MAUSAM, **72**, 2 (April 2021), 359-372

551.577.36 (540.27)

Annual maximum rainfall prediction using frequency analysis for Roorkee, Uttarakhand, India

LOVEPREET KAUR, ANVESHA, MANISH KUMAR#, SUMAN LATA VERMA and PRAVENDRA KUMAR

College of Technology, GBPUA&T, Pantnagar, Uttarakhand – 263145, India

(*Received 20 February 2020, Accepted 4 December 2020*)

#e mail : manishcae2k11@gmaill.com

सार – शुष्क क्षेत्रों में वर्षा, अल्प और महत्वपूर्ण जल वैज्ञानिक परिवर्ती है। जनसंख्या वृद्धि, आर्थिक विकास और शहरीकरण के कारण, इन क्षेत्रों में पानी की मांग दिन-प्रतिदिन बढ़ रही है और सभी उपलब्ध संसाधनों के साथ, जल प्रबंधन और अधिक महत्वपूर्ण हो जाता है। इसलिए, वर्षा संभाव्यता विश्लेषण कई जल प्रबंधन मुद्दों पर ध्यान देने और कम या अधिक वर्षा के कारण फसल की क्षति को रोकने के लिए आवश्यक है। किसानों के लिए आर्थिक रूप से बेहतर व्यवस्था हेतु वर्षा का वैज्ञानिक पूर्वानुमान और फसल नियोजन महत्वपूर्ण साबित हो सकता है। वर्षा के विभिन्न आँकड़ों की आवृत्ति का विश्लेषण करने का प्रयास किया गया। आवृत्ति विश्लेषण प्रत्याशित वर्षा को निर्धारित करने में मदद करता है। वर्षा की आवृत्ति का विश्लेषण करने में एक महत्वपूर्ण कदम वर्षा की जांच करने के लिए वर्षा की मात्रा का पता लगाना है। अधिकतम वार्षिक वर्षा के पूर्वानुमान के लिए Gumbel, Log Normal, Log Pearson Type III और Van Te Chow वितरण का उपयोग किया गया। 20 वर्ष (1991-2010) के वार्षिक वर्षा के आंकड़ों को नेशनल इंस्टीट्यूट ऑफ हाइड्रोलॉजी (NIH), रुड़की, भारत से एकत्र किया गया। इस शोध में अधिकतम वार्षिक वर्षा के लिए दैनिक, दो दिन लगातार, तीन दिन लगातार, विभिन्न सैद्धांतिक संभाव्यता वितरण के अनुरूप; और अधिकतम वार्षिक वर्षा के पूर्वानुमान के लिए दैनिक, दो दिन लगातार, तीन दिन लगातार सर्वोत्तम संभाव्यता वितरण का चयन करने के लिए कार्य किया गया। प्रेक्षित मानों के साथ अपेक्षित मानों की तुलना करके फिट काई-स्क्वेयर, प्रतिशत निरपेक्ष विचलन, इंटीग्रल स्क्वेयर एरर का निर्धारण किया गया। परिणामों से पता चला है कि लगातार एक, दो और तीन दिनों के लिए रुड़की के वार्षिक अधिकतम वर्षा मानों का पूर्वानुमान करने के लिए Gumbel वितरण सबसे उपयुक्त है। लगातार एक, दो और तीन दिनों के लिए दूसरा सबसे अच्छा वितरण Van Te Chow रहा।

ABSTRACT. Rain is a meagre and important hydrological variable in dryland areas. Due to population growth, economic development and urbanization, the demand for water in these areas is increasing day by day and with all available resources, water management becomes more and more crucial. Therefore, rain probability analysis is needed to address a number of water management issues and to prevent crop damage due to deficits or excess rainfall. Scientific rainfall forecasting and crop planning can prove to be an important tool in the hands of farmers for better economic returns. An attempt was made to analyze the frequency of different rainfall data. Frequency analysis helps in determining the expected rainfall. An important step in analyzing the frequency of rainfall is selecting a suitable distribution to represent the depth of the rainfall to investigate the precipitation. The distributions namely, Gumbel, Log Normal, Log Pearson Type III and Van Te Chow were used for the prediction of the maximum annual rainfall. Annual rainfall data for 20 years (1991-2010) were collected from the National Institute of Hydrology (NIH), Roorkee, India. This research attempts to fit several theoretical probability distributions to the maximum annual rainfall for daily, two consecutive days, three consecutive days; and to select the best probability distribution for the prediction of maximum annual rainfall for daily, two consecutive days, three consecutive days.For determination of goodness of fit chi-square, percentage absolute deviation, integral square errorwere carried out by comparing the expected values with the observed values. The results found showed that the Gumbel distribution emerged to be the best fit for the prediction of annual maximum rainfall values of Roorkee for one, two and three consecutive days. The second best fit was Ven Te Chow distribution for one, two and three consecutive days.

Key words – Gumbel, Log Pearson Type III, Log normal, Van TeChow, Frequency analysis, Rainfall prediction.

1. Introduction

Water is precious for survival of human and life and for the growth and development of plants. Whether it comes from rainfall or irrigation, water means life in agriculture. The term precipitation denotes all forms of water that reach the earth from the atmosphere. The usual forms are rainfall, snowfall, hails, frosts and dew of all these only the first two contribute the significant amount of water. Rainfall is the predominant form of precipitation that has a direct or indirect effect on agriculture. Most of the farmers have to deal with the problem of uncertain and variable rainfall and the effect it has on agriculture production. Rainfall is a serious risk factor in farming and farmers are always vulnerable to yield losses caused by extreme climate fluctuations. India receives an annual average rainfall of 852 mm for the whole country which would be 1200 mm for high rainfall years. The total precipitation is around 4000 billion cubic meters on the land surface. The knowledge of rainfall and its distribution throughout the year is also important for better crop planning, irrigation and drainage of crops. Intensity and distribution of rainfall may be used for design of hydrological structures, planning of soil conservation and flood programs.

Analysis of historic weather data helps to develop and modify management practices to increase agricultural production (Pearson, 1930), but there is no accurate method available to make seasonal, weekly or annual rainfall forecast. However, the hydrological frequency analysis can be used as a tool for predicting the occurrence of future events from the available data with the help of statistical methods (Bhakar *et al*., 2006). The determination of probability distribution of rainfall is of fundamental importance in many water resource design problems. Based on assumed distribution, it is possible to make probability statements of future rainfall of various magnitudes (Agarwal *et al*., 1988 and Dabral *et al*., 2009). Although precipitation is irregular and varies with time and space, it is generally possible to predict return periods using various probability distributions (Upadhaya and Singh, 1998). Therefore, an analysis of the probability of precipitation is necessary to solve various water management problems and to access crop failure due to a deficit or excess of precipitation. Scientific forecasting of rains and analytical planning of crops can be an important tool in the hands of farmers for better economic returns (Prajapati *et al*., 2002 and Panigrahi *et al*., 2001). Frequency analysis of rainfall data has been attempted for different return periods (Bhakar *et al*., 2006; Nemichandrappa *et al*., 2010 and Manikandan *et al*., 2011). The probability and frequency analysis of the pluviometric data allows us to determine the expected pluviometry at different chances. The most commonly used probability distribution functions for estimating the frequency of precipitation are the normal, log-normal, log-Pearson type III and Gumbel distributions (Chakraborty *et al*., 2012; Dingre *et al*., 2006 and Kumar *et al*., 1982;). There is no widely accepted procedure for forecasting one day, two consecutive days, three consecutive days of maximum precipitation. In the present study, an attempt was made to determine the statistical parameters and the maximum annual precipitation of one day, two consecutive days, three consecutive days using various probability levels using four distribution functions of probability, namely, normal, log-normal, log-Pearson distribution of type III and Gumbel and to select the best probability distribution system.

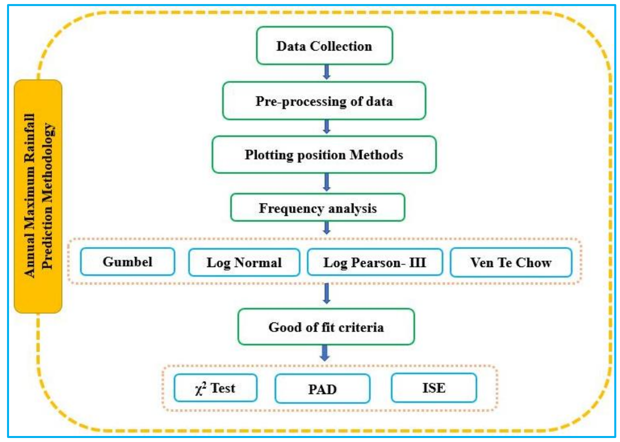

Fig. 1. A flowchart methodology for annual maximum prediction using frequency analysis for study area

2. Data sets and methodology

The annual rainfall data for 20 years (1991-2010) were collected from the National Institute of Hydrology (NIH), Roorkee which is located at 29.8667° N Latitude and 77.8833° E Longitude in state of Uttarakhand.

The rainfall data for years 1991-2010 were analyzed for extreme rainfall events. The extreme rainfall events for the years commencing from 1991-2010 were observed. The Gumbel distribution Log Pearson Type III, Log Normal and Van Te Chow method were adopted in the analysis, which are discussed in the following subsections. A flowchart methodology has presented in Fig. 1 for annual maximum prediction using frequency analysis for study area.

2.1. *Plotting position method*

The purpose of frequency analysis of an annual series is to obtain a relation between the magnitude of event and its probability of exceedance. The sample data are arranged in descending order of magnitude. Then each data is assigned an order number, m which starts from m = 1, for the first entry and so on till the last event for which m = N = number of years of records. The probability P of an event to or exceeded is given by the Weibull formula, as reported by Subramanaya (2009),

$$P = m / (N+1) \quad (1)$$

$$T = 1/P \quad (2)$$

in which T is the recurrence interval in years.

2.1.1. *Gumbel's extreme value distribution*

Gumbel's extreme value distribution was proposed by Gumbel (1941). This distribution is one of the most widely used probability distribution functions for extreme value in hydrology and meteorological studies for prediction of maximum rainfall, maximum wind velocity and maximum flood discharge.

Chow (1951) showed that most frequency distribution applicable in hydrology studies can be represented by the equation:

$$X_T = \overline{X} + K\sigma \quad (3)$$

where, the X_T is the value of the variate X of random hydrologic series with a return period T, $\overline{X}$ is the mean of the variate, σ is the standard deviation of the variate and K is the frequency factor. Equation 3 is known as the general equation of hydrologic frequency analysis.

Since practical data series of extreme events of rainfall depth have finite length of records, equation (3) is modified to account finite N (sample size) for practical use as given below:

$$X_T = \overline{X} + K.\sigma_{n-1} \quad (4)$$

in which σ_{n-1} represents the standard deviation of the sample andis expressed by the equation:

$$\sigma_{n-1} = \sqrt{\frac{\sum\left(X - \overline{X}\right)^2}{N-1}} \quad (5)$$

The frequency factor K, is expressed by the equation:

$$K = \frac{\left(Y_T - \overline{Y}_n\right)}{S_n} \quad (6)$$

where, Y_T is the reduced variate which depends on recurrence interval T and is expressed by the equation:

$$Y_T = -\left[\ln.\ln\left(T/(T-1)\right)\right] \quad (7)$$

where, $\overline{Y_n}$ is the reduced mean in equation (6) and depends upon sample size N. S_n is the reduced standard deviation in equation (6). $\overline{Y_n}$ and S_n corresponding to sample size N, are selected from tables that have been adopted from Subramanya (2009).

These equations (4 to 7) are used to estimate the extreme rainfall magnitude corresponding to a given recurrence interval based on annual rainfall series:

(*i*) The rainfall assembled for the sample size N in the present study is noted as 20 years. Annual maximum rainfall value is considered as the variate X, the mean of the variate, $\overline{X}$ and standard deviation of the sample, σ_{n-1} for the given data is calculated.

(*ii*) Using tables, the reduced mean $\overline{Y_n}$ and the reduced standard deviation S_n corresponding to the sample size, N equal to 20 are estimated. The reduced variate, Y_T for a given return period, T is computed by the equation (7).

(*iii*) The frequency factor is computed by equation (6).

(*iv*) The required value of variate X of random annual maximum rainfall series with a return period T is computed by the equation (4).

2.1.2. *Log Pearson type-III distribution*

Log-Pearson type-III distribution is used in frequency analysis. In this method, extreme rainfall magnitude of each year is first transformed into logarithmic form (base 10) and the transformed data is then analyzed. If X is the variate of random hydrologic series then the series of Z variates where,

$$Z = \log X \quad (8)$$

For the Z series determined by the equation (8), the equation (3) for the recurrence interval can be expressed as:

$$Z_T = \overline{Z} + K_Z.\sigma_Z \quad (9)$$

where, K_z is the frequency factor which depends on recurrence interval T and coefficient of skewness, C_s, $\overline{Z}$ is the mean of Z values and σ_z is standard deviation of the Z value sample. $\sum_z$ can be expressed by the following equation:

$$\sigma_z = \sqrt{\frac{\sum\left(Z - \overline{Z}\right)^2}{N-1}} \quad (10)$$

coefficient of skewness C_s, is expressed by the equation:

$$C_s = \sqrt{\frac{N\sum\left(Z - \overline{Z}\right)^3}{(N-1)(N-2)(\sigma_z)^3}} \quad (11)$$

TABLE 1

Estimation of one day maximum rainfall using Gumbel distribution

$\overline{X}$ = 57.035	$\overline{Y_n}$ = 0.5236	S_n = 1.0628	σ_{n-1} = 37.6014
Return period (T), years	Reduced variate Y_T = -ln ln [T/(T-1)]	Frequency factor $K = (Y_T - \overline{Y}_n)/S_n$	Estimated rainfall $X_T = \overline{X} + K_{\sigma_{n-1}}$
5	1.4999	0.9186	91.5793
10	2.2504	1.6248	118.1318
15	2.6739	2.0232	133.1125
20	2.9703	2.3022	143.6016
50	3.9022	3.1789	176.5697
100	4.6005	3.8360	201.2747

TABLE 2

Estimation of one-day maximum rainfall using Log Pearson type-III distribution

Return period (T), years	$\overline{Z}$ = 1.6918	σ_Z = 0.2266	C_S = 0.9632
	K_Z (From Table)	$Z_T = \overline{Z} + K_Z\sigma_Z$	X_T = antilog Z_T
5	0.388	1.7790	60.1284
10	1.34	1.9946	98.7782
15	1.583	2.0498	112.155
20	1.826	2.1047	127.290
50	2.593	2.2787	190.015
100	3.192	2.4143	259.626

TABLE 3

Estimation of one day maximum rainfall using Log Normal distribution

Return period (T), years	$\overline{Z}$ = 1.6918	σ_Z = 0.2266	C_S = 0
	K_Z (From table)	$Z_T = \overline{Z} + K_Z\sigma_Z$	X_T = antilog Z_T
5	0.4807	1.8008	63.2158
10	1.282	1.9824	96.0285
15	1.4383	2.0178	104.1880
20	1.5947	2.0532	113.0467
50	2.0542	2.1573	143.6749
100	2.3263	2.2190	165.5917

TABLE 4

Estimation of one day maximum rainfall using Ven Te Chow distribution

Return period (T), years	A = 11.1655		B = -75.6167	
	T-1	T/T-1	Z = log log (T/T-1)	$X = A+B \times Z$
5	4	1.25	-1.0136	87.7325
10	9	1.1111	-1.3395	112.3561
15	14	1.0714	-1.5234	126.2485
20	19	1.0526	-1.6521	135.9756
50	49	1.0204	-2.0568	166.5486
100	99	1.0101	-2.3600	189.4588

TABLE 5

Chi-Square test for goodness of fit for theoretical probability distribution for one day annual maximum rainfall data of Roorkee

Return period (T), years	Probability	Observed rainfall (R_O), mm	Expected rainfall (R_e), mm				$\chi^2 = (R_O - R_e)^2/R_e$			
			Gumbel	Log normal	Log Pearson type-III	Ven Te Chow	Gumbel	Log normal	Log Pearson type-III	Ven Te Chow
5	20	74.15	91.579	63.216	60.128	87.23	3.317	1.890	3.2697	1.961
10	10	106.65	118.13	96.032	98.778	112.356	1.115	1.173	0.6273	0.289
15	6.67	139.62	133.112	104.19	112.155	126.248	0.318	12.046	6.725	1.416
20	5	174.77	143.601	113.05	127.290	135.975	6.764	33.693	17.710	11.068
50	2	217.5	176.569	143.68	190.015	166.548	9.487	37.922	3.975	15.587
100	1	220.8	201.274	165.60	259.626	189.458	1.894	18.397	5.806	5.184
Mean							**3.816**	17.520	6.352	5.917

TABLE 6

Percent Absolute Deviation (PAD) values for goodness of fit for theoretical probability distribution for one day annual maximum rainfall data of Roorkee

Return period (T), years	Probability	Observed rainfall (R_O), mm	Expected rainfall (R_e), mm				$\text{PAD} = \frac{\lvert R_O - R_E \rvert}{R_o} \times 100$			
			Gumbel	Log normal	Log Pearson type-III	Ven Te Chow	Gumbel	Log normal	Log Pearson type-III	Ven Te Chow
5	20	74.15	91.5793	63.216	60.128	87.23	23.505	14.746	18.910	17.640
10	10	106.65	118.131	96.032	98.778	112.356	10.765	9.955	7.381	5.350
15	6.667	139.62	133.112	104.192	112.155	126.248	4.661	25.375	19.671	9.577
20	5	174.77	143.601	113.052	127.290	135.975	21.705	42.979	33.063	27.016
50	2	217.5	176.569	143.683	190.015	166.548	18.819	33.939	12.637	23.426
100	1	220.8	201.274	165.603	259.626	189.458	9.700	27.424	19.290	15.571
Mean							14.859	25.736	18.492	16.43

TABLE 7

Integral Square Error (I.S.E) values for goodness of fit for theoretical probability distribution for one day annual maximum rainfall data of Roorkee

Return period (T), years	Probability	Observed rainfall (R_O), mm	Expected rainfall (R_e), mm				$\text{ISE} = \frac{\left[\sum_{i=1}^{m}(R_{Ei}-R_{Oi})^2\right]^{\frac{1}{2}}}{\sum_{i=1}^{m}R_{Oi}}$			
			Gumbel	Log normal	Log Pearson type-III	Ven Te Chow	Gumbel	Log normal	Log Pearson type-III	Ven Te Chow
5	20	74.15	91.579	63.216	60.128	87.23	0.063	0.125	0.079	0.0791
10	10	106.65	118.131	96.032	98.778	112.356				
15	6.667	139.62	133.112	104.192	112.155	126.248				
20	5	174.77	143.601	113.052	127.290	135.975				
50	2	217.5	176.569	143.683	190.015	166.548				
100	1	220.8	201.274	165.603	259.626	189.458				

TABLE 8

Estimation of two consecutive days maximum rainfall using Gumbel distribution

$\overline{X}$ = 78.395	$\overline{Y_n}$ = 0.5236	S_n = 1.0628	σ_{n-1} = 43.184
Return period (T), years	Reduced variate $Y_T = -\ln\ln(T/(T-1))$	Frequency factor $K = (Y_T - \overline{Y_n})/S_n$	Estimated rainfall $X_T = \overline{X} + K\sigma_{n-1}$
5	1.4999	0.9118	117.7714
10	2.2504	1.6179	148.2661
15	2.6739	2.0163	165.4710
20	2.9703	2.2953	177.5174
50	3.9022	3.1721	215.3802
100	4.6005	3.8291	243.7530

TABLE 9

Estimation of two consecutive days maximum rainfall using Log Pearson type-III distribution

Return period (T), years	Z = 1.8431	σ_Z = 0.2266	C_S = 0.96322
	K_Z (From Table)	$Z_T = \overline{Z} + K_Z\sigma_Z$	X_T = antilog Z_T
5	0.4102	1.9276	84.6502
10	1.3389	2.1190	131.5230
15	1.5654	2.1656	146.4453
20	1.7915	2.2122	163.0298
50	2.4975	2.3577	227.9028
100	2.9563	2.4522	283.3302

TABLE 10

Estimation of two consecutive days maximum rainfall using Log normal distribution

Return period (T), years	$\overline{Z}$ = 1.8431	σ_Z = 0.2266	C_S = 0
	K_Z	$Z_T = \overline{Z} + K_Z\sigma_Z$	X_T = antilog Z_T
5	0.4807	1.9421	87.5297
10	1.282	2.1072	128.0196
15	1.4383	2.1394	137.8747
20	1.5947	2.1717	148.4955
50	2.0542	2.2664	184.6718
100	2.3263	2.3224	210.1219

TABLE 11

Estimation of two consecutive days maximum rainfall using Ven Te Chow distribution

Return period (T), years	A = 24.0050		B = -89.6631	
	T-1	T/T-1	Z = log log (T/T-1)	$X = A+B\times Z$
5	4	1.25	-1.0136	114.8903
10	9	1.1111	-1.3395	144.1121
15	14	1.0714	-1.5234	160.5987
20	19	1.0526	-1.6521	172.1423
50	49	1.0204	-2.0568	208.4246
100	99	1.0101	-2.3600	235.6130

TABLE 12

Chi-Square test for goodness of fit for theoretical probability distribution for two consecutive days annual maximum rainfall data of Roorkee

Return period (T), years	Probability	Observed rainfall (R_O), mm	Expected rainfall (R_e), mm				$\chi^2 = (R_O\text{-}R_e)^2/R_e$			
			Gumbel	Log normal	Log Pearson type-III	Ven Te Chow	Gumbel	Log normal	Log Pearson type-III	Ven Te Chow
5	20	120.2	117.771	87.59	84.650	114.890	0.050	12.140	14.929	0.245
10	10	156.46	148.266	128.091	131.523	144.112	0.452	6.283	4.728	1.057
15	6.667	174.34	165.471	137.81	146.445	160.598	0.475	9.683	5.313	1.175
20	5	188.72	177.517	148.48	163.029	172.142	0.706	10.905	4.048	1.596
50	2	202.56	215.380	184.67	227.902	208.424	0.763	1.733	2.818	0.165
100	1	222.7	243.753	210.121	283.330	235.613	1.818	0.753	12.975	0.707
			Mean				**0.711**	6.916	7.468	0.824

TABLE 13

Percent Absolute Deviation (PAD) values for goodness of fit for theoretical probability distribution for two consecutive days annual maximum rainfall data of Roorkee

Return period (T), years	Probability	Observed rainfall (R_o), mm	Expected rainfall (R_e), mm				$\text{PAD} = \frac{\lvert R_O - R_E \rvert}{R_o} \times 100$			
			Gumbel	Log normal	Log Pearson type-III	Ven Te Chow	Gumbel	Log normal	Log Pearson type-III	Ven Te Chow
5	20	120.2	117.7714	87.59	84.65026	114.8903	2.020	27.129	29.575	4.41
10	10	156.46	148.2662	128.091	131.523	144.1121	5.237	18.131	15.938	7.892
15	6.667	174.34	165.471	137.81	146.4454	160.5988	5.359	22.076	16.858	8.304
20	5	188.72	177.5175	148.48	163.0299	172.1423	5.936	21.322	13.613	8.784
50	2	202.56	215.3803	184.67	227.9029	208.4246	6.329	8.831	12.511	2.895
100	1	222.7	243.7531	210.121	283.3302	235.6131	9.453	5.648	27.225	5.798
						Mean	**5.722**	17.189	19.286	6.347

TABLE 14

Integral Square Error (I.S.E.) values for goodness of fit for theoretical probability distribution for two consecutive days annual maximum rainfall data of Roorkee

Return period (T), years	Probability	Observed rainfall (R_o), mm	Expected rainfall (R_e), mm				$\text{ISE} = \frac{\left[\sum_{i=1}^{m} (R_{Ei} - R_{oi})^2\right]^{1/2}}{\sum_{i=1}^{m} R_{oi}}$			
			Gumbel	Log normal	Log Pearson type-III	Ven Te Chow	Gumbel	Log normal	Log Pearson type-III	Ven Te Chow
5	20	120.2	117.771	87.59	84.650	114.890	**0.0121**	0.068	0.082	0.027
10	10	156.46	148.266	128.091	131.523	144.112				
15	6.667	174.34	165.471	137.81	146.445	160.598				
20	5	188.72	177.517	148.48	163.029	172.142				
50	2	202.56	215.380	184.67	227.902	208.424				
100	1	222.7	243.753	210.121	283.330	235.613				

TABLE 15

Estimation of three consecutive days maximum rainfall using Gumbel distribution

$\overline{X} = 92.755$	$\overline{Y_n} = 0.5236$	$S_n = 1.0628$	$\sigma_{n-1} = 44.45$
Return period (T), years	Reduced variate $Y_T = -\ln \ln (T/(T-1))$	Frequency factor $K = (Y_T - \overline{Y_n})/S_n$	Estimated rainfall $X_T = \overline{X} + K\sigma_{n-1}$
5	1.4999	0.9118	133.2877
10	2.2504	1.6179	164.6779
15	2.6739	2.0163	182.3880
20	2.9703	2.2953	194.7882
50	3.9022	3.1721	233.7628
100	4.6005	3.8291	262.9689

TABLE 16

Estimation of three consecutive days maximum rainfall using Log Pearson type-III distribution

Return period (T), years	Z = 1.9209	σ_Z = 0.2266	C_S = 0.96322
	K_Z (From Table)	$Z_T = \overline{Z} + K_Z\sigma_Z$	X_T = antilog Z_T
5	0.4538	2.0133	103.1151
10	1.3154	2.1905	155.0629
15	1.4996	2.2283	169.1952
20	1.6842	2.2663	184.6505
50	2.251	2.3829	241.4968
100	2.6008	2.4548	285.0012

TABLE 17

Estimation of three consecutive days maximum rainfall using Log normal distribution

Return period (T), years	$\overline{Z}$ = 1.9209	σ_Z = 0.2266	C_S = 0
	K_Z	$Z_T = \overline{Z} + K_Z\sigma_Z$	X_T = antilog Z_T
5	0.4807	2.0197	104.6592
10	1.282	2.1845	152.9545
15	1.4383	2.2167	164.7043
20	1.5947	2.2488	177.3651
50	2.0542	2.3433	220.4768
100	2.3263	2.3993	250.7954

TABLE 18

Estimation of three consecutive days maximum rainfall using Ven Te Chow distribution

Return period (T), years	A = 35.8819		B = -93.7565	
	T-1	T/T-1	Z	$X = A+B\times Z$
5	4	1.25	-1.0136	130.9165
10	9	1.1111	-1.3395	161.4723
15	14	1.0714	-1.5234	178.7117
20	19	1.0526	-1.6521	190.7822
50	49	1.0204	-2.0568	228.7209
100	99	1.0101	-2.3600	257.1506

TABLE 19

Chi-Square test for goodness of fit for theoretical probability distribution for three consecutive days annual maximum rainfall data of Roorkee

Return period (T), years	Probability	Observed rainfall (R_O), mm	Expected rainfall (R_e), mm				$\chi^2 = (R_O\text{-}R_e)^2/R_e$			
			Gumbel	Log normal	Log Pearson type-III	Ven Te Chow	Gumbel	Log normal	Log Pearson type-III	Ven Te Chow
5	20	129.19	133.287	104.659	103.112	130.916	0.125	5.749	6.593	0.022
10	10	163.42	164.67	152.954	155.06	161.472	0.009	0.716	0.450	0.023
15	6.667	179.44	182.388	164.704	169.195	178.711	0.047	1.318	0.620	0.002
20	5	192.49	194.788	177.365	184.650	190.782	0.027	1.289	0.332	0.015
50	2	230.43	233.762	220.476	241.496	228.720	0.047	0.449	0.507	0.012
100	1	265.97	262.968	250.795	285.001	257.150	0.034	0.918	1.270	0.302
			Mean				**0.048**	1.740	1.629	0.063

TABLE 20

Percent Absolute Deviation (PAD) values for goodness of fit for theoretical probability distribution for three consecutive days annual maximum rainfall data of Roorkee

Return period (T), years	Probability	Observed rainfall (R_O), mm	Expected rainfall (R_e), mm				$\text{PAD} = \frac{\lvert R_O - R_E \rvert}{R_o} \times 100$			
			Gumbel	Log normal	Log Pearson type-III	Ven Te Chow	Gumbel	Log normal	Log Pearson type-III	Ven Te Chow
5	20	129.19	133.287	104.659	103.115	130.916	3.074	23.438	25.287	1.318
10	10	163.42	164.67	152.954	155.063	161.472	0.763	6.842	5.389	1.206
15	6.667	179.44	182.388	164.704	169.953	178.711	1.616	8.946	6.054	0.407
20	5	192.49	194.788	177.365	184.606	190.782	1.179	8.527	4.245	0.895
50	2	230.43	233.762	220.476	241.498	228.720	1.425	4.514	4.582	0.747
100	1	265.97	262.968	250.795	285.001	257.150	1.141	6.050	6.677	3.429
			Mean				**1.153**	9.720	4.952	0.894

TABLE 21

Integral Square Error (I.S.E) values for goodness of fit for theoretical probability distribution for three consecutive days annual maximum rainfall data of Roorkee

Return period (T), years	Probability	Observed rainfall (R_O), mm	Expected rainfall (R_e), mm				$\text{ISE} = \frac{\left[\sum_{i=1}^{m}(R_{Ei} - R_{oi})^2\right]^{1/2}}{\sum_{i=1}^{m} R_{oi}}$			
			Gumbel	Log normal	Log Pearson type-III	Ven Te Chow	Gumbel	Log normal	Log Pearson type-III	Ven Te Chow
5	20	129.19	133.287	104.6592	103.115	130.916	**0.006**	0.033	0.032	0.008
10	10	163.42	164.678	152.9546	155.063	161.472				
15	6.667	179.44	182.388	164.7044	169.195	178.711				
20	5	192.49	194.788	177.3651	184.650	190.782				
50	2	230.43	233.762	220.4769	241.496	228.720				
100	1	265.97	262.968	250.7954	285.001	257.150				

The variation of K_z corresponding to C_s and T was obtained from table. After finding Z_T using equation (9), the corresponding value of variate, X_T, is obtained from the equation:

$$X_T = \text{antilog } Z_T \tag{12}$$

The computation of the theoretical rainfall magnitudes corresponding to different recurrence intervals was determined.

2.1.3. *Log Normal distribution*

Log Normal distribution is a special case of Log-Pearson type III distribution in which the coefficient of skewness, C_s, is zero. The other statistics like $\overline{Z}$ is calculated for the transformed rainfall data through equation 6 and σ_z can be calculated from equation (10), the values of K_z for a given return period T and $C_s = 0$ is read from table. Extreme rainfall values are estimated through equation 9.

2.1.4. *Van Te Chow distribution*

In this method, the plotting position is assigned to each annual maximum rainfall value arranged in decreasing order of magnitude. If the rank of X value is m, it's plotting position for return period.

$$T = \frac{N+1}{m} \tag{13}$$

in which N is equal to total no. of years of observations.

$$Z = \log\,[\log T - \log\,(T-1)] \tag{14}$$

$$Z = \log\log\frac{T}{T-1} \tag{15}$$

On substituting the value of $T = \frac{N+1}{m}$ in equation (15), the resulting expression becomes:

$$Z = \log\log\frac{N+1}{N+1-m} \tag{16}$$

The regression analysis between variable X and Y computed through equation

$$X_T = \text{A} + \text{B}Z \tag{17}$$

where, A and B are constants, which can be determined by the following equations:

$$\text{A} = \left\lfloor \sum \frac{X_i}{N} \right\rfloor - \left\lfloor \text{B} \sum \frac{Z_i}{N} \right\rfloor \tag{18}$$

$$\text{B} = \frac{\sum Z_i X_i - \sum Z_i \sum \frac{Z_i}{N}}{\sum Z_i^2 - \frac{\sum Z_i^2}{N}} \tag{19}$$

Once A and B are determined, the equation (17) can be used to determine X_T for any desired return period T. Based on equation (17), the theoretical extreme rainfall magnitudes were computed corresponding to selected recurrence intervals of 5, 10, 15, 20, 50 and 100 years.

2.2. *Goodness of fit criteria*

2.2.1. *Chi-square test*

This test is applicable to various problems of hydro-meteorological nature. It is primarily used for testing the agreement of the observed data with those expected upon a given hypothesis. The Chi-Square values, χ^2 can be calculated as:

$$\chi^2 = \frac{(R_o - R_E)^2}{R_E} \tag{20}$$

in which R_o and R_E are the observed and estimated rainfall magnitudes, respectively. The distribution with the least average of the Chi-Square values is adjudged to be the best. The $\chi^2 = 0$ indicates the R_o and R_E rainfall magnitudes agree exactly. The χ^2 values for each distribution are shown in Tables 5, 12 and 19.

2.2.2. *Percentage absolute deviation*

In order to test the goodness of fit of the computed and observed rainfall magnitudes, percentage absolute deviations (PAD) is determined by the equation which can be expressed as:

$$\text{PAD} = \frac{|R_O - R_E|}{R_o} \times 100 \tag{21}$$

where, PAD is the percentage absolute deviation of the computed extreme rainfall values with respect to the observed values are given in Tables 6, 13 and 20.

2.2.3. *Integral square error*

The integral square error (I.S.E) was used to measure the goodness of fit between the observed and estimated

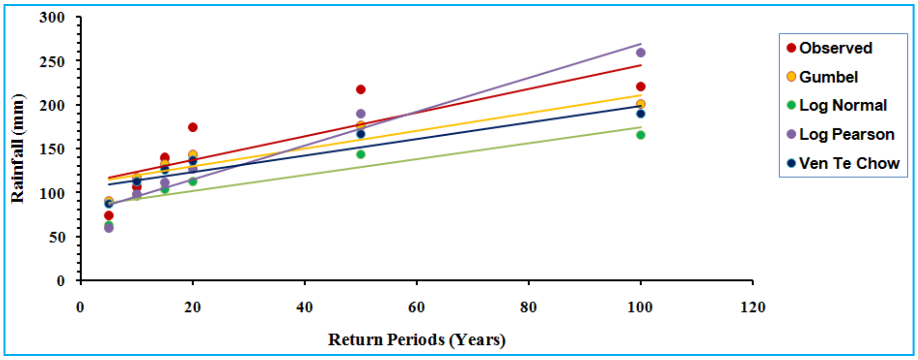

Fig. 2. Observed and estimated one day annual maximum rainfall using Gumbel, Log normal, Log Pearson type-III and Ven Te Chow

extreme rainfall. The integral square error values of distribution were estimated as reported by Disken *et al.* (1978).

$$\text{ISE} = \frac{\left[\sum_{i=1}^{m} \left(R_{Ei} - R_{oi}\right)^2\right]^{\frac{1}{2}}}{\sum_{i=1}^{m} R_{oi}} \tag{22}$$

where, R_{Oi} and R_{Ei} are the observed values of the estimated extreme rainfall magnitudes with respect to the observed values are given in Tables 7, 14 and 21.

Analysis of consecutive days maximum rainfall at different return periods is a basic tool for safe and economical planning and design of small dams, bridges, culverts, irrigation and drainage work, etc. Though the nature of rainfall is erratic and varies with time and space, yet it is possible to predict design rainfall fairly accurately for certain return periods using various probability distributions. The results of this study have been discussed in this section.

Frequency analysis is used to predict how often certain values of a variable phenomenon may occur and to assess the reliability of the prediction. It is a tool for determining design rainfalls and design discharges for drainage works and drainage structures, especially in relation to their required hydraulic capacity. Four different distributions were used to fit the observed maximum rainfall data for daily, two consecutive days and three consecutive days.

3. Results & discussion

3.1. *One day annual maximum rainfall*

Gumbel, Log Pearson type-III, Log normal and Ven Te Chow distributions were used to compute the extreme values of rainfall for one day as per the procedure explained. These computations are presented in tabular form in Tables 1 through 4. The observed and computed values for one day maximum annual rainfall obtained by using Gumbel, Log Pearson type-III, Log normal and Ven Te Chow distributions were plotted in Fig. 2. It is clear from the figure that the observed one-day annual maximum rainfall is very close to the theoretical values using Gumbel distribution. The best probability distribution was adjudged by comparing the average of Chi-Square, Percentage absolute deviation (PAD) and Integral square error (I.S.E.) values in percent obtained for these distributions corresponding to return period 5, 10, 15, 20, 50 and 100 years respectively as shown in Tables 5 through 7. The average of Chi-Square values for Gumbel, Log normal, Log Pearson type-III and Ven Te Chow was found to be 3.816, 17.520, 6.352 and 5.917 respectively. Average of PAD values for Gumbel, Log normal, Log Pearson type-III and Ven Te Chow distributions were observed to be5.120, 33.608, 15.431.

The average of Chi-Square values for Gumbel, Log normal, Log Pearson type-III and Ven Te Chow was found to be 3.816, 17.520, 6.352 and 5.917 respectively. Average of PAD values for Gumbel, Log normal, Log Pearson type-III and Ven Te Chow distributions were observed to be 5.120, 33.608, 15.431 and 11.0305 respectively and values of I.S.E for Gumbel, Log normal, Log Pearson type-III and Ven Te Chow distributions were 0.06343, 0.125, 0.0796 and 0.079. Hence, Gumbel Distribution gives the best fit for the predicted one day annual maximum rainfall values for Roorkee.

3.2. *Two days consecutive maximum annual rainfall*

The observed and estimated values of rainfall for two consecutive days were computed by Gumbel, Log Pearson type-III, Log normal and Ven Te Chow distributions as per the procedure explained in Chapter 3.

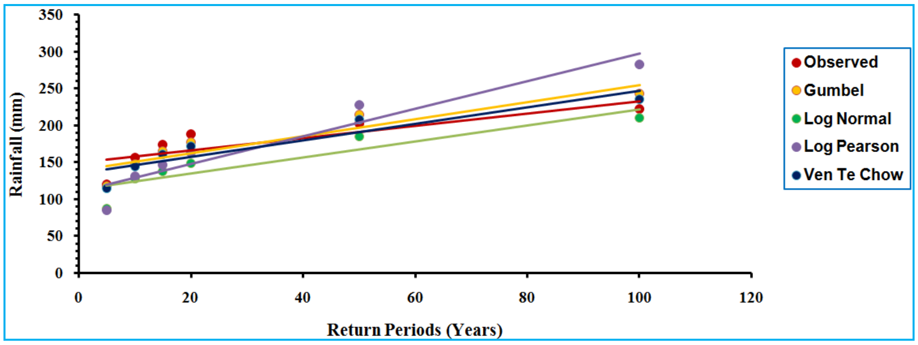

Fig. 3. Observed and estimated two consecutive days annual maximum rainfall using Gumbel, Log normal, Log Pearson type-III and Ven Te Chow

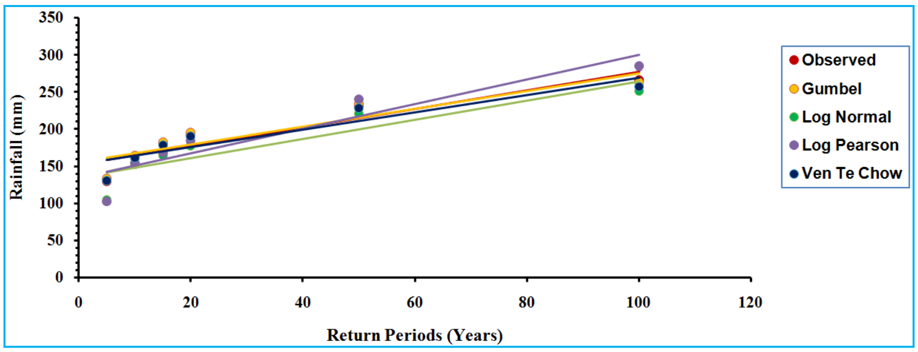

Fig. 4. Observed and estimated three consecutive days annual maximum rainfall using Gumbel, Log normal, Log Pearson type-III and Ven Te Chow

These computations are shown in tabular form in Tables 8 through 11. The observed and estimated values for two days consecutive maximum annual rainfall obtained by using Gumbel, Log Pearson type-III, Log normal and Ven Te Chow distributions were plotted in Fig. 3. It is evident from the figure that the observed two consecutive days maximum annual rainfall are very close to the theoretical values using Gumbel distribution. The best probability distribution was determined by comparing the average of Chi-Square, Percentage absolute deviation (PAD) and Integral square error (I.S.E) in percent values obtained for these distributions corresponding to return periods of 5, 10, 15, 20, 50 and100 years respectively as presented in Tables 12 through 14. The average of Chi-Square value for Gumbel, Log normal, Log Pearson type-III and Ven Te Chow distributions was determined to be 0.711, 6.916, 7.468 and 0.824 respectively. Average of PAD values for Gumbel, Log normal, Log Pearson type-III and Ven Te Chow distributions was found to be 0.778, 21.442, 10.540 and 3.846 respectively and I.S.E values for Gumbel, Log normal, Log Pearson type-III and Ven Te Chow distributions were 0.012, 0.068, 0.082 and 0.027. Hence, Gumbel distribution provides the best fit for the predicted two consecutive days annual maximum rainfall for Roorkee.

3.3. *Three days consecutive annual maximum rainfall*

The annual maximum value of rainfall for three consecutive days was determined by Gumbel, Log Pearson type-III, Log normal and Ven Te Chow distribution as per the procedure explained. These computations are given in tabular form in Tables 15 through 18. The observed and estimated values for three consecutive days annual maximum rainfall obtained by using Gumbel, Log Pearson type-III, Log normal and Ven Te Chow distributions were plotted in Fig. 4. It is seen from the figure that the observed three consecutive days maximum annual rainfall is very close to the theoretical values using Gumbel distribution. The best probability distribution was adjudged by comparing the average of Chi-Square, Percentage absolute deviation (PAD) and Integral square error (I.S.E) values in percent obtained for these distributions corresponding to return period at 5, 10, 15, 20, 50 and100 years respectively, shown in Tables 19 through 21. The average of Chi-Square value for Gumbel, Log Pearson type-III, Log normal and Ven Te Chow distributions were observed to be 0.048, 1.740, 1.629 and 0.063 respectively. Average of PAD values for Gumbel, Log Pearson type-III, Log normal and Ven Te

Chow distributions were found to be 01.153, 9.270, 4.952 and 0.894 respectively and I.S.E. values for Gumbel, Log Pearson type-III, Log normal and Ven Te Chow distributions were 0.006, 0.033, 0.032 and 0.008 respectively. Thus**,** Gumbel distribution gives the best fit for the predicted three consecutive days annual maximum rainfall values for Roorkee.

4. Conclusions

In this study, annual maximum rainfall values were estimated at return periods of 5, 10, 15, 20, 50 and 100 years for one day, two consecutive days and three consecutive days using four distributions namely, Gumbel, Log Normal, Log Pearson type-III and Ven Te Chow distributions. The data for 20 years (1991-2010), was collected from National Institute of Hydrology (NIH), Roorkee, which is located at 29.8667° N Latitude and 77.8833° E Longitude in state of Uttarakhand. On the basis of the present study the following conclusions have been drawn:

The observed rainfall magnitudes for the return periods of 5, 10, 15, 20, 50 and 100 years were fitted with the theoretical probability distributions namely, Gumbel, Log Pearson type-III, Log Normal, Ven Te Chow distributions. The values of Chi-Square, Percentage Absolute Deviation (PAD) and Integral Square Error (I.S.E.) were observed to be of very lower magnitudes. On the basis of this, it can be inferred that the Gumbel distribution emerged to be the best fit for the prediction of annual maximum rainfall values of Roorkee for one, two and three consecutive days. The second best fit was Ven Te Chow distribution for one, two and three consecutive days.

Acknowledgement

The authors express extreme reverence and profound sense of gratitude to their advisor during the research, Dr. Pravendra Kumar, Department of Soil and Water Conservation Engineering, GBPUA&T, Pantnagar for the invaluable guidance. Authors also thankful to Manish Kumar for his continuous encouragement and abundant counsel throughout the research work. We are highly grateful to the Director, National Institute of Hydrology, NIH (Roorkee), for providing the necessary data for the project.

Disclaimer : The contents and views expressed in this research paper/article are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

References

Agarwal, M. C., Katiyar, V. S. and Rambabu, 1988, "Probability analysis of annual maximum daily rainfall of U.P. Himalaya", *Indian J. Soil Cons*., **16**, 1, 35-43.

Bhakar, S. R., Bansal, A. N., Chhajed, N. and Purohit, R.C., 2006, "Frequency analysis of consecutive day's maximum rainfall at Banswara, Rajasthan, India", *ARPN Journal of Engineering and Applied Sciences*, **1**, 3, 64-67.

Chakraborty, S., Imtiyaz, M. and Issac, R. K., 2012, "Probability analysis for prediction of rainfall for Raipur region (Chhattisgarh*)", The Allahbad Farmer LXVII*, **2**, 6-15.

Chow, V. T., 1951, "General formula for hydrological frequency analysis", *Trans. Am. Geographic union*, **32**, 231-237.

Dabral, P. P., Pal Mautushi and Single, R. P., 2009, "Rainfall analysis for Doimukh (Itanagar), Arunachal Pradesh", *Journal of Indian Water Resources*, **29**, 9-15.

Dingre, S. and Shahi, N. C., 2006, "Consecutive days maximum rainfall prediction from one day maximum rainfall for Srinagar in Kashmir valley", *Indian J. Soil Cons*., **34**, 2, 153-156.

Gumbel, E. J., 1941, "The return period of flood flow", *Ann. Math. Statistics*, **12**, 2, 163-190.

Kumar, A. and Rastogi, D., 1982, "Analysis of rainfall data", B.Tech Thesis in Agricultural Engineering G.B Pant University of Agriculture and Technology, Pantnagar.

Manikandan, M., Thiyagarajan, G., Vijayakumar, G. and Vijayakumar, G., 2011, "Probability analysis for estimating annual one day maximum rainfall in Tamil Nadu", *Madras Agric. J*., **98**, 1-3, 69-73.

Nemichandrappa, M., Balakrishanan, P. and Senthilvel, S., 2010, "Probability and confidence limit analysis of rainfall in Raichur region", *Karnataka J. Agric. Sci*., **23**, 5, 737-741.

Panigrahi, B. and Panda, S. N., 2001, "Analysis of weekly rainfall for crop planning iun rain-fed region", *J. Agricultural Engineering*, **38**, 4, 47-57.

Pearson, K., 1930, "Tables for Statisticians and Biometricians, Part 1", The biometric Laboratory, University College, London, Printed by Cambridge Univ. Press, London.

Prajapati, R. N. and Kumar, S., 2002, "Probability analysis of rainfall data at Azamgarh Distt.", B.Tech. Thesis in Agricultural Engineering, G. B. Pant University of Agriculture and Technology, Pantnagar.

Subramanya, K., 2009, Engineering Hydrology. Chapter 7, McGraw Hill Book Co. Inc., New Delhi.

Upadhaya, A. and Singh, S. R., 1998 "Estimation of consecutive days maximum rainfall by various methods and their comparison.", *Indian J. Soil Conser*., **26**, 2,193-201.