Model Specification for Yearly Changes in Antarctica Land Ice Mass Data

Ting Li

s3912985

Introduction

Studying changes in Antarctica land ice mass is crucial in monitoring the impact of climate change such as the rising sea levels (EPA United States Environmental Protection Service, 2021). Therefore, it is important to identify appropriate models from studying the data to forecast future yearly changes in Antarctica land ice mass, such as the Auto Regressive Integrated Moving Average (ARIMA) model (Cryer & Chan, 2008). In this report, we demonstrate the process of specifying possible ARIMA models for the yearly changes in Antarctica land ice mass dataset through descriptive analysis, data transformation, differencing, statistical tests, and model specification tools.

Methods

Data

The dataset used in this study is the yearly changes of Antarctica land ice mass from years 2002 to 2020 compared to a baseline land ice mass level in 2001. The original data file contains two quantitative variables and 19 observations. Variable "Year" records the year of observation while "NASA...Annual.Antarctica.land.ice.mass" records the change in ice mass in that year, which we was renamed to "Changes" in this analysis.

Procedure

The TSA package and tseries package in R were used to investigate the dataset. Data is imported with the read.csv() function and created as time-series object. The plot() and summary() function were used to conduct descriptive analysis time-series data, with new functions created to avoid repeating R codes. The BoxCox.ar() function was used to transform raw time series, as an attempt to improve normality and stationarity. To check normality and correlation of neighbouring measurement for the raw and transformed time series, a plot and check function was created with plot(), qqnorm(), qqplot(), hist(), and shapiro.test(), etc. The acf() and pacf() functions are combined in an apacf() function to check for stationarity and significant lags. The diff() function was used to apply differencing to obtain a stationary time series, and to determine the d value for the ARIMA model. The stats.test() function was created to conduct Adf.test(), pp.test() and kpss.test(), to confirm stationarity of the time series data after differencing. For model specification, the eacf() and bic() functions were used to find possible p and q values of the ARIMA model.

Results and Discussion

Raw time series data

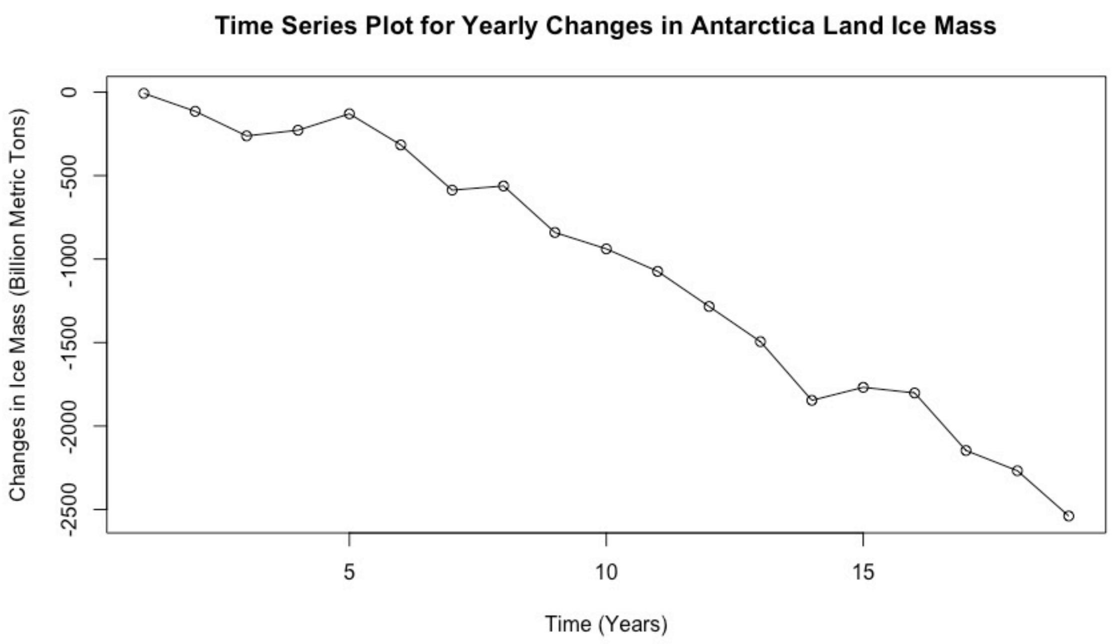

Figure 1. Time series plot for yearly changes in Antarctica land ice mass

Table 1 Summary statistics of raw ICEts

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-2539.055	-1785.137	-938.991	-1063.774	-289.390	-7.547

In the yearly changes in Antarctica land ice mass, there seems to be a negative trend over the 19 years of observation, suggesting a non-stationary time series data (Demirhan, 2022). There seems to be no evidence for changing variance or fluctuations over the given time period, with no indication of seasonality. There is hint of autocorrelation behaviour but no evidence of intervention in the observation. Summary statistics suggests a range from -2539.055 to -7.547 for the yearly decrease in ice mass. There is a possibility of a negatively-skewed data distribution given the mean is less than the median, so further investigation into the normality of data is needed (Baglin, 2016).

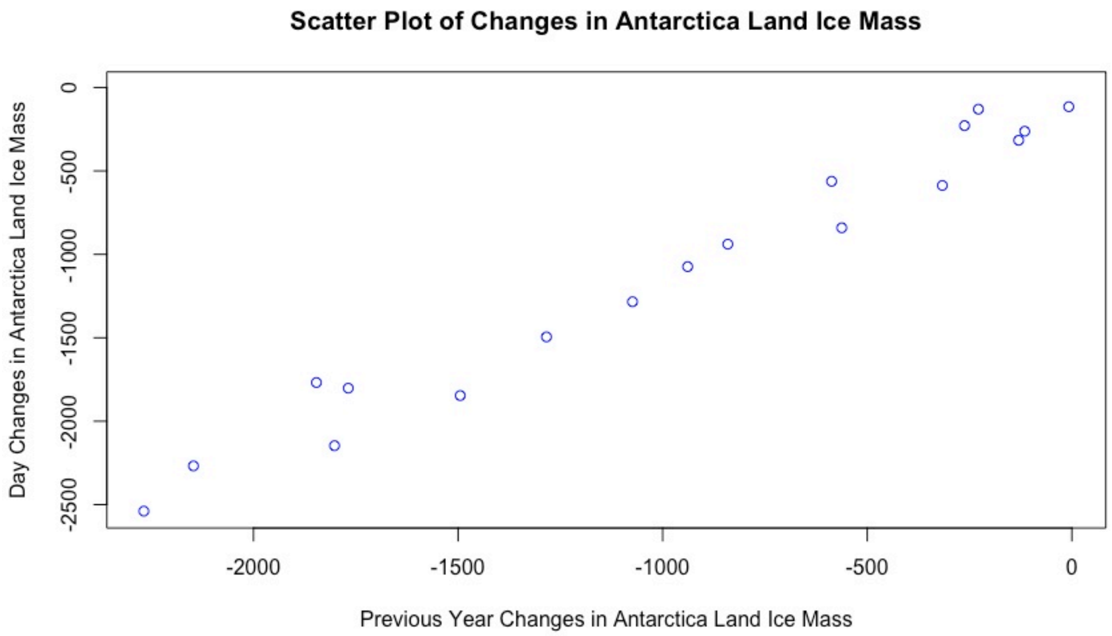

Figure 2. Scatter plot of changes in Antarctica land ice mass (ICEts data)

```
[1] "Correlation Index:"
[1] 0.985486

        Shapiro-Wilk normality test

data:  ts
W = 0.92616, p-value = 0.1471
```

Figure 3. Correlation index and Shapiro-Wilk test for ICEts data

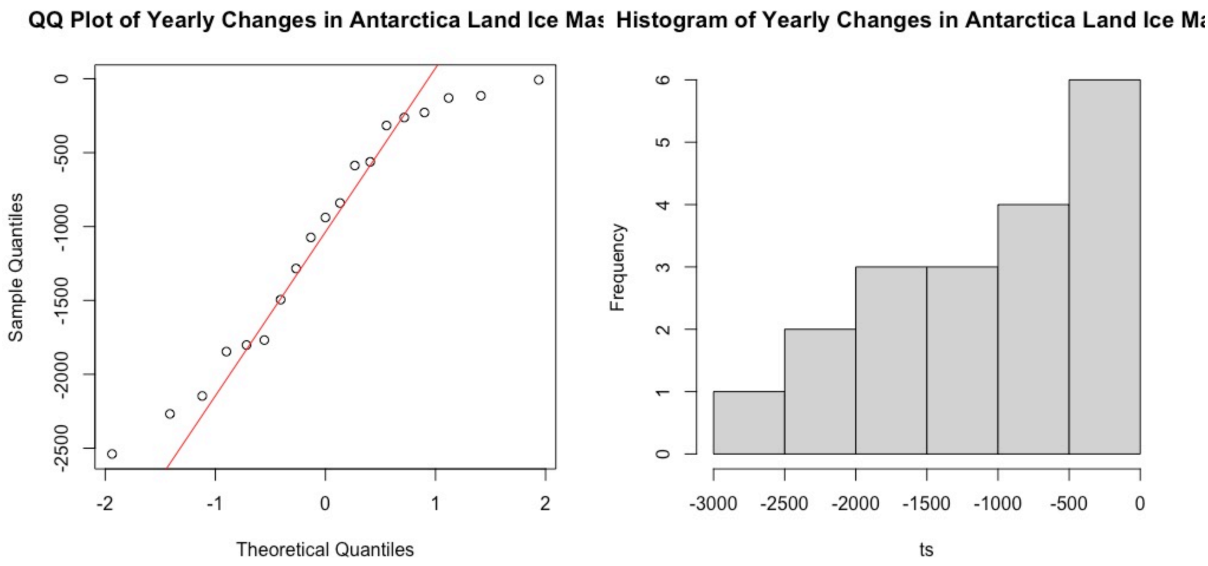

Figure 4. QQ plot and histogram of ICEts data

According to Figure 2 and Figure 3, neighbouring measurements are correlated with 0.985 correlation index. From the Shapiro-Wilk test result in Figure 3, we conclude normality for the raw time series data of yearly changes in Antarctica land ice mass, since we fail to reject the null hypothesis for normality with a p-value of 0.147 (Baglin, 2016). However, the data distribution demonstrates negative skewness with a QQ plot tailing more on the higher end and a negatively-skewed histogram, so we could attempt transformation to improve normality.

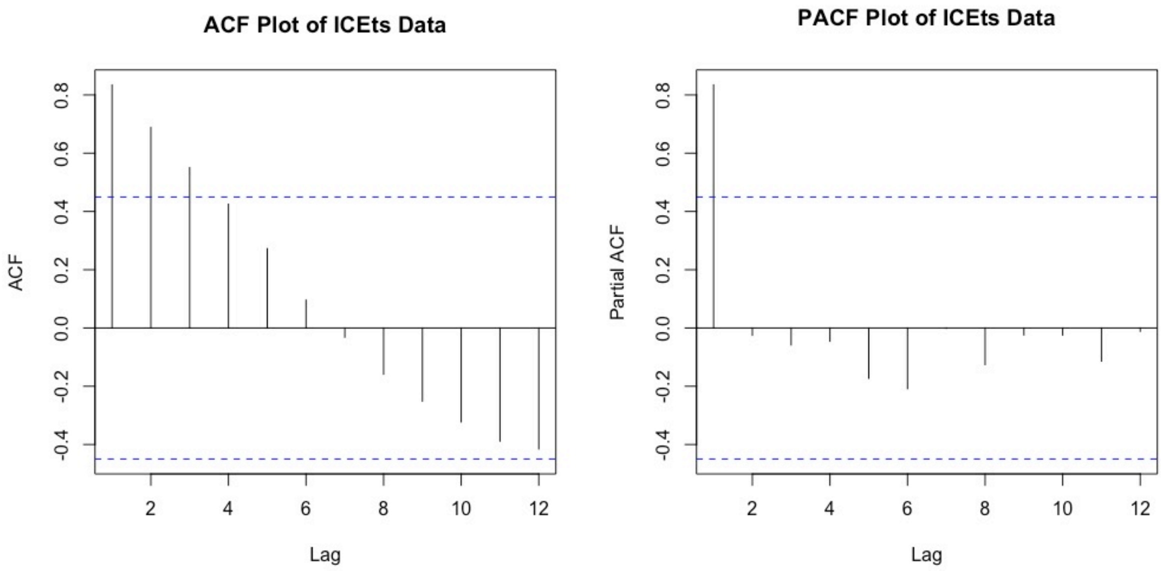

Figure 5. ACF and PACF plot of raw time series data (ICEts)

Table 2. Statistical test result of raw time series data (ICEts)

	Adf.test	pp.test	Kpss.test
p-value	0.3004	0.7516	0.01098

With a slowly decaying pattern in the ACF plot and a very high first lag in the PACF plot, Figure 5 indicates that the data is non-stationary (Demirhan, 2022). Moreover, from Table 2, p-values greater than 0.05 from the ADF test and the PP test suggests that we fail to reject the null hypothesis of non-stationarity, whilst a p-value less than 0.05 in the KPSS test suggests that we reject the null hypothesis of stationarity (Demirhan, 2022). Hence, our observation of non-stationarity in the raw time series data is supported by results in from the ACF plot, PACF plot in Figure 5 and statistical tests in Table 2, and transformation could be attempted to improve stationarity.

Transformation

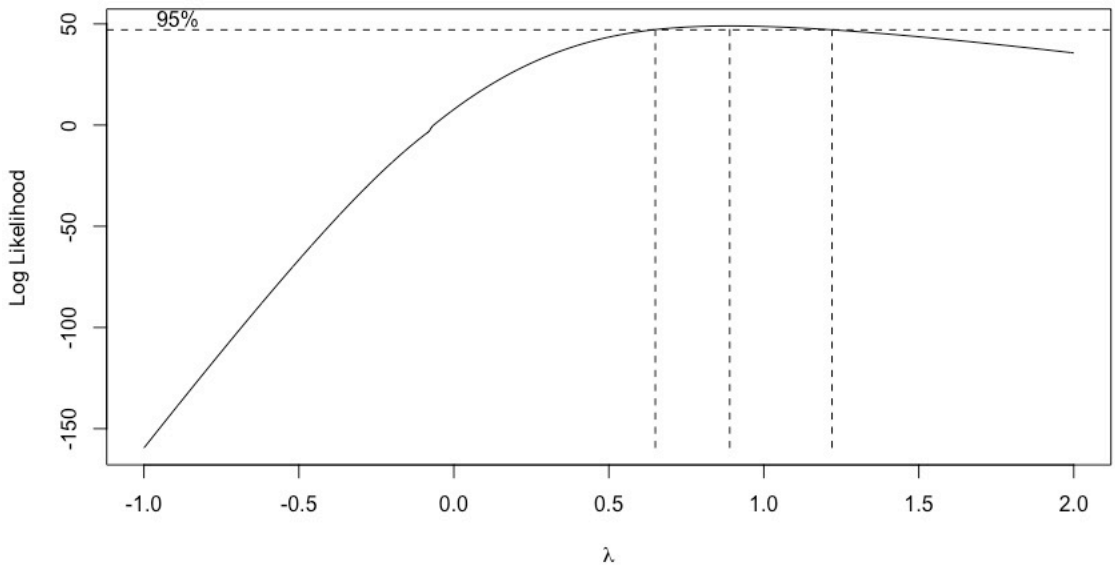

Figure 6. Lambda (middle vertical line) and confidence interval for lambda (left and right vertical line)

```
> BC$ci
[1] 0.65 1.22
> #find the middle vertical line -> lambda value
>
> lambda <- BC$lambda[which(max(BC$loglike)==BC$loglike)]
> lambda
[1] 0.89
```

Figure 7. Values for lambda and confidence interval for lambda

The Box-Cox transformation was applied as an attempt to improve stationarity and normality of the non-stationary raw time series data for yearly changes in Antarctica land ice mass

(Demirhan, 2022). Since all data values are negative in the dataset, a constant was added to make all values positive, in order to proceed with the Box-Cox transformation (see Appendix). After proceeding, the confidence interval for lambda was found to be [0.65, 1.22] with a lambda value of 0.89. Since the confidence interval for lambda includes 1, it is necessary to check shape of data distribution after transformation to see if it improves normality (Baglin, 2016).

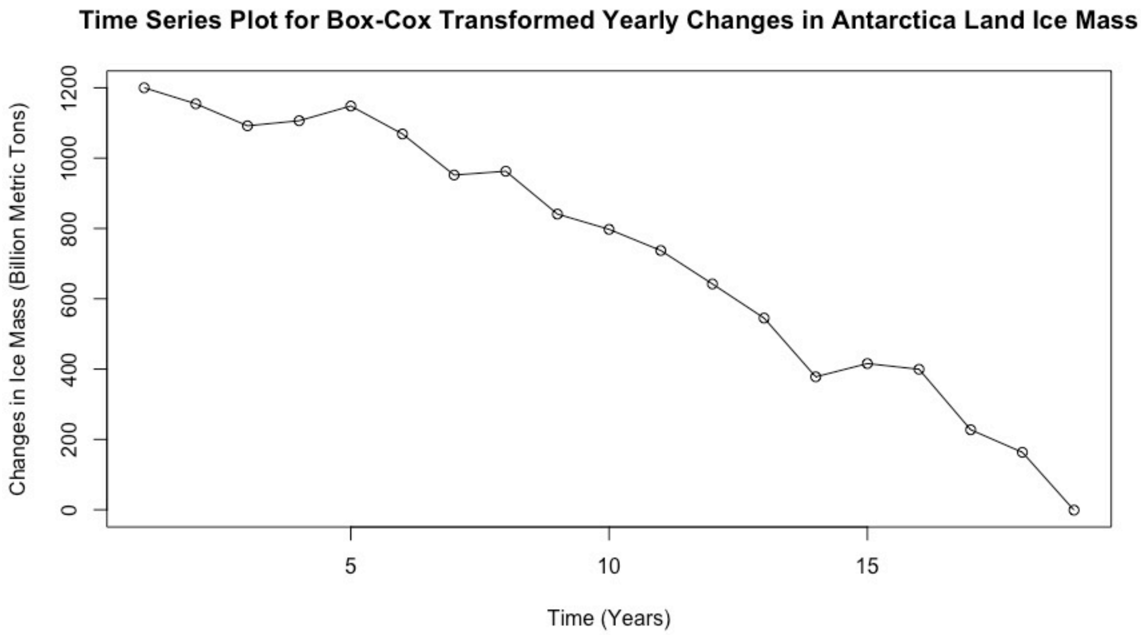

Figure 8. Time series plot of the Box-Cox transformed time series data (ICEtsBC)

The Box-Cox transformed data of yearly changes in Antarctica land ice mass seems to demonstrate similar characteristics to the raw time series data before transformation. There seems to be no evident change to the downward trend and autocorrelation behaviour, with no hint of changing variance, seasonality and intervention.

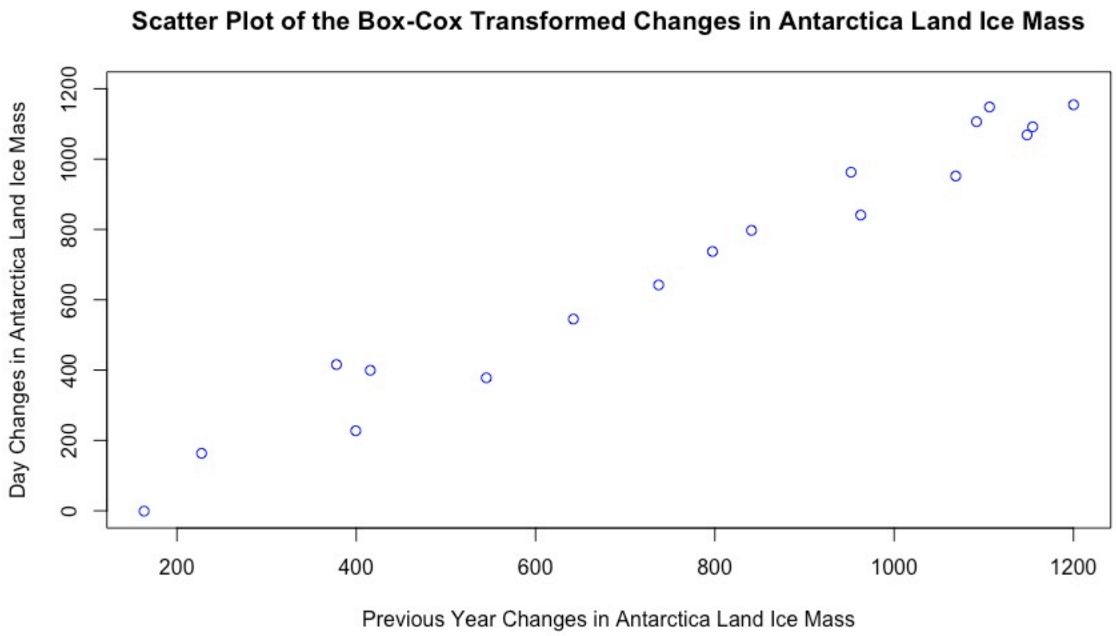

Figure 9. Scatter plot of the Box-Cox transformed time series data (ICEtsBC)

```
[1] "Correlation Index:"
[1] 0.9850052

        Shapiro-Wilk normality test

data:  ts
W = 0.92582, p-value = 0.145
```

Figure 10. Correlation index and Shapiro-Wilk test for ICEtsBC data

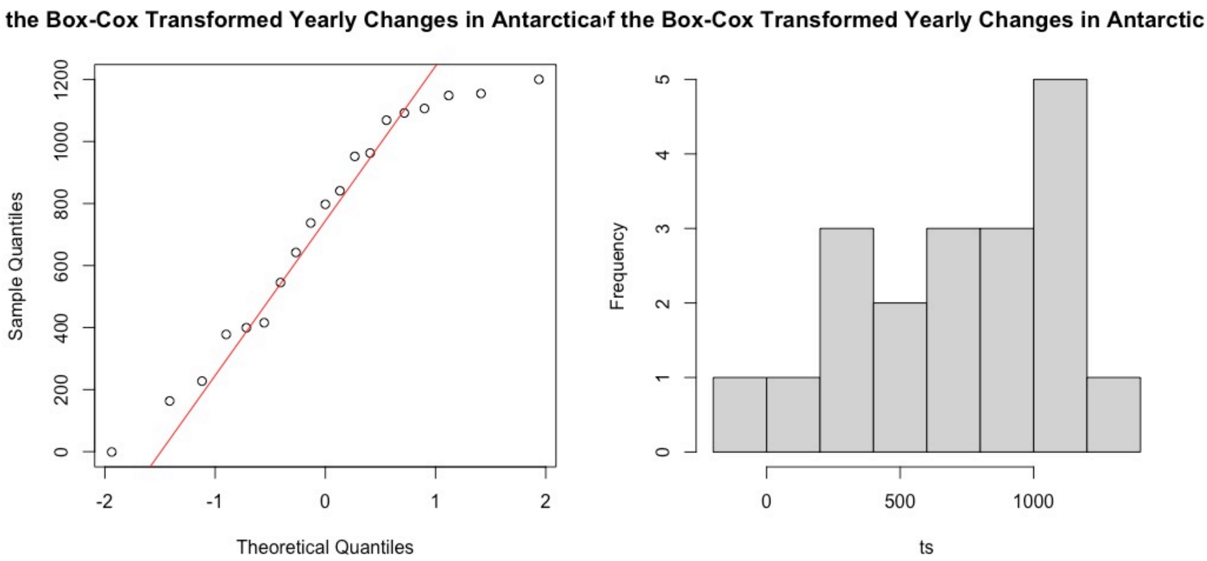

Figure 11. QQ plot and histogram of Box-Cox transformed time series data (ICEtsBC)

According to Figure 9 and Figure 10, similar to the raw time series data, neighbouring measurements of the Box-Cox transformed data are correlated with 0.985 correlation index, and Shapiro-Wilk test returns a p-value of 0.145, indicating normality with a similar level of statistical significance (Baglin, 2016). Figure 11 suggests that the shape of data distribution has not been evidently improved from the Box-Cox transformation with existence of negative-skewness.

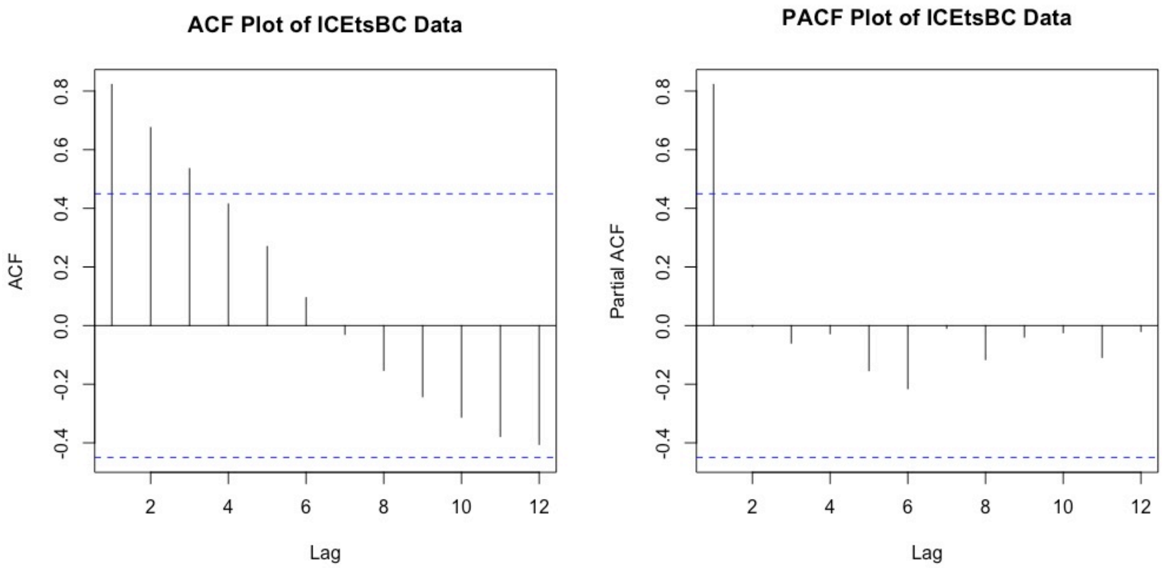

Figure 12. ACF and PACF plot of Box-Cox transformed time series data (ICEtsBC)

Table 3. Statistical test result of Box-Cox transformed time series data (ICEtsBC)

	Adf.test	pp.test	Kpss.test
p-value	0.5634	0.868	0.01108

In Figure 12, we observe a slowly decaying pattern in the ACF plot and a very high first lag in the PACF plot, indicating non-stationarity in the Box-Cox transformed data (Demirhan, 2022). The same assumption is supported by p-values in Table 3, as we fail to reject the null hypothesis of non-stationarity for the Box-Cox transformed data with the ADF test and the PP test p-values greater than 0.05 (Demirhan, 2022). A 0.01 p-value from the KPSS test also suggests that we reject the null hypothesis of stationarity. Therefore, we conclude that Box-Cox transformation has not improved normality and stationarity of the data, and that the raw time series data (ICEts) will be used to proceed with differencing and model specification.

Differencing

Differencing calculates the differences from one observation to another (Demirhan, 2022). It is applied to stabilize the mean and reduce trend, therefore changes the non-stationary raw time series data to stationarity, in order to proceed with model specification tools later (Demirhan, 2022). It starts with the first difference (d=1) of the time series data and stops when stationarity is confirmed, to avoid over-differencing (Demirhan, 2022).

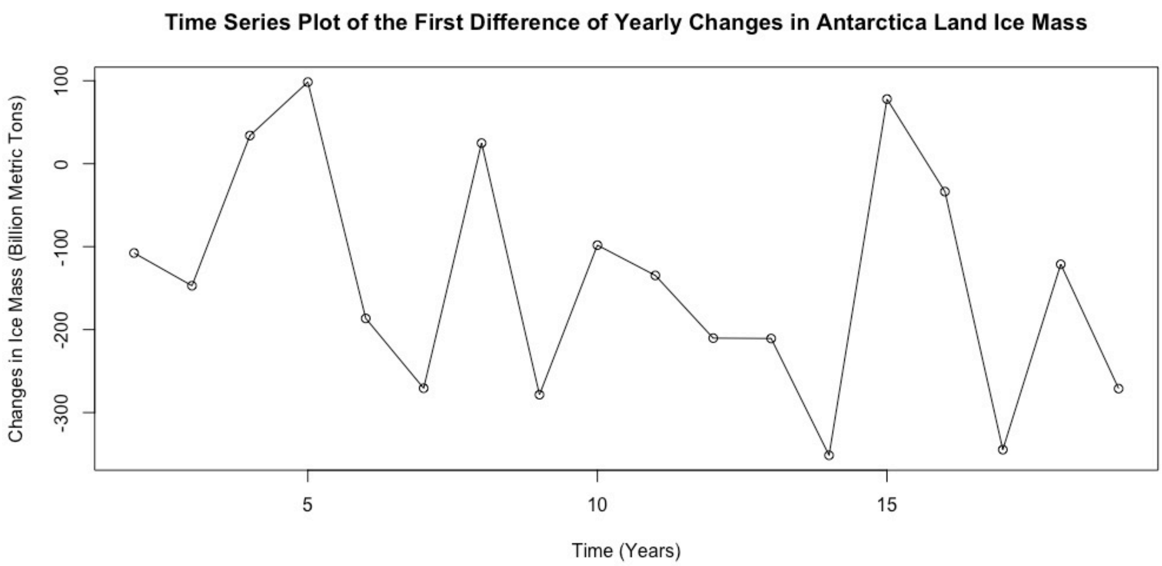

Figure 13. Time series plot of the first difference of yearly changes in Antarctica land ice mass data (ICEtsdif1)

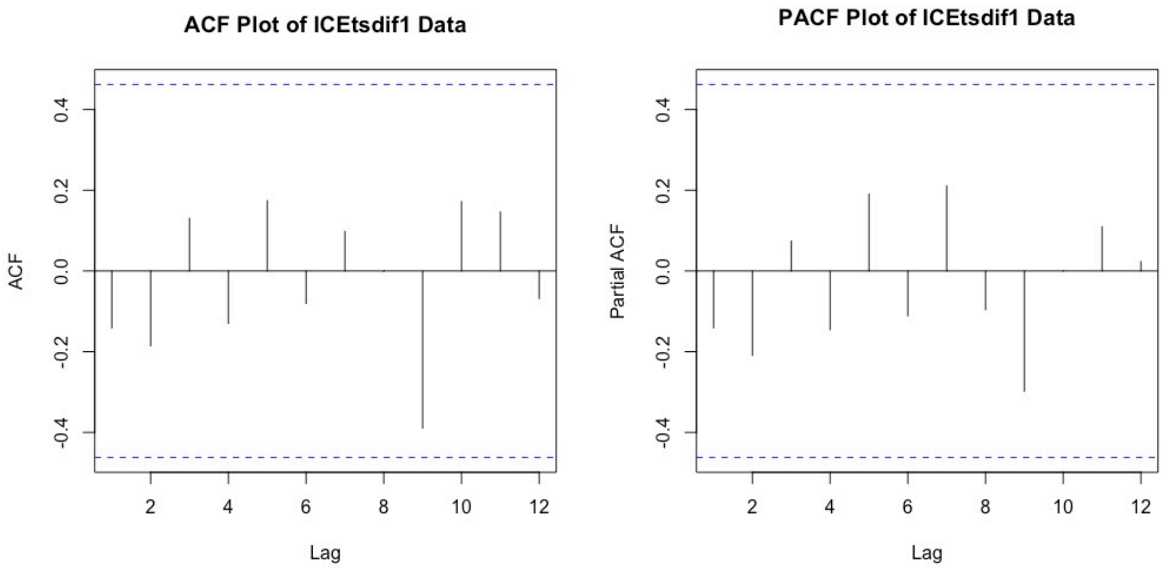

Figure 14. . ACF and PACF plot of the first difference of raw time series data (ICEtsdif1)

Figure 13 is the time series plot of the first difference of yearly changes in Antarctica land ice mass data (ICEtsdif1). The mean seems stable and there is no hint of trend in Figure 13. From Figure 14, no slowly decaying pattern was observed in the ACF plot, and the first lag in the PACF plot is not significantly high. Plots in Figure 13 and Figure 14 suggests that the time series data has become stationary after the first difference, and statistical tests are used to confirm the assumption for stationarity.

Table 4. Statistical test result of the first difference of raw time series data (ICEtsdif1)

	Adf.test	pp.test	Kpss.test
p-value	0.3566	0.04887	0.1

Table 4 shows statistical test results of the first difference of yearly changes in Antarctica land ice mass data. Although the p-value of 0.35 from ADF test suggests non-stationarity in the data after the first difference, results from the PP test and KPSS test support our observation of stationarity after the first difference. Since the default k value in the ADF test in Table 4 is 2, additional ADF tests were conducted with neighbouring k values to test for stationarity on lag 1 and 3, to see if they support the assumption of stationarity (Demirhan, 2022). Results of additional ADF tests below (Figure 15) suggest that the time series is stationary after the first difference, with a p-value of 0.01 when k=1.

```
> adf.test(ICEtsdif1, k=1)

	Augmented Dickey-Fuller Test

data:  ICEtsdif1
Dickey-Fuller = -4.2788, Lag order = 1, p-value = 0.01353
alternative hypothesis: stationary

> adf.test(ICEtsdif1, k=3)

	Augmented Dickey-Fuller Test

data:  ICEtsdif1
Dickey-Fuller = -2.9425, Lag order = 3, p-value = 0.2133
alternative hypothesis: stationary
```

Figure 15. Additional ADF test of the first difference of raw time series data (ICEtsdif1) with different k values

Although the ADF test result with default k value suggests non-stationarity, stationarity can be concluded for the yearly changes in Antarctica land ice mass data after the first difference from reading the time series plot, ACF plot, PACF plot, PP test, KPSS test and additional

ADF test (Demirhan, 2022). Since the value of d in the ARIMA model is determined by the order of differencing, we conclude that d=1 for all possible ARIMA models for this dataset. From reading ACF and PACF plot (Figure 14) we find p=0 and q=0, suggesting a random walk model of ARIMA(0,1,0). To specify other possible values of p and q, we proceed to model specification with the first difference of the raw time series data (ICEtsdif1).

Model Specification

Apart from reading the ACF and PACF plot in the previous section, the EACF table and BIC table are also used as tools to specify possible values of p and q for the ARIMA model (Demirhan, 2022). In the EACF table, we look for the top left 0 that is not distracted by x and its neighbouring models to find possible values for p from rows, and q from columns (Demirhan, 2022). The BIC table displays models in rows with possible values for p from p-lags and q from error-lags (Demirhan, 2022). We read the BIC table from the top row down since the top row is the best fitted model, and look for p and q values with darker colours. Maximum ar and ma orders are set to 3 (see Appendix) given the small size of data, to avoid unnecessary over fitting with large p and q values (Demirhan, 2022).

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

Figure 16. EACF table of first difference of raw time series data (ICEtsdif1)

Figure 16 shows the EACF table of first difference of raw time series data (ICEtsdif1). From reading the top left 0 and its neighbouring models, we find possible models of ARIMA(0,1,1), ARIMA(1,1,0), and ARIMA(1,1,1).

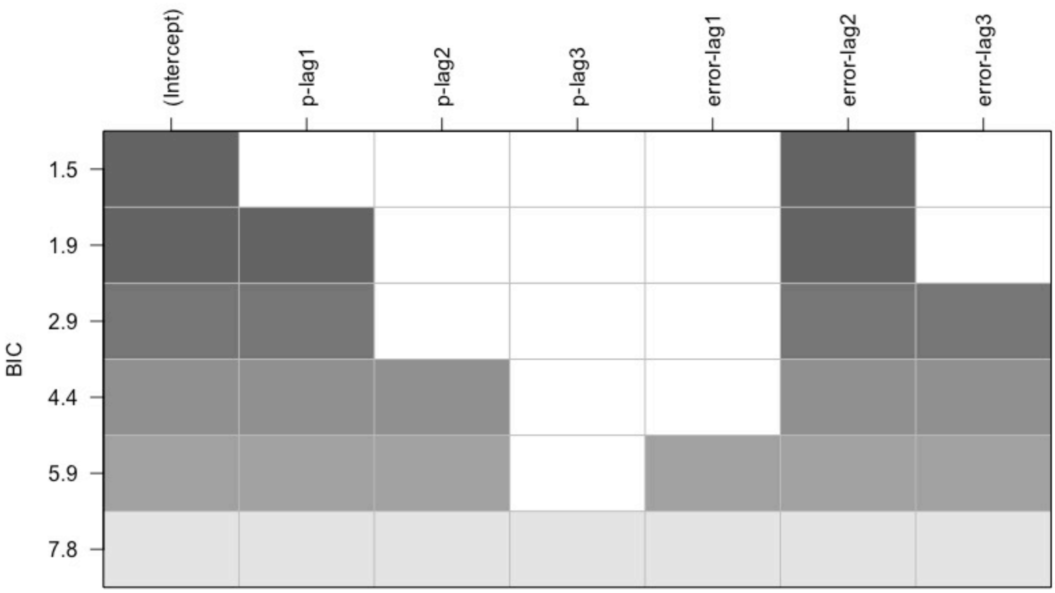

Figure 17. BIC table table of first difference of raw time series data (ICEtsdif1)

Figure 17 shows the BIC table of first difference of raw time series data (ICEtsdif1). We found ARIMA (0,1,2) from the best fitted model in the first row, and ARIMA (1,1,2) from the second best fitted model in the second row. Although q=2 is different from the q values we found in the EACF table, it is supported by all models fitted in the BIC table therefore will be included.

Overall, by reading ACF and PACF plot, EACF table, and BIC table, we found ARIMA(0,1,0), ARIMA(0,1,1), ARIMA(1,1,0), ARIMA(1,1,1), ARIMA (0,1,2), and ARIMA (1,1,2) as possible models for the yearly changes in Antarctica land ice mass data.

Conclusion

In conclusion, through the process of descriptive analysis, data transformation, differencing, statistical tests, and model specification, ARIMA(0,1,0), ARIMA(0,1,1), ARIMA(1,1,0), ARIMA(1,1,1), ARIMA (0,1,2), and ARIMA (1,1,2) were specified as possible models for forecasting the yearly changes in Antarctica land ice mass data. In further investigations, coefficient test and residual analysis could be performed to specify the best fitted model for forecasting.

References

Baglin, J. (2016). *Applied Analytics*. https://astral-theory-157510.appspot.com/secured/index.html

Cryer, J.D., & Chan, K.S. (2008). *Time Series Analysis with Applications in R (second edition)*. Springer. https://link-springer-com.ezproxy.lib.rmit.edu.au/book/10.1007/978-0-387-75959-3

EPA United States Environmental Protection Service (2021). *Climate Change Indicators: Ice Sheets.* https://www.epa.gov/climate-indicators/climate-change-indicators-ice-sheets

Demirhan, H. (2022, April 19). *MATH1318/MATH2204 QA Sessions/Online Versions/Module 1* [Lecture recording]. Canvas@RMIT University. https://rmit.instructure.com/courses/90825/external_tools/546

Demirhan, H. (2022, April 19). *MATH1318/MATH2204 QA Sessions/Online Versions/Module 1: Solution of practice tasks* [Lecture recording]. Canvas@RMIT University. https://rmit.instructure.com/courses/90825/external_tools/546

Demirhan, H. (2022, April 19). *MATH1318/MATH2204 QA Sessions/Online Versions/Module 3: Part 1* [Lecture recording]. Canvas@RMIT University. https://rmit.instructure.com/courses/90825/external_tools/546

Demirhan, H. (2022, April 19). *MATH1318/MATH2204 QA Sessions/Online Versions/Module 3: Part 2* [Lecture recording]. Canvas@RMIT University. https://rmit.instructure.com/courses/90825/external_tools/546

Demirhan, H. (2022, April 19). *MATH1318/MATH2204 QA Sessions/Online Versions/Module 3: Solution of practice tasks* [Lecture recording]. Canvas@RMIT University. https://rmit.instructure.com/courses/90825/external_tools/546

Demirhan, H. (2022, April 19). *MATH1318/MATH2204 QA Sessions/Online Versions/Module 4: Part 1* [Lecture recording]. Canvas@RMIT University. https://rmit.instructure.com/courses/90825/external_tools/546

Demirhan, H. (2022, April 19). *MATH1318/MATH2204 QA Sessions/Online Versions/Module 4: Part 2* [Lecture recording]. Canvas@RMIT University. https://rmit.instructure.com/courses/90825/external_tools/546

Demirhan, H. (2022, April 19). *MATH1318/MATH2204 QA Sessions/Online Versions/Module 4: Solution of practice tasks* [Lecture recording]. Canvas@RMIT University. https://rmit.instructure.com/courses/90825/external_tools/546

Demirhan, H. (2022, April 19). *MATH1318/MATH2204 QA Sessions/Online Versions/Module 5: Part 1* [Lecture recording]. Canvas@RMIT University. https://rmit.instructure.com/courses/90825/external_tools/546

Demirhan, H. (2022, April 19). *MATH1318/MATH2204 QA Sessions/Online Versions/Module 5: Part 2* [Lecture recording]. Canvas@RMIT University. https://rmit.instructure.com/courses/90825/external_tools/546

Demirhan, H. (2022, April 19). *MATH1318/MATH2204 QA Sessions/Online Versions/Module 5: Solution of practice tasks* – Part 1 [Lecture recording]. Canvas@RMIT University. https://rmit.instructure.com/courses/90825/external_tools/546

Demirhan, H. (2022, April 19). *MATH1318/MATH2204 QA Sessions/Online Versions/Module 5: Solution of practice tasks* – Part 2 [Lecture recording]. Canvas@RMIT University. https://rmit.instructure.com/courses/90825/external_tools/546

Appendix

```
#Assignment 2 model specification for ICE data

rm (list=ls())

library(TSA)

library(tseries)

#1. load data

setwd("/Users/momo/Documents/Master of Analytics/SEM1 2022/Time Series Analysis")

ICE <- read.csv("assignment2Data2022.csv")

#2. check data structure and rename variables

str(ICE)

names(ICE)[2] <- "Changes"

colnames(ICE)

#3. create ts object

ICEts <- ts(data=ICE$Changes, start = "1", frequency = 1)

#4. descriptive analysis

#4.1 data summary

summary(ICEts)

#4.2 plot time series data
```

```
plot(ICEts, ylab='Changes in Ice Mass (Billion Metric Tons)',xlab="Time (Years)", type =
"o",
     main = "Time Series Plot for Yearly Changes in Antarctica Land Ice Mass")

plotandcheck <- function(ts,  scatter_plot_title, qq_plot_title, hist_title){
   #scatter plot and check correlation of of neighbouring measurement
   plot(y=ts,x=zlag(ts),col = "blue", xlab = "Previous Year Changes in Antarctica Land Ice
Mass",
        ylab = "Day Changes in Antarctica Land Ice Mass", main = scatter_plot_title)


   y = ts
   x = zlag(ts)
   index = 2:length(x)
   print('Correlation Index:')
   print(cor(y[index],x[index]))
   #normality check
   par(mfrow=c(1,2))
   qqnorm(ts, main=qq_plot_title)
   qqline(ts, col = "red")
```

```
   hist(ts, main = hist_title)

   par(mfrow=c(1,1))

   print(shapiro.test(ts))

}

plotandcheck(ICEts,'Scatter Plot of Changes in Antarctica Land Ice Mass',

        'QQ Plot of Yearly Changes in Antarctica Land Ice Mass',

        "Histogram of Yearly Changes in Antarctica Land Ice Mass"

        )

apacf <- function(ts, acf_title, pacf_title){

  par(mfrow=c(1,2))

  acf(ts, main=acf_title)

  pacf(ts, main=pacf_title)

  par(mfrow=c(1,1))

}

apacf(ICEts, "ACF Plot of ICEts Data", "PACF Plot of ICEts Data")



stats.test <- function(ts)

{print(adf.test(ts))
```

```
  print(pp.test(ts))

  print(kpss.test(ts))

}

stats.test(ICEts)

#5.Box-Cox transformation to see if it improves stationarity and normality

#5.1 Box-Cox transformation: find values of the first and third vertical lines

ICEtsP <- ICEts + abs(min(ICEts)) + 0.01

BC = BoxCox.ar(y=ICEtsP, lambda = seq(-1,2,0.01))

BC$ci

#find the middle vertical line -> lambda value

lambda <- BC$lambda[which(max(BC$loglike)==BC$loglike)]

lambda

ICEtsBC <-((ICEtsP^lambda)-1)/lambda #apply Box-Cox transformation with lambda

plot(ICEtsBC, ylab='Changes in Ice Mass (Billion Metric Tons)', xlab="Time (Years)", type
= "o",

     main = "Time Series Plot for Box-Cox Transformed Yearly Changes in Antarctica Land
Ice Mass")

#5.2 plot Box-Cox transformed ts data
```

```
plotandcheck(ICEtsBC, 'Scatter Plot of the Box-Cox Transformed Changes in Antarctica
Land Ice Mass',
        'QQ Plot of the Box-Cox Transformed Yearly Changes in Antarctica Land Ice Mass',
        "Histogram of the Box-Cox Transformed Yearly Changes in Antarctica Land Ice Mass"
        )
apacf(ICEtsBC, "ACF Plot of ICEtsBC Data", "PACF Plot of ICEtsBC Data")
#5.3 statistical test on Box-Cox transformed ts data
stats.test(ICEtsBC)
#6. apply differencing for stationarity
ICEtsdif1=diff(ICEts, differences = 1)
plot(ICEtsdif1, ylab='Changes in Ice Mass (Billion Metric Tons)',xlab="Time (Years)", type
= "o",
     main = "Time Series Plot of the First Difference of Yearly Changes in Antarctica Land Ice
Mass")
apacf(ICEtsdif1, "ACF Plot of ICEtsdif1 Data", "PACF Plot of ICEtsdif1 Data")


stats.test(ICEtsdif1)
adf.test(ICEtsdif1, k=1)
adf.test(ICEtsdif1, k=3)
```

```
#{from acf q=0,from pacf p=0}

#7. model specification

#EACF

eacf(ICEtsdif1, ar.max = 3, ma.max = 3)

#BIC table

res = armasubsets(y=ICEtsdif1, nar = 3, nma = 3, y.name = "p", ar.method = "ols" )

plot(res)
```