Pak. J. Agri. Sci., Vol. 28, No. 3, 1991

FREQUENCY PREDICTION ANALYSIS OF FLOOD DATA OF CHENAB RIVER AT MARALA BY GENERALIZED EXTREME VALUE DISTRIBUTION

M. Adrees Ahmad & Shazia Arshad
Department of Mathematics & Statistics,
University of Agriculture, Faisalabad

Gumbel and GEV (generalized extreme value) flood frequency curves for river Chenab at Marala have been developed using 64 years (1925-1988) data concerning annual maximum flood series. The GEV distribution was estimated by the method of probability weighted moments. Its goodness of fit was determined by KS test. This distribution appeared to be appropriate for modelling the flood data. The flood estimates at various return periods i.e. 4, 5, 7, 10, 20, 50, 100 were worked out using GEV distribution. The estimated 100 years flood by this distribution was found to be 1.52 million cusecs.

INTRODUCTION

Floods cause many casualities and monetory losses in different countries of the world. Prediction of flood occurrence can save the destruction caused by floods. Forecasting can be beneficial in two ways. Firstly, the forecasting of an approaching flood can enable the people to leave the place where the risk is involved. The inhabitants of that area can transfer their valuables to a safer place and make preparations to cope with this situation. Secondly, the benefits through forecasting might be achieved by using the flood frequency analysis in the design of structures within the flood plan. Frequency analysis of flood data is a powerful source of providing plans of hydraulic structures, e.g. dams, spillways, bridges, etc. For efficient design of such structures the design engineers require estimation of flood quantities of various return periods. Estimation of the flood quantiles depend on previous flooding behaviour. We have taken annual maximum series for Marala Barrage from the Directorate of Hydrology, Lahore for flood frequency analysis. Some latest probability distributions, methods of estimation, goodness of fit tests are used for prediction of flood frequency. We have attempted to model the flood data by generalized extreme value distribution. This distribution was fitted by the method of probability weighted moment. To check goodness of fit, we have applied Kalmogorov Smirnov (KS) test and estimates of floods with different return periods have been worked out.

Source of data: Marala Weir situated at the confluence of the Chenab and Tawi was constructed in 1910-12 as a part of Tripple canals project (Ahmad and Chaudhary, 1988). It supplied water to upper Chenab canal having 16,500 cusec capacity. It was mainly a feeder canal supplying water to Balloki headworks. In 1954-56, Marala-Ravi Link canal was constructed to divert 22,000 cusecs to the Ravi to be utilized to feed Balloki-Sulemanki Link canal No. 1. Due to defective pond, it was not possible to feed both the canals. As remodelling of the old Weir was not possible, it was decided to construct a new barrage (about 1,200 feet down stream of the old Weir) and two head regulators for upper Chenab and Marala-Ravi Link. It is 4472.25 feet long and is designed to pass 11,00,000 cusecs through 66 bays each of 60 feet width with 10 bays on the left as undersluices. For feeding the canals, 6 bays each 40 feet wide, are constructed to feed upper Chenab canal and 8 bays of the same width to feed Marala-Ravi Link. This barrage is provided with radial type gates operated electrically. The undersluices have 13 bays each 60 feet wide. India has constructed a Dam at Salal as a hydroelectric project in Jammu territory about 49 miles upstream of Marala Barrage.

Table 1. Summary statistics of annual maximum series of river Chenab at Marala

N	Mean	Median	STDEV
64	342794	247564	262094
MIN	MAX	Q_1	Q_3
94383	1483592	170100	454577
C.V.	KURTOSIS	Coeff. of Sk	
76.4582	7.828836	0.61637	
TRMEAN	SEMEAN		
313207	32762		

Annual maximum series type data were taken up stream Marala Barrage from the Directorate of Hydrology at Feroze Pur Road, Lahore for flood frequency analysis for the period from 1925 to 1988. A look at these data revealed maximum flood occrruence during the months of July and August, indicating that the series have been generating water by a single phenomenon i.e. the torential rains. The data were first scrutinised and preliminary analysis was made for the summarisation. The basic statistics such as mean, media, standard deviation, standard error, skewness and kurtosis were calculated and the results have been presented in Table 1.

Frequency analysis: In order to model the flood frequency data, Gumbel and GEV (Jenkinson, 1955) distribution were considered and their parameters were estimated by the method of moments and by that of probability moments respectively.

Gumbel distribution: The Gumbel distribution has probability density function, cumulative distribution function and inverse cumulative distribution function as follows:

$$F(x) = 1/\alpha \exp[-(x-u)/ e^{-(x-u)/\alpha}] \quad (1)$$

$$F(x) = \exp[-e^{-(x-u)/\alpha}] \quad (2)$$

$$x(f) = u + \alpha Y \quad (3)$$

where, u is a location parameter, α is a scale parameter, and

$$Y = -\ln(-\ln F)$$

is the reduced form which does not depend on parameters.

GEV distribution: The generalized extreme value (GEV) distribution, introduced by Jenkinson (1955), combines into a single form the three possible types of limiting distribution for extreme values, as derived by Fishers and Tippett (1928). The distribution function is:

$$F(x) = \exp[-\{1-k(x-u)/\alpha\}^{1/k}] \quad k \neq 0$$
$$= \exp[-\exp\{-(x-u)/\alpha\}] \quad k = 0 \quad (4)$$

with x bound by $u + \alpha/k$ from above if $k>0$ and from below if $k<0$. Here u and α are location and scale parameters, respectively, and the shape parameter k determines which extreme value distribution is represented: Fisher and Tippett types 1, 11 and 111 correspond to $k = 0$, $k<0$ and $k>0$, respectively. When $k = 0$, the GEV distribution reduces to the Gumble distribution. The inverse distribution function is:

$$x(F) = u + \alpha[1-(-\log F)^k]/K, \quad K \neq 0$$
$$= u - \alpha \log(-\log F), \quad K = 0 \quad (5)$$

In practice the shape parameter usually lies in the range of $-1/2<k<1/2$. For example, the data base used in N.E.R.C. (1975 a) included 32 annual flood series with sample size of 30 or more. GEV distribution was fitted to these 32 samples by the method of maximum likelihood. The estimated shape parameter ranged from -0.32 to 0.48.

The probability-weighted moments of the GEV distribution for $k = 0$ are given by:

$$\beta r = (r+1)^{-1} [u + \alpha\{1-(r+1)^{-k}\}\,\Gamma(1+k)\}/k] \quad k > -1 \quad (6)$$

when $k \leq -1$ β_o (the mean of the distribution) and the rest of the βr do not exist. From (6), we have

$$\beta_o = u + \alpha\{1-\Gamma(1+k)\}/k \quad (7)$$
$$2\beta 1 - \beta_o = \alpha\,\Gamma(1+k)(1-2^{-k})/k \quad (8)$$

and

$$(3\beta_2 - \beta_o)/(2\beta 1 - \beta o) = (1-3^{-k})/(1-2^{-k}) \quad (9)$$

The PWM estimators $\hat{u}$, $\hat{\alpha}$, $\hat{k}$ of the parameters are the solution of (7) - (9) for u, α and k when the βr are replaced by their estimators br or $\hat{\beta}r[p_{j,n}]$. To obtain $\hat{k}$ the equation given below needs to be solved:

$$(3b_2 - b_o)/(2b_1 - b_o) = (1-3^{-k})/(1-2^{-k}) \quad (10)$$

The exact solution requires iterative methods, but because the function $(1-3^{-k})/(1-2^{-k})$ is almost linear over the range of values of k ($-1/2 < k < 1/2$), which is usually encountered in practice, low-order polynomial approximation for $\hat{k}$ are very accurate. Hosking *et al.* (1985) proposed the following approximate estimator:

$$\hat{k} = 7.8590c + 2.9554c^2,$$
$$c = 2b_1 - b_o/3b_2 - b_o - \log 2/\log 3 \quad (11)$$

The error in $\hat{k}$ due to using (11) rather than (10) is less than 0.0009 throughout the range of $-1/2 < k < 1/2$. Given $\hat{k}$, the scale and location parameters can be estimated successively from equations (8) and (7) as:

$$\hat{\alpha} = (2b_1 - b_o)^k/\Gamma(1+k)(1-2^{-k}),$$
$$u = b_o + \hat{\alpha}\{\Gamma(1+\hat{k}) - 1\}/\hat{k} \quad (12)$$

Equations (10) and (12) or their equivalent forms with br replaced by $\hat{\beta}r[p_{j,n}]$, define the PWM estimates of the parameters of the GEV distribution. Given the estimated parameters, the quantiles of the distribution are estimated using the inverse distribution function (5).

Goodness of fit: After the fitting of distribution our next step is to test the goodness of fit. In addition to theoretical justification, it is desirable to assess how well a distribution fits the observed flood series.

Goodness of fit tests are of several types and the test used here is based on empirical distribution function (edf) statistics. We have selected KS test out of various tests based on edf statistics. It is a nonparametric test based on the difference between the empirical distribution and the fitted distribution function.

edf = i/N & edf = F(x)
$D^+ = \sup [i/N - F(x)]$
$D^- = \sup [F(x) - (i-1)/N]$
$D = \max (D^+, D^-)$

Three test statistics are made by above D's.

$N/^{1/2}(D^+)$
$N^{1/2}(D^-)$
$N^{1/2}(D_{max})$

If the calculated value is greater than the tabulated value then the hypothesis of good fit is rejected (Stephen, 1977).

RESULTS AND DISCUSSION

The preliminary analysis of the annual maximum series is presented in Table 1. Various statistics presented in Table 1 reveal that there is a substantial difference between mean and median. Moreover, the coefficient of skewness is 0.61637, which indicates that the distribution is skewed. Thus, the first candidate distribution was Gumbel distribution. The estimated parameters of this distribution are presented in Table 2 along with KS statistic. The KS statistic rejected the hypothesis of Gumbel distribution suggesting that this distribution was not appropriate to model the flood series of river Chenab at Marala. Consequently the GEV distribution was considered. Its parameters were estimated by the method of probability weighted moments as given in Section 3 and presented in Table 2.

The goodness of fit was tested by KS test and the value of which is given in the same table. The KS test suggests that the GEV distribution fits the data well and hence can be used to predict the frequency of floods using the inverse distribution function

$$X(F) = u + (\alpha/K)[1-(-\ln F) \qquad K \neq 0$$

Table 2. Parameters estimating Gumble and GEV distributions fitted the 64 annual maximum flood series of river Chenab at Marala

Gumble			GEV		
$\hat{\alpha}$	=	0.5959	$\hat{\alpha}$	=	0.3452
$\hat{\mu}$	=	0.6560	$\hat{\mu}$	=	0.6315
D^+	=	0.1885	K	=	-0.3357
D^-	=	0.1612	D^+	=	0.116401
D	=	0.1885	D^-	=	0.0777575
Dcal	=	1.508	D	=	0.116401
D_{tab}	=	0.9570	D_{cal}	=	0.9312
			D_{tab}	=	0.9570

The quantities at various return periods i.e. 4, 5, 7, 10, 20, 50 and 100 years were worked out and are presented in Table 3. For both the distributions (GEV and Gumbel), the growth curves were developed. The KS test evidently favours the GEV distribution for modelling the flood data at this gauging station. It gives the estimates of higher floods at lower return periods desired by the hydrologists.

Conclusion: Gumble and GEV distributions are compared for modelling annual maximum series of river Chenab at Marala on the basis of KS test. GEV distribution is recommended for modelling flood frequency data. Estimates of floods at return periods 4, 5, 7, 10, 20, 50 and 100 have been derived and the growth curves to predict flood at any

Table 3. Flood estimates at various return periods by GEV

Row	Excedence probability	Return period	Quantile magnitude
1	0.25	4	$1.17\,\bar{X} = 399558.16$
2	0.20	5	$1.31\,\bar{X} = 447234.23$
3	0.15	7	$1.50\,\bar{X} = 512725.97$
4	0.10	10	$1.79\,\bar{X} = 614321.76$
5	0.05	20	$2.39\,\bar{X} = 819413.72$
6	0.02	50	$3.41\,\bar{X} = 1170264.61$
7	0.01	100	$4.42\,\bar{X} = 1515293.19$

return period both by Gumble and GEV distributions have been established.

REFERENCES

Fisher, R.A. and L.H.C. Tippet. 1928. Limiting forms of frequency distribution of the largest or smallest member of the sample. Proc. Cambridge Philosophical Society, 24: 180-190.

Hosking, J.R., J.R. Wallis and E.F. Wood. 1985. Estimation of the generalized extreme value distribution by the method of probability weighted moments. Technometrics, 27: 251-261.

Jenkinson, A.F. 1955. The frequency distribution of the annual maximum (or minimum) of meteorological elements. Quart. J. Royal Meteorol. Soc. 81: 158-171.

N.E.R.C. 1975. Flood Studies Report. Vol. 1. Natural Environmental Res. Council, H.M.S.O., London.

Stephens, M.A. 1977. Goodness of fit for the extreme value distribution. Biometrika, 64: 583-588.