MTH442A Group Project

Group ID – 5

Group Members –

Mihir 210607

Prajjwal Kumar 210736

Vala Yash Vipulbhai 211142

Anushka Panda 200174

Alok Yadav 210102

Project Topic – Global Temperature Change Prediction using Time Series Forecasting Techniques

Motivation – Global climate change is a very serious problem troubling the modern world to its core. We've already started observing the ill-effects of global warming, effects like shrinking of glaciers, rising sea level, changes in climate patterns across the world, etc. which in turn are having serious ecological effects, also leading up to natural disasters like forest fires, tsunamis, intense heatwaves, and other such disastrous events. All of these are effects that scientists had predicted in the past, but they are now becoming a reality. The amount of warming that has occurred globally since the end of the 19th century has only increased by about 0.75 degree Celsius. Although this may seem like a little change, scientific data indicates that it is already having a variety of effects on people and animals all around the world. However, something that is even more worrying is the fact that scientists are expecting the temperature rise to accelerate greatly in the coming years. Temperature, as we know, plays a great role in what happens on Earth, the only home we have. Shifting the temperature by only a couple of degrees can throw an entire ecosystem into chaos. As students of this course on Time Series Analysis we have the ability to forecast future temperature variations and hence influence the public's view on climate change and global warming, stressing on the need for urgent action, and then educating people about the necessary actions that need to be taken. We've chosen this problem statement to focus on the growing worry over global warming and to draw useful insights from real-world data that we can now access easily in this age of information.

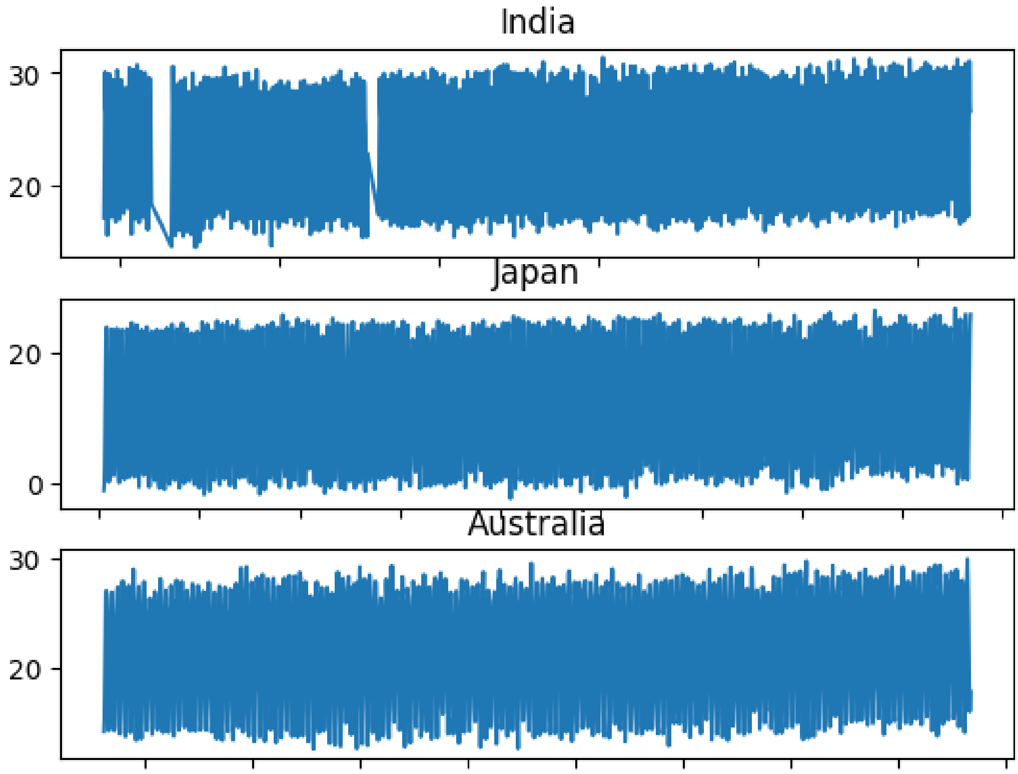
Problem Statement – Predicting the 10-year (2024) temperature change for 3 countries – India, Australia, Japan, using the dataset that lists the global temperature from the eighteenth century to the year 2013 for every month, country-wise

Dataset Used – Climate Change: Earth Surface Temperature Data (kaggle.com) (used the file GlobalLandTemperaturesByCountry.csv)

Models Used – Since our data is a time series, we tried 3 models - AR (Auto-Regressive), MA (Moving Average), ARIMA (Auto-Regressive Integrated Moving Average), and checked which one worked the best

Approach Explained –

We imported the data from Kaggle in a .csv file, and loaded it into our Jupyter Notebook. Below you can find the plot of the imported data. The data lists the global temperature from the 18th century to the year 2013 for every country, month-wise. We'll be carrying out the analysis for 3 countries- India, Australia, Japan.



Then we cleaned the data, doing actions like dropping NaN values, etc. and also calculating the average temperature for each month. For the sake of simplicity, we have used the average temperature for each month.

Then we checked for stationarity in the time-series of the 3 countries using the augmented dickey fuller (ADF) test. In the ADF test, if the p-value is less than 5% then we reject the null hypothesis. The null hypothesis in our case is that the time series is not stationary. However, the p-values obtained for all 3 countries were less than 0.05 (given in the Jupyter Notebook), hence we rejected the null hypothesis, and established the fact that all the 3 time series are in fact stationary.

If the time series had turned out to be non-stationary (i.e. p-value greater than 0.05) then we would have taken successive differences till it becomes stationary. However, we didn't need to do that in our case, as the time series were already stationary.

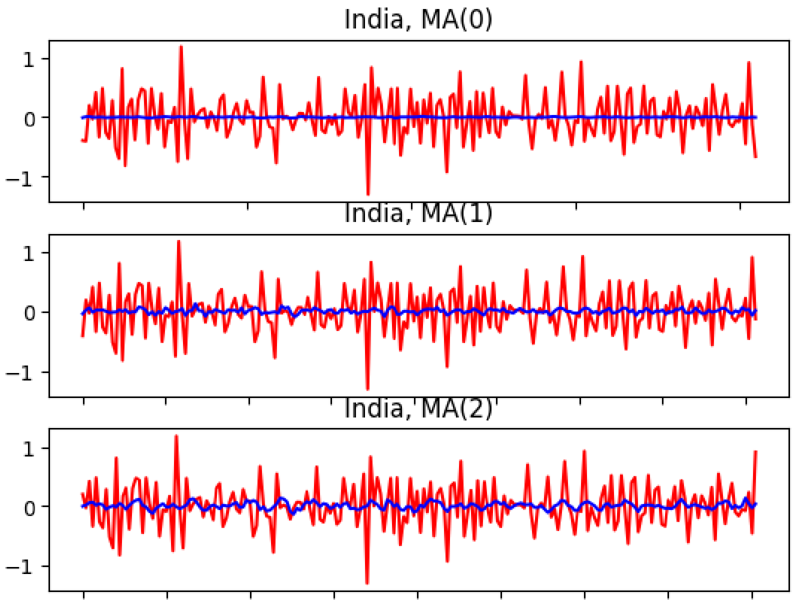
After this we removed the seasonal component of the data by using the moving average filter, taking the average of 12 month, that is a moving average with window size 12.

The we checked for trend in the data. For doing that we used the least square estimator method to find the best-fit line for the data. Then using the slope of that line, we inferred whether the time series had any significant trend or not. It turned out that the slope of the line in our case was between 0.006 and 0.01, which indicated a positive trend. Which indicates that the global temperatures are increasing, a matter for worry.

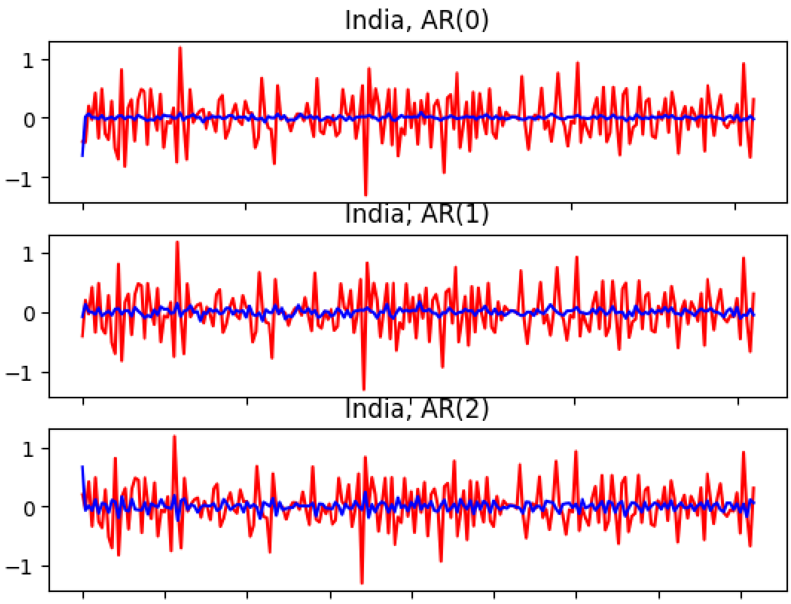
Then we plotted the ACF (Auto-Correlation Function) and PACF (Partial Auto-Correlation Function) for each of the 3 time series. The ACF and PACF are usually used for finding the parameters p and q of the AR and MA part of the ARMA (p,q) process respectively.

After this we proceeded to see if the MA or AR can represent our time-series well. De-trending wasn't really required as the trend in the time series was very subtle (0.006 to 0.01).

Then we tried fitting MA (0), MA (1) and MA (2) to our time series and plotted the predicted values with the actual values to see how well that process would represent our data (checking goodness of fit). The result was disappointing, as the MA process didn't really capture the details of the time-series sufficiently well. It can be seen in the MA plots for India below.



After this we tried fitting AR (0), AR (1) and AR (2) to our time series and plotted the predicted values with the actual values to see how well that process would represent our data (checking goodness of fit). The result was again slightly disappointing, as the AR process didn't really capture the details of the time-series sufficiently well. It can be seen in the AR plots for India below.



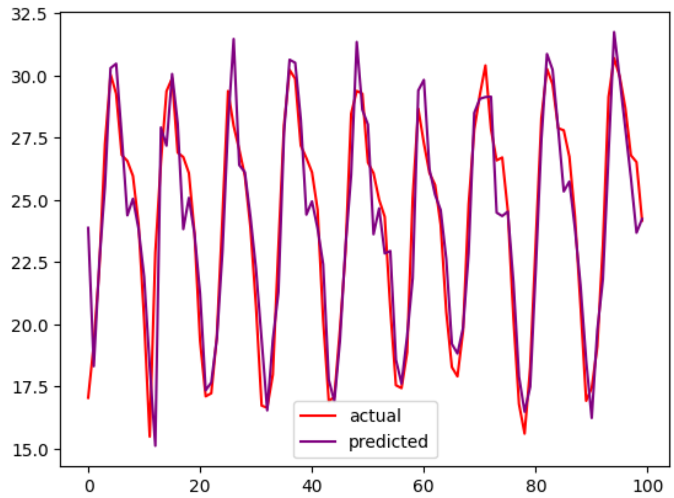
Now we can try with the original dataset, since the ACF of ARMA doesn't follow anything like a cutoff or a decay pattern. Also note that this is a stationary process if we take the threshold p-value to be 0.01, hence this ARMA process is definitely an ARIMA process!

In order to compare the results across combinations of p and q, we will use the Akaike and Bayesian Information Criteria

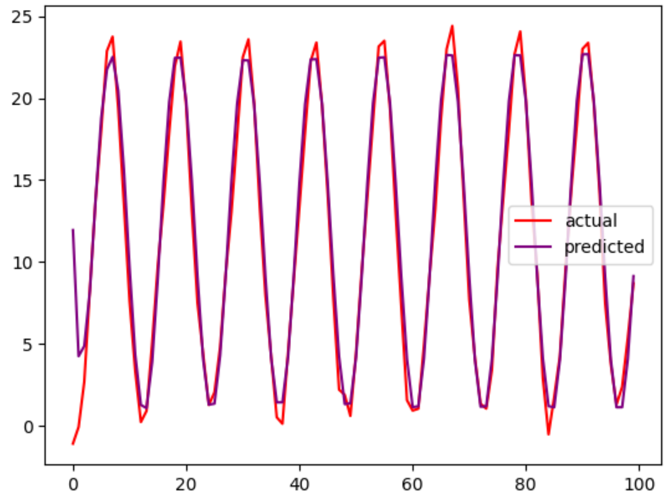
We used AIC (Akaike Information Criteria) and BIC (Bayesian Information Criterion) for calculating p and q in the ARIMA (p,q), we didn't use ACF or PACF.

Performance Evaluation –

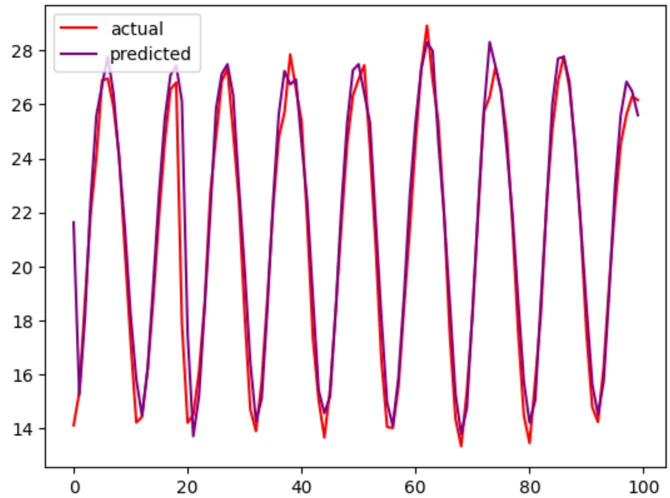
The actual and predicted (as predicted by the ARMA process) values for each of the 3 countries can be seen in the plots below.



plot for India



plot for Japan



plot for Australia

We also calculated the Mean-Squared Error and Mean-Absolute Error for each of the 3 countries and the results were as below-

India

MSE - 1.8866689749025167

MAE - 1.108009112888801

Japan

MSE - 1.5895905039744935

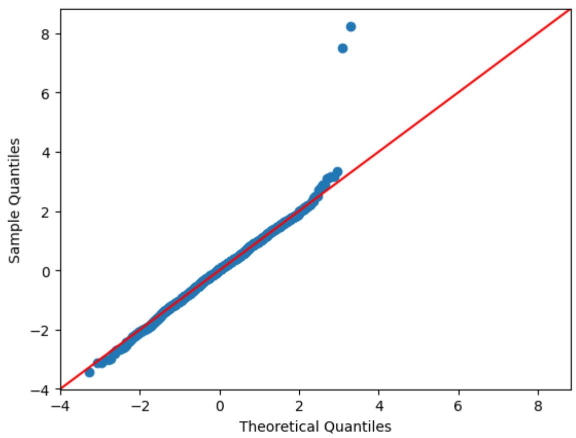
MAE - 1.0025434871936652

Australia

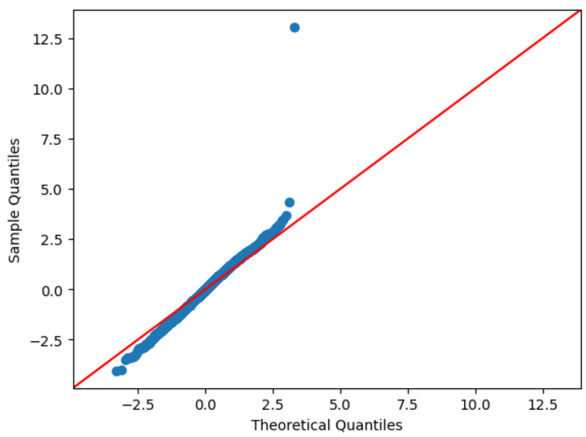
MSE - 1.081197655540619

MAE - 0.807466327703222

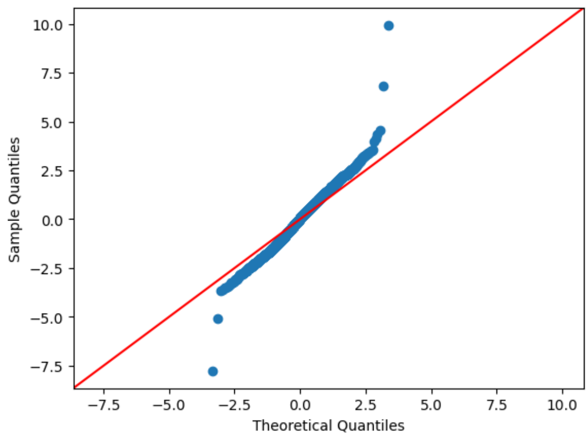
We also plotted the Q-Q plots for each of the 3 countries, they showed that the residue is as expected to be white noise, since the plots were mainly along the Y=X line.



plot for India



plot for Japan



plot for Australia

Results –

The average monthly predicted values for the year 2024 were calculated using our ARMA model for each of the 3 countries, and were plotted on a graph. The graphs can be seen below.