Frequency Analysis of Annual Maximum Flood Discharge Using Method of Moments and Maximum Likelihood Method of Gamma and Extreme Value Family of Probability Distributions

N. Vivekanandan*

Central Water and Power Research Station, Pune, India

Abstract

Estimation of Maximum Flood Discharge (MFD) for a given return period is important for planning, design and management of hydraulic structures for the project. This can be achieved through Flood Frequency Analysis (FFA) by fitting of probability distributions to the recorded Annual Maximum Discharge (AMD) data. In this paper, Gamma and Extreme Value family of probability distributions are adopted in FFA. Method of Moments and Maximum Likelihood Method are used for determination of parameters of six probability distributions. Goodness-of-Fit tests such as Chi-square and Kolmogorov-Smirnov are applied for checking the adequacy of fitting of probability distributions to the recorded AMD data. Diagnostic test of D-index is used for the selection of a suitable distribution for estimation of MFD. The study showed that the exponential distribution (using MLM) is found to be better suited amongst six distributions adopted in estimation of MFD at Dedtalai and gamma distribution (using MLM) for Ghala.

Keywords

Chi-square, D-index, Kolmogorov-Smirnov, Maximum Flood, Probability Distribution

1. Introduction

For proper planning and design of hydraulic structures like dams, spillways, culverts, etc., a reliable estimation of Maximum Flood Discharge (MFD) for a given return period at the site of interest is necessary. Since the hydrologic phenomena governing the MFD are highly stochastic in nature, the MFD can be effectively determined by fitting of probability distributions to the series of recorded Annual Maximum Discharge (AMD) data through Flood Frequency Analysis (FFA).

A number of probability distributions viz., Exponential (EXP), Gamma (GAM), Generalized Extreme Value (GEV), Generalized Pareto (GPA), Extreme Value Type-1 (EV1) and Pearson Type-3 (PR3) are commonly used in FFA (Khosravi et al., 2012). According to the theory of probability distributions, EXP, GAM and PR3 are called as gamma family of distributions whereas EV1, GEV and GPA are called as extreme value family of distributions. Generally, Method of Moments (MoM) is used for determination of parameters of the probability distribution because of (i) MoM is often simple to derive; (ii) MoM is consistent estimators for continuous type of probability distributions; and (iii) MoM provide initial values in search for maximum likelihood estimates (Hosking and Wallis, 1993). In view of the above, in the present study, MoM and Maximum Likelihood Method (MLM) are used for determination of parameters of probability distributions.

In the recent past, number of studies has been carried out by researchers adopting GAM and Extreme Value family of
probability distributions for FFA. Kumar et al. (2003) carried out regional FFA adopting twelve frequency distributions and found that the GEV is better suited distribution for eight gauging sites of middle Ganga plains. Lee (2005) expressed that the PR3 distribution is better suited amongst five distributions studied for analyzing the rainfall distribution characteristics of Chia-Nan plain area. Bhakar et al. (2006) studied the frequency analysis of consecutive day’s maximum rainfall at Banswara, Rajasthan, India. Study by Saf (2009) revealed that the PR3 distribution is better suited for modelling of extreme values in Antalya and Lower-West Mediterranean sub-regions whereas the Generalized Logistic distribution for the Upper-West Mediterranean sub-region. Mujere (2011) applied EV1 distribution for modelling flood data for the Nyanyadzi River, Zimbabwe. Baratti et al. (2012) carried out FFA on seasonal and annual time scales for the Blue Nile River adopting EV1 distribution. Esteves (2013) applied extreme value distributions to estimate the flood quantile estimators.

2. Methodology

The study is to assess the probability distribution for FFA. Thus, it is required to process and validate the data for application such as (i) select the Probability Density Functions (PDFs) for FFA (say, EXP, EV1,GAM, GEV, GPA and PR3); (ii) determine the parameters of distributions using MoM and MLM; (iii) select quantitative GoF and diagnostic tests and (iv) conduct FFA and analyse the results obtained thereof. Table 1 gives the PDFs with the corresponding flood quantile estimators of six probability distributions used in FFA.

Table 1. PDFs with flood quantile estimators \( Q_r \)

<table>
<thead>
<tr>
<th>Distribution</th>
<th>PDF</th>
<th>( Q_r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXP</td>
<td>( f(Q; \beta, \xi) = e^{-\beta Q - \xi} ), ( Q &gt; \xi ) for ( \beta &gt; 0 )</td>
<td>( Q_r = \xi - \beta \ln(1 - F) )</td>
</tr>
<tr>
<td>GAM</td>
<td>( f(Q; \alpha, \beta) = \frac{1}{\alpha \Gamma(1/\alpha)} \frac{Q^{\alpha - 1}}{e^{\beta Q} - 1} ), ( 0 &lt; Q &lt; \infty )</td>
<td>( Q_r = \beta + \frac{\alpha Q}{\sqrt{\alpha^2 + \beta^2}} )</td>
</tr>
<tr>
<td>PR3</td>
<td>( f(Q; \alpha, \beta, \xi) = \frac{\beta}{\alpha} [\beta (Q - \xi)^{\beta - 1} e^{-\beta (Q - \xi)}] ), ( Q &gt; \xi ) for ( \alpha &gt; 0 )</td>
<td>( Q_r = \xi - \alpha \ln(1 - F) )</td>
</tr>
<tr>
<td>EV1</td>
<td>( f(Q; \alpha, \xi) = \alpha e^{-\alpha Q} e^{-e^{-\alpha Q}} ), ( \alpha &lt; 0 )</td>
<td>( Q_r = \xi + (\alpha(1 - (\ln F)^{1/\alpha}) / \beta) )</td>
</tr>
<tr>
<td>GEV</td>
<td>( f(Q; \alpha, \beta, \xi) = (1/\alpha) [1 - (\beta(Q - \xi)/\alpha)^{\alpha/\beta}]^{-1} ), ( Q &gt; \xi ) for ( \beta &gt; 0 )</td>
<td>( Q_r = \xi + \alpha(1 - (1 - F)^{1/\beta}) / \beta )</td>
</tr>
<tr>
<td>GPA</td>
<td>( f(Q; \alpha, \beta, \xi) = (1/\alpha) [1 - (\beta(Q - \xi)/\alpha)^{\alpha/\beta}]^{-1} ), ( Q &gt; \xi ) for ( \beta &gt; 0 )</td>
<td>( Q_r = \xi + \alpha(1 - (1 - F)^{1/\beta}) / \beta )</td>
</tr>
</tbody>
</table>

where, \( F(Q) \) (or \( F \)) is the Cumulative Distribution Function (CDF) of \( Q \) and \( K_F \) is the frequency factor corresponding to \( C_S \). For GAM distribution, \( C_S \) is computed from \( C_S = 2 / \sqrt{\beta} \). Similarly, for PR3 distribution, \( C_S \) is computed from the series of AMD. \( \xi, \alpha \) and \( \beta \) are the location, scale and shape parameters respectively. \( Q_r \) is the estimated MFD by probability distribution corresponding to return period \( T \).

2.1. Theoretical Descriptions of MoM

MoM is a technique for constructing estimators of the parameters that is based on matching the sample moments with the corresponding distribution moments (Haktanir and Horlacher, 1993; Ghorbani et al., 2010). The \( r^{th} \) central moment \( \mu_r \) about the mean \( \bar{Q} \) of a random variable \( Q \) is defined by:

\[ \mu_r = E[(Q - \bar{Q})^r] = \int (Q - \bar{Q})^r f(Q)dQ, \]

if \( Q \) is continuous variable (1)
where, \( f(Q) \) is a PDF of a random variable \( Q \). The second moment (\( \mu_2 \)) about \( \bar{Q} \) is called as variance. Similarly, third and fourth moments (\( \mu_3 \) and \( \mu_4 \)) about \( \bar{Q} \) are used to define skewness (\( C_3 \)) and kurtosis (\( C_4 \)), which are as follows:
\[
c_i = \mu_i / \mu_2^2 \quad \text{and} \quad c_4 = (\mu_4 / \mu_2^2) - 3
\]  
\[ \quad \text{(2)} \]

### 2.2. Theoretical Descriptions of MLM

The probability of occurrence of an observed sample series of a random variable can be calculated by multiplying the PDFs of every single observed data of that series with each other on the assumption that the events of the random variable are independent, which results in the Likelihood Function (LF). The parameter values that make the LF maximum will be the most suitable ones for that sample series because it actually happened among so many other possible sample series of the population. The maximum values of the LF and the logarithm of the LF always coincide with the same magnitudes of the distribution parameters. Therefore, it is analytically more convenient to take the derivative of the logarithm of the LF, which consists of summations of logarithms of the PDF, namely, LLF. For example, LF and LLF for 2-parameter and 3-parameter probability distributions are as given below.
\[
\text{LF} = \prod_{i=1}^{N} f(Q_i; \xi, \alpha) \quad \text{and} \quad \text{LLF} = \sum_{i=1}^{N} \ln(f(Q_i; \xi, \alpha))
\]  
\[ \quad \text{(3)} \]
\[
\text{LF} = \prod_{i=1}^{N} f(Q_i; \xi, \alpha, \beta) \quad \text{and} \quad \text{LLF} = \sum_{i=1}^{N} \ln(f(Q_i; \xi, \alpha, \beta))
\]  
\[ \quad \text{(4)} \]

A system of non-linear equations can be obtained from the analytical expressions of the partial derivatives of each parameter through LLF. The roots that make all these equations zero simultaneously are the magnitudes of the parameters estimated by MLM (Seckin et al., 2010). The procedures involved in determination of parameters of probability distributions (using MoM and MLM) are briefly described in the text book of ‘Flood Frequency Analysis’ by Rao and Hamed (2000).

### 2.3. Goodness-of-Fit Tests

GoF tests are essential for checking the adequacy of probability distributions to the recorded series of AMD in the estimation of MFD. Out of a number GoF tests available, the widely accepted GoF tests are \( \chi^2 \) and KS, which are used in the study. The theoretical descriptions of GoF tests statistic are as follows:

\[ \chi^2 \text{ Statistic:} \]
\[
\chi^2 = \sum_{i=1}^{NC} \left( \frac{O_i(Q) - E_i(Q)}{E_i(Q)} \right)^2
\]  
\[ \quad \text{(5)} \]

where, \( O_i(Q) \) is the observed frequency value of \( j^{th} \) class, \( E_i(Q) \) is the expected frequency value of \( j^{th} \) class and \( NC \) is the number of frequency classes. The rejection region of \( \chi^2 \) statistic at the desired significance level (\( \eta \)) is given by \( \chi^2 \geq \chi^2_{\text{critical}} \). Here, \( m \) denotes the number of parameters of the distribution and \( \chi^2 \) is the computed value of \( \chi^2 \) statistic by PDF.

\[ K\text{S Statistic:} \]
\[
KS = \max_{i} \left( F_i(Q) - F_e(Q) \right)
\]  
\[ \quad \text{(6)} \]

where, \( F_i(Q) \) is the empirical CDF of \( Q \) and \( F_e(Q) \) is the computed CDF of \( Q \) (Zhang, 2002).

**Test criteria:** If the computed values of GoF tests statistic given by the distribution are lesser than that of the theoretical values at the desired significance level, then the distribution is considered to be acceptable for estimation of MFD.

### 2.4. Diagnostic Test

The selection of a suitable probability distribution for estimation of MFD is carried out through D-index, which is defined as:
\[
D\text{-index} = \left( \frac{1}{\eta \bar{Q}} \right) \sum_{i=1}^{6} \left| Q_i - \bar{Q} \right|
\]  
\[ \quad \text{(7)} \]

where, \( \bar{Q} \) is the average (or mean) of the recorded AMD, \( Q_i \)'s (i=1 to 6) are the first six highest sample values in the series and \( \bar{Q} \) is the estimated value by PDF. The distribution having the least D-index is identified as better suited distribution in comparison with the other distributions for estimation of MFD (Vivekanandan, 2014).

### 3. Application

In this paper, a study was carried out to estimate the MFD adopting six probability distributions on river Tapi at Dedtalai and Ghala gauging sites. Based on the water year (June-May), stream flow data related to the period 1977-78 to 2004-05 for Dedtalai and 1978-79 to 2004-05 for Ghala is used. The series of AMD is derived from the daily stream flow data and used in FFA. Table 2 gives the summary statistics of AMD.

<table>
<thead>
<tr>
<th>Gauging site</th>
<th>Statistical parameters (SD: Standard Deviation)</th>
<th>Mean (m³/s)</th>
<th>SD (m³/s)</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dedtalai</td>
<td></td>
<td>3441.9</td>
<td>3533.3</td>
<td>2.643</td>
<td>9.323</td>
</tr>
<tr>
<td>Ghala</td>
<td></td>
<td>3563.9</td>
<td>4901.4</td>
<td>1.801</td>
<td>1.994</td>
</tr>
</tbody>
</table>

Table 2. Summary statistics of AMD
4. Results and Discussions

Statistical software, namely VTFIT, is used in FFA. This software gives the parameters of the six probability distributions (using MoM and MLM), MFD estimates for different return periods, GoF and diagnostic tests results.

4.1. Estimation of MFD by Six Probability Distributions

The parameters of six probability distributions are determined by MoM and MLM; and further used for estimation of MFD. Tables 3 and 4 give the estimates of MFD for different return periods for river Tapi at Dedtalai and Ghala sites. The MFD estimates are used to develop the flood frequency curves and presented in Figures 1 and 2.

4.2. Analysis Based on GoF Tests

In the present study, the degree of freedom \((NC-m-1)\) was considered as one for 3-parameter distributions (PR3, GEV and GPA) and two for 2-parameter distributions (EXP, GAM and EV1) while computing the \(\chi^2\) statistic values for Dedtalai and Ghala. GoF tests statistic is computed using Eqs. (5) and (6); and the results are presented in Table 5.

From Table 5, it may be noted that the computed values of \(\chi^2\) statistic for EXP, GAM and EV1 distributions (using MoM and MLM) are lesser than the theoretical values at 5% significance level and thus these three distributions are acceptable at 5% significance level for Dedtalai. On the other hand, the computed values of \(\chi^2\) statistic by the distributions are greater than the theoretical values at 5% significance level and all six distributions are not acceptable at 5% significance level for Ghala when MoM and MLM is applied for determination of parameters of the distributions.

For Dedtalai, it may be noted that the computed values of \(KS\) statistic by six probability distributions (using MoM and MLM) are lesser than the theoretical value at 5% significance level and therefore all six distributions are acceptable for estimation of MFD. Also, from Table 5, it may be noted that the computed values of \(KS\) statistic by GAM, PR3 and GEV distributions (using MoM and MLM) are lesser than the theoretical value at 5% significance level and at this level these three distributions are acceptable for estimation of MFD at Ghala.

4.3. Analysis Based on Diagnostic Test

For the selection of the best suitable distribution for estimation of MFD, the D-index values of six probability distributions are computed from Eq. (7) and given in Table 6.

<table>
<thead>
<tr>
<th>Return period (year)</th>
<th>MoM Estimated MFD (m³)</th>
<th>MLM Estimated MFD (m³)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EXP</td>
<td>GAM</td>
</tr>
<tr>
<td>2</td>
<td>2358</td>
<td>2335</td>
</tr>
<tr>
<td>5</td>
<td>5595</td>
<td>5562</td>
</tr>
<tr>
<td>10</td>
<td>8044</td>
<td>8028</td>
</tr>
<tr>
<td>20</td>
<td>10493</td>
<td>10504</td>
</tr>
<tr>
<td>50</td>
<td>13731</td>
<td>13787</td>
</tr>
<tr>
<td>100</td>
<td>16180</td>
<td>16276</td>
</tr>
<tr>
<td>200</td>
<td>18629</td>
<td>18768</td>
</tr>
<tr>
<td>500</td>
<td>21867</td>
<td>22065</td>
</tr>
<tr>
<td>1000</td>
<td>24316</td>
<td>24562</td>
</tr>
</tbody>
</table>

Figure 1. Plots of recorded and estimated MFD by six probability distributions (using MoM and MLM) for river Tapi at Dedtalai

Figure 2. Plots of recorded and estimated MFD by six probability distributions (using MoM and MLM) for river Tapi at Ghala

Table 3. MFD estimates for different return periods for river Tapi at Dedtalai
Following observations are drawn from the study.

By using the diagnostic test results presented in Table 6, the following observations are drawn from the study.

i) The indices of D-index of 2.881 (using PR3) for Dedtalai and 3.965 (using GPA) for Ghala are comparatively minimum when MoM is applied for determination of parameters of the distributions.

ii) Likewise, the indices of D-index of 2.865 (using GEV) for Dedtalai and 5.057 (using EXP) for Ghala are comparatively minimum when MLM is applied for determination of parameters of the distributions.

iii) \( \chi^2 \) test results don’t support the use of PR3, GEV and GPA distributions (using MoM and MLM) for estimation of MFD at Dedtalai.

iv) Both \( \chi^2 \) and KS tests result don’t support the use of EXP, EV1 and GPA distributions (using MoM and MLM) for estimation of MFD at Ghala.

v) Based on the eliminations of the probability distributions have minimum D-index through GoF (\( \chi^2 \) and KS) tests results, it may be noted that:

a) D-index value of 3.307 computed by GAM (using MoM) is the next minimum when compared to the corresponding values of EXP and EV1 for Dedtalai.

b) For Ghala, it may be noted that the D-index value of 4.077 computed by PR3 distribution (using MoM) is the next minimum when compared to the corresponding values of GAM and GEV.

vi) From the research studies, it is observed that the estimated parameters of distributions fitted by MoM are often less accurate than those obtained by MLM. So, the selection of a suitable probability distribution is made through quantitative (using D-index) and qualitative (using probability plots) assessment.

a) The D-index values of 3.352 (for Dedtalai) and 5.556 (for Ghala) computed by EXP and GAM distributions (using MLM) are minimum when compared to the corresponding values of other probability distributions, which are supported by GoF tests.

b) By considering the trend lines of the fitted curves using estimated MFD values, the study identifies the EXP distribution (using MLM) is found to be a good choice for estimation of MFD for Dedtalai whereas GAM distribution (using MLM) for Ghala.

### Table 4. MFD estimates for different return periods for river Tapi at Ghala

| Return period (year) | EXP | GAM | PR3 | EV1 | GEV | GPA | EXP | GAM | PR3 | EV1 | GEV | GPA
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>60</td>
<td>100</td>
<td>145</td>
<td>190</td>
<td>230</td>
<td>270</td>
<td>310</td>
<td>350</td>
<td>400</td>
<td>450</td>
<td>500</td>
<td>550</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>240</td>
<td>280</td>
<td>320</td>
<td>360</td>
<td>400</td>
<td>440</td>
<td>480</td>
<td>520</td>
<td>560</td>
<td>600</td>
<td>640</td>
</tr>
<tr>
<td></td>
<td>700</td>
<td>750</td>
<td>800</td>
<td>850</td>
<td>900</td>
<td>950</td>
<td>1000</td>
<td>1050</td>
<td>1100</td>
<td>1150</td>
<td>1200</td>
<td>1250</td>
</tr>
</tbody>
</table>

### Table 5. Computed and theoretical values of GoF tests statistic

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Dedtalai</th>
<th>Ghala</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXP</td>
<td>4.143</td>
<td>0.119</td>
</tr>
<tr>
<td>GAM</td>
<td>4.143</td>
<td>0.122</td>
</tr>
<tr>
<td>PR3</td>
<td>9.143</td>
<td>1.052</td>
</tr>
<tr>
<td>EV1</td>
<td>3.786</td>
<td>1.246</td>
</tr>
<tr>
<td>GEV</td>
<td>4.500</td>
<td>0.452</td>
</tr>
<tr>
<td>GPA</td>
<td>6.286</td>
<td>0.587</td>
</tr>
</tbody>
</table>

### Table 6. D-index values of probability distributions

<table>
<thead>
<tr>
<th>Gauging site</th>
<th>EXP</th>
<th>GAM</th>
<th>PR3</th>
<th>EV1</th>
<th>GEV</th>
<th>GPA</th>
<th>EXP</th>
<th>GAM</th>
<th>PR3</th>
<th>EV1</th>
<th>GEV</th>
<th>GPA</th>
</tr>
</thead>
</table>
5. Conclusions

The paper describes briefly the study carried out for estimation of MFD by adopting FFA (using VTFIT software) for determination of parameters of six probability distributions (using MoM and MLM) for Dedtalai and Ghala. The following conclusions are drawn from the study:

i) The study presents the selection of suitable distribution evaluated by GoF (using $\chi^2$ and KS) and diagnostic (using D-index) tests.

ii) The $\chi^2$ test results showed that the EXP, EV1 and GAM distributions (using MoM and MLM) are acceptable for estimation of MFD at Dedtalai.

iii) The $\chi^2$ test results showed that the EXP, EV1, GAM, GEV, GPA and PR3 distributions are not acceptable for estimation of MFD at Ghala when MoM and MLM is applied for determination of parameters of distributions.

iv) The KS test results indicated that these six probability distributions are acceptable for estimation of MFD at Dedtalai whereas GAM, GEV and PR3 distributions are acceptable for Ghala when MoM and MLM is applied for determination of parameters of distributions.

v) By considering the trend lines of the fitted curves using estimated MFD values, the study presented that the EXP distribution (using MLM) is better suited amongst six distributions adopted for estimation of MFD for Dedtalai whereas GAM distribution (using MLM) for Ghala.

vi) The study suggested that the MFD values computed by EXP (for Dedtalai) and GAM (for Ghala) distributions (using MLM) could be considered as the design parameter for planning and design of hydraulic structures in the vicinity of the gauging sites.

Acknowledgements

The author is grateful to Shri S. Govindan, Director, Central Water and Power Research Station, Pune, for encouragement given for conducting the studies and also accordingly permission to publish this paper. The author is thankful to the Executive Engineer (Tapi Division), Central Water Commission for providing stream flow data used in the study.

References


[16] United States Water Resources Council (1981); Guidelines for determining flood flow frequency, Bulletin No. 17B.
