# IDS 476: BUSINESS FORECASTING

**GLOBAL TEMPERATURE FORECASTING**

Predicting Land Temperature based on historical data

**Team Members:**

**Komal Tejwani (UIN:** **663286022)**

**Manasa Koka (UIN: 656827978)**

**Mihir Nagwekar (UIN: 662419341)**

**Pronoy RK(UIN: 672528906 )**

**Srujan Kadanur (UIN: 659632968)**

# 1. DATASET

## Obtaining Data:

This data is obtained from <http://berkeleyearth.org/data/>

## Data Collection/ Information about the dataset:

The Berkeley Earth Surface Temperature Study has combined 1.6 billion temperature reports from 16 pre-existing archives. The data is nicely packaged and allows for slicing into interesting subsets (for example by country). The data consist of average land temperatures for 1850 – 2015.

## Characteristics of Dataset

The data is a data frame containing 3,193 observations of 9 variables.

1. **Description of Attributes:**

|  |  |  |
| --- | --- | --- |
| **Attributes** | **Data type** | **Description** |
| Dt | DateTime | * Starts in 1750 for average land temperature and 1850 for max and min land temperatures and global ocean and land temperatures. |
| LandAverageTemperature | Numeric | Global average land temperature in Celsius |
| LandAverageTemperatureUncertainty | Numeric | The 95% confidence interval around the average |
| LandMaxTemperature | String | Global avg maximum land temperature in Celsius |
| LandMaxTemperatureUncertainty | String | the 95% confidence interval around the maximum land temperature. |
| LandMinTemperature | String | Global avg. minimum land temperature in Celsius. |
| LandMinTemperatureUncertainty | String | The 95% confidence interval around the minimum land temperature |
| LandAndOceanAverageTemperature | String | Global average land and ocean temperature in Celsius |
| LandAndOceanAverageTemperatureUncertainty: | String | The 95% confidence interval around the global average land and ocean temperature |

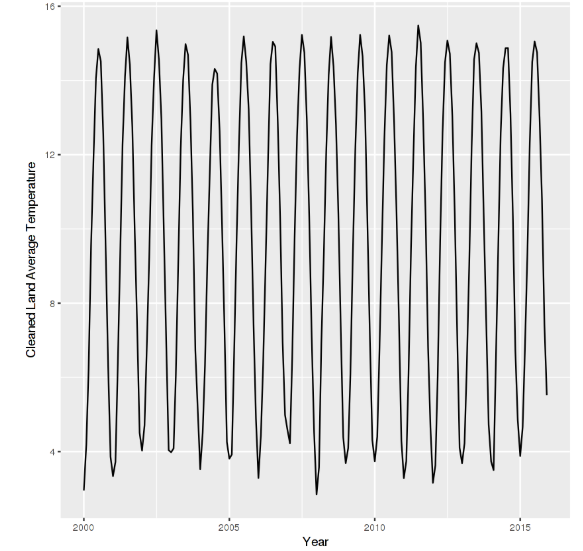
# 2. OBJECTIVE

To use ARIMA to forecast Land Average Temperature using data from the year 2000 – 2015.

# 3. STATISTICAL ANALYSIS

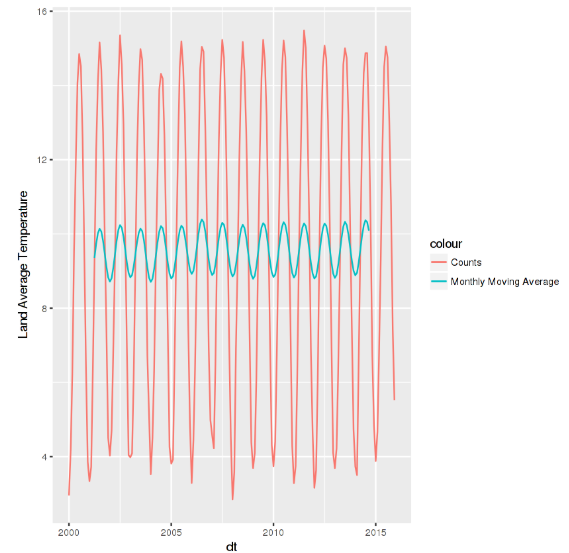
1. DATA CLEANING:

We observed that the data had various outliers and it we needed to remove those oultiers. sclean() was used to identify and replace outliers using series smoothing and decomposition.



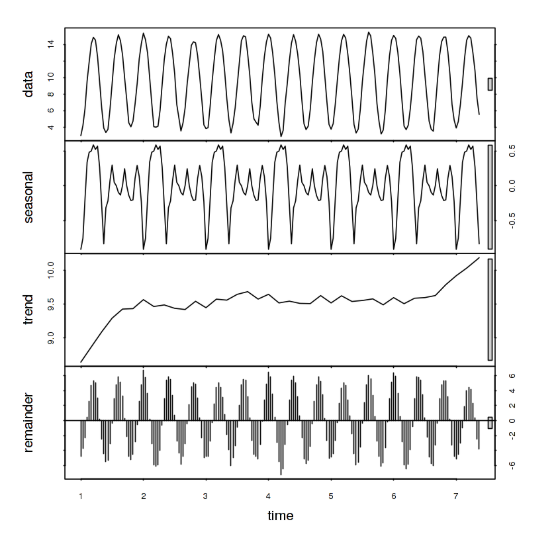
1. **DATA SMOOTHING FOR MA:**

Once the data is cleaned we will try Smoothing the data with Moving Average(MA). Here ma(q) refers to the error lags and combinations. In addition a monthly moving average is also calculated and plotted with the original data.



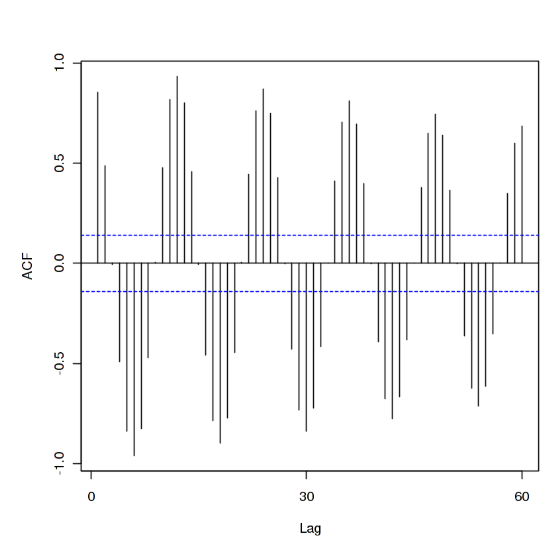
1. **DECOMPOSING THE DATA:**

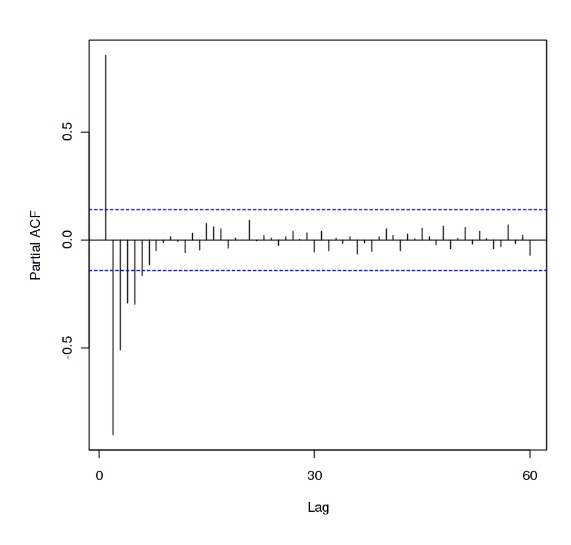
Then will Decompose the data which will break the seasonality, trend or cycle to make the forecasting accurate.



1. **STATIONARY:**

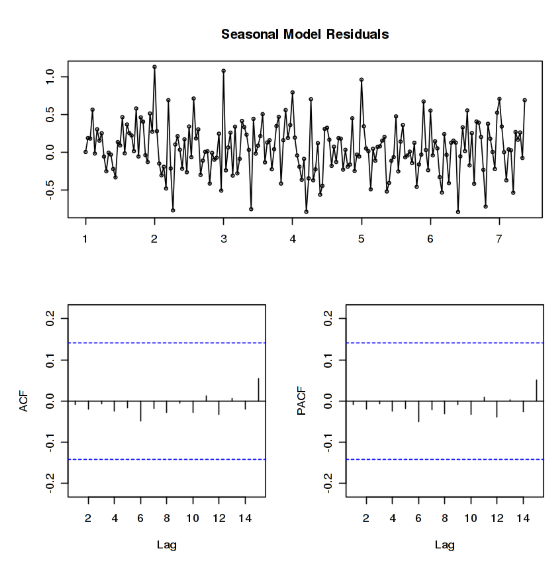
* ARIMA model requires time series data to be stationary. So, we will make the mean, variance and autocovariance of the data as time variant.
* We will use ACF and PACF to check if the data is stationary and this will also help us select the order of parameters for ARIMA model.
* ACF are a useful visual tool in determining whether a series is stationary. These plots can also help to choose the order parameters for ARIMA model. They display correlation between a series and its lags.
* PACF display correlation between a variable and its lags that is not explained by previous lags. PACF plots are useful when determining the order of the AR(p) model.





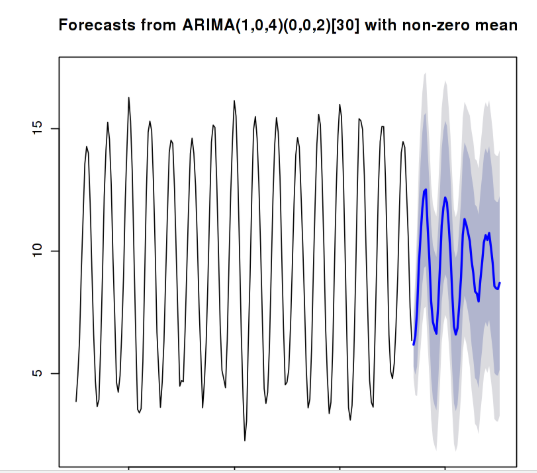
1. **SEASONALITY:**

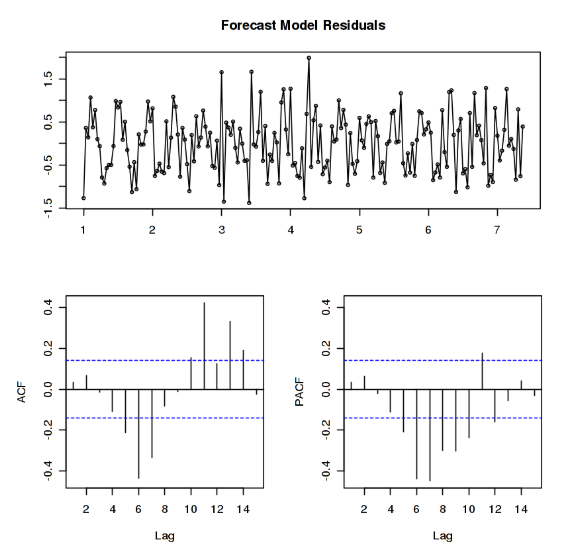
* Checking for any seasonality component still with the data.

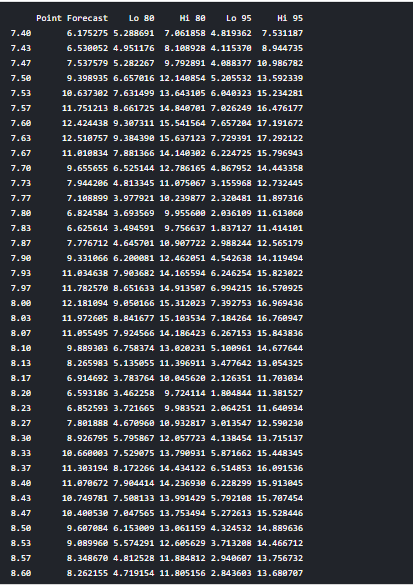


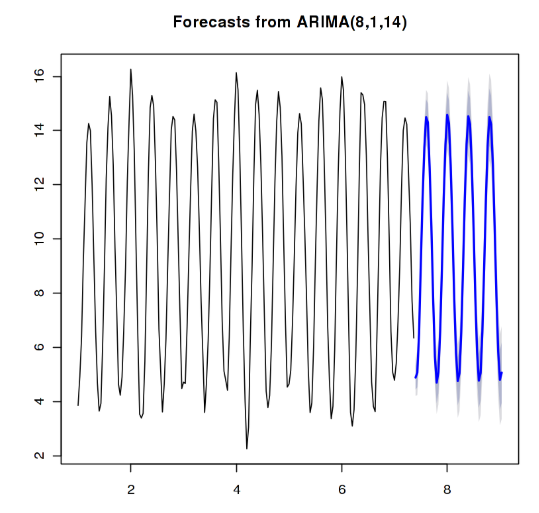
1. **FITTING MODEL:**

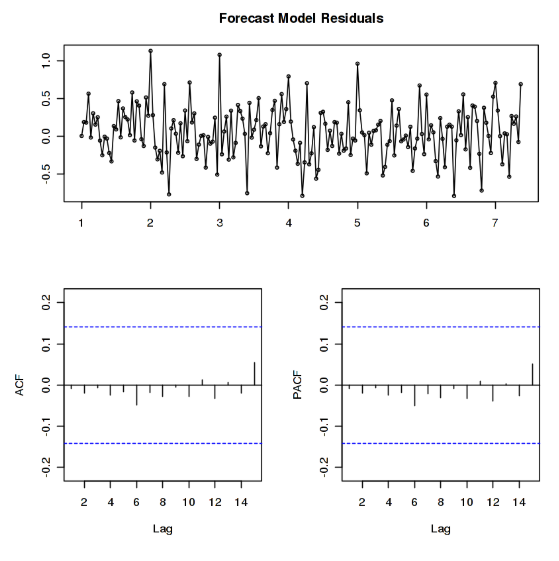
* Fitting the ARIMA model to forecast the global average temperature.
* Forecasting using a fitted model is straightforward in R. We can specify forecast horizon h periods ahead for predictions to be made and use the fitted model to generate those predictions.
* We tried different parameters for forecasting to get an accurate sense of the accuracy.
* Below is the output we got for each parameters used.



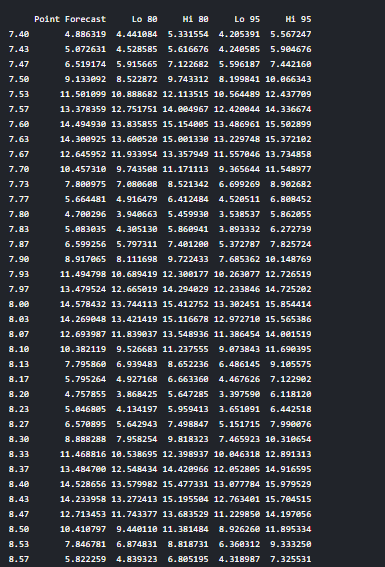




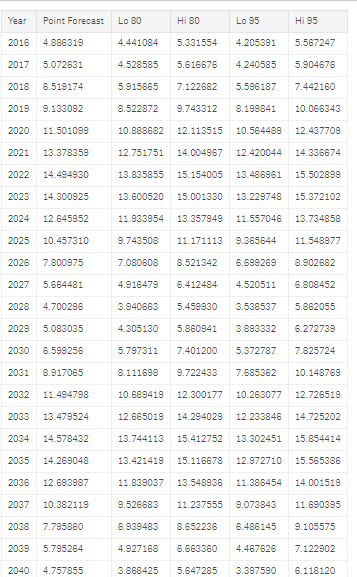




This shows very good results and the temperature forecasts also follow the actual pattern of history. The blue line is the temperature forecast of Land Average Temperature for next 50 years starting from 2015.



The output did not show Year values, so I added them. Here is what the final forecast average land temperature values of next 40 years look like.



# 4. CONCLUSION:

From the study it’s clear that there seems to be a seasonality and trend with temperature data. When we look at the graph, it suggests the Land Average temperature will trace the temperature levels as of past years. We were able to predict actual results, where the temperature trend in fact is increasing as years go by.