Transactions of the Nigerian Association of Mathematical Physics
Volume 18, (January – December, 2022 Issue), **pp69 –76**
© Trans. of NAMP

FLOOD FREQUENCY ANALYSIS OF OSSE RIVER WITH A THRESHOLD VALUE USING EXTREME VALUE TYPE 1 DISTRIBUTION

Akata I. U., Iduseri A. and Osemwenkhae J.E.

Department of Statistics, University of Benin, Benin City, Nigeria.

Abstract

This paper analyzes flood frequency using discharge data from Osse River with flow measurements carried out at-site station Iguoriakhi for 20 years' period (1994-2013). Some basic statistics and goodness of fit test of the datasets were examined. The recurrent interval (T) and the probabilities for different years return period were shown. Using the method of moment for parameters estimates and with a threshold value, the extreme value type1 distribution was used to predict the future maximum flood peak starting from 1.2years up to 50years.

***Keywords*: Extreme Value Type1, Method of Moment, Recurrent Interval, Flood Frequency Analysis, Probability.**

1.0 Introduction

In real life phenomena, most especially in hydrological and geomorphological studies, determining the return period and its magnitude of occurrence of hydrological measures or simply estimating the discharge value for a given return period is very paramount. Amidst the countless variety of principles governing statistics and probability that can be used to estimate return period and it magnitudes of given discharge outcome from the series of historical flows (Weibull, Exponential, Gumbel, generalized extreme values, generalized Pareto, generalized Logistic, Poisson distribution etc.), the Gumbel method was found to be appropriate for these types of estimates [1-6]. Flow discharge that are relatively high and overflow the banks of a river may be referred to as floods. Almost every year, a lot of resources are invested for flood mitigation and protection using either structural or non-structural. According to [7], meteorological forecast can only provide very short forecasts in a precise manner which may not allow enough time to reduce the impact of flood events in a locality. Given the above limitation of flood forecasting, this study attempts to estimate the probability values of T years return period and also the flood peaks of different magnitude from a recorded historical data using statistical methods. For this study, we adopted the extreme value type1 distribution due to its wide acceptance in flood frequency analysis.

2.0 Data and Methods

2.1 The Study Site

The study site where the flood frequency was carried out is Benin-Owena River Basin Development Authority (*BORDA*) and the gauging station location is Osse River (Iguoriakhi) as shown in Figure 1.

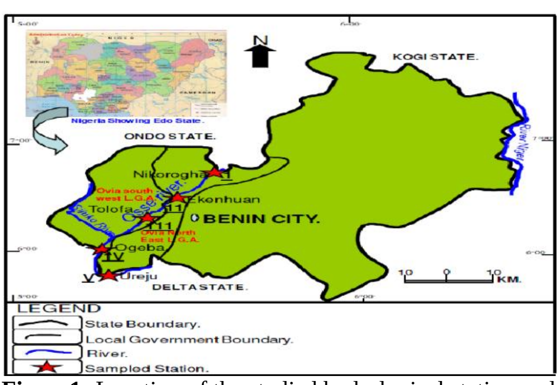

Figure1: Location of the studied hydrological station and simplified geological map of Osse River showing sample stations. Source [8].

Correspondence Author: Akata I.U., Email: innocent.akata@physic.uniben.edu, Tel: +2347063333338

Osse River transverse through Gelegele to Iziedema communities within the tropical rainforest belt in Ovia North-East Local Government Area of Edo State, Southern Nigeria. Osse River stretches between latitudes 06°12'N and 06°10'N and longitudes 05° 20'E to 05° 22'E in Ovia North-East Local Government Area of Edo State. Osse River drains into the Benin River which empties itself into the Atlantic Ocean. This study area falls within the rainforest belt of Nigeria, with a wet season (March to October) and a dry season period (November to March) [9].

2.2 Extreme Value Type 1 Distribution

The extreme value type 1 distribution also referred to as Gumbel distribution is a statistical method often used for predicting extreme hydrological event such as flood [10]. The cumulative distribution function (cdf) of two parameter Gumbel distribution is written as,

$$F(x) = \exp\left[-\exp-\left(\frac{x-\varepsilon}{\alpha}\right)\right] \tag{1}$$

and the corresponding probability density function (pdf) is defined as

$$f(x) = \alpha^{-1}\ \exp\ -\left(\frac{x-\varepsilon}{\alpha}\right)\exp\left[-\exp-\left(\frac{x-\varepsilon}{\alpha}\right)\right] \quad \alpha > 0,\ -\infty < x,\ \varepsilon < \infty \tag{2}$$

where ε is the location parameter and α is the scale parameter. From Equation (1) and (2), when the scale parameter $\alpha = 1$ and the location parameter $\varepsilon = 0$, we have the standardized Gumbel distribution.

2.2.1 Moment Generating Function (MGF) of the Extreme Value Type 1 Distribution

The mean and variance obtained through the MGF will be the basis of our computations through which the parameters of the distribution will be calculated. The moment generating function of a continuous random variable *X* is defined by,

$$M_X(t) = E\left[e^{tx}\right] = \int_{-\infty}^{\infty} e^{tx}\ f(x)\ dx \tag{3}$$

where $f(x)$ is the pdf defined in Equation (2), hence the MGF of a random variable Z having the Gumbel distribution with $Z = \frac{x-\varepsilon}{\alpha}$ Z , $\alpha = 1$ and $\varepsilon = 0$, is given as,

$$M_X(t) = E\left[e^{tx}\right] = \int_{-\infty}^{\infty} e^{tx}\ e^{-z}\ e^{-e^{-z}} dz = \int_{-\infty}^{\infty} \left(e^{-z}\right)^{-t} e^{-z}\ e^{-e^{-z}} dz \tag{4}$$

Let y = e^{-z}, when Z = ∞, y = 0, and when Z = -∞, y = ∞ hence –dy = $e^{-z}dz$. Therefore, $M_Z(t) = E\left[e^{tZ}\right] = \int_0^{\infty} y^{-t} e^{-y} dy$ = $\int_0^{\infty} y^{(1-t)-1} e^{-y} dy$

But $\Gamma_t = \int_0^{\infty} y^{t-1} e^{-t} dy$ from gamma function, then

$$E\left[e^{tZ}\right] = M_Z(t) = \Gamma(1-t),\ -\infty < t < 1. \tag{5}$$

Since [11] defined

$$\Upsilon = -\Gamma'(1) = -\int_0^{\infty} e^{-x} \ln x dx \approx 0.5772156649,$$

then the mean and variance of the Gumbel distribution can be calculated by differentiating Equation (5) once and twice respectively and equating t = 0. Therefore,

$$M'_Z(0) = -\Gamma'(1) = \Upsilon. \tag{6}$$

Similarly,

$$M''_Z(0) = \Gamma''(1) = \int_0^{\infty} (\ln x)^2 e^{-y} dy = \Upsilon^2 + \frac{\lambda^2}{6} \tag{7}$$

and the variance,

$$Var(Z) = \frac{\lambda^2}{6} \tag{8}$$

Since any Gumbel variable X with location parameter ε and scale parameter α can be expressed as a linear transformation of a standard Gumbel distribution Z i.e Z = $\frac{x-\varepsilon}{\alpha}$ which implies X= ε + α Z, then the two-parameter Gumbel distribution with location ε and scale parameter α is given by

$$M_X(t) = M_\varepsilon + \alpha Z(t) = E\left[e^{(\varepsilon+\alpha z)t}\right] = e^{\varepsilon t}.E\left[e^{\alpha z t}\right].$$

Based on the principle behind Equation (5),

$$M_X(t) = e^{\varepsilon t}\Gamma(1-\alpha t) \quad (9)$$

By taking the *log* of equation (9), we obtain the cumulant generating function denoted by $\phi(t)$. Then,

$$\log M_X(t) = \log e^{\varepsilon t}\Gamma(1-\alpha t) = \phi(t) \quad (10)$$

The mean and variance of the two-parameter Gumbel distribution can be calculated following the same approach stated above by differentiating Equation (10) once and twice respectively and equating t = 0. This yields

$$\phi^{/}(0) = E(x) = \varepsilon + 0.5772\alpha \quad (11)$$

and

$$Var(x) = \frac{\alpha^2 ƛ^2}{6} \quad (12)$$

where $\bar{\alpha} = \frac{S\sqrt{6}}{ƛ}$ see [12].

Substituting Equation (12) into Equation (11) yields

$$\bar{\varepsilon} = E(X) - 0.5772\frac{S\sqrt{6}}{ƛ} \quad (13)$$

where E(X) is the sample mean, S is the sample standard deviation and $ƛ$ = 3.142 respectively.

It is often the case of reality that hydrological systems are influenced by extreme events namely: a several magnitude of rainfall can cause excessive flooding, a minimum amount of rainfall can as well cause drought [13]. The magnitude of these extreme event is inversely proportional to their frequency of occurrence for example, it is evident that several flooding occurs less often than flood of a lesser magnitude, the main discourse of the hydrological experts is to perform an analysis on the frequency of these extreme events whether minimum or maximum.

2.3 The Return Method

An extreme event is defined to have occurred if a random variable X is greater than or equal to a particular level, say x_T. The time between the occurrences of $X > x_T$ is called the recurrence interval (τ).The expected value of τ, E(τ) is the average number of years in which the event $X > x_T$ return or happen again. Thus E(τ) is the return period "T" of the event $X > x_T$. Note that x_T is depending on T the return period because a flood of a given magnitude depends on the frequency of its occurrence happening within a period. This concept of return period is used to describe the likelihood of occurrences. The probability $P = P(X > x_T)$ of occurrence of the event $X > x_T$ in any observation is related to the return period as follows; For any observation, two outcomes are possible: "success denoted as P i.e. $X > x_T$" or "failure (1-P) i.e. $X < x_T$". Since the observations are independent, the probability of a recurrence interval of duration (τ) is the product of probability of $\tau - 1$ failures followed by a success i.e.

$$(1-P).(1-P)...(1-P).P = (1-P)^{\tau-1}.P \quad (14)$$

The expected value of τ is $E(\tau) = \sum_{\tau=1}^{\infty}\tau(1-P)^{\tau-1}.P$ When further simplified yields

$$E(\tau) = P\left[1 + 2(1-P) + 3(1-P)^2 + 4(1-P)^3 + \ldots\right] \quad (15)$$

The expression within the square bracket in Equation (15) has the form of power series expansion

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 + \frac{n(n-1)(n-2)}{3!}x^3 + ...$$

with $x = -(1-P)$ and n = -2, then equation (15) becomes

$E(\tau)=P(1-(1-P))^{-2} = \frac{1}{P}$.

Thus the return period

$$E(\tau)=T=\frac{1}{P}=P(X \geq x_T) \tag{16}$$

From Equation (16) we have that the probability of occurrence of an event in any observation is the inverse of its return period. Another useful question is what is the probability that a T years return period event will occur at least once in N years? Here we consider the situation that in N years the event $X > x_T$ does not occur at all, this gives the probability $(1-P)(1-P)(1-P)...$ N times = $(1-P)^N$.

The complementary event is that the T year event occurs at least once in N year period and the probability is $1-(1-P)^N$. Since $P=\frac{1}{T}$, then the required probability is

$$\zeta = 1-\left(1-\frac{1}{T}\right)^N \tag{17}$$

Equation (17) is very important especially when calculating the probability of an event equaling or exceeding a particular magnitude given a return period.

2.3.1 The Gumbel Model

Let y = $\frac{x-\varepsilon}{\alpha}$, then the cdf in Equation (1) becomes

$$F(y) = \exp[-\exp-(y)] \qquad -\infty \leq y \leq \infty \tag{18}$$

Taking the log of both side of Equation (18) gives

$\ln F(y) = -\exp-(y)$ Then $y=-\ln\left[\ln\left(\frac{1}{F(y)}\right)\right]$

But from Equation (16), $E(\tau)=T=\frac{1}{P}=P(X > x_T)$ which implies $F(x_T)=\frac{T-1}{T}$

Hence $y_T=-\ln\left[\ln\left(\frac{T}{T-1}\right)\right]$ (19)

If a given return period is known, the value of y can be calculated from (19) and then substituted into the relation

$$x_T = \varepsilon + \alpha y_T \tag{20}$$

3.0 Data Analysis and Results

Yearly maximum discharge of Osse River (Iguoriakhi) from the period of 1994 – 2013 collected from Benin Owena River Basin Development Authority, and the scatter plot of the flood peak discharge are presented in Table 1 and Figure 2 respectively.

Table 1: Annual peak flow data of Osse River (1994 to 2013)

S/N	YEAR	FLOOD PEAK (M^3/S)
01	1994	1978
02	1995	2009
03	1996	2010
04	1997	2101
05	1998	2136
06	1999	2150
07	2000	2210
08	2001	2211
09	2002	2365
10	2003	2310
11	2004	2442
12	2005	2547
13	2006	2539
14	2007	2661
15	2008	2650
16	2009	2573
17	2010	2594
18	2011	2603
19	2012	2616
20	2013	2607

Source: [13].

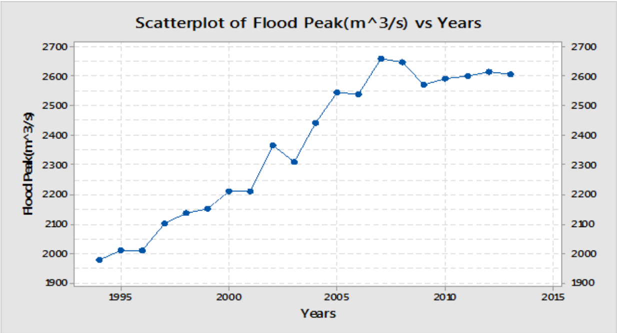

Figure 2: Plot showing the flood peak discharge of Osse River against the corresponding years.

From Figure 2, out of the 20 years' flood discharged recorded, Osse River experiences an upward trend within the first 14 years of the discharge and recorded the highest flood during 2007. The decline in the trend pattern happened during 2008 and thereafter maintain almost a steady level within 2009 - 2013.
To test whether the data follow a Gumbel distribution, an appropriateness of fitting of Gumbel distribution to the peak discharge data is presented in Figure 3.

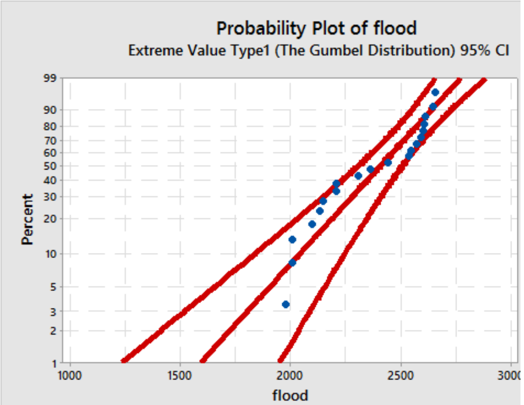

Figure 3: Probability plot showing the flood peak discharge of Osse River against the corresponding percentages.

From Figure 3, it was observed that majority of the dataset clustered around 40 to 90%, while only two data set fall bellows 10%. Also, a cursory look at Figure 3 shows that the data set lies between the upper and lower confidence interval (CI) which confirmed c. To further ascertain the suitability of the distribution for modelling the flood peak data, we computed some basic statistics and goodness of fits using our data set. The obtained results are presented in Table 2.

Table 2: Basic statistics and some goodness of fit test for the data set

Skweness	Kurtosis	Mean	S.D	AIC	BIC	Log
-0.248479	-1.63796	2365.6	244.3	282.0761	284.0675	-139.038

In Table 2, the obtained values of -0.248479 and -1.63796 for skewness and kurtosis are well within any known acceptable range. The skewness value of -0.248479 indicates that the dataset is fairly symmetrical, while the kurtosis value of -1.63796 indicates that the dataset has lighter tails than the normal distribution. Both results support the potential applicability of the Gumbel distribution to represent the dataset. The obtained values for AIC, BIC, and the Log serves as further confirmation of the suitability of the distribution for modelling the flood peak data.

3.1 Recurrent Interval for a given Threshold Value

We adopted the threshold value given by [13] which is the value at which the water overflows its bank and inundate the area. This threshold value is strictly for this particular catchment area. The various years at which the water overflows its bank and inundate the area is presented in Table 3.

Table 3. Flood peaks above a threshold value $(2600M^3/s)$

S/N	YEAR	FLOOD PEAKS (M^3/S)	Recurrent Interval (τ)
1	2007	2661	
2	2008	2650	1
3	2011	2603	3
4	2012	2616	1
5	2013	2607	1

From Table 3, the expected value of the recurrence interval (τ) is computed as

$E(\tau)=\frac{1+3+1+1}{5}$ = 1.2.

Hence, the return period (T) for a flood to exceed or equal the threshold value $(2600M^3/s)$ is 1.2 years.

3.2 Probability of a Flood Equaling or Exceeding the Threshold Value $(2600M^3/s)$

Recall from Equation (16) that $P=P(X\geq x_T)=\frac{1}{T}$, therefore

$$P=P(X\geq 2600)=\frac{1}{1.2}=0.83 \qquad (21)$$

Thus, the probability that a flood of magnitude $2600M^3/s$ will occur at a return period of 1.2 years is 0.83. Hence by using Equation (17), we computed the probabilities that a T years return period will occur at least once in N years. The obtained probabilities are shown in Table 4.

Table 4. The value of probability for different years

ζ	N
0.8300	1
0.9711	2
0.9951	3
0.9992	4
0.9994	5
0.9995	6
0.9996	7
0.9997	8
0.9998	9
0.9999	10

In Table 4, we observed that the probability that a T year return period event will occur in N year's increases as the number of year's increases.

3.3 Estimating Flood Magnitudes Using Extreme Value Type 1 Distribution for given Return Period

Consider the cumulative distribution function (cdf) of the extreme value type1 distribution define in Equation (1)

$$F(x)=\exp\left[-\exp-\left(\frac{x-\varepsilon}{\alpha}\right)\right] \qquad -\infty\leq x\leq\infty$$

Using equation (12) and (13), the sample standard deviation S = 238.1, the sample mean E(X) = 2365.6 and λ = 3.142. Hence,

$\bar{\alpha}=\frac{S\sqrt{6}}{\lambda}$ = 185.7

$\bar{\varepsilon}$ = E(X) – 0.5772 $\frac{S\sqrt{6}}{\lambda}$ = 2365.6 - 0.5772(185.7) = 2258.4.

Thus, the cdf in (1) becomes

$$F(x)=\exp\left[-\exp-\left(\frac{x-2258.4}{185.7}\right)\right]$$

Adopting Equation (19), we have

$$y_{1.2} = -\ln\left[\ln\left(\frac{1.2}{1.2-1}\right)\right] = -0.5832 \quad (22)$$

Substituting Equation (22) into Equation (20) we obtained the value of the flood peak corresponding to the return period of 1.2 years given as

$$x_{1.2} = 2258.4 + (185.7)(-0.5832)$$
$$= 2150 m^3/s$$

The values of the flood peak for the remaining ten return periods are presented in Table 5.

Table 5: Flood Magnitude Corresponding to Different Return Period T

Return period (T) in years	Flood Magnitude (x_T) / Expected Flood
1.2	$2150 m^3/s$
5	$2536 m^3/s$
10	$2676 m^3/s$
15	$2754 m^3/s$
20	$2809 m^3/s$
25	$2852 m^3/s$
30	$2886 m^3/s$
35	$2915 m^3/s$
40	$2941 m^3/s$
45	$2963 m^3/s$
50	$2982 m^3/s$

A cursory look at Table 5 also shows that the expected flood peaks for the given return periods increases as the number of years' increases.

4.0 Conclusion

It is shown in this work that using the method of moments as parameters estimate, the Extreme Value Type1 Distribution was well suited for the prediction of the flood peak. Based on the 20 years historic flood peak data of Osse River, it was observed that the maximum peak of about $2661 m^3/s$ was recorded in 2007. Given a threshold value of $2600 m^3/s$, the recurrent interval (T) of 1.2 was obtained from Table 3. The obtained results for basic statistics and goodness of fit test of the dataset shown in Table 2, and the probability plot showing the flood peak discharge of Osse River against the corresponding percentages confirmed the suitability of the Gumbel distribution for modelling flood peak data.

5.0 References

[1] Dubreuil, P. (1974). *Initiation à L'analyse Hydrologique*; ORSTOM (Office de la Recherche Scientifique et Technique Outre-Mer): Paris, France, ISBN 2-225-40140-3.

[2] Bravard, J. P., and Petit, F. (2000). *Les Cours d'eau—Dynamique du Système Fluvial*; Armand Colin, Collection U: Paris, France, ISBN 9782200251772.

[3] Assani, A.A., Petit, F., Mabille, G. (1999) Analyse des débits de la Warche aux barrages de Bütgenbach et de Robertville (Ardenne belge). *Bull. Soc. Géogr.* Vol. *36*, pp 17–30.

[4] Kidson, R., Richards, K.S. (2005). Flood frequency analysis: Assumptions and alternatives. *Prog. Phys. Geogr.* Vol. *29*, pp 392–410.

[5] Greenwood, J.A., Landwehr, J.M., Matalas, N.C., Wallis, J.R. (1979). Probability weighted moments: Definition and relation to parameters of several distributions express able in inverse form. *Water Resour. Res.* Vol. 15, pp 1049–1054.

[6] Ward, A., Moran, M. (2016). A novel approach for estimating the recurrence intervals of channel-forming discharges. *Water*, Vol 8, pp 269.

[7] Madamombe, R.K. (2005), Flood management practices, selected flood prone areas of Zambezi basin, Zimbabwe National Water Authority", Harare, Zimbabwe.

[8] Osagie, E., Joshua, I.I, and Odaro, S.O. (2015). Assessment of water quality index (WQI) for Osse River in Edo state, Southern Nigeria. *Nigeria Annals of Natural Science*, Vol. 15(1), pp 031-041.

[9] Uwaifo, O.P., Omogbeme, M.I., Olomukoro, J.O. (2018**).** Water quality assessment of Osse River, Gele-gele: A Tributary of Benin River, Southern Nigeria. *Journal of Appl. Sci. Environ. Manage*. Vol. 22 (8), pp 1349 –135.

[10] Haan, C.T. (1977). Statistical methods in Hydrology", Iowa State University Press, Iowa. Shaw, E.M. (1983), "Hydrology in Practice". Van Nostrand Reinhold, UK.

[11] Alexander, L and Verena, W. (2016). Generalized method of moment for estimating parameter of stochastic reaction network. *BMC Systems Biology*. Vol. 10(1):98 DOI:10.1186/s12918-016-0342-8

[12] Adebisi, A.A (1981). The physicochemical hydrology of a tropical seasonal River- Upper Ogun River. *Hydrobiologia.* Vol. 79, pp157-165. DOI:10.1007/BF00006123.

[13] BORDA (2014). "Benin Owena River Basin Development Athourity Hydrological year Book". 1994 to 2013.