Nigerian Journal of Technology (NIJOTECH)
Vol. 32. No. 2. July 2013, pp. 184-196

ISSN: 1115-8443
www.nijotech.com

EVALUATION OF SOME FLOOD PREDICTION MODELS FOR THREE FLOW GAUGING STATIONS IN UPPER BENUE RIVER BASIN IN NIGERIA – PART 1

O. C. Izinyon [a,*], H. N. Ajumuka [a,+]
[a] DEPARTMENT OF CIVIL ENGINEERING, UNIVERSITY OF BENIN, BENIN CITY. NIGERIA
*(Corresponding author): izinyon2006@yahoo.com , [+] henreo2kool84@yahoo.co.uk

Abstract
The annual maximum series of flow discharge data for three flow gauging stations located at River Donga at Manya,, River Donga at Donga and River Bantaji at Suntai within Upper Benue River Basin in Nigeria were fitted each with three probability distribution models namely ;Log normal, Extreme value Type 1 and Log Pearson Type III. The model results were subjected to four specific measures of error in prediction (i.e., RMSE, RRMSE, CC and MAE) and a scoring scheme on the basis of which best fit model for each station was selected. The best fit probability distribution models for the stations are Extreme value Type 1 (EV-I), Log Pearson Type III and Log normal for the stations at River Donga at Manya, River Donga at Donga and River Bantaji at Suntai respectively. The models can provide estimates of flood quantiles for planning, design, construction and operation of water resources projects within the river systems.

Keywords: Discharge, probability distribution model, return period, gauging station, prediction error levels

1. Introduction

Though flood water is an essential resource in many countries and flood plains hold many benefits for society, they can also be the causes of huge losses of lives, livelihoods and property; and can be a hindrance to socio-economic development. Floods are one of the most destructive natural disasters that occur in most parts of the world and have been identified as the most costly natural hazards having great propensity to destroy human lives and properties [1]

There is also the general concern that the risks resulting from hydrological extremes are on the increase and this is supported by evidence both from recent changes in frequency and severity of floods as well as droughts and outputs from climatic models which predict increases in hydrological variability [2].Thus, the need for preventive action to reduce unnecessary cost and economic loss as well as preventing the danger of overflow of water is urgent.

To manage flood risks successfully, knowledge is needed of both magnitude of any given flood and an estimate of likelihoods of occurrence. The design and construction of certain projects such as dams and urban drainage systems, the management of water resources and prevention of flood damage requires adequate knowledge of extreme events of high return periods [3]. Similarly, estimates of the magnitude of the flood in a certain return period which may be achieved by the method of flood frequency analysis is useful to the water resources engineer in the quantitative assessment of past flood events when evaluating future possibilities of such occurrence [4]. This type of information is used extensively for urban planning and development, flood plain management, establishment of insurance premiums and for efficient design and location of hydraulic structures [5, 6].

Flood frequency analysis is generally taken to denote a statistical analysis of flood, their magnitudes and or their frequency (recurrence intervals) because flood risk estimation is an inherently statistical problem. To derive the risk of occurrence of any flood event, the frequency distribution which can best describe the past characteristics of the magnitude and the possibility of such flood must be known and this requires determination of the most appropriate flood frequency model which can be fitted to the available historical data or record. The selection of the most appropriate distribution for annual maximum series has received widespread attention and a growing concern in flood studies is the choice of frequency distribution for fitting extreme flood series in a region, this is particularly challenging in developing countries because of dearth of data [7]. The main difficulty with short records is that conventional moment statistics are both highly biased and highly variable in small samples [8].

At present, there is no universally accepted frequency distribution model for frequency analysis of extreme floods, rather a whole group of models such as Gumbel (EV-1), Normal, Log normal, Pearson Type III, Log Pearson Type III etc have been suggested in the literature such as [7] and [9] for the prediction of extreme flood events. The selection of an appropriate model depends mainly on the characteristics of available discharge data at the appropriate site and also based on the outcome of some statistical tests to determine the best fitted distribution for any specific site.

In developing countries like Nigeria, basic planning data are scarce and the collation efforts are still at the infancy stages giving room for more research on the obtained data to avert the net effects of the uncertainties which have economic penalties resulting from imperfect planning, over or under design and wrong management decisions [10].

In this paper, the results of the first part of a study made to determine which flood frequency distribution model adequately fits the statistical characteristics of observed flood data for some flow gauging stations in the Upper Benue River Basin in Nigeria are presented. The main objective of this particular study was to apply three commonly utilized probability distribution models to flow or discharge data obtained from three flow gauging sites in the river basin with a view to evaluating their performance in accurately predicting extreme flood discharge estimates. The specific objectives of the study include:

(i) To fit Extreme value Type -1 (EV-I) , Log normal and Log Pearson Type III probability distribution models to observed peak flow data (1955 to 1986) obtained at three flow gauging stations within the river basin namely (River Donga at Manya, River Donga at Donga and River Bantaji at Suntai)
(ii) To apply specific measures of errors in prediction viz (RMSE,RRMSE, CC and MAE) to results obtained from (i) above and hence select best fit probability distribution model for observed data at each site
(iii) Based on selected best fit model, predict design floods for return periods of 2yrs,5yrs,10yrs,25yrs,50 yrs,100yrs,200yrs at each flow gauging station.

The three flow gauging sites whose data were utilized for this study are located in rivers situated within the Upper Benue Hydrological Area (HA-3) of Nigeria [11] which is one of the eight hydrological areas into which Nigeria is subdivided. Other important details relating to the study sites are given in Table 1

2. Basic Approach to Flood Prediction using Probability Distribution Functions

A probability density function (PDF) is a continuous mathematical expression that determines the probability of a particular event. If a prediction is to be based on a set of hydrologic data, then the distribution that best fits the set of data may be expected to give the best estimates usually an extrapolation of the probability of an event occurring. The three probability distributions selected for this study are Log normal, Extreme value type 1(EV-I) and Log Pearson Type III distributions. Their essential properties are given in Table 2.

Table 1: Important details relating to Gauging sites

Station	River	Latitude	Longitude	Drainage Area(Km[2])
Donga	Donga	7[0] 43'N	10[0] 05E	11,909
Manya	Donga	7[0] 19'N	10[0]14E	9040
Suntai	Bantaji	7[0] 55'N	10[0]21E	5815

Source: [11, 12]

Table 2: Probability distribution parameters in relation to sample moments [7, 13]

Distribution	Probability distribution function	Range	Equation of parameters in terms of sample moments
Log normal	$f(x) = \frac{1\sqrt{\pi}}{x\sigma} exp\left[\frac{-1}{2}\left(\frac{lnx-\mu}{\sigma}\right)^2\right]$	$x > 0$	$\mu_{y=\bar{y}}$ $\sigma_{y=s_y}$
Extreme value Type-1 (EV-I)	$f(x) = \frac{1}{\beta} exp\left[-\frac{x-u}{\beta} - exp\left(-\frac{x-u}{\beta}\right)\right]$	$(-\infty < x < \infty)$	$u = \bar{x} - 0.5772\beta$ $\beta = \frac{\sqrt{6}s_x}{\pi}$
Log Pearson Type III	$f(x) = \frac{(lnx-u)^{\gamma-1} exp\left[-\frac{(lnx-u)}{\beta}\right]}{\lvert\beta\rvert\,\Gamma(\gamma)}$	$lnx \geq u$	$u = \bar{y} - S_y\sqrt{\gamma}$ $\beta = \frac{\sqrt{\gamma}}{s_y}, \gamma = \left(\frac{2}{g_y}\right)^2$

2.1 Normal and Lognormal Probability distribution

The Normal distribution is the most familiar probability distribution [14]. Its PDF is given by:

$$f(x) = \frac{1}{\sigma\ \sqrt{2}\,\pi} \exp\left(-\frac{(x-\bar{x})^2}{2\sigma^2}\right) \quad (1)$$

It is defined by two distribution parameters; the mean ($\bar{x}$), and standard deviation (σ) evaluated by :

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (2)$$

where x_i is the magnitude of the i[th] event and N is the total number of events. The standard deviation (σ) which is a measure of the dispersion or spread of data set is given by:

$$\sigma = \left[\frac{\sum_{I=1}^{n}(x-\bar{x})^2}{N-1}\right]^{1/2} \quad (3)$$

The normal distribution describes many random processes it generally does not provide satisfactory fit for flood discharge and other hydrologic data [14]

A particular event x can be related to the probability of exceedence P by the following equation:

$$x = \bar{x} + k\sigma \quad (4)$$

where k is the frequency factor. Though the normal distribution is not well suited to hydrologic data, the related distribution; the lognormal distribution works reasonably well [14]. The Log normal distribution assumes that the logarithms of the discharge are themselves normally distributed.

The equation describing normal distribution is modified for use in the case of log normal distribution if the following substitution is made.

$$y_i = log x_i \quad (5)$$

With x replaced by y, the mean of the logarithms ($\overline{\log x}$) and standard deviation(σ_{logx}) becomes

$$\overline{\log x} = 1/N\sum_{i=1}^{N} log x_i \quad (6)$$

$$\sigma_{logx} = \left[\frac{\sum_{I=1}^{N}(logx-\overline{logx})^2}{N-1}\right]^{1/2} \quad (7)$$

The probability of exceedence is related to the occurrence of particular values if log values are used by the expression written as:

$$\log x = \overline{logx} + K\sigma_{logx} \quad (8)$$

2.2 Log Pearson Type III distribution

The problem with most hydrologic data is that an equal spread does not occur above and below the mean. The lower side is limited to the range from mean to zero while there is theoretically no limitation on the upper range thereby contributing to what is called a skewed distribution. The coefficient of skew (a) is defined mathematically by:

$$a_i = \frac{N\sum_{i=1}^{N}(x-\bar{x})^3}{(N-1)(N-2)\sigma^3} \quad (9)$$

To determine the skew(a_{logx}) when log values are used, equation (9) becomes:

$$a_{logx} = \frac{\sum_{i=1}^{N}(logx - \overline{logx})^3}{(N-1)(N-2)\sigma_{logx}^3} \quad (10)$$

It is to take account of the skew that may exist in data that the log Pearson type III distribution was developed to improve the fit [14]. The distribution uses three parameters namely: mean standard deviation and skew coefficient which are obtained using equations (6), (7) and (10) respectively. Equation (8) is used to define frequency factor.

2.3 Extreme value Type I (EV-I) distribution

The Extreme value Type 1 (Gumbel) distribution is one of the most commonly used distributions in flood frequency analysis. The distribution is based on theory of extremes and it is considered appropriate for this analysis as annual series data used for this study is composed of peak values (extreme values) for each year.

The PDF and other parameters relating to the distribution are given in Table 2

3. Materials and methods

3.1 Data and Analysis

The annual instantaneous flood peaks for flow gauging stations; River Donga at Manya, River Donga at Donga and River Bantaji at Suntai for the period (1955 -1986) were obtained from the publication [12] and analyzed. The data is presented in Table 3.

The observed data at each gauging station were ranked and evaluated with three probability distribution functions namely: Lognormal, EV-I and Log Pearson Type III with their corresponding plotting positions calculated using Blom, Gringorten and Cunnane formulae respectively as recommended in [7] in order to determine the best fit function. Four specific measures of errors in prediction (RMSE, RRMSE, CC, and MAE) and a scoring scheme were used for the selection of the best fit model.

3.1 .1: Lognormal distribution

Lognormal distribution was fitted to the observed data by first ranking the data and then taking logarithms of each variate to transform the original series of peak flow data into log domain. The mean $(\bar{y})$ and standard deviation (S_y) for the log transformed series were computed using equations (6) and (7) respectively. Blom plotting position formula was used since the logarithms of the data were to be fitted to a normal distribution [13]. The normal reduced variable (z) corresponding to an exceedence probability (p) was determined using the following equations [13]:

$$w = \left[ln\frac{1}{P^2}\right]^{1/2}, \quad 0<p \leq 0.5 \quad (11)$$

and

$$z = w - \frac{2.515517 + 0.802853w + 0.010328\,w^2}{1 + 1.432788w + 0.189269w^2 + 0.001308\,w^3} \quad (12)$$

where w is an intermediate variable defined by equation (11)

And for p> 0.5, 1-p is substituted for p in equation (11) and values of z computed in equation (12) are given a negative sign. MS EXCEL programming was utilized to facilitate the calculation process. The event magnitude with the same exceedence probability in the fitted lognormal distribution ; that is the flood for the T-year recurrence interval in the log domain was estimated using the frequency factor method using the equation; log Q = $(y_T) = \bar{y} + k_T\,S_y$ with $\bar{y}$ and S_y determined from the observed data and taking K_T = z . The estimated T-yr flood was transformed to the original domain by computing its exponent as thus:

$$X_T = 10^{(y_T)} \quad (13)$$

The results obtained are compared with log Q from the observed data. This was done for all 32 data set

Table 3: Annual Peak Discharge (m³/s) at Rivers in Upper Benue River Basin of Nigeria

S/N	Water Year	R.Donga at Manya (m^3/s)	R.Donga at Donga (m^3/s)	R. Bantaji at Sun tai (m^3/s)	S/N	Water Year	R.Donga at Manya (m^3/s)	R. Donga at Donga (m^3/s)	R. Bantaji at Suntai (m^3/s)
1	1955	1050	1776	710	17	1971	840	1350	600

S/N	Water Year	R.Donga at Manya (m^3/s)	R.Donga at Donga (m^3/s)	R. Bantaji at Sun tai (m^3/s)	S/N	Water Year	R.Donga at Manya (m^3/s)	R. Donga at Donga (m^3/s)	R. Bantaji at Suntai (m^3/s)
2	1956	1130	2047	750	18	1972	864	1510	630
3	1957	1080	1800	740	19	1973	850	1720	640
5	1959	602	1250	450	21	1975	846	1475	646
6	1960	735	1890	465	22	1976	845	1690	856
7	1961	726	1645	450	23	1977	584	1700	378
8	1962	982	1548	750	24	1978	854	2010	407
9	1963	970	2272	850	25	1979	965	1670	800
10	1964	858	1580	900	26	1980	940	2100	745
11	1965	1050	1730	508	27	1981	1440	2400	960
12	1966	1030	1850	670	28	1982	1550	1740	530
13	1967	1000	1560	632	29	1983	1020	1600	420
14	1968	947	1750	950	30	1984	1100	1490	600
15	1969	1040	2150	910	31	1985	900	1650	635
16	1970	835	1680	619	32	1986	1010	1930	700

3.1.2 Extreme value Type I (Gumbel) distribution

The fitting of EV-I distribution to the observed data was carried out using the following steps as given in [7]:

i. The variates of the annual flood series were ranked in descending order of magnitude
ii. Plotting position i.e. the probability of non exceedence corresponding to T-yr recurrence interval was assigned to each variate using Gringorten plotting position formula.
iii. The reduced variate (y_T) for the distribution corresponding to the different plotting position was computed using:

$$y_T = -\ln[ln\left(\frac{T}{T-1}\right)] \quad (14)$$

iv. The T-yr recurrence interval flood was estimated using the E V-I distribution given by

$$X_T = u + \beta y_T \quad (15)$$

v. For the EV-I fit, the frequency factor K_T is evaluated as:

$$K_T = \frac{-\sqrt{6}}{\pi}\left(0.5772 + ln(ln\left(\frac{T}{T-1}\right)\right) \quad (16)$$

3.1.3 Log Pearson Type III distribution

Log Pearson Type III distribution was fitted to the observed data by first ranking the data according to descending order of magnitude and then taking logarithms of each variate to transform the original series of peak flow data into log domain. Plotting position i.e. the probability of non exceedence corresponding to T-yr recurrence interval was assigned to each variate using the Cunnane Plotting position formula. The mean $(\bar{y})$, standard deviation (S_y) and coefficient of skewness (C_S) for the log transformed series were computed using equations (6) and (7) and (10) respectively. The frequency factor depends on the return period and coefficient of skewness (C_S). When (C_S) =0, the frequency factor (K_T is equal to the standard normal variable (z) and for $C_S \neq 0$, K_T was approximated using the equation in [15] as :

$$K_T = Z + (Z^2+1)K + \frac{1}{3}(Z^3-6Z)K^2 - (Z^2-1)K^3 + ZK^4 + \frac{1}{3}K^5 \quad (17)$$

where $K = \frac{C_S}{6}$. The value of z for a given return period was calculated using same procedure as was with log normal case, while K_T was obtained using equation (17) and $y_T = \bar{y} + K_T S_y$ and $X_T = 10^{y_T}$

3.1.4. Best fit Analysis (Statistical Test Criteria)

In order to determine the best probability distribution functions that describes the set of observed data at each gauging site, the three selected probability distribution models applied to the set of observed data at a particular station were subjected to statistical tests (specific measures of error in prediction). The tests chosen are Root mean square error (RMSE), Relative root mean square error (RRMSE), Maximum absolute error (MAE) and correlation coefficient (CC). The best fit is determined by means of a criterion depending on the differences between the observed and theoretical values obtained from using Probability distribution functions [14]. In order to judge the overall goodness of fit of each distribution a ranking scheme was utilized by comparing the four categories of test criteria based on the relative magnitude of the statistical test results. A distribution with the lowest RMSE, lowest RRMSE, lowest MAE or highest CC was given a score of 3. In the event of a tie, equal scores are given to the distributions and for each test category. In order to determine the best fit model at each station, the overall score of each distribution was obtained by summing the individual point score at each of the three stations and the distribution with the highest total score was chosen as the best fit distribution model.

3.1.4.1: Root mean square error (RMSE)

The root mean square error is the sum of the squares of the squares of the differences between the observed and predicted values and is given by:

$$RMSE = \left(\frac{\sum (x_i - y_i)^2}{(n-m)}\right)^{\frac{1}{2}} \quad (18)$$

where x_i , i=1,---------------,n are the observed values and y_i , i=1,-------,n are the values computed from the assumed probability distributions, the number of parameters estimated for the distribution is denoted by m.

3.1.4.2: Relative Root mean square error (RRMSE)

This is defined as;

$$RRMSE = \left(\frac{\sum \left(\frac{x_i - y_i}{x_i}\right)^2}{(n-m)}\right)^{\frac{1}{2}} \quad (19)$$

RRMSE calculates each error in proportion to the size of observation thereby reducing the influence of outliers which are common features of hydrological data [3] and thereby providing a better picture of the overall fit of a distribution.

3.1.4.3: Correlation Coefficient (CC)

The correlation coefficient (CC) is defined mathematically as:

$$CC = \frac{\sum[(X_i - \bar{X})(y_i - \bar{y})]}{[\sum (x_i - \bar{x})^2 \quad \sum (y_{i-} \quad \bar{y})^2]^{1/2}} \quad (20)$$

Where $\bar{x}$ and $\bar{y}$ represents the average value of the observations and predicted quantiles respectively

3.1.4.4: Maximum absolute error (MAE).

This represents the largest absolute difference between the observed and computed or predicted values. It is given by:

$$MAE = \max (x_i - y_i) \quad (21)$$

4.0 Presentation, analysis, application and discussion of results

The observed discharge data and the results obtained by fitting Lognormal, EV-I and Log Pearson Type III distributions each to the observed discharge data at the gauging sites at River Donga at Manya, River Donga at Donga and River Bantaji at Suntai are presented in Tables 4, 5 and 6 respectively. The summary of basic statistics of the discharge data at the gauge stations are given in Table 7.

Table 4 presents the observed discharge data at River Donga at Manya gauging station and the results obtained by fitting the probability distribution models to the data. From Table 4 it can be seen that the percentage deviation of the Log normal predicted values from the observed values ranges from -12.02% to 10.68% while for the EV-1 distribution, the percentage deviation of predicted values from observed values ranges from -15.98% to 8.86% and for the Log Pearson Type III distribution, the percentage deviation of predicted values from observed values ranges from -11.9% to 10.54%. The best fit model for the station is given in Table 10 and was selected based on statistical test results and a scoring scheme. Table 5 presents the observed discharge data at River Donga at Donga gauging station and the results obtained by fitting the probability distribution models to the data

Table 4: Peak discharge data and prediction results of Probability distribution models fitted to R.Donga at Manya discharge data.

Rank	Peak Discharge(m^3/s) R.Donga at Manya	Log Normal Prediction (m^3/s)	EV-1 Prediction (m^3/s)	Log Pearson III Prediction (m^3/s)
1	1550	1398.62	1462.21	1404
2	1440	1286.17	1312.39	1288.2
3	1130	1225.75	1236.94	1226.78
4	1100	1182.50	1185.48	1183.04
5	1080	1148.42	1146.03	1148.68
6	1051	1119.7	1113.76	1119.69
7	1050	1093.96	1086.28	1094.71
8	1040	1071.52	1062.20	1071
9	1030	1051.96	1040.63	1051.71
10	1020	1033.23	1021.02	1032.99
11	1010	1015.78	1002.92	1015.31
12	1000	997.7	986.05	997.70
13	982	983.10	970.17	982.43
14	977	967.83	955.08	967.16
15	965	953.01	940.64	958.35
16	947	938.42	925.74	937.97
17	940	924.27	913.24	923.63
18	920	910.12	900.07	909.49
19	900	896.18	887.13	895.57
20	864	882.26	874.36	881.65
21	858	868.16	861.63	867.56
22	852	853.88	862.41	853.29
23	850	839.46	836	838.68
24	846	824.32	822.88	823.76
25	845	808.72	809.37	808.16
26	840	792.32	795.27	791.59
27	837	774.64	780.35	773.82
28	835	755.26	764.19	754.74
29	735	733.50	746.15	732.82
30	726	707.62	724.97	707.13
31	602	674.34	697.77	673.59
32	584	620.15	652.49	618

From Table 5 it can be seen that the percentage deviation of the Log normal predicted values from the observed values ranges from -3.36% to 4.20% while for the EV-1 distribution, the percentage deviation of predicted values from observed values ranges from -7.8% to 2.6% and for the Log Pearson Type III distribution, the percentage deviation of predicted values from observed values ranges from -4.78% to 3.88%. The best fit model for the station is given in Table 10 and was selected based on statistical test results as presented in Table 8 and a scoring scheme.

Table 6 presents the observed discharge data at River Bantaji at Suntai gauging station and the results obtained by fitting the probability distribution models to the data.

.

Table 5: Peak discharge data and prediction results of Probability distribution models fitted to R.Donga at Donga discharge data.

Rank	Peak Discharge(m^3/s) R.Donga at Donga	Log Normal Prediction (m^3/s)	EV-1 Prediction (m^3/s)	Log Pearson III Prediction (m^3/s)
1	2400	2332.92	2455.79	2370.82
2	2272	2195.33	2250.75	2210.80
3	2150	2119.83	2147.47	2125.69
4	2100	2065.38	2077.06	2066.47
5	2047	2021.62	2023.06	2019.53
6	2010	1985.18	1978.90	1980.48
7	1930	1952.99	1941.30	1946.26
8	1910	1923.98	1908.34	1915.58
9	1890	1897.58	1878.83	1888.25
10	1850	1872.84	1851.98	1862.52
11	1800	1849.27	1827.22	1838.65
12	1776	1827.26	1804.12	1816.35
13	1750	1806.34	1782.39	1794.73
14	1740	1785.66	1761.74	1774.19
15	1730	1765.63	1741.97	1754.28
16	1720	1746.22	1722.95	1734.60
17	1700	1727.03	1704.47	1715.53
18	1690	1707.05	1686.45	1685.77
19	1680	1688.89	1668.74	1678.03
20	1670	1669.55	1651.26	1659.58
21	1650	1650.44	1633.84	1640.59
22	1645	1630.79	1634.91	1621.81
23	1632	1610.27	1598.77	1602
24	1600	1589.28	1580.81	1582.34
25	1580	1567.47	1562.32	1561.35
26	1560	1544.18	1543.02	1539.57
27	1548	1519.14	1522.59	1516
28	1510	1491.76	1500.48	1490.39
29	1490	1460.15	1475.80	1461.50
30	1485	1422.65	1446.84	1427.25
31	1350	1373.7	1409.58	1382.61
32	1250	1292	1347.62	1309.78

Table 6: Peak discharge data and prediction results of probability distribution models fitted to R.Bantaji at Suntai discharge data.

Rank	Peak Discharge(m^3/s) R.Bantaji at Suntai	Log Normal Prediction (m^3/s)	EV-1 Prediction (m^3/s)	Log Pearson III Prediction (m^3/s)
1	960	1083.93	1096.76	1087.67
2	950	972.52	969.07	973.64
3	910	913.48	904.76	914.11
4	900	871.98	860.90	872.57

Rank	Peak Discharge(m^3/s) R.Bantaji at Suntai	Log Normal Prediction (m^3/s)	EV-1 Prediction (m^3/s)	Log Pearson III Prediction (m^3/s)
5	856	839.46	827.28	839.85
6	850	812.45	799.77	812.83
7	800	789.04	776.36	789.41
8	751	768.24	755.83	768.42
9	750	749.37	737.45	749.55
10	745	732.15	720.73	732.32
11	740	715.97	705.31	716.14
12	710	700.65	690.93	700.81
13	700	686.27	677.40	686.28
14	670	672.35	664.54	672.51
15	646	659.02	652.23	659.17
16	645	646.10	640.38	646.10
17	640	646.69	628.87	633.43
18	635	621.012	617.65	621.01
19	632	608.69	606.62	601.69
20	630	596.34	595.73	596.35
21	619	584.11	584.88	583.98
22	601	571.61	585.55	571.61
23	600	558.98	563.04	558.98
24	562	546.13	551.84	546.
25	530	532.72	540.34	532.48
26	508	518.68	528.33	518.44
27	463	503.73	515.61	503.38
28	451	487.52	501.84	487.08
29	450	469.35	486.47	468.92
30	420	448.02	468.43	447.40
31	407	420.82	445.23	419.95
32	378	377.57	406.63	375.75

From the Table 6 it can be seen that the percentage deviation of the Log normal distribution predicted values from the observed values ranges from -12.9% to 6.83% while for the EV-1 distribution, the percentage deviation of predicted values from observed values ranges from -14.24% to 6.16% and for the Log Pearson Type III distribution, the percentage deviation of predicted values from observed values ranges from -13.29% to 6.83%.The best fit model for the gauging station is given in Table 10. It was selected based on statistical test results presented in Table 8 and a scoring scheme.

In order to determine the best fit model at each gauging station, the probability distribution model results were subjected to four statistical tests (specific measures of error in prediction) which include RMSE, RRMSE, CC and MAE. The results of these tests are presented in Table 8.
The best fit was determined by means of a criterion depending on the differences between the observed and the theoretical density functions or distributions [16]. In order to obtain the overall goodness of fit of each distribution at a station or gauging site, a ranking scheme was utilized based on the relative magnitude of the statistical test results. A distribution with the lowest RMSE, lowest RRMSE, lowest MAE or highest CC was given a score of 3, the next best was given the score 2, while the worst was given the score 1. The result of the scoring exercise at each station is presented in Table 9.

Table7: Summary of statistics for annual Peak discharge in stations in Upper Benue River Basin

Station	Mean ($\bar{X}$)	Standard deviation (σ)	Skew(a)	$\bar{X}_{logx}$	$\boldsymbol{\sigma_{logx}}$
Donga at Manya	948.94	191.12	0.01037	2.9690	0.08543
Donga at Donga	1753.9	254.37	0.2068	3.2396	0.061956
River Bantaji at Suntai	659.66	161.81	-00423	2.806	0.1108

Table 8: Results of the of Statistical tests applied the distribution models

Station	Distribution model	RMSE	RRMSE	CC	MAE
Donga at Manya	Lognormal	57.15	0.0535	0.9574	153.83
	EV-I	49.64	0.05527	0.9641	87.79
	Log Pearson III	56.6	0.05327	0.9582	151.8
Donga at Donga	Lognormal	35.54	0.01996	0.9910	76.67
	EV-I	30.96	0.02095	0.9931	33.23
	Log Pearson III	30.91	0.01874	0.9932	61.2
Bantaji at Suntai	Lognormal	31.80	0.04533	0.9818	41.02
	EV-I	38.48	0.06032	0.9723	50.23
	Log Pearson III	32.41	0.04578	0.9813	41.02

The overall score of each distribution was obtained by summing the individual point scores obtained from all the tests at each of the three stations and the distribution with the highest total score at each station was chosen as the best fit distribution model for the station. The best fit model for the discharge data at each station selected based on highest total score obtained at the station as shown in Table 9 is presented in Table 10.

The parameters of the best fit models at each station were estimated using equations given in Table 2. When the parameters of the distribution are estimated, the inverse distribution defines the Quantiles of the frequency curve and with the method of moments estimators, many distributions used in hydrologic engineering may be written in the general form; $Q_T = \bar{Q} + K_T S$, where Q_T is the quantile with specified return period T, $\bar{Q}$ is the sample mean,S is the sample standard deviation and K_T is the frequency factor and it depends on the distribution selected and is a function of return period and in some cases other population parameters. The frequency factor function can be tabulated or expressed in mathematical terms. The estimated model parameters and the corresponding flood quantile estimation equations at the gauging stations are given in Table 11a.

The selected best fit distribution models were used to predict flood quantiles (Q_T) for the three gauging stations for return periods of 2, 5,10,25,50,100,200 years. The quantile estimates are presented in Table 12

Table 9: Distribution model scoring scheme based on goodness of fit test results

Station	Distribution model	RMSE	RRMSE	CC	MAE	Total score
Donga at Manya	Lognormal	1	2	1	1	5
	EV-I	3	1	3	3	10
	Log Pearson III	2	3	2	2	9
Donga at Donga	Lognormal	1	2	1	1	5
	EV-I	2	1	2	3	8
	Log Pearson III	3	3	3	2	11
Bantaji at Suntai	Lognormal	3	3	3	3	12
	EV-I	1	1	1	1	4
	Log Pearson III	2	2	2	3	9

Table 10: Best fit model for discharge data at each station

station	Best fit distribution model
R. Donga at Manya	Extreme value Type I
R. Donga at Donga	Log Pearson Type III
R. Bantaji at Suntai	Lognormal

Table 11a: Best fit models and estimated parameters at gauging stations

Station	Best fit distribution model	Estimated model parameters and Quantile Estimation Equation
R.Donga at Manya	Extreme value type -1	u = 862.93, β = 149.05 $Q_T = 862.93 + 149.05Y_T$ where $Y_T = - \ln[\ln (T/T-1)]$ See Table 11b for Y_T values for different T values
R.Donga at Donga	Log Pearson Type III	γ= 93.9856, β = 156.47 u =2.6389 $\log Q_T = 3.239 + 0.06195 K_T$ $Q_T = 10^{(3.239+ 0.06195KT)}$ See Table11d for K_T values
R. Bantaji at Suntai	Log normal	$\bar{y} = 2.806$, $S_y = 0.1108$ $\log Q_T = \bar{Q} + K_T S_y$ $Q_T = 10^{(2.806 + 0.1108 KT)}$ See Table11c for K_T values

Table 11b: Reduced variate (Y_T) values for different T values for the Extreme value Type-1 distribution

T	2	5	10	25	50	100	200	500
Y_T	0.367	1.5	2.250	3.199	3.902	4.60	5.296	6.214

Table 11 c: K_T values for different T values for lognormal distribution

T	2	5	10	25	50	100	200
K_T	0.000	0.842	1.282	1.751	2.054	2.326	2.576

Table 11d: K values for Pearson Type III and log Pearson Type III distributions

T	2	5	10	25	50	100	200
K_T (γ = 0.2068)	-0.0341	0.8305	1.3015	1.8201	2.1624	2.4768	2.7693

Table 12: Quantile estimates for various return periods (yrs)

Station	Best fit model	2	5	10	25	50	100	200
Donga at Manya	EV-I	917.6	1086.5	1198.3	1339.6	1444.3	1548.5	1652.3
Donga at Donga	LP III	1742.2	1951.6	2087.4	2248.0	2260	2468.3	2573.4
Bantaji at Suntai	LN	639.7	793.0	887.	1000.0	1080.0	1158.0	1234

The reliability of the results of frequency analysis is dependent on how well the assumed or selected probability model applies to the given set of hydrologic data hence it was important to estimate confidence intervals for predicted return periods within which the true values are expected to lie. The procedure to estimate confidence interval involves as a first step, computing the standard error (SE) for the given distribution. The expression for the standard error is dependent on the probability distribution used. For the Gumbel (EV-1) distribution, SE is given by the equation [7]:

SE (Q_T) = $\frac{S}{\sqrt{N}}[1 + 1.1396K_T + 1.1K_T^2]^{1/2}$ (22)

where N is the number of annual maxima in the sample, s is standard deviation of the sample data and K_T is frequency factor for return period T (yrs). The upper and lower confidence limits (Q_{conf}) are calculated for particular confidence limits using equation (23).

$Q_{conf} = Q_T \pm$ f(c) SE (23)

where f(c) is the function of confidence probability. Values of f(c) for particular confidence intervals are given in Table 13.

The 95% confidence limits for 2yr, 5yr, 10yr, 25yr, 50yr, 100yr and 200yr predicted discharges for the gauging station at R.Donga at Manya using the selected best fit model (EV-1) is presented in Table 14. Comparing the discharge for upper and lower limits with the corresponding predicted discharge for each return period, it can be seen that the confidence intervals are not too wide and hence the model provide satisfactory fit for the data.

For lognormal and log Pearson distributions which were the selected best fit models for the stations at R.Bantaji at Suntai and R.Donga at Donga respectively; the upper confidence limits ($U_{T,\alpha}$) and lower confidence limits ($L_{T,\alpha}$) were obtained using the following equations [13] :

$U_{T,\alpha} = \bar{y} + S_y K_{T,\alpha}^U$ (24)

$L_{T,\alpha} = \bar{y} + S_y K_{T,\alpha}^L$ (25)

where $K_{T,\alpha}^U$ and $K_{T,\alpha}^L$ are the upper and lower confidence factors whose approximate values are obtained using the following relations [17]:

$$K_{T,\alpha}^U = \frac{K_T + \sqrt{K_T^2 - ab}}{a} \quad (26)$$

$$K_{T,\alpha}^L = \frac{K_T - \sqrt{K_T^2 - ab}}{a} \quad (27)$$

Where

$$a = 1 - \frac{Z_\alpha^2}{2(n-1)} \quad (28)$$

$$b = K_T^2 - \frac{Z_\alpha^2}{n} \quad (29)$$

where Z_α is the standard normal variable with exceedence probability, α

The results of the computation of 95% confidence limits for 2yr, 5yr, 10yr, 25yr, 50yr, 100yr and 200yr predicted discharges for the gauging stations at R.Bantaji at Suntai and R.Donga at Donga using the selected best fit models (Lognormal and log Pearson Type III) are presented in Tables 15 and 16 respectively. Comparing the discharge for upper and lower limits with the corresponding predicted discharge for each return period in each case, it can be seen that the confidence intervals are not too wide. Hence, the models provide satisfactory fits for discharge data at the respective stations.

Table 13: values of f(c) for particular confidence intervals [7]

C (%)	50	68	80	90	95	99
f(c)	0.674	1	1.282	1.645	1.96	2.58

Table 14: Calculation of 95% Confidence limits, f (c) =1.96 for EV-1 Distribution- R.Donga at Manya

\T(yrs)	2	5	10	25	50	100	200
K(T)	-0.16	0.72	1.30	2.04	2.61	3.14	3.68
Q_T(m^3/s)	917.6	1086.5	1198.3	1339.6	1444.3	1548.5	1650
S_E(m^3/s)	31.07	52.23	70.39	94.98	114.40	132.69	151.42
f_cS_E(m^3/s)	60.90	102.37	137.96	186.15	224.22	260	296.78
Upper Q_T(m^3/s)	978.50	1188.87	13 36.26	1525.75	1668.52	1808.5	1946.78
Lower Q_T(m^3/s)	856.70	984.13	1060.34	1153.45	1220.08	1288.5	1353.22

Table 15: Calculation of 95% Confidence limits for Lognormal Distribution- R.Bantaji at Suntai

T	K_T	K_T^U	K_T^L	$U_{T,\alpha}$	$L_{T,\alpha}$	Q_TU	Q_TL	Q_T
2	0	0.35775	- 0.35775	2.8456	2.7663	700.87	583.97	639.7
5	0.842	1.3195	0.4758	2.9522	2.8587	895.776	722.31	793.0
10	1.282	1.8605	0.8729	3.012	2.9027	1028.36	799.33	887.
25	1.751	2.453	1.2802	3.0778	2.9478	1196.26	886.84	1000.0
50	2.054	2.842	1.5376	3.1209	2.9763	1320.94	947.04	1080.0
100	2.326	3.193	1.7661	3.1598	3.002	1444.85	1003.90	1158.0
200	2.576	3.518	1.9745	3.1957	3.0247	1569.63	1058.71	1234

Table 16: Calculation of 95% Confidence limits for Log Pearson Type III Distribution R.Donga at Donga

T	$K_T(\gamma=0.2068)$	K_T^U	K_T^L	$U_{T,\alpha}$	$L_{T,\alpha}$	Q_TU	Q_TL	Q_T
2	-0.0341	0.3215	-0.3942	3.2589	3.2145	1815.17	1638.99	1742.2
5	0.8305	1.3056	0.4652	3.3199	3.2678	2088.73	1852.75	1951.6
10	1.3015	1.8848	0.8902	3.3558	3.2941	2268.65	1968.55	2087.4
25	1.8201	2.5416	1.3392	3.3965	3.3219	2491.45	2098.77	2248.0
50	2.1624	2.9817	1.6290	3.4237	3.3399	2652.88	2187.33	2260
100	2.4768	3.3890	1.8920	3.4489	3.3562	2811.58	2270.96	2468.3
200	2.7693	3.7699	2.1348	3.4725	3.3712	2968.56	2350.99	2573.4

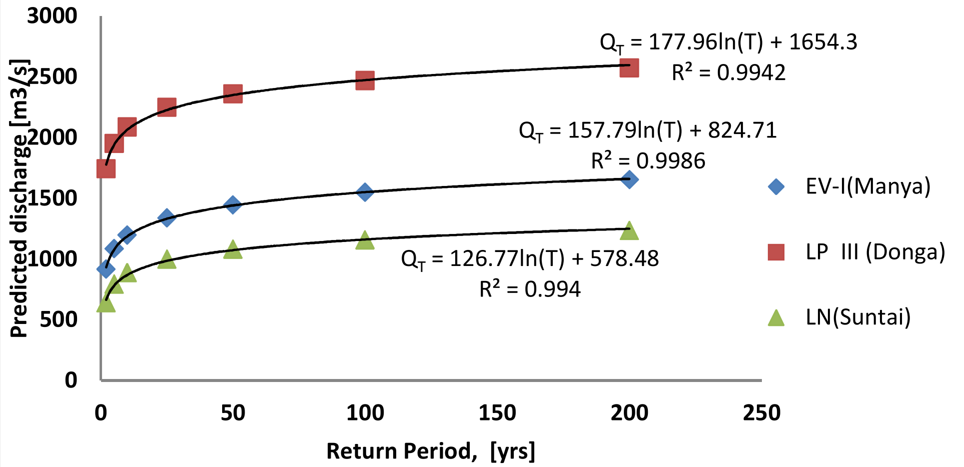

Figure 1: Plot of Predicted discharge vs return period at gauging stations using best fit models

These estimates of flood quantile magnitudes (Q_T) are useful in the planning, design, construction and operation of water resources projects or in decision processes relating to hydraulic works or flood alleviation programs and in general for water resources management within river systems. This may be for various hydraulic works such as design of weir, barrage, dam, irrigation facilities and flood control measures [18]

The plots of the predicted discharge against return periods for the different locations using best fit models are presented in figure1

Flood frequency analysis procedures have been used in this study to derive flood quantile estimates up to 200years return period. These quantile values can also be used for determining potential flood elevation and depths, areas of inundation, sizing of channels, levee heights and right of way limits [19] as well as for flood hazard and risk mapping. Flood risk map defines the susceptibility of a settlement to inundation and provides a means of assessment of flood risk in terms of loss of life, cropland and property.

In flood risk mapping, numeric values of these quantiles over the entire river network may be used as boundary conditions in hydraulic simulations carried out in determining flood prone areas for the given return periods [20]. Though no specific standards have been set for defining inundation zones in Nigeria, the 100 year flood is the basis for defining inundation zones in the United States [19] while 200 year flood is used in Norway [20].

5.0 Conclusion

From the results of the three probability distribution functions and specific measures of errors in prediction (RMSE, RRMSE, CC and MAE) applied in this study, it is concluded that the best fit models for the observed discharge data for the stations; River Donga at Manya, River Donga at Donga and River Bantaji at Suntai located within the upper Benue river basin (Hydrological Area-3) in Nigeria are Extreme value Type -1(Gumbel), Log Pearson Type III and Log normal respectively. These distributions have been utilized to predict flood quantile magnitudes (Q_T) at the stations using the obtained prediction equations. The flood quantiles find applications in design of hydraulic structures, dam safety assessments, flood hazard mapping and flood plain management.

The hydrologic problem in flood plain management is to define the area which will be flooded during the occurrence of a flood of specified return period.

References

[1] Ibrahim, M.H and Isiguzo, E.A. "Flood frequency analysis of Gurara River Catchment at Jere, Kaduna state, Nigeria". Scientific

Research and Essays, Vol.4, Number 6, 2009, pp.636- 646

[2] Intergovernmental Panel on Climate Change (IPCC). Climate Change Synthesis Report: An Assessment of Intergovernmental panel on Climate Change. IPCC Secretariat, World Meteorological Organization, Geneva, Switzerland. 2007.

[3] Tao, D.V., Nguyen, V.T. and Bourque.A. "On Selection of Probability Distributions for Representing Extreme Precipitations in Southern Quebec". Annual Conference of the Canadian Society for Civil Engineering. Montreal, Québec, Canada, June 5-8, 2002, pp 1-8

[4] Manandhar, B. Flood Plain Analysis and Risk Assessment of Lothar Khola. A Master of Science Degree thesis submitted to the Institute of Forestry, Tribhuvan University. Phokara, Nepal.2010

[5] Watson, I. and Burnett, A.D. Hydrology: An Environmental Approach, Lewis Publishers, Washington D.C., 1995

[6] Wurbs, R.A. and James, W.P. Water Resources Engineering, PHI Learning Private Ltd, New Delhi, India. 2009.

[7] Ojha, C.S.P., Berndtsson, R. and Bhunya, P. Engineering Hydrology. Oxford University Press, New Delhi. 2008

[8] Duggal, K.N and Soni, J.P. Elements of Water Resources Engineering, New Age International (P) Limited, Publishers, New Delhi, 2008

[9] Topaloglu, F. "Determining suitable probability distribution models for flow and precipitation series of Seyhan River Basin" Turk. Journal of Agric. (26), pp. 189 -194, 2002

[10] UBRBDA.Upper Benue River Basin Development Authority of Nigeria Hydrological year Book, Yola, Nigeria, 1987

[11] Akintola, J.O. Rainfall distribution in Nigeria, Impact Publishers Nigeria Ltd, Ibadan, 1986.

[12] Mustapha, S. and Yusuf, M.L.A. Text book of Hydrology and Water Resources, 1st Edition. Jenas prints and Publishing Company, Abuja, 1999.

[13] Chow, V.T., Maidment, D.R. and Mays, L.W. Applied Hydrology, McGraw Hill Book Company, New York, 1988.

[14] Prasuhn, A. Fundamentals of Hydraulic Engineering, Oxford University Press, New York, 1992.

[15] Kite, G.W. Frequency and Risk Analyses in Hydrology, Water Resources Publications, Fort Collins, Colorado, 1977.

[16] Kottegoda, N.T. Stochastic Water Resources Technology. Dept. of Civil Engineering, University of Birmingham, The Macmillan Press Ltd., London, 1980.

[17] United States Water Resources Council. Guidelines for determining flood flow frequency, Bulletin 17B, US Geological Survey, Reston. 1981

[18] Mishra, B.K., Takara, K., Yamashiki, Y. "An Assessment of Predictive Accuracy for two Regional Flood Frequency Methods", Annual Journal of Hydraulic Engineering, JSCE, Vol.54, 2010, pp. 7-12

[19] Mays, L.W. Water Resources Engineering, John Wiley & Sons Inc., Singapore, 2006.

[20] European Cooperation in Science and Technology (ECOST). Review of Statistical Methods for Flood frequency Analysis in Europe, Centre for Ecology and Hydrology, Wallingford, 2012.