# Statistical Analysis of 30 Years Rainfall Data: A Case Study

**G Arvind, P Ashok Kumar, S GirishKarthi and C R Suribabu***

School of Civil Engineering, SASTRA University, Thanjavur – 613 401, India.

*Email: suribabu@civil.sastra.edu

**Abstract.** Rainfall is a prime input for various engineering design such as hydraulic structures, bridges and culverts, canals, storm water sewer and road drainage system. The detailed statistical analysis of each region is essential to estimate the relevant input value for design and analysis of engineering structures and also for crop planning. A rain gauge station located closely in Trichy district is selected for statistical analysis where agriculture is the prime occupation. The daily rainfall data for a period of 30 years is used to understand normal rainfall, deficit rainfall, Excess rainfall and Seasonal rainfall of the selected circle headquarters. Further various plotting position formulae available is used to evaluate return period of monthly, seasonally and annual rainfall. This analysis will provide useful information for water resources planner, farmers and urban engineers to assess the availability of water and create the storage accordingly. The mean, standard deviation and coefficient of variation of monthly and annual rainfall was calculated to check the rainfall variability. From the calculated results, the rainfall pattern is found to be erratic. The best fit probability distribution was identified based on the minimum deviation between actual and estimated values. The scientific results and the analysis paved the way to determine the proper onset and withdrawal of monsoon results which were used for land preparation and sowing.

## 1. Introduction

Water is vital for any life process and there can be no substitute for it. Water is also used for transportation, is a source of power and serves many other useful purposes for domestic consumption, agriculture and industry. The main important source of water in any area is rain and it has a dramatic effect on agriculture. Plants get their water supply from natural sources and through irrigation. The yield of crops particularly in rain-fed areas depends on the rainfall pattern, which makes it important to predict the probability of occurrence of rainfall from the past records of hydrological data using statistical analysis. Frequency or probability distribution helps to relate the magnitude of the extreme events like floods, droughts and severe storms with their number of occurrences such that their chance of occurrence with time can be predicted easily. By fitting a frequency distribution to the set of hydrological data, the probability of occurrences of random parameter can be calculated. To fit the distribution, the hydrological data is analyzed and the variability in the data is studied from the statistical parameters. Suchit Kumar Rai et al., studied the change, variability and rainfall probability for crop planning in few districts of Central India [1]. Nyatuame et al. [2] performed the statistical analysis and studied the variability in the distribution of rainfall. Rajendran et al. [3] carried out the frequency analysis of rainy days and studied the rainfall variation.

## 2. Study Area

The present study is carried out for Musiri (Tiruchirapalli district, Fig.1) town situated at a distance of 29 km from Tiruchirapalli city. The region has a latitude and longitude of 10.9549°N and 78.4439°E respectively. Agriculture is the main occupation in this town situated on the northern bank of Cavery river. The crops include paddy, sugarcane, banana and vegetables. Musiri receives rainfall from both the northeast and southwest on an average (30 years) of 245.49 mm and 352.62 mm respectively. The daily rainfall data is collected from the Indian Meteorological Department (IMD), Musiri station, for a period of 30 years (1984-2013). This data is used for the Yearly, Monthly and Seasonal Rainfall-Probability analysis. Figure 2 presents the historical annual rainfall for the station.

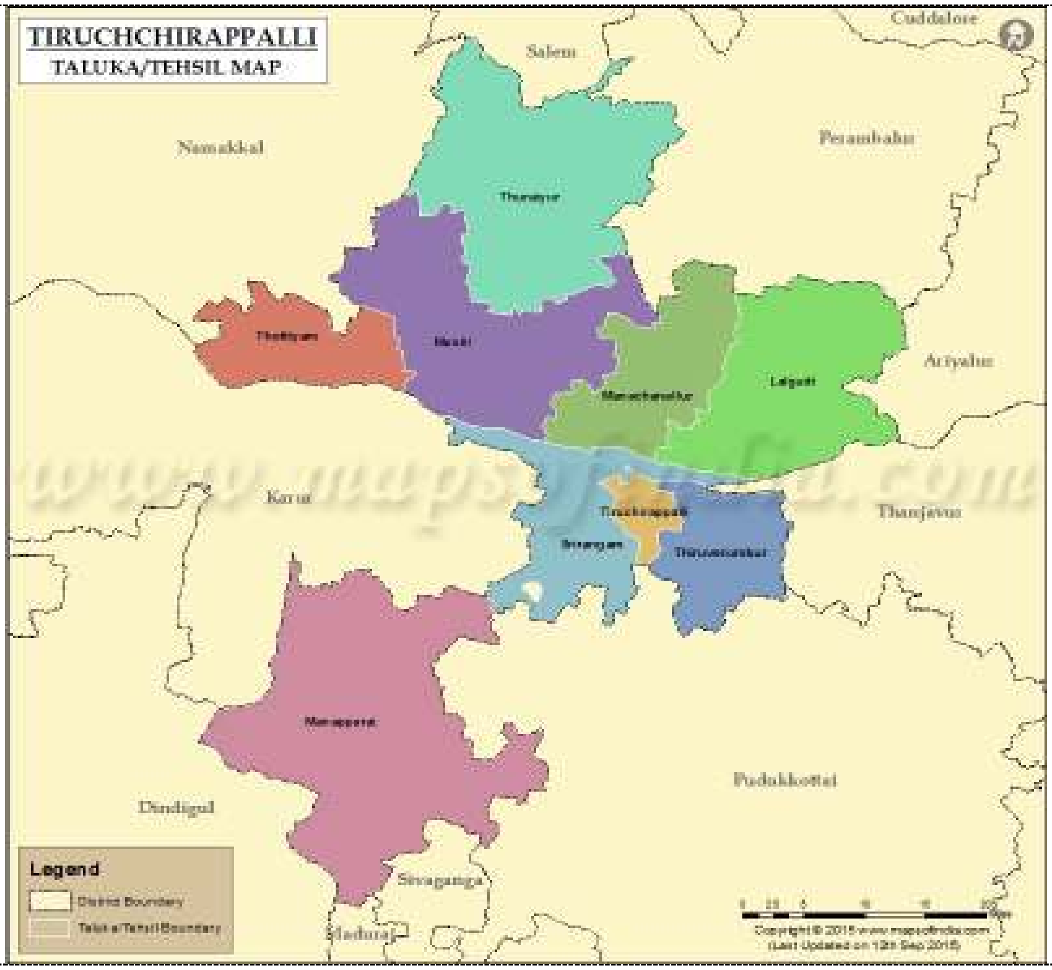


**Figure 1.** Location map of the study area

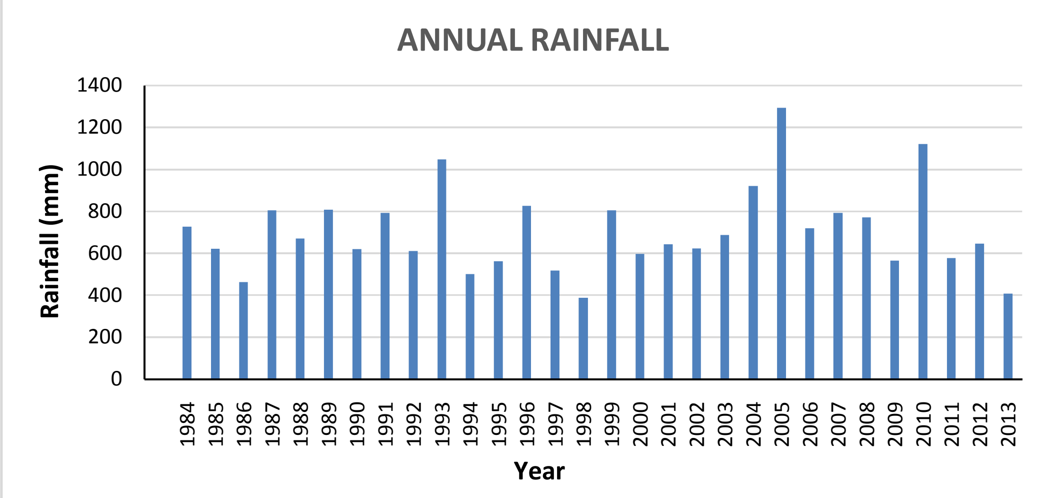


**Figure 2.** Annual Rainfall of Musiri town

## 3. Methodology

The methodology adopted in this study is Rainfall Statistics (Table 1), Probability analysis using plotting position and probabilistic methods (Table 2). From the Preliminary study and analysis, variation in results among the plotting position methods is found to be insignificant.

**Table 1.** Formulae for Statistical Parameters

| Description | Symbol | Formula | Explanation |
|---|---|---|---|
| Arithmetic Mean | $X_{avg}$ | $\sum X_i / n$ | X is the rainfall magnitude in mm, i=1, 2, to n and n is the length of the sample. |
| Standard deviation | Σ | $[\sum (X_i - X_{avg})^2 / (n-1)]^{1/2}$ | X is the rainfall magnitude in mm, i=1, 2, to n and n is the length of the sample. |
| Co-efficient of Variation | $C_v$ | $100 \times (\sigma / X_{avg})$ | $X_{avg}$ is the Mean<br>σ is the Standard deviation |
| Co-efficient of Skewness | $C_s$ | $(1 / \sigma^3) \times [(N / (N^2 - 3N + 2)] \times \sum (X_i - X_{avg})^3$ | σ is the Standard deviation<br>N = Total no. of years<br>$X_{avg}$ is the Mean<br>X is the rainfall magnitude in mm, i=1, 2 to n |

### *3.1 Annual Rainfall Analysis*

The annual rainfall data is analyzed and the variation in distribution over the area is studied with the statistical parameters. The best fit distribution method is found using various plotting position and probabilistic methods.

**Table 2** Plotting position and Probabilistic methods

| S. No. | Plotting position methods | Probabilistic methods |
|---|---|---|
| 1 | California = m/N | Normal Distribution |
| 2 | Hazen = (m-0.5)/N | Log-Normal Distribution |
| 3 | Weibull = m/(N+1) | Pearson Type-III Distribution |
| 4 | Beard = (m-0.31)/(N+0.38) | Log-Pearson Type-III Distribution |
| 5 | Chegodayev = (m-0.3)/(N+0.4) | Extreme Value Type-I Distribution |
| 6 | Blom = (m-3/8)(N+1/4) | |
| 7 | Tukey = (3m-1)/(3N+1) | |
| 8 | Gringorten = (m-0.44)/(N+0.12) | |
| 9 | Cunnane = (m-0.4)/(N+0.2) | |
| 10 | Adamowski = (m-1/4)/(N+1/2) | |

where, m is rank of the data and N = length of the sample (no. of years).

*3.2 Monthly Rainfall Analysis*

From the Preliminary study and analysis, variation in results among the plotting position methods is found to be insignificant and hence, only Weibull method is adopted for the analysis among them. From the Probabilistic methods, Gumbel and Normal distribution methods are used.The rainfall data are arranged into a number of intervals with definite ranges. Mean and standard deviation were found out for the grouped data. Chi-square values are calculated for the above methods, with the obtained probabilities. The method that gives the least Chi-square value is found to best fit the distribution.
Weibull Distribution is a continuous probability distribution type where in rainfall amounts are assigned with a rank and the corresponding probabilities are found out using probability density function:

$$(X) = m/\ (n + 1) \quad (1)$$

Where,

m and n represents the rank and total number of data used in the analysis

*Gumbel Distribution* is used to model the distribution of the extremities of a number of samples of various distributions.

P(X) = exp. ((- (a + x)/c) – e – (- (a + x) / c))

$$a = 0.450055\ \sigma - Xavg \,\&\, c = 0.7797\ \sigma \quad (2)$$

Where,

P(X) is the probability density function for Gumbel method and Xavg represents the average rainfall in mm

Normal Distribution is a very common continuous probability distribution. Normal distributions represent real-valued random variables whose distributions are not known.

$$B = 0.5[1 + 0.196854\ |Z| + 0.115194\ |Z|^2 + 0.000344\ |Z|^3 + 0.015927\ |Z|^4$$

$$Z = (X - Xavg)\ /\ \sigma,\ F\ (Xi) = B for Z< 0\ \&F\ (Xi) = 1 - B for Z> 0 \quad (3)$$

The Probability Density function for the Normal Distribution method is as follows:

$$(X) = F\ (Xi + 1) – (X) \quad (4)$$

Where,

Xi is the rainfall at any instant i = 1,2,3 to n

*Goodness of Fit* is a test used to find out the best fit probability distribution. The best fit distribution varies for different time period. Chi-squared test is used in the determination of best fit distribution for weekly and seasonal rainfall in this study. Chi-Squared Test is used for continuously sampled data only and is used to determine if a sample comes from a population with a specific distribution.

$$Xc^2 = \sum(O-E)^2 / E \quad (5)$$

Where,

$\sum$ from i = 1 to k, O = Observed frequency,
E = Expected frequency,
i = Number of observations and k = the total number of data used

Chi-Square Formula adopted in the study is as follows:

$$Xc^2 = [(fs\ (Xi) - P\ (Xi))^2 / P\ (Xi)]\ x \sum Ni$$
$$fs\ (Xi) = Ni / \sum Ni \quad (6)$$

Where,

Xc is the Chi – squared value,
P(X) is the probability density function

*3.3 Seasonal Rainfall Analysis*

In this analysis, the variation of distribution of rainfall is studied with the statistical parameters using formulae mentioned in Table 1.

*3.4 Effective Rainfall*

Water requirement for various crops is found and they are related with the effective rainfall, which is calculated from the rainfall data. Effective rainfall is the amount of rainfall effectively used by the crops. The effective rainfall in the study area is calculated using the formula:

$$R_e = 0.8*P-25, \text{ if } P \geq 75mm \quad (7)$$
$$R_e = 0.6*P-10, \text{ if } P < 75mm \quad (8)$$

where,

$R_e$ is the Effective Rainfall (mm),
P is the Total Monthly Rainfall (mm)

## 4. Results and Discussions

*4.1 Annual Rainfall Analysis*

The rainfall data are ranked in descending order and various plotting position and probabilistic methods are applied to determine the return period. Rainfall magnitudes were calculated for different return periods using the rainfall-return period equation obtained from the graphs for all plotting position methods (Table 3).

**Table 3.** Maximum Annual Rainfall based on Plotting Position methods

| Method/Return period | 10 | 30 | 50 | 100 | 300 | 500 |
|---|---|---|---|---|---|---|
| WEIBULL | 1019.914 | 1275.352 | 1394.124 | 1555.288 | 1810.726 | 1929.499 |
| CALIFORNIA | 1027.544 | 1282.982 | 1401.754 | 1562.918 | 1818.356 | 1937.129 |
| HAZEN | 973.8311 | 1199.124 | 1303.879 | 1446.022 | 1671.315 | 1776.07 |
| **CHEGODAYEV** | **994.2638** | **1233.036** | **1344.059** | **1494.708** | **1733.48** | **1844.503** |
| BLOM | 987.0367 | 1221.063 | 1329.879 | 1477.533 | 1711.56 | 1820.376 |
| GRINGORTEN | 980.3763 | 1210.008 | 1316.781 | 1461.663 | 1691.295 | 1798.067 |
| BEARD | 993.3244 | 1231.482 | 1342.218 | 1492.479 | 1730.636 | 1841.373 |
| TUKEY | 991.092 | 1227.788 | 1337.845 | 1487.184 | 1723.88 | 1833.937 |
| CUNNANE | 984.5198 | 1216.887 | 1324.932 | 1471.54 | 1703.907 | 1811.952 |
| ADAMOWSKI | 998.8729 | 1240.656 | 1353.078 | 1505.626 | 1747.408 | 1859.831 |
| **AVERAGE MAX RAINFALL** | **995.078** | **1233.838** | **1344.855** | **1495.496** | **1734.256** | **1845.274** |

For Musiri, California method gives the maximum value for rainfall for different return periods and Hazen method is found to give the least value and is hence not acceptable for the analysis. It is seen that Chegodayev method gives a maximum rainfall which is approximately 99.9% to that of the average maximum rainfall unlike other methods of distribution and is hence the best fit distribution for annual rainfall data. It is also seen that whenever there is an increase in return period, the rainfall amount also increases and vice-versa. Hence, Rainfall and return period are proportional to each other.

**Table 4.** Maximum Annual Rainfall based on Probabilistic methods

| Distribution/Return Period | 50 years (x) in mm | 100 years (x) in mm |
|---|---|---|
| NORMAL | 1119.037 | 1173.951 |
| **GUMBEL** | **1315.274** | **1441.859** |
| PEARSON TYPE-III | 1222.606 | 1322.139 |
| LOG-NORMAL | 1193.561 | 1286.205 |
| LOG-PEARSON TYPE-III | 1225.471 | 1334.455 |

Considering the results of Plotting positions to be actual, Gumbel distribution (Extreme value type-I) gives a value that is closer to the actual value and is hence the best method of fit for the annual rainfall data (Table 4).

**Table 5.** Statistical Parameters for Annual Rainfall Analysis

| S. No. | Description | Normal series | Log-transferred series |
|---|---|---|---|
| 1 | Mean Rainfall (mm) | 704.35 | 2.83168 |
| 2 | Standard deviation (mm) | 201.88 | 0.119355 |
| 3 | Co-efficient of Variation | 0.2866 | 0.042150 |
| 4 | Co-efficient of skewness | 1.0582 | 0.181678 |

*4.1.1 Statistical parameters.* The above Table 5 shows that the Standard deviation value is considerable large which indicate there is larger variation in rainfall pattern. Skewness represents the distribution of data about the mean. It is equal to zero in the case of normal distribution. When the peak of the sample is towards the right of a plotted graph, it is said to be negatively skewed and when the peak of the sample is towards the left of a plotted graph, it is said to be positively skewed. From the above table, it is clear that Normal series and Log-transferred series data are positively skewed.

*4.2 Monthly Rainfall Analysis*

The rainfall data are arranged into a number of intervals with a range of 25 mm and the frequency of occurrence is found out initially, to convert the normal data into a grouped data. Mean and Standard Deviation were found out for the same to check the variation in rainfall. Chi-square values obtained are compared for all the methods and the Least Chi-square value in all the cases is given by Gumbel distribution.

**Table 6.** Chi-square values for Monthly rainfall analysis

| Distribution | Chi-Square value |
|---|---|
| WEIBULL | 2443.51 |
| **GUMBEL** | **258.93** |
| NORMAL | 367.4 |

Thus, from the above Table 6, it is inferred that the chi-square value is least for Gumbel distribution, showing that it best fits the monthly rainfall data.

**Table 7.** Statistical Parameters for Monthly Rainfall Analysis

| Months | Mean (mm) | Standard Deviation (mm) | Coefficient of Variation | Coefficient of Skewness |
|---|---|---|---|---|
| January | 8.877 | 23.637 | 2.663 | 2.823 |
| February | 6.563 | 18.887 | 2.877 | 3.350 |
| March | 13.047 | 27.319 | 2.094 | 2.698 |
| April | 25.250 | 35.078 | 1.389 | 2.239 |
| May | 52.493 | 49.681 | 0.946 | 2.100 |
| June | 23.327 | 38.466 | 1.649 | 2.399 |
| July | 37.213 | 41.139 | 1.105 | 0.852 |
| August | 66.147 | 48.218 | 0.729 | 0.250 |
| September | 118.820 | 74.452 | 0.626 | 0.896 |
| October | 145.003 | 81.348 | 0.561 | 0.630 |
| November | 156.077 | 108.673 | 0.696 | 0.856 |
| December | 51.537 | 64.968 | 1.260 | 2.694 |

*4.2.1 Statistical parameters.* From the above Table 7, it is observed that maximum average rainfall is received in the months of August, September, October and November during the monsoons, having the standard deviation values less than their corresponding mean values. Whereas, the standard deviation values for the remaining months are higher than their corresponding mean values showing larger variation in the distribution of rainfall over the months. From the above table it is clear that, Normal series data are positively skewed.

**Table 8.** Statistical Parameters for Seasonal Rainfall Analysis

| S. No. | Statistical Parameters | SW Monsoon | NE Monsoon |
|---|---|---|---|
| 1 | Mean Rainfall (mm) | 245.49 | 352.62 |
| 2 | Standard deviation (mm) | 116.12 | 166.70 |
| 3 | Co-efficient of Variation | 0.0157 | 0.0157 |
| 4 | Co-efficient of skewness | 0.406 | 0.896 |

*4.3 Seasonal Rainfall Analysis*

From the above Table 8, it is shown that Musiri receives the rainfall seasonally on an average of 245.49 mm and 352.62 mm during south-west (SW) and north-east (NE) monsoons respectively and it also shows that to how much extent the distribution of rainfall is varied over the study area during the respective monsoons. The monthly rainfall series data of both the monsoons are positively skewed.

*4.4 Crop Planning*

Average effective rainfall for 30 years were found and tabulated. From table 8, it is observed that the average effective rainfall is higher in the months of August, September, October and November. Whereas, the average effective rainfall in the months of January, February and March is zero. Based on the effective rainfall the crop planning is done which is shown in Table 9.

**Table 9.** Average effective rainfall for 30 years

| Months | Average Rainfall (mm) |
|---|---|
| January | 0 |
| February | 0 |
| March | 0 |
| April | 5.90 |
| May | 23.64 |
| June | 4.96 |
| July | 13.53 |
| August | 32.94 |
| September | 72.15 |
| October | 92.05 |
| November | 101.42 |
| December | 23.76 |

Musiri taluk has net cultivable area of 27,344 hectares which constitutes 13.4% of total agricultural area in Trichy district. The soil type in Trichy district is found to be Red sandy soil and black soil. Crops like Paddy, Sugarcane, Maize, Pulses (Black-gram and Soybean), Groundnut, Sesame(Gingelly) and Sunflower are very much suitable to be grown in Red Sandy soil and Black soil is suitable for growing Cotton. Cultivation of crop varies according to the climate and water requirement. From the collected data, it has been found that Paddy constitutes about 39% the total cultivable area and stands top in net production followed by Groundnut which is cultivated for about 2338 ha, Maize for about 1315 ha, Sugarcane for about 1002 ha, Cotton for 442 ha, Gingelly for merely 57 ha and other cash crops like Cholam, Cumbu etc for the remaining area. Based on the effective rainfall data, crop water requirement for various crops were identified and segregated into rain-fed and irrigated. The Table 8 above shows the best sowing and harvesting time for cultivation of various crops and water requirement through rain and irrigation.

**Table 10.** Crop planning for the Musiri Region

| S. No. | Crops | TWR* | best suited month | | Rain-Fed (mm) | Irrigated (mm) |
|---|---|---|---|---|---|---|
| | | | Sowing | Harvesting | | |
| 1 | Maize | 500mm for 100 days | Early August | November | 299 | 201 |
| 2 | Pulses: | | | | | |
| | 1. Blackgram | 280mm for 65 days | Mid August | October | 197 | 83 |
| | 2. Soybean | 320mm for 85 days | September | November | 266 | 54 |
| 3 | Groundnut | 510mm for 105 days | September | December | 290 | 220 |
| 4 | Sesame (Gingelly) | 150mm for 85 days | July | September | 119 | 31 |
| 5 | Cotton | 600mm for 165 days | Mid July | Mid December | 335 | 265 |
| 6 | Sunflower | 450mm for 110 days | August | November End | 299 | 151 |

*TWR-Total Water Requirement

## 5. Conclusion

From the Rainfall Probability analysis on the Annual and Monthly rainfall for Musiri Region, it is evident that Gumbel Distribution (Extreme Value Type – I) is ascertained as the best fit distribution type considering its Least Chi-Square Value among all other methods of analysis. Chegodayev Distribution from Plotting Position methods is found to best fit the Annual rainfall data. The present statistical analysis provides clear picture on rainfall data and it is found that rainfall available in the region is insufficient to carry out wet crop. Conjunctive use of surface water, available rainfall and ground water is essential for better agricultural and irrigation management for this area, Thus, the analysis helps in understanding the rainfall pattern of Musiri region and also in efficient crop planning and water availability of the region.

## Acknowledgements

The authors acknowledge the Vice Chancellor, SASTRA University, Thanjavur, India for providing the facilities to carry out the work and the encouragement in completing this work.

## References

[1] Suchit Kumar Rai, Sunil Kumar, Arvind Kumar Rai, Satyapriya and Dana Ram Palsaniya 2014 Climate Change Variability and Rainfall Probability for Crop Planning in Few Districts of Central India. *Atmos. Climate Sci.* **4** 394-403.
[2] Nyatuame M, Owusu-Gyimah V and Ampiaw F 2014 Statistical Analysis of Rainfall Trend for Volta Region in Ghana. *Int. J. Atmos. Sci.* **67**(2) 1-11.
[3] Rajendran V, Venkatasubramani R and Vijayakumar G 2016 Rainfall variation and frequency analysis study in Dharmapuri district (India). *Indian J. Geo. Mar. Sci.* **45**(11) 1560-5.