RESEARCH PAPER | International Journal of Agricultural Engineering | Volume 13 | Issue 1 | April, 2020 | 67-73
ISSN–0974–2662 ■ Visit us : *www.researchjournal.co.in* ■ **DOI: 10.15740/HAS/IJAE/13.1/67-73**

Frequency analysis of daily rainfall data of Udaipur district

Harshvardhan Baghel

Received : 09.01.2020; **Revised :** 17.02.2020; **Accepted :** 04.03.2020

Author for Correspondence :
Harshvardhan Baghel
Office of Watershed Development and Soil Conservation, Panchayat Samiti, Pindwara, **Sirohi (Rajasthan) India**
Email : harshvicky11@gmail.com; pia.pin2014@gmail.com

ABSTRACT : Rainfall is a prime input for various engineering design such as hydraulic structures, water conservation 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. The present study comprises statistical analysis *i.e.* frequency analysis of daily maximum rainfall data of Udaipur district. The daily rainfall data for a period of 56 years is collected to evaluate designed value of rainfall using probability distribution models. The different probability distributions *viz.,* Gamble's extreme value type I, Logpearson type III, Lognormal, Normal, Exponential, Pearson type III and Gamma distribution were used to evaluate maximum daily rainfall. Kolmogorov-Smirnov and Chi-square tests were used to examine the goodness of fit of the probability distributions. Results showed that Lognormal and Gumbel distributions, found to be having least critical values for both the tests, hence consider as the best fit distribution for given sample rainfall data. Also maximum daily expected value of rainfall for various return periods were evaluated using all distribution model under consideration.

KEY WORDS : Chi-square test, Kolmogorov-smirnov test, Probability distribution model

HOW TO CITE THIS PAPER : Baghel, Harshvardhan (2020). Frequency analysis of daily rainfall data of Udaipur district. *Internat. J. Agric. Engg.,* **13**(1) : 67-73, **DOI: 10.15740/HAS/IJAE/13.1/67-73.**

Analysis of daily maximum rainfall of different return periods is a basic tool for safe and economical planning and design of small dams, bridges, culverts, irrigation and drainage system etc. Though the nature of rainfall is erratic and varies with time and space, yet it is possible to predict design rainfall more accurately for certain return periods using various probability distributions (Upadhyay and Singh, 1998). Design Engineers and Hydrologists require one day maximum rainfall at different frequencies or return periods for appropriate planning and design of small and medium hydraulic structures like small dams, bridges, culverts, etc. (Agarwal *et al.,* 1998). Probability analysis can be used for predicting the occurrence of future events of rainfall from the available data with the help of statistical methods (Kumar, 1989). Al-suhili and Khanbilvardi (2014) analyzed monthly rainfall data in Sulaimania region, north Iraq for the period (1984-2010). The distributions models fitted are of Normal, Log-normal, Weibull, Exponential and Two parameters Gamma type. The Kolmogorov-Smirnov test was used to evaluate the goodness of fit. The Gamma, Exponential and Weibull distributions were found as the best fit, Arvind *et al.* (2017) Collected daily rainfall data for a period of 30 years from the raingauge station located closely in Trichy district. Collected historic data are used to understand normal rainfall, deficit rainfall. From the calculated results, the rainfall pattern is found to be erratic. Anaya Kalita *et al.* (2017) worked on frequency analysis of daily rainfall data of 24 years to determine the annual one day maximum rainfall and discharge of Ukiam (Brahmaputra River). Weibull's plotting position Gumbel, Log Pearson and Log normal probability distribution functions were fitted. For determination of goodness of fit chi-square test was carried out. The results reveals that the Log Pearson and Log Normal were the best fit probability distribution. Esberto (2018) determined the best fit frequency distribution of rainfall patterns for event forecasting in order to address potential disasters using 60 Probability Distribution Functions (PDF). Rainfall data were analyzed using Chi-Square and K-S goodness of fit tests. Amin *et al.* (2016) analyzed to find the best fit probability distribution of annual maximum rainfall based on a twenty four hour sample in the northern regions of Pakistan using four probability distributions: normal, log-normal, log Pearson type-III and Gumbel max. Based on the scores of goodness of fit tests, the normal distribution was found to be the best fit probability distribution at the Mardan rainfall gauging station. The log-Pearson type-III distribution was found to be the best-fit probability distribution at the rest of the rainfall gauging stations. This project is an effort to summarize the rainfall features for the Udaipur district. The total rainfall received in a given period at a location is highly variable from one year to another. The variability depends on the type of climate and the length of the considered period. The statistical inferences found in this study are important for designing optimum flood control facilities. Basically frequency analysis of rainfall is used for different purposes as mentioned above.

■ METHODOLOGY

Udaipur district is situated between $23^0 40'$ and $25^0 30'$ north latitude and $73^0 0'$ and $74^0 35'$ east longitude. It is located in the south eastern part of Rajasthan and lies in Aravali ranges. The district is having 1, 89,746 ha area surrounded by hills (Google map, cited on 25 May, 2019). 56 years of daily mean rainfall data from 12 raingauge stations of Udaipur district have collected from 'Rainfall Profile of Udaipur' manual published by Indian Meteorological Department Jaipur (2014).

$\overline{X}$ is the arithmetic Mean, X_i is Variate, N is the total number of observations, S is Standard Deviation, C_v is the coefficient of Variation and C_s is the co-efficient of skewness.

Table A : Formula of statistical parameters

Sr. No.	Parameter name	Formula
1.	Arithmetic mean	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{N}$
2.	Standard deviation	$S = \sqrt{\frac{\sum_{i=1}^{N} X_i}{N-1}}$
3.	Co-efficient of variation	$C_v = \frac{S}{\overline{X}}$
4.	Co-efficient of skewness	$C_s = \frac{N \sum (X-\overline{X})^3}{(N-1)(N-2)S^3}$

Tests for goodness of fit (Verification of sample population):

The goodness of fit of a statistical model describes how well it fits a set of observations (Ghanshyamdas, 2014). Measures of goodness of fit typically summarize the discrepancy between observed values and the values expected under the model in question.

In stochastic hydrology, there are two ways whether or not a particular distribution adequately fits a set of observations.

– Compare observed relative frequency with theoretical relative frequency.

– Using probability papers

Two tests were used to compare observed relative frequency with theoretical relative frequency (Ghanshyamdas, 2014).

Chi-square test:

The chi-square test is used to determine whether there is a significant difference between the expected and the observed frequencies in one or more categories

$$\chi_c^2 = \frac{\sum_{i=1}^{N}(N_i - E_i)^2}{E_i} \qquad (1)$$

where N is the total number of observations, N_i is the observed relative frequencies, and E_i is the theoretical or probable relative frequencies. If $\chi_c^2 = 0$, it indicates that observed and theoretical frequencies agree exactly while if $\chi_c^2 > 0$, they do not agree exactly. The hypothesis that the data follows a specific distribution is accepted if,

$$\chi^2_{data} \le \chi^2_{\alpha-1,k-p-1} \qquad (2)$$

where α is the significance level and K-P-1 is the degree of freedom. Test is carried out at 10% significance level. Critical values of chi-square test for a particular degree of freedom and at particular significance level can be obtained from Chi-square distribution table.

Kolmogorov-smirnov test:

In statistics, the Kolmogorov–Smirnov test is a nonparametric test of the equality of continuous (or discontinuous), one dimensional probability distributions that can be used to compare a sample with a reference probability distribution (like Chi-square Test), This is the alternative to Chi-square test. The absolute difference between theoretical cumulative probability F(x) and calculated cumulative probability P(x) is calculated. The Kolmogorov-smirnov test statistics Δ is the maximum of this absolute difference calculated in step 4.

$$\Delta = \text{Maximum}\,|P(x) - F(x)| \qquad (3)$$

The critical value of Kolmogorov-smirnov test statistics Δ_α is obtained from the Kolmogorov-smirnov table for 10% significance level. If $\Delta<\Delta\alpha$, accept the hypothesis. For sample size more than 50, use following formula for critical values of Kolmogorv-smirnov test statistics.

$$\Delta_\alpha = \frac{1.22}{\sqrt{N}\,(\alpha = 10\%)} \qquad (4)$$

Table B : Plotting position parameters for probability plotting			
Sr. No.	Probability distribution	Parameter Plotted on Abscissa	Parameter plotted on ordinate
1.	Normal distribution	Z (Normal Z Value)	(x) Rainfall in mm
2.	Log normal distribution	Z (Normal Z Value)	(Log x) Rainfall in mm
3.	Gumbel's distribution	Y_t (Reduced vaiate)	(x) Rainfall in mm
4.	Log Pearson Type III distribution	K_t (Frequency Factor)	(Log x) Rainfall in mm
5.	Gamma distribution	Γ^{-1}(p)(Gamma Parameter)	(x) Rainfall in mm
6.	Exponential distribution	-Log [1-f(x)]	(x) Rainfall in mm
7.	Pearson Type III Distribution	K_t (Frequency Factor)	(x) Rainfall in mm

Probability plot method:

Frequency distribution models:

Gumbel's extreme value distribution model:

Gumbel found that the probability of occurrence of an event, equal or larger than a value is given by the equation,

$$P(X \geq x_0) = 1 - e^{-e^{-y}} \qquad (5)$$

$$y_t = -\left[\ln \ln \frac{T}{T-1}\right] \qquad (6)$$

$$X_T = \bar{X} + K\sigma_{n-1} \qquad (7)$$

For N=56 the values for $\bar{y}_n$ and σ_n are 0.551 and 1.1696, respectively from standard tables (Ghanshyamdas, 2014) .

Log-Pearson type III distribution:

$$z = \log x \qquad (8)$$

For any recurrence interval T above equation can be expressed as

$$Z_t = \log x_t \qquad (9)$$

Applying general equation chow, Z_T data series can be expressed a

$$z_T = \bar{z} + K_f \sigma_z \qquad (10)$$

where, K_f is the frequency factor, c_z is the co-efficient of skewness, $\bar{z}$ is the mean of the representative variate sample z, σ_z is the standard deviation of the representative variate sample z. value of K_f can be determined by using the standard table for a specific value of c_z and recurrence interval T.

Log normal probability distribution method:

The flood or rainfall of any return period which follows the log normal probability law is computed from:

$$Q_T = \bar{Q} + K\sigma_n \qquad (11)$$

where K is log normal frequency factor. A function of skewness co-efficient, given by

$$C_S = 3C_v + C_v^3 \qquad (12)$$

where C_v is a co-efficient of variation and given by

$$C_v = \frac{\sigma}{Q} \qquad (13)$$

The value of K can be determined from the normal probability table.

Normal distribution:

It is also a most widely used method in extreme value distributions.

$$X_T = \bar{X} = K_T\sigma \qquad (14)$$

$$K_T = Z = \frac{X_T - \bar{X}}{\sigma} \qquad (15)$$

$$K_T = w - \frac{2.515517 + 0.80285w + 0.010328w^2}{1 + 1.432788w + 0.189269w^2 + 0.001308w^3} \qquad (16)$$

Gamma distribution:

Gamma distribution – a distribution of sum of b independent and identical exponentially distributed random variables.

$$f(x) = \frac{\lambda^{\beta}(x-\varepsilon)^{\xi-1}\, e^{-\lambda(x-\varepsilon)}}{\Gamma(\beta)} \quad (17)$$

Γ = **Gamma function**

$$\Gamma(\cdot) = \int_0^{\alpha} t^{n-1} e^{-t}\, dt$$

Pearson type III:

Named after the statistician Pearson, it is also called three-parameter gamma distribution. A lower bound is introduced through the third parameter (e).

$$f(x) = \frac{\lambda^{\beta}(x-\varepsilon)^{\xi-1}\, e^{-\lambda(x-\varepsilon)}}{\Gamma(\beta)} \quad (18)$$

Exponential distribution:

In hydrology, the inter arrival time (time between stochastic hydrologic events) is described by exponential distribution.

$$f(x) = \lambda e^{-\lambda\lambda} \quad x \geq 0,\ \lambda = \frac{1}{\bar{x}}$$

$$\text{Variance} = 1/\lambda^2$$

■ RESULTS AND DISCUSSION

Probability plot results:

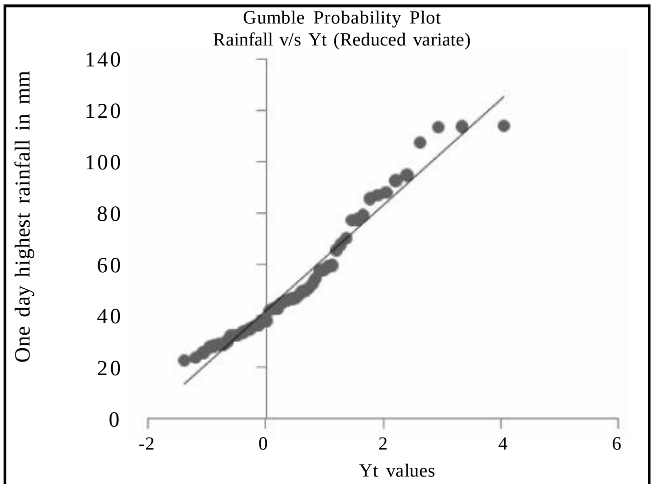

Fig. 1 : Gumble probability plot

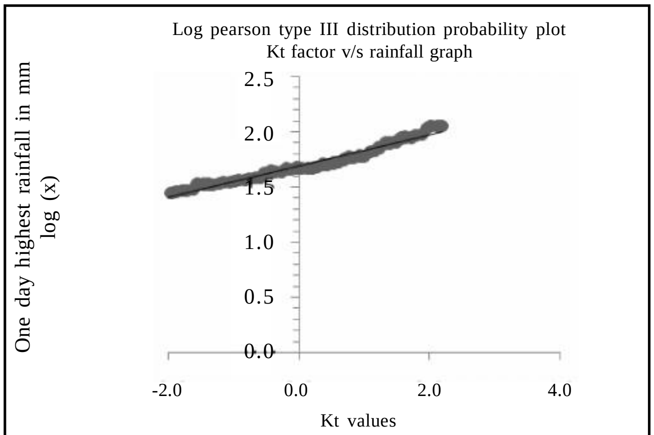

Fig. 2 : Log pearson type III probability plot

Table 1 : Goodness of fit result summary

Sr. No.	Distribution Model	Test Performed	Calculated values for χ_c^2 and KS test	Degree of freedom	Critical values at 10 % significance level	Result
1.	Gumbel's distribution	Chi-square Test	9.406	7	12.02	Accepted
		Kolmoorov-Smirnov Test	0.092		0.163	Accepted
2.	Log-Pearson Type-III	Chi-square Test	22.793	6	10.64	Rejected
	distribution	Kolmoorov-Smirnov Test	0.175		0.163	Rejected
3.	Normal distribution	Chi-square Test	20.851	7	12.02	Rejected
		Kolmoorov-Smirnov Test	0.159		0.163	Accepted
4.	Lognormal distribution	Chi-square Test	8.444	6	10.64	Accepted
		Kolmoorov-Smirnov Test	0.082		0.163	Accepted
5.	Exponential distribution	Chi-square Test	48.331	8	13.362	Rejected
		Kolmoorov-Smirnov Test	0.338		0.163	Rejected
6.	Pearson-III distribution	Chi-square Test	54.742	6	10.64	Rejected
		Kolmoorov-Smirnov Test	0.248		0.163	Rejected
7.	Gamma distribution	Chi-square Test	10.163	7	12.02	Accepted
		Kolmoorov-Smirnov Test	0.098		0.163	Accepted

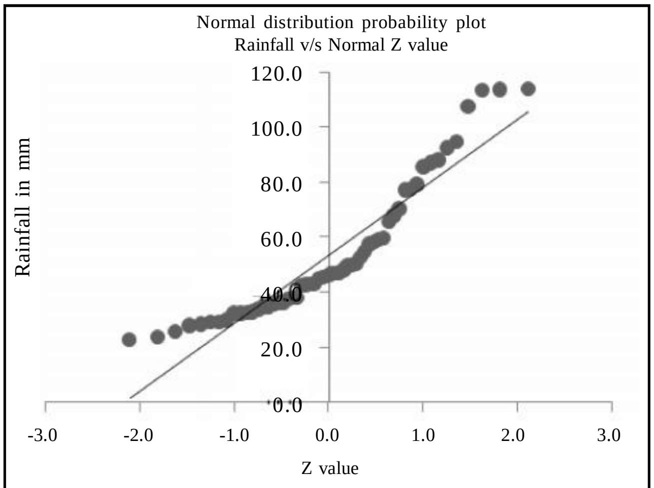

Fig. 3 : Normal probability plot

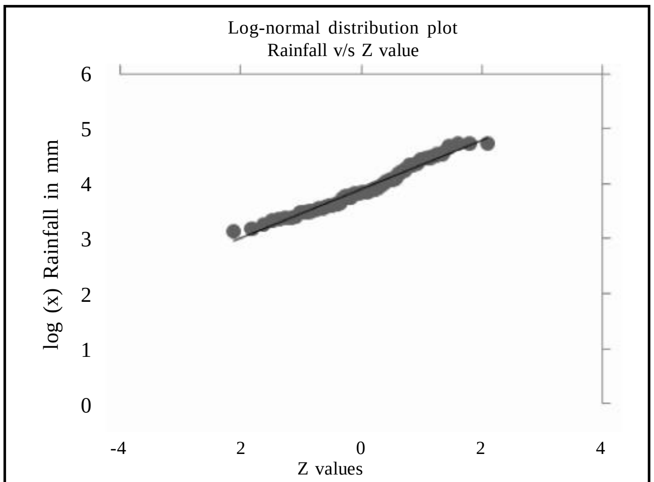

Fig. 4 : Lognormal probability plot

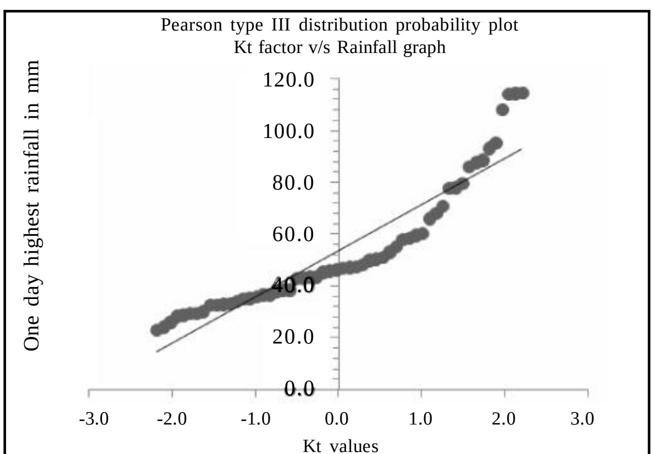

Fig. 5 : Pearson type III probability plot

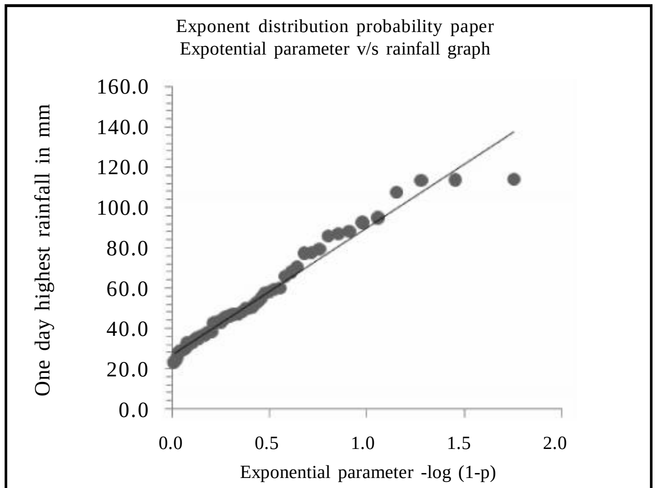

Fig. 6 : Exponential probability plot

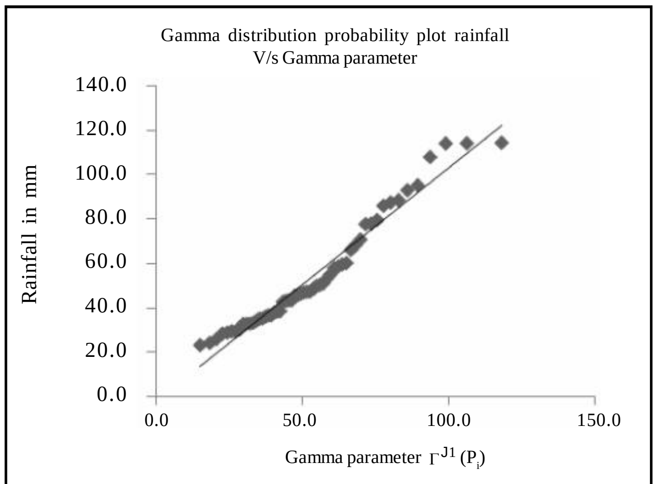

Fig. 7 : Gamma probability plot

Probability plot result summary:

Table 2 : Probability plot result summary

Sr. No.	Probability Plot	Correlation co-efficient	Result
1.	Gumbel's distribution	0.981	Accepted
2.	Logpearson type III distribution	0.984	Accepted
3.	Normal distribution	0.942	Accepted
4.	Log-normal distribution	0.986	Accepted
5.	Exponential distribution	0.984	Accepted
6.	Pearson type III distribution	0.977	Accepted
7.	Gamma distribution	0.980	Accepted

Magnitude of daily rainfall (mm) for various distribution models:

Table 3 : Magnitude of designed value of daily rainfall

Distribution model	Return period in years									
	5	10	25	50	100	200	300	400	500	1000
Gumbel distribution	73.85	89.77	109.88	124.80	139.61	154.37	162.99	169.10	173.84	188.55
Log-Pearson Type-III distribution	69.40	86.03	109.52	128.96	150.09	173.21	180.03	187.12	188.73	235.95
Normal distribution	74.60	85.52	97.17	104.69	111.45	117.64	121.04	123.37	125.14	130.40
Lognormal distribution	70.16	84.74	103.63	118.01	132.65	147.62	156.57	163.00	168.05	184.04
Exponential distribution	86.53	123.79	173.06	210.32	247.59	284.86	306.65	322.12	334.12	371.38
Pearson-III distribution	72.53	86.97	104.41	116.79	128.70	140.28	143.41	146.54	149.66	165.29
Gamma distribution	72.74	86.95	103.95	115.96	127.48	138.62	144.99	149.46	152.89	163.42

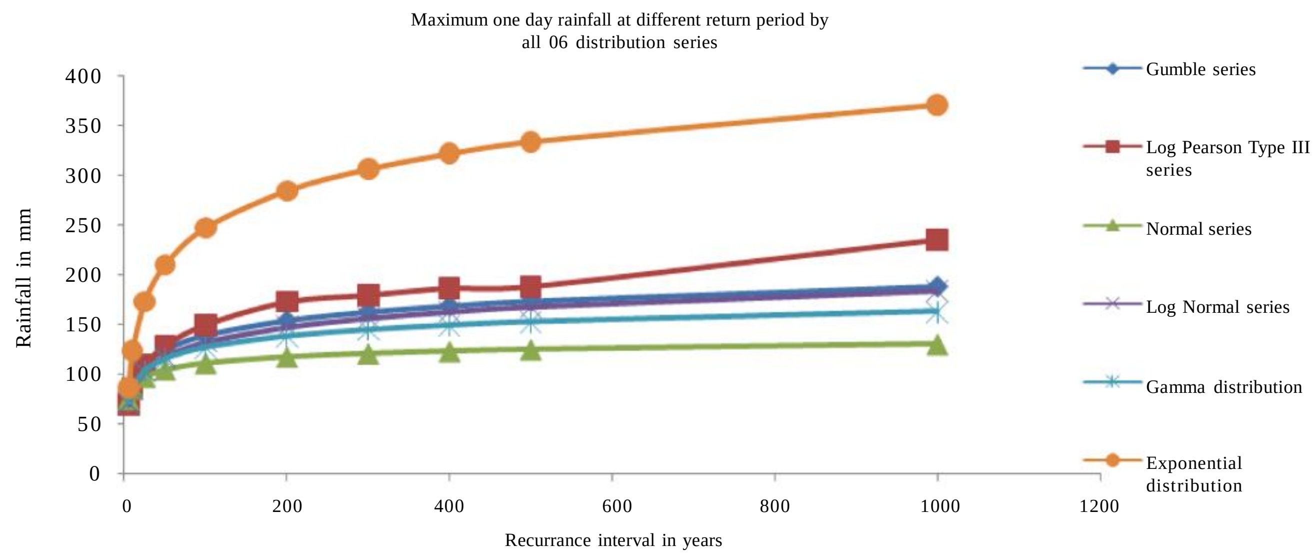

Fig. 8 : Comparison of different Probability distribution model of annual maximum daily rainfall

Conclusion:

56 years of daily rainfall data is taken from the IMD manual published in 2014. For the series of daily rainfall data, annual maximum daily rainfall data is arranged. The seven probability distributions were subjected to test from two goodness of fit tests (Kolmogorov-smirnov test and Chi-squared test). Further sample data is also tested by probability plotting *i.e.* plotting sample data with distribution parameter and calculate correlation coefficient. The purpose of the study was to find the best-fit probability distributions for district Udaipur for designing various hydraulic structures. The maximum values of expected rainfall or rainfall estimates calculated using a probability distribution that does not provide the best-fit may yield values that are higher or lower than the actual values. These calculations may be used to influence decisions relating to local economics and hydrologic safety systems.

Both the tests were performed at 10% significance level. Out of 07 models 04 models have passed in one or more tests. The Log-normal distribution and Gumbel distribution provided the best-fit probability distribution with the least score for both the test. The expected values of designed rainfall or rainfall estimates calculated using the best-fit probability distributions at the rainfall gauging stations might be used by design engineers to safely and feasibly design hydrologic projects.

■ REFERENCES

Amin, T.A., Rizwan, M. and Alazba, A.A. (2016). A best-fit probability distribution for the estimation of rainfall in northern regions of Pakistan. *Open Life Sci.*, **11** : 432-440.

Arvind, G., Kumar, A.P., Girishkarthi, S. and Suribabu, C.R. (2017). Statistical analysis of 30 Years rainfall data: A case study. *IOP Conf. Series: Earth & Environ. Sci.*, **80**(2017):01-09.

Esberto, M.D.P. (2018). Probability distribution fitting of rainfall patterns in Philippine regions for effective risk management. *Environ. & Ecol. Res.*, **6**(3): 178-186.

Ghanshyamdas (2014). Hydrology and Soil Conservation Engineering:Including Watershed Management, PHI Learning Private Limited.99-123p.

Kalita, A., Bormudoi, A. and Saikia, M.D. (2017). Probability distribution of rainfall and discharge of Kulsi River Basin. *Internat. J. Engg. & Adv. Technol. (IJEAT)*, **6**(4): 31-37.

Kumar, A. (1989). Probability analysis for prediction of annual maximum daily rainfall for Pantnagar. *Indian J. Soil Conservation*, **27**(2):171-173.

Latitude and longitude of Udaipur. (https://www.google.co.in/maps/place/Udaipur,+Rajasthan/@24.5873424,73.6407606,23152m/data=!3m1!1e3!4m2!3m1!1s0x3967e56550a14411:0xdbd8c28455b868b0!6m1!1e1).

Upadhaya, A. and Singh, S.R. (1998). Estimation of consecutive day's maximum rainfall by various methods and their comparison. *Indian J. Soil Conservation*, **26**(3):193-200.

13th Year of Excellence