Establishment of a Stochastic Model for Sustainable Economic Flood Management in Yewa Sub-Basin, Southwest Nigeria

Agbede O.A [a], Aiyelokun O.O [a*]

[a] Department of Civil Engineering, University of Ibadan, Ibadan, Nigeria.

Abstract

Of all natural disasters, floods have been considered to have the greatest potential damage. The magnitude of economic damages and number of people affected by flooding have recently increased globally due to climate change. This study was based on the establishment of a stochastic model for reducing economic floods risk in Yewa sub-basin, by fitting maximum annual instantaneous discharge into four probability distributions. Daily discharge of River Yewa gauged at Ijaka-Oke was used to establish a rating curve for the sub-basin, while return periods of instantaneous peak floods were computed using the Hazen plotting position. Flood magnitudes were found to increase with return periods based on Hazen plotting position. In order to ascertain the most suitable probability distribution for predicting design floods, the performance evaluation of the models using root mean square error was employed. In addition, the four probability models were subjected to goodness of fit test besed on Anderson-Darling (A^2) and Kolmogorov-Smirnov (KS). As a result of the diagnostics test the Weibul probability distribution was confirmed to fit well with the empirical data of the study area. The stochastic model $Q = 1.497 ln(Tr) - 0.151$ generated from the Weibul probability distribution, could be used to enhance sustainable development by reducing economic flood damages in the sub-basin.

Keywords: Sustainable; Development; Economic Flood; Stochastic Model.

1. Introduction

Flood has caused tremendous losses to properties and sometime life. There is a continuous interest in determining the most appropriate data distribution for flood frequency analysis, since this information is crucial for hydraulic analysis and designing hydraulic structure [1]. The problems of hydrological extremes such as floods damage and risk could be avoided, if adequate and precise flood forecasting mechanisms are put in place. Engineering designs for flood management involves the construction of minor and major hydraulic structures such as barrages, bridges, culverts and dams, spillways, road/railway bridges, urban drainage systems, flood plain zoning and flood protection projects. These constructions are designed and mechanically fit for managing and utilizing water resources to the best advantage using the records of past events [2].

It is possible to estimate the frequency of a given magnitude event by using an empirical distribution function, whereas in situations where too few data are available, the empirical distribution produced would not be suitable, since it would be required to estimate the frequency of occurrence of events larger than the maximum records. It has established that an alternative is to fit the empirical data to a theoretical frequency distribution been.

In projects involving hydraulic and hydrologic designs, several types of theoretical probability distributions have been applied to stream records. Some of the probability distributions commonly used are Normal Distribution, Lognormal Distribution, Exponential Distribution, Gamma Distribution, Pearson Type III Distribution, Log-Pearson Type III Distribution and Extreme Value Distribution which is further subdivided into three form that include EVI (Gumbel Distribution), EVII (Frechet Distribution) and EVIII. The most popular theoretical probability distributions have been the lognormal, log Pearson Type III and Gumbel distributions. In the United States and Australia the log

---

* Corresponding author: aiyelokuntobi@gmail.com

Pearson Type III (LPIII) distribution has been selected as a standard by federal agencies whereas, General Extreme Value (GEV) distribution is the standard method for flood-frequency analysis in the U.K.

For economic consideration, engineering structures are designed to drain flood of high magnitudes of up to 500 years return period. Sustainable flood control ensures that drainage channels are designed to dispose both excess rainfall and runoff at maximum rate and velocity. Uncertainty is always present when planning, developing, managing and operating water resources systems. It arises because many factors that affect the performance of water resources systems are not and cannot be known with certainty when a system is planned, designed, built, managed and operated [3]. The success and performance of each component of a system often depends on future meteorological, demographic, economic, social, technical, and political conditions, all of which may influence future benefits, costs, environmental impacts, and social acceptability [3]. Floods events cannot be described with certainty due to their stochastic nature; hence, they cannot be properly understood using empirical data. A classical way of describing the frequency and magnitude of floods is fitting annual peak instantaneous discharge of streams or annual maximum daily rainfall of an area to probability distributions. A probability distribution is merely useful if it does not fit the data of interest accurately, which inform the need to probability distribution adequacy assessment. Therefore, within the framework of this research, emphases were placed on describing the best probability distribution for flood risk management to establish a flood prediction model for the Yewa sub-basin, which would serve as a standard for flood risk reduction within the study area. This was achieved by analyzing the frequency of annual peak flood, estimating the design flood of various return periods and employing Goodness of Fit test to select most suitable distribution model for sustainable flood management.

2. Literature Review

Numerous works have been done using Probability distributions for flood risk management and planning. Ahmed et al. [1] compared five probability models for Johor River Basin by estimating the average recurrent interval (ARI) of flood event based on the distributions of annual peak flow. The study employed distribution models, namely Generalized Extreme Value (GEV), Lognormal, Pearson 5, Weibull and Gamma were tested. The goodness fit test (GoF) of Kolmogorov-Smirnov (K-S) was used to evaluate and estimate the best-fitted distribution. The results reaffirm the current practice that GEV is still the best-fitted distribution model for fitting the annual peak flow data. On the other hand, gamma distribution showed the poorest result.

Vivekanandan [4] compared eight probability distributions used for estimation of PFD for Malakkara and Neeleswaram. Maximum likelihood method was used for determination of parameters of the probability distributions. Goodness-of-Fit tests such as Anderson-Darling and Kolmogorov-Smirnov were applied for checking the adequacy of fitting of the distributions to the recorded annual maximum discharge. A diagnostic test of D-index was used for the selection of a most suitable distribution for FFA. Based on GoF and diagnostic test results, the study showed the EV1 distribution was better suited for estimation of PFD for Malakkara whereas LP3 for Neeleswaram.

Kochanek et al. [5] performed a data-based comparison of flood frequency analysis methods used in France. Results from this comparative exercise suggest that two implementations dominate their competitors in terms of predictive performances, namely the local version of the continuous simulation approach and the local-regional estimation of a GEV distribution. More specific conclusions include the following: (i) the Gumbel distribution is not suitable for Mediterranean catchments, since this distribution demonstrably leads to an underestimation of flood quantiles; (ii) the local estimation of a GEV distribution is not recommended, because the difficulty in estimating the shape parameter results in frequent predictive failures; (iii) all the purely regional.

Vivekanandan [6] dealt with the fitting of Extreme Value Type-1, Gamma, 2-parameter Lognormal (LN2) and Log Pearson Type-3 (LP3) distributions to the annual maximum data; and examined the use of goodness-of-fit tests and diagnostic analysis in assessing the adequacy of suitable probability distribution for estimation of design flood. Results of the study showed that LN2 distribution was better suited for modelling flood data for Tapi at Burhanpur, Girna at Dapuri and Bori at Malkheda sites; and LP3 was the best for Purna at Lakhpuri.

3. Materials and Method

3.1. Description of Study Area

Yewa River is a trans-boundary river between Republic of Benin and Nigeria. It lies approximately within latitudes 6^0 22′ and 6^0 36′ N and longitudes 2^0 50′ and 2^0 54′ E of the Greenwich Meridian. The basin has a total catchment area of approximately 5000 km^2 and it lies west of the Ogun, Ona and Oshun basins. The Yewa river discharges into the Lagos lagoon. The Yewa River basin is located within the West African tropical climate, which is under the influence of the tropical continental air mass and the tropical maritime air mass. However, the basin is classified as belonging to the equatorial hot, wet climate, with distinct dry and wet seasons. The mean annual rainfall varies between 800 and 1150 mm, in the north to 1500 mm in the south, while the annual mean temperature is about 28ºC with a range of ±4ºC [7]. The scope of this study is limited to the upper part of the Yewa drainage area gauged at Ija-Oke (Figure 1).
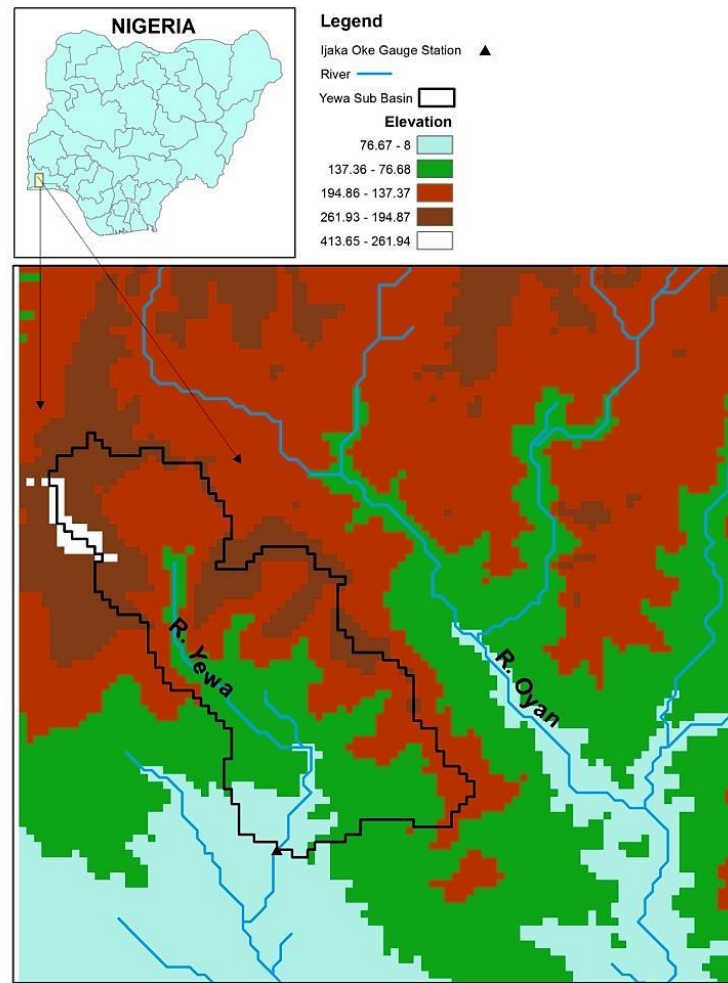
Figure 1. Map of Study Area

3.2. Materials

A time series of 13 years (1988-2000) hydrologic record of river Yawa gauged at Ijaka-Oke was obtained from the Ogun-Oshun River Basin Development Authority (OORBDA), Abeokuta Nigeria to generate an annual hydrograph for the study area. Annual peak floods were selected and arranged in descending order of magnitude to form an annual maximum series and the probabilities that ranked annual maximum will be equalled or exceeded in any year were determined by the Hazen's plotting position.

The Hazen plotting position was used in the study and is represented by the following equation:

$$Tr = \frac{2n}{2m-1} \tag{1}$$

Where m is the order or rank while n is number of years of study.

3.3. Methods

3.3.1. Normal Distribution (NOR)

For a symmetrically distributed data, the most appropriate distribution of continuous variable is the normal distribution which is also called the Gaussian distribution [8]. The probability density function (PDF) of this distribution model according to Vivekanandan [4] is given by:

$$f(x, \mu x, \sigma x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}, x, sx > 0 \tag{2}$$

3.3.2. Gamma (GAM)

The gamma probability distribution describes the number of events in Poisson process; it assumes the sum of independent and identical exponentially distributed random variables. The probability density function (PDF) under this distribution is given as:

$$f(x, \alpha, \beta) = \frac{1}{\alpha^\beta \Gamma \beta} x^{\beta-1} e^{-\left(\frac{x}{\alpha}\right)}, x, \alpha, \beta > 0 \tag{3}$$

3.3.3. Extreme Value Type-1 (EV1)/ Gumbel

The Gumbel distribution also referred to as the extreme value type I distribution [9] has two forms, one is based on the smallest extreme (minimum case), and the other is based on the largest extreme (maximum case). In this study, the maximum case is used. The probability density function (PDF) under this distribution is given as:

$$f(x; \alpha, \beta) = \frac{e^{-(x-\beta)/\alpha} e^{-e^{-(x-\beta)/\alpha}}}{\alpha}, -\infty < x < \infty, \alpha > 0 \tag{4}$$

3.3.4. Frechet (EV2)/ Weibull

The Weibull distribution, also known as extreme value type III distribution, is still a two-parameter distribution with parameters α and β. The probability density function (PDF) under this distribution is given as:

$$f(x; \alpha, \beta) = \frac{\beta}{\alpha} \left(\frac{\alpha}{x}\right)^{\beta+1} e^{-(\alpha/x)^\beta}, \quad -\infty < x < \infty, \quad \alpha > 0 \tag{5}$$

The Weibull distribution is a versatile distribution that can take on the characteristics of other types of distributions, based on the value of the shape parameter, β.

3.4. Performance Evaluation

This study employed the use of three statistical procedures for evaluating the performance of the distributions. They include coefficient of determination, root mean square error (RMSE) and correlation coefficient. The RMSE is expressed by the equation:

$$RSME = \left[n^{-1} \sum_{i-1}^{n} (P - O)^2 \right]^{0.5} \tag{6}$$

Where:
RMSE is root mean square error (m^3/s),
P is predicted discharges under each distribution (m^3/s),
Q is observed discharges (m^3/s), and n is as previously defined.

3.5. Goodness of Fit (Gof) Test

In order to check for the adequacy of fitting of the probability distributions to the recorded annual peak data, two goodness of fit test was applied for the study. GoF tests include Anderson-Darling (A^2) and Kolmogorov-Smirnov (KS).

3.6. Anderson-Darling Test

The Anderson-Darling test compares an observed CDF to an expected CDF. This method gives more weight to the tail of the distribution than KS test, which in turn leads to the AD test being stronger, and having more weight than the KS test. The test rejects the hypothesis regarding the distribution level if the statistic obtained is greater than a critical value at a given significance level (α) [10]. The significance level most commonly used is $\alpha=0.05$, producing a critical value of 2.5018. This number is then compared with the test distributions statistic to determine if it can be rejected or not. The AD test statistic (A^2) is:

$$A^2 = n - \frac{1}{n} \sum_{i=1}^{n} (2i - 1) . \left[lnF(x_i) + \ln\left(1 - F(x_{n-i+1})\right) \right] \tag{7}$$

3.7. Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov test statistic is based on the greatest vertical distance from the empirical and theoretical CDFs. Similar to the AD test statistic, a hypothesis is rejected if the test statistic is greater than the critical value at a chosen significance level. For the significance level of $\alpha=0.05$, the critical value calculated is 0.12555 [10]. The samples were assumed to be from a CDF F(x). The test statistic (D) is:

$$D = \max\left(F(x_i) - \frac{i-1}{n}, \frac{1}{n} - F(x_i) \right) \tag{8}$$

3.8. Diagnostic Test

The selection of a most suitable probability distribution for estimation of PFD was performed through D-index, which is defined as:

$$D - index = \left(\frac{1}{\mho}\right) \sum_{i=1}^{6} |x_i - x^*_i| \tag{9}$$

Here, $\mho$ is the average value of the recorded annual peak data, x_i is the ith sample of the first six highest values in the series of annual peak data and x_i^* is the corresponding estimated value by probability distribution. The distribution having the least D-index is considered as the better suited distribution for estimation of PFD [4].

4. Results

4.1. Flood Frequency Analysis

The return periods estimated from the Hazen plotting position of each instantaneous peak flows of the ranked years between 1988 and 2000 are presented in Table 1. The highest flood magnitude of 5.07 m^3/s for the years of study was estimated to have a return period of 26 years, with a low probability of been equaled or exceeded of 0.04. Moderate flood magnitude of 3.46 m^3/s to 1.19 m^3/s was found to have a return period of 8.67 to 2.89 years, while floods with low magnitude of 0.62 m^3/s to 0.37 m^3/s is expected to occur every 0.42 to 0.88 years. While, least flood of 0.02 m^3/s has a return period of 0.96 year, and a high probability of occurrence. It can also be seen from Table 1. that as the magnitude of floods increases the returning periods increases, while the probability of exceedance decreases. This goes to show that floods with great magnitude were not frequently experienced in the sub basin; however, such floods would have high risk when they occur.

Table 1. Frequency analysis of river Yewa flood data

Rank	Water year	Ranked max streamflow, q, (m^3/s)	Return period [(2n/(2m-1)]	Exceedence probability (1/Tr)
1	2000	5.07	26.00	0.04
2	1999	3.46	8.67	0.12
3	1991	2.49	5.20	0.19
4	1989	1.93	3.71	0.27
5	1988	1.19	2.89	0.35
6	1997	0.62	2.36	0.42
7	1995	0.59	2.00	0.50
8	1994	0.58	1.73	0.58
9	1992	0.55	1.53	0.65
10	1998	0.47	1.37	0.73
11	1993	0.4	1.24	0.81
12	1990	0.37	1.13	0.88
13	1996	0.02	1.04	0.96

4.2. Estimation of Design Flood Using Probability Distributions

Table 2. shows the estimated design floods that were computed for different return periods. Based on maximum likelihood method of parameter estimation, The table shows that Normal distribution predicted the highest magnitude of floods for return period of 2-year and 5-year, while it predicted lowest value for return period of 100-years and above out of the four distribution model.

Table 2. Estimated design flood

Return Period (yr)	Gumbel	Gamma	Normal	Weibul
2	1.09	0.92	1.37	0.9
5	2.06	2.21	2.57	2.2
10	2.7	3.2	3.2	3.21
20	3.52	4.52	3.87	4.58
50	4.12	5.52	4.31	5.63
100	4.72	6.53	4.69	6.69
200	5.32	7.53	5.06	7.76
500	6.1	8.87	5.49	9.18

4.3. Analysis Based on Goodness of Test (GoF) and Diagnostic Test

As shown in Table 3, the Weibul distribution can be considered to be the most suitable probability distribution for estimating design floods in Yewa sub basin, since it has the lowest Anderson-Darling, Kolmogorov-Smirnov (KS) and Root Mean Square Error. The suitability of the Weibul Distribution was also confirmed since it has highest correlation coefficient of 0.999 and lowest D-index of 0.2.

Table 3. Summary of GoF and diagnostic test

Probability Distributions	As	Ks	RMSE	Correlation Coefficient	D-Index	RANK
Weibul	0.52	0.23	0.31	0.999	0.2	1
Gamma	0.53	0.24	0.45	0.999	0.25	2
Normal	1.28	0.31	2.13	0.979	0.95	3
Gumbel	1.05	0.31	2	0.998	1.15	4

The adequacy of the four distributions could be further investigated in the GoF plots depicted in Figure 2. to Figure 5. From the density plot, which represents the density function of the fitted distribution along with the histogram of the empirical distribution (Figure 2), it could be observed that Gamma and Weibull distribution fitted well to the empirical distribution of the annual peak flood of river Yewa at Ijaka-Oke gauging station. Figure 3 depicts the Q-Q plot, which represents the empirical quantiles (y-axis) against the theoretical quantiles (x-axis) and emphasizes the lack of fit along the distribution tail; it could be observed that the Weibull and the Gamma fitted well along the tail. Furthermore, as depicted in Figure 3; the Cumulative Distribution Function (CDF) of Weibull and Gamma fitted well with the CDF of the empirical distribution. While Figure 4, which emphasizes the lack of fit at the distribution center, shows that, none of the distribution fits well at the center, however, Weibull and gamma distributions are preferred for their better description of the right tail of the empirical distribution. Based the quantitative assessment of the distribution the Weibull distribution was considered to be the most adequate distribution for modeling design flood for sustainable flood management in the upper part of the Yewa river basin.
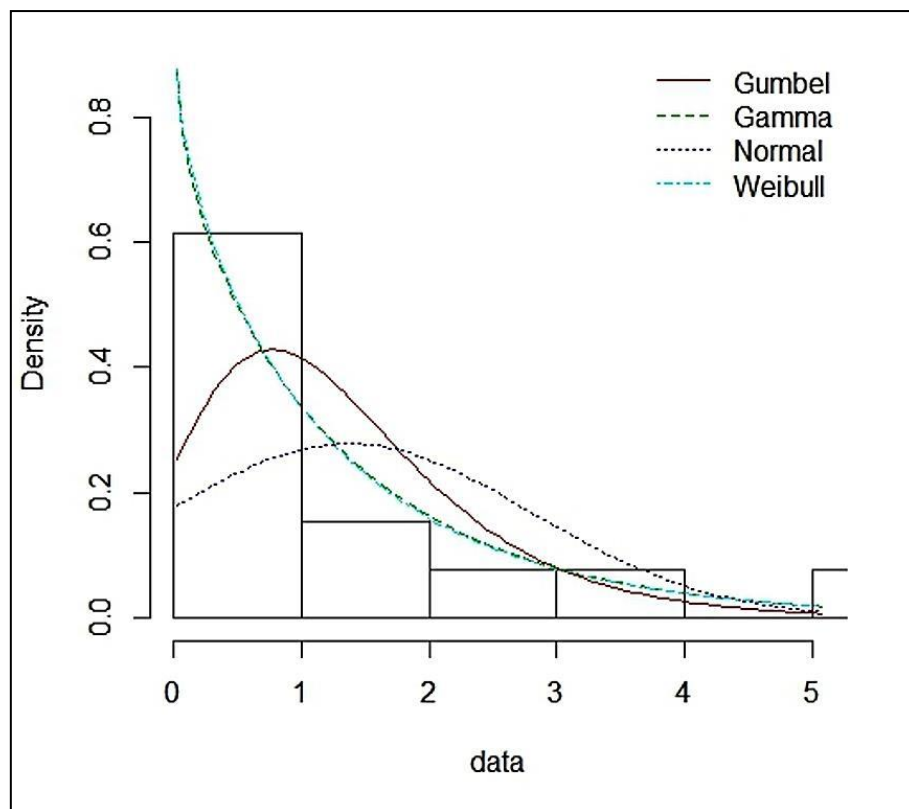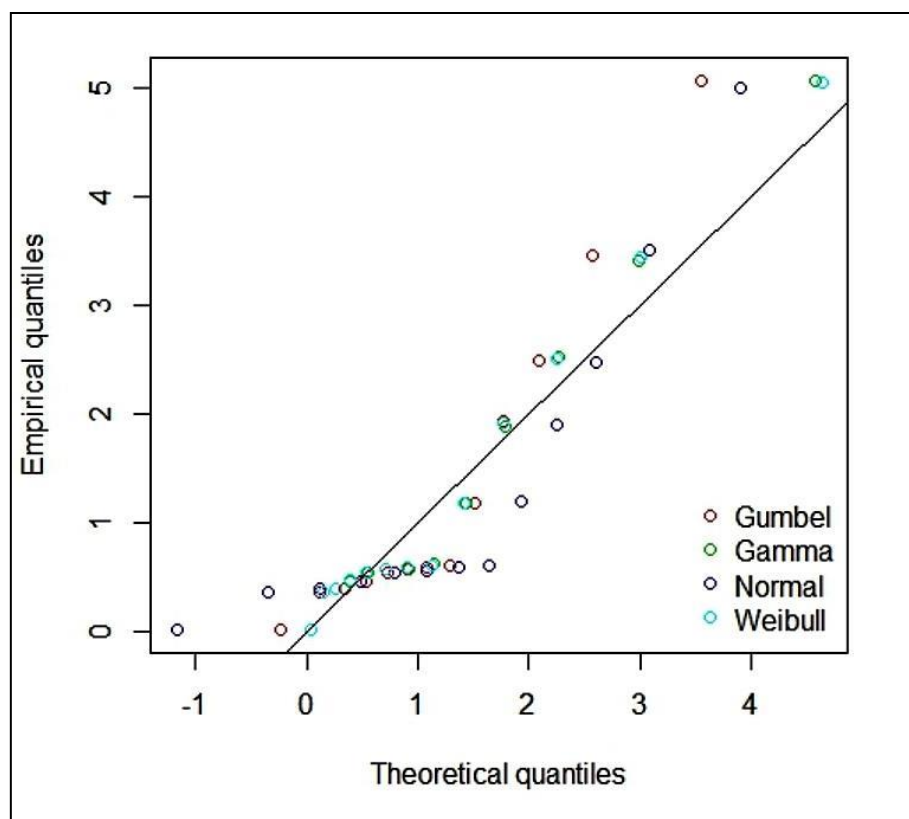
Figure 2. Histogram and Theoretical Densities Plot

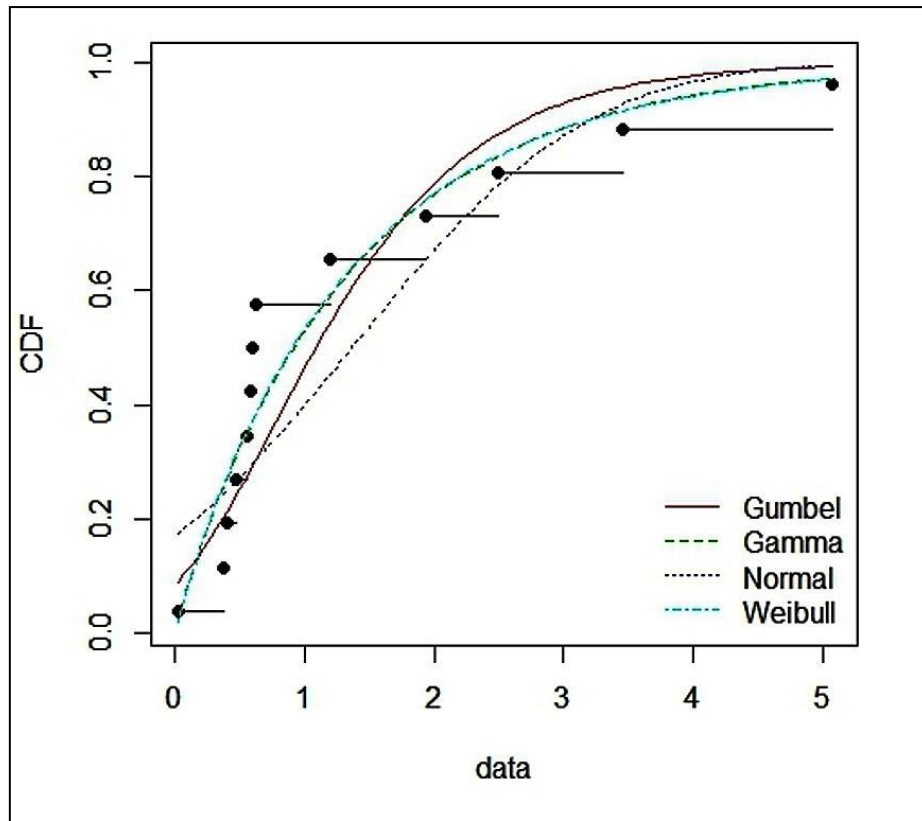**Figure 3. Q-Q Plot**

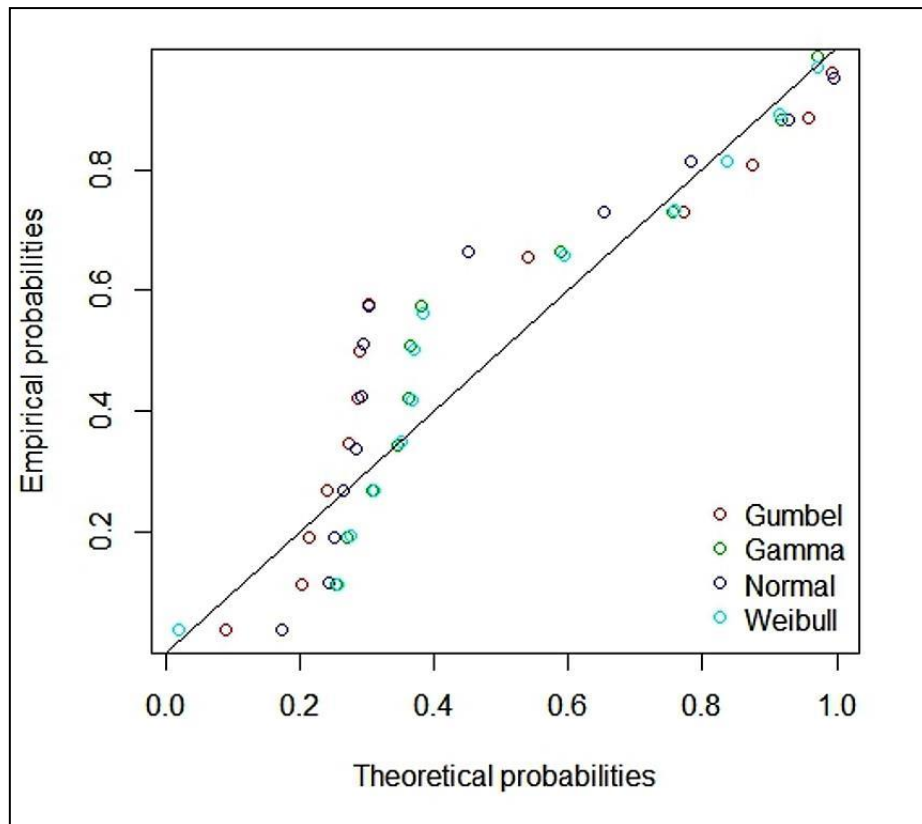**Figure 4. Empirical and Theoretical CDFs Plot**

Figure 5. Empirical and Theoretical Probabilities Plot

Floods of different magnitude can be computed for different return periods for Yewa sub basin based on Figure 6. The prediction model can represented by $Q = 1.497 \ln(Tr) - 0.151$, where Q is discharge in m^3/s, and Tr is the return periods in Years; a coefficient of determination (R^2) of 0.998 shows that there is a strong indication that the flood prediction model would give accurate flood for any return periods.
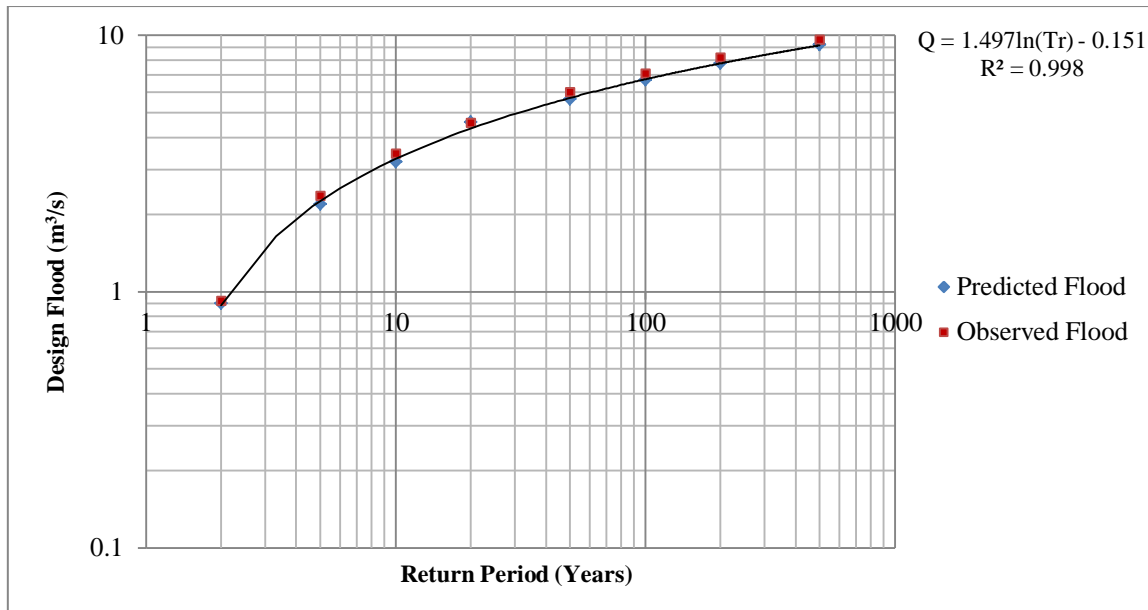
Figure 6. Logarithmic Probability Plot of Annual Peak Flood of River Yewa at Ijaka-Oke

4.4. Discussion of Findings

Flood frequency analysis for determining efficient designs of hydraulic structures is one method of decreasing flood damages and economic losses. The study was concerned with the establishment of a prediction model for economic flood management in Yewa sub-basin. The flood frequency analysis shows that floods of high magnitudes were not common in Yewa sub-basin, but are likely to have high risks when they occur. The estimated flows of the selected return periods of 2, 5, 10, 25, 50, 100, 200 and 500 years fitted into the four distribution shows an increase with discharge, which corroborates many other studies such as Fasinmirin and Olufayo [11], Adeboye and Alatise (2007).

In order to ascertain the suitability of distribution models for economic flood management, the predicted floods based on each model were firstly subjected to quantitative goodness of fit test, which revealed that Weibull distribution had the lowest Anderson-Darling, Kolmogorov-Smirnov (KS) and Root Mean Square Error values of 0.52, 0.23 and 0.31 respectively (Table 3). This implies that the cumulative distribution function of the recorded annual floods and the Weibull distribution were similar, in addition, a low estimate of RMSE shows that Weibull distribution performed more accurately in predicting design floods that other models. Qualitative test disqualifies the suitability of Normal and Gumbel distributions. While the diagnostic test confirmed that Weibull distribution was the most suitable model for economic flood management. As a result both quantitative and qualitative aspect of the Goodness of Fit test indicated that the Weibull distribution fits well to the empirical data of the floods.

The Weibull distribution was further used for the generation of the generation of the prediction model $Q = 1.497 \ln(Tr) - 0.151$ for the sub-basin. This model could assist hydrologist and engineers in sustainable planning for flood regulation and protection measures.

5. Conclusion

Sustainable economic flood management could be ensured when institutional and physical infrastructures are properly design to convey flood of high magnitudes. This study presents a stochastic model for flood planning of up to 500 years return periods. Major facts that emerged from the study are that high magnitudes of flood are not to be expected frequently, however, their occurrence could be very devastating; Weibul distribution had high collinear relationship with the recorded flood data and that GoF test confirmed Weibul as most suitable model for reducing flood risks.

6. Acknowledgment

The authors deeply appreciate the staff members of Hermtech Research and Consultancy for their support, which made this research a success.

7. References

[1] Ismail, Ahmad Zuhdi, Zulkifli Yusop, and Zainab Yusof. "comparison of flood distribution models for johor river basin." Jurnal Teknologi 74, no. 11 (2015): 123–128.

[2] Agbede A.O and Abiona. Plotting Position Probability Fittings to Lagos Metropolitan Precipitation: Hydrological Tools for Hydraulic Structures Design in Flood Control. International Journal of Pure and Applied Sciences and Technology. 10 (2012): 37-43.

[3] UNESCO. Concepts in Probability, Statistics and Stochastic Modelling. Water Resources Systems Planning and Management Report. 2005.

[4] Vivekanandan, N. Comparison of Probability Distributions for Estimation of Peak Flood Discharge. Open Access Library Journal, 1 (2014): e498.

[5] Kochanek, K. B. Renard, P. Arnaud, Y. Aubert, M. Lang, T. Cipriani, and Sauquet, E. A data based comparison of flood frequency analysis methods used in France. Nat. Hazards Earth Syst. Sci., 14 (2014): 295–308.

[6] Vivekanandan, N. Assessing Adequacy of a Probability Distribution for Estimation of Design Flood. Bonfring International Journal of Industrial Engineering and Management Science. 2(2012): 22-27.

[7] Oyebande, L. and Adeaga, O. Flow simulation in an ungauged basin using a digital elevation model. Predictions in Ungauged Basins: PUB Kick-off (Proceedings of the PUB Kick-off meeting held in Brasilia, 20–22 November 2002). IAHS Publ. 309, 2007.

[8] Tilahun K. The Characterization of Rainfall in the arid and semi-arid regions of Ethiopia, Water SA, 32(2006), 429-439.

[9] Coles, S. An Introduction to Statistical Modeling of Extreme Values. New York: Springer Verlag. 2001.

[10] Millington, N., Samiran Das and S. P. Simonovic. The Comparison of GEV, Log-Pearson Type 3 and Gumbel Distributions in the Upper Thames River Watershed under Global Climate Models. Water Resources Research Report. 2011.

[11] Fasinmirin J.T. and Olufayo A.A. Comparison Of Flood Prediction Models For River Lokoja, Nigeria. Department of Agricultural Engineering, Federal University of Technology, Akure, Nigeria. 2006.

[12] Adeboye, O. B. and Alatise, M. O. Performance of Probability Distributions and Positions in Estimating the Flood of River Osun at Apoje Sub-basin Nigeria. Agricultural Engineering International: the CIGR Ejournal. 9(2007).