TIME SERIES ANALYSIS

Author: Bharat Thakur

Background

Time series is a sequence of discrete-time data. Time series is useful to analyze how a given variable changes over time. Now, we will perform time series analysis of temperature and precipitation data of two cities.

Data Source

There are two datasets used in this analysis:

1.) Average temperature by month of Welland from January 1985 to December 1994.
2.) Total precipitation in Waterloo for the years 1970 to 1995.

This report is divided into two main sections, section 1 and section 2. In section 1 we analyze Welland's temperature data and in section 2 we analyze Waterloo's rainfall data.

Section 1: Welland Temperature

1. Data Transformation and Cleaning (Description)

Dates

The dates column was removed since we only need temperature data to perform time series analysis.

Temperature

The monthly temperature data was converted to time series data using function ts().

```
           Jan       Feb       Mar       Apr       May       Jun
1985 -6.616129 -4.328571  1.367742  8.926667 14.151613 16.450000
```

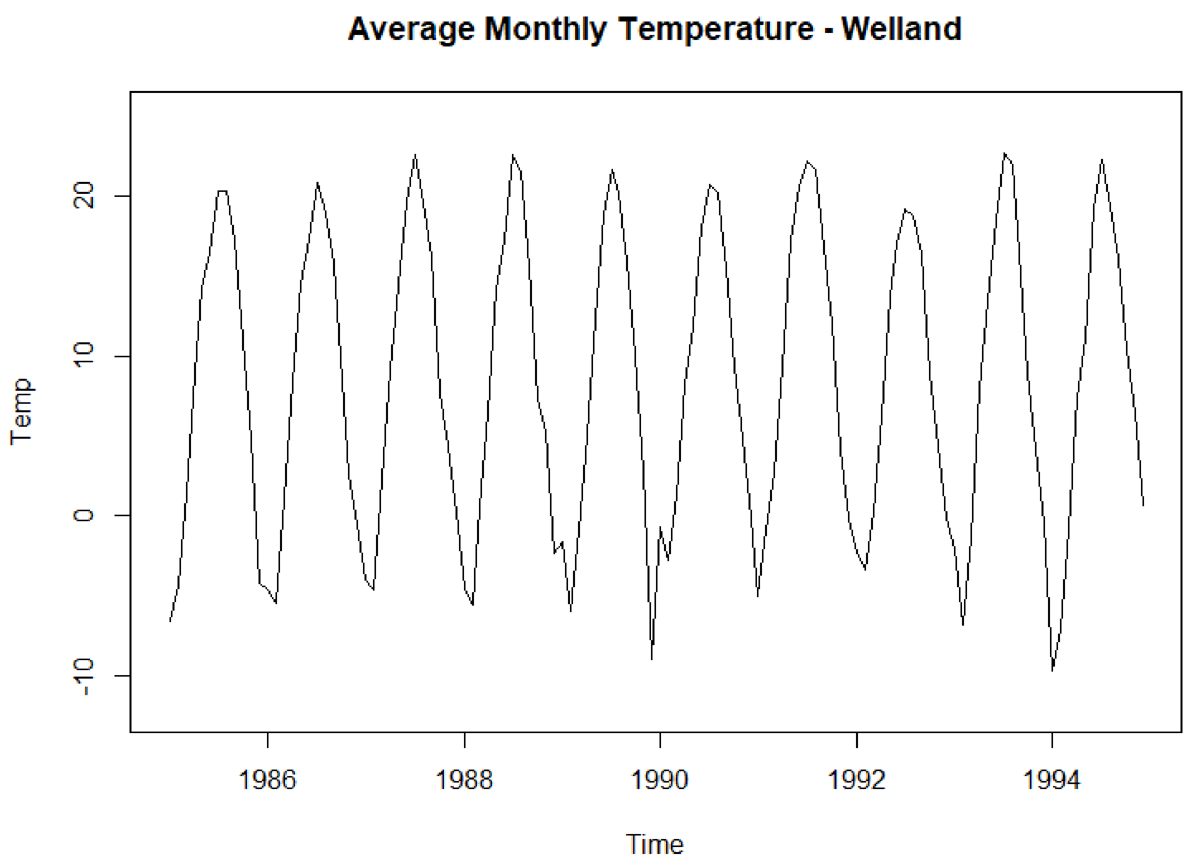
2. Descriptive Data Analysis

2.1 Summary of the temperature information:

```
                     Temp
nbr.val       120.0000000
nbr.null        0.0000000
nbr.na          0.0000000
min            -9.7129032
max            22.6387097
range          32.3516129
sum          1015.5660173
```

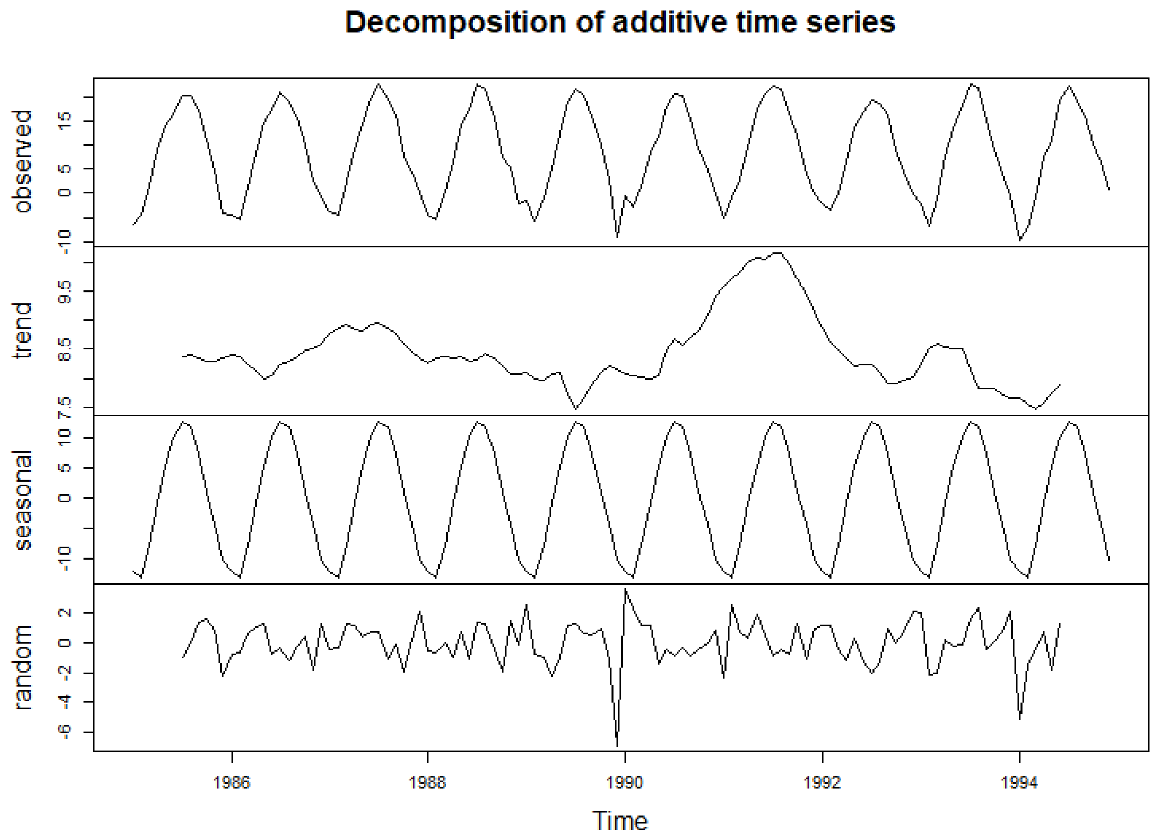
```
median           8.9326882
mean             8.4630501
SE.mean          0.8500028
CI.mean.0.95     1.6830905
var             86.7005765
std.dev          9.3113144
coef.var         1.1002315
```

2.2 Plot of time series :



Temperature data shows strong seasonal tendencies although the mean temperature appears to be relatively constant.

2.3 Decomposition of the times series data into the constituent components:



As we noted earlier, the trend is relatively constant except for a peak around 1992.

2.4 Stationarity:

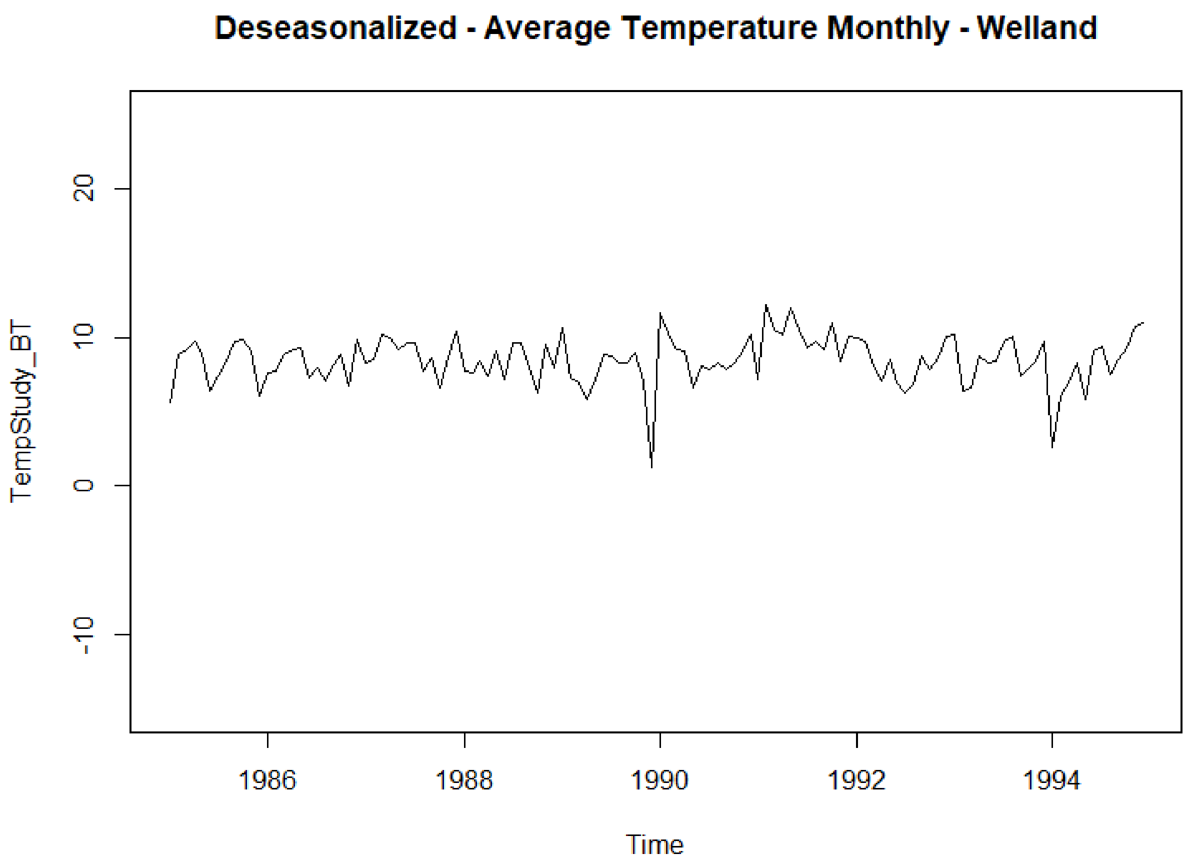
```
Warning in adf.test(TempStudy_BT): p-value smaller than printed p-value


 Augmented Dickey-Fuller Test

data:  TempStudy_BT
Dickey-Fuller = -13.342, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

Augmented Dickey-Fuller test resulted in a p-value less than .05 indicating that the time series is stationary.

2.5 Deseasonalized Temperature Data :



2.6 Comments :

1. Overall trend component is not varying much, except for a peak around 1992.
2. As one would expect seasonal component has a strong cyclic nature.
3. The random component seems to have constant variance.

Section 2: Waterloo Precipitation

1. Data Transformation and Cleaning (Description)

Dates

The dates column was removed since we only need precipitation data to perform time series analysis.

Precipitation

The total precipitation data was converted to time series data using function ts().

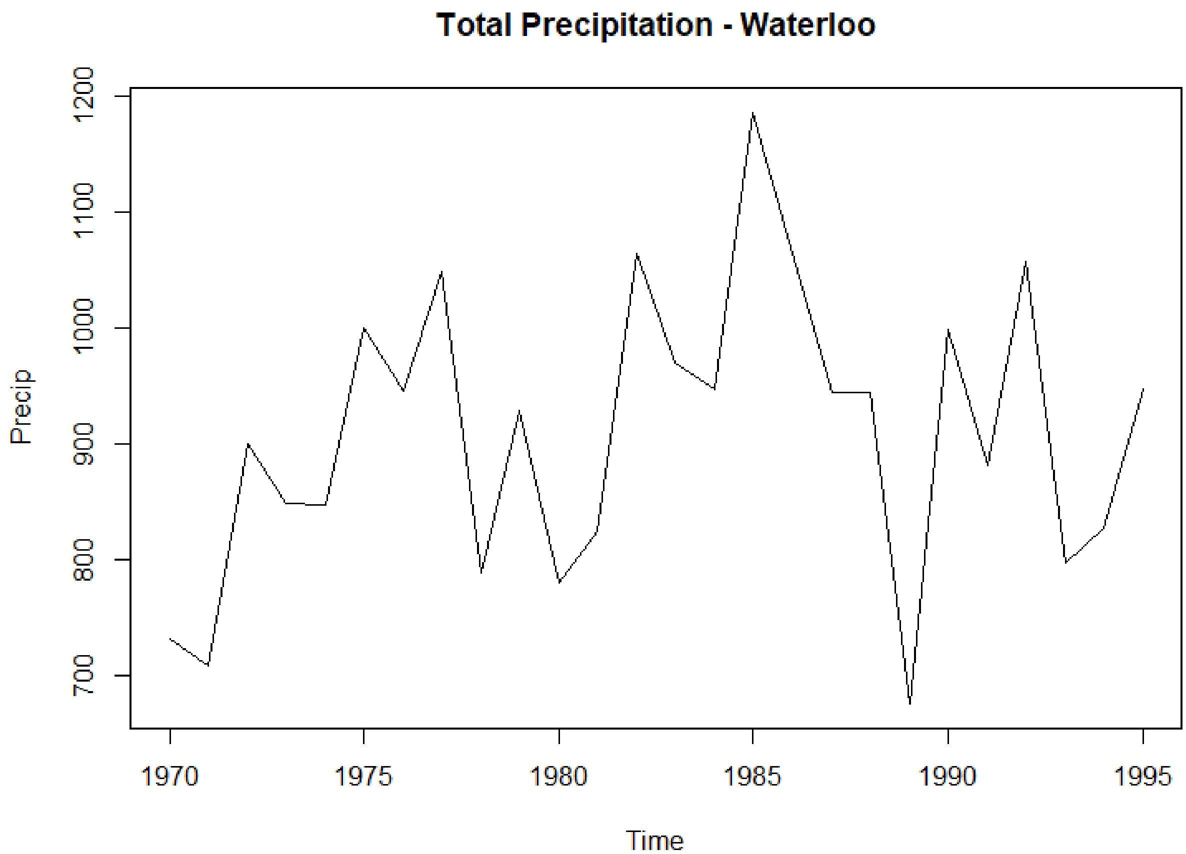
```
Time Series:
Start = 1970
End = 1975
Frequency = 1
     Precip
[1,]  731.3
[2,]  707.7
[3,]  899.5
[4,]  847.9
[5,]  846.4
[6,]  999.8
```

2. Descriptive Data Analysis

2.1 Summary of the precipitation information:

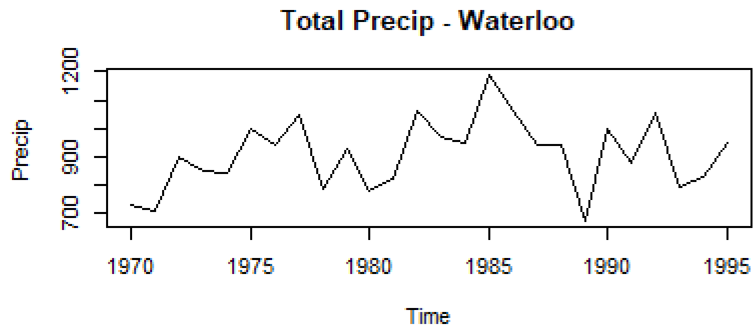
```
                  Precip
nbr.val         26.00000
nbr.null         0.00000
nbr.na           0.00000
min            674.80000
max           1186.40000
range          511.60000
sum          23648.20000
median         936.10000
mean           909.54615
SE.mean         24.39122
CI.mean.0.95    50.23467
var          15468.22738
std.dev        124.37133
coef.var         0.13674
```

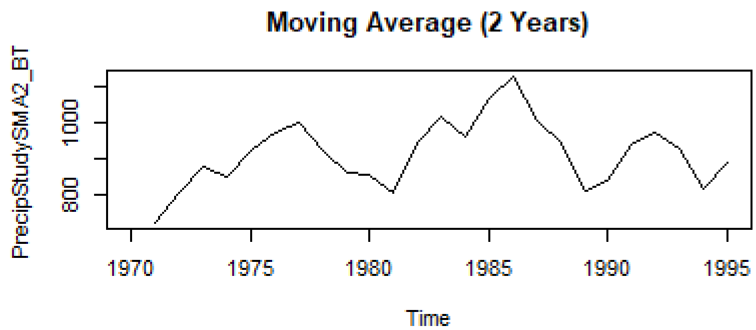
2.2 Plot of time series :

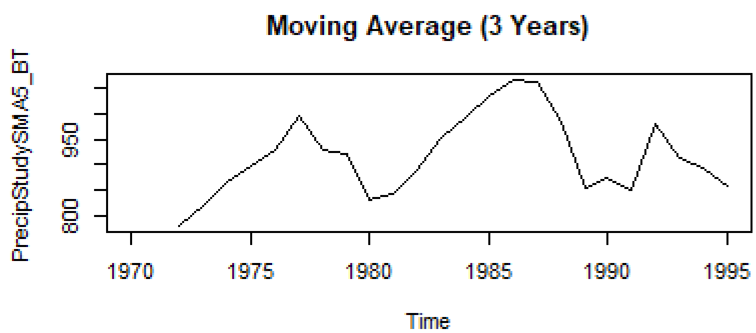


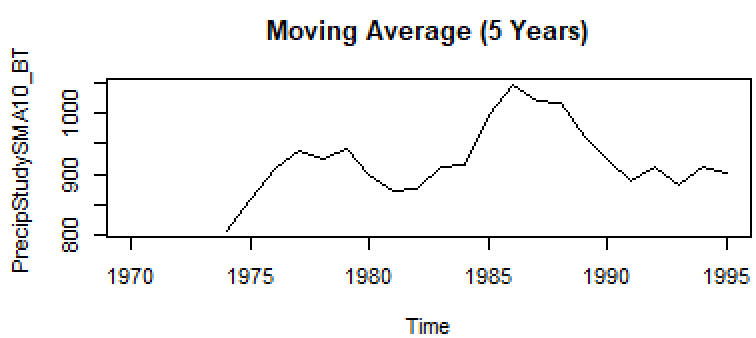
Precipitation significantly varies year-to-year with maxima around 1985 and minima around 1989.

2.3 Decomposition of the times series data into the constituent components:









In my opinion, moving average using 2 years window best conveys the trend since the window size of 3 years and 5 years appears to over smoothen the data (leading to loss of details).

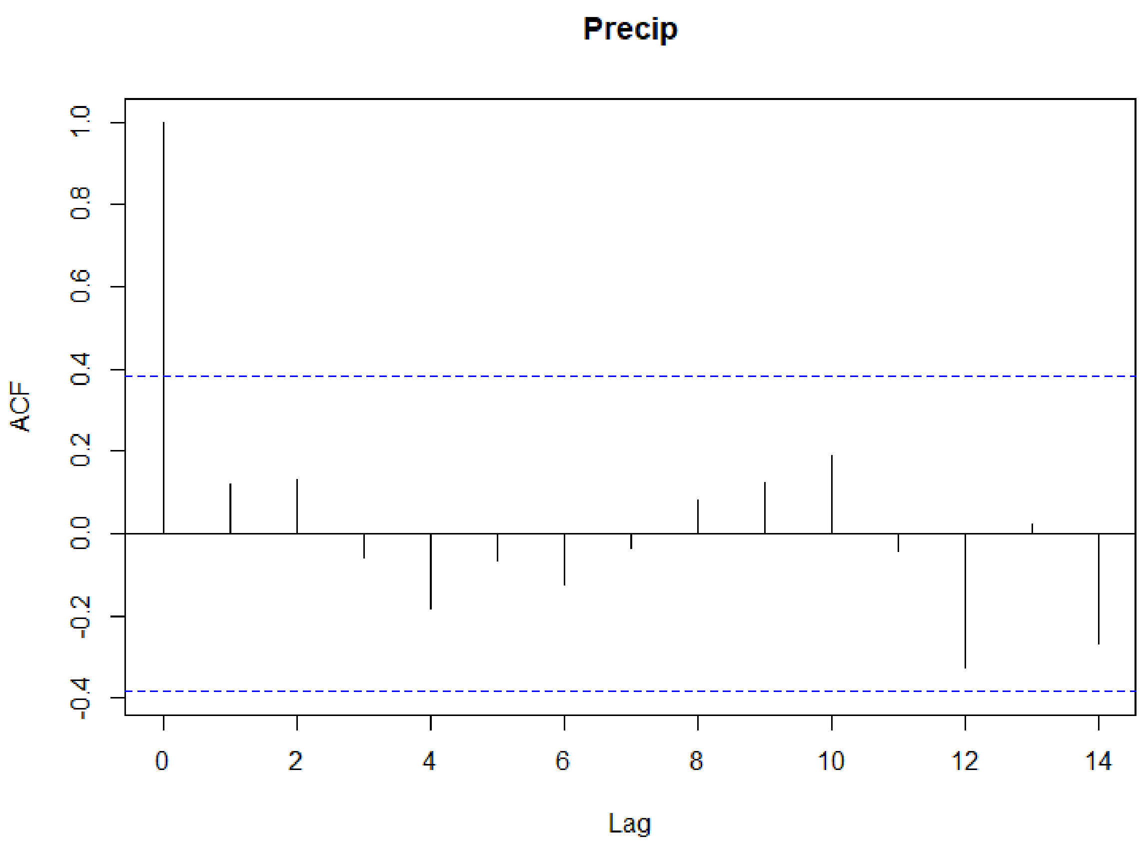
2.4 Stationarity:

```
 Augmented Dickey-Fuller Test

data:  PrecipStudy_BT
Dickey-Fuller = -2.5257, Lag order = 2, p-value = 0.3721
alternative hypothesis: stationary
```

Augmented Dickey-Fuller test resulted in a p-value greater than .05 indicating that time series is non-stationary.

2.5 Autocorrelation chart:

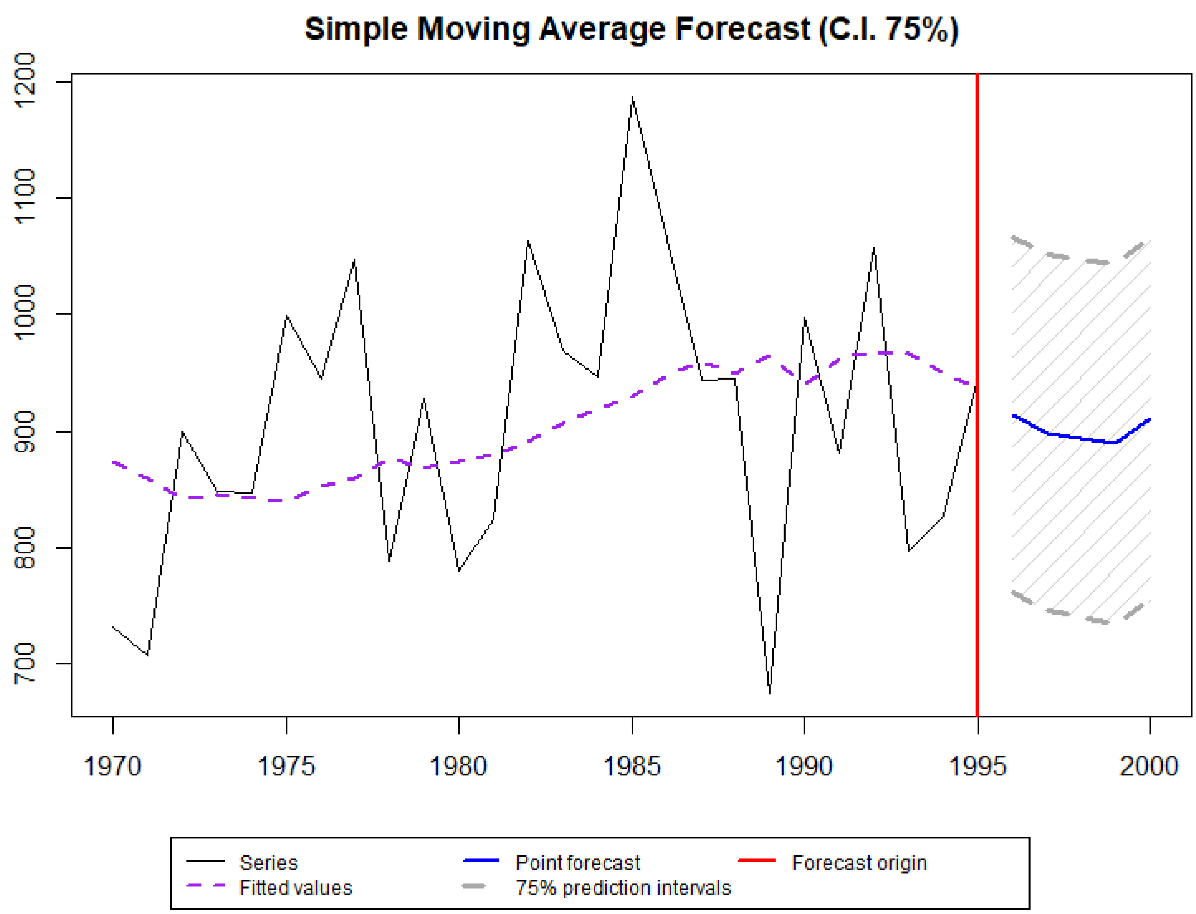


All the lags, except lag 0 (i.e., the correlation of a time series with itself), are not significant as they fall within horizontal blue lines which are the approximate 95% confidence intervals. This means that previous values do not seem to be influencing the current values.

3. Forecast

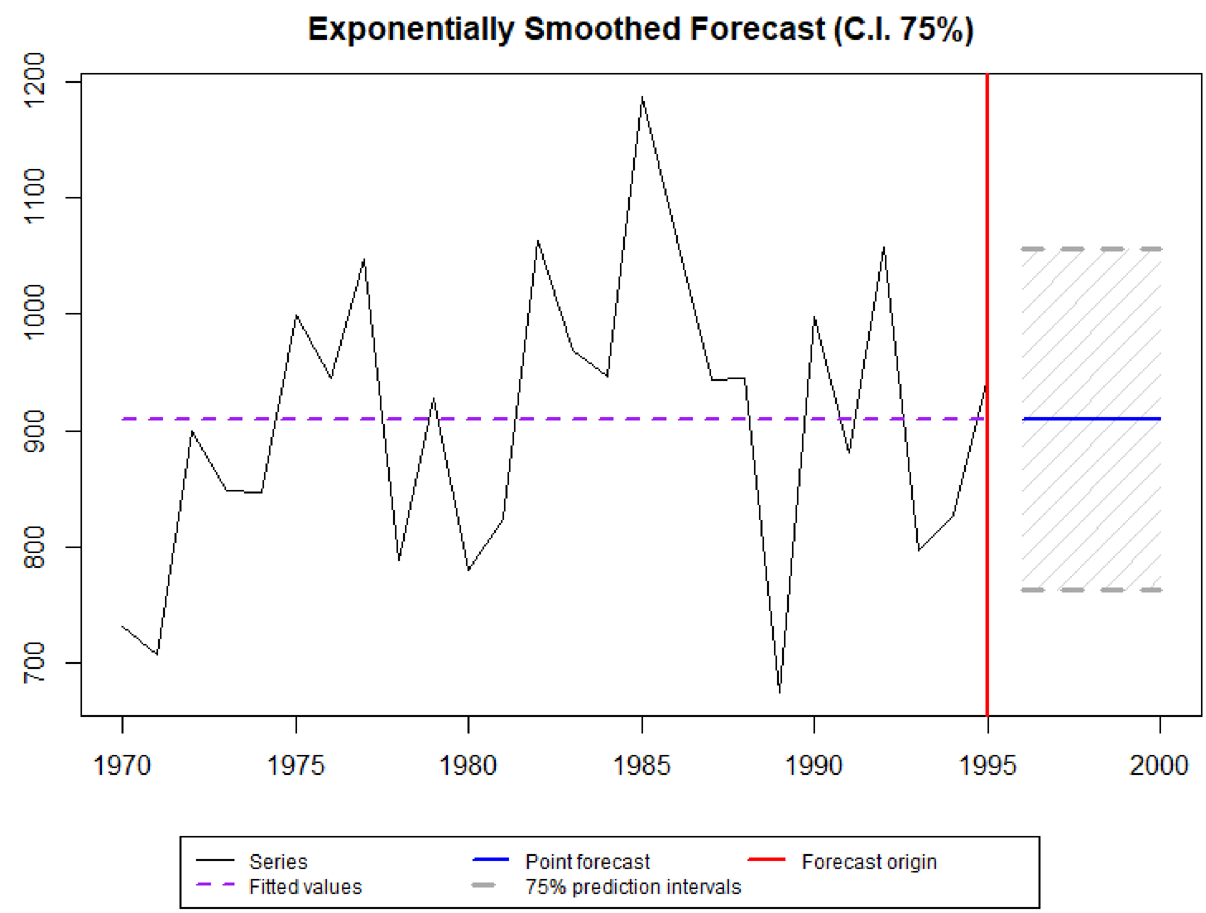
3.1 Simple Moving Average Forecast

```
Time Series:
Start = 1996
End = 2000
Frequency = 1
     Point forecast Lower bound (12.5%) Upper bound (87.5%)
1996       913.6000            761.3417            1065.858
1997       898.5000            745.4823            1051.518
1998       893.9700            740.0385            1047.902
1999       888.8970            733.8669            1043.927
2000       910.3067            753.9577            1066.656
```



3.2 Exponentially Smoothed Forecast

```
Time Series:
Start = 1996
End = 2000
Frequency = 1
     Point forecast Lower bound (12.5%) Upper bound (87.5%)
1996       909.5462            763.072             1056.02
1997       909.5462            763.072             1056.02
1998       909.5462            763.072             1056.02
1999       909.5462            763.072             1056.02
2000       909.5462            763.072             1056.02
```



3.3 Conclusion

In general, recent years are better for predicting future years. So, in this case, exponential smoothing forecast seems to stand better for forecasting temperature in Waterloo. It appears that the overall trend is relatively constant. It is not going upward or downward. In the case of moving average, we can observe that the trend appears to go first downwards and then upwards, for which we have no reasoning.