Probability Distribution and Frequency Analysis of Consecutive Days Maximum Rainfall at Sambra (Belagavi), Karnataka, India

P. M. Hodlur and R. V. Raikar

April 8, 2021

Probability Distribution and Frequency Analysis of Consecutive Days Maximum Rainfall at Sambra (Belagavi), Karnataka, India

P. M. Hodlur[1] and R. V. Raikar[2]

[1]*Asst. Professor, Department of Civil Engineering, K.L.E College of Engineering & Technology, Bananthikodi Road,Chikodi, District Belgaum-591201, Karnataka, India*
[2]*Professor, Department of Civil Engineering and Dean, R & D, KLE Dr. M. S. Sheshgiri College of Engineering & Technology, Udyambag, Belgaum-590008, Karnataka, India*
Email: pmhodlur@klecet.edu.in,rvraikar@klescet.ac.in,Mobile No.: +91 9482402860, +91 9980272152

Abstract

In this case study, Daily Rainfall Data (1984-2019) of SambraRaingauge station in North Karnataka is used. An attempt was made to fit various probability distribution functions to the datasets of 1 day and 2 to 5 consecutive days annual maximum rainfall. The goodness of fit of probability distribution functions were tested by comparing the Chi-square (χ^2) values. No single probability distribution was adequate to describe the entire datasets. Various trendlines were also fitted to the rainfall datasets mentioned above; the best fit was decided based on the value of coefficient of determination R^2, no single trendline equation was able to describe the entire datasets. The magnitudes of 1 day as well as 2 to 5 consecutive days annual maximum rainfall corresponding to 2 to 100 years return period were estimated best fit distribution function, it was found that even though Normal distribution function had low Chi-square value comparatively, it cannot be used overall for estimation of rainfall values of different return periods for all the datasets. Rainfall was also estimated by best fit trendline equation i.e.polynomial 3rd order, for all the datasets corresponding to 2 to 100 years return period. It was observed the rainfall values predicted for 100 years return period for 1 to 5 consecutive days maximum rainfall were extremely high and unrealistic with respect to climate conditions of Sambra region. Chi-square test (χ^2) was conducted between observed rainfall and predicted rainfall by different trendline equations to ascertain the bestfit as determined by R^2, it was not able to establish the same results as determined by coefficient of determination.

Keywords:Rainfall, Frequency Analysis, Probability Distribution, trendline equation, Chi-square test*

1. Introduction

In India, rain is the principal form of precipitation, except in the Himalayan region where there is snowfall (Subramanya, 2011). The major portion of the country gets more than 75% of its annual rainfall due to monsoon winds, which extends generally from June to September and little rainfall during retreating monsoon season in the months October and November. The rainfall data is of prime importance for all hydrologic studies (Reddy, 2014) and the variation in rainfall distribution both spatially and temporally causes serious hydrological problems (extreme events) such as floods and droughts (Subramanya, 2011).

The magnitude of an extreme event and its frequency of occurrence are inversely related to each other; like very severe event occurs less frequently than more moderate events (Chow et al., 2010). For economic planning, the design engineer associated with water infrastructure projects such as dams, flood control structures, irrigation and drainage work and others often require estimates of extreme maxima events with recurrence interval of 2-100 years, for this they resort to frequency analysis of rainfall /streamflow data.

1.1*Frequency Analysis*

The frequency analysis is the estimation of frequency of occurrence of a hydrological event (Bhakar et al., 2006) relating the magnitude of extreme events to their frequency of occurrence using probability distributions. The hydrologic data analyzed are assumed to be independent and identically distributed, and the hydrologic system producing them is considered to be stochastic, space independent and time independent (Chow et al., 2010).

Adlouni and Ouarda(2010) mentioned the steps to carry out the frequency analysis in accordance to Annual Maximum Series (AMS) approach. The steps involved in AMS are (i) selection of a sample in the form of a data series, which satisfies certain statistical criteria, (ii) fitting of the best theoretical probability distribution to represent this sample using the best fitting technique available for the distribution, and (iii) use of this fitted distribution to make statistical inferences about the underlying data.

1.2 Studies Carried out on Frequency Analysis

Frequency analysis of rainfall data has been studied extensively in different parts of India at various temporal scales such as daily rainfall(annual daily maximum, seasonal daily maximum etc.) and consecutive days maximum rainfall, varying from 2days to 7 days (May, 2004; Guhathakurta et.al., 2005; Bhakar et al., 2006; Ramesh et.al., 2008; Patel and Shete, 2008; Deka, et.al., 2009; Vivekanandan and Mathew, 2010; Singh, 2012; Mandal and Choudhury, 2014; Singla et al., 2014; Kandpal, 2015 and Sabarish, 2017), weekly rainfall (Sharda and Das, 2005; Bhakar et al., 2008; Nemichandrappa, 2010; Kusre and Singh, 2012; Singh et al., 2016; Rajeshkumar, 2016), monthly rainfall, seasonal rainfall, annual rainfall (Bhakar et al., 2008; Kusre and Singh, 2012; Singh et al., 2016; Kumar, 2017; Sukrutha, 2018).

The commonly used probability distributions were Normal, Lognormal, Gamma, Weibull, Log-Pearson type III, and Gumbel distributions. On the other hand, the goodness of fit were tested by comparing the Chi-square values, Akaike Information Criterion and Bayesian Information Criterion, Kolmogorov-Smirnov tests or by using combination of these. Attempts were also made to compare the different forms of distribution functions viz. Lognormal (2P, 3P),Gamma (2P, 3P), Weibull (2P, 3P), log-logistic(2P, 3P), generalized gamma (3P, 4P) etc. and the goodness of fit were tested by comparing Kolmogorov–Smirnov test, Anderson Darling test and Chi-Square test (Sharda and Das, 2005; Mandal and Choudhury, 2014; Kumar, 2017).

Based on Likelihood ratio (LR) test, Sharda and Das (2005) revealed that three parameter distributions did not significantly improve the fit over two-parameter distributions within the same family. Even though the three-parameter probability distributions provided a better fit over two-parameter distributions in certain cases, the estimated percentiles and/or bounds of 95% confidence interval were found to be inadmissible and/or physically unrealistic whenever improvement in the fit was observed.

The exceedance probability of an event is obtained by the use of empirical formula, known as plotting position. Various plotting-position formulae have been listed (Rao and Hamed, 2000). To analyze the rainfall data by plotting position method, the conclusion of Cunnane (1978) as cited by Chow et.al.(2010) is duly acknowledged. He concluded that the Weibull plotting formulae is biased and plots the largest values of a sample at a too small return period for normally distributed data. He also found the Blom (1958) plotting position (b=3/8) is closest to being unbiased, while for data distributed according to the Extreme Value Type-I distribution the Gringorten (1963) formula (b=0.44) is the best. Makonnen (2008) mentioned that Weibull's plotting position formula is the correct plotting position in the extreme value analysis. The various other methods for determining the plotting positions, suggested during the last 90 years, such as the formulas by Blom, Jenkinson, and Gringorten, the computational methods by Yu and Huang (2001), as well as the modified Gumbel method, are incorrect when applied to estimation of return periods (Makkonen, 2005).

2 Datasets and Methodology

Sambra is a suburban area located in Belagavi taluk of Belagavi district in Karnataka, India. The normal rainfall of Belagavi taluk is 1504 mm with an average of 68 rainy days. The region experiences pleasant winters and dry hot summers. The taluk falls under Northern transition zone according to agro-climatic zones of Karnataka. Most parts of Belagavi district contributes runoff to the Krishna river basin except small catchments of Khanapur, Belagavi and Bailhongal taluks which contribute runoff to the Mahadayi and Kalinadi rivers that flow towards the west. The daily rainfall data (1984-2019) of Sambra Observatory obtained from Indian Meteorological Department, Pune is used in present study. The Sambra observatory is located at 15.84˚N 74.53˚E, at an elevation of 747 m above mean sea level. Figure 1 presents the location map of Sambra observatory station.

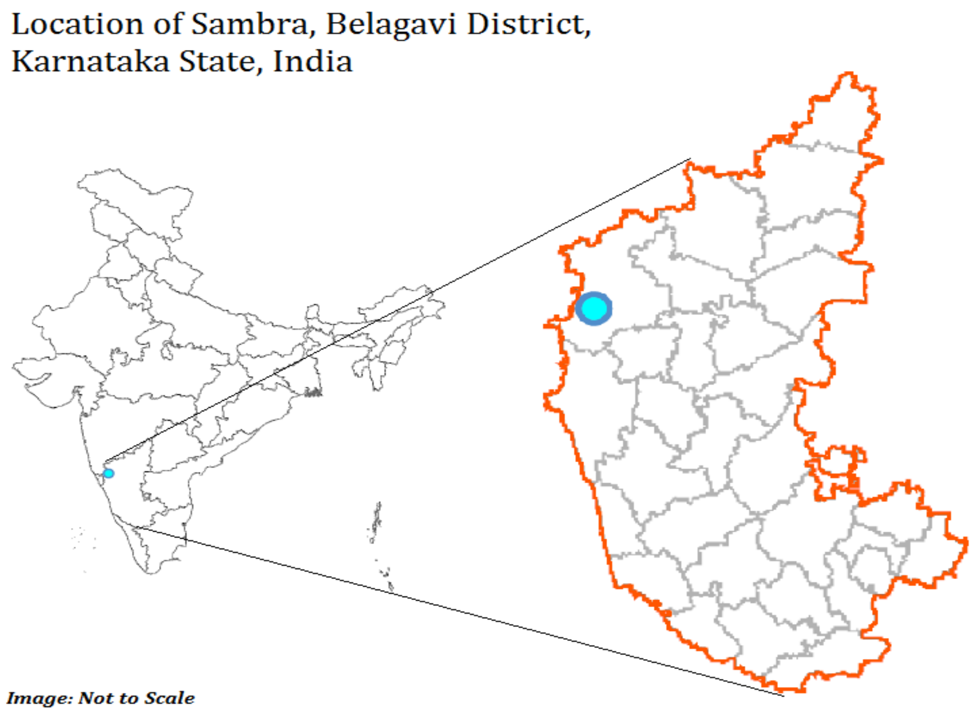

Figure1: Location map of Sambra, Belagavi District, Karnataka State, India
(Image Source: https://bharatmaps.gov.in)

The daily data in a particular year is converted to 2 to 5 days consecutive days rainfall by summing up the rainfall of corresponding previous days. The maximum amount of 1 day and 2 to 5 consecutive days rainfall for each year was taken for analysis (Bhakar et al., 2006). The statistical parameters of 1 day and 2 to 5 consecutive days annual maximum rainfall are furnished in Table-1.

Table 1: Statistical Parameters of Annual 1day and 2 to 5 Consecutive Days Maximum Rainfall

S. No	**Parameters**	**1 day**	**2 days**	**3 days**	**4 days**	**5 days**
1	Minimum (mm)	150.0	273.1	382.2	434.5	506.6
2	Maximum (mm)	41.0	56.4	59.0	62.6	66.2
3	Mean (mm)	76.7	107.7	129.6	146.3	159.6
4	Standard deviation (mm)	29.9	44.8	58.3	64.8	74.8
5	Coefficient of variation (%)	37.7	41.6	45.0	44.3	46.9
6	Coefficient of skewness	1.04	1.74	2.26	2.38	2.71

The probability of exceedance of rainfall is computed using the Weibull's plotting position formula and was applied to the prepared dataset of 1 day and 2 to 5 consecutive days annual maximum rainfall. The probability of exceedance of rainfall is given by $p = M/(N+1)$, where M is the order or rank and N is the total number of events. The recurrence interval or return period T is computed as inverse of probability p (T = 1/p).

In the present study, an attempt is being made to fit various probability distribution functions viz. Normal (2P), Lognormal (2P), Gumbel (EVI), Pearson Type III and Log Pearson Type III to the datasets of 1 day and 2 to 5 consecutive days annual maximum rainfall. The summary of probability distribution functions is given in Table-2.The goodness of fit of probability distribution functions will be tested by comparing the Chi-square (χ^2) values. Also, various trendlines will be fitted to these datasets mentioned above; the best fit will be decided based on the value of coefficient of determination R^2. The magnitudes of 1 day and 2 to 5 consecutive days annual maximum rainfall corresponding to 2 to 100 years return period will be estimated using best fit probability distribution function and compared with estimated values of rainfall with best fit trendline equation.

Table 2: Probability Distribution Functions (as adopted from Chow et.al.2010)

S. No	Distribution	Probability Distribution Function	Parameters in terms of sample moments
1	Normal	$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$\mu=\bar{x}, \sigma=s_x$
2	Lognormal*	$f(x)=\frac{1}{x\sigma\sqrt{2\pi}}\exp\left(-\frac{(y-\mu_y)^2}{2\sigma_y{}^2}\right)$	$\mu=\bar{y}, \sigma_y=s_y$
3	Gumbel (EVI)	$f(x)=\frac{1}{\alpha}\exp\left[-\frac{x-u}{\alpha}-\exp\left(-\frac{x-u}{\alpha}\right)\right]$	$\alpha=\frac{\sqrt{6}s_x}{\pi}$ $u=\bar{x}-0.5772\alpha$
4	Pearson Type III	$f(x)=\frac{\lambda^\beta(x-\epsilon)^{\beta-1}e^{-\lambda(x-\epsilon)}}{\Gamma(\beta)}$	$\lambda=\frac{s_x}{\sqrt{\beta}}, \beta=\left(\frac{2}{C_S}\right)^2$ $\epsilon=\bar{x}-s_x\sqrt{\beta}$
5	Log Pearson Type III**	$f(x)=\frac{\lambda^\beta(y-\epsilon)^{\beta-1}e^{-\lambda(y-\epsilon)}}{x\Gamma(\beta)}$	$\lambda=\frac{s_y}{\sqrt{\beta}}, \beta=\left(\frac{2}{C_S(y)}\right)^2$ $\epsilon=\bar{y}-s_y\sqrt{\beta}$

* $y=\log_e x$ and ** $y=\log_{10} x$

3 Frequency Analysis Using Frequency Factors

According to Chow et.al. (2010), the variable X_T of a hydrologic event is expressed as in equation (1):

$$X_T=\mu+K_T\sigma \quad (1)$$

where μ is the mean, σ is the standard deviation and K_T is the frequency factor, which is the function of return period and type of probability distribution used for analysis. For Normal distribution function, the frequency factor can be expressed as:

$$K_T=\frac{(X_T-\mu)}{\sigma} \quad (2)$$

Equation (2) is same as the standard normal variate z. The value of z corresponding to an exceedance of p (= 1/T) can be calculated by finding the value of an intermediate variable w given by

$$w = \left[\ln\left(\frac{1}{p^2}\right)\right]^{1/2} \qquad (0 < p \leq 0.5) \tag{3}$$

The standard normal variate z is computed using the equation (4) given by Abramowitz and Stegun (1965). When $p > 0.5$, (1-p) is substituted for 'p' in equation (3) and the value of z is computed by equation (4), however it is given a negative sign. The frequency factor K_T for normal distribution is taken equal to variable z(Chow et.al. 2010).

$$z = w - \left[\frac{2.515517 + 0.802853w + 0.010328w^2}{1 + 1.432788w + 0.189269w^2 + 0.001308w^3}\right] \tag{4}$$

In case of Lognormal distribution, the same procedure of normal distribution applies except that the logarithms of the variables Y_T is used in place of X_T, and their mean and standard deviation are used in equation (5). The required value of X_T is found by taking the antilogarithm of Y_T.

$$Y_T = \mu_y + K_T\sigma_y \tag{5}$$

The equation (6) given by Chow (1953) was used for determination of frequency factor in case of Extreme Value Type I:

$$K_T = -\frac{\sqrt{6}}{\pi}\left\{0.5772 + \ln\left[\ln\left(\frac{T}{T-1}\right)\right]\right\} \tag{6}$$

For Pearson Type III Distribution, frequency factor is computed using equation (7) given by Kite (1977) as mentioned in Chow et.al. (2010).

$$K_T = z + (z^2 - 1)k + \frac{1}{3}(z^3 - 6z)k^2 - (z^2 - 1)k^3 + zk^4 + \frac{1}{3}k^5 \tag{7}$$

where $k = C_S/6$, and Cs = coefficient of skewness.

For Log-Pearson Type III Distribution, the logarithms to the base 10 of the hydrologic data was computed. The mean, standard deviation and coefficient of skewness C_S were computed for the logarithmic values of the data and the frequency factor was computed using equation (7).

4 Goodness of Fit

The goodness of fit of a probability distribution can be tested by comparing the theoretical and sample values of the relative frequency or the cumulative frequency function (Chow et.al., 2010).In case of the relative frequency function, the χ^2test is used. The relative frequency of interval i is given by

$$f_s(x_i) = \frac{n_i}{n} \tag{8}$$

where n_i is number of observations in the interval *i* and n is the total number of observations.

The theoretical value of the relative frequency function, called the incremental probability function is computed by equation (9)

$$p(x_i) = F(x_i) - F(x_{i-1}) \tag{9}$$

The χ^2 test statistic $\chi_c{}^2$ is given by

$$\chi_C{}^2 = \sum_{i=1}^{m} \frac{n\left[f_s(x_i) - p(x_i)\right]^2}{p(x_i)} \tag{10}$$

where m is the number of intervals. It may be noted that $nf_s(x_i)$ $=n_i$, the observed number of occurrences in interval *i*, and $np(x_i)$ is the corresponding expected number of occurrences in interval *i*.

In χ^2 test, degree of freedom is υ= m - p - 1, where m is the number of intervals and p is the number of parameters used in fitting the proposed distribution. A confidence level is chosen for the test; it is often expressed as (1 – α), where α is the significance level. A typical value for the confidence level is95 percent (α = 5%). The null hypothesis for the test is that the proposed probability distribution fits the data adequately. This hypothesis is rejected (i.e., the fit is deemed inadequate) if the value of $\chi_c{}^2$ in equation (10) is larger than a limiting value,$\boldsymbol{\chi^2_{\upsilon,1-\alpha}}$determined from the χ^2distribution with υ degrees of freedom as the value having cumulative probability (1–α).

4.1 *Curve Fitting or Trendline to Frequency Analysis Datasets*

The recurrence interval and rainfall values from the datasets of 1day and 2 to 5 days consecutive maximum rainfall was used to plot the variation of rainfall versus return period (in logarithmic scale). Various trendlines such as exponential, linear, logarithmic, polynomial (order-2), polynomial (order-3),and power were fitted to the data. The best fit was determined based upon the value of coefficient of determination R^2. The table mentioned in Annexure I gives information about the different trendlines fitted, trendline equation along with coefficient of determination R^2 for the datasets of 1day and 2 to 5 days maximum rainfall.

5 Results and Discussions

Figure 2 shows the variation between maximum rainfall and probability for observed 1 day and 2 to 5 days consecutive maximum rainfall and estimated rainfall values by various probability distributions. The estimated values of rainfall by normal distribution follow the trend of straight line except at the extremities where the trend deviates to curvilinear as also observed by Christopoulos and Liakopoulos (1963). The normal distribution underestimates the observed rainfall values (both high and low) at boundaries. On the other hand, lognormal and Log Pearson type III distributions are the special cases of one another at low skewness coefficient and are in close agreement to each other (Sharda and Bhushan, 1985) and in turn with observed values of rainfall except at high boundary. At high boundary, both distributions estimate less rainfall compared to the observed value with increase in consecutive days of rainfall. Further, the Extreme Value Type I distribution underestimates the observed rainfall values (both high and low) at boundaries. The estimated rainfall by Pearson Type III distribution is close to observed values for 1 and 2 days maximum rainfall except at higher boundary where it under estimates the observed rainfall. For 3-5 consecutive days maximum rainfall, Pearson Type III overestimates observed rainfall at lower boundary and underestimates observed rainfall at lower boundary.

The Chi-square(χ^2)values of different probability distribution have been furnished in Table-3. For 1 day annual maximum rainfall and 2 days consecutive maximum rainfall, the normal distribution had least chi-square values of 0.137 and 0.518, respectively whereas the extreme value type (I) distribution exhibit maximum chi-square values of 9.331 and 10.04, respectively, in these series. In case of 3 days consecutive maximum rainfall, lognormal distribution had least chi-square value of 0.534 while extreme value type (I) distribution yields maximum chi-square value of 9.583. For 4 days consecutive maximum rainfall log-Pearson (III) distribution results least chi-square value of 0.649, on the other hand lognormal distribution gives maximum chi-square value of 9.386. Further for 5 days consecutive maximum rainfall, normal distribution shows least chi-square value of 0.557 while extreme value type (I) distribution results maximum chi-square value of 4.818. Based on the statistical comparison of chi-square values for goodness of fit, normal distribution is the best fit for the observed values of 1 day annual maximum rainfall and 2 days and 5 days consecutive maximum rainfall, whereas lognormal distribution is the best suited for the observed values of 3 days maximum rainfall and log-Pearson type (III) distribution is best fit for the observed values of 4 days maximum rainfall. In all the series, extreme value type (III) exhibits the highest χ^2 values.

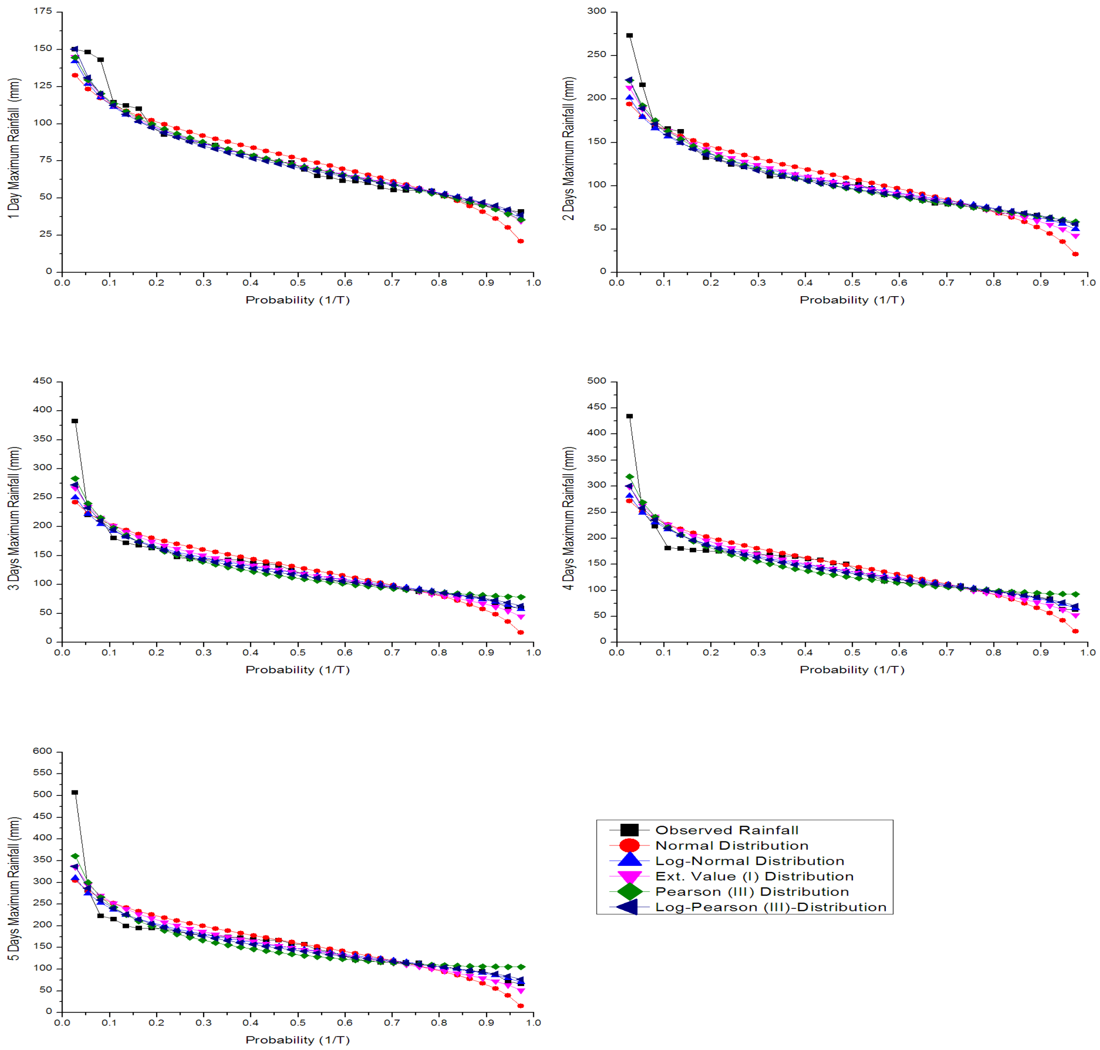

Figure 2: Variation of maximum rainfall with probability for observed 1day and 2 to 5 days consecutive Maximum Rainfall and estimated Rainfall values by various Probability Distributions

The Chi-square (χ^2) test values for different probability distribution functions mentioned in Table-3 viz. Normal, Lognormal, Extreme Value Type I, Pearson Type III and Log-Pearson Type III are calculated using equation (10) were found to be less than the limiting value of Chi-square at 95% confidence level i.e.,$\chi^2_{\upsilon,1-\alpha}$for all the data series. Hence, the null hypothesis for the test i.e., the proposed probability distribution fits the data adequately and is well accepted at 95% confidence level.

Table 3: Chi-Square test Values for Various Distribution Function

Data Series	Probability Distribution Function				
	Normal	Log-Normal	Extreme Value Type(I)	Pearson (III)	Log-Pearson (III)
1 Day MR	**0.137**	0.313	***9.331***	6.174	1.159
2 Days MR	**0.518**	0.560	***10.04***	1.747	1.413
3 Days MR	0.724	**0.534**	***9.583***	2.892	0.675
4 Days MR	0.778	1.062	***9.386***	3.348	**0.649**
5 Days MR	**0.557**	0.654	***4.818***	2.670	0.649

As illustrated in Table-4, logarithmic trendline is the best suited for 1 day and 2days consecutive maximum rainfall with respective R^2 values of 0.976 and 0.971. On the other hand, 3rd order polynomial gives highest R^2values of 0.953, 0.917, and 0.934 for 3 to 5 consecutive days maximum rainfall respectively. The details of best-fit trendline type, corresponding equation along with coefficient of determination for the datasets of 1day and 2 to 5 days consecutive maximum rainfall are given in Table 4.

Table 4: Details of Best fit Trendline Equation and Coefficient of Determination

Data Series	Best Fit- Trendline	Trendline Equation	R^2
1 Day MR	Logarithmic	$y = 33.53\ln(x) + 44.81$	$R^2 = 0.976$
2 Days MR	Logarithmic	$y = 51.8\ln(x) + 58.38$	$R^2 = 0.971$
3 Days MR	Polynomial 3rd Order	$y = 0.031x^3 - 1.700x^2 + 28.95x + 57.11$	$R^2 = 0.953$
4 Days MR	Polynomial 3rd Order	$y = 0.032x^3 - 1.704x^2 + 29.04x + 72.36$	$R^2 = 0.917$
5 Days MR	Polynomial 3rd Order	$y = 0.035x^3 - 1.858x^2 + 31.64x + 77.97$	$R^2 = 0.934$

As the χ^2 values of normal distribution, log-normal distribution and log-Pearson distribution were small and comparable, hence, it is decided to estimate the rainfall values for 2, 5,10, 20, 50, and 100 years return period using all the three distribution functions.Table-5gives predicted values of rainfall for 1 day and 2to 5consecutive days maximum rainfall by all the three distribution functions. It is observed that normal distribution function estimates high values of rainfall for smaller return periods 2, 5,10 years of return period(except for 1 day). However, log-Pearson type III distribution estimates high rainfall values for larger return periods of 20, 50 and 100 years. Hence even though normal probability distribution function had low Chi-square value, it cannot be used to estimate rainfall for different return periods in general for all the time periods.

Table 5: Predicted Rainfall using Normal, Lognormal and Log-Pearson type III Distribution function

Return Period (years)	Estimated Values of Maximum Rainfall (mm)														
	1 Day			2 Days			3 Days			4 Days			5 Days		
	N	LN	LP	N	LN	LP	N	LN	LP	N	LN	LP	N	LN	LP
2	**76.7**	72.0	70.40	**107.7**	100.4	96.6	**129.6**	119.8	115.9	**146.3**	135.5	132.2	**159.6**	147.3	142.6
5	**101.1**	96.8	96.0	**145.4**	136.0	133.7	**178.6**	165.3	163.1	**200.8**	186.4	184.6	**222.6**	203.8	201.1
10	113.9	113.1	**114.4**	**165.2**	159.4	162.2	**204.3**	195.6	198.7	**229.3**	220.3	223.2	**255.5**	241.6	245.4
20	124.4	128.5	**133.1**	181.5	181.8	**192.7**	225.4	224.8	**236.4**	252.8	252.8	**263.0**	282.7	278.0	**292.2**
50	136.3	148.4	**158.9**	199.8	210.7	**237.3**	249.3	262.9	**290.9**	279.3	295.2	**319.2**	313.3	325.5	**359.8**
100	144.1	163.4	**179.8**	212.0	232.5	**275.1**	265.2	291.9	**336.3**	296.9	327.4	**365.2**	333.7	361.7	**416.2**

*****N-Normal Distribution, LN-Lognormal Distribution, LP-Log-PearsonType III Distribution

Table-6 shows predicted values of rainfall by 3rd order polynomial for 1 day and 2 to 5 consecutive days maximum rainfall respectively. A maximum of 64.8mm in 1 day, 89.9 mm in 2 days, 108.5mm in 3 days, 123.9mm in 4 days and 134.1 mm is expected to occur in Sambra for every 2 years. For recurrence interval of 100 years the maximum predicted for 1day, 2 to 5 days consecutive maximum rainfall are 4701.6 mm, 8031.2mm, 16952.1mm,17936.mm and 19662.0 mm respectively.

The rainfall values predicted for 100 years return period for 1 to 5 consecutive days maximum rainfall were extremely high and unrealistic with respect to climate conditions of Sambra region.

Table 6: Predicted Rainfall Values using Polynomial 3rd order Trendline Equation

Return Period (years)	Maximum Rainfall (mm) 1 Day	2 Days	3 Days	4 Days	5 Days
2	64.8	89.9	108.5	123.9	134.1
5	99.3	134.4	163.2	179.0	194.1
10	132.3	178.7	207.6	224.4	243.6
20	136.0	199.2	204.1	227.6	247.6
50	243.1	578.7	1129.6	1264.4	1390.0
100	4701.6	8031.2	**16952.1**	**17936.4**	**19662.0**

The trendlines are also curves, fitted to the datasets of observed values of rainfall, the bestfit being decided based upon the highest value of coefficient of determination R^2. To ascertain the bestfit as determined by R^2 values it was also decided to conduct Chi-square test(χ^2) between observed rainfall and predicted rainfall by different trendline equations mentioned in earlier sections. Table-7 shows chi-squares values between observed rainfall and predicted rainfall by different trendline equations. The logarithmic trendline was the best fit as it had lowest χ^2 values among all the trendlines and in turn in all the observed datasets. The 3rd order Polynomial trendline, which was used to overall estimate the rainfall for 1 day maximum rainfall and 2 to 5 consecutive days maximum rainfall and for return periods for 2, 5, 10, 20, 50, and 100 years had more Chi-square value than logarithmic trendline. Thus, it can be concluded that Chi-square (χ^2) test is an important tool, which can be used instead of coefficient of determination R^2 in determining the bestfit.

Table 7: Chi-square values- Observed rainfall and Predicted rainfall by different trendline equations

Observed Datasets	Trendline Equation Type* Linear	Logarithmic	Polynomial 2nd order	Polynomial 3rd Order	Power
1 Day Maximum Rainfall	33.07	**6.05**	6.64	7.63	12.71
2 Days Maximum Rainfall	14.65	**0.24**	15.80	0.34	0.56
3 Days Maximum Rainfall	11.12	**1.92**	11.45	5.86	6.03
4 Days Maximum Rainfall	42.71	**12.24**	44.92	28.36	18.20
5 Days Maximum Rainfall	22.04	**10.28**	17.19	15.41	31.18

* Exponential trendline equation was not considered

6 Conclusions

An attempt is made to fit different probability distribution functions to 1 day and 2 to 5 consecutive days annual maximum rainfall data of SambraRaingauge station of Belagavi, Karnataka, India. The distributions included in the study are Normal (2P), Lognormal (2P), Gumbel (EVI), Pearson Type III and Log Pearson Type III. The goodness of fit of probability distribution functions is tested by comparing the Chi-square values. Following are the conclusions drawn from the study:

1. It is found that no single probability distribution is adequate to describe the annual maximum rainfall of different durations. Normal distribution is best suited for the observed values of 1 day, 2 days and 5 days consecutive maximum rainfall with chi-square values of 0.137. 0.518 and 0.557 respectively. Lognormal distribution is best fit for 3 days maximum rainfall with chi-square value of 0.534 and log-Pearson type (III) distribution for 4 days maximum rainfall with chi-square value of 0.649. All the chi-square values are found to be less than the limiting value of Chi-square at

95% confidence level and hence the proposed probability distribution fit the data adequately and are accepted at 95% confidence level.

2. Various trendlines are also fitted to the rainfall datasets. Based on the value of coefficient of determination R^2,logarithmic trendline is the best one for 1 day and 2days consecutive maximum rainfall with R^2 values of 0.976 and 0.971, respectively. For 3 to 5 consecutive days maximum rainfall, 3rd order Polynomial trendline with highest R^2values of 0.953, 0.917, and 0.934 respectively, is best suited.
3. The magnitudes of 1 day as well as 2 to 5 consecutive days annual maximum rainfall corresponding to 2 to 100 years return period were estimated using normal distribution, log-normal distribution and log-Pearson type III distribution as their Chi-square (χ^2) values were small and comparable. Normal distribution function estimated high values of rainfall for smaller return periods 2, 5 and 10 years return period(except for 1 day) while log-Pearson type III distribution estimated high rainfall values for larger return periods of 20, 50 and 100 years. In spite of low Chi-square value, normal distribution function cannot be used for overall estimation of rainfall values of different return periods.
4. Rainfall was also estimated by 3rd order polynomial equation for all the data range corresponding to 2 to 100 years return period. It was observed the rainfall values predicted for 100 years return period for 1 to 5 consecutive days maximum rainfall are extremely high and unrealistic with respect to climate conditions of Sambra region.
5. Chi-square test (χ^2) was conducted between observed rainfall and predicted rainfall by different trendline equations to ascertain the bestfit as determined by R^2. The logarithmic trendline was the best fit as it had lowest chi-square values (χ^2) among all the trendlines and in turn in entire datasets. The 3rd order polynomial trendline, which was used to overall estimate the rainfall for 1 day maximum rainfall and 2 to 5 consecutive days maximum rainfall and for return periods for 2, 5, 10, 20, 50, and 100 years had more Chi-square value than logarithmic trendline. This indicates that the Chi-square (χ^2) test is an important tool to determine the goodness of fit rather than coefficient of determination.
6. The results will facilitate the design engineers and hydrologist, who require information about annual daily maximum rainfall and consecutive days maximum rainfall of different frequencies or return period for planning and design of the small and medium hydraulic and soil and water conservation structures, irrigation, drainage works.

7Acknowledgement

We thank Indian Meteorological Department (IMD), Pune for providing us Rainfall Data of SambraRaingauge Station, Belagavi. We duly acknowledge the help rendered by District Statistical Office, Belagavi and the Director, KSNDMC, Bengaluru. We acknowledge the support and encouragement extended by the Principals of KLE College of Engineering and Technology, Chikodi and Dr. M. S. Sheshgiri College of Engineering and Technology, Belagavi. We thank the organizers of HYDRO 2020, ISH Pune and NIT Rourkela for giving us an opportunity to present our work. Finally, we thank each and everyone for their timely support.

8References

Adlouni, S. E., and Ouarda, T. B. M. J. (2010). Frequency analysis of extreme rainfall events, Rainfall: State of the Science, Geophysical Monograph Series,Vol-191, 171-188.

Bhakar, S. R., Bansal, A. K., Chhajed, N., and Purohit, R. C. (2006). Frequency analysis of consecutive days maximum rainfall at Banswara, Rajasthan, India. ARPN Journal of Engineering and Applied Sciences, Vol-01(03), 64-67.

Bhakar, S. R., Iqbal, M., Devanda, M., Chhajed, N., and Bansal, A. K. (2008). Probability analysis of rainfall at Kota. Indian J. Agric. Res., Vol-42 (3), 201 -206.

Chow, V. T., Maidment, D. R., and Mays, L. W. (2011). Applied Hydrology. Tata McGraw Hill Education Private Limited, New Delhi, India.

Christopoulos, C., andLiakopoulos, A. (1963). A method for frequency analysis of rainfall and runoff data, Hydrological Sciences Journal, Vol-8(4), 59-70, DOI: 10.1080/02626666309493357.

Deka, S., Borah, M., and Kakaty, S. C. (2009). Distribution of annual maximum rainfall series of North-East India. European Waters, Vol-27/28, 3-14.

Guhathakurta, P., Menon, P. A., Dikshit, S. K., and Sable, S.T. (2005). Extreme rainfall analysis of Andhra Pradesh using probability distribution model: A Regional Estimate. Mausam,Vol-56(4), 785-794.

Haktanir, T. (1991) Practical computation of gamma frequency factors, Hydrological Sciences Journal, Vol-36(6), 599-610.

Harter, H. L. (1969). A new table of percentage points of the Pearson Type III distribution, Technometrics, Vol-11(1), 177-187.

Kandpal, A., Kanwal, S., and Gosain, A. (2015). Estimation of consecutive days maximum rainfall using different probability distributions and their comparsion. Proceedings of National Conference on Innovative Research in Agriculture, Food Science, Forestry, Horticulture, Aquaculture, Animal Sciences, Biodiversity, Environmental Engineering and Climate Change (AFHABEC-2015), 100-106, ISBN: 978-93-85822-05-6.

Kumar, V., Shanu, and Jahangeer (2017). Statistical distribution of rainfall in Uttarakhand, India. Applied Water Science. Vol-7, 4765–4776.

Kusre, B. C., and Singh, S. (2012). Frequency analysis of different time series rainfall at Kohima, Nagaland, India. J. Inst. Eng. India Ser. A, Vol-93(2), 105–109.

Makkonen, L. (2005). Plotting positions in extreme value analysis. Journal of Applied Meteorology and Climatology, Vol-45(2), 334-340. https://doi.org/10.1175/JAM2349.1.

Makkonen, L., (2008) Bringing closure to the plotting position controversy. Communications in Statistics - Theory and Methods, Vol- 37(3), 460-467, DOI: 10.1080/03610920701653094.

Mandal, S., and Choudhury, B.U. (2015). Estimation and prediction of maximum daily rainfall at Sagar Island using best fit probability models. Theoretical and Applied Climatology. Vol-121, 87–97.

May, W. (2004). Variability and extremes of daily rainfall during the Indian summer monsoon in the period 1901–1989. Global and Planetary Change, Vol- 44, 83–105.

Nemichandrappa, M., Balakrishnan, P., and Senthilvel, S. (2010). Probability and confidence limit analysis of rainfall in Raichur region, Karnataka. J. Agric. Sci., Vol-23 (5), 737-741.

P. Jaya Rami Reddy. (2011). A Text Book of Hydrology. University Science Press (Laxmi Publication Pvt. Ltd.), New Delhi, India.

Patel N. R., and Shete, D. T. (2008) Probability distribution analysis of consecutive days rainfall data for Sabarkantha district of North Gujarat region, India, ISH Journal of Hydraulic Engineering, Vol-14(3), 43-55, DOI:10.1080/09715010.2008.10514921.

Rajeshkumar, N. K., Balakrishnan, P, G., V. Srinivas Reddy, Polisegowdar, B. S., and Satishkumar, U. (2016). Rainfall probability distribution analysis in selected lateral command area of Upper Krishna Project (Karnataka), India. Hydrologic Modeling, Select Proceedings of ICWEES-2016, Vol-81, 3-12.

Ramesh, C., Vivekanandan, N., Surwade, K. B., Bapat, A. D., Govindan, S., and Mathew, F. T. (2008). Extreme value analysis of rainfall in Mumbai region to aid estimation of severe flood, ISH Journal of Hydraulic Engineering, Vol-14(2), 102-117, DOI:10.1080/09715010.2008.10514908.

Rao, A. R., and Hamed, K. H. (2000).Flood frequency Analysis, 1st Edition, CRC Press, New Delhi.

Sabarish, R.M., Narasimhan, R., Chandhru, A.R., Suribabu, C. R., Sudharsan, J., and Nithiyanantham, S., (2017). Probability analysis for consecutive-day maximum rainfall for Tiruchirapalli City (South India, Asia). Appl. Water Sci., Vol-7, 1033–1042 (2017).

Sharda, V.N., and Bhushan, L.S. (1985). Probability analysis of annual maximum daily rainfall for Agra. Indian J. Soil Cons., Vol-13 (1), 16-20.

Sharda, V.N., and Das, P.K. (2005). Modelling weekly rainfall data for crop planning in a sub-humid climate of India. Agricultural Water Management, Vol-76, 120-138.

Singh, A.K., Singh, Y.P., Mishra, V. K., Arora, S., Verma, C.L., Verma, N., Verma, H.M., and Srivastav, A. (2016). Probability analysis of rainfall at Shivri for crop planning. Journal of Soil and Water Conservation. Vol-15(4), 306-312.

Singh, B., Rajpurohit, D., Vasishth, A., and Singh, J. (2012). Probability analysis for estimation of annual one day maximum rainfall of Jhalarapatanarea of Rajasthan, India. Plant Archives. Vol-12(2), 1093-1100.

Singla, S., Haldar, R., Khosa, R., Singla, R., and Rajeev, R. (2014). Frequency analysis of annual one day to five consecutive days maximum rainfall for Gandakriver basin. International Journal of Engineering and Technology, Vol-3 (2), 93-98.

Subramanya, K. (2011).Engineering Hydrology. Tata McGraw Hill Education Private Limited, New Delhi, India.

Sukrutha, A., Dyuthi, S.R.,and Desai, S. (2018). Multimodel response assessment for monthly rainfall distribution in some selected Indian cities using best-fit probability as a tool. Applied Water Science. Vol-8:145 1-10.

Vivekanandan, N., and Mathew, F. T. (2010). Probabilistic modelling of annual D-day maximum rainfall. ISH Journal of Hydraulic Engineering, Vol-16:sup1, 122-133,

Annexure I

Table A Details of Trendline Type, Trendline Equation and Coefficient of Determination (R^2)

Data Series	Trendline Type	Trendline Equation	R^2
1 Day Maximum Rainfall	Exponential	$y = 61.05e^{0.038x}$	$R^2 = 0.508$
	Linear	$y = 3.547x + 61.51$	$R^2 = 0.646$
	logarithmic	***$y = 33.53\ln(x) + 44.81$***	***$R^2 = 0.976$***
	Polynomial 2nd Order	$y = -0.209x^2 + 10.47x + 44.66$	$R^2 = 0.924$
	Polynomial 3rd Order	$y = 0.011x^3 - 0.800x^2 + 16.67x + 34.57$	$R^2 = 0.961$
	Power	$y = 49.35x^{0.396}$	$R^2 = 0.918$
2 Days Consecutive Maximum Rainfall	Exponential	$y = 83.32e^{0.043x}$	$R^2 = 0.624$
	Linear	$y = 6.231x + 80.96$	$R^2 = 0.832$
	logarithmic	***$y = 51.8\ln(x) + 58.38$***	***$R^2 = 0.971$***
	Polynomial 2nd Order	$y = -0.195x^2 + 12.7x + 65.23$	$R^2 = 0.933$
	Polynomial 3rd Order	$y = 0.016x^3 - 1.015x^2 + 21.3x + 51.23$	$R^2 = 0.963$
	Power	$y = 67.77x^{0.412}$	$R^2 = 0.951$
3 Days Consecutive Maximum Rainfall	Exponential	$y = 98.51e^{0.045x}$	$R^2 = 0.608$
	Linear	$y = 8.246x + 94.18$	$R^2 = 0.863$
	logarithmic	$y = 64.94\ln(x) + 67.73$	$R^2 = 0.904$
	Polynomial 2nd Order	$y = -0.126x^2 + 12.44x + 83.97$	$R^2 = 0.888$
	Polynomial 3rd Order	***$y = 0.031x^3 - 1.700x^2 + 28.95x + 57.11$***	***$R^2 = 0.953$***
	Power	$y = 79.81x^{0.426}$	$R^2 = 0.901$
4 Days Consecutive Maximum Rainfall	Exponential	$y = 112.1e^{0.044x}$	$R^2 = 0.585$
	Linear	$y = 9.101x + 107.2$	$R^2 = 0.851$
	logarithmic	$y = 70.26\ln(x) + 79.39$	$R^2 = 0.857$
	Polynomial 2nd Order	$y = -0.091x^2 + 12.11x + 99.90$	$R^2 = 0.862$
	Polynomial 3rd Order	***$y = 0.032x^3 - 1.704x^2 + 29.04x + 72.36$***	***$R^2 = 0.917$***
	Power	$y = 91.76x^{0.409}$	$R^2 = 0.850$
5 Days Consecutive Maximum Rainfall	Exponential	$y = 120.8e^{0.046x}$	$R^2 = 0.611$
	Linear	$y = 10.68x + 113.8$	$R^2 = 0.879$
	logarithmic	$y = 80.36\ln(x) + 83.12$	$R^2 = 0.840$
	Polynomial 2nd Order	$y = -0.064x^2 + 12.83x + 108.5$	$R^2 = 0.883$
	Polynomial 3rd Order	***$y = 0.035x^3 - 1.858x^2 + 31.64x + 77.97$***	***$R^2 = 0.934$***
	Power	$y = 98.70x^{0.420}$	$R^2 = 0.861$