MATH1318 Time Series Analysis - Assignment - 2

Zhen Qian

2022-04-24

Introduction

This report aims to propose a set of possible ARIMA(p, d, q) models for the time series of Antarctica land ice mass. The data set contains yearly changes in Antarctica land ice mass from 2002 to 2020. This report is going to use model specification tools such as ACF-PACF, EACF, BIC table to find possible choices of (p, d, q). Transformation techniques, such as Box-Cox transformation and differencing, will be used when applicable.

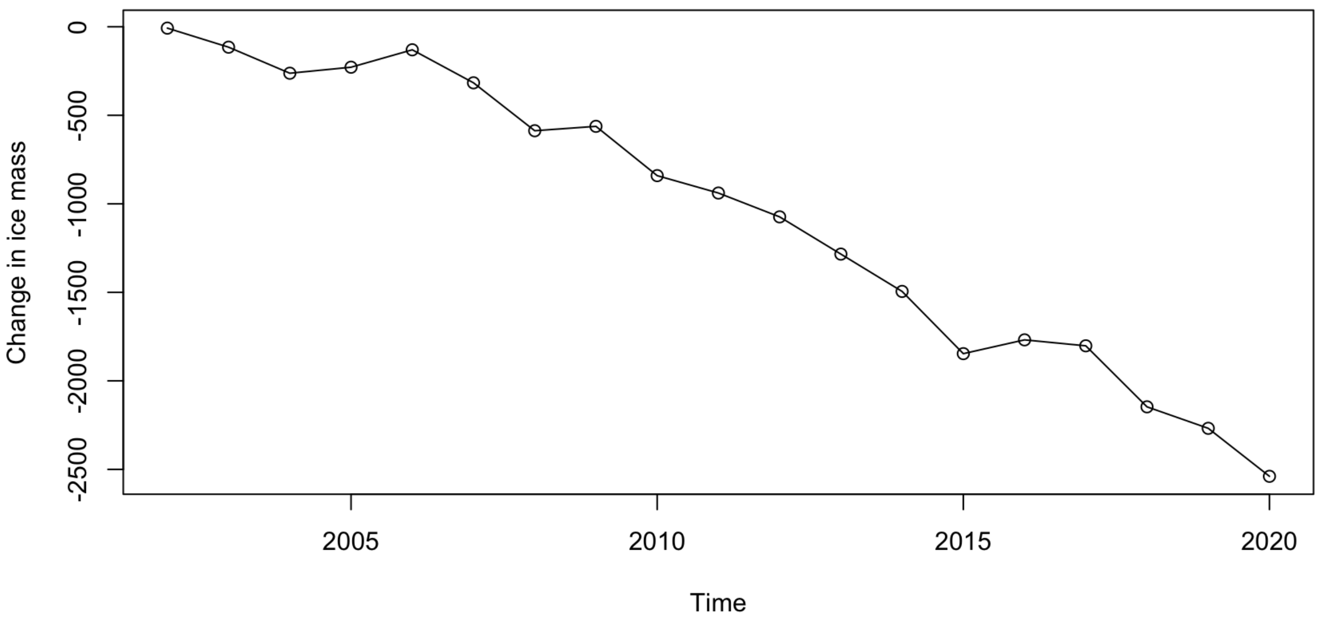
Decriptive Analysis

The data set of Antarctica land ice mass is imported into R. It is transformed to time series object and plotted in Figure 1. We can see that the time series have a downward trend between 2002 and 2020. There is no clear sign of seasonal repeating pattern in the time series. It fluctuates a little bit but we cannot observe changing variance from the figure. It terms of behavior, it is quite obvious that several successive points are moving in the same direction, indicating auto-regressive process, but moving average process is not obvious. Speaking of change point, there is no obvious intervention in the time series and it seems that there is no change point.

```
library(TSA)
library(tseries)

ice <- read.csv('assignment2Data2022.csv', skip=1, header=FALSE)
ice_ts <- ts(ice$V2, start=2002, end=2020)
plot(ice_ts, type='o', ylab='Change in ice mass', main='Figure 1. Time series plot of
yearly changes in Antarctica land ice mass')
```

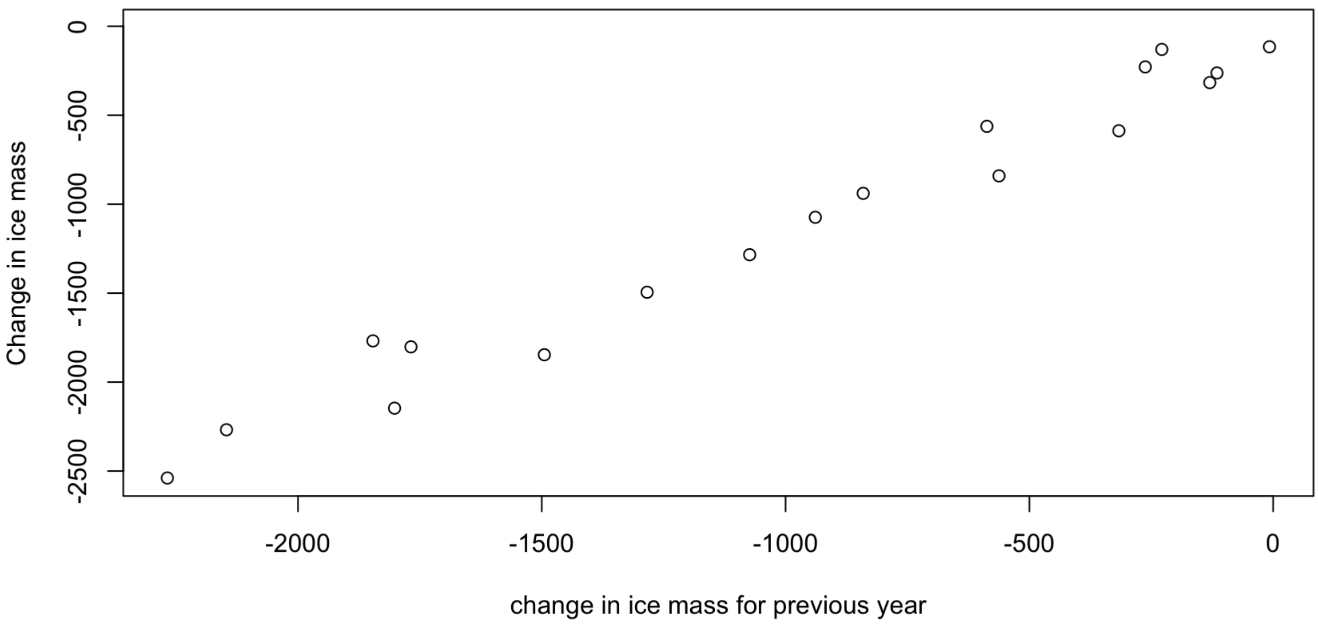
Figure 1. Time series plot of yearly changes in Antarctica land ice mass



The correlation between consecutive values is further examined in figure 2. The change in ice mass is plotted against previous year's value and it shows a linear relation, indicating auto-correlation for lag 1.

```
plot(y=ice_ts, x=zlag(ice_ts), ylab='Change in ice mass', xlab='change in ice mass fo
r previous year', main='Figure 2. Scatter plot of neighbouring change in ice mass')
```

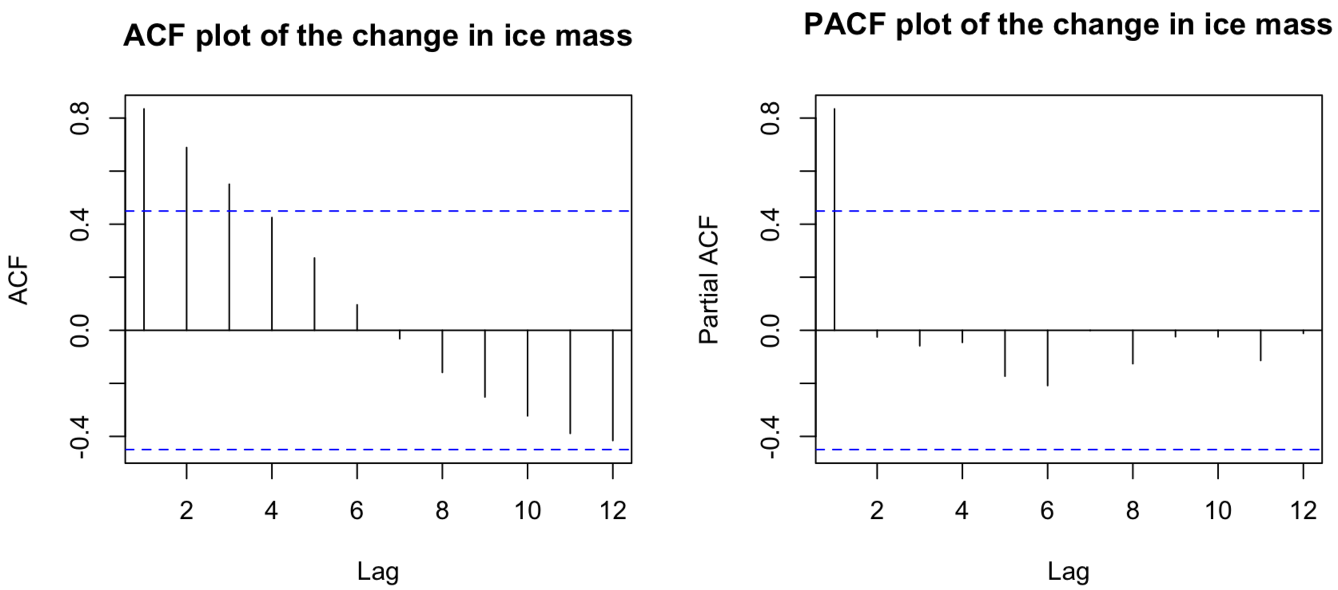
Figure 2. Scatter plot of neighbouring change in ice mass



ACF and PACF are plotted in figure 3. It shows that auto-correlation is slowly decaying in ACF and partial auto-correlation is significant for lag 1. This implies that the time series may have trend and non-stationarity.

```
par(mfrow=c(1,2), mar=c(5,4,4,2), oma=c(0,0,4,0))
acf(ice_ts, main='ACF plot of the change in ice mass')
pacf(ice_ts, main='PACF plot of the change in ice mass')
par(mfrow=c(1,1))
mtext('Figure 3. ACF-PACF analysis', line=2, cex=2, side=3, outer=TRUE)
```

Figure 3. ACF-PACF analysis

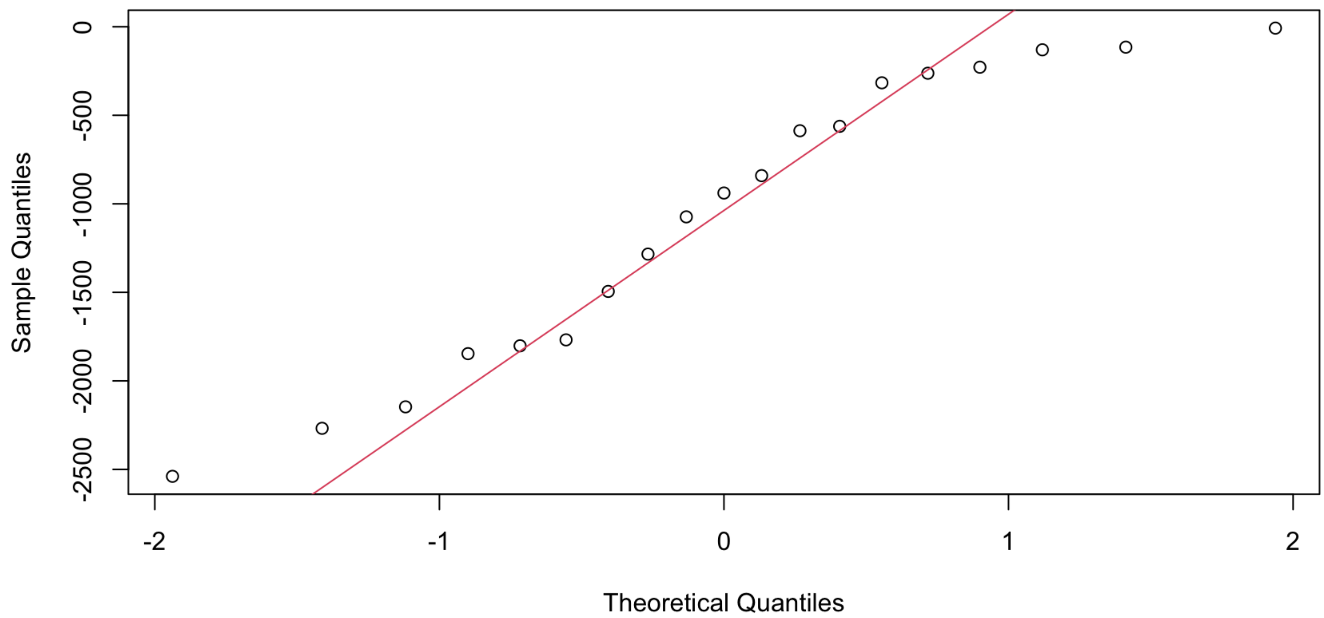


Transformation and differencing We have observed from figure 1 that there is no changing variance, so transformation may not have effect on the series. However, we will still perform the Box-Cox transformation to see if the normality of the observations can be improved. We will also perform unit root test to see if the series is non-stationary. We will difference the series if it is not stationary.

The normality of the original series is examined in figure 4. It shows that most of the sample points are close to the straight line. The p-value from the Shapiro-Wilk test is 0.1417 (greater than 0.05). We cannot reject the non-hypothesis that the observations are normally distributed.

```
qqnorm(ice_ts, main='Figure 4. QQ plot of the original series')
qqline(ice_ts, col=2)
```

Figure 4. QQ plot of the original series

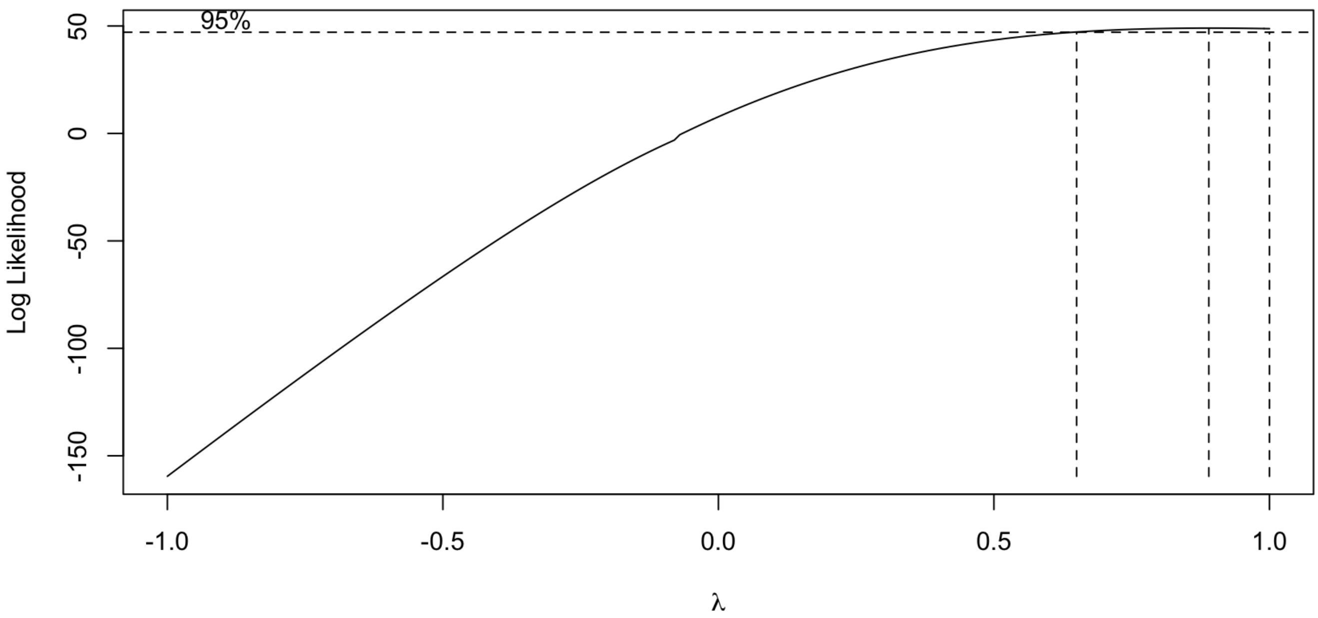



```
shapiro.test(ice_ts)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  ice_ts
## W = 0.92616, p-value = 0.1471
```

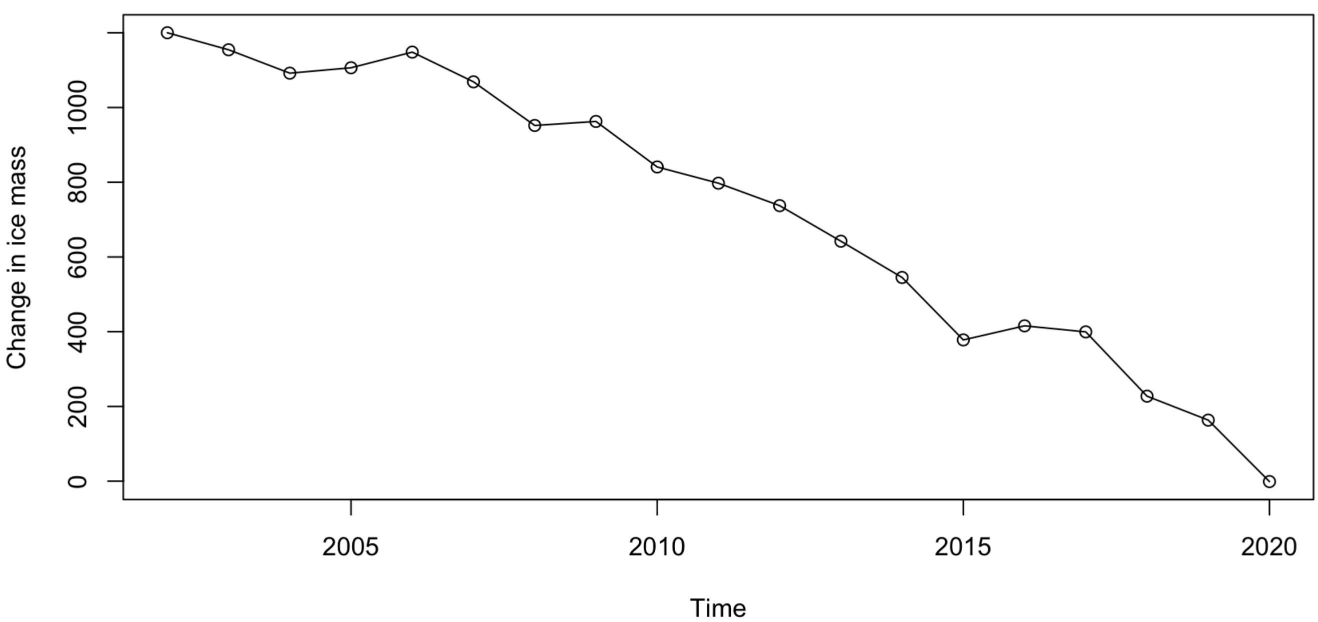
The BC transformed series is plotted in figure 5. As what we have expected, the transformation does not have effect on the original time series in terms of stabilizing the variance, because the original series do not show changing variance. The QQ plot of the transformed series is showed in figure 6. The sample points do not better fit the straight line and the p value for Shapiro-Wilk test is not increased. It seems that the transformation does not improve the normality of the original series. We will use the original series for unit root test and differencing.

```
ice_ts2 <- ice_ts + abs(min(ice_ts)) + 0.01
BC <- BoxCox.ar(y=ice_ts2, lambda=seq(-1,1,0.01))
```



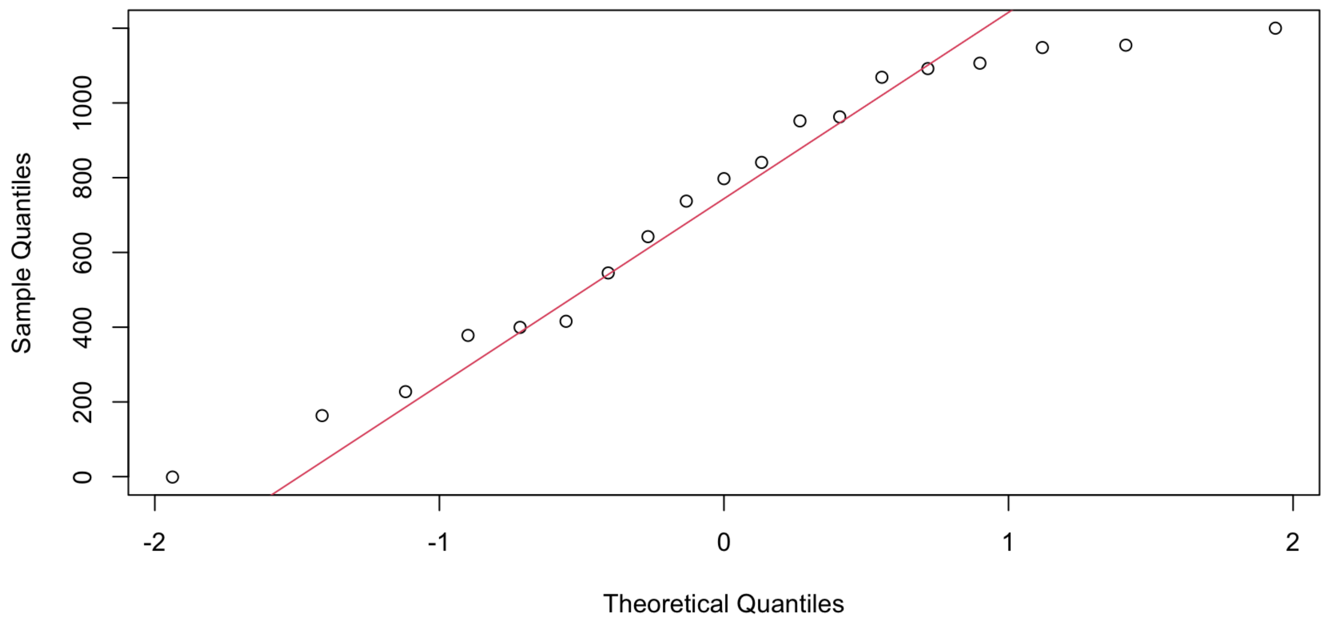

```
lambda <- BC$lambda[which(BC$loglike == max(BC$loglike))]
ice_ts_BC <- (ice_ts2^lambda - 1)/lambda
plot(ice_ts_BC, type='o', ylab='Change in ice mass', main='Figure 5. Time series plot
of BC transformed yearly changes in Antarctica land ice mass')
```

Figure 5. Time series plot of BC transformed yearly changes in Antarctica land ice mass




```
qqnorm(ice_ts_BC, main='Figure 6. QQ plot of the BC transformed series')
qqline(ice_ts_BC, col=2)
```

Figure 6. QQ plot of the BC transformed series




```
shapiro.test(ice_ts_BC)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  ice_ts_BC
## W = 0.92582, p-value = 0.145
```

The Augmented Dickey-Fuller Test is performed and the p-value is 0.3004. We cannot reject the non-hypothesis that the series is non-stationary at 5% significance level. Differencing is needed.

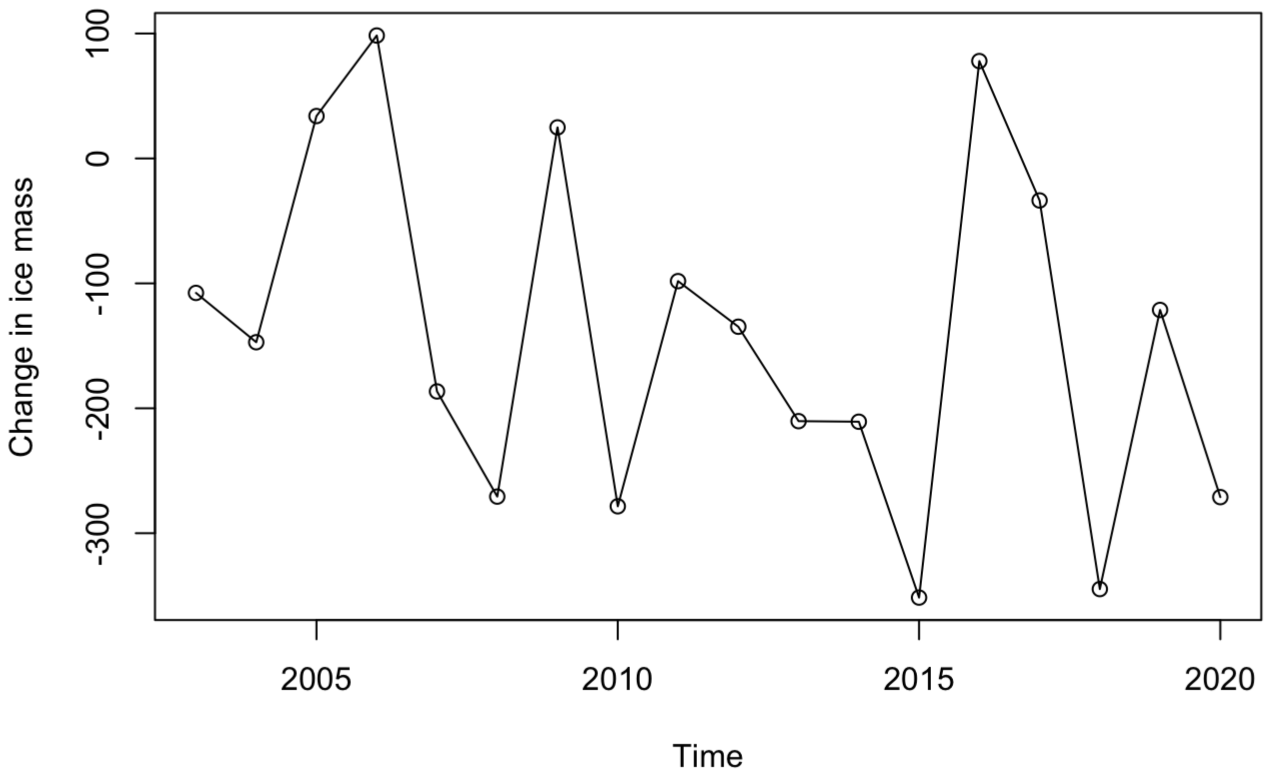
```
adf.test(ice_ts)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  ice_ts
## Dickey-Fuller = -2.7141, Lag order = 2, p-value = 0.3004
## alternative hypothesis: stationary
```

The differenced time series is plotted in figure 7. It seems that the downward trend still exists. The p value from the ADF test is 0.3566. We still cannot reject the non-hypothesis that the series is non-stationary. Second differencing is needed.

```
ice_ts_diff <- diff(ice_ts)
plot(ice_ts_diff, type='o', ylab='Change in ice mass', main='Figure 7. Time series pl
ot of the first differenced yearly changes')
```

Figure 7. Time series plot of the first differenced yearly changes



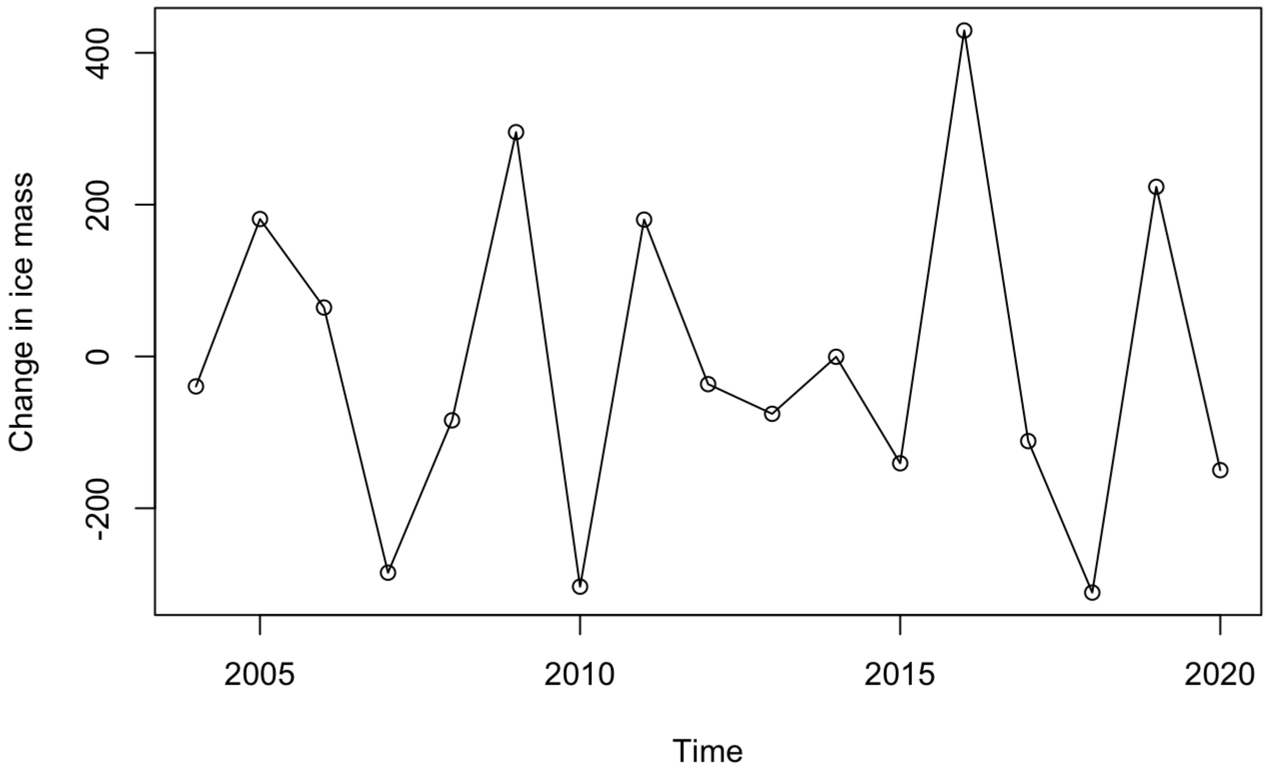

```
adf.test(ice_ts_diff)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  ice_ts_diff
## Dickey-Fuller = -2.5664, Lag order = 2, p-value = 0.3566
## alternative hypothesis: stationary
```

The second differenced time series is plotted in figure 8 and the trend seems removed. The p value from ADF test is 0.0725. The series is still non-stationary at 5% significance level, but the p value is quite close to the threshold. We will perform PP test and have a look at the ACF and PACF.

```
ice_ts_diff2 <- diff(ice_ts, differences = 2)
plot(ice_ts_diff2, type='o', ylab='Change in ice mass', main='Figure 8. Time series p
lot of the second differenced yearly changes')
```

Figure 8. Time series plot of the second differenced yearly changes



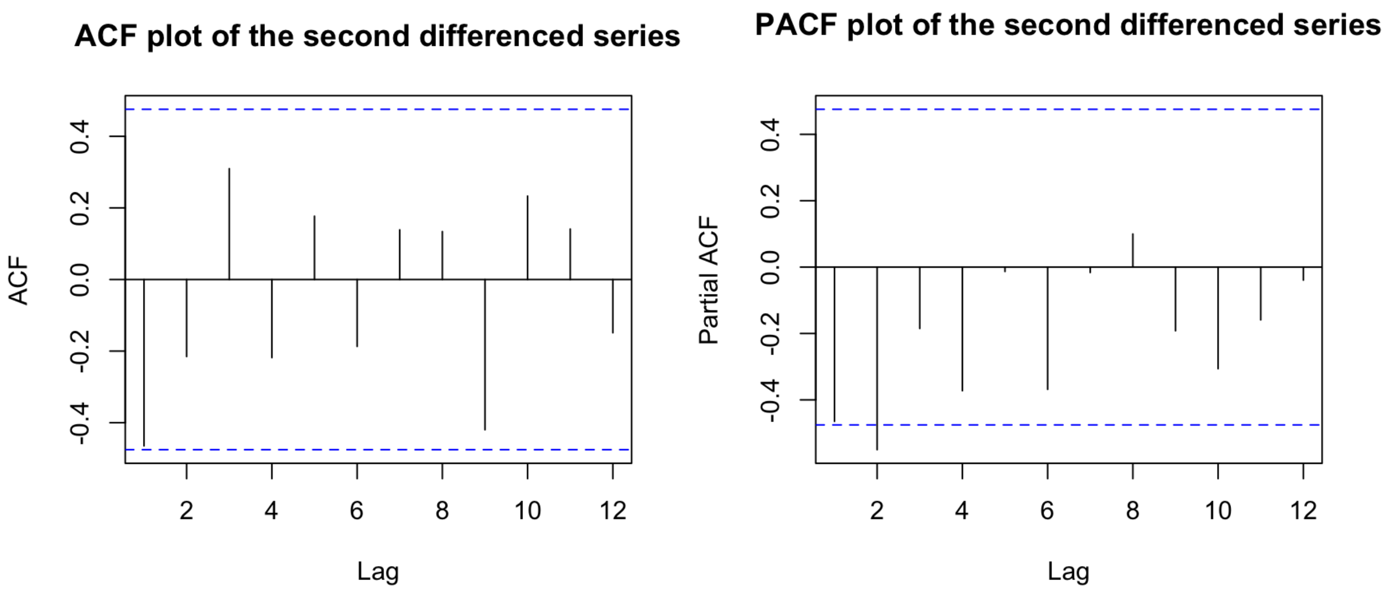

```
adf.test(ice_ts_diff2)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  ice_ts_diff2
## Dickey-Fuller = -3.438, Lag order = 2, p-value = 0.0725
## alternative hypothesis: stationary
```

The ACF and PACF plots are displayed in figure 9. It implies that there is no significant autocorrelation for all lags and partial autocorrelation might exist for lag 2. The p value from the PP test is 0.02604 (less than 0.05). We can conclude that the second differenced series is stationary.

```
par(mfrow=c(1,2), mar=c(5,4,4,2), oma=c(0,0,4,0))
acf(ice_ts_diff2, main='ACF plot of the second differenced series')
pacf(ice_ts_diff2, main='PACF plot of the second differenced series')
par(mfrow=c(1,1))
mtext('Figure 9. ACF-PACF analysis', line=2, cex=2, side=3, outer=TRUE)
```

Figure 9. ACF-PACF analysis




```
pp.test(ice_ts_diff2)
```

```
##
##  Phillips-Perron Unit Root Test
##
## data:  ice_ts_diff2
## Dickey-Fuller Z(alpha) = -19.817, Truncation lag parameter = 2, p-value
## = 0.02604
## alternative hypothesis: stationary
```

Model specification

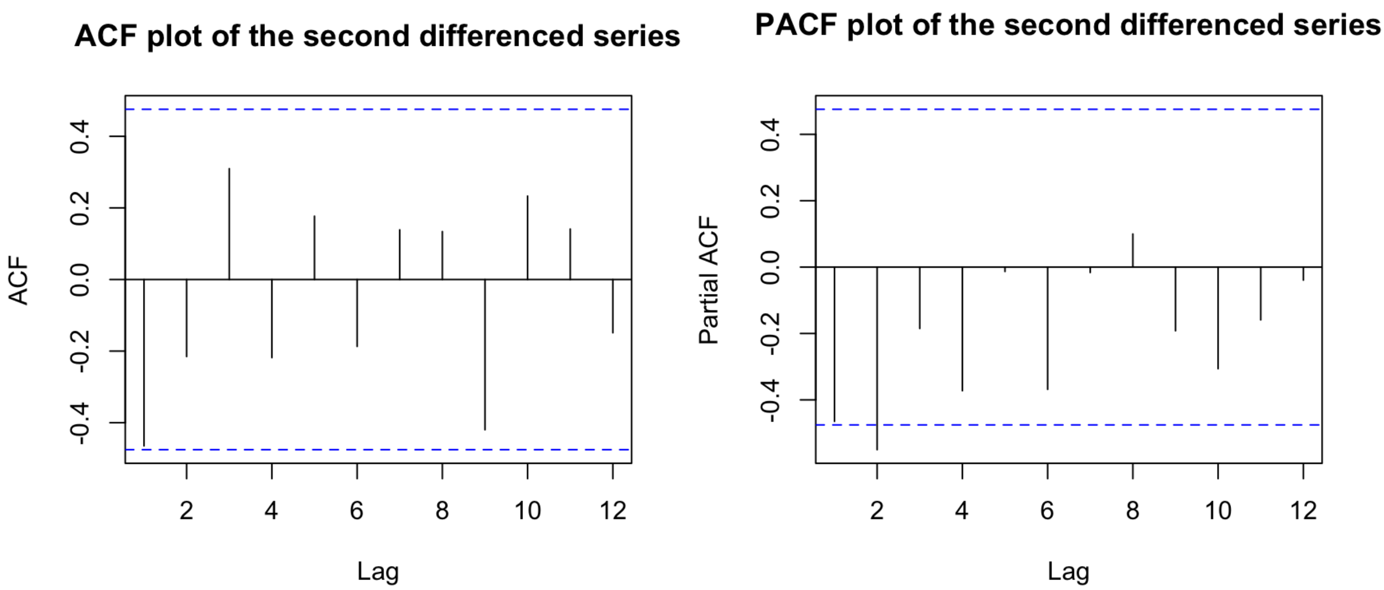
In this part, we will specify the ARIMA(p, d, q) by using ACF-PACF plots, EACF plot, and BIC table. We already know from the previous analysis that the time series of the yearly changes in Antarctic land ice mass can achieved stationary by second differencing. We can make the conclusion that d=2. We will try to determine a set of p and q in the following section.

ACF and PACF plots

We can refer back to figure 9. The ACF plot indicate that there is not significant autocorrelation for all lags. The autocorrelation for lag 1 is quite close to the confidence interval. The MA process might have an order of 0 or 1. The PACF plot indicate that the autocorrelation for lag 2 is significant. The AR process might have an order of 2. The proposed set of possible models is {ARIMA(2,2,0), ARIMA(2,2,1)}

```
par(mfrow=c(1,2), mar=c(5,4,4,2), oma=c(0,0,4,0))
acf(ice_ts_diff2, main='ACF plot of the second differenced series')
pacf(ice_ts_diff2, main='PACF plot of the second differenced series')
par(mfrow=c(1,1))
mtext('Figure 9. ACF-PACF analysis', line=2, cex=2, side=3, outer=TRUE)
```

Figure 9. ACF-PACF analysis



EACF plot

The possible top-left ’o’s om tje EACF plot appear at (0,0) and (2,0), implying the possible set of ARIMA model to be {ARIMA(0,2,0), ARIMA(2,2,0)}

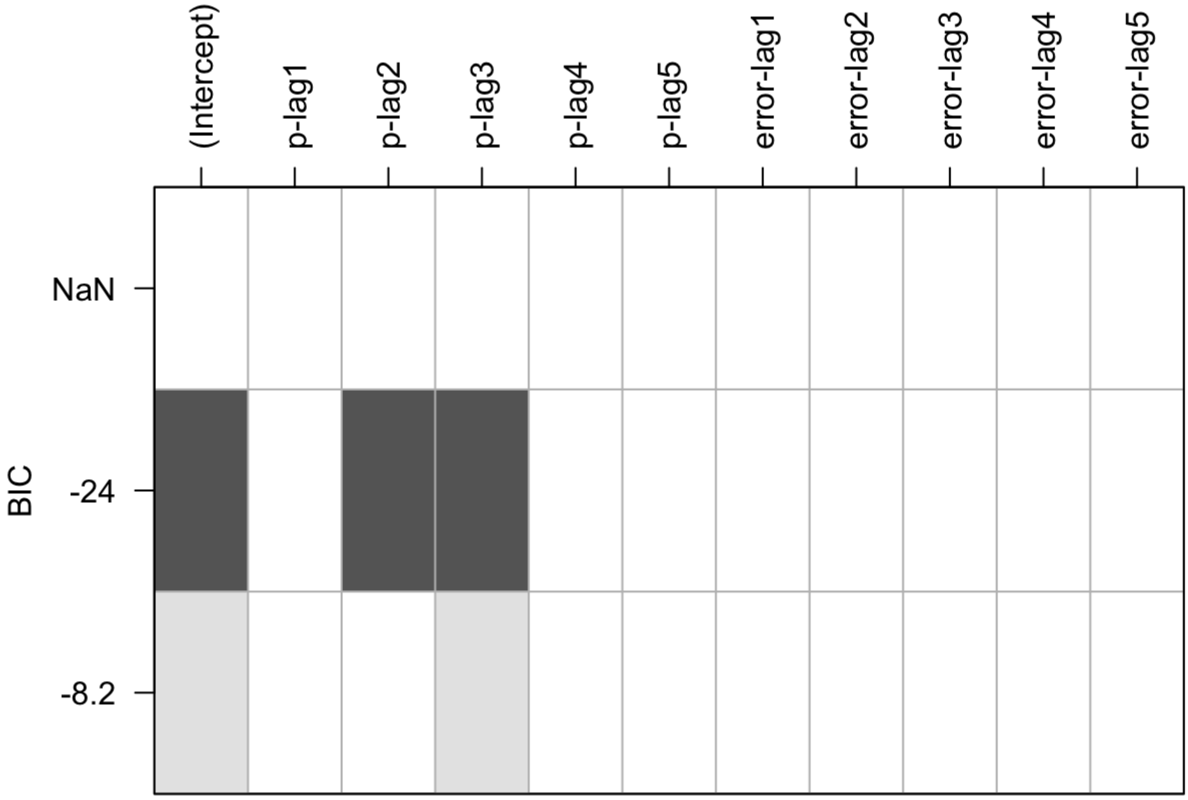
```
eacf(ice_ts_diff2, ar.max=3, ma.max=3)
```

```
## AR/MA
##   0 1 2 3
## 0 o o o o
## 1 x x o o
## 2 o x o o
## 3 o o o o
```

BIC table

The BIC table indicates that the subset ARMA(5,5) model with smallest BIC contains only lag 2 and lag 3 for AR process and no log for MA process, so the possible set of ARIMA model is {ARIMA(2,2,0), ARIMA(3,2,0)}

```
res <- armasubsets(y=ice_ts_diff2,nar=5,nma=5,y.name='p',ar.method='ols')
plot(res)
```



Conclusion

In conclusion, the time series of yearly changes in Antarctic land ice mass can achieve stationary by second differencing. By applying the model specification tools such as ACF-PACF plots, EAFC plot, and BIC table, we conclude and propose to use following set of models: {ARIMA(2,2,0), ARIMA(2,2,1), ARIMA(0,2,0), ARIMA(3,2,0)}