ISYE-6402 Time Series Analysis

Final Project

Forecasting Hourly Electricity Consumption in Georgia

VIDYUT RAO

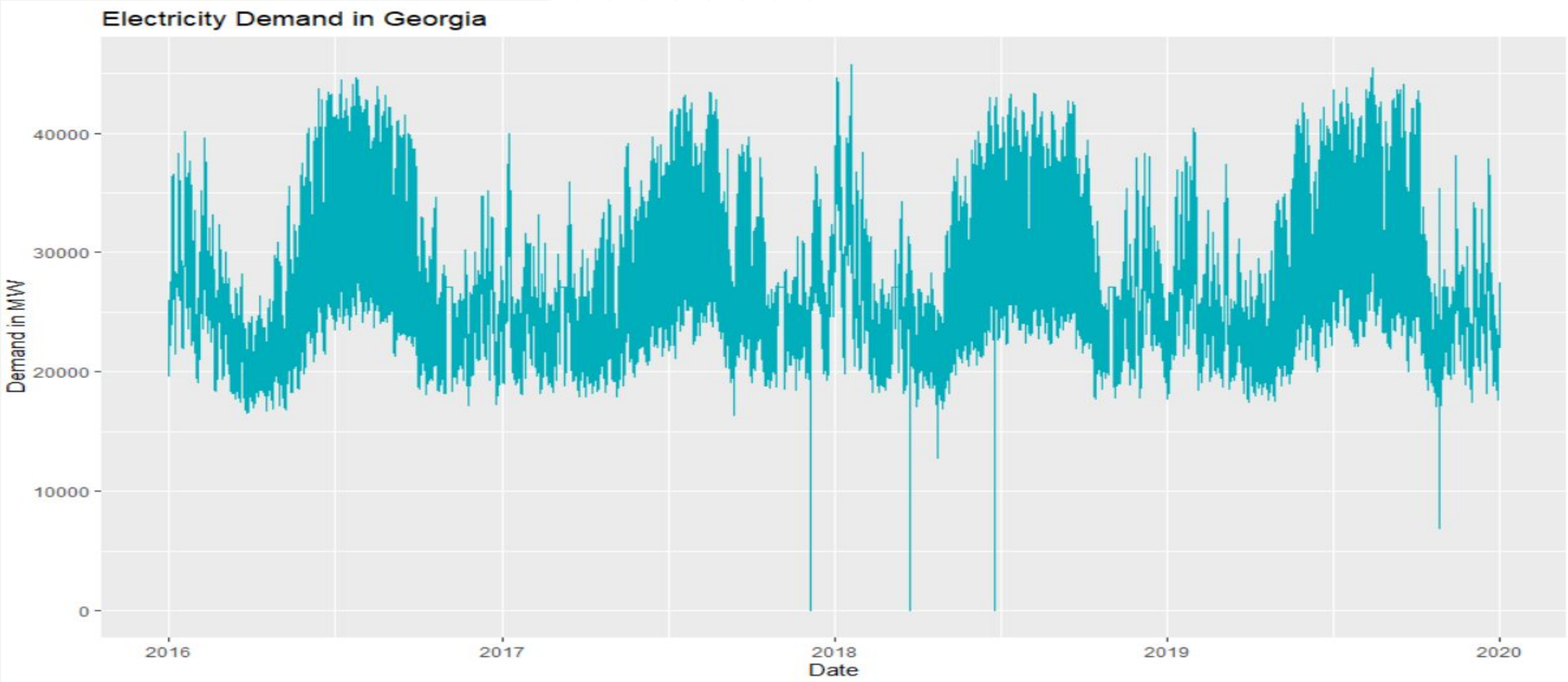
903511348

CREATING THE NEXT®

Introduction and Data

- This project aims to study the hourly electricity demand (or consumption) in Megawatts in the state of Georgia and develop an approach to forecast it.
- The electricity demand data is available from Jan 2016 to the present (April 2020). However, due to the ongoing COVID-19 pandemic, the data generated for the last couple of months have been erratic and do not follow the patterns we see in the preceding years. To make for a smooth analysis, I have only utilized data till December 2019.
- The data was obtained from the US Energy Information Administration website which provides a centralized and comprehensive source for hourly operating data about the high-voltage bulk electric power grid in the Lower 48 states. The data is collected from the electricity balancing authorities (BAs) that operate the grid.

Exploratory Data Analysis



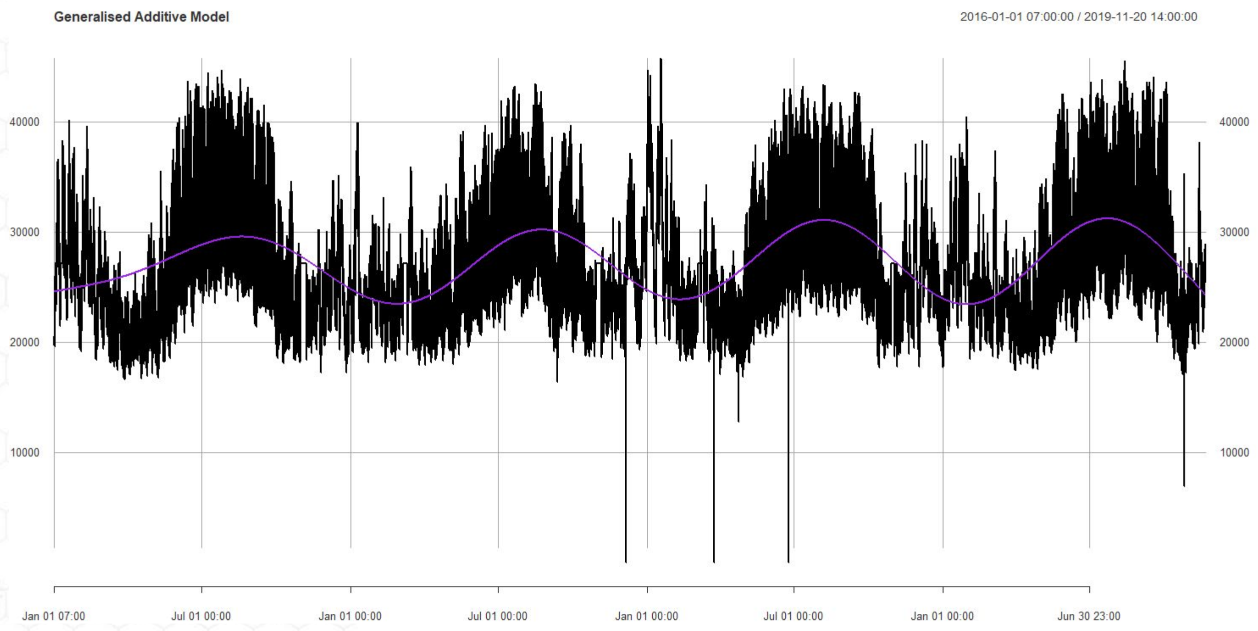
The plot of electricity demand in MW over time above clearly displays a yearly seasonality. Additionally, hourly and weekly seasonality might be present too, unnoticeable at this scale.

Approach

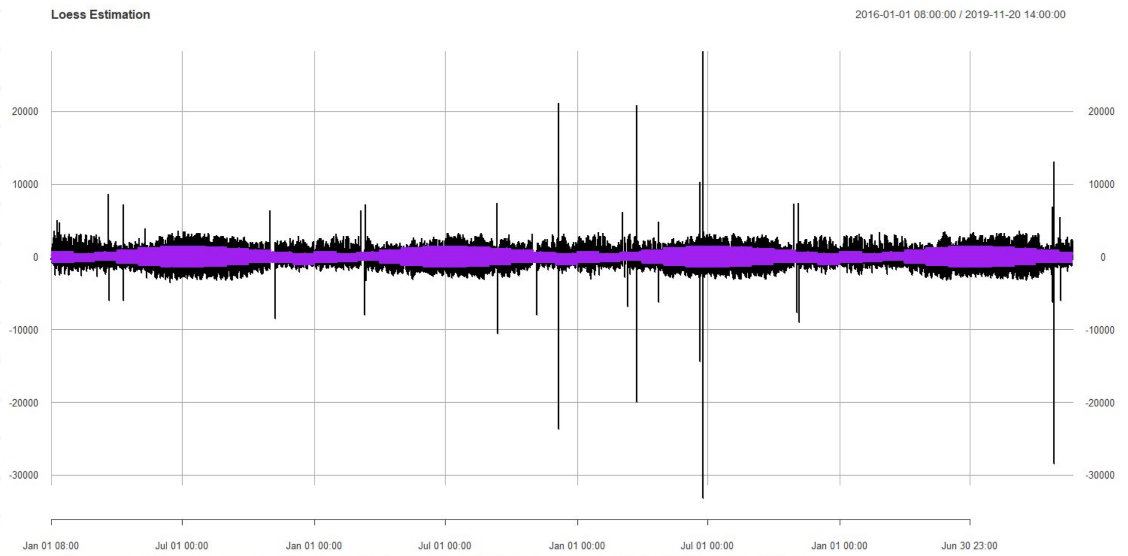
- After some initial data cleaning and wrangling, we first test multiple models to capture the trend and seasonality of the time series and remove those elements of the time series.
- We then find the parameters of the ARIMA model that best fits our detrended and deseasonalized time series.
- Noticing that the residuals of the above model isn't stationary, by studying the ACF and PACF plots, and that the variance has a time varying trend to it, we fit a ARCH/GARCH model to it.
- With the last 1000 hours of the data as the validation set, we evaluate the performance of our model using Root Mean Squared Error as the metric.

Trend and Seasonality

1. To the left, we have a generalized additive model fitted to the training set to showcase the seasonality and trend (in purple color).



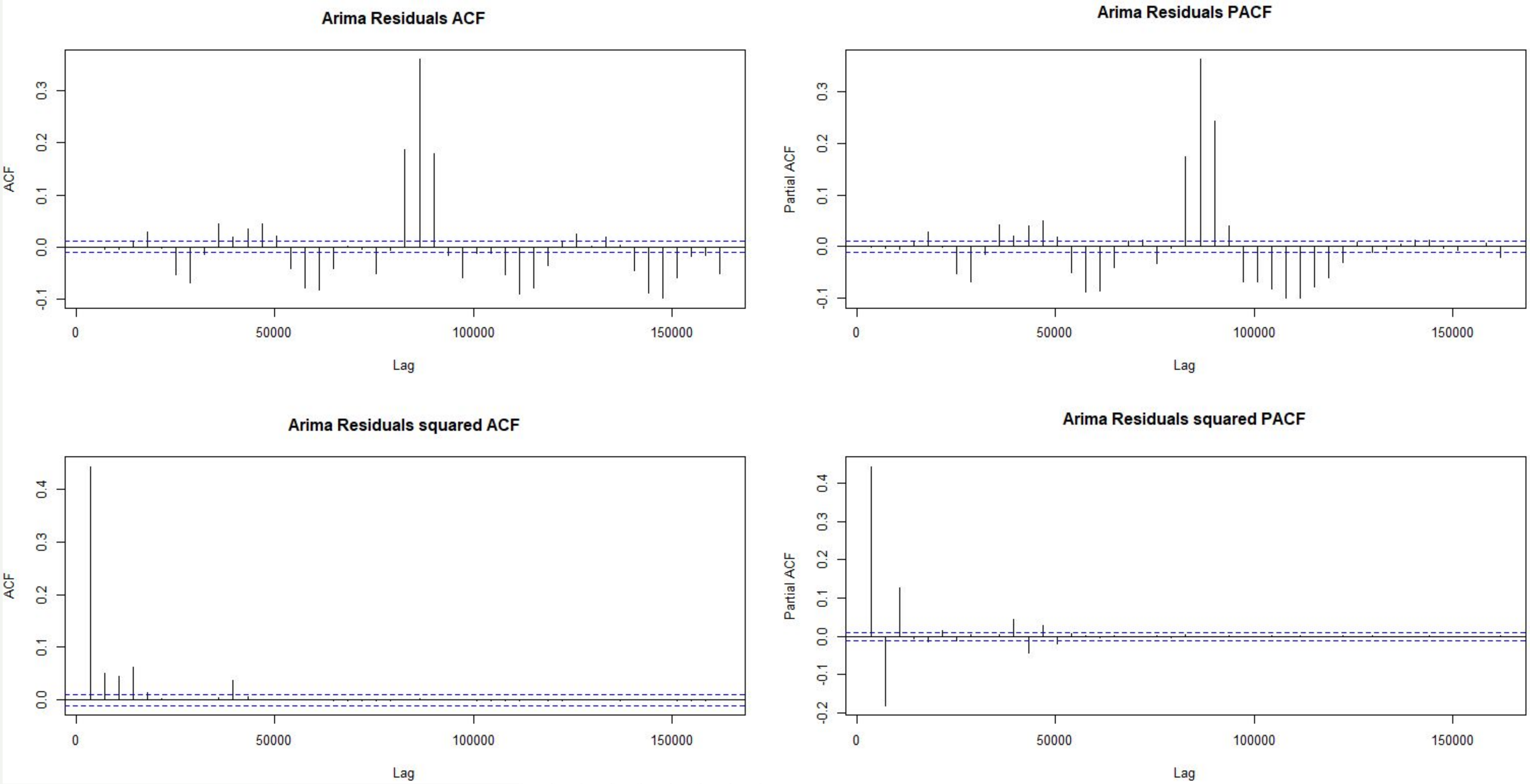
2. Realizing that the data needs differencing of order 1, we fit the best model, Loess, from empirical testing to the differenced data. We remove this estimate and proceed.



ARIMA

- We then fit an ARIMA model to the data using the *auto.arima()* function available in R. This relaxes the need for us to manually search for optimal (p,d,q) parameters. The best parameters were found to be (6,0,4) [Note: the data was already differenced in the previous step].
- In the next slide I have added the ACF and PACF plots of the residuals of the fitted ARIMA model.
- While the ACF and PACF plots of the residuals themselves don't give us clear information about the nature of the volatility, the same plots for the squared residuals tell us that the variance is autocorrelated to some lags.

ARIMA (continued)



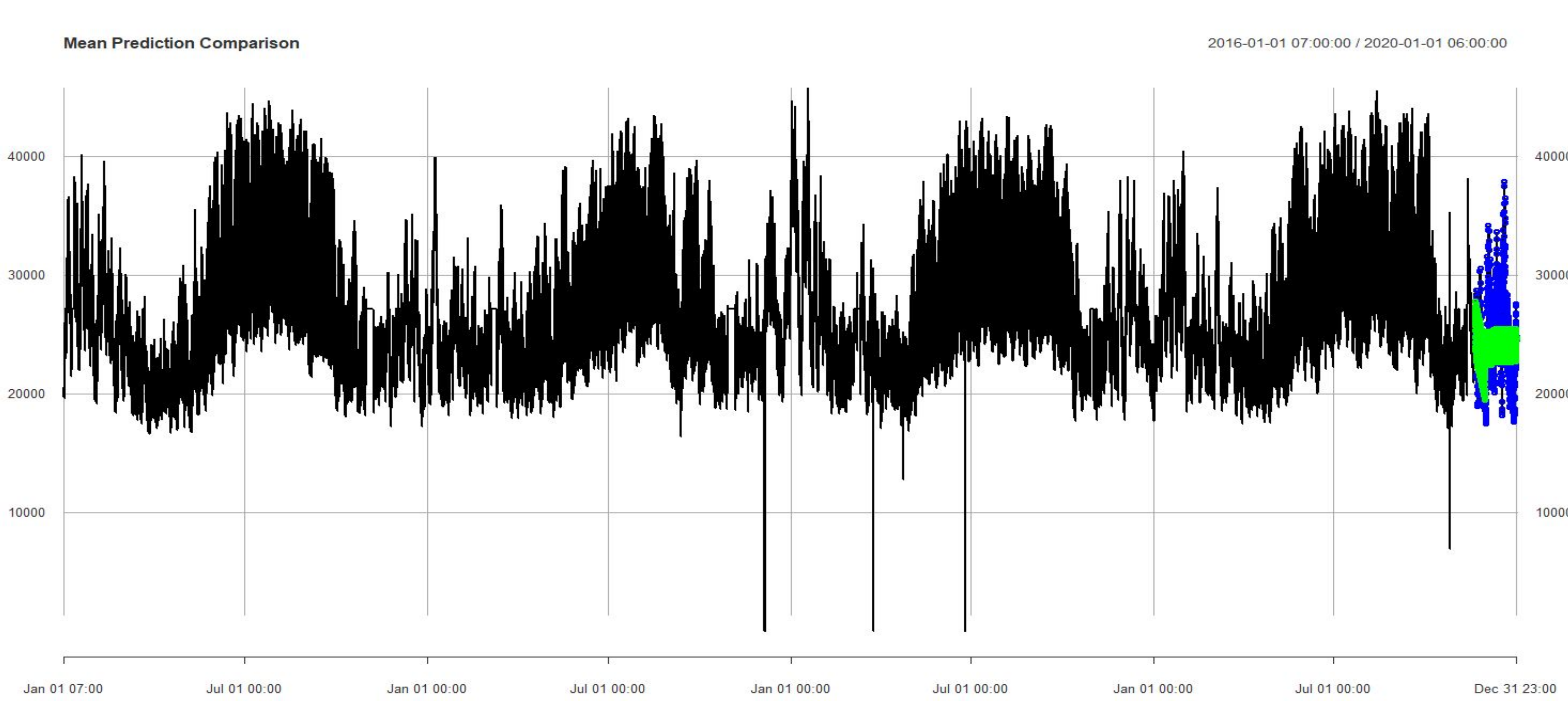
GARCH

- The results from the ARIMA section and conducting *Box-Pierce* tests tells us that a variance model is required.
- We run a grid search to find an optimal set of parameters for the GARCH(p,q) model. This was found to be (1,3).
- The parameters were chosen on the basis of their Bayesian Information Criterion (BIC), the lowest one being 15.0047.

Forecasting

- With all our models in place, we forecasting the last 1000 data points from 2019/11/20 09:00:00 in steps of 50 i.e. we forecast 50 points into the future with the training data and then add these forecasts to the training data and forecasts the next 50 and so on.
- We also add the seasonal and trend components back to these forecasts and undifference the data.
- The resulting forecasts are plotted in the next slide.
- The final RMSE of our model is **3650.836**

Forecasting (continued)



Green: Forecasted Values Blue: Actual Values

Results and Conclusion

- In order to gauge the relative magnitude of the RMSE obtained, I also built a model using the popular *xgboost()* function in R with predictors such as hour of day, month of the year, day of week etc, as well as another model which only predicts the mean of the training data.
- With the mean model we see an RMSE of **4511.758** and the xgboost model gives us a RMSE of **4472.886**.
- We see that the ARIMA/GARCH model we built has a **19.08%** improvement over the mean model and a **18.38%** improvement over the xgboost model.
- While this performance is great, there is much room for improvement. With better computational power we can calculate the test data in steps of 1 as opposed to the step of 50 I used, and use each forecasted point to predict the next point. This will give us far more accurate forecasts.
- Additionally, we can also build vector autoregressive models by studying the interaction of the energy demand in Georgia with that of neighboring states or even the weather (since temperature has a high correlation with the utilization of electricity consuming devices)