JANS

Best fitting probability distributions for annual maximum discharge data of the river Kopili, Assam

Abhijit Bhuyan and Munindra Borah*

Department of Mathematical Sciences, Tezpur University, Napaam, Tezpur-784 028, INDIA
*Corresponding author. E-mail : mborah@tezu.ernet.in

_____

Abstract: In this study our main objective is to determine the best fitting probability distribution for annual maximum flood discharge data of river Kopili, Assam. Various probability distributions i.e. Gumbel (G), generalized extreme value (GEV), normal (N), log-normal (LN3), generalized logistic (GLO), generalized pareto (GPA) and Pearson type-III (PE3) have been used for our study. The L-moments methods have been used for estimating the parameters of all the distributions. The root mean square error (RMSE), model efficiency and D-index (fit in the top six values) together with L-moment ratio diagram is used for goodness of fit measure. It has been observed that Generalized Pareto is the best fitting probability distribution for annual maximum discharge data of river Kopili.

Keywords: L-moments, Probability distributions, Quantiles, L-moment ratio diagram
_____

INTRODUCTION

In many water resources design problem, determination of exact probability distribution is an important task. Flood frequency analysis mainly focused on the estimation of magnitude of floods for desired return periods. Based on an assumed distribution, it is possible to make probability statements of future flows of various magnitudes. Therefore, finding out suitable probability distributions for annual maximum flood discharge data is essential for designing structures and proper management of water resource. As per Indian design criteria, frequency based floods find their applications in estimation of design floods for almost all the types of hydraulic structures viz. small size dams, barrages, weirs, road and railway bridges, cross drainage structures, flood control structures etc., excluding large and intermediate size dams (Kumar *et al.*, 2003).

In recent years L-moment based flood frequency analysis introduced by Hosking (1990) used by various researchers world wide. The advantages of L-moment over conventional moments are that they are relatively insensitive to outliers and do not have sample size related bounds. The parameter estimations by method of L-moment are more reliable than the conventional method of moment estimates, particularly from small samples, and are usually computationally more tractable than maximum likelihood estimates.

In this study L-moment parameter estimation method is used for estimating parameters of various distributions viz. Gumbel , generalized extreme value, normal, log-normal, generalized pareto, generalized logistic and pearson type-

III. The three goodness-of-fit tests viz. RMSE, Model efficiency and D-index along with another graphical goodness-of-fit test i.e. L-moment ratio diagram are used for identification of best fitting probability distribution for our study area.

MATERIALS AND METHODS

Study area and data availability : The river Kopili is one of the most important south bank tributaries of the river Brahmaputra in Assam. The annual maximum discharge data for 31 years, collected from Brahmaputra Board, Guwahati, Assam, has been used for our study. The sample L-moments (details are defined later) of our data sets are given in Table 1.

Table 1. Sample L-moments of river Kopili

Sample Size	l_1	l_2	t	t_3	t_4
31	1196.6981	243.8254	0.2037	0.3114	0.1557

Probability distributions : The proper choice of a distribution is necessary to describe the design flood in a particular site. In flood frequency analysis, the commonly used probability distributions are gumbel, normal, generalized extreme value (GEV), generalized Pareto (GPA), generalized logistic (GLO), lognormal (LN3) and Pearson Type III (PE3) distributions. The probability density functions (pdfs) and quantile functions of these distributions can be obtained from the references such as Hosking and Wallis (1997), Rao and Hamed (2000) and are given in Table 2.

Table 2. Probability density functions and Quantiles of the probability distributions.

Sl.No.	Distributions	Probability density function, $f(x)$	Quantile function, $x(F)$
1	Gumbel	$a^{-1}\exp\{-(x-\pmb{x})/\pmb{a}\}\exp[-\exp\{-(x-\pmb{x})/a\}]$	$\pmb{x}-\pmb{a}\log(-\log F)$
2	Normal	$\pmb{a}^{-1}(2\pmb{p})^{-1/2}\exp[-1/2\{(x-\pmb{x})/\pmb{a}\}^2]$	explicit analytical form is not available
3	LN3	$\dfrac{\exp^{-\log\{1-k(x-\pmb{x})/\pmb{a}\}-\frac{1}{2}\{-\frac{1}{k}\log\{1-k(x-\pmb{x})/\pmb{a}\}\}^2}}{\pmb{a}\sqrt{2\pmb{p}}}$	explicit analytical form is not available
4	GEV	$\dfrac{1}{\pmb{a}}\left[1-k\left(\dfrac{x-\pmb{x}}{\pmb{a}}\right)\right]^{\frac{1}{k}-1}\exp\left[-\left\{1-k\left(\dfrac{x-\pmb{x}}{\pmb{a}}\right)\right\}^{\frac{1}{k}}\right]$	$\pmb{x}+\dfrac{\pmb{a}}{k}[1-(-\log F)^k]$
5	GPA	$\dfrac{1}{\pmb{a}}\left[1-k\left(\dfrac{x-\pmb{x}}{\pmb{a}}\right)\right]^{\frac{1}{k}-1}$	$\pmb{x}+\dfrac{\pmb{a}}{k}\left\{1-(1-F)^k\right\}$
6	GLO	$\dfrac{1}{\pmb{a}}\left[1-k\left(\dfrac{x-\pmb{x}}{\pmb{a}}\right)\right]^{\frac{1}{k}-1}\left[1+\left\{1-k\left(\dfrac{x-\pmb{x}}{\pmb{a}}\right)\right\}^{\frac{1}{k}}\right]^{-2}$	$\pmb{x}+\dfrac{\pmb{a}}{k}\left[1-\left\{\dfrac{1-F}{F}\right\}^k\right]$
7	PE3	$\dfrac{1}{\pmb{a}\Gamma(k)}\left(\dfrac{x-\pmb{x}}{\pmb{a}}\right)^{k-1}\exp\left\{-\left(\dfrac{x-\pmb{x}}{\pmb{a}}\right)\right\}$	explicit analytical form is not available

where $\pmb{x}$, α and k are the location, scale and shape parameters of the distributions

In this study to fit the model to our data set, we used L-moment parameter estimation method and a detail of the methods are given in the following subsection.

L-moments : Hosking (1990) introduce L-moments of a random variable as a linear combination of Probability Weighted Moments (PWMs). An unbiased estimator of PWMs of an order statistics $X_{(i)}$ is defined by Greenwood *et al.*,(1979) as:

$$b_r=\frac{1}{n}\sum_{i=1}^{n}\frac{(i-1)(i-2)\ldots(i-r)}{(n-1)(n-2)\ldots(n-r)}X_{(i)}$$

For any distribution, the first four L-moments can be calculated from:

$$l_1=b_0$$
$$l_2=2b_1-b_0$$
$$l_3=6b_2-6b_1+b_0$$
$$l_4=20b_3-30b_2+12b_1-b_0$$

L-coefficient of variation, L-CV $(t)=l_2/l_1$
L-coefficient of skewness, L-skew $(t_3)=l_3/l_2$
L-coefficient of kurtosis, L-kurtosis $(t_4)=l_4/l_2$
The sample L-moments of our data and the parameters of the distributions are calculated by using Fortran-77

subroutines provided by Hosking (1991) and are given in Table 1 and Table 3 respectively.

Goodness-of-fit : The three goodness-of-fit tests used are Model efficiency, Root mean square error (RMSE) and D-index (fit in the top six values). The formulas for these are given as follows:

Model efficiency = ((FIV-FRV)/FIV)*100,

where, FIV= Initial variance = $\sum_{i=1}^{n}(Q_i-\overline{Q})^2$

FRV = Remaining variance = $\sum_{i=1}^{n}(Q_i-\hat{Q}_i)^2$

RMSE = $\left(\sum_{i=1}^{n}(Q_i-\hat{Q}_i)^2/n\right)^{1/2}$, and

D-index = $\sum_{i=1}^{6}\left|Q_i-\hat{Q}_i\right|*(1/\overline{Q})$,

where, Q_i and $\hat{Q}_i$ are i^{th} observed and computed values of the annual maximum flood series, $\overline{Q}$ is mean of observed annual maximum flood series and n is length of data. The Weibul plotting position has been used for calculating

Model efficiency, RMSE and D-index for annual maximum discharge data's of river Kopili and are given in Table 4. If Model efficiency is used to compare the models, the highest value of it will indicate best fitting distribution. However, when RMSE and D-index are used, the model with smallest value of both is the best.

A graphical test which has been introduced by Hosking (1990), known as L-moment ratio diagram, is also used to determine the best fitting probability distribution for our study. The details are given in the following subsection.

L-moment ratio diagram : The L-moment ratio diagram (Fig.1) is a graph between L-kurtosis and L-skewness and can be used to select a best fitting probability distribution for a particular site. Theoretical curves of various distributions as well as the L-skewness and L-kurtosis of annual maximum flood discharge data of river Kopili are plotted on same graph to select the best fitting distribution.

RESULTS AND DISCUSSION

From Table 4, it has been observed for all the three goodness-of-fit tests that the GPA distribution is the best fitting probability distribution for the annual maximum

Table 3. Parameters of various distributions.

Distributions	Parameters		
	Location	Scale	Shape
Gumbel	993.6534	351.7657	
GEV	964.3325	278.3841	-0.2089
Normal	1196.6981	432.1693	
Log-normal	1065.1760	361.7343	-0.6525
GLO	1116.6555	227.5207	-0.2037
GPA	696.8133	524.9667	0.0502
PE3	1196.6981	480.5930	1.8687

Table 4. Model efficiency, RMSE and D-index of different distributions for river Kopili.

S. No.	Distribution	Model Efficiency(%)	RMSE	D-index
1	Gumbel	93.57	113.2838	1.0556
2	GEV	93.98	109.5355	1.0890
3	Normal	86.09	166.5740	1.4283
4	Log-normal	94.87	101.1578	1.0010
5	GLO	92.68	120.7982	1.1512
6	GPA	96.21	86.8861	0.8251
7	PE3	95.96	89.7862	0.8597

discharge data of the river Kopili. Again, in L-moment ratio diagram (Fig. 1) it has been observed that the L-skeness and L-kurtosis of our data lies near to the GPA distribution. Thus, based on L-moment ratio diagram also the GPA distribution is identified as the best fitting probability distribution of the data. Therefore, GPA distribution can be used for estimating design floods for various return periods viz. 2, 5, 10, 100, 500 and 1000 etc. for our annual maximum discharge data of river Kopili. The quantile estimates of the river Kopili for various return periods are given in Table 5.

Table 5. Quantile estimates for various return periods.

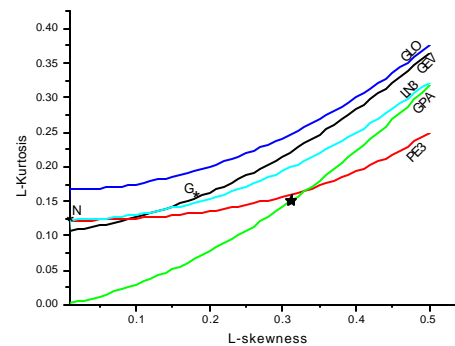| Return periods (in years)	2	5	10	100	500	1000
GPA	1054.43	1508.48	1838.35	2855.27	3499.41	3761.19

Fig. 1. L-moment ratio diagram

REFERENCES

Greenwood, J. A., Landwehr, J.M., Matlas, N.C. and Wallis, J.R. (1979). Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form, *Water Resour. Res.* , 15(5):1049-1054.

Hosking, J.R.M. and Wallis, J.R. (1997). In: *Regional frequency analysis- An approach based on L-moments.* New York: Cambridge University Press.

Hosking, J.R.M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics, *J.R. Stat. Soc. Ser. B. Methodol.*, 52 (2):105-124.

Hosking, J.R.M. (1991). *Fortran routines for use with the method of L-moments*, Version 2. Research Report RC-17097, IBM Thomas J. Watson Research Center, Yorktown Heights, NY.

Kumar, R., Chatterjee, C., Kumar, S., Chani, A.K.L. and Singh, R.D. (2003). Development of regional flood frequency relationships using L-moments for middle Ganga Plains subzone 1(f) of India, *Water Resources Management*, 17: 243-257.

Rao, A.R. and Hamed, H.K. (2000). In: *Flood frequency analysis*, Boca Raton, FL, USA: CRC.