
TIME SERIES OF ANALYSIS ANNUAL TEMPERATURE ANOMALIES (1850–2021) FOR THE NORTHERN HEMISPHERE

Hengyuan Liu

Department of Statistics and Data Science
University of California, Los Angeles
Los Angeles, CA 90095
hengyuanliu@g.ucla.edu

Kathy Mo

Department of Statistics and Data Science
University of California, Los Angeles
Los Angeles, CA 90095
kathymo24@g.ucla.edu

Weilin Cheng

Department of Statistics
University of California, Davis
Davis, CA 95616
wncheng@ucdavis.edu

February 19, 2025

ABSTRACT

Today, climate change is one of the most substantial issues in the world. Therefore, annual temperature anomalies are the problem we are concerned about, since anomalies are how we see how the climate changes. Our data set is about the annual temperature anomalies (1850–2021) for the northern hemisphere [1]. A temperature anomaly means a deviation from a reference value or long-term average. The data set contains two columns which are Year and Temperature Anomalies. Based on our knowledge of time series, we want to predict the trend of Temperature Anomalies in the future. As we know, a sequence captured at successive, equally spaced points in time is referred to as a time series. Since Temperature Anomalies are taken over time, which is Years, it is a time series.

Because of prior climatic history and its effects on humans, we believe that it is vital to analyze and forecast future anomalies. To better prepare and safeguard the environment, it is crucial to anticipate what these potential future anomalies might be. People can then prepare to mitigate it based on how severe it is.

1 Introduction

There are two methods we are going to use to forecast future temperature anomalies: 1) the ARIMA model on stationary data, and 2) Estimating the trend and forecasting the rough part. For the second method, we will use the formula $Y_t = m_t(\text{trend}) + X_t(\text{rough})$ to make our predictions.

For the first method, we are going to use the stationary data to get the preliminary analysis of the ARIMA model by ACF plot and PACF plot, then use the AICc criterion to define our final ARIMA model and diagnose our residuals and their properties. Also, we are going to use the Box-Ljung test to test if our residuals are identically and individually distributed. Finally, we will use the data from 1850-2015 to predict 2016-2021 by the final ARIMA model.

For the second method, we are going to find our prediction of the trend and rough part first. To get the trend, we will use the spline trend method from the year 1850-2015 temperature anomalies, then use the `approxExtrap` function on our data from the years 2016-2021. To predict our rough part from years 2016-2021, we will first use our observations to subtract trends to get our rough part from years 1850-2015. Then, we will get the preliminary analysis of the ARIMA model by ACF plot and PACF plot. After, we will use the AICc criteria to define the final ARIMA model of our rough part. Next, we will use the rough part from 1850-2015 to predict the rough part from 2016-2021 by the final ARIMA

model. Lastly, we will add our trend and rough part from 1850-2015 together to get our prediction of observations from the years 2016-2021.

In the end, we will get our conclusion by visualizing these two methods of prediction points and comparing them with real observations to see which method is better, more accurate, or more reasonable.

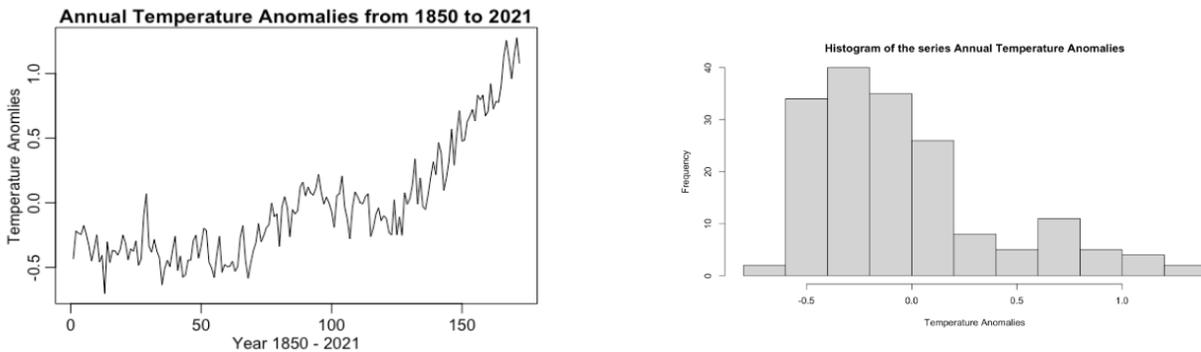
2 Research Statments

To properly analyze our dataset, we decided to base our research project on the following questions:

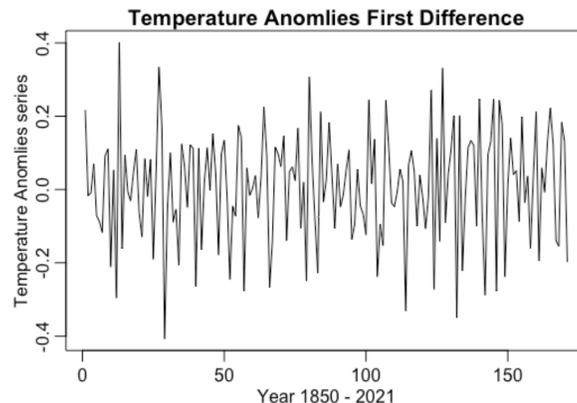
- Which method is best for forecasting future temperature anomalies?
- What is the relationship between Temperature Anomalies and Year?

3 Method I — ARIMA modeling

3.1 Data Visualization



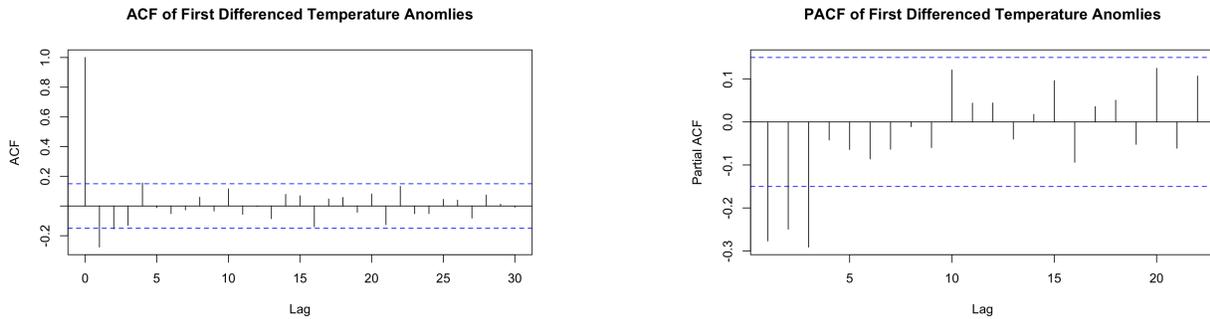
Before applying the methods we are going to use, we first check if our series plot is stationary. A time series model is not stationary if it has a trend or seasonal effects. Therefore, we first check the series of observations to determine if it has a trend. From the plot of annual temperature anomalies above, we find that it is not stationary. To understand the nature of variation in the data, we plot the series of Temperature Anomalies. The plot above shows the annual temperature anomalies data from the year 1850 to 2020. We can observe that there is a clear trend where Temperature Anomalies increased from 1900 to 2020 and Temperature Anomalies became positive, which means it became warmer over time. Also, the temperature anomalies data does not clearly have seasonality since this plot shows there are no regular and predictable changes that recur every year. The average temperature anomaly from 1850 to 2021 is negative which is -0.03455814 and the median is -0.124 . We also get our histogram of the temperature anomaly, the distribution of temperature anomaly is right skewed which tells us the mean of average temperature anomaly is greater than the median of temperature anomaly. The average temperature anomaly is overestimating the most common values in a positively skewed distribution.



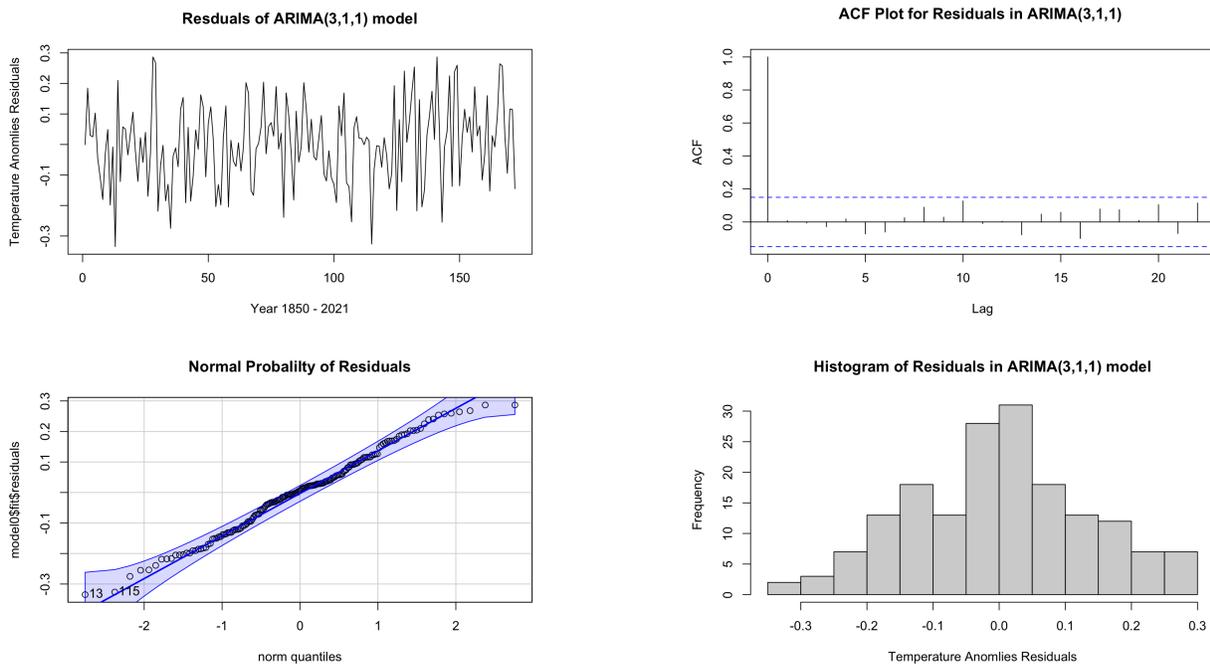
Since subsequent analyses such as modeling are done on stationary series, we need to take the difference of the series to achieve stationarity, which the first differenced series plot shows. After taking the first difference, the trend seems to be gone. As a matter of fact, the series looks stationary as seen above in the plot and the mean of the series looks constant because it fluctuates about zero. Our original series increased temperature as time increased, while the first difference series stayed constant as time increased. Now, we can continue by performing subsequent analyses of the ARIMA model Methodology and Diagnostics.

3.2 ARIMA model Methodology and Diagnostics

3.2.1 Preliminary Identification — AR(3) or MA(1)



Next, we want to obtain the appropriate ACF and PACF plots and use these to make a preliminary identification of a time series model. From the ACF and PACF plots after the first difference in the Temperature Anomalies data(yd1) above, we get our preliminary identification of time series models as AR(3) and MA(1) because the ACF plot cuts off after significant lag at 1 so we choose MA(1), while the PACF plot cuts off after a significant lag of 3 so we choose AR(3). Based on this, it seems like the best model according to the ACF plot is MA(1), while for the PACF plot it is AR(3). Since both of the plots show that the larger orders are insignificant, we can use the AR(3) or MA(1) based on our preliminary results.



By checking the residuals through the top left plot, we can see how good the model fit is. This plot shows constant variance and constant mean because it fluctuates around zero.

We examine the residuals by the ACF plot on the top right, and we find that the residuals from the ARIMA model are identically and individually distributed because the ACF plot shows that none of the lags are significant.

The last property we need to check is the normality assumption. The QQ plot of the residuals on the bottom left shows that they are normally distributed because it follows the fitted line.

Finally, the histogram of residuals on the bottom right is bell-shaped, which further confirms and gives strong evidence that our residuals follow a normal distribution.

χ^2	Degree of Freedom	p-value
6.591	10	0.7634

Table 1: Box-Ljung Test Results

To determine if our residuals are independent and identically distributed, we use the Box-Ljung test. From the table above, our P-value is 0.7634, which is very large. This means we cannot reject the assumption of independent and identically distributed residuals. Therefore, we conclude our model residuals are identical and individually distributed, and have strong evidence that our model fits well.

3.2.2 Model Selection — AICc

Next, we want to select the final model using the model selection criterion AICc. We want to use AICc instead of AIC because using AICc gives a more accurate answer for smaller samples while giving the same answer as AIC for larger samples. It may be enough to consider the following 16 models. Here is a table with values of the criterion function below.

	MA(0)	MA(1)	MA(2)	MA(3)
AR(0)	-0.9228030	-1.093587	-1.119037	-1.106939
AR(1)	-0.9918207	-1.113262	-1.106827	-1.107563
AR(2)	-1.0447157	-1.113521	-1.110429	-1.111585
AR(3)	-1.1219870	-1.114391	-1.107756	-1.099317

Table 2: AICc values for AR(0–3) and MA(0–3).

From the AICc criterion table, MA(0) and AR(3) are the best models with the first difference because through the AICc we have ARIMA(3,1,0). By looking at our AICc criterion values, we find that the smallest value is -1.1219867, which means that our final and best model is ARIMA(3,1,0).

For coding, we use sarima instead of arima because sarima does not assume that the mean is zero, unlike arima. This means it is better to use since even if the mean is close to zero, it should not be taken as zero. Furthermore, since it takes seasonality into account, it deals with more complex data better. So although our dataset does not have seasonality, we chose to use the sarima function after taking the above reasons into consideration. Thus, we use the sarima function to fit our model.

The estimated parameters and the standard errors are given in the following table. These estimated parameters are important because they are our estimated Φ_i , Φ_1 , Φ_2 , Φ_3 , which we plug into the above equation. It gives us an estimated model of our chosen model. The standard errors are important since we can determine how accurate the coefficients are.

	Φ_1	Φ_2	Φ_3	mean()
Coefficients	-0.4234	-0.3557	-0.2947	0.0085
Standard Error	0.0735	0.0756	0.0735	0.0050

Table 3: Estimated parameters and standard errors for ARIMA(3,1,0) model.

ARIMA(3,1,0) Model Expression:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \epsilon_t,$$

where $\phi_0 = \mu(1 - \phi_1 - \phi_2 - \phi_3)$.

With a first difference:

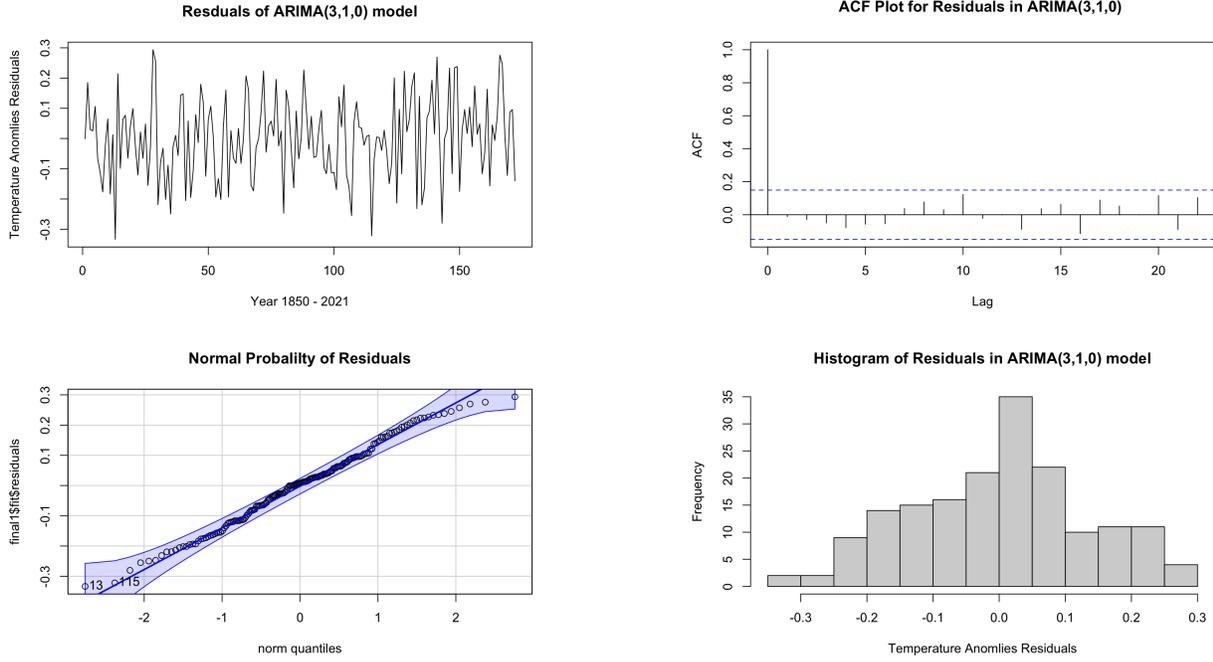
$$X_t = Y_t - Y_{t-1}, X_{t-1} = Y_{t-1} - Y_{t-2}, X_{t-2} = Y_{t-2} - Y_{t-3}, X_{t-3} = Y_{t-3} - Y_{t-4}$$

Final model expression with estimated parameters:

$$\hat{X}_t = 0.0176256 - 0.4234 \hat{X}_{t-1} - 0.3557 \hat{X}_{t-2} - 0.2947 \hat{X}_{t-3}.$$

3.2.3 Diagnostics of Residuals

Next, we perform diagnostics on the residuals of the final model ARIMA(3, 1, 0).



By checking the residuals through the top left plot, we can see how good the model fit is. This plot shows constant variance and constant mean because it fluctuates around zero.

Next, we examine the residuals by the ACF plot on the top right, and we find that the residuals from the ARIMA model are independent and identically distributed because the ACF plot shows that none of the lags are significant.

The last property we need to check is the normality assumption. The QQ plot of the residuals on the bottom left shows that they are normally distributed because it follows the fitted line.

Finally, the histogram of residuals on the bottom right is bell-shaped, which further confirms and gives strong evidence that our residuals follow a normal distribution.

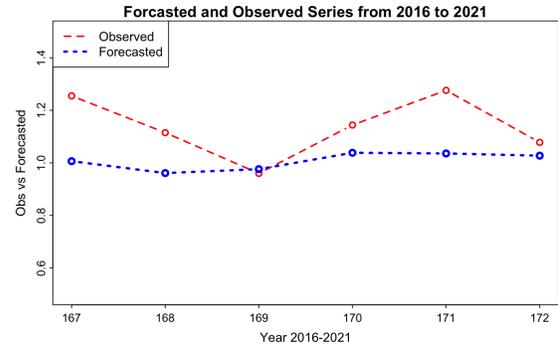
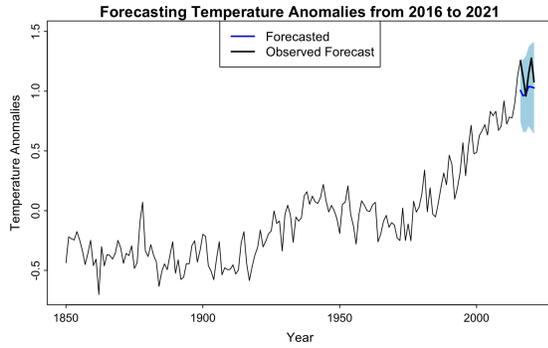
χ^2	Degree of Freedom	p-value
7.1741	10	0.7089

Table 4: Box-Ljung Test Results

To diagnose if the residuals of our model are independent and identically distributed, we use the Box-Ljung test. From the table above, our P-value is 0.7089, which is very large. This means we cannot reject the assumption of independent and identical residuals. Therefore, we conclude our model residuals are identical and individually distributed, and we have strong evidence that our model fits well.

3.3 Method I Results

Finally, we refit the final model by sarima.for function (we use AR and MA orders of the final model which is ARIMA(3,1,0), but not the parameter estimates) using all the data except for the last 6 years and use this model to



Year	Forecasted (ARIMA)	Forecasted (SARIMA)	Observed
2016	0.9931122	1.0064209	1.255
2017	0.9391939	0.9612019	1.115
2018	0.9459326	0.9760345	0.960
2019	1.0004318	1.0384229	1.144
2020	0.9901110	1.0358330	1.276
2021	0.9740973	1.0274494	1.078

Table 5: Forecasted and Observed Temperature Anomalies (2016-2021)

forecast temperature anomalies for the last 6 years. We plot the available observed and the forecasted values against time (6 years) on the same graph.

The right plot is plotting observed and forecasted in the same graph from the year 2016 to 2021. As you can see, the prediction points and observations are pretty close to each other which reflects our ARIMA(3,1,0) model is a good fit. The forecasted is around the observed, but lower and more constant.

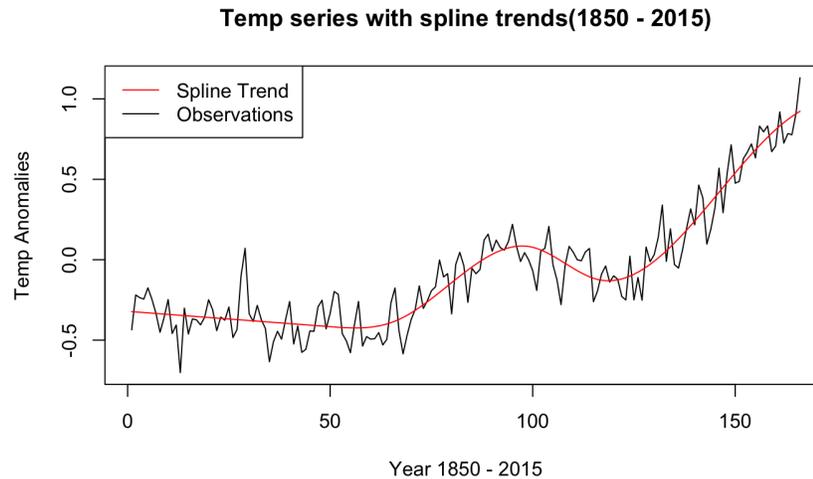
The left plot shows true vs fitted observations. It predicts what the true observations will be from 2016 to 2021. These predictions are given in the blue confidence interval which means an approximate 95% prediction interval. The way to get the prediction interval is to use the predicted values plus and minus the value of 1.95 multiplied by the standard errors for each predictor. As we can see, the blue confidence band area fully encompasses the true observed values, further confirming that our model is a good fit which means our prediction ARIMA model is good at predicting.

Also, from the table shown above, we can see the observed and the predicted values for the year 2016-2021 are quite similar which also implies that our ARIMA model predicts well. The table also shows the difference in predicted values from the sarima and arima functions. The sarima function is more accurate than the arima, which confirms why we used the sarima function in our code to make our ARIMA model.

4 Method II — Trend and Rough Forecasting

We are going to apply the formula of (Obs = trend + rough). If we want to forecast observations from Y_{2016} to Y_{2021} , we first need to predict the trend part which is m_{2016} to m_{2021} and then we predict the rough part which is X_{2016} to X_{2021} . After getting both the prediction of the trend and rough parts from 2016 to 2021, we can add them together to get our prediction of observations from Y_{2016} to Y_{2021} .

4.1 Estimation and Forecasting Trend



First, we use the spline method to estimate the trend from our Temperature Anomalies from the years 1850 - 2015. The graph above, Temp series with spline trends(1850-2015), goes from 1850 to 2015. It does not include the last 6 years of our dataset. We start from time zero, which represents our first year, 1850. The red line is an estimate of the trend using the spline method.

The R^2 of the spline trend we find is .89997913 which means that our model explains 89.997913% of the variance in our observations can be explained by the trend. In other words, the trend fits our observations very well.

4.1.1 Diagnostics of Residuals

chi^2	Degree of Freedom	p-value
17.706	10	0.06013

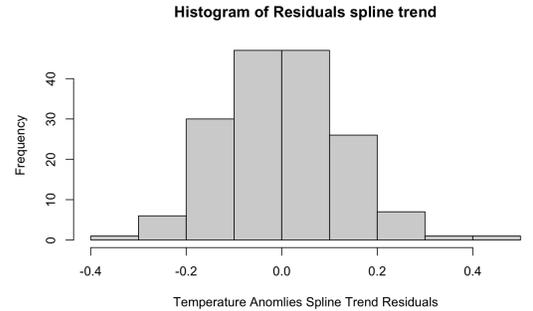
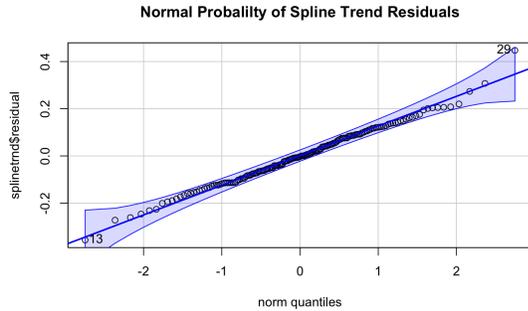
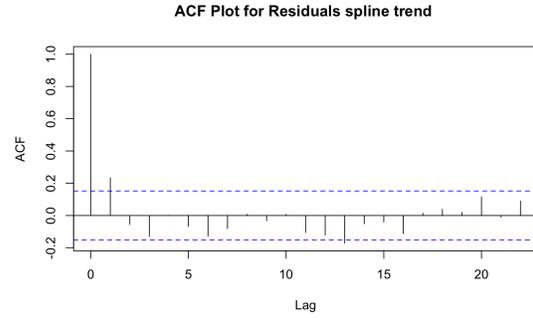
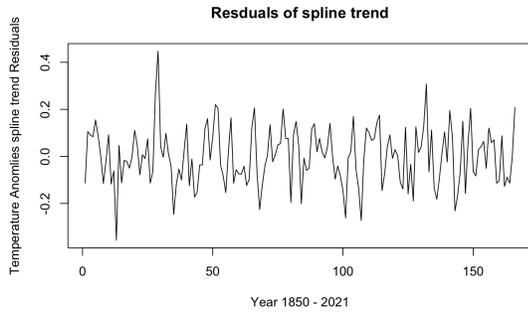
Table 6: Box-Ljung Test Results

By checking the residuals through the top left plot, we can see how good the model fit is. This plot shows constant variance and constant mean because it fluctuates around zero.

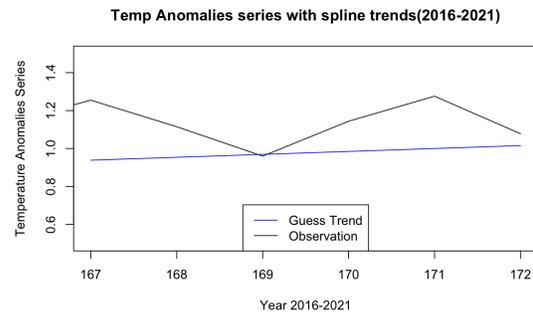
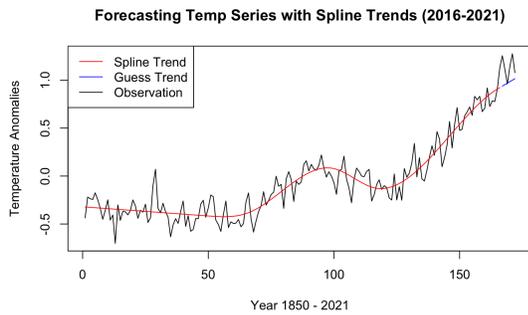
Next, we examine the residuals by the ACF plot on the top right, and we find that it is hard to determine if the residuals are independent and identically distributed since lag 1 and lag 13 are significant. Therefore, we decided to use the box-ljung test, which is shown in the table above. The p-value is larger than the usual alpha of 0.05, so therefore we conclude our model residuals are identical and individually distributed, and we have strong evidence that our model fits well.

The last property we need to check is the normality assumption. The QQ plot of the residuals on the bottom left shows that they are normally distributed because it follows the fitted line except there are two outliers which are located at 13 and 29.

Finally, the histogram of residuals on the bottom right is bell-shaped, which further confirms and gives strong evidence that our residuals follow a normal distribution.



4.1.2 Trend Forecasting

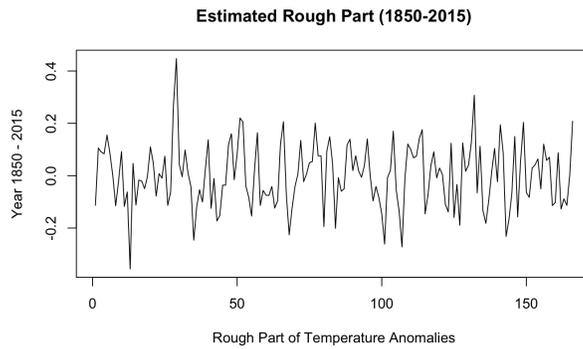


The graph on the left, Temp series with spline trends, includes the years 1850 to 2015 which is the red line, and also includes the last 6 years(2016 -2021) that we guess by using the approxExtrap function, which is our blue line.

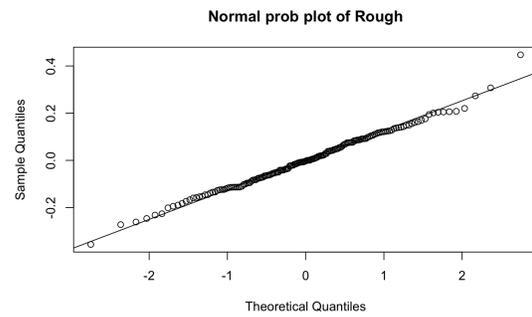
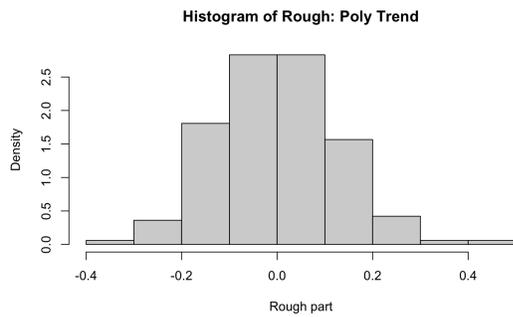
The graph on the right just looks at the last six years, 2016 to 2021, from the previous graph on the top right. We take a closer look at these last six points to better observe the fitted versus observed line and how well the spline method estimates the actual observations. The spline trend values from 2016 to 2021 are 0.9383427, 0.9537709, 0.9691991, 0.9846274, 1.0000556, and 1.0154839.

Both of the graphs include the guess and actual observed values of the Temperature Anomalies Series.

4.2 Estimating and Forecasting of Rough Part

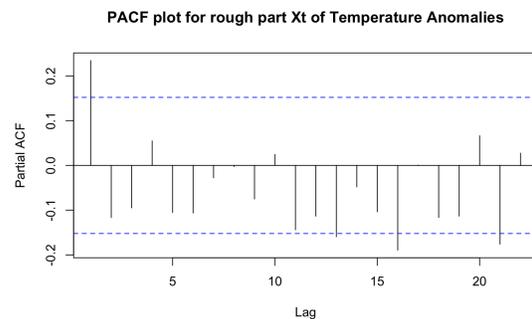
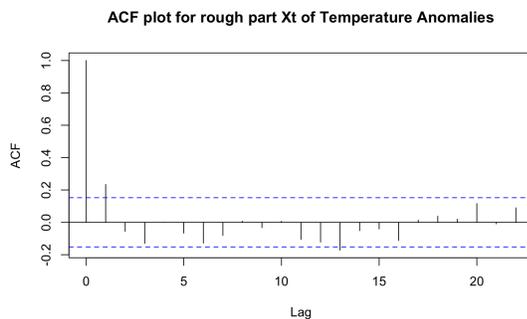


First, we get our rough estimate by using the Temperature Anomalies observations to subtract the trend of Temperature Anomalies from the years 1850 to 2015. As you can see, the rough part of the Temperature Anomalies is already stationary.



Both the histogram and the normal probability plots are given above, we can see that our histogram of the rough part is normally distributed and the QQ plot of the residuals on the bottom left shows that they are normally distributed because it follows the fitted line.

4.2.1 Preliminary Identification of Rough Part



Next, we want to obtain the appropriate ACF and PACF plots, and use these to make a preliminary identification of the rough. Since our rough part is always stationary, we don't need to take the first difference. From the ACF and PACF plots, we can get our preliminary identification of the rough part as MA(1) and AR(1) because the ACF plot and PACF plot show that all lag orders above one are insignificant. Therefore, we can use MA(1) and AR(1) as our preliminary identification.

4.2.2 Model Selection Rough Part — AICc

Next, we want to select the final model using the model selection criterion AICc. It may be enough to consider the following 16 models. Here is a table with values of the criterion function below.

	MA(0)	MA(1)	MA(2)	MA(3)
AR(0)	-1.350098	-1.402409	-1.390188	-1.405467
AR(1)	-1.395323	-1.390030	-1.390864	-1.476548
AR(2)	-1.397435	-1.485000	-1.393911	-1.468015
AR(3)	-1.394342	-1.389045	-1.468693	-1.455545

Table 7: AICc values for different ARMA model configurations

From the table, MA(1) and AR(2) are the best models. By looking at our AICc criterion values, we find that the smallest value is -1.485, which means that our final and best model is ARIMA(2,0,1).

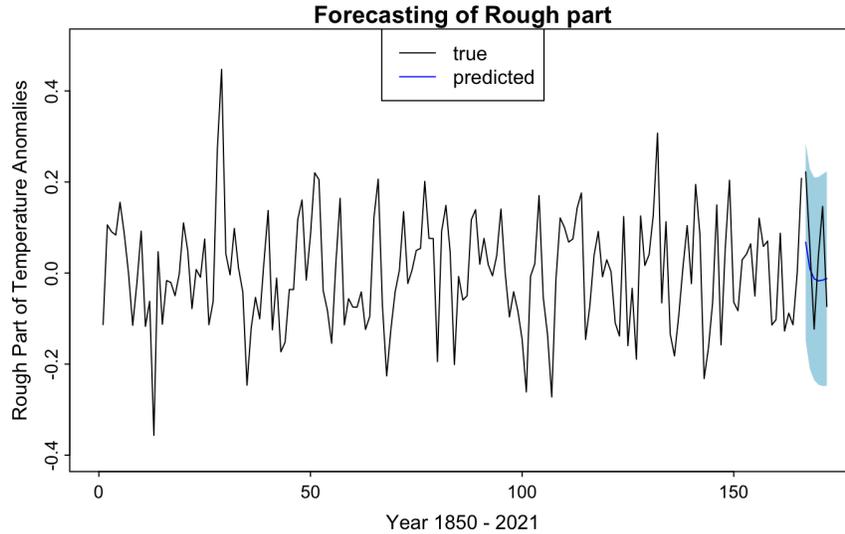
	Φ_1	Φ_2	θ_1	mean()
Coefficients	1.1433	-0.3303	-1.000	-0.0001
Standard Error	0.0735	0.0737	0.0157	0.0009

Table 8: Estimated parameters and standard errors for ARIMA(2,0,1) model.

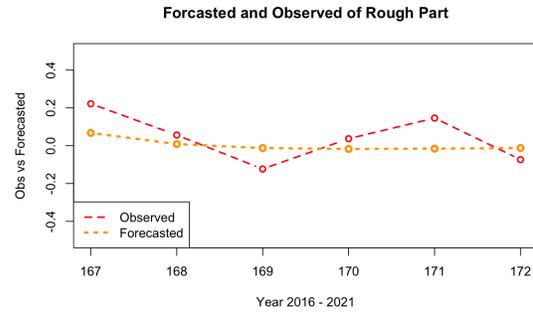
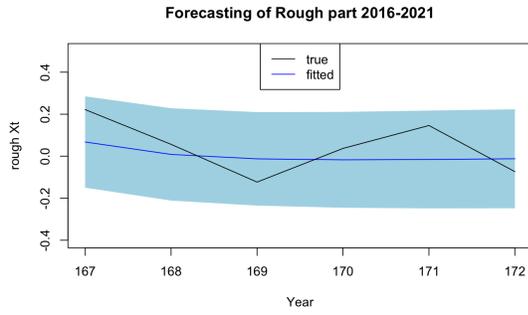
The expression of ARMA(2,1) for rough part:

$$\begin{aligned}
 X_t - \mu &= \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + \varepsilon_t + \theta_1\varepsilon_{t-1} \\
 X_t &= \phi_0 + \phi_1 \cdot X_{t-1} + \phi_2 \cdot X_{t-2} + \varepsilon_t + \theta_1\varepsilon_{t-1} \quad , \text{where } \phi_0 = \mu(1 - \phi_1 - \phi_2) \\
 X_t &= -0.0000187 + 1.1433 \cdot X_{t-1} - 0.3303 \cdot X_{t-2} - \varepsilon_{t-1}
 \end{aligned}$$

4.2.3 Rough Part Forecasting



We use the year 1850 - 2015 Temperature Anomalies rough part to forecast our last 6 years' Temperature Anomalies rough part. First, we use the sarima.for function with AR(2) and MA(1) to forecast last year's Temperature Anomalies. Then, we plot the six forecasting values and the real values in one graph. By looking at the graph on the top, we can see that the blue confidence band fully encompasses the fitted values, which means it is a good prediction. The blue area means an approximate 95% prediction interval. The way to get the prediction interval is to use the predicted values plus and minus the value of 1.95 multiplied by the standard errors for each predictor. The bottom left graph is

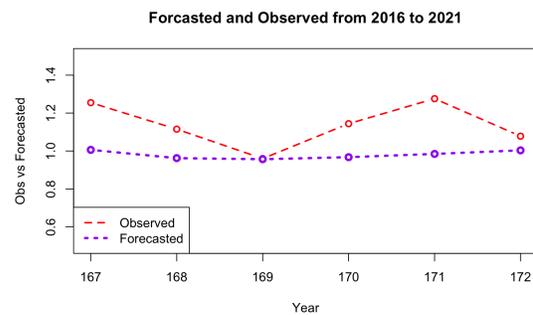
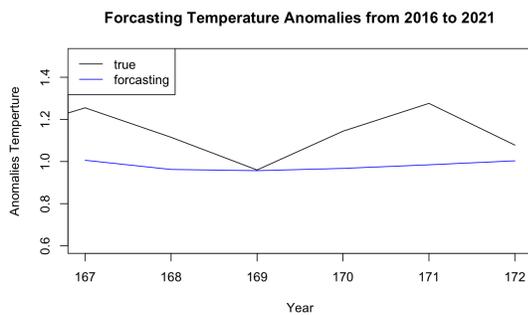
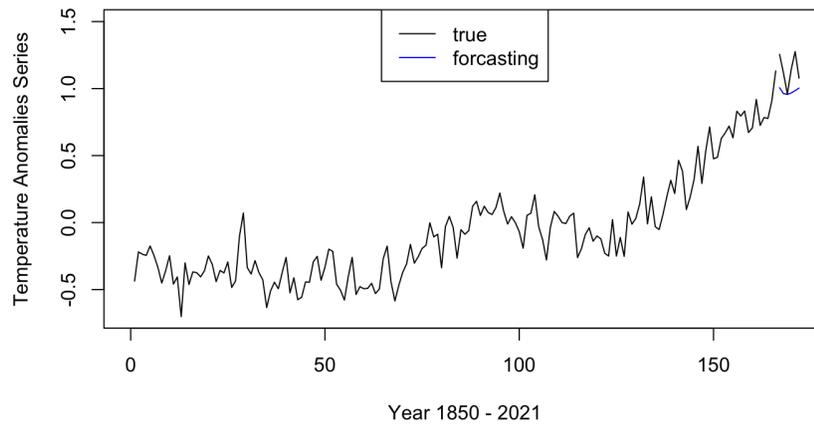


just a zoomed-in graph of our top graph for the last six years. The last six years' temperature anomalies values of the rough part we predict by ARMA(2,1) are 0.067679785, 0.008626265, -0.012512317, -0.017174109, -0.015521631, and -0.012092486.

The bottom right graph is the forecasted and observed values of the rough from the last 6 years from 2016 to 2021. We can see that the forecasted values go through the middle of the observed values. The three above graphs show that our forecast of the rough part is a good fit.

4.3 Method II Results

Forecasting Temperature Anomalies from 2016 to 2021



The above two plots show our results from forecasting. The top center plot shows our forecasting for the rough and trend. For the very last six years, there is a small horizontal blue line that predicts what the values from 2016 to 2021 are. Meanwhile, the actual observed values are not constant as the fitted line suggests.

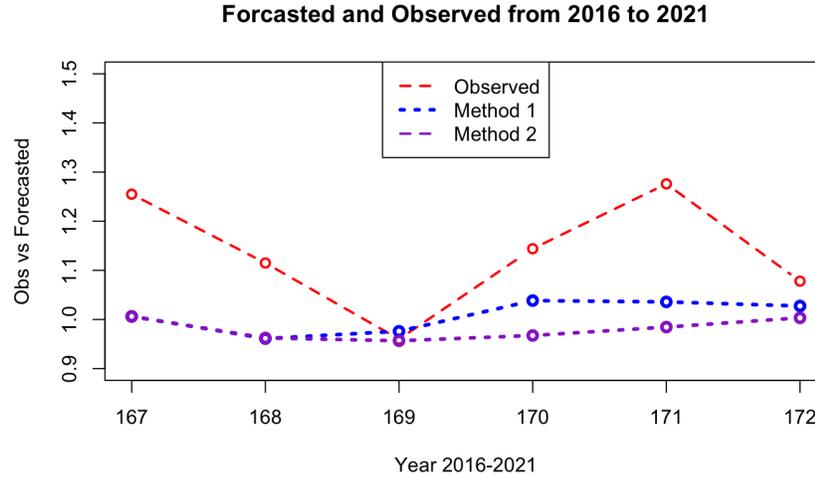
The bottom left plot zooms in on the last six years of the top center plot and looks at the fitted versus observed values. We can see that the forecasted values are approximately horizontal and constant throughout time, which matches the forecasted line in the bottom right graph.

Year	m_t	X_t	$Y_t = m_t + X_t$	Observed
2016	0.9383427	0.067679785	1.0060224	1.255
2017	0.9537709	0.008626265	0.9623972	1.115
2018	0.9691991	-0.012512317	0.9566868	0.960
2019	0.9846274	-0.017174109	0.9674533	1.144
2020	1.0000556	-0.015521631	0.9845340	1.276
2021	1.0154839	-0.012092486	1.0033914	1.078

Table 9: Forecasted and Observed Temperature Anomalies (2016-2021)

The table above shows the values we got from the spline method. The m_t row is our trend, the X_t row is our rough, the Y_t column is our predicted observations, and the Observed column is the true observed values.

5 Final Results



Year	Method 1	Method 2	Observed
2016	1.0064209	1.0060224	1.255
2017	0.9612019	0.9623972	1.115
2018	0.9760345	0.9566868	0.960
2019	1.0384229	0.9674533	1.144
2020	1.0358330	0.9845340	1.276
2021	1.0274494	1.0033914	1.078

Table 10: Comparison of Forecasted Temperature Anomalies by Method 1 and Method 2 with Observed Values (2016-2021)

These two methods look to have a similar result in the prediction, however, we prefer method 1 which is the blue trend. We prefer this method because it is closer to the actual observed values overall. Looking at the blue trend, which is method I, and the purple trend, which is method II, we see that the first two predicted years are the same, and after that, the two method predictions diverge. Meanwhile, the table shows the values of each of the observed and fitted values for the last six years. Therefore, the model we chose from method one is our final overall model, which is ARIMA(3,1,0).

6 Conclusion and Discussion

Based on our findings, we found these both two methods did good predictions for the last 6 years, the two methods predict very similar values for the years 2016 and 2017. However, overall we think that the ARIMA model gives better predictions because it provides a closer prediction value to the real value of temperature anomalies. We applied two methods, the first of which is ARIMA to forecast the observations of the last 6 years. The second is to forecast the last 6 years to add the Rough Part and Trend Part together. Through our graph to compare these two methods, we find the two methods have similar results in the prediction. We think these two methods are good at predicting, but we prefer method one which is the ARIMA(3,1,0) model to forecasting the Temperature Anomalies because it is closer to our real data of the last 6 years from 2016 to 2021. Based on the visualization of Temperature Anomalies, we find there is clearly an increasing trend which means that as the years go by, temperature becomes warmer and warmer. Thus, it is important to forecast based on the potential consequences of a warmer climate. With this, we successfully answered both of our research questions.

Another thing we want to discuss is that our method I and method II predictions are all underestimates compared with the observed. Although the values are relatively close to the observed, both are still underestimating the last six values. Perhaps it is because the previous values that the forecasting was based on did not indicate such a large increase in temperature. We can see the steep increase from our observed value graphs, so even if the forecasted values predicted an increase, it was just not high enough.

References

- [1] National Oceanic and Atmospheric Administration (NOAA). Climate Data and Information. Available at: <https://www.climate.gov/>. Accessed: February 18, 2025.