STA 137 Project

Riley Adams, Samuel Van Gorden, Phillip Yu

March 2022

Riley Adams ID: 918343985

Samuel Van Gorden ID: 995998955

Phillip Yu ID: 915922558

1 Introduction

Our given data is annual temperature anomalies from the northern hemisphere, recorded from 1850 to 2019. We find importance in analyzing temperature anomaly data due to our changing climate. We plot the data across time. Since the data is a series indexed by time it is considered a time series. The plot of annual temperature anomalies from 1850-2019 is shown below.

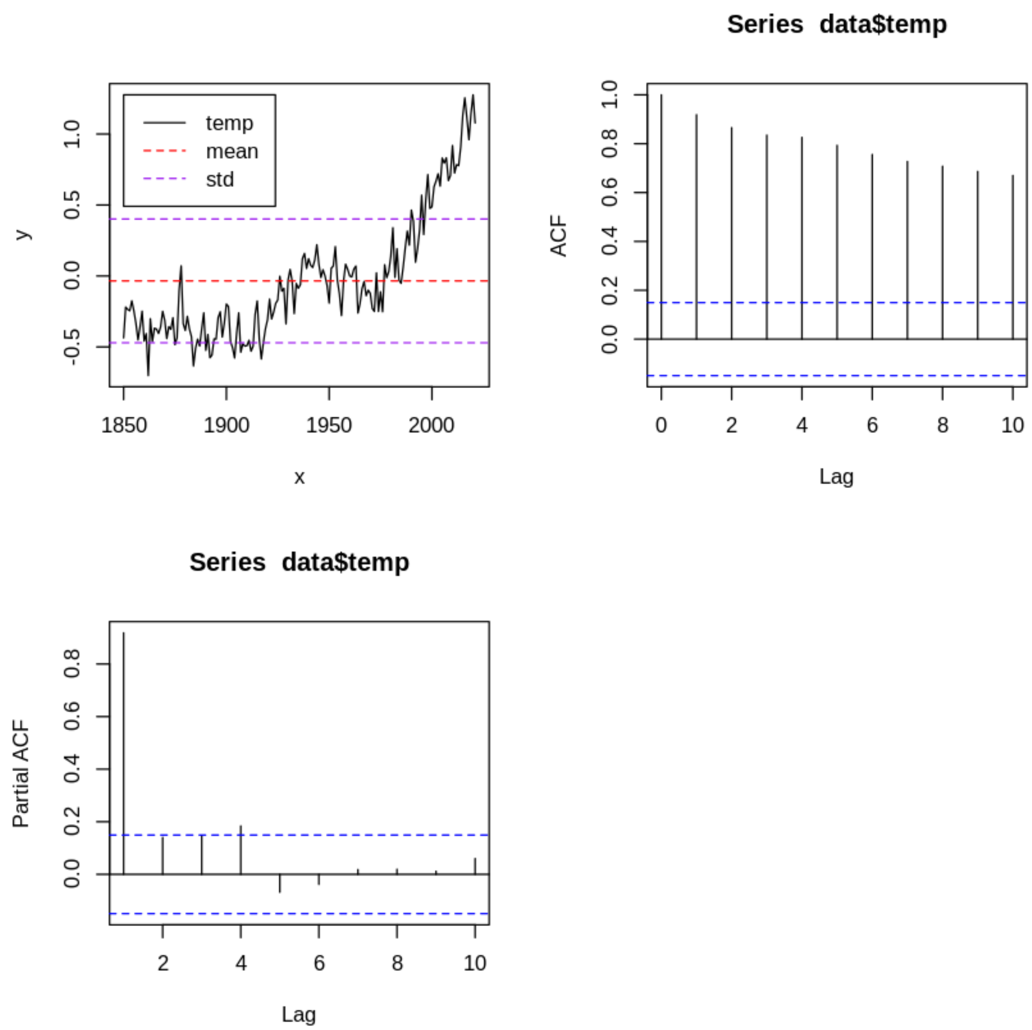

Figure 1: We plot annual temperature anomalies from 1850 to 2019.

Based on the time series plot in the upper left, it suggests that the mean and variance is not constant. The ACF plot also shows significant auto-correlation across many lags. Therefore, the temperature anomalies data does not follow weak stationarity.

2 Trend Fit of Data

Since the temperature anomalies data does not follow stationarity we decide to examine the trend using local polynomial regression fitting. The following plots show the trend fitting across a span of 0.3.

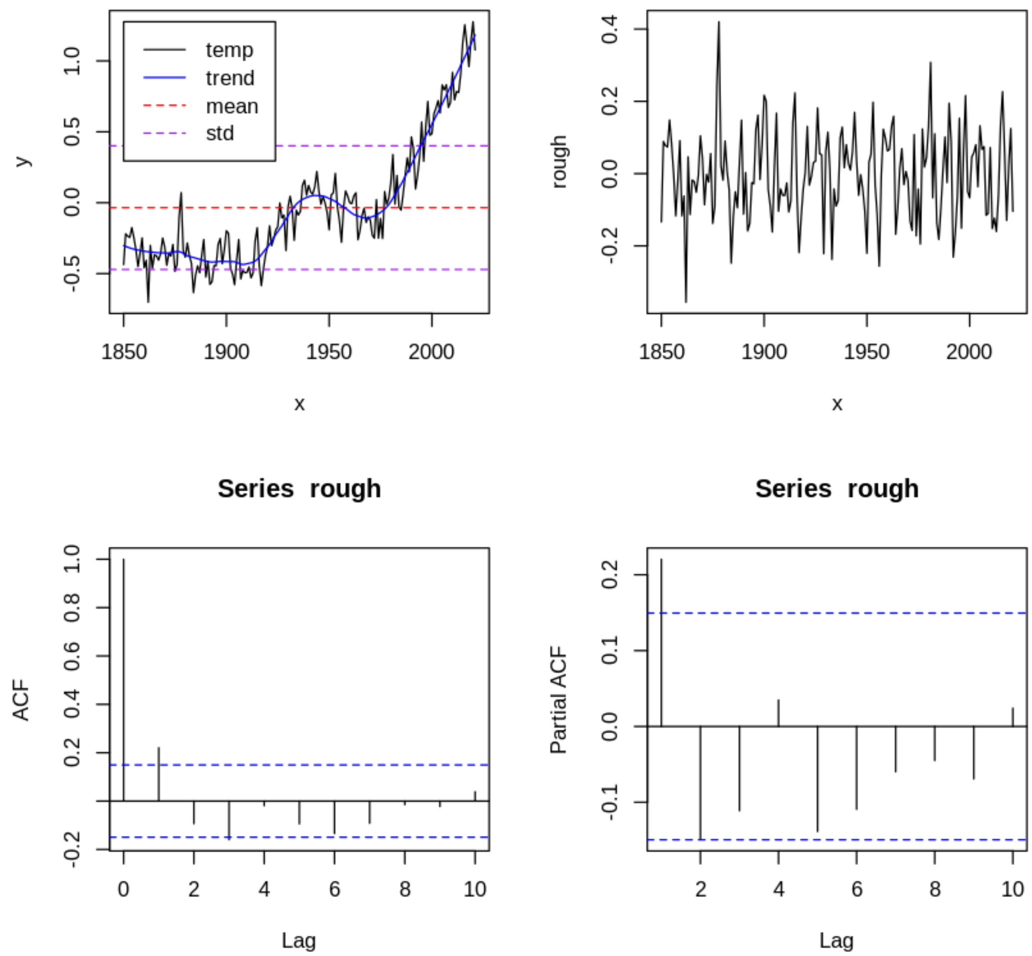

Figure 2: Fitting trend of temperature anomalies data using local polynomial regression fitting.

We find that the trend fit is a relatively good fit. The residuals appear to exhibit stationarity with constant mean and variance. The ACF does show significant auto-correlation of our trend residuals at lag of 1; however, we still conclude that the residuals are approximately stationary and our trend fit is good.

3 Difference of Data

Since the data is not stationarity we decide to take the difference of the series with one lag.

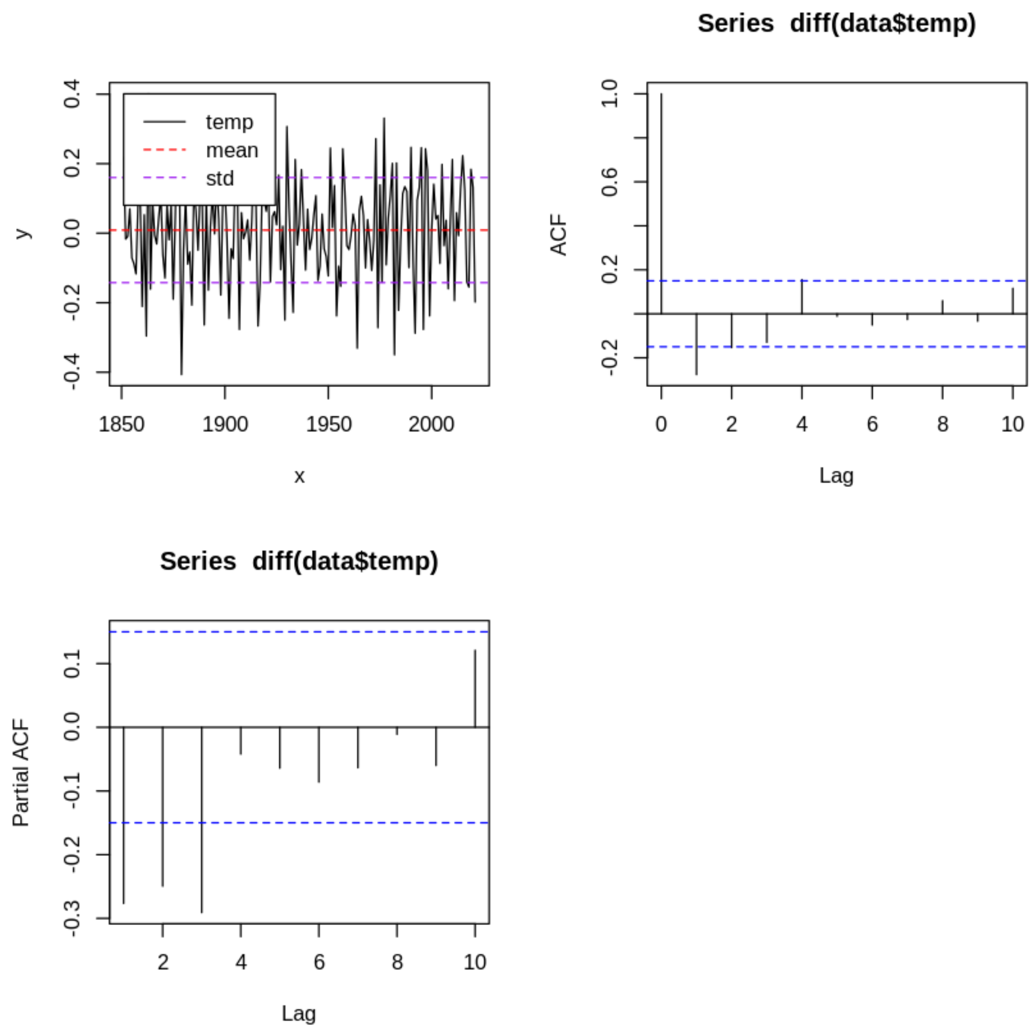

Figure 3: We plot the difference with one lag of the annual temperature anomalies from 1850 to 2019.

Based on the differenced time series plot in the upper left, it suggests that there is a constant mean and variance. Our ACF plot does exhibit significant auto-correlation at lag of one; however, we still believe that the differenced temperature anomalies data somewhat follow weak stationarity.

4 Model Selection

Since the differenced temperature anomaly data is approximately stationarity, we decide to model our time series using an ARIMA model with an difference order of one. Additionally, the ACF plot leads us to believe that an MA of order 1 would be appropriate, and the PACF plot leads us to believe that an AR order of 3 would be appropriate. Therefore, an ARIMA(3, 1, 1) might be appropriate; however, we should be cautious that the ACF and PACF plots alone cannot determine the exact orders of AR and MA. Therefore, we fit all possible models with AR order of 0 to 3 and MA order of 0 to 3, using the one with the lowest AIC.

	MA0	MA1	MA2	MA3
AR0	-0.9229414	-1.094005	-1.119878	-1.108348
AR1	-0.9922384	-1.114102	-1.108236	-1.109690
AR2	-1.0455561	-1.114930	-1.112556	-1.114580
AR3	-1.1233958	-1.116518	-1.110752	-1.103335

Figure 4: Table of AIC values for different orders of MA and AR

Based on the AIC table, it appears that an ARIMA(3, 1, 0) obtains the lowest AIC. Therefore, we proceed with our analysis with an ARIMA(3, 1, 0) as our final model. Fitting an ARIMA(3, 1, 0) we obtain the following estimates for our coefficients.

	AR1	AR2	AR3
Estimates	-0.4113	-0.3430	-0.2837
Std. Error	0.0738	0.0759	0.0739

Table 1: Fitted Estimates for ARIMA(3, 1, 0)

With these estimates our final model follows

$$X_t = Y_t - Y_{t-1} \tag{1}$$

$$\text{where} \quad Y_t = -0.4113Y_{t-1} - 0.3430Y_{t-2} - 0.2837Y_{t-3} \tag{2}$$

5 Model Diagnostics

Given our chosen model, we perform graphical diagnostics to determine if the fit is appropriate. We expect that our residuals from the ARIMA(3, 1, 0) model to be stationarity and the residuals to be normally distributed.

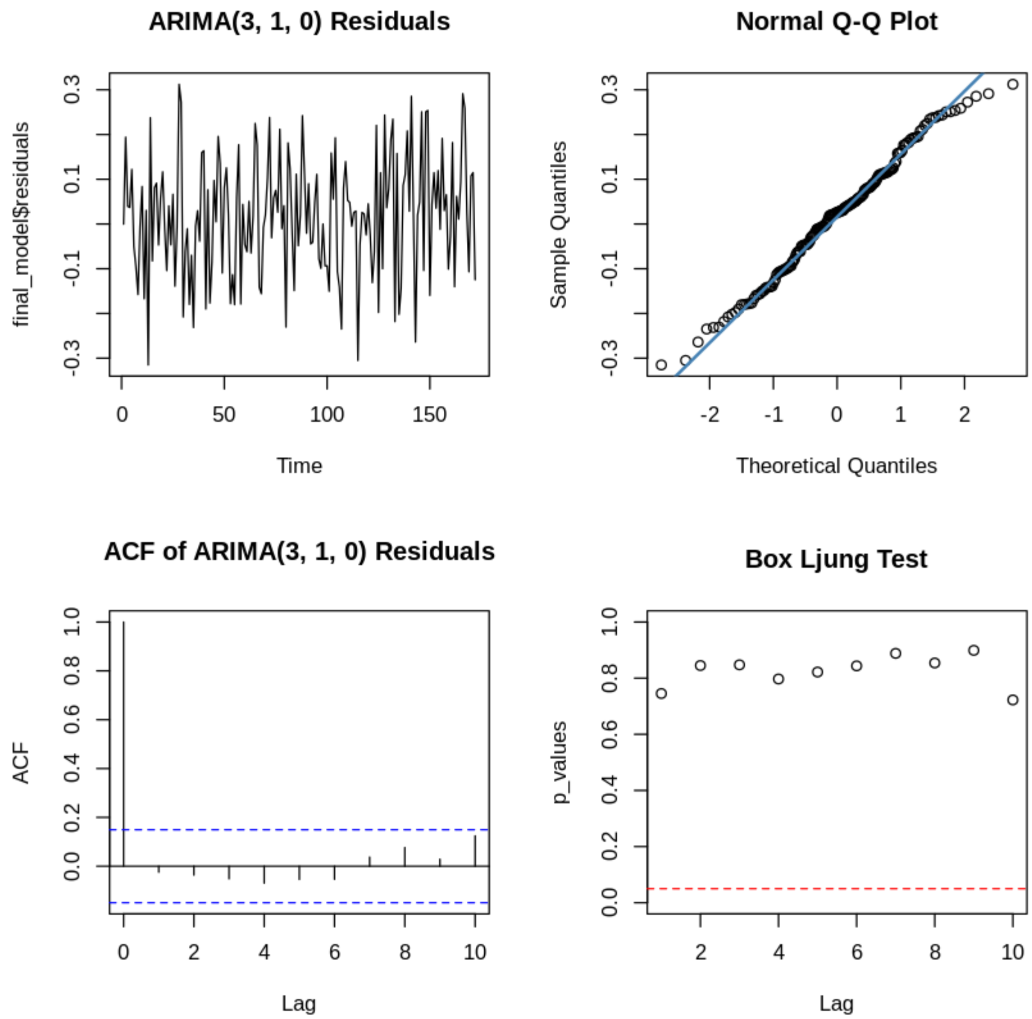

Figure 5: Graphical Diagnostics for ARIMA(3, 1, 0) Model.

We find that the residuals for our ARIMA(3, 1, 0) model follows stationarity. There is constant mean and variance over time. The ACF plot of our residuals does not show any significant auto-correlation for any lags. The Normal Q-Q Plot shows that the residuals are approximately normal. And the p-values for our Box Ljung Test are high, which means we cannot reject that the auto-correlation is zero for any lags. Since the residuals for our model are constant over time, we determine that the ARIMA(3, 1, 0) fits well.

6 Periodogram of Model

We plot the spectral density and smoothed periodogram for our ARIMA(3,1,0) fitted values. The spectral density will show the theoretical frequency decomposition of an ARIMA(3,1,0) model. The periodogram shows the actual frequency decomposition of the fitted values of our ARIMA(3,1,0) model.

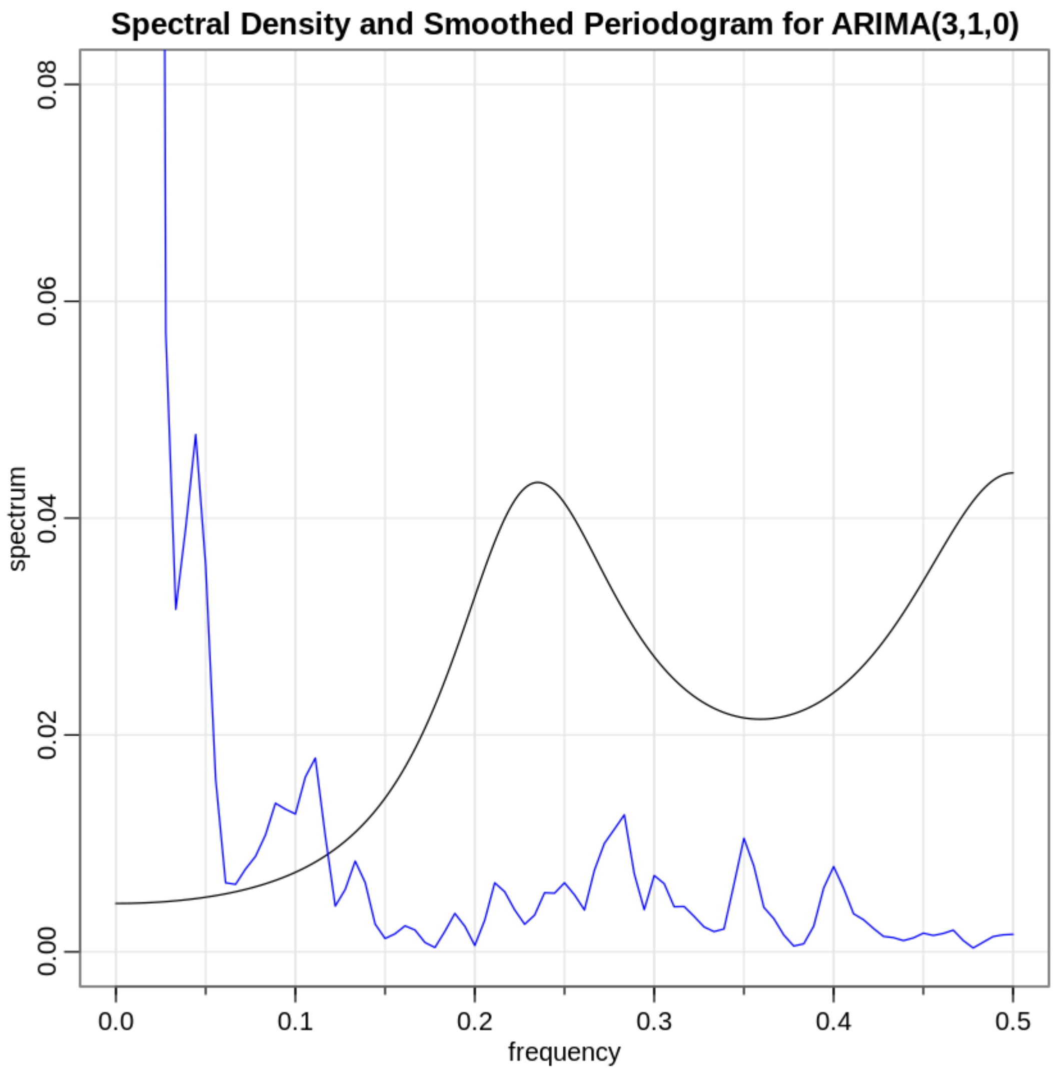

Figure 6: Plot with both Spectral Density and Smoothed Periodogram.

We observe there is a significant difference between the theoretical and actual frequency decomposition of the fitted values of our ARIMA(3,1,0) model. The fitted values of our ARIMA(3,1,0) model do not experience periodicity while the theoretical frequency decomposition shows that we should expect periodicity with frequency ≈ 0.23 and 0.5.

7 Results, Conclusion, & Discussion

Given our final model ARIMA(3, 1, 0), we refit the model using all the data except the last six years. We then use this model to forecast temperature anomalies for the last six years.

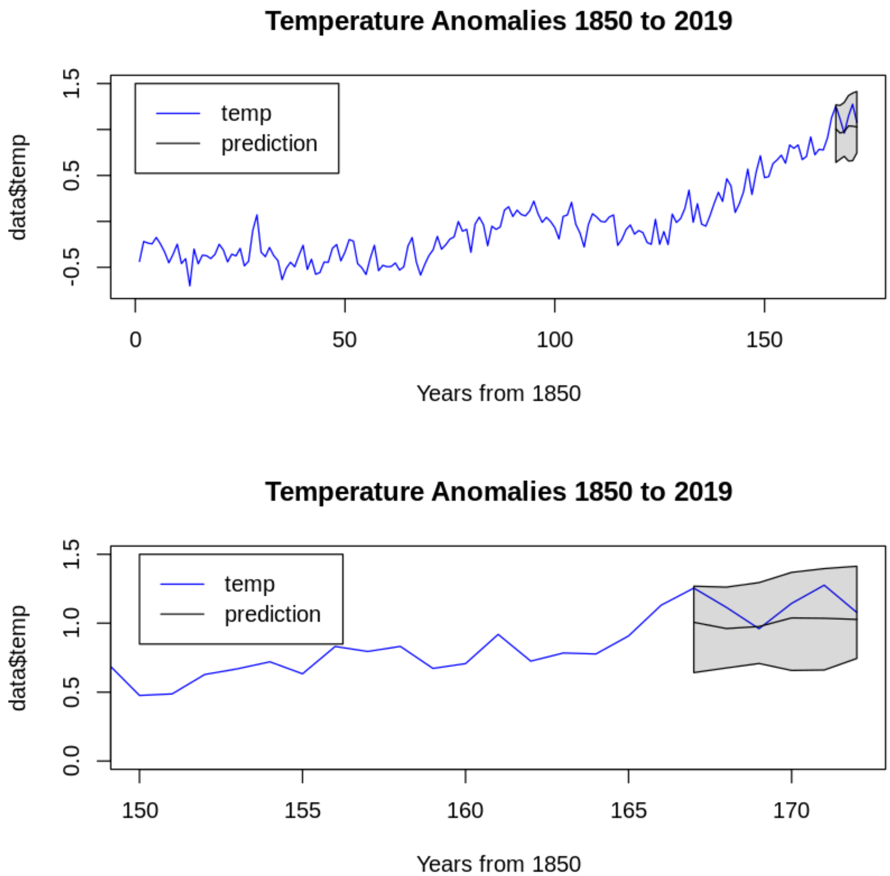

Figure 7: ARIMA(3, 1, 0) predictions of last six years.

We find that the point estimates for our ARIMA(3, 1, 0) model underestimate the true temperature anomaly; however, the confidence interval $\pm 1.96 * SE$ sufficiently captures the true temperature anomaly for the last six years.

We believe that the model underestimates the true temperature anomaly due to the rapid increase in climate temperatures across the Northern Hemisphere. This reaffirms our initial importance of analyzing temperature anomaly data as climate change will have a significant impact on our lives.