ETM 510 Demand Forecasting

Final Project: Journey through Energy Demand Forecasting using ARIMA model

Portland State University
Maseeh College of Engineering & Computer Science
Department of Engineering & Technology Management

Author:
Adrian Reyes
March 16, 2017

ETM 510
Demand Forecasting

Contents

List of Figures

1 Introduction

Over the time, the electricity industry has become the most important factor impacting the development of socioeconomic status worldwide. Accurate demand forecasting decreases the operating cost of utility companies, especially in a market environment [1]. Electricity demand forecasting is a useful policy tool for decision-makers; thus, it is valuable in allowing both power generators and consumers to make their plans. Accurate forecasting of electricity demand is not only valuable for power generators, allowing them to schedule operation of their power stations to match generation capacity with demand, but is also a fundamental piece of information used for trading in the energy market [2]. Consequently, electricity systems can be optimized by developing a scheduling algorithm for electricity demand. Because energy demand has increased in the last decade in several (not all) energy sectors worldwide, we must pay more attention to improvement of electricity demand forecasting.

There are other significant reasons that necessitate research on electricity demand. The first is global carbon dioxide emissions. Coal and coal-fired electricity are typical up- and downstream industries [3]. Also, grids in many countries fail frequently, causing widespread power outages that affect people's lives and cause huge economic losses. Therefore, accurate and timely electricity demand forecasting can provide a reliable basis for a reasonable plan for power generation and an adjustment of power policy and structures, which will decrease carbon dioxide emissions and outages. For example: The Northeast Blackout of 2003, which was a widespread power outage that occurred throughout eight U.S. states and Ontario, Canada on August 14, 2003 at 4:11 P.M. Eastern Daylight Time (EDT), was the biggest blackout in North American history and resulted in the loss of 61,800 MW of power to over 50 million people [4]. Electricity systems, along with power systems, the electricity market, and residents' lives, are affected by internal factors of power system and by the socio-economic environment, natural environment, law and policy, technical progress and population growth [5]. These factors make more difficult forecast demand, however, they are challenges to overcome with forecasting techniques.

Because of the electricity demand needs for accurate planning to avoid electricity shortages, many studies has been conducted focused on the electricity demand forecasting using different techniques. [6] used a neuro-fuzzy approach to model electricity demand and concluded that the neuro-fuzzy system performed better than neural networks, ARIMA model and Victorian Power Exchange (VPX) forecasts. [7] introduced the application of the triple seasonal methods in short-time electricity demand forecasting. [8] estimated changes in California's annual electricity demand using end-use energy models produced by each warming scenario. [9] used co-integration and ARIMA modeling to predict the electricity demand in Turkey.

Among the statistics forecast methods, non-structural prediction methods only consider the changes of data rather than the structural changes in the system, and because of the complexity of systems, structural prediction method cannot achieve accurate predicting results in many situations [10]. Therefore, using non-structural prediction models to forecast the complex systems may be a better choice. In this paper, we used the non-structural model Autoregressive Integrated Moving Average (ARIMA) method to estimate the future primary energy demand of nine different areas.

2 Methodology

The ARIMA demand model has been analysed extensively by researchers and used widely by forecasting practitioners due to its attractive theoretical properties and empirical evidence in its support. Since 1970s, when ARIMA models were developed, they have been studied extensively by many researchers. It should be noted that evidence from the forecasting competitions has shown the ARIMA methodology to be competitive in terms of forecast accuracy [11], hence, it provides support for the assumption of such processes.

The ARIMA has been originated from the autoregressive model (AR), the moving average model (MA) and the combination of the AR and MA. The ARIMA model can be used when the time series is stationary and there is no missing data within the time series. In the ARIMA analysis, an identified underlying process is generated based on observations to a time series for generating a good model that shows the process-generating mechanism precisely [12].

To start applying ARIMA model, the SMAlags (seasonal moving average) should be determined. Several methods determined the lags by dividing the number of observations by 4. Since in this study we used 14 observation (days), 2003-2016, to forecast the next value on the corresponding day in 2017, SMAlags was set to 11. Obtaining SMAlags by dividing by 4 with that small number of observation, the results would be wrong. The SMAlags and SMA values were determined by training the function so it imitates closely the actual data. As for obtaining MA, first autocorrelation graph should be determined. Autocorrelation graph, which provide information about the MA orders, is then drawn based on the specified lag number. The variance was determine from the actual data. Everyone of these values was recalculated to adjust the ARIMA function to seasonality. This means the function did not work the same way for the whole month in analysis. Thus, changes were needed in order to overcome this issue.

As part of the training ARIMA function part, we studied the data for the 24 hours of every day of the month in study. Each individual graph in 1 represent one day data, 24 hours, from 2011 to 2016. Similarly, figure 2 represent the days 16 to 24 of March, 10-18. This study was conducted for the whole month. It was interesting to see that through the years every hour of the same day followed a certain pattern or function. This finding was very helpful for the calculation of the values introduced in the paragraph above.

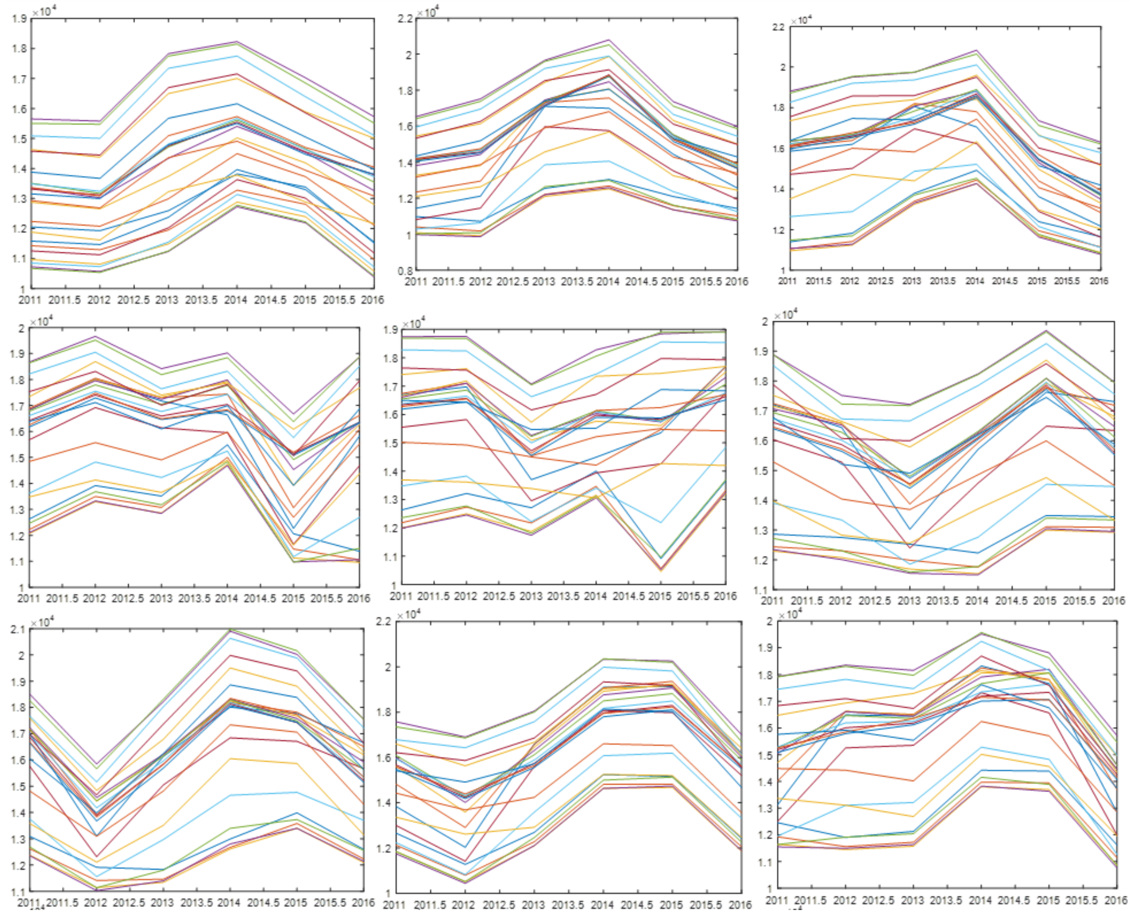

Figure 1: 24 hours data for the first 9 days of March from 2011-2016

On the other hand, percentiles were obtained from the mean square errors returned by the ARIMA function. The ARIMA function returns two values, the main value, also interpreted as the 50 percentile, and the mean square error. The 4 percentiles to the left of 50 (40, 30, 20, and 10) were calculated as the difference between the 50 percentile and (10, 20, 30, 40) times the mean square error respectively. Similarly, the percentiles to the right (60, 70, 80, 90) were calculated as the sum of the 50 percentile and (10, 20, 30, 40) times the mean square error respectively (see Appendix: Matlab script).

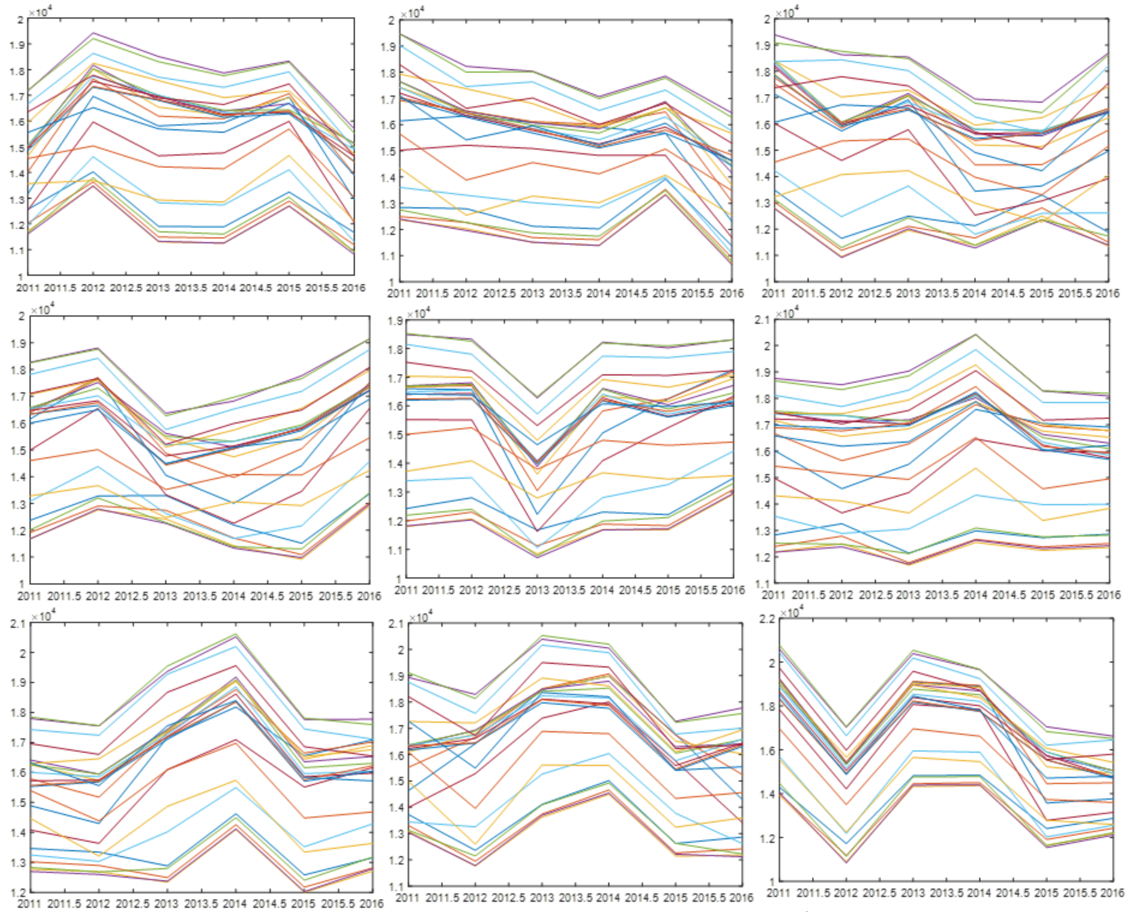

Figure 2: 24 hours data for the day 16 to 24 of March from 2011-2016

Matlab was the computer language used to create the program. All graphs were processed using Matlab R2015a. As mentioned before, the code developed is presented in the Appendix. The output is an excel sheet with the nine percentiles energy demand for the nine zones (including New England) analyzed in the competition for the month in analysis. The program also output graphs if it is required by the user. Some code modifications are needed to change month in analysis and to get an specific graph.

3 Analysis and Discussion

For the purpose of our study, we initially considered an ARIMA with MA =(0.9,0.7), SMA=0.52, SMALags=11, and Variance=0.2. Figure 3 shows the first results obtained in our journey. When the function was used to validate (recreate) the actual data, there was a huge difference between the forecasted and real data (unfortunately, we did not keep the very first code validation results).

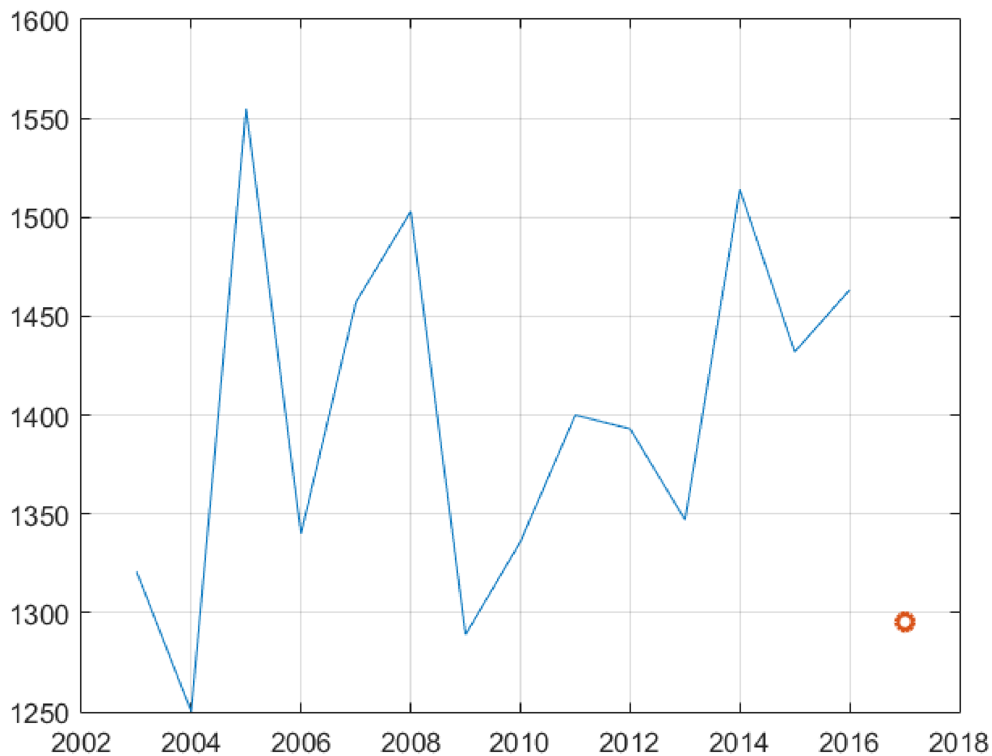

Figure 3: Forecasting for March 28th, 2017, using the same day data from previous years for New England

We then proceed to modify the ARIMA function accordingly and study seasonality. We thought that demand would have a close relationship with the day of the week. Figure 4 shows this study. The results was that there was not real evidence about the existence of that relation. As seen, we could get a lower demand on a weekend day than any other day of the week. The next potential relation to study would be between energy demand and temperature (weather). However, it would be a recommendation for future continuation to this study.

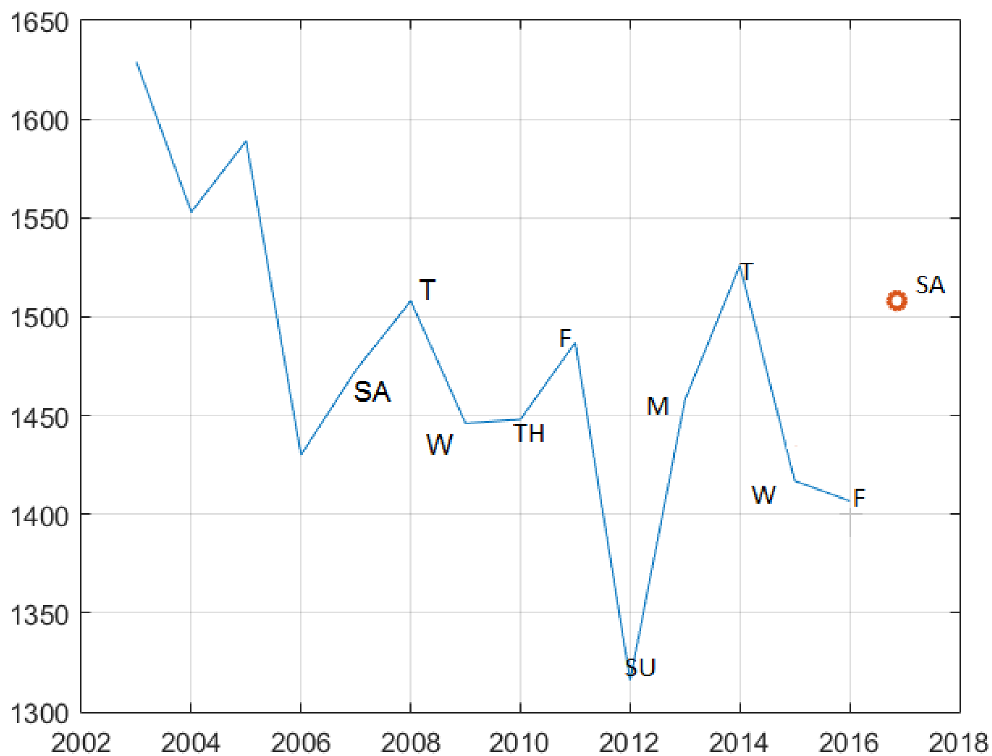

Figure 4: Seasonality study based on weekdays for March 11th, using the same day data from previous years for New England

In our third attempt we introduced two ARIMA function to forecast the future demand based on our previous seasonality results. In this case, for the first ARIMA function we used the calculated and trained values MA=(0.94,0.72), SMA=0.59, SMALags=11, and Variance=0.2. For the second function we used MA=(0.94,0.76), SMA=0.74, SMALags=11, and Variance=0.1. Figure 5 shows the validation of the function for the 2015 and 2016 data. As seen, the validation data match closely the actual data. It can be also noted that the forecasted demand for 2017 using this new approach was lower than the previous one.

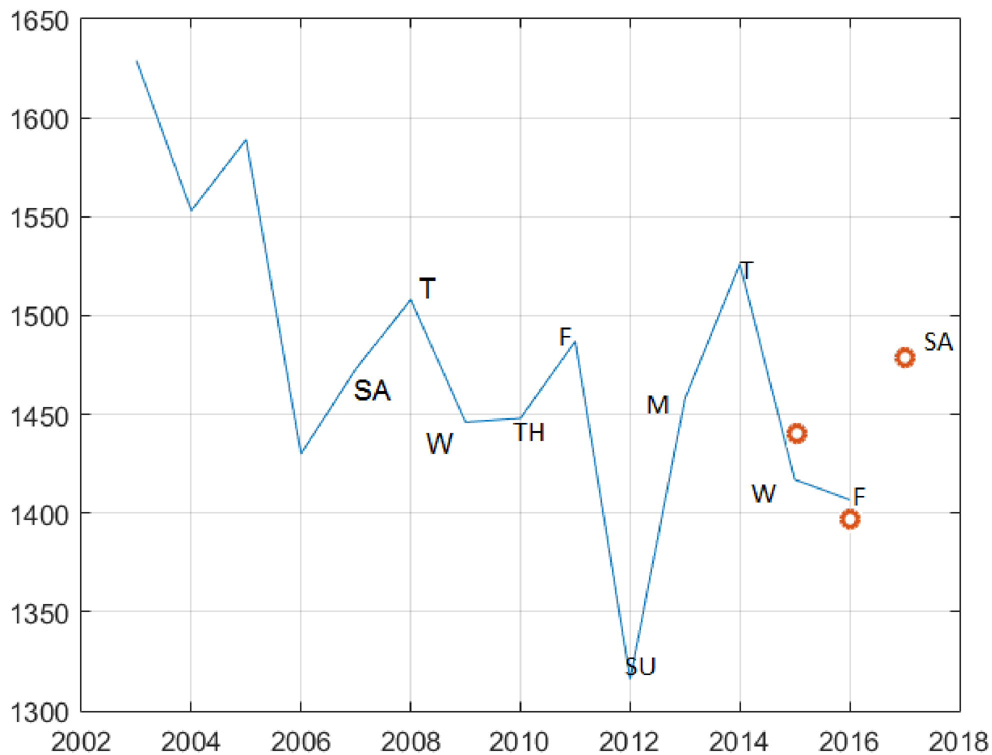

Figure 5: Validation and new forecasting results using the same day data from previous years for New England, March 11th

As requested for the energy demand competition, the output should be an excel sheet with the 9 percentiles for the 9 zones (including New England) analyzed. Figure 6 presents the nine percentiles for the first 24 hours of April 1st for New England.

Date	Hour	Q10	Q20	Q30	Q40	Q50	Q60	Q70	Q80	Q90
4/1/2017	1	9251.0	9255.507	9259.98	9264.45	9268.92	9273.40	9277.87	9282.34	9286.81
4/1/2017	2	8775.3	8779.747	8784.22	8788.69	8793.16	8797.64	8802.11	8806.58	8811.05
4/1/2017	3	8567.1	8571.55	8576.02	8580.49	8584.97	8589.44	8593.91	8598.38	8602.85
4/1/2017	4	8468.0	8472.451	8476.92	8481.39	8485.87	8490.34	8494.81	8499.28	8503.76
4/1/2017	5	8601.2	8605.72	8610.19	8614.66	8619.14	8623.61	8628.08	8632.55	8637.02
4/1/2017	6	9217.8	9222.297	9226.77	9231.24	9235.71	9240.19	9244.66	9249.13	9253.60
4/1/2017	7	10416.0	10420.52	10424.99	10429.47	10433.94	10438.41	10442.88	10447.35	10451.83
4/1/2017	8	11585.2	11589.7	11594.18	11598.65	11603.12	11607.59	11612.06	11616.54	11621.01
4/1/2017	9	12627.8	12632.31	12636.78	12641.25	12645.72	12650.20	12654.67	12659.14	12663.61
4/1/2017	10	13310.8	13315.23	13319.70	13324.18	13328.65	13333.12	13337.59	13342.06	13346.54
4/1/2017	11	13736.0	13740.46	13744.93	13749.40	13753.88	13758.35	13762.82	13767.29	13771.76
4/1/2017	12	13887.1	13891.56	13896.04	13900.51	13904.98	13909.45	13913.92	13918.40	13922.87
4/1/2017	13	13931.4	13935.84	13940.32	13944.79	13949.26	13953.73	13958.21	13962.68	13967.15
4/1/2017	14	13887.9	13892.36	13896.83	13901.30	13905.77	13910.24	13914.72	13919.19	13923.66
4/1/2017	15	13837.7	13842.13	13846.61	13851.08	13855.55	13860.02	13864.49	13868.97	13873.44
4/1/2017	16	13860.8	13865.24	13869.71	13874.18	13878.66	13883.13	13887.60	13892.07	13896.55
4/1/2017	17	14176.7	14181.2	14185.68	14190.15	14194.62	14199.09	14203.56	14208.04	14212.51
4/1/2017	18	14474.3	14478.75	14483.22	14487.70	14492.17	14496.64	14501.11	14505.58	14510.06
4/1/2017	19	14280.3	14284.81	14289.28	14293.75	14298.22	14302.70	14307.17	14311.64	14316.11
4/1/2017	20	14442.7	14447.18	14451.66	14456.13	14460.60	14465.07	14469.54	14474.02	14478.49
4/1/2017	21	14122.0	14126.45	14130.93	14135.40	14139.87	14144.34	14148.82	14153.29	14157.76
4/1/2017	22	13155.3	13159.8	13164.28	13168.75	13173.22	13177.69	13182.17	13186.64	13191.11
4/1/2017	23	11917.8	11922.29	11926.76	11931.23	11935.70	11940.17	11944.65	11949.12	11953.59
4/1/2017	24	10746.4	10750.84	10755.32	10759.79	10764.26	10768.73	10773.21	10777.68	10782.15
4/2/2017	1	9902.8	9907.241	9911.71	9916.19	9920.66	9925.13	9929.60	9934.07	9938.55
4/2/2017	2	4968.8	4973.263	4977.74	4982.21	4986.68	4991.15	4995.62	5000.10	5004.57
4/2/2017	3	9237.1	9241.538	9246.01	9250.48	9254.95	9259.43	9263.90	9268.37	9272.84
4/2/2017	4	9052.5	9056.979	9061.45	9065.92	9070.40	9074.87	9079.34	9083.81	9088.28
4/2/2017	5	9126.3	9130.808	9135.28	9139.75	9144.22	9148.70	9153.17	9157.64	9162.11
4/2/2017	6	9640.7	9645.126	9649.60	9654.07	9658.54	9663.01	9667.49	9671.96	9676.43
4/2/2017	7	10620.7	10625.2	10629.68	10634.15	10638.62	10643.09	10647.57	10652.04	10656.51
4/2/2017	8	11445.5	11449.93	11454.41	11458.88	11463.35	11467.82	11472.29	11476.77	11481.24

... NewEng | ME | NH | VT | CT | RI | SEMASS | WCMASS | NEMASSBOST

Figure 6: Percentiles data

4 Conclusion

This paper presented a journey through the Energy Demand Forecasting process. The project allowed us to learn and familiarize with forecasting methods. It also helped us gain physical insight into the ARIMA function, so we can relate published information to the algorithm developed.

Three different updates were conducted to the algorithm developed. The results for each phase were presented. The seasonality study based on day of the week to demand relationship showed that there was not real evidence about the existence of that relation. As seen, we could get a lower demand on a weekend day than any other day of the week. Our recommendation to this study would be to study and implement a method based on temperature (weather) - energy demand relationship.

5 Appendix:

5.1 MATLAB SCRIPT

```
clear all; clc;
% check if month has 30 days i,j:720, if 31 days :744. Change also ranges
% Excel Ranges (Modify accordingly, unless month data show first)
xlRangesp = 'D746:D1465';
xlRangel = 'D2186:D2905';
xlRanges = 'D2162:D2881';
% Reading cycle
for j=1:9
sheet = j+1;
% 2003 Data
filename03 = '2003_smd_hourly.xls';
data(j,:,1) = xlsread(filename03,sheet,xlRangesp);
% 2004 Data
filename04 = '2004_smd_hourly.xls';
data(j,:,2) = xlsread(filename04,sheet,xlRangel);
% 2005 Data
filename05 = '2005_smd_hourly.xls';
data(j,:,3) = xlsread(filename05,sheet,xlRanges);
% 2006 Data
filename06 = '2006_smd_hourly.xls';
data(j,:,4) = xlsread(filename06,sheet,xlRanges);
% 2007 Data
filename07 = '2007_smd_hourly.xls';
data(j,:,5) = xlsread(filename07,sheet,xlRanges);
% 2008 Data
filename08 = '2008_smd_hourly.xls';
data(j,:,6) = xlsread(filename08,sheet,xlRangel);
% 2009 Data
filename09 = '2009_smd_hourly.xls';
data(j,:,7) = xlsread(filename09,sheet,xlRanges);
% 2010 Data
filename10 = '2010_smd_hourly.xls';
data(j,:,8) = xlsread(filename10,sheet,xlRanges);
```

```
% 2011 Data
filename11 = '2011_smd_hourly.xls';
data(j,:,9) = xlsread(filename11,sheet,xlRanges);
% 2012 Data
filename12 = '2012_smd_hourly.xls';
data(j,:,10) = xlsread(filename12,sheet,xlRangel);
% 2013 Data
filename13 = '2013_smd_hourly.xls';
data(j,:,11) = xlsread(filename13,sheet,xlRanges);
% 2014 Data
filename14 = '2014_smd_hourly.xls';
data(j,:,12) = xlsread(filename14,sheet,xlRanges);
% 2015 Data
filename15 = '2015_smd_hourly.xls';
data(j,:,13) = xlsread(filename15,sheet,xlRanges);
% 2016 Data
filename16 = '2016_smd_hourly.xls';
data(j,:,14) = xlsread(filename16,sheet,xlRangel);
end
% Creating date and hour data
cont=0;
for t=1:30
  f=1;
    while f <= 24
        cont=cont+1;
   date(cont,:) = {' 4/' num2str(t) '/2017'};
   str(cont) = strcat(date(cont,1),date(cont,2),date(cont,3));
   hour(cont) = f;
      f=f+1;
    end
end
% ************************Forecasting*********************************
%ToEstMdl = arima('MA',{0.9,0.7},'SMA',0.52,'SMALags',11,'Constant'...
%,0.04,'Variance',0.2);
for k=1:9
 sheetf=k+1;
 % Seasonality based on data validation using ARIMA
for i=1:577
```

```
    ToEstMdl = arima('MA',{0.94,0.72},'SMA',0.59,'SMALags',11,...
        'Constant',0.04,'Variance',0.2);
 tempp(1:14) = data(k,i,:);
EstMdl = estimate(ToEstMdl,tempp(:));
[YF,YMSE(i)] = forecast(EstMdl,1,'Y0',tempp(:));
%
perc(i,5) = YF;
end
for i=578:720
    ToEstMdl = arima('MA',{0.94,0.76},'SMA',0.74,'SMALags',11,...
        'Constant',0.07,'Variance',0.1);
 tempp(1:14) = data(k,i,:);
EstMdl = estimate(ToEstMdl,tempp(:));
[YF,YMSE(i)] = forecast(EstMdl,1,'Y0',tempp(:));
%
perc(i,5) = YF;
end
%************filling percentiles****************
temp=0;
for i=1:9
    temp=temp+10;
  if i<=4
    for j=1:720
perc(j,i)=perc(j,5)-(50-temp)*sqrt(YMSE(j));
    end
  end
    if i>=6
    for j=1:720
perc(j,i)=perc(j,5)+(temp-50)*sqrt(YMSE(j));
    end
  end
end
% Cleaning content
filenamew = 'C:\Users\adria\OneDrive\Documents\R\Trackcopy.xls';
sheetname(1:9)={'NewEng' 'ME' 'NH' 'VT' 'CT' 'RI' 'SEMASS' 'WCMASS'...
    'NEMASSBOST'};
Excel = actxserver('Excel.Application');
Workbook = Excel.Workbooks.Open(filenamew);
```

```
sheettemp=char(sheetname(k));
Workbook.Worksheets.Item(sheettemp).Range('A2:K745').ClearContents
% Closing book
Workbook.Save;
Excel.Workbook.Close;
Excel.Quit;
delete(Excel)
% Writing data
xfRange = 'C2:K721';
xlswrite(filenamew,perc(:,:),sheetf,xfRange)
dateRange = 'A2:A721';
xlswrite(filenamew,str(:),sheetf,dateRange)
hourRange = 'B2:B721';
xlswrite(filenamew,hour(:),sheetf,hourRange)
%}
end
%% ***********graph for analysis************************************
%
% 579-up ToEstMdl = arima('MA',{0.94,0.76},'SMA',0.74,'SMALags',11,...
%'Constant',0.04,'Variance',0.2);
ToEstMdl = arima('MA',{0.94,0.72},'SMA',0.59,'SMALags',11,...
    'Constant',0.04,'Variance',0.2);
figure
%for i=625:648
i=431;
k=2;
 ppll(1:14)=data(k,i,:);
EstMdl = estimate(ToEstMdl,ppll(:));
[YF,YMSE] = forecast(EstMdl,1,'Y0',ppll(:));
plot(2003:1:2016,ppll(:));
hold on
plot(2017,YF,'o','LineWidth',2);grid;
%end;
%}
```

References

[1] H.R. Sadeghi Keyno, F. Ghaderi, A. Azade, J. Razmi *Forecasting electricity consumption by clustering data in order to decrease the periodic variable's effects and by simplifying the pattern*, Energy Conversion and Management, 50 (2009), pp. 829–836

[2] H.T. Nguyen, L.T. Nabney *Short-term electricity demand and gas price forecasts using wavelet transforms and adaptive models*, Energy, 35 (2010), pp. 3674–3685.

[3] B. Wang *An imbalanced development of coal and electricity industries in China*, Energy Policy, 35 (2007), pp. 4959–4968

[4] Wikipedia *Northeast Blackout of 2003*, http://en.wikipedia.org/wiki/Northeast Blackout of 2003.

[5] M.J. Xu, Z.G. Hu, B.G. Shan, J.Y. Wu *Analysis on influencing factors of electricity demand by using interpretative structural modeling*, Electric Power, 42 (2009), pp. 1–4

[6] A. Abraham, B. Nath *A neuro-fuzzy approach for modeling electricity demand in Victoria*, Applied Soft Computing, 1 (2001), pp. 127–138

[7] J.W. Taylor *Triple seasonal methods for short-term electricity demand forecasting*, European Journal of Operational Research, 204 (2010), pp. 139–152

[8] L.W. Baxter, K. Calandri *Global warming and electricity demand: a study of California*, Energy Policy, 20 (1992), pp. 233–244

[9] E. Erdogdu *Electricity demand analysis using cointegration and ARIMA modeling: a case study of Turkey*, Energy Policy, 35 (2007), pp. 1129–1146

[10] F.M. Tseng, H.C. Yu, G.H. Tzeng *Combining neural network model with seasonal time series ARIMA model*, Technological Forecasting and Social Change, 69 (2002)

[11] S Makridakis, M Hibon *The M3-competition: results, conclusions and implications*, International Journal of Forecasting, 16 (2000), pp. 451–476

[12] S.L. Ho, M. Xie, T.N. Goh *A comparative study of neural network and Box-Jenkins ARIMA modeling in time series predictions*, Computers and Industrial Engineering, 42 (2002), pp. 371–375