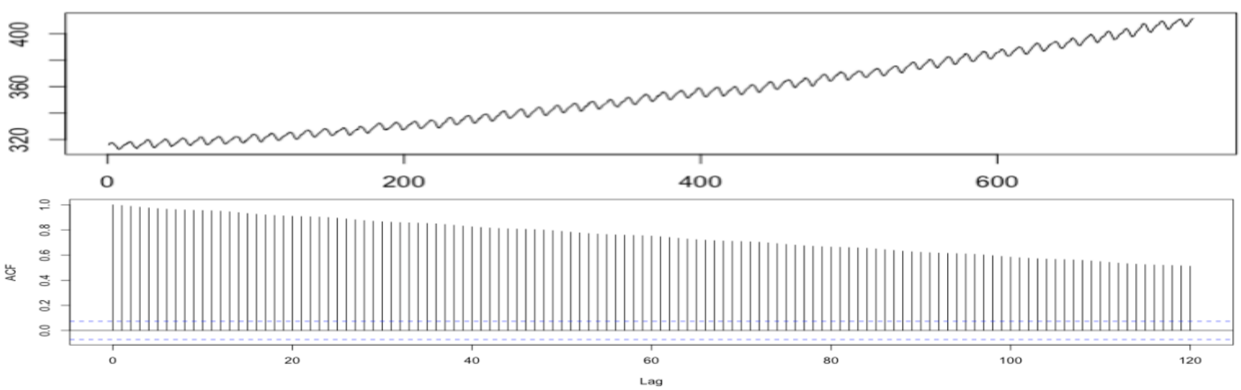
Forecasting Time Series

Individual Final Assignment – Diego Cuartas

In the following report I am using the Box-Jenkins methodology to find two time-series models analyze monthly mean CO2 mole fraction at Mauna Loa Observatory, Hawaii, with 732 records from March 1958 until February 2019.

CO2 mole fraction at Mauna Loa Observatory 1983-2009 (Monthly data):

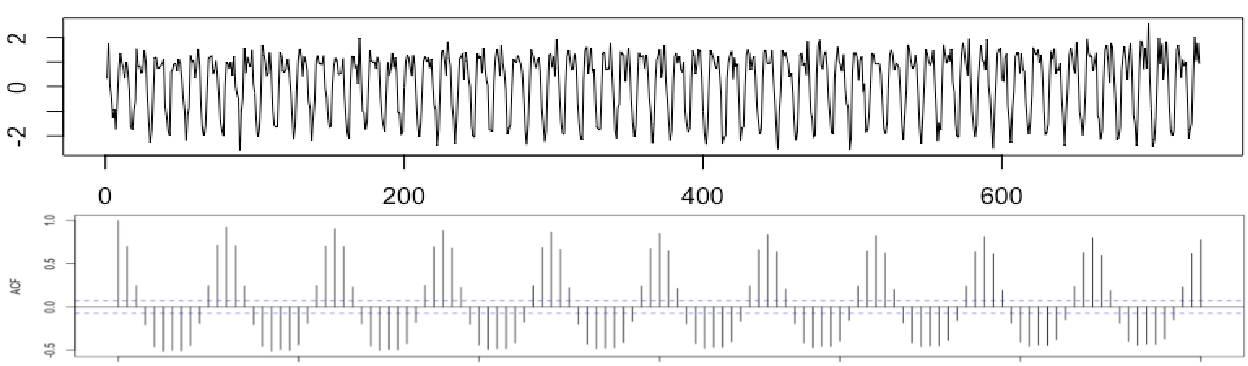


By plotting the CO2 data, it was observed that the series (Yt) is not stationary in the mean but seems to be stationary in the variance. Moreover, by plotting the ACF, its slow decrease to 0 is a strong indication of not stationarity at least in the mean. The Box-Jenkins method is now applied.

Box-Jenkins methodology:

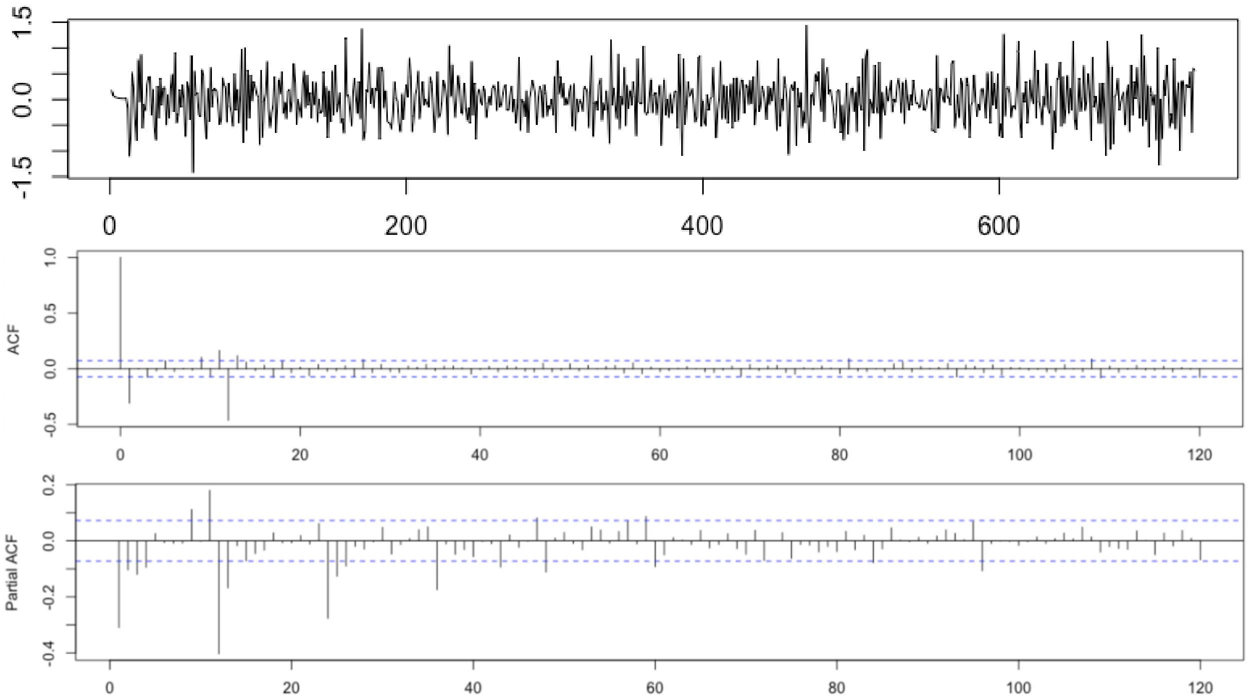
a) (Yt) - S: 12

1. Seasonal Differences: **0** - Regular Differences: **1**



The series now seems stationary in the mean. Executing the *Dickey Fuller Test,* the recommendation of the formal test is to not take any additional differences, but after plotting the ACF and PCAF uncover the indication for a cyclical/seasonal trend still present in the data (lags =12, 24, ...). This can be covered applying a seasonal difference:

1. Seasonal Differences: **1** - Regular Differences: **1**



The ACF and PACF have structure to it, so there to possible models. Thereafter and check if their coefficients were statistically significant.

2. Possible **SARIMA (p, d, q) x (P, D, Q)**, parameter **significance** and **residuals**:

The seasonal and regular lags are examined to explore several autoregressive and moving average models. Although many models were proposed they were subsequently disregarded as their parameters were not statistically significant (e.g. {(0,1,1) x (1,1,0)} or {(0,1,0) x (0,1,1)}).

Thereafter I checked if the residuals for white noise. More proposed models were removed from the list. I then selected 4 models with a simple structure on the PACF and ACF which allowed me to capture the seasonal trend with a small number of significant parameters and white noise residuals.

SARIMA Model	Sig. Param	Residuals WN	Box Test
Regular(0,1,1) Seasonal (3,1,0)	Yes	Yes	0.06344
Regular(1,1,0) Seasonal (0,1,1)	Yes	Yes	0.1611
Regular(2,1,0) Seasonal (0,1,1)	Yes	Yes	0.2377
Regular(3,1,0) Seasonal (0,1,1)	Yes	Yes	0.4429

Final Forecast with Errors:

I chose the Rolling Scheme to assess the forecasting error, to avoid the potential problem with the model's stability, and it is more robust against structural breaks in the data as observed in this case when plotting the CO_2 data.

SARIMA Model	Sig. Param	Residuals WN	Box Test	MSFE (H=1)	MAPE (H=1)
Regular(3,1,0) Seasonal (0,1,1)	Yes	Yes	0.4429	0.1287021	0.06986877
Regular(1,1,0) Seasonal (0,1,1)	Yes	Yes	0.1611	0.1288252	0.07102978
Regular(2,1,0) Seasonal (0,1,1)	Yes	Yes	0.2377	0.1297264	0.07018261
Regular(0,1,1) Seasonal (3,1,0)	Yes	Yes	0.06344	0.1392039	0.07119498

The best model I chose for forecasting the monthly mean CO_2 mole fraction time series is **SARIMA (1,1,0) x (0,1,1) s=12**. because My decision criteria are mainly based on the simplicity of the model to achieve similar results compared with the best model in terms of forecasting error. With a lower number of parameters to estimate it achieves the second lowest **MSFE (0.1288252)** for the first horizon predicted (only determined 1 horizon considering the time limitation and computer power required, further analysis should review 12 horizons for monthly predictions).

Below is a plot of my final model prediction: