Performance of Probability Distributions and Plotting Positions in Estimating the Flood of River Osun at Apoje Sub-basin, Nigeria

Omotayo B. Adeboye[*] and Michael O. Alatise

Department of Agricultural Engineering, Federal University of Technology, P.M.B 704, Akure, Ondo State, Nigeria; +234(0)8025306602; Email:[*] adeboyeomotayo@yahoo.com

ABSTRACT

The occurrence of flood and its attendant damage to both life and property in recent times especially in rural and urban dwellings and also in agricultural establishments have generated a lot of concern in Nigeria. The search for solution to problems on flood damage and risk which can probably be avoided if adequate and precise flood forecasting mechanisms are put in place has provoked this study. For the purpose of statistical analysis, 18-year streamflow record of River Osun at Apoje gauging station was obtained from Ogun-Osun River Basin and Rural Development Authourity Abeokuta, Nigeria. The annual hydrograph was drawn for the station by plotting the peak discharges against their hydrologic years. The peak discharges were fitted to the three major statistical distributions namely normal, log-normal and log-Pearson type (iii) while seven plotting positions of Hazen, Weibull, Blom, Cunnane, California, Gringorton and Chegodajev were used in determining their probabilities of exceedance. Results obtained showed that the annual maximum discharges at Apoje station varied from 169m^3/s to 400m^3/s within duration of 18 years. Weibull's plotting position combined with normal distribution gave the highest fit, most reliable and accurate predictions of the flood in the study area having the coefficient of determination R^2 and root mean square error RMSE of 0.9950 and 35.09 m^3/s respectively.

Keywords: Apoje, flood, plotting positions, statistical distributions, Ogun-Osun river basin annual maximum discharges.

1. INTRODUCTION

Water is the most vital requirement for human survival and it has been described as the elixir of life and a major agent that ensures continuity in an ecosystem (Tadesse, 2006). Scarcity of water and the presence of it in abundance especially if not controlled have their consequences in every environment. Hydrologist finds it difficult to make accurate prediction of flood estimates using limited historic information of runoff, rainfall, river stages. These can be attributed to lack of trained personnel and equipment for adequate assessment of these quantities on systematic basis. Where these data are available, they are characterised by gaps, which in most cases limit the inferences from their statistical analysis and eventually lower their usefulness for planning purposes (Viessman and Lewis, 1996). Using mathematical equations and statistical models, these gaps can be filled or the series extended to a longer

period of time (Claps and Laio, 2003). In stochastic hydrology, the common and result oriented probability distributions in use are the normal, lognormal and the log Pearson type (iii), gamma, Weibull' and Gumbell (Hromadka and Whitley, 1989; Moughamian, et al, 1987; Robert, 1987). The normal and the lognormal distributions fit adequately to the peak rainfall and streamflow while for the extreme hydrologic variables, the Weibull and Gumbell distributions are used (Aksoy, 2000). The uses of direct statistical and regional techniques in flood frequency estimation have long been advocated for, but have witnessed little attention. The regional methods allow the design event to be estimated at ungauged site and the reliability of the estimated events increases due to the increase in the incorporation of additional information (Hoskins and Wallis, 1997). However, the reliability of the predicted events for the design purposes at ungauged sites is significantly lower when compared with the values at a gauged site (Kjeldsen *et al*, 2001). Hebson and Cunnane (1987) showed that the values of the mean annual floods (MAF) at ungauged sites are less precise than the estimates at gauged site even with only one year of reliable data. For the purpose of efficiency, most hydrologists prefer the Maximum Likelihood, and Bayesian methods (Hirsch and Stedinger, 1987). Most engineering decisions are made with the use of graphic display, but for the purpose of extrapolation of the curves, the probability curves are used.

The subject of plotting positions is not new to hydrologists and statisticians. Plotting positions have been used in estimating magnitudes of hydrological events and their corresponding return periods, detecting outliers, fitting distributions to the data and in evaluating the adequacy of fit of the alternative parametric floods frequency models. A variety of plotting position formulae have been proposed in hydrological and statistical literature during the past 50 years. The optimum formulae are selected based on the purpose of investigation and the distribution under consideration (Van-Thanh-Van Nguyen, et al, 1989).

At present, there is no single universally accepted model, rather, a whole group of model such as the Gumbel, the Log-Logistic, the lognormal, and the log-Pearson type (iii) distributions have been suggested for predicting the magnitudes of such extreme events (Van-Thanh-Van Nguyen, et al, 1989). Therefore the question of better fit among numerous models is always welcome (Topaloglue, 2002). The performances of models have been assessed using statistical tools such as root mean square error (RMSE) and coefficient of determination (Van Bladeren, 1993 and Ekanayake and Cruise, 1993). The objective of this study was to determine the flood characteristics of the study area by using three major statistical methods and seven (7) different plotting positions.

2. MATERIALS AND METHODS

2.1 The Study Area

The Osun River Basin is located in an area whose boundaries are approximately latitudes $8^0 20$'N and $6^0 30$'N and longitudes $5^0 10$'E and $3^0 25$'E (Fig.1). The area of the basin is approximately 16,700Km^2 (FRN, 1982). Generally, the Osun basin's climate is influenced by the movement of the inter-tropical convergence Zone (ITCZ), a quasi-stationary boundary

zone which separates the sub-tropical continental air mass over the Sahara and the equatorial maritime air mass over the Atlantic Ocean as reported by Adeboye, (2005). In the wet season, the mean rainfall ranges between 1,020 and 1,520 mm in the South of the basin, but in the North, it is less than 1,020 mm. In the North and South, the mean dry season rainfall varies from 127 to 178 mm and 178 to 254 mm respectively. Water resources in Osun basin include the surface water and groundwater. Surface water plays the prominent role in the basin but the limited stream flow records create problems in water resources assessment. Generation of long record is done by streamflow synthesis, of rainfall records. Apoje sub basin is selected for this study because it is an agrarian community which serve as the food basket for a major part of the south-western Nigeria.
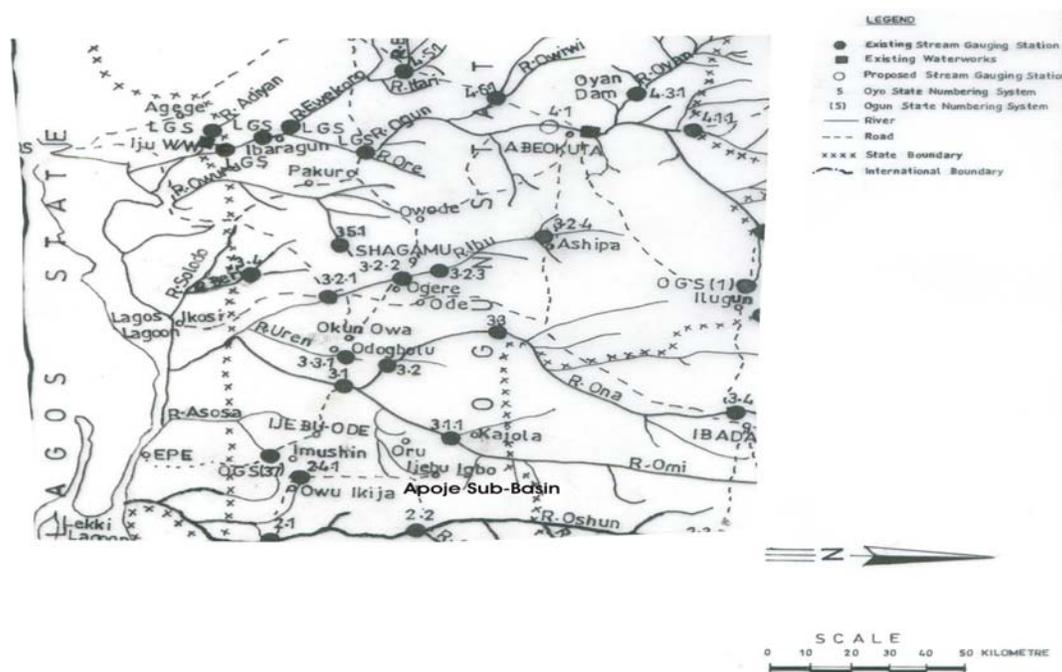

Figure 1: Hydrological networks of Ogun-Osun River Basins.

The Methods of Data Analysis

Daily steamflow data for the duration of 18-years (1982-1999) at Apoje gauging station were obtained from Ogun-Osun River Basin and Rural Development Authority, Abeokuta, Nigeria for the purpose of this study. The annual peak discharges were selected and ranked in descending order of magnitudes based on the recommendations of the (USWRC, 1981). The peak discharges were plotted against their hydrologic years (1982-1999) in order to determine the spatial variations of the annual maximum discharges in the sub-basin. The mean of the

annual flows were also plotted against the water years. The probabilities of exceedence of the discharges were determined using the seven plotting positions shown in Table 1.

Table 1: Plotting positions used in determining the flood probabilities of exceedence

Plotting Positions	Formulae
Hazen (1914),	$P(Q \geq Q_T) = \dfrac{m - 0.5}{n}$
Weibull (1939),	$P(Q \geq Q_T) = \dfrac{m}{n+1}$
Blom (1958),	$P(Q \geq Q_T) = \dfrac{m - 0.375}{n + 0.25}$
Cunnane (1978),	$P(Q \geq Q_T) = \dfrac{m - 0.4}{n + 0.2}$
California (1923),	$P(Q \geq Q_T) = \dfrac{m}{n}$
Gringorton(1963),	$P(Q \geq Q_T) = \dfrac{m - 0.44}{n + 0.12}$
Chegodajev(1955)	$P(Q \geq Q_T) = \dfrac{m - 0.3}{n + 0.4}$

From the Table 1
Q is anticipated streamflow, (m³/s),
Q_T is streamflow of estimated return period to be equalled or exceeded, (m³/s),
m is rank, and
n is number of observations.

The return periods of the anticipated discharges were determined by finding the reciprocals of the exceedence probabilities and is expressed by

$$T_r = \frac{1}{p} \tag{1}$$

where
T_r is return period;
p is probability of exceedence that is, the probability that a given flood is equalled or exceeded.

Three distribution models were considered for the purpose of accurately estimating the flood magnitudes and frequencies of River Osun at Apoje gauging station, Nigeria.

3. FREQUENCY ANALYSIS METHODS

3.1 Normal Distribution

For a symmetrically distributed data, the most appropriate distribution of continuous variable is the normal distribution which is also called the Gaussian distribution (Tilahun, 2006). The probability density function (PDF) of this distribution model according to (Chow *et al*, 1988) is given by

$$f(z) \quad = \frac{1}{\sqrt{2\pi}} \ell^{-z^2/2} \tag{2}$$

where

       z is standard normal variable, and
       e is exponential.

The statistical parameters, mean and standard deviation of the annual maximum discharges were determined using the method of moment and their respective equations are expressed as

$$\overline{Q} = \frac{1}{n} \sum_{i=1}^{n} Q_{max} \tag{3}$$

$$S_Q = \sqrt{\frac{\sum_{i=1}^{n} \left(Q_{max} - \overline{Q} \right)^2}{n-1}} \tag{4}$$

Where

$\overline{Q}$ is mean of the annual maximum discharges (m^3/s),

Q_{max} is annual maximum discharges (m^3/s),

S_Q is standard deviation of the annual maximum discharges (m^3/s), and

n is as previously defined.

In this distribution, the intermediate variable was determined using the expression

$$w = \sqrt{\ln\left(\frac{1}{p}\right)} \qquad 0 < p \le 0.5 \tag{5}$$

where
w is intermediate variable, and
p is as previously defined.
The frequency factors corresponding to the return periods of the ranked annual maximum discharges were determined using the expression

$$z = w - \frac{2.515517 + 0.802853w + 0.010328w^2}{1 + 1.432788w + 0.189269w^2 + 0.001388w^3} \qquad (6)$$

where,

p is probability of exceedence,

w is intermediate variable,

z is standard normal variable or the frequency factor.

The predicted flood at various return periods were determined using the mathematical expression

$$Q_T = \overline{Q} + z \cdot S_Q \qquad (7)$$

Q_T, Q, z, and S_Q are as previously defined.

3.2 Lognormal Distribution

Large numbers of hydrological continuous variable random variables tends to be asymmetrically distributed. It is advantageous to transform the distribution to a normal distribution by taking the logarithms of the annual maximum discharges (Tilahun, 2006; USWRC, 1981). The probability density function (PDF) under this distribution is given as

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(\frac{-y - \mu_y}{2\sigma_y^2}\right) \qquad (8)$$

In this flood analysis, the logarithms of the annual maximum discharges were taken to base 10. Using the method of moment, the mean and the standard deviation of the ranked annual maximum discharges were determined by the expression:

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} \log Q_{max} \qquad (9)$$

$$S_y = \sqrt{\frac{\sum_{i=1}^{n}(y - \overline{y})^2}{(n-1)}} \qquad (10)$$

respectively,

where

$\overline{y}$ is mean of y (m³/s),

S_y is standard deviation (m³/s),

$y = \log Q_{max}$ (m³/s), and

n is as previously defined,

The intermediate variables and frequency factors corresponding to the ranked annual maximum discharges were determined by Equations (5) and (6) respectively. The statistical variate and the predicted discharges under this distribution were determined respectively by the expressions

$$y_T = \bar{y} + z \cdot S_y \tag{11}$$

$$Q_T = 10^{(\bar{y} + z \cdot S_y)} \tag{12}$$

Where

y_T is variate of the annual maximum discharges at return periods T (years) ,

$\bar{y}$ is mean of the logarithm of the historic annual maximum discharges (m^3/s),

S_y, Q_T and z are as previously defined.

3.3 Log Pearson Type (iii) Distribution

This is referred to as the three parameter fit. Due to its performance in stochastic hydrology, it has been adopted in some countries as the standard distribution for flood frequency analysis (Sumioka et al, 1997). The probability model is given as

$$f(x) = \frac{\lambda^\beta (y - \varepsilon)^{\beta-1} \ell^{-\lambda 9 y - \varepsilon)}}{x\Gamma(\beta)} \tag{13}$$

where,

$$y = \log x, \quad \lambda = \frac{S_y}{\sqrt{\beta}}, \quad \beta = \left[\frac{2}{C_s(y)}\right]^2, \quad \varepsilon = \bar{y} = S_y\sqrt{\beta} .$$

In addition to the mean and the standard deviation of the historic annual maximum discharges determined using Equations (9) and (10), the third parameter, the skew coefficient, was determined using the expression.

$$C_s = \frac{n\sum_{i=1}^{n}(y - \bar{y})^3}{(n-1)(n-2)(n-3)} \tag{14}$$

Where

C_s is the coefficient of skewness; and n, y and $\bar{y}$ are as previously defined.

Using the Log Pearson type (iii) distribution, the frequency factors corresponding to the predicted annual maximum discharges was given by (Kite, 1977) as

$$K_T = z + (z^2 - 1)k + \frac{1}{3}(z^3 - 6z)k^2 - (z^2 - 1)k^3 + zk^4 + \frac{1}{3}k^5 \tag{15}$$

where

K_T is frequency factor,

k is expressed as $C_s/6$, and

z is as previously defined.

At each return period, the predicted discharges were determined by

$$Q_T = 10^{\left(\bar{y} + K_T \cdot S_y\right)} \tag{16}$$

In order to compare a model output to the observed data, criteria for making such a comparison must be identified as suggested by Green and Stephenson (1985). Visual comparison of the plotted predicted and observed discharges can be very useful in assessing the accuracy of the model output. However, additional statistical analysis is needed because visual comparison usually tends to be very subjective (Tewolde and Smithers, 2006). Three statistical procedures were used in these analyses for evaluating the performance of the distributions and these are coefficient of determination, absolute differences between the predicted and observed discharges and root mean square error (RMSE). The RMSE between the predicted and observed discharges were determined using the equation given by (O'Donnell, 1995) and is expressed as

$$RMSE = \left[n^{-1} \sum_{i=1}^{n} (P - O)^2 \right]^{0.5} \tag{17}$$

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.

4. RESULTS AND DISCUSSION

4.1 Hydrograph

The time series of the annual hydrograph is shown in Fig.2 The highest discharge of 400m^3/s was observed in 1987 and declined to 169m^3/s in the year 1999.The minimum peak flow in the Apoje sub-basin was 72m^3/s. The minimum and maximum Mean Annual Flows (MAF) were 82 and 189m^3/s respectively. The difference in magnitudes can be attributed to diversion of water to the control structures at Asejire and Ede which are along the course of River Osun. According to Kottegoda, (1980), emphasis in time series analyses should be laid on the mechanisms that generate the data but not necessarily on the future sequence of events over a period of time.

4.2 Flood Frequency Analysis

The annual maximum discharges were fitted to the normal, lognormal and log Pearson type iii distributions and the exceedence probabilities determined using the various equations in Table 1. The determined probabilities of exceedence of the peak discharges before subjecting them to the statistical distributions are shown in Table 2.
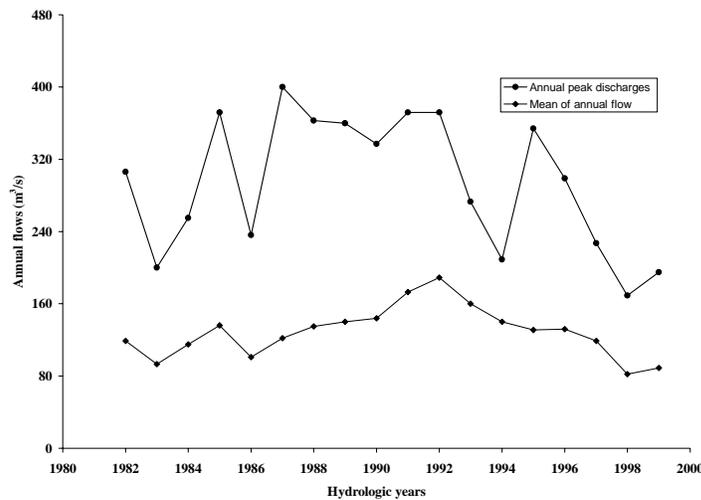
Figure 2 Annual flows of River Osun at Apoje gauging station from 1982 to 1999.

Table 2: Probabilities of exceedence of the annual maximum discharges using seven plotting positions

Years	Annual maximum discharges (m^3/s)	Hazen	Weibull	Blom	Cunnane	California	Gringorton	Chegodajev
1982	306	0.0278	0.0526	0.0342	0.0329	0.0556	0.0309	0.0380
1983	200	0.0833	0.1053	0.0890	0.0879	0.1111	0.0861	0.0924
1984	255	0.0833	0.1053	0.0890	0.0879	0.1111	0.0861	0.0924
1985	372	0.0833	0.1053	0.0890	0.0879	0.1111	0.0861	0.0924
1986	236	0.1389	0.1579	0.1438	0.1429	0.1667	0.1413	0.1467
1987	400	0.1944	0.2105	0.1986	0.1978	0.2222	0.1965	0.2011
1988	363	0.2500	0.2632	0.2534	0.2528	0.2778	0.2517	0.2554
1989	360	0.3056	0.3158	0.3082	0.3077	0.3333	0.3068	0.3098
1990	337	0.3611	0.3684	0.3630	0.3626	0.3889	0.3620	0.3641
1991	372	0.4167	0.4211	0.4178	0.4176	0.4444	0.4172	0.4185
1992	372	0.4722	0.4737	0.4726	0.4725	0.5000	0.4724	0.4728
1993	273	0.5278	0.5263	0.5274	0.5275	0.5560	0.5278	0.5272
1994	209	0.5833	0.5789	0.5822	0.5824	0.5611	0.5828	0.5815
1995	354	0.6389	0.6616	0.6370	0.6374	0.6667	0.6379	0.6359
1996	299	0.6944	0.6842	0.6918	0.6923	0.7222	0.6932	0.6900
1997	227	0.7500	0.7368	0.7466	0.7423	0.7778	0.7483	0.7446
1998	169	0.8056	0.7895	0.8016	0.8022	0.8333	0.8035	0.7989
1999	195	0.9167	0.8421	0.8562	0.8571	0.8889	0.8587	0.8533

4.3 Hazen's Plotting Position

The observed discharges in Figs.3a, b and c at return periods of 25, 50 and 100 years were 404, 412 and 417m^3/s respectively and the coefficient of determination R^2 was 0.9761. Figure 3a shows the graphical illustration of the annual maximum discharges against the non-exceedence probability for the normal distribution using the Hazen's plotting position. From this figure, it can be seen that at return periods of 25, 50 and 100 years, the predicted discharges were 410, 416 and 420m^3/s respectively. The coefficient of determination R^2 was 0.9908. These estimates are higher and compares well with observed discharges. Minimum absolute differences between the predicted and the observed discharges were obtained under this distribution. The root mean square error (RMSE) was 35.84m^3/s (see Table 3). Figure 3b shows similar graphical illustration as in Fig.3a but under the lognormal distribution. At return periods of 25, 50 and 100 years, the predicted discharges were 434, 441 and 445m^3/s respectively and R^2 was 0.9752. These flood estimates are higher and do not compare well with the observed discharges. The RMSE was 43.09m^3/s, the highest under this plotting position. The highest absolute differences between the observed and predicted discharges were obtained under this distribution. Figure 3c gives the same graphical illustration as Figs.3a and 3b but this time under the log Pearson type (iii) distribution. At return periods of 25, 50, and 100 years, the predicted discharges were 427, 435 and 438m^3/s respectively and R^2 was 0.9890. The normal distribution gave the highest fit using this plotting position.

4.4 Weibull's Plotting Position

The observed discharges in Figs 4a, b and c at return periods of 25, 50 and 100 years were 417, 426 and 431m^3/s respectively and coefficient of determination R^2 was 0.9761. Figure 4a shows the graphical illustration of the annual maximum discharges against the probabilities of non-exceedence for the normal distribution using Weibull's plotting position. From the figure, at return periods of 25, 50 and 100 years, the predicted discharges were 408, 414 and 417m^3/s respectively and R^2 for the predicted discharges was 0.9950. The flood estimates compared well with the observed discharges having root mean square error (RMSE) of 35.09m^3/s which was the smallest under this plotting position (see Table 3). Similar graph of lognormal distribution is shown in Fig.4b. At return periods of 25, 50 and 100 years, the predicted discharges were 428, 435 and 439m^3/s respectively and R^2 for the predicted discharges was 0.9860. Minimum absolute differences between the observed and predicted discharges were obtained under this distribution. Figure 4c shows the graphical illustration for log-Pearson type (iii) distribution. At return periods of 25, 50 and 100 years, the predicted discharges 433, 440 and 444m^3/s respectively and R^2 values was 0.9942. By comparing the three statistical distributions, it is evident that the normal distribution gave the highest degree of correlation for the predicted discharges.

4.5 Blom's Plotting Position

The observed discharges in Figs 5a, b and c were 409, 422 and 427m^3/s at return periods of 25, 50 and 100 years respectively. The graph of annual maximum discharges against the non-exceedence probabilities for the normal distribution using Blom's plotting position is shown

in Fig.5a. At the return periods of 25, 50 and 100 years, the predicted discharges were 413, 419 and 422m^3/s respectively and R^2 was 0.9924. The root mean square error (RMSE) was 35.75m^3/s (see Table 3). These predicted discharges compares well with the observed discharged. Similar graphical illustration under lognormal distribution is shown in Fig.5b. The predicted discharges were 431,439 and 443 m^3/s respectively while the R^2 for the predicted discharges was 0.9790. The forecasted discharges are significantly higher and do not compare well with the observed discharges. Figure 5c shows the graphical illustration for log-Pearson type (iii) distribution. The predicted discharges were 425, 432 and 436m^3/s respectively at the same return periods and the coefficients of determination R^2 0.9909. The normal distribution gave the best fit and minimum absolute differences between the predicted and observed discharges and is therefore selected as the best distribution using this plotting position.

4.6 Cunnane's Plotting Position

The observed discharges in Figs 6a, b and c were 410, 419 and 423m^3/s at return periods of 25, 50 and 100 years respectively. Figure 6a shows the graphical illustration of the annual maximum discharges against the non-exceedence probability using the normal distribution. At the return periods of 25, 50 and 100 years, the predicted discharges were 412, 418 and 421m^3/s respectively and coefficient of determination R^2 was 0.9921. Figure 6b shows the graph similar graph for lognormal distribution. The predicted discharges at return periods of 25, 50 and 100 years were 435, 443 and 447m^3/s respectively and R^2 for the predicted discharges was 0.9784. Figure 6c shows graphical illustration for log-Pearson type (iii) distribution. The predicted discharges at the same return periods were 425, 432 and 435m^3/s respectively and R^2 was 0.9906. The observed and predicted discharges in Fig 6a compare favourably having the least absolute differences unlike Fig.6b where the observed and predicted discharges were significantly higher. In the light of this, normal distribution is taken as the best under this plotting position
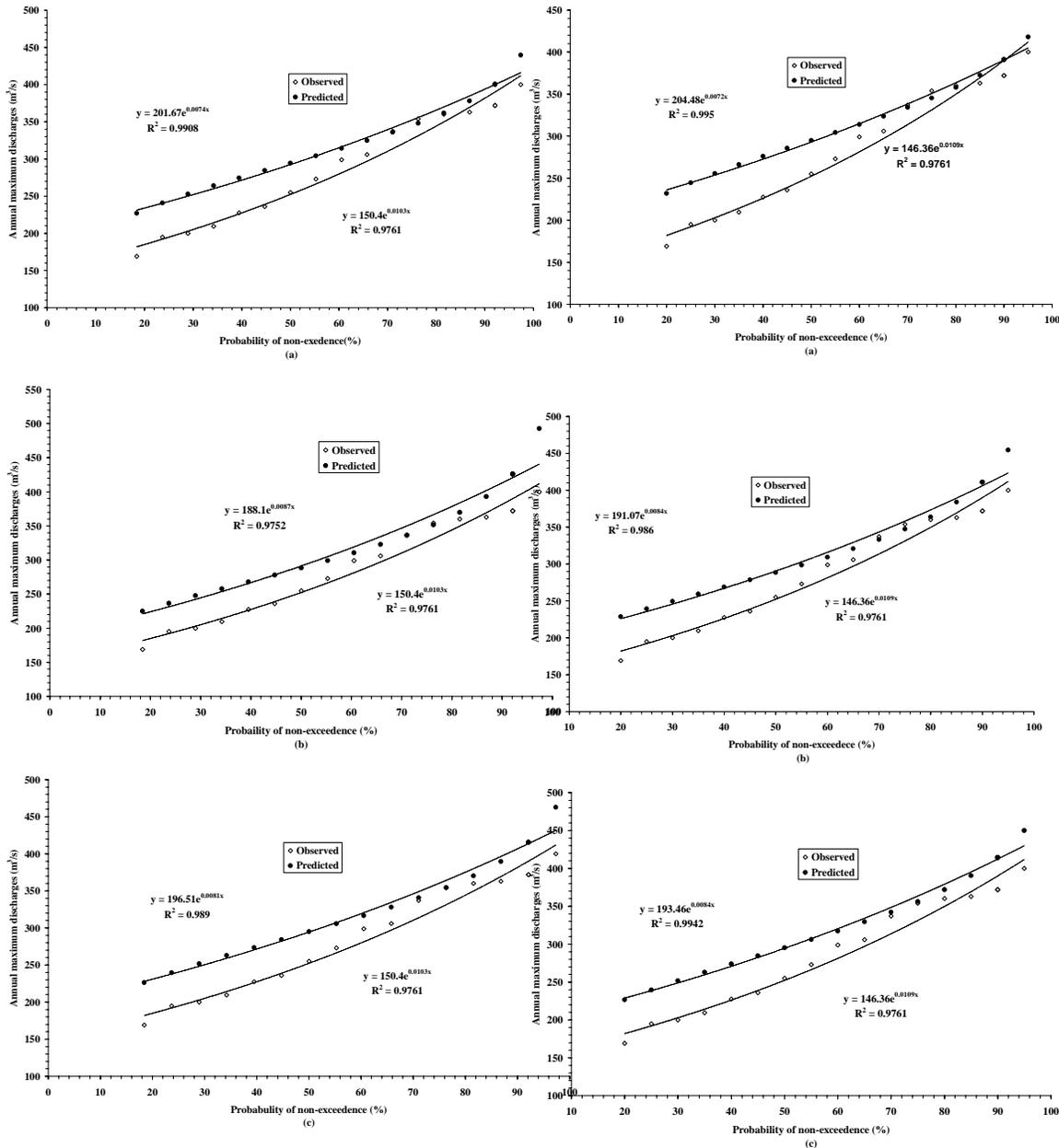
Figure 3. Normal (a), Lognormal (b) and Log-Pearson type (iii) (c), distributions using Hazen's plotting position

Figure 4. Normal (a), Lognormal (b) and Log-Pearson type (iii) (c) distributions using Weibull's plotting position
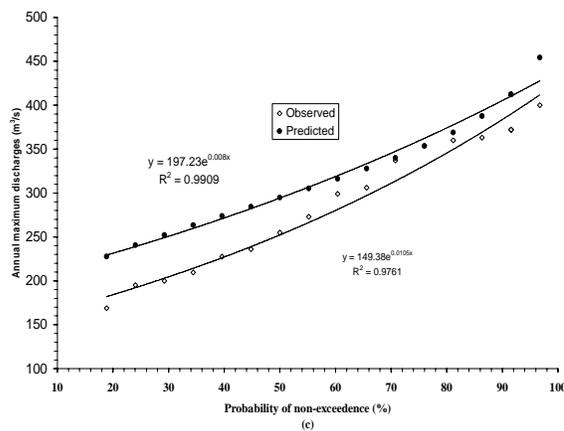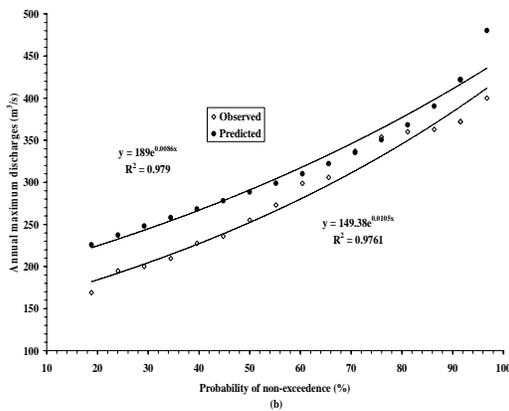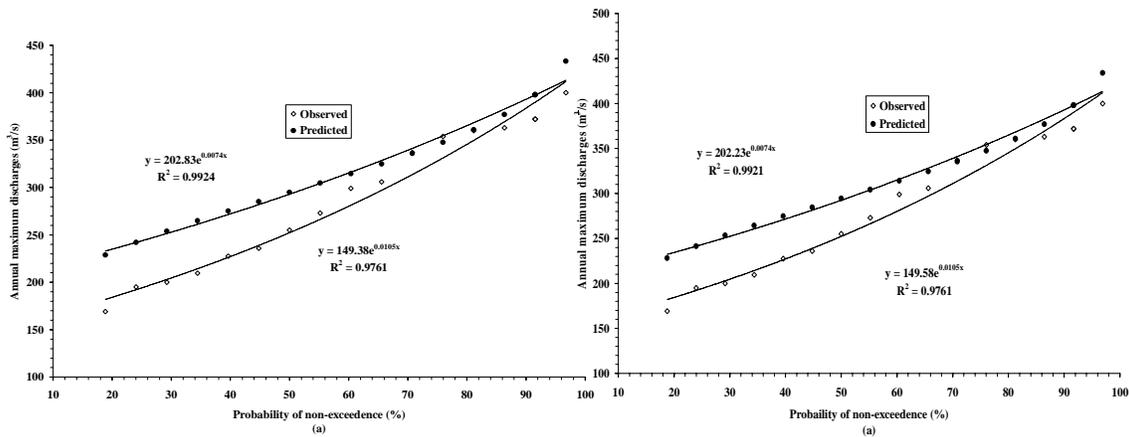
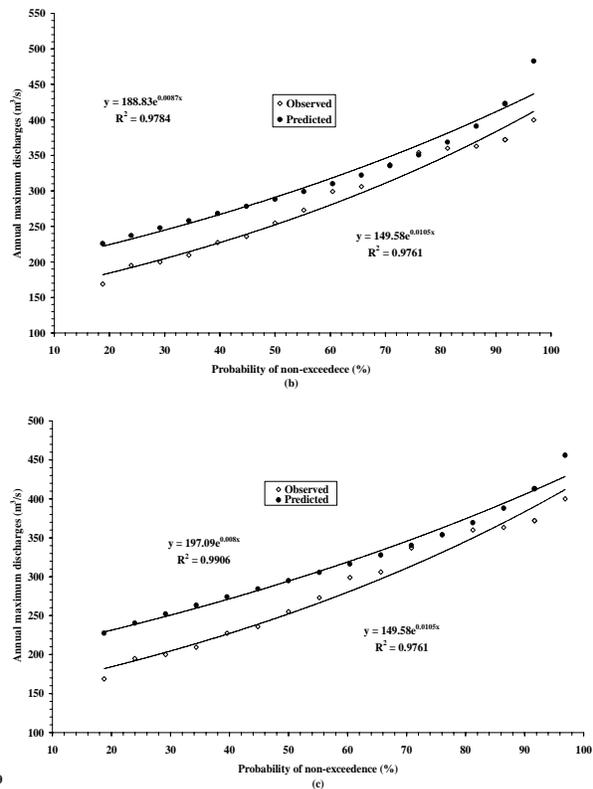Figure 5. Normal (a) Lognormal (b) and Log-Pearson type (iii) (c) distributions using Blom's plotting position

Figure 6. Normal (a), Lognormal (b) and Log-Pearson type (iii) (c) distributions using Cunnane's plotting position

4.7 California Plotting Position

The observed discharges at return periods of 25, 50 and 100 years were 415,424 and 429 m^3/s respectively (see Figs 7a, b and c). Figure 7a shows the graphical illustration of the annual maximum discharges against the probabilities of non-exceedence for the normal distribution using California plotting position. The predicted discharges at return periods of 25, 50 and 100 years were 407, 413 and 416m^3/s respectively and coefficient of determination R^2 was 0.9939. Although of smaller magnitudes, the predicted discharges compares well observed discharges Also, the minimum root mean square error (RMSE) of 29.06m^3/s was estimated under this distribution (see Table 3). Figure 7b shows the similar graph for lognormal distribution. At return periods of 25, 50, and 100 years, the predicted discharges were 425, 432 and 435m^3/s respectively and R^2 was 0.9887. Figure7c gives the same graphical illustration for log–Pearson type (iii) distribution. At the return periods of 25, 50 and 100 years, the predicted discharges were 422, 429 and 432m^3/s respectively and R^2 was 0.9946. The log-Pearson type (iii) distribution had the highest fit and the minimum absolute differences between the observed and predicted discharges and is selected as the best using this plotting position.

4.8 Gringorton's Plotting Position

The observed discharges at return periods of 25, 50 and 100 years were 415,424 and 429 m^3/s respectively as shown in Figs. 8a, b and c. Figure 8a gives the graphical illustration of the normal distribution using Gringorton's plotting position. At return periods of 25, 50 and 100 years, the predicted discharges were 411, 417 and 420m^3/s respectively and coefficient of determination R^2 was 0.9916.These predicted and observed discharges compares well and gave the minimum absolute differences when compared with all other plotting positions used in this statistical analysis. The root mean square error (RMSE) was 35.57m^3/s (see Table 3). Figure 8b shows similar graph for lognormal distribution. The predicted discharges at the same return periods were 435, 442 and 446m^3/s respectively and R^2 was 0.9772. Figure 8c gives graphical illustration for log Pearson type (iii) distribution. The predicted discharges were 424, 431 and 435m^3/s respectively at the same return periods and R^2 for the predicted discharges was 0.9900. The regression coefficient coupled with the minimum absolute difference show that normal distribution gave the best fit to the annual peak flow using this plotting position.

4.9 Chegodajev's Plotting Position

The observed discharges at return periods of 25, 50 and 100 years were 412, 420 and 425 m^3/s respectively as shown in Figs. 8a, b and c. Figure 9a shows the normal distribution of the annual maximum discharges against the. non-exceedence probabilities using Chegodajev's plotting position. At return periods of 25, 50 and 100 years, the predicted discharges were 409, 415 and 418m^3/s respectively and coefficient of determination R^2 was 0.9931. The root mean square error (RMSE) was 33.32m^3/s (see Table 3). Similar graphical illustration for lognormal distribution is shown in Fig.9b. At return periods of 25, 50, and 100

years the predicted discharges were 433, 440 and 444m^3/s respectively and R^2 was 0.9809. Figure 9c shows the graphical illustration for log-Pearson type (iii) distribution. The predicted discharges were 426,432 and 436m^3/s at return periods of 25, 50 and 100 years respectively. The coefficient R^2 was 0.9918. Minimum absolute differences were obtained between the predicted and observed discharges in Fig. 9a and the highest coefficient of determination was obtained under normal distribution.

Table 3 shows the coefficients of determination, root mean square errors (RMSE) and absolute differences between observed and predicted discharges. The best distribution was determined by considering the absolute differences between the observed and the predicted discharges, coefficients R^2 of their regression equations and the root mean square errors (Kottegodal, 1980; Benjamin, Garry and Jean, 2001). The minimum absolute differences between the predicted and observed discharges under each plotting position were obtained using normal distribution. However, the Weibull's plotting position that has the highest coefficient of determination R^2 of 0.9950 under normal distribution had higher absolute differences when compared with those of Cunnane and Gringorton under the same normal distribution. California plotting position had the minimum (RMSE) under the normal distribution. As indicated earlier, the best formula is selected based on purpose of investigation and the distribution under investigation. The Weibull's plotting position had the highest R^2 and is hereby selected as the best under normal distribution. This is in agreement with Alatise, (1998) and Chen, *et al*, (2002) where the same indicators were used in selecting optimum formulae in flood analyses. Similarly, log-Pearson type (iii) and normal distributions had the highest coefficient of determination R^2 using the California formula. Generally, the performances of the probability distributions were satisfactory as none of them had coefficient of determination R^2 that was less than 0.98. For the purpose of accurate flood forecast which is instrumental to decisions making on flood mitigation measures, graphical approaches are good and is encouraged.
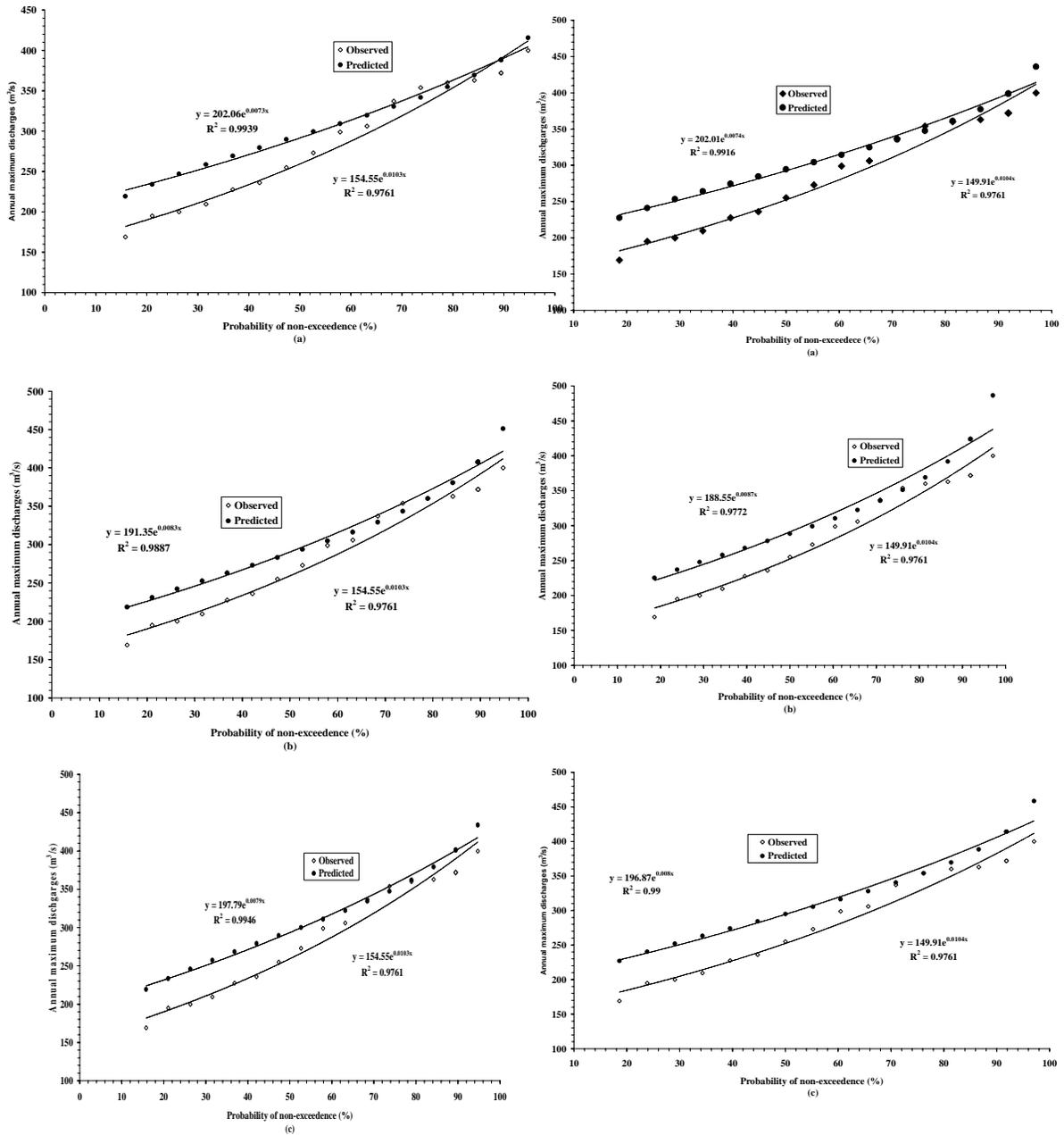
Figure 7. Normal (a), Lognormal (b) and (c) Log-Pearson type (iii) (c) distributions using California plotting position

Figure 8. Normal (a), Lognormal (b) and (c) Log-Pearson type (iii) (c) distributions using Gringorton's plotting position
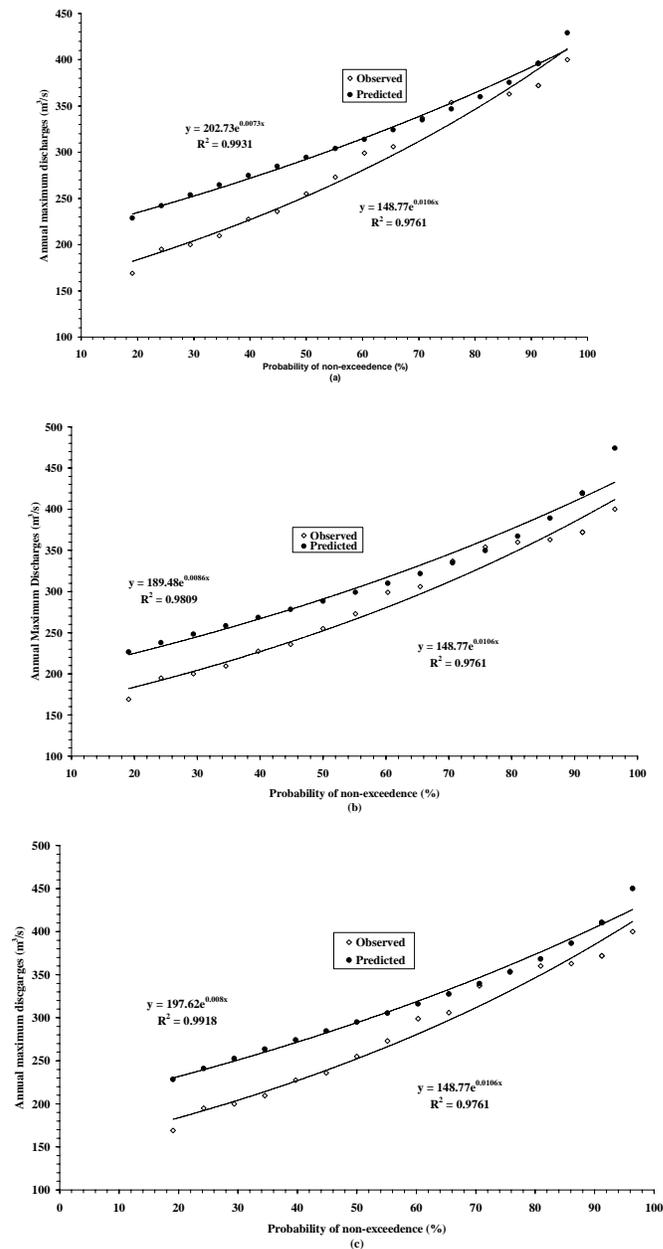
Figure 9. Normal (a), Lognormal (b) and (c) Log-Pearson type (iii) (c) distributions using Chegodajev's plotting position

Table 3: Coefficients of determination, root mean square errors (RMSE) and absolute differences between observed and predicted discharges

	Plotting Positions	Probability Distributions								
		Normal			Lognormal			Log-Pearson type (iii)		
Correlation coefficients R^2	Hazen	0.9908			0.9752			0.9890		
	Weibull	0.9950			0.9860			0.9942		
	Blom	0.9924			0.9790			0.9909		
	Cunnane	0.9921			0.9784			0.9906		
	California	0.9939			0.9887			0.9946		
	Gringorton	0.9916			0.9772			0.9900		
	Chegodajev	0.9931			0.9809			0.9918		
Root Mean Square Errors (RMSE) (m^3/s)	Hazen	35.84			43.09			40.24		
	Weibull	35.09			36.59			39.25		
	Blom	35.75			40.86			39.12		
	Cunnane	35.45			41.26			39.32		
	California	29.06			31.87			31.89		
	Gringorton	35.57			41.95			39.67		
	Chegodajev	33.32			37.73			36.60		
Return Periods (years)		25	50	100	25	50	100	25	50	100
Absolute differences (m^3/s)	Hazen	06	04	03	30	29	28	23	23	21
	Weibull	09	12	14	11	09	08	16	14	13
	Blom	04	03	05	22	17	16	16	10	09
	Cunnane	02	01	02	25	24	21	15	13	12
	California	08	11	13	10	08	06	07	05	03
	Gringorton	04	02	00	28	27	26	17	16	15
	Chegodajev	03	05	07	21	20	19	14	10	11

O.B. Adeboye and M.O. Alatise and "Performance of Probability Distributions and Plotting Positions in Estimating the Flood of River Osun at Apoje Sub-basin, Nigeria". Agricultural Engineering International: the CIGR Ejournal. Manuscript LW 07 007. Vol. IX. July, 2007.

4. CONCLUSION

In this study, the peak discharges were plotted against their hydrologic years. Three probability distributions and seven plotting positions were fitted to the annual maximum discharges of River Osun at Apoje gauging station. The performances of the probability distributions were assessed using the coefficients of determination, root mean square errors (RMSE) and absolute differences between predicted and observed discharges. The following conclusions were drawn from the study:

- The annual maximum discharges of River Osun at Apoje gauging station vary in magnitude and space, ranging from 169 to $400\text{m}^3/\text{s}$ within between 1982 and 1999. The minimum "peak flow" in the dry season was $72\text{m}^3/\text{s}$. The minimum and maximum mean of the annual flows were 82 and $189\text{m}^3/\text{s}$ respectively.
- The normal distribution had the highest coefficient of determination using Welbull's plotting position while lognormal and log-Pearson type (iii) distributions had the highest fit when matched with California plotting positions.
- Normal distribution had minimum (RMSE) when matched with California and Weibull's plotting positions. The lognormal distribution had the minimum (RMSE) when matched with California plotting position.
- The minimum absolute differences at return periods of 25, 50 and 100 years were obtained under the normal distribution when matched with Cunnane plotting position.

The flood frequency analysis shows that under the normal distribution, the Weibull's formula gives the best fit while California formula gives the best fit under the log-Pearson type (iii) distribution and can therefore be proposed for the river basins in the rain forest belt. Generally, Apoje sub-basin has abundant water potential and if well harnessed can be used for various purposes in the basin.

5. REFERENCES

Adeboye O. B. 2005. Flood Characteristics and Potential Reservoir Capacity of River Osun at Apoje Gauging Station, M. Eng Thesis, Unpublished, Department of Agricultural Engineering, Federal University of Technology, Akure, Nigeria, 110pp

Akso, H. 2000. Use of Gamma Distribution in Hydrological Analysis, *Turks Journal of Engineering and Environmental Sciences*, **24:**419-428

Alatise, M. O. 1998. Comparison of the Log-Normal and Pearson type (iii) Methods in Analyzing the Flood Characteristics of River Owena in Ondo State, *N.S.E Technical Transactions*, **33**(4), 18 pp

Benjamin, F. P., D. T. Gary, and C. R. Jeane. 2001. Estimating the Magnitudes and Frequency of Floods in the Rural Basin of the North Carolina-Revised, U.S.G.S, *Water Resources Investigation Reports* 01-4207, Raleigh, North Carolina, 49pp

Blom, G. 1958. Statistical Estimates and Transformed Beta-variable, John Willey & Sons, N.Y

Chegodajev, N. N. 1955. Formulas for Calculation of the Confidence of Hydrologic Quantities, Handbook of Applied Hydrology, V.T. Chow, ed, McGraw Hill, N.Y

Chen.Y. U. Y., V. G. Pieter, and Z. Sha, 2002, A Study of the Parameter Estimation Methods for Pearson Type (iii) Distributions in Flood Frequency Analysis, *IASH Publications,* No 271, Iceland, 350-407.

Cunnane, C. 1978. Unbiased Plotting Position- A review, *Journal of Hydrology,* 37:205-222

Flows in California stream, 1923. Bulletin Number 5, Division of Engineering and Irrigation, California, Department of Public works, California state Printing office, Sacramento, California

Chow, V. T., R. D. Maidment, and W. M. Larry, 1988. *Applied Hydrology,* International edition, McGraw Hill Inc, Singapore, 350-407

Claps, P. and F. Laio. 2003. Can Continuous Streamflow Data Support Frequency Analysis? An Alternatives to the Partial Duration Series Approach, *Water Resources Research,* **39** (8), 1-11

Ekanayake. S. T., J. F. Cruise, 1993. Comparisons of Weibull and exponential based partial duration stochastic flood models, Stochastic Hydrology and Hydraulics, 7: 283-297

Federal Republic of Nigeria. 1982. Semi Detailed Soil Survey of Osun, Ona and Ogun River Basins, Prepared by Department of Soil and Science, Obafemi Awolowo University, Ile-Ife, Nigeria.

Gringorton, I. I. 1963. A Plotting Position for Extreme Probability Paper, *Journal of Geophysics Research,* **68**(3), 813-814

Green I. R. A., and D. Stephenson. 1985. Comparison of Urban Drainage Models for use in South Africa, WRC Report No 115/6/86., *Water Resources commission,* Pretoria RSA.

Hazen, A. 1914. Storage to be Provided in Impounding Reservoir for Municipal Water Supply, Transaction of ASCE, 77:1547-1550

Hebson, C. S and C. Cunnane. 1987. Assessments of the Site and Regional Flood Frequency Estimation, In VP Singh(ed), *Hydrology frequency modelling*, Reidel Norwell, Massachusetts, U.S.A, 433-448

Hirsch, R. M. and J. R. Stedinger, 1987. Plotting Positions for Historic Flood and their Precision, *Water Resources Research,* **23**(4), 715-727

Hoskin J. R. M and J. R. Wallis. 1997. Regional Frequency Analysis: An approach based on L-Moments, Cambridge University press, UK, 22pp

Hromadka, 11 T. V and R. J. Whitley, 1989. Checking Flood Frequency Curves using Rainfall Data, *Journal of Hydraulic Engineering,* ASCE, **115**(4), N.Y, 544-548

Kite, G. M. 1977. Frequency and Risk Analysis in Hydrology, *Water Resources Publications,* Fort Collins, N.Y

Kjeldsen, T. R., J. C. Smithers and R. E. Schulze, 2001. Flood Frequency Analysis at ungauged sites in the KwaZulu-Nata Province, South Africa, *Water SA*, **27**(3), 315-324

Kottegodal, N. T. 1980. Stochastic Water Resources Technology, 1st edition, Macmillan Press Ltd, London, 208-215

Moughamiam, M. S., D. B. Mclaughlin, and R. E., Bras. 1987. Estimation of Flood Frequency: An Evaluation of two Derived Distribution Procedures, *Water Resources Research*, **23**(7), 1309-1319

O' Donnell. T. 1985. A Direct three-parameter Muskingum procedure incorporating lateral inflow, *Journal of Hydrological Sciences*, **30**(4), 494-495

Robert, M. H. 1987. Plotting Positions for Historic Floods and their Precision, American *Geophysical Union,* **23**(4), Washington, D.C, U.S.A, 715-719

Sumioka, S. S., D. L Kresh, and K. D Kasnick. 1997. Magnitudes and Frequencies of Flood in Washington, US Geological Survey, *Water Resources Investigation Report,* 97-4277 W. A, 15 pp

Tadesse N. 2006. Surface Water Potentials of the Hantebet Basin, Tigray, Northern Ethiopia", Agricultural Engineering International: *CIGR Ejournal*, Vol. VIII.

Tewolde. M. H. and J. C. Smitthers. 2006. Flood Routing in Ungauged Catchments using Muskingum methods, WRC, *Water S.A* **32**(3), RSA

Tilahun K. 2006. The Characterization of Rainfall in the arid and semi arid regions of Ethiopia, *Water SA*, **32**(3), 429-439, http://www.wrc.org.za

Topaloglu, F. 2002, Determining Suitable Probability Distributions Models for Flow and Precipitation Series of Seyhan River Basin*, Turk Journal Agriculture,* 26, 187-194

Van Bladeren, D. 1993. Application of Historic data in Flood Frequency Analysis for Natal and Transkei Regions. In proceeding of the 6th South African National Hydrologic Symposium Vol.1, Pietermaritzburg, RSA

Van Thanh-Van-Nguyen, V. N., I. Nophadel., B. Bernarb. 1989. New Plotting Position Formulae for Pearson Type (iii) distribution, *Journal of Hydraulic Engineering,* Transaction of ASCE, **115**(6), N.Y, 709–730.

Viessman, W. J and L. Lewis. 1996. Introduction to Hydrology, 4th edition, HapercollinCollege Publisher, N.Y, 760pp

Weibull, W. 1939. A Statistical Theory of Strength of Materials Ing. Vet. A.K, Handl., 151, Genelstabens Litografiska Anstals Forlg Stocklholm, Sweden

U.S. Water Resources Council, 1981, Guidelines for Determining Flood Flow Frequency, Bulletin 17A*, U.S Geological Survey,* Washington, D.C