Time series analysis on the pattern and trend of global occean temperature

Abstract:

Affected by the global warming, the global ocean temperature has presented an upward trend recently, rising by around 1℃ from 1990 to 2015. This project is aimed to apply an appropriate model to investigate the future trend of global ocean temperature based on existing data, helping the government to timely take measures in terms of any possible marine disasters in future. The data used for this project is the time series gtemp_ocean including 138 observations from 1880 to 2017. I model global ocean temperature with the ARIMA(2,1,0) model, using its two previous values(differences) of global ocean temperature as predictors. The predictor that is explained as first past difference[1] of global ocean temperature ((**temperature value with lag =1**) – (**temperature value with lag = 2**)) has an estimate = -0.0519 and it is is not significant. The predictor that is explained as second past difference[2] of global ocean temperature ((**temperature value with lag = 2**) – (**temperature value with lag =3**)) has an estimate = -0.4375 and it is statistically significant. Relying on this model, I also conduct forecasting and peridogram analysis on the data. Based on results, I predict that the global ocean temperature will rise from 0.64 to 0.71 ℃ in next 10 years. I also detect there appears to exist periodic behavior in global ocean temeprature change, with the first three predominant periods 0.5357, 0.2127 and 0.141. However, only the first predominant period is statistically signifcant. It

[1]The first past difference refers to the difference between the **first past value of global ocean temperature (with lag =1)** and the **second past value of global occean temperature (with lag =2)**. For example, if currently t = 2016, then the **first past difference** refers to the difference between global ocean temperature in 2015 and global ocean temperature in 2014. In this report, 'first past difference' is in the abbreviated form used for whole report, which aims to help me present results.

[2] The second past difference refers to the difference between **the second past value of global ocean temperature (with lag =2)** and **the third past value of global occean temperature (with lag =3).** For example, if currently t = 2016, then the **second past difference** refers to the difference between the global ocean temperature in 2014 and the global ocean temperature in 2013. In this report, 'second past difference' is in the abbreviated form used for whole report, which aims to help me to present results.

indicates that there is signifcant dominant periodocity of 144. 92754 years in global ocean temperature change. Based on forecasting and detected cyclical behavior, I conclude that in furture the increase in global ocean temperature will slow out but continues, and the government should still be aware of it and make preparations for possible marine disasters. Further study and modelling shoud be improved by taking account into the seasonal behavior and addressing the related outliers in order to increase forecasting precision.

Introduction:

Since 1990, greenhouse gas emissions have increased significantly due to industrial development and excessive human activities. The heat of greenhouse gases emitted by industrial productions and human activities is mainly absorbed by the oceans, resulting in the rising global ocean temperature. From 1990 to 2015, the yearly global ocean temperature rose from - 0.4 ℃ to 0.5 ℃, a total of about 1 ℃. If not timely alleviating or addressing it, the constant ocean temeprature rise, sea level rise may lead to typhoon, tsunami, endangered marine species and other marine disasters. To help the government make sufficient preparations and timely take effective measures to address predictable marine disasters, this project is aimed to obtain useful information about future trend of global occean temperature by applying an appropriate model to conduct forecasting and periodogram analysis on existing data.

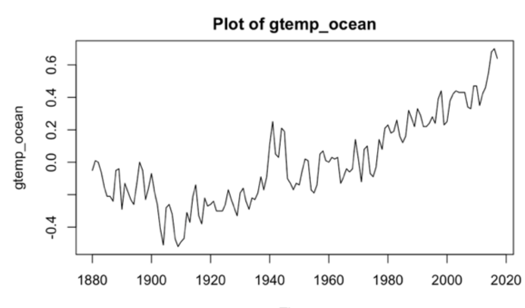
Statistical method:

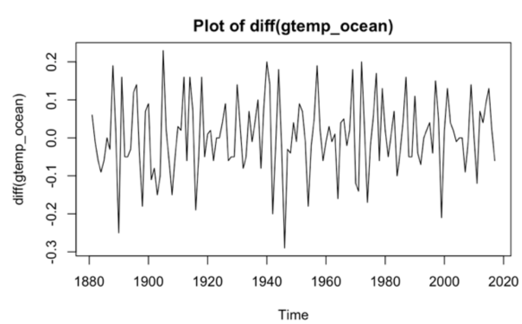
In this project, firstly I construct the time plot of data gtemp_ocean and inspect any possible anomailies. If there are anomalies inspected, then I suggest the appropriate transformation on data. After inspecting possible anomailies and making transformation on data, I am going to identify the dependence orders of ARIMA models and propose two models for preliminary analysis.

After finishing proposing two models, I build two proposed models and fit two proposed models on data respectively. I estimate the parameters and test the significance of paramerters for two proposed models respectively. I also perform all necessary diagnostics including standardized residuals, ACF of residuals, normal qq plot and p-values for Ljung-Box statistic to check if the model follows the model assumption and how well the model fits into data for two proposed models respectively. If both two models follow the model assumption, then I will choose the appropriate final model based on the comparisons of AIC, BIC of two proposed models. If one of two proposed models does not follow model assumption, then I will directly choose the other model as a final model.

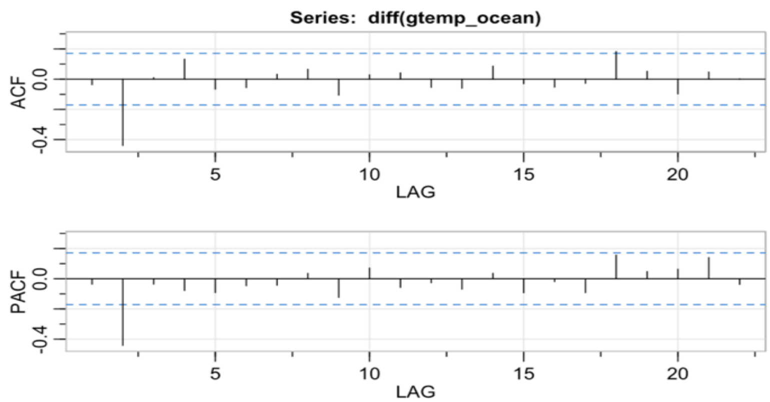
After deciding the final model, I apply the final model to conduct the forecasting for global ocean temperature in next 10 years. I also construct the 95% prediction intervals for each forecast point of global ocean temperature in next 10 years. Last but not least, I perform the periodogram analysis on the data and identify the first three predominant periods to detect the possible possible periodic behavior in the global ocean temperature. I also obtain the 95% confidence intervals for previously identified predominant periods and see whether they are significant predominant periods.

Result:





I construct the time plot of data gtemp_ocean as seen in the ‘plot of gtemp_ocean’, it presents an upward trend. Hence, I take a difference of it to make it stationary as seen in the ‘plot of diff(gtemp_ocean)’.



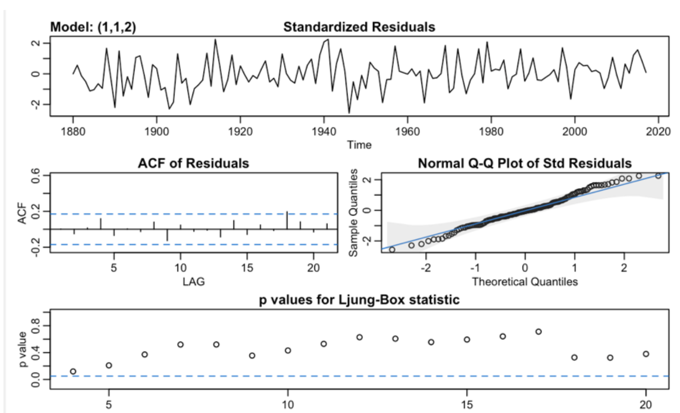
After appropriately transforming the data, I identify the dependence orders of ARIMA models and propose two models used for preliminary analysis. Seen from the PACF of diff(gtemp_ocean), PACF is cutting off at lag = 2. I may suggest gtemp_ocean follows ARIMA(2,1,0) model. Rather than focusing on one model, seen from the ACF of diff(gtemp_ocean), ACF is cutting off at lag = 2. I may also suggest gtemp_ocean follows ARIMA (1,1,2) model.

After proposing two models, I fit two proposed models on data respectively. For each model, I estimate the parameters and conduct the significance testing by looking at estimates and corresponding p-values. For the previously proposed ARIMA(2,1,0) model, as seen in the ‘relevant summary statistics for ARIMA (2,1,0) model’ below, the estimate of coefficient ar1 is -0.0519. It explains that one more increase in the first past difference of global ocean temperature is associated with 0.0519 ℃ more decrease in the current difference of global ocean temperature on average. The corresponding p-value is 0.5001 and it is larger than the chosen significance level = 0.05. Hence, ar1 is not significant. The estimate of coefficient ar2 is -0.4375.

It explains that the one more increase in the second past difference of global ocean temperature would be associated with 0.4375℃ more decrease in the current difference of global ocean temperature on average. The corresponding p-value is 0.0000 and it is smaller than the chosen significance level = 0.05. Hence, ar2 is statistically significant. For the previously proposed ARIMA(1,1,2) model, as seen in the ‘relevant summary statistics for ARIMA (1,1,2) model’ below, the estimate of ar1 is 0.0851. It explains that one more increase in the first past difference of global ocean temperature is associated with 0.0851℃ more increase in the current difference of global ocean temperature on average. Ar1 is not statistically significant since its p-value(0.6064) is larger than 0.05. The estimate of ma1 is -0.1834. It explains that one more increase in the first past difference of forecast errors is associated with 0.1834℃ more decrease in the current difference of global occean temperature on average. Ma1 is not significant since its p-value(0.1993) is larger than 0.05. The estimate of ma2 is -0.4346. It explains that one more increase in the second past difference of forecast errors is associated with 0.4346 ℃ more decrease in the current difference of global occean temperature on average. Ma2 is significant since its p-value(0.000) is smaller than 0.05.

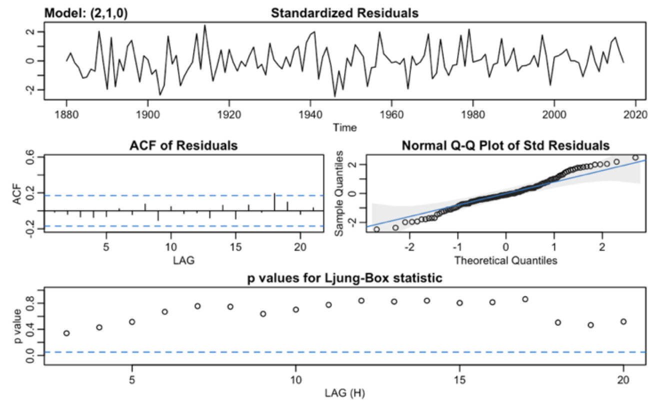
Relevant summary statistics and diagnostics of ARIMA(1,1,2) model

```
         Estimate     SE t.value p.value
ar1        0.0851 0.1648  0.5165  0.6064
ma1       -0.1834 0.1422 -1.2900  0.1993
ma2       -0.4346 0.0722 -6.0230  0.0000
constant   0.0050 0.0032  1.5407  0.1258

$AIC
[1] -1.9348

$AICc
[1] -1.932588

$BIC
[1] -1.828231
```



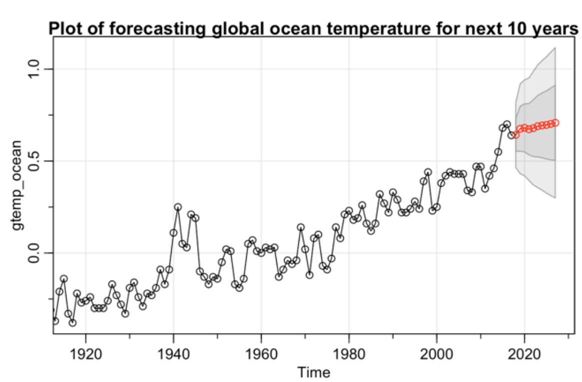
Relevant summary statistics and diagnostics of ARIMA(2,1,0) model

```
         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

$AIC
[1] -1.946634

$AICc
[1] -1.945317

$BIC
[1] -1.861379
```

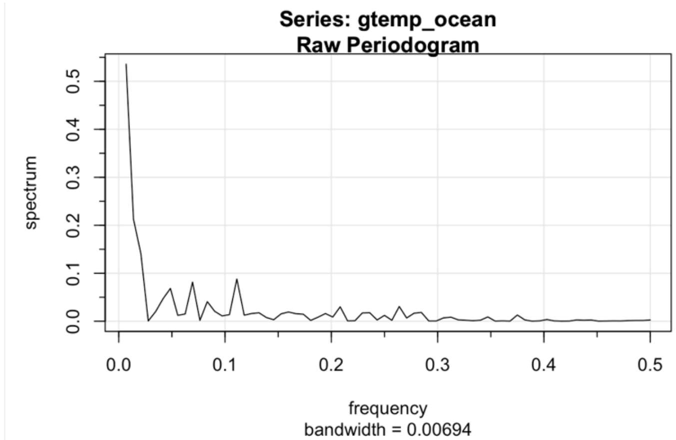


I also check the model assumption for each model by looking at the diagnostics. Seen from 'diagnostics of ARIMA(2,1,0) and ARIMA(1,1,2)' above, the pattern in corresponding diagnostics plots for both models are very similar. Both plots of standard residuals have no obvious pattern. Residuals seems to have mean zero and constant variance except small fluctuations due to outliers. It indicates residuals seem to behave as iid sequence in both models. Both plots of the ACF of residuals show almost no sample ACFs exceeding the limit (blue dashed line), which indicate the randomness is reasonable and residuals are uncorrelated. Both plots of normal qq plot show that almost all points fall on the qq line except 'tails' a little deviated due to outliers, which indicate the assumption of normality is reasonable. Both plots of p-values for Ljung-Box statistic show all p-values are large and Q-statistics are not significant, which indicates residuals are independent. Based on above analysis, both ARIMA(1,1,2) and ARIMA(2,1,0) model follow the model assumption and appear to fit data well. Then I compare the AIC and BIC of them to choose the optimal model. The AIC for ARIMA(1,1,2) is -1.9348 and AIC for ARIMA(2,1,0) is -1.946634. The BIC for ARIMA(1,1,2) is -1.828231 and BIC for ARIMA (2,1,0) is -1.861379. Since both AIC and BIC of ARIMA (2,1,0) are smaller than ARIMA (1,1,2) model, I finally choose the ARIMA (2,1,0) model as my final model.

After choosing the final model: ARIMA(2,1,0) model, I apply it on data to forecast the global ocean temperature in next 10 years and construct the 95% prediction interval for each forecasting point. Seen from the 'plot of forecasting global ocean temperature for next 10 years' below, each red circle represents the forecasting of global ocean temperature in each of next ten years. Overall, it shows the slightly increasing trend in global ocean temperature in next 10 years, from around 0.64 ℃ to 0.71℃. This predicted increasing velocity is relatively slighter compared with current/previous years'. The specific prediction value for each of next 10 years is shown in the table beside the 'the plot of forecasting global occean temperature for next 10 years'. Besides point forecasting, I also construct the 95% prediction intervals for each forecasting point, which aims to predict global ocean temperature for next 10 years by accounting for possible variability. As seen in the same table beside 'the plot of forecasting global occean temperature for next 10 years', the prediction interval for global ocean temperature next year is (0.4681407, 0.8156601). It explains we are 95% confident that the next-year global ocean temperature will fall within this range. The explanation for latter nine 95% prediction intervals are same as above. These 95% prediction intervals express how much uncertainty is related to each forecast, which helps to evaluate the reliability of each forecasting more accurately and correctly. By considering any uncertainty, this helps the government to make more comprehensive preparations and reduce the possibility of being misled by prediction due to information asymmetry.



Prediction <dbl>	Lower.limit.for.95..PI <dbl>	Upper.limit.for.95..PI <dbl>
0.6419004	0.4681407	0.8156601
0.6755890	0.4361438	0.9150341
0.6805481	0.4250288	0.9360674
0.6730906	0.3997584	0.9464228
0.6788458	0.3763010	0.9813906
0.6893478	0.3621612	1.0165345
0.6938232	0.3484632	1.0391833
0.6965346	0.3329936	1.0600756
0.7019740	0.3192947	1.0846533
0.7080436	0.3076157	1.1084716



Last but not least, I also perform the peridogram analysis to identify the first three predominant periods and construct the 95% confidence intervals for identified first three predominant periods to test their significance. Seen from the 'Series: gtemp_ocean', the first three predominant periods correspond with the first three highest peaks in the periodogram, which are 0.5357, 0.2127 and 0.141. The first highest peak occurs at the predominant frequency 0.0069 and corresponds with period 144.92754 years. The second highest peak occurs at the predominant frequency 0.0139 and corresponds with period 71.94245 years. The third highest peak occurs at the predominant frequency 0.0208 and corresponds with period 48.07692 years. These are first three prominent periodic components of gtemp_ocean data. By identifying the predominant periods, I could potentially identify the dominant cyclical behavior in the series of gtemp_ocean, especially for those uncommon periodicity(not quarterly, monthly etc.) . Based on above analysis, since gtemp_ocean is the yearly data, it illustrates that there appears to be a periodicity of about 48.07692 years, 71.94245 years and 144.92754 years in global ocean temperature change potentially. Then, I also contruct 95% confidence intervals for aforementioned first three predominant periods to test their significance and identify the significant predominant periods. Seen from the belowing 'Table 1: Identified first three predominant periods', the 95% confidence interval for the first peak/highest spectrum 0.5357 (f(0.0069)) is (0.14522025,21.15902). This confidence interval is too wide. However, the lower limit of 0.14522025 is higher than all other periodogram ordinates (By checking, the maximum peridogram ordinate is 0.08501315), so it is reasonable and safe to say the first peak/highest spectrum 0.5357 is statistically significant. The 95% confidence interval for the second peak/second highest spectrum 0.2127 (f(0.0139)) is (0.05765979,8.401201). This confidence interval is too wide and we cannot establish the signifcance of second peak/second highest spectrum 0.2127. The 95% confidence interval for the third peak/third highest spectrum 0.141 (f(0.0208)) is (0.03822299, 5.569203). The confidence interval is again too wide and we cannot establish the significance of third peak/third highest spectrum 0.141. Hence, specifically there is only one significant predominant period and the corresponding significant periodicity in global ocean temperature change is 144.92754 years.

Table 1: Identified first three predominant periods

period	frequency	spectrum	lower.for.95..CI	upper.for.95..CI
144.92754	0.0069	0.5357	0.14522025	21.159020
71.94245	0.0139	0.2127	0.05765979	8.401201
48.07692	0.0208	0.1410	0.03822299	5.569203

Discussion:

In conclusion, it is appropriate for us to use the ARIMA(2,1,0) model to conduct the valid forecast of global ocean temperature based on two previous values of global ocean temperature. Through forecasting, I predict that in next 10 years, the global ocean temperature will rise from 0.6419004 ℃ to 0.7080436 ℃, around 0.06 ℃ amount of increase in total. It is slower than current/past year's increase in global ocean temperature. The slower but continuous increase in the global ocean temperature in future reflects that the global ocean warming will be alleviated a little in future. However, the government should still be alert to the possible marine disasters caused by this slow but continuous increase and take timely measures. I also detect there appears to have a potential dominant periodocity of 48.07692, 71.94245 and 144. 92754 years in global ocean temperature change pattern. After further investigating, the only significant periodicity is 144.92754 years. Overall, by investigating the future trend and cyclical behavior of global ocean temperature change, the government could have a overview of future global ocean temperature. Furthermore, the govenrment could make sufficient preparations and take effective measures timely to respond any predictable marine disasters due to global ocean temperature changes in future.

However, there are some limitations in my chosen ARIMA(2,1,0) model. My model does not address the existing outliers well. It affects the normality to a certain extent. Due to those outliers, fitting model on data does not give totally precise estimate of parameter estimates. Furthermore, in ARIMA(2,1,0) model, the predictor that is explained as the first past difference of global ocean temperature (ar1) is not statistically signifcant. Last but not least, my model does not take account into those periodic/seasonal behavior in the global ocean temperature. Due to these limitations, my model cannot achieve the 100% accuracy of prediction. Further study and modeling should focus on addressing those outliers and takes account into the periodic/seasonal behavior in global ocean temperature by fitting an appropriate seasonal ARIMA model.