Time Series Project

Analysis of Average Annual Temperatures within US

Name: **Zain Attawala**

Email: attawalazain@gmail.com

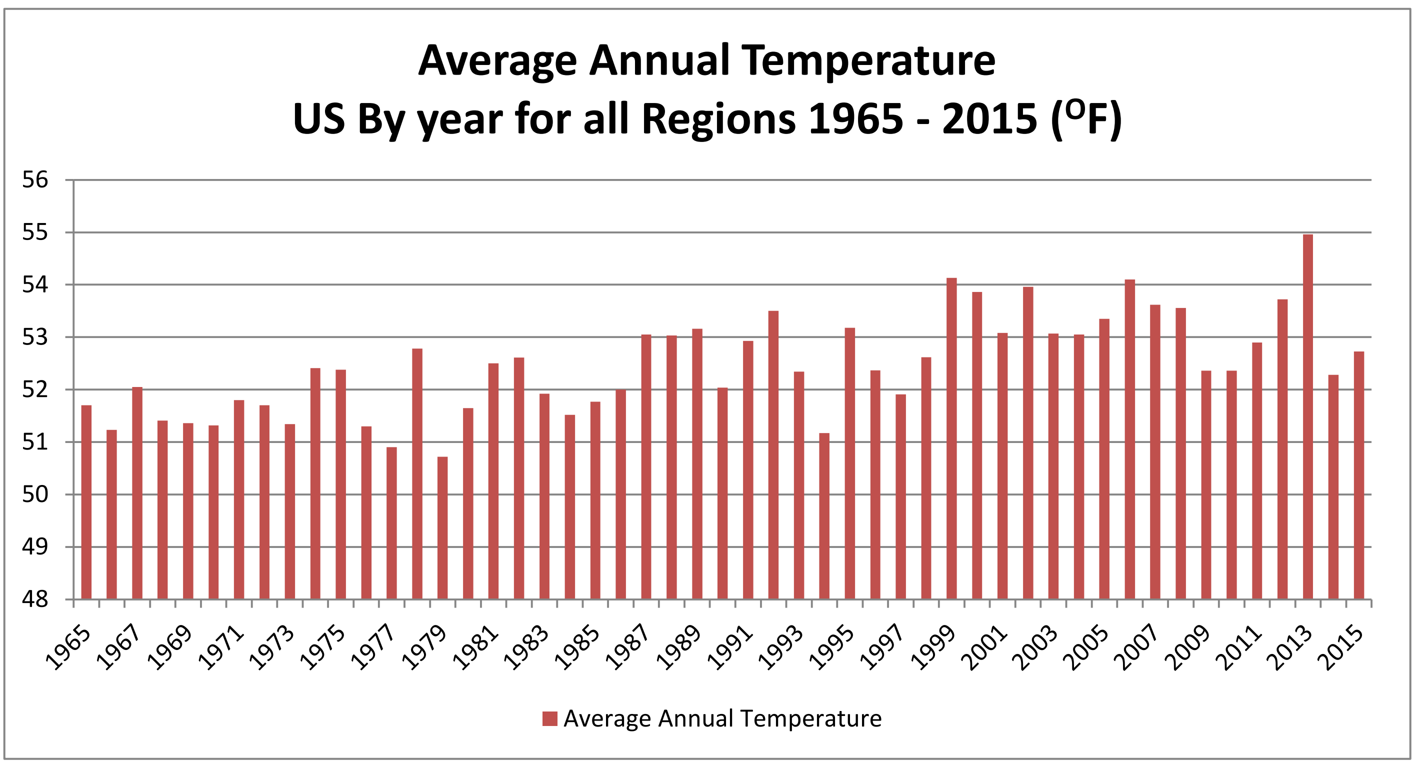
Session: Summer 2014

1. Introduction:

The goal of this project was to determine whether past information on the average annual temperature within United States is reasonable to predict future annual temperatures. This project is a process to determine optimal model and number of parameters to be used in the model. Using the data from the national climatic center, I did some analysis of the average temperatures observed in America for over 50 years beginning 1965. The hope is to discern if the pattern of average past temperatures annually is an accurate predictor for the future years.

2. Data:

The data was acquired for years form 1965 – 2015 to have a wide range of years (50 years). The reason for this selection is the range of factors, that developed in these years, that can contirbute to the patterns of temperatures. An average annual temperature, for each day of the year and for each region in US, is taken and graphed onto a bar chart. Additionally, the attached excel file contains the information, charts and the analysis performed.



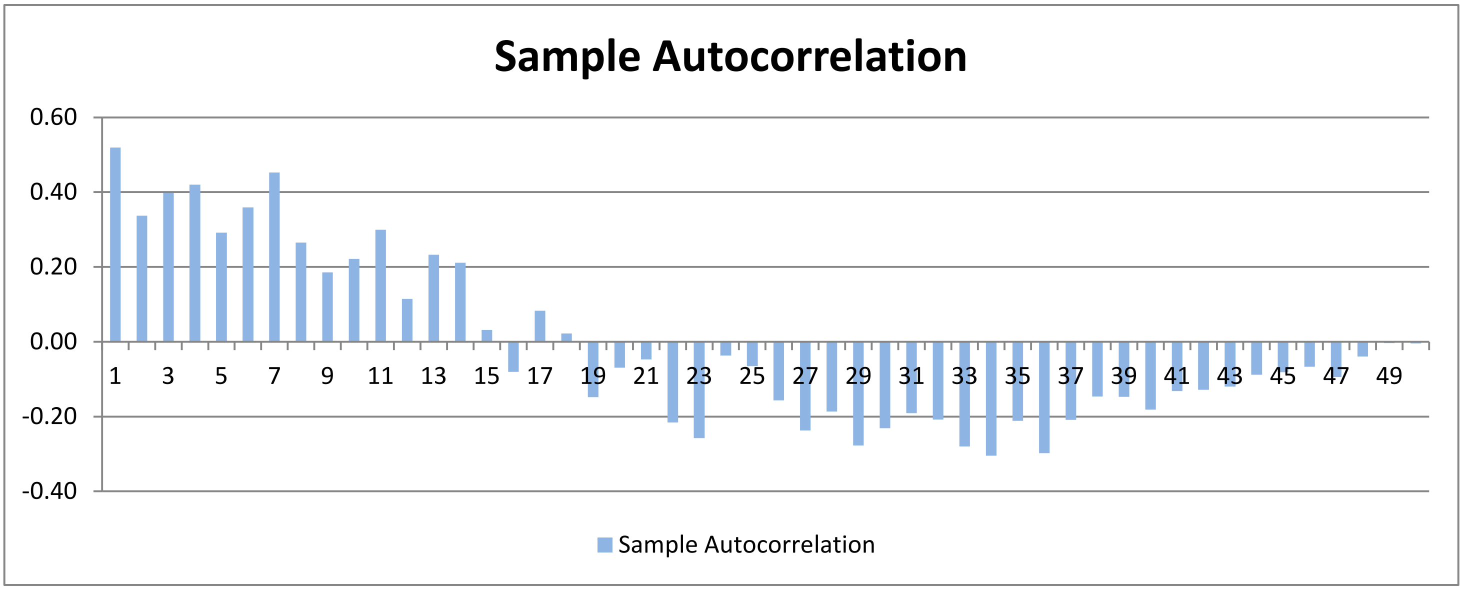
Data Source: http://www.ncdc.noaa.gov/cag/time-series/us

3. Analysis of Data:

The data has a general upwards trend in bigger picture. Althought this trend has swinging periods, in general we can observe relatively warmer temperatures being approached in later recent years. If we observe the the first 10 bars, the range was between 50 °F to 52 °F on average. While the bars mid-ninetees and two thousands from 1999 to 2009 were on average swinging between 53 °F to 54°F. On more close analysis of the warmest temperatures, it was noticed hat few highest average tempeatures annually occured in years 2013, 2006, 1999, 2002) all in past 15 years. However, on the whole the range of average temperatures, dominantly, have been between a very close range, giving my initial hypothesis of this process being stationary.

4. Sample Autocorrelation Function

To prove this idea/hypothesis of stationary, I calculated the sample autocorrelation function at every lag of the given data. The results are displayed in chart below.



The autocorrelation chart above shows the gradual decrease for the lags from 1 to 18. The values of autocorrelation go from 0.52 to 0.02 between lags 1 and 18. The decrement trend continues until age 36, where the chart then increases and goes back to an autocorrelation of 0 at lag 49. At this point, an autoregressive model of first two orders seems to be the model to fit the underlying data.

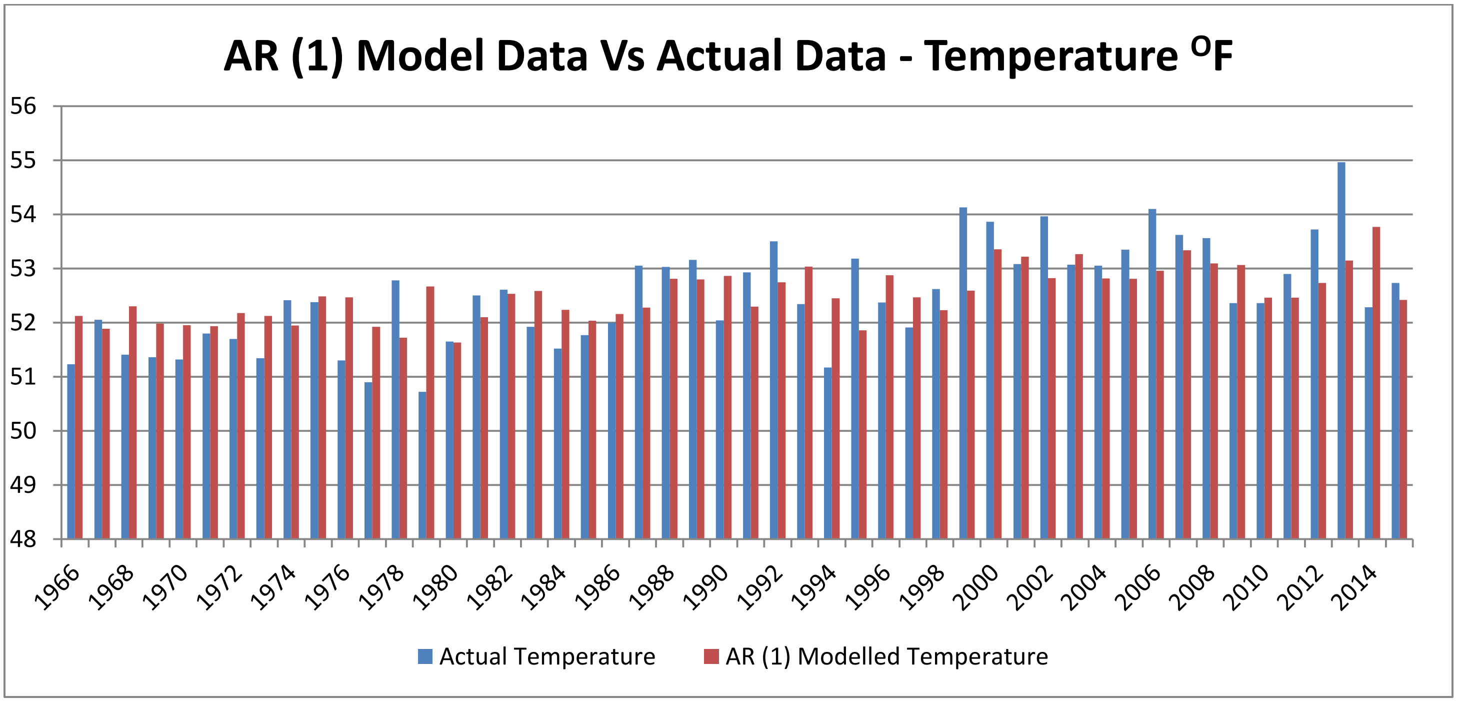
To determine if first or second order Auto Regressive model represents the data in the best possible manner, I will test the fit of both the AR (1) and AR (2) models.

5. First Order Autoregressive Model

To calculate ϕ (phi), method of methods will be used. I will be estimating the value of phi to be the lag 1 sample autocorrelation which according to table in excel file and chart is 0.504. We can then use the formula θ_0=u(1- ϕ), to find θ_0 which according the values equals 26.048. Thus, using the values and observations we derive the following equations:

$Y_t = 26.048 + 0.504Y_{t-1} + e_t$

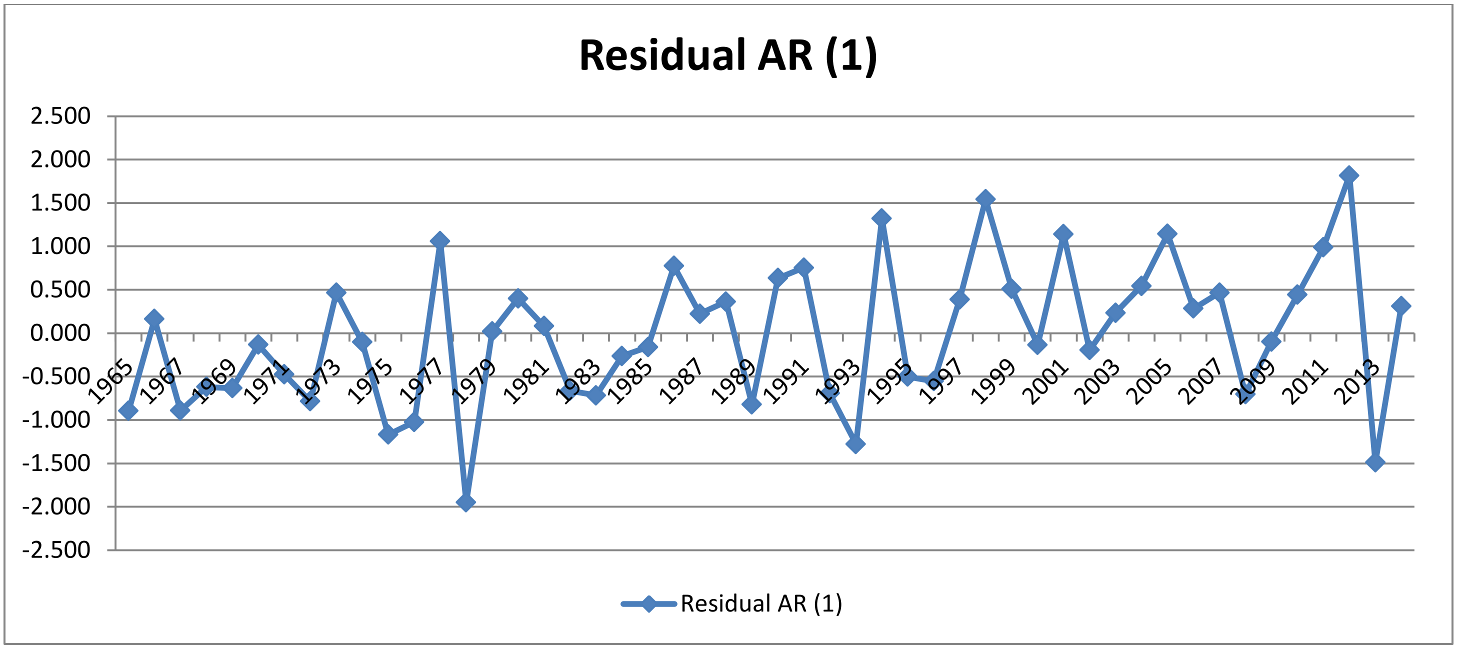
The chart below shows the comparison of the actual data with the estimated data using AR (1) model.



The chart above shows a high correlation between actual temperatures and AR (1) model temperatures. The high correlation can be seen especially in the first 15 – 20 years. The model seems to be little away from actual data during recent ten years, where we can see the largest differences between both figures for same year.

In fact, the absolute change in temperature from year to year has increased dramatically from first 15 years to last 15 years. Thus, an updated model should be tested for better fitting.

The chart below shows the residual of the AR (1). The residuals seem to be fairly evenly distributed reflecting lack of any noticeable pattern, except for the tendency of having a slightly higher residual in the recent years.

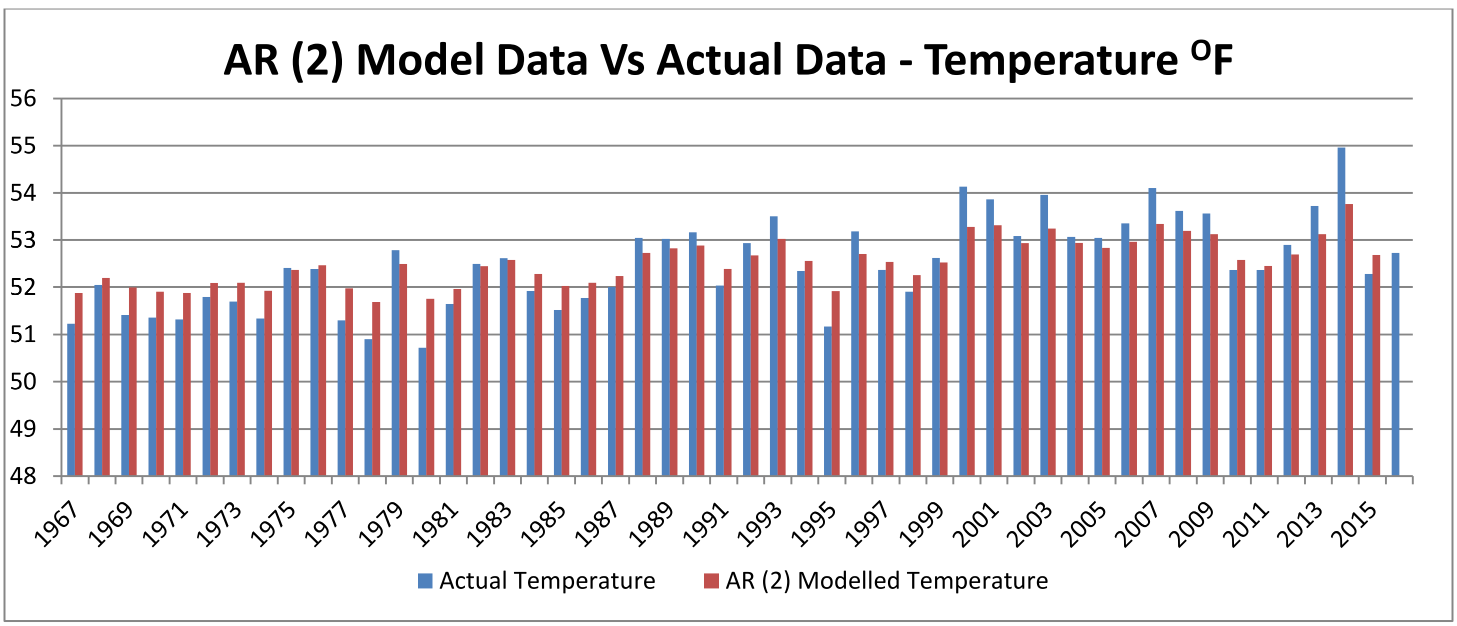


6. Second Order Autoregressive Model

The method of moments is then used to estimate both ϕ_1 and ϕ_2, also called Yule-Walker estimates. The resulting model is as follows:

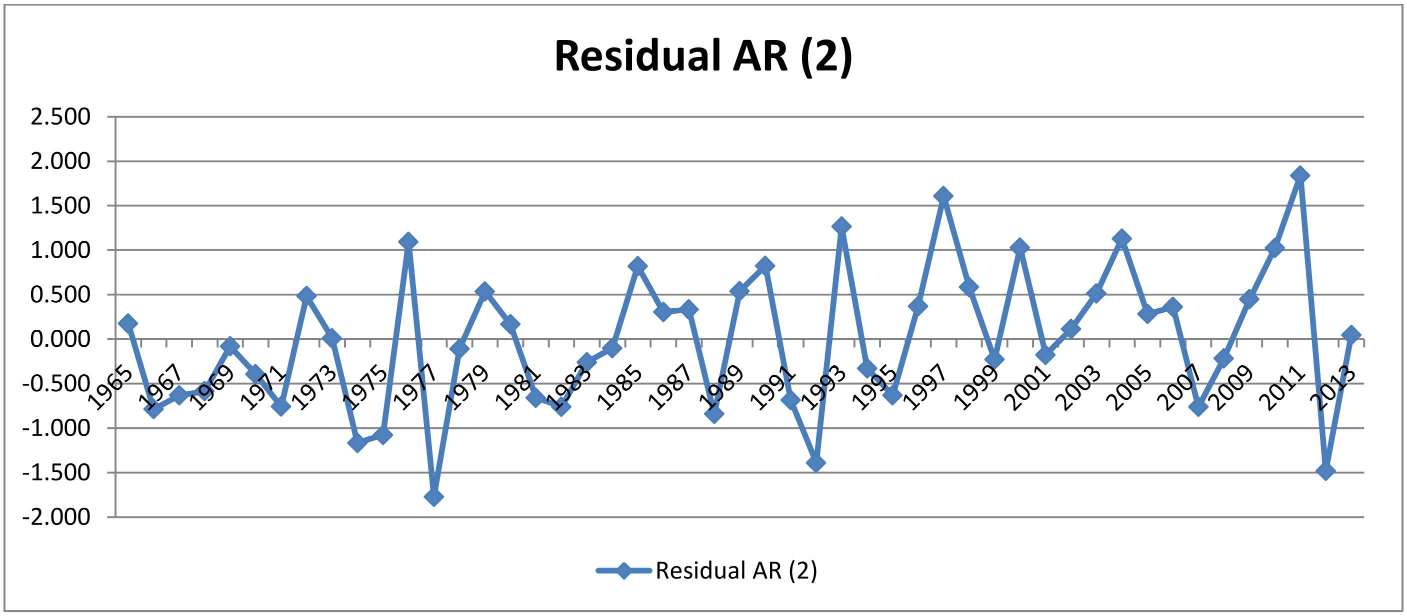
$Y_t = 23.502 + 0.450Y_{t-1} + 0.102Y_{t-2} + e_t$

Below is a graphic display of the actual data alongside the estimated data using the AR (2) model.



Again, the model seems to match the actual data quite well. This model however seems to be less approproate than the first-order model, as much of the model data seem to be slightly less accurate.

The residual analysis reaffirms this idea. The addition of the second parameter adds nothing to the model accuracy, and in fact, could decrease it.



Conclusion

Based on the analysis above, the most appropriate model for the average annual temperature in the United States over the last 50 years is the first-order autoregressive model:

$$Y_t = 26.048 + 0.504Y_{t-1} + e_t$$

This first order auto regressive model was able to predict average annual temperatures over the last 50 years with a great deal of accuracy. The second order model was not up to the mark accuracy, and may have decreased it. Therefore in my opinion AR (1) is the most appropriate model