Civil and Environmental Research
ISSN 2224-5790 (Paper) ISSN 2225-0514 (Online) DOI: 10.7176/CER
Vol.11, No.4, 2019

Flood Frequency Analysis of Niandan River at Baro, Nigeria

Itolima Ologhadien
Department of Civil Engineering, Rivers State University, Port Harcourt, Nigeria

Abstract
The determination of appropriate probability distribution is crucial for flood – risk reduction and mitigation of flood –induced damages in flood plains. This paper investigates the selection of an appropriate probability distribution for at – site flood frequency analysis using annual maximum series of Niandan River at Baro. The model results were subjected to five goodness of fit –tests(i.e. Probability plots, Probability Plot Correlation Coefficient (PPCC), Percent bias (PBIAS), Nash- Sutcliffe efficiency(NSE) and RMSE-observations standard deviation (RSR) and performance scoring/ranking to identified the best – fit distribution.The Normal, and Weibull distributions have been identified as the top two distributions for Niandan at baro. Assessment of other of flood frequency analysis studies across Nigeria show that no single distribution has been found appropriate for whole country. Consequently, research work for the search of best – fit distribution for Nigeria will continue into the future.
Keywords: Flood flow, Flood frequency analysis, PPCC, Flood quantiles, efficiency criteria,
DOI: 10.7176/CER/11-4-03
Publication date:May 31st 2019

Introduction

Flood frequency analysis is concerned with the assessment of flood magnitudes of stated frequency (or degree of rarity) for use as input into the process of flood risk assessment and management. Flood risk assessment is needed in the design of flood relief and protection works and in the assessment of the safety of existing and planned infrastructure. Assessment of hydrologic risk requires the estimation of the probability of occurrence of the flood magnitude that would inundate or damage the planned infrastructure. It is these probabilities that are estimated through flood frequency analysis. Consequently, the determination of appropriate probability distribution is of fundamental importance for flood-risk reduction and mitigation of flood-induced damages, as a wrong choice could lend to significant error bias in and design flood estimates, particularly at higher return periods, leading to either under-or over-over-estimation, which may have detrimental impacts on engineering practice and economy (Rahman et. al., 2013). Hydrologic data for flood frequency analysis are based on a set of fundamental assumptions, namely; the series (i) consistent,(ii) is trend-free and (iii) constitute a stochastic the processes whose random component follows the appropriate probability distribution function (**Adeloye and Montaseri, 2002).**

Consistency implies the data at a certain observation station are generated by the same mechanism that generated similar flood (stream flow) data at other stations. In order to determine the relative consistency, one compares the observations from a station with the mean of observation from several nearby stations. The double mass curve is the most widely used technique for evaluating a time series for consistency.

Similarly, the spearman's rank correlation method (SRCM) is applied to verify the absence of trend in the data series. The SRCM method is preferred over other trend detection methods such as the Mann-Kendall test, because it is recommended by the World Meteorological organization for tend detection in flow volumes (WMO, 1988). If the hydrological data does have a trend, it cannot be used for frequency analysis or modelling (Dahmen and Hall, 1990; Adeteoye and Montaseri, 2002). The serial correlation test is applied to identify the independence of a time series. If time series is completely random, the population auto correlation function will be zero for all lags other than zero, and the sample serial-correlation coefficients will deviate slightly from zero only because of sampling effects. The rank von Neumann ratio test and Sample auto correlation function are commonly used in testing the independence of a time series. In flood frequency analysis, three different data models are employed; (i) annual maximum series (AMS) (ii) partial duration series and (iii) time series. Detailed treatment on data models may be found in standard texts (e.g. Loveridge 2013, Rao and Hamed, 2000).

In this study, the annual maximum series (AMS) model is used. Following the adoption of AMS model the next important step in Flood Frequency Analysis is the selection of the best probability distribution(s). The distributions most commonly used are the Gumbel (EV1), General Extreme value (GEV), Pearson type III, log-Pearson type III (LP3). Lognormal (LN2), Lognormal (LN3), and Weibull. The selection of the best-fit distribution among the plethora of distributions has been a challenge and research concern in statistical hydrology. The World Meteorological Organization (WMO-No. 718, 1989) presents an excellent review of the issues involved in the selection of the appropriate distribution.

Furthermore, the WMO (No.718, 1989) also presented the global utilization of distribution types as standard tools for flood frequency analysis. It stated that the number of countries where EV1 is utilized as a standard flood frequency analysis tool is 10, while 8,7, 3 for lognormal, Pearson type 3, log Pearson type 3, and

EV2 respectively . This implies that there is no probability distribution universally accepted as standard tool for flood frequency analysis. Consequently, each country or agency need to establish the best-fit distribution(s) peculiar to its geographic setting and the prevalent stochastic processes. Rahman et al., (2013) studied the selection of probability distributions for at-site flood frequency analysis in Australia and identified the Log-Pearson 3, generalized extreme value, and generalized Pareto distribution as the top three best-fit distributions. Bayazit (1995) evaluated various statistical distributions for determination of best-fit probability distribution from 19 stations mainly in Europe and the GEV was found superior to others. Vogel and Wilson (1996) studies the probability distribution of annual maximum, mean, and minimum stream flows in the United States. They found the generalized extreme value (GEV), three-parameter lognormal (LN3) and the log Pearson Type III (LP3) distributions good approximations to the distribution of annual maximum series. The COST Action ESOPO1 Flood Freq (2013) study undertook a pan-European comparison and evaluation of methods of flood frequency estimation and found no standardized European flood frequency estimation approach. The study also observed that in a number of countries (e.g Austria, Germany, Italy and Spain) the Generalized Extreme value distribution is among the recommended choices, but a variety of 2- or 3 – parameter distributions are also used (e.g., Gumbel EV1) Generalized Pareto, (GPA), 3-parameter lognormal LN3) Gumbel EV1, Generalized praetor, (GPA), 3-parameter log Normal (LN3), etc) depending on the region. Abida and Ellouze (2007) analyzed the probability distribution of flood flows in Tunisia and found the Generalized Extreme Value (GEV) and Generalized Logistic (GLO) superior to the various candidate distributions.

Similarly, some of the Flood Frequency studies under the various climatic and geographic conditions found in Nigeria includes, Ehiorobo and Akpejiori (2016); Gbadebo et al. (2014); Ibrahim and Isiguzo (2009); Izinyon and Ajumuka (2013); Ibrahim et al., 2016); Mamman et al., (2017), etc. Based on their findings, the lognormal, LP3, and EVI were identified as the best-fit distributions. The flood flow data employed ranges between 17 and 52 years.

The suitability of a given model is usually evaluated using goodness-of-fit tests, such as Probability Plot Correlation Coefficient (PPCC); Root Mean Square Error (RMSE), Relative root mean square error (RRMSE), Mean Absolute error (RRMSE), Chi-square text (χ^2), Kolmogorov-Smirnov (KS) GoF test, Anderson-Darling (AD) GoF tests, Akaike's Information criterion (AIC), L-moments diagrams, Bayesian Information criterion (BIC) (Abida and Euoeeze, 2007, Rahman et al., (2013), Rao and Hamed, 2000). Once a model has selected, the parameters that fit the data need to be identified. The most commonly used methods are (1) the method of moments (MOM), (ii) the maximum likelihood method (MLM),(iii) the probability weighted moments (PWM). The MLM approach is considered the most efficient method since it provides the smallest sampling variance of the estimated parameters, hence of the estimated quintiles, compared to other methods. The MOM is natural and relatively easy parameter estimation method. The PWM method gives parameter estimates comparable to the method of maximum Likelihood (Rao and Hamed, 2000).

This research study investigate the best-fit probability distribution model of flood flow for River Niger at Baro. The probability distributions used are, Normal, Lognormal, Weibull and Gumbel EV1 while the appropriateness of the distributions is investigated by the PPCC methods and probability plots. The goodness of fit –tests applies to select the best –fit distribution are probability plots, probability plot correlation coefficient (PPCC), coefficient of determination(R^2), Percent bias (PBIAS), Nash- Sutcliffe efficiency(NSE) and RMSE-observations standard deviation (RSR).

This paper is organized into four sections, section 1 gives a concise background understanding of flood frequency analysis, comprising the fundamental assumptions on data quality, data models, probability distributions, parameter estimation, and goodness-of-fit. Section 2 contains the materials and methods including consistency checks, data and site description, PPCC and probability Plots methods. Section 3 presents results and Discussions of the results in section. Section 4 contains the conclusion and recommendations.

2.0 Materials and Methods

2.1Data and Study Area

The Baro hydrological station (Gauge station) is located between latitude 08°35' and longitude 06 °23'. The size of its drainage area is 729,510km^2. The altitude of the gauge is 57.22m above mean sea level. The station is 698km from the coast. The annual maximum series (AMS) used for the study were obtained from Mahe and Olivry, 1991 and Sangare (2001). The record length of the AMS of Niandan River at Baro is between 1984 and 2000 (54 years).

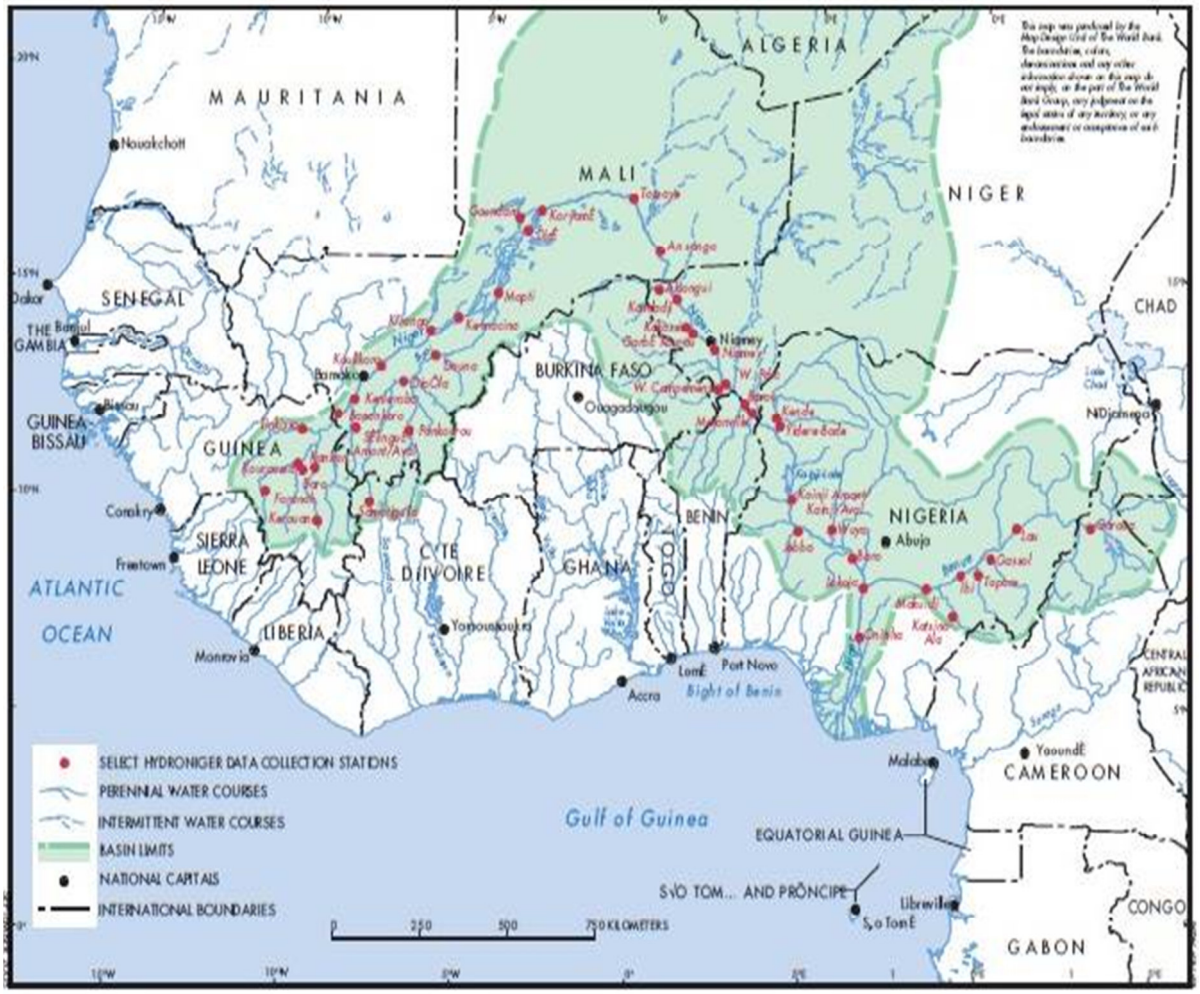

Fig. 1:Selected Hydrological Stations of the Niger River(IBR&D/WB)

2.2 Preliminary Test of Flood Data

The hydrologic data series , was subjected to the following tests; (i) Spearman Rank Order Correlation (ii) Theil Sen Trend (TSA) test and (iii) Trend – Free Pre-Whitening (TFPW) method.

2.2.1 Spearman's rho (SP) test

The Spearman's rho test like the Mann – Kendall test is a non – parametric test. The method distribution - free and has almost uniform power for linear and non-linear trends (Dahmen and Hall, 1990; Hamed, 2016). The Spearman's Rank statistics R_{sp} and the standardized test statistics Z_{SR} are:

$$R_{sp} = 1 - \frac{6 * \sum_{i=1}^{n} (R_i - m)^2}{n(n^2 - 1)} \quad \ldots\ldots\ldots\

$$\rho_k = \frac{\frac{1}{n-1}\sum_{t=1}^{n-k}(x_t - x_t)(x_{t+1} - x_t)}{\sum_{i=1}^{n}(x_t - x_t)^2}$$

r_k = ... 7

$$x_t = \frac{1}{n}\sum_{t=1}^{n} x_t$$

here 8

Where x_i is an observation, x_{t+1} is the following observation, x is the mean of the time series and n is the number of data. Based on the value of the first order autocorrelation (r_1), the hypothesis Ho: r_1(that there is no correlation between two consecutive observations) against the alternate hypothesis, H1: $r_1 < > 0$. (Dahmen and Hall, 1990): defined the critical region, u at the 5% level of significance as;

$$\frac{-1-1.96\sqrt{n-2}}{n-1} \leq r1_i \leq \frac{-1+1.96\sqrt{n-2}}{n-1}$$ 9

2.2.3 Theil Sen Trend (TSA) Test.

The Theil-Sen trend line (Helsel, 2005) is a non-parametric alternative to linear regression which can be used in conjunction with the Mann-Kendall test to estimate the magnitude of the detected trend. The Sen's method uses a linear model to estimate the slope of the trend and the variance of the residuals should be constant in time (Da Silva et al., 2015). The Theil – Sen (Q_{ik}) is calculated as:

$$Q_{ik} = \frac{Q_j - Q_k}{j-k} \quad \text{for i = 1, ... n} \quad .....10$$

Where Q_j and Q_k are the data points at time j and k (j - k), respectively. With a sample size of n, there should be a total of N = n(n-1)/2 such pairwise estimates Q_{ik}.

The Theil-Sen's Estimator of the slope is the median of the N values of Q_{ik}. The N values of Q_{ik} are ranked in ascending order, smallest to the largest and the Sen's estimator is calculated as:

$$Q_{ik} = \begin{cases} Q\left[\frac{n+1}{2}\right]; \ If\ N\ is\ odd \\ Q\left[\frac{n}{2}\right] + Q\left[\frac{n+1}{2}\right]; \ If\ N\ is\ even \end{cases}$$ 11

The sign of Q_{ik} reflects the data trend, while its numerical value indicates the steepness of the trend. Positive or negative slope is obtained as upward (increasing) or downward (decreasing) trend.

2.2.4 Trend-Free Pre-Whitening – Mann – Kendall Procedure (TFPW-MK)

The trend - free - pre – whitening procedure was applied to remove AR(1) process as follows (Yue et al., 2003);

1) The slope (β) of a trend in sample data is estimated using the approach proposed by (Theil, 1950 and Sen, 1968). The trend is assumed to be linear, and the sample data are detrended by:
   $Q_d = X_t - T_t = Q_o - \beta t$ 12
   Where Q_d is detrended time series, Q_o is original time series, β is slope estimated by Theil-Sen approach, t is time.
2) The lag-1 serial correlation (r_1) of the detrended time series Q_t is computed using (Yue et al, 2003), thereafter the AR(1) is removed from Q_d, resulting in pre-whitened times series.
   $Q_{pw} = Q_d - r_1 * Q_{d-1}$ 13
   Where Q_{pw} is pre – whitened time series, the application of Equation 13 is know as trend-free prewhitening (TFPW) procedures. The residual serves after applying the TFPW procedure (Q_{pw}) should be an independent series.
3) The identified trend T_t (=βt) and Equation 13 are blended by
   $Q_{bd} = Q_{pw} + T_t = Q_d - r_1 Q_{d-1} + \beta^* t$.....14
   Where Q_{bd}, is blended series, could preserve the true trend and is no longer influenced by the effect of serial correlation. In Equation 13, the trend (β*t) was added back to the autocorrelation data before applying the MK test in order for its significance.

2.3 Probability Density Functions

In flood frequency analysis, a random variable is used to fit probability distribution which in turn is used to extrapolate recorded events and design events either graphically or analytically by estimating the parameters of the distribution. The probability density functions, cumulative density functions the associated unbiased plotting position formulae are shown in table1.

Table1: Cumulative Density Function, Probability Plots and Unbiased Plotting Position

S/N	Distribution of random variable x	Cumulative Density Func.(CDF)	Coordinates of Probability Plot		Recommending Plotting Position Formulae	Relationship(Pi)
			X- axis	Y- axis		
1.	Gumbel EV1	$F(x)=\exp\left[-\exp\left(-\frac{Q-\mu}{\alpha}\right)\right]$	-Ln(-Ln(Pi))	$Y = Q_i$	Gringorten	(i + 0.4)/(N + 0.12)
2.	Weibull	$F(x) =1- \exp[-(Q/\alpha)^k]$	Ln(-Ln(1 - Pi))	Y=LnQi	Weibull	i/(N + 1)
3.	Normal	$F(x)=\Phi[(x-\mu)/\sigma]$	$X = \Phi^{-1}(Pi)$	$Y = Q_i$	Blom	(i–0.375)/(N+0.25)
4.	Lognormal	$f(x)=\frac{1}{x\sigma y\sqrt{2\Pi}}-\exp\left[\frac{-(y-\mu)^2}{2\sigma^2}\right]$ $F(x)=\Phi\left[(LnQ-\mu L_{nQx})\right]/\sigma L_{nx}$	$X = \Phi^{-1}(Pi)$	$Y=LnQ_i$	Blom	(i-0.375)/(N+ 0.25)

F(x) is CDF, parameters α and μ are function of the mean, and standard deviation, standardized form: μ = 0, σ = 1 and Φ is CDF of standardized Normal (0,1) distribution and Φ^{-1} is inverse of the cumulative standard normal distribution

2.4 Probability Plots

2.4.1 Probability Plot

Probability plot is a graphical techniques for assessing whether or not a data set follows a given distribution. The data are plotted against a theoretical distribution in such a way that the points should form approximately a straight line (El-Shanshoury and El-Hemamy, 2013). The mathematical relation of a typical probability plot is given by Equation 15;

$$\hat{Q} = \beta * g(Fi) + \alpha \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad 15$$

If the estimate of the observed Q is $\hat{Q}$ then the relationship between the observed Q and g(Fi) should be linear. If the fitted distribution is exactly the parent distribution. Then by plotting the observed data Q against g (Fi), these distributions which give a straight line relationship on the probability plot can be selected. Table 1 contains the probability distributions and coordinates of the x-axis and Y-axis on the probability plots of the evaluated probability distributions and unbiased plotting position formulae. The slope(β) and intercept (α) in Equation15 are parameters of the model.

2.4.2 Probability Plot Correlation Coefficient (PPCC)

The PPCC test was developed by Filliben (1975) as a powerful test among many goodness fit tests for normality and non-normality. The test measures the linearity of the plot under an assumed distribution and provides a quantitative measure for comparing the relative goodness of fit of a fitted distribution and also with sufficient power to discriminate between different distribution hypothesized (Vogel, 1986l Stedinger et al., 1993) Aminataee and Montaseri, 2013).

The equation for the probability plot correlation coefficient is expressed as:

$$r = \frac{\sum_{i=1}^{n} (Q_i - \bar{Q})(Q_m - \bar{Q}_m)}{\sqrt{\sum_{i=1}^{n} (Q_i - \bar{Q}_m)^2 \sum_{i=1}^{n} (Qs_{im} - \bar{Q}_m)^2}}$$

Given the level of significance (α), determine the critical point (r_{cp}) the Filliben's test which depends on the sample size n. Compare the probability plot correlation coefficient (r) against the critical point (r_{cp}). If $r \geq r_{cp}$, conclude that normality is a reasonable model for the underlying population at the α-level of significance. If, however, $r < r_{cp}$, reject the null hypothesis and conclude that another distributional model would provide a better fit.

Where $\bar{Q}$ and $\bar{Q}_m$ are the mean values of the observation X_i and the fitted quantiles Q_i, respectively, and n is the sample size.

3.0 Model Evaluation Statistics

The percent bias (PBIAS); RMSE – Observation standard deviation ratio (RSR) and Nash-Sutchiffe efficiency (NSE) where used to evaluate the quantiles of the data against the quantiles of the standardized empirical distribution. (Moriasi *et al*., 2007)

i) The Percent bias (PBIAS) measures the average tendency of the standard data (Q_{sim}) to be larger or smaller than the observed data (Q_{obs}). The optimal value of PBIAS is 0.0, with the low-magnitude values indicating accurate model simulation. PBIAS is calculated with Equation 16:

$$PBIAS = \left[\frac{\sum_{i=1}^{n}(Q_{obs} - Q_{sim}) * 100}{\sum_{i=1}^{n}(Q_{obs})}\right] \quad \ldots\ldots 16$$

Where PBIAS is the deviation of data being evaluated, expressed as a percentage.

ii) Nash-Sutcliffe efficiency (NSE) determines the relative magnitude of the residual variance ("Noise") compared to the measure data variance (observed data). NSE is computed according to Equation 17.

$$NSE = I - \left[\frac{\sum_{i=1}^{n}(Q_{obs} - Q_{sim})^2}{\sum_{i=1}^{n}(Q_{obs} - \bar{Q})^2}\right] \quad \ldots\ldots 17$$

Where Q_{obs}, Q_{sim} are observation and simulated times, $\bar{Q}$ is the mean of the observed data, and n is the total number of observation.

iii) RMSE – observations standard deviation ratio (RSR), standardized the RMSE using the observations standard deviation and it combines both an error index and the additional information. RSR is calculated as the ratio of the RMSE and standard deviation of measured data, as shown in the Equation 18.

$$RSR = \frac{RMSE}{STDEV_{obs}} = \left[\frac{\sum_{i=1}^{n}(Q_{obs} - Q_{sim})^2}{\sqrt{\sum_{i=1}^{n}(Q_{obs} - \bar{Q}sim)^2}}\right] \quad \ldots\ldots 18$$

The lower RSR; the lower the RMSE, and the better the model simulation. The best probability distribution is one with the lowest RSR value.

4.0 Analysis, Results and Discussion

4.1 Analysis

The main results comprises the descriptive statistics, auto-correlation analysis, Spearman Rank Order (SROC), Probability Plot Correlation Coefficient (PPCC) test, probability plots, Sen's slope estimation of the Annual Maximum Series (AMS). The time series of AMS is shown in Figure 2 with a discernable negative trend typical of most time series in the Niger River Basin. Figures 3 – 6 show the probability plot of Niandan River while Figures 7 – 10 presents the probability plot correlation coefficient tests. The abscissas and ordinates of the probability plots is represented by the expression (Φ^{-1}(Pi), Y_i). Thus, the probability plot is a plot of the sample quantiles Y_i against theoretical quantiles of Qi, where Φ^{-1} is the inverse of the cumulative standard normal distribution, and Pi denote the appropriate quantiles probabilities. In the paper, Φ^{-1} (P_i) and the theoretical quantile (Q_m) in Equation 15 were calculated using the MS Excel built-in functions; NormInv (P_i) and Norm. Inv (P_i, mean, standard deviation). The results of the descriptive statistical analysis are summarized in Table 4.1, while Table 4.2 shows the PPCC test results and efficiency criteria and Table 4.3, the performance ranking of model evaluation indices. Finally , Figures 1 – 10, shows the PPCC and Probability plots.

4.2 Results

Table 4.1: Descriptive Statistics

S/No	Parameters	Values
1	Mean discharge (Q)	698.15m^3/sec
2	Standard deviation (σ)	188.71m^3/sec
3	Skewness	0.145
4	Kurtosis (C_k)	-0.889
5	Sample auto-correlation (r)	-0.320
	Coefficient (r) After TFPW	-0.025
	Confidence interval (LCL-ULC) range	$-0.360 \leq r \leq 0.302$
6	Sen's Estimation (β)	-7.61m^3/s/yr
7(a)	Coefficient of variation	-0.263
7(b)	Coefficient of Trend (r_s)	0.570
8(a)	SROC Critical value at $t_{0.005,51}$	-2.0
8(b)	T - statistic	4.99

In Table 4.1, the coefficient of variation (CV) lie in the range;f $0.17 \leq CV \leq 0.352$, indicating that the AMS is moderate to highly variable and also positively skewed thus a non-normal distribution. The Kurtosis coefficient and excess coefficient (E) defined as $E = C_K - 3$, gave negative value, indicating a platykurtic and non-normal type frequency distribution. The auto-correlation coefficient (r_1) of the original data ($r_1 = 0.320$) fall outside the critical region given by Equation 9, thus indicating that the data is serially correlated at significance level of 5%.The SROC test statistic (t) is greater than the critical point of the t – distribution $t_{0.005,51}$ at the 5% significant level as shown in Table 4.1. Consequently, there is no statistically significant long-term trend in the AMS. However, since the auto-correlation coefficient (r_1) falls outside the critical region. Trend free Pre-Whitening approach was applied to the original data following the steps in sub-section 3.1.5. Thus, log-one auto-regressive AR(1) component was removed and the data became fit for flood frequency analysis.

The probability plot is a graph of quantiles of sample data plotted against quantiles of the standardardized theoretical distribution. The distribution which causes the data to be most like a straight line on its probability plot is the one which most closely resembles the distributional shape of the data. The PPCC test is an aided instrument to augment the visual determination of linearity on the plot(Helsel and Hirsch, 2002).Table 4.2 show the results of the PPCC test, Probability plots and model evaluation indices. The best – fit distribution will be selected based on the following criteria; probability plots, probability plot correlation coefficient (PPCC), coefficient of determination(R^2), Percent bias (PBIAS), Nash- Sutcliffe efficiency(NSE) and RMSE-observations standard deviation (RSR).(Moriasi et. Al 2007, Raju and Kumar, 2017).

The acceptance region for each of the evaluation statistics are (i) R^2 is a measure of the explanatory power of the model. R^2 ranges from 0 to 1.0 and a value near to 1.0 indicates good model performance.(ii) The PPCC test statistic (r) is the linear correlation between the empirical and theoretical quantiles. If the PPCC test statistic(r) > $r_{crit,0.05}$ the normality is an acceptable model but if (r) < $r_{crit,0.05}$, then reject the null hypothesis.(iii) Nash – Sutcliffe and PBIAS were used for evaluating the model error, by quantifying the efficiency of the model. NSE ranges from –minus infinity to 1.0. If a model simulates observed conditions perfectly, NSE will be 1.0.The optimal value of PBIAS is 0.0, with low – magnitude values indicating accurate model simulation. Positive values indicate model underestimation bias, and negative values indicate model overestimation bias (Gupta et al 1999).

Table 4.2: Results of PPCC test and Probability Plots

S/No	Probability Distribution	R^2	PPCC test Statistics (r)	$r_{Crit, 0.05}$	PBIAS	NSE	RSR
1	Normal	0.9644	0.9820	0.97745	2.6×10^{-14}	0.9644	0.189
2	Lognormal	0.9563	0.9779	0.97745	1.44×10^{-14}	0.9563	0.209
3	Gumbel EVI	0.9345	0.9815	0.9658	11.79	0.661	0.582
4	Weibull	0.9392	0.9843	0.9660	11.45	0.9797	0.144

The normal distribution is rated best in terms of R2 having a value of 0.9644, seconded by lognormal distribution. The PPCC test statistic (r) > $r_{crit,0.05}$ for all the distributions evaluated. Consequently, the four distributions in Table 4.2 are plausible models of the underlying population at the 5% level of significance. The lognormal distribution is best rated in terms of PBIAS, seconded by normal distribution. The Weibull is rated best in terms of NSE and RSR indices, seconded by Normal distribution. Based on the five model evaluation criteria and the performance scoring of Table 4.3, the best – fit probability distribution model for Niandan river at Baro is Normal distribution, seconded by Weibull, Lognormal- third and Gumbel EV1, the fourth position. .

Table 4.3: Performance Ranking of Probability distributions

Plotting Position Formulae	Probability Distribution	Model Evaluation Indices					TOTAL SCORE
		R^2	PPCC	PBIAS	NSE	RSR	
Blom	Normal	1	1	2	2	2	8
Blom	Lognormal	2	1	1	3	3	10
Gringorten	Gumbel	4	1	4	4	4	17
Weibull	Weibull	3	1	3	1	1	9

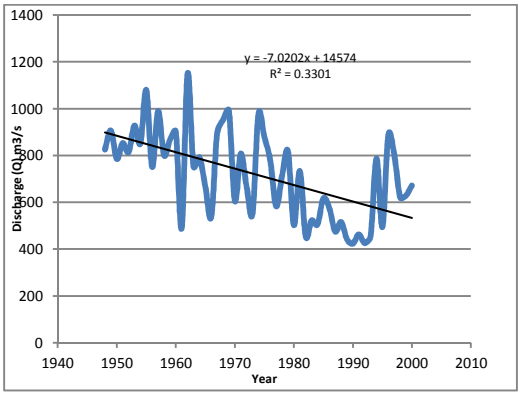

Figure 2. Annual Maximum Series of Niandan River at Baro.

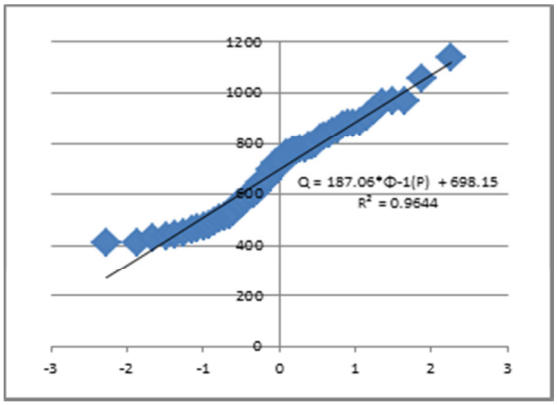

Figure 3. Normal Probability Plot, Niandan River at Baro.

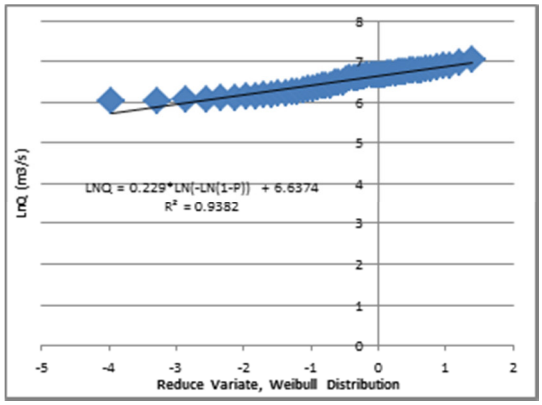

Figure 6. Weibull Probability Plot, Niandan River at Baro

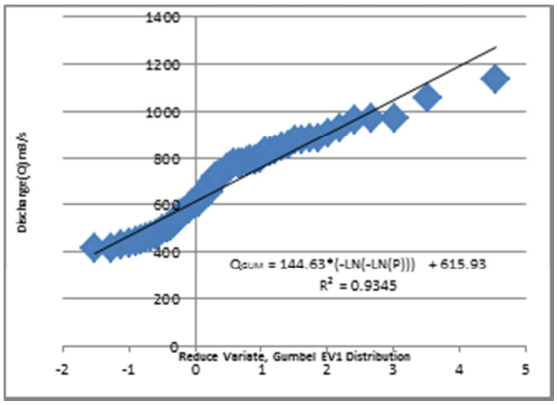

Figure 7: PPCC 's GoF For Normal Distribution

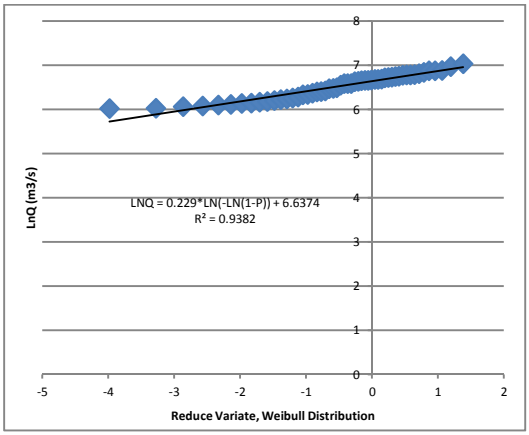

Figure 4. Lognormal Probability Plot, Niandan River at Baro.

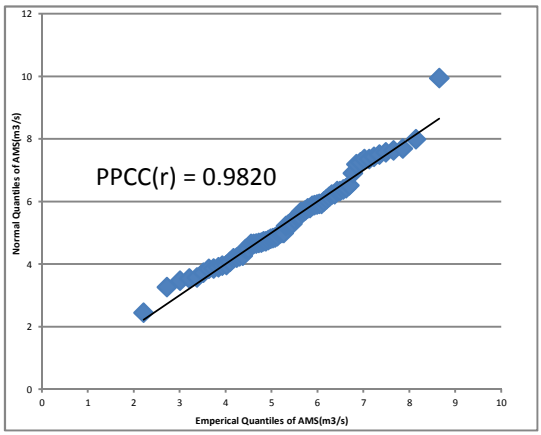

Figure 5. Gumbel EV1 Probability Plot, Niandan River at Baro.

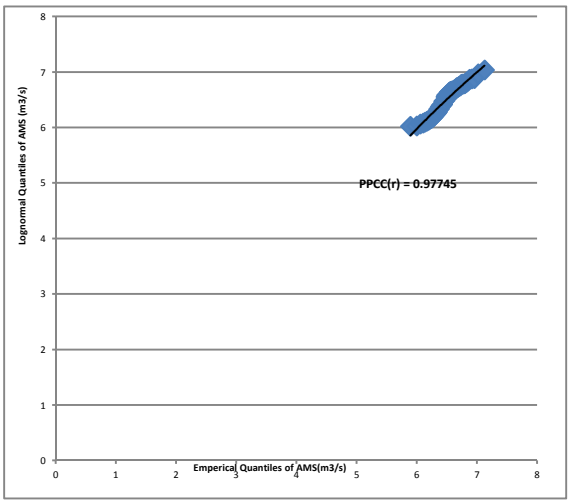

Figure 8: PPCC Filliben's GoF For Lognormal Distribution

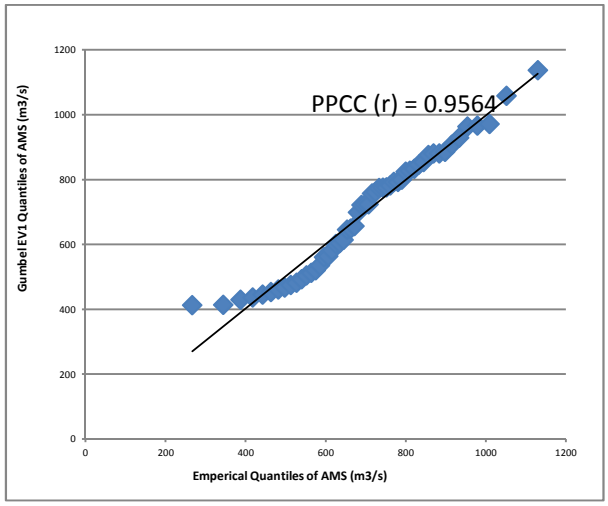

Figure 9: PPCC Filliben's GoF For Gumbel EV1 Distribution

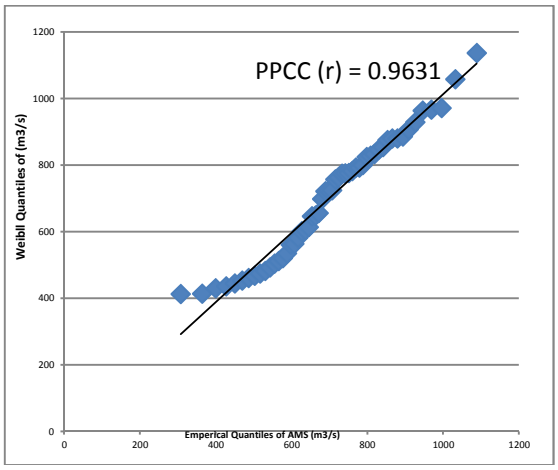

Figure10: PPCC Filliben's GoF For Weibull Distribution

Discussions

These findings disagree with Ehiorobo and Akpejiori, (2016) and Ibrahim et al., (2016) who found Lognormal distributions most suitable for flooding frequency analysis at Agenebode on the Niger River and at Hadejia-Jama'are River Basin respectively in Nigeria. Also Ibrahim *et al*., (2009) found Pearson type 3 distribution the best-fit for Guruara River at Jere while Mamman *et al*., (2017) found Gumbel EVI the Best-fit probability distribution model for the prediction of inflows at Kanji dam reservoir. Similarly,Izinyon and Ajumuka (2013) evaluated flood prediction models for three flow gauging stations in upper Benue River Basin and found Gumbel EV1, log Pearson Type 3, and Lognormal for the stations at River Donga at Manya,River Donga at Donga and River Bantaji at Suntai.

In conclusion, no single distribution can be specified as the best-fit for the whole of Nigeria, Finally, the results of Table 4.2 indicate that the Normal distribution is the best-fit distribution while Weibull is the second - best distribution for Niandan River at Baro.

Conclusion

This paper investigates the selection of an appropriate probability distribution for at- site flood frequency analysis using AMS of Niandan River at Baro. A total of four good – of – fit tests were employed, i.e. probability plots, probability plot correlation coefficient (PPCC), Percent bias (PBIAS), Nash- Sutcliffe efficiency(NSE) and RMSE-observations standard deviation (RSR) to identify the best – fit distribution model. The Normal and Weibull distributions have been identified first and second – best distributions. Literature review of flood frequency analysis across Nigeria show that no single distribution is found superior to the others as the case in countries like Australia and United States of America. Consequently, research work for the search of best – fit distribution for Nigeria will continue in the future.

References

Abida, H.. & Ellouze, M.(2007), Probability distribution of flood flows in Tunisia, Hydrol. Earth Syst. Sci. Discuss., 4, 957 – 981

ADELOYE, A.J., & MONTASERI, M. (2002),Preliminary streamflow data analyses prior to water resources planning study, Hydrological Sciences Journal, 47(5), 679 -602.

Bayazit, O.B. (1995), Best-fit distributions largest available flood samples, Journal of Hydrology, 167 (1995), 195 – 208.

Berz, G.(2000), Flood disasters: lessons from the past-worries for the future, Proc. Instn Civ. Engrs Water & Mar. Engng, 142, Mar., 3 – 8. Paper 12212.

Chatfield, C. (2004). The Analysis of Time Series: An Introduction (6th Edition). Boca Raton, FL: Chapman and Hall.

Dahmen, E.R., & Hall, M. J. (1990). Screening of Hydrological Data: Tests for Stationarity and Relative consistency, Intern. Inst. for Land Reclamation and Improvement/ILRI, P. O. Box 45, 6700 AA Wageningen, The Netherlands, Publication 49.

Di Baldassarre, G; Montanari, A., Lins, H., Koutsoyiannis, D., Brandimarte, L., & Bloschl G. (2010). Flood Fatalities in Africa: From diagnosis to mitigation, Geophysical Research letters, 37, L22402, doi:10.1029/2010GL045467, 2010.

Ehiorobo, J.O. & Akpejiori I.J (2016),Flood Frequency Analysis of River Niger at Agenebode, Edo State, Nigeria , Journal of the Nigerian Association of Mathematical Physics, Volume 38, pp.309 – 318.

FLOODFREQcost Action ES0901, REVIEW OF APPLIED-STATISTICAL METHODS FOR FLOOD – FREQUENCY ANALYSIS IN EUROPE(WG2).

GHORBABI, M.A., RUSKEPP, A.A. H., SINGH, V.P., & SVAKUMAR, B. (2010), Flood frequency analysis using Mathematica, Turkisk J.Eng. Env. Sci. 34(2010), 171 -188

Ibrahim, U.A., Yadima, S.G.,& Nur Alkali A(2016) Flood Frequency Analysis at Hadejia River in Hadejia-Jama'are River Basin, Nigeria, Civil and Engineering Research, Vol.8. No.9. , ISSN 2224-5790(Paper) ISSN 2225-0514(Online).

Mahe, G., & Olivry, J.C. 1991. "Les changements climatiques et variations des ecoulements en Afrique occidentale et Centrale, du mensuel a l' internuel" in Hydrology for the water management of Large River Basins. Proceedings of the Vienna Symoposium, August, 1991. IAHS series of Proceeding and Reports 201: 163 – 172.

Mamman, M.J., Otache, Y.M., Ibrahim, J., & Shaba M.I (2017), Evaluation of Best –Fit Probability Distribution Models for the Prediction of Inflows of Kanji Reservoir, Niger State, Nigeria, Air, Soil and Water Research, Volume 10: 1-7, DOI:10.1177/117862211768 1034

Manta, I.H, & Isiguzo E.A (2009) Flood frequency analysis of Gurara River catchment at Jere, Kaduna State, Nigeria, Scientific Research and Essay Vol. 4(6), pp.636 – 646.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R.L., Harmel, R. D., & Veith, T. L. (2007). Model evalutionguildline for systematic Quantification of Accuracy in Watershed Simulation. Trans, ASABE, 50(3), 885 – 901

Rahman, A.S., Rahman, A., Zaman, M.A., Haddad, K., Ahsan, A.,& Imteaz M(2013), 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.

Raju, S. K. & Kumar (2018), Impact of climate Change on Water Resources, with Modeling Techniques and Case Studies, Springer Nature Singapore pte Ltd. 2018, ISBN 978-981-10-6109-7.

Rao, A.A & Hamed, K.H (2000), FLOOD FREQUENCY ANALYSIS, CRC Press,ISBN 0-412-55280-9.

Richard, M.V. & Wilson I (1996), Probability distribution of Annual Maximum, Mean, and Minimum streamflows in the United States, Journal of Hydrologic Engineering, Vol. 1No.2, pp.69-76.

Sangare, S.(2001)"Bilan Hydrologique du Niger en Guinee". Actes du Colloque Friend AOC, Capetown, March, 2001.

The International Bank for Reconstruction and Development (IBRD)/TheWorld Bank(WB):The Niger River Basin: A Vision for Sustainable Development; Andersen, I., Dione, O., Jarosewich – Holder, M., Olivvry, J.C.Editor: Katherine, George Golitzen. DOI: 10.1596/978-0-8213-6203-7.

Yue, S., Pilon, P., & Phinney, B. (2003). Canadian Streamflow trend Detection: Impact of serial and cross-correlation; Hydrological Science Journal, 48:1, 51-63, DOI: 10.1623/hysj.48.1.51.43478.

The World Meteorological Organization (WMO-No. 718, 1989). " Statistical Distributions for Flood Frequency Analysis, Operational Hydrology Report No. 33.