Vol. 44, No. 1, pp 73-78, 2016
Indian Journal of Soil Conservation
Online URL:http://indianjournals.com/ijor.aspx?target=ijor:ijsc&type=home

Flood frequency prediction using Gumbel's and Log-Pearson method for Siang river

Saroj Kumar Pattanaaik[1], Ajai Kumar Pandey and Prankanu Debnath
College of Horticulture and Forestry, Central Agricultural University, Pasighat, Arunachal Pradesh.
[1]E-mail:saroj_swce@rediffmail.com

ARTICLE INFO

Article history :
Received : July, 2014
Revised : July, 2015
Accepted : August, 2015

Key words :
Flood frequency,
Gumbel's distribution,
Log-Pearson distribution,
Recurrence interval

ABSTRACT

Siang is the major river flowing within the Pasighat town of Arunachal Pradesh and poses flood inundation danger to the town. Flood frequency analysis is carried out for Siang river on the basis of peak flow data (46 years) collected from Komlighat gauging station. The hydrologic flood flow data is found to be fitted well to the Gumbel and Log-Pearson Type-III distribution function. Flood magnitudes of the Siang river have also been predicted and information generated for policy makers to adopt flood protection measures. The hydrologic parameters like design discharge of 30,520 cum at 100 years recurrence interval, HFL=156.3 m above msl, a free board of 2 m are recommended for design of required flood protection measures. Levee, stone spur, turfing, revetment and apron works and countryside plantation are among these measures. These measures are required for preventing the Pasighat town and its adjoining areas including the three existing hydroelectric plants *viz.*, Ymbung, Sille and Rina from future flood havoc.

1. INTRODUCTION

The Siang is the major river flowing within the Pasighat town of Arunachal Pradesh. It flows, generally single channel course through the mountain and enters the broad valley plains of Pasighat, where the river course becomes unstable and unpredictable. The river appears to be shifting to the southwards of Pasighat eroding valuable agricultural and urban lands, displacing people and reducing the economic potential of the people. Extensive areas of the flood plain that have been under agriculture in the past have been destroyed by sand casting when Siang river spilled over its banks downstream of Pasighat and spread fine sediments into the adjacent fields. The river has eroded a considerable portion of Pasighat township land along the bank. The river has encroached an area of 27.40 ha from the year 1976 to 1990 and it has eroded another 27.60 ha of land from 1990 to 2008 (WRD, Govt. of Arunachal Pradesh). In the year 2000 there was a flood in the Pasighat town affecting the people. The flood had also caused total collapse of the Ymbung hydroelectric plant, which is located at the upstream side of Pasighat town and was providing electricity to General Hospital. The flood is reported to have been a landslide lake outburst originating in China. In the year 2004, the flash flood in the Siang river was due to a very intense monsoon rainfall. The daily rainfall recorded by IMD for 18th and 19th July, 2004 was 450.3 and 421.2 mm, respectively. The flooding in the Siang's right bank downstream of the existing embankment in future may affect the urban, semi-urban and rural areas and the establishments like College of Horticulture and Forestry, General Hospital, market, airport and a number of Govt. offices. Hence, the flood level should be predicted in long term basis, based on which the flood protection measures may be taken.

Hydrologic processes such as floods are exceedingly complex natural events. They are resultants of a number of complex parameters and are, therefore, very difficult to model analytically. The flood in a catchment depend upon the characteristics of the catchment, rainfall and antecedent conditions, each of these factors in turn depend upon a host of constituent parameters. This makes the estimation of the flood peak a very complex problem leading to many different approaches. The empirical formulae and unit hydrograph method are some of them. Another approach to the prediction of flood flows, and also applicable to other hydrologic processes such as rainfall, *etc.* is the statistical method of frequency analysis. Gumbel's extreme value distribution and Log-Pearson Type-III distribution are two common methods of frequency distribution functions used for prediction of extreme flood values (Subramanya, 2008). Flood frequency analysis involves the fitting of a probability model to the sample of annual flood peaks recorded over a period of years for a catchment of a given region. The model parameters established can then be used to predict the extreme events of large recurrence interval (Pegram and Parak, 2004). Reliable flood frequency estimates are vital for flood plain management; to protect the public, minimise flood related costs to government and private enterprises, for designing and locating hydraulic structures and assessing hazards related to the development of flood plains (Tumbare, 2000). Nevertheless, to determine flood flows at different recurrence intervals for a site or group of sites is a common challenge in hydrology. Although studies have employed several statistical distributions to quantify the likelihood and intensity of floods, none had gained worldwide acceptance and is specific to any country (Law and Taskear, 2003). In India flood frequency studies have also been conducted on region basis (Throvat and Majumdar, 2011). They have established a regional relationship between mean annual peak flood and the catchments area of Krishna basin. Taking this into account a study on the flood frequency analysis was made by using 46 years of base data on the maximum discharge of the Siang river at Komlighat Gauge Discharge station, maintained by WRD, Govt. of Arunachal Pradesh.

2. MATERIALS AND METHODS

Study Area

Pasighat, the headquarter of East Siang district, is the earliest administrative centre, which was established in the year 1912 A.D. in Arunachal Pradesh. Geographically, the town is located at longitude 95º18′ 12″ E to 95º 20′ 27″ E and latitude 28º4′38″ N to 28º2′45″ N. The town of Pasighat is located on the right (South-western) bank of the Siang river at the confluence with its tributary *viz.,* Sibo-korong river. Siang river flows from its source in Tibet, where it is known as '*Yarlung Zangbo*' or '*Tsang Po*' river. Siang river basin includes three districts *viz.,* Upper, West and East Siang which has extension of snow covered glacial topography with barren and degraded forest belts of the Greater Himalayan high lands of North. The degraded mountain ridges and escarpments of Lesser Himalayas are dominated by degraded forest cover owing to severe soil erosion & shifting cultivation, which is one of the causes of flood in the downstream flood plain of Pasighat. There is increasing pressure of population in open flat lands of depositional alluvial plains of East Siang district where intensive settled cultivation is practiced. In Assam, it has become Bramahaputra. The tributary river Sibo-korong originates from Bapi hills at about 13 km upstream of its outfall in Siang river as shown in the Fig.1. The town is affected by flooding from both of these rivers. However, the river Siang plays an important role in flooding. The Siang river has a length of 250 km and catchment area of 2,49,000 sq km. (within the state of Arunachal Pradesh) and Sibo-korong has length of 13.5 km and catchment area of 39 sq km.

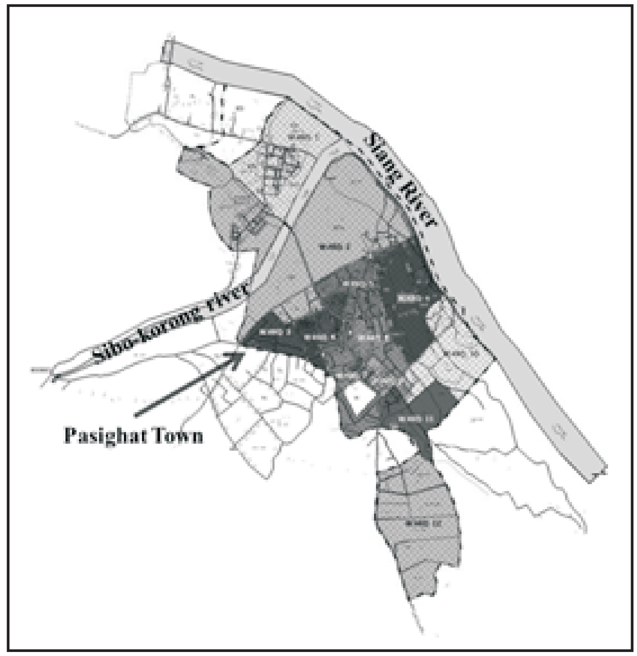

Fig.1. Location of Pasighat at the confluence of Siang river and Sibo-korong tributary

Data Collection and Interpretation

Hydrologic data series *i.e.* maximum discharge data (Q in cumec) for 46 water years of record at Pasighat gauging station on Siang river were collected from the water resource department of Government of Arunachal Pradesh. This is depicted in Table 1. Flow data were expressed in terms of exceedance probabilities and recurrence intervals. The data are then arranged in decreasing order of magnitude and the probability P of each event being equalled to or exceeded is calculated by the plotting position formula:

$$P = \frac{m}{N+1} \qquad ...(1)$$

Where, m = order number of the event and N = total number of events in the data. The recurrence interval, T (also called return period or frequency) is calculated as (Hann, 1977 and Shaw, 1983) below:

$$T = \frac{1}{P} \qquad ...(2)$$

Then Q was plotted versus T in semi-logarithmic paper, which yielded probability distribution as shown in Fig. 5. In frequency analysis of floods, the usual problem is to predict extreme flood events. Towards, this, specific extreme value distributions are assumed and the required statistical parameters calculated from the available data. Using these the flood magnitude for specific return period is estimated (Subramanya, 2008). Chow (1951) has shown that most frequency distribution function applicable in hydrologic studies can be expressed by general equation of hydrologic frequency analysis.

$$x_T = \bar{x} + K\sigma \qquad ...(3)$$

Where, x_T is the value of variate x of a random hydrologic series with a return period T, $\bar{x}$ is the mean of the

Table: 1
Maximum discharge of Siang river in various years

Year	Flood discharge (cum)	Year	Flood discharge (cum)	Year	Flood discharge (cum)	Year	Flood discharge (cum)
1949	11361	1961	21679	1986	12138	1998	27705
1950	24541	1962	29643	1987	22836	1999	17151
1951	18506	1976	16747	1988	20195	2000	44106
1952	9476	1977	24019	1989	14417	2001	17874
1953	6004	1978	14866	1990	12061	2002	19682
1954	29327	1979	18600	1991	23562	2003	19580
1955	14498	1980	16917	1992	17850	2004	17169
1956	14271	1981	12344	1993	18374	2005	16105
1957	17309	1982	14866	1994	13628	2006	12608
1958	14079	1983	16993	1995	19497	2007	14497
1959	13819	1984	28412	1996	19267		
1960	19725	1985	24824	1997	20778		

Source: WRD of Govt. of Arunachal Pradesh and NHPC

variate, σ is standard deviation of the variate and K is frequency factor, which depends on the return period and the assumed frequency distribution. Flood frequency analysis was carried out by fitting the data into both Gumbel's and Log-Pearson Type-III distribution.

Gumbel's Distribution

It is one of the most widely used probability distribution functions for prediction of flood peaks (Zelenhasic, 1970; Haan, 1977; Shaw, 1983). Gumbel (1941) defined flood as the largest of the 365 daily and the annual series of flood flows constitute a series of largest values of flows. In this study, it has been applied for flood frequency analysis because (i) peak flow data are homogeneous and independent hence lack long term trends; (ii) the river is less regulated, hence is not significantly affected by reservoir operations, diversions or urbanisation; and (iii) flow data cover relatively long record (more than 10 years) and is of good quality (Mujere, 2006 and 2011). As per the distribution, following equations are used to predict the flood peaks based on an annual series of flood.

$$x_T = \bar{x} + K\sigma_{n-1} \qquad ...(4)$$

Where, σ_{n-1} = standard deviation of the sample of size

$$N = \sqrt{\frac{\sum (x-\bar{x})^2}{N-1}} \qquad ...(5)$$

$$K = \text{Frequency factor} = \frac{y_T - \bar{y}_n}{S_n} \qquad ...(6)$$

y_T = Reduced variate corresponding to a recurrence interval = $-[\ln . \ln \frac{T}{T-1}]$ ...(7)

$\bar{y}_n$ = Reduced mean, a function of sample size N

S_n = Reduced standard deviation, a function of sample size N (Subramanya, 2008)

Log-Pearson Type-III Distribution

The Log-Pearson Type III distribution is a statistical technique for fitting frequency distribution data to predict the design flood for a river at some sites. Once the statistical information is calculated for the river site, a frequency distribution can be constructed. The probabilities of floods of various sizes can be extracted from the curve. The advantage of this particular technique is that extrapolation can be made of the values for events with return periods well beyond the observed flood events. This technique is the standard technique used by Federal Agencies in the United States. Here, variate is first transformed into logarithmic form and the transformed data is then analysed. If x is the variate of a random hydrologic series, then the series of Z variates is calculated as:

$$Z = \log x \qquad ...(8)$$

For this Z series, for any recurrence interval T,

$$Z_T = Z + Kz\,\sigma z \qquad ...(9)$$

Where, K_Z = a frequency factor which is a function of recurrence interval T and the coefficient of skew C_S of z variate:

σ_z = Standard deviation of Z variate sample

$$= \sqrt{\frac{\sum (z-\bar{x})^2}{N-1}} \qquad ...(10)$$

$$C_S = \frac{N\sum(z-\bar{z})^3}{(N-1)(N-2)(\sigma z)^3} \qquad ...(11)$$

$$X_T = \text{antilog}\,(Z_T) \qquad ...(12)$$

Test of Fitting Flood Data with Gumbel's Distribution

To verify whether the given data follow the assumed Gumbel's distribution, the procedure as developed by Subramanya (2008) was followed. Gumbel's distribution has the property which gives T=2.33 years for the average of the annual series when the sample size is large. Thus, the value of a flood with T=2.33 years is called the mean annual flood. In the graphical plot of predicted discharge (x_T) vs. return period (T), this gives a mandatory point through which the line showing variation of x_T and T must pass.

Anderson-Darling Goodness of fit test

The Anderson-Darling is used to test if a sample of data came from a population with a specific distribution. It is a modification of the Kolmogorov-Smirnov (KS) test, which is one of the most widely used goodness-of-fit test, used for hydrological data analysis test (Shin *et al.*, 2011 and Stephens, 1974; Kumar *et al.,* 2011). The Anderson-Darling (AD) test was first proposed to reduce errors in the stochastic progress by Anderson and Darling (1952). This test used a test statistic of the sum of the squares of the differences between the empirical and theoretical distribution functions with a weight function that emphasized discrepancies in both tails. Stephens (1976) showed that the empirical distribution function (EDF) tests were more powerful than other tests through power studies, and AD test was one of the most powerful of the EDF tests, in contrast with the KS test. However, this test also assigned equal weights to both tails of the distribution (Stephens, 1974, 1976). Software developed by Kevin Otto in 2005 was used. The Anderson-Darling test is still a good method of testing goodness of fit for the hydrologic data (Van Nooijen *et al*., 2010).

3. RESULTS AND DISCUSSION

The 46 years record of annual discharge on the Siang river provided a reasonable basis of estimation of the 2, 10, 25, 50, 100 and 200 years discharge. The peak flow data are homogeneous and independent hence lack long term trends. This is obtained from the Petit test. Pettit test was used to assess the homogeneity of the flood discharge data, which has been used to test the rainfall homogeneity in India (Panda *et al.*, 2013; Goyal, 2014).

Under null hypothesis, the annual values of the flood discharge (x_i) are independent and identically distributed and the series is considered as homogeneous. Under alternative hypothesis, test statistics assumes that the series has break in the mean and, thus can be considered as inhomogeneous. The Pettit test statistic is given by:

$$U_{t,n} = U_{t-1,n} + \sum_{i=1}^{n} \mathrm{sgn}\,(x_t - x_t), \text{ for the year } t = 2, \ldots, N. \quad ...(13)$$

For testing the hypothesis, the statistic K_t and the associated probability are given by:

$$K_t = \mathrm{Max}_{1 \leq t \geq n} |U_{t,n}|, \text{ and} \quad ...(14)$$

$$P \cong 2\exp\{-6(K_t)^2(n^2+n^3)^{-1}\}, \text{ respectively} \quad ...(15)$$

The null hypothesis is rejected at the significance level 0.05 if $p < 0.05$, and this suggests that a change point has occurred at time t. Application of this test to the flood discharge data indicated that p is equal to 0.95, thus the null hypothesis of homogeneous time series is accepted.

The maximum instantaneous flow of 44,106 cum was due to the landslide lake outburst flood originated in China. The instantaneous flow of 27,705 cum was recorded in the year 1998 and the minimum flow of 12061cum was recorded in 1990. The mean instantaneous maximum flow is 18000 cum. The principal uncertainty in the analysis arises from the difficulty of assessing the frequency and magnitude of landslide lake outburst flood. The flood data for the year 2000 *i.e.* 44,106 cum was not considered as it is an outlier (Fig. 2).The variation in observed and predicted flood flows is shown in Fig. 3. The observed and predicted flow shows the data followed Gumbel's distribution in Anderson-Darling test. Anderson-Darling test results test statistic 0.2028 which is less than the critical value 0.856 at 5% level of significance. The variation in the observed maximum discharge and predicted discharge is given in Fig. 3 and the values are given in Table 2. The predicted values follow the same trend of variation as the observed values.

Another test as suggested by Subramanya, 2008 was carried out for fitting the hydrologic flood data into both the distributions. The semi-log graphical plot of predicted discharge (x_T) vs. return period (T) is given in Fig. 5. The predicted mean annual flood discharge point passes through the line showing variation of x_T and T. This shows that the hydrologic flood data well fit to these distributions. Hence, the flood flow data of Siang river can be used for predicting future flood by using Gumbel and Log-Pearson distributions.

The hydrologic flood flow data were also fitted into Log-Pearson Type-III distribution. The predicted values were used for their goodness of fit by Anderson-Darling test. The test statistic was found to be 0.3318 and the critical value was 0.3978. Since the test statistic is less than the critical value the predicted data followed also the Log-Pearson Type-III distribution. The comparison of predicted

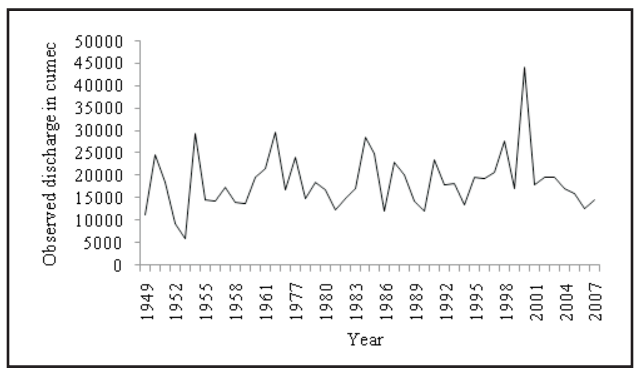

Fig. 2. Presence of outlier data in the hydrologic flood series

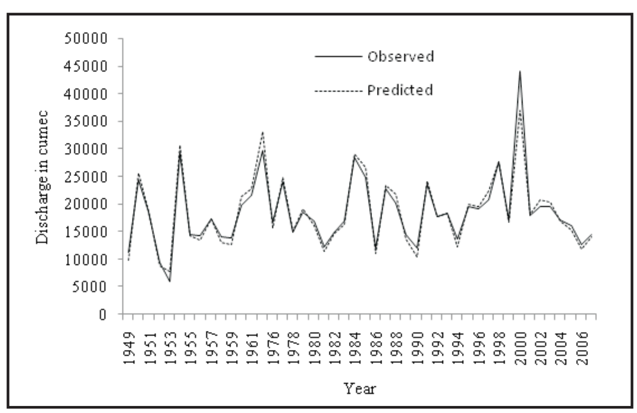

Fig. 3. Observed and predicted flood flows

Table: 2
Observed and predicted floods in corresponding years

Year	Observed flood (cum)	Predicted flood (cum)	Year	Observed flood (cum)	Predicted flood (cum)
1949	11361	9843.39	1961	21679	22892.63
1950	24541	25715.59	1962	29643	33077.70
1951	18506	18787.90	1976	16747	15701.86
1952	9476	8844.23	1977	24019	24906.06
1953	6004	7788.99	1978	14866	15081.71
1954	29327	30740.07	1979	18600	19157.92
1955	14498	14367.48	1980	16917	16039.16
1956	14271	13423.49	1981	12344	11481.34
1957	17309	17368.47	1982	14866	14738.36
1958	14079	13020.91	1983	16993	16354.92
1959	13819	12690.44	1984	28412	29060.04
1960	19725	21316.90	1985	24824	26649.95
Year	Observed flood (cum)	Predicted flood (cum)	Year	Observed flood (cum)	Predicted flood (cum)
1986	12138	10963.76	1998	27705	27742.33
1987	22836	23505.50	1999	17151	16699.81
1988	20195	21814.76	2000	44106	37030.51
1989	14417	13701.91	2001	17874	18071.30
1990	12061	10344.37	2002	19682	20835.47
1991	23562	24166.40	2003	19580	20403.16
1992	17850	17692.56	2004	17169	17022.72
1993	18374	18424.63	2005	16105	15401.78
1994	13628	12329.53	2006	12608	11930.37
1995	19497	19959.37	2007	14497	14047.22
1996	19267	19559.68			
1997	20778	22338.42			

data by Gumbel's and Log-Pearson Type-III distributions (Table 3) indicates no significant variation in the predicted discharge between the two well established distributions. However, the predicted value increases as the recurrence interval increases in both the cases. This indicates the use of both Gumbel and log-Pearson distributions for predicting flood data of Siang river. The hydrologic parameters like design discharge of 30,520 cum (at 100 years recurrence interval), HFL=156.3 m above msl (obtained from Fig. 4), a free board of 2 m are recommended for design of required hydraulic structures like levee, stone spur, turfing, revetment and apron works and countryside plantation works. These flood protection measures are required for preventing the Pasighat town and adjoining areas from future flood havocs.

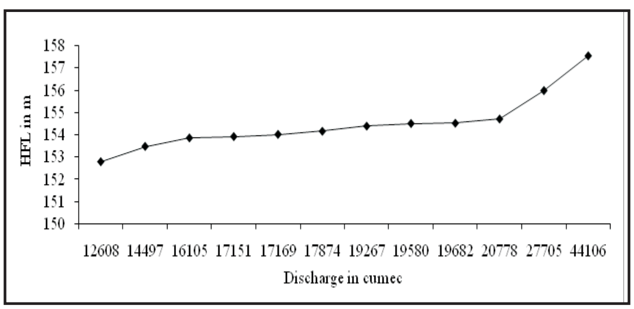

Fig. 4. Rating curve of Siang river at Komlighat Gauge site

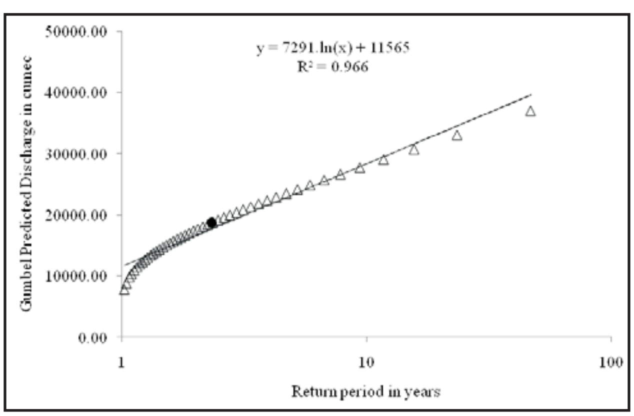

Fig. 5. Fitting of data in Gumbel's distribution

4. CONCLUSIONS

Flood frequency analysis had been carried out for Siang river at Komlighat gauging station using 46 years of peak flow data. The outcome of the analysis clearly indicates the good capability of the Gumbel and Log-Pearson Type-III distribution function to predict flood magnitudes of the Siang river flood flow.The information is useful for policy

Table: 3
Comparison of flood estimation using Gumbel and Log-Pearson Type-III distributions

Return period, T (yrs)	Flood flow by Gumbel's distribution (cum)	Flood flow by Log-Pearson Type-III distribution (cum)
2	17449.07	17845.55
2.33	18217.18	18035.16
10	24245.32	23136.62
25	27758.11	25376.98
50	30211.11	26915.35
100	32681.40	28358.28
200	35160.50	29742.67

makers to adopt suitable flood control measures especially to protect the Pasighat town from future floods. Among the various control measures, hydraulic structures like levee, stone spur, turfing, revetment and apron works and countryside plantation works are important. These structures are suggested for preventing the Pasighat town and adjoining areas including the existing three hydroelectric plants *viz.,*Ymbung, Sille and Rina from future flood havocs. The Ymbung hydro-electric plant was completely devastated due to 2000 year flood which has come in the Siang river.

ACKNOWLEDGEMENTS

The authors are grateful to the Executive Engineer of the Water Resources Department of East Siang district of Arunachal Pradesh for providing the base data for the analysis.

REFERENCES

Anderson, T.W. and Darling, D.A. 1952. Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes*. Ann. Math. Stat*., 23: 193-212.

Chow, V.T. 1964. Handbook of applied hydrology, McGraw-Hill, New York.

Goyal, Manish Kumar. 2014. Statistical analysis of long term trends of rainfall during 1901-2002 at Assam, India*. Water Resour. Manage.,* 28: 1501-1515.

Hann, C.T. 1977. Statistical methods in hydrology, IOWA State University Press, Ames, Iwoa. 1st Edition.

http://www.kevinotto.com/RSS/templates/Anderson-Darling Normality Test Calculator.xls

Kumar Ambrish, Sharma, H.C., Singh, R. And Prasad, S. 2011. Modelling of spring discharge in mid hills of north-west Himalayas. *Ind. J. Soil Cons.,* 39(2): 95-99.

Law, G.S. and Tasker, G.D. 2003. Flood-Frequency prediction methods for unregulated streams of Tennessee, Water resources investigations report 03-4176, Nashville, Tennessee.

Mujere, N. 2006. Impact of river flow changes on irrigation agriculture: A case study of yanyadzi irrigation scheme in Chimanimani district, MPhil Thesis, University of Zimbabwe, Harare.

Mujere, N. 2011.Flood frequency analysis using the Gumbel distribution*. Int. J. Comp. Sci. Eng.*, 3: 2775-2778.

Panda, Dileep K., Kumar, A., Singandhupe, R.B. and Sahoo N. 2013. Hydroclimatic changes in a climate-sensitive tropical region*. Int. J. Climatol.*, 33: 1633-1645.

Pegram, G. and Parak, M. 2004. A review of regional maximum flood and rational formula using geomorphological information and observed floods*. Water SA*, 30 (3): 377-388.

Shaw, E.M. 1983. Hydrology in Practice, Van Nostrand Reinhold, UK.

Shin, H., Jung, Y., Jeing, C. and Heo, J.H. 2011. Assessment of modified Anderson-Darling test statistics for the generalised extreme value and generalised logistic distributions, Stoch Environ Res Risk Assess, doi1-.1007/s00477-011-0463-y.

Stephens, M.A. 1974. EDF statistics for goodness of fit and some comparisons*. J. Am. Stat. Assoc.*, 69: 730-737.

Stephens, M.A. 1976. Asymptotic results for goodness-of-fit statistics with unknown parameters*. Ann. Statistics*, 4: 357-369.

Subramanya, K. 2008. Engineering Hydrology, TATA McGraw Hill, New Delhi-110063.

Throvat, A.R. and Majumdar, M.M. 2011. Design flood estimation for upper Krishnabasin through RFFA. *Int. J. Engg. Sci. Tech.*, 3(6): 5252-5259.

Tumbhare, M.J. 2000. Mitigating floods in Southern Africa, Paper presented at the 1st WARSFA/Water Net Symposium: Sustainable use of water resources, 1-2 November, Maputo.

Van Nooijen, R.R.P., Gubareva, T. and Kolechkina, A.G. 2010. A comparison of fitting methods and tests for severaldistributions on hydrological data, International Workshop Advances in Statistical Hydrology, May 23-25, 2010 Taormina, Italy: 01-09.

Zelenhasic, E. 1970. Theoretical Probability Distributions for Flood Peaks, Colorado University Press, Colorado.