Analysis of load forecast of Zone J in NYISO

Abhishek Jain, Rajat Pungaliya, Arunika Chandra

Keywords: Forecast, Zonal Load, Day-Ahead LBMP, TBATS, MSTS, NYISO

1. Abstract:

As a part of this study, we forecasted daily trends in load that are as part of the day-ahead demand in the New York ISO region for a year from 2020 – 2021. The analysis and forecasts were done using R Studio where we start with diagnostic tests and analyze the seasonal components of the electricity demand data. From year to year, the load forecast data does not show any significant trend, neither upward nor downward. The data is stable around the mean of 6021 MW. The load forecast data for New York City has a bigger peak in the summer (June to August) and a smaller peak in the winter (December to February). The weekly trend clearly shows higher peaks during weekdays and lower peaks for weekends. The hourly trend shows a recognizable pattern with two daily peaks, smaller one around noon and a bigger one in the evening. We explore the following models to forecast - I) exponential smoothing, ii) SARIMA, and iii) TBATS. The SARIMA and SARIMA model performed relatively better than the simple exponential smoothing model. Furthermore, our analysis suggests that TBATS performed the best as the residues left behind were relatively scattered around zero with lowest magnitudes. The forecast was able to capture all levels of seasonality. While TBATS has the potential to provide a fair and just assessment into load forecasts a lot more parameters would help the models do more accurate analyses. These parameters include - Consumption behavior, energy efficiency practices, change of energy policies, change of weather conditions, influence of natural disasters and various political actions.

2. Introduction:

2.1 NYISO:

New York ISO (NYISO) is one of the 7 ISOs/RTOs in the United States and is responsible for operating the power grid in the state of New York. It is responsible for managing wholesale energy and capacity markets, while ensuring that the electric supply is reliable and stable through its ancillary services market; all this at least costs. NYISO has totally 11 zones as shown in Figure 1.

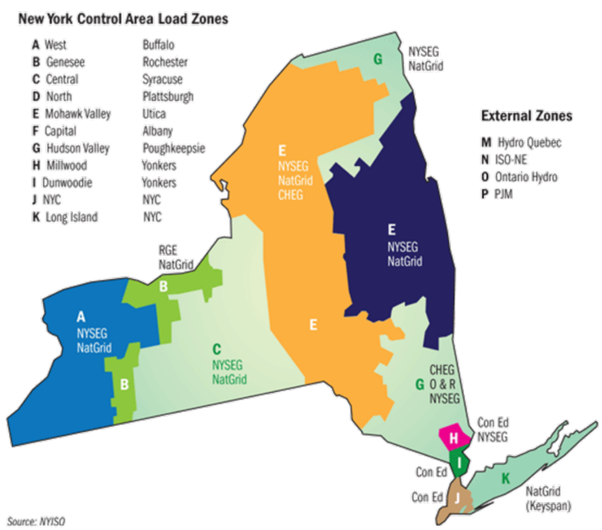

Figure 1: NYISO load zones

These zones are grouped according to different pricing zones. Zones A, B, C, D and E are in the West pricing zone. Zone F is in the East pricing zone. Zones G, H and I are in the Lower Hudson Valley pricing zone. Two-thirds of New York's electricity is used in Long Island (Zone K), NYC (Zone J), and the Lower Hudson Valley (Zones GHI). Because of the high demand for electricity in these areas, it is required that a certain percentage of electricity generation be physically located in those areas.

2.2 Current trends in NYISO:

NYISO has been pioneering in improving system efficiency by leading a shift towards clean energy resources since 1999[1.] In today's time as well, NYISO has increasingly set an example in clean energy penetration with new technologies such as storage and solar beginning to enter wholesale markets. These trends are also being catalyzed by the recently set renewable energy goal set in 2018 that 50% of the electricity consumed by New York would come from renewable energy sources by 2030. Interestingly, this goal was made even more ambitious the very next year in 2019 with the goal now set at 70% renewable energy by 2030 and 100% "clean energy sources" by 2040[2]. New policies have also been considered to accelerate energy storage penetration. Several policies have also affected reliability standards in all three competitive wholesale markets. These trends are not only affecting the way power is generated and managed but also how electricity is consumed.

One critical trend is the impact of a shift in energy mix on reliability, or in other words how supply and demand are maintained. For instance, in today's time, Distributed energy resources (DERs) such as Behind-the-meter (BTM) solar and energy storage are not allowed to participate in wholesale markets. Additionally, New York, as aforementioned, has witnessed a surge in large scale renewable energy resources. These shifting landscapes not only present opportunities but also present risks to managing the overall electric grid system.

BTM DERs are displacing peaking plants and traditional generators to a great extent. This eventually means that the load addressed by the grid has been lowering. As shown in Figure 2 below, NYISO demand trends clearly indicate a decline that is consistent with national trends. This trend can also be attributed to NYISO's increased efforts in building energy efficiency. In fact, without Energy efficiency and DERs: the state's Electricity demand to grow at 3.03% YoY between 2018 and 2028 whereas with Energy efficiency and DERs, demand shall decline at 0.13%[2].

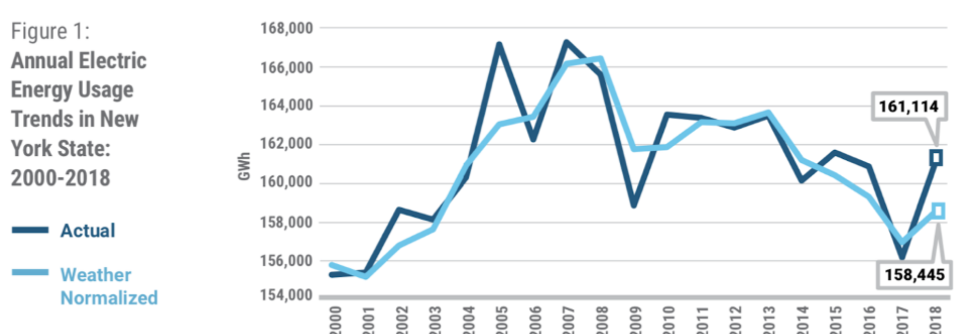

Figure 2: Annual electric energy usage trends

Interestingly, we see that the change in demand is not consistently same across all zones. Without considering the success of energy efficiency programs and DERs, Figure 3 shows the annual electric energy usage between 2017 and 2018.

REGION	2017 GWh	2018 GWh	% CHANGE
New York State (NYCA)	156,370	**161,114**	3.03% ↑
Upstate (zones A-E)	52,938	**55,211**	4.29% ↑
Downstate (zones F-I)	30,351	**31,218**	2.86% ↑
New York City (zone J)	52,266	**53,360**	2.09% ↑
Long Island (zone K)	20,815	**21,326**	2.46% ↑

Figure 3: Annual electric consumption by zone

We see that the highest % change in demand is in the upstate zones of A – E. However, NYC (Zone J) alone has a significantly high YoY. Additionally, we also know that the over 51% of the load comes from the NYC and Long island control zones. Hence, to make most use of our study, it revolves around Zone J i.e, NYC only.

2.2 Significance of forecasting load:

As aforementioned, addition of BTM sources and DERs has transformed the supply and demand scenario of NYISO's grid. However, it has also added significant complexity. For instance, shift from traditional generators and Front-of-the-meter peaking plants also brings the issue of reliability into question. Elimination or reduction of load from the grid will also need to account for the times when these resources are unable to operate either due to the resources' operational and physical characteristics or due to other unforeseen circumstances. This brings NYISO's resource adequacy and capacity market transactions into question. Another aspect of reliability is the concern of peak demand. Peak demand is the maximum amount of energy used in a one-hour period during the year. And though it represents a small fraction of the overall electricity consumed in a year, it still is an important metric in grid operation. Such concerns are governed by the projected peak demand without considering DERs and other load reduction strategies and go a long way in affecting reliability standards, installed capacity requirements, and operational reserves. These would also be critical in developing a road map for achieving the clean energy goals as set forth by NYISO.

3. Data

3.1 Overview

All data was obtained from New York Independent Operator. The flexible dashboard on their website offers a flexible and wide array of options to download the data in different formats. For load, the Load Forecast (MW) data was downloaded. Forecasted data was used and not the real-time data because the dashboard kept throwing blank files. Hourly data for five years was used for Load Forecast for New York City zone from 11 zones of the NYISO. This was from January 1, 2015 to December 31, 2019. These gave us 43,824 data points. For the prices, Day-Ahead Zonal LBMP ($/MWh) was downloaded. LBMP is Location Based Marginal Price, and to correspond with the load data, this data was used with the same granularity of hourly prices for five years from January 1, 2015 to December 31, 2019. These prices were for the New York City zone as well. These gave us 43,608 data points. Some data for prices were missing so the data for equivalent load points had to be omitted as well.

3.2 Diagnostics

All of this data had to be cleaned and diagnosed for zeros and NAs before importing to RStudio where all the analyses were done. Five separate data files were generated for each of the two variables and had to be bonded into single variables. These were then transformed from data structures to time series variables with a frequency of 24 hours. Summary of New York City Load Forecast (MW) data between 2015 and 2019 is as follows:

Min	1st Quartile	Median	Mean	3rd Quartile	Max
3031	5103	5936	6021	6621	11028

3.3 Seasonality

The Load Forecast (MW) data exhibits multiple levels of seasonality (Figure 4). Each level needs to be understood to better judge the variations and hence use appropriate tools for forecasting.

1. *Annual Seasonality*- From year to year, the load forecast data does not show any significant trend, neither upward nor downward. The data is stable around the mean of 6021 MW

2. *Monthly Seasonality*- The load forecast data for New York City has a bigger peak in the summer (June to August) and a smaller peak in the winter (December to February)

3. *Weekly Seasonality*- The weekly trend clearly shows higher peaks during weekdays and lower peaks for weekends.

4. *Daily Seasonality*- The hourly trend shows a recognizable pattern with two daily peaks, smaller one around noon and a bigger one in the evening.

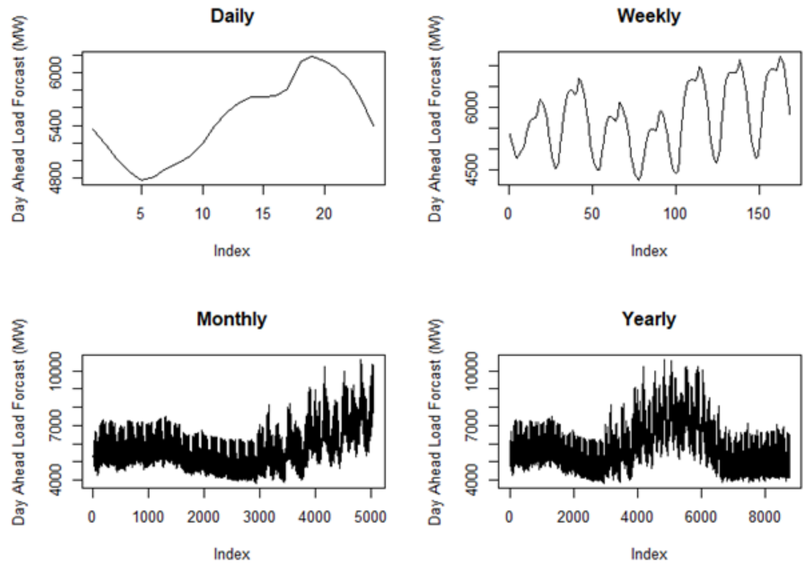

Figure 4: Various levels of seasonality in the Load Forecast (MW) data

4.Methodology

4.1 Data Pre-Processing

The analysis was done using RStudio. The following packages were used to perform numerous sets of analysis: “forecast”, “tseries”, “Kendall”, “dplyr”, “Hmisc”, “graphics”, “XLConnect”, “xlsx” and “readxl”. The allocated memory in JAVA was increased to 2 Gigabytes as our datasets were large and required susbstanial space. The data was imported into a data frame for basic analysis. For data cleaning, we examined if there was any missing data or outliers which could represent errors for our analysis. However, the data set was clean and processed.

4.2 Diagnostic Tests

For time-series analysis, the load data from the data frame was taken and exported into two functions: ts() and msts(). Given the hourly resolution of the data, there were multiple seasonal components in the data. Accurate forecasting techniques require us to identify and decompose them accurately. A standard ts() function was given frequency=24 to identify the daily seasonality. Decomposition function revealed that the ts() function could only identify the daily seasonality. It captured the annual and weekly seasonality in the trend and random components as shown in Figure 5.

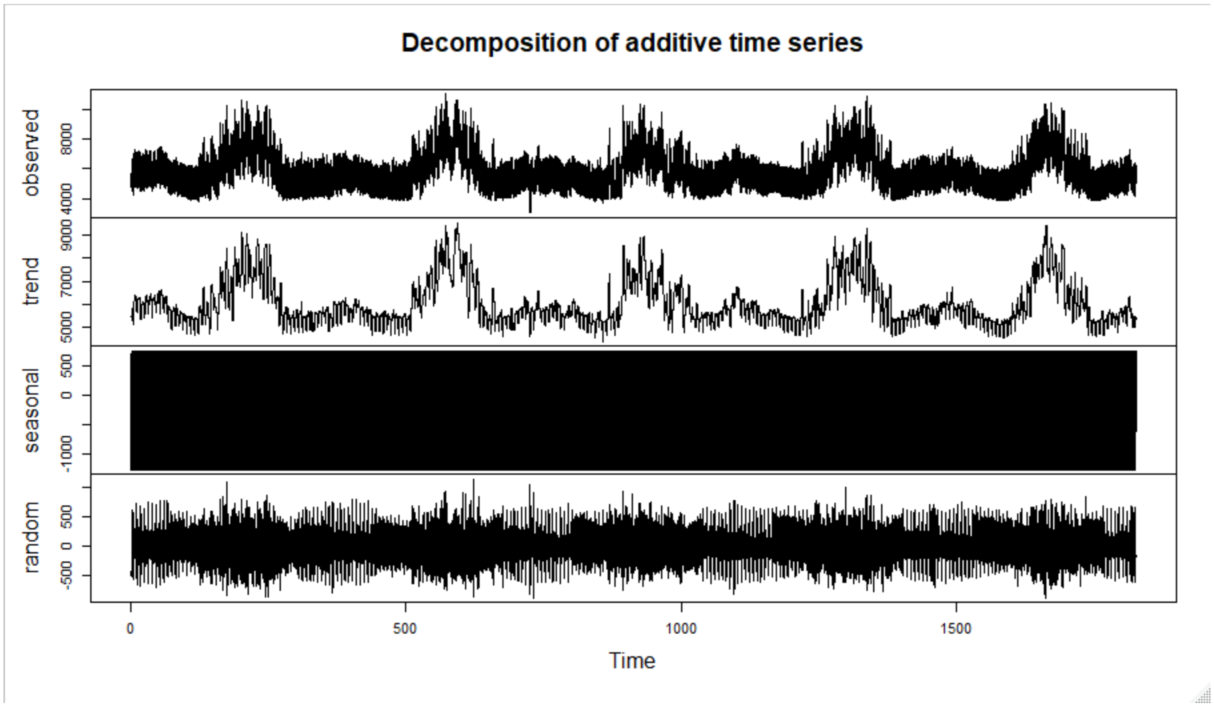

Figure 5: Decomposition of NYC Load Time Series

The msts() function lets the user define multiple seasonal components. Two msts() time series were created to see if there is any monthly seasonality as well. The frequencies supplied here were: c(24, 168, 24*365.25). The components obtained from the msts function are shown below: Figure 6.

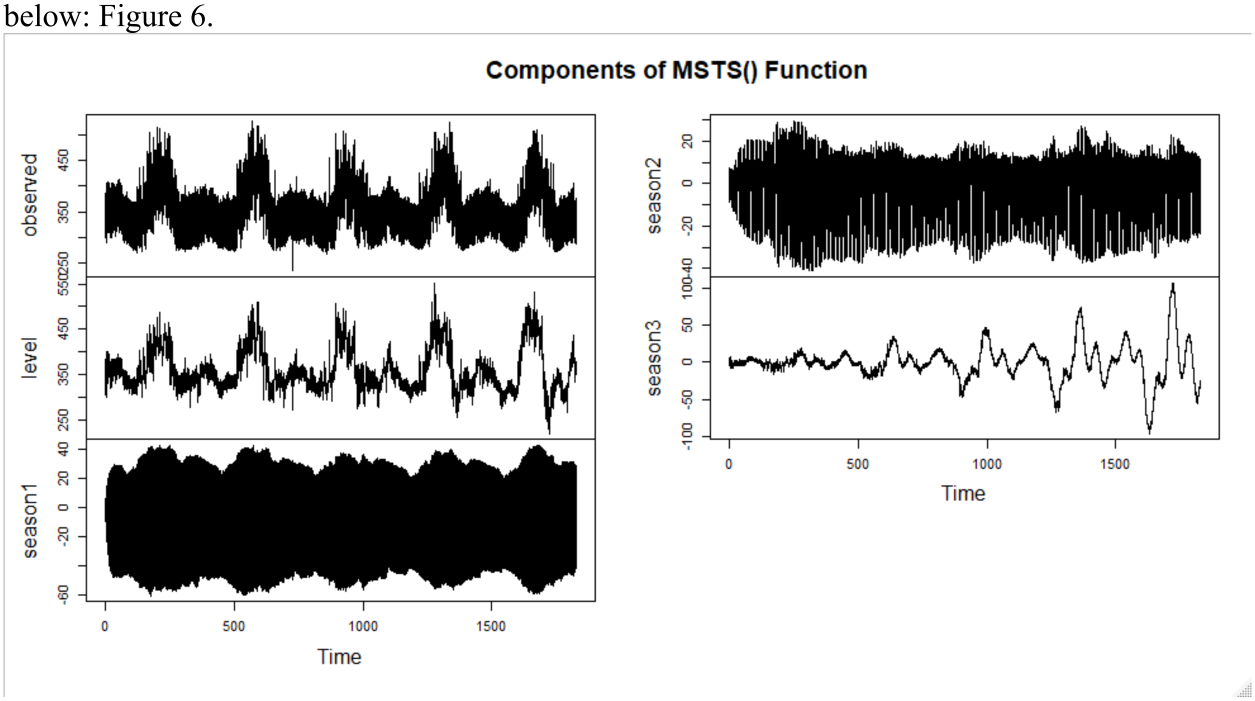

Figure 6: Multiple Seasonality in NYC Load Data

The time series constructed above were used for multiple forecasting models. These models are explained below. Our aim is to understand which model would best represent load data on hourly resolution.

5. Modeling

Multiple models were considered to fit the load data and forecast one year into the future. In this section, we compare the models used based on their forecast and accuracy.

1. **Simple Exponential Smoothing**: This function was accessed using the ses() function. R could assign the alpha value. This method was chosen as a baseline forecasting model as it is used in cases when there is no clear trend or seasonality. The alpha value introduces a weightage component on the previous observations. Larger the value of the alpha, more weight is given to recent observations. Alpha value can vary from 0 to 1.

2. **SARIMA**: The Seasonal Autoregressive Integrated Moving Average (SARIMA) model was run by using the auto.arima() function from the forecast package. In this function, R performs multiple iterations to assign AR, difference, MA components to the trend and seasonality components of the series based on model performance parameters like AIC and BIC. R allows users to add an external regressor as an argument to improve the model quality. We supplied the Zonal LBMP as the external regressor, however, the AIC of the model did not improve significantly.

3. **TBATS**: The Trigonometric seasonality (T), Box-cox transformation (B), ARMA Errors (A), Trend and Seasonal components (TS) model is capable of modeling with multiple seasonality unlike the other models discussed here. The TBATS model is very similar to the exponential smoothing model. Unlike linear terms in exponential smoothing, TBATS can represent timesteps as trigonometric terms based on Fourier series. TBATS model lets us model with Box-Cox transformation. State-space model assumes a normal distribution of the data used; however, our data violated this assumption. Therefore, we transformed our data such data it resembles a normal distribution before analysis.

6.Results

6.1 SES Model: As expected, the SES model performed poorly. The model automatically selected its alpha value as 0.999, which indicates, the SES simply turned to a naïve forecast in which the next timestep value was equal to the previous one. Therefore, we received a straight line for the forecast which does not represent the trend or seasonality accurately. Figure 8 represents the forecast for one year ahead into 2020. Figure 7 represents the residuals from the forecasted value when compared to the original data. The ACF plot of the residuals exhibit some seasonality indicating that there were components of seasonality that was not captured in the model.

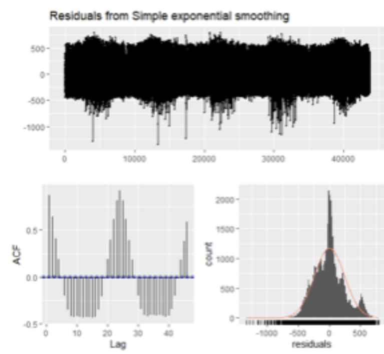

Figure 7: Residuals from Simple Exponential Smoothing

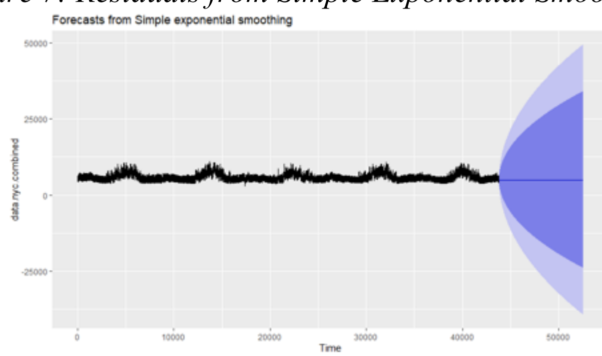

Figure 8: Forecasts from Simple Exponential Smoothing

6.2. SARIMA Model: Given the auto.arima() function picks the best model on the basis of lowest AIC value, the function returned the following model: (5,0,1)(2,10). As expected, we see that the model has multiple AR components. However, to limit the AR and MA seasonal components to one, we introduced the limit of max.P = 1 and max. Q =1. The results indicated an increase in the MA trend component as: (5,0,2)(1,1,0) with drift. The forecast is indicated in the Figure 9. Since the SARIMA can model only one seasonal component, we can see that the model picked the daily components (given time series frequency). The annual and weekly seasonality will be picked in the residuals of the forecast. The residuals of this model are indicated in the Figure 10. These results emulate the actual data much better than the simple exponential smoothing model.

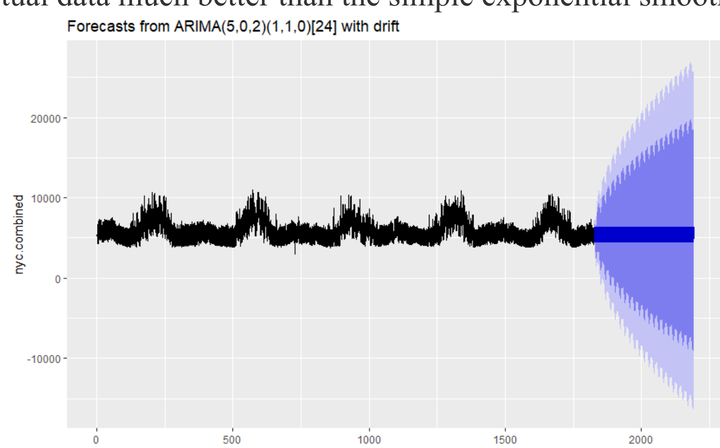

Figure 9: Forecasts from SARIMA model

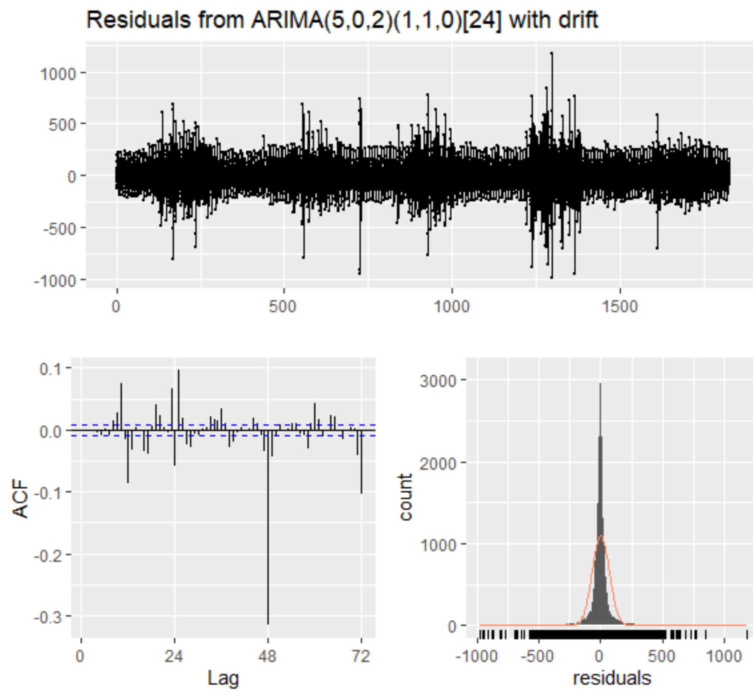

Figure 10: Residuals from SARIMA model

6.3.TBATS Model: The TBATS model emulates the actual data the best. Given the multiple seasonality modeled, the daily, weekly, and annual seasons are seen in the forecast. As expected, we notice that peak in the following year is lower than the previous years. Figure 11 indicates the forecast by the TBATS model. The residuals from the TBATS model is indicated in the Figure 12. The magnitude of the residuals is much lesser than what we observed in the previous two models with no seasonal components in the ACF of the residuals.

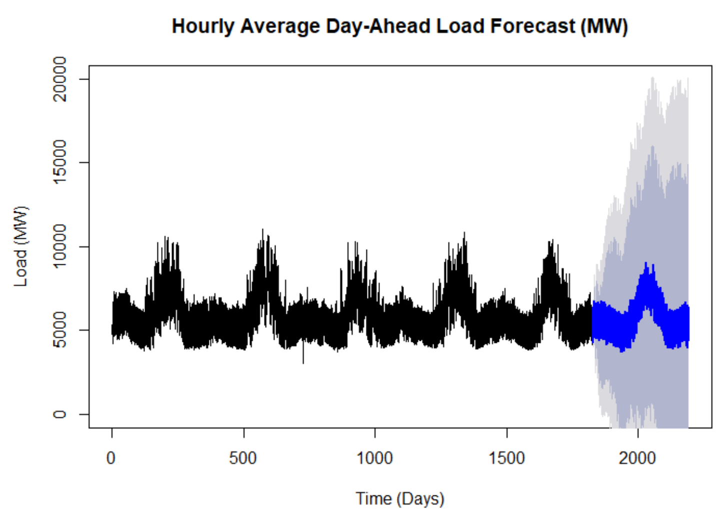

Figure 11: Forecasts from TBATS model

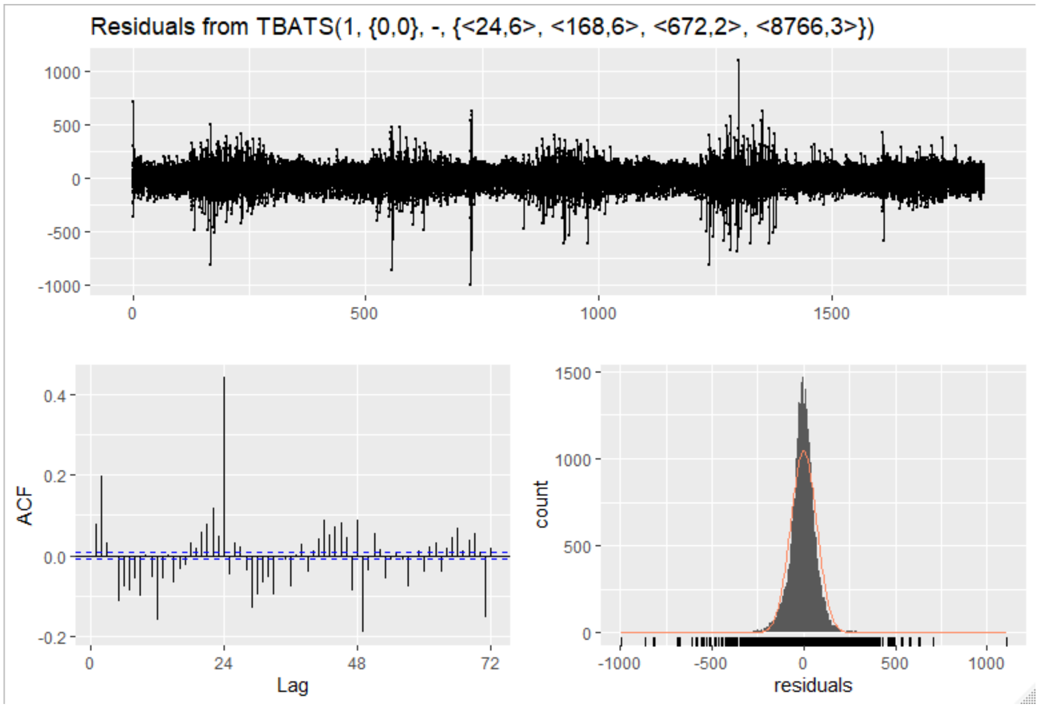

Figure 12: Residuals from TBATS model

7. Discussion:

7.1 Diagnostic Tests

The correlation matrix in Figure 13 suggests that there is a very weak correlation between load forecast data and price, a meagre 0.374. The slow decay of auto correlation function (Figure 14) suggest that this data is not a white noise series but is autoregressive, meaning that the step in this time series is dependent on previous values of the same times series. A strong seasonality can be observed at every 24-hour lag. The Seasonal Man-Kendall test suggested that the series is non-stationary with a weak downward monotonic trend.

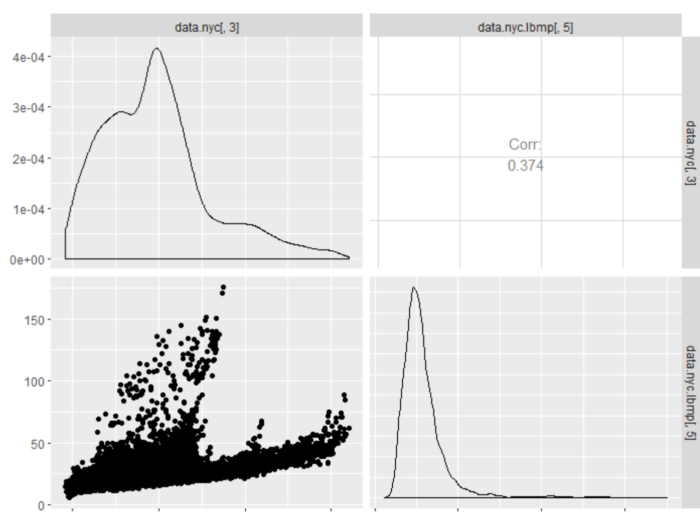

Figure 13: Correlation Matrix of Load and Price

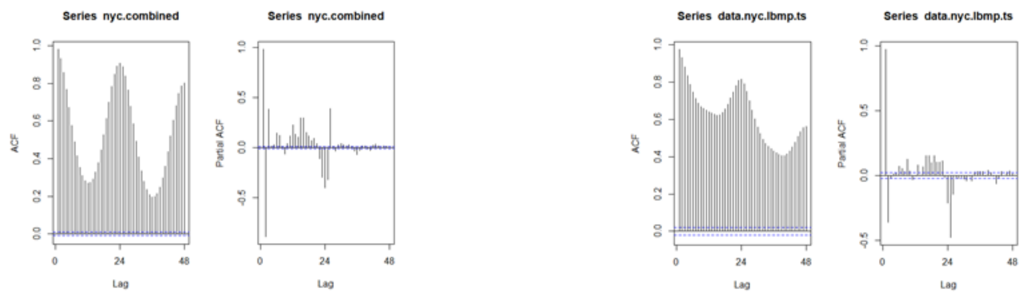

Figure 14: ACF and PACF plots of Load and LBMP respectively

7.2 Seasonal Exponential Smoothing

This model was not able to capture all the seasonality. The residues left behind were significantly scattered away from zero with the highest values amongst all the three models. The forecast returned was a straight line which is because of its inherent nature of predicting from one single value of actual data.

7.3 Seasonal ARIMA

This model did a better job at capturing the seasonality but not the best. The residues left behind were also significantly scattered away from zero and relatively higher values as well. The forecast only contained daily seasonality and not any monthly variations.

7.4 TBATS

This model did the best job amongst all three studied. The residues left behind were relatively scattered around zero with lowest magnitudes. The forecast was able to capture all levels of seasonality. Interestingly, the overall nature and the peaks were lower in magnitude than the actual data, suggesting a total overall reduction in load.

8. Conclusion

The data obtained for both load and price was comprehensive enough to do a good analysis but the lack of external regressors was felt which could have explained all the different levels of variations better, in turn giving a better forecast. The data is limited to New York City, a zone which has one of the highest population densities in the US and a wide range of consumption practices. Even though the model considered a large amount of historical data, a lot more parameters would help the models do more accurate analyses.

Consumption behavior, energy efficiency practices, change of energy policies, change of weather conditions, influence of natural disasters and various political actions are very likely to affect the load usage and we suggest analysts to find ways to consider these factors in their models.

Since this type of forecasting, with multiple levels of trend, is influenced from a significant period of historical data, machine learning algorithms like various kinds of neural networks can probably do a better job at understanding these hidden relationships and capturing them to make stronger forecasts. They can also use more factors with greater flexibility and ease of computational power in doing such analyses and forecasting.

Change of energy sources and novel demand-response programs are also likely to have unprecedented changes in load consumption behavior. Participation of novel energy storage technologies in the NYISO market are also likely to influence the prices. Hence, it is important for grid planners, policymakers, analysts and regulators to work in conjunction to build accurate forecasts.

9. References:

1. TBATS Model: https://pkg.robjhyndman.com/forecast/reference/tbats.html
2. https://rdrr.io/cran/forecast/man/tbats.html
3. https://rdrr.io/cran/forecast/man/msts.html
4. NYISO Demand response programs: https://www.nyiso.com/documents/20142/1398619/NYISO+Demand+Response+FAQs+for+Prospective+Resources.pdf/eb272dd5-4ac3-4d23-d7c0-16bf3cda61bb
5. Renewable Energy insights: https://www.renewableinsights.com/2019/06/new-york-adopts-the-nations-most-aggressive-clean-energy-goals/
6. Power grid international, Demand Response Optimization and Optionality https://www.power-grid.com/2014/04/29/demand-response-optimization-and-optionality/#gref
7. Reliability and Greener grid, 2019 https://www.nyiso.com/documents/20142/2223020/2019-Power-Trends-Report.pdf/0e8d65ee-820c-a718-452c-6c59b2d4818b?t=1556800999122
8. Load and Price data source https://www.nyiso.com/load-data