Analysis and Forecasts on PJM Electricity Consumption

Yiwen Wang, Yike Xu, Mingyu Zhang, Siyuan Zhang
Nicholas School of the Environment, Duke University

Abstract
Eastern connection grid in the U.S. is partly covered by one of the world's largest competitive wholesale electricity markets, PJM, which is short for Pennsylvania-New Jersey-Maryland Interconnection (PJM). This report aims to understand the electricity consumption pattern in PJM service territory in the past decade, identify the best forecast model for future electricity consumption among SARIMA, SNative, Smoothing model. Specifically, we target to analyze the electricity consumption pattern in PJM during 2010 to 2015 based on historical data. Throughout this timeline, we compared the pattern in PJM west region and east region with winter and summer seasons divided up. Furthermore, we used autocorrelation function plots and Augmented Dickey-Fuller tests to test the seasonality. With the overview of electricity patterns analyzed in these two subregions and seasons, we then applied forecast models to predict their electricity consumption from 2015 to 2017 and assessed the accuracy with historical data. Results indicated that the forecast models work better with annual data overall. More specifically, the SARIMA model works more accurately with annual PJM electricity consumption, while the SNaive model works more accurately with seasonal PJM electricity consumption.

Keywords: Electricity Consumption, PJM, Time Series Forecast

1. Introduction

PJM Interconnection is part of the eastern interconnection grid operating an electric transmission system serving all or parts of Delaware, Illinois, Indiana, Kentucky, Maryland, Michigan, New Jersey, North Carolina, Ohio, Pennsylvania, Tennessee, Virginia, West Virginia, and the District of Columbia (Figure 1). As a regional transmission organization (RTO), PJM plays a complex and balancing game for the demand and supply of electricity. They monitor the electric power system for all of the parts of 13 states and Washington D.C., which are completed through a variety of markets.

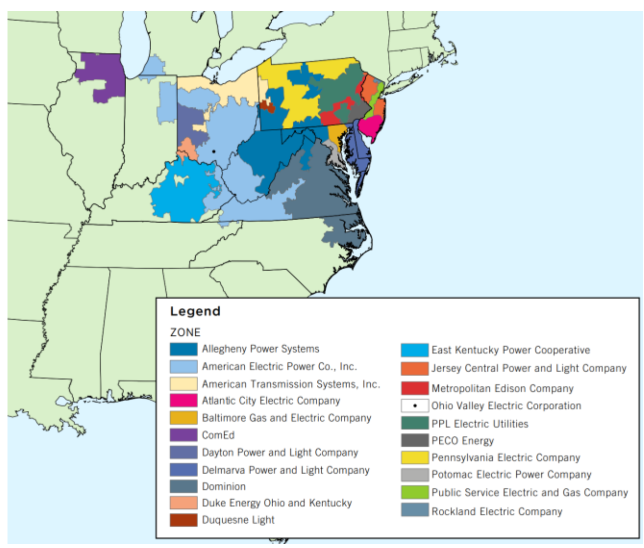


Figure 1. PJM-zones, by Dec. 03, 2018[1]

[1] "Map of PJM Territory Served." *PJM*, 3 Dec. 2018, www.pjm.com/-/media/about-pjm/pjm-zones.ashx?la=en.

The operation mechanisms at PJM electric power markets have evolved through changes in time. As increasingly more utility companies joined the PJM network from 1927 to 2013, PJM east and west regions have shown different evolutions[2]. Specifically, PJM west has the bulk of vertically integrated utilities, where generation, transmission, and distribution (GTD) are monopolized by their single entities instead of outsourcing part of investment and management to another company. PJM east, on the contrary, relies mostly on power pools for sharing capacity and single dispatch, where utility companies could partly share the generation and transmission network for their own customers. After 1990, the wholesale electricity market and the creation of a real-time market came into the picture and profoundly changed the way customers buy their electricity and the end-use consumption.

Given the uncertainty of the electricity consumption pattern based on flexible operation mechanisms at PJM territory, researchers have been analyzing optimized pathways to better the forecast of future consumption. Forecasts of hourly electricity loads and weekly overall loads are important for market players with utility at each stage, including generators, retail sellers, customers. Through preliminary research, we understand that among a variety of time series modeling techniques, some models work better for long-term forecasting, others work better with relatively short-term forecasting. Considering the availability of the data year span, we aimed to research on the following aspects:

I. Compare electricity consumption patterns of PJMW (west) and PJME (east) from 2010 to 2015, and summer versus winter.
II. Assess forecast methods by comparing the forecast results of 2016 and summer/winter.
III. Identify the best forecast model annually and on the season.
IV. Recommend improvements in forecasts.

2. Data

The dataset used for the analysis is obtained as hourly energy consumption data of PJM regions in megawatts. The dataset is published on the *Kaggle* website[3] And originally organized from information at PJM's official website. It involves hourly energy consumption data of different PJM zones. We chose PJMW and PJME subsets for analysis.

For the data cleaning process, we first sorted the data ascendingly based on hours. We identify several missing values of certain days at the time of 0, which is likely to be the missingness during data collection. To handle this missingness, we excluded consumption data at time 0 for

[2] Glazer, Craig. "The Evolution of the PJM market in the United States: Looking Back to Look Forward." *PJM*, 5 Sept. 2018, csis-prod.s3.amazonaws.com/s3fs-public/event/180926_Glazer.pdf.
[3], “Hourly Energy Consumption.” *Kaggle*, 30 Aug. 2018, www.kaggle.com/robikscube/hourly-energy-consumption.

each day. Moreover, because the timeframe we are looking at is for years, we further calculated the weekly average for hourly data for analysis in order to narrow down the original data volume in thousands.

For the forecast analysis based on separate season data (summer/winter), we define summer to be June-September and winter to be October-January. We used this division based on energy consumption peaks, although they are not the usual definition of summers and winters. The summer and winter data subsets were separated from each year and combined via Python-based on months.

Name	Period	Observations
PJME	2010-2015	448
PJMW	2010-2015	448
PJME_summer	June to September in 2010-2015	149
PJME_winter	October to January in 2010-2015	141
PJMW_summer	June to September in 2010-2015	149
PJME_winter	October to January in 2010-2015	141

Table 1. Sub-datasets Used for Forecast

3. Methodology

Except for using Python for data preparation, all the analysis is conducted with R. The R packages we used for the analysis include "forecast", "tseries," and "smooth." The data was first converted into a time series format with the ts() function. Then, we have an overview of each sub-datasets by plotting the original series, decompositions, ACFs, and PACFs to understand the seasonality and trends.

To see the differences in seasonality and trend between summer and winter periods from 2010 to 2015, the PJMW and PJME hourly datasets were firstly compiled into average weekly datasets and then separated into winter datasets and summer datasets through using the *subset* function. Therefore, in total, we have six different datasets performed in our project. Similar to using the annual datasets, decompositions, ACFs, and PACFs are performed for winter and summer datasets, giving initial assessments on their trends and seasonalities.

We conducted two types of forecast: The seasonal forecast and the annual forecast to compute the 2016 summer and winter electricity usage. We conducted these two forecasts to determine which one is better. As for the forecast models, we compared the three models, i.e., SARIMA, Exponential Smoothing, and Seasonal Naive models, and used the function accuracy() to obtain ME, RMSE, MAE, MPE, and MAP and evaluate forecast outputs of each model.

1) SARIMA Model

SARIMA model is fitted by the auto.arima() function in the forecast package. It can estimate the trend and seasonal elements of the series. The prediction of SARIMA is a weighted linear sum of recent past observations or lags.

2) Exponential Smoothing Model

Exponential Smoothing model is a forecasting method for univariate data. Exponential Smoothing forecast is completed by the es() function in the forecast package. The prediction of es() is also a weighted sum of past observations, but the model uses an exponentially decreasing weight for past observations[4]. The parameter alpha was taken as default values in the analysis.

3) Seasonal Naive Model[5]

The seasonal Naive model can be used for highly seasonal data, which is conducted by snaive() function in the forecast package. The forecast is set to be equal to the latest observation from the same season of the year. In the case of our dataset, the forecast value is equal to the same week of the previous year.

The seasonal forecast is computed by combining the 2010 to 2015 summer/winter and then use the model to forecast the 2016 summer/winter. Thus there are four seasonal forecasts conducted (two for PJMW and two for PJME). The annual forecast is computed by combining all data(whole year) from 2010 to 2015, and then we use the model to forecast 2016 summer and winter. Similar to the seasonal forecast, there are four annual forecasts conducted.

The final step is to compute the accuracy test. Our two major questions are : (i) which model is best for (seasonal/annual) forecast and (ii) is the seasonal forecast better than the annual forecast?

4. Results and Discussion

4.1 Annual trend and seasonality

To start understanding the consumption pattern in the PJM service territory, we first developed the time series data for the east region and west region, respectively, along the timeline of 2010 to 2016.

[4] Brownlee, Jason. “A Gentle Introduction to Exponential Smoothing for Time Series Forecasting in Python.” *Machine Learning Mastery*, 12 Apr. 2020, machinelearningmastery.com/exponential-smoothing-for-time-series-forecasting-in-python.

[5] “3.1 Some Simple Forecasting Methods.” *Forecasting: Principles and Practice*, otexts.com/fpp2/simple-methods.html. Accessed 25 Apr. 2020.

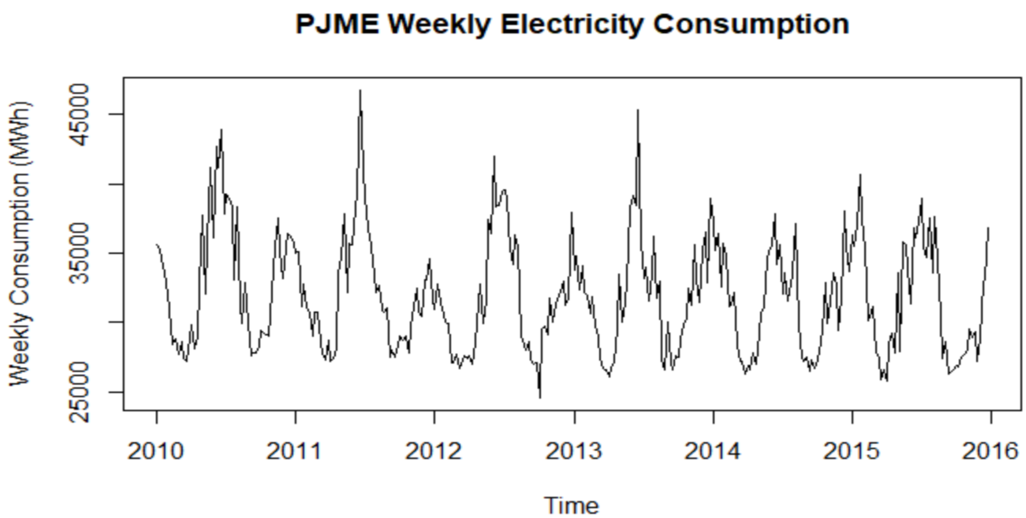


Figure 2. Weekly Electricity Consumption in PJM East region

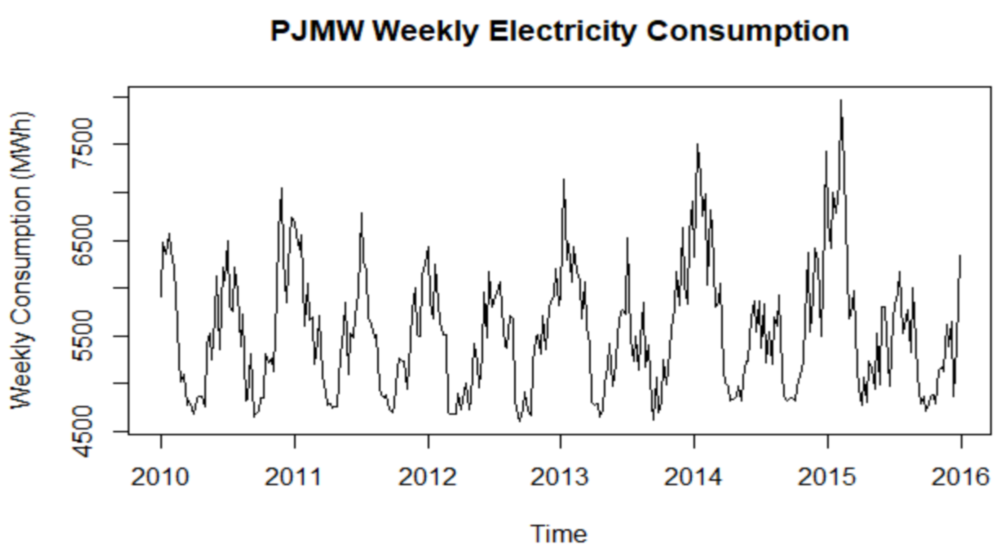


Figure 3. Weekly Electricity Consumption in PJM West region

From the initial time series plots we generated for PJM west and east regions (Figure 3 &4), we can see that both regions are showing clear seasonality throughout the timeline. Similarly, two peaks for each year are shown in both plots indicating that potential peak electricity consumption typically shows during the summer and wintertime for PJM interconnection, which will be discussed in later sections regarding summer and winter comparison.

4.1.2 Decomposition of annually average west and east PJM energy consumption

Another phenomenon we witnessed is that east region energy consumption shows smoother peak consumption compared to that of the west region. In other words, the west region peak period shows more fluctuation as opposed to peak hours in the east region, which can be seen in seasonality sections in Figure 5 and Figure 6. The reason for the differentiates might be relevant to the historical power market operation discrepancies among two regions.

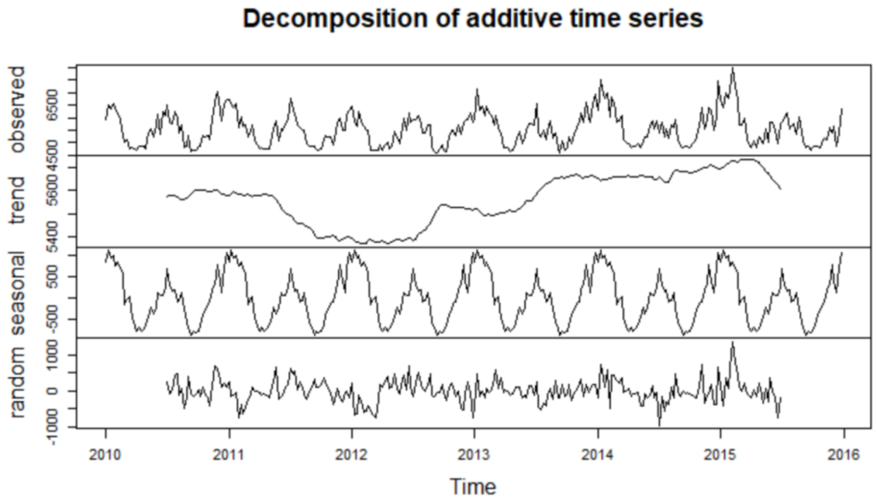


Figure 4. Decomposition of Weekly Electricity Consumption in PJM West region

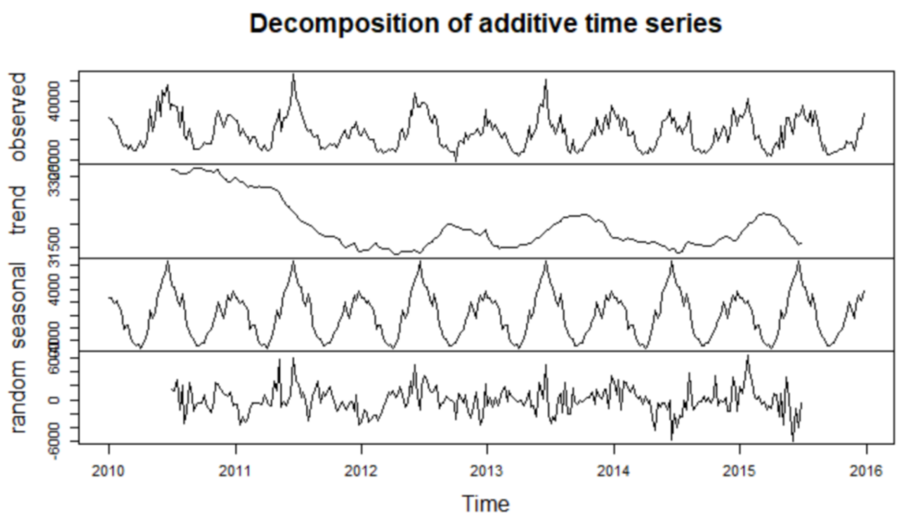


Figure 5. Decomposition of Weekly Electricity Consumption in PJM East region

PJM west region has been adopting a vertically integrated system as mentioned above, while PJM east region is following a traditional power pool arrangement. The mechanism with power pool arrangement refers to the cost-sharing of transmission lines, which has a higher chance for utility companies to operate with smoothed investment. Without "lumpy" power flow from single entities, the peak consumption would result in more consistency and concentration as opposed to more diverse operations in the west with more fluctuated peak consumption.

4.1.3 ADF test results

```
        Augmented Dickey-Fuller Test

data:  PJMW_t
Dickey-Fuller = -6.883, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary

        Augmented Dickey-Fuller Test

data:  PJME_t
Dickey-Fuller = -7.3413, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
```

Figure 6. Augmented Dickey-Fuller Test Results

To better understand the energy consumption in both regions regarding their seasonality patterns, we conducted the Augmented Dickey-Fuller Test (ADF test) to test out the stationarity. From the results (Figure. 7), p-values for both datasets show significance (<0.05) that indicates the fact we should reject the null hypothesis and accept the alternative hypothesis of being stationarity. With that, we understand that both time series have their seasonal pattern remain stable over time. Another thing we notice is that the test statistics for east regions are a little bit smaller than that of the west region (-7.3413 < -6.883), which indicates that this is a more restrictive test, and we can reject the null with a higher significance level. Essentially, we have greater confidence to say that east region energy consumption over time has a stable seasonal pattern.

4.2 Summer & Winter Trend & Seasonality

To understand the differences between summer months and winter months, decompositions were performed with results showing in table 2. From 2010 to 2015, both PJME and PJMW experienced downward trends for their summer electricity consumption. After 2015, both PJME and PJMW showed dramatically increasing electricity consumption. There is not much difference between PJME's and PJMW's summer trend and seasonality, which means the potential influential factors that drive the change have little impact from geographic characteristics.

Compared with two figures from the summer periods, the average electricity consumption is significantly higher during winter periods. The reasons behind the difference are mainly due to shorter days during winter, meaning increasing use of lightning, and excessive usage of electricity for heating purposes.[6] However, the two figures from winter datasets show more clear overall decreasing trends, showing decreasing average electricity consumption in winter over the years from both PJME and PJMW areas. From 2014 to 2016, both winter figures

[6] “U.S. Energy Information Administration - EIA - Independent Statistics and Analysis.” *Winter Residential Electricity Consumption Expected to Increase from Last Winter - Today in Energy - U.S. Energy Information Administration (EIA)*, www.eia.gov/todayinenergy/detail.php?id=29112.

show bumps followed with dramatically decreasing trends, indicating a slightly increasing trend of electricity usage from 2014 to 2015 and a massive decreasing trend from 2015 to 2016.

These moves from two winter figures are aligned with corresponding summer trends. Plus, there is no difference between PJME and PJMW, so that it is assumed the factor that drives the change is related to the overall climate and national economy. According to the EIA website, most areas in the U.S. experienced significantly increasing temperatures from winter 2015 to 2016 because of a strong El Niño weather pattern over the Pacific Ocean.[6] That explains the dramatically increasing electricity consumption trend from 2015 to 2016 summers and a decreasing trend from 2015 to 2016 winters. People increase the use of cooling equipment during summer and reduce the use of heating equipment during winter due to the overall higher temperature.

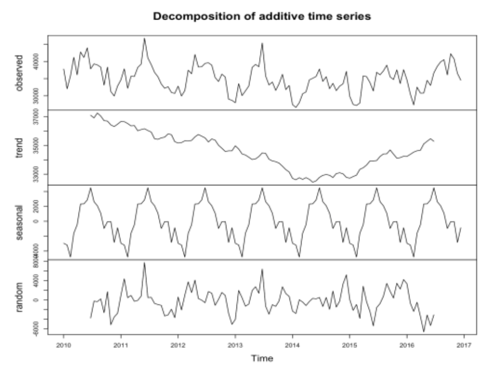
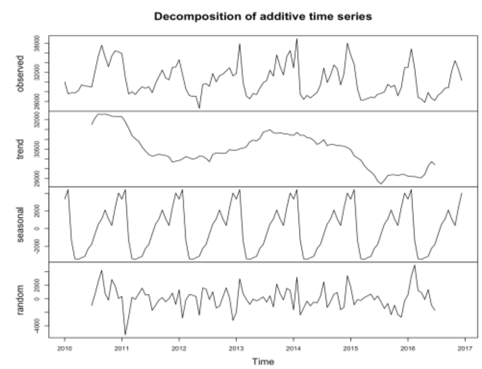
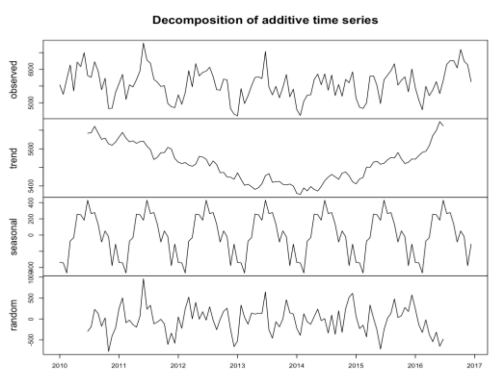
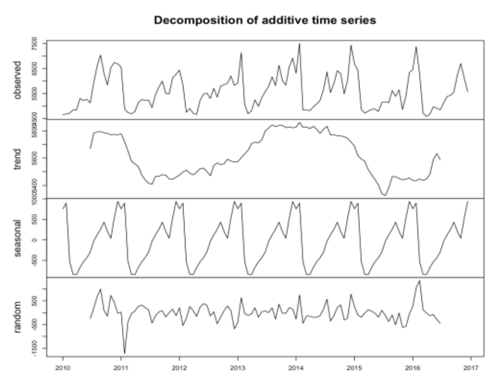
	Summer	Winter
PJME		
PJMW		

Table 2. PJME and PJMW Summer and Winter Weekly Average Electricity Consumption Decompositions.

4.3 Forecast Models

The forecasting models applied to our projects are SARIMA Model, Exponential Smoothing Model, and Seasonal Naive Model. From the figures below in table 3, forecasting results by using annual weekly PJME and PJMW electricity consumption datasets are more accurate than using summer/winter datasets. Among three models in PJME and PJMW annual forecasting plots, SARIMA performed the best.

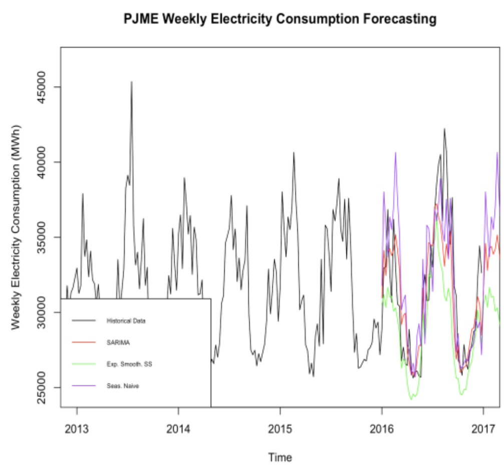
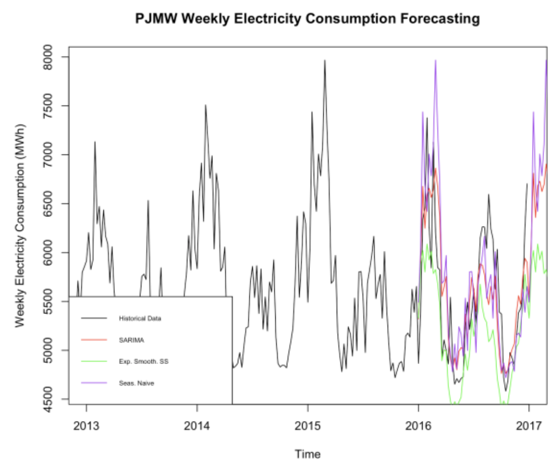
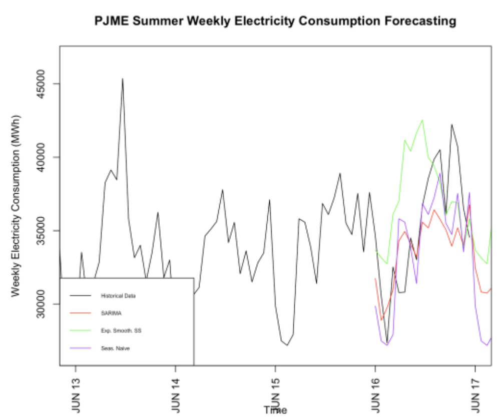
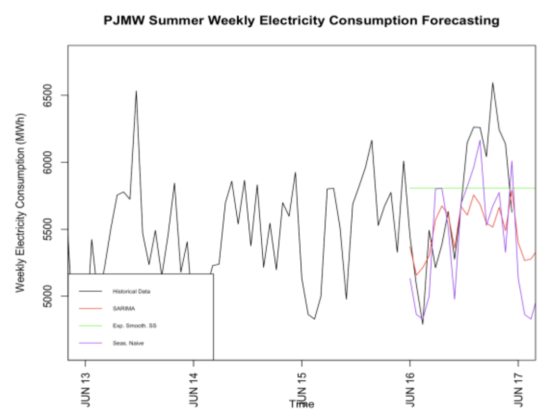
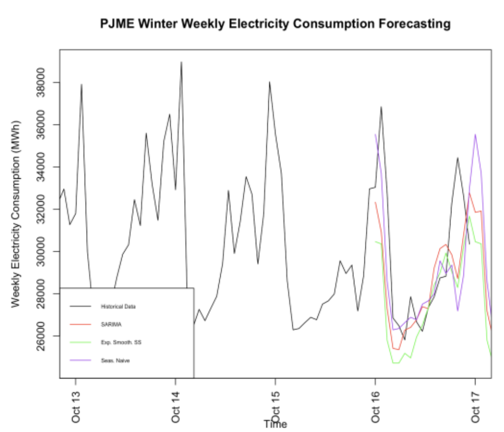
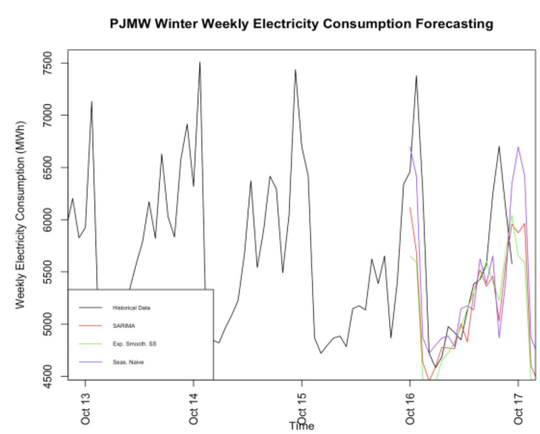
	PJME	PJMW
Annual		
Summer		
Winter		

Table 3. Annual, Summer, and Winter Weekly Average Forecasting models

As for the error analysis, we used the accuracy-test in R and produced the results as below. Table 4 illustrates our result for the error analysis.

	PJMW Summer	PJMW Winter	PJME Summer	PJME Winter	PJMW Annual Summer	PJMW Annual Winter	PJME Annual Summer	PJME Annual Winter
Best Method	SARIMA	SNaïve	SNaïve	SNaïve	SARIMA	SARIMA	SNaïve	SARIMA
Error								
ME	299.304	97.84	1341	256.1	250.42	22.1	1170	492.5
RMSE	476	637.9	3410	2500	401.42	401.67	3339	1577.112
MAE	393.78	410.5	2872	1684	318.7262	301.19	2809	1189.22
MPE	4.78	1.36	3.3	0.58	4.04	-0.65565	2.75	1.46
MAPE	6.55	6.95	8.13	5.34	5.328	5.13	8.044	3.81

Table 4 Error Analysis of the Forecast
Note: White form indicates the error analysis result for the seasonal forecast. Grey form indicates the error analysis result for the annual analysis.

From Table 4, we could conclude that SNAIVE is the most effective model for the seasonal forecast, and SARIMA is the most effective model for the annual forecast. From Table one, we can tell that SNAIVE is good for three out of four seasonal forecasts, and SARIMA is good for three out of four annual forecasts. What is also interesting is that PJME has much larger error than PJMW, this could be due to the fact that the electricity usage for PJME is larger than the PJMW, and these errors measure the absolute error instead of relative error.

	Difference(Annual - Seasonal)			
	PJMW Summer	PJMW Winter	PJME Summer	PJME Winter
Error				
ME	-48.8848	-75.74	-171	236.4

RMSE	-74.58	-236.23	-71	-922.888
MAE	-75.0538	-109.31	-63	-494.78
MPE	-0.74	-2.01565	-0.55	0.88
MAPE	-1.222	-1.82	-0.086	-1.53

Table 5: Difference in Errors between the annual forecast and seasonal forecast

From Table 5, we could see that most values are negative, and this result indicates that the seasonal forecast model has a larger error than the annual forecast model. Except for PJME winter, all the other columns show negative results. The only positive values are the ME and MPE for the PJME winter. Thus overall, we could conclude that the annual forecast model is better than the seasonal forecast model.

There are several underlying reasons that the annual forecast model is better. First, SARIMA itself is the most accurate model for the annual forecast[7], and the SNAIVE is the most accurate one for the seasonal forecast. However, SARIMA is, in general, more accurate than SNAIVE, and this could explain to some extent. Secondly, the annual model can pick up more underlying cycles than the seasonal forecast. Since the seasonal forecast only consists of four months, it cannot pick up much of other cycles, which might be crucial for forecasting. Thirdly, longer-term factors like climate impacts(a warmer year) can be represented better with the annual forecast than the seasonal forecast. Last but not least, the seasonal forecast consists of fewer data points. Since we only use five years (2010- 2015) of summer/winter to model, the seasonal forecast might have too few data points and therefore have poor accuracy. Thus, overall, these are the possible reasons that the annual forecast is a better model than the seasonal forecast.

5. Conclusion

Regarding the annual time series trend in west and east regions for PJM interconnection. The east region shows slightly downward throughout the year 2010 to 2015, while the west region shows a bit of fluctuation with average consumption remaining stable. And east region energy consumption over time has a stable seasonal pattern with a higher chance. Also, the west region shows more fluctuated peak times compared to east region energy consumption.

By separating PJME and PJMW annual weekly electricity consumption datasets into summer datasets and winter datasets, the results show differences between each summer trend and the corresponding winter trend. Even though for PJME and PJMW, both summer and winter trends are slightly decreasing over time, summer trends rise sharply around 2015, and winter trends

[7] Brownlee, Jason. “A Gentle Introduction to Exponential Smoothing for Time Series Forecasting in Python.” *Machine Learning Mastery*, 12 Apr. 2020, machinelearningmastery.com/exponential-smoothing-for-time-series-forecasting-in-python.

drop around 2015 due to climate factors such as the El Niño impact.[8] There are no clear differences between PJME and PJMW summer trends/seasonalities and their winter trends/seasonalities, meaning geographic factors do not play an important role in influencing the electricity consumption pattern here.

Overall the annual forecast is better than the seasonal forecast, and we should use the annual forecast to achieve better accuracy. As for future studies, there are two potential ways to extend this research and optimize our results. We would like to assess the impact of climate change and EV adoption on electricity demand. A lot of research has revealed the challenges of predicting electricity demand due to increasing penetration of renewable sources, broader EV adoption, and the uncertainty due to climate change .[9] Therefore, we would assume air temperature and EV adoption as two determinant factors, use air temperature as a factor to increase the accuracy of our current results, and generate scenario analysis based on the low and high predictive trajectories of EV adoption in PJMW and PJME areas.

[8] "U.S. Energy Information Administration - EIA - Independent Statistics and Analysis." *Winter Residential Electricity Consumption Expected to Increase from Last Winter - Today in Energy - U.S. Energy Information Administration (EIA)*, www.eia.gov/todayinenergy/detail.php?id=29112.
[9] Nagbe, Komi, Jairo Cugliari, and Julien Jacques. "Short-term electricity demand forecasting using a functional state-space model." *Energies* 11.5 (2018): 1120.