Primary productivity in Chatham Rise

(New Zealand)

Final Project

Time Series Analysis (MATH1318)

Submitted By: GROUP DREAM

Zhuoming Li (s3815870)

Sarthak Sirari (s3766477)

Lei Wang (s3412072)

Mingyi Tang (s3414342)

# Table of Content

[Table of Content 1](#_Toc42303949)

[Introduction 2](#_Toc42303950)

[Data 2](#_Toc42303951)

[Data Analysis 3](#_Toc42303952)

[Model Specification 6](#_Toc42303953)

[Finding the Best Seasonal ARIMA Model 11](#_Toc42303954)

[Residual Analysis of all the Chosen Competitive Models 18](#_Toc42303955)

[Analysis of Final Competitive Model 22](#_Toc42303956)

[Forecasting 25](#_Toc42303957)

[Conclusion 28](#_Toc42303958)

[References 29](#_Toc42303959)

[Appendix 29](#_Toc42303960)

[R Code 29](#_Toc42303961)

# Introduction

The most primary productivity growth in New Zealand’s marine environment is phytoplankton. The growth of phytoplankton is affected by the nutrients and sunlight. It is important to study phytoplankton growth because it supports the marine organisms, such as fish, mammals, and seabirds. It will help find the changing of ecosystems so that the economic, cultural and recreational purposes for the marine services will be affected. For example, fisheries.

The phytoplankton growth is measured by the concentration of a pigment Chlorophyll-a in milligram/cubic. This data contains records for the phytoplankton growth in oceanic regions between 1998 and 2017, also records for that in coastal areas between 2003 and 2017. We observe that the increasing speed of phytoplankton in oceanic regions is faster than in coastal areas. Thus, we plan to subgroup the dataset to only research in oceanic water areas. Moreover, the Chatham Rise region in this area keeps the highest offshore amounts of phytoplankton in New Zealand.

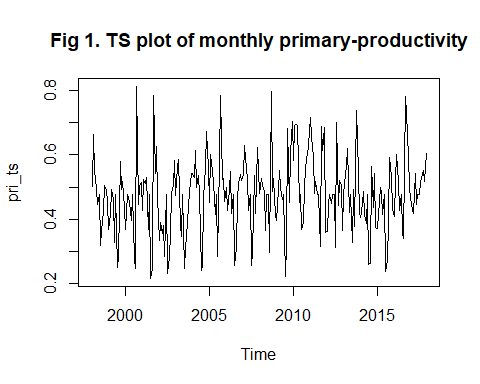
Overall, we are interested in the change of phytoplankton growth in the Chatham Rise region. Our aim is to forecast the phytoplankton growth in the next 10 years.

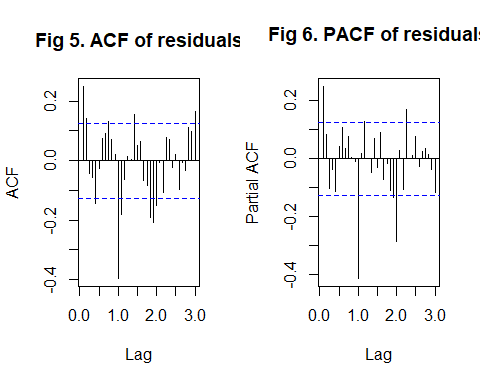
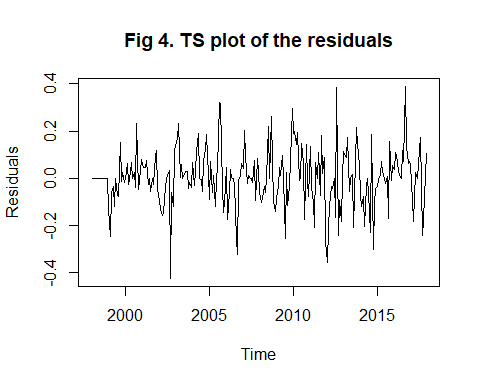
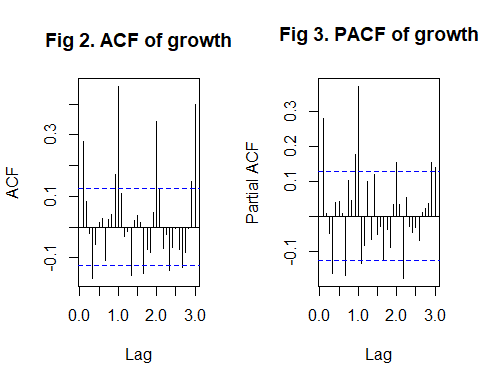
# Data

Our data set is made available by Statistics New Zealand and collected by National Institute of Water and Atmospheric Research (NIWA), National Oceanic and Atmospheric Administration (NOAA) and National Aeronautics and Space Administration (NASA). It contains a total of 9 columns and 10680 observations.

# Data Analysis

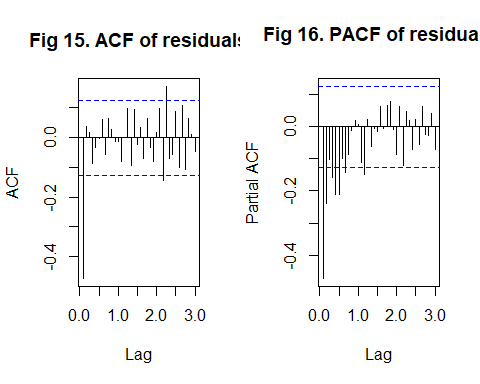
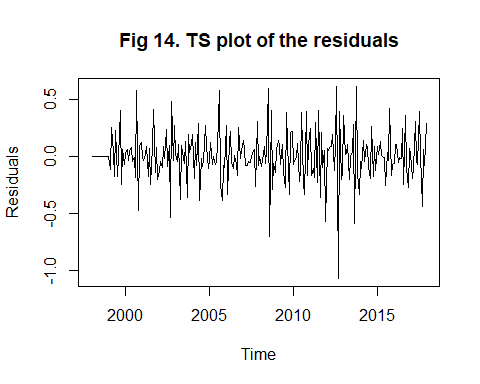
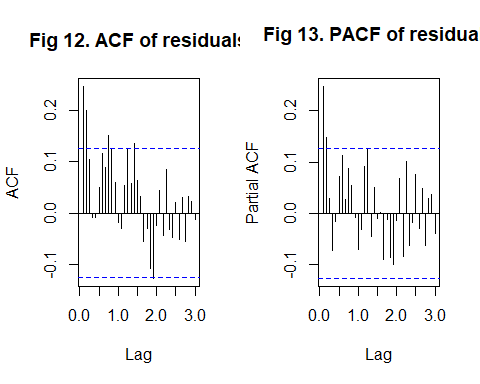
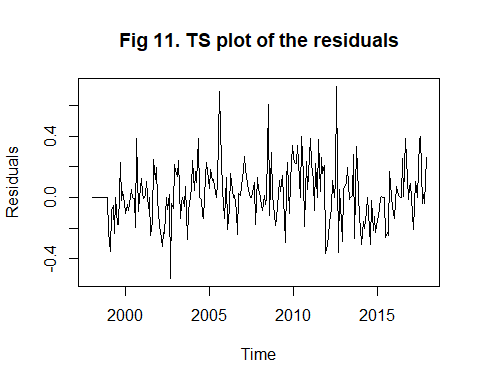
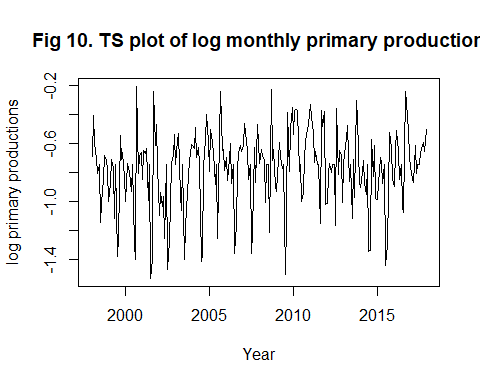
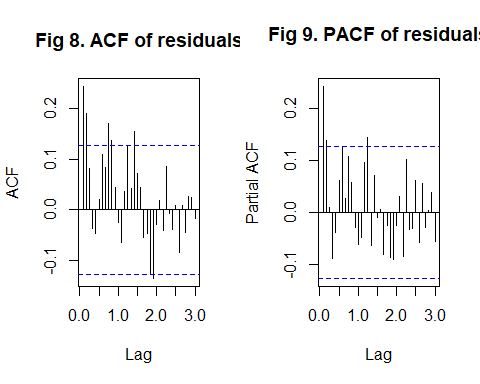
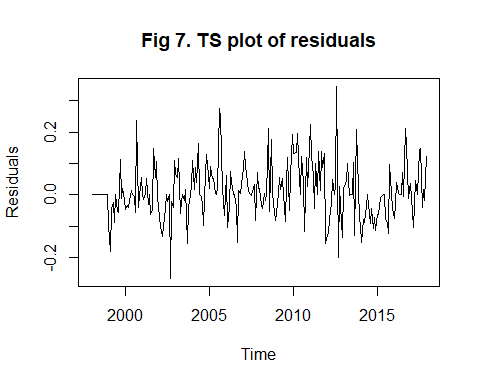
* Since we are only interested with the phytoplankton growth in the Chatham Rise region in oceanic water,we subdivide the data and focus on the first 240 rows which records the concentration from Jan1998 to Dec2017 in Chatham Rise. However, we observe some missing values for concentrations in some years, thus we replaced the missing values by the calculated average amount to minimise the bias.
* We create a monthly time-series data and visualise the plot(Fig 1). It shows a clear repeated pattern through 1998 to 2017. The trend between 2010 and 2015 is slightly moving down which might indicate the phytoplankton starts decreasing.
* Next, we check the ACF AND PACF for the original time-series (Fig 2 and Fig 3) and we can see the curve pattern in the ACF plot. Thus, the seasonality and existence of trend should be reasonably considered in this time-series data.
* To remove the seasonal trend effect, we fit a plain model with the first seasonal difference (D=1) and check the residuals(Fig 4) and autocorrelations, we found that the trend is removed, but in the plot “ACF of residuals”(Figure.1 Sample ACF), it shows significant autocorrelation at lag 12 and weak autocorrelation at lag 24; in the PACF of residuals(Fig 6 and Fig 6), both lag12&24 are significantly outstanding.
* Therefore, the next step is fitting SAR(2),SMA(2) with the first seasonal difference(D=1) on the plain model.





# Model Specification

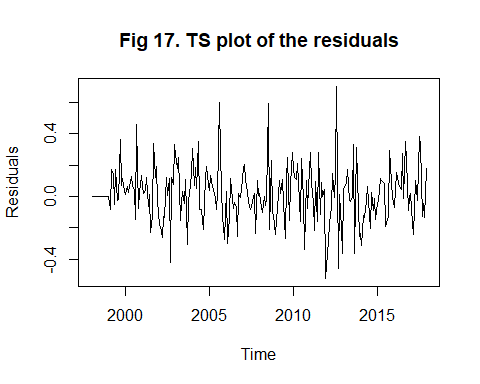
* In this step, we find all the possible SARIMA models that could fit our data.
* First, we started with plotting the time series for the residuals of the SARMA(2,2) model on our data set.
* In the time series plot (in Fig 7), the trend can still be seen.
* Then we plotted the ACF and PACF plot to check if the seasonality is still present.
* From the ACF (in Fig 8) and PACF (in Fig 9), we can see no significant lags at the first seasonal lags in either ACF and PACF, therefore, we conclude that the seasonality is filtered out.
* The significant correlation in PACF is probably due to the change point in the series.
* Then we move ahead with applying the logarithmic transformation on our data.
* In the time series plot (in Fig 10), variance seems to be stabilized.
* Now, we plot the time series for the residuals of SARMA(2, 2) model on our log transformed data set.
* In the time series plot (in Fig 11), the trend can still be seen.
* Then we plotted the ACF and PACF plot to check if the seasonality is still present.
* From the ACF (in Fig 12) and PACF (in Fig 13), we found no clear outstanding significant lags at lag 1 in ACF and PACF.
* This indicates non-stationary, hence we move ahead to apply first differencing and plot the time series for the residuals of SARIMA(2,1,2) model on our data set.
* In the time series plot (in Fig 14), the plot seems to be detrended.
* Then we plotted the ACF and PACF plot to check if the seasonality is still present.
* From the ACF (in Fig 15) and PACF (in Fig 16), we can observe one significant lag at lag 1 in ACF and two significant lags at lag 1 and lag 2 in PACF.
* So we include SARIMA(2,1,1)x(2,1,2)\_12 and SARIMA(0,1,1)x(2,1,2)\_12 (if you think PACF is tailing off and so this is only MA(1)) models into the set of possible SARIMA models.
* Further, we plot the EACF table to check for more possible SARIMA models.
* We select three points at the top left corner of table, i.e. ARMA(0,1), ARMA(0,2) and ARMA(1,2).
* Therefore, we include SARIMA(0,1,1)x(2,1,2)\_12, SARIMA(0,1,2)x(2,1,2)\_12 and SARIMA(1,1,2)x(2,1,2)\_12 models into the set of possible SARIMA models.
* We added SARIMA(2,1,2)x(2,1,2)\_12 model into the set of possible SARIMA models, just to overfit SARIMA(2,1,1)x(2,1,2)\_12 and SARIMA(1,1,2)x(2,1,2)\_12 models.
* Finally, the set of possible SARIMA models includes SARIMA(0,1,1)x(2,1,2)\_12, SARIMA(0,1,2)x(2,1,2)\_12, SARIMA(1,1,2)x(2,1,2)\_12, SARIMA(2,1,2)x(2,1,2)\_12 and SARIMA(2,1,1)x(2,1,2)\_12 models.

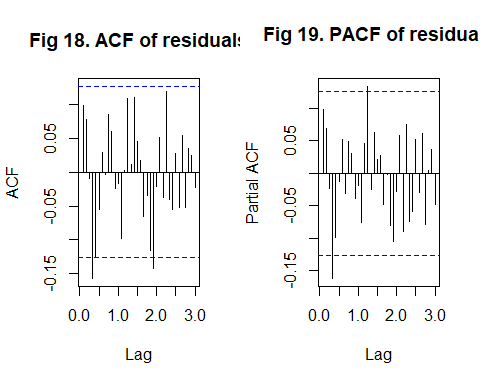


## AR/MA  
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13  
## 0 x o o o o o o o o o o o o o   
## 1 x x o o o o o o o o o o o o   
## 2 x o x o o o o o o o o o o o   
## 3 x o o x o o o o o o o o o o   
## 4 x o o x o o o o o o o o o o   
## 5 x x o x o x o o o o o o o o   
## 6 x x x x x o x o o o o o o o   
## 7 x x x x x o x x o o o o o o

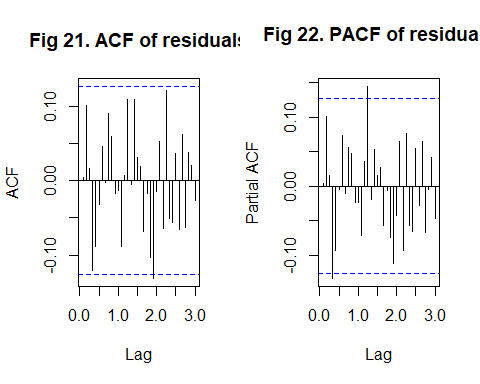
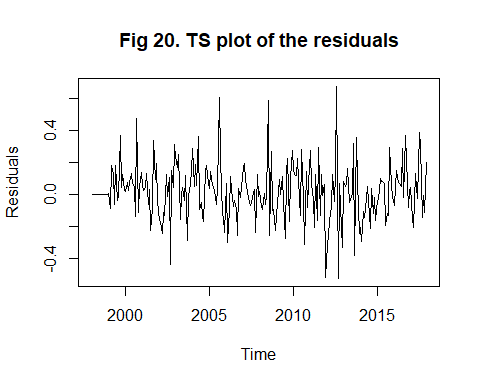
# Finding the Best Seasonal ARIMA Model

* Next step, we will fit those models above from small to larger and check the residuals plot to see which one performs white noise on residuals.
* The Fig 17, below, shows the time series plot of the residuals.
* There is constant variance, which means around zero.
* There are few residuals of [0.4,-0.4].the PACF and ACF for the SARIMA(0,1,1)x(2,1,2)\_12.
* There are two significant lags in both PACF and ACF (Fig 18 and Fig 19). We doubt the residual is white noise.

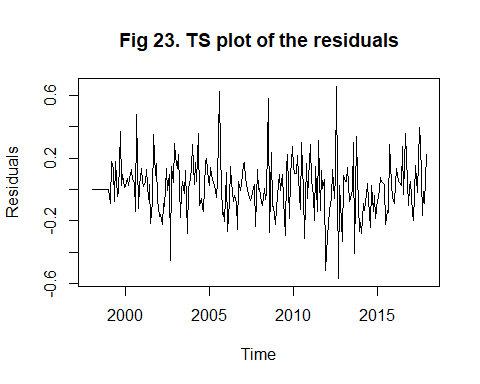


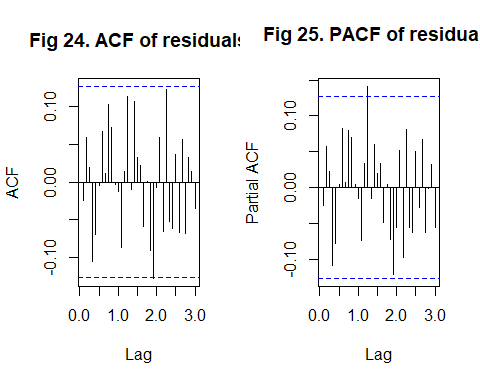


* Next, we increase the Moving Average term to MA(2), SARIMA(0,1,2)x(2,1,2)\_12 model is applied.
* The residuals plot(Fig 20) looks almost the same as the previous model.
* The PACF and ACF(Fig 21 and 22) are performing better, there are still two significant lags in PACF but closer to the blue dash line, and only one correlated lags on ACF.

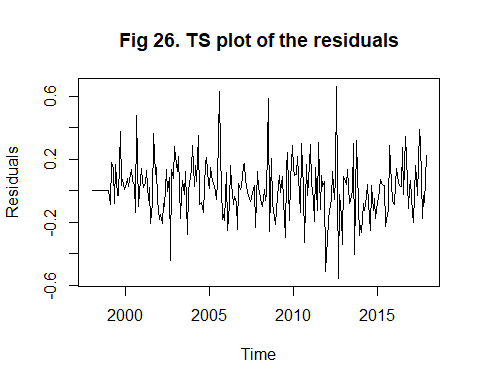


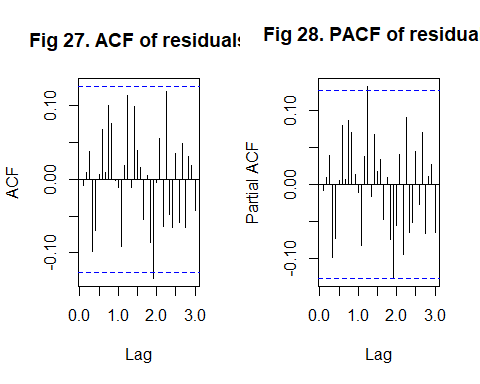
* Then, we increase the AR term to AR(1), the model SARIMA(1,1,2)x(2,1,2)\_12 is applied.
* We check the residual plot (Fig 23) which looks slightly squeezing and more constant variance. It performs slightly better.
* To check the PACF and ACF (Fig 24 and Fig 25), there is one outstanding auto-correlated lags on ACF and one significant auto-correlated lags in PACF, but it is closer to the blue-dashed line, which performs better than the model SARIMA(0,1,2)x(2,1,2)\_12.



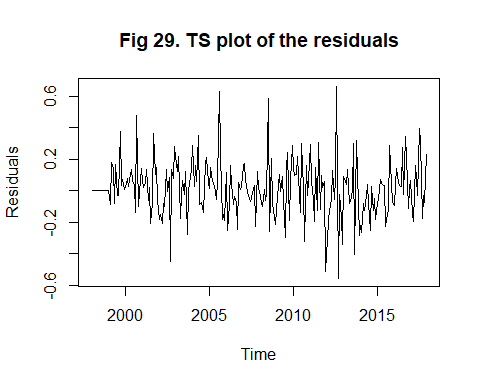


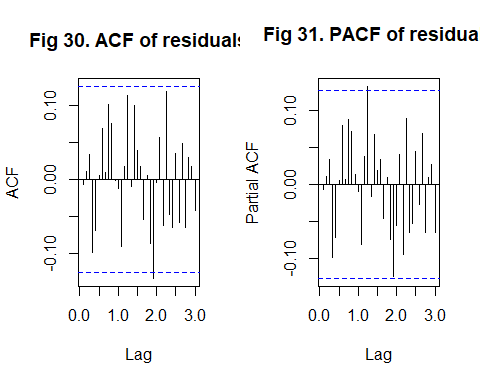
* Similar process to check the performance of model SARIMA(2,1,1)x(2,1,2)\_12.
* The residuals plot (Fig 26) looks approximately the same as the previous one.
* Besides, the ACF (Fig 27) shows no significant lags except at lag 0 and the PACF (Fig 28) shows only 1 weak significant lag, which indicates the model SARIMA(2,1,1)x(2,1,2)\_12 outperformed than model SARIMA(1,1,2)x(2,1,2)\_12.





* The model SARIMA(2,1,2)x(2,1,2)\_12 is an overfitting model of SARIMA(2,1,1)x(2,1,2)\_12.
* By visualising the residual plot (Fig 29) and ACF and PACF plots (Fig 30 and Fig 31), the overfitting model has reasonably the same performance.
* So, because of the principle of parsimony, SARIMA(2,1,1)x(2,1,2)\_12 is preferred.



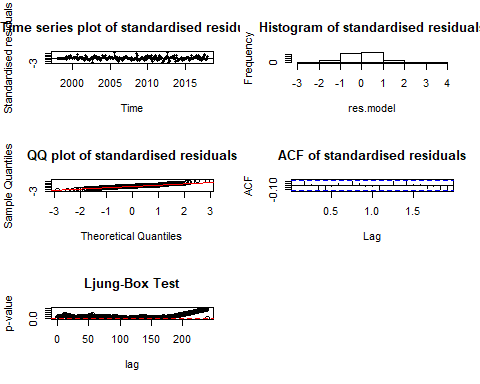


# Residual Analysis of all the Chosen Competitive Models

* Overall, we have three competitive tested models SARIMA(0,1,2)x(2,1,2)\_12, SARIMA(1,1,2)x(2,1,2)\_12 and SARIMA(2,1,1)x(2,1,2)\_12. These 3 models are almost performing white-noise. We also compare the adequacy of the models using an overfitting approach.This approach is based on significance tests. Therefore we should be concerned about normality.
* By the result of the Shapiro-wilk test for these 3 models, the p-values are less than 0.05 which indicates evidence to reject the normality. Also, we check the standardised residuals which look satisfied. The Ljung-Box test shows evidence to support independence. Since we have a large sample size and plots look good, by Central Limit Theorem (in Math Stat and Prob Theory), the significance test results are valid even if the original data/residuals are non-normal.
* Then we check the estimated coefficients in these 3 models, we found the SARIMA(2,1,1)x(2,1,2)\_12 is containing the most significant terms than others. However, only AR(1), AR(2), MA(1), SAR(1) and SMA(2) are significant which reject the SAR(2) fitting in the model. Therefore, it is highly likely to decrease the level of SAR term and use the model SARIMA(2,1,1)x(1,1,2)\_12.

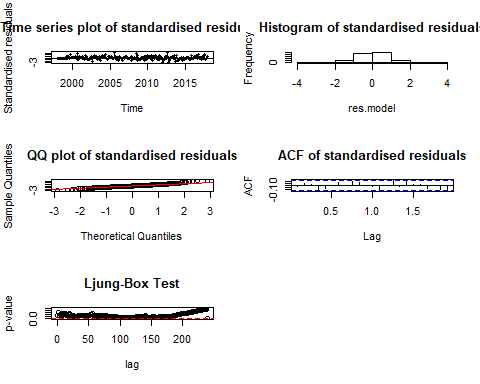
##   
## Shapiro-Wilk normality test  
##   
## data: res.model  
## W = 0.97664, p-value = 0.0005392

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ma1 -0.780717 0.062006 -12.5910 < 2e-16 \*\*\*  
## ma2 -0.107588 0.063140 -1.7040 0.08839 .   
## sar1 -0.907632 0.080878 -11.2222 < 2e-16 \*\*\*  
## sar2 -0.072556 0.083465 -0.8693 0.38468   
## sma1 0.092925 0.109299 0.8502 0.39522   
## sma2 -0.907036 0.103315 -8.7793 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1



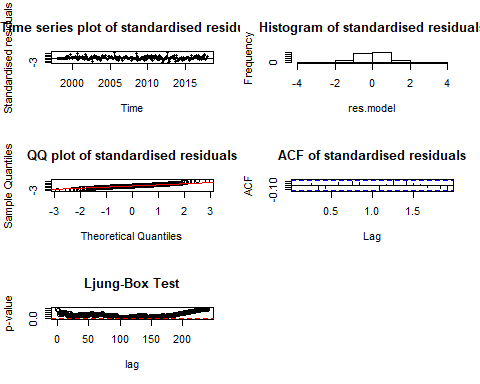
##   
## Shapiro-Wilk normality test  
##   
## data: res.model  
## W = 0.97387, p-value = 0.0002082

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ar1 0.536403 0.237591 2.2577 0.02397 \*   
## ma1 -1.294259 0.250409 -5.1686 2.359e-07 \*\*\*  
## ma2 0.330091 0.219458 1.5041 0.13255   
## sar1 -0.900183 0.082045 -10.9718 < 2.2e-16 \*\*\*  
## sar2 -0.057012 0.085447 -0.6672 0.50463   
## sma1 0.089479 0.103242 0.8667 0.38611   
## sma2 -0.910417 0.098393 -9.2528 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1



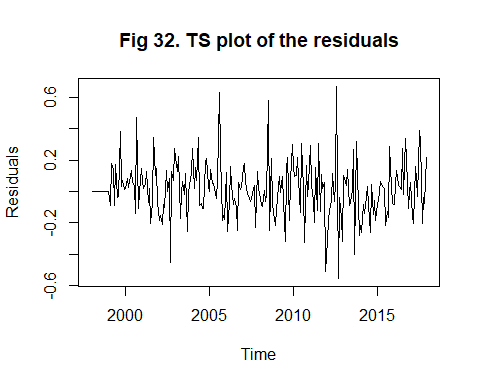
##   
## Shapiro-Wilk normality test  
##   
## data: res.model  
## W = 0.97298, p-value = 0.0001548

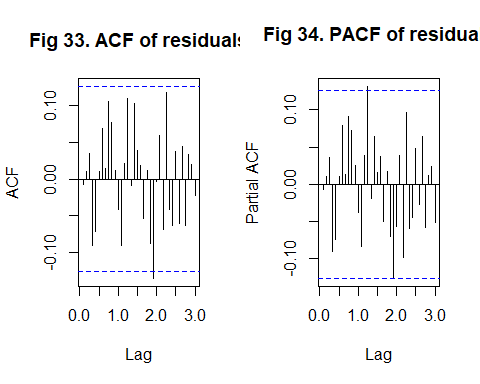
##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ar1 0.170102 0.079314 2.1447 0.03198 \*   
## ar2 0.126470 0.076415 1.6550 0.09792 .   
## ma1 -0.946702 0.046078 -20.5458 < 2e-16 \*\*\*  
## sar1 -0.906901 0.082864 -10.9444 < 2e-16 \*\*\*  
## sar2 -0.057599 0.085593 -0.6729 0.50099   
## sma1 0.093920 0.098574 0.9528 0.34070   
## sma2 -0.906035 0.093693 -9.6703 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1



# Analysis of Final Competitive Model

* From the above information, we need to verify whether SARIMA(2,1,1)x(1,1,2)\_12 is the most competitive model. Firstly, we check the residual plot (Fig 32), it shows the residuals randomly allocated around zero line which indicates zero mean and constant variance. Then we check the ACF and PACF (Fig 33 and Fig 34) where there is a weak significant lag in PACF, but it is near to the dashed line so it could be ignore; in the ACF, there is one significant lag, but it indicates the autocorrelation is smaller than -0.01 which could be ignore. It is reasonably considered this model as White-Noise.
* Next, we use the Shapiro-Wilk test and it shows the p-value is less than 0.05 which should indicate that there is evidence to reject the normality. However, from the standardised residuals plot shows reasonably constant variance, the histogram of standardized residuals is showing a symmetric bell-shape and most of data are allocating on the line in the normal QQ plots, which indicates normal distribution. Moreover, there is no significant lag in the ACF of Standardised residuals. Also, the Ljung-Box shows independence. We believe that the normality and independence in this model are satisfied, even the result of the Shapiro-Wilk test is less than 5%. Thus, SARIMA(2,1,1)x(1,1,2)\_12 is a feasible competitive model.
* To confirm it, we found that SARIMA(2,1,1)x(1,1,2)\_12 has the lowest AIC and BIC scores among the above competitive models. Therefore, we should believe that SARIMA(2,1,1)x(1,1,2)\_12 is the best competitive model for prediction.



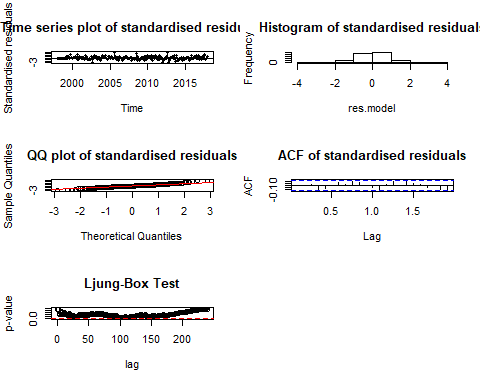


##   
## Shapiro-Wilk normality test  
##   
## data: res.model  
## W = 0.97275, p-value = 0.0001434

##   
## z test of coefficients:  
##   
## Estimate Std. Error z value Pr(>|z|)   
## ar1 0.180388 0.076833 2.3478 0.01888 \*   
## ar2 0.128614 0.075868 1.6952 0.09003 .   
## ma1 -0.954859 0.044457 -21.4781 < 2e-16 \*\*\*  
## sar1 -0.861830 0.048958 -17.6036 < 2e-16 \*\*\*  
## sma1 0.070275 0.094382 0.7446 0.45653   
## sma2 -0.929694 0.091576 -10.1522 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## df AIC  
## sp10.product 7 -88.03452  
## sp8.product 8 -86.47973  
## sp7.product 8 -85.33750  
## sp6.product 7 -84.82260

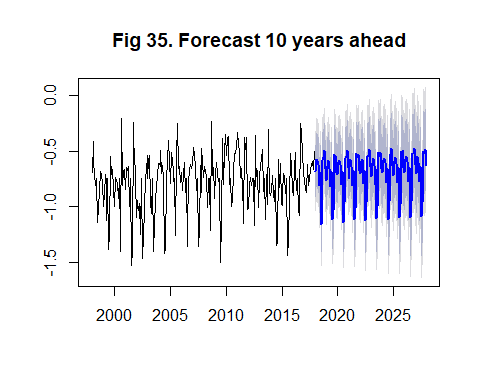
## df BIC  
## sp10.product 7 -64.05987  
## sp6.product 7 -60.84795  
## sp8.product 8 -59.08013  
## sp7.product 8 -57.93790



# Forecasting

* After above sorting, the best competitive model SARIMA(2,1,1)x(1,1,2)\_12 with a LOG-Transform on the original monthly time-series data, with method maximum-likelihood is used to forecast 10 years ahead. Since this is a monthly data, each 12 points(months) is equal to 1 year. For 10 years forwards, there are h=120 ahead.
* The plot(Fig 35) represents that the productivity in Chatham region will approximately remain the same pattern moving forward as previously, but there is a slightly upward trend for the intervals, which indicates the phytoplankton possibly is increasing.

## Point Forecast Lo 80 Hi 80 Lo 95 Hi 95  
## Jan 2018 -0.6825375 -0.9190708 -0.4460042 -1.0442840 -0.32079097  
## Feb 2018 -0.5765345 -0.8190192 -0.3340499 -0.9473828 -0.20568622  
## Mar 2018 -0.5853496 -0.8330958 -0.3376033 -0.9642448 -0.20645439  
## Apr 2018 -0.6184896 -0.8676727 -0.3693065 -0.9995822 -0.23739690  
## May 2018 -0.8084866 -1.0586419 -0.5583312 -1.1910662 -0.42590698  
## Jun 2018 -0.6910694 -0.9418772 -0.4402616 -1.0746468 -0.30749197  
## Jul 2018 -1.1564252 -1.4077927 -0.9050578 -1.5408586 -0.77199189  
## Aug 2018 -0.9649726 -1.2168505 -0.7130947 -1.3501865 -0.57975864  
## Sep 2018 -0.5805633 -0.8329333 -0.3281933 -0.9665299 -0.19459668  
## Oct 2018 -0.5584615 -0.8113154 -0.3056076 -0.9451681 -0.17175484  
## Nov 2018 -0.4934645 -0.7468021 -0.2401269 -0.8809110 -0.10601802  
## Dec 2018 -0.6499811 -0.9038004 -0.3961617 -1.0381642 -0.26179789  
## Jan 2019 -0.7654655 -1.0259906 -0.5049404 -1.1639042 -0.36702675  
## Feb 2019 -0.6362858 -0.8980505 -0.3745212 -1.0366204 -0.23595132  
## Mar 2019 -0.5849047 -0.8478554 -0.3219539 -0.9870531 -0.18275617  
## Apr 2019 -0.6546754 -0.9184168 -0.3909339 -1.0580331 -0.25131760  
## May 2019 -0.8200770 -1.0845399 -0.5556140 -1.2245382 -0.41561579  
## Jun 2019 -0.6803626 -0.9454912 -0.4152340 -1.0858418 -0.27488340  
## Jul 2019 -1.1163820 -1.3821572 -0.8506067 -1.5228501 -0.70991384  
## Aug 2019 -1.0314021 -1.2978128 -0.7649914 -1.4388422 -0.62396203  
## Sep 2019 -0.5239056 -0.7909472 -0.2568641 -0.9323104 -0.11550086  
## Oct 2019 -0.5722372 -0.8399070 -0.3045675 -0.9816028 -0.16287164  
## Nov 2019 -0.5417879 -0.8100877 -0.2734880 -0.9521170 -0.13145870  
## Dec 2019 -0.6864343 -0.9553629 -0.4175057 -1.0977251 -0.27514352  
## Jan 2020 -0.7040016 -0.9734614 -0.4345417 -1.1161048 -0.29189833  
## Feb 2020 -0.6004233 -0.8704748 -0.3303719 -1.0134314 -0.18741532  
## Mar 2020 -0.5824263 -0.8530693 -0.3117833 -0.9963391 -0.16851352  
## Apr 2020 -0.6180168 -0.8892621 -0.3467714 -1.0328507 -0.20318284  
## May 2020 -0.8017651 -1.0736138 -0.5299164 -1.2175218 -0.38600841  
## Jun 2020 -0.6804170 -0.9528699 -0.4079640 -1.0970978 -0.26373612  
## Jul 2020 -1.1411991 -1.4142562 -0.8681420 -1.5588039 -0.72359433  
## Aug 2020 -0.9642561 -1.2379174 -0.6905948 -1.3827849 -0.54572732  
## Sep 2020 -0.5627348 -0.8370019 -0.2884676 -0.9821902 -0.14327935  
## Oct 2020 -0.5503209 -0.8251959 -0.2754458 -0.9707059 -0.12993579  
## Nov 2020 -0.4900764 -0.7655704 -0.2145824 -0.9114081 -0.06874465  
## Dec 2020 -0.6449431 -0.9210564 -0.3688297 -1.0672220 -0.22266417  
## Jan 2021 -0.7468925 -1.0285885 -0.4651966 -1.1777094 -0.31607572  
## Feb 2021 -0.6212485 -0.9042109 -0.3382861 -1.0540022 -0.18849484  
## Mar 2021 -0.5744795 -0.8586628 -0.2902962 -1.0091003 -0.13985870  
## Apr 2021 -0.6395267 -0.9245909 -0.3544624 -1.0754948 -0.20355850  
## May 2021 -0.8074633 -1.0933469 -0.5215796 -1.2446845 -0.37024200  
## Jun 2021 -0.6702867 -0.9569394 -0.3836339 -1.1086842 -0.23188912  
## Jul 2021 -1.1097278 -1.3971323 -0.8223234 -1.5492750 -0.67018064  
## Aug 2021 -1.0120402 -1.3001861 -0.7238943 -1.4527213 -0.57135913  
## Sep 2021 -0.5191877 -0.8080709 -0.2303045 -0.9609964 -0.07737903  
## Oct 2021 -0.5591253 -0.8487437 -0.2695069 -1.0020584 -0.11619215  
## Nov 2021 -0.5245587 -0.8149172 -0.2342002 -0.9686237 -0.08049371  
## Dec 2021 -0.6706174 -0.9617148 -0.3795200 -1.1158125 -0.22542235  
## Jan 2022 -0.6998449 -0.9920062 -0.4076836 -1.1466670 -0.25302274  
## Feb 2022 -0.5932175 -0.8861601 -0.3002750 -1.0412345 -0.14520058  
## Mar 2022 -0.5712447 -0.8649633 -0.2775261 -1.0204485 -0.12204092  
## Apr 2022 -0.6109056 -0.9053701 -0.3164410 -1.0612503 -0.16056089  
## May 2022 -0.7924690 -1.0876724 -0.4972656 -1.2439437 -0.34099436  
## Jun 2022 -0.6689336 -0.9648698 -0.3729975 -1.1215290 -0.21633830  
## Jul 2022 -1.1267668 -1.4234332 -0.8301005 -1.5804788 -0.67305481  
## Aug 2022 -0.9607756 -1.2581706 -0.6633805 -1.4156021 -0.50594909  
## Sep 2022 -0.5466338 -0.8447589 -0.2485087 -1.0025768 -0.09069081  
## Oct 2022 -0.5414540 -0.8403112 -0.2425968 -0.9985167 -0.08439133  
## Nov 2022 -0.4847578 -0.7843624 -0.1851531 -0.9429635 -0.02655200  
## Dec 2022 -0.6384073 -0.9387598 -0.3380549 -1.0977568 -0.17905787  
## Jan 2023 -0.7303079 -1.0355316 -0.4250841 -1.1971073 -0.26350844  
## Feb 2023 -0.6072916 -0.9138254 -0.3007579 -1.0760946 -0.13848867  
## Mar 2023 -0.5639490 -0.8717527 -0.2561454 -1.0346941 -0.09320400  
## Apr 2023 -0.6254882 -0.9342629 -0.3167135 -1.0977184 -0.15325805  
## May 2023 -0.7953078 -1.1049979 -0.4856177 -1.2689379 -0.32167772  
## Jun 2023 -0.6600162 -0.9705759 -0.3494565 -1.1349763 -0.18505612  
## Jul 2023 -1.1019988 -1.4134120 -0.7905857 -1.5782642 -0.62573350  
## Aug 2023 -0.9948728 -1.3071298 -0.6826158 -1.4724287 -0.51731697  
## Sep 2023 -0.5128968 -0.8259940 -0.1997996 -0.9917377 -0.03405589  
## Oct 2023 -0.5465999 -0.8605359 -0.2326640 -1.0267235 -0.06647634  
## Nov 2023 -0.5089754 -0.8237580 -0.1941929 -0.9903938 -0.02755705  
## Dec 2023 -0.6560831 -0.9717111 -0.3404550 -1.1387945 -0.17337160  
## Jan 2024 -0.6939709 -1.0111709 -0.3767708 -1.1790865 -0.20885529  
## Feb 2024 -0.5850789 -0.9032313 -0.2669264 -1.0716510 -0.09850669  
## Mar 2024 -0.5601531 -0.8792481 -0.2410581 -1.0481668 -0.07213936  
## Apr 2024 -0.6028372 -0.9228101 -0.2828642 -1.0921936 -0.11348076  
## May 2024 -0.7827778 -1.1036146 -0.4619410 -1.2734554 -0.29210024  
## Jun 2024 -0.6576179 -0.9793068 -0.3359291 -1.1495985 -0.16563730  
## Jul 2024 -1.1132608 -1.4357969 -0.7907247 -1.6065372 -0.61998445  
## Aug 2024 -0.9554037 -1.2787843 -0.6320232 -1.4499717 -0.46083581  
## Sep 2024 -0.5318884 -0.8561144 -0.2076624 -1.0277493 -0.03602757  
## Oct 2024 -0.5320816 -0.8571549 -0.2070083 -1.0292383 -0.03492484  
## Nov 2024 -0.4780207 -0.8039593 -0.1520822 -0.9765007 0.02045926  
## Dec 2024 -0.6307663 -0.9575704 -0.3039623 -1.1305701 -0.13096263  
## Jan 2025 -0.7152032 -1.0464164 -0.3839901 -1.2217500 -0.20865643  
## Feb 2025 -0.5941387 -0.9267144 -0.2615631 -1.1027693 -0.08550813  
## Mar 2025 -0.5533410 -0.8872430 -0.2194391 -1.0640000 -0.04268204  
## Apr 2025 -0.6122747 -0.9472353 -0.2773142 -1.1245527 -0.09999680  
## May 2025 -0.7834929 -1.1194610 -0.4475248 -1.2973118 -0.26967408  
## Jun 2025 -0.6496013 -0.9865348 -0.3126679 -1.1648966 -0.13430608  
## Jul 2025 -1.0934716 -1.4313552 -0.7555879 -1.6102200 -0.57672311  
## Aug 2025 -0.9793354 -1.3181601 -0.6405106 -1.4975231 -0.46114765  
## Sep 2025 -0.5054376 -0.8452004 -0.1656749 -1.0250599 0.01418467  
## Oct 2025 -0.5345103 -0.8752099 -0.1938106 -1.0555654 -0.01345511  
## Nov 2025 -0.4946145 -0.8362618 -0.1529672 -1.0171190 0.02788990  
## Dec 2025 -0.6425012 -0.9850951 -0.2999074 -1.1664533 -0.11854915  
## Jan 2026 -0.6868213 -1.0314233 -0.3422194 -1.2138445 -0.15979816  
## Feb 2026 -0.5762473 -0.9219525 -0.2305421 -1.1049578 -0.04753680  
## Mar 2026 -0.5491283 -0.8959232 -0.2023334 -1.0795053 -0.01875126  
## Apr 2026 -0.5940578 -0.9418501 -0.2462654 -1.1259603 -0.06215528  
## May 2026 -0.7727931 -1.1215634 -0.4240229 -1.3061912 -0.23939506  
## Jun 2026 -0.6464267 -0.9961584 -0.2966950 -1.1812952 -0.11155817  
## Jul 2026 -1.1004428 -1.4511295 -0.7497561 -1.6367718 -0.56411380  
## Aug 2026 -0.9486272 -1.3002650 -0.5969894 -1.4864109 -0.41084355  
## Sep 2026 -0.5181499 -0.8707393 -0.1655605 -1.0573889 0.02108910  
## Oct 2026 -0.5223337 -0.8758765 -0.1687910 -1.0630307 0.01836328  
## Nov 2026 -0.4702302 -0.8247458 -0.1157146 -1.0124150 0.07195457  
## Dec 2026 -0.6223044 -0.9777930 -0.2668159 -1.1659773 -0.07863158  
## Jan 2027 -0.7011979 -1.0607911 -0.3416046 -1.2511483 -0.15124744  
## Feb 2027 -0.5815829 -0.9425957 -0.2205702 -1.1337044 -0.02946153  
## Mar 2027 -0.5426754 -0.9050745 -0.1802763 -1.0969169 0.01156614  
## Apr 2027 -0.5996740 -0.9632154 -0.2361326 -1.1556626 -0.04368538  
## May 2027 -0.7719309 -1.1365673 -0.4072945 -1.3295941 -0.21426771  
## Jun 2027 -0.6390792 -1.0047712 -0.2733872 -1.1983568 -0.07980150  
## Jul 2027 -1.0843514 -1.4510846 -0.7176181 -1.6452215 -0.52348126  
## Aug 2027 -0.9650085 -1.3327744 -0.5972427 -1.5274579 -0.40255923  
## Sep 2027 -0.4971108 -0.8659065 -0.1283150 -1.0611352 0.06691364  
## Oct 2027 -0.5227442 -0.8925691 -0.1529193 -1.0883425 0.04285417  
## Nov 2027 -0.4811615 -0.8520288 -0.1102943 -1.0483540 0.08603093  
## Dec 2027 -0.6296269 -1.0015355 -0.2577183 -1.1984120 -0.06084182



# Conclusion

* We analysed the date collected for investigating phytoplankton growth in oceanic waters around the Chatham Rise Region from Jan 1998 to Dec 2017, on a monthly basis.
* We performed the data modelling and concluded that our data gives best results for model fitting, when we do the first differencing of the Logarithmic transformed data.
* Further, we have proposed a set of possible ARIMA models, using each and every model specification tool such as ACF, PACF, EACF as SARIMA(0,1,1)x(2,1,2)\_12, SARIMA(0,1,2)x(2,1,2)\_12, SARIMA(1,1,2)x(2,1,2)\_12, SARIMA(2,1,2)x(2,1,2)\_12 and SARIMA(2,1,1)x(2,1,2)\_12.
* Going forward, we performed Residual Analysis and conclude that our chosen model SARIMA(2,1,1)x(1,1,2)\_12 is the best fit model.
* Finally, we use time series to do the forecasting to find the change of ecosystems so that the economic, cultural and recreational purposes for the marine services will be affected.
* For background research, we know that the increasing speed of phytoplankton in oceanic regions is faster than in coastal areas.
* Furthermore, we use our fitted model to predict the growth in next 10 years. The result is it will have a slight increase in future years. The reason for the rise might be a result of a small potential climate change.

# References

* Statistics New Zealand: Overall, the Chatham Rise region had the highest offshore primary productivity in New Zealand, between 1997 and 2018. (Last updated Oct 17, 2019) <https://www.stats.govt.nz/indicators/primary-productivity>
* MATH1318 Time Series Analysis notes prepared by Dr. Haydar Demirhan.
* Time Series Analysis with application in R by Cryer and Chan.

# Appendix

### R Code

knitr::opts\_chunk$set(echo = FALSE, warning = FALSE,message=FALSE)

library(TSA) library(fUnitRoots) library(lmtest) library(s20x) library(FitAR) library(forecast) sort.score <- function(x, score = c(“bic”, “aic”)){ if (score == “aic”){ x[with(x, order(AIC)),] } else if (score == “bic”) { x[with(x, order(BIC)),] } else { warning(‘score = “x” only accepts valid arguments (“aic”,“bic”)’) } } residual.analysis <- function(model, std = TRUE){ library(TSA) library(FitAR) if (std == TRUE){ res.model = rstandard(model) }else{ res.model = residuals(model) } par(mfrow=c(3,2)) plot(res.model,type=‘o’,ylab=‘Standardised residuals’, main=“Time series plot of standardised residuals”) abline(h=0) hist(res.model,main=“Histogram of standardised residuals”) qqnorm(res.model,main=“QQ plot of standardised residuals”) qqline(res.model, col = 2) acf(res.model,main=“ACF of standardised residuals”) print(shapiro.test(res.model)) k=0 LBQPlot(res.model, lag.max = length(model$residuals)-1 , StartLag = k + 1, k = 0, SquaredQ = FALSE) }

pri=read.csv(“primary\_productivity\_clean.csv”) pri=pri[seq(1:240),]#Take the first 240, which Jan1998-Dec2017,in Chatham Ocean cal\_mean=mean(pri$data\_value,na.rm=T) pri[is.na(pri)]<-cal\_mean#Convert the missing value to the mean of data\_value pri\_ts=ts(pri$data\_value,start = c(1998,1),end=c(2017,12),frequency = 12) plot(pri\_ts,main=“Fig 1. TS plot of monthly primary-productivity”) par(mfrow=c(1,2)) acf(pri\_ts, lag.max = 36, main=“Fig 2. ACF of phytoplankton growth”) pacf(pri\_ts, lag.max = 36,main=“Fig 3. PACF of phytoplankton growth”) p1.product = arima(pri\_ts,order=c(0,0,0),seasonal=list(order=c(0,1,0), period=12)) res.p1 = residuals(p1.product) par(mfrow=c(1,1)) plot(res.p1,xlab=‘Time’,ylab=‘Residuals’,main=“Fig 4. TS plot of the residuals”) par(mfrow=c(1,2)) acf(res.p1, lag.max = 36, main = “Fig 5. Sample ACF of the residuals”) pacf(res.p1, lag.max = 36, main = “Fig 6. Sample PACF of the residuals”)

p2.product = arima(pri\_ts, order=c(0, 0, 0), seasonal = list(order = c(2, 1, 2), period = 12)) res.p2 = residuals(p2.product) par(mfrow = c(1, 1)) plot(res.p2, xlab=‘Time’, ylab=‘Residuals’, main=“Fig 7. TS plot of residuals”) par(mfrow = c(1, 2)) acf(res.p2, lag.max = 36, main = “Fig 8. ACF of the residuals”) pacf(res.p2, lag.max = 36, main = “Fig 9. PACF of the residuals”)

log.pri\_ts = log(pri\_ts) par(mfrow=c(1, 1)) plot(log.pri\_ts, ylab=‘log primary productions’, xlab=‘Year’, main = “Fig 10. TS plot of log monthly primary production”)

p3.product = arima(log.pri\_ts, order = c(0, 0, 0), seasonal = list(order = c(2, 1, 2), period = 12)) res.p3 = residuals(p3.product) par(mfrow = c(1, 1)) plot(res.p3,xlab=‘Time’,ylab=‘Residuals’,main=“Fig 11. TS plot of the residuals”) par(mfrow = c(1, 2)) acf(res.p3, lag.max = 36, main = “Fig 12. Sample ACF of the residuals”) pacf(res.p3, lag.max = 36, main = “Fig 13. Sample PACF of the residuals”)

p4.product = arima(log.pri\_ts, order = c(0, 1, 0), seasonal = list(order = c(2, 1, 2), period = 12)) res.p4 = residuals(p4.product) par(mfrow = c(1, 1)) plot(res.p4, xlab = ‘Time’, ylab = ‘Residuals’, main = “Fig 14. TS plot of the residuals”) par(mfrow = c(1, 2)) acf(res.p4, lag.max = 36, main = “Fig 15. Sample ACF of the residuals”) pacf(res.p4, lag.max = 36, main = “Fig 16. Sample PACF of the residuals”) eacf(res.p4)

sp5.product = arima(log(pri\_ts),order=c(0,1,1),seasonal=list(order=c(2,1,2), period=12)) res.sp5 = residuals(sp5.product) par(mfrow=c(1,1)) plot(res.sp5,xlab=‘Time’,ylab=‘Residuals’,main=“Figure.5 TS plot of the residuals”) par(mfrow=c(1,2)) acf(res.sp5, lag.max = 36, main = “Figure.5 Sample ACF of the residuals”) pacf(res.sp5, lag.max = 36, main = “Figure.5 Sample PACF of the residuals”)

sp6.product = arima(log(pri\_ts),order=c(0,1,2),seasonal=list(order=c(2,1,2), period=12)) res.sp6 = residuals(sp6.product) par(mfrow=c(1,1)) plot(res.sp6,xlab=‘Time’,ylab=‘Residuals’,main=“Figure.6 TS plot of the residuals”) par(mfrow=c(1,2)) acf(res.sp6, lag.max = 36, main = “Figure.6 Sample ACF of the residuals”) pacf(res.sp6, lag.max = 36, main = “Figure.6 Sample PACF of the residuals”)

sp7.product = arima(log(pri\_ts),order=c(1,1,2),seasonal=list(order=c(2,1,2), period=12)) res.sp7 = residuals(sp7.product) par(mfrow=c(1,1)) plot(res.sp7,xlab=‘Time’,ylab=‘Residuals’,main=“Figure.7 TS plot of the residuals”) par(mfrow=c(1,2)) acf(res.sp7, lag.max = 36, main = “Figure.7 Sample ACF of the residuals”) pacf(res.sp7, lag.max = 36, main = “Figure.7 Sample PACF of the residuals”)

sp8.product = arima(log(pri\_ts),order=c(2,1,1),seasonal=list(order=c(2,1,2), period=12)) res.sp8 = residuals(sp8.product) par(mfrow=c(1,1)) plot(res.sp8,xlab=‘Time’,ylab=‘Residuals’,main=“Figure.8 TS plot of the residuals”) par(mfrow=c(1,2)) acf(res.sp8, lag.max = 36, main = “Figure.8 Sample ACF of the residuals”) pacf(res.sp8, lag.max = 36, main = “Figure.8 Sample PACF of the residuals”)

sp9.product = arima(log(pri\_ts),order=c(2,1,2),seasonal=list(order=c(2,1,2), period=12)) res.sp9 = residuals(sp9.product) par(mfrow=c(1,1)) plot(res.sp9,xlab=‘Time’,ylab=‘Residuals’,main=“Figure.9 TS plot of the residuals”) par(mfrow=c(1,2)) acf(res.sp9, lag.max = 36, main = “Figure.9 Sample ACF of the residuals”) pacf(res.sp9, lag.max = 36, main = “Figure.9 Sample PACF of the residuals”)

residual.analysis(model = sp6.product)#Check normality and residuals coeftest(sp6.product)#Check significant terms

residual.analysis(model = sp7.product) # Check normality and residuals coeftest(sp7.product) # Check significant terms

residual.analysis(model = sp8.product)#Check normality and residuals coeftest(sp8.product)#Check significant terms

sp10.product = arima(log(pri\_ts),order=c(2,1,1),seasonal=list(order=c(1,1,2), period=12)) res.sp10 = residuals(sp10.product) par(mfrow=c(1,1)) plot(res.sp10,xlab=‘Time’,ylab=‘Residuals’,main=“Figure.10 TS plot of the residuals”) par(mfrow=c(1,2)) acf(res.sp10, lag.max = 36, main = “Figure.10 Sample ACF of the residuals”) pacf(res.sp10, lag.max = 36, main = “Figure.10 Sample PACF of the residuals”)

residual.analysis(model = sp10.product)

coeftest(sp10.product)#SARIMA(2,1,1)x(1,1,2)\_12

sc.AIC = AIC(sp6.product,sp7.product,sp8.product,sp10.product) sort.score(sc.AIC, score = “aic”) sc.BIC = BIC(sp6.product,sp7.product,sp8.product,sp10.product) sort.score(sc.BIC, score = “bic”)

sp10.product\_2 = Arima(log(pri\_ts),order=c(2,1,1),seasonal=list(order=c(1,1,2), period=12),method = “ML”) future = forecast(sp10.product\_2, h = 120) print(future) plot(future,main=“Figure.10 Forecast 10 years ahead”)