Best fitted distribution for estimation of future flood for Rapti river systems in Eastern Uttar Pradesh

R Yadav
B R Ambedkar Environmental & Water Management Centre, 399, Zaruri Bazar, Ranikhet 263 645, India
&
B B Lal Pande
M. M. M. Engineering College, Gorakhpur 273 010, India
Received 30 July 1996; accepted 12 February 1997

The flood frequency analysis studies are the most important means of estimating floods for various return periods. The expected frequency for normal, log-normal, extreme value, gamma and Pearson type III distributions have been calculated and histograms have been plotted. Goodness of fit criterion has been tested by Chi-square test and best fitted distribution has been recommended for the estimation of future flood for Rapti river systems.

The magnitude of floods and their frequencies are often needed for planning and design of various hydraulic structures. All rivers are subjected to a non-zero probability that any particular level of flow will be equalled or exceeded. With each level of flow, there is a frequency with which a flow at that quantity or greater will be experienced over a very long time interval. The frequency can be used to define a return period which represents an average time elapsed between floods of that magnitude or greater. Any hydraulic structure whose failure would seriously endanger human lives should be designed to pass safely the greatest flood that will probably ever occur at that point. The maximum flood that any hydraulic structure can safely pass is called the design flood.

Flood Frequency Analysis
The flood frequency analysis (FFA) studies are the most important means of estimating floods in systematic manner which may be carried out at such gauging sites, where the historical peak discharges are available for sufficiently long period. The peak flood data used for frequency analysis are normally supposed to satisfy the following assumptions. (i) The data should be random, homogeneous and good quality. (ii) The sample size should be such that the population parameters can be estimated from it.

Procedure for FFA
In flood frequency analysis procedures generally the following steps are involved.
Process the historical records from frequency analysis point of view.
Choose various theoretical frequency distributions.
Fit the chosen frequency distributions with the historical flood records. Estimate the parameters of the distributions using one or more parameter estimation techniques.
Choose some of goodness of fit criteria and select the best fit distribution based on those criteria.
Estimate the floods for different recurrence intervals using the estimated parameters of best fit distribution.

The present study is based on the annual flood series (AFS) model.
There are various distributions and methods of parameter estimation techniques available in the flood frequency analysis literature for fitting the peak flood data for the purpose of flood frequency analysis. A large number of peak flow distributions available in literature among them the Normal, lognormal, extreme value, gamma distribution with 2 parameters, Pearson Type III have been commonly used in most of flood frequency studies.
Frequency Histograms for River Rapti at Birdghat

River Rapti emerges from Gaunra range in Nepal Himalaya at an elevation of 3048 m above mean sea level and outfalls into river Ghagra near Kaparwarghat in Deoria district. The catchment of Rapti river is shown in Fig. 1. The annual peak discharge data from 1960 to 1989 of river Rapti at Birdghat is given in Table I.

Frequency histograms for the data for river Rapti at Birdghat was first prepared to test the distributions. Various frequency distributions mentioned earlier will be studied with their frequency histograms and suitability of these distributions or goodness of fit of the distributions with data have been tested by Chi-Square Test. Once the goodness of particular distribution is established reasonably well, the discharge values predicted by the equation can be recommended for use with proper justification to river Rapti and tributaries and other rivers of Eastern Uttar Pradesh as the hydrological and morphological characteristics of the rivers are likely the same.

Normal Distribution

The general equation for the normal curve is

\[ y = f(Q) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\left(Q - \bar{Q}\right)^2 / 2\sigma^2}, \]

\[ -\infty \leq Q \leq +\infty \]  

(1)

Where \( y = f(Q) \) is the height of the curve of probability density distribution of \( Q \)

\( Q \) = annual peak discharge in the river

\( \bar{Q} \) = average of annual peaks

\( N \) = number of observations available

\( \sigma \) = standard deviation of series of annual maxima

This curve has a smooth symmetric bell shape whose highest ordinate the mode corresponds to the mean of the population \( Q \). Its coefficient of Skewness is zero and its coefficient of Kurtosis is 3.0.

Fitting of Normal Distribution to Plotted Histograms—The expected frequency can be calculated by formula

\[ y_i = \left( N_i / \sigma \sqrt{2\pi} \right) e^{-\left(Q_i - \bar{Q}\right)^2 / 2\sigma^2} \]  

(2)

where

\( N_i \) = frequency in class interval \( i \)

\( y_i \) = the frequency for a discharge \( Q_i \)

Table 1—Annual peak flood discharge of river Rapti at Birdghat in descending order of magnitude

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Discharge (cumecs)</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5346.50</td>
<td>1989</td>
</tr>
<tr>
<td>2</td>
<td>5256.00</td>
<td>1974</td>
</tr>
<tr>
<td>3</td>
<td>4217.93</td>
<td>1988</td>
</tr>
<tr>
<td>4</td>
<td>4154.70</td>
<td>1982</td>
</tr>
<tr>
<td>5</td>
<td>4072.00</td>
<td>1975</td>
</tr>
<tr>
<td>6</td>
<td>4000.00</td>
<td>1962</td>
</tr>
<tr>
<td>7</td>
<td>3747.73</td>
<td>1984</td>
</tr>
<tr>
<td>8</td>
<td>3700.00</td>
<td>1978</td>
</tr>
<tr>
<td>9</td>
<td>3648.30</td>
<td>1981</td>
</tr>
<tr>
<td>10</td>
<td>3600.00</td>
<td>1961</td>
</tr>
<tr>
<td>11</td>
<td>3513.00</td>
<td>1969</td>
</tr>
<tr>
<td>12</td>
<td>3163.46</td>
<td>1983</td>
</tr>
<tr>
<td>13</td>
<td>2908.00</td>
<td>1971</td>
</tr>
<tr>
<td>14</td>
<td>2903.87</td>
<td>1986</td>
</tr>
<tr>
<td>15</td>
<td>2857.00</td>
<td>1963</td>
</tr>
<tr>
<td>16</td>
<td>2800.00</td>
<td>1980</td>
</tr>
<tr>
<td>17</td>
<td>2800.00</td>
<td>1964</td>
</tr>
<tr>
<td>18</td>
<td>2668.93</td>
<td>1987</td>
</tr>
<tr>
<td>19</td>
<td>2507.00</td>
<td>1973</td>
</tr>
<tr>
<td>20</td>
<td>2326.12</td>
<td>1985</td>
</tr>
<tr>
<td>21</td>
<td>2291.00</td>
<td>1970</td>
</tr>
<tr>
<td>22</td>
<td>2227.00</td>
<td>1968</td>
</tr>
<tr>
<td>23</td>
<td>2163.82</td>
<td>1979</td>
</tr>
<tr>
<td>24</td>
<td>2154.00</td>
<td>1976</td>
</tr>
<tr>
<td>25</td>
<td>2000.00</td>
<td>1972</td>
</tr>
<tr>
<td>26</td>
<td>1972.00</td>
<td>1965</td>
</tr>
<tr>
<td>27</td>
<td>1864.00</td>
<td>1967</td>
</tr>
<tr>
<td>28</td>
<td>1776.00</td>
<td>1977</td>
</tr>
<tr>
<td>29</td>
<td>1690.00</td>
<td>1966</td>
</tr>
</tbody>
</table>

NORMAL DISTRIBUTION

The observed and calculated frequencies are given in Table 2 and plotted in Fig. 2.

Log-normal distribution

If \( Q \) is replaced by its logarithmic values in Eq. (1), one can write

\[
f (Q) = (1/2.303) \varphi (\log Q) \tag{6}
\]

Expressing \( \varphi (\log Q) \) in terms of normal distribution by analogy of Eq. (1) one can write

\[
f (Q) = (0.4342/Q \sigma_m \sqrt{2 \pi}) e^{-(\log Q - \log \bar{Q})^2/2\sigma^2} \tag{7}
\]

Where \( \sigma_m \) is the standard deviation of \( \log Q \).

A theoretical basis is provided for the log-normal distribution by considering causative factors as having positive and multiplicative, rather than additive effects. Hence, the logarithms of factors should satisfy the basic four conditions for normal distribution.

Fitting of Log-normal distribution to plotted histogram—The frequency can be calculated by the formula

\[
y_i = (N/2.303 \sigma_m ) e^{-(\log Q - \log \bar{Q})^2/2\sigma^2} \tag{8}
\]

The calculated and observed frequencies are given in Table 3 and plotted in Fig. 3.

EXTREME VALUE DISTRIBUTION

Type I extreme value distribution in hydrology has come to be known as Gumbel's distribution. Gumbel made use of distribution of largest values as suggested by Fisher and Tippet. Type I extreme value distribution is also known as Double exponential distribution

Gumbel's Law of Extreme Value Distribution Assumptions

The distribution of extreme value is of exponential type. Number of observations are sufficiently large. Observations are independent.

This distribution is represented as

\[
F(Q) = e^{-e^Q} \quad -\infty < Q < \infty \tag{9}
\]
Where \( y = a (Q - u) \)

For a particular return period \( T \), this equation can be written as

\[
Y_T = a (Q - u) \quad \ldots \quad (10)
\]

where, \( Q_T \) = estimated discharge for return period \( T \) and

\[
Y_T = -\ln[\ln F(Q)] \quad \ldots \quad (11)
\]

The value \( F(Q) \) is taken to be the probability \( P \) of a variate \( Q \) and this is equal to \((1 - 1/T)\).

Where \( T \) is the return period.

Thus the value \( Y_T = \ln[\ln T/(T-1)] \quad \ldots \quad (12) \)

For estimation of parameters \( a \) and \( u \) of this, method of probability paper, method of moment and maximum likelihood (MLH) method, approximate MLH has been used and given in Table 4 and the comparison between MLH and approximate MLH method is given in Table 5.

**Histogram for extreme value distribution (Gumbel distribution) by probability paper**

<table>
<thead>
<tr>
<th>Name of River</th>
<th>Parameter</th>
<th>Probability paper</th>
<th>Method of MLH</th>
<th>Approximate MLH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rapti</td>
<td>( a )</td>
<td>1.14x10^3</td>
<td>2.9x10^3</td>
<td>2.9x10^3</td>
</tr>
<tr>
<td></td>
<td>( u )</td>
<td>2574</td>
<td>2606</td>
<td>2569</td>
</tr>
<tr>
<td>Birdghat</td>
<td>( a )</td>
<td>1.29x10^3</td>
<td>2.9x10^3</td>
<td>2.9x10^3</td>
</tr>
<tr>
<td></td>
<td>( u )</td>
<td>2569</td>
<td>2606</td>
<td>2569</td>
</tr>
</tbody>
</table>

**Histogram for Gamma Distribution**

The calculated frequency is given by

\[
y_i = \frac{N_i}{\beta^\alpha \Gamma(\alpha)} Q_i^{\alpha-1} e^{-Q_i/\beta} \quad \ldots \quad (16)
\]

The calculated frequencies are given in Table 7 and are shown in Fig. 5.

<table>
<thead>
<tr>
<th>Class interval (1000 Cumec)</th>
<th>Mid value ( Q_i ) (1000 Cumec)</th>
<th>Observed frequency</th>
<th>Calculated frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-1</td>
<td>0.5</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>1-2</td>
<td>1.5</td>
<td>5</td>
<td>6.2</td>
</tr>
<tr>
<td>2-3</td>
<td>2.5</td>
<td>12</td>
<td>11.4</td>
</tr>
<tr>
<td>3-4</td>
<td>3.5</td>
<td>7</td>
<td>7.4</td>
</tr>
<tr>
<td>4-5</td>
<td>4.5</td>
<td>3</td>
<td>2.9</td>
</tr>
<tr>
<td>5-6</td>
<td>5.5</td>
<td>2</td>
<td>0.8</td>
</tr>
</tbody>
</table>

**Fig. 4—Comparison of observed and calculated frequencies by Gumbel distribution**
Pearson type III Distribution

The function

\[ f_\alpha(q) = \frac{1}{\beta^\alpha \Gamma \alpha} (q - r)^{\alpha - 1} e^{\frac{q-r}{\beta}} \]

\( f_\alpha \) is the three parameter probability density function. The parameter \( r \) is known as location parameter. In hydrology for frequency distribution, the use of Gamma distributions with three parameters becomes very common in the form of Pearson Type III distribution.

Pearson has derived a series of probability density function to fit virtually any distribution. They have been widely used in practical statistical work to define the shape of many distribution curves. Pearson Type III distribution is a special case of this distribution which is often used in hydrological frequency analysis. In this the skewness coefficient is given by

\[ r = \frac{\sum (Q_i - \bar{Q})^3}{(N - 1)\sigma^3} \]

And the value of \( \alpha, \beta \) and \( r \) in Eq. (17) is given by

\[ \alpha = 4 / r_1 \]

\[ \beta = \sigma \sqrt{\frac{r}{2}} \]

\[ r = \frac{\bar{Q} - (2\sigma / r_1)}{2} \]

Testing of the Distributions

To select a particular distribution for the purpose of prediction of future floods, tests of the goodness of fit have been done. The Chi-Square Test of fit is commonly used.

The Chi-Square Test—It has been shown that the Chi-Square parameter

\[ \chi^2 = \sum_{i=1}^{k} \frac{(b_i - c_i)^2}{c_i} \]

where

\( b_i \) = number of observations actually in a given class interval

\( c_i \) = Expected number of observations in a given class interval

\( i = 1, \ldots, k \) = class interval covering the range of data.

The Chi-Square test prescribes the critical value of \( \chi^2 \) for a given significance level \( \alpha \), so that for \( \chi^2 < \chi^2_\alpha \) the null hypothesis of a good fit is accepted, otherwise for \( \chi^2 \geq \chi^2_\alpha \), it is rejected.
In general, the number of class interval \( k \) should be greater than 5 and the expected values of absolute frequencies should be \( c_i > 5 \). If the parameters of distribution function estimated from the sample data has a number \( h \), then the number of degree of freedom \( \text{N.D.F} = k - h - 1 \). The values of \( \chi^2 \) can be obtained for a given N.D.F at a particular level of significance (\( \alpha \)) usually taken as 5% from the standard text in statistics. Values of \( \chi^2 \) calculated are listed in Table 9.

**Conclusion**

It can be seen from Table 9 that all the theoretical distribution pass the test of acceptance. The Log-normal distribution in each of these tests gives lowest value as compared with other distributions. Also from perusal of histograms (Figs 2-6), Log normal distribution is best fitted.

Hence, Log-normal distribution method can be inferred to be the best distribution for calculation of future floods at various sites for river Rapti and its tributaries and other river systems in Tarai region of Eastern Uttar Pradesh.

**References**