Forecasting Daily Power Consumption for Households

Deepak Mahapatra Kunal Punjabi Harshal Zalke
(Graduate Students, University of Texas at Austin)

Advised by
Dr. Dragan Djurdanovic (Professor, University of Texas at Austin)

May 15, 2017

Abstract

Due to the increase in the available power consumption data and the changing pattern of electricity consumption, distribution companies are faced with the challenging task of forecasting the electricity consumption. We use the data from UCI machine learning repository to build a forecasting model. The Power consumption problem involves building an ARMA model, first a stationary model with trends and seasonalities and then a non stationary model with trends and seasonalities. We then predict for all the models. We conclude by selecting the best ARMA model among these also obtain 30 day ahead predictions.

1 Introduction

In recent years, addition of Advanced Metering Infrastructure System to the US energy market has led to a tremendous increase in the available power consumption data [1]. This has opened up new challenges and opportunities for the utility companies to enhance their day to day operations using advanced forecasting techniques [2]. Also with the advent of renewable energy resources and distributed generation technologies, such as roof top solar panels, the consumption pattern of the customers have changed drastically. This has further increased the importance of accurate day to day forecasts, to meet daily demand. Furthermore, the distribution companies maximize their profits by bidding on the forecasted "day ahead" demand in the energy market. In addition to that, the power generation systems are built to meet the peak demand, thus the distribution companies study "long term" demand for the purpose of capacity planning. Thus the entire energy market is dependent on accurate forecasting. For this purpose, we look at the electricity consumption data, provided by UCI Machine Learning Repository. The dataset has information for 370 houses (in KW) from January 2011 to December 2014. Based on this, we build an ARMA model that will help gain a better understanding of future consumption pattern and help in optimal planning for both "day ahead" and "long term" demands.

2 Data Compilation

The data set used for this project was taken from UCI machine learning data repository [3]. This data gives the household energy consumption for 370 individual houses in Portuguese. The energy consumption values are given in kW for each 15 min.The data-set used the date time format for Portuguese hour time and each day has 96 measures. Every year in March, on the day-light savings shifting day (which has only 23 hours) the values between 1:00 am and 2:00 am are zero for all points. Similarly, every year in October, on the day-light saving shifting day (which has 25 hours) the values between 1:00 am and 2:00 am are aggregate consumption of two hours. In order to ease the computational complexity of our models we have taken the sum of the energy consumption of the all 370 houses given in the data-set. Then for further using these values in our analysis we converted the frequency of observation of the consumption by taking the hourly total and the daily total values. On observing the values of the individual 370 clients we found that some of the clients joined after the year 2011, so these clients were having zero consumption in that year. So in order to get a coherent data-set we removed the individual houses which were not having any data in the year 2011. After removing the individual units and finding the total for all the units we plotted the raw data to see if there is any particular deterministic trend or seasonality in the data. As it can be seen in the Fig. 1 the daily production level almost stayed in the same range with varying values depending on the days of the year, with possible fluctuations in demand over some period of the years.

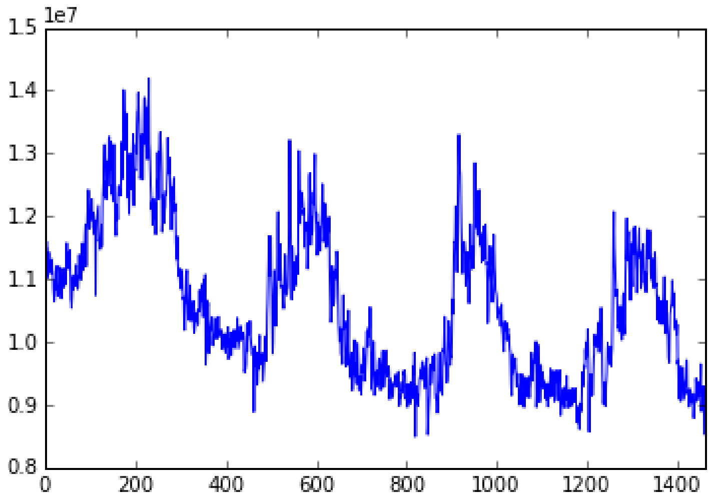


Figure 1: Daily energy consumption of all the Households

To see if we can model the data on hourly basis we tried to visualize the total hourly consumption of the filtered individual households. As given in the Fig. 2, we can see that the same seasonality which were seen in the total daily consumptions were present in the hourly data with some seasonality also in the hourly levels.

If we zoom in on the total daily energy consumption, as seen in the Fig. 3 we can see that there is a seasonality of seven days among the data. So in order to analyze that seasonality we will utilize our methodology for applying ARMA models to find the significance of this seasonality period.

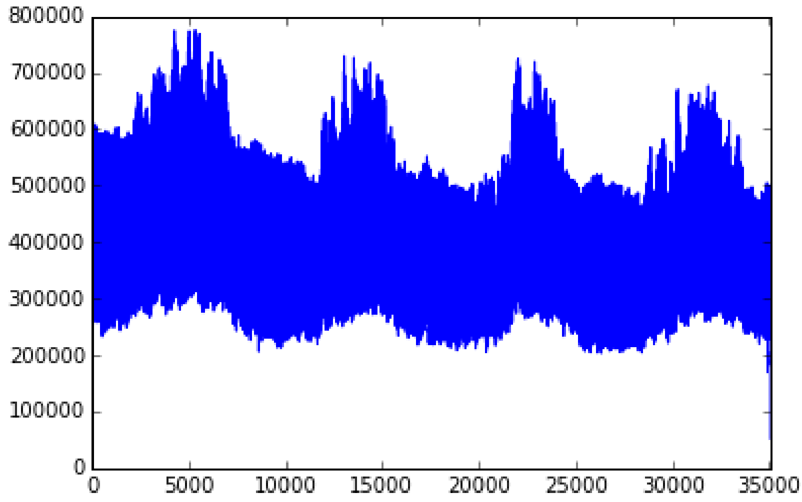


Figure 2: Hourly energy consumption of all the Households

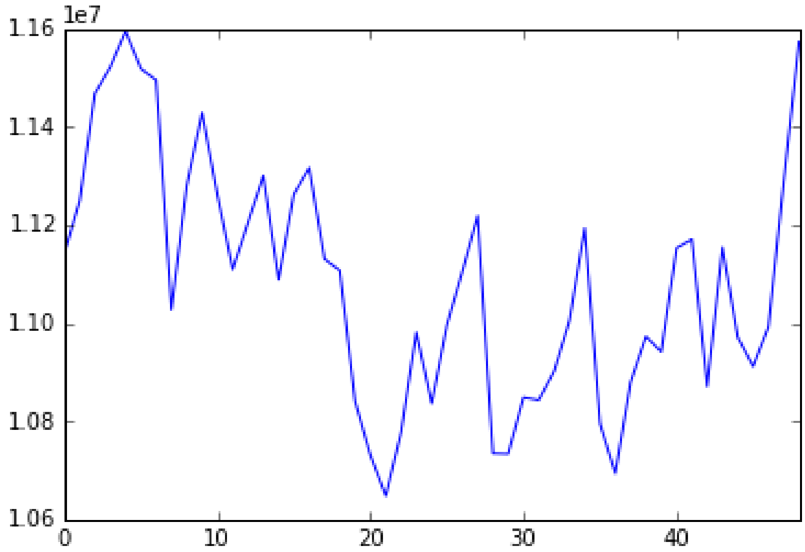


Figure 3: Daily energy consumption of all the Households for the first 50 days

3 Methodology

3.1 Stationary model without trend and seasonality

In order to get the stationary model, the mean of our time series was subtracted from each data point in the series. Then we found the optimal ARMA model by implementing "ARMAX" function in MATLAB. The optimal model that resulted in our case is ARMA(10,9). The residual sum of squares(RSS) obtained from this model is $7.2444e + 13$. The F-value for each $(2n, 2n - 1)$ model were checked with an iterative algorithm for finding a significant decrease in RSS in each step. The Fig. 13a gives correlation among the residuals, which shows that the ACF among the residuals of final ARMA(10,9) model is within the bounds, and thus we can conclude that the residuals are random white noise.

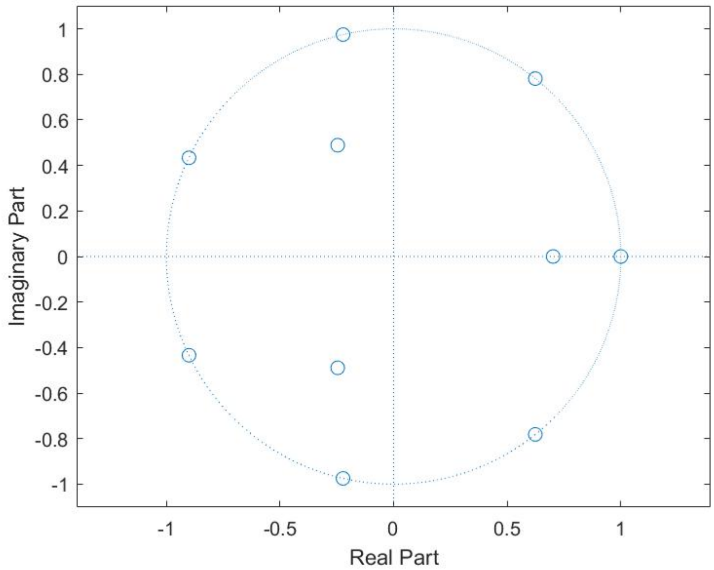


Figure 4: AR roots of the ARMA(10,9) model not considering any seasonality and trend

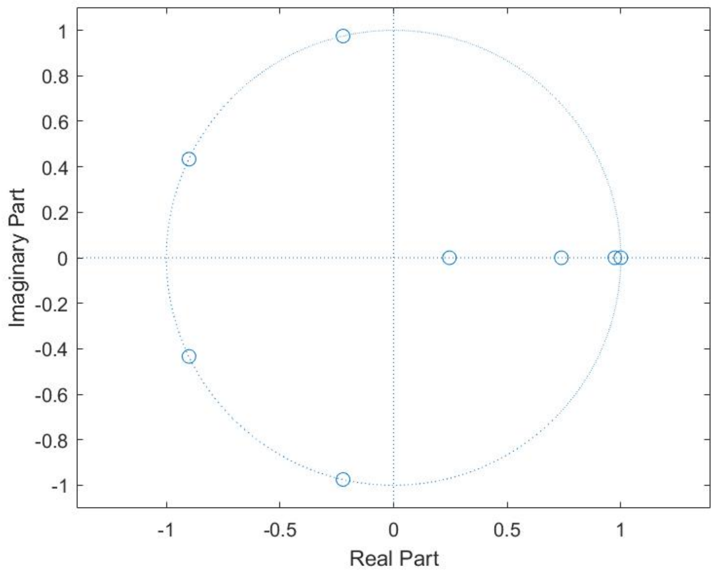


Figure 5: AR roots of the parsimonious ARMA(8,9) model using a period of 7 days and no trend

3.2 Stationary model with trend

In order to check for stochastic trend, we checked the real roots of AR part of our model. For this we plotted the roots of the AR part of ARMA(10,9) model, as shown in Fig. 4. We can see that one of the real roots is on unit circle. This implies that we have a potential stochastic trend in the data.

In order to check if there is a trend, we restrict one parameter of the AR part of ARMA(10,9) model to get a parsimonious model. We compute the RSS for this parsimonious ARMA(9,9) model and perform the F-test with respect to the original model. The parsimonious model is $(1-\beta)\cdot X_t$ and its RSS is $7.3834e+13$ and the computed F-score is 27.6364. This F value is less than F critical value ($F_{2,\infty} = 2.99$). So, we conclude that there is no trend in our data.

3.3 Stationary model with seasonality

In order to check for seasonality in our model, we analyze the roots of AR part. We look for complex conjugate pairs with magnitude close to 1 [4]. As shown in Fig. 4, there are 3 complex conjugate pairs on the unit circle. These 3 complex conjugate pairs have the potential to cause stochastic seasonality. Now, in order to find the pairs that cause seasonality, we restrict two parameters of the AR part of the model and perform the F-test.The restricted model in case of each pair of roots is given by $(1-2\cdot\cos\omega\beta+\beta^2)\cdot X_t$. The real part of each of the restricted roots were used to compute the ω value which gives us the time period of the possible seasonality in the data. The RSS, F-value and time period for the three possible parsimonious models, ARMA (8,9) are give in Table 1. As shown, the F-value obtained for the third pair, i.e., "Time Period = 7.001", is less than F-critical ($F_{2,\infty} = 2.99$). Now, we check correlation among residuals. As shown in Fig. 13b , the ACF is within bounds and thus we conclude that the residuals are random white noise. Thus we conclude that our data has a seasonality of 7 days. This is in accordance with the assumption we made with reference to the patterns in the raw data.

Table 1: RSS and F values for parsimonious models when checking seasonality

Roots	$-0.9002 \pm 0.4337i$	$-0.2225 \pm 0.9748i$	$0.6235 \pm 0.7816i$
RSS	$1.0e + 13 * 7.6044$	$1.0e + 13 * 7.2473$	$1.0e + 13 * 7.3494$
F_values	25.7700	-9.6958	0.4476
Time Period	2.33	3.5	7.001

3.4 Non-stationary Model

In order to build a non-stationary time series, first we need to fit a polynomial or exponential or harmonic function for the deterministic part of the series. In order to check for the harmonic function, we used the optimizer available in scipy library in python to fit a sine-cosine curve to our data. We also fit a polynomial function along with the harmonic function using the python optimizer. We calculate the RSS between actual data and the fit data for each of these possible deterministic trends. These RSS values are shown in the Table 2. As seen, the fit function with a first order polynomial and two sine-cosines of period 7 and 360 give the least RSS. Also, it can be seen in Fig. 7, that the fit function of our optimizer gives a good result. So we take this deterministic model and subtract it from the original time series to find a remainder series. Now we use ARMAX in MATLAB, to model our data, as explained in section 3.1. As shown in Fig. 13c. the ACF is within bounds and thus we conclude that the residuals are random white noise.

The best model obtained from the step wise procedure is ARMA (8,7) with a RSS value of $7.1653e + 13$. The AR roots of the models are plotted on unit circle, as shown in Fig. 6

3.5 Non-stationary model with stochastic trend

In order to check for stochastic trend, we check the real roots of AR part of our model. For this, we plot the roots of the AR part of ARMA(8,7) model, as shown in Fig. 6. We can see that one of the real roots is on unit circle. This implies that we have a potential trend in the data. Similar to section 3.2, we build the parsimonious ARMA(7,7) model and perform the F-test. We compute the RSS for this parsimonious ARMA(7,7) model and perform the F-test

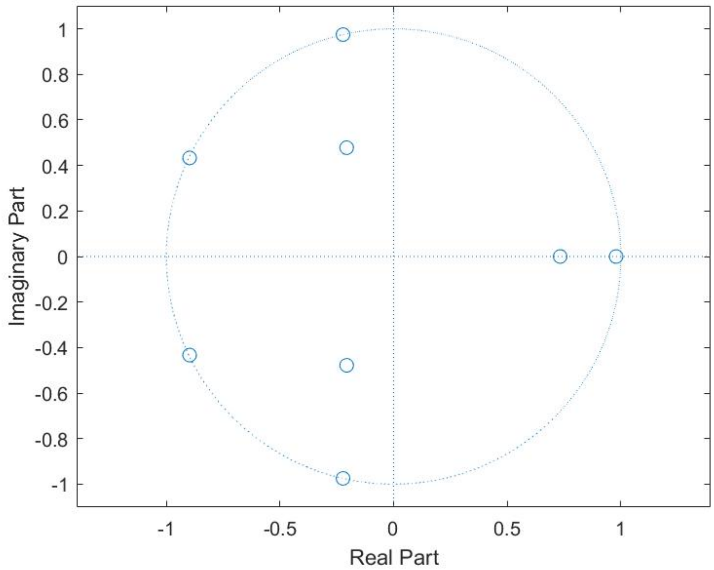


Figure 6: AR roots of the ARMA(8,7) model when deterministic trend is fit

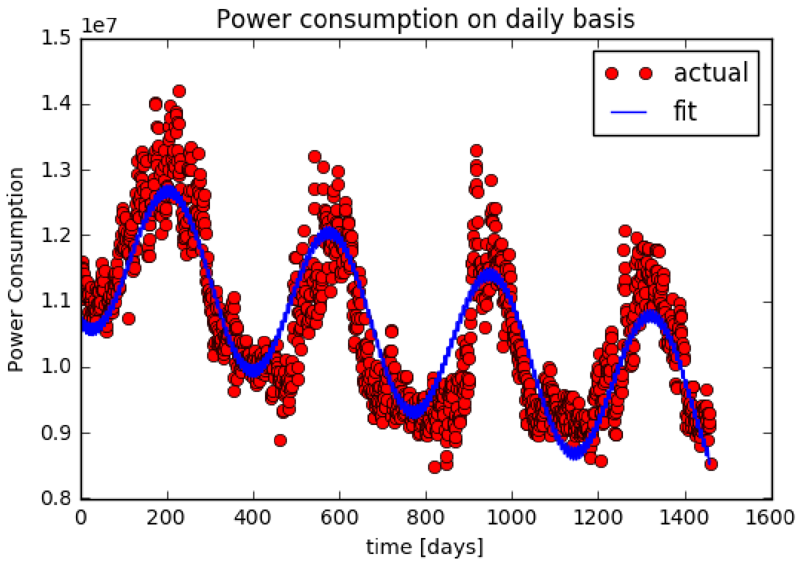


Figure 7: Fitting a deterministic trend to an original series

Table 2: RSS of different fit functions for the non-stationary model

Function Type	RSS	Function parameters
Harmonic with Polynomial order 2	583287242910196.44	$x, x^2$and sin 360, 7
Harmonic with Linear trend	**493550211962205.62**	x and sin 360, 7
Linear	1533604112750618.2	x
Polynomial order 2	1436502738677743.0	x and x^2
Ploynomial order 3	1426308917505729.7	$x, x^2 and x^3$

with respect to the original model. The F-score is 10.754. This F-value is less than F critical value ($F_{2,\infty} = 2.99$). So, we conclude that there is no trend in our data.

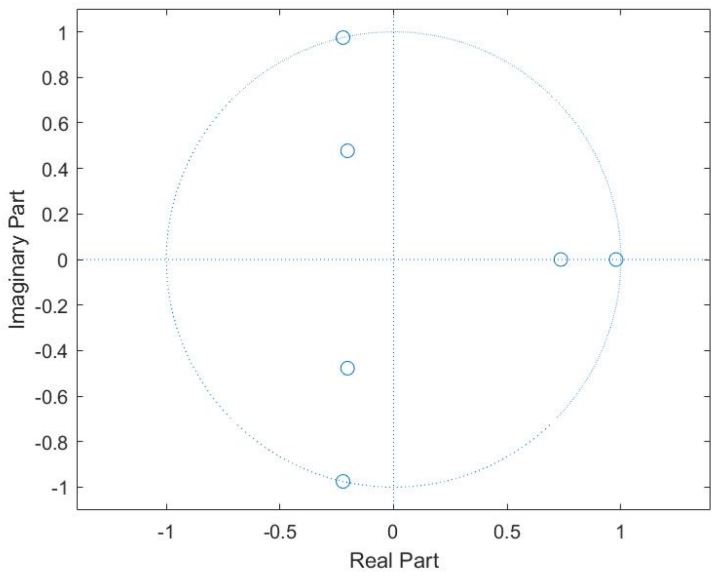


Figure 8: Roots of parsimonious model when deterministic trend is fit

3.6 Non-stationary model with stochastic seasonality

In order to check for seasonality in our model, similar to section 3.3, we analyze the complex conjugate pairs with magnitude close to 1. As shown in Fig. 6, there are 2 complex conjugate pairs on the unit circle, of which only the first pair gives F-value less than the F critical value. As shown in the Table 3 , only the first roots' F-value gives an adequate parsimonious (6,7). Now, we check correlation among residuals. As shown in Fig. 13d, the ACF is within bounds and thus we conclude that the residuals are random white noise. Thus we conclude that our data has a seasonality of 2.34 days.

Table 3: RSS and F values for parsimonious models when checking seasonality for the non stationary models

Roots	-0.8976 ± 0.4331	$-0.2226 \pm 0.9747i$
RSS	$1.0e + 13 * 7.1719$	$1.0e + 13 * 7.5658$
F_values	0.6659	40.3309
Time Period	2.34	3.49

4 Results

The predictions obtained from the ARMA(10,9) stationary model are shown in Fig. 9. As can be seen, they are within the 95% confidence interval. The RSS for the model is $7.2444e + 13$.

The predictions obtained from the ARMA(8,9) parsimonious stationary model are as shown in Fig. 10. As can be seen, they are within the 95% confidence interval. The RSS for the model is $7.3494e + 13$.

The predictions obtained from the ARMA(8,7) non stationary model are shown in Fig. 11. As can be seen, they are within the 95% confidence interval. The RSS for the model is $7.1653e + 13$.

The predictions obtained from the parsimonious ARMA(6,7) parsimonious non stationary models are as shown in Fig. 12. As can be seen, they are within the 95% confidence interval. The RSS for the model is $7.1719e+13$. Thus this is the best model.

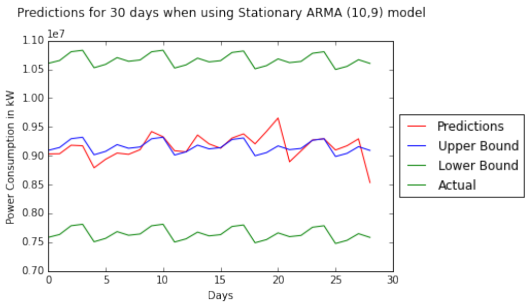


Figure 9: Prediction of ARMA(10,9)

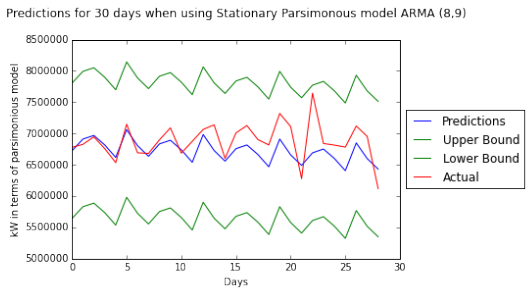


Figure 10: Prediction of ARMA(8,9)

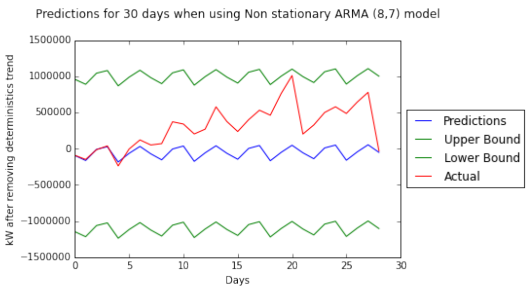


Figure 11: Prediction of ARMA(8,7)

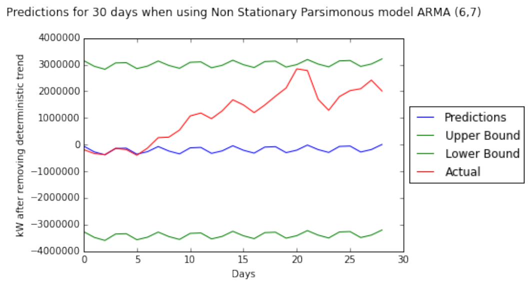
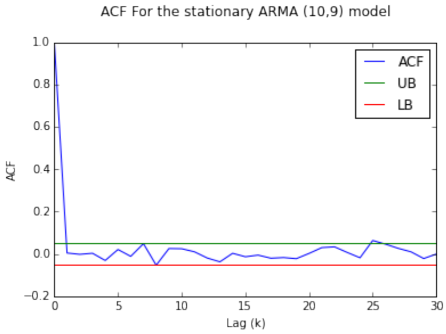
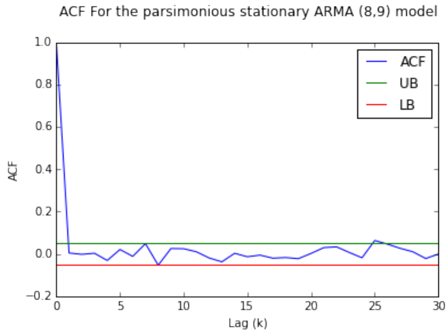


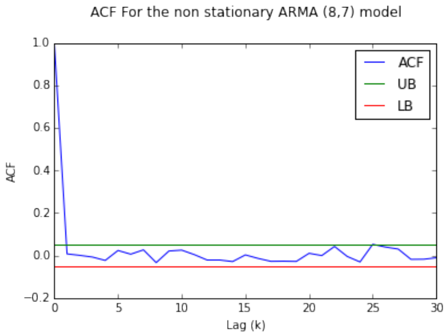
Figure 12: Prediction of ARMA(6,7)



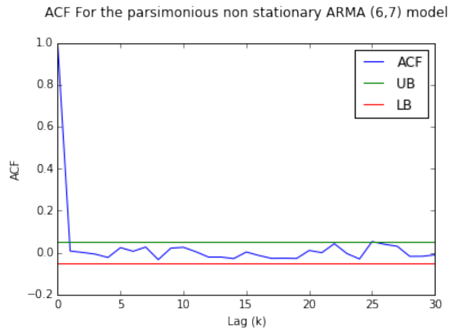
(a) ACF ARMA(10,9)



(b) ACF parsimonious ARMA(8,9)



(c) ACF Deterministic ARMA(8,7)



(d) ACF Deterministic parsimonious ARMA(6,7)

Figure 13: Illustration of various images

5 Conclusion

From the analysis performed in this project we conclude that the best predictions were obtained by a parsimonious model of ARMA (8,9) order with a seasonality of period 7 days when we use the data as a stationary series. However when we remove the deterministic trend from the series with a harmonic function and a first order term, we see that the 7 day seasonality has been removed from our model. The parsimonious model of ARMA(6,7) obtained in this deterministic trend case performs best as compared to the other ARMA models. The 30 step ahead predictions using the above models were found to be within 95% confidence interval.

References

[1] N. M. G. Strategy, "Advanced metering infrastructure," *US Department of Energy Office of Electricity and Energy Reliability*, 2008.

[2] T. Hong and D. A. Dickey, "Electric load forecasting: fundamentals and best practices," 2012.

[3] M. Lichman, "UCI machine learning repository," 2013. [Online]. Available: http://archive.ics.uci.edu/ml

[4] S. M. Pandit and S.-M. Wu, *Time series and system analysis, with applications.* John Wiley & Sons, 1983.