STA457 Final Report: Analysis of Global Mean Ocean Temperature Deviations

Peiyu Li 1004975627

April 16, 2021

Abstract

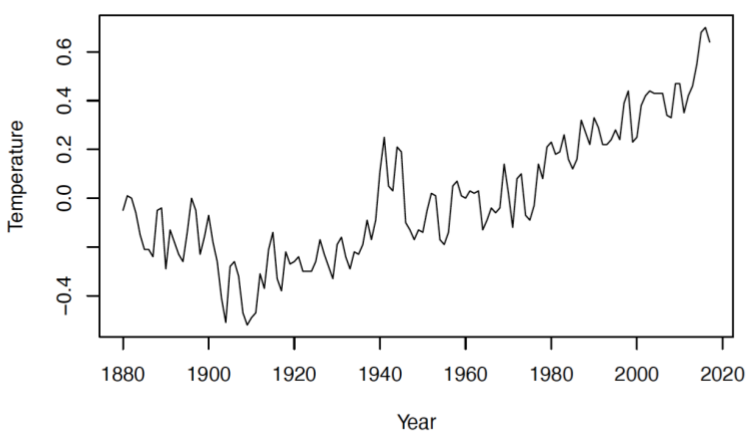
The analysis is focus on the global ocean temperature data, and it is motivated by the ocean warming issue. The data contains 138 observations, measured from 1880 to 2017. The first three dominant frequencies in this series occur at 0.0069, 0.0139, and 0.028. Differencing is applied to the data to eliminate the upward trend. Two models, ARIMA(2,1,0) and ARIMA(2,1,2) are proposed based on their autocorrelation function plots. After diagnostics, ARIMA(2,1,0) is used to make predictions for the next ten years ocean temperature. The prediction shows that there is an increasing trend for the global ocean temperature from 2018 to 2027.

Introduction

Global warming has always been a concern for human beings. It is a trend that the world is becoming warmer everywhere. Much of the extra heat from greenhouse gas pollution is absorbed by the ocean, resulting in increasing ocean temperatures (IUCN, 2018). Over the last 100 years, the average global sea surface temperature has risen by 0.13°C per decade, according to data from the US National Oceanic and Atmospheric Administration (NOAA, 2017). Increasing ocean temperatures would negatively affect marine species and the ecosystem. Thus, it is important for statisticians to get some helpful information from the global mean temperature data to make a prediction and implement appropriate mitigation strategies.

In this report, the analysis of the ocean surface temperature will be performed. The data was provided in a R package called “astsa” and the data source was from NASA. It contains global average ocean temperature deviations from 1880 to 2017 and the temperature deviations were measured in Celsius. Figure 1 shows how the global mean ocean temperature changes over time. It is not surprised that there is an increasing trend, especially after 1900.

Figure 1: Global Mean Ocean Temperature



Statistical Methods

Since there is a trend for the data, the differences between consecutive observations should be performed. This is because we want to stabilize the mean of the time series so that the properties do not depend on time. Because the goal is to predict future ocean surface temperature, an ARIMA model will be fitted. ARIMA model can include autoregressive terms, differencing operations, and moving average terms. For the differenced data, the order of the differencing will be zero. To decide the orders of AR and MA terms, we should consider both the auto-correlation function (ACF) and partial auto-correlation function (PACF) of the differenced data. The rule is that if the ACF plot is cut-off at lag h, the orders of MA terms will be h. If the PACF plot is cut-off at lag h, then the order of AR terms will be h.

I propose two ARIMA models, ARIMA(2,1,0) and ARIMA(2,1,2) for the original data, after choosing the orders of AR and MA terms. There is one of the parameter estimates in ARIMA(2,1,0) is statistically significant, and none of the estimates in the other model is significant, so ARIMA(2,1,0) is preferred. The two models are examined based on the assumptions of the residuals. The variance of the residuals is checked using a standardized residuals plot and ACF of residuals. The normal Q-Q plot shows the normality of residuals. Then, p-values for Ljung-Box statistics can help us to check the residuals are independent. All the assumptions are satisfied by these two models. The AIC and BIC values are also calculated for both models. AIC and BIC are the criteria that help us to find a better model for prediction, and lower values for both AIC and BIC are preferred.

After considering models' significance, diagnostics, AIC and BIC values of these two models, the best model ARIMA(2,1,0) will be used to make predictions over the next ten years of global mean ocean temperature. This model can be expressed as: $y_t = \mu + \phi_1\, y_{t-1} + \phi_2\, y_{t-2} + W_t,\ y_t = Y_t - Y_{t-1}$, where Y_t and Y_{t-1} represents the mean global ocean temperature at time t and t-1, ϕ_1 and ϕ_2 are the parameter of the AR(2) model, μ is the mean average global ocean temperature and W_t is the random error. Since y_t depends on y_{t-1} and y_{t-2}, we know that the last two differences of the ocean temperature can affect the ocean temperature on the current year. The prediction will base on the estimates of the parameters in the model, and the significance of the estimates is also important for the predictions to be meaningful. A 95% prediction interval can tell people that the expected true value at a certain time will be within the interval. In addition, a periodogram is a useful tool to identify any dominant periods in the time series data. I perform a periodogram analysis including finding the first three dominant frequencies and their confidence intervals. The 95% confidence interval for the dominant frequency can help to get significance of the frequency. However, if the confidence interval is too wide, then it is not helpful with getting significance.

Results

In Figure 2, the time series is stationary after differencing, which means the temperature does not depend on time anymore. The differencing works very well from the plot, since all the observations variates around zero.

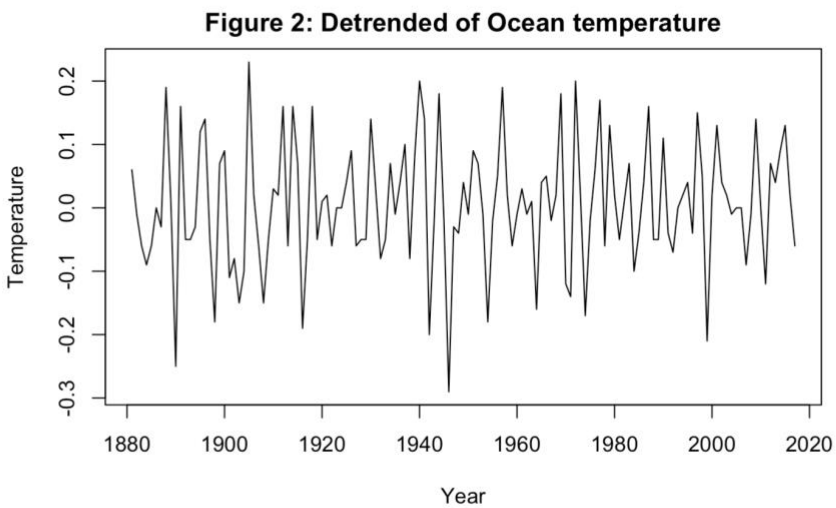
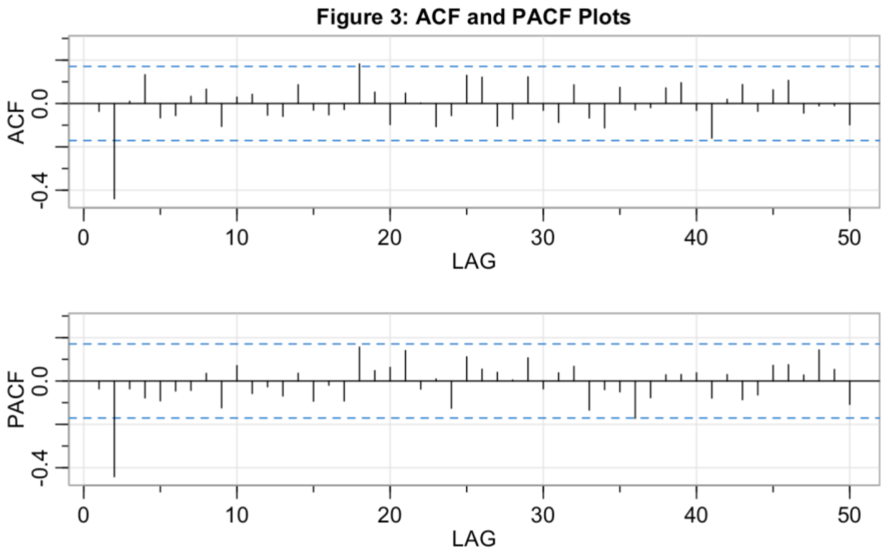


Figure 3 is used to decide the orders of AR and MA terms. According to the plot, PACF and ACF are both cutting off at lag 2, so it seems appropriate to propose ARIMA(2,1,0) and ARIMA(2,1,2) for the global ocean temperature data.



The estimates and p-values are shown in the following tables. From Table 1, the p-value for the second estimate (ar2) of ARIMA(2,1,0) model is less than 0.05, showing that this parameter is statistically significant. For ARIMA(2,1,2), none of the estimates are significant, which means this model is not very helpful.

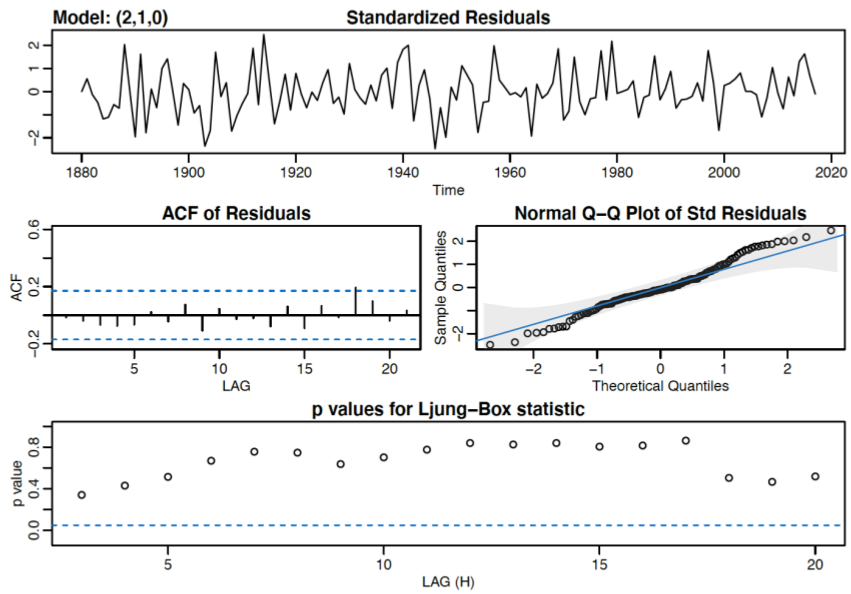
Table 1: t-table for ARIMA(2,1,0)

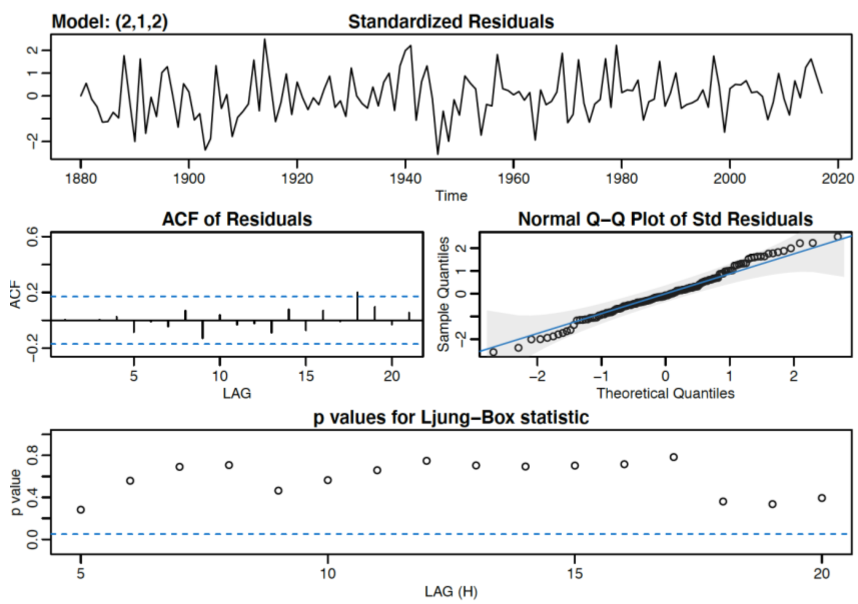
	Estimate	SE	t.value	p.value
ar1	-0.0519	0.0767	-0.6762	0.5001
ar2	-0.4375	0.0761	-5.7522	0.0000
constant	0.0051	0.0051	0.9905	0.3237

Table 2: t-table for ARIMA(2,1,2)

	Estimate	SE	t.value	p.value
ar1	0.0845	0.1855	0.4556	0.6494
ar2	-0.2280	0.1919	-1.1882	0.2369
ma1	-0.1751	0.1864	-0.9396	0.3491
ma2	-0.2489	0.1916	-1.2988	0.1963
constant	0.0050	0.0038	1.3048	0.1942

Then, we need to inspect the diagnostic plots below for ARIMA(2,1,0) and ARIMA(2,1,2). The standardized residuals show both models have no obvious patterns. The ACF residuals plot show a significant spike at lag 18 in both models, but it is not significant and all the other ACF are within the limits. This shows that the randomness assumption for the models is satisfied. The normal Q-Q plot of residuals indicates the normality assumption are met by both models, except for the possible outliers. There are some outliers detected at the tails which show a deviation from normality. The p-values for Ljung-Box statistics are all above the significance level for all lags for both models. It means we do not reject the null hypothesis that the residuals are independent. Therefore, both models meet all the assumptions for residuals.





In addition, the AIC and BIC values for ARIMA(2,1,0) are -1.95 and -1.86, while the AIC and BIC for ARIMA(2,1,2) are -1.93 and -1.80. In this case, AIRMA(2,1,0) seems to be the better one. Also, all the estimates in ARIMA(2,1,2) are not significant, while one of the estimates in ARIMA(2,1,0) is significant. Based on this evidence, ARIMA(2,1,0) should be better for prediction. This model can be expressed as: $y_t = 0.0051 - 0.0519\, y_{t-1} - 0.4375\, y_{t-2}, y_t = Y_t - Y_{t-1}$, where Y_t and Y_{t-1} represents the mean global ocean temperature at time t and t-1, -0.0519 and -0.4375 are the parameter estimates of the AR(2) model, 0.0051 is the mean average global ocean temperature. This model shows that the last two differences in ocean temperature have an effect on the current ocean temperature. According to this model, the prediction of the average global ocean surface temperature from 2018 to 2027 is shown in the plot below. Table 3 presents the 10 prediction values and their prediction intervals. For example, the prediction for 2018 is 0.6419 and the prediction interval is (0.8157, 0.4681), this means we have 95% confidence that the true average ocean temperature would be between 0.8157 and 0.4681.

From the plot below, the overall trend of the next 10-year-prediction shows a upward trend. However, the confidence interval for the next five years are very similar, which means there is not much difference in global ocean temperature for the next five years. The same conclusion applies for the second five years later due to the similar and overlapping confidence interval according to Table 3.

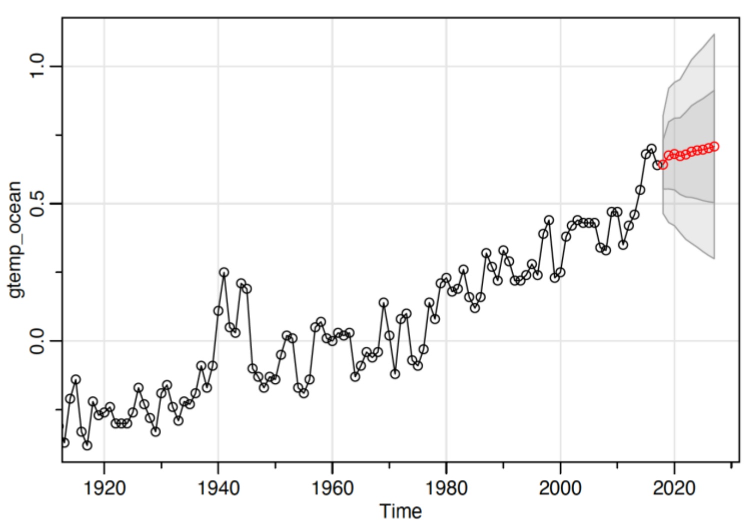


Table 3: 95% Prediction Interval

Year	Prediction	Lower.Bound	Upper.Bound
2018	0.6419004	0.8156601	0.4681407
2019	0.6755890	0.9150341	0.4361438
2020	0.6805481	0.9360674	0.4250288
2021	0.6730906	0.9464228	0.3997584
2022	0.6788458	0.9813906	0.3763010
2023	0.6893478	1.0165345	0.3621612
2024	0.6938232	1.0391833	0.3484632
2025	0.6965346	1.0600756	0.3329936
2026	0.7019740	1.0846533	0.3192947
2027	0.7080436	1.1084716	0.3076157

After performing spectral analysis on the original time series, the periodogram plot for frequency and spectrum is shown in the plot below, and the first three dominant frequencies are identified as in Table 4. The dominant frequencies are $\omega = 0.0069, 0.0139,$ and 0.028, where ω is the frequency.For the global mean ocean temperature series, a 95% confidence interval for the spectrum $f(0.0069)$ is [0.1452, 21.1590], and f is the spectrum density. Similarly, the 95% confidence intervals for the spectrum $f(0.0139)$ and $f(0.0208)$ are [0.0577, 8.4012] and [0.0382, 5.5692]. We cannot establish significance of the first peak, since the periodogram ordinate is 0.5357, which lies in the confidence intervals of the second and third peaks. We cannot establish significance of the second peak, because the periodogram ordinate is 0.2127 and it lies within the confidence intervals of the first and third peak. The periodogram ordinate for the third peak is 0.1410 and it lies within the confidence interval of the second peak. All the three confidence intervals are not very wide, so they are acceptable. However, the confidence interval for the first dominant frequency is wider than the other two, it means the other two spectrums are easier to identify than the first one.

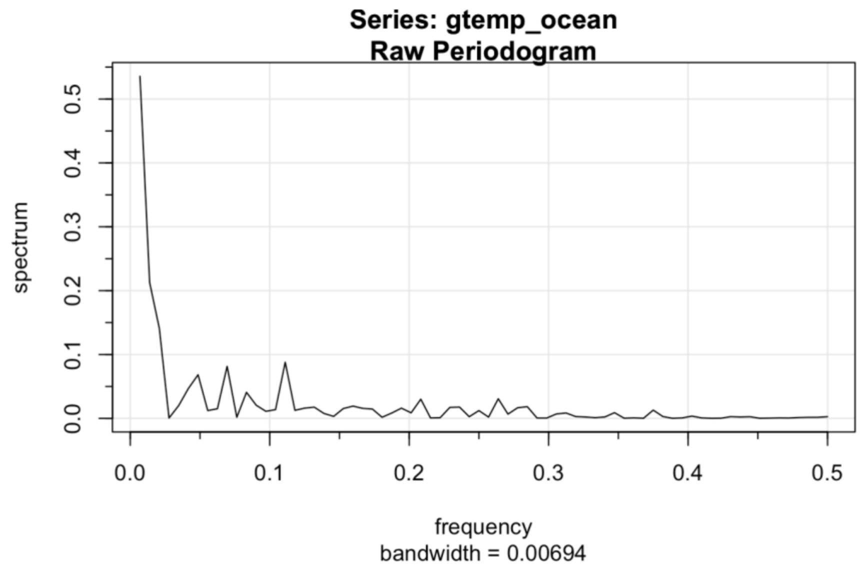


Table 4: The First 3 Dominant Frequency and CI

frequency	period	spectrum	Lower	Upper
0.0069	144	0.5357	0.1452203	21.159020
0.0139	72	0.2127	0.0576598	8.401201
0.0208	48	0.1410	0.0382230	5.569203

Discussion

It is good to see that we finally get a model that could be used to make the prediction for the mean global ocean temperature. However, this model is not perfect, and there are still some limitations that need to be worried about. For ARIMA(2,1,0) model, only one of the parameter estimates is significant, and this may be the problem. If all the estimates are statistically significant, then this model is really meaningful and helpful. Also, there are some outliers that appear at the tails of the Q-Q plot, and it limits the model prediction. In addition, there is some seasonal trend in this series, which is typical for temperature data. The ocean temperature may be affected by seasons, but the model I chose did not account for it. Thus, a SARIMA model could have been a better choice for this data since it can explain the seasonal effects much better.

Overall, the final model ARIMA(2,1,0) helps us to predict the mean ocean temperature for the next ten years. The prediction indicates that there is still an increasing trend in the temperature from 2018 to 2027, and it is not a good sign. People should be more aware of the global warming issues, and it is urgent to limit greenhouse gas emissions. We should also protect and restore marine and coastal ecosystems.

References

IUCN. (2018). Ocean warming. https://www.iucn.org/resources/issues-briefs/ocean-warming

NOAA. (2017). Climate at a glance. https://www.ncdc.noaa.gov/cag/global/time-series/globe/ocean/ytd/12/1880-2017