Modeling and forecasting global mean temperature time series

Final project: STA 6857, Applied Time Series Analysis

Sotuyo, Carlos

Abstract

An ARIMA time series model was developed to analyze the yearly records of the change in global annual mean surface air temperature, one of the most relevant indicators of global warming. Using R statistical software functions an ARIMA (autoregressive integrated moving-average) model was developed in order to obtain an optimal forecasting for the next five years. The development of the model was based on the characterization of the variation patterns of a global dataset, the methodology may be applied to any time series analysis with similar characteristics. The ARIMA model developed in this study may be used as a tool for short term, five or less years, for environmental planning and decision making.

Introduction:

A change in the worlds climate will have a major impact on the lives of many people, as global warming is likely to lead to an increase in natural hazards, such as floods and droughts. It is likely that the world economy will be severely effected as governments from around the globe try to enforce a reduction in fossil fuel usage and measures are taken to deal with any increase in natural disasters.

To determine whether the Earth is warming or cooling, scientists look at annual mean temperatures. At a single station, the warmest and the coolest temperatures in a day are averaged. Averages are then calculated at stations all over the Earth, over an entire year. The change in global annual mean surface air temperature is calculated from a base established from 1880 to 2015, and the result is reported as an *anomaly*. (Montgomery et al, p17).

ARIMA models are powerful tools in the analysis of time series as they are capable of modeling a very wide range of series. Differentiating a series $\{X_t\}$ can remove trends, whether those trends are stochastics, as in a random walk, or deterministic, as in the case of a linear rend. The latter is the case we examine.

The first order differenced series is white noise $\{Z_t\}$ given by $\Delta_t = X_t - X_{t-1}$ and therefore becomes stationary. The KPSS Test for Level Stationarity in R was used in our study.

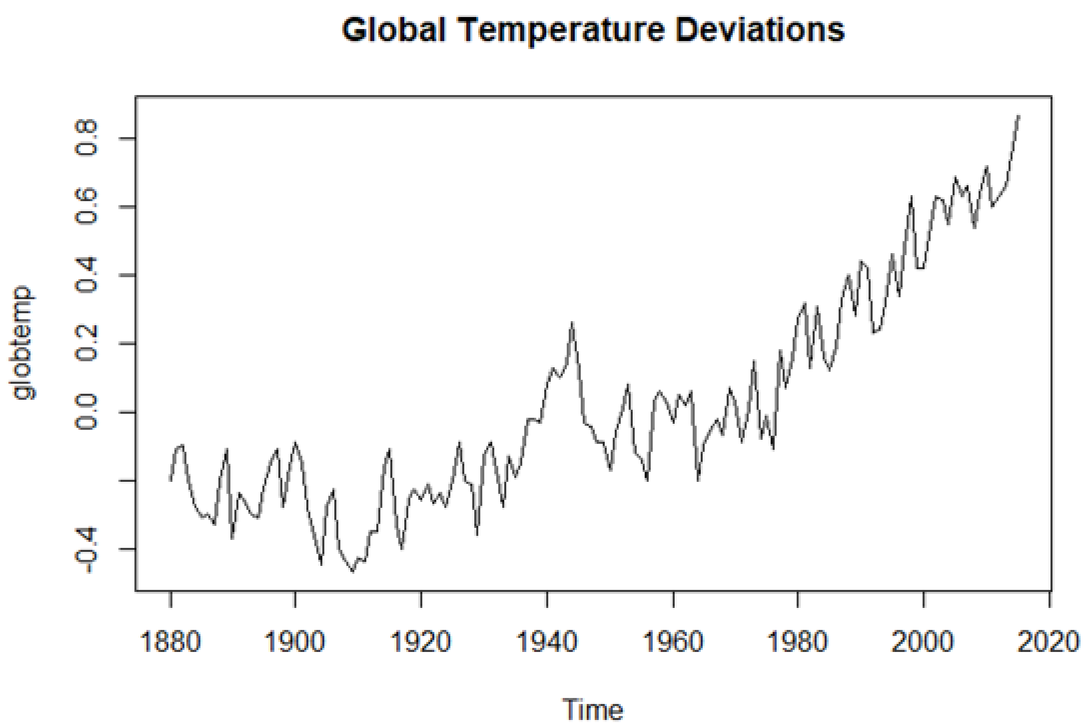
An ARIMA(p, d, q) process can be fitted to data using the R function arima with the parameter order set to c(p,d,q). The auto.arima function finds the best, lower AICc model given a range of p and q. It is equivalent to $model > estimation > autofit$ in ITSM. In absence of auto.arima identification of the appropriate ARIMA model requires skills obtained by experience. Several excellent examples of the identification process are given in Montgomery et al, p235.

Once an appropriate time series model has been fit, it may be used to generate forecasts of future observations. The standard criterion to use in obtaining the best forecast is the mean squared error for which the expected value of the squared forecast errors is minimized. It should be noted that the variance of the forecast error gets bigger with increasing forecast lead times. This intuitively makes sense as we should expect more uncertainty in our forecasts further into the future.

Statistical methods and case studies:

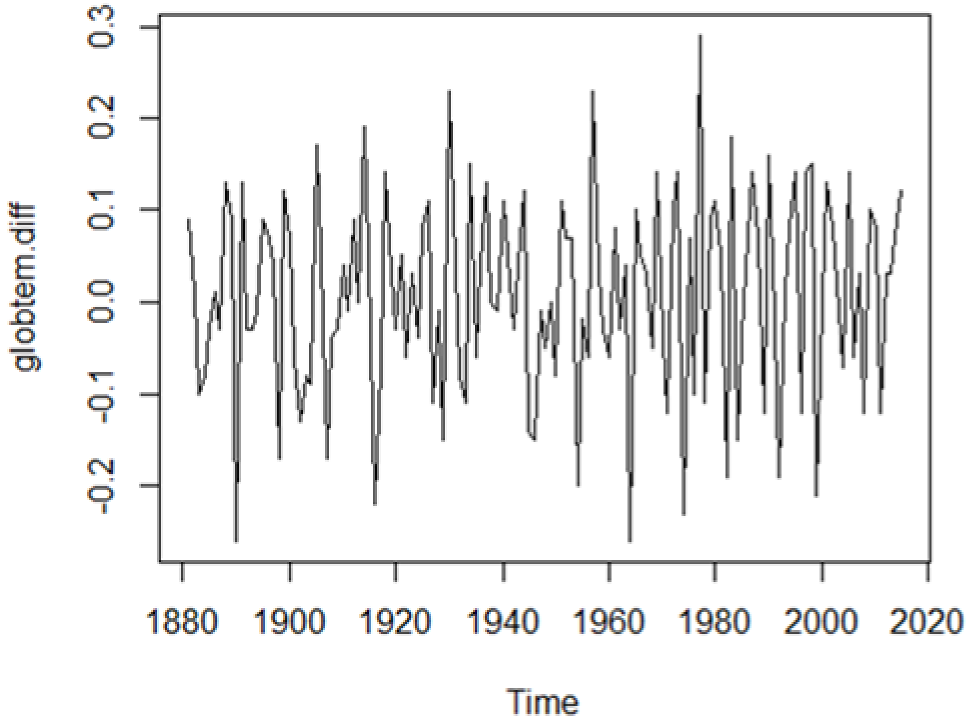
The time series global mean land-ocean temperature deviations measured in degrees centigrade, for the years 1880-2015 was used in this study. In climate change studies global mean land-ocean temperature deviations measured in degrees centigrade has been using in developing prediction models (see Jones and Moberg, et al. 2003).

A plot of the time series data shows that the mean is non-constant and there is an obvious upward trend which become stronger after late 1970s or early 1980s. The variance appears to be pretty consistent however. In order to run a Breusch-Pagan test a linear model of temperatures vs time we fit the model in R, the bptest yields a p value of 0.388; therefore, we fail to reject the null hypothesis that the variance of the residuals is constant.

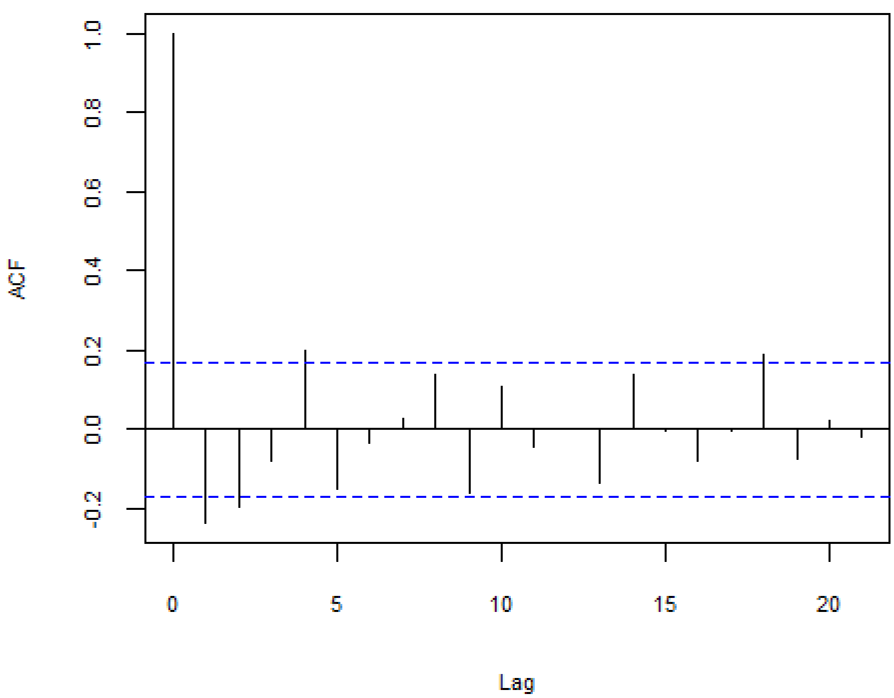


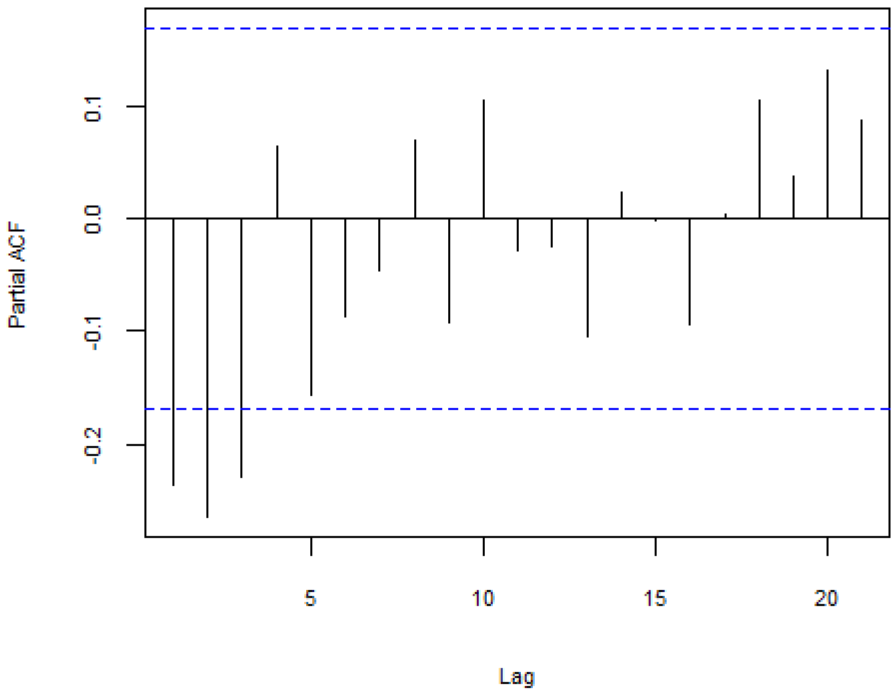
By analyzing the ACF plot of the time series, it reveals a highly auto correlated time series, at the same time there is no evidence of a seasonal component. The stl function in R, which decomposes a time series into seasonal, trend and irregular components using loess, shows that the series is not periodic which proves that the series lacks of seasonality.

The KPSS Test for Level Stationarity was performed in R. A p value of 0.01, the null hypothesis of stationarity is rejected, which confirms our interpretation of the time series plot which shows an upward trend. Once the time series is differentiated, the first order differenced series is white noise and therefore the series becomes stationary.



Our differentiated series ACF/PACT were examined in order to determine an ARIMA model. The ACF/PACF plot give us suggestions on what degree of parameters to utilize. We choose d = 1 as our degree of differencing, evidenced by our earlier call of the diff (difference) function.





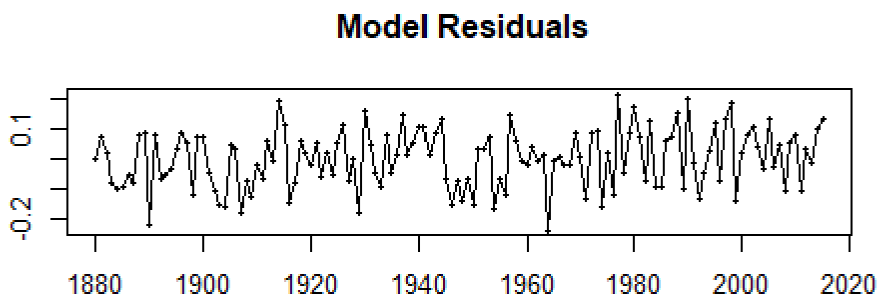
The ACF is used to identify the order of a MA model. Note that the ACF plot has a significant spike only at lags 2 and 4. The PACF is used to identify the order of a RA model. The PACF plot has a significant spike only at lag 3. Three arima models were fit: arima.fit1: (1,1,1); arima.fit2: (1,1,3) and arima.fit3: (3,1,2).

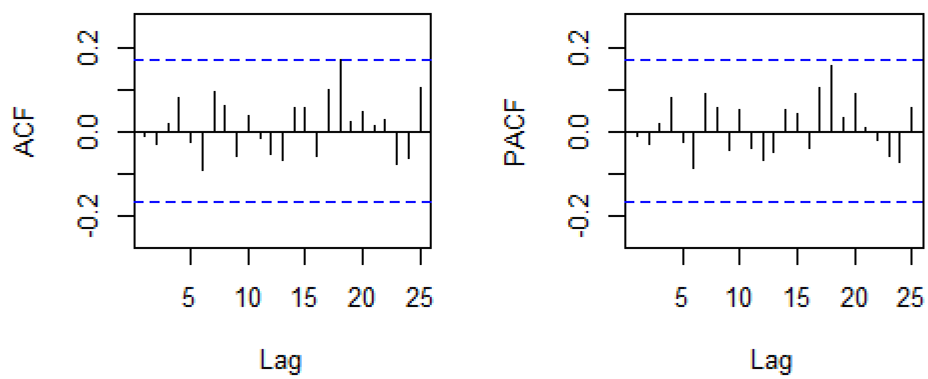
The auto.arima function in R yields arima (1,1,3) as the best fitted model (lower AICc); ITSM software autofit model also finds arima(1,1,3) to be the most optimal model. The AICC for all three models are as follow: arima.fit1: (1,1,1), AICC = -231.46; arima.fit2: (1,1,3), AICC = -233.47 and arima.fit3: (3,1,2), AICC =-231.23. Thus, the arima.fit2(1,1,3)model has the minimum AICC value. The difference of AICC values between the (1,1,3) model and the other two models is slightly greater than two. An AICC value of less than two units between two satisfactory models may be ignored in the interest of model simplicity (Brockwell and Davis, p158.)

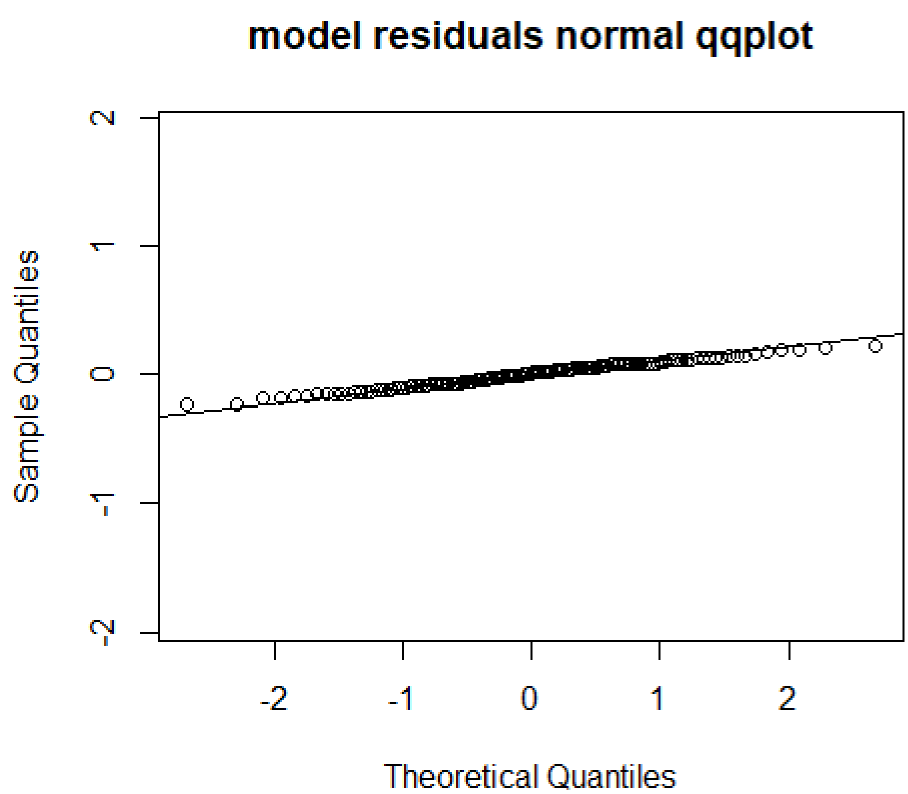
In our case the difference, as previously emphasized, is slightly greater than two. That is the reason our chosen model was model arima (1,1,3) in order to perform the forecast calculations:

$$X_{(t)} = -0.9448X_{(t-1)} + Z(t) + 0.6081Z_{(t-1)} - 0.5681Z_{(t-2)} - 0.31Z_{(t-3)}$$

No coefficient is close to zero. The intercept, 0.0072, which is close to zero, was omitted. The residuals from the chosen model were checked by plotting the ACF of the residuals and by the qq normal plot. The plots of the residuals show a zero mean and all autocorrelations within the limits indicating that the residuals are behaving like white noise. The qq plot of the residual falls on a straight line.



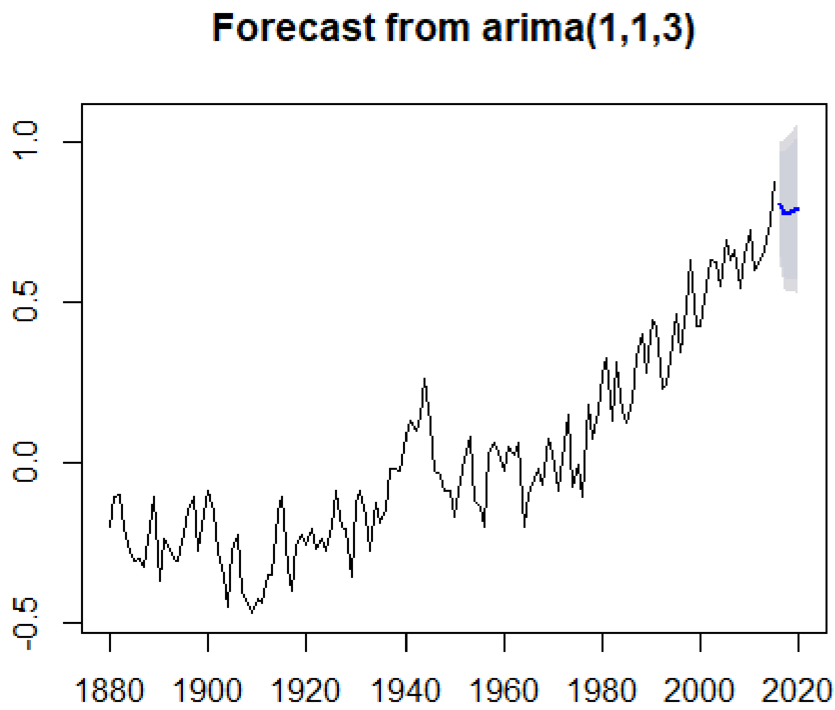




Since the residuals of our model look like white noise, we calculate forecasts.

By using the forecast function in R, the forecast for the next five years were generated.

Point Forecasts: 2016: 0.80; 2017: 0.77; 2018: 0.77; 2019: 0.79; 2020: 0.79



The forecast for five years is shown in blue. The shaded area indicates the 0.90 and 0.95 confidence intervals of the forecasts. A forecast for more than five years may lack of interest due the constant changes the global climate conditions.

Conclusions:

The global annual mean temperature deviations time series shows a homoscedastic, non-seasonal, upward trend. The series first difference is becomes stationary (de-trend). An ARIMA model (1,1,3) yields the minimum AICC value. The model fits well the time series, showing the residuals are white noise. The forecast for the next five years show the mean global temperature deviations fluctuate between 0.80, 0.77 and 0.78 degrees Celsius which is consistent with the time series model.

References

[1] Brockwell, P. J. and Davis, R. A. Introduction to Time Series and Forecasting, third edition (2016), Springer-Verlag, New York.

[2] Cowpertwait, P. S. Metcalfe, A. V. Introductory Time Series with R. Springer, New York-London, 2009.

[3] Jones, P. D and Moberg, A. Hemispheric and Large-Scale Surface Air Temperature Variations: An Extensive Revision and an Update to 2001. American Meteorological Society (AMS). Retrieve from https://journals.ametsoc.org/doi/pdf

[4] Montgomery, D. C, Jennings, Ch. Kulahci, M: Introduction to Time Series Analysis and Forecasting, first edition, John Wiley & Sons, New Jersey, 2008.