Frequency Analysis of Annual Flow Series of River Opeki at Abidogunin Oyo State, Nigeria

Aho M.I[1], Akpen G.D[2] and Godwin-joachim A.A.[3]
[1,2,3]*Civil Engineering Department, University of Agriculture Makurdi, Nigeria.*

Abstract **- Frequency analysis procedure was performed on annual discharge data from River Opeki in Oyo State, Nigeria from 1988 to 2017, utilizing five probability distribution models, to predict flow discharge estimates at 2, 5, 10, 25, 100, 200, 500 and 1000 years return periods. The five models are; Normal, Log-Normal, Gumbel, Log-pearson Type III and Gamma. From the results, Normal distribution predicted discharge values ranging from 8633.35m^3/s to 18265.98 m^3/s for return periods of 2 to 1000 years. Log- Normal distribution predicted discharge values ranging from 7894.05m^3/s to 37748.53 m^3/s for return periods of 2 to 1000 years. Gumbel distribution predicted discharge values ranging from 7622.76 m^3/s to 24295.78 m^3/s for return periods of 2 to 1000 years. Log–Pearson Type III distribution predicted discharge values ranging from 7050.18 m^3/s to 72493.66 m^3/s for return periods of 2 to 1000 years. Gamma distribution predicted discharge values ranging from 7623.47 m^3/s to 21402.47 m^3/s for the same return periods. From the results, Gamma distribution was found most suitable for flow estimate of the River, based on the goodness of fit test using chi-squared analysis. Further research on extreme values of rainfall for the study area is recommended.**

Key words: **River Opeki, Flow Discharge, Frequency Distribution, Goodness of Fit.**

I. INTRODUCTION

Flow which is the rate of discharge of water from seas, lakes, streams or other water bodies due to excessive rainfall often leads to flooding. Rivers can overflow their banks to cause flooding. This happens when there is more water upstream than usual, and it flows downstream to the adjacent low-lying areas (also called floodplains), where there is a burst and water gets into the land. High amounts of water flowing in a river often leads to flooding, and flooding is one of the most common and costly types of natural disasters [1]

Hydrological process like flood is one of the most destructive natural disasters that occur in most parts of the world, and have been identified as the most costly natural hazards having great propensity to destroy human lives and properties [2] There is also the general concern, that the risks resulting from hydrological extremes are on the increase, and this is supported by evidences, both from recent changes in frequency and severity of floods as well as droughts and outputs, from climatic models which predict increases in hydrological variability [3] To manage flood risk successfully, knowledge is needed of both magnitude of any given flow discharge and an estimate of likelihood of its occurrence. It is in line with the above objectives that River Opeki in Oyo State is chosen for frequency analysis.

II. BASIC FREQUENCY ANALYSIS PRINCIPLES

Normal distribution

The general formula for the Probability Density Function (PDF) of the normal distribution is given by [4] as:

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

Where$\bar{x}$ = mean σ = standard deviation, which is evaluated by Equation 2 and 3

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

Where, $Xi$is the magnitude of the ith event and N is the total number of events.

The standard deviation (σ) which is a measure of the dispersion or spread of data set is given by the Equation 3 thus:

$$\sigma = \left[\frac{\sum_{i=1}^{N}(x-\bar{x})^2}{N-1}\right]^{½} \qquad (3)$$

Log-normal distribution

The log-normal distribution assumes that the logarithms of the discharge are themselves normally distributed. It is obtained from the normal distribution if the following substitution is made (2).

$$Y_i = logX_i \qquad (4)$$

Based on logarithm of each value, the mean and standard deviation becomes:

$$log\bar{x} = \frac{1}{N}\sum_{i=1}^{N} logX_i \qquad (5)$$

$$\sigma_{logx} = \left[\frac{\sum_{i=1}^{N}(logX_{i-} log\,\bar{x})^2}{N-1}\right]^{½} \qquad (6)$$

The probability of exceedance is related to the occurrence of particular values by the expression given by:

$$logX = log\bar{X} + K\sigma_{logx} \qquad (7)$$

Where, k is the frequency factor.

Gumbel or Extreme value Type I distribution

Extreme value Type I or Gumbel distribution is one of the most commonly used distributions in flow frequencies. This was first proposed by [5] and was based on the theory of extremes. Its probability distribution function (PDF) is given by (2) as:

$$f(x) = \exp\left(-e^{-\alpha(x-\mu)}\right) - \infty \le x \le \infty \quad (8)$$

The parameters are estimated by the equation 9 and 10

$$\alpha = \frac{\sqrt{6}}{\pi}\sigma \quad (9)$$

$$u = \bar{x} - 0.5772\alpha \quad (10)$$

Where $\propto is$ a scale parameter, and u is the mode of distribution.

A reduced variate y is defined as:

$$y = \frac{x-u}{\propto} \quad (11)$$

For a given return period T, the reduce variate (Y_T) is given as:

$$Y_T = -ln\left[ln\left(\frac{T}{T-1}\right)\right] \quad (12)$$

For the extreme value type 1 distribution, $X_T$is related to $Y_T$by the equation given by:

$$X_T = u{+}\propto y_T \quad (13)$$

Log-pearson Type III distribution

The Log-Pearson Type III distribution is a statistical technique for fitting frequency distribution data to predict the design flow for a river at some site. Once the statistical information is calculated for the river site, a frequency distribution can be constructed. The probabilities of floods of various sizes can be extracted from the curve. The advantage of this particular technique is that extrapolation can be made of the values for events with return periods well beyond the observed flood events. This technique is the standard technique used by Federal Agencies in the United States. It is calculated using the formula:

$$logx = log\bar{x} + k\sigma_{logx} \quad (14)$$

Where x is the flood discharge value of some specified probability, $log\bar{x}$ is the average of the *log x* discharge values, *K* is a frequency factor, and σ is the standard deviation of the *log x* values. The frequency factor *K* is a function of the skewness coefficient and return period and can be found using the frequency factor table. The flow magnitudes for the various return periods are found by solving the general equation. The mean, variance, and standard deviation of the data can be calculated using Equations (15), (16) and (17) below.

$$\text{Mean } log\bar{x} = \frac{\sum(logx_i)}{n} \quad (15)$$

$$\text{Variance} = \frac{[\sum_i^n(logQ - avg(logQ)]^2}{n-1} \quad (16)$$

Standard deviation $\sigma_{logx} = \frac{\sqrt{\sum(logx - \log\bar{x})^2}}{n-1}$ or $\sigma_{logx} = \sqrt{variance}$ (17)

The skewness coefficient C_s can be calculated as follows

$$C_S = \frac{n\sum(logx - \log\bar{x})^2}{(n-1)(n-2)(\sigma_{logx})^3} \quad (18)$$

Where n is the number of entries, x the flow of some specified probability and σ_{logx} is the standard deviation [6]

Gamma or Pearson Type III distribution

This is one of the commonly used distributions in hydrology because of its shape, and it's well known mathematical properties. It has three parameters and is bound in the left with positive skewness. The three parameters are mean, variance and skewness, but the CDF (cumulative distribution function) can be evaluated using the frequency factor. The frequency factor K is a function of the skewness and the return period. To evaluate the T-yrs flood, the probability distribution function given below by [6] is used.

$$Q_T = \bar{x} + k\sigma \quad (19)$$

Where $\bar{x}$ is the mean and σ is the standard deviation

Test of Goodness of Fit

The chi-squared testx^2 is one of the versatile and popular statistics available for data analysis and significance testing. It is used to compare differences between observed and expected frequencies. The observed frequencies are those obtained after an observation or experiment has been carried out (data actually collected from field). The expected frequencies are generated on the basis of hypothesis or speculation (what would be expected, all things being well) as stated by [7].

According to [8] Chi-Square test is used to determine the goodness of fit of distribution of annual peak flow. It is also used to determine how closely actual data fit expected data and is given as:

$$x^2 = \sum_{i=1}^{k}\frac{(b-c)^2}{c} \quad (20)$$

b = Number of observed frequencies, c= Expected frequencies

Where $i = 1, 2, 3 \dots\dots, k$ class interval covering the range, b_i = Number of observations actually in class interval, c_i = Expected number of observations in a given class interval.

In a Chi-Square test, a critical value x_o^2 of $x^2$for a significance level $\propto$, so that$x^2 < x_o^2$, Null hypothesis of good fit is accepted. For $x^2 \geq x_o^2$, Null hypothesis of good fit is rejected. The value of x_o^2 is usually obtained for a given number of degree of freedom (NDF) at a particular level of significance $\propto$ usually taken as 5% from standard statistics.

III. MATERIALS AND METHODS

Study Area

River Opeki,is also known as River Awpekki at Abidogun village located in the Southwestern Nigeria in Oyo State. The village is one of the fast growing rural areas in Nigeria. It is located at an elevation of 71 meters above sea level. The Rivers linking the Ogun River are the Ofiki and Opeki Rivers; two seasons are distinguishable in the region basin; a dry season from November to March and a wet season between April and October. Mean annual rainfall ranges from 900 mm in the north to 2000 mm towards the south. The total annual potential evapotranspiration is estimated to range from1600 and 1900 mm. [9] .A map showing River Opekiis shown below in Figure 1. [10]

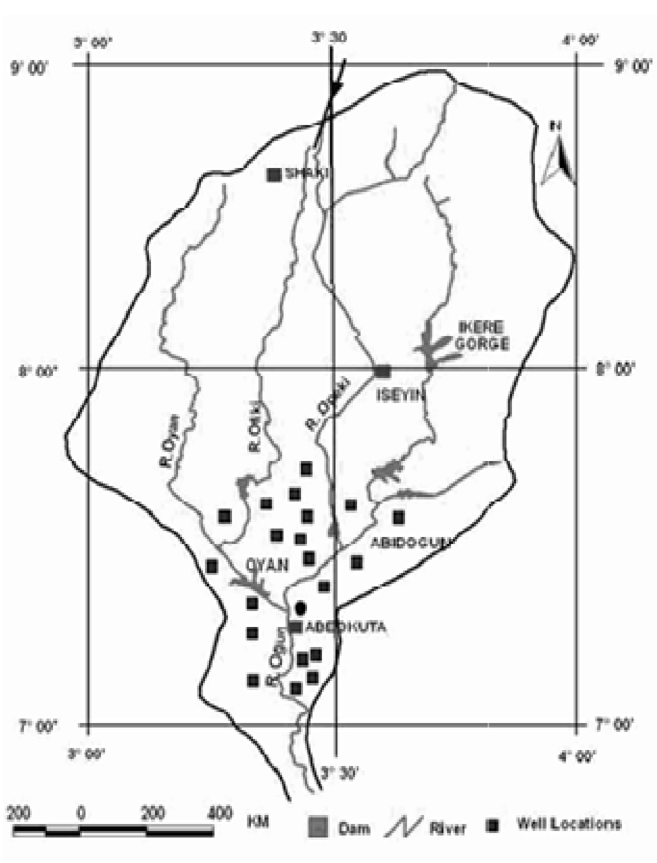

Figure 1: Map showing River Opeki

Data Collection/ Analysis

The discharge data for River Opeki for 30 years was obtained from Ogun-Osun River Basin Development Authority, and subjected to flow frequency analysis utilizing the models as described in section 2.0.
The mean, variance, standard deviation, and skew coefficients of the data were calculated.

The obtained data was used to estimate discharge using recurrence interval (T) of 2, 5, 10, 25, 50, 100, 200, 500, and 1000 years. Which were obtained using Equation (12) and the Weibull formula as

$$T = \frac{n+1}{m} \quad (21)$$

where, n = number of years of record and m = rank obtained by arranging the annual flow series in descending order of magnitude with the maximum being assigned the rank of 1 as seen in Table 1.

Test for goodness of fit was determined by using the Chi-Squared test.

The Z value for the different T (yrs) for normal distribution and its discharge were calculated using the formula's below.

$$f(z) = 1 - \frac{1}{T} \qquad (22)$$

$$Q = \bar{x} + z \ (23)$$

Where Z is the standard normal variate and corresponds to the frequency factor.

IV. RESULTS AND DISCUSSIONS

Tables 3, 4, 5 and 6 are the results for the log-normal, Gumbel extreme value type 1, log Pearson type III and Gamma or Pearson type III respectively. The results are in agreement with the previous hydrological study on the river [9].

Based on the result of the chi-square test using SPSS as presented in Table 7, the computed values for Normal, Log normal and Gumbel's Extreme type I were 139, 95 and 50 respectively, and were all greater than the tabulated chi-squared value of 42.56, an indication of no best fit probability distribution for the models. But Log-Pearson type III and Gamma or Pearson type III indicated low chi-squared values as 38.750 and 24.043 respectively. Therefore, Gamma or Pearson type III having the least chi-squared value is said to be the best fit probability distribution model in the estimation of flow discharge in River Opeki.

Table 1: Annual Flow Discharge Data from River Opeki

RANK	X or Q (m^2/s)		Log X
1	18478.99		4.2666
2	12527.98		4.0978
3	12343.67		4.0914
4	1128.61		4.0838
5	11696.65		4.0680
6	10961.25		4.0398
7	10190.19		4.0081
8	10126.45		4.0054
9	9594.27		3.9820
10	9649.36		3.9844
11	9444.95		3.9751
12	8597.1		3.9343
13	7833.26		3.8939
14	758.55		3.8897
15	7672.36		3.8849
16	7468.16		3.8732
17	7403.63		3.8694
18	6988.22		3.8443
19	6971.8		3.8433
20	6833.15		3.8346
21	6437.6		3.8087
22	6165.03		3.7899
23	5859.73		3.7678
24	5541.1		3.7435
25	5441.85		3.7336
26	5219.9		3.7176
27	4827.1		3.6836
28	4162.25		3.6193
29	3484.69		3.5421
30	2987.79		3.4753
Mean $\bar{x}$	8159.853	Mean log $\bar{x}$	3.8929
STDEV σ	3270.59	STDEV $\sigma log x$	0.2293
Skew C_s	1.0263	Skew C_s	1.2279

Table 2: Application of normal distribution to the obtained data

T (yrs)	Z	σ	$Z\sigma$	$\bar{x}$	$Q = \bar{x} + Z\sigma$ (m^3/s)
2	0.0	3513.355	0.0	8159.853	8159.853
5	0.85	3513.355	2986.352	8159.853	10939.85
10	1.30	3513.355	4567.362	8159.853	12411.62
25	1.80	3513.355	6324.039	8159.853	14046.92
50	2.10	3513.355	7378.046	8159.853	15028.09
100	2.33	3513.355	8186.117	8159.853	15780.33
200	2.60	3513.355	9134.723	8159.853	16663.39
500	2.87	3270.59	9386.593	8159.853	17546.45
1000	3.09	3270.59	10106.123	8159.853	18265.98

Table 3: Application of Log-normal distribution to the obtained data

T (yrs)	K_T	$\sigma log x$	$K_T \sigma log x$	$log \bar{x}$	$Y_T = log \bar{x} + K \sigma log x$	$Q = 10^{Y_T}$ (m^3/s)
2	0	0.2293	0	3.8929	3.8929	7814.48
5	0.842	0.2293	0.1930	3.8929	4.0859	12187.09
10	1.282	0.2293	0.2939	3.8929	4.1868	15374.46
25	1.751	0.2293	0.4015	3.8929	4.2944	19696.99
50	2.054	0.2293	0.4709	3.8929	4.3638	23110.00
100	2.326	0.2293	0.5333	3.8929	4.4262	26680.87
200	2.576	0.2293	0.5906	3.8929	4.4835	30443.87
500	2.633	0.2293	0.6037	3.8929	4.4966	31376.18
1000	2.982	0.2293	0.6838	3.8929	4.5769	37748.53

Table 4: Application of Extreme Value Type 1 Distribution to the obtained data

T (yrs)	Y_T	α	αY_T	μ	$Q_T = \mu + \alpha Y_T$ (m^2/s)
2	0.3665	2549.08	934.24	6688.52	7622.76
5	1.4999	2549.08	3823.37	6688.52	10511.89
10	2.2504	2549.08	5736.45	6688.52	12424.97
25	3.1985	2549.08	8153.23	6688.52	14841.75
50	3.9019	2549.08	9946.26	6688.52	16634.78
100	4.6001	2549.08	11726.02	6688.52	18414.54
200	5.2958	2549.08	13499.42	6688.52	20187.94
500	6.2136	2549.08	15838.96	6688.52	22527.48
1000	6.9073	2549.08	17607.26	6688.52	24295.78

Table 5: Application of Log-Pearson Type III distribution to the obtained data

T (yrs.)	K_T	$\sigma log x$	$K_T \sigma log x$	$log \bar{x}$	$Y_T = log \bar{x} + K \sigma log x$	$Q = 10^{Y_T}$ (m^2/s)
2	-0.195	0.2293	-0.0447	3.8929	3.8482	7050.18
5	0.732	0.2293	0.1678	0.2293	4.0607	11500.06
10	1.340	0.2293	0.3073	0.2293	4.2002	15856.23
25	2.087	0.2293	0.4785	0.2293	4.3714	23517.98
50	2.626	0.2293	0.6021	0.2293	4.495	31260.79

100	3.149	0.2293	0.7221	0.2293	4.615	41209.75
200	3.661	0.2293	0.8395	0.2293	4.7324	54000.78
500	3.519	0.2293	0.8069	0.2293	4.6998	50095.65
1000	4.219	0.2293	0.9674	0.2293	4.8603	72493.66

Table 6: Application of Gamma distribution to the obtained data

T (yrs)	Z	σ	Kσ	$\bar{x}$	$Q_T = \bar{x} + K\sigma$ (m^2/s)
2	-0.164	3270.59	-536.38	8159.853	7623.47
5	0.758	3270.59	2479.11	8159.853	10638.97
10	1.340	3270.59	4382.59	8159.853	12542.44
25	2.043	3270.59	6681.82	8159.853	14841.67
50	2.542	3270.59	8313.84	8159.853	16473.69
100	3.022	3270.59	9883.72	8159.853	18043.57
200	3.489	3270.59	11411.09	8159.853	19570.94
500	3.406	3270.59	11139.63	8159.853	19299.48
1000	4.049	3270.59	13242.62	8159.853	21402.47

Table 7: Chi-Square test of Fit for various Distribution

Distribution	Chi-Square χ^2 for 5% Level NDF=29, χ^2_0= 42.56
Normal	139.000
Log-Normal	95.000
Gumbel' Extreme Value Type I	70.000
Log-Pearson Type III	38.759
Gamma or Pearson Type III	23.043

V. CONCLUSION AND RECOMMENDATIONS

From the results of the study conducted, the following conclusions and recommendations are drawn:

I. The estimated discharge values were within the same range for all the five distributions at lower return periods. But for higher return periods, the results gave higher magnitudes.

II. The log-pearson Type III probability distribution recorded the highest estimated flow discharge value of 72,493.66 m^3/s at 1000 years return periods.

II. Log-pearson Type III and Gamma or Pearson type III models are appropriate for frequency analysis of Rive Opeki, based on their fitness tests as indicated by their chi-squared values.

However, Gamma or Pearson Type III model is recommended as the best probability distribution for the River based on its goodness of fit.

Furthermore, probability distribution for extreme rainfall for the River is suggested.

REFERENCES

[1]. Nelson, S. A. Natural Disasters. Tulane University Department of *Earth and Environmental Science.* 204pp, 2016.

[2]. Jacob, O. E. and Osadolor, C. I. "Flood Frequency Analysis at Oshun River in Asejire Dam Site", *Nigeria Journal of Earth Science and Engineering*, vol. 3, pp. 292-300, 2013.

[3]. IPCC."Climate Change Synthesis Report", An Assessment of Intergovernmental panel on Climate Change, *IPCC Secretariat World Meteorological Organization, GenevaSwitzerland*, 104pp, 2007.

[4]. NIST/SEMATECH. "E-Handbook of *Statistical* Methods" *www.itl.nist.gov/div 898/ handbook.*Accessed on July 18, 2006.

[5]. umbel, E. I.*Statistic of Extremes Book* 2nd edition, New York: Columbia University Press, 357pp, 1958.

[6]. nalysis techniques."Flood frequency analysisCivil Construction and Environmental Engineering DepartmentOregon State UniversityCorvallis"*www.streamflow.engr.oregonstate.edu/analysis/floodfreq/*. Accessed 2005

[7]. Odu, E. N and Ihejiamaiza, E. C. *Stastical and Basic Research Methods in Education andSocial Science* 2nd Edition, Calabar: University of Calabar press, 235pp, 2001.

[8]. Yadav, R. and Lal, B.B. "Pande Best fitted Distribution for estimation of future flood For Rapti river system in eastern Uttar Pradesh", *Indian Journal of Engineering and Material Science*, vol. 5, 22-27pp, 1998.

[9]. ORBDA. Osun River Basin Development Authority, *Hydrological Year Book*, 2011.
[10]. Ogun-Osun River Basin Development Authority. "Oshun River Basin master plan for development of Water Resources", Draft Final Report, *Tahal consultants Nigeria Ltd in association with associated Engineers and Consultants Nigeria Ltd*, Ibadan, May 1980.