Civil and Environmental Research
ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online)
Vol.3, No.2, 2013

Probability distribution models for flood prediction in Upper Benue River Basin – Part II

Osadolor Christopher Izinyon[1]* and Henry Nwanne Ajumuka[1]

[1]Department of Civil Engineering, University of Benin, PMB 1154, Benin City. NIGERIA

*E-mail: izinyon2006@yahoo.com

Abstract

The annual maximum series of discharge or flow data for 32 years (1955 to 1986) at three flow gauging stations namely; River Katsina Ala at Serav, River Taraba at Garsol and River Mayokam at Mayokam located within Upper Benue river basin of Nigeria were each fitted with three probability distribution models viz ;Log normal, Extreme value Type 1 and Log Pearson Type III and subjected to four specific measures of errors in prediction i.e., RMSE, RRMSE, CC and MAE in order to select the best probability distribution model that fits the observed flow data at the stations. The best fit distribution model at each station was then utilized to predict return period floods for each station for return periods of 2, 5, 10, 25, 50, 100, 200, and 500 years. The best fit probability distribution models obtained for the different stations are Log Normal, Log Normal and Log Pearson Type III for the stations at River Katsina Ala at Serav, River Taraba at Garsol and River Mayokam at Mayokam respectively. The corresponding return period flood prediction equations useful in the estimation of extreme flood discharge for the stations were also obtained .This type of information is used for urban development planning, flood plain management, establishment of insurance premiums and for efficient design and location of hydraulic structures.

Key words: Discharge, probability distribution models, return period, gauging station goodness of fit tests,

1. Introduction

Though flood water are an essential water 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.

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 (IPCC, 2007).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 this occurring. 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 (Tao *et al*, 2002). Similarly, estimates of the magnitude of the flood in a certain return period which may be achieved by the method of flood frequency analysis, one of the most important studies of river hydrology is very useful to the water resources engineer in the design of hydrological projects for the quantitative assessment of flood events as it is essential to interpret past record of flood events in order to evaluate future possibilities of such occurrence (Manadhar, 2010). This type of information is used extensively for planning urban development, flood plain management, establishment of insurance premiums and for efficient design and location of hydraulic structures (Watson and Burnett; Wurbs and James).

Flood frequency analysis is generally taken to denote a statistical analysis of flood, their magnitudes and or their frequency (occurrence rates in time) 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 shortness of flood records. The main difficulty with short records is that conventional moment statistics are both highly biased and highly variable in small samples.

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 (Topaloglu,2002) and (Ojha *et al*,2008) 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.

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 (UBRBDA, 1987).

In this paper, we present the results of the second part of a study made to determine which flood frequency distribution model adequately fits the statistical characteristics of observed flood data in some flow gauging stations in the Upper Benue river basin of Nigeria. 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 predicting accurate 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 Katsina Ala at Serav, River Taraba at Garsol and River Mayokam at Mayokam)

(ii) To apply specific measures of error 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 and 500yrs at each flow gauging station**.**

1.1 : Study Area

The three flow gauging sites utilized for this study are located in rivers situated within the Upper Benue Hydrological Area (HA-3) of Nigeria (Akintola,1986) which is one of the eight hydrological areas into which Nigeria is subdivided. Other important details and hydrological statistics relating to the study sites are given in Tables 1and 2

Table 1: Important details relating to Gauging sites

Station	River	Latitude(N)	Longitude(E)	Drainage Area(Km^2)
Serav	Katsina Ala	7^0 47' 59”	8^0 52' 13”	22,000
Garsol	Taraba	8^0 34' 0”	10^0 15' 0”	20,513
Mayokam	Mayokam	5^0 15’ 0”	11^0 05’	2986

Source :(Akintola, 1986; Mustapha and Yusuf, 1999)

Table 2**:** Summary of Hydrological statistics for annual Peak discharge in stations in Upper Benue River Basin

Station	Mean ($\bar{X}$) (m3/s)	Standard deviation (σ) (m3/s)	Skew(a)	σ_{logx}	$\bar{X}_{logx}$
Katsina Ala at Serav	2614.53	555.25	-1.00363	0.09421	3.407
Taraba at Garsol	1791.16	302.8	-0.4551	0.07527	3.2469
Mayokam at Mayokam	357.46	104.91	-1.1326	0.14825	2.5314

2: Basic Theory of Probability Distributions used for the Study

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 3.

Table 3: Probability distribution parameters in relation to sample moments (Ojha *et al*, 2008; Chow *et al,* 1988)

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 (Prasuhn, 1992), 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 but it generally does not provide satisfactory fit for flood discharge and other hydrologic data (Prasuhn, 1992)
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 (Prasuhn,1992)] 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 and standard deviation 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 is:

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

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 when log values are used, equation (29) 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 (Prasuhn, 1992). 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, one of the most commonly used distribution 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 3

3.0: Materials and Methods

3.1: Data and Analysis

The annual instantaneous flood peaks for three study sites or flow gauging stations in the upper Benue River Basin; River Katsina Ala at Serav, River Taraba at Garsol and River Mayokam at Mayokam for the period (1955 - 1986) were obtained from the publication (Mustapha and Yusuf,1999) and analyzed. The data is presented in Table 4.

Table 4: Annual Peak flood discharge in Rivers in Upper Benue River Basin Nigeria (Mustapha and Yusuf, 1999)

S/N	Water Year	R. Katsina Ala at Serav	R. Taraba at Garsol	R.Mayokam at Mayokam	S/N	Water Year	R. Katsina Ala at Serav	R. Taraba at Garsol	R.Mayokam at Mayokam
1	1955	2600	1930	400	17	1971	2000	1188	170
2	1956	3708	2038	420	18	1972	1680	1675	320
3	1957	2900	2000	410	19	1973	2254	1700	370
4	1958	2475	1710	330	20	1974	1520	1730	267
5	1959	1820	1120	140	21	1975	2060	1410	254
6	1960	2875	1885	390	22	1976	2300	1605	350
7	1961	2790	1725	340	23	1977	2420	1670	400
8	1962	2200	1635	300	24	1978	2360	1650	410
9	1963	3080	2280	500	25	1979	2700	1690	300
10	1964	2450	1680	320	26	1980	3356	2047	504
11	1965	2550	1880	390	27	1981	3950	2500	580
12	1966	2615	1785	350	28	1982	2885	1865	380
13	1967	2810	1720	340	29	1983	2150	1460	200
14	1968	3250	2195	480	30	1984	2350	1552	150
15	1969	3200	2400	540	31	1985	3123	1685	325
16	1970	2240	2017	420	32	1986	3000	1890	380

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 (Ojha *et al,* 2008) in order to determine the best fit function. Four types of goodness of fit tests were used for selection of the best fit model.

3.1 .1: Lognormal distribution fit to data

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 are being fitted to a normal distribution (Chow *et al*, 1988). The normal reduced variable (z) corresponding to the probability of non exceedence was determined using the following equations (Chow *et al*, 1988):

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

and

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

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:

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

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

3.1.2: Extreme value Type I (Gumbel) distribution fit to data

The fitting of EV-I distribution to the observed data was carried out using the following steps as given in (Ojha *et al*, 2008):

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 for the distribution corresponding to the different plotting position was computed using

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

iv. The T-yr recurrence interval flood was estimated using the E V-I distribution given by $X_T = u + \beta y_T$ (15)

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

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

3.1.3: Log Pearson Type III distribution

Log Pearson Type III distribution was fit 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) =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 given in (Kite, 1977) as :

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

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: Statistical Test Criteria (Measures of error in Prediction)

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 (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 the theoretical density functions or distributions (Kottegoda, 1980). 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:

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

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 $\text{RRMSE} = \left(\frac{\sum\left(\frac{x_i-y_i}{x_i}\right)^2}{(n-m)}\right)^{\frac{1}{2}}$ (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 [Tao *et al*, 2002] 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:

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

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:

$MAE = \max (x_i - y_i)$ (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 Katsina Ala at Serav, River Taraba at Garsol and River Mayokam at Mayokam are presented in Tables 5, 6 and 7 respectively.

Table 5: observed discharge data and prediction results of Probability distribution models fitted to R.Katsina Ala at Serav discharge data.

Rank	Observed discharge **R. K.Ala at Serav**	Log Normal Prediction	EV-1 Prediction	Log Pearson III Prediction
1	3950	4003.13	4139.74	3540.78
2	3708	3650.06	3693.97	3398.59
3	3350	3460.98	3469.48	3305.98
4	3250	3327.36	3316.36	3232.21
5	3200	3221.06	3198.97	3168.84
6	3123	3132.56	3102.95	3111.72
7	3080	3056	3021.21	3059.85
8	3000	2987.45	2949.56	3010.93
9	2900	2924.83	2885.39	2964.15
10	2885	2867.47	2827.03	2918.77
11	2875	2813.19	2773.19	2875.41
12	2810	2762.48	2722.97	2832.69
13	2790	2713.94	2675.74	2789.97
14	2700	2667.47	2630.84	2748.53
15	2615	2622.40	2587.87	2706.45
16	2600	2578.69	2546.50	2665.01
17	2550	2535.12	2506.34	2623.01
18	2475	2492.87	2467.17	2580.48
19	2450	2450.76	2428.67	2536.88
20	2420	2408.79	2390.65	2492.29
21	2360	2366.46	2352.78	2446.25
22	2350	2323.80	2355.10	2398.28
23	2300	2279.82	2276.53	2348.01
24	2254	2235.11	2237.49	2295.09
25	2240	2188.26	2197.28	2238.72
26	2200	2138.94	2155.34	2178.21
27	2150	2086.89	2110.93	2111.06
28	2060	2029.55	2062.86	2035.64
29	2000	1964.72	2009.19	1940.43
30	1820	1888.86	1946.23	1841.62
31	1680	1791	1865.24	1698.24
32	1520	1633	1730.50	1453.45

Table 5 presents the observed discharge data at River Katsina Ala at Serav gauging station and the results obtained by fitting the probability distribution models to the data. From the table it can be seen that the percentage deviation of the Log normal predicted values from the observed values ranges from -7.43% to 2.94% while for the EV-1 distribution, the percentage deviation of predicted values from observed values ranges from -13.84% to 4.10% and for the Log Pearson distribution, the percentage deviation of predicted values from observed values ranges from -4.26% to 10.36%.

The observed discharge data at River Taraba at Garsol gauging station and the results obtained by fitting the probability distribution models to the data are presented in Table 6.

Table 6: Observed discharge data and prediction results of Probability distribution models fitted to R.Taraba at Garsol discharge data.

Rank	Observed discharge **TARABA at Garsol**	Log Normal Prediction	EV-1 Prediction	Log Pearson III Prediction
1	2500	2527.55	2611.18	2422.14
2	2400	2347.47	2371.65	2293.19
3	2280	2249.57	2251.02	2218.19
4	2195	2179.72	2168.74	2162.22
5	2047	2124.22	2105.66	2116.90
6	2038	2077.30	2054.07	2077.30
7	2017	2036.57	2010.14	2042.21
8	2000	1999.86	1971.65	2009
9	1930	1966.53	1937.17	1979.70
10	1890	1935.53	1905.81	1951.19
11	1885	1906.34	1876.87	1924.42
12	1880	1878.88	1849.89	1898.45
13	1865	1852.25	1824.51	1873.27
14	1785	1826.88	1800.39	1848.84
15	1730	1802.18	1777.29	1824.32
16	1725	1777.87	1755.07	1800.53
17	1720	1754.28	1733.49	1776.64
18	1710	1730.61	1712.44	1753.07
19	1700	1707.26	1691.75	1729.01
20	1690	1683.84	1671.32	1704.51
21	1685	1659.97	1660.97	1679.57
22	1680	1636.06	1652.22	1654.20
23	1675	1613.62	1610.00	1627.42
24	1670	1585.98	1589.02	1599.92
25	1650	1559.55	1567.42	1570.36
26	1635	1531.44	1544.88	1539.22
27	1605	1501.41	1521.01	1505.22
28	1552	1468.25	1495.18	1466.89
29	1460	1430.88	1466.35	1423.44
30	1410	1386.43	1432.51	1370.25
31	1188	1328.62	1388.91	1299.57
32	1120	1233	1316.59	1179.29

From the table it can be seen that the percentage deviation of the Log normal predicted values from the observed values ranges from -11.84% to 6.45% while for the EV-1 distribution, the percentage deviation of predicted values from observed values ranges from -17.55% to 5.51% and for the Log Pearson distribution, the percentage deviation of predicted values from observed values ranges from -9.39% to 6.21%.

Table 7 presents the observed discharge data at River Mayokam at Mayokam gauging station and the results obtained by fitting the probability distribution models to the data.

Table 7: Observed discharge data and prediction results of probability distribution models fitted to R. Mayokam at Mayokam discharge data.

Rank	Observed discharge **RIVER MAYOKAM**	Log Normal Prediction	EV-1 Prediction	Log Pearson III Prediction
1	580	688.65	637.37	555
2	540	595.38	555.63	525.05
3	504	547.52	514.45	505.01
4	500	514.51	486.37	488.99
5	480	489.10	464.84	475.25
6	421	468.05	447.23	462.81
7	420	450.09	432.24	451.44
8	411	434.31	419.10	440.65
9	410	420.15	407.33	430.53
10	401	407.19	396.62	420.73
11	400	395.18	386.75	411.24
12	391	383.97	377.54	401.98
13	390	373.42	368.88	392.92
14	381	363.83	360.64	383.97
15	380	353.75	352.76	375.06
16	370	344.50	345.15	366.10
17	351	335.42	337.81	357.19
18	350	326.66	330.62	348.26
19	341	318.05	323.56	339.13
20	340	309.45	316.59	329.84
21	330	300.95	309.64	320.31
22	325	292.41	310.07	310.46
23	321	283.79	295.66	300.26
24	320	275.04	288.50	289.60
25	301	266.07	281.12	278.36
26	300	256.74	273.43	266.32
27	267	246.88	265.28	253.34
28	254	236.26	256.47	238.89
29	200	224.59	246.63	222.54
30	170	211.05	235.08	203.00
31	150	194.08	220.22	177.78
32	140	167.8	195.51	137.44

From the table it can be seen that the percentage deviation of the Log normal distribution predicted values from the observed values ranges from -29.38% to 14.42% while for the EV-1 distribution, the percentage deviation of predicted values from observed values ranges from

-46.24% to 9.84% and for the Log Pearson distribution, the percentage deviation of predicted values from observed values ranges from -19.41% to 11.22%.

In order to determine the best fit model at each gauging station, the probability distribution model results were subjected to four statistical tests (goodness of fit tests) namely: RMSE, RRMSE, CC and MAE. The results of these tests are presented in Table 8.

Table 8: Results of the of goodness of fit tests applied to the distribution models

Station	Distribution model	RMSE	RRMSE	CC	MAE
Katsina Ala at Serav	Lognormal	53.12	0.024798	0.9956	113
Katsina Ala at Serav	EV-I	83.35	0.04007	0.9890	114.26
Katsina Ala at Serav	Log Pearson III	106.79	0.03259	0.9842	409.22
Taraba at Garsol	Lognormal	61.11	0.041	0.9804	140.6
Taraba at Garsol	EV-I	69.23	0.05150	0.9745	90.12
Taraba at Garsol	Log Pearson III	63.76	0.03857	0.9783	106.81
Mayokam at Mayokam	Lognormal	98.10	0.1160	0.9600	108.65
Mayokam at Mayokam	EV-I	29.87	0.1482	0.9599	31.5
Mayokam at Mayokam	Log Pearson III	19.97	0.07246	0.9824	33.68

The best fit was determined by means of a criterion depending on the differences between the observed and the theoretical density functions or distributions [14] 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.

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

Station	Distribution model	RMSE	RRMSE	CC	MAE	Total score
Katsina Ala at Serav	Lognormal	3	3	3	3	**12**
Katsina Ala at Serav	EV-I	2	2	2	2	8
Katsina Ala at Serav	Log Pearson III	1	1	1	1	4
Taraba at Garsol	Lognormal	3	2	3	1	**9**
Taraba at Garsol	EV-I	1	1	1	3	6
Taraba at Garsol	Log Pearson III	2	3	2	2	**9**
Mayokam at Mayokam	Lognormal	1	2	2	1	6
Mayokam at Mayokam	EV-I	3	1	1	3	8
Mayokam at Mayokam	Log Pearson III	2	3	3	2	**10**

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.

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

Station	Best fit distribution model(s)
R. Katsina Ala at Serav	Lognormal
R. Taraba at Garsol	Lognormal and Log Pearson Type III
R. Mayokam at Mayokam	Log Pearson Type III

For the River Taraba at Garsol gauging station, Log Pearson Type III and Log Normal probability distributions have the same total of 9. Hence the best fit model for the station was determined after plotting of the predicted discharge against return periods for the different distributions and the Log Normal distribution plot with a higher coefficient of determination (**R^2**) of 0.9877 was selected as the best model for the station.
The selected best distribution model for discharge data at each station was used to predict maximum annual discharge for return periods as given in Table 11.
Table11: Quantile estimates (Q_T) for various return periods (T yrs)

Station	Best fit model	2 yrs	5 yrs	10yrs	25yrs	50yrs	100yrs	200yrs	500yrs
K/Ala at Serav	LN	2552.7	3064.3	3371.1	3732.2	3985.7	4227.9	4463.6	4765.5
Taraba at Garsol	LN	1765.6	2043.0	2204.9	2391.6	2520.6	2642	2759.3	2907.5
Taraba at Garsol	LP III	1788.8	2047.8	2182.4	2323.6	2413.1	2492.0	2562.4	2764.1
Mayokam at Mayokam	LPIII	362.08	453.82	494.79	531.26	551.26	566.28	577.88	642.57

Table 12: Return period flood prediction equations at gauging stations

Station	**Best fit distribution (s)**	**Quantile flood Prediction Equation**
R.K/Ala at Serav	Log Normal	$Q_T = 392.13\ln(T) + 2405.2$, $R^2 = 0.9915$
R.Taraba at Garsol	Log Normal	$Q_T = 201.7\ln(T) + 1702.2$, $R^2 = 0.9877$
R Taraba at Garsol	Log Pearson Type III	$Q_T = 161.86 \ln(T) + 1758.2$, $R^2 = 0.965$
R.Mayokam at Mayokam	Log Pearson Type III	$Q_T = 43.94\ln(T) + 369.5$, $R^2 = 0.9047$

5.0: Conclusion

From the results of the three probability distribution functions and goodness of fit tests applied in this study, it is concluded that the best fit models for the discharge data obtained in the stations; River Katsina Ala at Serav, River Taraba at Garsol and River Mayokam at Mayokam located within the upper Benue river basin (Hydrological Area-3) of Nigeria are Log Normal, Log Normal and Log Pearson Type III respectively and which distributions could be utilized to predict the return period flood estimates at the stations using the prediction equations obtained.

References

Akintola, J.O. (1986). Rainfall distribution in Nigeria. Impact Publishers Nigeria Ltd, Ibadan

Chow, V.T., Maidment, D.R. and Mays, L.R. (1988). Applied Hydrology. McGraw Hill Book Company, New York.

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

Kite, G.W. (1977). Frequency and Risk Analysis in Hydrology. Water Resources Publications, Fort Collins. Colorado

Kottegoda, N.T. (1980). Stochastic Water Resources Technology. Department of Civil Engineering, University of Birmingham. The Macmillan Press Ltd. London

Manandhar, B. (2010). 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.

Mustapha, S. and Yusuf, M.L. (1999). A Text book of Hydrology and Water Resources. First Edition. Jenas prints and Publishing Company, Abuja

Ojha, C.S.P., Berndtsson, R. and Bhunya, P. (2008). Engineering Hydrology. Oxford University Press, New Delhi.

Prasuhn, A. (1992). Fundamentals of Hydraulic Engineering. Oxford University Press, New York

Tao, D.V., Nguyen, V.T. and Bourque.A. (2002). On Selection of Probability Distributions for Representing Extreme Precipitations in Southern Quebec. *Proceedings of Annual Conference of Canadian Society for Civil Engineering. June 5-8*

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

UBRBDA. (1987). Upper Benue River Basin Development Authority Hydrological year Book for 1987

Watson, I. and Burnett, A.D. (1995). Hydrology: An Environmental Approach, Lewis Publishers, Washington D.C

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