# SYDE 631 – TIME SERIES MODELLING

**ASSIGNMENT 2**

**Keshitha Babu Suresh Kumar (20802870)**

***5.6 Examine the graph of a time series which is of direct interest to you. Describe general statistical properties of the series that you can detect in the graph.***

ANS: Given below is a graph showing the annual average of phosphorus concentrations in water taken from the year 1972 to 2016. There is a noticeable downward trend in the concentrations and there seems to be one significant spike between 1975 and 1980(further inspection revealed it to be the year 1977 with an average of approximately 0.087). The values seem to be correlated and the series does not resemble a white noise series. It also appears that the variance decreases with increase in time. The graph also indicates that the time series is non-stationary as the average seems to be continuously changing.

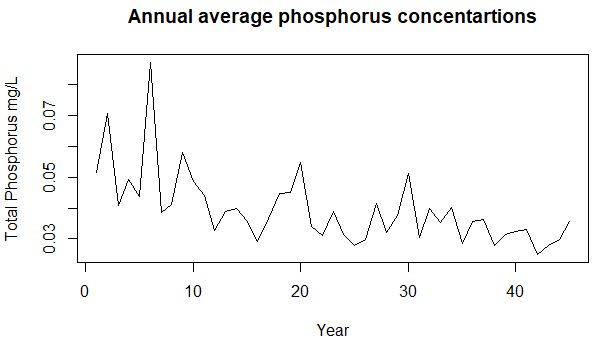
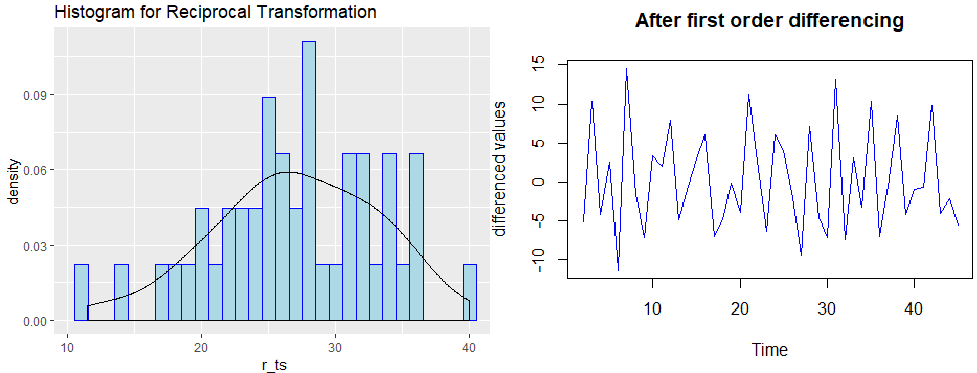


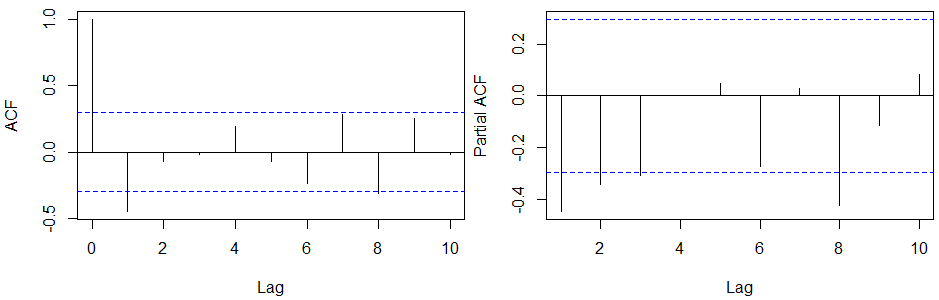
Figure 1 : A plotted graph of the data set

***5.11 Select a non-seasonal time series that is of interest to you. Obtain each of the identification graphs in Section 5.3.3 to 5.3.7 for the series. Based upon these identification results, what is the most appropriate type of ARMA or ARIMA to fit the data set?***

ANS: The data was discovered to be non-stationary (due to a downward trend). To eradicate non-stationarity, data transformation and differencing were carried out on the raw data. Of all transformation techniques, the reciprocal transformation resulted in transforming the data into a seemingly normally distributed series (which Shapiro-Wilk tests also confirm (with a p-value of 0.9448). We first-order differenced the resultant data, which then resulted in a stationary time series. We then proceeded to plot correlograms for the ACF and the PACF,



The Figure below shows the ACF and the PACF correlogram for 10 lags. The ACF clearly cuts off after lag 1. So, the value of *q* for the MA(*q*) can be 1. Similarly, the PACF cuts off after lag 3 and hence, the *p* value for the corresponding AR(*p)* process is clearly 3. Since these correlation functions were obtained after differencing once, the value of *d* in an ARIMA(*p,d,q*) model is 1. Hence, a possibly good model for this series is an ARIMA(3,1,1). Another interpretation of the ACF and the PACF is that the ACF seems to dies down(possibly in a sinusoidal manner) and the PACF cuts off after lag 3. So maybe an AR(3) on the differenced data or an ARIMA(3,1,0) on the reciprocal data.



***6.11 Two approaches for applying the AIC in conjunction with model construction are described in Section 6.3.3. Using an annual time series of your choice, employ these two procedures for determining the best overall ARMA or ARIMA model to fit to the series.***

MODEL SELECTION

ANS : MAICE(Minimum AIC estimation)asks to calculate two values. 1) The AIC and 2) The BIC for all possible models. From the previous question, we decided that an ARIMA(3,1,1) or ARIMA(3,1,0) would be suitable for the model. We can consider also ARIMA(2,1,1) and ARIMA(1,1,4) to examine the effects of changing the number of parameters. The table listed below articulates our findings.

|  |  |  |
| --- | --- | --- |
| **Model** | **AIC** | **BIC** |
| ARIMA(3,1,1) | 279.5676 | 288.4886 |
| ARIMA(3,1,0) | 277.572 | 284.7087 |
| ARIMA(2,1,1) | 282.1183 | 285.7712 |
| ARIMA(1,1,4) | 281.2532 | 291.9583 |

The lowest AIC value belongs to the ARIMA(3,1,0) model and the highest AIC is that of the ARIMA(2,1,1) model. The lowest BIC value again belongs to the ARIMA(3,1,0) model and the highest BIC is that of the ARIMA(1,1,4) model. Hence, considering low values for both the AIC and BIC, ARIMA(3,1,0) seems like the best out of all options.

***7.1 Select an average annual time series that is of interest to you. Following the three stages of model construction and using an available time series program such as the MH package mentioned in Section 1.7, fit the most appropriate ARMA(p,q) model to the dataset. Overspecify the fitted model by adding an additional MA or AR parameter. Estimate the parameters of the overspecified model and comment upon the size of the SE’s. Employ the likelihood ratio test of [7.2.1] to ascertain if overfitting is needed to start with and also to determine if the overfitted model is better than the simpler model***

ANS :

PARAMETER ESTIMATION

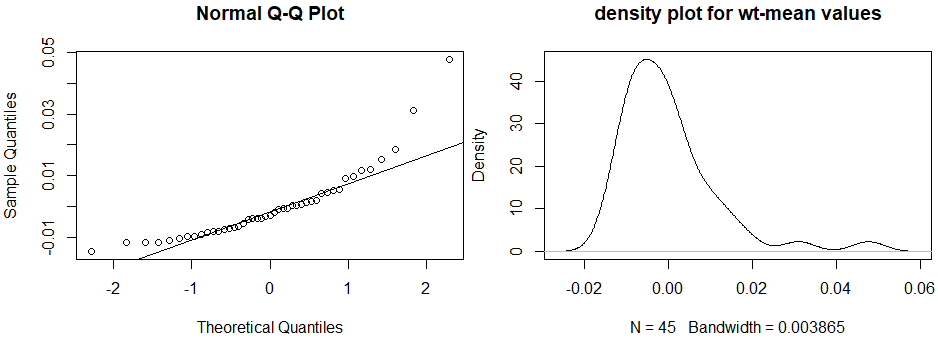
The parameter estimation for the models mentioned previously is as given below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **MODEL** | **PARAMETERS** | **MLE** | **SE** | **σa2** |
| ARIMA(3,1,1) | ɸ1  ɸ2  ɸ3  θ1 | -0.7542  -0.5758  -0.3506  0.0236 | 0.3446  0.2475  0.1806  0.3530 | 26.28 |
| ARIMA(3,1,0) | ɸ1  ɸ2  ɸ3 | -0.7333  -0.5632  -0.3431 | 0.1444 0.1623 0.1449 | 26.28 |
| ARIMA(2,1,1) | ɸ1  ɸ2  θ1 | -0.0523  -0.1073  -0.6968 | 0.2550  0.2108  0.2171 | 26.93 |
| ARIMA(1,1,4) | ɸ1  θ1  θ2  θ3  θ4 | -0.4803  -0.1943  -0.4121  -0.2448  0.2503 | 0.2574  0.2720  0.1707  0.2406  0.2090 | 26.35 |

It is clear from the table that the SEs of the previously selected model, ARIMA(3,1,0) are much lesser compared to the other models. This means that the model is a good fit for the model without any bias.

***7.8(iii)normality plot***

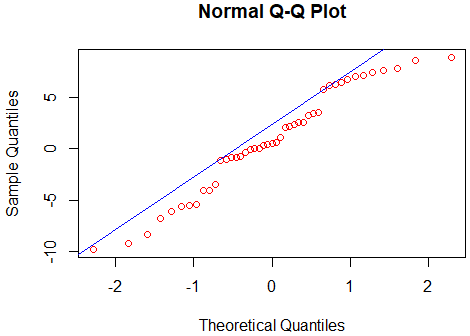
ANS : A QQ normality plot and a density plot were generated for the values obtained after subtracting µ from each wt.



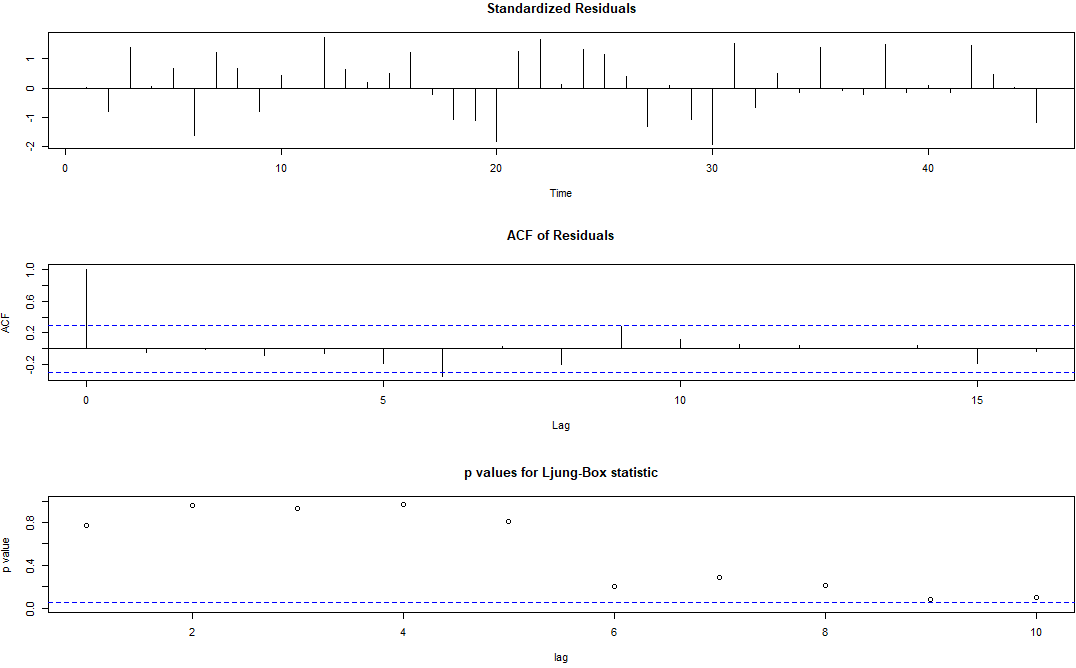
***7.12 select a yearly hydrological time series to model. Using a time series package, follow the three stages of model construction to ascertain the best ARMA or ARIMA model to fit to the data. Clearly explain all your steps and show how identification and diagnostic check graphs.***

DIAGNOSTIC CHECKING

After plotting the residuals in QQ plot, we can observe that they seem to vaguely follow a trend line. A Shapiro-Wilk test was carried out for confirming the presence of normality and the test confirmed(p-value = 0.09593) our assumptions. To check if the residuals were independent, one of the portmanteau tests were carried out. The Ljung-Box test confirmed that the residuals can be assumed to be independent, with a p-value of 0.7706.



The Residual ACF and the Ljung-Box statistic are given below.



After trying to forecast the residuals, the plotted forecasts show that forecasting was not done for unknown data. This further proves that the residuals are pure white noise.

