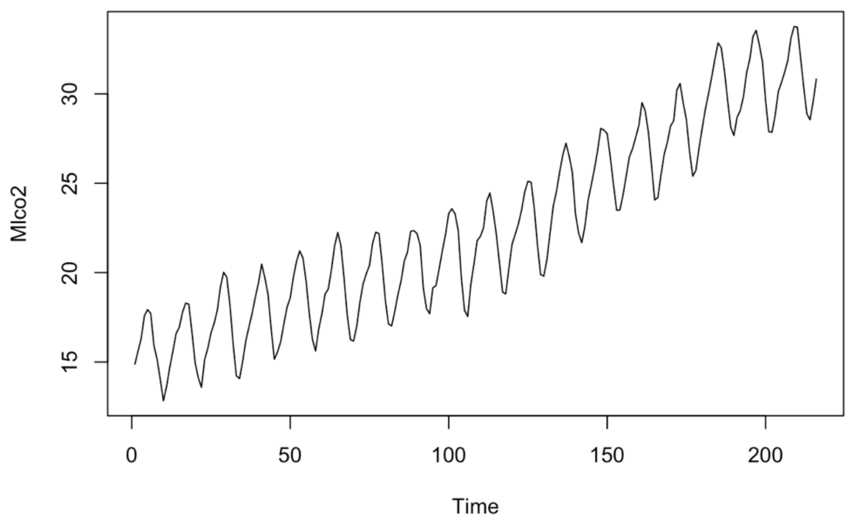
NHUT CAO 906939
COMPUTER EXERCISE 4

4.3.a) We first read and study the data, we see that this time series has a linear trend and seasonality, therefore, to find a proper model, we should take these two into consideration

```
# 4.3 a
library(forecast)
```

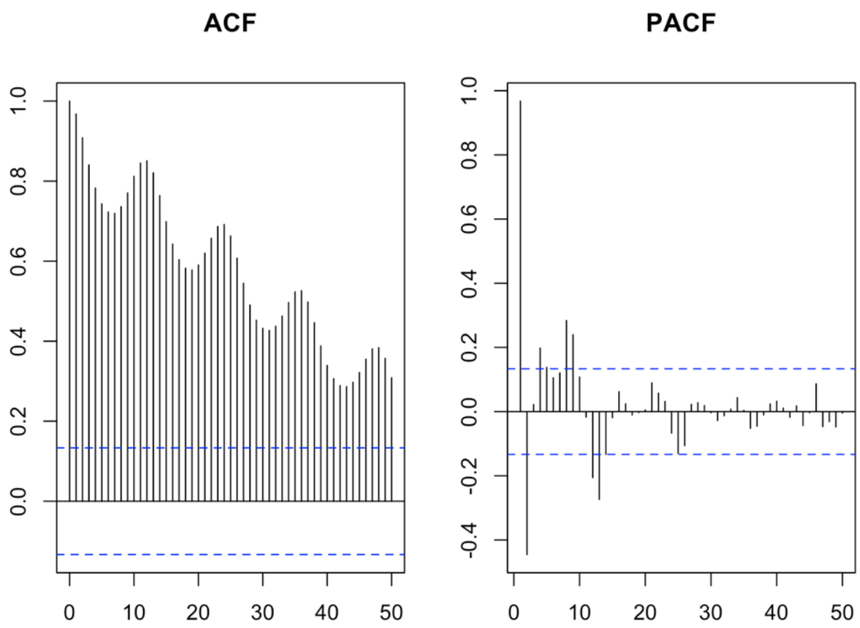
```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
MLCO2 <- read.table("MLCO2.txt", header=T)
Mlco2 <- ts(MLCO2$MLCO2)
plot(Mlco2)
```



Below we study the ACF and PACF of the original data, it can be seen that there are no exponentially decays or elements that suggest this is a suitable model

```
par(mfrow=c(1,2),mar=c(2.5,2.5,3.5,1.5))
acf(Mlco2,main="ACF", lag.max=50)    # No exponentially decay
pacf(Mlco2,main="PACF", lag.max=50) #cuts off after lag 2.
```




```
# remove linear upward trend and seasonality
seasonal = list(order = c(0,1,2), period=12)
test.model <- arima(Mlco2, order=c(2,1,1),seasonal)
test.model
```

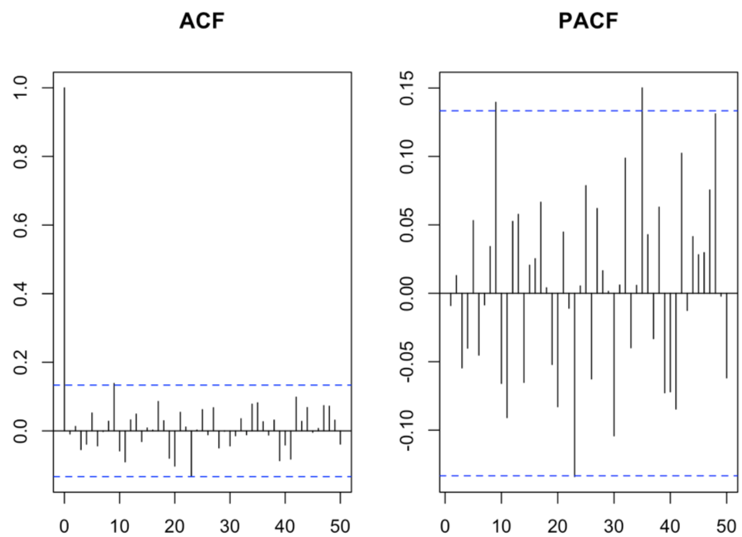
So, to find ARIMA model, we consider two elements that can affect the result, which are the linear upward trend and the seasonality. Here, when apply with the order parameters, we get AR(1), AR(2), MA(1), SMA(1), and SMA(2)

```
# remove linear upward trend and seasonality
seasonal = list(order = c(0,1,2), period=12)
test.model <- arima(Mlco2, order=c(2,1,1),seasonal)
test.model
```

```
##
## Call:
## arima(x = Mlco2, order = c(2, 1, 1), seasonal = seasonal)
##
## Coefficients:
##          ar1      ar2      ma1     sma1    sma2
##       0.3890  -0.0134  -0.7067  -1.0078  0.0080
## s.e.  0.1374   0.0868   0.1189   0.1656  0.0834
##
## sigma^2 estimated as 0.09389:  log likelihood = -65.44,  aic = 142.89
```

We then study the ACF and PACF of the models, we see that there are no seasonality of lags, and some spikes are significant.

```r
#par(mfrow=c(1,2),mar=c(2.5,2.5,3.5,1.5))
acf(test.model$res,main="ACF", lag.max=50)
pacf(test.model$res,main="PACF", lag.max=50)
```



Now use the Ljung-Box Test to test the models. The result has shown that the model passed all the test. Therefore, these models are potentially correct models.

```r
mlco2_bl <- rep(NA,45)
k <- 5
for(i in 1:45){
  mlco2_bl[i] <- Box.test(test.model$res,lag=(i+k),fitdf=k,
                          type="Ljung-Box")$p.value
}
mlco2_bl
```

```
##  [1] 0.1478548 0.3507944 0.5170311 0.1562532 0.1913266 0.1576421 0.2165961
##  [8] 0.2584561 0.3251842 0.4113739 0.5001573 0.4388728 0.5035348 0.4635427
## [15] 0.3591399 0.3815785 0.4481114 0.2625803 0.3182021 0.3247577 0.3809342
## [22] 0.3757072 0.3988561 0.4564598 0.4861898 0.5400064 0.5771302 0.6286171
## [29] 0.5952317 0.5539655 0.5952986 0.6424980 0.6763590 0.6248907 0.6491902
## [36] 0.6078577 0.5293443 0.5657537 0.5523514 0.5969657 0.6395501 0.6154567
## [43] 0.5939493 0.6239648 0.6468343
```

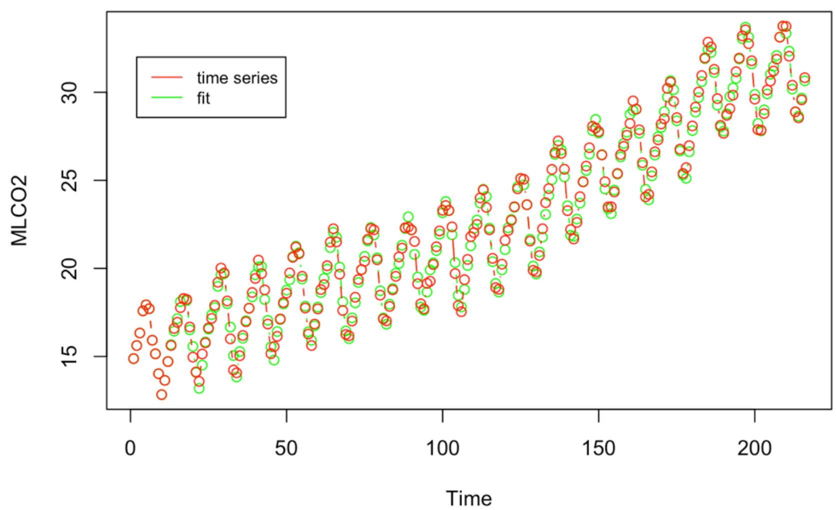
```r
which(mlco2_bl <= 0.05) + k
```

```
## numeric(0)
```

We then test if the models match with the original time series. The result is pretty good, the models are really close to the original data.

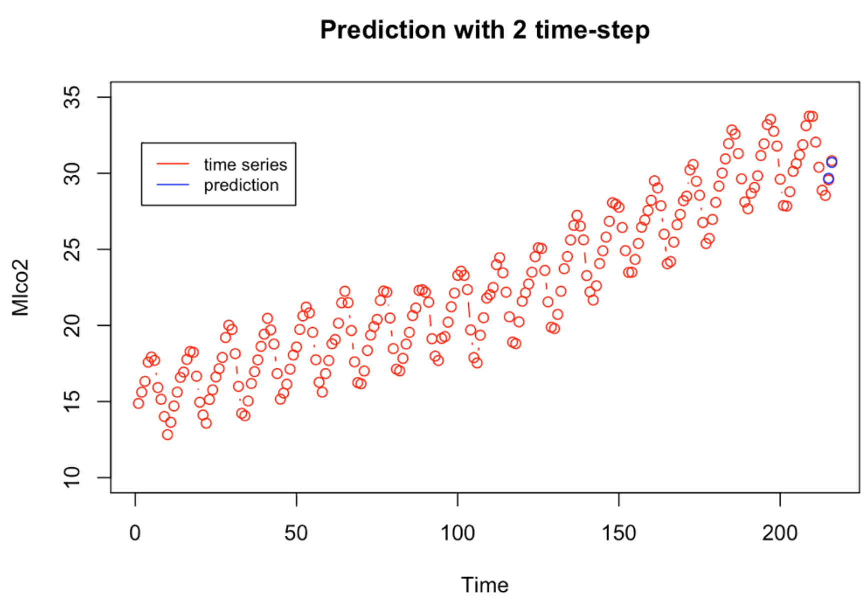
```
seasonal = list(order = c(0,1,2), period=12)
test.model <- arima(Mlco2, order=c(2,1,1),seasonal)

fit.model <- fitted(test.model)
plot(fit.model, type="b",col="green",xlab="Time", ylab='MLCO2')
lines(Mlco2,type='b',col='red')
legend(2,32, legend=c("time series", "fit"), col=c("red","green"),lty=c(1,1),cex=0.8)
```

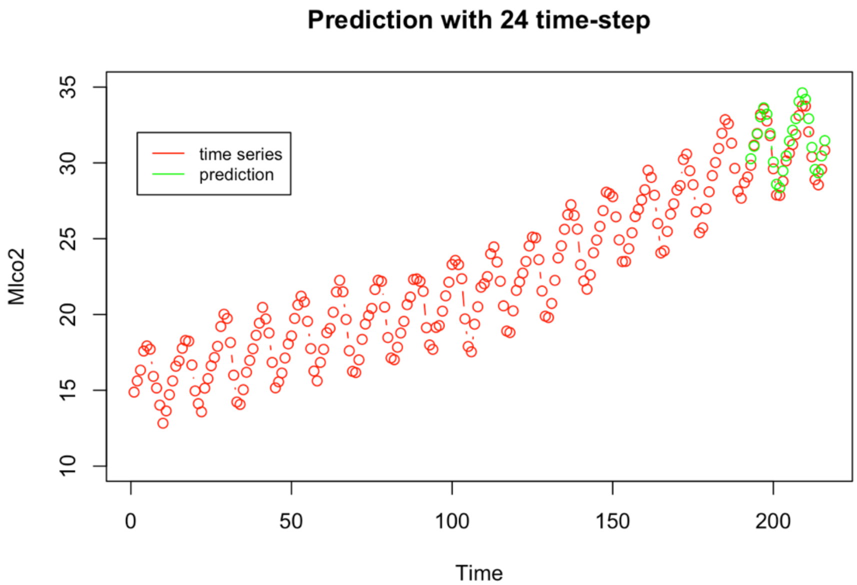


Next, we test the prediction of 2-time step and 24-time step

```
model_plot1 <- arima(Mlco2[1:214], order=c(2,1,1),seasonal)
prediction <- forecast(model_plot1,h=2, level=FALSE)$mean
plot(Mlco2, col='red', type='b', ylim=c(10,35), main='Prediction with 2 time-step')
lines(prediction,col='blue',type='b')
legend(2,32, legend=c("time series", "prediction"), col=c('red','blue'),lty=c(1,1), cex=0.8)
```



```
model_plot2 <- arima(Mlco2[1:192], order=c(2,1,1),seasonal)
prediction <- forecast(model_plot2,h=24, level=FALSE)$mean
plot(Mlco2, col='red', type='b', ylim=c(10,35), main='Prediction with 24 time-step')
lines(prediction,col='green',type='b')
legend(2,32, legend=c("time series", "prediction"), col=c('red','green'),lty=c(1,1), cex=0.8)
```



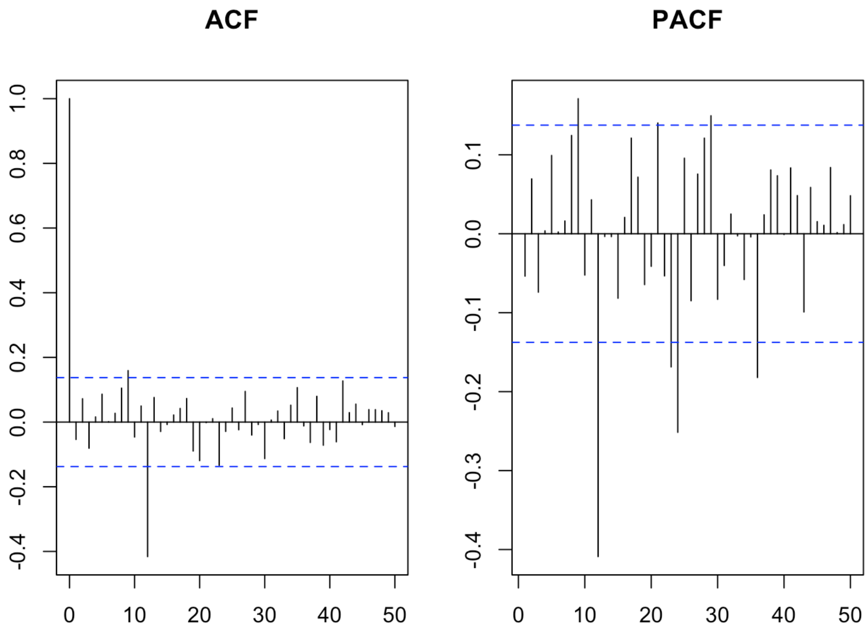
The model is SARIMA(2,1,1)(0,1,2)_12 predict almost correctly, the predictions are close to the time series and show minor differences.

Now if we study the auto.arima() models, the result is not as good as the models we obtained by hand above. Those models don't pass the Ljung-Box Tests. So, the auto models are not always the best choice in choosing appropriate models for the time series.

```
new_mlco2 <- diff(diff(Mlco2),lag=12)
auto.model <- auto.arima(new_mlco2)
auto.model
```

```
## Series: new_mlco2
## ARIMA(1,0,1) with zero mean
##
## Coefficients:
##          ar1      ma1
##       0.6796  -0.9588
## s.e.  0.0842   0.0470
##
## sigma^2 estimated as 0.1827:  log likelihood=-115.04
## AIC=236.08   AICc=236.2   BIC=246.02
```

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




```
mlco2_test <- rep(NA,48)
k <- 2
for(i in 1:48){
  mlco2_test[i] <- Box.test(auto.model$res,lag=(i+k),fitdf=k,
                            type="Ljung-Box")$p.value
}
round(mlco2_test,3)
```

```
##  [1] 0.082 0.214 0.199 0.325 0.440 0.306 0.082 0.110 0.137 0.000 0.000 0.000
## [13] 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## [25] 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
## [37] 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.001
```

```
which(mlco2_test > 0.05) + k
```

```
## [1]  3  4  5  6  7  8  9 10 11
```