Frequency Analysis of Rainfall Data of Dharamshala Region

Nitish Kumar Sharma[1], Swedika Sharma[2]

[1]Assistant Professor, Department of Civil Engineering, Chandigarh University, Gharuan

[2]Assistant Professor, Department of Computer Science Engineering, Chandigarh University, Gharuan

Abstract: ***The present study evaluates the rainfall magnitude for different return periods and also to ascertain the type of probability distribution that best fits the rainfall data of Dharamshala (H.P.), India. The study uses the 20 years of annual rainfall data that are useful for the prediction of annual one day maximum rainfall and two to seven days consecutive days maximum rainfall corresponding to return period varying from 2 to 20 years are to be used for the economic planning, by design engineers and by hydrologists, design of small and medium hydrologic structures and determine the drainage coefficient for agricultural fields. Most probability distributions (viz: Normal, Log Normal and Gamma distribution) and transformations are applied to estimate one day and two to seven consecutive days annual maximum rainfall of various return periods in the Dharamshala (H.P.) region. The mean value of one-day annual maximum rainfall at Dharamshala is found to be 142.9 mm with standard deviation and coefficient of variation of 54.8 and 51.34 respectively. The coefficient of skewness is 1.1. For 2 to 7 days consecutive annual maximum rainfall range values for mean, standard deviation, coefficient of variation and coefficient of skewness are 201 – 393.4 mm , 70.17 – 146.5, 41.65 – 30.47 and 0.726 – 1.593. It is observed that all distribution fitted function significantly. A maximum of 191.1 mm in 1 day, 262.7 mm in 2 days, 314.27 mm in 3 days, 370.24 mm in 4 days, 419.7 mm in 5 days, 477.4 mm in 6 days and 522.32 mm in 7 days is expected to occur at Dharamshala, Himachal Pradesh every 2 years. For a recurrence interval of 20 years, the maximum rainfall expected in 1 day, 2, 3, 4, 5, 6 and 7 days is 277.7 mm, 373.6 mm, 445.11 mm, 518.62 mm, 589 mm, 680.3 mm and 753.79 mm respectively. Various probability distributions and transformations are applied to estimate one day and two to seven consecutive days annual maximum rainfall of various return periods. Three commonly used probability distributions (viz: Normal, Log Normal and Gamma distribution) are tested by comparing the Chi-square, Kolmogorov-Smirnov and Anderson Darling values.***

Keywords: rainfall, consecutive, normal, lognormal, gamma, distribution, probability, frequency, analysis

1. Introduction

Analysis of rainfall and determination of annual maximum daily rainfall would enhance the management of water resources applications as well as the effective utilization of water resources. Probability and frequency analysis of rainfall data enables us to determine the expected rainfall at various chances. Such information can also be used to prevent floods and droughts, and applied to planning and designing of water resources related to engineering such as reservoir design, flood control work, soil and water conservation planning. The primary source of water for agricultural production for most of the world is rainfall. Useful for forecasting the floods to downstream towns and villages. Probability analysis of rainfall is necessary for solving various water management problems and to access the crop failure due to deficit or excess rainfall. Chi square, Anderson-Darling and Kolmogorov–Smirnov goodness of fit tests are used to judge the applicability of the distributions for modeling of the recorded rainfall data. The standard procedure for estimating the frequency of occurrence of hydrological event is frequency analysis. The objective of frequency analysis of hydrologic data is to relate the magnitude of extreme events to their frequency of occurrence using probability distributions. The extreme rainfall event involves the selection of a sequence of the maximum observations from the respective data series. Information of the amount, intensity and distribution of monthly or annual rainfall for the most important places in the world is generally available. Both flood and drought occurrence probabilities can be interpreted by using various statistical methods. Frequency analysis is used in hydrology as one such tool. In frequency analysis, the magnitude of extreme events is related to their frequency of occurrence through the use of probability distributions. Three main characteristics of rainfall are its amount, frequency and intensity, the values of which vary from place to place, day to day, month to month and also year to year. In general, three types of probability distributions (Normal, Log Normal and Gamma distribution) Walck (2007) are commonly used to determine the best fit probability distribution using the comparison of Chi-square, Anderson-Darling and Kolmogorov–Smirnov values. Therefore investigations were conducted in the past for study of the rainfall analysis by Suribabu et al. (2015), Olofintoye et al. (2009), Bhakar et al. (2006), Nyarko et al. (2002), Singh et al. (2001), Rizvi et al.(2001), Mohanty et al. (1999), Upadhaya et al. (1998), Hirose et al. (1994), Kulandaivelu et al. (1984), among others.

2. Methods and Material

Study area and collection of data

Licensed Under Creative Commons Attribution CC BY

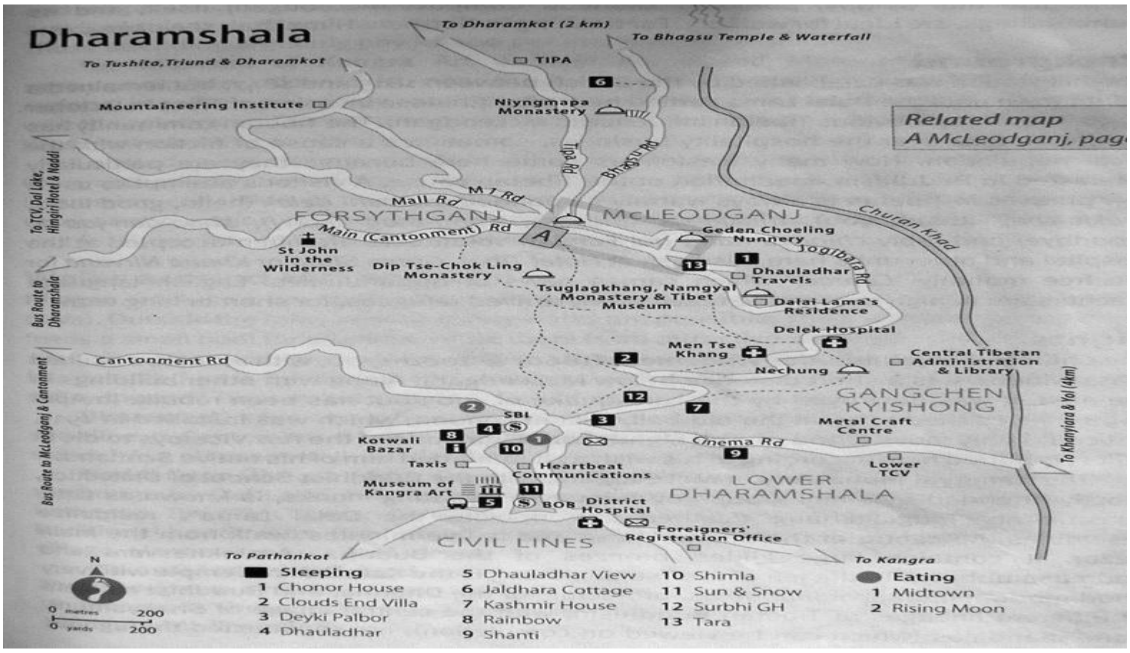

Figure 1: Area map of Dharamshala region

Daily rainfall data recorded at Dharamshala (Himachal Pradesh) for a period of 20 years (1992-2012) by India Meteorological Department, Shimla (H.P.), India were used for this study. The daily data, in a particular year, is converted to 2-7 days consecutive rainfall of corresponding previous days. The maximum amount of 1-day and 2 to 7 consecutive day's rainfall for each year was then taken for the analysis. Figure 1 shows the study area used in this research.

Statistical analysis of data

The statistical behavior of any hydrological series can be described on the basis of certain parameters, generally mean, variance, standard deviation, coefficient of variation and coefficient of skewness were taken as measures of variability of any hydrologic series. All these parameters have been used to describe the variability of rainfall in the present study.

Mean

Mean represents measures of central tendency. Arithmetic mean is given by

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i \qquad (1)$$

Where,

$\bar{X}$ = mean

X_i = variate

N = total number of observations

Standard deviation

This parameter, as a measure of variability is most adoptable to statistical analysis. It is computed by

$$S = \sqrt{\frac{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2}{(N-1)}} \qquad (2)$$

Where, S = standard deviation

Coefficient of variation

This is a dimensionless dispersion parameter usually expressed as percent. It is given by

$$C_v = \frac{S}{\bar{X}} \qquad (3)$$

Where, C_v = coefficient of variation

Coefficient of skewness

The coefficient of skewness was determined for each set of data from the formula given below:

$$C_s = \frac{N^2 M_3}{(N-1)(N-2)S^3} \qquad (4)$$

Where,

C_s = coefficient of skewness

$$M_3 = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^3 \qquad (5)$$

3. Indentations and Equations

Goodness of fit

Goodness of fit can be analysed by comparing the theoretical and sample values of the relative frequency of the cumulative frequency function of a probability distribution (Dabral et al. 2009). In case of the relative frequency function, the Chi-square test is used. The sample value of the relative frequency of interval *i* is calculated by the following equation:

$$f_s(x_i) = \frac{n_i}{n} \qquad (6)$$

The theoretical value of the relative probability function is

$P(X_i) = F(X_i) - F(X_{i-1})$

Chi Square

The Chi-square test statistic χ_c^2 is given by the equation below

Licensed Under Creative Commons Attribution CC BY

$$\chi_c^2 = \sum_{i=1}^{m} n\left[\frac{(f_s(X_i) - P(X_i))^2}{P(X_i)}\right] \quad (7)$$

Where, $n\, f_s\,(x_i) = n_i$ is the observed number of occurrences in interval i, $n\, P\,(x_i)$ is the corresponding expected number of occurrences in interval i

The χ^2 distribution functions are tabulated in many statistics texts. In the χ^2 test,

$v = m - p - 1$

Where, v = degree of freedom, m = number of intervals, p = number of parameters used in fitting the proposed distribution

A confidence level is chosen for the test, it is often express as 1-α, where 'α' is termed as the significant level. A typical value for the confidence level is 95 per cent. The null hypothesis for the test is that the proposed probability fits the data adequately. This hypothesis is rejected if the value of χ_c^2 is larger than a limiting value, $\chi^2_{v,\ 1-\alpha}$ (which is determined from the χ^2 distribution with ν degree of freedom at 5 % level of significance. Otherwise it was rejected.

Kolmogorov Smirnov

In statistics , the Kolmogorov Smirnov test (Walck 2007) is a non-parametric test of the equality of continuous, one-dimensional probability distributions that can be used to compare a sample with a reference probability distribution (one-sample K–S test), or to compare two samples (two-sample *K–S* test). The empirical distribution function F_n for n observations, X_i is defined as

$$F_n(x) = \frac{1}{N}\sum_{i=1}^{n} I\{x_i \le x\} \quad (8)$$

where, $I\{x_i \le x\}$ is the indicator function, equal to 1 if and equal to 0 otherwise. The Kolmogorov–Smirnov statistic for a given cumulative distribution function $F(x)$ is:

$T = sup_x\, |\, F^*(x) - S\,(x)\, |$

where, sup_x is the supre mum of the set of distances.

Anderson Darling

The Anderson-Darling Test (Walck 2007) will determine if a data set comes from a specified distribution, in our case, the normal distribution. The test makes use of the cumulative distribution function. The Anderson-Darling statistic is given by the following formula:

$$A^2 = \left[\frac{\sum_{i=1}^{n}(2i-1)(\ln(u_i) + \ln(1-u_{n+1-i}))}{n} - n\right] \quad (9)$$

where , $u_i = F(x_i)$

where, n = sample size, $F(X_i)$ = cumulative distribution function for the specified distribution and i = the i[th] sample when the data is sorted in ascending order

$AD = -n - S$

Summation term in the Anderson-Darling equation:

$$S = (2i - 1)[\ \ln\ F(X_i) + \ln(1- F(X_{n-i+1}))\ ] \quad (10)$$

Probability analysis using frequency factors

Chow (1988) has shown that many frequency analyses can be reduced to the form

$$X_T = X\ (\ 1 + C_V\ K_T\ ) \quad (11)$$

For Normal and Log Normal distribution, the frequency factor can be expressed by the following equation (Chow, 1988)

$$K_T = \frac{X_T - \mu}{\sigma} \quad (12)$$

This is the same as the standard normal variable z. The value of z corresponding to an exceedance of p ($p = 1/T$) can be calculated by finding the value of an intermediate variable w:

$$\text{Where,}\quad w = \left[\ln\left(\frac{1}{p^2}\right)\right]^{\frac{1}{2}} \quad (0 < p \le 0.5) \quad (13)$$

Then calculating z using the equation

$$z = w - \left[\frac{(2.515517 + 0.802853w + 0.010328w^2)}{(1 + 1.432788w + 0.189269w^2 + 0.001308w^3)}\right] \quad (14)$$

When $p > 0.5$, $1-p$ is substituted for p in equation (13) and the value of z is computed by equation (14) is given a negative sign. The frequency factor K_T for the normal distribution is equal to z, as mentioned above.

In case of Gamma distribution, frequency analysis was done by the method as described by Hann (1994) as given in Table-1

Table 1: Description of various probability distribution functions (Walck 2007)

Distribution	Probability density function
Normal	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
Log Normal	$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-\mu_y)}{2\sigma_y^2}\right)$
Gamma	$f(x) = \frac{\lambda^{\beta} x^{\beta-1} e^{-\lambda}}{T(\beta)}$

4. Figures and Tables

Statistical parameters of annual 1 day to 7 consecutive day's maximum rainfall

The statistical parameters of annual 1-day as well as 2 to 7 consecutive day's annual maximum rainfall are shown in Table 2. The mean value of one-day annual maximum rainfall at Dharamshala is found to be 142.9 mm with standard deviation and coefficient of variation of 54.8 and 51.34 respectively. The coefficient of skewness is 1.1. For 2 to 7 days consecutive annual maximum rainfall range values for mean, standard deviation, coefficient of variation and

Licensed Under Creative Commons Attribution CC BY

coefficient of skewness are 201 – 393.4 mm , 70.17 – 146.5, 41.65 – 30.47 and 0.726 – 1.593. It is observed that all distribution fitted function significantly. On the basis of these tests, the results will be helpful in soil and water conservation planning and design of small and medium hydraulic structures such as small dams, bridges, culverts drainage works, etc. A common use of rainfall data is in the assessment of probabilities or return periods of given rainfall at a given location. Such data can then be used in assessing flood discharges of given return period through modeling or some empirical system and can thus be applied in schemes of flood alleviation or forecasting.

Table 2: Statistical parameters of annual 1 day to 7 consecutive day's maximum rainfall

Parameters	1 day	2 days	3 days	4 days	5 days	6 days	7 days
Maximum (mm)	279.6	334.4	397.2	475.8	574.0	691.6	770.4
Minimum (mm)	84.1	104.8	128.7	156.9	203	218.4	240.9
Mean (mm)	142.9	201.0	241.4	287.6	325.5	364.4	393.4
Standard deviation (mm)	54.8	70.17	82.81	93.91	107.1	128.4	146.5
Coefficient of variation	51.34	41.65	37.18	34.56	32.77	31.47	30.47
Coefficient of skewness	1.1	0.726	0.886	0.785	1.12	1.279	1.593

Fitting of various probability distribution functions

One day annual maximum, 2 to 7 consecutive days annual maximum rainfall data in its original form was fitted to different probability distribution function i.e. Normal, Lognormal. As per the Chi-square value, Normal distribution function is found to be best fit function for 1 to 3 days annual maximum rainfall data. The computed Chi-square values for the probability distribution i.e. Normal, Log Normal and Gamma are found to be less than the critical value of Chi-square at 95% confidence level for 1 day as well as consecutive days maximum rainfall series. The statistical comparison by Chi-square test for goodness of fit clearly shows that Log Normal distribution is the best fitting representative function for rainfall frequency analysis in this region as shown in table 3.

Table 3: Chi-square values for different distribution

Consecutive days	Normal	Log Normal	Gamma
1 Day	1.20	4.83	2.08
2 Days	1.04	2.19	6.59
3 Days	3.67	1.86	1.53
4 Days	1.68	1.14	2.44
5 Days	3.78	1.23	7.12
6 Days	1.82	2.34	2.64
7 Days	6.35	3.72	5.09

The graphical plot as shown in figure 2 below shows Chi-square values for different distribution at Dharamshala region. Log Normal distribution gave minimum value of χ^2 for annual 1 day and 2 to 7 days consecutive maximum rainfall series at Dharamshala region. The statistical comparison by Chi-square test for goodness of fit clearly shows that Log Normal is the best fitting representative function for rainfall frequency analysis at Dharamshala region.

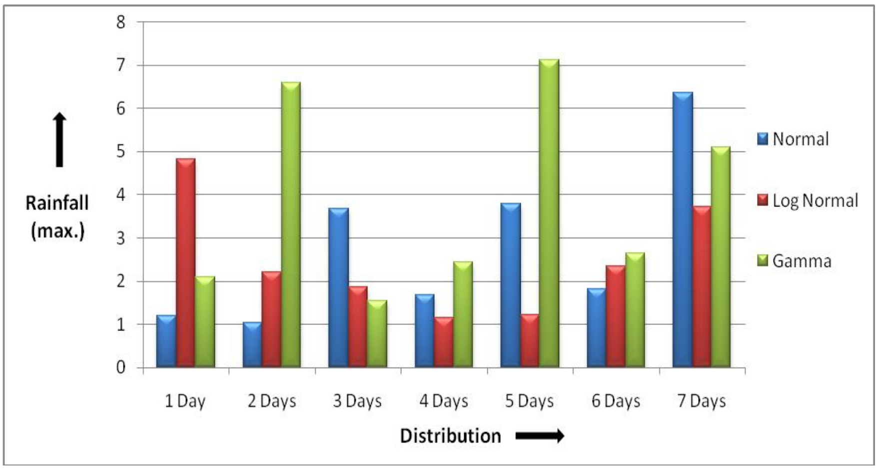

Figure 2: Graphical plot showing Chi-square values for different distribution

The data presented in Table 4 revealed that the computed K-S values for three probability distribution i.e. Normal, Log Normal and Gamma are found to be less than the critical value of K-S at 95% confidence level for 1 day as well as consecutive days maximum rainfall series. Gamma distribution gave minimum value of K-S for annual 1 day and 2 to 7 days consecutive maximum rainfall series at Dharamshala region. The statistical comparison by K-S test for goodness of fit clearly shows that is the best fitting representative function for rainfall frequency analysis at Dharamshala region.

Table 4: K-S values for different distribution

Consecutive days	Normal	Log Normal	Gamma
1 Day	1.05	2.16	1.82
2 Days	3.89	1.24	3.49
3 Days	2.16	3.96	2.16
4 Days	1.26	2.22	1.34
5 Days	2.08	1.32	2.82
6 Days	3.09	4.62	2.96
7 Days	4.28	5.16	3.72

The graphical plot as shown in figure 3 below shows K-S values for different distribution at Dharamshala region. Gamma distribution gave minimum value of K-S for annual

Licensed Under Creative Commons Attribution CC BY

1 day and 2 to 7 days consecutive maximum rainfall series at Dharamshala region. The statistical comparison by K-S test for goodness of fit clearly shows that Gamma is the best fitting representative function for rainfall frequency analysis at Shimla and Dharamshala region.

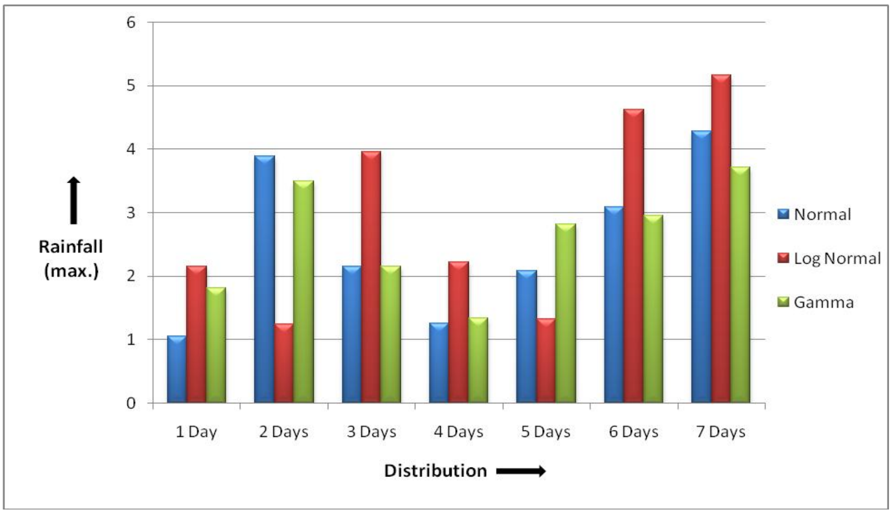

Figure 3: Graphical plot showing K-S values for different distribution

The data presented in Table 5 revealed that the computed AD values for three probability distribution i.e. Normal, Log Normal and Gamma are found to be less than the critical value of AD at 95% confidence level for 1 day as well as consecutive days maximum rainfall series. Log Normal distribution gave minimum value of AD for annual 1 day and 2 to 7 days consecutive maximum rainfall series at Dharamshala region. The statistical comparison by AD test for goodness of fit clearly shows that Log Normal is the best fitting representative function for rainfall frequency analysis at Dharamshala region.

The graphical plot as shown in figure 4 below shows AD values for different distribution at Dharamshala region. Log Normal distribution gave minimum value of AD for annual 1 day and 2 to 7 days consecutive maximum rainfall series at Dharamshala region. The statistical comparison by AD test for goodness of fit clearly shows that Log Normal is the best fitting representative function for rainfall frequency analysis at Dharamshala region.

Table 5: AD values for different distribution

Consecutive days	Normal	Log Normal	Gamma
1 Day	2.84	1.12	3.78
2 Days	3.67	2.43	2.92
3 Days	2.11	3.92	4.76
4 Days	1.42	1.78	1.47
5 Days	2.69	2.91	3.99
6 Days	4.78	3.24	4.89
7 Days	5.6	1.48	1.24

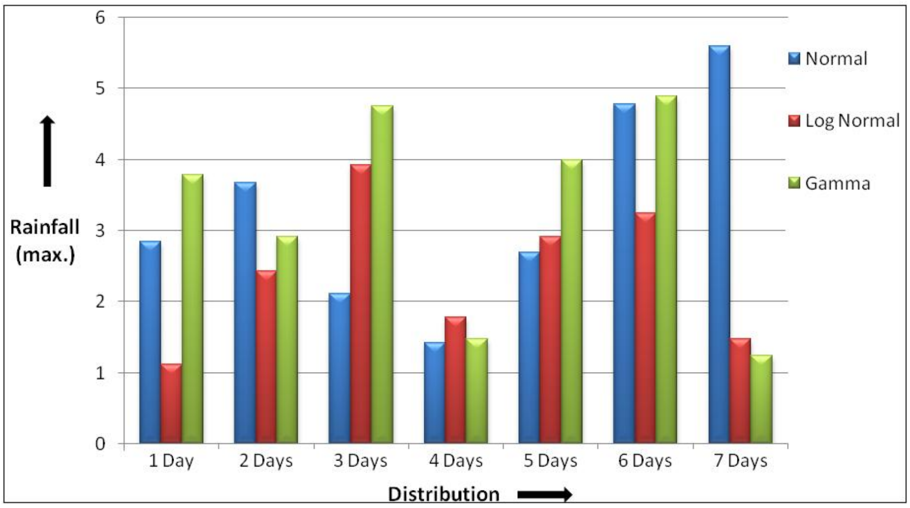

Figure 4: Graphical plot showing AD values for different distribution

Estimation of 1-day as well as 7 consecutive day's maximum rainfall for various return periods

Annual one day maximum rainfall and two to seven days consecutive days maximum rainfall corresponding to return period varying from 2 to 20 years are to be used by design engineers and hydrologists for the economic planning, design of small and medium hydrologic structures and determination of drainage coefficient for agricultural fields. For Dharamshala region, a maximum of 191.1 mm in 1 day, 262.7 mm in 2 days, 314.27 mm in 3 days, 370.24 mm in 4

Licensed Under Creative Commons Attribution CC BY

days, 419.7 mm in 5 days, 477.4 mm in 6 days and 522.32 mm in 7 days is expected to occur at Dharamshala, Himachal Pradesh every 2 years. For a recurrence interval of 20 years, the maximum rainfall expected in 1 day, 2, 3, 4, 5, 6 and 7 days is 277.7 mm, 373.6 mm, 445.11 mm, 518.62 mm, 589 mm, 680.3 mm and 753.79 mm respectively. Various probability distributions and transformations are applied to estimate one day and two to seven consecutive days annual maximum rainfall of various return periods. Three commonly used probability distributions (viz: Normal, Log Normal and Gamma distribution) are tested by comparing the Chi-square, Kolmogorov-Smirnov and Anderson Darling values. Table 6 gives the 1 day and consecutive days maximum rainfall for different return periods at Dharamshala region as determined by selected distribution.

Table 6: 1 day as well as consecutive days maximum rainfall for various return periods

S. No.	Return Period	1 Day	2 Days	3 Days	4 Days	5 Days	6 Days	7 Days
1	1.25	183.5	252.9	302.68	357.09	404.8	459.4	501.81
2	2	191.1	262.7	314.27	370.24	419.7	477.4	522.32
3	5	211.9	289.4	345.74	405.93	460.4	526.2	577.99
4	10	242.6	328.7	392.11	458.52	520.4	598.1	660.03
5	20	277.7	373.6	445.11	518.62	589	680.3	753.79

5. Conclusion

Important conclusions of the study are as follows:

1) The present study concluded that the data of twenty years (1992-2012) is sufficient to obtain annual maximum rainfall (mm) distribution of Dharamshala region. As per the Chi-square values, Normal distribution function is found to be best fit function for 1 to 3 days and Log Normal for 6 to 7 days annual maximum rainfall data. As per K-S values, Log Normal distribution function is found to be best fit function for 1 to 3 days and Gamma for 6 to 7 days annual maximum rainfall data. As per AD values, Log Normal distribution function is found to be best fit function for 1 to 3 days and for 6 to 7 days annual maximum rainfall data.
2) The study of frequency analysis will be used as a rough guide by engineers and hydrologists to prevent floods and droughts, and applied to planning and designing of water resources related to engineering such as reservoir design, flood control work. Analysis of rainfall and determination of annual maximum daily rainfall would enhance the management of water resources applications as well as the effective utilization of water resources. The results of this study would be useful for agricultural scientists, decision makers, policy planners and researchers in order to identify the areas where agricultural development and construction of drainage systems takes place.
3) The present study aims to evaluate the rainfall magnitude for different return periods and also to ascertain the type of probability distribution that best fits the rainfall data of Dharamshala (H.P.), India. From the above results, Gamma and Log Normal distributions are found to be best fit for the region. In order to achieve the objectives in present study, the following tests were performed.
4) On the basis of these tests, the results will be helpful in soil and water conservation planning and design of small and medium hydraulic structures such as small dams, bridges, culverts drainage works, etc. A common use of rainfall data is in the assessment of probabilities or return periods of given rainfall at a given location. Such data can then be used in assessing flood discharges of given return period through modeling or some empirical system and can thus be applied in schemes of flood alleviation or forecasting.
5) Frequency analysis usually involves the fitting of a theoretical frequency distribution using a selected fitting method, although empirical graphical methods can also be applied. The fitting of a particular distribution implies that the rainfall samples of annual maxima were drawn. For each distribution one can obtain the following: Estimation of parameters of the distribution, a table of rainfalls of specified exceedance probabilities or return periods with confidence limits, results of goodness of fit tests, a graphical plot of the data fitted to the distribution.
6) The mean value of one-day annual maximum rainfall at Dharamshala is found to be 142.9 mm with standard deviation and coefficient of variation of 54.8 and 51.34 respectively. The coefficient of skewness is 1.1. For 2 to 7 days consecutive annual maximum rainfall range values for mean, standard deviation, coefficient of variation and coefficient of skewness are 201 – 393.4 mm , 70.17 – 146.5, 41.65 – 30.47 and 0.726 – 1.593. A maximum of 191.1 mm in 1 day, 262.7 mm in 2 days, 314.27 mm in 3 days, 370.24 mm in 4 days, 419.7 mm in 5 days, 477.4 mm in 6 days and 522.32 mm in 7 days is expected to occur at Dharamshala, Himachal Pradesh every 2 years. For a recurrence interval of 20 years, the maximum rainfall expected in 1 day, 2, 3, 4, 5, 6 and 7 days is 277.7 mm, 373.6 mm, 445.11 mm, 518.62 mm, 589 mm, 680.3 mm and 753.79 mm respectively.

References

[1] Suribabu, C. R., Sudarsan, J. S., Ramanan, S. R. *"A study on consecutive day maximum rainfall for disaster mitigation in Tiruchirapalli City, India."* Jordan Journal of Civil Engineering, *9(2)*, 2015, *197–202*.
[2] Olofintoye, O. O, Sule, B. F. and Salami, A. W. *"Best–fit Probability distribution model for peak daily rainfall of selected Cities in Nigeria."* New York Science Journal, *2(3)*, 2009, *1554-2000*.
[3] Bhakar, S. R., Bansal, A. N., Chhajed, N., and Purohit, R. C. *"Frequency analysis of consecutive day's maximum rainfall at Banswara, Rajasthan, India."*ARPN Journal of Engineering and Applied Sciences, *1(3)*, 2006, *64-67*.
[4] Nyarko, K. B. *"Application of a rational model in GIS for flood risk Assessment in Accra, Ghana."* Journal of Spatial Hydrology, *2(1)*, 2002, *12-14*.

Licensed Under Creative Commons Attribution CC BY

[5] Singh, R. K. *"Probability analysis for prediction of annual maximum daily rainfall of Eastern Himalaya (Sikkim Mid Hills)."* Indian Journal of Soil conservation, *29(3)*, 2001, *263-265*.
[6] Rizvi, R. H., Singh, R., Yadav, R. S., Tewari, R. K., Dadhwal, K. S. and Solanki, K. R. *"Probability analysis of annual maximum daily rainfall for Bundelkhand region of Uttar Pradesh."* Indian Journal of Soil Conservation, *29(3)*, 2001, *259-262*.
[7] Mohanty, S., Marathe, R. A. *"Probability analysis of annual maximum daily rainfall for Amravati*." Indian Journal of Soil Conservation, *43(1)*, 1999, *15-17*.
[8] Upadhaya, A., and Singh, S. R. *"Estimation of consecutive days Maximum Rainfall by Various Methods and their Comparison."* Indian Journal of Soil Conservation, *26(2)*, 1998, *1993-2001*.
[9] Hirose, H. *"Parameter Estimation in the Extreme-Value distributions using the Continuation Method."* Information Processing Society of Japan, *35*, 1994, *9*.
[10] Kulandaivelu, R. *"Probability analysis of rainfall and evolving cropping system for Coimbatore."* Mausam, *5(3)*, 1984, *257-258.*

Licensed Under Creative Commons Attribution CC BY