A STUDY ON LONG TERM RAINFALL OF GUWAHATI IN ASSAM

"A Time Series Approach"

SUBMITTED TO

Mr. Ashish Kumar
Regional Meteorological Centre,
Guwahati

Submitted by

Dhritishman Sarmah
B. Tech, Electronics and Communication Engineering,
National Institute of Technology, Silchar

Acknowledgements

A summer project is a golden opportunity for learning and self-development. I consider myself very honoured and lucky to have so many wonderful people lead me through in completion of this project.

My grateful thanks to Mr. Ashish Kumar who in spite of being busy with his duties, took time out to show us how the department works, how the instruments necessary for weather forecasting are used and giving us a tour for every step of it, heard and guided to the correct path.

Also thanks to the department and its staff for giving an opportunity to have the much needed internship here and for the help whenever needed.

Dhritishman Sarmah

Contents

1. Abstract
2. Introduction
3. Study Area
4. Data and Methodology
5. The Python Environment
6. Results and discussion
    1. Descriptive statistics
    2. Time series analysis
    3. Probability Distribution
7. Conclusion
8. License
9. Reference

Abstract

Understanding the rainfall pattern is necessary for finding the solution of several regional environmental issues of water resources management for agriculture, climate change, and natural calamity such as floods and droughts. The data was collected from Regional Meteorological Centre, Guwahati and it consist of daily rainfall data for Guwahati station for a period of 1988-2017. Trends in monthly, annual and seasonal rainfall were examined in this study. Trend analysis showed some significant falling trends in rainfall analysis for annual, monthly and seasonal rainfall data. The average annual rainfall of Guwahati station is 1737.02 mm.

Introduction

Water is one of the most important natural resource and one of the main source of water is rainfall. Rainfall in a region affects livelihood and agriculture, functioning of various industries, hydroelectric power generation and many other aspects of the economy. It is the meteorological phenomenon that has the greatest impact on human activities.

Understanding rainfall variation for a region is essential to optimally manage the available water resources. India is an agriculture based country as most of the country consists huge plains surrounding mighty Ganga and Brahmaputra and their tributaries. These crops depend much on rain. The more crops we get the better for us and for the economy of our country and for other countries as well.

The climate of North-East India is distinct from rest of India due to special features such as orography, the altering pressure cells of the region and that of Bay of Bangle. The region is covered by the mighty Brahmaputra Barak river systems and their tributaries. Assam is the largest state of North-East India. The climate of Assam is typically “tropical monsoon rainfall” type with high levels of humidity and heavy rainfall.

Guwahati, one of the fastest growing cities in India, is situated on the south bank of Brahmaputra river, is the largest city of Assam and North-East India, a major riverine port city. Here, total annual average rainfall is 1737.02 mm. Most of the times, 2-3 hours of constant heavy rainfall causes flood in the city.

Keeping all these points in mind, time series analysis has been carried out to find out any trend in the long term rainfall record of Guwahati station. The objective of this work is to analyse the variation in annual, seasonal and monthly rainfall data for the period 1988-2017 of the station. Finally determine the best fitted distribution for the given rainfall data.

Rainfall type	Short	Amount in mm
No Rain	NR	0.0
Very light Rain	VLR	0.1- 2.4
Light Rain	LR	2.5 – 7.5
Moderate Rain	MR	7.6 – 35.5
Rather Heavy	RH	35.6 – 64.4
Heavy Rain	HR	64.5 – 124.4
Very Heavy Rain	VHR	124.5 – 244.4
Extremely Heavy Rain	EX.HR	>244.5

Table 1 Classification of rainfall based on amount in millimetres.

STUDY AREA

North-East India has predominantly humid sub-tropical climate with hot, humid summers, severe monsoons, and mild winters. Part of Indian sub-continent's last remaining rain forests can be found here which supports diverse flora and fauna. The region is covered by he mighty Brahmaputra-Barak river systems and their tributaries. The south-west monsoon is responsible for bringing 90% of the annual rainfall to the region. April to late October are the months where most of the rainfall occurs. The region's high rainfall creates problems such as flood.

Guwahati sits on the valley of the river Bharalu, a small tributary of the Brahmaputra. It is surrounded by hills except where the Bharalu discharges into the Brahmaputra. It is one the most important cities of Assam because it is close to the seat of power in Assam, is a commercial centre, and is the node that connects six other north-eastern states. The climate in Guwahati is warm and temperate. Compared to winters, summers here have much more rainfall. Average annual temperature in Guwahati is 24.6°C. The average annual rainfall is 1737.02 mm.

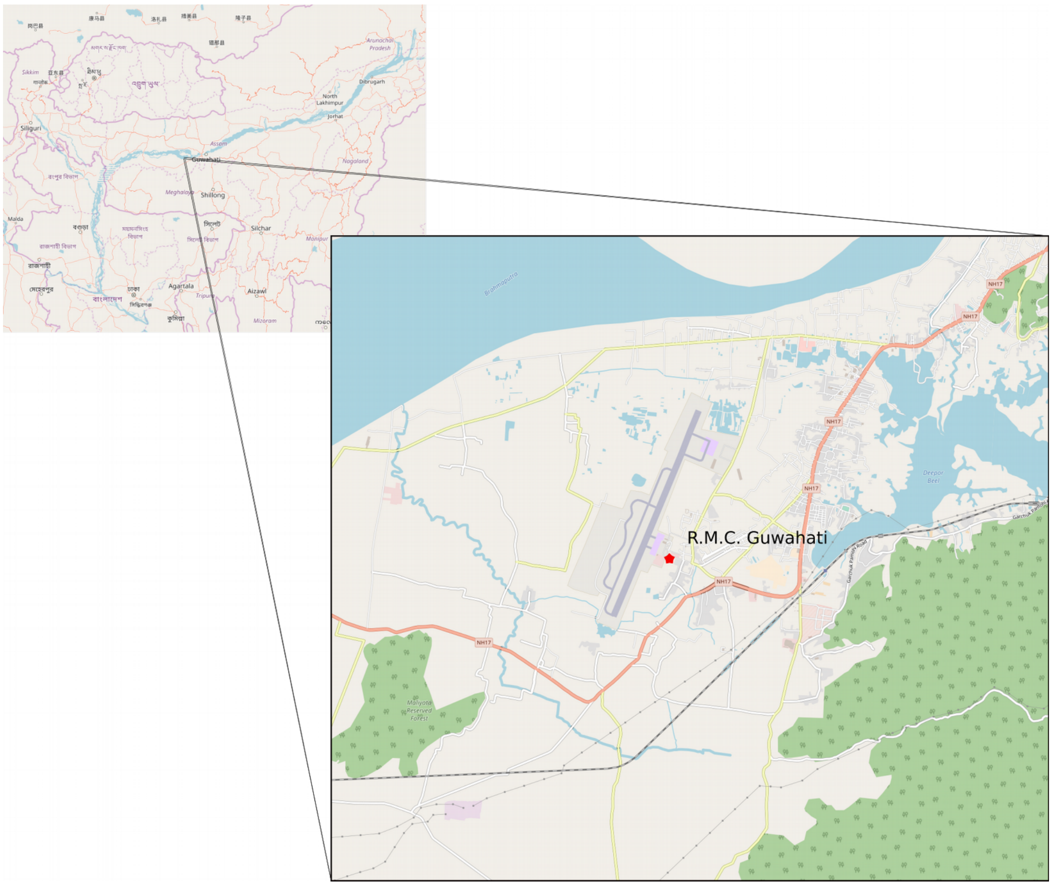

Figure 1 Regional Meteorological Centre, Guwahati. *Source: www.openstreetmap.org*

DATA AND METHODOLOGY

The data for analysis is collected from Regional Meteorological Centre, Guwahati. The data consists of rainfall data on a daily basis for the period 1988-2017 for Guwahati station. This raw data is further simplified and classified into types of rain as described in *table 1*. It should be noted that for those days where rainfall is below 2.5 mm, also called very light rain (VLR), are not considered as a rainy day.

Descriptive Statistics: The following statistical parameters were calculated for a better understanding of the data.

- Maximum
- Mean
- Median
- Variance
- Standard Deviation
- Coefficient of Variation
- 1^{st} quartile
- 3^{rd} quartile

Coefficient of variation is equal to the ratio of standard deviation to the mean in percentage. It is a simple measure of relative event dispersion. The 1^{st} quartile is the median of the first half of the dataset while the 3^{rd} quartile is the median for the second half of the dataset where the data is separated into two halves by the median. 25 % of data points lie below the 1^{st} quartile while 75 % of data points lie below 3^{rd} quartile.

Time Series Analysis: Trends were analysed for annual, monthly and seasonal scale for the data. Annual rainfall was obtained by the summation of monthly data, while seasonal data was the summation of specific set of months which represented the season. Time series forecasting is the use of a model to forecast future events based on known past events to predict data points before they are measured. There are different types of time series analysis

Probability distribution: Based on the value of the chi-square goodness-of-fit value the annual rainfall and the types of rainfall classification data, the best fit theoretical probability distributions are indicated in the table 2. Theoretical probability distributions were superimposed on respective frequency histograms of annual rainfall and types of rainfall data.

Functions	Definition
Normal Distribution	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
Log normal Distribution	$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{lnx-\mu}{\sigma}\right)^2}, x > 0$
Rectangular Distribution	$f(x) = \frac{1}{b-a},\ for \quad a \le x \le b$ $= 0,\ otherwise$
Gamma Distribution	$f(x) = x^{\alpha-1}\frac{e^{-\frac{x}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)}\ for\ x > 0\ and\ \alpha,\beta > 0$
Exponential Distribution	$f(x) = \lambda e^{-\lambda x},\ x \ge 0$ $= 0\ x < 0$

Table 2 Various popular probability distribution with their definition

THE PYTHON ENVIRONMENT

Python is a very powerful programming language that is much easier to learn than some other languages. It’s a great language to start with if you’ ve never programmed before and easy to pick up if you’ve migrated from other language. Python has an excellent interactive shell and has a large collection of free and open source packages, a very simple syntax which makes writing and debugging easy and less time consuming. Python is a multi-purpose language which brings together people from different backgrounds.

Python has become the language of choice for data analytics. One of the major reasons for this is the availability of some free and amazing libraries which makes it interesting to work and analyse large data sets. Some of them are:

- *Numpy* is a library for the python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.
- *Scipy* is a free and open source python library that used for scientific computing. It has a collection of algorithms and high level commands for manipulating and visualising data. It also contains modules for linear algebra, integration, optimization, Fast Fourier transform, image processing and much more.
- *Pandas* is a library for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.
- *Matplotlib* is a Python 2D and 3D plotting library which produces publication quality figures in a variety of formats and interactive environments. Plots such as histograms, power spectra, bar charts, box plots, and many more complex plots can be generated with just a few lines of codes. Matplotlib easily integrates with Pandas dataframe to make visualisations quickly and conviniently.
- *Seaborn* is a Python visualisation library based on matplotlib. It provides a high level interface for drawing attactive statistical graphics.

For this work, we have mostly used the Seaborn library for visualisation and Pandas for data manipulation. The easy to code and simple to read feature made it possible for us to complete the project in time.

RESULTS AND DISCUSSIONS

Descriptive statistics:

Parameters	Annual Rainfall(mm)	Total Rainy Days
Mean	1737.02	90.63
Variance	68702.71	99.55
Standard Deviation	262.11	9.98
Coefficient of Variation	15.09	11.01
Minimum	1296.70	65
1st quartile	1550.62	85.25
Median	1746.25	92
3rd quartile	1872.90	97
Maximum	2250.20	108

Table 3 Descriptive Statistics for annual rainfall

Table 3 shows some of the statistics for total annual rainfall of Guwahati station. The average annual rainfall is 1737.02 mm and average no. of total rainy days in a year is 90.63 . The coefficient of variation is higher for total annual rainfall than total no. of rainy days.

Time Series Analysis:

From the time series plots for total annual rainfall we observe that there is a significant *downward* trend in total annual rainfall and total no. of rainy days in a year, while maximum rainfall for a year shows *upward* trend.

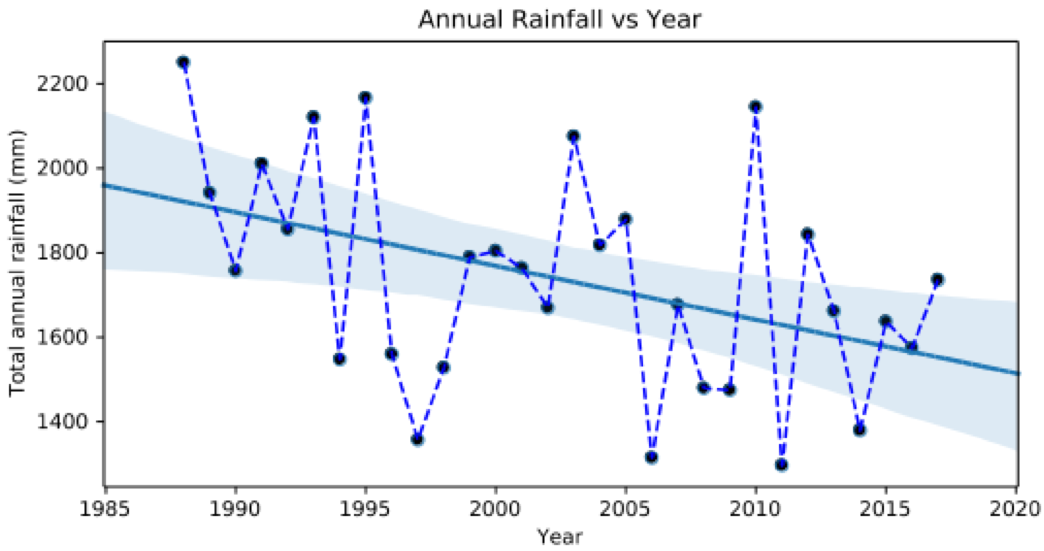

Figure 2 Time series plot for annual rainfall with regression line.

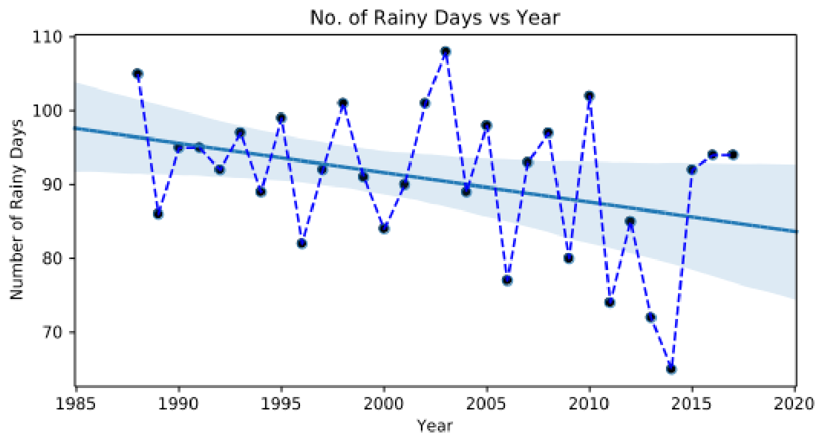

Figure 3 Time series plot for no. of rainy days in a year with regression line.

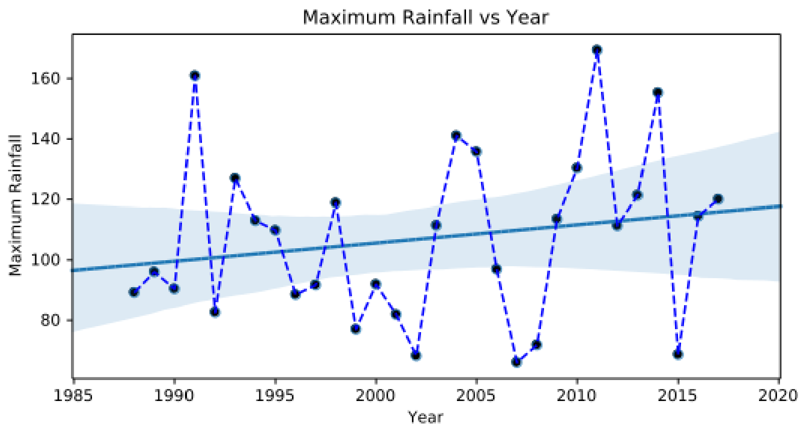

Figure 4 Time series plot for Maximum annual rainfall with regression line.

A boxplot is an efficient and concise way to graphically represent parameters such as max, min, median , 1st and 3rd quartiles all in a single graph. From the boxplot analysis on monthly rainfall, we can clearly see that months from *April* to *September* has higher rainfall the rest of the year. This argument is aided by the fact that , it is the same part of the year where monsoon usually occurs.

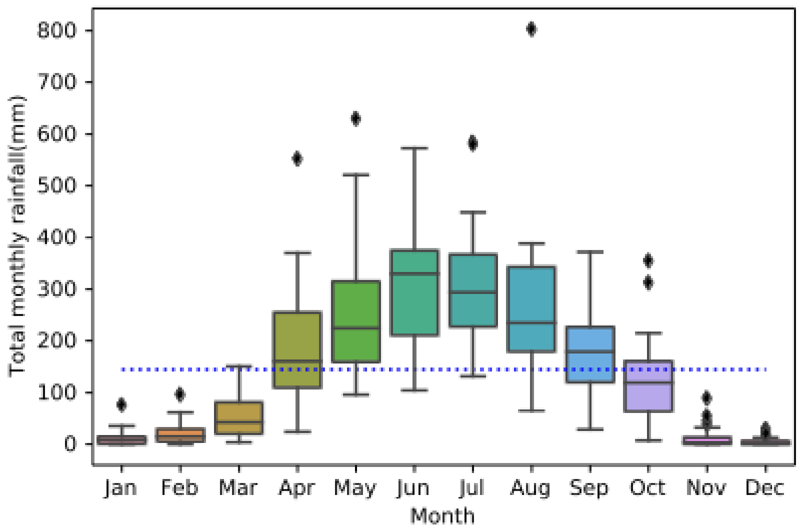

Figure 5 Boxplot for total monthly rainfall where the horizontal line denotes average monthly rainfall.

Time series analysis for monthly rainfall for only two months are showed here as they showed significant trends in their analysis. The month of July seems to receiving less rainfall over the years whereas April is receiving more rainfall over the years.

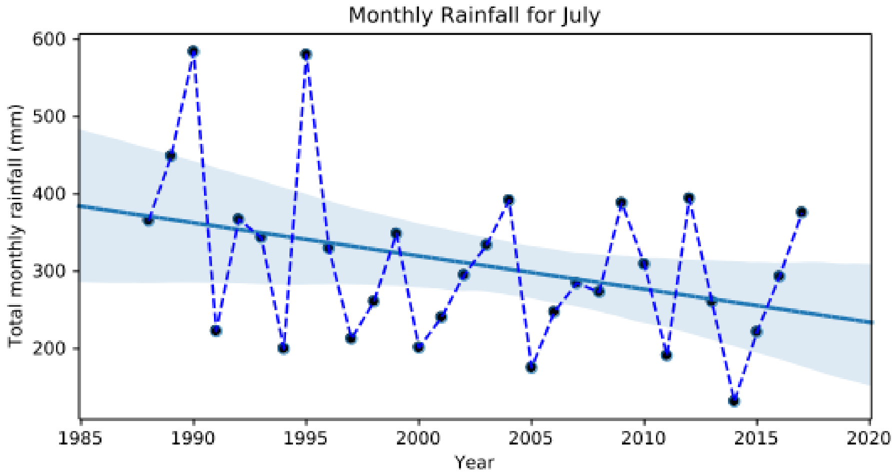

Figure 6 Time series plot for monthly rainfall for July with regression line.

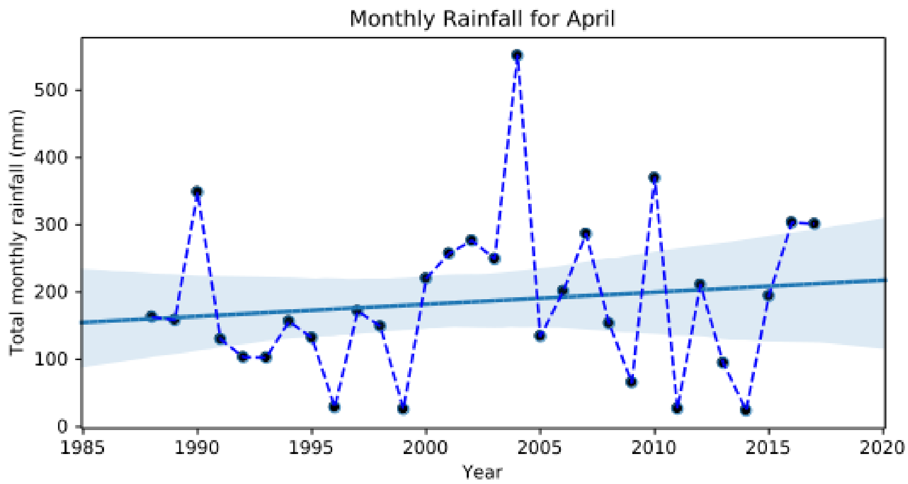

Figure 7 Time series plot for monthly rainfall for April with regression line.

A countplot for seasonal rainfall represents how number of rainy days is distributed among the four seasons.

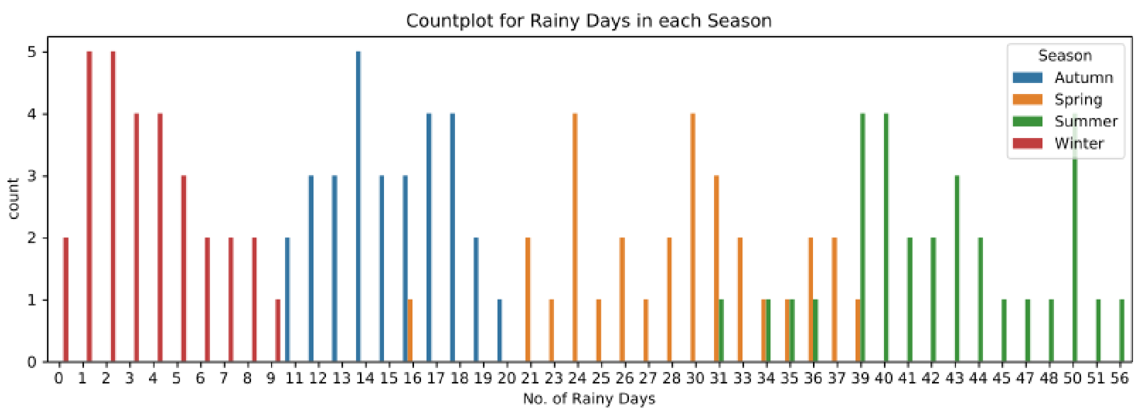

Figure 8 Countplot for total no. of rainy days in each season for various years.

Interstingly, total no. of rainy days for winter season never went above 10 days, while that for summer season never came below 31 days. The autumn season stays inside the interval 10-20 days and spring seasons between 21-39 days with one extreme data point at 16. This plot prooves to be very significant in predicting the season based on only the total no. of rainy days in the season.

Time series analysis on seasonal raindfall tells us that their is *downward* trend in total seasonal rainfall and no. of rainy days while maximum seasonal rainfall doesn't show any significant trend.

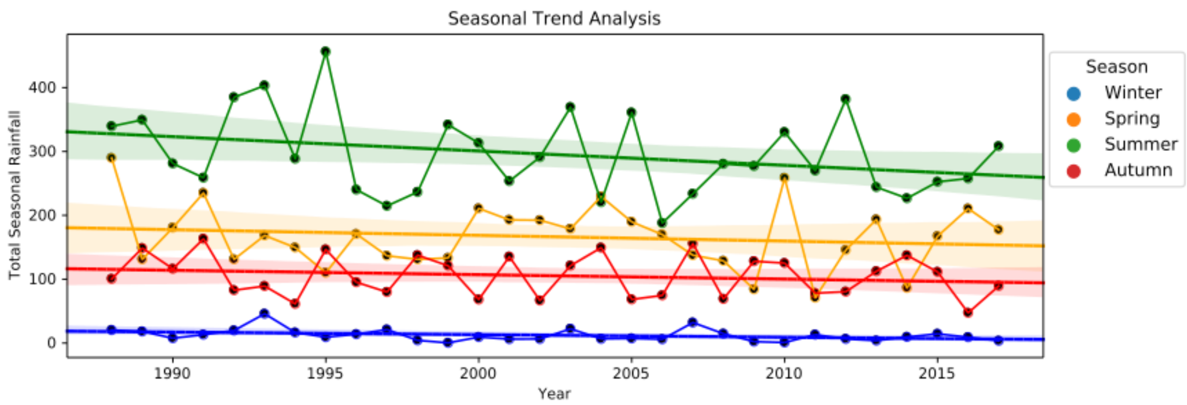

Figure 9 Time series plot for total seasonal rainfall with regression line for each season.

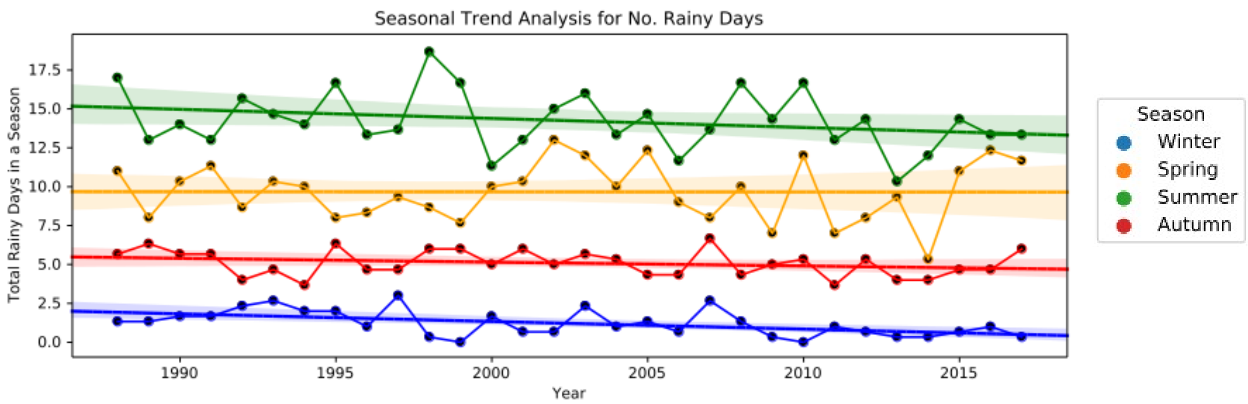

Figure 10 Time series plot for total no. of rainy days with regression line for each season.

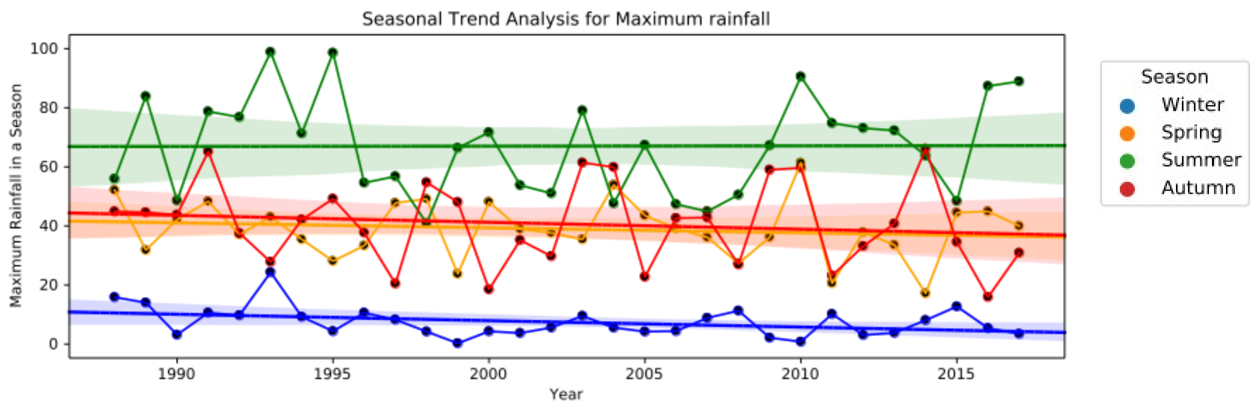

Figure 11 Time series plot for maximum seasonal rainfall with regression line for each season.

Probability Distribution:

From the histogram plots fitted with their respective best-fit probability distribution, it can be clearly seen that heavy rainfall(1500mm - 2000mm) has higher frequencies of occurence. Annual trends mostly follow normal distribution while monthly trends best fits to exponential distribution.

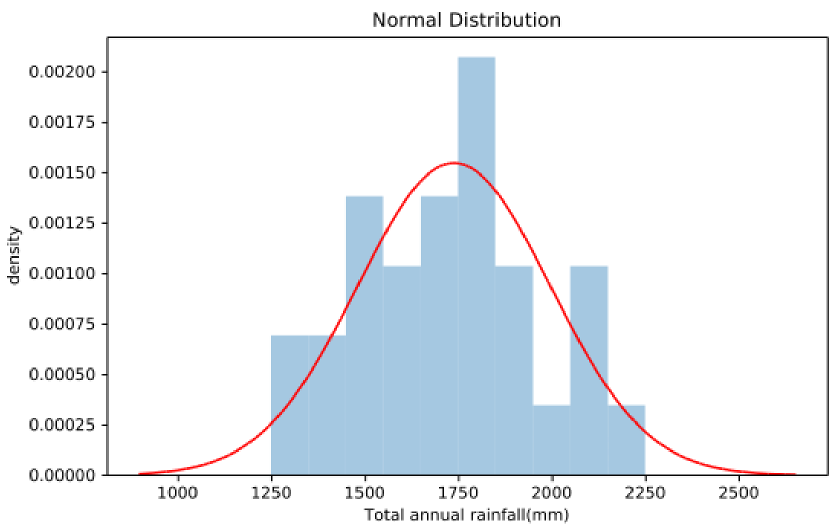

Figure 12 Histogram plot for total annual rainfall fitted with normal distribution.

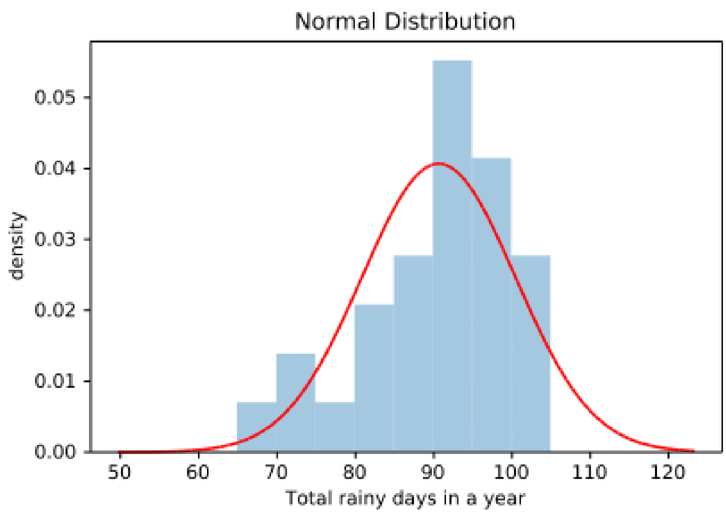

Figure 13 Histogram plot for total rainy days ina year fitted with normal distribution.

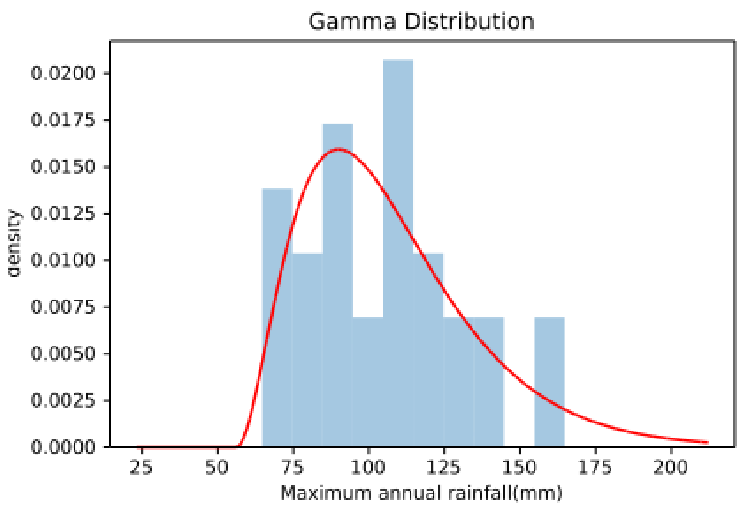

Figure 14 Histogram plot for maximum rainfall for a year fitted with gamma distribution.

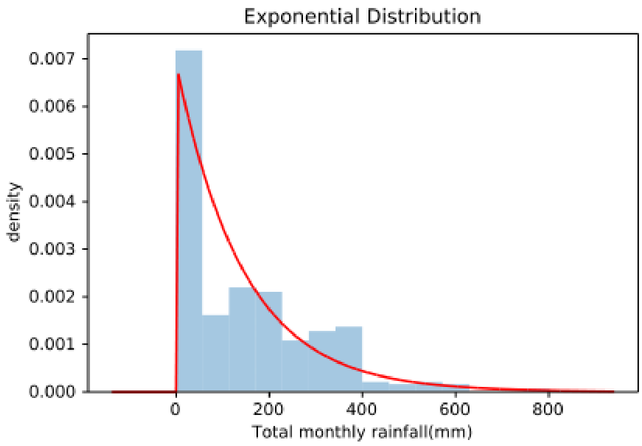

Figure 15 Histogram plot for total monthly rainfall fitted with exponential distribution.

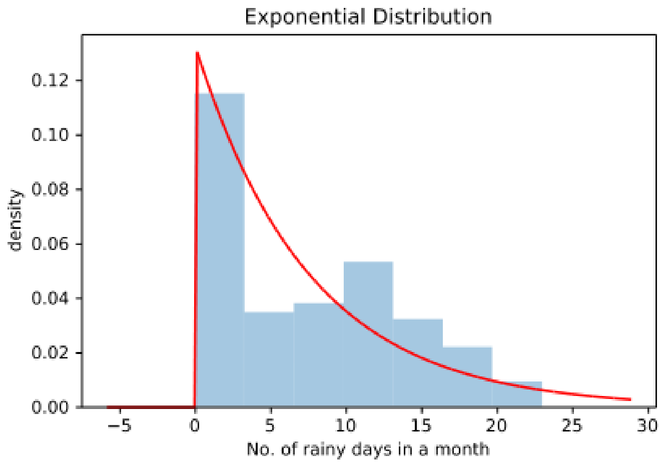

Figure 16 Histogram plot for no.of rainy days in a month fitted with exponential distribution.

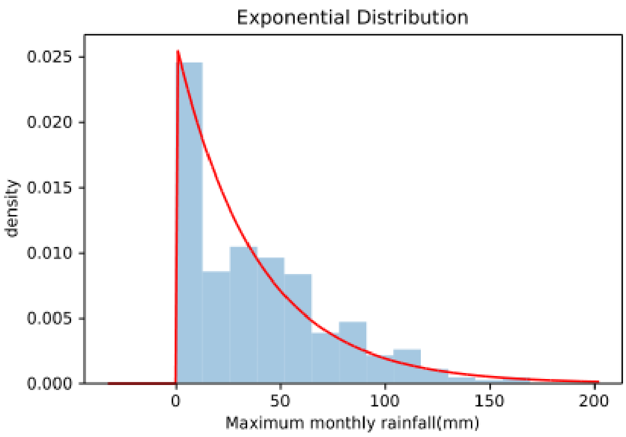

Figure 17 Histogram plot for maximum monthly rainfall fitted with exponential distribution.

CONCLUSIONS

The rainfall characteristics especially variability and trend are necessary for the proper planning and management of water resources. This report deals with the study of long term rainfall trends for Guwahati station for the period 1988-2017. From the results of the time series analysis it can broadly concluded that:

1. There is a significant downward trend in the total annual rainfall and in the number of rainy days in a year.
2. A rising trend was observed in time series analysis for the maximum rainfall for a year.
3. Seasonal analysis showed that winters receive very less rainfall i.e below 10 rainy days in the whole season while summer receives at least 30 days of rainfall.
4. Time series analysis for seasonal rainfall showed a falling trend in total seasonal rainfall in summer season as well as winter season.
5. Time series analysis for no. of rainy days in a season as well as maximum seasonal rainfall also showed falling trend for summer and winter seasons.
6. However, no significant trend could be seen for autumn and spring season during time series analysis for total seasonal rainfall, no. of rainy days and maximum seasonal rainfall.
7. Time series analysis for monthly rainfall for two months April and July were observed. Significant falling trend was for the month of July while a rising trend was observed for that in April.
8. Probability distribution analysis showed that annual rainfall mostly follows normal distribution while monthly rainfall data follows exponential distribution.

LICENSE

Reference

- http://imd.gov.in/section/nhac/termglossary.pdf for types of rainfall.
- http://python-graph-gallery.com for different types of visualisation examples.
- Stackoverflow for all those small queries.
- http://www.researchgate.net for those example reports.
- Google Search for its never-ending support.
- Wikipedia for definitions.