A Project Report

ON

ANALYSIS OF INDIAN RAINFALL

BY

DHRUV GOYAL (1PI10IS030)
DHRUPAD C KANERIA (1PI10CS037)
SUMEDHA GOVINDARAYA (1PI10CS104)

Guide
DR. KAVI MAHESH
Director, KAnOE
PESIT-CSE
Bangalore

August 2013 – Dec 2013

CENTER FOR KNOWLEDGE ANALYTICS AND ONTOLOGICAL ENGINEERING

PES INSTITUTE OF TECHNOLOGY
(An autonomous institute under VTU)
100 FEET RING ROAD, BANASHANKARI III STATE
BANGALORE-560085

CONTENTS

ABSTRACT

Average monthly rainfall data was obtained from Indian Institute of Tropical Meteorology (IITM) for various regions of India. This data was available for 141 years from 1871 to 2011. Out of this data, only the total rainfall during the Southwest Monsoon period of June to September was considered for each region. The data was treated as a time series from Year 1 (= 1871) to Year 141 (= 2011). No trend was observed in the data. The hypothesis was that rainfall exhibits a strong periodic pattern. A periodogram was constructed by transforming to the frequency domain. Periods of 10 to 70 years were considered (i.e., up to half of the total length of time for which data was available). The results show a very strong period of 60 years. This is especially significant since a period of 60 years has astronomical significance: it is the least common multiple of the orbital periods of Mars, Jupiter and Saturn (in addition to Mercury and Venus). It may also be noted that the traditional Indian calendar (known as *panchanga*) has always used a cycle of 60 years (called *samvatsara*s) to name the years and to predict rainfall. The result also suggests a strong period of 40 years. A confidence level test was conducted for the obtained periodogram for a confidence test of 95 %. ACF plots were also made to verify result.

1. INTRODUCTION

Rainfall is one of the most important elements in affecting the climate, economy and the lives of the people of the country. Being an agricultural country, good rainfall is the key to bountiful agriculture and food for the Indians. The success or failure of the harvest and water scarcity in any year is always considered with the greatest concern. These problems are closely relinked with the behavior of the summer monsoon rains in India (Rajeevan, 2001).

The term monsoon seems to have been derived either from the Arabic 'mausim' or from the Malayan 'monsin'. As first used, it was applied to southern asia and the adjacent waters, where it referred to the seasonal surface air streams which reverse their direction between winter and summer, southwest in summer and northeast in winter in this area.

2. PROBLEM DEFINITION

We are trying to analyze the Indian Rainfall data for the past 141 years (1871 – 2011) and researching for the possibility of any trend or cycle in the data if exists.

3. SYSTEM REQUIREMENT SPECIFICATION

3.1 HARDWARE REQUIREMENT

- NA

3.2 SOFTWARE REQUIREMENT

3.2.1 Operating System

- Microsoft Windows XP or newer version
- Any Linux flavored Operating System such as Ubuntu/Fedora/MAC OS

3.2.2 Tools

- Weka
- Microsoft Office
- Google Fusion Table

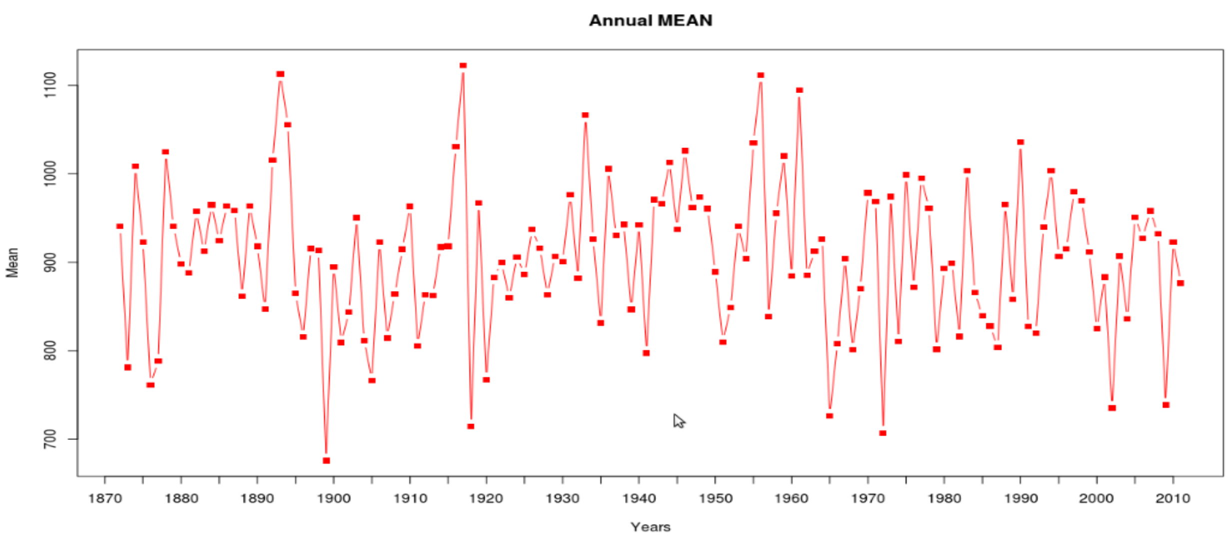
3.2.3 Programming

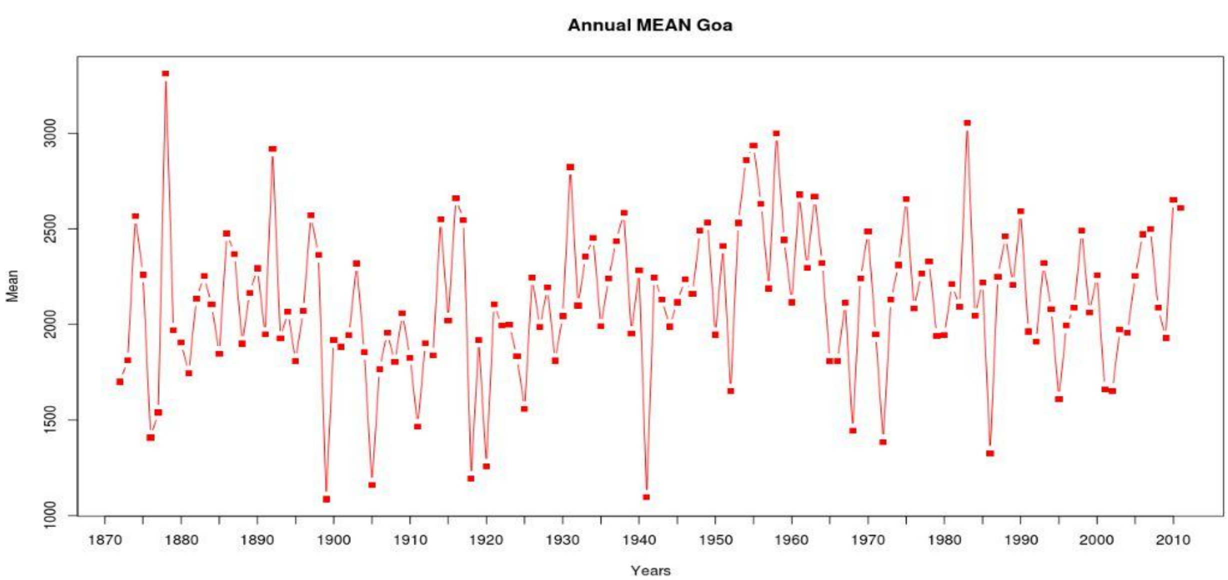
- R
- Weka API
- Google API

4. METHODOLOGY

4.1 Primary Approach

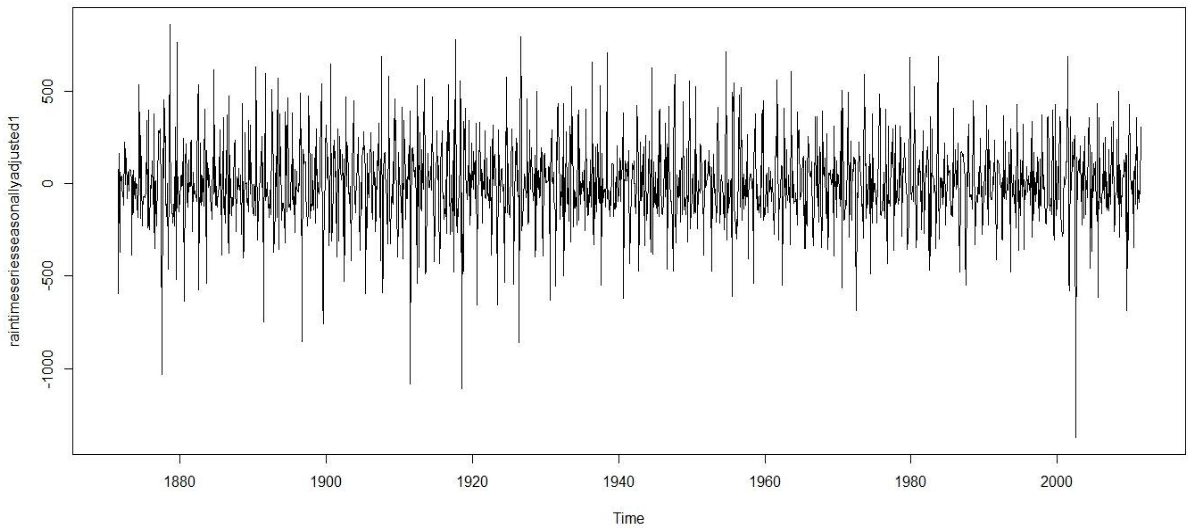
The Indian rainfall series data from 1871 to 2011 from the IITM website was used for the analysis purpose. Indian continent is divided into 36 meteorological subdivisions. Initially the mean values of annual rainfall using R was plotted for all sub divisions. The same was repeated for monsoon rainfall data also. But no notable trend was observed and the graph was random.

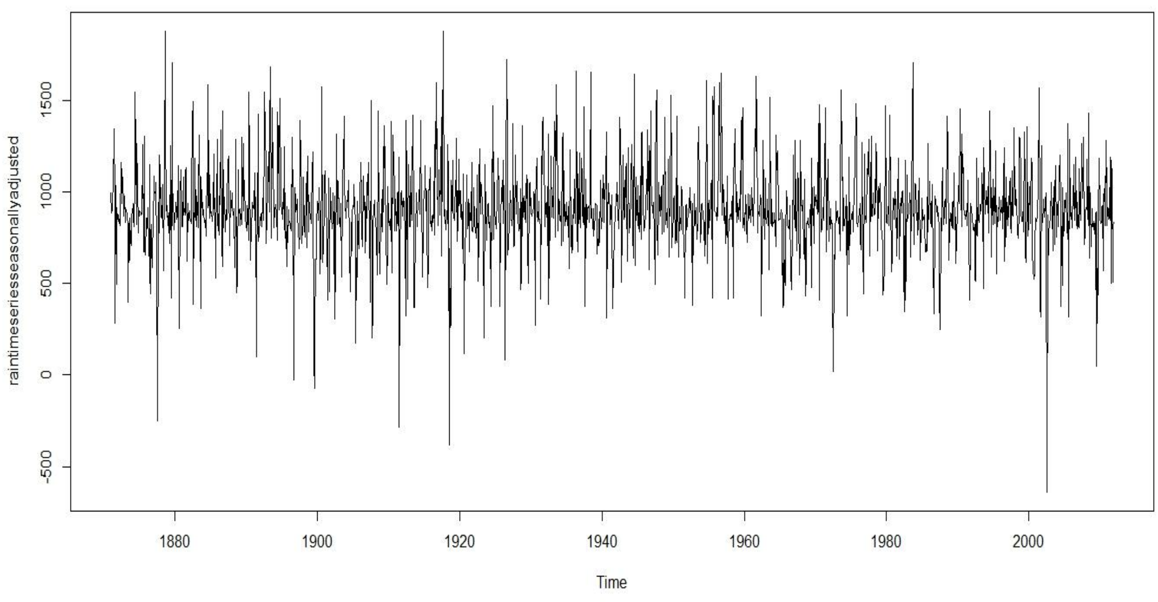




4.2 Time-series Analysis

Different components of the time-series like seasonal, trend and random were analyzed using R but no significant result was obtained.

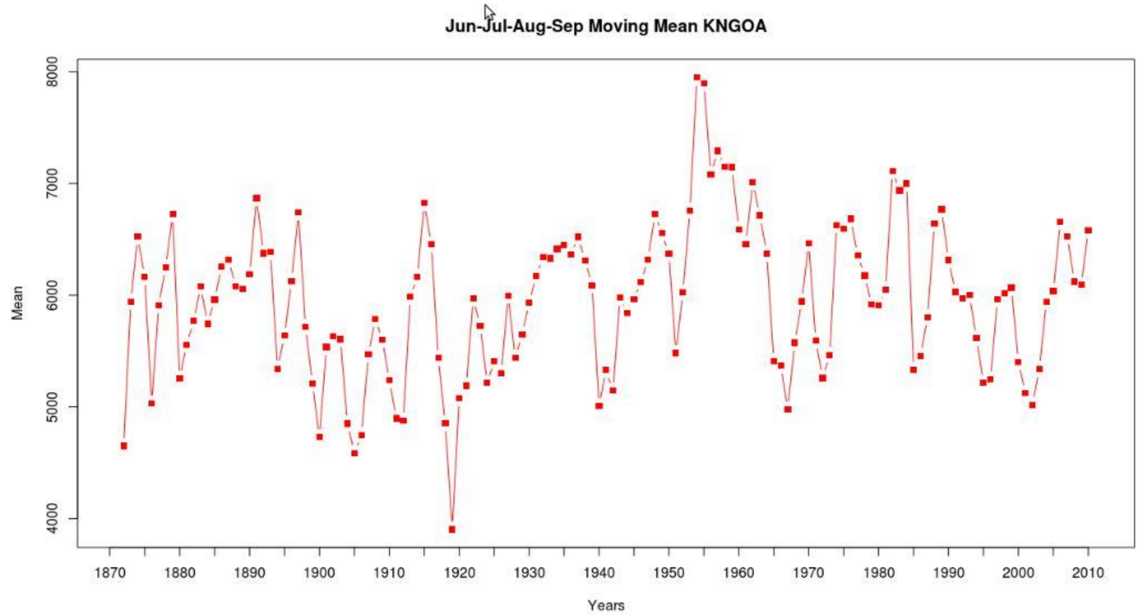





```
> demo <- read.csv("D:/SummerRainfallProject/Updated/demo.csv")
> year <- demo$Year
> mean_kngoa <- demo$Mean_kngoa
> plot.ts(mean_kngoa)
```

4.3 Moving Average (3 years)

The moving average of 3 years for all the years, of all the subdivisions was plotted. The plot obtained here was smoother and showed increasing and decreasing trends for random number of years but still no significant conclusions could be made from the moving averages plot.



4.4 Weka Tool

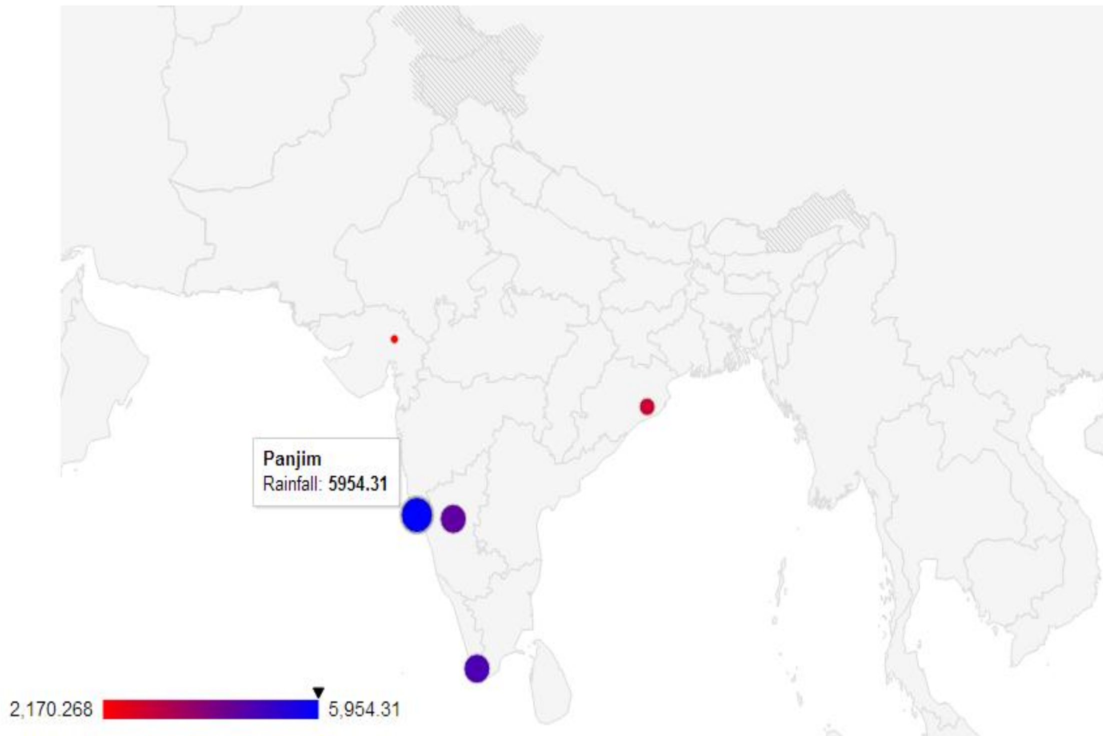
Weka tool was used to find any possible correlation between the annual rainfalls of all the sub divisions. We observed a high correlation in the sub divisions which are closely located. This shows that the rainfall of closely located geographical areas has higher correlation but this was not true in all cases and therefore was superficial for our research.

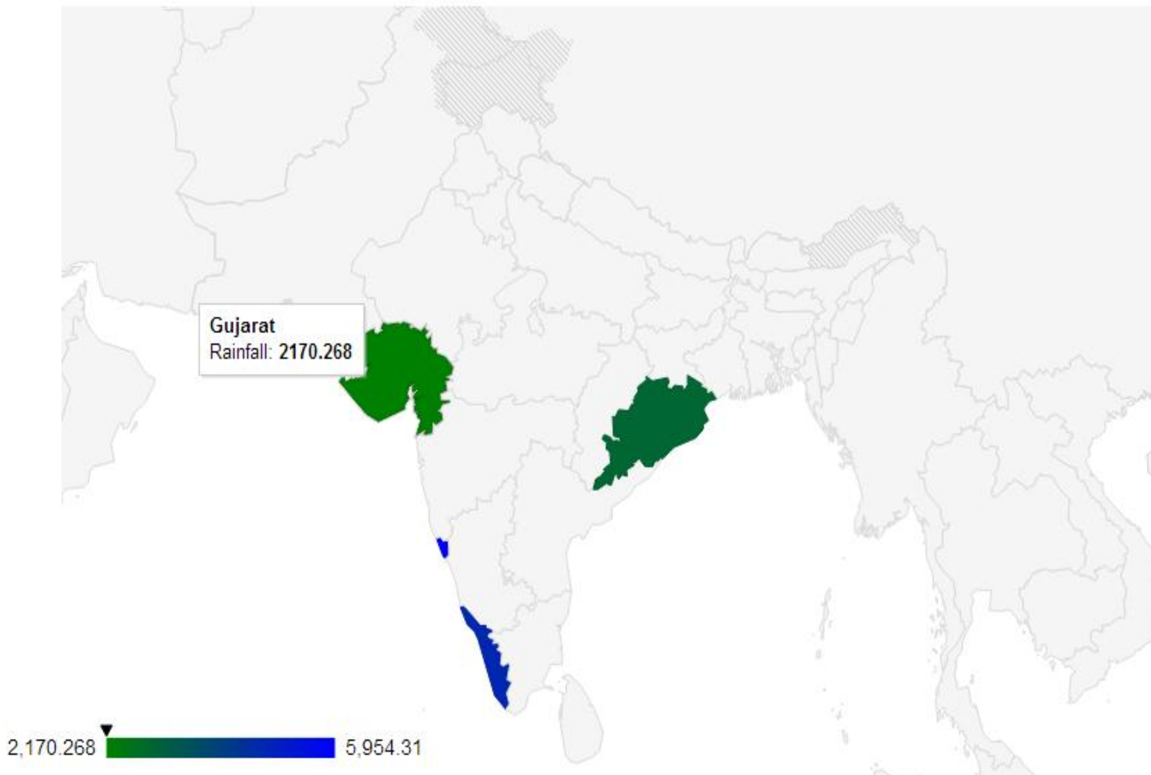
4.5 RMS (Root Mean Square)

To find the 60 years cycle relationship in the data,we tried using a formula for RMS for different sub divisions and obtained the range between 400 to 2100 for them, which was not consistent and hence was not sufficient to prove the hypothesis of 60 years cycle.

4.6 Visualization

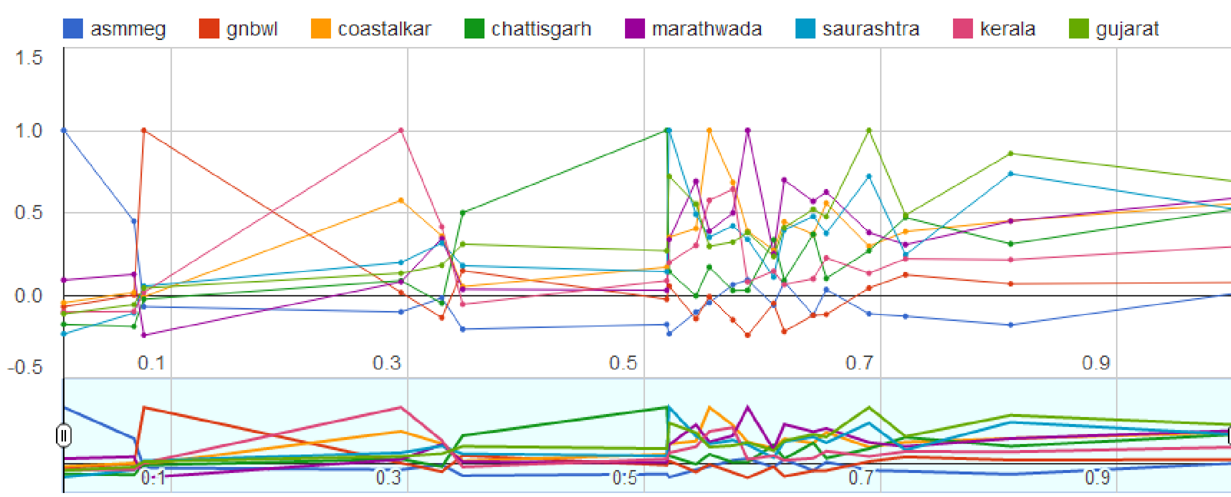
The data was visually represented. The annual rainfall of the sub divisions was graphically represented on the map with the help of geo-charts and geo-maps (Google Apps). (The regions were approximately plotted and were not up to scale.) As we needed internet connectivity to view these maps, we tried to write code which turned out to be cumbersome and useless as we still needed Google APIs to run them. So we moved to fusion tables provided by Google to represent and plot the required maps and charts.





4.7 Weka API

Weka APIs were used to calculate the correlation values for all the subdivisions and from that, the correlogram was made using fusion tables.



4.8 Periodogram Analysis

The highest common factor of the rotation periods of all the planets in the solar system is approximately 60 years. Considering this fact, we plotted a periodogram with periods as 10,20,30,40,50,60,70 years for both annual rainfall and monsoon rainfall of all sub divisions. The graph showed a high peak value for 40 and 60 years.

$$\frac{T}{2}a = \int_{t_1}^{t_1+T} f(t)\cos(kt)dt$$
$$\frac{T}{2}b = \int_{t_1}^{t_1+T} f(t)\sin(kt)dt$$

$$r = \sqrt{a^2 + b^2}$$

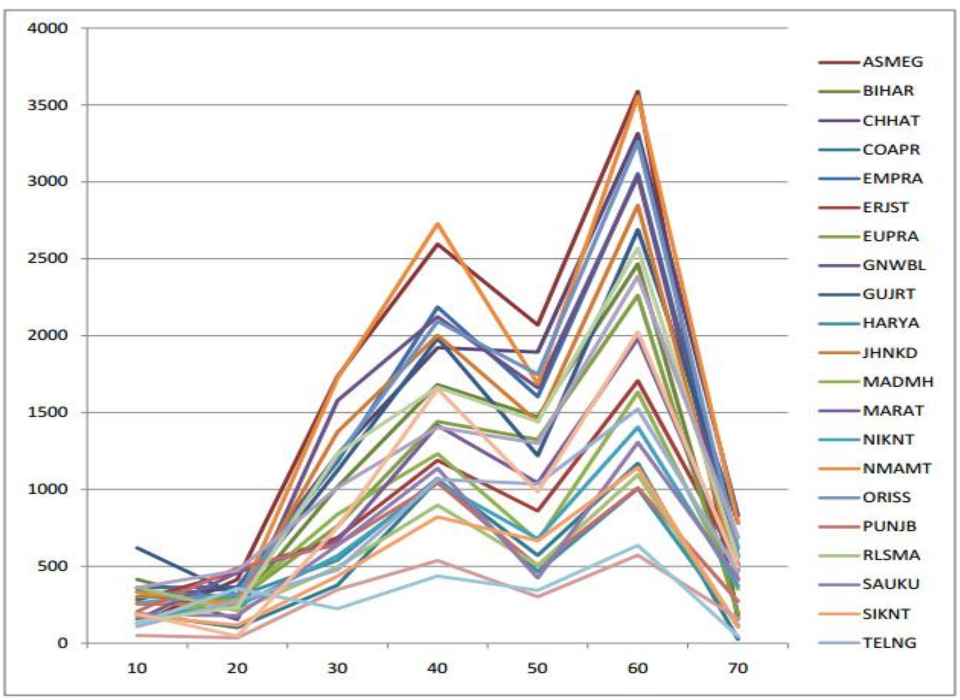
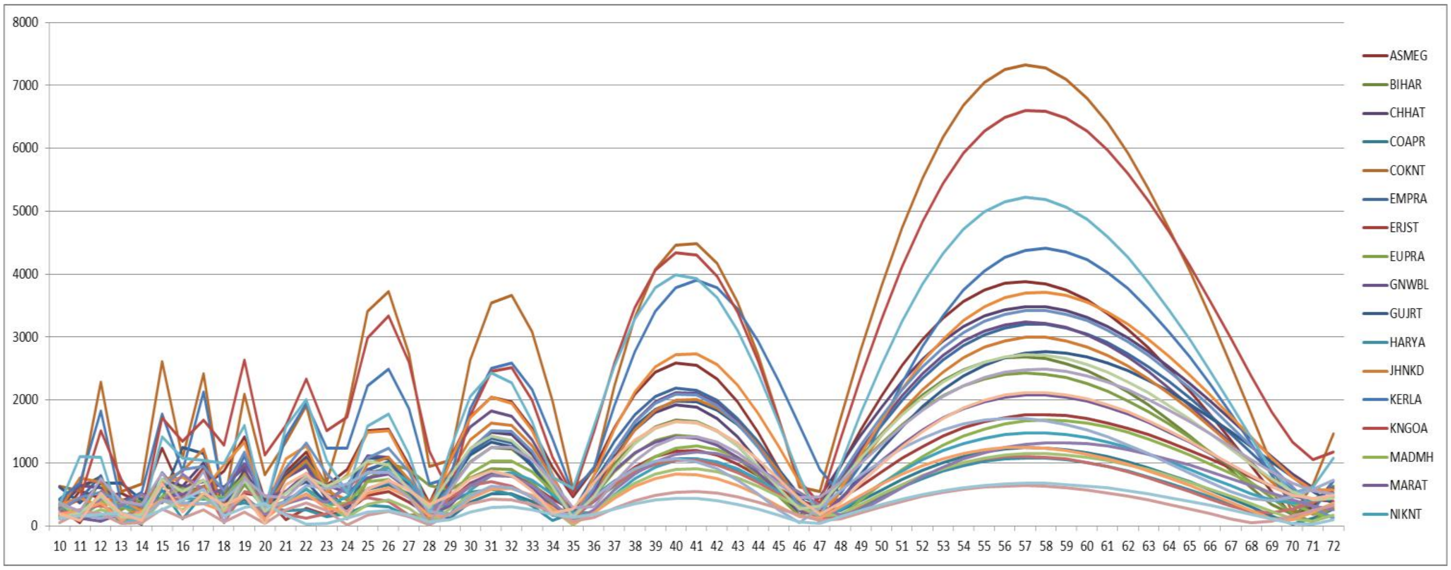
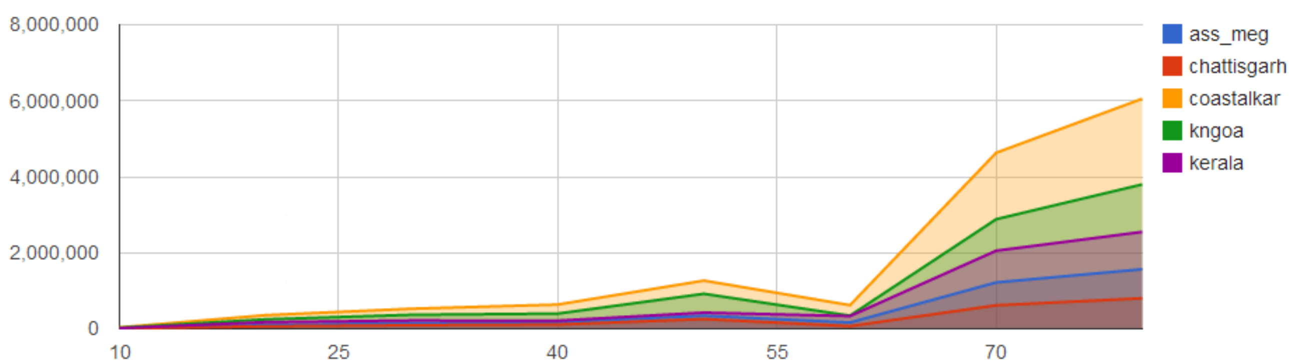


Fig. 1 Periodicity in Monthly Average Rainfall in Regions during Jun-Sep (1871-2011)



4.9 Moving Average (30 years)

The possible existence of a cycle of 30 year wet and 30 year dry period is already published in a paper. So, moving average for 30 years was obtained to see how it affects the 60 years cycle. We found that it diminished the 60 year cycle which it shouldn't have. Hence this approach dint lead to any outcome.



4.10 Confidence Level Test

The significance of 40 years cycle was tested using confidence level intervals for the results obtained from periodogram. A confidence level test for a confidence rate of 95% was performed. A confidence level of 95-96 % and 90-95% was obtained for the 60 year cycle and for the 40 year cycle respectively.

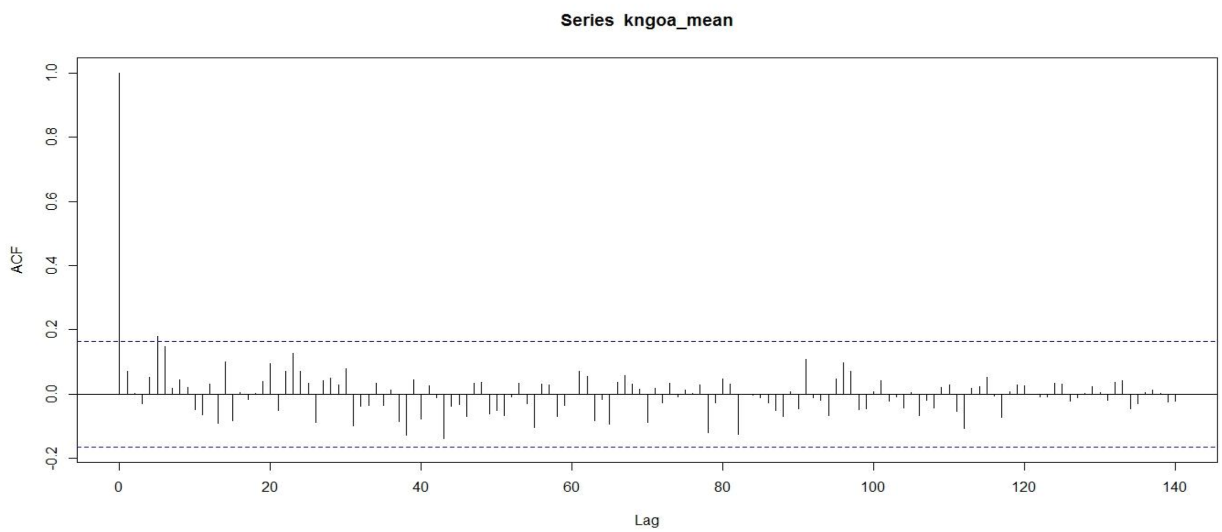
$$\text{If } n \geq 30, \text{ then}$$
$$\text{Confidence interval} = x \pm z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$$

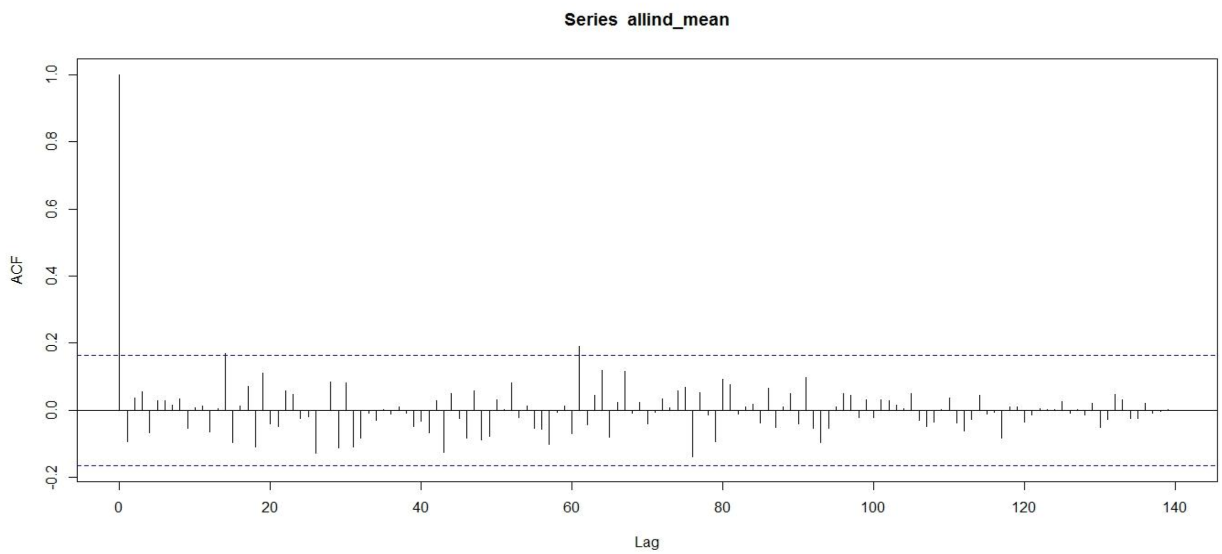
$$\text{If } n < 30, \text{ then}$$
$$\text{Confidence interval} = x \pm t_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right)$$

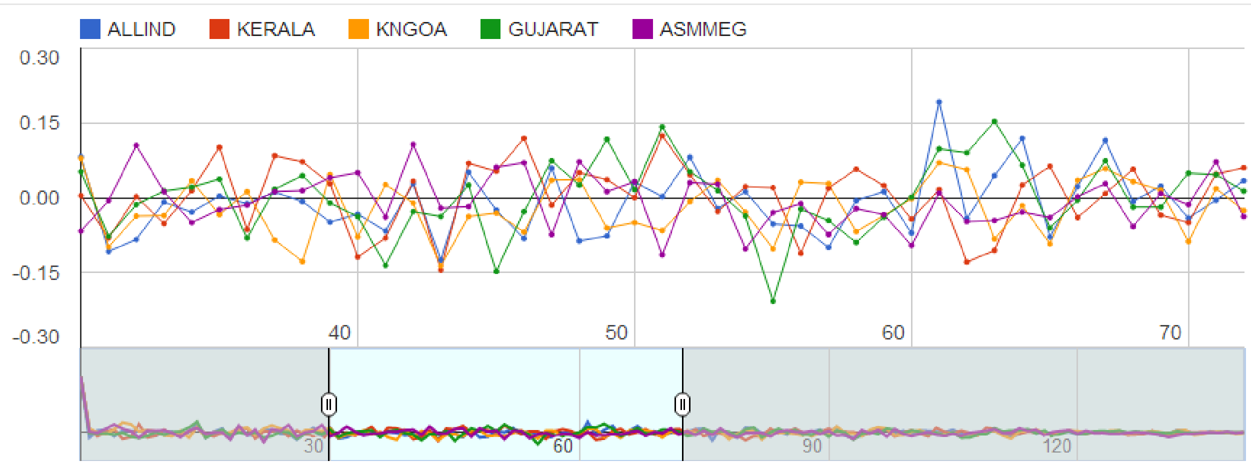
4.11 ACF

ACF plots were also made to verify the results. Autocorrelation is the cross-correlation of a signal with itself. Informally, it is the similarity between observations as a function of the time lag between them. ACF plot was obtained to check whether there is any correlation between the rainfalls of the same region with a time lag. The values obtained were in the range of -1 to +1.

```
> demo <- read.csv("D:/SummerRainfallProject/Updated/demo.csv")
> year <- demo$Year
> mean_kngoa <- demo$Mean_kngoa
> meandiff1_kngoa <- diff(mean_kngoa, differences=1)
> plot.ts(meandiff1_kngoa)
> acf(meandiff1_kngoa, lag.max=20)
> acf(meandiff1_kngoa, lag.max=20,plot=FALSE)
```


Series kngoa_mean
ACF
Lag


Series allind_mean
ACF
Lag


ALLIND
KERALA
KNGOA
GUJARAT
ASMMEG

5. Further Enhancement

- Supplementing the result of the paper that depicted the 30 year wet and dry period can be performed if the data is extrapolated to obtain 5 cycles of 30 years each (According to Hindu calendar).

6. Bibliography

References:

- A Little book of R.
- A wavelet based significance test for periodicities in Indian Monsoon Rainfall by Sarita Azad, R Narasimha, and S.K Sett.
- Periodicities in Indian monsoon rainfall over spectrally homogeneous regions by Sarita Azad, T.S Vignesh, R Narasimha.
- Prediction of Indian summer monsoon: Status, problems and prospects by Rajeevan.