding to 19 years data we used in the project. Energy consumption is usually higher in winter and summer compared to the other two seasons. Summer though exhibits a slightly different behaviour compared to other seasons as the peaks are much sharper and the total load is also significantly higher than winter resulting in higher variance of the data which might help explain the higher error in load prediction as reported in Results. Bottom panel shows the same data though with rolling mean for a period of around 4 days. Please note the paucity of data in the initial years, specially upto 2006 which gives an impression of varying ***α*** (transparency parameter) while plotting. This is only because of missing data though cadence is the same as any other period.

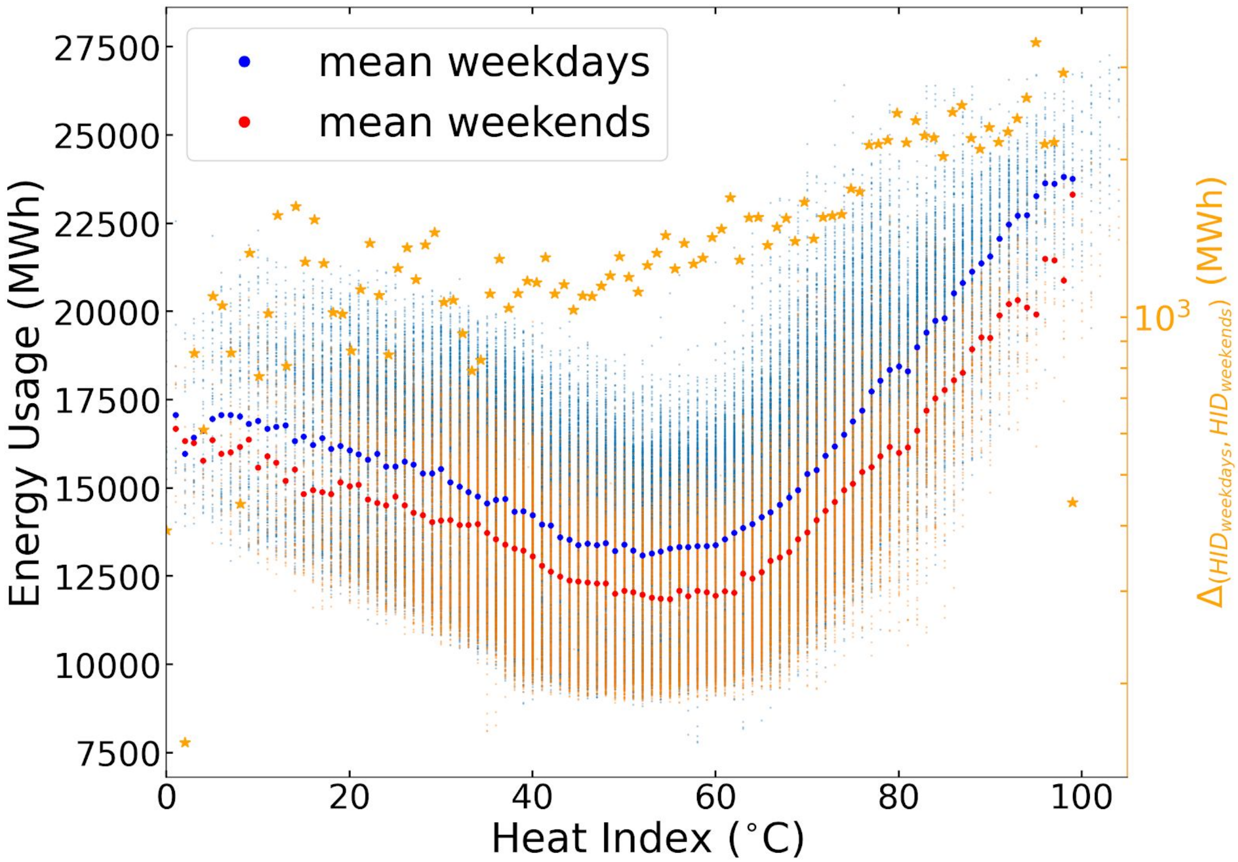


Figure A.2 : Plot of Boston heat index and energy consumption for the New England area. Blue dots correspond to weekdays whereas red dots correspond to weekends. Energy consumption during the weekend is slightly lower than those during weekdays for all seasons. The difference of average energy consumption (shown by green dots) during weekdays and weekends is always positive and shows a log-linear relation with heat index. Please note the vertical stripe like feature of the data points is the consequence of digitization of data as heat index is recorded only as integers.

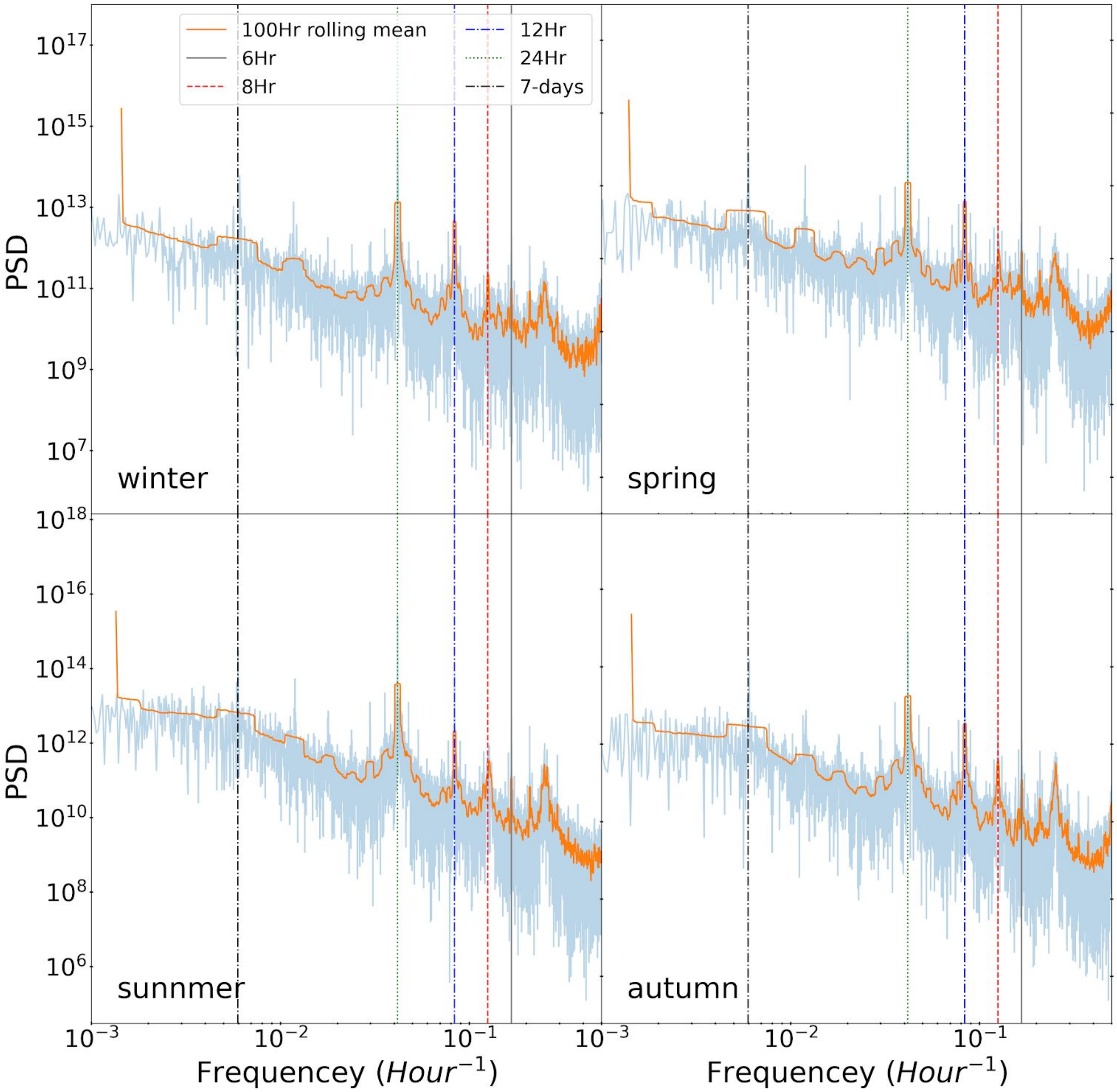


Figure A.3: Power spectral density of load during different seasons for the New England area. The figure shows peaks along expected periods of 24 Hrs (dotted green line) and a small broad peak centered around frequency corresponding to 7 days (dash-dot black line). Broadness of the peak at week long frequency can be attributed to the drop in power consumption over weekends as compared to a weekday. There are other peaks present at frequency corresponding to 8 hr, which relates to the usual 8 hr long office hours in the corporate sector.

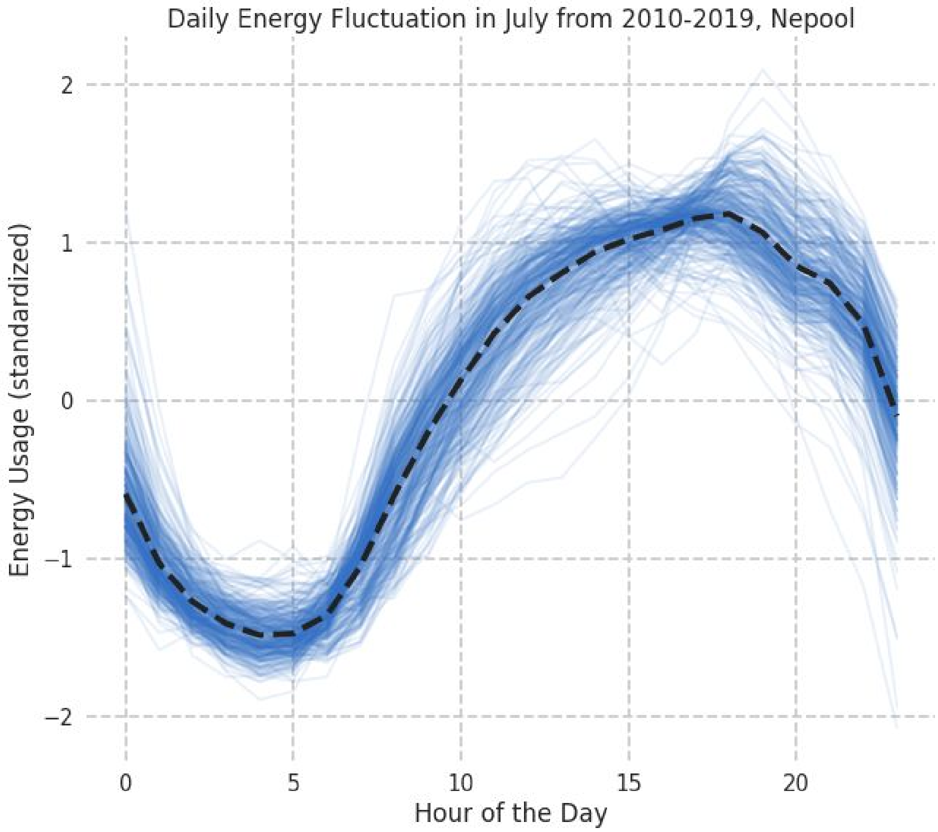


Figure A.4: Ensemble of daily oscillations in the month of July of the energy demand data for 9 years. Daily oscillations in the summer are distinguishable from the winter by their single hump--a steady increase through the day, peaking around 5PM, then steadily decreasing into the night.

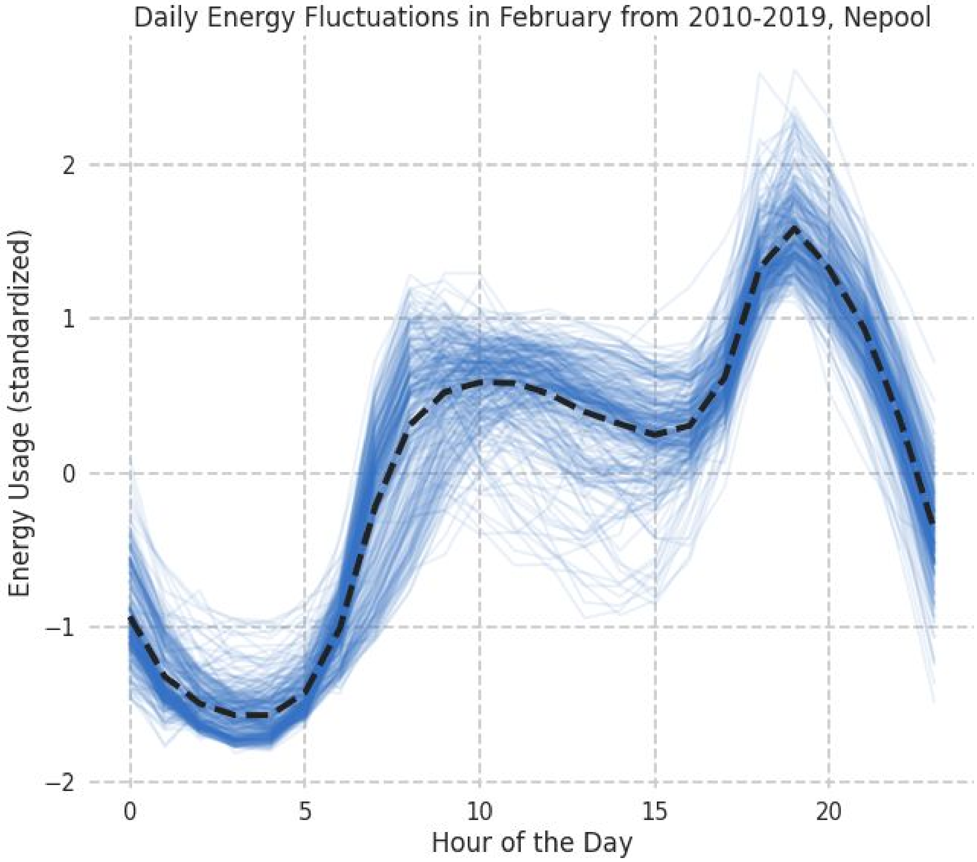


Figure A.5: Similar to figure A.4, except during the winter. Note that the daily oscillations are now defined by a double hump: an increase through the early hours of the day, slightly decreasing until 3PM, then spiking sharply at 7PM until decreasing through the night.

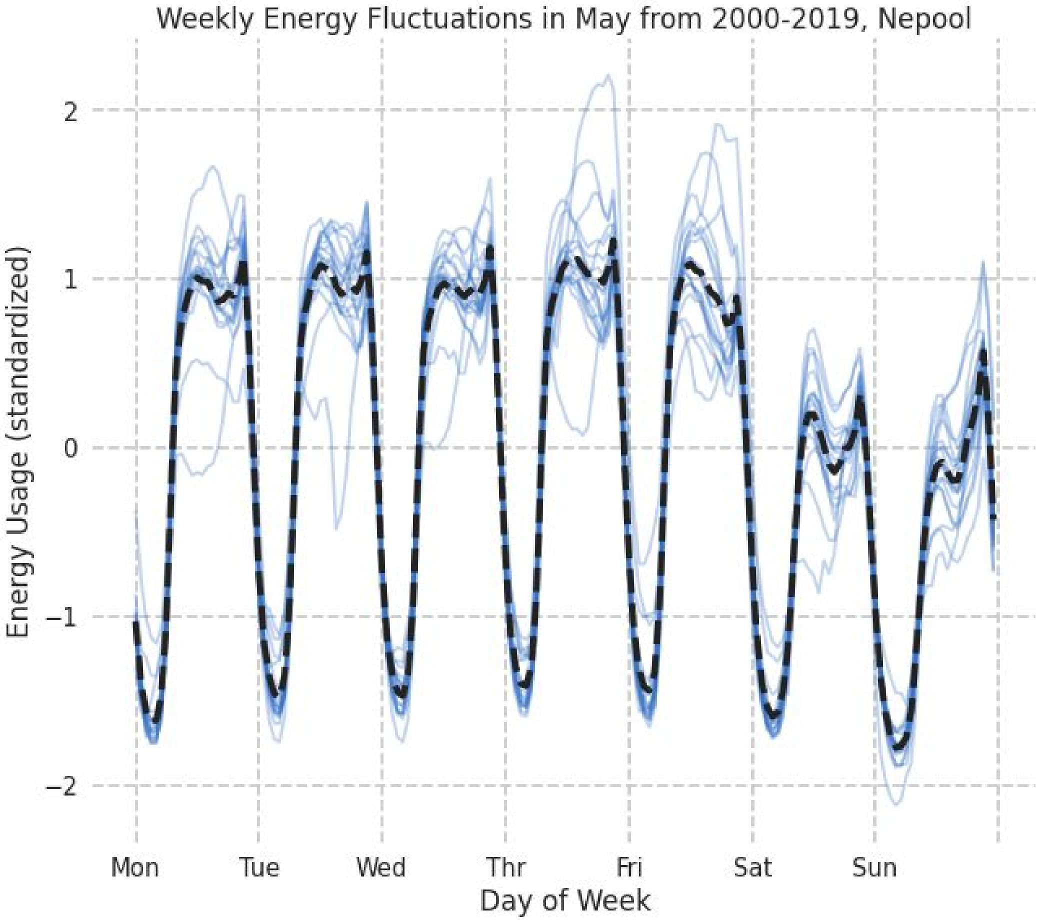


Figure A.6: Behavior of energy demand in May over 19 years of data. Energy demand over the weekends tends to be lower due to industry and commercial buildings shutting down. Therefore, a typical weekly oscillation will show energy demand higher during the weekdays and lower in the weekends. Although there are exceptions to this behavior, the average, as shown in the dotted black line, illustrates this trend to be true.

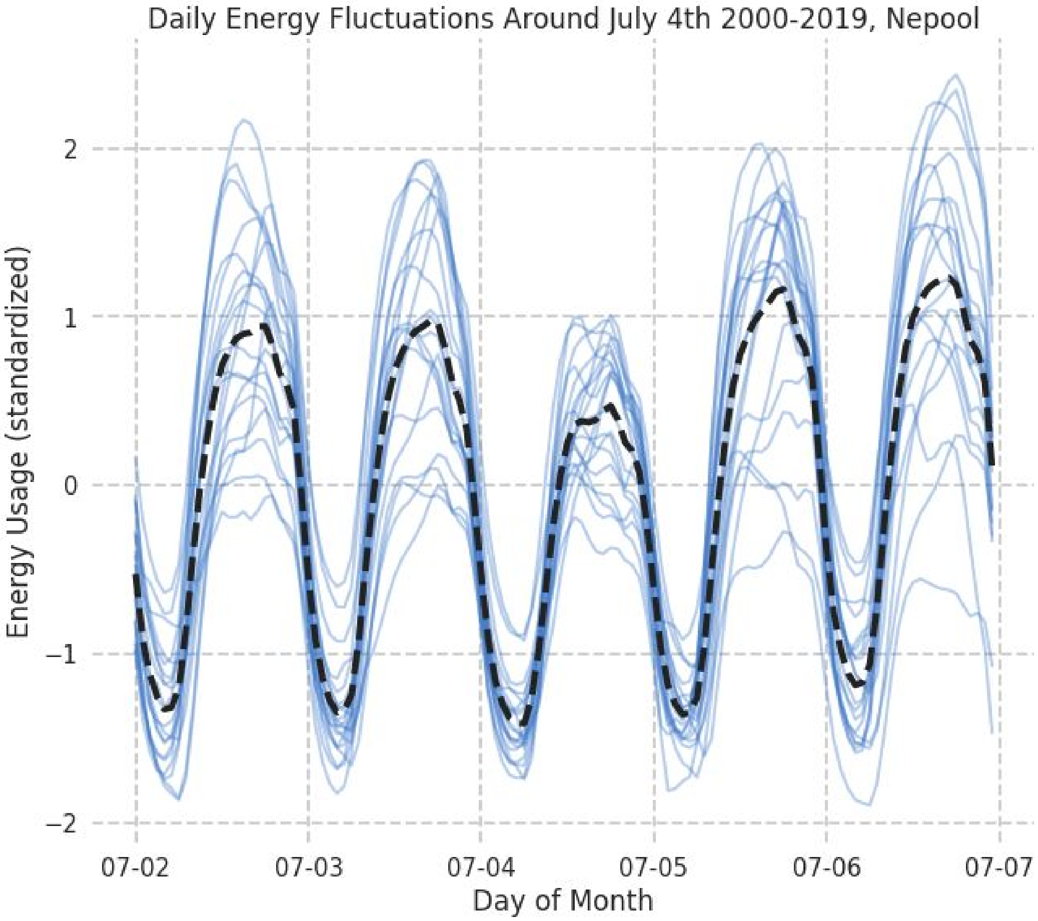


Figure A.7: Behavior of the energy demand for Nepool around July 4th. Typically, holidays act identical to weekends, as businesses and industry close down.

Appendix B: Principal Component Analysis Results

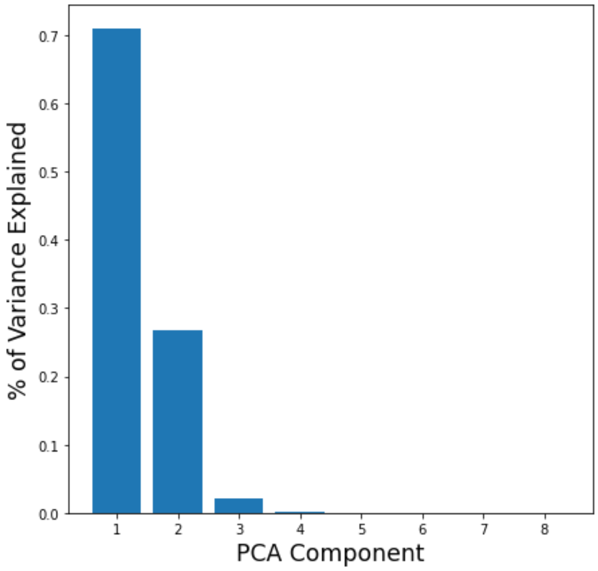


Figure B.1: The bar chart shows how much of the total variance in the original weather dataset is explained by each principal component. The first two components explain 97.65%, the first three explain 99.70%, and the first six explain 100%.

Appendix C: Additive Regression with Facebook's Prophet Results

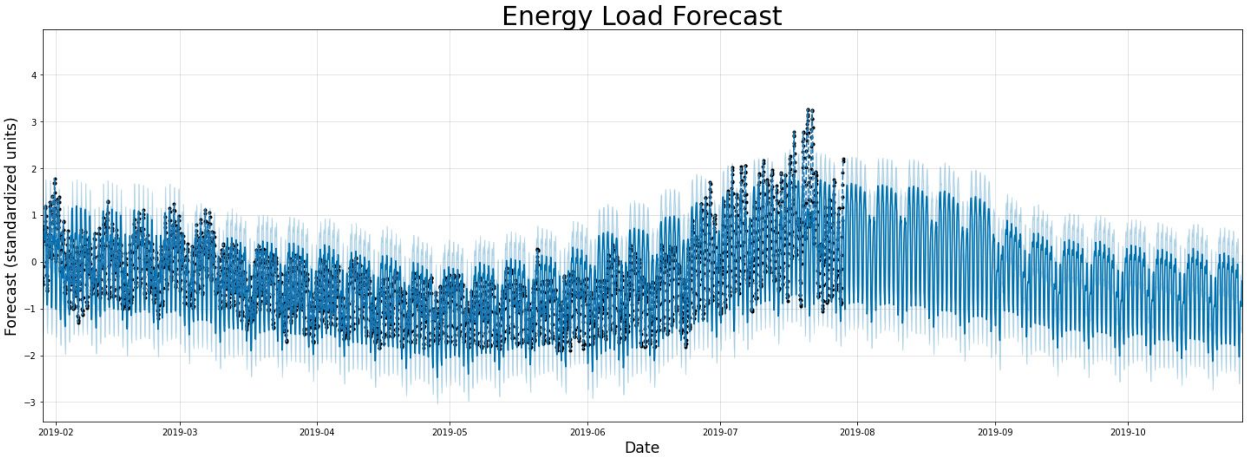


Figure C.1: A high level view of the model's forecast shows that the uncertainty does not grow over time. The seasonalities learned by the model are strong enough to maintain a tight confidence interval throughout the duration of the prediction period.

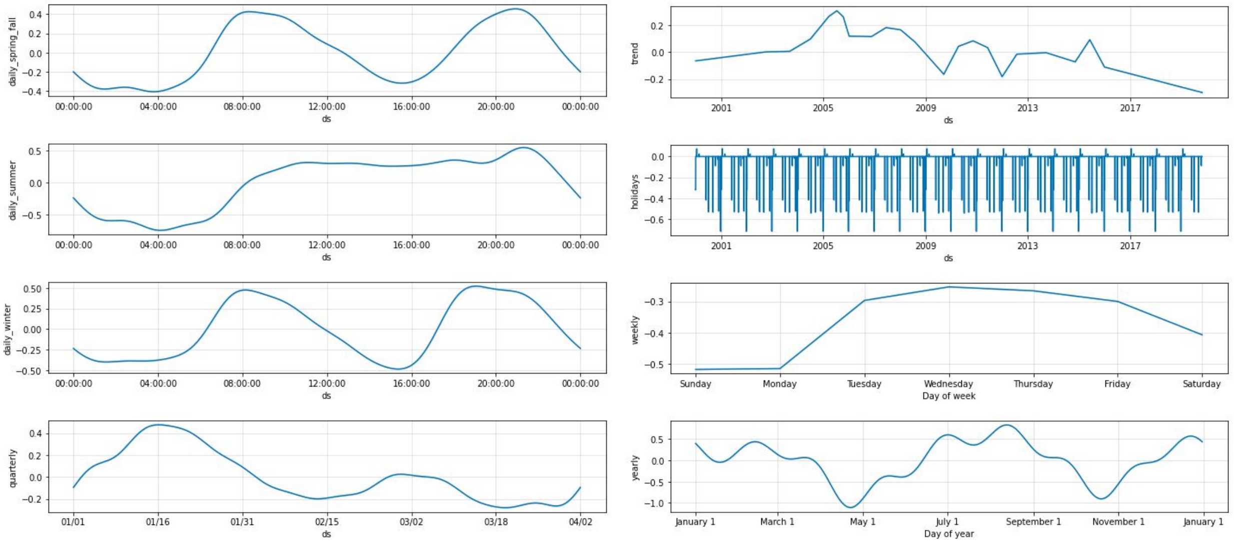


Figure C.2: These are the different components of the additive regression model. From top to bottom, right to left: *daily spring/fall, daily summer, daily winter, quarterly, trend, holidays, weekly, yearly.* Though the methodology behind Prophet's optimization is a relatively "black box", we can plot the different seasonality components of the trained model and explore what trends have been learned. By observing the above figures, some glaring observations can be made. First, the captured daily seasonality was vastly different between the summer months versus the winter and spring/fall months. This was expected, as we saw the winter months

exhibit a “double peak” in energy load, while the summer months did not. The overall trend (top right figure) is not very telling; there appears to be a gradual decrease in energy consumption over the last 15 years, but the chart is noisy and comes after a few years of increasing demand. The yearly trend is very noisy and very likely overfit the training data. The fourier order is too high, even though this component was learned using Prophet’s default setting. This will have implications on the results and marks an area for future models to be improved.

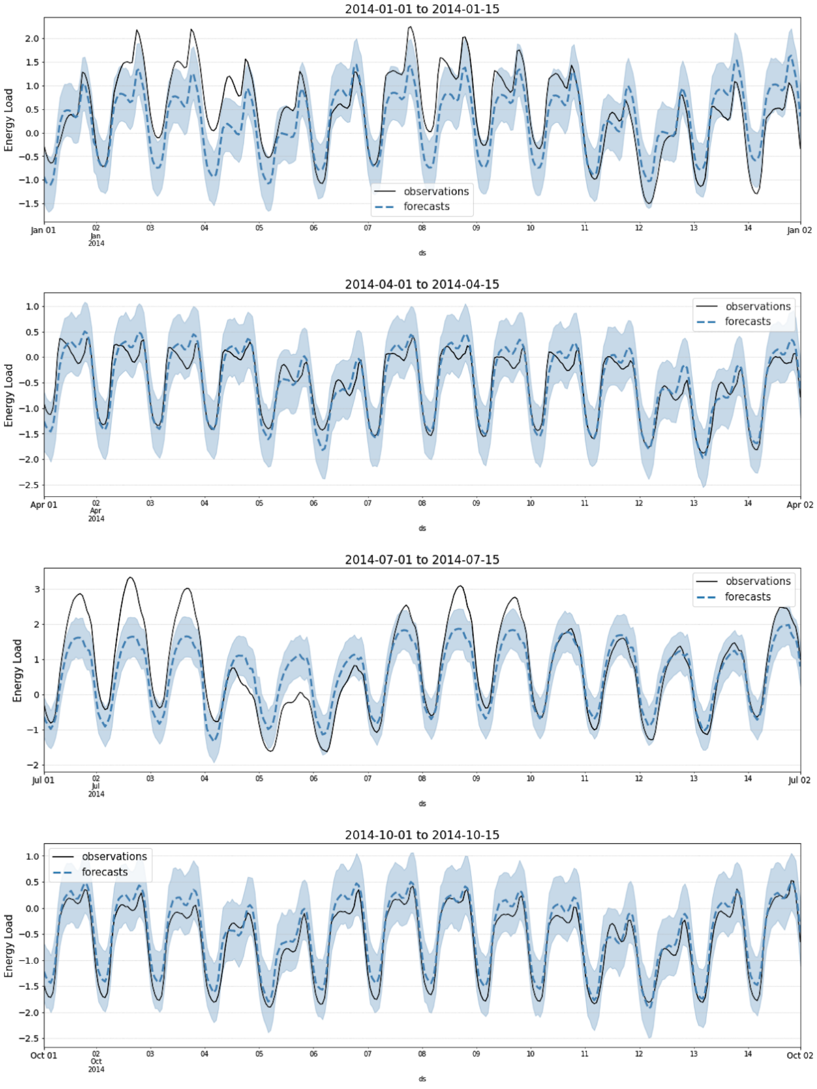


Figure C.3: A zoomed in look at the energy forecasts versus their actual values for the different seasons of the year. These plots show segments of the predictions in January, April, July, and October (listed from top to bottom), to get an idea of the yearly accuracy of the model. The best predictions came in the spring and fall months, while the least accurate predictions came in the summer months.The model was able to properly associate “double peaks” with the winter, spring, and fall months and a “single peak” with the summer months, but the amplitudes of the summer peaks were not predicted with as much precision.

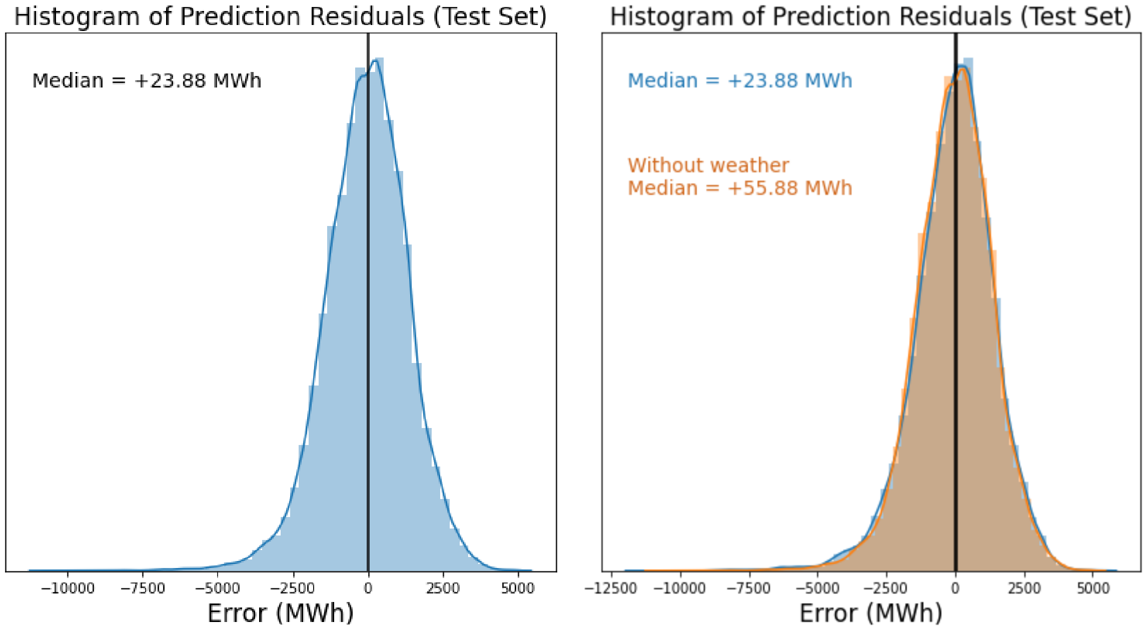


Figure C.4: A histogram of the residuals using the model (left), and a histogram comparing the residuals of the main model with a model that does not consider the weather features in its prediction-making (right). The distributions are both well-behaved and nearly identical. Note that 24 MWh is smaller than 0.2% of the mean hourly energy load.

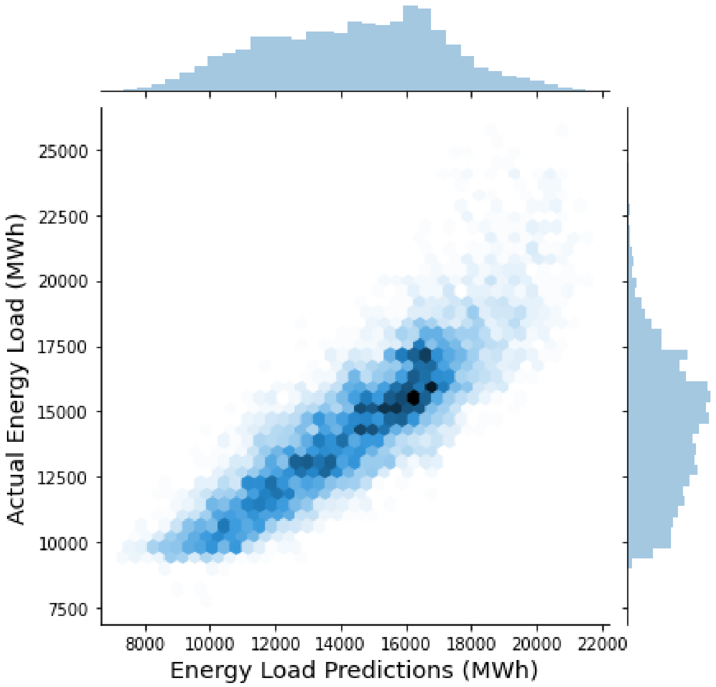


Figure C.5: The joint plot shows the actual energy loads versus the energy load predictions. The marginal histograms show the distributions of the energy loads. The correlation is linear for most of the range, tailing upwards at the right end of the plot. This implies that the additive model tended to underpredict larger energy demands.

Appendix D: Neural Network Results

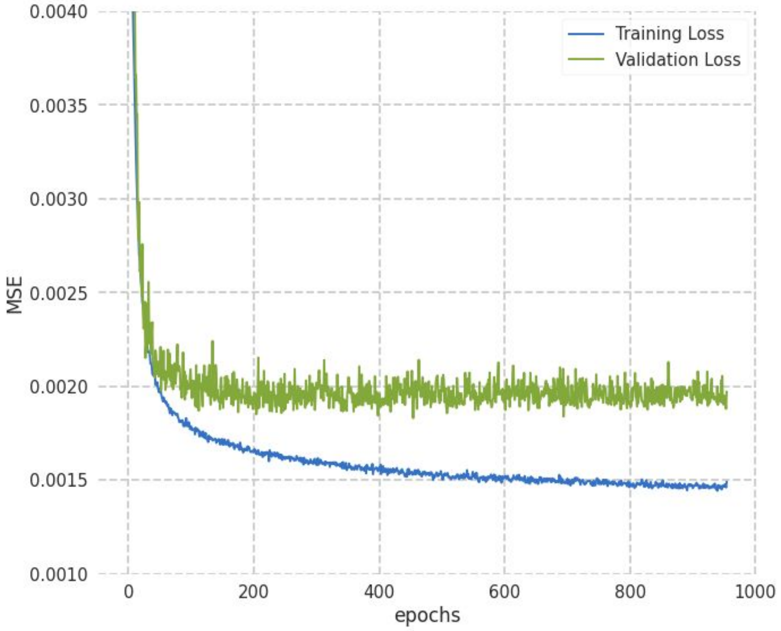


Figure D.1: Plot of the training and validation MSE decreasing as epochs continue. Training data will continue to decrease as the model overfits the training data, so minimization of external validation data is necessary to properly fit the model. Note for our purposes that training past 1000 epochs was unobtainable due to computational limits, thus it was common to need to stop before the limit, even if the validation shows potential signs of further decrease.

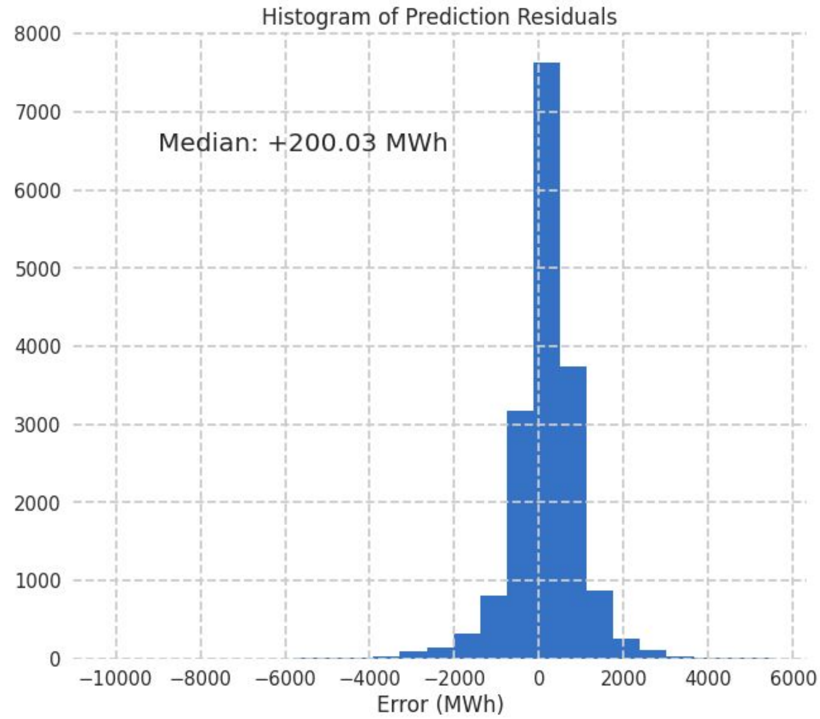


Figure D.2: Histogram of residuals of resulting fits in the feedforward neural network. Note a heavy right side on the histogram, indicative of the model's tendency to overpredict results, explaining the resulting high median.

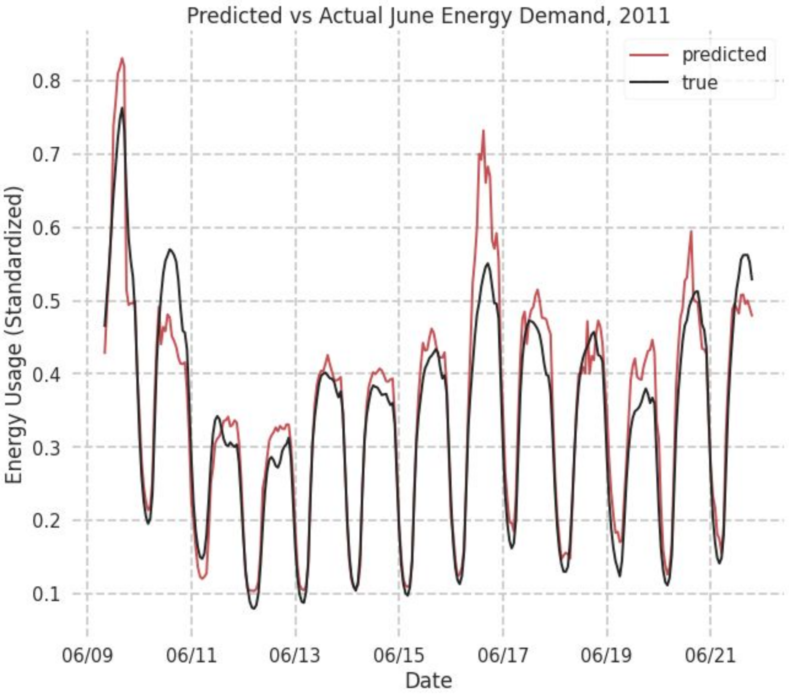


Figure D.3 Line plot showing the true series (black) overlayed on the predicted series (red) over June, 2011. Previous models notably had a tendency to overfit on weekends--the current

model overfits arbitrarily. No explanation besides a lack of a good fit explain why this occurs in the model.

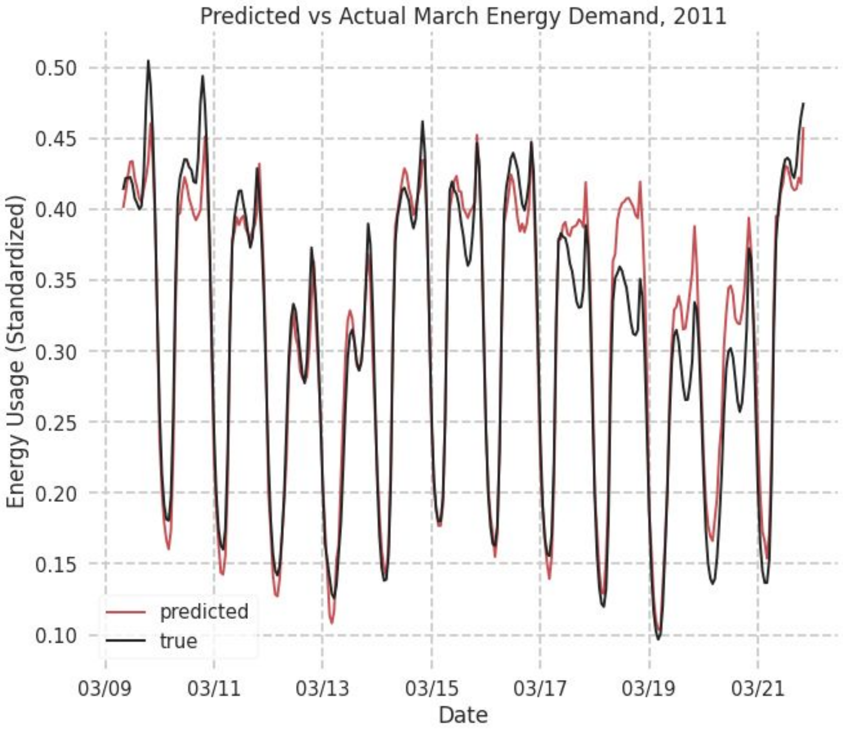


Figure D.4 Similar to Figure D.3, over March 2011. This plot shows the model accurately forecasts lower energy demand on weekends (March 19th-20th), thus cannot fully explain its tendency to overpredict the model.

Appendix E: Recurrent Neural Network Results

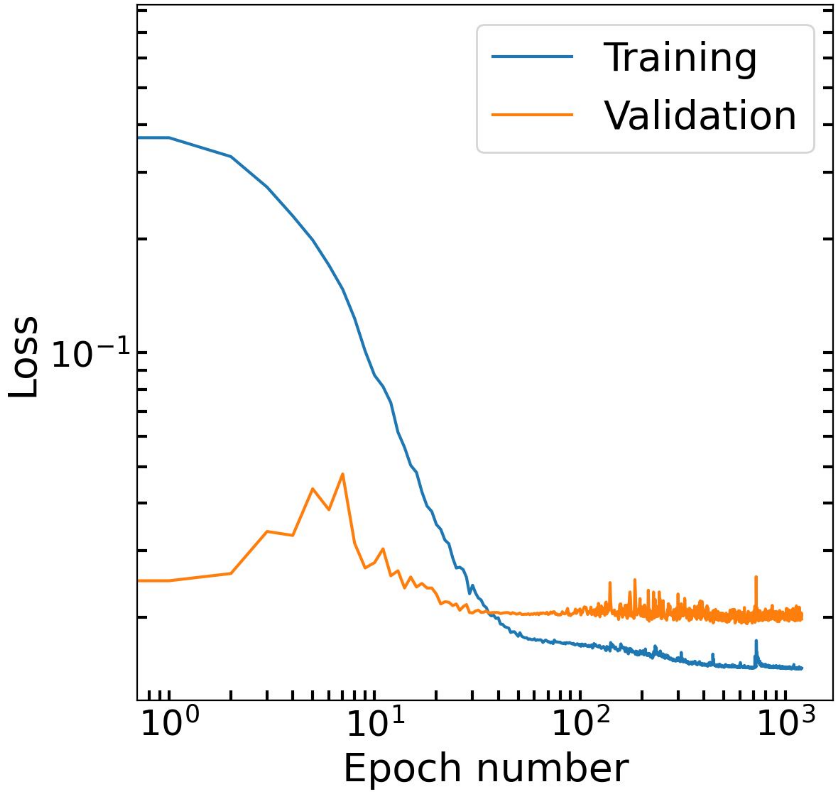


Figure E.1 : Plot of loss rate for training and validation sets for an RNN.

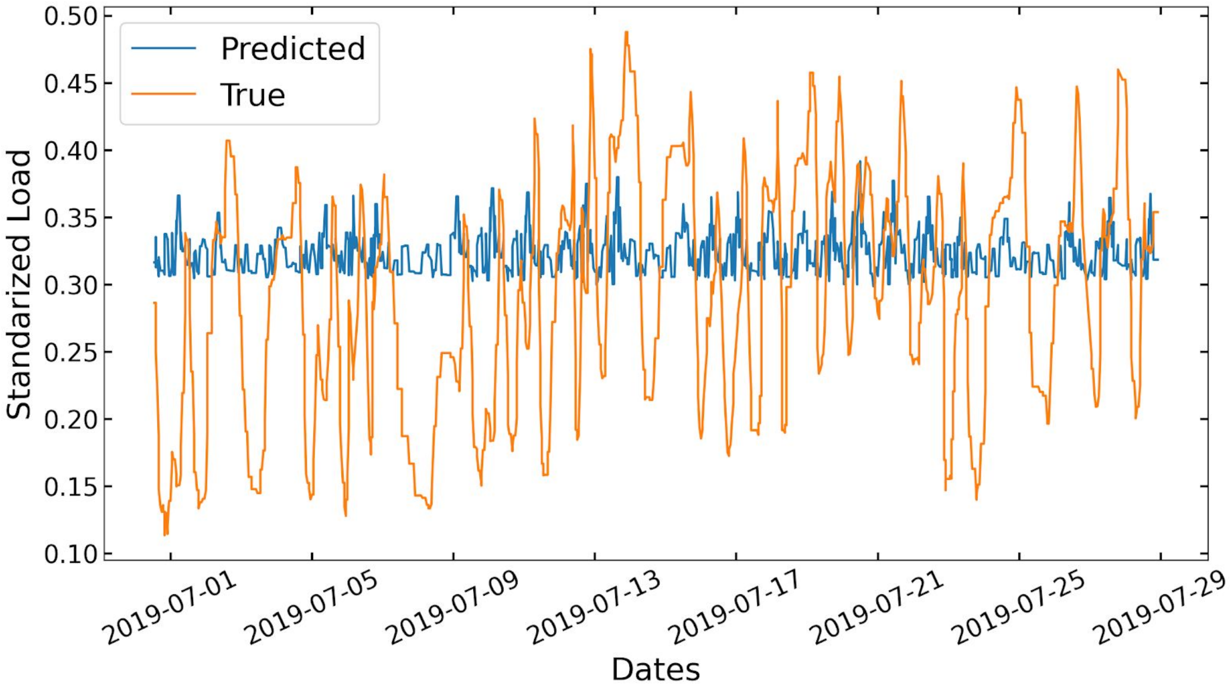


Figure E.2: Plot of predicted and true values for load (standardized) based on the RNN model using future dataset.

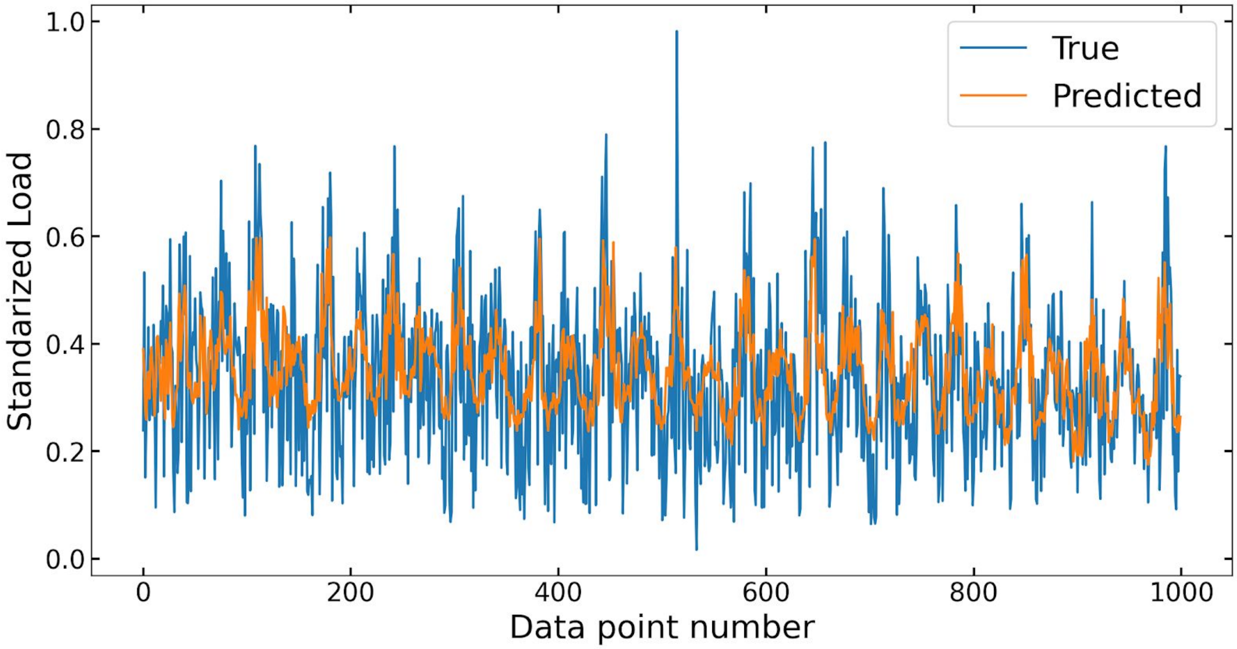


Figure E.3: Plot of predicted and true values for load (standardized) based on the RNN model using a subset of data from the training set.