Global Association of Risk Professionals

University of Texas at Dallas

Department of Finance

Electricity Load Forecasting Using Machine Learning and Air Thermodynamics applied to New York City

Gabriel Castaneda

1. Introduction

This project seeks to provide a method to forecast electricity loads for a defined homogeneous region, for example, the city of New York. Homogeneous region means an area that is affected by the same variables and about the same values. An area which population is rapidly growing and is been developed would be hard to track. If I chose to model a full state, i.e. the state of California, weather conditions would be largely different within the state. There are states such as Tennessee that cover two time zones, so time, a key variable, would not have the same value along the state.

With the use of first principles, such as air thermodynamics, the use of accurate weather forecasts, such as the ones from the Weather Prediction Center, and the latest machine learning algorithms, it is intended to improve the price of the electricity that is bought or sold for a certain region. If companies can forecast electricity loads more accurately, over a longer time window, they may be able to reduce their energy expenditures by avoiding the premium paid by last-minute electricity purchases. At the same time, if a power company knows it will have extra capacity in the future, it will have more time to find clients and obtain a better price. This ability is expected to be very useful for power companies, as they would minimize costs to load customers.

The literature has well documented that power consumption is correlated with many factors. Hiroto Shiraki et al[1] proposed to include demand hour, hourly temperatures, house characteristics in the hybrid model and conduct a conditional demand analysis based on appliances used. The conditional demand analysis introduced the concept that electricity demand changed by the hours according to people's activities. Differences between residential and industrial demand, building characteristics and climate were proposed by Mahmound Safari et al[2]. Socio-economic data was also taken into consideration. Dynamic models of residential electricity demand were proposed by Nina Boogen et al [3]. They estimate short and long run elasticities of electricity demand using a dynamic model of residential electricity consumption. Yanming Sun et al [4] estimated a dynamic partial adjustment model for electricity demand elasticities on price and income in the residential sector. It acknowledged a difference between short and long run elasticity behaviors. The integration of long term economic scenarios into peak load forecasting was introduced by Julian Moral-Carcedo et al [5]. These authors proposed a methodology that allowed to produce accurate long-term forecasts of electricity loads an hourly frequencies based on multiple linear regression models that account for variables such as economic activity or temperature. Siyan Wang et al [6] used a Bayesian regression model for predicting summer electricity demand. Sigauke et al [7] predicted daily peak electricity demand using volatility forecasting models such as a seasonal ARIMA or a SARIMA-GARCH model. The use of quantiles of day ahead electricity loads were explored by Li, Hurn and Clemens[8]. Neural Networks without using weather as input, but using advanced wavelet decomposition and nonlinear algorithms were tried by Rana and Koprinska [9]. Hybrid gene expression programming and cloud computing patterns were explored by Song Deng et al. [10] . The use of probabilistic load forecasting to generate residential forecasts was studied by Shepero [11] A hybrid model using empirical mode decomposition (IEMD), ARIMA and wavelet neural networks optimized by fruit fly optimization algorithms (FOA) was also proposed by Jinlinag Zhang et al. [12]

Nowadays, air conditioning systems consume a major proportion of the electricity load, approximately 18% in US homes, while overall residential cooling consumes 24%, according to 2018 data from the Energy Information Administration, EIA[13].

This study proposes the use of accurate weather forecasts as feed to machine learning algorithms to forecast electricity demand. The variables pulled from the forecast models are hourly air data such as air humidity, barometric pressure and temperature, since they are the main variables of air conditioning. Day lengths are also fed to the mix. This study compares the results of three different machine learning models. A comparison to other forecasting methods such as ARIMA, based on time series, and multiple linear regression is done. The paper is organized in the following sections. Section 2 will give an overview of electricity markets in US. Section 3 will discuss the weather forecast accuracy issues. Section 4 will describe the data and variables. Section 5 will introduce the models. Section 6 will discuss the model results. Section 7 will review virtual contracts in the state and their relation with uncertainty. The paper's conclusions are presented in the last section.

2. Overview of electricity markets in the USA

In a deregulated electricity market, competitors buy and sell electricity in both financial and physical markets.. Generation owners sell electricity to the organized markets such as an ISO or sell through the bi-lateral market. Retail suppliers set price for consumers and consumers may choose between available suppliers.

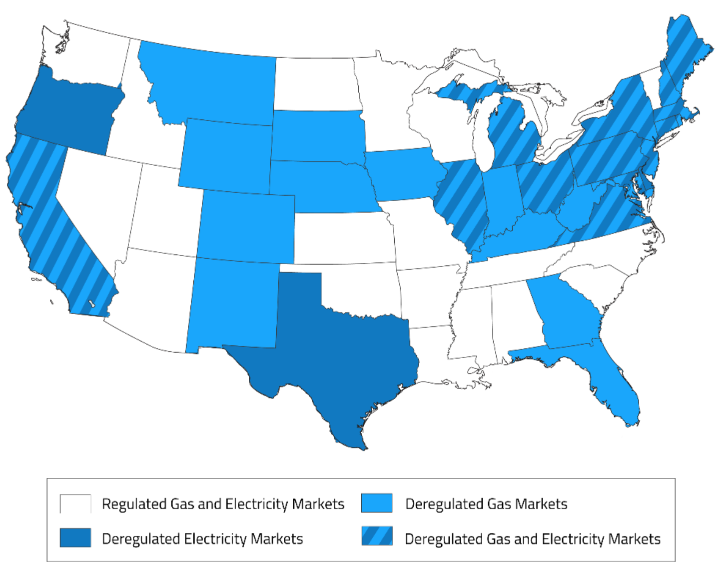

Figure 1. Map of Deregulated Energy States and Markets, 2018 [14]

As it is shown on Figure 1, the electricity and gas markets in the USA vary from state from state. Some states, like CA and NY, have deregulated gas and electric markets. In others, like TX, only the electricity market is deregulated. In FL only the gas market is deregulated, while in OK both the electricity and the gas markets are fully regulated.

Under 89 FERC 61,285 18 CFR Part 35 Docket No RM99-2-000, Regional Transmission Organizations, the Federal Energy Regulatory Commission, FERC, proposed a non-profit ISO system which would direct the operation of the transmission system and run day-ahead and real-time power markets coupled with a grid entity that owns and maintains the transmission in the area operated by the different ISOs. [15]

The intraday and real time markets are managed and operated by independent system operators (ISO). There are seven ISOs in the USA (see Figure 2 below), they are non-profit and their area of services vary. New York ISO, NYISO, covers mainly one state, while Midcontinent ISO, MISO, covers several. Figure 2 also shows that some ISOs cover one state, while others cover several regions which may not be attached to one another.

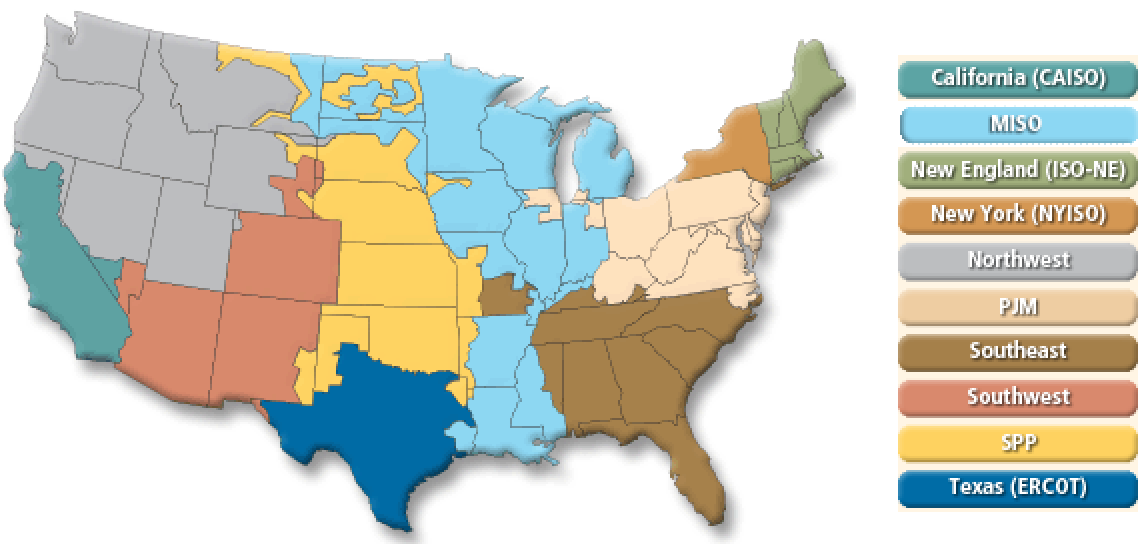

Figure 2. USA Energy Power Markets, FERC.GOV

At the wholesale level, electricity is constantly balanced in real time. The lack of storage and other more complex factors lead to very high volatility of spot prices.

All FERC jurisdictional RTOs and ISOs, deregulated markets, including NYISO, have what is called a dual settlement system, which include a combination of the following:

Day Ahead Energy Market, DAM

In order to hedge some of the volatility, generators and load servicing entities bid 85% - 95% of their load in the Day-Ahead Market. This market produces one financial settlement. If market participants have no deviations from day-ahead schedules, the real-time markets impact is minimal. The day ahead market process balances supply offers (physical and virtual) against demands (physical and virtual).[16]

Real Time Energy Market

The real time energy market is operated by the ISO's and allows market participants to buy and sell wholesale electricity during the course of the operating day, typically in five minute intervals. Prices are determined by real time dispatch conditions. It balances the differences between day ahead commitments and the actual real time demand for and production of electricity. The real time energy market produces a second and separate financial settlement.

In February 2018, members of the Mid Continent ISO, MISO, recommended the development of a multi-day market forecast with target implementation of 2021, the reasoning is that the current day-ahead market construct is not designed to forecast economic commitments beyond the next day, resulting in the inability to economically commit long-lead (or high startup cost) units and could cause uneconomic cycling of certain units across a certain period. [17] In their presentation they quantified significant economic benefits of changing to a multi-day market forecast.

3. Weather Forecasting

Weather forecasting has greatly improved in the last few decades thanks to numerical weather predictions. The 3D, multigrid computer model starts with a current state of the atmosphere and uses mathematical equations that describe both horizontal and vertical air motions, temperature changes, moisture processes, etc. to calculate what the atmosphere might look like at some future time.

In the United States, the National Weather Service's array of supercomputers run several models such as the Weather Research and Forecasting model, the Antarctic Mesoscale Prediction System (AMPS), the Model for Predicting Across Scales (MPAS), and the NCAR Nested Regional Climate Model (NRCM). In forecasting weather, the big computers run the models using all observational data that the National Weather Service (NWS), collects. The data resources are from satellites, weather balloons, buoys, radar and more.

The output variables of these models are quite large including temperature, barometric pressure, relative humidity and other features. These variables will be used as inputs in the machine learning algorithms proposed in this paper.

According to the US Weather Prediction Center, (WPC), temperature forecasts have improved in the last few years.

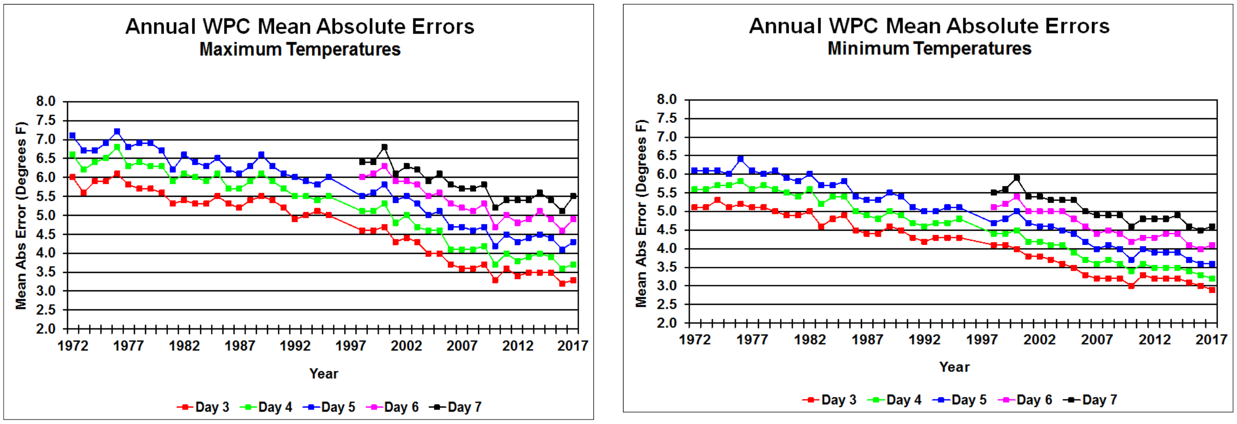

Figure 3. WPC Mean Absolute Errors, Maximum and Minimum Temperatures

As is shown in figure 3, the mean maximum temperature error for a 3 day forecast has decrease from 4.5 to 3.2°F from 2002 to 2017 and from 6.7 to 6 °F for the 7 day forecast. Similar results happen for the minimum temperatures. Having such accurate weather forecasts will permit prediction of power loads based on weather.

4. Data Description

To test the model of coupling a weather forecast to a machine learning algorithm and then predict electricity demand, I decided to use it over data that would be somewhat homogeneous, so I focused on a city, the city of New York, as its data was readily available.

I merged two different datasets, one set that contained electricity loads by the hour and another one with weather data, also by the hour, the sets were joined by date and hour to form a larger data set to be fed to the different algorithms.

The electricity load data was gathered from the NYISO, website. The data was taken on hourly intervals from January 2010 to December 2017. The weather measurements were taken from data obtained from the National Oceanic and Atmospheric Administration (NOAA) reflecting weather conditions at La Guardia Airport in New York City. The weather data used include hourly intervals for temperature, relative humidity, atmospheric pressure and day length variations.

As previously mentioned, much of the energy consumed in the United States goes to power air conditioning systems and their power requirements vary according to weather.

From a thermodynamics and energy consumption perspective, the major variables of air conditioning systems are all air properties and include barometric pressure, relative humidity and temperature.

Barometric pressure is also related to sunny or rainy/snowy weather. High barometric pressure is related to sunny weather, low barometric pressure to precipitation in the form of rain or snow. While there were other variables available within the NOAA dataset such as wet bulb temperature and dew point temperature, they were not taken into consideration as they were not independent from one another or, in statistical terms, they would exhibit multicollinearity. Other variables, such as wind speed and wind direction, were not taken into the model as they did not belong to the air conditioning variables.

Relative humidity increases or decreases the energy content on air, and affects the power consumption in air conditioning systems. A high relative humidity will make air conditioning systems run for longer periods of time to cool down the air.

Temperature directly affects the demand for energy. During winter time, people use electricity and gas heating needs and in the summer air conditioning is used to cool down dwellings.

The hour of a day also affects electricity loads as people use less energy when they are sleeping and more when they are awake and performing activities.

Literature published by Hiroto Shiraki, Mahmoud Salari and Yanming Sun among others [1-12] mention that electric loads vary according to day hour and temperature, and that data is highly seasonal. In order to corroborate such information, we plotted the 7 years data and generated a 3D plot.

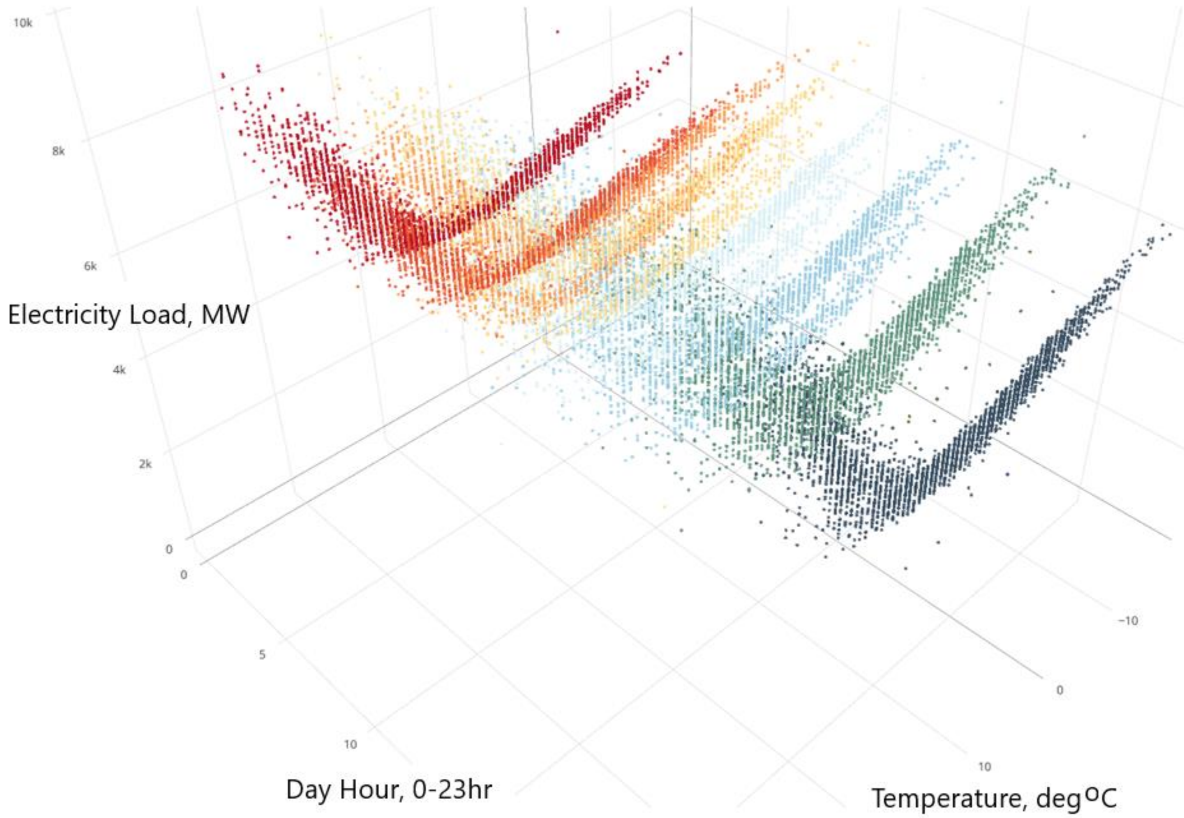

Figure 4. 3D Weather plot.

Figure 4 shows that the hourly curve shapes are similar when compared against temperature changes. We use color code to the difference of electricity consumption difference over daily hours and temperature. There is a minimum at approximately 15ºC/60ºF and the slopes are different on each side.

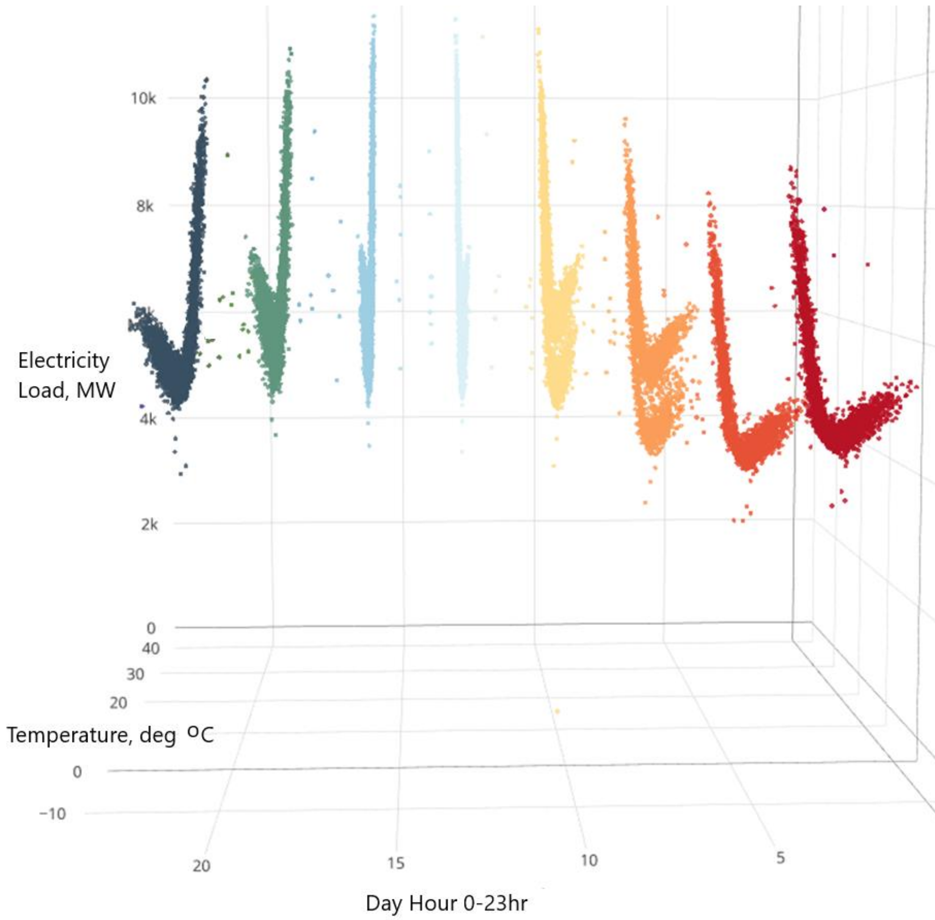

Figure 5. 3D Weather data, Day Hour axis

In Figure 5, the data is observed through its day hour axis, the demand peak happens at about 12pm, which minimum is at about 4K MW. During the night hours the electricity consumption diminishes to about 3K MW as a minimum, when temperature is around 15ºC/60ºF. The electricity usages show some clustering pattern on an hourly basis, all the clusters show a similar behavior related to temperature as they all show a minimum at around 15ºC/60ºF.

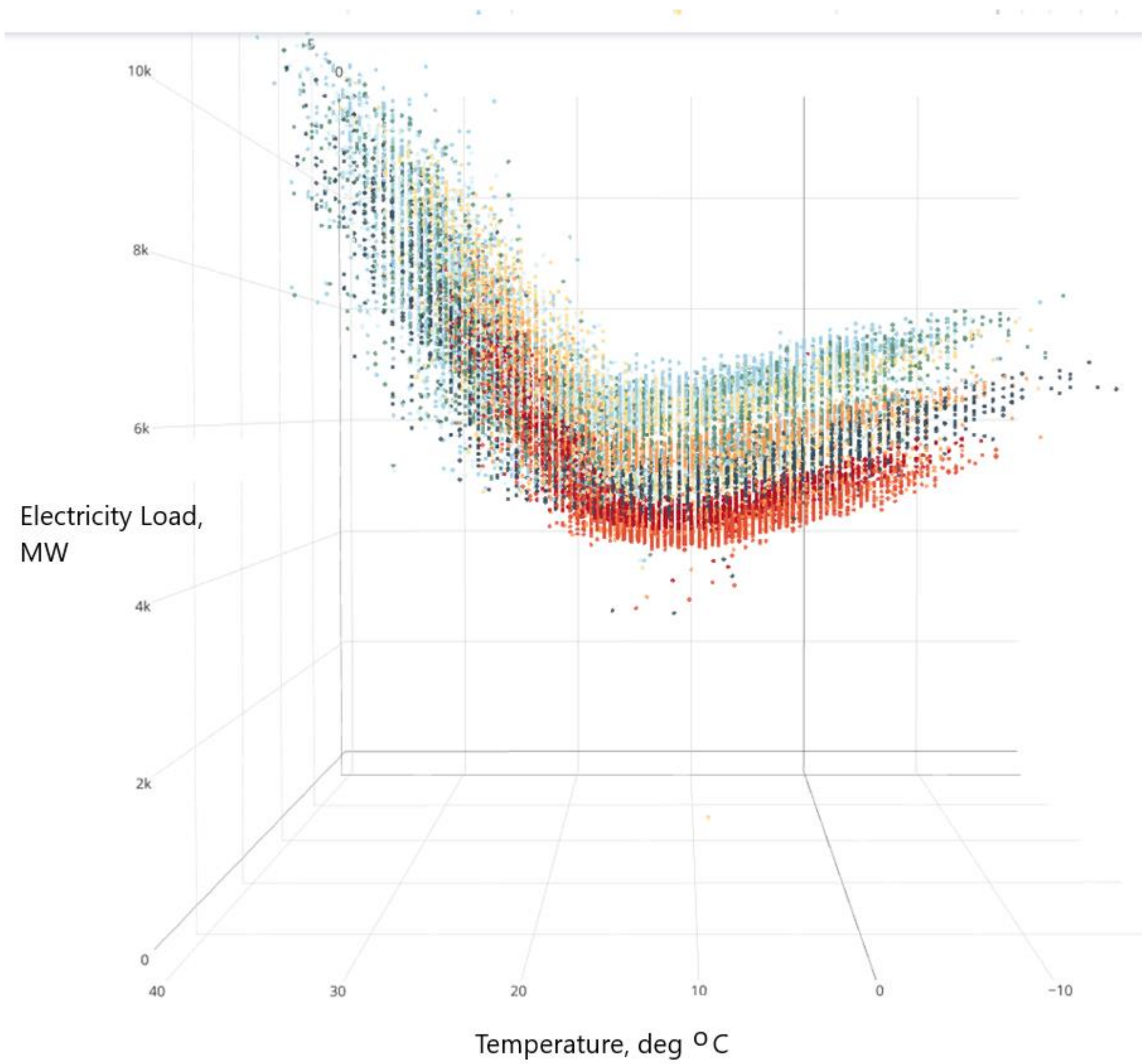

Figure 6. 3D Weather data. Temperature axis

In Figure 6, the data is observed from the temperature axis; the minimum temperature is approximately 15ºC / 60ºF. At temperatures lower than 15ºC/ 60ºF, heating starts and the load slowly increases. When the temperature rises above 20ºC/ 70ºF, air conditioning starts being used and the electricity consumption quickly increases. There is a large difference in the electricity slopes. . While the electricity demand during cold temperatures gradually increases, the power demand when the temperature goes up rapidly increases. It also can be seen the maximums/minimum are different according to the hour so distinguishing the electrical load according to the hour of the day is quite important.

5. Estimation Methods

Several different methods will be used to compare results with the goal of determining which would be the best one for this specific cause. First traditional forecasting models such as a multiple linear regression model and an ARIMA model are used, then machine learning algorithms including a Random Forest regressor, Gradient Booster and Neural Networks/Deep Learning method with 2 layers will also be used. Electricity companies generally use proprietary models to create their own forecasts.

It is important to notice something important about the different methods. Time series methods, such ARIMA, only require a trend, and then they will provide a mean value plus a range. They will not require any more inputs, so they will not change. Models such as multi-linear regression model and all the machine learning models require input variables and an output trend to mimic. If the data is varied enough, including extreme weather, the models will take all those points in their regression. In this study we took 7 years of continuous data.

5.1 Linear Regression Model[18]

For the multiple regression model we used temperature, daily hour, atmospheric pressure, relative humidity, and day length as independent variables to predict electricity load. The data is then regressed with the consumption data to provide the forecasted results. The graphs provided in the data description section clearly shows the relationship among the different variables and the energy demand is not linear, still multilinear regression is a method used by many as an initial model.

Multiple linear regression models the relationship between several variables and a response variable by fitting a linear equation. Every value of the independent variable x is associated with a value of the dependent variable y. The slope of the line is given by

$$\mu_y=\beta_0+\beta_1x_1+\beta_2x_2+\dots\beta_px_p \qquad (1)$$

The model for multiple linear regression, given n observations is:

$$y_i=\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\dots+\beta_px_{ip}+e_i \quad \text{for } i=1,2,\dots n \qquad (2)$$

5.2 ARIMA model

The ARIMA model is one of time series models to examine the time dependence of the predicted variable over time periods. In the model, AR is Auto Regressive term which represent how the lagged values affect the dependent variable itself, and MA is Moving Average term denoting the effects of lagged error terms. The important assumption in time series model is that the time series is covariate stationary. If the data has some time trend or integrated over time, we need to take a difference between the current values and the lagged values to stabilize the data. The ARIMA model only uses the load trend, so no weather data is used during the modeling of an ARIMA.

An ARIMA (p,d,q) (Auto Regressive Integrated Moving Average with orders p,d, q) model is a discrete time linear equations with noise, of the form [19]

$$\left(1-\sum_{k=1}^{p}\alpha_k L^k\right)(1-L)^d X_t=\left(1+\sum_{k=1}^{q}\beta_k L^k\right)\varepsilon_t. \tag{3}$$

Where p and q are the orders of the autoregressive moving average, d is the number of differentiations, $\alpha 1..k$ and $\beta 1..k$ are the parameters or coefficients (real numbers) for the autoregressive and moving average coefficients respectively, ε_t is an error term (usually white noise), L is a time lag operator or backward shift and Xt are the initial conditions.

5.3 Random Forest

Random Forest is a supervised learning algorithm. It builds multiple, random decision trees and merges them together to get a more accurate and stable prediction. An in-depth model explanation can be found at the PhD thesis by Louppe[20] and the paper by Klusowski.[21] In essence, Random Forest is a bagging model of trees where each tree is trained independently on a group of randomly sampled instances with randomly selected features.

The training of a Random Forest is as follows:

For *t* = 1, …, *T*,

1. Sample n_{try} instances from the dataset with replacement.
2. Train an unpruned decision or regression tree ft on the sampled instances with the following modification: at each node, choose the best split among m_{try} features randomly selected rather than among all features.

Both n_{try} and m_{try} are predefined constants.

The Random Forest comes from integrating all the trees together [21]

$$f_n^M(X,\theta_1,\dots\theta_M)\triangleq\frac{1}{M}\sum_{m=1}^{M}f_n(X;\theta_m,D_n) \tag{4}$$

When M (number of trees) is large, the law of large numbers justifies using

$$f_n(X) \triangleq E_\theta\left[f_n(X, \theta, D_n)\right] \quad (5)$$

where E_θ is the expected value with respect to the random parameter θ, conditionally on X (the desired values) and the dataset Dn. The sequence $\{\theta_m)_{1<m<M}$ consists of i.i.d. realizations of a random variable θ, which governs the probabilistic mechanisms that builds each tree.

5.4 Gradient Boosting Algorithm [22, 23]

Gradient Boosting algorithm is based on gradient descent plus boosting, where the predictors are not made independently but sequentially. This technique employs the logic in which the subsequent predictors learn from the mistakes of the previous predictors. Therefore, the observations have an unequal probability of appearing in subsequent models and the ones with the highest error appear most.

Gradient Boosting is considered an ensemble method and combines different predictors in a sequential manner with some shrinkage on them and also provides variable selection.

Gradient Boosting, as an ensemble method, may be described as follows:

$$\mathrm{y} = \mu + \sum_{m=1}^{M} v h_m(\mathrm{y}; \mathrm{X}) + e \quad (6)$$

where y is the vector of observed phenotypes, μ is a population mean, v is a shrinkage factor, h_m is a predictor model as described earlier, X is the matrix of corresponding genotypes and e is the vector of residuals. Each predictor is added in a sequential manner, and is applied consecutively to the residuals from the committee formed by the previous ones, weighted by $c_{i\neq 0}=v$.

5.5 Neural Networks/Deep Learning [24]

Neural networks/deep learning algorithms are designed to recognize patterns and are loosely based on neurons. They are formed by networks composed of several layers, the layers are made of nodes, which is the place where computation happens. Nodes combine input from the data with a set of coefficients or weights that either amplify or dampen that input. The network combines all the nodes and the final output is determined using activation functions.

A neural network is usually made of several units, also known as neurons, of the form:

$$h_j(\mathrm{x}) = \sigma(w_j + \sum_{i=1}^{n} w_{ij} x_i) \quad (7)$$

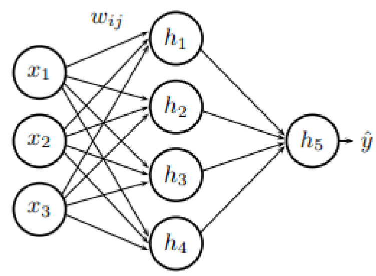

Figure 7. Neural Network[20]

Where σ is a non-linear activation function, such as the sigmoid function. These units are typically structured into successive layers, where the outputs of a layer are directed through weighted connections, or synapses, to the inputs of the net layer. In the figure a 3 layer neural network is shown. The first layer is the input layer, which transmits the input values x={x_1,..x_p) to the second layer. The second layer is made of activation units h_j, taking as inputs the weighted values of the input layers and producing non-linear transformations as outputs. The third layer is made of a single activation unit. It takes as inputs the weighted outputs of the second layer and procures the predicted value y.

6. Model results

Below are the results of the linear regression, ARIMA model and the three machine learning models: Random Forest, Gradient Boosting and Neural Networks/Deep Learning.

6.1 Linear Regression Model

The linear regression model was generated using R.

We regressed the hourly consumption with the weather data. We expected that as weather changes occur, energy consumption would react positively and fluctuate over time. In Figure 8 we can observe that the hourly data is quite noisy, and that there is a positive relationship between the weather variables and consumption as residuals are kept within a certain range.

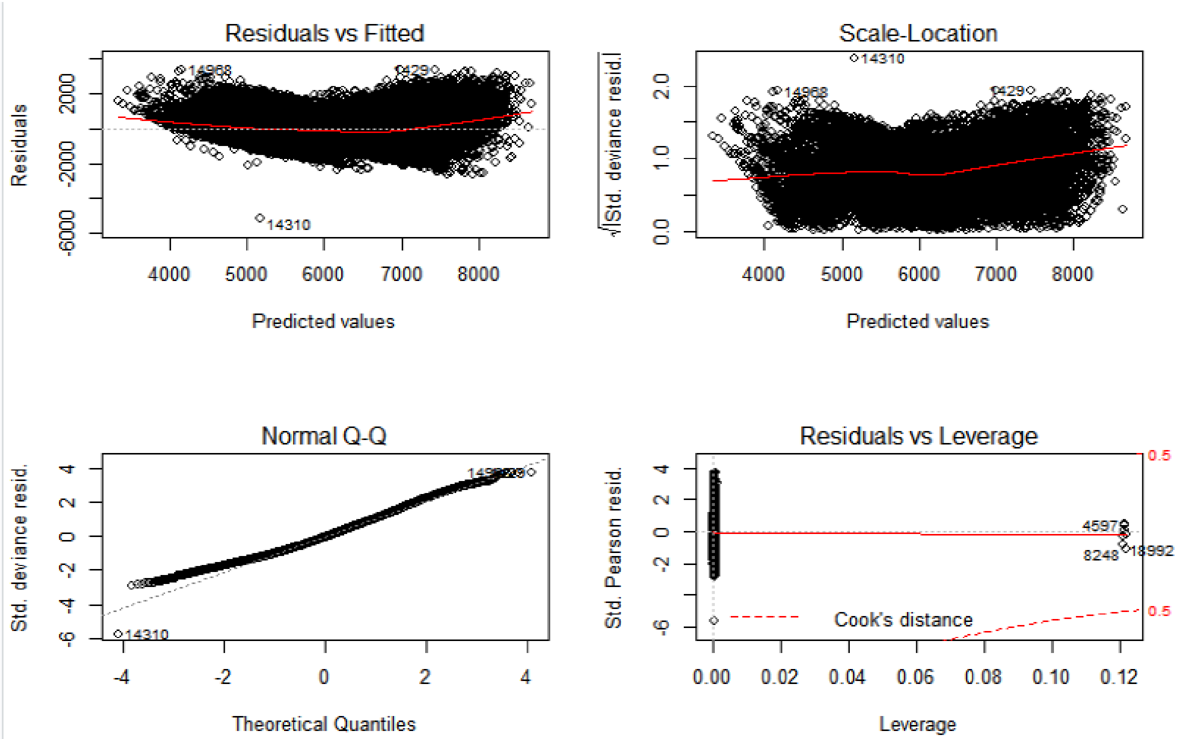

Figure 8. Linear Regression model graphs

The graph “Residuals vs Fitted” in Figure 8 shows a linear model, which residuals increase in magnitude as the electricity values become higher. The scale location plot shows if the residuals are spread equally along the predictor range, the line does move up and deviates from its horizontal line, which means it is not homoscedastic. The QQ plot shows that there’s a non-normal distribution with the end quantiles as the plot diverges from the line. The Cook’s Distance shows that only a few points could be influential outliers.

In conclusion, a linear regression model is not able to simulate with precision the electricity demand changes as the underlying system, air thermodynamics, is highly non-linear so a different method is required. This same conclusion could be drawn by looking at the original data.

6.2 ARIMA model

The ARIMA model and forecast were generated using R.

We performed an auto-ARIMA using R to the electricity load data to obtain the ARIMA coefficients. Then we compute an Autocorrelation Function, (ACF), to review the residuals of the model found.

The ARIMA model obtained by using an AUTOARIMA function in R and the model results are reported in Figure 9. The model ARIMA (5,1,4) means the data had some time trend and are not stationary, as expected.

	AR1	AR2	AR3	AR4	AR5	MA1	MA2	MA3	MA4
Value	0.4068	-0.4887	-0.6038	0.3587	-0.8286	-0.1707	-0.0196	-0.5902	-0.3766
standard error	0.0046	0.0053	0.0047	0.0049	0.0042	0.0071	0.0066	0.0062	0.0065

Figure 9. AUTOARIMA results of electricity load data

After performing the ARIMA (5,1,4) simulation, we compute an ACF of the residuals which is reported in Figure 10.

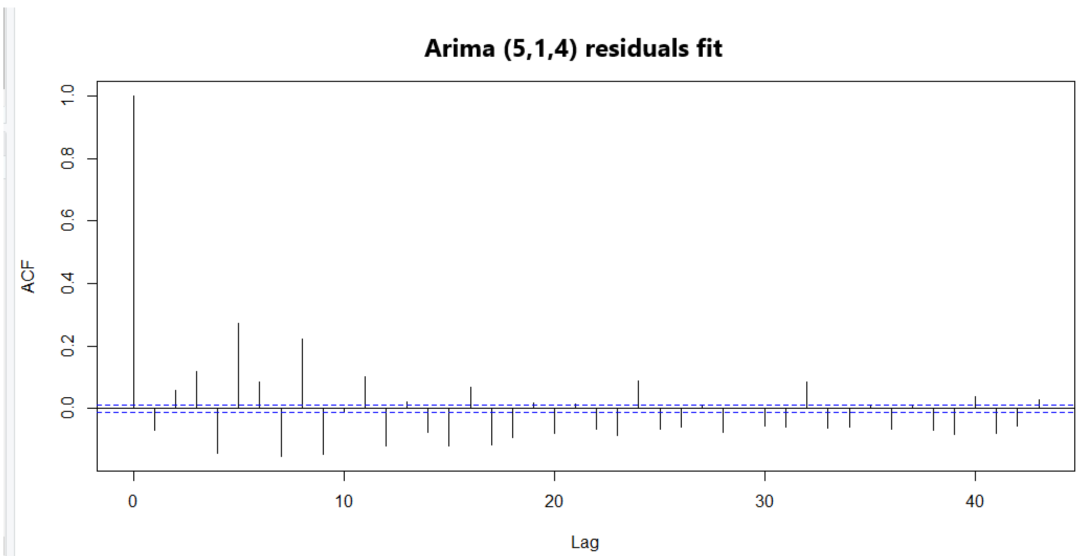

Figure 10. AR graph of ARMA residuals, no seasonality correction.

Figure 10 shows that there is some correlation in the data, so the -Ljung Box test using the residues was performed to corroborate the results of the graph. The test gave a p-value less than 2.2E-16. As the p-value is lower than the significant level of 0.05, therefore the test failed and the ARIMA model is not a good way to forecast electricity in this case, as the data is correlated with each other.

For illustration purposes, a forecast of an ARIMA (5,1,4) model was obtained and shown on figure 11

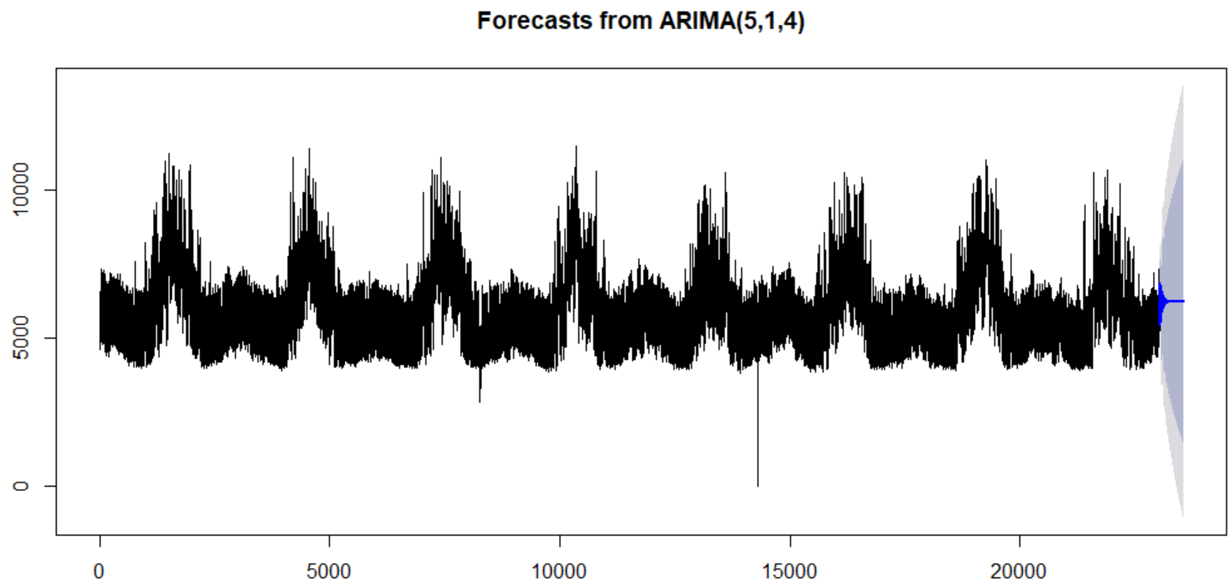

Figure 11. Forecasts from ARIMA (5,1,4)

ARIMA forecasts provide a central value and then a range where the next points may end. The inner range is at 80% confidence at the outer range is at 95% confidence. As it can be seen, the forecast of an ARIMA gives a central value surrounded by a confidence interval that expands with time. It does not vary according to inputs, the obtained forecast is not a high precision forecast.

6.3 Machine Learning Algorithms General Procedure and Numerical Results

The machine learning algorithms were run using weather data on an hourly basis as model input and electricity demand per hour as output.

Good machine learning algorithms, such as the ones used in this paper, function like high-power regression models and can handle non-linear variables if enough data is provided. Regression data needs to encompass extreme situations that are wanted to be predicted, for instance, very hot and humid summers and cold winters.

The algorithms were programmed using Python and machine learning libraries such as Sklearn, Tensorflow, and Keras. In this project I used 1000 trees for both the Random Forest and the Gradient Boosting algorithms and 40 epochs (loops), 2 layers and RELU activation functions for the Neural Networks algorithm.

Initially, all the data was standardized, that is, each datapoint was divided by its variable mean. The objective is to make the magnitudes of the different variables comparable to each other, essentially homogenizing the magnitudes of all variables being fed to the algorithm. This is done to prevent variables with bigger magnitudes skewing the training process.

The data is then randomly divided in two datasets, a training dataset and a testing dataset. In machine learning algorithms, data needs to be partitioned in two and used as training and testing data. The training data is used by the algorithm to fit the model parameters to predict the trend. The testing data is used by the algorithm to compare the output from the algorithm against data it has never seen before. If too much data is used in the training section, we would get what is known as overfitting, which means the algorithm will fit all of the points with the training data, but it will not be able to perform well when new data is given.

The mean square error (MSE) and mean absolute percentage error (MAPE) were calculated for all three algorithms in order to compare their results. Testing and training scores were calculated for both Random Forest and Gradient Boosting algorithms.

Figure 12 shows the results of running the different machine learning algorithms using 10% to 80% of the available data for training. Random Forest and Gradient Boosting were found to perform best at approximately 70% of the data, as overfitting started at 80%. Neural Networks performed at its best at about 40% of the data. The MSE and MAPE score reflected the difference between the calculated values and showed that the error in the predictions was diminishing.

Algorithm	Random Forest		Gradient Boosting		Neural Networks		Random Forest		Gradient Boosting	
training split	MSE	MAPE	MSE	MAPE	MSE	MAPE	train	test	train	test
0.8	446	5.32	441	5.3	486	5.59	0.83	0.81	0.85	0.84
0.7	439	5.33	439	5.38	483	6.13	0.83	0.83	0.85	0.85
0.6	446	5.36	445	5.45	496	5.72	0.83	0.83	0.85	0.85
0.5	445	5.36	447	5.48	484	5.89	0.82	0.83	0.85	0.86
0.4	455	5.47	449	5.51	481	5.76	0.82	0.83	0.85	0.85
0.3	462	5.65	451	5.55	483	5.92	0.81	0.83	0.84	0.85
0.2	478	5.81	462	5.7	507	6.29	0.81	0.83	0.84	0.85
0.1	489	5.98	489	6.06	510	6.08	0.81	0.83	0.84	0.85

Figure 12. Machine learning runs at different training splits.

The three algorithms, which are among the best in machine learning, gave very similar results, Gradient Boosting and Random Forest are found to be almost equivalent and show slightly better results than Neural Networks.

The resulting graphs of the three different algorithms were also found to be almost identical. The graphs are going to be shown according to their numerical accuracy using train and test data plus MSE/MAPE, first Gradient Boosting, then Random Forest and last Neural Networks.

The small sensitivity seen by the MSE and MAPE numbers when the data is split from 10% to 80% to train the algorithm shows that the amount of data provided even at 10% split is enough to train the different algorithms. If the provided data would not be enough, we would see high sensitivity in the indicators. The total dataset has about 25,000 points.

6.4 Ranking of Variables in Gradient Boosting and Random Forest

Using the Gradient Boosting algorithm, the major variables were ranked; temperature and day length were major contributors, followed by relative humidity and finally the hour of the day, as seen in Figure 13.

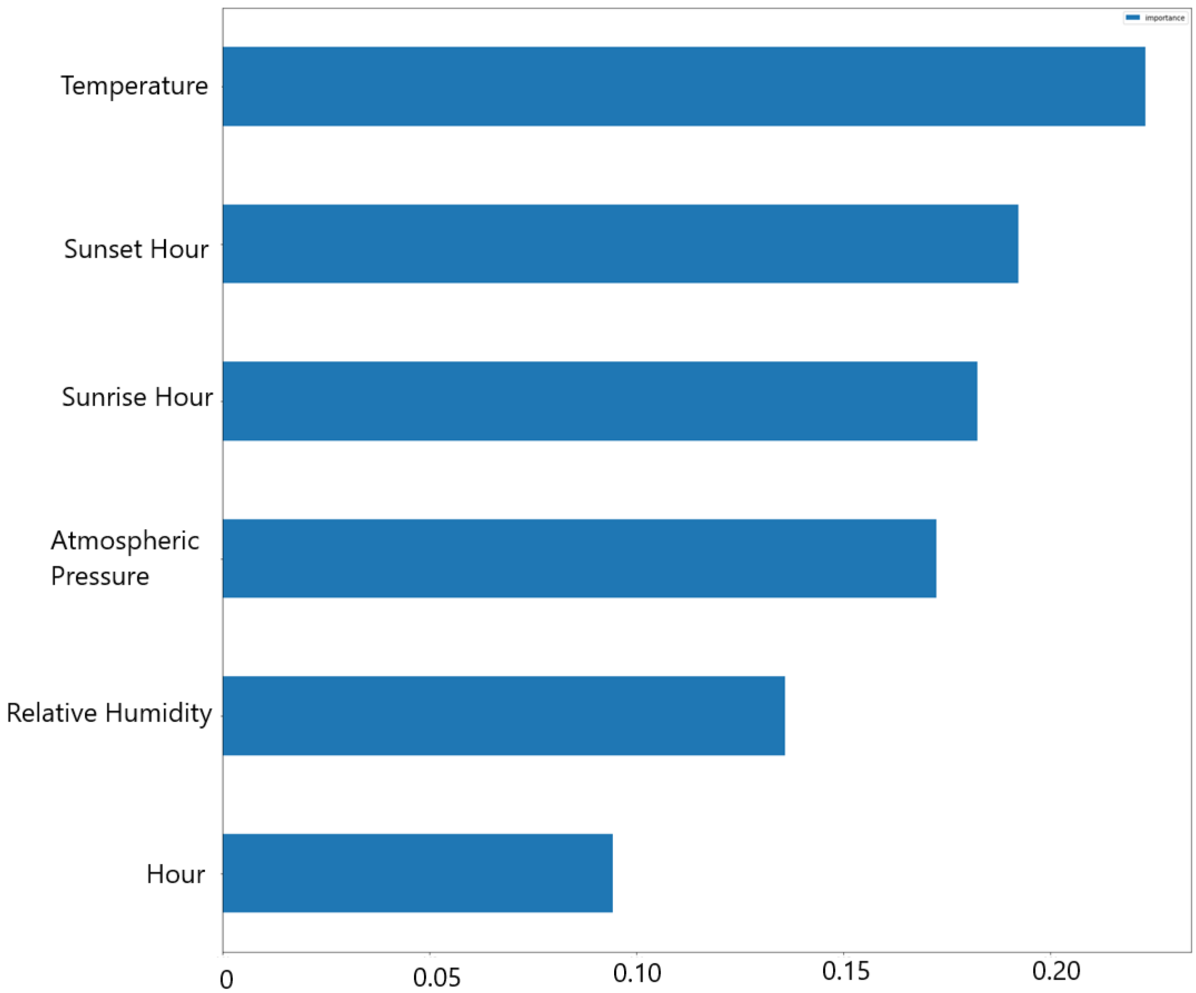

Figure 13. Variable importance, Gradient Boosting Algorithm

Interesting enough, the ranking of variables turned out differently for Random Forest as it is shown on Figure 14.

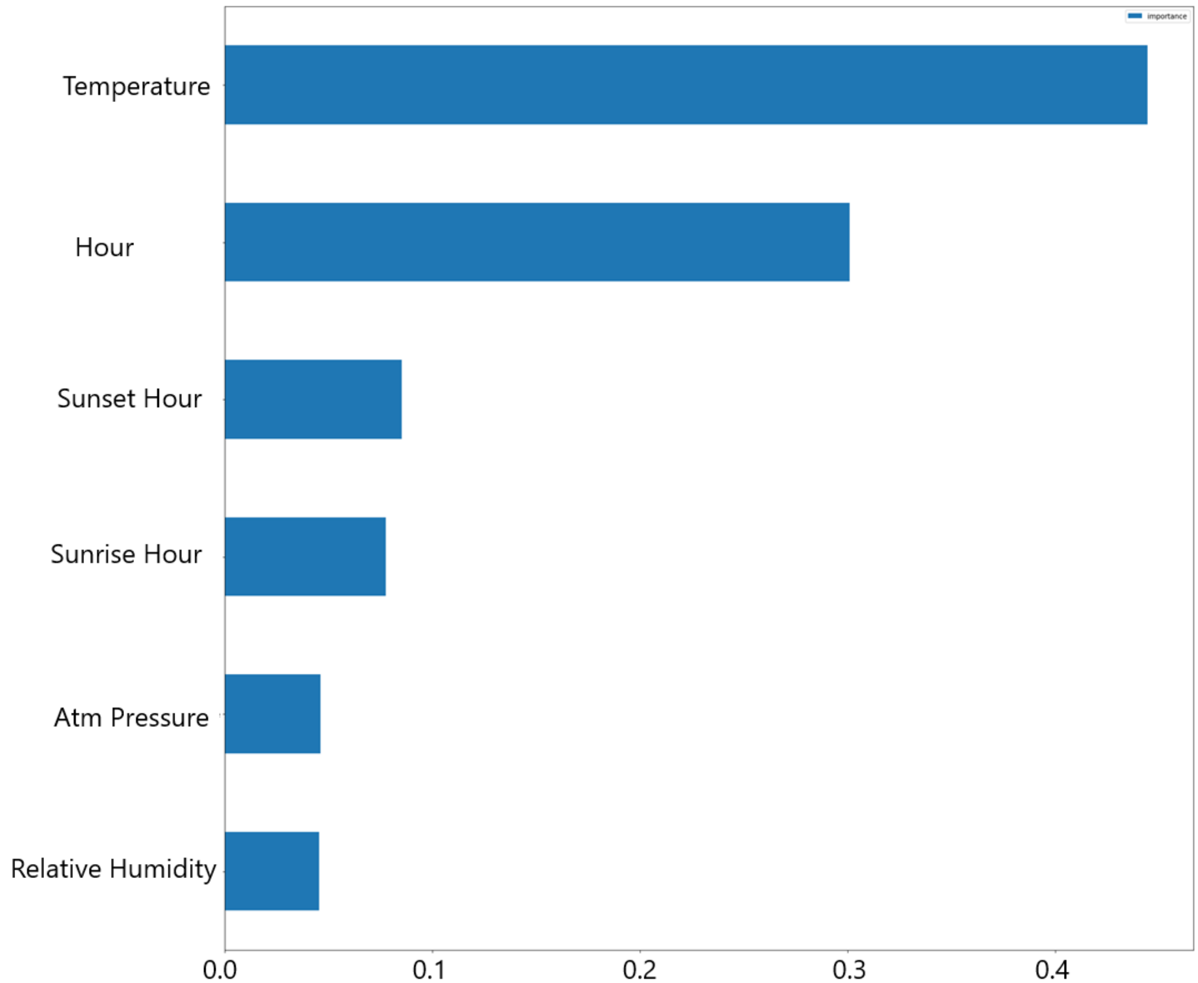

Figure 14. Random Forest Importance Chart

Using the Random Forest algorithm, see Figure 14, temperature and hour are the major contributors by far, followed by day length and barometric pressure and air humidity at the end. Gradient Boosting sometimes performs better than Random Forest because the algorithm retrains on the data where it got the worst performance.

6.5 Graphs Comparison among Gradient Boosting, Random Forest and Neural Networks

Several graphs were created for all three machine learning algorithms. All three compare the predicted values against the original test values and all of them look quite similar.

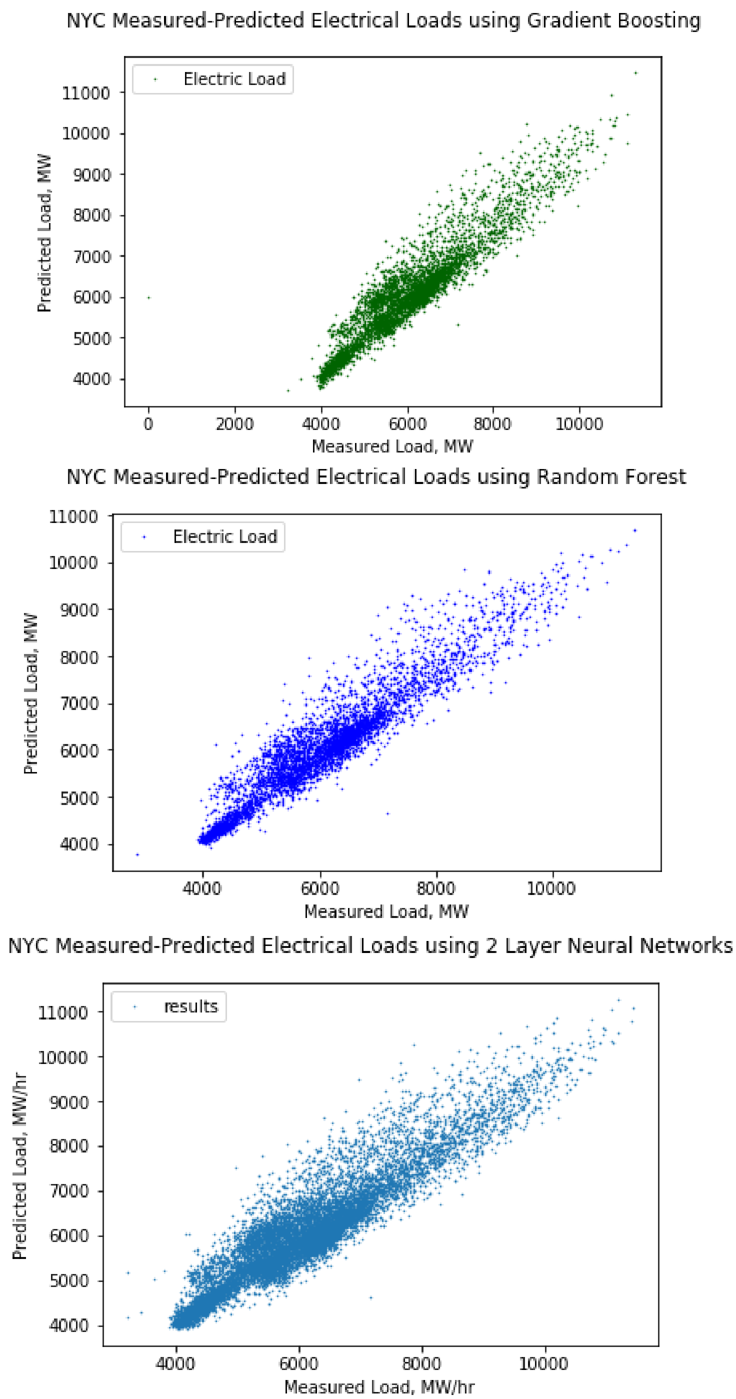

Figure 15. NYC Measured vs. Predicted Electric Loads using Machine Learning

Figure 15 shows an XY plot of the relationship between the measured load and the electricity load predicted by the model. It shows the models are effective within the range. It also shows that between 4000 and 8000 MW the forecast is more accurate, while its predictions are less accurate for values higher than 8000 MW as the difference between the predicted and the measured value increases.

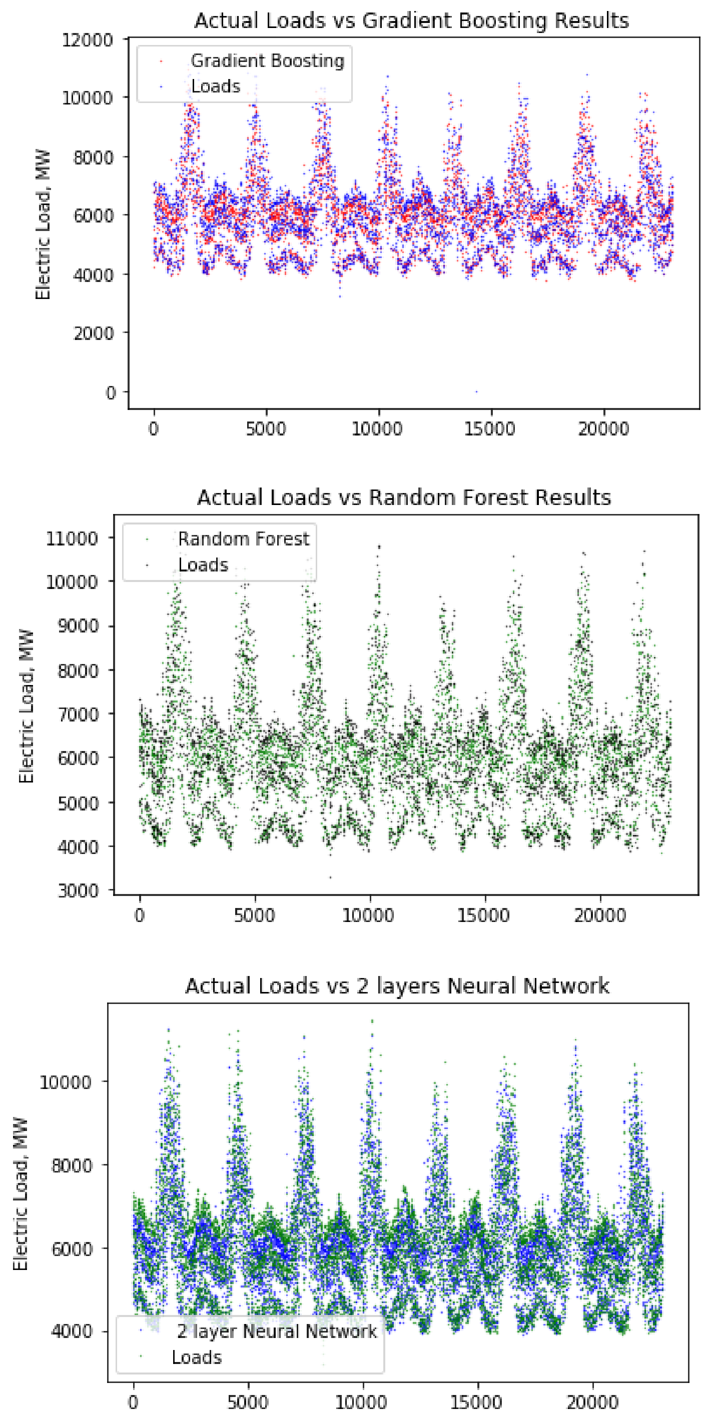

Figure 16. Actual Loads vs. Machine Learning Results

Figure 16 illustrates that the results of all algorithms resemble the behavior of the actual load over the 7 year period, going up and down, behaving just like the original data does. Neither time series nor multiple linear regression can predict the data with such precision.

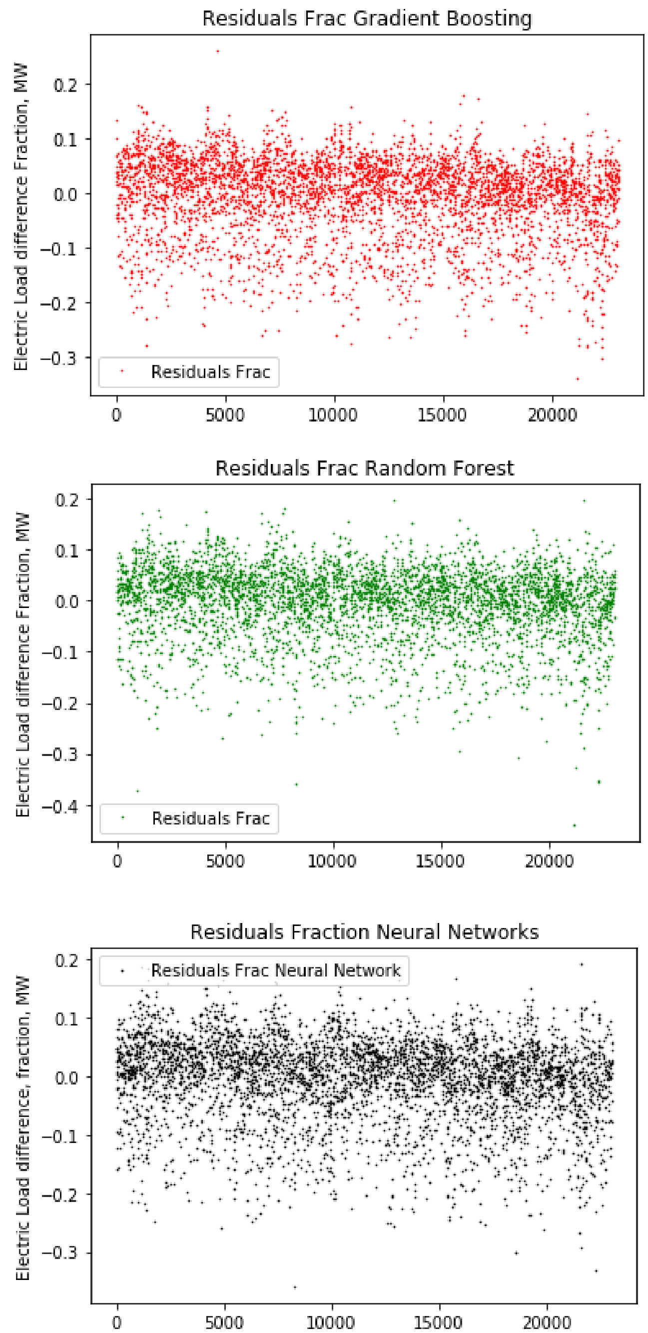

Figure 17. Residuals Fraction of Machine Learning results

Figure 17 shows how large the fraction difference between the calculated results and the loads are, the tendency to negative values shows that calculated loads are typically higher than the actual load when there is a difference, which for a power company is very relevant data..

6.6 Cumulative Differences among Machine Learning Algorithms

I reviewed all the data pairs and found that the difference between the calculated and the original values was quite small most times, the values found are shown in Figure 18.

Cumulative Difference	Gradient Boosting	Random Forest	Neural Networks
less than 5%	58.17	58.92	59.83
less than 10%	85.65	85.72	85.63
less than 20%	98.53	98.23	98.52
less than 30%	98.53	98.23	98.52

Figure 18. Percentage of population, maximum difference among machine learning algorithms

Figure 18 shows that all three methods perform well and deliver very similar results. Almost 60% of the results vary at most 5% of the original data. Around 99% of the results vary at most 20%.

7. Conclusions

Energy forecasts based on multilinear regressions are not accurate as the underlying phenomena, air enthalpy variation, is not linear. Air enthalpy changes in a non-linear manner when temperature, barometric pressure and relative humidity are taken into consideration. Multilinear regressions are typically used in data science as a first approach to model outcomes, but it is important to verify if the underlying phenomena is linear or not.

Typical forecasting methods such as ARIMA are not quite applicable as temperature, a main variable in electricity demand, is not a purely random phenomena and follows a trend, this violates the randomness requirement for the use of time series. Differentiation was tried to improve results but did not work. In addition, forecasts based on time series gives a median value plus a range, these forecasts are not meant to give accurate numbers as they only receive a trend as an input. In addition, they don’t adapt to new conditions.

Air conditioning is responsible for much of the variation in power demand in the summer for in the United States. For this reason, variables that influence the demand for air conditioning, such as temperature, barometric pressure and relative humidity should be incorporated into power forecasting models.

While I cannot compare my results against the methods used by the industry, as they are proprietary, the integration of state of the art weather forecasting given by the Weather

Prediction Center in the USA plus machine learning algorithms such as Random Forest, Gradient Boosting or Deep Learning/Neural Networks create a viable tool to produce accurate hour by the hour forecasts. It will allow power companies to propose FERC the use of a multiday market forecast to improve their economic commitment decisions, just like the proposal given in 2018 by MISO [32], expand their energy hedging contracts from 1 to maybe 3 days and supply most of its energy based on these hedging contracts. Real time traders will still be required, but the major sourcing of energy will be bid in day-ahead market. Generators with a multi-day offer means less risk and therefore lower prices as unit with longer lead times or high startup costs or other factors may be taken into consideration in the mix. All these factors, as shown in the MISO study, will improve the profits of electricity companies.

References

[1] Hiroto Shiraki, Shogo Nakamura, Shuichi Ashina, Keita Honjo, Estimating the hourly electricity profile of Japanese household – coupling of engineering and statistical methods Energy 114 (2016) 478-491

[2] Mahmoud Salari, Roxana J. Javid, Residential energy demand in the United States: Analysis using static and dynamic approaches, Energy Policy 98 (2016) 637-649

[3] Nina Boogen, Souvik Datta, Massimo Filippini , Dynamic models of residential electricity demand: Evidence from Switzerland, Energy Strategy Reviews 18(2017) 85-92,

[4] Yanming Sun, Yihua Yu, Revisiting the residential electricity demand in the United States: A dynamic partial adjustment modelling approach The Social Science Journal 54 (2017) 295–304

[5] Julian Moral-Carcedo, Julian Perez-García Integrating long-term economic scenarios into peak load forecasting: An application to Spain, Energy 140 (2017) 682-695

[6] Siyan Wang, Xun Sun, Upmanu Lall, A hierarchical Bayesian regression model for predicting summer residential electricity demand across the U.S.A. , Energy 140 (2017) 601e611

[7] C. Sigauke, D. Chikobvu. Prediction of daily peak electricity demand in South Africa using volatility forecasting models. Energy Economics 33 (2011) 882–888

[8] A.S. Hurn, A.E. Clements, Forecasting quantiles of day-ahead electricity load, Energy Economics 67 (2017) 60–71

[9] Mashud Rana, Irena Koprinska Forecasting electricity load with advanced wavelet neural networks, Neurocomputing 182(2016)118–132

[10] Song Deng, Changan Yuan, Lechan Yang , Liping Zhang, Distributed electricity load forecasting model mining based on hybrid gene expression programming and cloud computing Pattern Recognition Letters 109 (2018) 72–80

[11] Mahmoud Shepero, Dennis van der Meer, Joakim Munkhammar, Joakim Widén Residential probabilistic load forecasting: A method using Gaussian process designed for electric load data, Applied Energy 218 (2018) 159–172

[12] Jinliang Zhang a, Yi-Ming Wei b, Dezhi Li c, Zhongfu Tan a, Jianhua Zhou Short term electricity load forecasting using a hybrid model, Energy 158 (2018) 774-781

[13] US Energy Information Administration website https://www.eia.gov/tools/faqs/faq.php?id=96&t=3

[14] Map of deregulated energy states and markets, 2018 https://www.electricchoice.com/map-deregulated-energy-markets/

[15] 89 FERC 61,285, 18 CFR part 35 Regional Transmission Organizations, Federal Energy Regulatory Commission

[16] Price Formation in Organized Wholesale Electricity Markets Docket No. AD14-14-000 Operator Initiated Commitments in RTO and ISO Markets, FERC, December 2014

[17] Using Market Optimization Software to Develop a MISO Multi Day Market Forecast, Chuck Hansen et al, MISO FERC Technical Conference, AD10-12-009, June 26th 2018

[18] Yale University Department of Statistics www.stat.yale.edu/courses/1997-98/101/linmult.htm

[19] Dipartimento di Matematica, Universita di Pisa, Franco Fladoli, Arima models http://users.dma.unipi.it/~flandoli/AUTCap4.pdf

[20] Gilles Louppe, Understanding Random Forests from Theory to Practice, PhD Thesis, University of Liege, July 2014

[21] Jason M. Klusowski ,Complete Analysis of a Random Forest Model, September 24.2018, Rutgers University, New Brunswick, Jason.klusowski@rutgers.edu

[22] Gradient Boosting from scratch, accessed August 17, 2018 https://medium.com//mlreview/gradine-boosting-from-scratch-1e317ae4587d

[23] O. Gonzalez-Recio, J.A. Jimenez-Montero, The Gradient Boosting algorithm and random boosting for genome-assisted evaluation in large data sets, J. Dairy Sci, 96:614-624

[24] CMU 15-781 Lecture 17 Deep Learning/ Neural Networks, Gianni Di Caro http://www.cs.cmu.edu/~arielpro/15381f16/c_slides/781f16-17.pdf

[25] 2017 State of the Market Report for the New York ISO Markets

[26] Assessing Forecast Accuracy

https://www.e- education.psu.edu/meteo3/node/2285

[27] Weather Prediction Center https://www.wpc.ncep.noaa.gov/html/hpcverif.shtml#medmin

[28] Weather Research and forecasting model

https://www.mmm.ucar.edu/weather-research-and-forecasting-model

[29] National Weather Service, https://www.nco.ncep.noaa.gov/

[30] 2017 State of the Market Report for the New York ISO Markets

[31] Electricity Power Markets: National Overview

https://www.ferc.gov/market-oversight/mkt-electric/overview.asp