Course Project
Applied Time Series Analysis

Geomagnetic Pole Intensities

Zherong Ye
1005459754

2023-04-01

Abstract

The Earth's magnetic field is a dynamic phenomenon that plays a crucial role in various aspects of our planet. It undergoes changes in strength and orientation over time, known as "geomagnetic secular variation." In recent years, the geomagnetic poles have been drifting towards Siberia at an accelerating rate, posing challenges to global navigation systems and industries. In this project, we develop a forecasting methodology based on the Box-Jenkins framework to predict the magnetic intensities of the North and South poles at their current locations. Our analysis aims to understand the behavior of the Earth's magnetic field and address critical questions related to geomagnetic secular variation. This research is essential for mitigating and adapting to the impacts of magnetic field changes on navigation systems, industries, and satellite performance.

Contents

Introduction

The Earth's magnetic field plays a crucial role in various aspects of our planet. It is represented as a time series data set that captures the changes in magnetic field intensity at the North and South poles over time. The data set includes historical measurements of magnetic field intensity, typically collected from magnetic observatories and satellite measurements, covering a significant time span.

Understanding the behavior of the Earth's magnetic field is of paramount importance due to its significant impact on our planet and its inhabitants. The magnetic field acts as a shield, protecting the Earth from harmful solar flares, magnetic storms, and cosmic rays. It also plays a critical role in navigation systems, industries, and satellite technology, making it a topic of great relevance and interest.

Investigating changes in the Earth's magnetic field, particularly the phenomenon of "geomagnetic secular variation," is essential for several reasons. First, the recent acceleration in the drift of the geomagnetic poles from Canada towards Siberia has raised concerns about the potential implications for global navigation systems and industries that rely on precise magnetic field information. This drift has been attributed to a strong positive geomagnetic anomaly near Lake Baikal in Siberia, which makes it crucial to understand and predict such variations.

Second, changes in the magnetic field intensity at higher altitudes can impact satellite performance and equipment that depend on them. Accurate prediction of these changes can help mitigate potential risks and optimize satellite operations. Furthermore, studying the geomagnetic secular variation can provide insights into the dynamic behavior of the Earth's core and geophysical processes that drive these changes.

Through a comprehensive analysis of the data set and the application of appropriate forecasting techniques, this study aims to provide valuable insights into the behavior of the Earth's magnetic field, its changes over time, such as whether there is seasonality within these behavior or there is a significant trend. The goal of this analysis, is to find models which may effectively predict the future trends of such behavioral changes in the magnetic fields of the North and South poles.

Model Specification

The model that has been derived using 90% of the Geomagnetic Intensity Data as training data.

North:

Model specification: ARIMA(5,3,4)

- Autoregressive (AR) component of order 5
- Differencing (D) of order 3
- Moving average (MA) component of order 4

An ARIMA(5,3,4) model was used to fit the Geomagnetic field of the North Pole, which included five autoregressive terms and four moving average terms, after differencing the series three times to achieve stationarity. The model is deduced from simple differencing, and estimated using EACF to select the lowest significant AIC & BIC values for the AR and MA components of this model. The AIC and BIC values for the model were 1966.435 and 2004.084, respectively, indicating relatively good fit. The residuals of the model have a near 0 p-value for a Shapiro-Wilk normality test, it suggests that the model may not adequately capture all the patterns in the data and there may be remaining systematic variation in the residuals. This potentially indicates that the model needs to be improved, by incorporating additional predictors or a more complex model structure. A seasonal component is not included in the model, since there is no evidence of it being a good fit.

South:

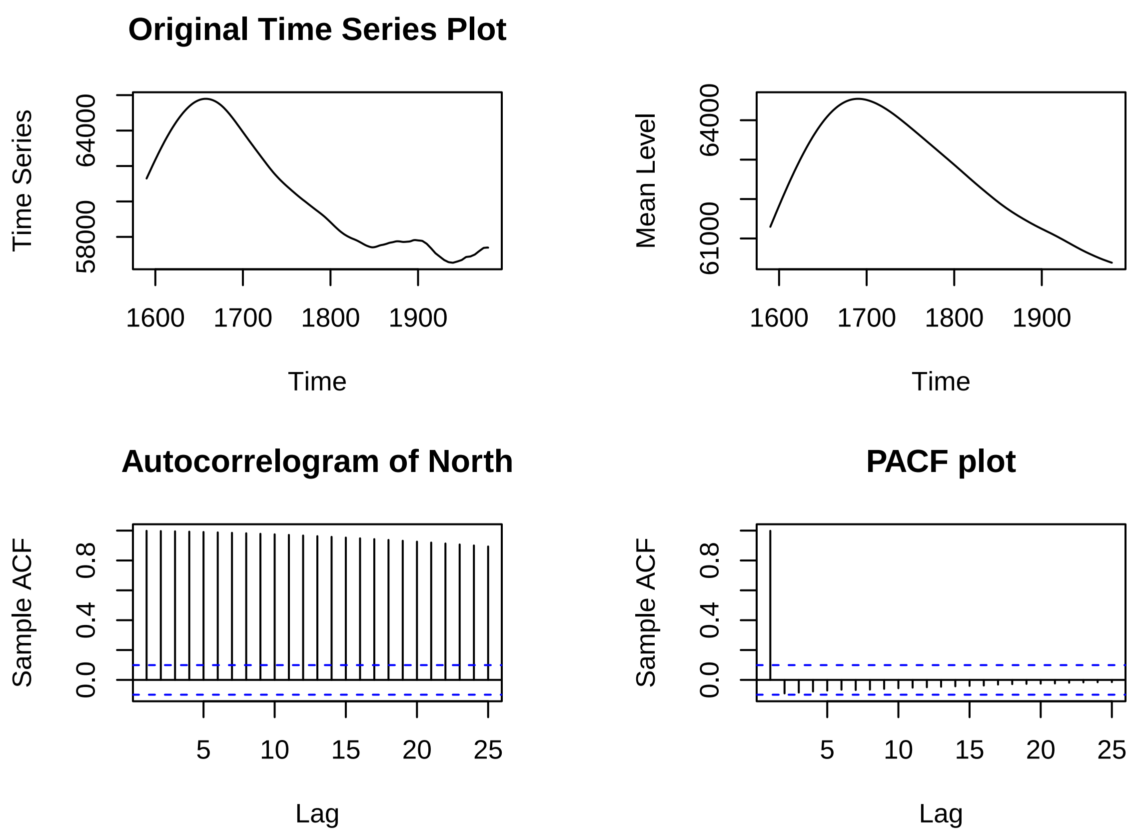
Model specification: ARIMA(5,2,5)

- Autoregressive (AR) component of order 5
- Differencing (D) of order 2
- Moving average (MA) component of order 5

An ARIMA(5,3,4) model was used to fit the Geomagnetic field of the South Pole, which included five autoregressive and moving average terms, after differencing the series two times to achieve stationarity. Similar to the North Model, it is deduced from simple differencing, and estimated using EACF to select the lowest significant AIC & BIC values for the AR and MA components of this model. The AIC and BIC values were the lowest for this model, which is 2304,767 and 2346.403, respectively, indicating the best fit among other values. The residuals of the model have a near 0 p-value for a Shapiro-Wilk normality test, it suggests that the model may not adequately capture all the patterns in the data and there may be remaining systematic variation in the residuals. This potentially indicates that the model needs to be improved, by incorporating additional predictors or a more complex model structure. A seasonal component is not included in the model, since there is no evidence of it being a good fit.

Fitting and Diagnostics

North:



From the Time Series above, we can see that there is no evidence of a pattern that repeats itself over a fixed interval from the time series plots. Thus, we may eradicate the possibility of using seasonal differencing method in our model, but, we could roughly see a downward trend, which mean that a Log Transformation may be reasonable. From the Mean Level Plot above, we may see that the values reach its apex at t=100, and are getting smaller when the time approaches the most right-hand side, thus, it suggests that the underlying process generating the data may not be stable over time.

ACF and PACF analysis

The ACF plot of the original time series displays spikes above the threshold level, indicating a significant amount of autocorrelation in the data. This implies that past values of the time series are strongly related to current values, and that differencing terms must be applied to make the time series stationary and remove any autocorrelation. Our aim is to obtain a time series with constant statistical properties to enable the modeling of future values. Conversely, the PACF plot indicates that all values except lag0 fall within the threshold, suggesting that an AR model may be appropriate for this time series.

Unit root and Stationarity Tests

Table 1: Data in Appendix A

Tests	Test.Statistic	Lag.Order	P.Value
ADF	-3.1294	7	0.1003
KPSS Level	5.9459	5	0.01
KPSS Trend	0.67056	5	0.01

ADF Test

$$\begin{cases} H_0 : Non - stationary \\ H_a : Stationary \end{cases}$$

The P-Value of the ADF test is 0.1003 > 0.05, therefore we need to reject H0. Thus, the Original Time Series is non-stationary.

KPSS Level Test

$$\begin{cases} H_0 : Stationary \\ H_a : Non - Stationary \end{cases}$$

The P-Value of the ADF test is 0.01 < 0.05, therefore we need to reject H0. Thus, the Original Time Series is non-stationary.

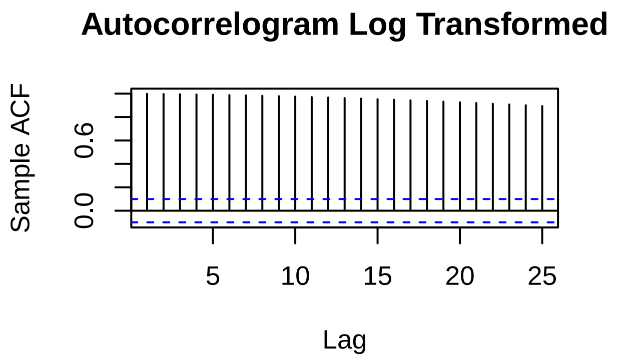
KPSS Trend Test

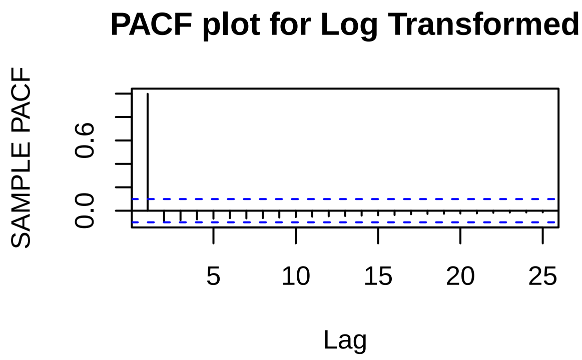
$$\begin{cases} H_0 : Trend - Stationary \\ H_a : Non - Trend - Stationary \end{cases}$$

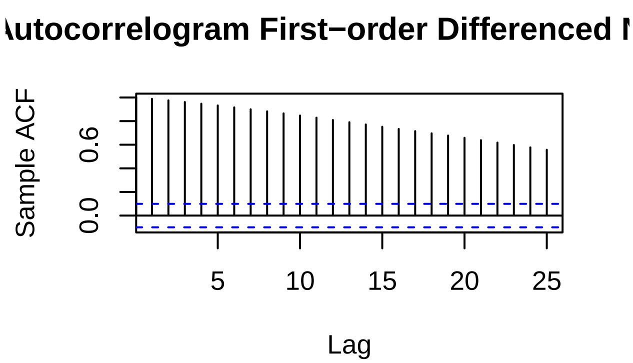
The P-Value of the ADF test is 0.01 < 0.05, therefore we need to reject H0. Thus, the Original Time Series Not Trend-stationary.

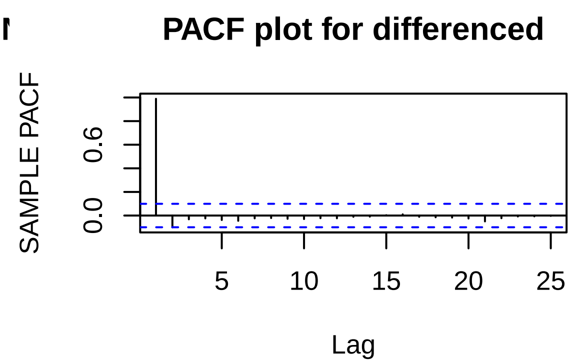
The results of the ADF and KPSS tests provide enough evidence to conclude that the time series is non-stationary. To make the time series stationary, we need to apply transformations or differencing techniques. Since there are no apparent patterns that repeat over time. We can first explore mathematical transformations and differencing techniques to create a stationary time series. In the following section, we will compare and determine which technique is more appropriate.

Log Transformation vs First-order Differenced









After analyzing the ACF and PACF plots of the Log Transformed and First-order Differenced Series, it was observed that the First-order Differenced Series had a significant tail off pattern in the ACF plot, especially starting at lag 5. This indicates that differencing the series has removed more of the correlation from the data, making it a better fit for modeling than the log-transformed series. Additionally, the tail off pattern in the ACF at lag 5 suggests that there is a possible (AR) process of order 5.

However, the PACF plot does not show any values lying outside the threshold, indicating that an AR(5) model may not be appropriate, contradicting the previous assumption. Furthermore, since the Original Time Series is non-stationary, an ARIMA model will be considered.

Plot of Log Transformed Series see Appendix B

Plot of First-order Differenced Series see Appendix C

First-Order Difference

Table 2: Data in Appendix C.1

Tests	Test.Statistic	Lag.Order	P.Value
ADF	-2.5263	7	0.3549
KPSS Level	1.1898	5	0.01
KPSS Trend	1.0953	5	0.01

Consistent with the findings of the Unit Root and Stationarity Tests performed on the original time series, we find that the First-order Differenced series is also non-stationary, as shown in Table 2.

Second-Order Difference

The Time Series plot reveals instability as the observations approach the right-hand side of the plot (see Appendix D), it suggests that the series is becoming more unpredictable. Furthermore, there is spikes at lag5 and lag 20 in the ACF plot (see Appendix D.1), which may indicate a seasonal AR process of order 20.

Table 3: Data in Appendix D.2

Tests	Test.Statistic	Lag.Order	P.Value
ADF	-4.5429	7	0.01
KPSS Level	0.69819	5	0.01371
KPSS Trend	0.17089	5	0.02926

The results presented in Table 3 are contradictory, as the p-value of the ADF test is 0.01, which is less than the significance level of 0.05, indicating that we can reject the null hypothesis and conclude that the model is stationary. However, the KPSS test yields results that lead to the rejection of the null hypothesis, suggesting that the model is non-stationary. To resolve this conflict, we will take a third order difference of the model.

Third-Order Difference

The Time Series plot reveals even more instability as the observations approach the right-hand side of the plot in comparison to the Second-Order Difference, but the Mean Level plot dips to near 0 as we move rightwards. This indicates that the series has a stationary behavior, and the data points in the series do not show any significant trend or seasonality and are randomly fluctuating around a constant mean value (see Appendix E).

Table 4: Data in Appendix E.2

Tests	Test.Statistic	Lag.Order	P.Value
ADF	-11.433	7	0.01
KPSS Level	0.034423	5	0.1
KPSS Trend	0.017776	5	0.1

The results presented in Table 4 show that the model is stationary, as the p-value of the ADF test is 0.01, which is less than the significance level of 0.05, indicating that we can reject the null hypothesis and conclude that the model is stationary. Furthermore, the KPSS test yields similar results which now fail to reject the null hypothesis, suggesting that the model is stationary.

However, similar to Second-order Differencing, the presence of a spike at lag 20 (see Appendix E.1) in the ACF plot, suggests that there may be some presence of a seasonal pattern or some other type of repeating pattern in the data that occurs every 20 time periods.

Selecting ARMA Model

I will select the model with the lowest (AIC) and (BIC) values as it is a commonly used method for model selection in time series analysis.

Table 5: Data in Appendix E.4

Model	AIC	BIC
ARIMA(4,0,4)	1965.756	1999.444
ARIMA(3, 0, 5)	1965.889	1999.577
ARIMA(5,0,4)	1966.435	2004.084

Based on Table 5, which shows that the ARIMA(4,0,4) model has the lowest AIC and BIC values, making it the most promising candidate model. However, further investigation using the EACF table (refer to Appendix E.3) reveals that both ARIMA(4,0,4) and ARIMA(3,0,5) are non-significant, indicating that the correlation coefficient is not significantly different from zero. The only model left as a potential candidate is ARIMA(5,0,4), which is shown as significant in the EACF, suggesting that the correlation coefficient is significantly different from zero. Therefore, we will proceed the analysis using ARIMA(5,0,4).

Residual analysis

The standardized Residuals become unstable as we move towards the right-hand side, notably after t=200 (see Appendix E.7). This means that the model has adequately captured the trend of the data before t=200, but the instability after t=200 means that there is a possibility that we were not able to effectively more recent data. Also another possibility for this instability is due to natural variations, thus, we can not be certain if it is purely due to the ineffectiveness of this model (see Appendix E.5). Furthermore, by inspecting the Ljung-Box statistic, we see that $lag_{1},_{2},_{3},_{4}$ are insignificant as the P-Values are greater than 0.05, but lag_5 on wards are all significant. This indicates that the past values of the time series are important for predicting the future values, and hence the lagged value can be included in the model as a predictor variable.

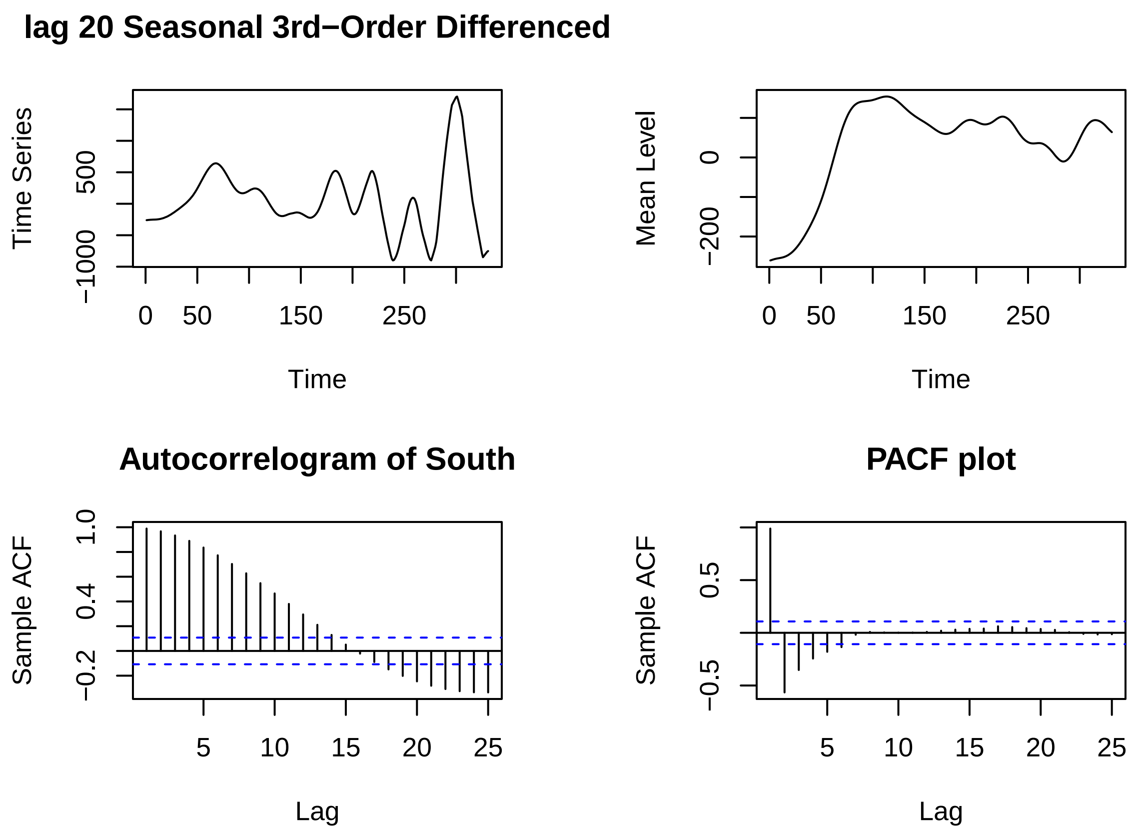
Table 6: Data in Appendix E.6

Shapiro.Wilk.normality.test	Value
W	0.67722
P-Value	$<$ 2.2e-16

By inspecting the QQ Plot (see Appendix E.6), we may see that the middle values of residuals lie on the line of best-fit, with left and right ends tailing downwards and upwards away from the line of best-fit respectively. This indicates that the residuals may be somewhat normally distributed. We will use the Shapiro Wilk Normality test to confirm this.

From Table 6 we see that, Shapiro Wilk Normality test W has a value of 0.67722, which indicates some deviation from normality and the low p-value suggests that the deviation from normality is statistically significant. which means that there is sufficient evidence to reject the null hypothesis of normality in favor of an alternative hypothesis that the data is not normally distributed.

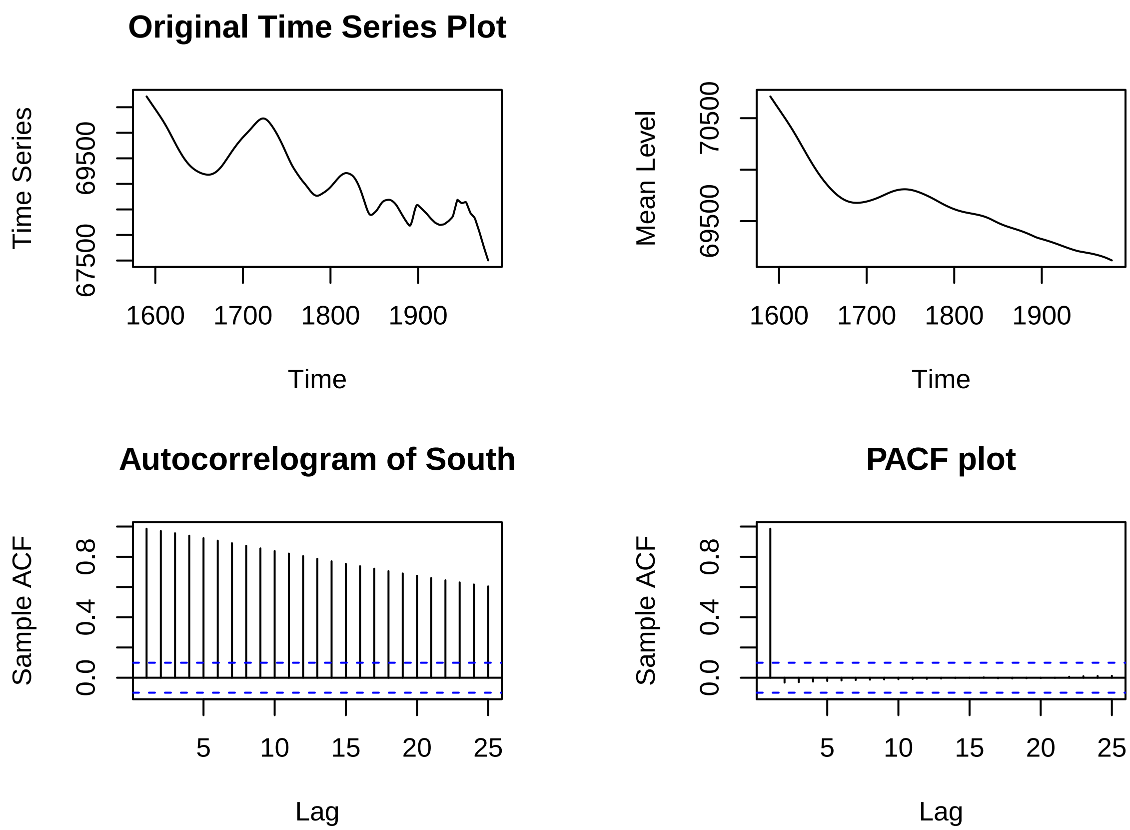
Seasonality



When we implement Seasonal Thrid-order Differenced of Order 20, we may see that the PACF plot spikes at Lag2 then a upward tail, then no other spikes at future lags, this means that there is no seasonality in our data. Thus, we may conclude that seasonality is not an appropriate method.

Even though the ARMA(5,0,4) does not follow white-noise nor normal distribution. There is enough evidence to believe that it is in fact a good model and the instability that exists within the model may possibly be caused by natural variation.

South:



From the Time Series above, we can see that there possibility of a pattern that repeats itself on the right-hand side of the time series plot. Thus, there is possibility of using seasonal differencing method in our model. We could roughly see a downward trend, which mean that a Log Transformation may be reasonable.

ACF and PACF analysis

The PACF plot indicates that all values except lag_0 fall within the threshold, suggesting that an AR model may be appropriate for this time series.

The ACF plot of the original time series displays a tail off pattern at Lag_5, suggests that there is a possible (AR) process of order 5, with all values above the threshold. Differencing terms must be applied to make the time series stationary and remove any autocorrelation. Our aim is to obtain a time series with constant statistical properties to enable the modeling of future values.

Unit root and Stationarity Tests

Table 7: Data in Appendix F

Tests	Test.Statistic	Lag.Order	P.Value
ADF	-3.4834	7	0.04433
KPSS Level	5.0584	5	0.01
KPSS Trend	0.2705	5	0.01

ADF Test

$$\begin{cases} H_0 : Non-stationary \\ H_a : Stationary \end{cases}$$

The P-Value of the ADF test is 0.04433 $<$ 0.05, therefore we fail to reject H0. Thus, the Original Time Series is stationary.

KPSS Level Test

$$\begin{cases} H_0 : Stationary \\ H_a : Non-Stationary \end{cases}$$

The P-Value of the ADF test is 0.01 $<$ 0.05, therefore we need to reject H0. Thus, the Original Time Series is non-stationary.

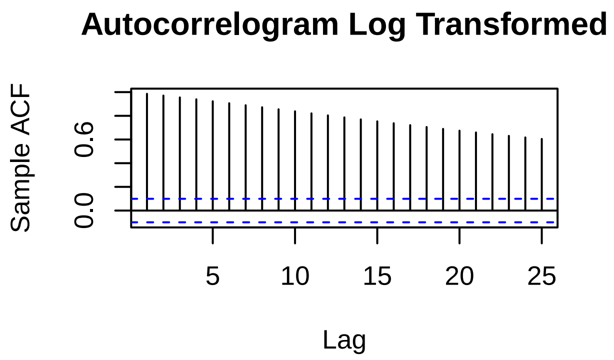
KPSS Trend Test

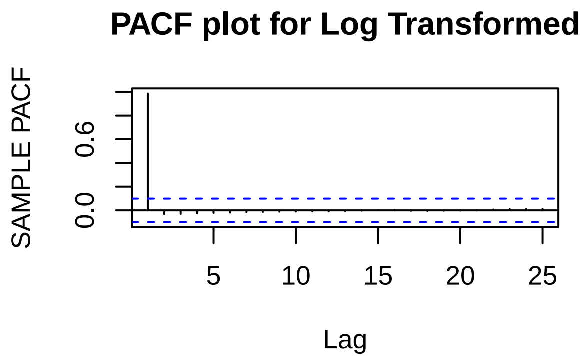
$$\begin{cases} H_0 : Trend-Stationary \\ H_a : Non-Trend-Stationary \end{cases}$$

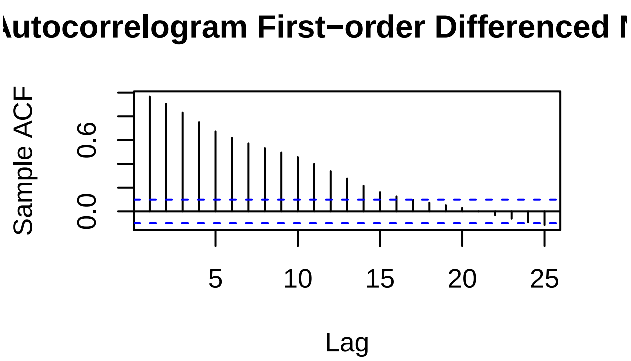
The P-Value of the ADF test is 0.01 $<$ 0.05, therefore we need to reject H0. Thus, the Original Time Series Not Trend-stationary.

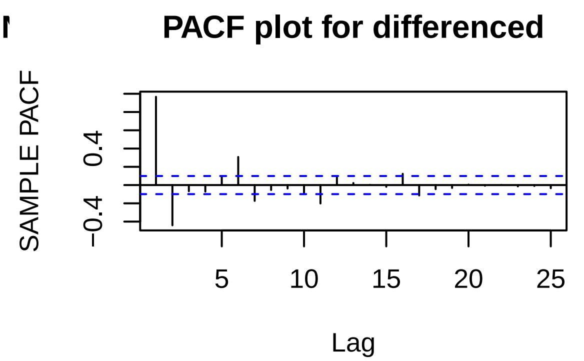
The results of the ADF and KPSS tests contradict each other. To make the time series stationary, we need to apply transformations or differencing techniques. Since there are no apparent patterns that repeat over time. We can explore mathematical transformations and differencing techniques first. In the following section, we will compare and determine which technique is more appropriate.

Log Transformation vs First-order Differenced









After analyzing the ACF and PACF plots of the Log Transformed and First-order Differenced Series, it was observed that the First-order Differenced Series has a steeper tail off pattern in the ACF plot, This indicates that differencing the series has removed more of the correlation from the data, making it a better fit for modeling than the log-transformed series. The PACF plot has values lying outside the threshold, notably at lag 1 and lag 6 Series is non-stationary, an ARIMA model will be considered.

Plot of Log Transformed Series see Appendix G

Plot of First-order Differenced Series see Appendix H

First-Order Difference

Table 8: Data in Appendix H

Tests	Test.Statistic	Lag.Order	P.Value
ADF	-3.0157	7	0.1483
KPSS Level	0.17975	5	0.1
KPSS Trend	0.17941	5	0.02372

Consistent with the findings of the Unit Root and Stationarity Tests performed on the original time series, we find that the First-order Differenced series is also non-stationary, as shown in Table 7.

Second-Order Difference

The Time Series plot reveals instability as the observations approach the right-hand side of the plot, near t=300 (see Appendix I), it suggests that the series is becoming more unpredictable. Furthermore, there is spikes at lag 5 and lag 10 in the ACF plot (see Appendix I.1), which may indicate a seasonal AR process of order 5 or 10.

Table 9: Data in Appendix I.2

Tests	Test.Statistic	Lag.Order	P.Value
ADF	-6.8955	7	0.01
KPSS Level	0.058337	5	0.1
KPSS Trend	0.024827	5	0.1

The results presented in Table 8 show that the model is stationary, as the p-value of the ADF test is 0.01, which is less than the significance level of 0.05, indicating that we can reject the null hypothesis and conclude that the model is stationary. Furthermore, the KPSS test yields similar results which now fail to reject the null hypothesis, suggesting that the model is stationary. However, the presence of a spike at lag 5 and 10 in the ACF plot, suggests that there may be some presence of a seasonal pattern or some other type of repeating pattern in the data that occurs every 5 or 10 time periods.

Selecting ARMA Model

The lowest (AIC) and (BIC) models will be selected.

Table 10: Data in Appendix I.4

Model	AIC	BIC
ARIMA(5,0,5)	2304.767	2346.403
ARIMA(3, 0, 5)	2319.594	2353.303
ARIMA(5,0,4)	2333.485	2371.157

Based on Table 9, which shows that the ARIMA(5,0,5) model has the lowest AIC and BIC values, making it the most promising candidate model. Furthermore, EACF table (see Appendix I.3) confirms its significance. Therefore, we will proceed the analysis using ARIMA(5,0,5).

Residual analysis

The standardized Residuals become unstable as we move towards the right-hand side, near t=300 (see Appendix I.7). This means that the model has adequately captured the trend of the data before t=300, but the instability near t=300 means that there is a possibility that we were not able to effectively more recent data. Also another possibility for this instability is due to natural variations, thus, we can not be certain if it is purely due to the ineffectiveness of this model. Furthermore, by inspecting the Ljung-Box statistic (see Appendix I.5), we see that only lag10 is significant. Thus, we may consider a Seasonal differencing of order 10

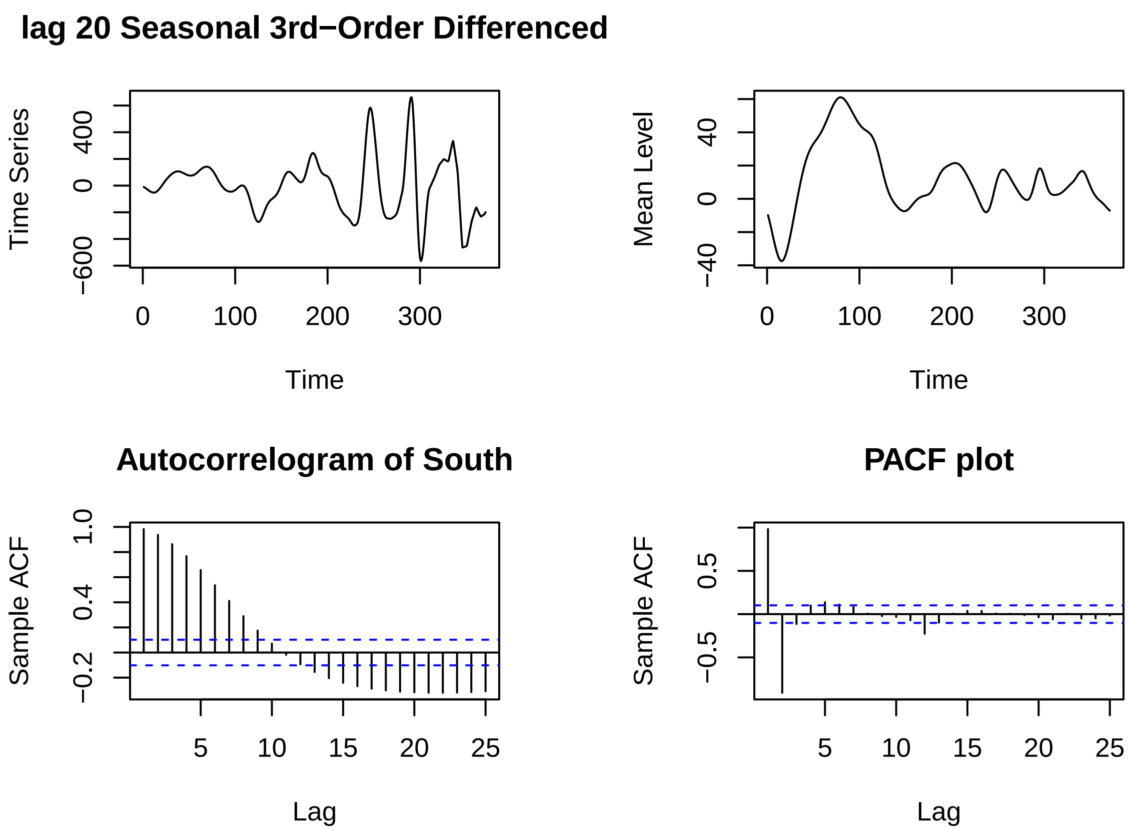
Table 11: Data in Appendix I.6

Shapiro.Wilk.normality.test	Value
W	0.58212
P-Value	< 2.2e-16

By inspecting the QQ Plot (see Appendix I.6), we may see that the middle values of residuals lie on the line of best-fit, with left and right ends tailing downwards and upwards away from the line of best-fit respectively. This indicates that the residuals may be somewhat normally distributed. We will use the Shapiro Wilk Normality test to confirm this.

From Table 6 we see that, Shapiro Wilk Normality test W has a value of 0.58212, which indicates some deviation from normality and the low p-value suggests that the deviation from normality is statistically significant. which means that there is sufficient evidence to reject the null hypothesis of normality in favor of an alternative hypothesis that the data is not normally distributed.

Seasonality



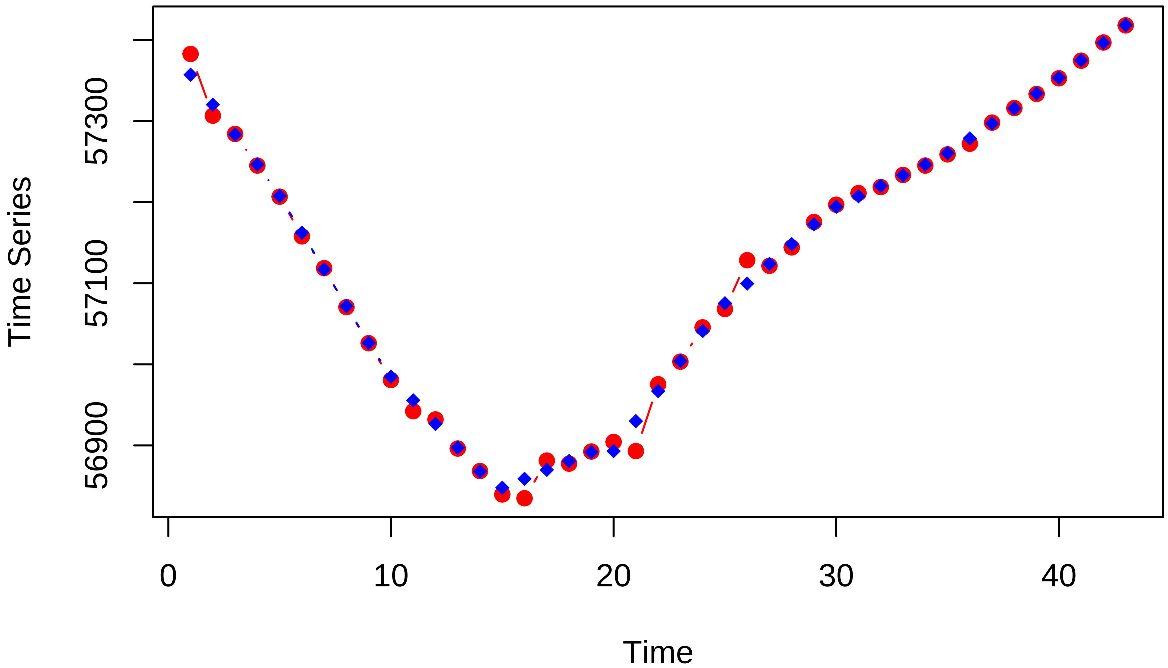
When we implement Seasonal Thrid-order Differenced of Order 10, we may see that the PACF plot spikes at Lag2 then a upward tail, then no other spikes at future lags, this means that there is no seasonality in our data. Thus, we may conclude that seasonality is not an appropriate method.

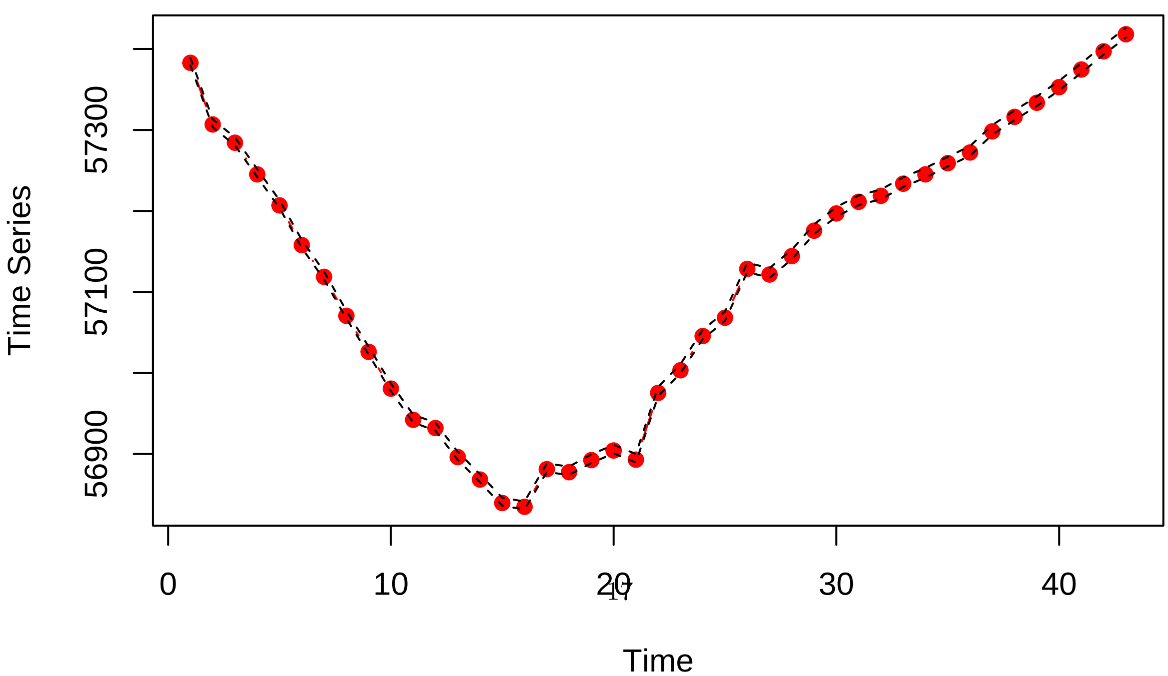
Even though the ARIMA(5,0,5) does not follow white-noise nor normal distribution. There is enough evidence to believe that it is in fact a good model and the instability that exists within the model may possibly be caused by natural variation.

Forecasting

North:

Model vs Data

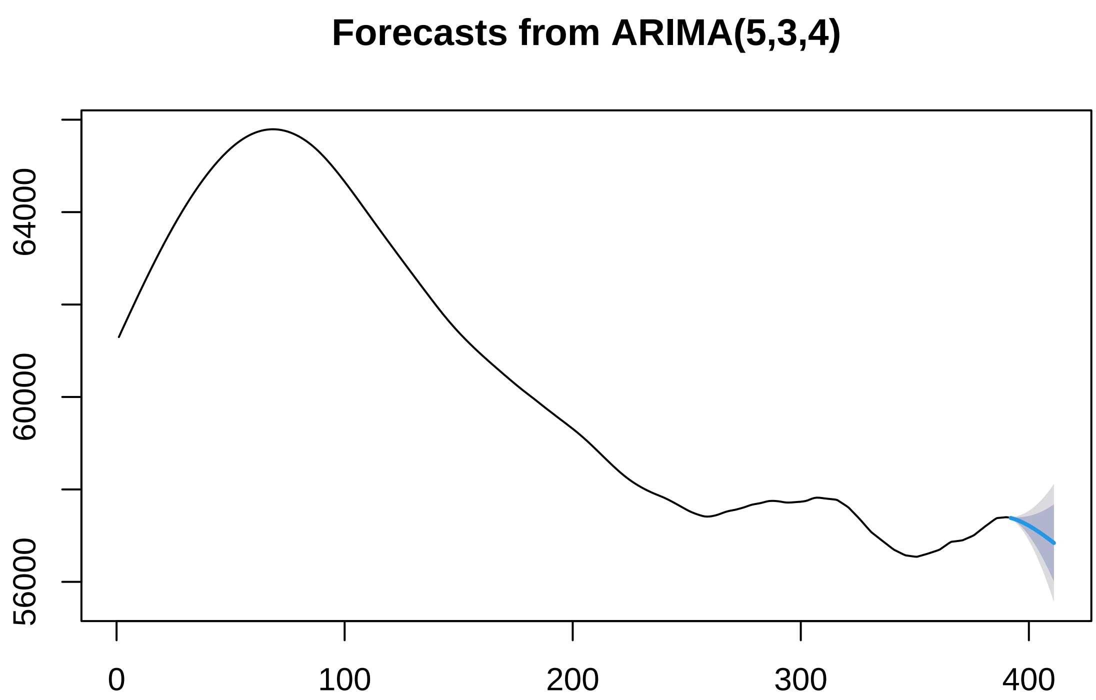




As mentioned, the deriving process used 90% of the data as training data.

In this graph, it uses the model to fit points onto the test data set and provides insight to how ARIMA(5,3,4) has performed in the test data set. Overall, the blue dots have closely followed the red dots, only with minor deviations neat t=16, t=21 and t=27. Thus, we mar see that ARIMA(5,3,4) has performed very well against this data set.

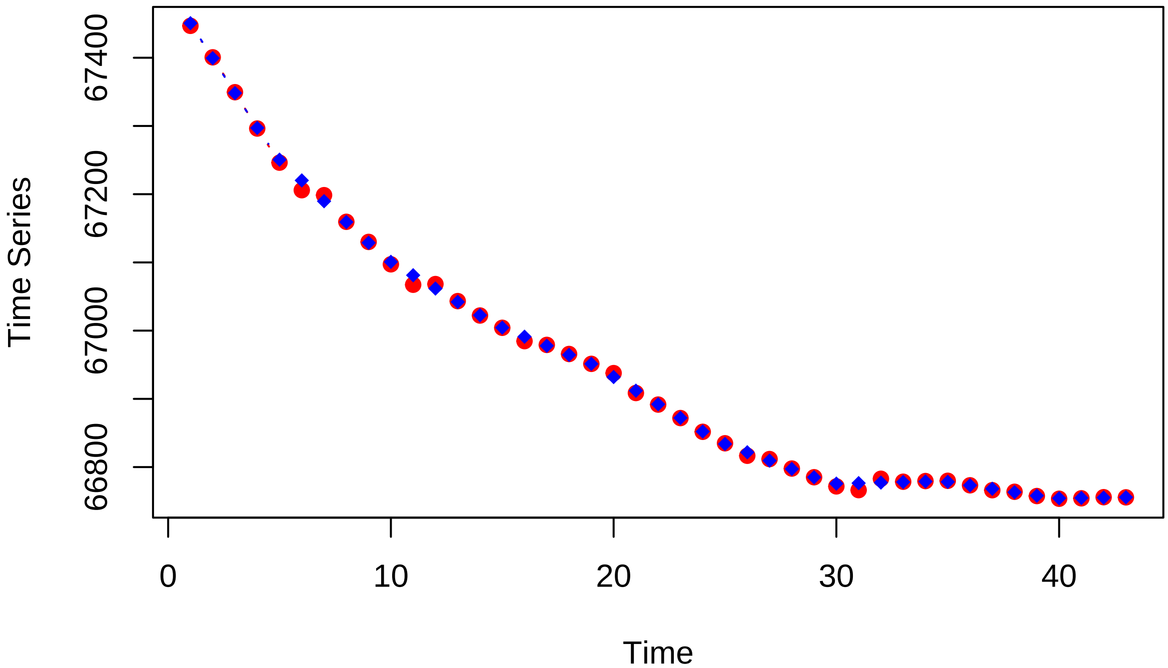
Forecast ARIMA(5,3,4)

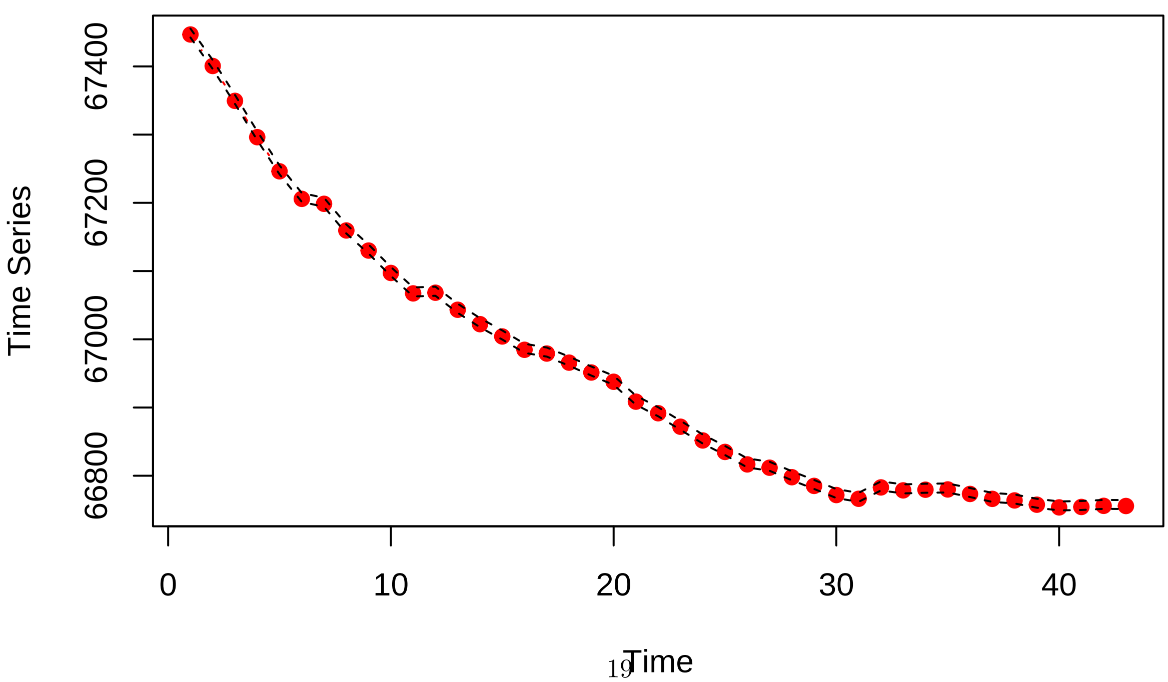


This graph only has the training data set as input, the remaining 10% are not given. The blue line is the prediction of how the North Pole Magnegtic field would behave, and the dark grey and grey areas are its prediction intervals. With the darker grey with more confidence. We may see that in this graph there is a tight prediction interval, indicating that there is more certainty and less deviation moving forward. Lastly, the last 10% of the Original time Series graph do indeed fall in the prediction interval. Thus, the model may be a good fit.

South:

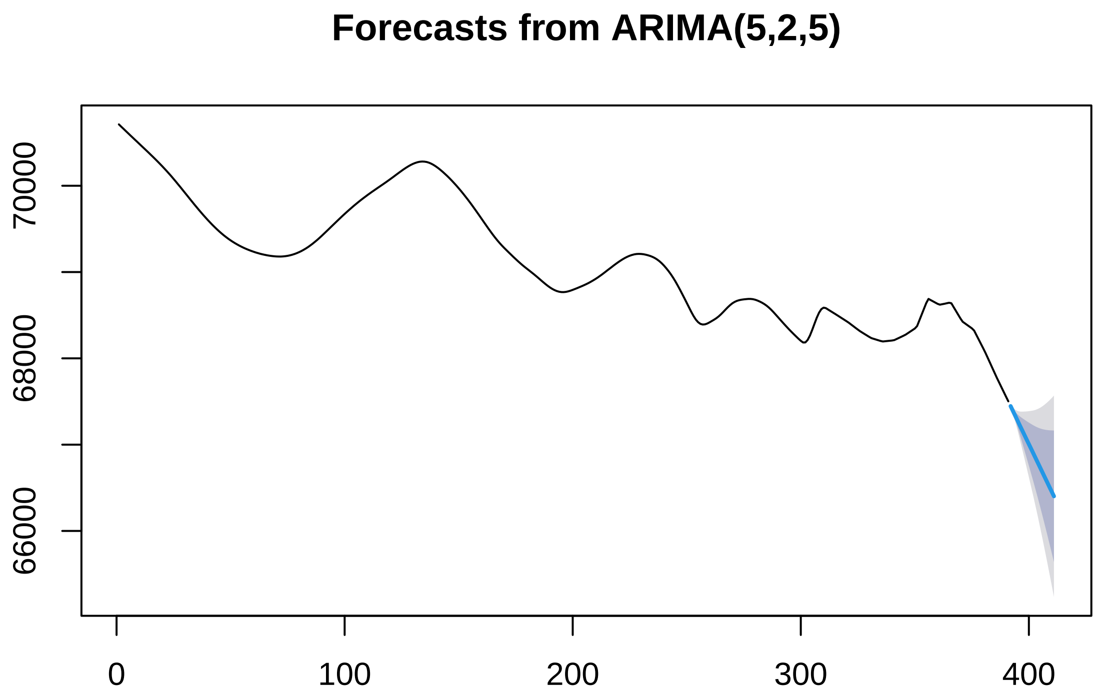
Model vs Data





In this graph, it uses the model to fit points onto the test data set and provides insight to how ARIMA(5,3,5) has performed in the test data set. Overall, the blue dots have closely followed the red dots, only with no obvious deviations. This model has performed better on South Pole data set compared to ARIMA(5,3,4) performed on the North Pole data set.

Forecast ARIMA(5,2,5)



This graph only has the training data set as input, the remaining 10% are not given. The blue line is the prediction of how the North Pole Magnegtic field would behave, and the dark grey and grey areas are its prediction intervals. With the darker grey with more confidence. We may see that in this graph there is a wider prediction interval when compared to the North Pole model, indicating that there is more turbulence and deviation moving forward. Lastly, the last 10% of the Original time Series graph do indeed fall in the prediction interval. Thus, the model may also be a good fit.

Discussion

Even though these models seem to have a good fit on both the North and South Pole data, there are still limitations and major flaws that are present. Firstly, the ARIMA model has its limitation in capturing complex patterns. Even though these two time series may not seem to have a very complex pattern, there is some important trends and patterns that were not captured due to the nature of ARIMA models. They are best suited for linear patterns, while these two data sets showed a downward oscillating trend. Thus, important information is lost during the deifferncings. Secondly, ARIMA models rely heavily on past observations, which makes predicting future values difficult and inaccurate when the data experiences shocks or sudden changes. Lastly, the fixed parameter nature that is determined by past data sets, do not provide the flexibility to adapt to future changes that is going or has already occurred.

Thus, even though the two ARIMA models has successfully provided accurate prediction intervals for the future 10% of data values. It will most possibly not provide accurate predictions when asking to predict 20% of data instead of 10%. Therefore, this model is very short sighted and requires a lot more complex components and tweaking in order for this model to perform up to industrial standards.

Bibliography

A chapter on Unit Root TestsDownload A chapter on Unit Root Tests (from Applied Econometric Time Series)

Summary of ADF and KPSSDownload Summary of ADF and KPSS, Dr. Masoud Ataei

Danielsson, Jon, Casper G. de Vries, and Björn Jónsson. "On the Distribution of Test Statistics in Unit Root Testing." Econometrica 67, no. 5 (1999): 1027-58. doi:10.1111/1468-0262.00064.

Brockwell, P. J., and R. A. Davis. Introduction to Time Series and Forecasting. 3rd ed., Springer, 2016.

Tsay, R. S. (2014). Multivariate time series analysis: With R and financial applications (3rd ed.). Wiley. Chapter 4: Stationarity and differencing.

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and practice (2nd ed.). OTexts.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. Time Series Analysis: Forecasting and Control. Wiley, 2015.

Appendices

N: North

Appendix A: Unit root and Stationarity Tests

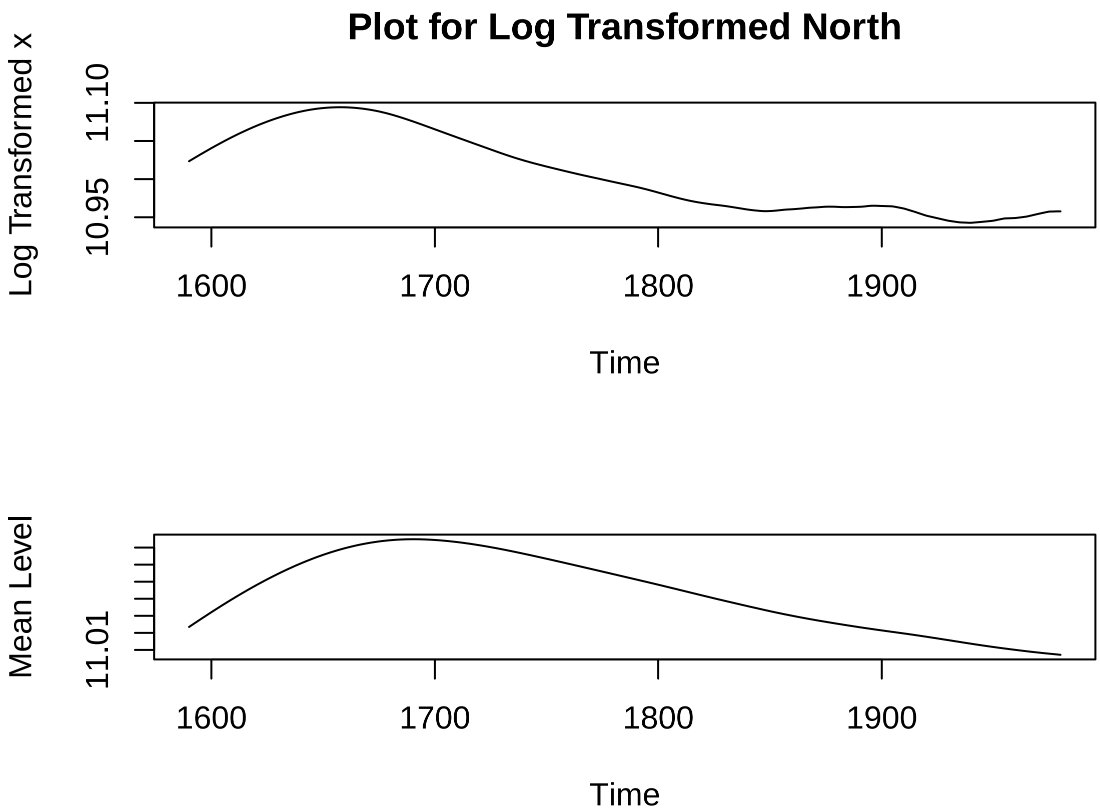
```
##
##  Augmented Dickey-Fuller Test
##
## data:  north
## Dickey-Fuller = -3.1294, Lag order = 7, p-value = 0.1003
## alternative hypothesis: stationary
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  north
## KPSS Level = 5.9459, Truncation lag parameter = 5, p-value = 0.01
```

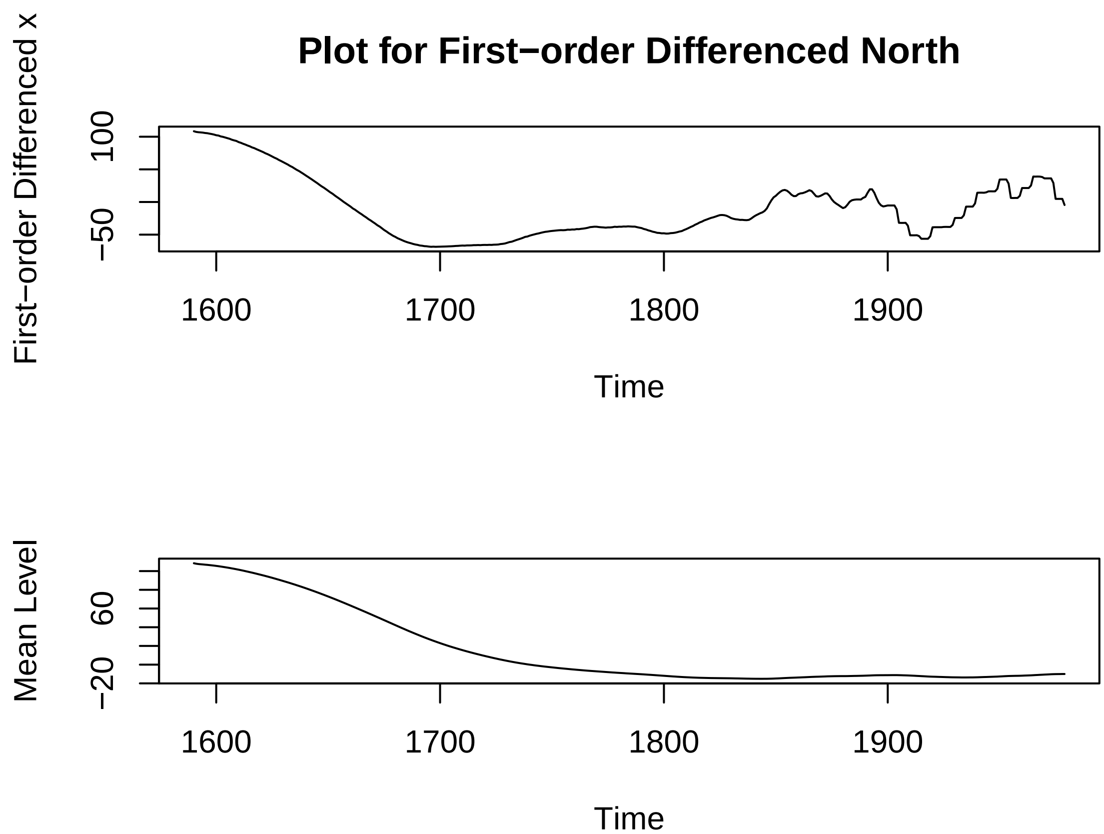
```
##
##  KPSS Test for Trend Stationarity
##
## data:  north
## KPSS Trend = 0.67056, Truncation lag parameter = 5, p-value = 0.01
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  north
## KPSS Level = 5.9459, Truncation lag parameter = 5, p-value = 0.01
```

Appendix B: Log Transformed series



Appendix C: First-order Differenced series



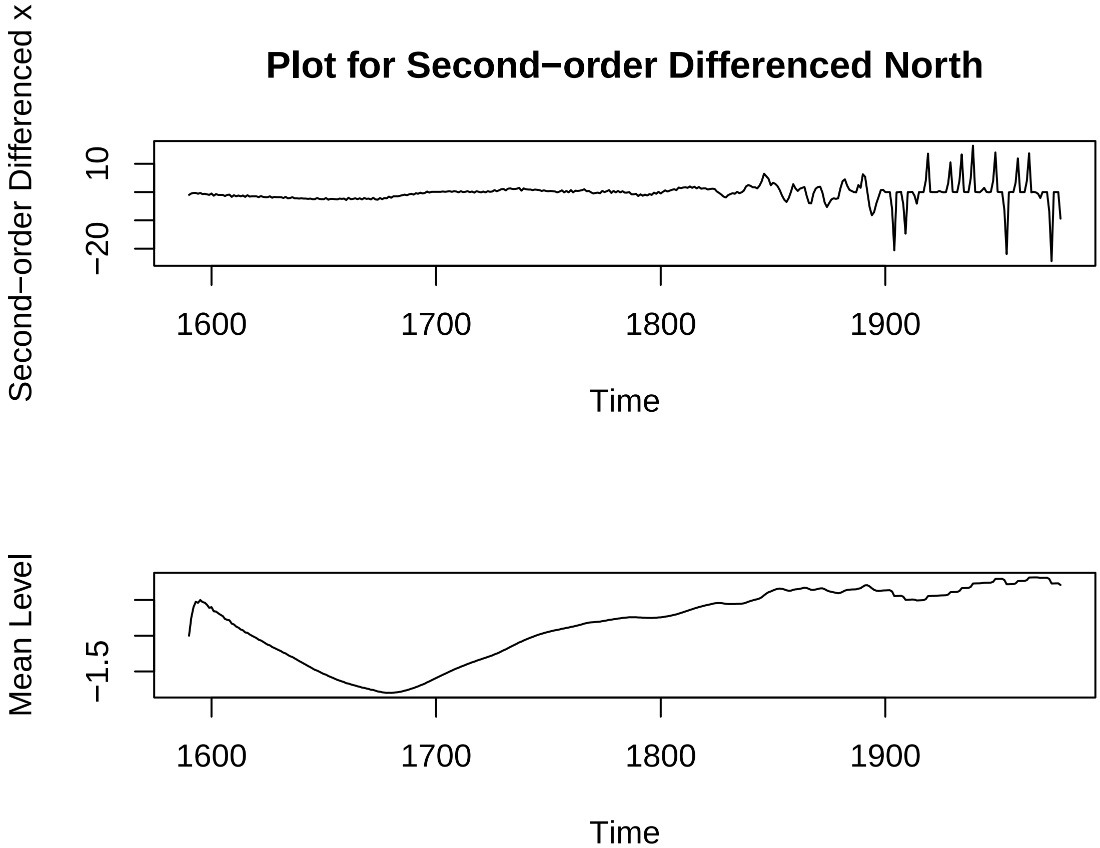
C.1: Unit root and Stationarity Tests

```
##
##  Augmented Dickey-Fuller Test
##
## data:  north_1
## Dickey-Fuller = -2.5263, Lag order = 7, p-value = 0.3549
## alternative hypothesis: stationary
```

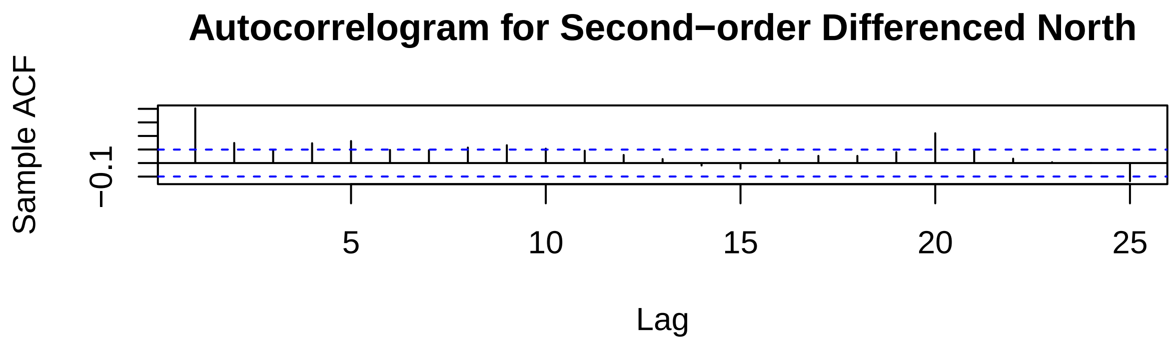
```
##
##  KPSS Test for Level Stationarity
##
## data:  north_1
## KPSS Level = 1.1898, Truncation lag parameter = 5, p-value = 0.01
```

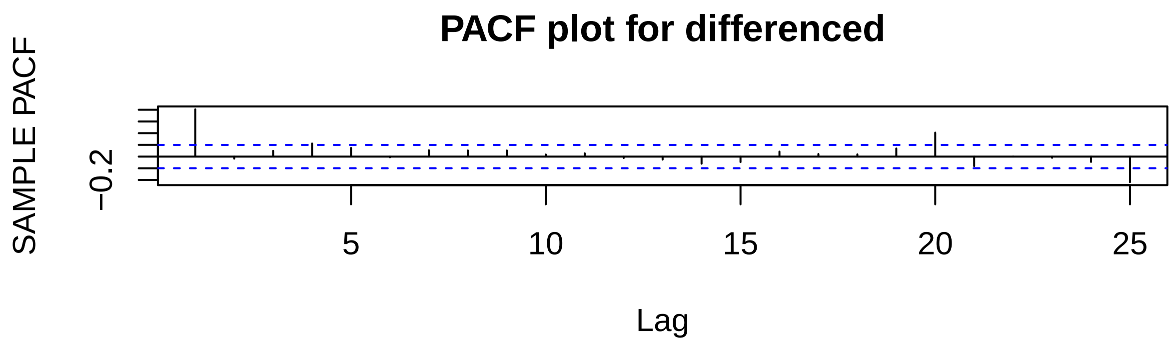
```
##
##  KPSS Test for Trend Stationarity
##
## data:  north_1
## KPSS Trend = 1.0953, Truncation lag parameter = 5, p-value = 0.01
```

Appendix D: Second-order Differenced



D.1: Sample ACF & PACF Plot





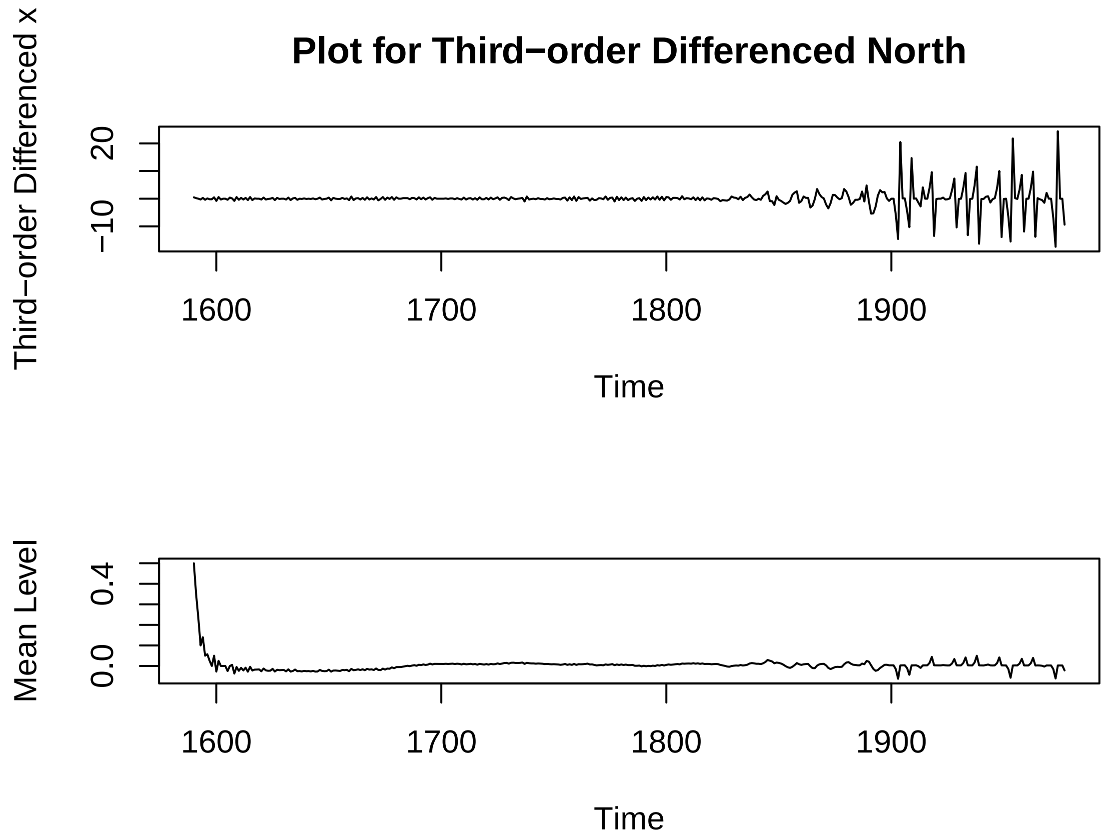
D.2: Unit root and Stationarity Tests

```
##
##  Augmented Dickey-Fuller Test
##
## data:  north_2
## Dickey-Fuller = -4.5429, Lag order = 7, p-value = 0.01
## alternative hypothesis: stationary
```

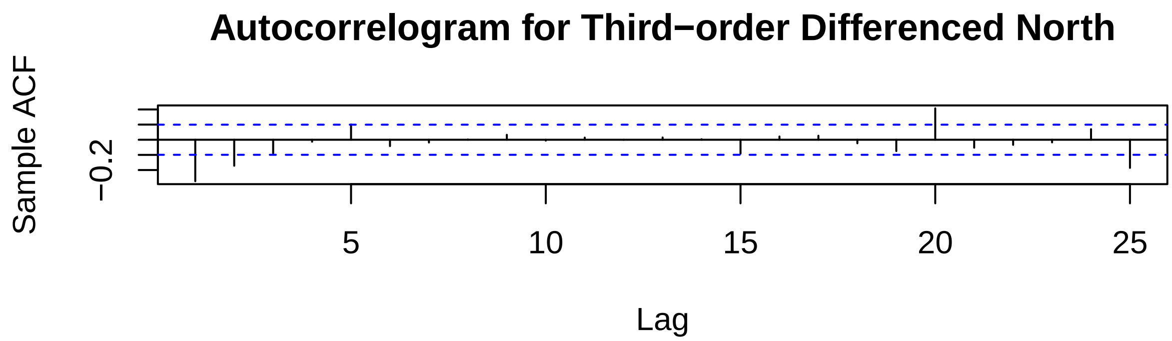
```
##
##  KPSS Test for Level Stationarity
##
## data:  north_2
## KPSS Level = 0.69819, Truncation lag parameter = 5, p-value = 0.01371
```

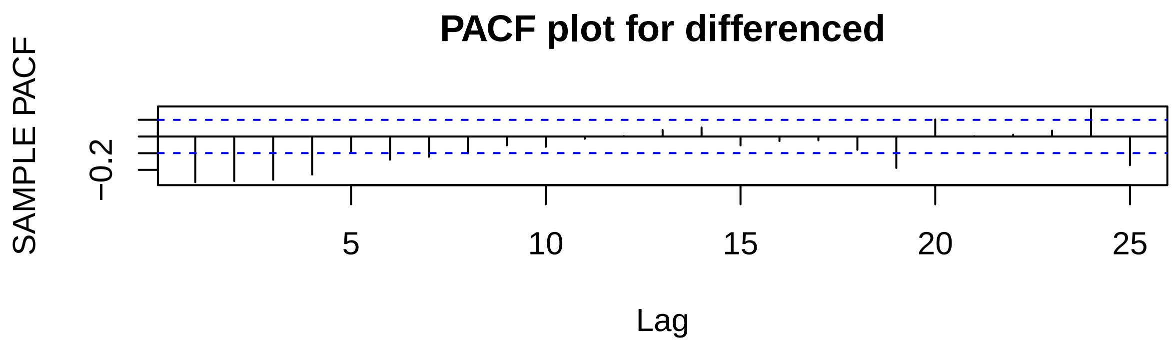
```
##
##  KPSS Test for Trend Stationarity
##
## data:  north_2
## KPSS Trend = 0.17089, Truncation lag parameter = 5, p-value = 0.02926
```

Appendix E: Third-order Differenced



E.1: Sample ACF & PACF Plot





E.2: Unit root and Stationarity Tests

```
##
##  Augmented Dickey-Fuller Test
##
## data:  north_3
## Dickey-Fuller = -11.433, Lag order = 7, p-value = 0.01
## alternative hypothesis: stationary
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  north_3
## KPSS Level = 0.034423, Truncation lag parameter = 5, p-value = 0.1
```

```
##
##  KPSS Test for Trend Stationarity
##
## data:  north_3
## KPSS Trend = 0.017776, Truncation lag parameter = 5, p-value = 0.1
```

E.3: EACF

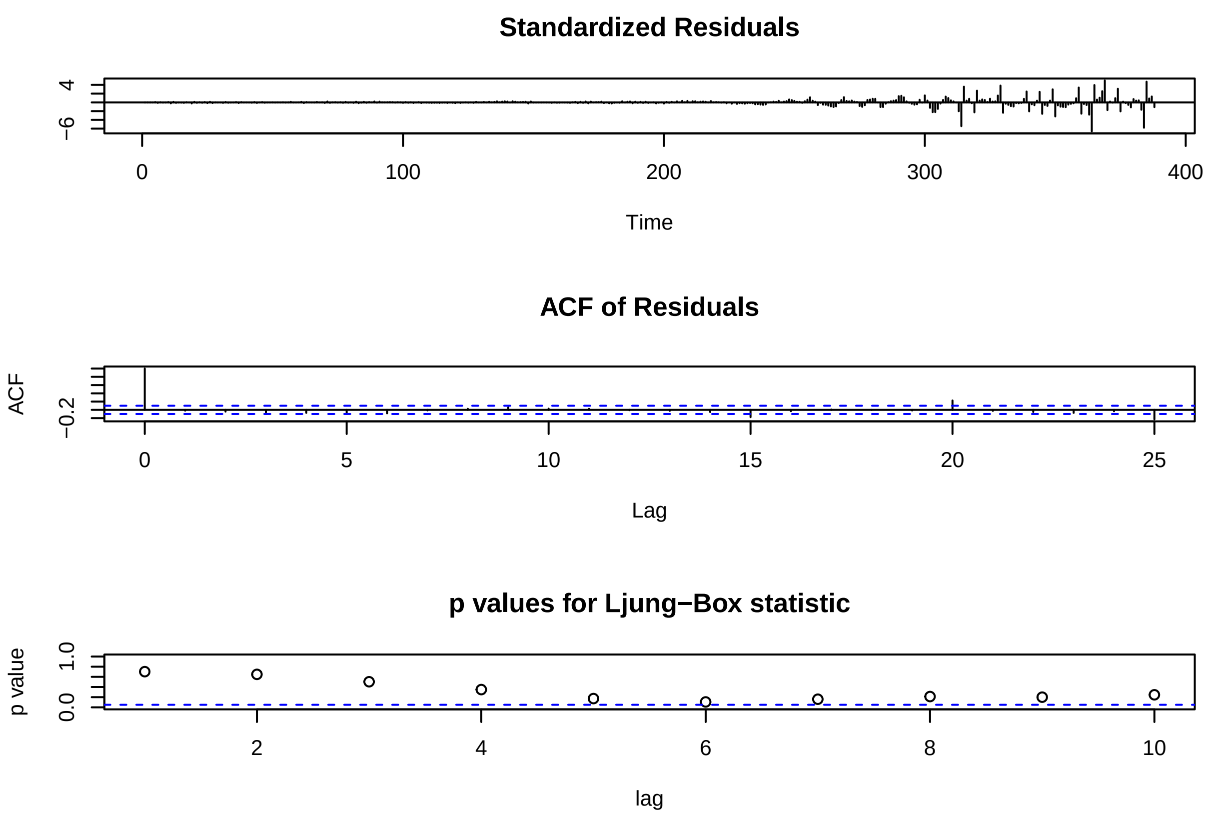
```
## AR/MA
##   0 1 2 3 4 5
## 0 x x o o o o
## 1 x o o o o o
## 2 x x o x o o
## 3 x o o x x o
## 4 x x x x o o
## 5 x x x x x o
```

E.4: AIC & BIC

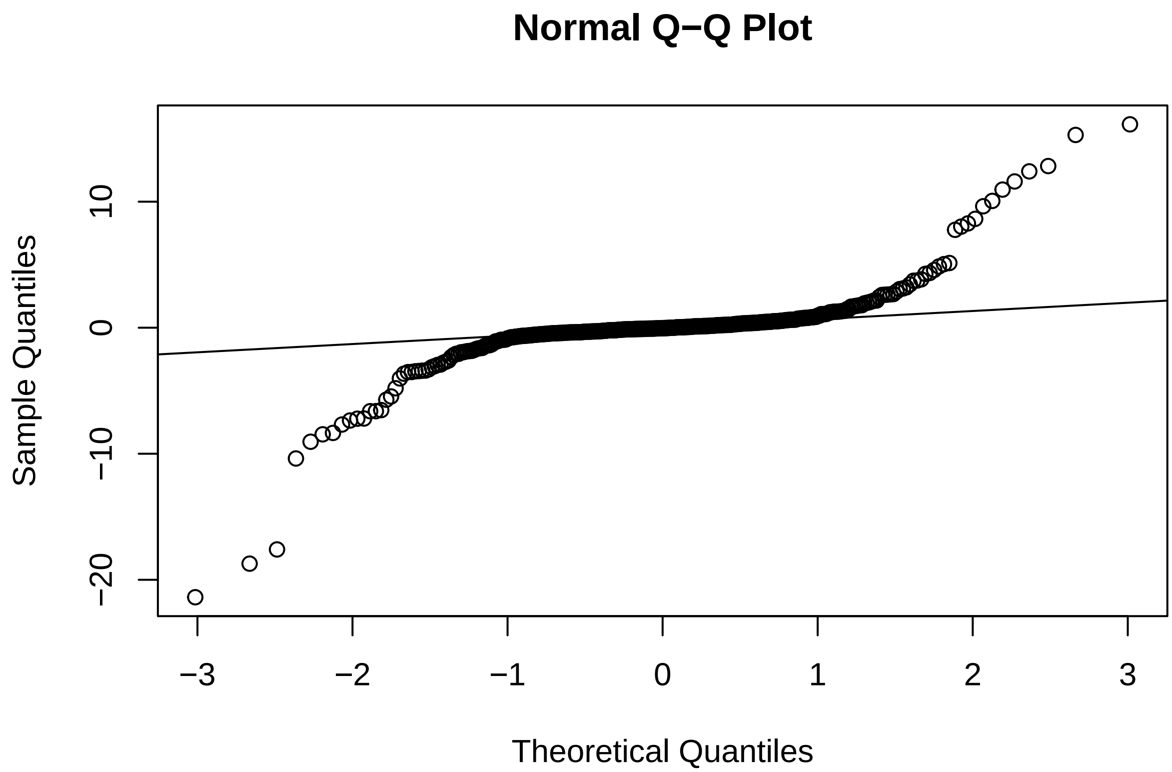
```
##    r4:c4    r3:c5    r5:c4    r4:c5    r5:c5    r5:c3    r1:c3    r2:c2
## 1965.756 1965.889 1966.435 1967.628 1968.935 1976.428 1976.604 1976.931
##    r3:c2    r1:c4    r2:c3    r5:c2    r1:c5
## 1977.239 1977.259 1977.578 1977.919 1978.908
```

```
##    r1:c3    r2:c2    r3:c2    r1:c4    r2:c3    r4:c4    r3:c5    r5:c4
## 1994.448 1994.775 1999.044 1999.064 1999.383 1999.444 1999.577 2004.084
##    r1:c5    r4:c2    r2:c4    r3:c3    r4:c5
## 2004.674 2004.852 2004.877 2004.958 2005.277
```

E.5: Standardized Residuals

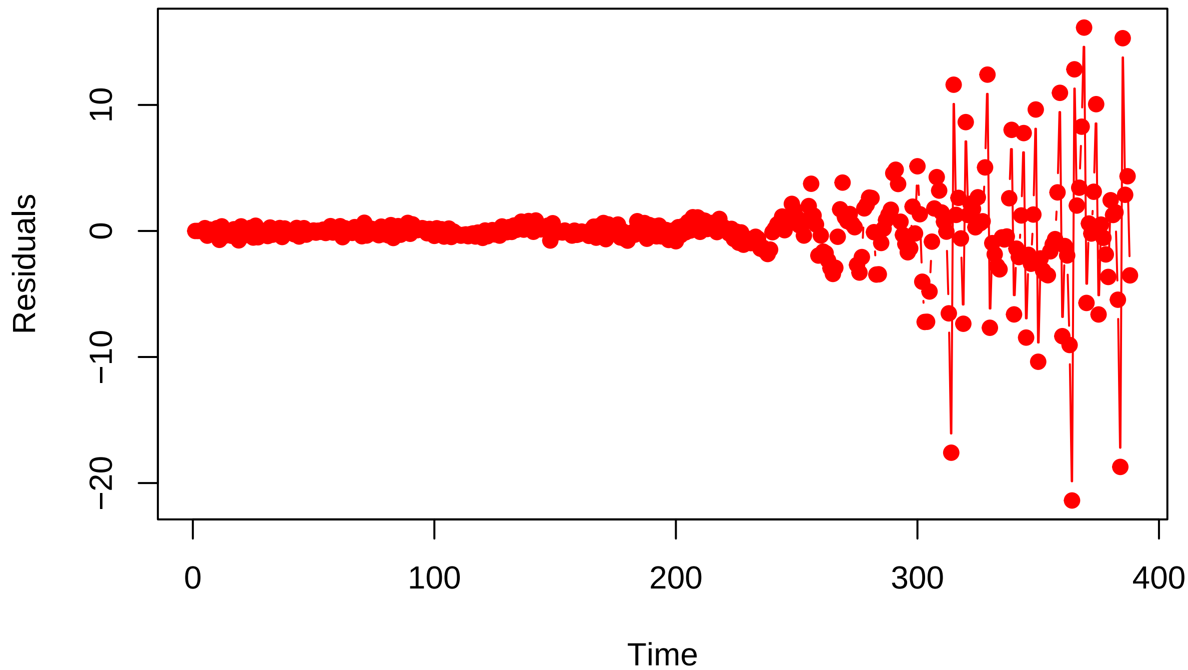


E.6: QQ Plot & Sgapiro Test




```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(north_3.fit_1)
## W = 0.67722, p-value < 2.2e-16
```

E.7: Residual Plot



South

Appendix F: Unit root and Stationarity

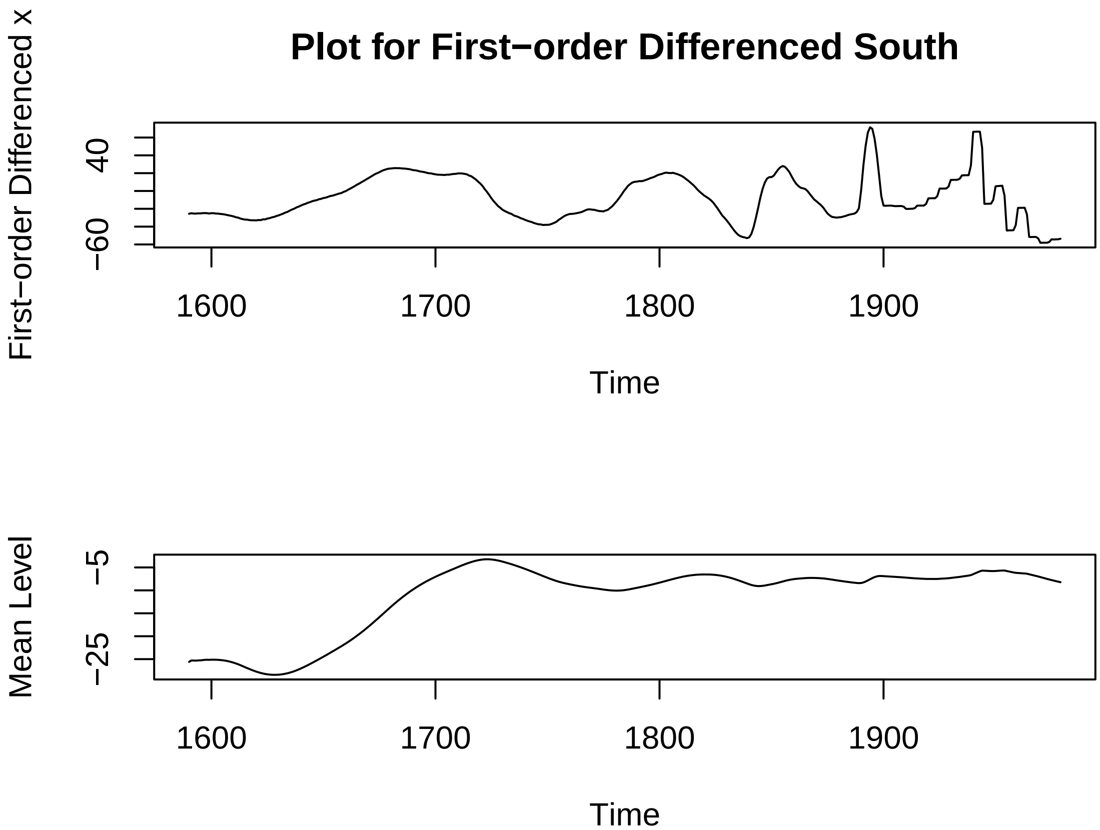
```
##
##  Augmented Dickey-Fuller Test
##
## data:  south
## Dickey-Fuller = -3.4834, Lag order = 7, p-value = 0.04433
## alternative hypothesis: stationary
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  south
## KPSS Level = 5.0584, Truncation lag parameter = 5, p-value = 0.01
```

```
##
##  KPSS Test for Trend Stationarity
##
## data:  south
## KPSS Trend = 0.2705, Truncation lag parameter = 5, p-value = 0.01
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  south
## KPSS Level = 5.0584, Truncation lag parameter = 5, p-value = 0.01
```

Appendix G: First-order Differenced series



G.1: ACF & Unit root and Stationarity Tests

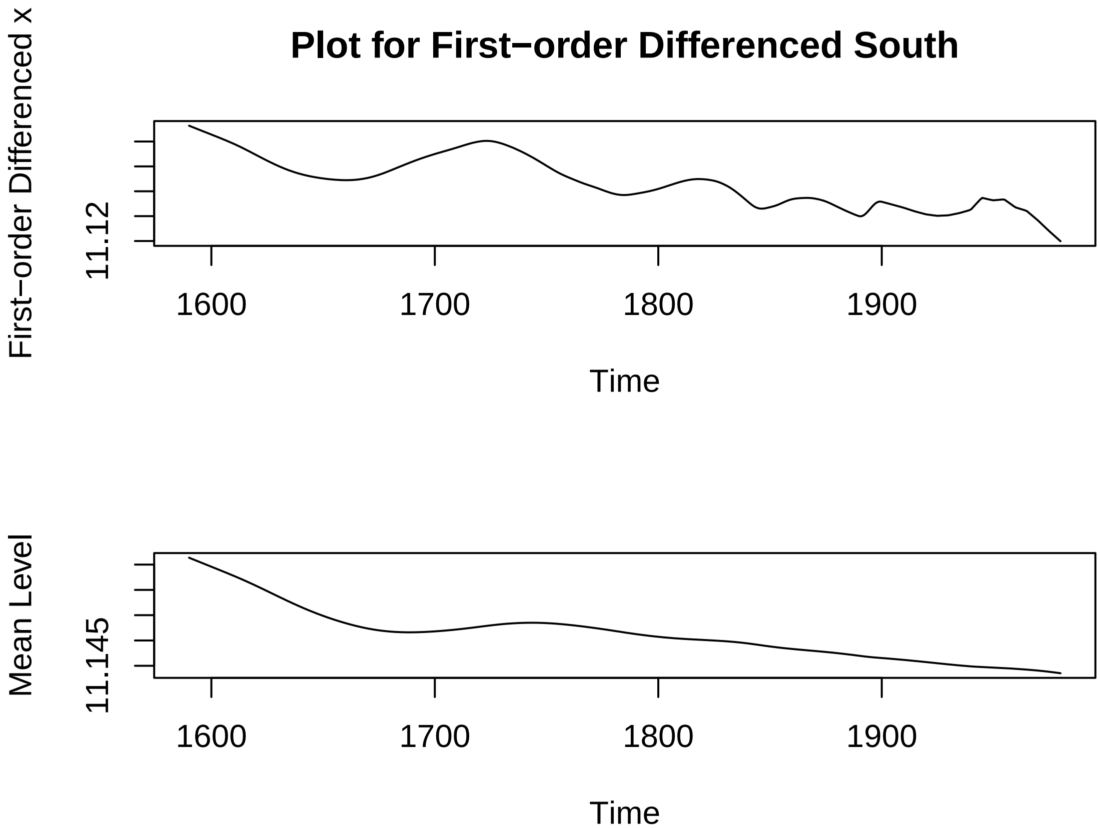
```
##
##  Augmented Dickey-Fuller Test
##
## data:  south_1
## Dickey-Fuller = -3.0157, Lag order = 7, p-value = 0.1483
## alternative hypothesis: stationary
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  south_1
## KPSS Level = 0.17975, Truncation lag parameter = 5, p-value = 0.1
```

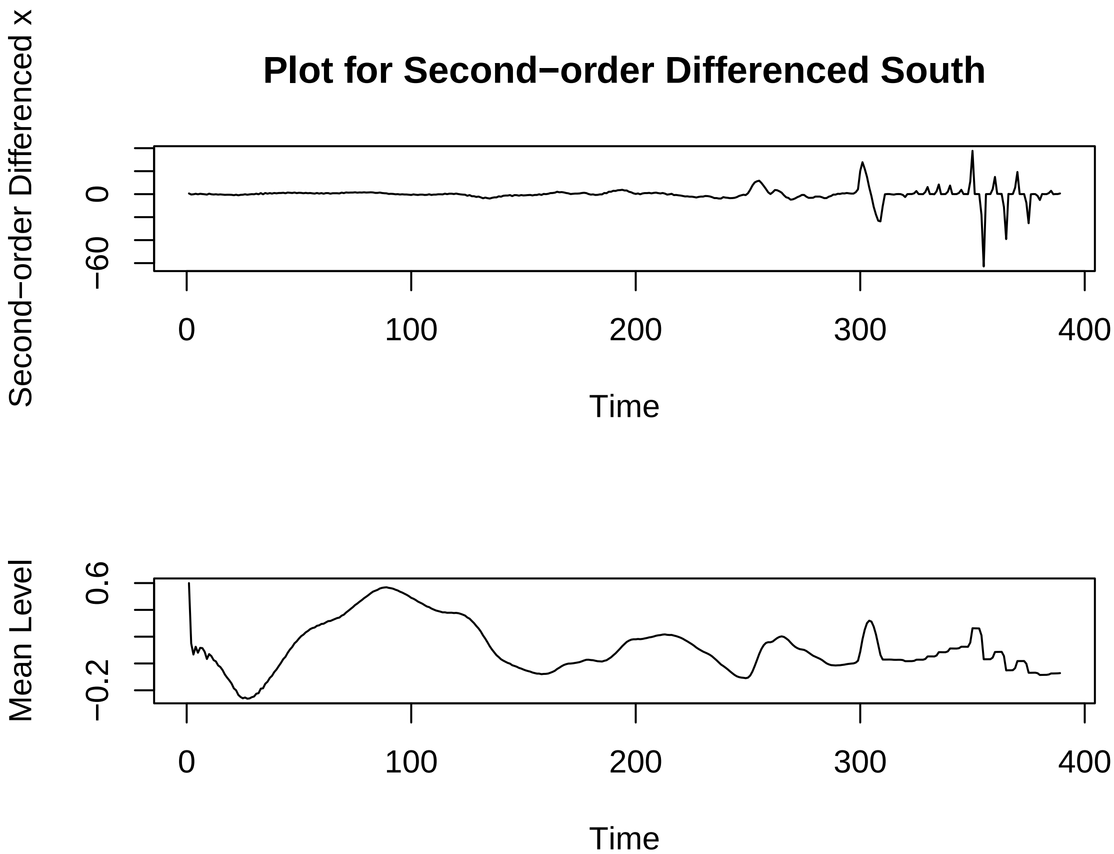
```
##
```

```
##  KPSS Test for Trend Stationarity
##
## data:  south_1
## KPSS Trend = 0.17941, Truncation lag parameter = 5, p-value = 0.02372
```

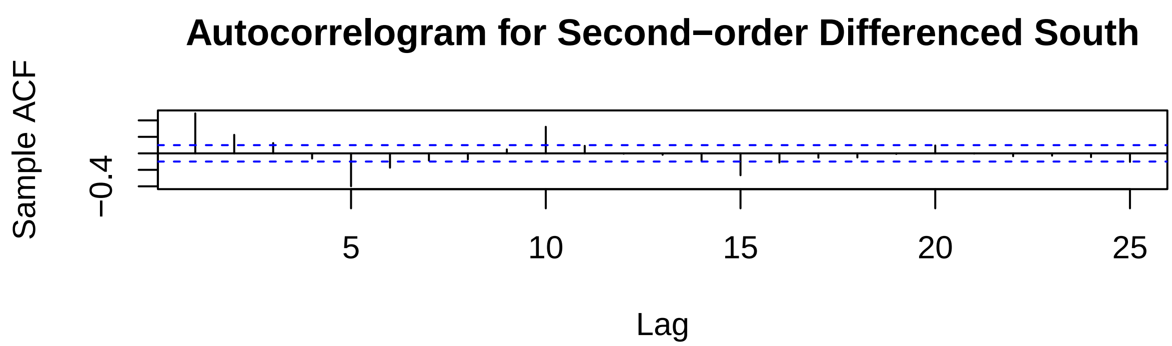
Appendix H: Log Transformed Series

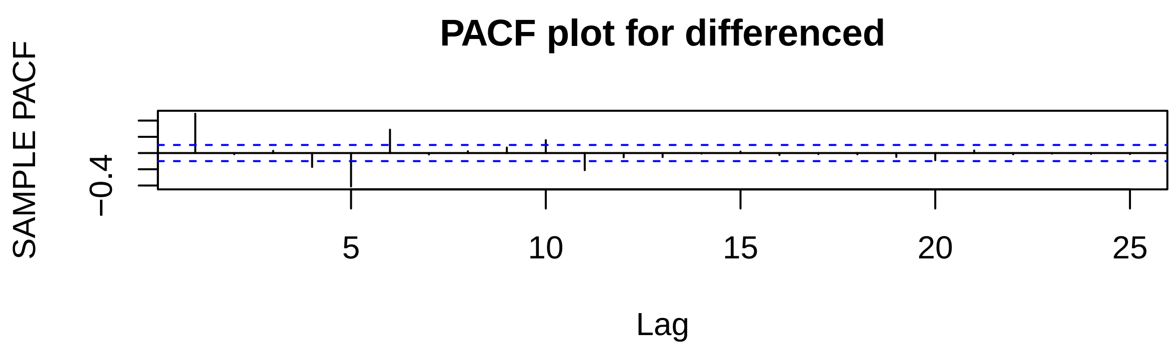


Appendix I: Second-order Differenced series



I.1: Sample ACF & PACF Plot





I.2: Unit root and Stationarity Tests

```
##
##  Augmented Dickey-Fuller Test
##
## data:  south_2
## Dickey-Fuller = -6.8955, Lag order = 7, p-value = 0.01
## alternative hypothesis: stationary
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  south_2
## KPSS Level = 0.058337, Truncation lag parameter = 5, p-value = 0.1
```

```
##
##  KPSS Test for Trend Stationarity
##
## data:  south_2
## KPSS Trend = 0.024827, Truncation lag parameter = 5, p-value = 0.1
```

I.3: EACF

```
## AR/MA
##   0 1 2 3 4 5
## 0 x x x o x x
## 1 o o x o x o
## 2 x o o o x x
## 3 x o o x x x
## 4 x x x x x x
## 5 x x x x x x
```

I.4: AIC & BIC

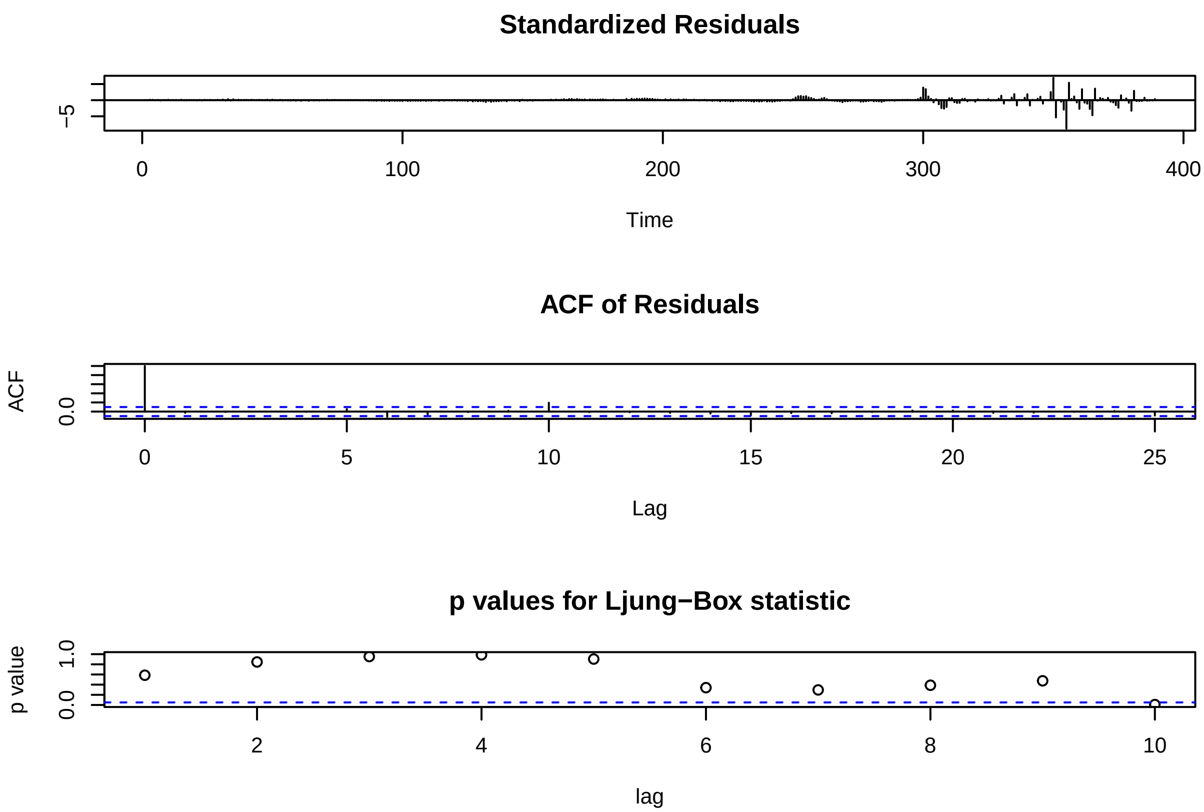
```
south_2.aic <- matrix(0,5,5)
south_2.bic <- matrix(0,5,5)

for (i in 0:4) for(j in 0:4) {
  south_2.fit <- arima(south_2, order=c(i,0,j), method = "ML", include.mean = TRUE)
  south_2.aic[i+1,j+1] <- south_2.fit$aic
  south_2.bic[i+1,j+1] <- BIC(south_2.fit)
}

south_2.aic_vec <- sort(unmatrix(south_2.aic, byrow=FALSE))[1:13]
south_2.bic_vec <- sort(unmatrix(south_2.bic, byrow=FALSE))[1:13]

south_2.aic_vec
```

```
##    r5:c5    r3:c5    r5:c4    r4:c5    r2:c5    r5:c3    r1:c5    r4:c4
## 2304.767 2319.594 2333.485 2337.723 2346.214 2348.318 2348.827 2350.886
##    r4:c3    r3:c4    r5:c2    r5:c1    r2:c4
## 2351.432 2355.928 2376.635 2391.971 2394.115
```

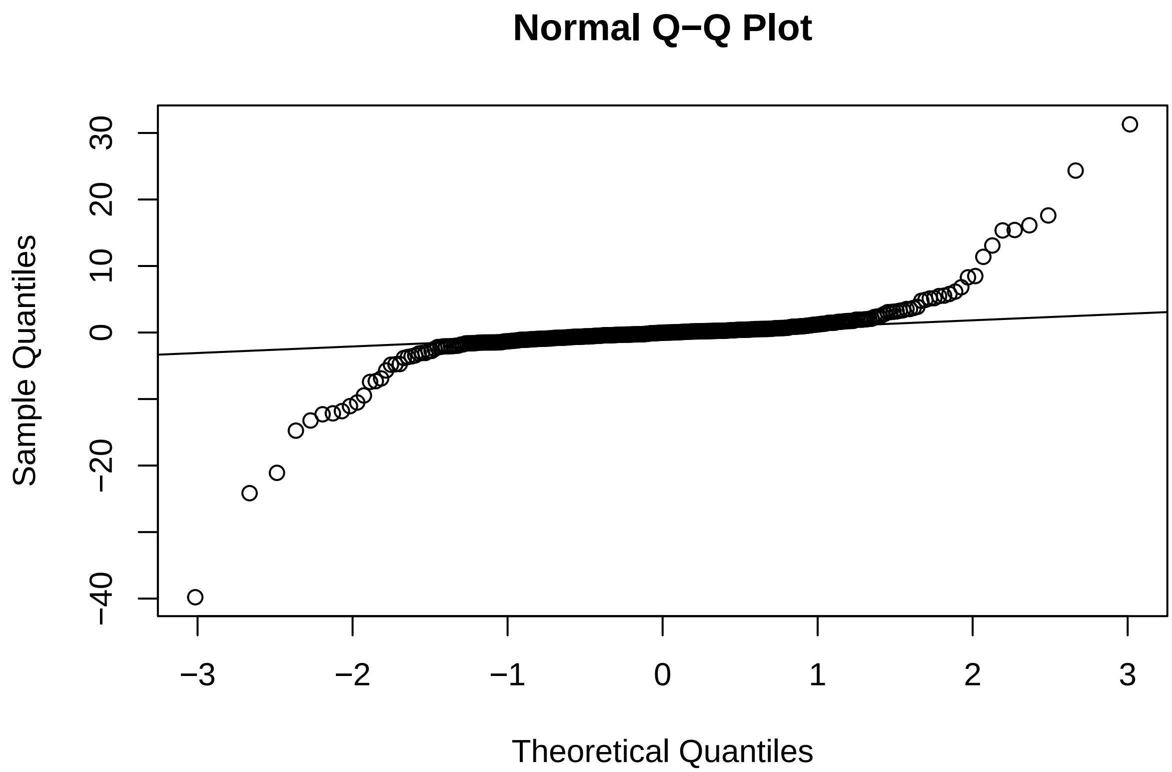
```
south_2.bic_vec
```

```
##    r5:c5    r3:c5    r5:c4    r1:c5    r4:c5    r2:c5    r4:c3    r5:c3
## 2346.403 2353.303 2371.157 2374.609 2375.395 2375.959 2381.177 2382.027
##    r4:c4    r3:c4    r5:c2    r2:c1    r2:c2
## 2384.594 2385.673 2406.380 2411.852 2417.723
```

I.5: Standardized Residuals

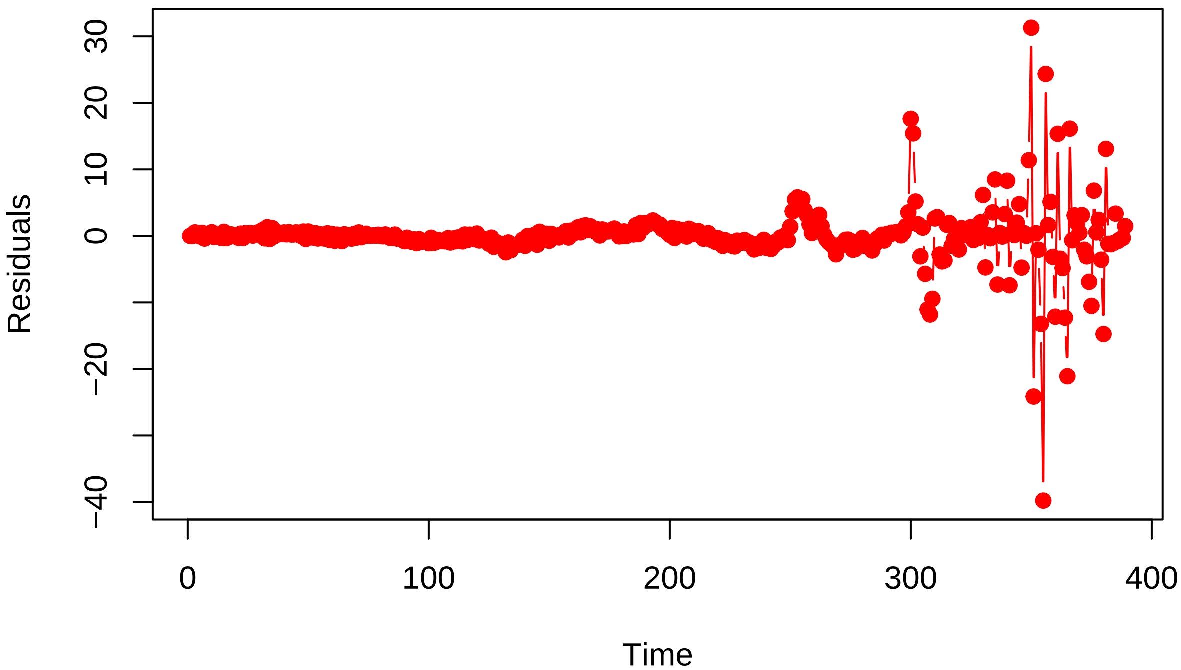


I.6: QQ Plot & Shapiro test




```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(south_2.fit_1)
## W = 0.58212, p-value < 2.2e-16
```

I.7: Residual Plot



Code:

```
knitr::opts_chunk$set(echo = TRUE)
# Library Used
library("TSA")
library("urca")
library("readxl")
library("tseries")
library("gdata")
library(forecast)
library(kableExtra)
library(knitr)
# Reading file
Geomagnetic_Intensity_Data <- read_excel("C:/Users/BBBBBB/Desktop/STA457 Project/Geomagnetic_Intensity_I
x <- Geomagnetic_Intensity_Data$North_Geomagnetic_Pole
n <- length(x)
pTest = 0.1
nTrain <- n - floor(n*pTest)
nTest <- n - nTrain

#Training data
north <- Geomagnetic_Intensity_Data$North_Geomagnetic_Pole[1:391]
```

```
south <- Geomagnetic_Intensity_Data$South_Geomagnetic_Pole[1:391]
t <- Geomagnetic_Intensity_Data$Year[1:391]

#Log Transformation
north_log <- log(north)
south_log <- log(south)

#First-oder Differenced
north_1 <- diff(north)
north_2 <- diff(north_1)
north_3 <- diff(north_2)

south_1 <- diff(south)
south_2 <- diff(south_1)

#Test data
north_t <- Geomagnetic_Intensity_Data$North_Geomagnetic_Pole[1:434]
south_t <- Geomagnetic_Intensity_Data$South_Geomagnetic_Pole[1:434]
#Time plot of Original series
cummeannorth <- cumsum(north)/seq_along(north)

par(mfrow=c(2,2),mar=c(4,4,4,4))
plot(t, north, type='l', xlab="Time", ylab="Time Series", main = "Original Time Series Plot")
plot(t, cummeannorth, type = 'l', xlab = "Time", ylab = 'Mean Level')
acf(north,xlab="Lag", ylab="Sample ACF", main="Autocorrelogram of North")
acf(north,type="partial",xlab="Lag", ylab="Sample ACF", main="PACF plot")
testTable <- data.frame('Tests' = c('ADF', 'KPSS Level', 'KPSS Trend'), 'Test Statistic' = c('-3.1294',
testTable %>% kable(caption = "Data in Appendix A", format = "markdown", booktabs = TRUE) %>% kable_styl
#ACF & PACF for Log Transformed North vs Original
par(mfrow=c(2,2))
acf(north_log, xlab="Lag", ylab="Sample ACF", main="Autocorrelogram Log Transformed")
acf(north_log, type = 'partial', xlab = "Lag", ylab = 'SAMPLE PACF', main="PACF plot for Log Transforme
acf(north_1, xlab="Lag", ylab="Sample ACF", main="Autocorrelogram First-order Differenced North")
acf(north_1, type = 'partial', xlab = "Lag", ylab = 'SAMPLE PACF', main="PACF plot for differenced")
testTable <- data.frame('Tests' = c('ADF', 'KPSS Level', 'KPSS Trend'), 'Test Statistic' = c('-2.5263',
testTable %>% kable(caption = 'Data in Appendix C.1', format = "markdown", booktabs = TRUE) %>% kable_s
testTable <- data.frame('Tests' = c('ADF', 'KPSS Level', 'KPSS Trend'), 'Test Statistic' = c('-4.5429',
testTable %>% kable(caption = 'Data in Appendix D.2', format = "markdown", booktabs = TRUE) %>% kable_s
testTable <- data.frame('Tests' = c('ADF', 'KPSS Level', 'KPSS Trend'), 'Test Statistic' = c('-11.433',
testTable %>% kable(caption = 'Data in Appendix E.2', format = "markdown", booktabs = TRUE) %>% kable_s
AICBICTable <- data.frame('Model' = c('ARIMA(4,0,4)', 'ARIMA(3, 0, 5)', 'ARIMA(5,0,4)'), 'AIC' = c('196
AICBICTable %>% kable(caption = 'Data in Appendix E.4', format = "markdown", booktabs = TRUE) %>% kable
testTable <- data.frame('Shapiro-Wilk normality test' = c('W', 'P-Value'), 'Value' = c('0.67722', '< 2.
testTable %>% kable(caption = 'Data in Appendix E.6', format = "markdown", booktabs = TRUE) %>% kable_s
north_s <- diff(north, differences = 3, lag = 20)
cummeannorth_s <- cumsum(north_s)/seq_along(north_s)
par(mfrow=c(2,2),mar=c(4,4,4,4))
plot( north_s, type='l', xlab="Time", ylab="Time Series", main = "lag 20 Seasonal 3rd-Order Differenced
plot(cummeannorth_s, type = 'l', xlab = "Time", ylab = 'Mean Level')
acf(north_s,xlab="Lag", ylab="Sample ACF", main="Autocorrelogram of South")
acf(north_s,type="partial",xlab="Lag", ylab="Sample ACF", main="PACF plot")
cummeansouth <- cumsum(south)/seq_along(south)
```

```
par(mfrow=c(2,2),mar=c(4,4,4,4))
plot(t, south, type='l', xlab="Time", ylab="Time Series", main = "Original Time Series Plot")
plot(t, cummeansouth, type = 'l', xlab = "Time", ylab = 'Mean Level')
acf(south,xlab="Lag", ylab="Sample ACF", main="Autocorrelogram of South")
acf(south,type="partial",xlab="Lag", ylab="Sample ACF", main="PACF plot")
testTable <- data.frame('Tests' = c('ADF', 'KPSS Level', 'KPSS Trend'), 'Test Statistic' = c('-3.4834',
testTable %>% kable(caption = "Data in Appendix F", format = "markdown", booktabs = TRUE) %>% kable_styl
#ACF & PACF for Log Transformed North vs Original
par(mfrow=c(2,2))
acf(south_log, xlab="Lag", ylab="Sample ACF", main="Autocorrelogram Log Transformed")
acf(south_log, type = 'partial', xlab = "Lag", ylab = 'SAMPLE PACF', main="PACF plot for Log Transforme
acf(south_1, xlab="Lag", ylab="Sample ACF", main="Autocorrelogram First-order Differenced North")
acf(south_1, type = 'partial', xlab = "Lag", ylab = 'SAMPLE PACF', main="PACF plot for differenced")
testTable <- data.frame('Tests' = c('ADF', 'KPSS Level', 'KPSS Trend'), 'Test Statistic' = c('-3.0157',
testTable %>% kable(caption = "Data in Appendix H", format = "markdown", booktabs = TRUE) %>% kable_styl
testTable <- data.frame('Tests' = c('ADF', 'KPSS Level', 'KPSS Trend'), 'Test Statistic' = c('-6.8955',
testTable %>% kable(caption = "Data in Appendix I.2", format = "markdown", booktabs = TRUE) %>% kable_st
AICBICTable <- data.frame('Model' = c('ARIMA(5,0,5)', 'ARIMA(3, 0, 5)', 'ARIMA(5,0,4)'), 'AIC' = c('230
AICBICTable %>% kable(caption = 'Data in Appendix I.4', format = "markdown", booktabs = TRUE) %>% kable_
testTable <- data.frame('Shapiro-Wilk normality test' = c('W', 'P-Value'), 'Value' = c('0.58212', '< 2.
testTable %>% kable(caption = 'Data in Appendix I.6', format = "markdown", booktabs = TRUE) %>% kable_st
north_s <- diff(south, differences = 2, lag = 10)
cummeannorth_s <- cumsum(north_s)/seq_along(north_s)
par(mfrow=c(2,2),mar=c(4,4,4,4))
plot( north_s, type='l', xlab="Time", ylab="Time Series", main = "lag 20 Seasonal 3rd-Order Differenced
plot(cummeannorth_s, type = 'l', xlab = "Time", ylab = 'Mean Level')
acf(north_s,xlab="Lag", ylab="Sample ACF", main="Autocorrelogram of South")
acf(north_s,type="partial",xlab="Lag", ylab="Sample ACF", main="PACF plot")
x <- Geomagnetic_Intensity_Data$North_Geomagnetic_Pole
x.test.forecast  <- matrix(0, nTest,1)
x.test.res <- matrix(0,nTest,1)
x.test.se <- matrix(0,nTest,1)
for (i in (nTrain : (n-1)) ){
  x.test.fit <- arima(x[1:i], order=c(5,3,4), method = "ML", include.mean=TRUE)
  x.test<- predict(x.test.fit, 1)
  x.test.forecast[i-nTrain+1,1] <- as.numeric(x.test$pred)
  x.test.res[i-nTrain+1,1] <- as.numeric(x.test$pred) - x[i+1]
  x.test.se[i-nTrain+1,1] <- as.numeric(x.test$se)
}
plot(unlist(x.test.forecast), type="b", pch=19, col="red", xlab="Time", ylab="Time Series")
lines(x[(nTrain+1):n], pch=18, col="blue", type="b", lty=2)

plot(unlist(x.test.forecast), type="b", pch=19, col="red", xlab="Time", ylab="Time Series")
lines(unlist(x.test.forecast)+unlist(x.test.se*1.96), pch=2, col="black", type="l", lty=2)
lines(unlist(x.test.forecast)-unlist(x.test.se), pch=2, col="black", type="l", lty=2)
final_model <- Arima(north, order=c(5,3,4), method = "ML", include.mean=TRUE)
forecast <- forecast(final_model, 20)
plot(forecast)
y <- Geomagnetic_Intensity_Data$South_Geomagnetic_Pole
y.test.forecast  <- matrix(0, nTest,1)
y.test.res <- matrix(0,nTest,1)
y.test.se <- matrix(0,nTest,1)
for (i in (nTrain : (n-1)) ){
```

```
  y.test.fit <- arima(y[1:i], order=c(5,2,5), method = "ML", include.mean=TRUE)
  y.test<- predict(y.test.fit, 1)
  y.test.forecast[i-nTrain+1,1] <- as.numeric(y.test$pred)
  y.test.res[i-nTrain+1,1] <- as.numeric(y.test$pred) - y[i+1]
  y.test.se[i-nTrain+1,1] <- as.numeric(y.test$se)
}
plot(unlist(y.test.forecast), type="b", pch=19, col="red", xlab="Time", ylab="Time Series")
lines(y[(nTrain+1):n], pch=18, col="blue", type="b", lty=2)

plot(unlist(y.test.forecast), type="b", pch=19, col="red", xlab="Time", ylab="Time Series")
lines(unlist(y.test.forecast)+unlist(y.test.se*1.96), pch=2, col="black", type="l", lty=2)
lines(unlist(y.test.forecast)-unlist(y.test.se), pch=2, col="black", type="l", lty=2)
final_model <- Arima(south, order=c(5,2,5), method = "ML", include.mean=TRUE)
forecast <- forecast(final_model, 20)
plot(forecast)
#Unit root and Stationarity Tests for Original Series North
north.adf <- adf.test(north, alternative="stationary")

north.kpss_level <- kpss.test(north, null="Level")
north.kpss_trend <- kpss.test(north, null="Trend")

north.adf
#pvalue 0.1003 > 0.05 -> reject H0 -> non-stationary

north.kpss_level
#Assuming that the null hypothesis is stationarity without drift: reject the null hypothesis and conclu

north.kpss_trend
# we reject the null hypothesis of trend-stationarity and conclude that the time series is non-stationa

kpss.test(north)
#Less then 0.05 -> Fail to reject H0 -> non-stationary
#Time plot for Log Transformed series North

cummeannorth_log <- cumsum(north_log)/seq_along(north_log)

par(mfrow=c(2,1),mar=c(4,4,4,4))
plot(t, north_log, type='l', xlab="Time", ylab="Log Transformed x", main="Plot for Log Transformed North
plot(t, cummeannorth_log, type = 'l', xlab = "Time", ylab = 'Mean Level')
#Time plot for First-order Differenced series North
cummeannorth_1 <- cumsum(north_1)/seq_along(north_1)

par(mfrow=c(2,1),mar=c(4,4,4,4))
plot(t[1:390],north_1, type='l', xlab="Time", ylab="First-order Differenced x", main="Plot for First-ord
plot(t[1:390],cummeannorth_1, type = 'l', xlab = "Time", ylab = 'Mean Level')
#Unit root and Stationarity Tests for First-order Differenced Series
north_1.adf <- adf.test(north_1, alternative="stationary")

north_1.kpss_level <- kpss.test(north_1, null="Level")
north_1.kpss_trend <- kpss.test(north_1, null="Trend")

north_1.adf
north_1.kpss_level
```

```
north_1.kpss_trend
#Time plot for Second-order Differenced series
north_2 <- diff(north_1)

cummeannorth_2 <- cumsum(north_2)/seq_along(north_2)

par(mfrow=c(2,1),mar=c(4,4,4,4))
plot(t[1:389],north_2, type='l', xlab="Time", ylab="Second-order Differenced x", main="Plot for Second-o
plot(t[1:389],cummeannorth_2, type = 'l', xlab = "Time", ylab = 'Mean Level')
par(mfrow=c(2,1))
acf(north_2, xlab="Lag", ylab="Sample ACF", main="Autocorrelogram for Second-order Differenced North")
acf(north_2, type = 'partial', xlab = "Lag", ylab = 'SAMPLE PACF', main="PACF plot for differenced")
#Unit root and Stationarity Tests for Second-order Differenced Series
north_2.adf <- adf.test(north_2, alternative="stationary")

north_2.kpss_level <- kpss.test(north_2, null="Level")
north_2.kpss_trend <- kpss.test(north_2, null="Trend")

north_2.adf
north_2.kpss_level
north_2.kpss_trend
#Time plot for Third-order Differenced series
north_3 <- diff(north_2)

cummeannorth_3 <- cumsum(north_3)/seq_along(north_3)

par(mfrow=c(2,1),mar=c(4,4,4,4))
plot(t[1:388], north_3, type='l', xlab="Time", ylab="Third-order Differenced x", main="Plot for Third-or
plot(t[1:388], cummeannorth_3, type = 'l', xlab = "Time", ylab = 'Mean Level')
par(mfrow=c(2,1))
acf(north_3, xlab="Lag", ylab="Sample ACF", main="Autocorrelogram for Third-order Differenced North")
acf(north_3, type = 'partial', xlab = "Lag", ylab = 'SAMPLE PACF', main="PACF plot for differenced")
#Unit root and Stationarity Tests for Third-order Differenced Series
north_3.adf <- adf.test(north_3, alternative="stationary")

north_3.kpss_level <- kpss.test(north_3, null="Level")
north_3.kpss_trend <- kpss.test(north_3, null="Trend")

north_3.adf
#Reject H0 -> Model is stationary
north_3.kpss_level
#Fail to reject H0 -> Model is stationary
north_3.kpss_trend
#EACF for Third-order Differenced
north_3.eacf <- eacf(north_3, ar.max = 5, ma.max = 5)
north_3.aic <- matrix(0,5,5)
north_3.bic <- matrix(0,5,5)

for (i in 0:4) for(j in 0:4) {
  north_3.fit <- arima(north_3, order=c(i,0,j), method = "ML", include.mean = TRUE)
  north_3.aic[i+1,j+1] <- north_3.fit$aic
  north_3.bic[i+1,j+1] <- BIC(north_3.fit)
}
```

```
north_3.aic_vec <- sort(unmatrix(north_3.aic, byrow=FALSE))[1:13]
north_3.bic_vec <- sort(unmatrix(north_3.bic, byrow=FALSE))[1:13]

north_3.aic_vec
north_3.bic_vec
#Fitted model ARIMA(5,3,4)
north_3.fit_1 = arima(north_3, order = c(5,3,4), method = "ML", include.mean = TRUE)
tsdiag(north_3.fit_1)
qqnorm(residuals(north_3.fit_1))
qqline(residuals(north_3.fit_1))

shapiro.test(residuals(north_3.fit_1))
plot(unlist(north_3.fit_1$residuals), type="b", pch=19, col="red", xlab="Time", ylab="Residuals")
#Unit root and Stationarity Tests for Original Series South
south.adf <- adf.test(south, alternative="stationary")

south.kpss_level <- kpss.test(south, null="Level")
south.kpss_trend <- kpss.test(south, null="Trend")

south.adf
#pvalue 0.1003 > 0.05 -> reject H0 -> non-stationary

south.kpss_level
#Assuming that the null hypothesis is stationarity without drift: reject the null hypothesis and conclu

south.kpss_trend
# we reject the null hypothesis of trend-stationarity and conclude that the time series is non-stationa

kpss.test(south)
#Less then 0.05 -> Fail to reject H0 -> non-stationary
#Time plot for First-order Differenced series North
cummeansouth_1 <- cumsum(south_1)/seq_along(south_1)

par(mfrow=c(2,1),mar=c(4,4,4,4))
plot(t[1:390],south_1, type='l', xlab="Time", ylab="First-order Differenced x", main="Plot for First-or
plot(t[1:390],cummeansouth_1, type = 'l', xlab = "Time", ylab = 'Mean Level')
#Unit root and Stationarity Tests for First-order Differenced Series
south_1.adf <- adf.test(south_1, alternative="stationary")

south_1.kpss_level <- kpss.test(south_1, null="Level")
south_1.kpss_trend <- kpss.test(south_1, null="Trend")

south_1.adf
south_1.kpss_level
south_1.kpss_trend
#Time plot for First-order Differenced series North
cummeansouth_log <- cumsum(south_log)/seq_along(south_log)

par(mfrow=c(2,1),mar=c(4,4,4,4))
plot(t[1:391],south_log, type='l', xlab="Time", ylab="First-order Differenced x", main="Plot for First-
plot(t[1:391],cummeansouth_log, type = 'l', xlab = "Time", ylab = 'Mean Level')
#Time plot for Second-order Differenced series South
south_2 <- diff(south_1)
```

```
cummeansouth_2 <- cumsum(south_2)/seq_along(south_2)

par(mfrow=c(2,1),mar=c(4,4,4,4))
plot(south_2, type='l', xlab="Time", ylab="Second-order Differenced x", main="Plot for Second-order Dif
plot(cummeansouth_2, type = 'l', xlab = "Time", ylab = 'Mean Level')
par(mfrow=c(2,1))
acf(south_2, xlab="Lag", ylab="Sample ACF", main="Autocorrelogram for Second-order Differenced South")
acf(south_2, type = 'partial', xlab = "Lag", ylab = 'SAMPLE PACF', main="PACF plot for differenced")
#Unit root and Stationarity Tests for Second-order Differenced Series South
south_2.adf <- adf.test(south_2, alternative="stationary")

south_2.kpss_level <- kpss.test(south_2, null="Level")
south_2.kpss_trend <- kpss.test(south_2, null="Trend")

south_2.adf
south_2.kpss_level
south_2.kpss_trend
#EACF for Second-order Differenced South
south_2.eacf <- eacf(south_2, ar.max = 5, ma.max = 5)
south_2.aic <- matrix(0,5,5)
south_2.bic <- matrix(0,5,5)

for (i in 0:4) for(j in 0:4) {
  south_2.fit <- arima(south_2, order=c(i,0,j), method = "ML", include.mean = TRUE)
  south_2.aic[i+1,j+1] <- south_2.fit$aic
  south_2.bic[i+1,j+1] <- BIC(south_2.fit)
}

south_2.aic_vec <- sort(unmatrix(south_2.aic, byrow=FALSE))[1:13]
south_2.bic_vec <- sort(unmatrix(south_2.bic, byrow=FALSE))[1:13]

south_2.aic_vec
south_2.bic_vec
#Fitted model ARIMA(5,2,5)
south_2.fit_1 = arima(south_2, order = c(5,2,5), method = "ML", include.mean = TRUE)
tsdiag(south_2.fit_1)
qqnorm(residuals(south_2.fit_1))
qqline(residuals(south_2.fit_1))

shapiro.test(residuals(south_2.fit_1))
plot(unlist(south_2.fit_1$residuals), type="b", pch=19, col="red", xlab="Time", ylab="Residuals")
```