STATISTICS FOR SPATIO-TEMPORAL DATA

Professor: Gaetan Carlo
Student: Dao Quang Hoan(871510)
Ca'Foscari, Venice 2017/2018

I. Introduction

It is widely known that the global average temperature has increased, especially in recent decades. We are going to do a basic exploration of the global temperature time series as available in the official website NASA - Global Climate Change. The global temperature dataset reports the deviation in degrees centigrade from 1880-2017 global mean land-ocean temperature. The dataset contains 3 columns: the first column shows sequence of year from 1880 to 2017; the second column is the most important part, which displays the global average temperature deviation; the third column indicates the lowess smoothing which is locally weighted scatterplot smoothing that can help to discover non-linearities.

II. Methodology

1. Statistical indices

We can see a remarkable increase of the temperature deviations in the last decades. The global temperature time series appears to be non stationary due basically to the last decades upward trend. A plot of global temperature against its lowess smoothing helps understand better (Figure 1).

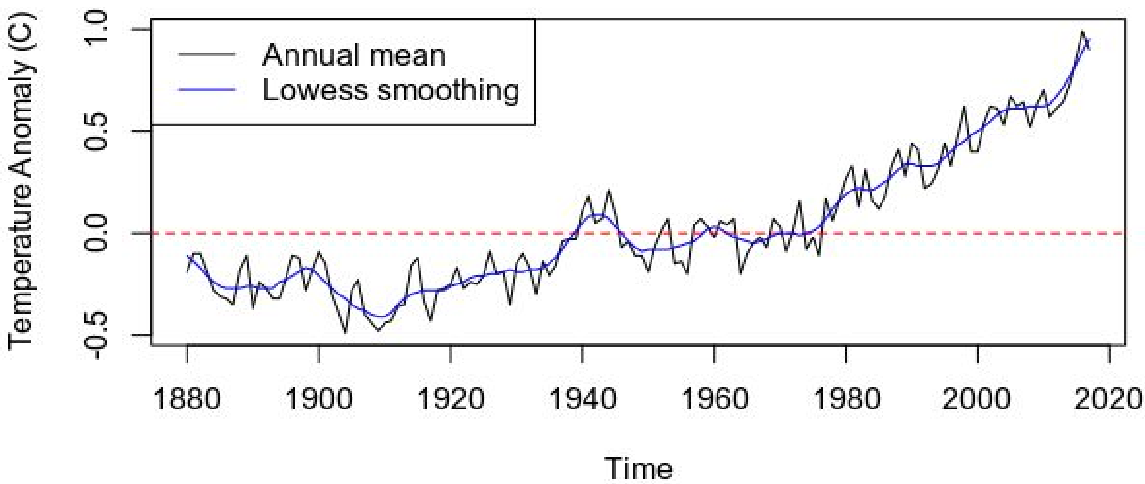

Figure 1: Yearly average temperature with Lowess smoothing

Figure 2 shows the 10 warmest years on record from 1880 to 2017. The observed warmest year on record was 2016, followed by 2017. This explicitly implies that the global temperature has risen significantly in recent years. Therefore, its trend could be predicted as an upward movement in the future (part II.4).

Ranking	Year	Anomaly(°C)
1	2016	0.99
2	2017	0.90
3	2015	0.86
4	2014	0.73
5	2010	0.70
6	2005	0.67
7	2013	0.64
8	2007	0.64
9	2009	0.63
10	2006	0.62

Figure 2: The 10 warmest years 1880–2017

The period 138 years of data are used to display the following statistical concepts: mean, variance, standard deviation. The mean gives the average of samples. The variance and standard deviation measure the spread of samples. They are large when the samples have a broad spread.

Mean: $\mu(x) = \frac{1}{n} \sum_{k=1}^{n} x_k \approx 0.026$

Variance: $\sigma^2(x) = \frac{1}{n-1} \sum_{k=1}^{n} (x_k - \mu(x))^2 \approx 0.11$

Standard deviation: $\sigma(x) = (\sigma^2(x))^{1/2} \approx 0.332$

2. Polynomial Regression

The global average annual mean temperature apparently does not vary linearly with time. It is thus useful to examine the underlying nonlinear variation of the annual temperature time series. The simplest nonlinear trend exploration is thorough a polynomial fit. The polynomial regression formula is given as belows:

$$y_t = b_0 + b_1 t + b_2 t^2 + \ldots + b_q t^q + \varepsilon_t$$

where b_0 , b_1 ,..., b_q are regression coefficients, and ε_t is a random error or noise.

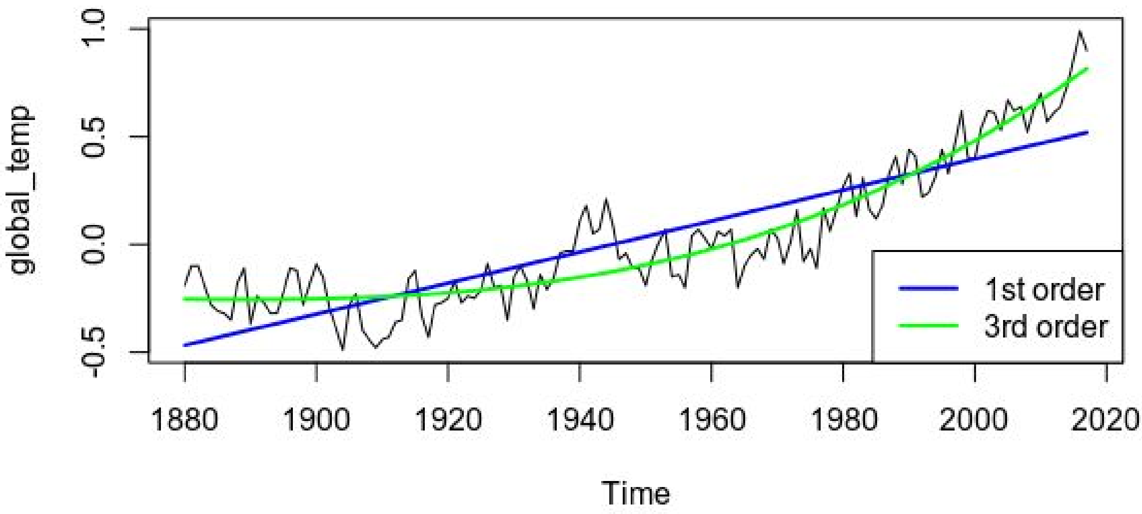

Figure 3: Polynomial regression with orders

Figure 3 illustrates two fits by the 1st order and 3rd order polynomials. I have tried higher order polynomials which often show an unphysical overfit. Notice that, the 1st order of polynomial is linear trend.

3. Differencing Global Temperature

The global temperature series shown in Figure 1 appears to behave more as a random walk than a trend stationary series. Hence, rather than detrend the data, it would be more appropriate to use differencing to coerce it into stationarity. The detrended data are shown in Figure 4 along with the corresponding ACF (Figure 5). In this case it appears that the differenced process shows minimal autocorrelation, which may imply the global temperature series is nearly a random walk with drift. It is interesting to note that if the series is a random walk with drift, the mean of the differenced series, which is an estimate of the drift, is about 0.008, or an increase of about 0.8 degree centigrade per 100 years.

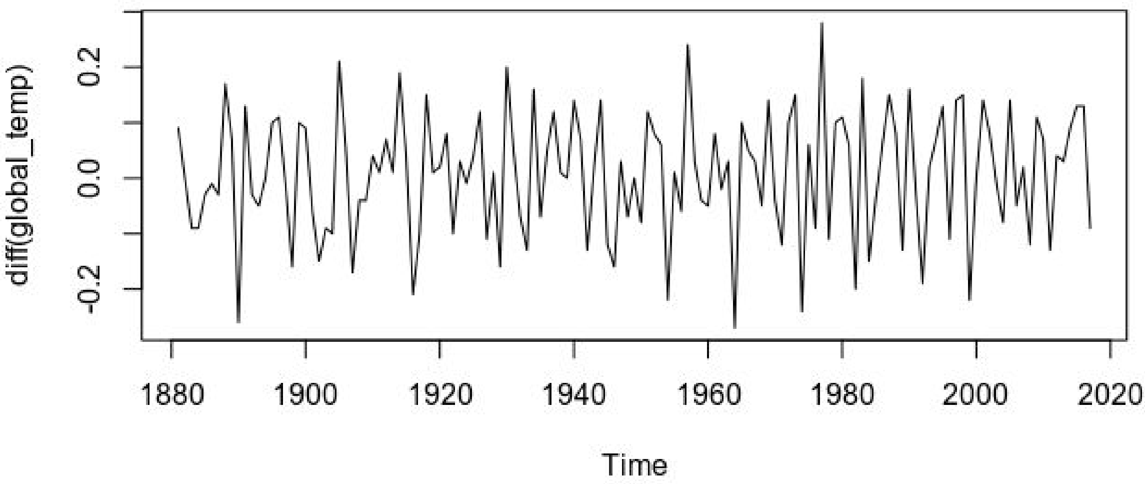

Figure 4: Differenced global temperature series

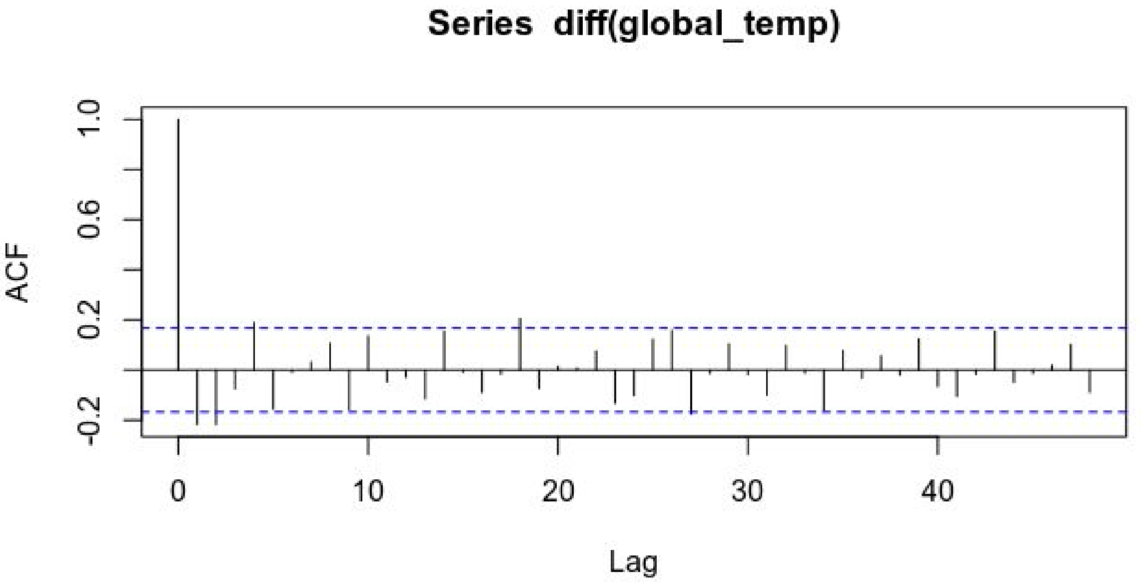

Figure 5: ACF of differenced data

4. Prediction using ARIMA(p,d,q) model

Now it is time to fit an ARIMA model and predict the future 10 years (Figure 6). ARIMA, which determines the optimal autoregressive integrated moving average. The ARIMA model: (p, d, q) refers to the AR order, the degree of difference, and the MA order respectively.

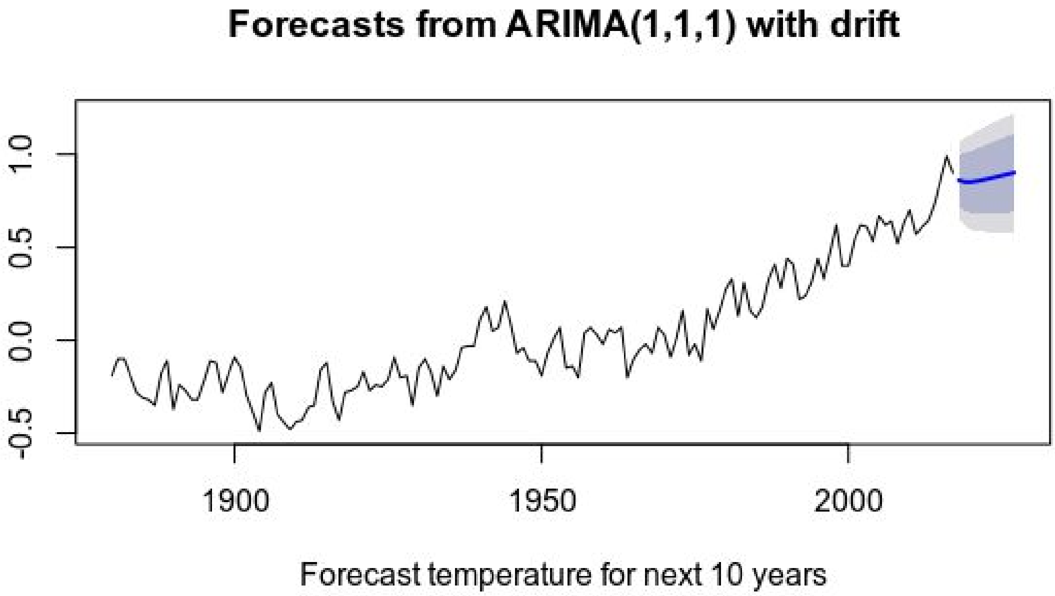

Figure 6: A prediction for time series in next 10 years

We estimate parameter (p,d,q) using the R auto.arima() function for figuring out the optimal selection of exponential which dedicates (1,1,1). The forecasts are shown as a blue line, with the 80% prediction intervals as a dark shaded area, and the 95% prediction intervals as a light shaded area.

III. Discussion and Conclusions

After a basic exploration of the global average temperature dataset, we leveraged on structure change of a typical time series, besides with lowess smoothing. The year 2016 was the warmest year since records began in 1880. Actually, all 10 warmest years have come in the 21st century. Additionally, there are statistical indices such as mean, variance to standard deviation for statistical reference. Trend and 3rd order of polynomial regression were identified.

We also remove trend by using difference to force time series into stationary, with its corresponding ACF. At the end, using ARIMA model is the optimal method for forecasting the future 10 years in how the global temperature anomalies walk. This is the overall process by which we can analyze time series data and forecast values from existing series using ARIMA model.