Flood Flow Probability Distribution Model Selection on Niger/Benue River Basins in Nigeria

Itolima Ologhadien[1*]

[1]*Department of Civil Engineering, Rivers State University of Science and Technology, P.M.B, 5080, Nkpolu, Port Harcourt, Nigeria.*

Author's contribution

The sole author designed, analyzed, interpreted and prepared the manuscript.

Original Research Article

ABSTRACT

Flood frequency analysis is a crucial component of flood risk management which seeks to establish a quantile relationship between peak discharges and their exceedance (or non-exceedance) probabilities, for planning, design and management of infrastructure in river basins. This paper evaluates the performance of five probability distribution models using the method of moments for parameter estimation, with five GoF-tests and Q-Q plots for selection of best –fit- distribution. The probability distributions models employed are; Gumbel (EV1), 2-parameter lognormal (LN2), log Pearson type III (LP3), Pearson type III(PR3), and Generalised Extreme Value(GEV). The five statistical goodness – of – fit tests, namely; modified index of agreement (Dmod), relative root mean square error (RRMSE), Nash – Sutcliffe efficiency (NSE), Percent bias (PBIAS), ratio of RMSE and standard deviation of the measurement (RSR) were used to identify the most suitable distribution models. The study was conducted using annual maximum series of nine gauge stations in both Benue and Niger River Basins in Nigeria. The study reveals that GEV was the best – fit distribution in six gauging stations, LP3 was best – fit distribution in two gauging stations, and PR3 is best- fit distribution in one gauging station. This study has provided a significant contribution to knowledge in the choice of distribution models for predicting extreme hydrological events for design of water infrastructure in Nigeria. It is recommended that GEV, PR3 and LP3 should be considered in the development of regional flood frequency using the existing hydrological map of Nigeria.

_____

Corresponding author: Email: itolima2000@yahoo.com;

1. INTRODUCTION

Floods are the overflowing of the normal confines of a stream or other body of water or the accumulation of water over areas that are not normally submerged. Floods include river (fluvial) floods, flash floods, urban floods, pluvial floods, sewer floods, coastal floods, and glacial lake outburst floods [1]. The focus of this paper is fluvial flooding. Intense and long – lasting rainfall is the common cause of river (fluvial) floods in large river basins, for example, the Niger river basin, while floods in small basins are generated by short – duration, highly intense rainfall [2]. Floods are responsible for 20 – 30% of economic losses caused by natural hazards globally and they are also responsible for more than 50% of all fatalities due to natural disasters [1] and [3]. Similarly, [4] reported that the proportion of the world's population living in flood – prone river basins has increased by 114% , while those living on cyclone – exposed coastlines have grown by 192% since the 1980s and flooding is the most frequent and greatest hazard for the 633 largest cities or urban agglomerations. Furthermore [5] and [6] reported that since 1990, the United States floods have caused more than 10,000 deaths and losses over US470 billion. In Europe, 264 flood disasters were reported in the last 30 years, with the number increasing from 31 in the period of 1973 – 1982 to 179 during the last decade [7]. Also [3] reported that in 2010, a destructive flood in Pakistan affected up to 20 million people and left more than 1500 dead. The impacts of floods are expected to increase due to population growth, population migration to coastal areas, and climate change effects [8].

Furthermore, [3] reported that in 2012, "killer floods", inducing more than 50 fatalities each, occurred in Madagascar, Niger and Nigeria in Africa; Bangladesh, China, India, North and South Korea, the Philippines and Russia in Asia; and Argentina, the United States and Haiti in the Americas. Similarly, [9] reported that the 2012 floods in Nigeria affected 7.7 million people, 363 fatalities were recorded and approximately 600,000 houses were damaged or destroyed. This disaster greatly worsened an already existing housing deficit thereby placing huge pressure on all levels of government to address the sharp increase in housing demand.

The application of probability distribution model to annual flood flow presupposes screening and application of non – parametric tests of randomness, independence, homogeneity and stationarity. It is only when empirical evidence was found to rule out the non – parametric tests, before the available data is considered fit for flood frequency analysis.

The two data types commonly used in flood frequency analysis; Annual Maximum Series (AMS) and Partial-duration series (POT). The AMS is used when only one damaging flood event per year is possible, while the POT is used when more than one damaging flood event per year is possible. The AMS is adopted in this study, because it is consistent with the occurrence of floods in Nigeria wherein one damaging fluvial flood event occurs annually [10].

The choice of suitable probability distribution model and parameter estimation method is crucial to a successful model selection exercise , because the parameters of the distribution and its uncertainty assessment are determined by the candidate distribution model.

Several probability distribution models have been considered in different situations, for the probabilistic modelling of extreme events, such as GEV, LN2 and LN3), Gamma (Gam), EV1, LP3, PR3. Reviews of pertinent works include [11 - 21]. Assessment of cited literatures, show that there is no distribution that universally fits all the long - term series of flood data. Therefore, a number of commonly used distributions are evaluated, and then the best – fit distribution is selected amongst the candidate distributions. [22] presented the state-of-the-art review of current practice with regard to use of distribution types for frequency analyses on extremes of precipitation and floods. The Review reported six most frequently used distribution type for flood frequency analysis in the order of popularity as: EV1, LP3, LN2, P3, GEV and Gamma. Furthermore, [22] and [14] observed that there has been no consensus about a globally accepted probability distribution model for flood frequency analysis across various sites, thus the true model of the data at a site is unknown. Consequently, for flood frequency analysis to be of practical use, commonly used distributions have been evaluated to establish the best – fit distribution. [15] studied the selection of probability distribution for at-site flood frequency

analysis in Australia and identified LP3,GEV and generalized pareto-distribution (GPO) as the top three best-fit distributions. [11] evaluated various statistical distributions for determination of best-fit probability distribution from 19 stations all over the world and found GEV distribution superior to other six distributions examined. [13] studied the probability distribution of annual maximum, mean, and minimum stream flows in the United States and (LN3) and LP3 distributions good approximations to the distribution of annual maximum series. [23] undertook a pan-European comparison and evaluation of methods of flood frequency estimation and found no standardized European flood frequency estimation approach. It reported that in a number of countries (i.e Australia, Germany, Italy and Spain), GEV distribution is among the recommended choices, and a variety of 2- or 3- parameter distributions. [24] studied regional flood frequency distributions for different zones in Tunisia and found the GEV and GLO superior to the other candidate distributions. Previous flood frequency studies conducted in Nigeria worthy of mention are [25] who found LN2, LP3 and EV1 the best-fit distributions for the Benue River Basin.

Several methods have been developed to estimate the parameters of probability distribution functions. They are: (i) Methods of Moments (MoM), (ii) The Maximum Likelihood Method (MLM) (iii) The Probability Weighted Moments Method (PWM) and method of L – moments, see [26] and [27] for details. A brief review by [28] and [29] revealed that PWM or L- moments is preferred to other parameter estimators while MLE are not recommended for small sample sizes less than 25). Furthermore, MLE is generally best for fitting LN2 for sample sizes longer than 25 years. Also, MoM performs best for Log – normal distributions with low skewness coefficient. Similarly LP3 applies to hydrologic frequency analysis only when λ (shape) › 1.0 and 1/(scale factor) › 0.0. [30] and [31], reported that the MoM was more suitable for data with lower skewness values and small sample sizes, whereas the method of L-moments was more suitable for data with higher skewness values and is appropriate for all sample sizes. In this study, MoM is used because of its simplicity and being relatively easy to apply by equating the sample moments with the moments of the population distribution functions.

The performances of fitted probability distribution models are compared using different accuracy measures, to identify the best – fit model among the employed probability distribution models. These measures may be categorised as follows; (i) graphical assessments (ii) statistical goodness –of – fit tests (iii) hypothesis-based oodness -of – fit tests and (iv) information –based criteria. The commonly used statistical indices are (i) Relative Root Mean Square error (RRMSE) (ii) Nash – Sutcliffe efficiency (NSE) (iii) Percent bias (PBIAS) (vi) ratio of the root mean square error to the standard deviation of measured data (RSR)[32-33]. Each method has its strengths and weaknesses when applied to model selection. Therefore the selection of the best efficiency measures should reflect the intended use of the model and should concern model quantities which are deemed relevant for the study at hand. [34] recommended calculating statistical goodness – of – fit for quantitative evaluation of the differences between observed and simulated discharges. The statistical goodness –of – fit tests and PPCC test is adopted in this study. Flood flow model selection studies in Nigeria is still at infancy level as there is no established probability distribution model for accurate prediction of flood quantiles, while the country continues to suffer devastating floods. Besides the 2012 "killer floods" the frequency of flooding has been [9]. The objective of this paper is to evaluate the performances of five commonly used probability distribution models to find the best – fit distribution(s) that could be adopted in practice to accurately simulate or model flood flow in Nigeria. The introductory section presents a literature review and background information on flood frequency analysis, together with a statement of the problem. The study area, brief description of Benue and Niger river basins and data description are presented in section 2. Section 3 provides a synthesis of the methodology comprising the probability distribution functions used, parameter estimation method, goodness -of – fit tests, and procedure for estimation of flood quantiles.. Section 4 contains results and discussion comprising derived flood quantiles, results of goodness – of – fit tests. The conclusion derived in this study and recommendations are presented in section 5. The results of this study will be useful in hydraulic engineering and design of infrastructure to control the devastating impacts of floods.

2. DATA DESCRIPTION AND STUDY AREA

The Niger River Basin covers a total area of approximately 2,156,000km^2, only about

1,270,000km^2 actively contributes to runoff and river discharge. The whole basin is spread over the territory of ten countries. Nigerian shares about 28.3% of the active Niger River Basin (424,500km^2). The Niger Basin extends across 20 out of 36 states of Nigeria and comprises two main rivers; the Niger and Benue [35]. Table 1 shows the distribution of the gauging stations, the geographical and background information and the descriptive statistics of both Niger and Benue river basins. The data length of the gauging station is 30 years each. For the Niger river basin the Coefficient of Variation (CV) ranges between 0.219 and 0.321 except at Asamabiri with 0.151. Thus the year-to-year variation of the AMS is moderately variable. The coefficient of skewness ranges between 0.162 and 0.480, all positive values, which implies non – normal probability distribution.

The Benue River Basin is a major tributary of the River Niger forming a confluence at Lokoja and it contributes more than the actual Niger River discharges at the confluence. It originates from the Adamawa Plateau in Cameroun and has a total length of about 1200km from origin to the confluence at Lokoja with about 4.4 percent (66,000km^2) of the Benue Basin lies in Cameroun [35]. In terms of the descriptive statistics the coefficient of variation (CV) ranges between 0.194 and 0.313, which implies that the year-to-year variation of the AMS is also moderately variable. The coefficient of skewness (CS) ranges between 0.301 and 0.575, all

positive values, which also implies non – normal probability distribution. The AMS data used in this study, was obtained from the Nigerian Inland Waterways Authority (NIWA), Lokoja Nigeria. The NIWA authority operates the river gauging stations in Nigeria.

3. METHODOLOGY

3.1 Probability Distribution Functions(PDFs)

The five PDFs employed in this study comprised; EVI, LN2, LP3, PR3 and GEV. Table 2 shows the probability distribution models, sample parameters and quantile estimators. Detailed procedures for flood frequency analysis may be found in [26,36] and [37].

3.2 Parameter Estimation Method

The central moments of a distribution are given by:

$$\mu_r = E(Q - \mu)^r = \int(Q - \mu)^r f_Q (Q)dQ \qquad (1)$$

Variance: $\sigma^2 = \mu_2$

Skewness:
$$\sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{3/2}} \qquad (2)$$

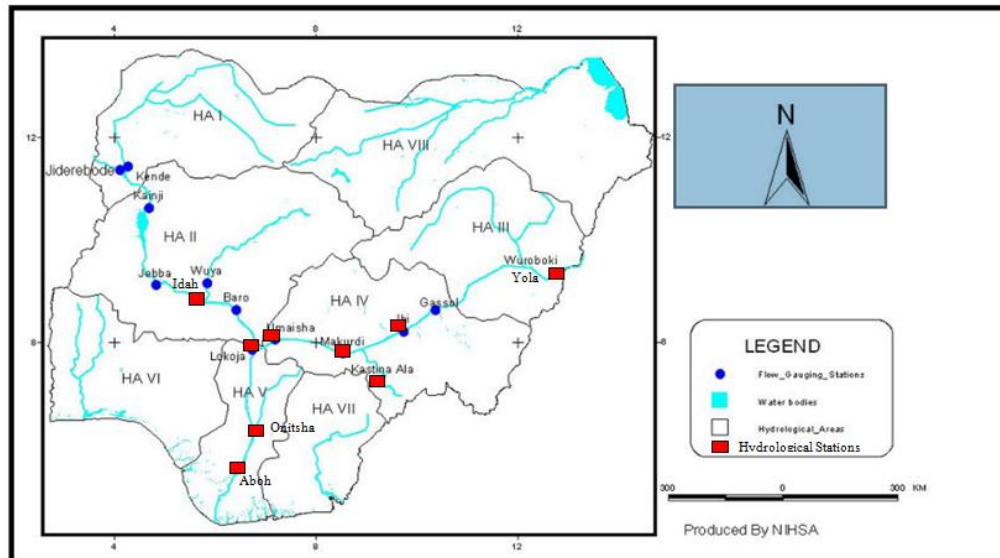**Fig. 1. Map of Nigeria showing the studied stations**

Table 1. Characteristics of selected gauge stations

(1)	(2)	(3)	(4)	(5)	(6)	(7)		(8)	(9)
S/N	Station	Latitude	Longitude	River	Catchment	Annual Streamflow		Co. of Var.	Skewness
		(N)	(E)		Km^2	Max. m^3/s	Min. m^3/s	(cv)	(cs)
1	Asamabiri	05°32'	06°31'	Niger	1,112,830	18671.41	12281.48	0.151	0.480
2	Baro	08°35'	06°23'	Niger	729,510	8852.21	103.45	0.321	0.162
3	Idah	07°06'	06°43'	Niger	1,105,780	26,760.24	826.32	0.252	0.174
4	Lokoja	07°49'	06°44'	Niger	750,790	28,360	248.75	0.219	0.337
5	Onitsha	06°10'	06°45'	Niger	1,125,170	26,607.53	426.84	0.237	0.164
6	Ibi	08°11'	09°45'	Benue	275,370	12,454.94	12.68	0.251	0.575
7	Makurdi	07°45'	08°32'	Benue	317,430	16,034.93	30.48	0.194	0.469
8	Umaisha	08°00'	07°14'	Benue	343,210	18,408.97	7.71	0.254	0.349
9	Yola	09°14'	12°28'	Benue	112,680	6641.30	8.93	0.313	0.301

Extracted from hydrological year books (1914 – 1989); NIWA, Lokoja, Nigeria

Table 2. Probability models, sample parameters and moments

Probability Model	Probability Density Function	Range	MoM	Quantile Function (Q_T)
Lognormal (LN2)	$f(Q) = \dfrac{1}{Q\sigma_Q\sqrt{2\Pi}} \exp*\tfrac{1}{2}\left[-\left(\dfrac{Q-\mu_Q}{2\sigma_Q^2}\right)\right]$ $Q = LnQ$	$-\infty \le Q \le \infty$ $(0 \le x \le \infty)$	$\sigma_Q = \sqrt{Ln(C_V^2 + 1)}$ $\mu_y = LN\vec{Q} - \dfrac{\sigma_Q^2}{2}$	$LnQ_T - \mu_Q + Z\sigma_Q$ $Q_T = e^{(\mu_Q + z\sigma_Q)}$
Gumbel (EV1)	$f(Q) = \dfrac{1}{\alpha}\exp\left[-\dfrac{Q-\beta}{\alpha} - \exp\left(-\dfrac{Q-\beta}{\alpha}\right)\right]$ $F(Q) = \exp\left[-\exp\left(-\dfrac{Q-\beta}{\alpha}\right)\right]$	$-\infty < Q < \infty$ $0 \le \beta \le \infty$	$\mu = \beta + 0.5772\alpha$ $\sigma_Q^2 = 1.645\,\alpha^2$	$Q_T = \beta + \alpha\, Ln\left[-Ln\left(1-\dfrac{1}{T}\right)\right]$
Log Pearson (LP3)	$f(Q) = \dfrac{(y-u)^\alpha}{\beta^2 Q\Gamma(\alpha+1)}\exp\left[\dfrac{-(y-u)}{\beta}\right]$ $y = LnQ$	$-\infty \le Q \le \infty$ $0 \le Q \le \infty$	Indirect MoM, $\mu_z = \mu + \alpha\,\beta$ $\sigma_Z^2 = \alpha^2\beta$ $\beta = \left(\dfrac{2}{Cs}\right)^2$	$Z_T = LnQ_T = \mu_Z + K_T\,\sigma_Z$ $Q_T = e^{Z_T}$
Pearson Type III (PR3)	$f(Q) = \dfrac{1}{\alpha\Gamma(\beta)}\left(\dfrac{Q-\beta}{\alpha}\right)^{\beta-1}\exp\dfrac{-(Q-\beta)}{\alpha}$	$u < Q < \infty$	$\beta = \left(\dfrac{2}{Cs}\right)^2$ $\alpha = \left(\dfrac{\sigma_Q^2}{\beta}\right)^{\tfrac{1}{2}}$ $u = \overline{Q} - \sqrt{\sigma_Q^2\beta}$	$Q_T = \alpha\beta + u + K_T\sqrt{\alpha^2\beta}$
Generalized Extreme Value (GEV)	$f(Q) = \dfrac{1}{\alpha}\left[1 - K\left(\dfrac{Q-\beta}{\alpha}\right)\right]^{\tfrac{1}{K}-1}$ $e - \left[1 - K\left(\dfrac{Q-\beta}{\alpha}\right)\right]^{\tfrac{1}{K}}$	$\beta + \dfrac{\alpha}{k} > Q < \infty$ $-\infty < Q < \beta + \dfrac{\alpha}{k}$	Equations 31 – 33	$Q_T = \beta + \dfrac{\alpha}{K}\left\{1 - \left[-Ln\left(1-\dfrac{1}{T}\right)^K\right]\right\}$

Where β,α, and k are the location, scale, and shape parameters of the distributions

Kurtosis: $\beta_2 = \dfrac{\mu_4}{\mu_2^2}$

The sample moment are calculated using Equations 1 and 2:

$$\bar{Q} = n^{-1} \sum Q_i \tag{3}$$

$$m_r = n^{-1} \sum (Q_i - \bar{Q})^r$$

The sample moments are corrected for bias following [26]. The unbiased estimators for some central moments are calculated as follows:

$$S^2 = (n-1)^{-1} \sum (Q_i - \bar{Q})^2$$

$$m_3^* = \frac{n^2}{(n-1)(n-2)} m_3 \tag{4}$$

$$m_4^* = \frac{n2}{(n-2)(n-3)} \left\{ \frac{n+1}{n-1} m_4 - 3m_2^2 \right\}$$

The conventional moment ratio defined by (i) coefficient of variation (cv); (ii) Skewness (C_s) and (iii) coefficient of kurtosis (C_k) are respectively:

$$C_v = S / \bar{Q} \tag{5}$$

$$C_s = m_3^* / S^3 \tag{6}$$

$$C_k = k_4^* / S^4 + 3 \tag{7}$$

The moments for the selected distributions are shown in Table 2.

3.3 Goodness-of-fit-tests

The GoF measures are selected to give sound comparative evaluation study with quantities deemed relevant to objective estimates of the "closeness" of the simulated discharges to observed flood flow [34]. Details about the selected GoF measures may be found in [33,36,38-40].

3.3.1 Nash-sutcliffe efficiency (NSE)

$$NSE = 1 - \left[\frac{\sum_{i=1}^{n} (Q_{i,obs} - Q_{i,sim})^2}{\sum_{i=1}^{n} (Q_{i,obs} - Q_{obs})^2} \right] \quad \dots 0 < NSE < 1.0$$

$$\dots \dots (8)$$

3.3.2 RMSE – observation standard deviation ratio (RSR)

$$RSR = \frac{RMSE}{STDEV_{obs}} = \frac{\left[\sqrt{\sum_{i=1}^{n} (Q_{i.obs} - Q_{i,sim})^2} \right]}{\sqrt{\sum_{i=1}^{n} \left(Q_{i.obs} - \bar{Q}_{sim} \right)^2}} \tag{9}$$

3.3.3 Modified index of agreement (Dmod)

$$D\,mod = 1 - \frac{\sum_{i=1}^{n} \left\| (Q_{sim,i} - Q_{obs,i}) \right\|}{\sum_{i=1}^{n} \left(\left\| Q_{sim,i} - \bar{Q}_{obs,} \right\| + \left\| Q_{obs,i} - \bar{Q}_{obs,} \right\| \right)} \tag{10}$$

3.3.4 Relative-root-mean-square error (RRMSE)

$$RRMS = \left[\frac{1}{n-m} \sum_{i-1}^{n} \left(\frac{Q_{obs,i} - Q_{sim,i}}{Q_{obs,i}} \right)^2 \right]^{\frac{1}{2}} \tag{11}$$

3.3.5 Percent bias (PBIAS)

$$PBIAS = \left[\frac{\sum_{i=1}^{n} (Q_{i.obs} - Q_{i,sim})^2 * 100}{\sqrt{\sum_{i=1}^{n} (Q_{i.obs})}} \right] \tag{12}$$

3.3.6 Probability plot correlation coefficient (PPCC) test

The PPCC (r) statistic is calculated as:

$$r = \frac{\sum\limits_{m=1}^{N}(Qobs.i - \overline{Q}_{obs})(q_m - \overline{q})}{\sqrt{\sum\limits_{m=1}^{N}(Qobs.i - \overline{Q}_{obs})^2 \sum\limits_{i=1}^{N}(q_m - \overline{q})^2}} \qquad (13)$$

3.3.7 Plotting position formulae

The unbiased plotting position formulas have the general formula [41].

$$Pi = \frac{i-a}{n+1-2a} \qquad (14)$$

Where "a" varies from 0 to 0.5; Pi is the plotting probability and i is the rank in ordered observation with i = 1 for the smallest observation in data sample.

In Equation 14; when a = 0.375; Blom formula was used for LN2 distribution, when a = 0.44, Gringorten formula was used for Gumbel (EVI) distribution; for a = 0.4, Cunnane formula , was used for LP3 and GEV distributions. The unbiased plotting formula for PR3are Cunnane and IN-NA and Ngugen formulas: [42]. Where Cs is coefficient of skewness, N is sample size and i is rank with i = 1 indicating the smallest sample number.

3.4 Flood Quantile Estimation

The estimation of flood quantiles and related equations for various life expectancies of civil engineering systems can be found in the following texts [26,36,37,43] and [44]. The exceedance probability is $P(Q_T \rangle Q) = 1/T$. The cumulative probability of non – exceedance is $F(Q_T)$ is given by;

$$F(Q_T) = P(Q_T \langle Q) = 1 - P(Q_T \rangle Q) = (1 - 1/T) \qquad (15)$$

Equation 15 is the basis for estimating the magnitude of a flood , Q_T given its exceedance or non – exceedance probabilities.

The quantile estimate for T years is calculated by substituting the value of F = (1-1/T) into the quantile functions in Table 2.

3.5 Uncertainty Assessment

The confidence interval specify the probability that the quantiles estimates lie within the upper and lower confidence interval coefficients; $K^U_{Tr,\beta}$ and $K^L_{Tr,\beta}$ using the non – central t distribution. Confidence limits are computed as follows;

$$U_{Tr,\beta}(Q) = \overline{Q} + K^U_{Tr,\beta} * \sigma \qquad (16)$$

$$L_{Tr,\beta}(Q) = \overline{Q} + K^L_{Tr,\beta} * \sigma \qquad (17)$$

Where $\overline{Q}$, and σ are the log base – 10 mean and standard deviation, $U_{Tr,\beta}(Q)$ and $L_{Tr,\beta}(Q)$ are the upper and lower limits respectively. More details may be found in [26], p200.

4. RESULTS AND DISCUSSION

4.1 Quantile Estimates

The estimated parameters for each distribution across the hydrological stations and the quantile relations expressed in the form of Q_T – T reationships are presented in Table 3. The quantile relations for the probability models were derived via unbiased position formulas given in subsection 3.3.7 using the cumulative probability of non – exceedance. The predictive performances of the probability distribution models were evaluated using the statistical performance evaluation criteria, stated in section 3.3. These indices are recommended standardized guidelines for judging model performance and comparing various models [45] and [32]. A ranking scheme was devised to rank the distributions based on their test values. Ranking scores are assigned to each distribution according to the optimal value of the statistical criteria. For example, the distribution with the lowest RRMSE, RSR or PBIAS values close to zero, and highest NSE, Dmod, and total accuracy of 1.0 is given a rank of 5. Accordingly, for each criterion, the overall ranks associated with each distribution is computed by summing the individual ranks obtained for each study station. The highest score implies the best – fit distribution. Using this ranking scheme, it was plausible to find the best distribution for each station [32]. Tables 4 and 5 show the ranking scores of probability distributions models for both Niger and Benue river basins.

Columns 9 and 10 of Table 4 present the PPCC calculated test statistics and critical test statistics, at the 5% significance level of probability distributions across the study stations. The critical test statistics were obtained from various approximating equations, see [36], pp. 299 -300).

Table 4 also shows that the 2-parameter distributions; LN2 and EV1 performed better than their 3-parameter counterpart; LP3, PR3, GEV. The decision on the 3-parameter distributions just satisfactory, as the PPCC calculated test statistics are very close to their critical values. Figs. 2 – 4 show the graphical plots of observed and simulated discharges of Lokoja station, while Figs. 5 and 6 show the corresponding plots for Umaisha station . Due to lack of space, only the

plots for GEV, PR3 and LP3 for Lokoja and Umaisha stations are displayed. Figs. 9 – 15 show the 95% confidence interval for GEV, PR3 and LP3 distributions. The total ranked scores presented in Column 8 of both in Tables 4 and 5 are graphical displayed in Fig. 7 for Niger river basin and Fig. 8 for Benue river basin. Figs. 7 and 8 both indicate that GEV is best best – fit – distribution, seconded by PR3 and thirdly LP3 for both Niger and Benue river basins.

Table 3. Distribution parameter and quantile relations

Station	PDF	β	α	K (or γ)	$Q_T - T$ Models
			Distribution Parameters		
Lokoja	EV 1	146.77	3343.02	–	$Q_T = 14677.07 + 3343.02Y_T$
	LN 2	9.69	0.26	–	$L_{nQT} = 9.69 + 0.26Z_T$
	LP 3	9.749	0.0899	8.806	$Q_T = 9.68 + 0.281K_T$
	PR 3	105.47	417.55	– 27431.25	$Q_T = 16606 + 4288.15K_T$
	GEV	15211.62	4421.78	0.353	$Q_T = 15211.62 + 12909.76\{1 - -LNT - 1T0.353$
Baro	EV 1	4343.91	1272.03	–	$Q_T = 84343.91 + 1272.05Y_T$
	LN 2	8.49	0.32	–	$L_{nQT} = 8.48 + 0.32K_T$
	LP 3	2.64	0.24	7.85	$Q_T = 8.47 + 0.383K_T$
	PR 3	152.35	132.18	– 15059.25	$Q_T = 5058.03 + 1631.48 K_T$
	GEV	4538.50	1674.12	0.3313	$Q_T = 4538.50 + 5052.68\{1 - -lnT - 1T0.331$
Idah	EV 1	14480.30	3209.98	–	$Q_T = 14480.30 + 3209.98Y_T$
	LN 2	9.67	0.25	–	$L_{nQT} = 9.67 + 0.25Z_T$
	LP 3	6.31	0.110	8.97	$Q_T = 9.67 + 0.28K_T$
	PR 3	132.82	357.23	– 31113.95	$Q_T = 16332.85 + 4116.94 K_T$
	GEV	14979.09	4231.93	0.335	$Q_T = 14979.09 + 12623.09 \left\{1 - \left[-ln\left(\frac{T-1}{T}\right)\right]^{0.335}\right\}$
Onitsha	EV 1	15077.08	3126.02	–	$Q_T = 15077.08 + 3126.02Y_T$
	LN 2	9.71	0.24	–	$L_{nQT} = 9.71 + 0.24Z_T$
	LP 3	41.87	0.038	8.12	$Q_T = 9.71 + 0.25K_T$
	PR 3	148.29	329.24	– 31940.62	$Q_T = 16.881.17 + 4009.25K_T$
	GEV	15364.91	3883.13	0.227	$Q_T = 15364.91 + 17147.41 \left\{1 - \left[-ln\left(\frac{T-1}{T}\right)\right]^{0.227}\right\}$
Asamabiri	EV 1	14347.69	1812.51	–	$Q_T = 14347.69 + 1812.51Y_T$
	LN 2	9.63	0.15	–	$L_{nQT} = 9.63 + 0.15Z_T$
	LP 3	3.70	0.084	9.32	$Q_T = 9.63 + 0.16K_T$
	PR 3	17.33	558.46	5717.31	$Q_T = 15393.73 + 2324.63K_T$
	GEV	14755.66	5584.09	0.444	$Q_T = 14755.66 + 5584.09 \{1 - -lnT - 1T0.444$

Station	Distribution Parameters				
	PDF	β	α	K (or γ)	$Q_T - T$ Models
Umaisha	EV 1	8879.041	1989.390	–	$Q_T = 8879.04 + 39Y_T$
	LN 2	9.18	0.260	–	$L_{nQT} = 9.18 + 39K_T$
	LP 3	141.97	0.022	6.10	$L_{nQT} = 9.18 + 0.259K_T$
	PR 3	32.819	445.378	– 4589.79	$Q_T = 10027.17 + 2551.48K_T$
	GEV	9007.563	13729.01	0.173	$Q_T = 9007.563 + 137291.81$ $\left\{1 - \left[-ln\left(\frac{T-1}{T}\right)\right]^{0.173}\right\}$
Makurdi	EV 1	9197.077	1526.42	–	$Q_T = 9197.077 + 1526.42Y_T$
	LN 2	9.20	0.193	–	$L_{nQT} = 9.22 + 0.104 K_T$
	LP 3	8853.63	0.0021	– 8.98	$L_{nQT} = 9.20 + 193K_T$
	PR 3	18.31	457.48	1700.29	$Q_T = 10078.01 + 1957.71K_T$
	GEV	9273.04	1777.89	0.142	$Q_T = 9273.04 + 12545.67\left\{1 - -lnT-1T0.142\right.$
Ibi	EV 1	7415.658	1657.342	–	$Q_T = 7415.658 + 1657.34Y_T$
	LN 2	9.0	0.25	–	$L_{nQT} = 9.0 + 0.25K_T$
	LP 3	1.67	0.236	8.60	$L_{nQT} = 8.994 + 0.305K_T$
	PR 3	12.109	610.844	975.426	$Q_T = 8372.148 + 2125.616K_T$
	GEV	7826.864	2285.596	0.479	$Q_T = 7826.864 + 4771.643$ $\left\{1 - \left[-ln\left(\frac{T-1}{T}\right)\right]^{0.479}\right\}$
Yola	EV 1	2987.238	1111.29	–	$Q_T = 2987.258 + 1111.283Y_T$
	LN 2	8.125	0.379	–	$L_{nQT} = 8.125 + 0.379 K_T$
	LP 3	18.337	0.0997	6.287	$L_{nQT} = 8.115 + 0.427 K_T$
	PR 3	44.141	214.524	–5840.754	$Q_T = 3628.605 + 1425.273K_T$
	GEV	3066.43	1342.044	0.187	$Q_T = 3066.43 + 7192.97\left\{1 - -lnT-1T0.187\right.$

4.2 Results

Tables 4 and 5 shows the performance ranking of the five distributions across the hydrological stations. The best –fit distribution is identified based on the total score obtained using the GOF tests. Column 8 of Table 4 and column 9 of Table 5 show the total scores for each distribution across the nine hydrological stations. On assessment of the total scores in Table 4 shows that GEV is best-fit distribution for Lokoja, Baro and Asamabiri stations with a total score of 25 each. LP3 is best-fit distribution in Idah and Onitsha stations scoring 19 and 20 respectively. In terms of PPCC Gof test, the 2-parameter distributions have larger margins between the calculated test statistics and critical values than the 3-parameter distributions. Consequently, the decision to reject the null hypothesis for EV1 and LN2 is unquestionable but just satisfactory for

GEV. Table 5 shows that GEV is the best – fit distribution for Umaisha, Makurdi and Yola hydrological stations with total scores of 29 and 27 respectively, while PR3 is best – fit distribution for Ibi station. The total scores for each distribution across the 9 hydrological stations are graphically displayed in Figs. 7 and 8, for the two basins. Finally, Tables 4 and 5 show that GEV is best – fit distribution for Lokoja, Baro, Asamabiri, Umaisha, Makurdi, and Yola (total 6), LP3 is best – fit distribution for Idah and Onitsha(Total 2), and PR3 is best – fit distribution fir Ibi station (1 station). The results also show that no single probability distribution model emerged the best – fit distribution across all the stations. Figs. 9 – 15 show the 95% confidence interval for the hydrological stations for Lower Niger basin. It shows the estimated quantiles lie within the computed ranges of upper and lower confidence limits.

Table 4. Ranking scores of probability distribution models and total scores

Station	PDF	dmod	RRMSE	NSE	PBIAS	RSR	Total Score	PPCC$_{crit}$	PPCC$_{cal}$	Decision
	EV1	2	2	2	2	2	10	0.9218	0.9826	Reject
	LN2	3	3	3	3	3	15	0.9345	0.9965	Reject
Lokoja	LP3	1	1	1	1	1	5	0.9844	0.9910	Ha
	PR3	4	4	4	4	4	20	0.9738	0.9880	Ha
	GEV	5	5	5	5	5	25	0.9295	0.9815	Reject
	EV1	1	1	2	1	1	6	0.9325	0.9995	Reject
	LN2	2	2	3	2	2	11	0.9203	0.9901	Reject
Baro	LP3	3	3	4	3	3	16	0.9899	0.9865	Ha
	PR3	4	4	5	4	4	21	0.9900	0.9780	Ha
	GEV	5	5	5	5	5	25	0.9245	0.9655	Reject
	EV1	4	5	1	4	1	15	0.9552	0.9838	Reject
	LN2	5	4	3	1	3	16	0.9680	0.9839	Reject
Idah	LP3	3	3	5	3	5	19	0.9845	0.9841	Ha
	PR3	1	1	4	5	4	15	0.9765	0.9839	Ha
	GEV	2	2	2	4	2	12	0.9675	0.9900	Reject
	EV1	5	5	4	1	4	19	0.9526	0.9924	Reject
	LN2	5	4	3	2	3	17	0.9741	0.9925	Ha
Onitsha	LP3	4	3	5	3	5	20	0.9736	0.9937	Ha
	PR3	1	1	2	4	2	10	0.9748	0.9924	Ha
	GEV	3	2	1	5	1	12	0.9720	0.9938	Reject
	EV1	1	1	1	2	1	6	0.9658	0.9680	Ha
	LN2	2	2	2	3	2	11	0.9613	0.9678	Ha
Asamabiri	LP3	2	2	4	1	4	17	0.9735	0.9636	Ha
	PR3	3	3	3	4	3	16	0.9672	0.9681	Ha
	GEV	5	5	5	5	5	25	0.9462	0.9786	Reject

Table 5. Evaluation of probability distribution models

Station	PDF	Dmod	NSE	PBIAS	Tot. Accuracy	RSR	RRMSE	Total Score
	EV1	1	2	2	1	1	2	9
	LN2	2	3	3	2	2	3	15
Umaisha	LP3	3	3	1	3	3	4	17
	PR3	4	4	5	4	4	4	25
	GEV	5	5	4	5	5	5	29
	EV1	2	2	2	1	1	1	9
	LN2	4	4	4	4	5	5	26
Makurdi	LP3	3	2	1	2	3	2	13
	PR3	4	4	3	3	4	3	21
	GEV	5	5	5	5	5	4	29
	EV1	1	1	3	1	1	1	8
	LN2	2	2	4	2	2	2	14
Ibi	LP3	3	3	1	3	3	3	16
	PR3	5	5	2	5	5	5	27
	GEV	4	4	5	4	4	4	25
	EV1	2	2	3	2	2	2	13
	LN2	1	1	1	1	1	1	6
Yola	LP3	5	3	2	4	4	5	23
	PR3	3	4	5	3	3	3	21
	GEV	4	5	4	5	5	4	27
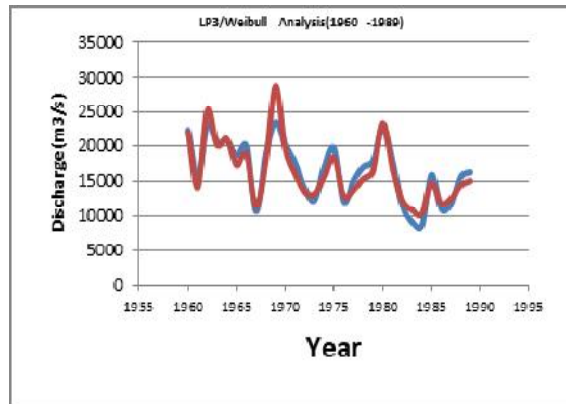
Fig. 2. Observed and simulated discharge(lokoja)
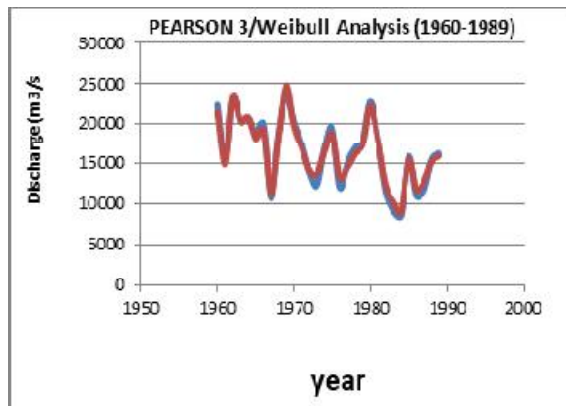
Fig. 3. Observed and simulated discharge(lokoja)
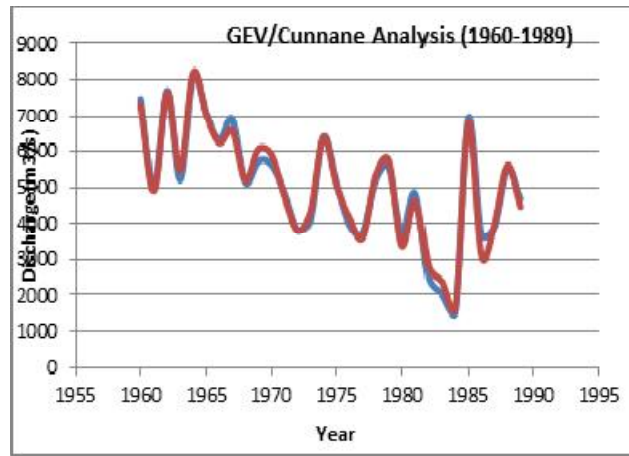
Fig. 4. Observed and simulated discharge(lokoja)
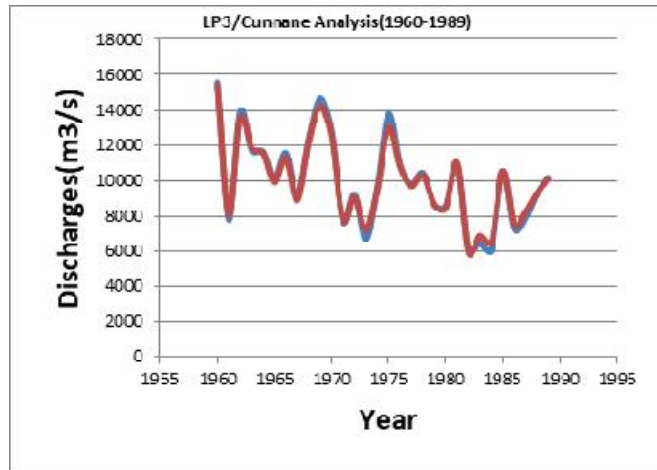
Fig. 5. Observed and simulated discharges (Umaisha)

Fig. 6. Observed and simulated discharges(Umaisha)
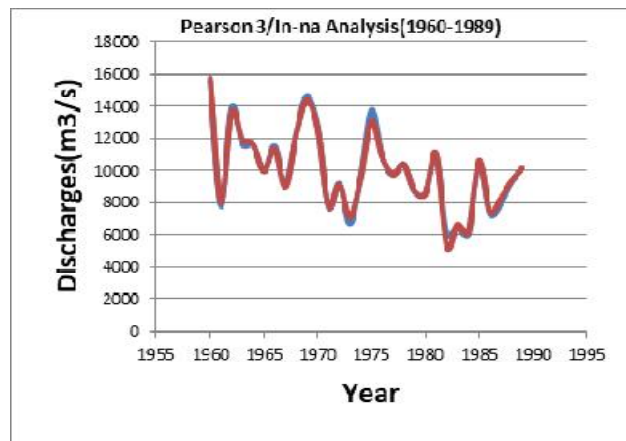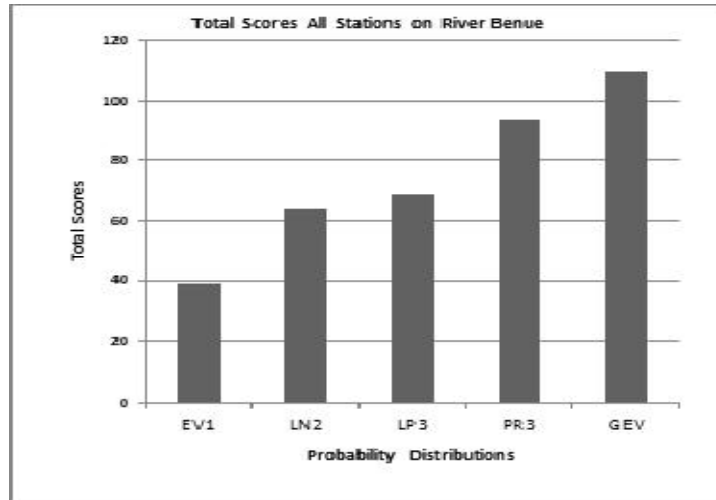
Total Scores All Stations on River Niger
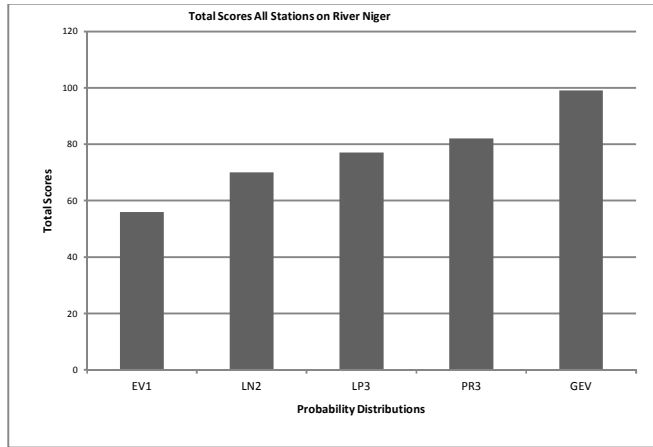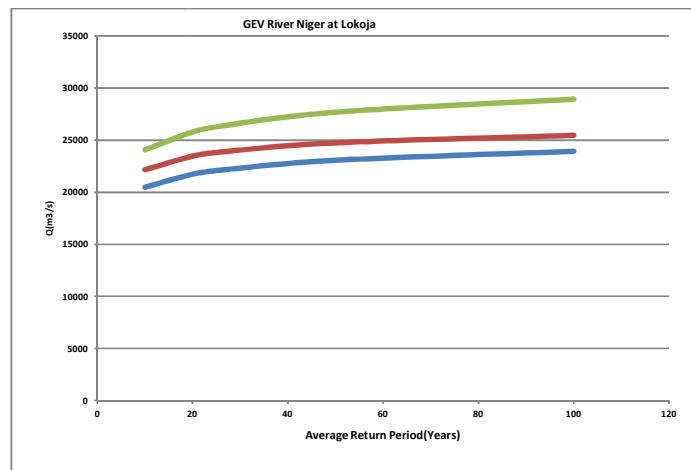
Fig. 7. Total scores all distribution on river Niger

Total Scores All Stations on River Benue

Fig. 8. Total scores all distribution on river Benue

GEV River Niger at Lokoja

Fig. 9. 95% confidence band for GEV at Lokoja Station
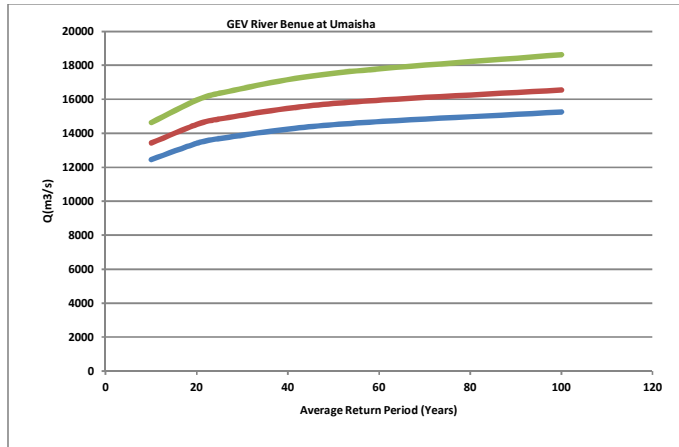
Fig. 10. 95 % confidence band for GEV at Umaisha Station
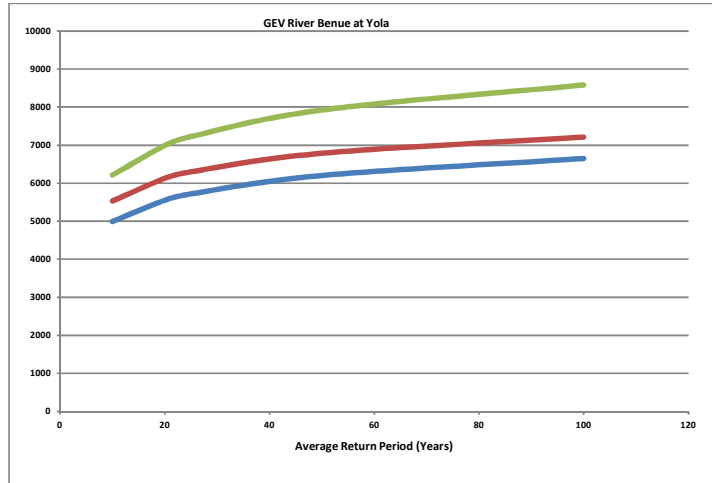
Fig. 11. 95 % confidence band for GEV at Baro Station
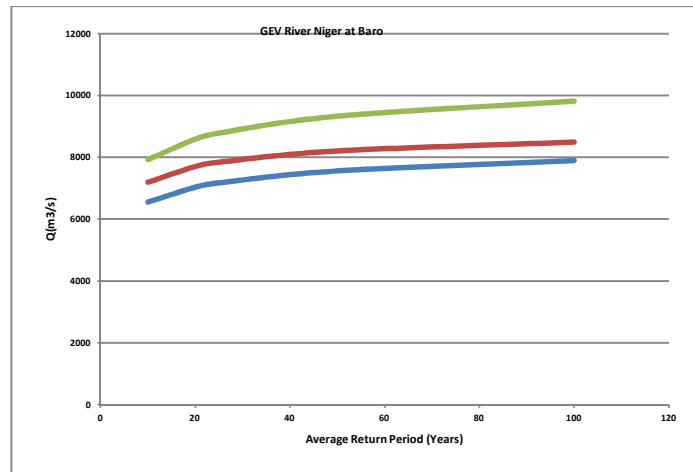
Fig. 12. 95 % confidence band for GEV at Yola Station
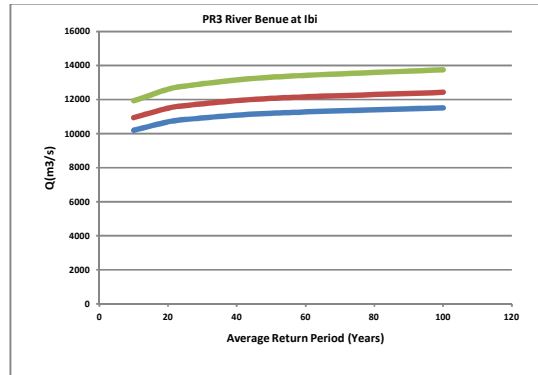
Fig. 13. 95 % confidence band for PR3 at Ibi Station
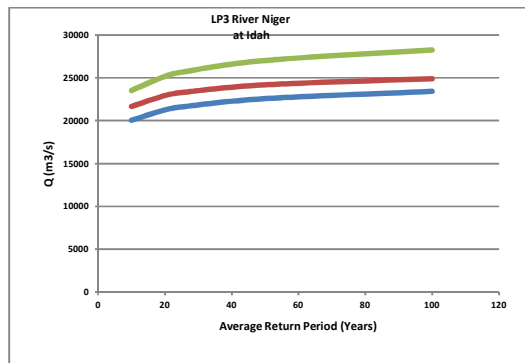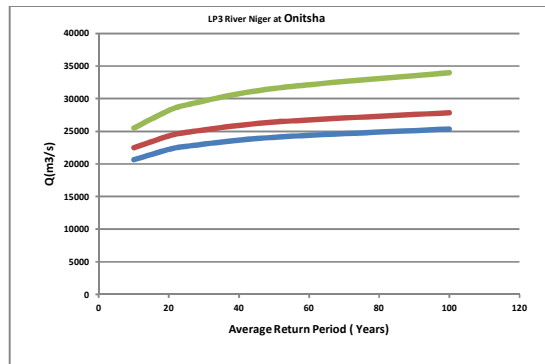

Fig. 14. 95 % confidence band for LP3 at Idah Statio

Fig. 15. 95 % confidence band for LP3 at Onitsha

4.3 Discussion

The discussion is presented in the following order:

4.3.1 Comparison with previous studies in Nigeria

The hydrological stations evaluated have not been systematically studied. However, two previous studies worthy of mention in the Niger river basin, are [46], who conducted flood frequency analysis of Niger River at Shintaku and found LP3, the best - fit distribution [47] conducted a similar study at Agenebode using LN2, LP3 and EV1 distributions and found LN2, the best – fit distribution. This study aggress with [46] who found LP3, the best – fit distribution for Shintaku gauging station. But disagrees with [47] who found LN2, the best – fit model for Agenebode.

4.3.2 Suitability of selected distributions

The best-fit distributions found in this study are GEV, PR3 and LP3 distributions. The choice of GEV distribution agrees with [48] who reported that GEV has a convincing relevance to the peak of floods, as most other probability distributions are not true depictions of flood peaks from the theoretical cause – effect standpoint. The GEV distribution is used as standard probability distribution model in 1 country. PR3 is a standard probability model in 7 countries while LP3 in 7 countries. The optimum distribution found in this study are in line with global practice for utility of probability distributions according to [22,16] and [11].

5. CONCLUSION

This paper presents the evaluation of probability distribution models of flood flow in Nigeria, comprising Niger and Benue river basins, EV1, LN2, LP3, PR3 and GEV. The performances of the five probability distribution models are compared using the GOF tests, namely Dmod, RRMSE, NSE, PBIAS, RSR and PPCC to identify the best-fit probability distribution model with MoM for parameter estimation. The study was conducted using AMS of five hydrological stations in Niger river basin and four hydrological stations in Benue river basin. The study found GEV is best – fit distribution for Lokoja, Baro, Asamabiri, Umaisha, Makurdi, and Yola (total 6), LP3 is best – fit distribution for Idah and Onitsha(total 2), and PR3 is best – fit distribution fir Ibi station (1 station).The study recommends the development of a regional GEV, PR3 and LP3 distributions using the existing hydrological map of Nigeria which had demarcated the country into eight homogenous hydrological regions. The results of this study would be useful for at – site flood frequency analysis on other stations in the lower Niger river basin in Nigeria.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Kundzewicz ZW, et al. Flood risk and climate change: global and regional perspectives. Hydrological Sciences Journal. 2014;59(1):1–28.
2. UNISDR (United Nations International Strategy for Disaster Reduction), revealing risk, redefining development. Geneva: UNISDR; 2011.
3. Berz G. Flood disasters: Lessons from the past – worries for the future. Proceeding Institution of Civil Engineers, Water and Marine Engineering, Vol. 2000;142:3–8,12212.
4. Panda A, Amaratunga D. Resilience Cities, oxford research encyclopedia of hazard science; 2019.
   DOI: 10.1093/acrefore/ 9780199389407.013.321
5. Cunha LK, Krajewski WF, Mantilla R, Cunha L. A framework for food risk assessment under nonstationary conditions or in the absence of historical data, Journal of flood risk management; 2011.
   DOI: 10.1111/j.1753-318X.2010.01085.x
6. National Weather Service (NWS); 2007.
7. Hoyois P, Guha Sapir D. Three decades of floods in Europe: A preliminary analysis of EMDAT data. Working draft prepared for Centre for Research on the Epidemiology of Disasters; 2003.
   Available:http://www.cred.be/sites/default/files/FLOOD-EUR.pdf
8. Feyen L, Barredo JI, Dankers R. Implications of global warming and urban land use change on flooding in Europe. Water and urban development paradigms—towards an integration of engineering. In: J. Feyen, K. Shannon, andM. Neville, eds. Design and management approaches. Boca Raton, FL: CRC Press. 2009;217–225.
9. Bilau AA, Witt E, Lill I, Bustani SA. Housing reconstruction following the 2012 Nigerian floods: Was it built back better? in Prins M, Wamelink H, Giddings B, Ku K, Feenstra M(Eds.)(2016).Proceeding of the CIB World Building Congress 2016: Volume II-Environmental opportunities; Constructing Commitment and Acknowledging Human Experiences .(Tampere University of Technology. Department of Civil Engineering. Construction Management and economics. Report; Vol. 18). Tampere University of Technology; 2012.
10. EM 1110-2-1450 Hydrologic Frequency Estimates, U.S. Army Corps of Engineers, Washington; 1994.
    DC 20314-1000.
11. Onoz B, Bayazit M. Best-fit distributions largest available flood samples, Journal of Hydrology. 1995;167:195–208.

12. Chen X, Shao Q, Xu CYu, Zhang J, Zhang L, Ye C. Comparative study on the selection criteria for fitting flood frequency distribution models with emphasis on upper – tail behavoiur, water. 2017;9:320.
DOI: 10.3390/w9050320

13. Vogel RM, Wilson I. Probability distribution of annual, maximum, mean, and minimum streamflows in the United States. Journal of Hydrologic Engineering. 1996;1(2):69–76.

14. Ahmad I, Faward M, Mahmood I. At –site flood frequency analysis of annual maximum stream flows in Pakistan using robust estimation methods. Pol. J. Environ. Stud. 2015;24(6):2345–2353.
DOI: 10. 15244/pjoes/59585

15. Rahman AS, Rahman A, Zaman MA, Haddad K, Ahsan A, Imteaz M. A study on selection of probability distributions for at – site flood frequency analysis in Australia, Nat Hazards. 2013;69:1803-1813.
DOI: 10.1007/s11069-013-0775-y

16. Morlot M, Brilly M, Sraj M. Characterisation of the floods in the Danude River Basin through flood frequency and seasonality analysis. Acta hydrotecnica. 2019;32/57:73–89.
Available:http://doi.org/10.15292/acta.hydro.2019.06

17. Strupczewski WG, Kochanck K, Bogdanowicz E. Flood frequency analysis supported by the largest historical flood. Nat.Hazards Earth Syst. Sci. 2014;14:1543–1551.
DOI: 10.5194/nhess-14-1543-2014

18. Prieto C, Patel D, Han D. Preface: Advances in flood risk assessment and management. Natural Hazards and Earth System Sciences. 2020;20(4):1045–1048.
Available:http://doi:org/10.5194/nhess-20-1045- 2020

19. Kousar S, Khan AR, Ul Hassan M, Noreen Z, Bhatti SH. Some best-fit probability distributions for at-site flood frequencyanalysis of the Ume River. J Flood RiskManagement. 2020;13:e12640.
Available:https://doi.org/10.1111/jfr3.12640

20. Ul Hassan M, Hayat O, Noreen Z. Selecting the best probability for at – flood frequency analysis; A study of Torne River. SN Applied Sciences. 2019;1:1629.
Available:http://doi.org/10.10.1007/s42452-019-1584-z

21. Philip Kibet Langat 1, Lalit Kumar 1, Richard Koech 2. Identification of the most suitable probability distribution models for maximum, minimum, and mean streamflow. Water. 2019;11:734.
DOI: 10.3390/w11040734

22. Cunnane C. Statistical distribution for flood frequency analysis. operational hydrol. Rep. 33, World Meteorological Organisation, Geneva; 1989.

23. Floodfreqcost Action ES0901, review of applied-statistical methods for flood – frequency analysis in Europe(WG2).

24. Abida H, Ellouze M. Probability distribution of flood flows in Tunisia, Hydrol. Earth Syst. Sci. Discuss. 2007;4:957–981.

25. Izinyon OC, Ajumuka HN. Probability distribution models for flood prediction in Upper Benue River Basin- Part II. Civil and Environmental Research. 2013;3(2):62–74.

26. Rao AA, Hamed KH. Flood frequency analysis, CRC Press; 2000.
ISBN 0-412-55280-9.

27. Van Gelder PHAJM, Wang W, Vrijling JK. Statistical estimation methods for extreme hydrological events in vasiliev, O.F et al. (eds.) Extreme Hydrological Events : New Concepts for Security, Springer. 2007;199-252.

28. Hansen C. Frequency analysis of annual maximum rainfall for selected sites in Alberta, using at – site and regional approaches. Edmonton, Alberta; 2011.

29. AMEC Environmental & Infrastructure. Frequency analysis procedures for stormwater design manual, CW2138; 2014.

30. Sankarasubramanian A, Srinivasan K. Investigation and comparison of sampling properties of L-moments and conventional moments. J Hydrol. 1999;218(1–2):13–34.

31. Bezak N, Brilly M, Sraj M. Flood frequency analysis, statistical trends and seasonality analysis of discharge data: A case study of the Litija station on the Sava River. Journal of Flood Risk Management. 2014;9:154–168.

32. Son K, Lin L, Band L, Owens EM. Modelling the interaction of climate, forest ecosystem, and hydrology to estimate catchment dissolved organic carbon export. Hydrological processes. 2019;33:1448-1464.
Available:https://doi.org/10.1002/hyp.13412

33. Moriasi DN, Gitau MW, Pai N, Daggupati P. Hydrologic and water quality models: Performance measures and evaluation criteria. Transactions of the American

Society of Agricultural and Biological Engineers. 2015;58(6):1763-1785.

34. Bennett ND, Croke BFW, Guariso G, Guillaume JHA, Hamilton SH, Jakeman AJ, et al. Characterising performance of environmental models. Environ. Model. Software. 2013;40:1-20.
Available:http://dx.doi.org/10.1016/j.envsoft.2012.09.011

35. Golitzen KG (eds), Inger A, Dione O, Jaroewich-Holder M, Olivry JC. The Niger River Basin: A vision for sustainable management; 2005.
DOI: 10. 1596/978-0-8213-6203-7
ISBN-10: 0-8213-6203-8.

36. Naghettini M (ed). Fundamentals of statistical hydrology.
ISBN 978 – 3 – 319 – 43561-9, Springer.

37. Naghavi B, Yu Xin F. Selection of parameter- estimation method for LP3 distribution. Journal of Irrigation and Drainage Engineering. 1996;122:24–30.

38. Krause P, Boyle DP, Base F. Comparison of different efficiency criteria for hydrological model assessment, advances in Geosciences. 2005;5:89–97.

39. Willmott CJ, Robeson SM, Matsuura K. Short communication: A refined index of model -performance. Int. J. Climatol. 2012;32:2088–2094.

40. Gupta HV, Sorooshian S, Yapo PO. Status of automatic calibration for hydrologic models: Comparison with multilevel calibration . J. hydrologic Eng. 1999;4(2):135–143.

41. Cunnane C. Unbiased plotting positions – A review, journal of hydrology. 1978;37:205-222.

42. IN-NA N, Ngugen VTV. An unbiased plotting position formula for the general extreme value distribution. Journal of hydrology. 1989;106: 193–209.

43. Wilson EB, Hilferty MM. "The distribution of Chi-squre", Proceedings. National Academic of Science (New York). 1931;17(12):684-688.

44. Abramawitz M, Stegun IA. Handbook of mathematical functions. Dover Publications, New York; 1965.

45. ASCE. Criteria for evaluation of watershed models. J. Irrigation Drainage Eng. 1993;119(3):429–442.

46. Ibeje, Andy O. Flood frequency analysis of River Niger, Shintaku Gauging Station, Koji State, Nigeria. Journal of Engineering and Technology(FUOYEJET). 2020;5(2):198–203.
ISSN: 2579-0625(Online), 2579 – 0617(Paper).

47. Ehiorobo JO, Akpejiori IJ. Flood frequency analysis of River Niger at Agenebode, Edo State, Nigeria. Journal of the Nigerian Association of Mathematical Physics. 2016;38:309–318.

48. Haktanir T. Comparison of various flood frequency distributions using annual flood peaks data of rivers in Anatolia. Journal of Hydrology. 1992;136:1–31.