

Modelling of hourly energy consumption in the state of Virginia using ARMA models.

Written by: Harshita Vemula,

University of Texas at Austin, Time series and system analysis,

Spring 2019.

1.1 INTRODUCTION

Energy lies at the heart of the economy. Energy demand or Energy consumption forecasting is crucial for policy makers to formulate efficient electricity supply policies. Energy market being a free market has a lot of manipulators who cause fluctuations in the price which in turn affects the demand. The presence of complex linear and non linear patterns makes this predictive modelling problem challenging. Demand forecasting can be broadly divided into three main categories: short-term forecasts which are usually from one hour to one week, medium forecasts which are usually from a week to a year, and long-term forecasts which are longer than a year. In this report an attempt was made to forecast energy demands at an hourly level.

2.1 CASE STUDY

Dominion Power also known as Virginia Electric and Power company (VEPCO) is the major supplier of electricity in the state of Virginia. Dominion Energy serves about 2.5 million homes and businesses or about 2/3rd of the electric customers in the state. Using the data of electricity supplied by Dominion energy in the December of 2005, which is available on kaggle repository, an attempt was made to build a model to forecast hourly energy demand.

3.1 METHODS

The objective of this report is to understand the dynamics of the system of energy consumption in the state of Virginia. The report compares different ARMA models used for modelling and attempts to make accurate forecasts of the energy that would be consumed in the next few hours.

The ARMA models were fit using the F criterion and ARMA (2n,2n-1) modelling strategy was used to arrive at the right ARMA model. Preprocessing of the data, such as normalizing the data and removing of seasonal patterns, were also carried out in an attempt to achieve better results. The results of the four models that would be described below are of: a simple ARMA model, a parsimonious model, an ARMA model fitted to normalized data and an ARMA model fitted to deseasonalized data.

4.1 RESULTS

4.1.1 A Simple ARMA model- ARMA(32,31)

Used F-criterion and ARMA(2n,2n-1) modelling strategy to arrive at the right ARMA Model of order (32, 31). The RSS of the model is 8.5882e+06 and the standard deviation of the residuals is 130.6.

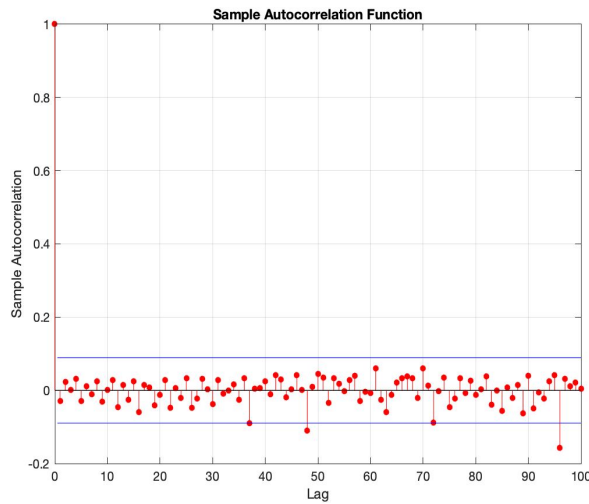


Fig.1.1

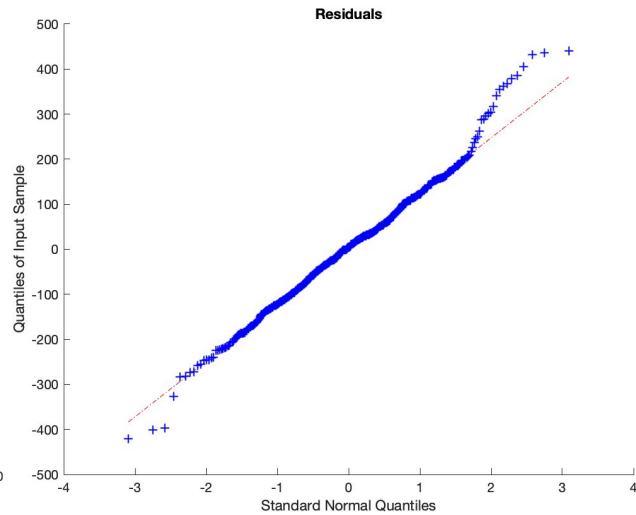


Fig.1.2

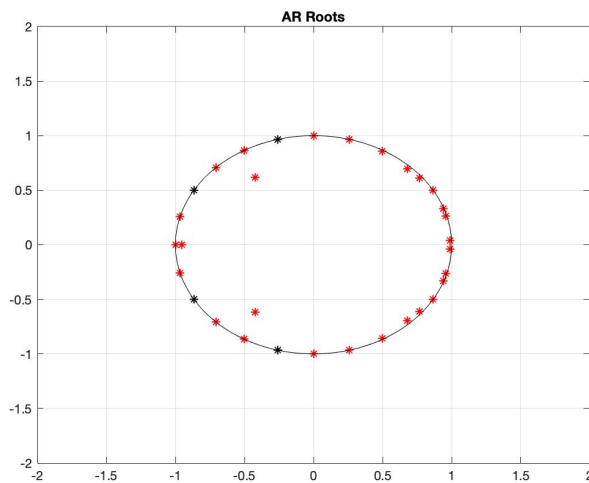


Fig.1.3

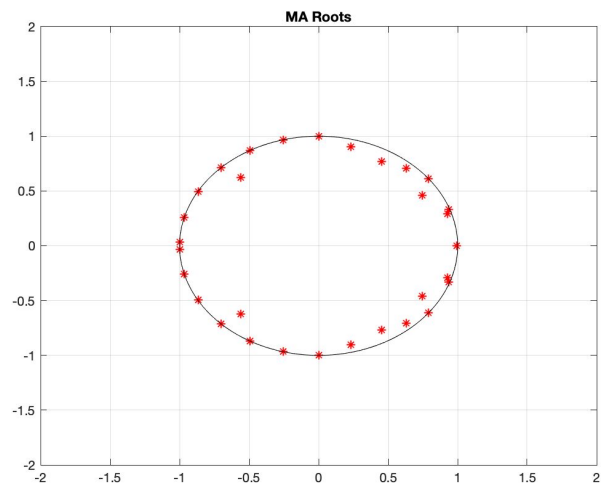


Fig.1.4

In the figure 1.3, the AR roots colored in black have an absolute value greater than 1 and have periodicities of 2.5 and 3.5. From the fig1.3 ,one can see the presence of a real root close to -1 and complex roots (with absolute value greater than one) with seasonalities of 2,3,4,5,6,12,24 and 168. We will later check if these trends and seasonality exist while building a parsimonious model.

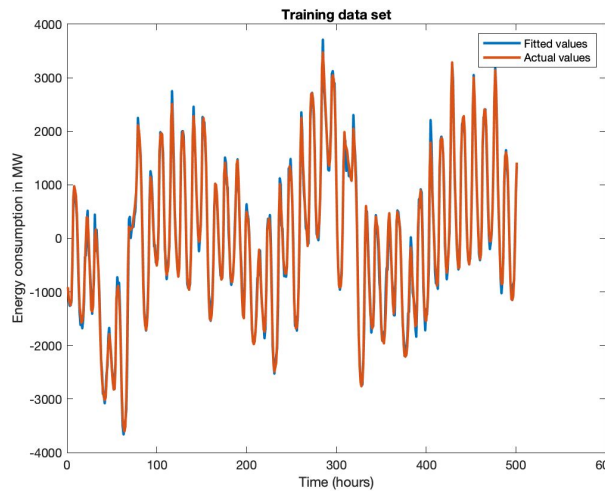


Fig.2.1

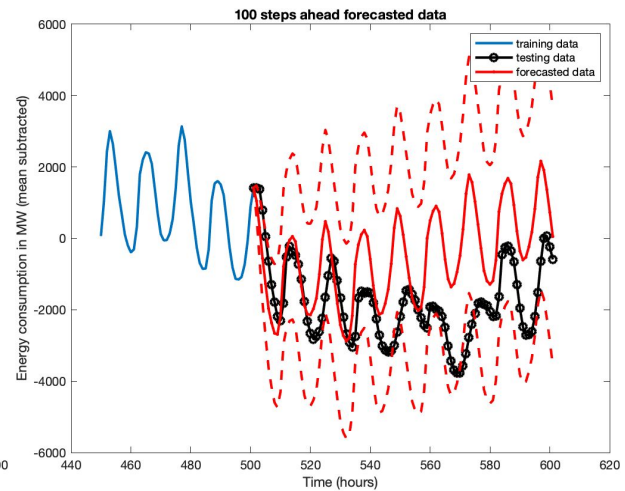
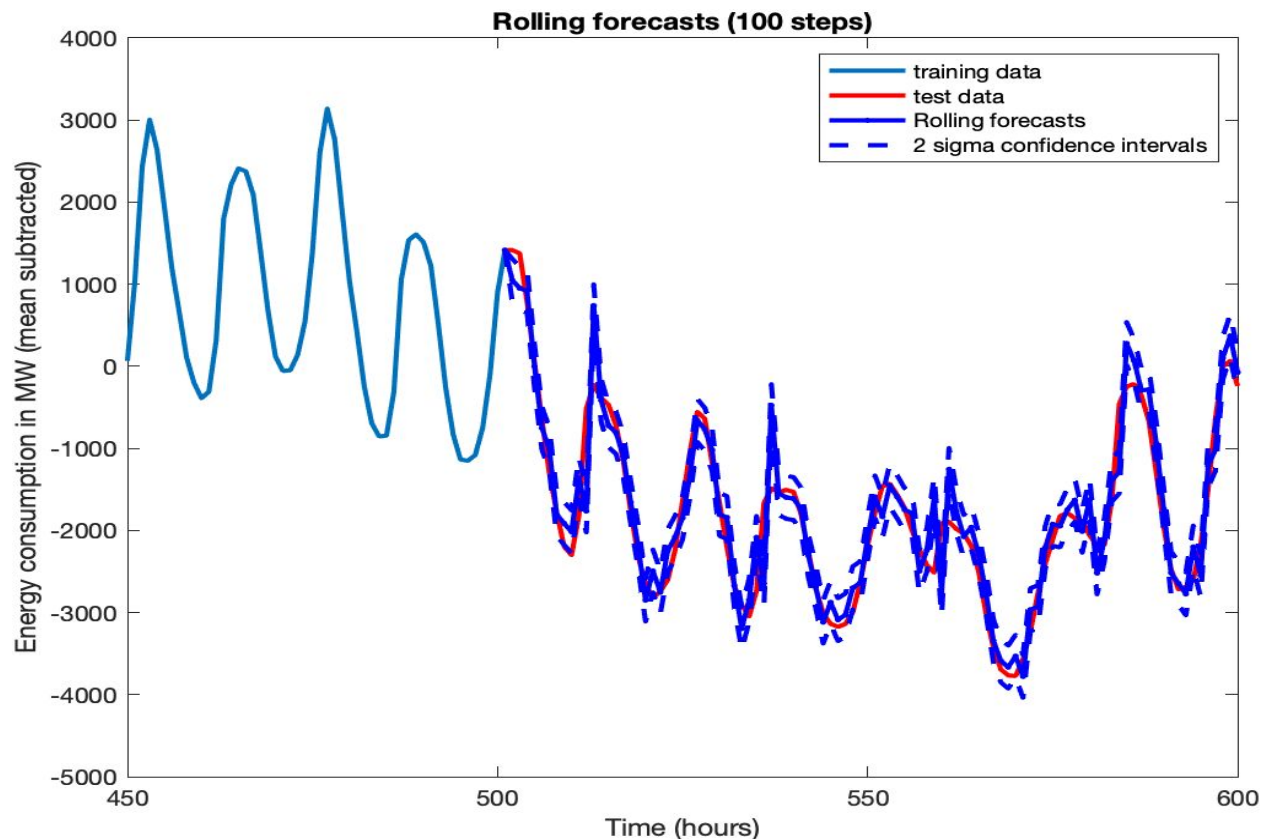


Fig.2.2

The figure 2.1 shows how the ARMA(32,31) fits the model and the figure 2.2 shows that the forecasts upto 20 steps ahead are accurately predicted using the model.

Rolling forecasts

The figure 3.1 shows the rolling forecasts of the energy consumption for 100 hours which seem to be accurate. The plot of the residuals of forecast will help detect anomalies if present and also give an indication when the model is changing. The distribution of the residuals of the training data and the residuals of rolling forecasts for 100, 200 and 300 hrs are plotted below.



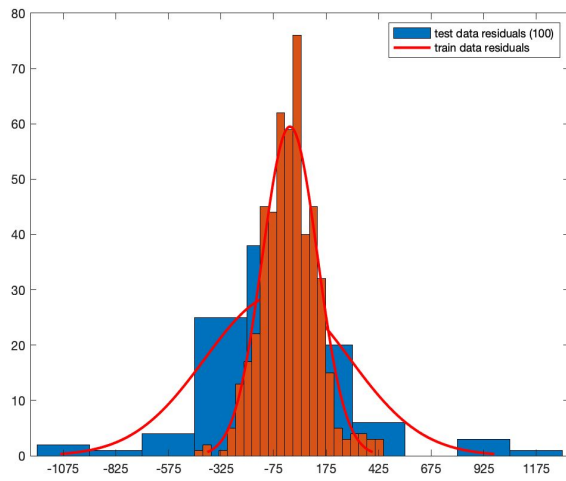


Fig.4.1

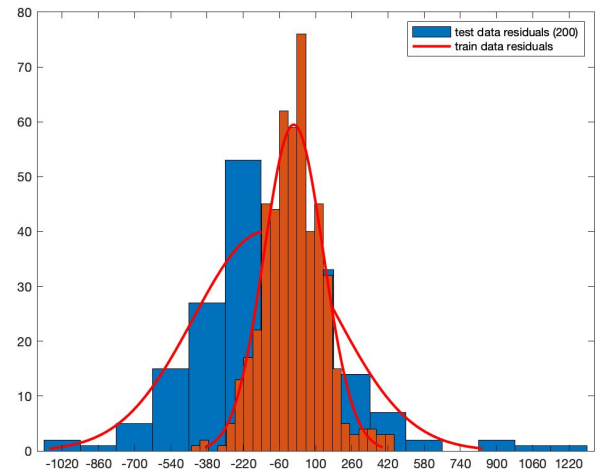


Fig.4.2

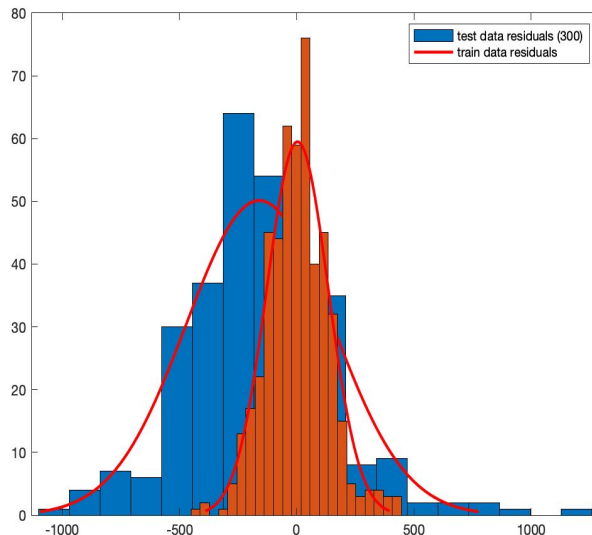


Fig.4.3

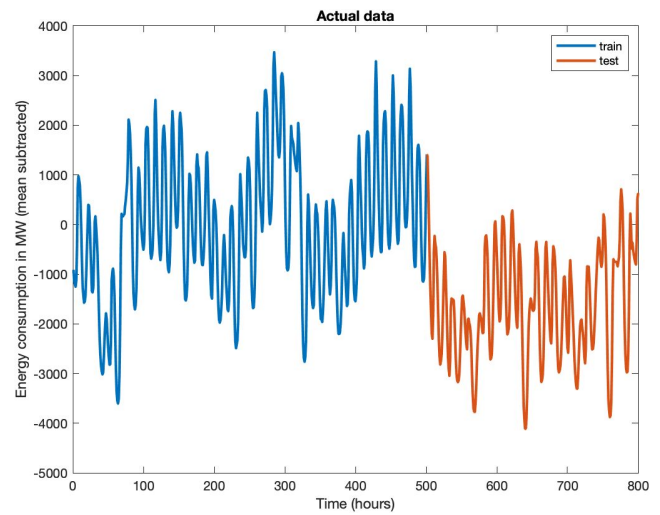


Fig.5.1

From the figures 4.1, 4.2 and 4.3, one can see that the distribution of the residuals of one step ahead forecast or rolling forecasts are drifting away (to the left) from the distribution of the residuals of the training data, indicating that the model is changing gradually. The drifting of the distribution to the left indicates that the test data has a mean different from that of the data used for training and that mean is less than that of the training data. Our hypothesis is further confirmed by the figure 5.1. This means that consumption of electric power tends to decrease towards the end of December in Virginia. This seems reasonable as it is the time of Christmas, commercial buildings and schools, which consume a lot of power, would be closed and people might go away for holidays.

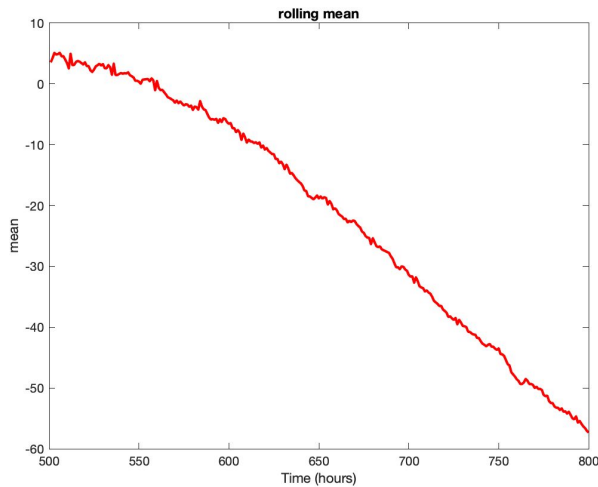


Fig.6.1

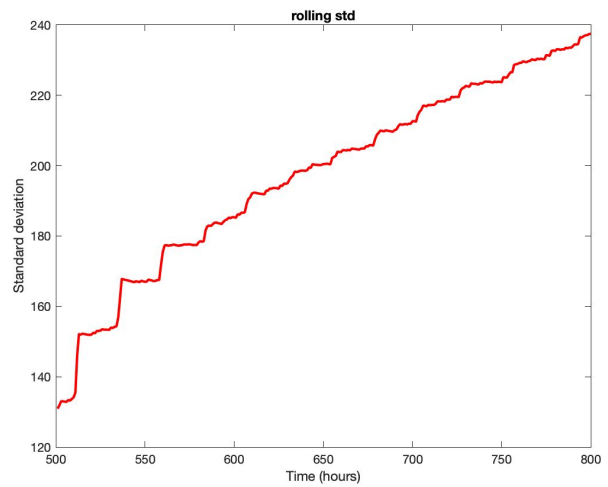


Fig.6.2

The figures 6.1 and 6.2 show the changing rolling mean and standard deviation of the residuals of the one step ahead forecasts. These plots clearly indicate how the mean and standard deviation of the residuals changed over time, hinting a gradual change in the model and that the model should be retrained. The model obtained upon retraining by including the first 100 data points from the test data set in the training data set is ARMA (30,29) and the model obtained upon retraining by including the first 300 data points from the test data set in the training data set is ARMA (27,23). This repeated change in models over time indicates that the data is non stationary in nature and the underlying system is unstable. The underlying system in our case is the energy market. It is highly volatile and has high risks associated with it, which explains the continuous change in models.

4.1.2 Parsimonious Model, ARMA(30,31)

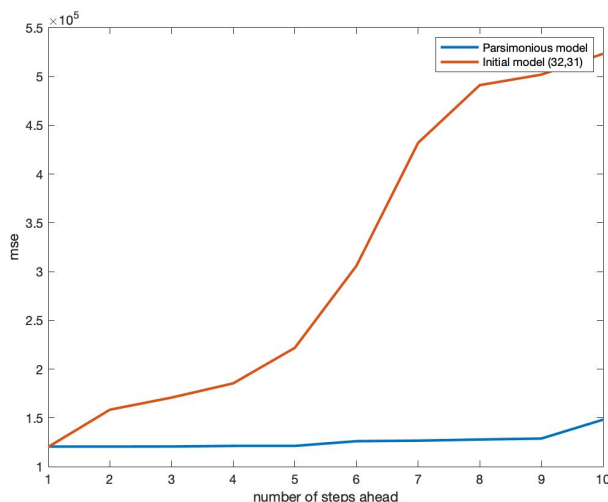
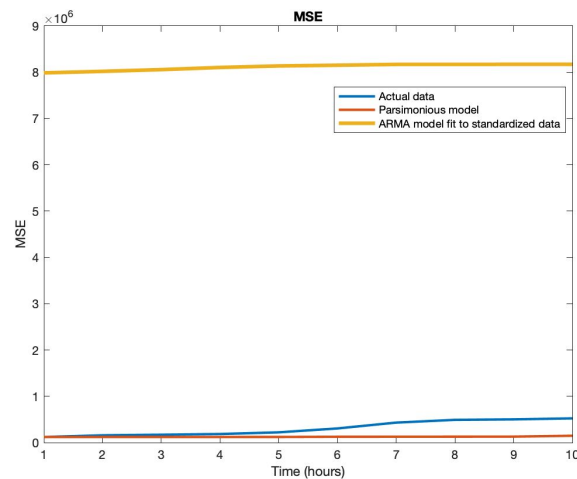


Fig.7.1

Complex roots with periods close to 2,3,4,5,6,12,24 and 168 lie close to the unit circle. Of these periodicities, the ones that are meaningful for the hourly data are 168, 24 and 12. The parsimonious models with periodicity of 168 and 12 were adequate. Of these two, minimum increase in residual sum of squares was achieved using the model with periodicity of 12. The parsimonious model with the two periodicities taken together was not adequate. Also checked if the real root lies on the unit circle. It turns that the parsimonious model

corresponding to this root is not adequate. The results obtained in figure 7.1 are quite interesting. Even though the rss of the actual model on the training data is less than that of the parsimonious model, the mean square errors of the forecasts by the parsimonious model are much better than that of the actual model, indicating that the actual model has overfit the data. The parameters, AR roots and ACF plots of this model can be seen in the Appendix.

4.1.3 ARMA Model fit to normalized data

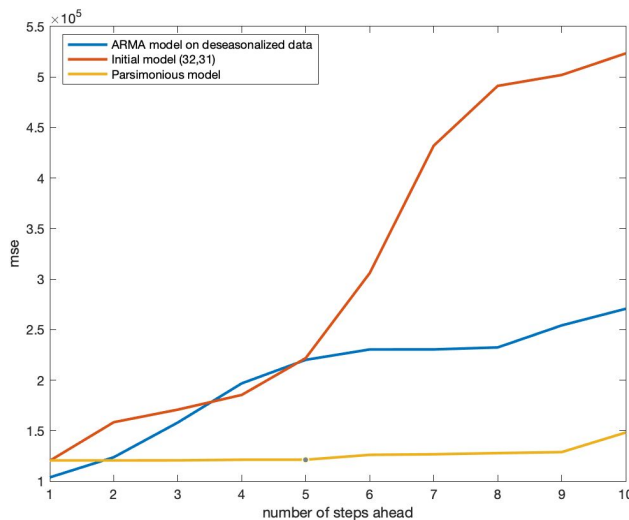


The data that we have has an hourly resolution. Assuming that the data points obtained on a given day of the week and at a given hour would be identically distributed, (for example: The data points obtained on any monday at 1pm would all come from same distribution.) the sample mean and sample standard deviations of each of these distributions are calculated and the data points are normalized using their corresponding distributions mean and variance. ARMA model of (3,2) is then fit to this normalized data using F criterion.

Fig 8.1

Fig 8.1 compares the mean square errors of this model with the mses of the previous parsimonious model and initial ARMA(32,31) model. Even though standardizing the data helped stationarize it, the figure 8.1 clearly indicates that standardizing did not help get better forecasts. The parameters, AR roots and ACF plots of this model can be seen in the Appendix.

4.1.4 ARMA model fit to deseasonalized data



Taking a look at the underlying system of energy consumption, one can hypothesize that deterministic seasonalities of periods 168 (As one would expect the energy consumption pattern to repeat every week, which corresponds to 168 hours) and 12 (As the parsimonious model indicates so) would be present. The drop in RSS achieved by fitting a sine wave with period 168 is significant (The rss dropped from $9.9945e+08$ to $6.661e+08$). Fitting 2 sine waves with periods of 168 and 12 reduced the rss further by

Fig.9.1

40%. Even though the f test indicated that this drop is not significant, the F value and the F cutoff value were close, and also as the drop of 40% in rss is significant, fitting two sine waves would be a rational choice to make. ARMA

Model was fit to the deseasonalized data using F criterion to arrive at ARMA(27,24) model. The improvement in mse of forecasts is significant compared to the actual model, but this model could not outperform the parsimonious model. This can be seen in fig 9.1. This may mean that the periodicities of 12 and 168, though present, are more stochastic than deterministic in nature i.e. the periods of 12 and 168 are discernible but the periodicity appears to be changing from one part of the data to the other. However, this model performed the best for one step ahead forecasts. This is evident from the table below and from the figure 9.1.

The parameters, AR roots and ACF plots of this model can be seen in the Appendix.

Note: The residuals obtained after deseasonalizing the data are of considerable magnitude when compared with the actual data. Hence the parameters obtained by fitting the separate models should be taken as initial estimates and the two models should be optimized together to get the final parameter estimates.

5.1 CONCLUSION

Upon comparing the different models built, we arrive at a conclusion that the parsimonious model significantly outperformed other models and is better when more than one step ahead forecasts are to be made. Deseasonalizing the data and then fitting ARMA (27,24) model would give better results when one step ahead forecasts are to be made.

Model	RSS of one step ahead forecasts on the testing data set
ARMA (32,31)	3.6532e+07
Parsimonious model-ARMA (30,31)	3.5840e+07
ARMA(27,24) on deseasonalized data	2.9036e+07

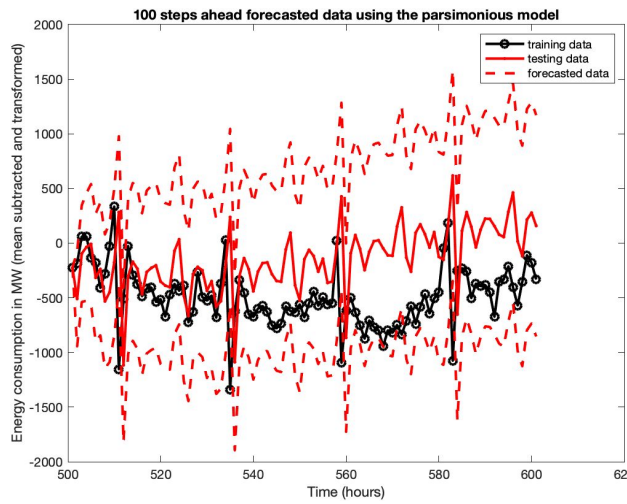


Fig.10.1

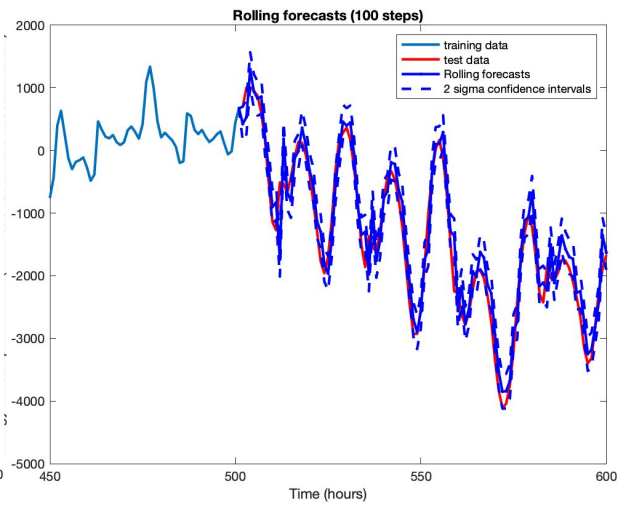


Fig.10.2

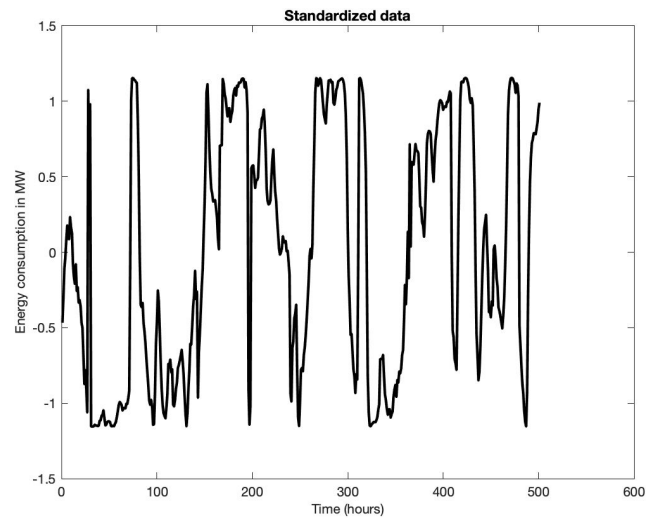
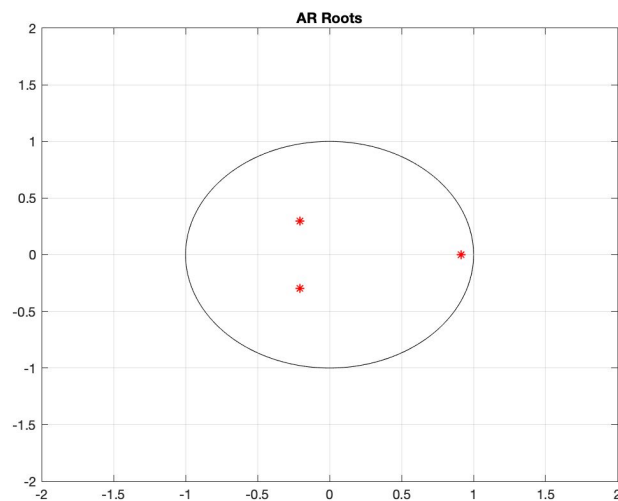
Fig.10.1 indicates 100 steps ahead forecast using the parsimonious model. One can see that this model accurately forecasts up to about 35 steps ahead. Fig.10.2 indicates the rolling forecasts of up to 100 hours using the ARMA (27,24) model on deseasonalized data.

Appendix

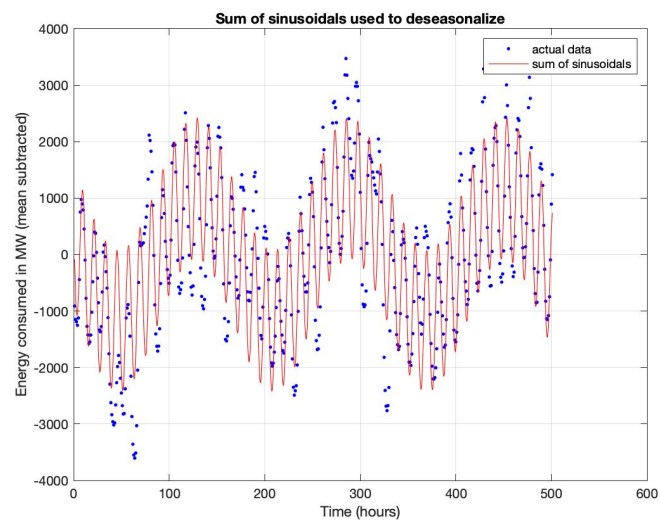
Data

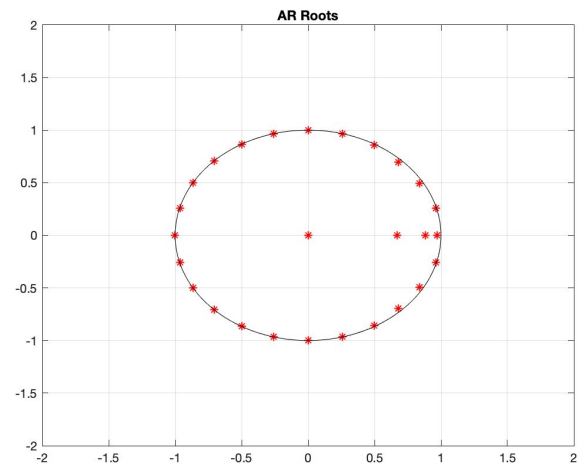
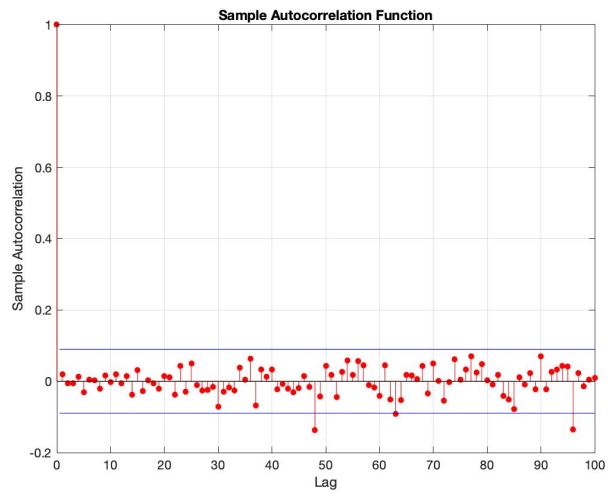
Data source: <https://www.kaggle.com/robikscube/hourly-energy-consumption>

X.2.1 ARMA model on standardized data



X.2.2 ARMA model on deseasonalized data





X.2.3 Parsimonious model

