THE BEST FITTED DISTRIBUTION FOR ESTIMATING
ANNUAL PEAK FLOWS IN BENUE RIVER,
UPPER BENUE RIVER BASIN TROUGH

BY

Tanko¹, J.A. and Agunwamba², J.C.
1. Department of Civil Engineering
   Federal University of Technology, Yola
2. Department of Civil Engineering,
   University of Nigeria, Nsukka

ABSTRACT
Frequency analysis is widely applied for estimating extreme flows and precipitation. However, its merits for applicability based on the various distributions available for different data and purposes have not been clearly established especially for River Benue in the Upper Benue River Basin. The annual peak flows of river Benue were taken at the Jimeta – Yola, bridge gauging station in Adamawa State for a 45 – years record from 1960 – 2004. The record was fitted with Normal, Log-Normal, Gumbel’s Extreme values, Gamma, and Pearson type III distributions. Log-Normal distribution was found most suitable for annual peak flow estimation in the Benue River based on the goodness of fit test using Chi-Square test.

Key words: Peak flows, Frequency analysis distributions and goodness of fit

1. INTRODUCTION
Large peak flows (floods) usually impact man and his property. Several hydrological models are in use among which are LISFLOOD.FP and GLUE models (1-3) used for flood forecasting and for assessing predictive uncertainty respectively. Frequency analysis is used for analyzing stations with more than 10 years peak flow data. However, the choice of the distribution to be fitted for analysis remains a problem for the Hydrologist. WMO (4) outlined the importance of adopting the best fitted distribution for each river Basin in various regions. Yadav and Lal (1998) selected Log-normal distribution as the best fitted distribution for Rapti river system in Eastern Urtta Padesh, India. Several distributions are available but the applicability of each to a particular situation had not been checked especially for the Upper Benue Basin Rivers. The use of appropriate distribution would ensure good estimation of design floods for hydraulic structures design.

Today more than 40% of the catastrophes are caused by flooding alone with an average of about 10,000 deaths per year as observed by Lafittle and Bartle (5) and Hunt (6). Notable flood associated
2. Theory of Frequency Analysis

The following are the theory of some selected distributions and their fitting techniques:

2.1 Normal Distribution: The general equation for the Normal distribution curve is given by (4, 13, 14) as

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

For $-\infty < x < +\infty$

Where, $f(x)$ = the height of the curve of probability density distribution of variable $x$

$x$ = random variable, $\mu$ = variable mean, $\sigma$ = standard deviation of variables and $f$ = frequency of discharge $Q_i$

The above equation was fitted to the peak flow variables as

$$y_i = \frac{N_i}{\sigma\sqrt{2\pi}} e^{-\frac{(Q_i - \mu)^2}{2\sigma^2}}$$

Where $N_i$ = frequency in class interval, $i$; $y_i$ = frequency for discharge; $Q_i$ and $Q_m$ = the discharge in class interval $i$ and mean discharge respectively.

$$Q_m = \frac{\Sigma iq_i}{\Sigma f}$$

$$\sigma = \left(\frac{\Sigma f(Q_i - Q_m)^2}{\Sigma f}\right)^{1/2}$$

data ... (4)

2.2. Long-Normal Distribution: the peak flow discharge values ($Q$) were replaced by their logarithmic values in equation (2) as suggested by Yevjevich (15)

$$f(Q)|\text{ABS}(dQ) = \Phi(log Q)|\text{ABS}(log dQ)$$

Where, $\text{ABS}$ = absolute value of the parenthesis.
For $Q \geq 0$ and $f(Q) = \frac{1}{2.303} \Phi(\log Q)$

And by analogy, $\Phi(\log Q)$ could be expressed in terms of normal distribution and equation (5) was written as

$$f(Q) = \frac{0.4342}{Q^2} e^{-\log Q - \log Q_m^2/2\sigma^2}$$

Where, $\sigma_m$ = standard deviation of log $Q$

$logQ_m$ = mean of log $Q$

A theoretical basis exists for the log – normal distribution by considering causative factors as having positive and multiplicative, rather than additive effects. Therefore, the logarithms of factors should satisfy the basic four conditions for normal distributions as suggested by Yadav and Lal (5). After simplification, equation (7) was fitted to the flood discharges as

$$y_i = \frac{N_i}{2.303} \frac{e^{-\log Q - \log Q_m^2/2\sigma^2}}{Q^2}$$

Where, all terms retain their earlier meaning.

2.3. Gumbel’s Extreme Values: The following assumptions were made for this distribution.

P The distribution is of exponential type.

P The number of observations should sufficiently large, and

P Observations are assumed to be independent variables.

The Gumbel’s Extreme Value distribution is given as suggested by Yadav and Lal (5) and Subramanya (13) as

$$f(Q) = e^{-\alpha - e^{-\alpha}}$$

For $-\infty < Q < \infty$ and $Y = \alpha(Q - U)$

The parameters $\alpha$ and $U$ were estimated using the method of Maximum Likelihood (MLH). The values of $\alpha$ and $U$ estimated from the peak flow discharge record were $1.15 \times 10^{-3}$ and 3620 respectively.

The annual peak flow discharges were fitted with

$$y_i = N_i e^{-\alpha(Q - U)} e^{-e^{-\alpha(U - \alpha)}}$$

2.4 Gamma Distribution with 2 Parameters

The general form of this function is given by

$$f(x) = \frac{1}{\beta \Gamma(a)} x^{a-1} e^{-x/\beta}$$

Equation (12) was fitted to the peak flow record by

$$y_i = \frac{N_i}{\beta \Gamma(a)} Q_i^{a-1} e^{-Q_i/\beta}$$

Where, $N_i$ and $y_i$ retain their earlier meaning, $\alpha =$ shape parameter and $\beta =$ scale parameter

The values of $\alpha$ and $\beta$ were estimated in terms of $Q_m$ and $\sigma$ using the relations suggested by Yadav and Lal (5) as

$$\alpha = \frac{Q_m^2}{\sigma^2}$$

$$\beta = \frac{\sigma^2}{Q_m}$$

Where, $Q_m$ and retain their earlier meanings.

Pearson Type III Distribution: The general function for this distribution is as suggested by Mustafa and Yusuf (14) Sabramanya (13).

$$f(x) = \frac{1}{\beta \Gamma(a)} (x_i - \alpha)^{a-1} e^{-(x_i - \alpha)/\beta}$$
\[ f_a(x) = \frac{1}{\beta \Gamma(a)} (x_i - r)^{a-1} e^{-\frac{(x-r)}{\beta}} \]

Where \( f_a \) = three parameter probability density function, \( r \) = location parameter. The Pearson type IIII distribution is a special case of distribution often used in hydrological frequency analysis. The equation used for this distribution is given by:

\[ y_i = \frac{N_i}{\beta \Gamma(a)} (Q_i - r)^{a-1} e^{-\frac{(Q_i-r)}{\beta}} \]

The skewness coefficient is given by Pearson (15) as:

\[ r_i = \sum (Q_i - Q_m)^2 / (N - 1) \]

Where, \( a = \frac{4}{r_1^2} \)

\[ \beta = \sigma r_i / 2 \]

\[ r = Q_m - (2\sigma / r_1) \]

RESULTS

4.0

The observed and estimated frequencies of the peak flow discharges are shown in Table 1.0. The plot is shown in Figure 1.0. The values for \( \sigma \) and \( Q_m \) were estimated from the peak flow data as 1.8276 ×10^3 and 4.1911 ×10^3 m^3/s respectively.

The estimated and observed frequencies of the peak flow discharges and the plot are shown in Table 1.0 and Figure 1.0 respectively. The frequencies were re-estimated using equation (11). The observed and the estimated frequencies for the peak flow discharges are shown in Table 1.0. The plot is shown in Figure 1.0.

The parameters \( \alpha \) and \( \beta \) were estimated from the peak flow record as 5 and 0.7978 respectively. The estimated and the observed frequencies for peak flows in the Benue River are shown in Table 1.0. The plot of estimated frequencies using various distributions is shown in Figure 1.0.

The parameters \( \alpha, \beta, r \) and \( r_1 \) were estimated from the peak flow record as 2.0587, 1.2744, 1.5674 and 1.3939 respectively. The observed and estimated frequencies for the peak flow and the plot are shown in Table 1.0 and Figure 1.0 respectively.

3. Materials and methods

Flood frequency analysis using Annual Flood Series (AFS) was adopted because of its suitability for sites with available historical peak flow records. A peak flow record of 45 years from 1960 to 2004 taken at the Jimeta – Yola Bridge gauging station was used for the analysis.

Guidelines for using frequency analysis were also suggested by Yadav and Lal (5) as follows:

After processing historical flow records, a theoretical frequency distribution is chosen.

Parameters for the distributions are estimated using the available techniques.

Goodness of fit criterion chosen for the best fitted distribution based on the criterion selected.

The floods for different reoccurrence intervals are then chosen using the estimated parameters of the best fitted distribution.

Peak flow estimation distributions selected in literature for this study include Normal, Log-Normal, Gumbel’s Extreme Values, Gamma with 2 parameters, and Pearson Type III (5).
Table 1.0: The observed and estimated frequencies for the Benue River for the various distributions.

<table>
<thead>
<tr>
<th>S/No</th>
<th>Class (x1000 m³/s)</th>
<th>Mid values (x1000 m³/s)</th>
<th>Observed Freq. (1)</th>
<th>Estimated Frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Normal (2)</td>
</tr>
<tr>
<td>1</td>
<td>0-1.59</td>
<td>0.80</td>
<td>3.00</td>
<td>2.80</td>
</tr>
<tr>
<td>2</td>
<td>1.60-2.59</td>
<td>2.10</td>
<td>7.00</td>
<td>4.00</td>
</tr>
<tr>
<td>3</td>
<td>2.60-3.59</td>
<td>3.10</td>
<td>6.00</td>
<td>5.50</td>
</tr>
<tr>
<td>4</td>
<td>3.60-4.59</td>
<td>4.10</td>
<td>11.0</td>
<td>12.00</td>
</tr>
<tr>
<td>5</td>
<td>4.60-5.59</td>
<td>5.10</td>
<td>9.00</td>
<td>8.50</td>
</tr>
<tr>
<td>6</td>
<td>5.60-5.59</td>
<td>6.10</td>
<td>4.00</td>
<td>2.60</td>
</tr>
<tr>
<td>7</td>
<td>6.60-7.59</td>
<td>7.10</td>
<td>3.00</td>
<td>0.90</td>
</tr>
<tr>
<td>8</td>
<td>7.60-8.59</td>
<td>8.10</td>
<td>2.00</td>
<td>0.20</td>
</tr>
</tbody>
</table>

5. TEST OF GOODNESS OF FIT
The goodness of fit of the distributions to the annual peak flows of the Benue River in Upper Benue River Basin trough was done using Chi-Square test as used by Yadav and Lal (5). It is given by

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

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

\( b_i = \) Number of observations actually in a
given class interval, \( c_i = \text{Expected number of observations in a given class interval.} \)

In a Chi-Square test, a critical value \( \chi^2_0 \) of \( \chi^2 \) for a significance level \( \alpha \) so that for \( \chi^2 < \chi^2_0 \), Null hypothesis of good fit is accepted. And for \( \chi^2 \geq \chi^2_0 \), Null hypothesis of good fit rejected. The value of \( \chi^2_0 \) is usually obtained for a given number of degree of freedom (NDF) at a particular level of significance \( \alpha \) usually taken as 5% from standard statistics text book. The estimated values of \( \chi^2 \) are shown in Table 2.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Chi-Square ( \chi^2 ) for 5% level NDF = 2, ( \chi^2_0 = 5.99 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Normal</td>
<td>24.27</td>
</tr>
<tr>
<td>2. Log-Normal</td>
<td>3.35</td>
</tr>
<tr>
<td>3. Gumbel’s extreme Values</td>
<td>172.56</td>
</tr>
<tr>
<td>5. Pearson Type III</td>
<td>20.30</td>
</tr>
</tbody>
</table>

6. CONCLUSION
This research determined the best distribution for estimation of the annual peak flows in Benue River. Records on annual peak flows of River Benue were collected at Jimeta – Yola, Adamawa State for a 45 – years record from 1960 – 2004. the data was fitted to five different distributions and their fitness compared using Chi-Square.

From the goodness of fit test, the Log-Normal distribution would be more suitable for annual peak flow estimation in the Benue River, Upper Benue Basin through.

Acknowledgment: This is to acknowledge the staff and management of the Upper Benue River Basin Development authority, Yola for providing all the necessary data for the research. My gratitude also goes to Engr. (Prof.) N. Egbuniwe for guidance and encouragement.

REFERENCES
9. Ojiako G.U. “Floods in Nigeria and their optimal management strategy”. Third World strategies for Technological Development, Federal University of