ARID ZONE JOURNAL OF ENGINEERING, TECHNOLOGY & ENVIRONMENT
AZOJETE March 2022. Vol. 18(1):143-158
Published by the Faculty of Engineering, University of Maiduguri, Maiduguri, Nigeria.
Print ISSN: 1596-2490, Electronic ISSN: 2545-5818
www.azojete.com.ng

ORIGINAL RESEARCH ARTICLE

EVALUATION OF BEST- FIT PROBABILITY DISTRIBUTION MODELS FOR PREDICTION OF RAINFALL IN SOUTHERN NIGERIA

A. I. Agbonaye*. and O. C. Izinyon
Department of Civil Engineering, University of Benin, Benin City. Nigeria
**Corresponding author's email address: augustine.agbonaye@uniben.edu*

ARTICLE INFORMATION

Submitted 15 Sept., 2021
Revised 19 October, 2021
Accepted 24 October, 2021

Keywords:
Best- Fit
Maximum-Rainfall
Probability-Distribution
Goodness-of–fit
Return-level

ABSTRACT

Suitable and adequate hydrological design data especially for flood design, are not readily available. When data is available, it is often outdated and irrelevant to current events especially in this era of global warming and climate change. This study presents rainfall frequency analysis for some cities in Southern Nigeria using the annual maximum series of daily rainfall data for the stations. The objective of the study was to select the probability distribution model from among six commonly used probability distribution models namely: Generalized Extreme value distribution (GEV), Extreme value type I distribution (EVI), Generalized Pareto distribution (GPA), Pearson Type III (PIII), log Normal (LN) and Log Pearson Type III (LP111) distributions. These distributions were applied to the annual maximum series of daily rainfall at each station using the parameters of the distributions estimated by the method of moments. The best fit probability distribution model at each location was selected based on the results of seven goodness of fit tests values with a scoring and ranking scheme. Our results indicate that the best-fit distribution models at the study locations are PIII for Ibadan and Benin City; GEV for Onitsha, Enugu, Owerri, Calabar, and Port Harcourt; EVI and LN for Uyo; EVI for Akure and LP111 for Ikeja. This implies that GEV performed better by occupying 50% of the studied area, followed by EVI and PIII which performed by occupying 20% each. These best fit probability distribution models are recommended for use for necessary design at each location for flood hazard mitigation.

1.0 Introduction

The design and construction of certain projects such as dams and urban drainage systems, management of water resources, and prevention of flood damage require adequate knowledge of extreme events of high return periods as worst conditions are normally considered in designs (Tao *et al*., 2002; Olofintoye *et al*., 2009). In most cases, the return periods of interest sometimes exceed the period of available records and as such cannot be extracted directly from the recorded data. It is necessary to extrapolate design parameters from the data and this cannot be achieved by other methods. That is why in current engineering practice, the estimation of extreme rainfall depths used for design is accomplished based on statistical frequency analysis of maximum precipitation records. Rainfall frequency analysis is the estimation of how often rainfall of a specified magnitude will occur. Such analyses help define policies relating to water resources management. It serves as the source of data for flood hazard mitigation and the design of hydraulic structures aimed at reducing losses due to flood action. The choice of an appropriate probability distribution and parameter estimation method plays a vital role in rainfall frequency analysis. Further stimulation for this study resulted from a recent presentation by NIMET of evidence of climate variation especially on the trend of maximum rainfall in 2012 (Okoloye *et al*., 2013). As the climate is changing, the necessity to accurately estimate extreme events such as the maximum rainfall frequency has become urgent as it will contribute to the design of safer and more efficient hydraulic structures. Thus, the application of probability distributions to rainfall data have been investigated by several researchers from different regions of the world**.** The Best Fitted Probability Distribution for Monthly Rainfall Data was obtained by Ghosh *et al*. (2016) for three Bangladeshi stations from 1979 to 2013. They revealed that the best suited distribution model of monthly rainfall data was the generalized extreme value. Mohamed and Ibrahim (2016) analyzed annual rainfall data for fourteen rainfall stations in Sudan from 1971 to 2010. Normal and gamma distribution models were chosen by them as the best match probability distribution models. For 35 locations in Bangladesh, Alam *et al*. (2018) determined the best-fit probability distributions for maximum monthly rainfall using 1984–2013 data. After applying different statistical analysis and distribution types, they discovered the best-fitting probability distributions to be Generalized Extreme Value, Pearson type 3, and Log-Pearson type 3. Also, they calculated 10-year, 25-year, 50-year and 100-year return periods of maximum monthly rainfall for all locations studied. Kumar (2000) and Singh *et al*. (2012) found that the Log normal type 2 distribution is the best-fit probability distribution for annual maximum daily rainfall in India, Amin et al. (2016) used annual maximum rainfall based on a daily rainfall and found that the log-Pearson type 3 distribution was the best-fit distribution in the northern regions of Pakistan. Kousar *et al*. (2020) determined the best-fitted probability distributions for at-site flood frequency analysis of the Ume River in Sweden. The generalised extreme value distribution with the L-moments estimation provided the best fit to maximum annual streamflow at gauging sites Solberg and Stornor-Krv. Also, the best-fitted distribution for each gauging site was used to predict the maximum flow of water for return periods of 5, 10, 25, 50, 100, 200, 500, and 1000 years. Mamman *et al*. (2017) Evaluated the Best-Fit Probability Distribution Models for the Prediction of Inflows of Kainji Reservoir, Niger State, Nigeria and selected the Gumbel (EVI) model as the best fitted distribution model. Masereka *et al.* (2015) identified the best fit probability distribution models for Frequency Analysis of Extreme Mean Annual Rainfall Events in South Africa to be Log Pearson 3, Generalised Logistic and Extreme Generalised Value. Langat *et al*. (2019) Identified the most suitable probability distribution models for maximum, minimum, and mean streamflow for Tana River in Kenya. The log-normal and GEV distribution functions were the best-fit functions for the annual mean flows of the Tana River.

However, Probability distribution models utilized for the analysis of hydrological data are numerous but the six commonly used models by researchers in Nigeria are: Generalized Extreme value distribution (GEV), Extreme value type I distribution (EVI), Generalized Pareto distribution (GPA), Pearson Type III (PIII), log Normal (LN) and Log Pearson Type III (LP111). Interestingly, no particular model is considered superior for all practical purposes. WMO (2008) recommended that available models are screened based on the problem to be solved and the nature of available data. Data for frequency analysis should be independent and identically distributed. Hence, it is necessary to screen candidate distributions for best fit to available data at a location (Agbonaye and Izinyon, 2017). Therefore, this study is needed to screen and select the best- fit probability distribution models for prediction of rainfall in southern Nigeria.

2.0 Materials and Method

2.1: *Data and Analysis*

This study entails frequency analysis of the annual maximum series of daily rainfall depths of selected cities in Southern Nigeria. The cities selected are: Ikeja, Akure, Ibadan, Benin City, Port Harcourt, Uyo, Calabar, Onitsha, Enugu, Owerri. The Map of Nigeria Showing Ten Selected Cities is presented in Figure.1.

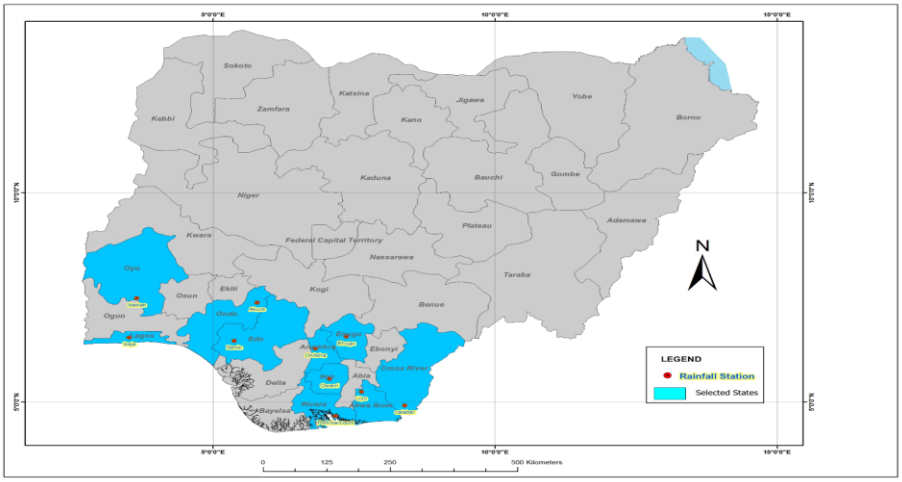

Figure.1 Map of the study area.

Source: Adapted from Office of the Surveyor General of the Federation (OSGOF, 2011)

The daily rainfall data for the selected cities were obtained from the Nigerian Meteorological Agency (NIMET) Oshodi, Lagos for the period between 1965 and 2014 (50 years). The annual maximum series is the series formed from the selection of the highest rainfall value for each of the years under consideration for the stations. The annual series data for each station was tested for homogeneity using Rainbow Software and checked for outliers using HEC-SSP Software. Six probability distribution models namely: Gumbel Extreme Value type I (EVI), Generalized Extreme Value (GEV), Generalized Pareto (GPA), Log-Normal, and Pearson type III and log Pearson type III were fitted to the annual maximum series data at each location. The Quantile Function and Parameters of Probability Distribution models are shown in Table 1. The best fit model at a location was selected based on results of the goodness of fit tests with the application of a scoring and ranking scheme. Subsequently, the best fit model at a location was used to forecast rainfall return levels for return periods of engineering design significance (5 to 500 years).

Table 1: Quantile Function and Parameters of Probability Distribution (Vivekananda, Koutsoyiannis 2014)

S. No.	Distribution	Quantile function (R_T)	Parameter by MOM
1	EV1	$R_T = \xi - \alpha \ln(-\ln F)$	$\xi = \bar{R} - 0.5772157\alpha$ $\alpha=(\sqrt{6}/\pi)\, S_R$
2	GEV	$R_T = \xi + \frac{\alpha}{k}(-\ln F)^{-k} - 1$ Where $F = 1 - \frac{1}{T}$	$K = \frac{1}{3} - \frac{1}{0.31+0.91Csx+\sqrt{(0.91Csx)^2+1.8}}$ $\alpha = C_1 S_x$ $C1 = \frac{\lvert k \rvert}{\sqrt{\Gamma(1-2k) - \Gamma^2(1-k)}}$ $C3 = \frac{(\Gamma(1-k)-1)}{k}$ $\xi = R_m - C_3$
3	GPA	$R_T = \xi + \alpha(1-(1-F)^k)/k$	$\bar{R} = \xi + \alpha/(1+k)$; $S_R = \alpha^2/(1+2k)(1+k)^2$ $C_s = 2(1-k)(1+2k)^2/(1+3k)$
4	LN-NOR	$R_T = e^{\alpha+z_f k}$	$k=\sqrt{[\ln(1+\{Sx/Rm\}^2}$, $\alpha = k^2/2$
5	PIII	$R_T = \xi + Z_F \alpha$ where Z_F is obtained from a table	$k=4/c_{sx}^2$, $\alpha = S_R/\sqrt{k}$, $\xi=R_m-k\alpha$ where C_{sx} is the coefficient of skewness of obtained data, S_R is the standard deviation
6	LN-PIII	$R_T=e^{\alpha+z_f{}^* k}$	$K=4/C_{sy}^2$, $\alpha=S_y/\sqrt{k}$, $\xi=R_{Lm}-k\alpha$

F(R) (or F) is the cumulative distribution function (CDF) of R; P is the probability of exceedance ξ, α and k are the location scale and shape parameters respectively; $\mu(or\,R), \sigma(or\,S_R)$ and $C_s(or\,\psi)$ are the average, standard deviation and coefficient of skewness of the recorded rainfall data; sign (k) is plus or minus 1 depending on the sign of k; R_T is the estimated rainfall by the probability distribution for a return period).

2.2 *Goodness of fit test criteria*

The goodness of fit test used to check the adequacy of fit of probability distribution models to the series of recorded rainfall data for the stations are Root mean square error (RMSE), Relative root mean square error (RRMSE), Mean absolute deviation index (MADI), Maximum absolute error (MAE), Probability plot correlation coefficient (PPCC), Chi-square (x^2) test.

It is to be noted that in evaluating the performance of a probability distribution model at a location, the lower the value of the Goodness-of-fit test results the better the distribution except for the PPCC criterion in which the nearer the value is to 1 numerically the better the distribution.

2.2.1 *Root mean square error (RMSE)*

The root means square was used to measure the difference between values predicted by a model and the observed values. These individual differences are called residuals. The RMSE of a model prediction concerning the estimated variable x model was obtained using the formula (Tao et al., 2002)

$$RMSE = \left(\frac{\sum (R_o - R_f)^2}{(n-m)} \right)^{\frac{1}{2}} \quad (1)$$

Where $R_i, i = 1,....,n$ are observed values while R_f are the corresponding values computed from the assumed probability distribution and m is the number of parameters estimated for the distribution. The smaller the value of RMSE obtained for distribution, the more fitted it is to the observed data (Sabri and Arif, 2009; Ahmad et al., 2011). Hence, a smaller value of RMSE for candidate distributions indicates that it is more fitted to the observed data.

2.2.2 Relative Root Mean Square Error (RRMSE)

Relative root means the square error was computed by dividing the root mean square error by the mean observed data. The relative root means square error provides a good picture of the overall fit of a distribution. It computes each error in proportion to the size of observation thereby reducing the effect of outliers that are

Commonly found in hydrological data. RRMSE was computed using the formula (Tao et al., 2002)

$$RRMSE = \left(\frac{\sum \left(\frac{R_{oi} - R_{fi}}{R_{oi}} \right)^2}{(n-m)} \right)^{\frac{1}{2}} \quad (2.)$$

Where

$R_{oi}, i = 1,...,n$ is observed values while R_f are the corresponding values computed from the assumed probability distribution and m is the number of parameters estimated for the distribution. . The smaller the value of RRMSE obtained for distribution, the more fitted it is to the observed data (Sabri and Arif, 2009; Ahmad et al., 2011). Hence, a smaller value of RRMSE for candidate distributions indicates that it is more fitted to the observed data.

2.2.3 Mean Absolute Deviation Index (MADI)

The mean absolute deviation index is the average distance between each data value and the mean. It was computed using the formula (Sabri and Arif, 2009; Ahmad *et al*., 2011):

$$MADI = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{R_o - R_f}{R_o} \right| \quad (3)$$

Where R_0 is the observed value, R_f the value forecasted by the distribution, and N the number of data points. The smaller the value of MADI obtained for distribution, the more fitted it is to the observed data (Sabri and Arif, 2009; Ahmad et al., 2011). Hence, a smaller value of MADI for candidate distributions indicates that it is more fitted to the observed data.

2.2.4 Maximum Absolute Error (MAE)

The maximum absolute error gives the largest absolute difference between the observed (Filliben, 1975) and values predicted by the distributions. MAE was computed using the relationship given by the formula).

$$MAE = \max\left(\left|R_o - R_f\right|\right) \tag{4}$$

Where R_o are the observed values, R_f are the values predicted by the probability distribution model. . The smaller the value of MAE obtained for distribution, the more fitted it is to the observed data. (Filliben, 1975). Hence, a smaller value of MAE for candidate distributions indicates that it is more fitted to the observed data.

2.2.5 Probability Plot Correlation Coefficient (PPCC)

PPCC is a measure of the correlation between the ordered observations and corresponding fitted values determined by a plotting position equation. PPCC was computed using the relation (Filliben, 1975).

$$PPCC = \frac{\sum\left\{\left(R_{0i} - R_m\right)\left(R_{f_i} - \overline{R_{fm}}\right)\right\}}{\left[\sum\left(R_{0_i} - \bar{R}_m\right)^2 \sum\left(R_{f_i} - \overline{R_{fm}}\right)^2\right]^{\frac{1}{2}}} \tag{5}$$

Where R_m, R_{mf} represent the mean values of the observed and forecasted values respectively.
A value of PPCC close to 1 suggests that the observed data could have been drawn from the fitted distributed at a location.

2.2.6 Chi-Square Test

The Chi-square test is a statistical test commonly used to compare observed data with the data we would expect to obtain according to a specific hypothesis. It is computed using the formula (Vivekananandan, 2014).

$$x^2 = \sum_{i=1}^{N} \frac{\left(R_i - R^*\right)^2}{R^*} \tag{6}$$

Where R_i, i, N, R^* are the year recorded data, year, total year, and the expected rain, respectively. The smaller the value of Chi-square test value obtained for distribution, the more fitted it is to the observed data.

The following steps were taken to determine the size of our Chi-Square values.:
(a) Calculating the number of "degrees of freedom" (d.f.).
(b) Choosing a level of likelihood (p=0.05) for P≤ 0.05;
(c) Obtaining a list of "critical Chi-Square values "from Tables.
(d) The critical Chi-Square values for this test were determined. (At the intersection of the appropriate d.f. row and probability column), and compared to the obtained value.

If $\chi^2 = 0$, it indicates that prediction with the model is exact match with the observed rainfall.
If $\chi^2 > 0$, they do not agree exactly.
The following is the stated hypothesis:
Null Hypotheses ($H_{0)}$ Probability Distribution Model fits the data
Alternative Hypotheses (Ha): Probability Distribution Model do not fit the data
Accept Ho): if χ^2 test statistics is less than χ^2 Critical values at 5% significance level
Or reject Ho): if χ^2 test statistics is greater than χ^2 Critical values at 5% significance level (P>.05)

2.2.7 *Diagnostic Test (D-Index)*

D – Index for distribution was computed using the relationship given by Vivekananandan (2014):

$$D - Index = \left(\frac{1}{R} \sum_{i-1}^{6} \left| R_i - R_i * \right| \right) \tag{7}$$

Where $\overline{R}$ is the average (or mean) of recorded Annual maximum Density (AMD) $R_i(i = 1 + 6)$ are the first six highest sample values in the series and R_i* is estimated value by probability distribution function (PDF). The distribution having the least D – index is identified as better suited, in Comparison with the other distribution for estimation of maximum rainfall values (Vivekananda, 2014).

2.2.8 *Scoring and ranking scheme*

Based on the results of the goodness of fit tests, a scoring scheme in which the best performing distribution concerning a test criterion was assigned a score of six (6), the next test was assigned five, and the worst test assigned one. The distribution with the highest total score at a location based on the Goodness of fit criteria was adjudged the best distribution model at the station and was selected as the best distribution at the station.

2.2 9 Model validation

The best-fit probability models were validated using Chi- square Goodness of fit Statistic explained in section 2.2.6.

2.2.10 Forecasting of rainfall Return Levels at the station

The rainfall return levels at a station were computed by substituting the appropriate return period into the appropriate Quantile function in Table 1 for the applicable best distribution

3. Results and Discussion

The summary of statistics means, standard deviation, skewness and kurtosis of the annual maximum rainfall series are presented in Table 2.

Table 2: Summary of Descriptive Statistics of Annual Maximum Rainfall (AMR)

Serial Number	Station Location	Mean	Standard Deviation	Skewness	Kurtosis
1	Ikeja	107.67	44.56	1.26	1.38
2	Akure	86.03	24.40	1.35	1.36
3	Ibadan	69.71	27.35	-1.16	1.72
4	Benin City	103.53	48.79	-0.25	0.78
5	Port-Harcourt	99.14	42.87	-0.64	0.82
6	Uyo	98.54	25.66	0.64	0.03
7	Calabar	111.73	49.47	-0.58	0.72
8	Onitsha	97.77	38.28	-0.70	1.21
9	Enugu	85.35	39.02	-0.33	0.70
10	Owerri	111.02	36.48	0.16	0.18

Applying these values in Table 2, to the equations for Parameter estimation by MOM in Table1, the values for ξ, α and k (location scale and shape parameters) respectively were obtained. These were fitted into the quantile functions in Table 1 to have the quantile rainfall values.

3.1 Goodness of fit test

The result of the Goodness of fit test for the distributions at the stations are shown in Tables 3. The Values in Table 3 were obtained from excel sheets of analysis of observed rainfall ($R_{o)}$. The corresponding computed quantile rainfall values Q(F) were for F= 0.999, 0.99, 0.9 and 0.5, 0.02, 0.01, 0.002, for each of the assumed probability distributions. The computation was done using equation 1 to 7.

Table 3: Goodness-of-Fit Test for the Distributions at the Locations

Station	GOF Test criteria	GOF Test values					
		EVI	GEV	GPA	LN	PIII	LPIII
	RMSE	7.44	12.34	14.23	7.71	5.28	4.70
	RRMSE	0.075	0.139	0.76	0.05	0.032	0.031
Ikeja	MADI	0.0076	0.0696	0.540	0.028	0.012	0.009
	MAE	25.76	35.78	98.62	31	22.5	21.02
	PPCC	0.9865	0.9890	0,7100	0.990	0.9965	0.9959
	CHI- SQ	23.03	66.84	1075	18.10	7.90	6.80
	D-Index	0.595	0.908	2.61	0.82	0.55	0.404
Enugu	RMSE	13.49	10.19	33.96	21.48	11.8	14.76
	RRMSE	2.67	1.46	7.16	2.41	1.57	1.81
	MADI	0.78	0.357	2.23	0.75	0.38	0.55
	MAE	35.9	18.5	76.6	74.1	28.43	47.62
	PPCC	0.9687	0.9661	0.9893	0.9393	0.9552	0.9945
	CHI- SQ	12.36	37.68	619.8	757	70.95	155.6
	D-Index	0.7986	0.827	3.41	2.855	1.074	1.598
	RMSE	15.97	12.1	42.9	19.89	8.79	16.27
Benin City	RRMSE	3.43	2.004	9.38	5.69	2.06	4.7
	MADI	1.028	0.50	3.85	1.8	0.59	1.48
	MAE	43.99	39.08	3.85	60.26	35	46.8
	PPCC	0.35	0.969	0.945	0.076	0.99	0.96
	CHI- SQ	187.1	71.8	80.18	479.2	64.1	326.7
	D-Index	0.806	0.45	3.6	0.411	0.4837	0.426

3.2 Assessment of Probability Distribution Models by Scoring Goodness- of- Fit Tests

The assessment of the probability distribution models was based on the total score obtained from all the tests. The test scores ranging from one to six are awarded to each distribution model based on the criteria that the distribution model with the highest score is chosen as the best distribution model for the data of a particular city. The distribution best supported by a test is awarded a score of six (6), the next best is awarded a score of five (5) and so on in descending order. Using the scoring scheme outlined, Tables 4 was computed. The overall ranking results are presented in Table 4.

Table 4: Scoring and Ranking Scheme for Distribution at the Locations

Station	Test Criteria	Distribution Score					
		EVI	GEV	GPA	LN	PIII	LPIII
Ikeja	RMSE	4	2	1	3	5	6
	RRMSE	3	2	1	4	5	6
	MADI	6	2	1	3	4	5
	MAE	4	2	1	3	5	6
	PPCC	2	3	1	4	6	5
	CHI-SQUARE	4	2	1	4	5	6
	D-INDEX	4	2	1	3	5	6
	Total score	26	15	7	24	35	40
	RANK	3rd	5th	6th	4th	2nd	1st
Enugu	RMSE	4	6	1	2	5	3
	RRMSE	2	6	1	3	5	4
	MADI	2	6	1	3	5	4
	MAE	4	6	1	2	5	3
	PPCC	4	3	5	1	2	6
	CHI- SQUQRE	1	6	2	3	5	4
	D-Index	6	5	1	2	4	3
	Total score	23	39	12	16	31	27
	Rank	4th	1st	6th	5th	2nd	3rd
Benin City	RMSE	4	5	1	2	6	3
	RRMSE	4	6	1	2	5	3
	MADI	4	6	1	2	5	3
	MAE	3	4	6	1	5	2
	PPCC	2	5	3	1	6	4
	CHI- SQUQRE	3	5	4	2	6	1
	D-Index	2	4	1	6	3	5
	Total score	22	35	17	16	36	21
	Rank	3rd	2nd	5th	6th	1st	4th

A detail study of Table 4 indicates that GPA performed poorly in the study area, taking the 6th position in Ikeja and Benin City while it took 5th in Enugu. This implied that GPA distribution model is not useful in the study area.

Ranking the total scores, a summary of the performance of the six probability models was obtained as shown in Table 5. It shows in detail the Best-fit-model and the second-best fit model for each location

Table 5: Goodness of fit test and the selected model for the peak rainfall

S/N	Location	Best Fit Model	Total Max Score	Second Best- Fit Model	Total Max Score
1	Ikeja	Log-Pearson III	40	Pearson III	35
2	Akure	Gumbel Extreme Value I	42	Pearson III	31
3	Ibadan	Pearson III	39	Generalized Extreme Value	33
4	Benin City	Pearson III	36	Generalized Extreme Value	35
5	Port Harcourt	Generalized Extreme Value	39	Log-Normal	31
6	Uyo	Extreme Value	36	Log-Normal	36
7	Calabar	Generalized Extreme Value	38	Pearson III	36
8	Onitsha	Generalized Extreme Value	40	Pearson III	29
9	Enugu	Generalized Extreme Value	39	Pearson III	31
10	Owerri	Generalized Extreme Value	39	Log-Pearson III	29

Table 5 shows that Generalized Extreme Value is the best-fit probability distribution model in five locations namely: Port Harcourt, Uyo, Calabar, Onitsha, Enugu and Owerri. This result is in harmony with that of Ghosh *et al*., (2016) and Alam *et al*., (2018) who found generalized extreme value distribution to be the most appropriate distribution of the monthly rainfall data for the three selected station in Bangladesh. However, Olufintoye *et al*., (2009) in similar research in the above locations recommended P111 and LP111 as GEV considered best fit in this study was not utilized by them.

Gumbel Extreme Value I is the best-fit probability distribution model in two locations namely: Akure and Uyo. Pearson III is the best-fit probability distribution model in two other locations namely: Ibadan and Benin City. Log-Pearson III is the best-fit probability distribution model for Ikeja. The model validation using Chi Square goodness of is presented in Table 6.

Table 6. Model validation Result with Chi- square Goodness of fit Statistic

Location	Best-Fit Distribution	Test Performed	Calculated values for χ2 test	Degree of freedom	Critical values at 5% significance level	Decision
Ikeja	Log-Pearson III	Chi- square	6.80	49	66.326	Accept Ho
Akure	Gumbel Extreme Value I	Chi- square	9.68	49	66.326	Accept Ho
Ibadan	Pearson III	Chi- square	56.14	49	66.326	Accept Ho
Benin City	Pearson III	Chi- square	64.1	49	66.326	Accept Ho
Port Harcourt	Generalized Extreme Value	Chi- square	10.39	49	66.326	Accept Ho
Uyo	Extreme Value	Chi- square	3.35	49	66.326	Accept Ho
Calabar	Generalized Extreme Value	Chi- square	14.8	49	66.326	Accept Ho
Onitsha	Generalized Extreme Value	Chi- square	7.94	49	66.326	Accept Ho
Enugu	Generalized Extreme Value	Chi- square	37.68	49	66.326	Accept Ho
Owerri	Generalized Extreme Value	Chi- square	12.4	49	66.326	Accept Ho

The χ2 test statistics in Table 6 were Calculated using equation 6. The Critical values of χ2 at 5% significance level were interpolated in Chi- square Table in Standard textbooks. The Implication of the decision column of the table is that we have to accept the null hypotheses ($H_{o):}$ that the Probability Distribution Model fits the data. This has validated our models outlined in the table and therefore they can be reliably used for future rainfall prediction as was carried out in Table 7.

Table 7: Forecasted Rainfall (mm) for Varying Return Periods at Each Location.

S/NO	Location	Best Fit Distribution	Return Period (Years) 5	10	25	50	100	200	500
1	Ikeja	LPIII	132.03	158.11	216.3	222.09	252.21	284.31	330.16
2	Akure	EVI	103.58	117.85	134.39	149.27	162.54	175.78	193.23
3	Ibadan	PIII	92.85	98.25	104.03	107.90	109.9	111.41	112.85
4	Benin City	PIII	142.2	164.41	183.60	197.38	208.4	218.2	229.72
5	Port Harcourt	GEV	143.7	159.17	171.90	179.11	182.43	185.00	188.1
6	Uyo	EVI	116.99	131.98	149.39	165.03	179.99	192.92	211.27
		LN	115.33	127.5	138.52	152.05	161.81	171.28	183.51
7	Onitsha	GEV	132	145	151	153.94	160.42	162.52	164.35
8	Enugu	GEV	112	135	148.57	155.95	161.68	165.88	169.95
9	Owerri	GEV	144.14	161.63	178.89	191.33	200.98	209.23	218.35
10	Calabar	GEV	150.33	174.04	199.86	216.42	230.95	243.77	258.46

The best fit probability distribution shown in Table 5 was used to compute the Quantile values in Table 7. The results of the various analyses culminating in the selection of the best fit probability distribution model for each station and the rainfall return levels (mm) for selected return periods of between 5years and 500 years are as presented in Table 7.

3.2 *Rainfall Frequency Curves (RFCS)*

The AMR estimates obtained for the probability distribution models in Table 7 were used to develop the RFCs. These were implemented by HEC-SSP Software, and they are presented in Figures 2-4.

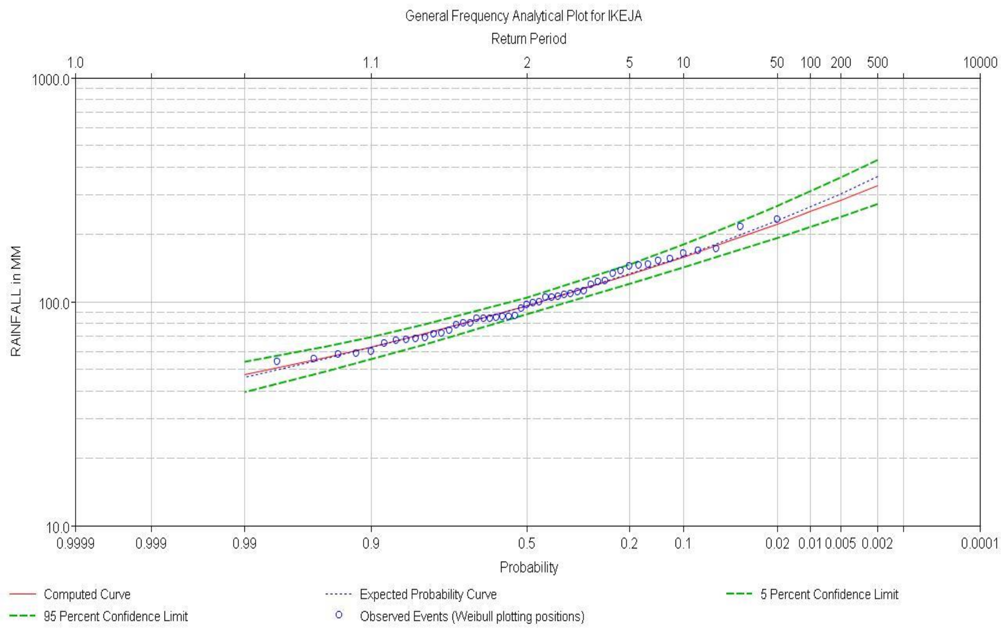

Figure 2 Rainfall frequency curve (RFC) for Ikeja

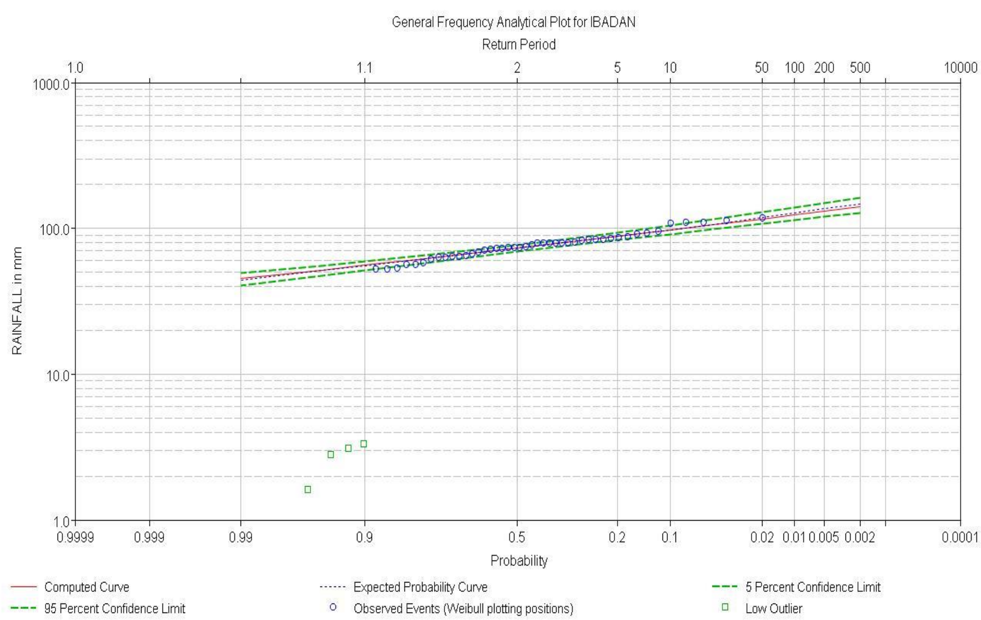

Figure 3: Rainfall frequency curve (RFC) for Ibadan

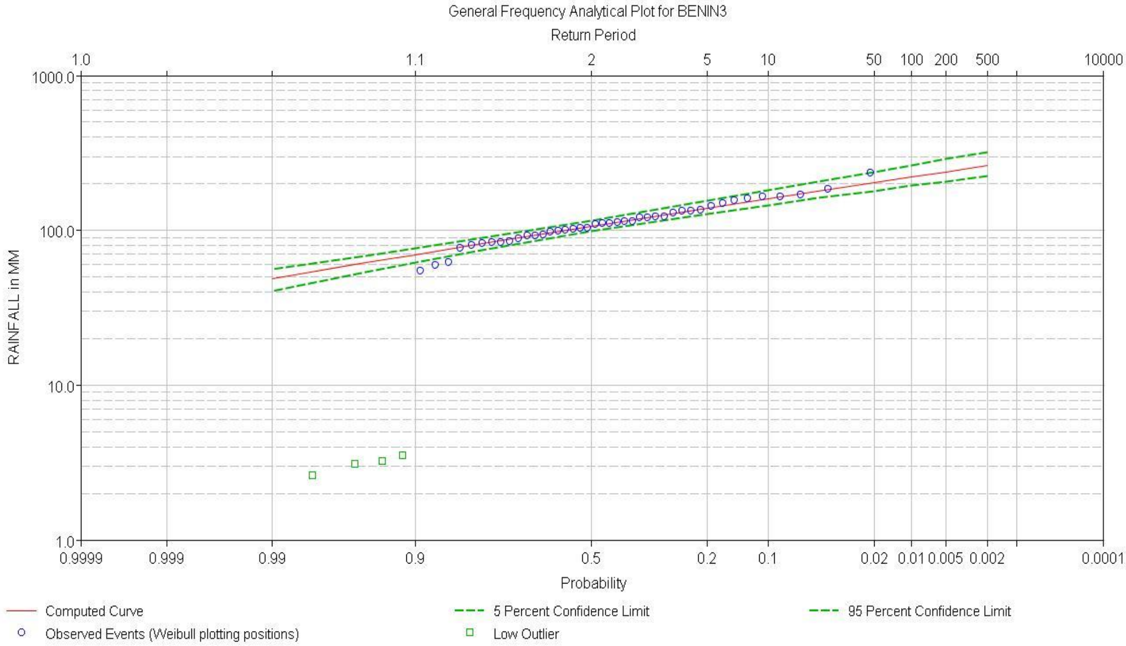

Figure 4: Rainfall Frequency Curve (RFC) for Benin City

These Rainfall Frequency Curves (RFC) which are representatives of the ten curves in the study area are useful and powerful tools for estimating the rainfall frequency distribution and calculating T-years rainfall . They are also useful in the planning, design and management of hydraulic structures for flood mitigation and prevention of flood damage at the locations.

These Rainfall Frequency Curves (RFC) contain five main lines which are:

1) The zero-line representing the observed events on Weibull plotting positions.
2) Red line representing the computed curve.
3) Black dotted line representing the expected probability curve.
4) The upper and lower green lines representing 95% and 5% confidence limits.
5) The square line representing the outliers.

In figure 2,3 and 4, the curves are near linear depicting a continuous increase of rainfall values with increasing return periods. The computed curves almost merged with the expected probability curves which are the best -fit curves for each location. These confirm that the curves adequately represent the most appropriate distribution model for each location and could be used for forecasting of the rainfall at the required return periods Almost all the points in the plots were withing the 95% confidence interval There were low outliers (4 each) for Ibadan and Benin city data and none for Ikeja data.

4. Conclusions

In this study, the best fit probability distribution model applicable to each location was selected for the cities in Southern Nigeria from six probability distributions and used to predict rainfall return levels of engineering importance for the location. Based on the results the following conclusions are made:

1. Generalized Extreme value distribution (GEV) model is the best-fit probability model for Onitsha, Enugu, Owerri, Calabar, and Port Harcourt. EVI and LN are the best-fit probability models for Uyo. Also, the best-fit probability models are EVI and LP111 for Akure and Ikeja respectively. The Quantile estimates or forecasted rainfall values for different return periods as shown in Table 6 are considered adequate design parameters for planning, design of hydraulic structures for flood mitigation and flood precautions in the various locations.

2. Rainfall frequency curves have been provided for Ikeja, Ibadan and Benin City. They were drawn with the forecasted rainfall values that were derived from the best fit probability distribution models. These are guides in hydrological designs.
3. This study has provided useful Engineering Design Parameters for Planning and improved Hydrological design needed for Efficient Hydraulic design of Structures needed for flood control, mitigation, and flood precaution in the various location. Utilization of these parameters by engineers will greatly reduce or advert design failures.

References

Agbonaye, AI. and Izinyon, OC. 2017. Best-Fit Probability Distribution Model for Rainfall Frequency Analysis of Three Cities in South Eastern Nigeria, Nigerian Journal of Environmental Sciences and Technology (NIJEST), 1(1): 34 – 42.

Alam, A., Emura, K., Farnham, C. and Yuan, J. 2018. Best-Fit Probability Distributions and Return Periods for Maximum Monthly Rainfall in Bangladesh, Climate, 6(1): 9; https://doi.org/10.3390/cli6010009

Amin, MT., Rizwan, M. and Alazba, AAA. 2016. Best-fit probability distribution for the estimation of rainfall in northern regions of Pakistan. Open Life Science, 11: 432–440.

Filliben, JJ. 1975. The Probability Plot Correlation Coefficient Test for Normality. Technimetrics, 17(1): 111-117.

Ghosh, S., Roy, M K. and Biswas, SC. 2016. Determination of the Best Fit Probability Distribution for Monthly Rainfall Data in Bangladesh. American Journal of Mathematics and Statistics 6(4): 170-174.

Kousar, S., Khan, AR., Hassan, M., Noreen, Z. Bhatti, SH. 2020. Some best-fit probability distributions for at-site flood frequency analysis of the Ume River. Journal of Flood Risk Management, 13(e12640): 1-11.

Koutsoyiannis, D. 1997. Statistical Hydrology Edition 4, Chapter 6. National Technical University of Athens, Athens, pp: 131-175, doi:10.13140/RG.2.1.5118.2325.

Kumar, A. 2000. Prediction of annual maximum daily rainfall of Ranichauri (Tehri Garhwal) based on probability analysis. Industrial Journal of Soil Conservation, 28: 178–180.

Langat, PK, Kumar, L. and Koech, R. 2019. Identification of the Most Suitable Probability Distribution Models for Maximum, Minimum, and Mean Streamflow. Water, 11(4): 734. doi:10.3390/w11040734

Mamman, MJ., Martins, OY., Ibrahim, J. and Shaba, M. 2017. Evaluation of Best-Fit Probability Distribution Models for the Prediction of Inflows of Kainji Reservoir, Niger State, Nigeria. Air, Soil and Water Research, 10: 1-7.

Masereka, EM., Otieno, FAOGM. and Snyman, OJ. 2015. Best Fit and Selection of Probability Distribution Models for Frequency Analysis of Extreme Mean Annual Rainfall Events. International Journal of Engineering Research and Development, 11(04): 34-53.

Mohamed, TM. and Ibrahim, AAA. 2016. Fitting Probability Distributions of Annual Rainfall in Sudan. SUST Journal of Engineering and Computer Sciences (JECS), 17(2): 34-39.

Okoloye, CU., Aisiokuebo, N., Ukeje, J., Anuforom, AC. and Nnodu, I. 2013. Rainfall Variability and the Recent Climate Extremes in Nigeria. NIMET, National Weather Forecast, and climate research center. Journal of Meteorology and Climate Science, 11: 49-57.

Olofintoye, OO., Sule, BF. and Salami, AW. 2009. Best–fit probability distribution model for peak daily of selected cities in Nigeria. New York Science Journal, 2(3): 1-12.

OSGOF, 2011. Map of Nigeria Showing Ten Selected Cities. Office of the Surveyor General of the Federation, Abuja, Nigeria.

Singh, B., Rajpurohit, D., Vasishth, A. and Singh, J. 2012. Probability Analysis for Estimation of Annual One Day Maximum Rainfall of Jhalarapatan Area of Rajasthan. Plant Archives, 12: 1093-1100.

Tao, DQ., Nguyen, VT. and Bourque, A. 2002. On selection of probability distribution for representing extreme Precipitations in southern Quebec. Annual conference of the Canadian Society for Civil Engineering, 5-8 June, Montreal, Quebec, Canada, pp 1 – 8.

Vivekananandan, N. 2014. Comparison of probability distributions for frequency analysis of annual maximum rainfall. International Journal of Research and Innovative Technology, 1(3): 50–55.

WMO 2009. WMO Report No 168. Guide to Hydrological Practices Volume II: Pp 112 Management of Water Resources and Application of Hydrological Practices. World Meteorological Organization, Geneva