Prediction and Time Series Computer Exercise Week 4

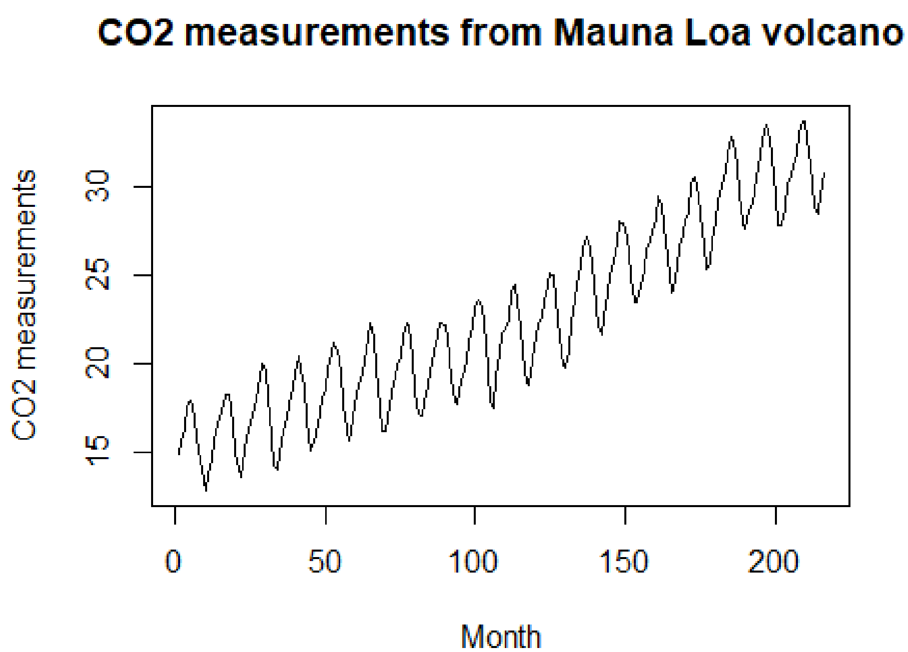
Name: Nguyen Xuan Binh Student ID: 887799

11/25/2021

Exercise 4.3: A time series of carbon dioxide measurements from the Mauna Loa volcano is given in the file MLCO2.txt. The length of the time series is 216 months. Recall that, we studied this time series during the third computer exercises.

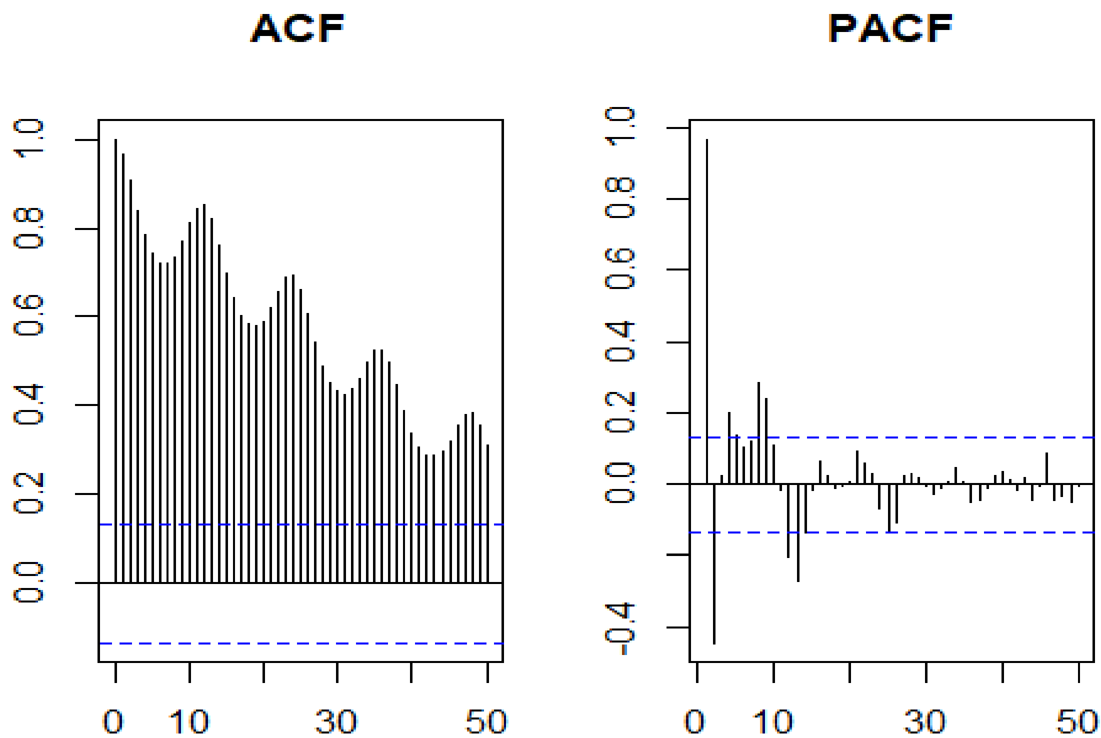
a) Using SARIMA processes, find the best possible model to describe the time series MLCO2.

```
library(forecast) # Run every time you want to use functions of the forecast package

## Warning: package 'forecast' was built under R version 4.1.2

## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo

MLCO2 <- read.table("MLCO2.txt",header=T)

mlco2ts <- ts(MLCO2$MLCO2, frequency = 1)


plot(mlco2ts, xlab = "Month" ,ylab = "CO2 measurements", main = "CO2 measurem
ents from Mauna Loa volcano")
```



According to the time series, the CO2 measurements display seasonality and also a regular upward trend. We will now test which time series model will fit best by studying the time series' ACF and PACF

```
par(mfrow=c(1,2),mar=c(2.5,2.5,3.5,1.5))
acf(mlco2ts,main="ACF", lag.max=50)
pacf(mlco2ts,main="PACF", lag.max=50)
```



Analysis: The time series is seasonal ACF decays exponentially (although it may appear as linear decay at first 50 lags) along the seasonal lags and PACF has cut off at lag 2. The sample PACF with lags 4, 5, 8, 9, 12, 13 reach over the blue lines indicating statistical significance => AR(2) should be a component of SARIMA model. Now we use auto sarima model on the time series.

```
model_mlco2 <- auto.arima(mlco2ts)
model_mlco2

## Series: mlco2ts
## ARIMA(2,1,1) with drift
##
## Coefficients:
##          ar1      ar2      ma1   drift
##       1.5336  -0.8358  -0.9357  0.0778
## s.e.  0.0368   0.0364   0.0191  0.0093
##
## sigma^2 estimated as 0.3561:  log likelihood=-193.69
## AIC=397.38   AICc=397.66   BIC=414.23
```

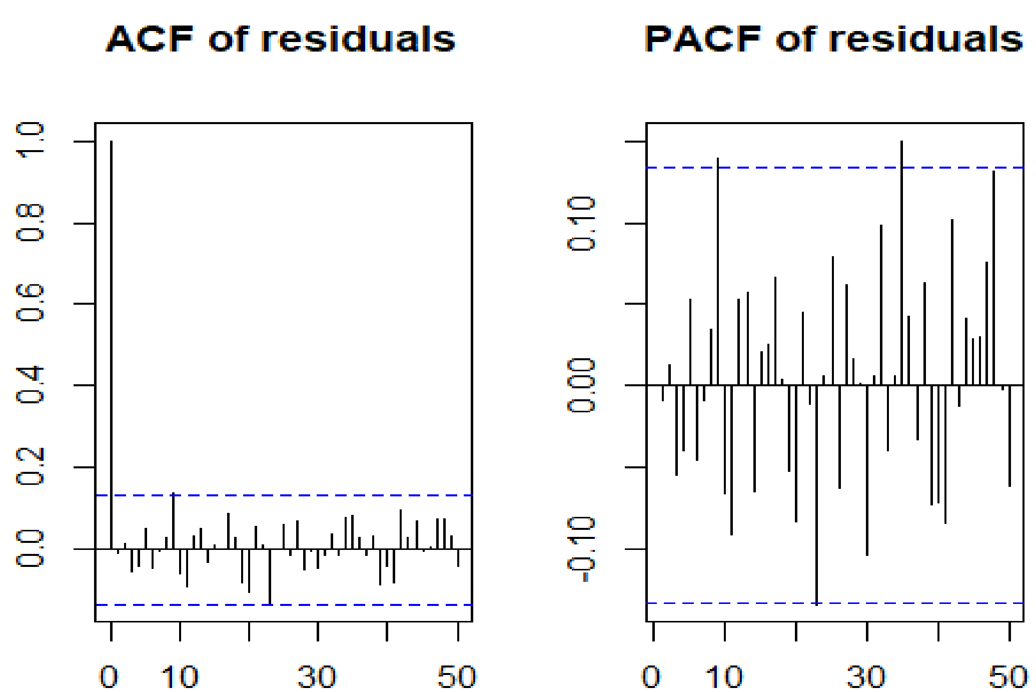
From the auto.arima function, it needs first order differencing of non-seasonal component in order to remove the linearly upward trend of the time series. Also, AR(2) is present in the prediction => SARIMA(2, 1, 1)(P, H, Q)[12] is a candidate. We will try the auto.arima function for the time series of frequency = 12

```
mlco2SeasonTs <- ts(MLCO2$MLCO2, frequency = 12)
model_mlco2Season <- auto.arima(mlco2SeasonTs)
model_mlco2Season
```

```
## Series: mlco2SeasonTs
## ARIMA(0,1,2)(2,1,0)[12]
##
## Coefficients:
##          ma1      ma2     sar1     sar2
##      -0.3609  -0.1808  -0.6505  -0.3520
## s.e.  0.0679   0.0672   0.0679   0.0705
##
## sigma^2 estimated as 0.1299:  log likelihood=-82.17
## AIC=174.34   AICc=174.65   BIC=190.91
```

The auto model predict with the order part of (0,1,2). I will choose this ordering to be seasonal order instead. The candidate model now is SARIMA(2,1,1)(0,1,2)[12]. We need to verify if it is correct model

```
modelSARIMA <- Arima(mlco2ts,order=c(2,1,1), seasonal=list(order=c(0,1,2),per
iodc=12))
par(mfrow=c(1,2),mar=c(2.5,2.5,3.5,1.5))
acf(modelSARIMA$res,main="ACF of residuals",lag.max=50)
pacf(modelSARIMA$res, main="PACF of residuals", lag.max=50)
```

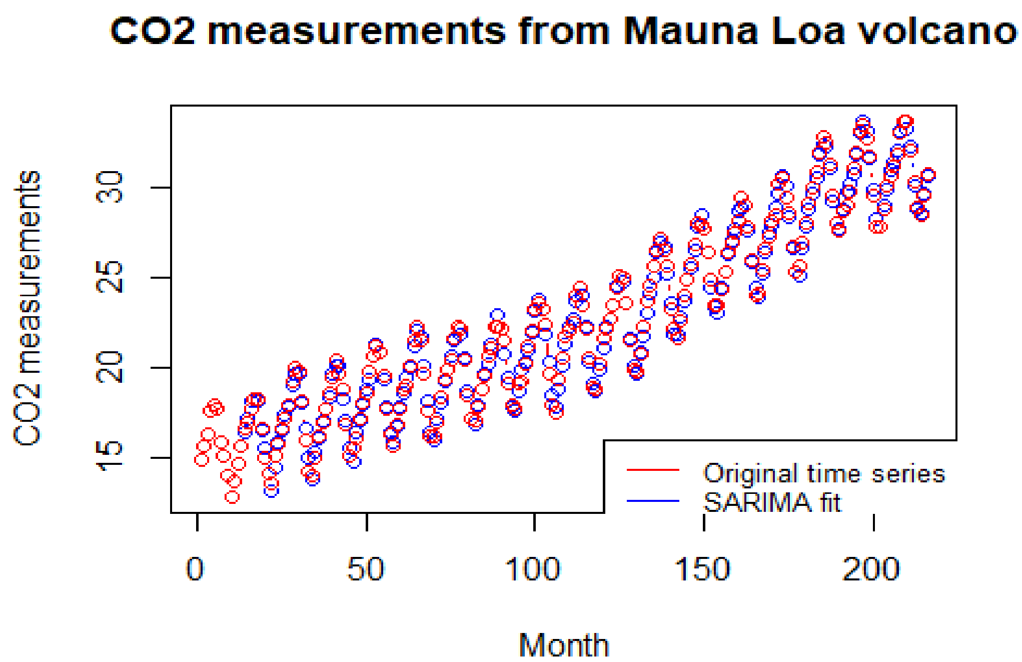


There are about 3 significant value in residual ACF and 3 significant values in residual PACF, which is few compared to the large number of sample data. It could be a sign that this SARIMA model is correct Now we utilize Ljung Box test Number of fitted parameters: k = 2+1+2 = 5

```
k <- 5
mlco2_bl <- rep(NA,47)
for (i in 1:47) {
  mlco2_bl[i]=Box.test(modelSARIMA$res,lag=(i+k),fitdf=k, type="Ljung-Box")$p
.value
}
round(mlco2_bl, 3)

##  [1] 0.148 0.351 0.517 0.156 0.191 0.158 0.217 0.258 0.325 0.411 0.500 0.4
39
## [13] 0.504 0.464 0.359 0.382 0.448 0.263 0.318 0.325 0.381 0.376 0.399 0.4
56
## [25] 0.486 0.540 0.577 0.629 0.595 0.554 0.595 0.642 0.676 0.625 0.649 0.6
08
## [37] 0.529 0.566 0.552 0.597 0.640 0.615 0.594 0.624 0.647 0.662 0.684

which(mlco2_bl < 0.05)

## integer(0)
```

There are no p-values of the SARIMA model residuals that is smaller than 0.05 in the Ljung-Box test => The null hypothesis of Ljung-Box is rejected. The SARIMA(2, 1, 1)(0, 1, 2)[12] will be chosen as the model for future prediction. Now we compare original time series and the SARIMA fit model Checking if the SARIMA model fits into the original time series, we can see that SARIMA model approximates the time series quite close

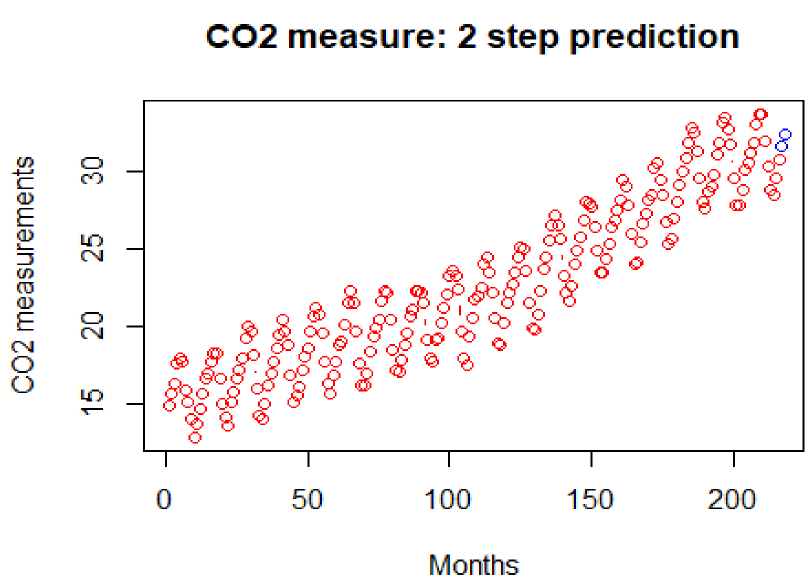
```
fit.mlco2 <- fitted(modelSARIMA)

plot(fit.mlco2,type="b",col="blue", ylab="CO2 measurements",xlab="Month", mai
n="CO2 measurements from Mauna Loa volcano")

lines(mlco2ts,col="red",type="b")
legend(120,16, legend=c("Original time series", "SARIMA fit"),
col=c("red","blue"),lty=c(1,1),cex=0.8)
```



b) Make 2 and 24 time step predictions by using the model chosen in a). Study the goodness of the predictions 2 time step predictions by using the model in a)

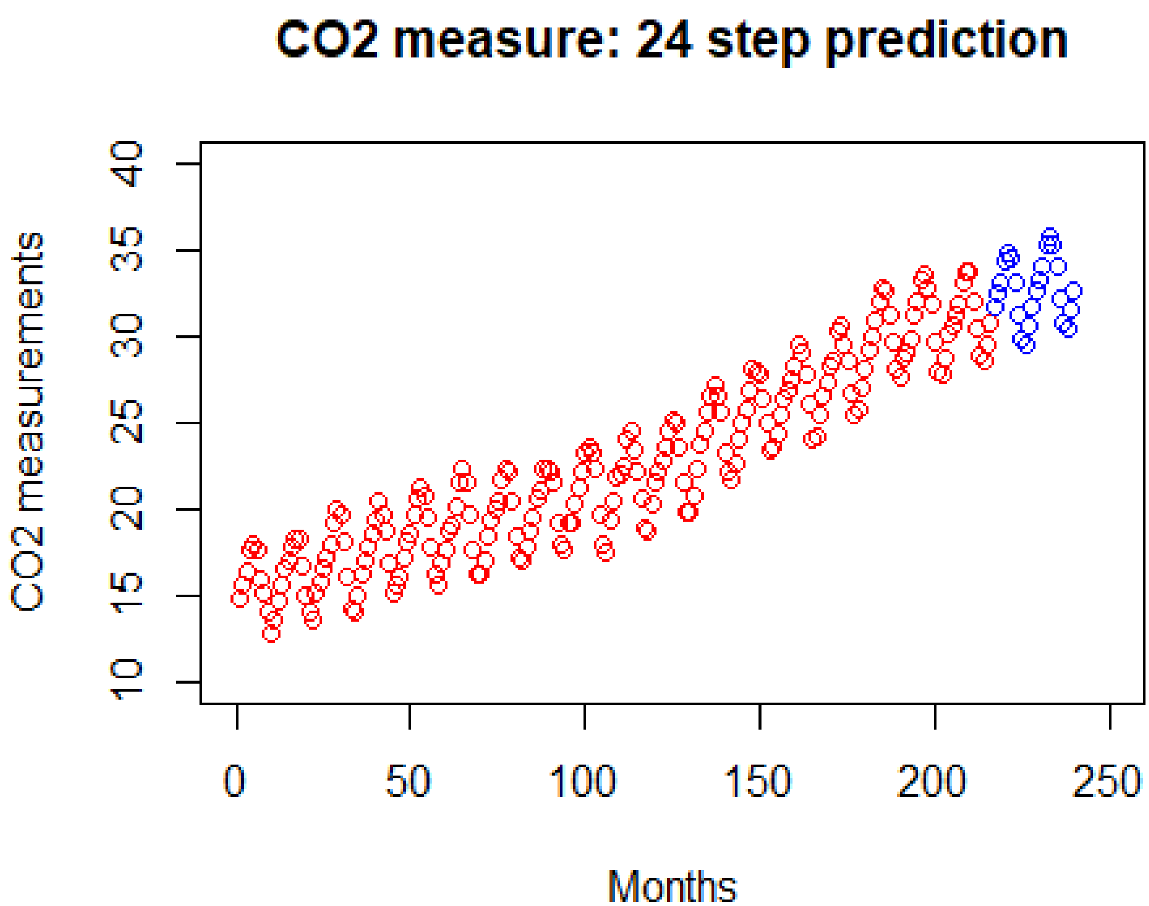
2 time step predictions by using the model in a)

```
currentSARIMA = arima(mlco2ts[1:216],order=c(2,1,1), seasonal=list(order=c(0,
1,2),period=12) )
prediction2steps <-forecast(currentSARIMA,h=2,level=FALSE)$mean
plot(mlco2ts,col="red",type="b", ylab="CO2 measurements",xlab="Months",main="
CO2 measure: 2 step prediction")
lines(prediction2steps,col="blue",type="b")
```

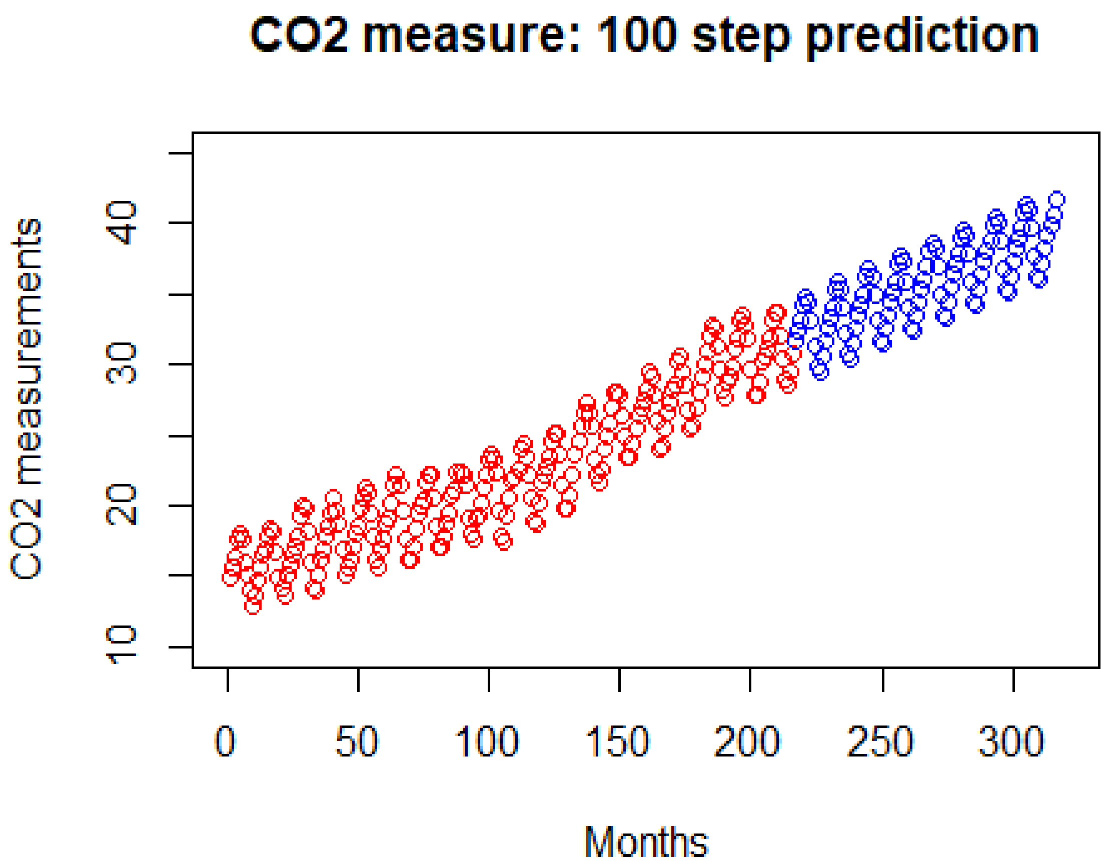


24 time step predictions by using the model in a)

```
currentSARIMA = arima(mlco2ts[1:216],order=c(2,1,1), seasonal=list(order=c(0,
1,2),period=12) )
prediction2steps <-forecast(currentSARIMA,h=24,level=FALSE)$mean

plot(mlco2ts,col="red",type="b", xlim=c(0, 250), ylim =c(10,40),
     ylab="CO2 measurements",xlab="Months",main="CO2 measure: 24 step predict
ion")
lines(prediction2steps,col="blue",type="b")
```



```
currentSARIMA = arima(mlco2ts[1:216],order=c(2,1,1), seasonal=list(order=c(0,
1,2),period=12) )
prediction2steps <-forecast(currentSARIMA,h=100,level=FALSE)$mean

plot(mlco2ts,col="red",type="b", xlim=c(0, 320), ylim =c(10,45),
     ylab="CO2 measurements",xlab="Months",main="CO2 measure: 100 step predic
tion")
lines(prediction2steps,col="blue",type="b")
```



We can see that prediction of SARIMA(2,1,1)(0,1,2)[12] is quite good in predicting the future of amount of CO2, since this model has passed the Ljung-Box test. Therefore, I believe the prediction for time steps of 2 and 24 are likely to be correct under this SARIMA model