Journal of Energy Technology and Environment Vol. 6(4) 2024 pp. 58 - 70 ISSN-2682-583x

NIPES

Journal of Energy Technology and Environment

Journal homepage: www.nipesjournals.org.ng

Best Fit Probability Distribution /Parameter Estimation Procedure for Precipitation Frequency Analysis at Some Stations in Southern Nigeria

Okagbare, O. A*., Izinyon, O. C. and Agbonaye, A. I.
Department of Civil Engineering, University of Benin, Benin City, Nigeria
*Corresponding Author: fadaokagbare@gmail.com

Article information

Article History
Received 24 September 2024
Revised 30 October 2024
Accepted 8 November 2024
Available online 19 Dec 2024

Keywords:
Probability Distribution; Parameter Estimation; Precipitation Frequency Analysis; Method of Moments; L-Moments.

OpenAIRE

https://doi.org/10.5281/zenodo..14523597

https://nipes.org

Abstract

This study investigates the best-fit probability distribution model and parameter estimation procedure namely EV1-MOM, EV1-MOM, GEV-LM, GPA-LM and GLO-LM suitable for at-site precipitation frequency analysis at rainfall gauging stations in Ikeja, Owerri and Port Harcourt in Nigeria. The best fit probability distribution - parameter estimation procedure at each station was selected based on a scoring and ranking scheme using the relative magnitudes of five goodness of fit measures namely: root mean square error (RMSE), relative root mean square error (RRMSE), mean absolute deviation index (MADI), maximum absolute error (MAE) and probability plot correlation coefficient (PPCC) obtained for the procedures. The results of the analyses indicate that the Generalized Extreme Value - L-Moments model (GEV-LM) procedure which is highest ranking is the best fit distribution model / estimation procedure for Ikeja and Port Harcourt stations. While Extreme Value type 1 – Method of Moments (EV1 MOM) is the best fit distribution model / estimation procedure for Owerri station. They were subsequently used to predict rainfall return values for return periods of engineering importance (from T = 2 years to 200 years) at the study stations. The obtained values were used to develop rainfall frequency curves for the stations. The rainfall frequency curves developed for the stations are useful for design of engineering infrastructure for flood mitigation at the studied stations and in addition, the results provide valuable insights for hydrological analysis and policymaking, particularly in addressing the escalating risk of extreme flood events in the southern Nigerian cities due to changing precipitation patterns and climate impacts

1. Introduction

Extreme precipitation events such as heavy rainfall at the southern part of Nigeria, pose significant risks to human society, causing destructive floods that lead to the loss of lives, properties, and crops, and contributing to waterborne diseases affecting humans, plants, and animals [1]. These events are becoming increasingly critical due to climate change, which has heightened the frequency and intensity of such occurrences. Effective hazard assessment and risk management are essential to mitigate these impacts, necessitating a detailed understanding of extreme rainfall patterns [2].

This knowledge is crucial for designing hydraulic structures like dams and urban drainage systems, which help reduce economic losses and prevent flood damage [3]. Accurate information on flood magnitudes and frequencies is vital for public policy, water resources management, and agricultural planning, making the study of extreme precipitation events highly significant [4].

It is known that frequency analysis of floods and extreme precipitation events involves using probability distributions to relate the magnitude of these events to their frequency of occurrence [5]. This analysis is essential for water resources planning, hydraulic design, and mitigating socio-economic impacts associated with climate variability. Various statistical models and distributions, such as two- and three-parameter models, are used to estimate the magnitude and probability of extreme rainfall events. However, there is no consensus among hydrologists on the best distribution to use, as hydro-climatic regimes differ by region. The selection of suitable models and parameter estimation procedures remains a key area of research, with challenges arising from the random nature of extreme events and the lack of long-term data records, particularly in developing countries such as Nigeria [6]. Thus, while significant progress has been made in understanding and modelling these events, uncertainties persist in accurately predicting their occurrence and impact.

The hydrological challenge lies in selecting the most appropriate probability distribution and parameter estimation technique from the numerous options available in the literature that are suitable for a specific location. Therefore, it is essential to evaluate a variety of probability distribution models and parameter estimation techniques to identify the most accurate method for estimating extreme rainfall at a given site. This research aims to investigate the use of widely adopted two- and three-parameter probability distribution models, along with two parameter estimation techniques (Method of Moments and L-Moments), for at-site precipitation frequency analysis at five rainfall gauging stations in Southern Nigeria. Consequently, the study seeks to determine the best-fitting probability distribution model and parameter estimation technique for the precipitation frequency analysis.

2. Methodology

2.1 Data Collection and Processing

This study utilized time series data of annual maximum daily rainfall at rainfall gauging stations in Ikeja, Owerri and Port Harcourt respectively (Figure 1). Stations with at least 30 years of data close to the present time were selected to ensure a stable distribution for estimating future rainfall probabilities [7]. A record length of 25 years was suggested as sufficient for extreme precipitation analysis in humid regions, as the required length is related to the area's general humidity and physiographic conditions [8]. The annual maximum daily rainfall series was created by extracting the maximum daily rainfall for each year from the daily data, resulting in as many extreme values as the total number of years recorded. The daily rainfall data for these stations was obtained from the Nigeria Meteorological Agency (NIMET) and processed using the MS-EXCEL program to compile the annual maximum rainfall series, representing the highest rainfall received in a single day each year.

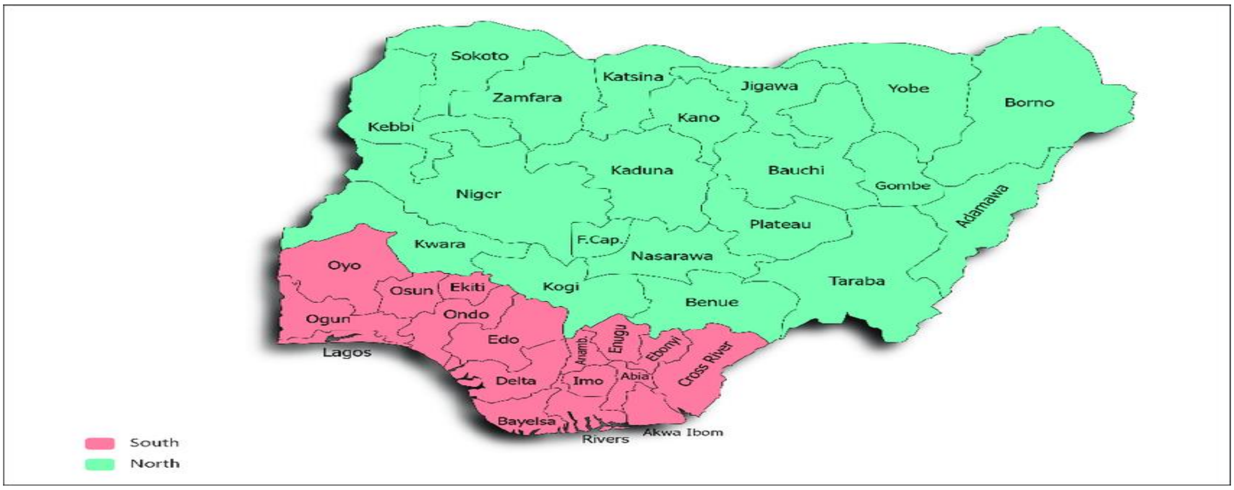

Figure 1: Map of Southern Nigeria Source: [9]

2.2 Preliminary Data Analysis

2.2.1 Descriptive Statistics

The applicable formulae for computing descriptive statistics was utilize to compute same from annual daily maximum rainfall at each station (Mean, Standard deviation, Skewness, Kurtosis and coefficient of variation). The MS-Excel software programme was used for computations and the formula is as presented in Table 1.

Table 1: Equations for Descriptive statistics

S/N	Statistics Text	Applicable Formula	Remark
1	Mean	$\bar{R} = \frac{\Sigma(R)}{N}$	It is the total of all the variates, divided by the total number of variates.
2	Standard Deviation	$б = \sqrt{\frac{\Sigma(R-\bar{R})^2}{N}}$	It is the square root of the mean-squared deviation of individual observations from their mean Standard deviation.
3	Variance	$б^2 = \frac{\Sigma(R-\bar{R})^2}{N}$	
4	Skewness	$\alpha = \frac{1}{N}\ \Sigma(R-\bar{R})^3$	Where R is a variate, $\bar{R}$ is the mean of the data set and N is the total number of variates
5	Kurtosis	$K = n\,\frac{\sum_{i=1}^{n}(R_i-\bar{R})^4}{\left(\sum_{i=1}^{n}(R_i-\bar{R})^2\right)^2}$	Kurtosis indicates how peaked a distribution is, usually considered relative to a normal distribution. Leptokurtic is a measure of a normal distribution which is not very peak. Also isocratic is defined as a normal distribution that is not very peaked or very flat-topped

2.2.2 Homogeneity Test

The data used for precipitation frequency analysis is required to be homogeneous and independent [10]. Homogeneity of data assumes that the observations are from the same population. This test of homogeneity to be adopted for this study is that proposed in [11]. This test of homogeneity is based on cumulative deviations from the mean as given below:

$$S_K = \sum_{i=1}^{k}(X_i - \bar{X}) \quad (1)$$

k =1, -, -. ----- -, n

Where X_i are the records from the series and $\bar{X}$ is the mean. The initial value and the last value are equal to zero. To test the homogeneity of the data set, the cumulative deviations are often rescaled by dividing the standard deviation value (S). By evaluating the maximum (Q) or the range (R) of the rescaled cumulative deviation from the mean, the homogeneity of the data of a time series can be tested.

$$Q = \max\left[\frac{S_K}{S}\right] \tag{2}$$

$$R = \max\left(\frac{S_K}{S}\right) - \min\left(\frac{S_K}{S}\right) \tag{3}$$

High values of Q or R indicate the data of the time series do not come from the same population. The homogeneity of the annual maximum rainfall data at the stations will be tested using the Rainbow software.

Rainbow is a software tool that analyses agro-meteorological and hydrological records using frequency analysis and tests the homogeneity of the record [12].

2.3 Precipitation Frequency Analysis Procedure

The procedure utilized in the precipitation frequency analysis are as follows below.

a. The parameters of the probability distributions (EV1, GEV, GPA, and GLO) are estimated using MOM, and L-Moments as designed for the study. Before computation of L-Moments, the Probability weighted moments (PWMs) are determined.
b. The adequacy of fitting the different probability distribution/ parameter estimation procedures to data at each station, was determined by use of Goodness of Fit and Diagnostic tests and the best fit distribution-parameter estimation procedure was selected.
c. By means of the selected best fit probability distribution/ parameter estimation procedure determines precipitation magnitudes of recurrence intervals of engineering importance (T = 2 years, 5 years, 10 years, 25years, 50 years, 100 years, 200 years) at each station.
d. Construct Rainfall frequency curves at each station based on the best procedure at each station.

2.3.1 Probability Weighted Moments (PWMs) and L-Moment Equations

PWMs are needed for calculation of L-Moments. The dataset was first arranged in ascending order and PWMs then computed by use of the following equations [13].

1) The probability weighted moments, b, are estimated for a given data set by use of the following equations:

$$b_0 = \frac{1}{N}\sum_{J=1}^{N} Q_j \quad \text{(That is the sample mean)} \tag{4}$$

$$b_1 = \frac{1}{N}\sum_{J=2}^{N} \frac{(j-1)}{(N-1)} Q_j \tag{5}$$

$$b_2 = \frac{1}{N}\sum_{J=3}^{N} \frac{(j-1)\,(j-2)}{(N-1)(N-2)} Q_j \tag{6}$$

$$b_3 = \frac{1}{N}\sum_{J=4}^{N} \frac{(j-1)\,(j-2)(j-3)}{(N-1)(N-2)(N-3)} Q_j \tag{7}$$

Where Q_j is the jth element of a sample of annual maximum series (precipitation in the case of this study) arranged in ascending order and N is the sample size (the number of annual maxima in the record)

2) The sample L-moments are defined as:

$$l_1 = b_0 \tag{8}$$

$$l_2 = 2\,b_1 - b_0 \tag{9}$$

$$l_3 = 6b_2 - 6\,b_1 + b_0 \tag{10}$$

$$l_4 = 20b_3 - 30b_2 + 12\,b_1 - b_0 \tag{11}$$

3) The sample L-moment ratios are defined as:

$$L - CV = t_2 = \frac{l_2}{l_1} \quad (12)$$

$$L - skewness = t_3 = \frac{l_3}{l_2} \quad (13)$$

$$L - kurtosis = t_4 = \frac{l_4}{l_2} \quad (14)$$

2.4 Determination of Best Fit Distribution/ Parameter Estimation Procedure

Scoring and Ranking Scheme for Selection of Best Fitting Distribution at a Station

The best fit distribution - parameter estimation procedure at a station was selected based on a scoring and ranking scheme. Ranking of the distributions at a station will be based on the relative magnitude of the results of goodness of fit measures through the formula presented in Table 4. The distribution with the lowest RMSE, lowest RRMSE, lowest MADI, lowest MAE, or highest PPCC at a station was considered as the best fitting distribution with respect to the test criteria [14]. The five performance measures were used in this study and the best procedure at a station was assigned a score of 5, the next best will be given the score 4 and, in that order, while the worst will be given the score 1. The overall score of each distribution/ parameter estimation procedure at a location based on all GOF test criteria will be obtained by summing the individual point scores of the different tests at the stations and the procedure with the highest total score at each station was judged as the best fit distribution model / estimation procedure for the station.

Table 4: Summary of Goodness of Fit (GOF) Tests

S/N	Test	Abbreviation	Mathematical Equations
1	Root Mean square error	RMSE	$RMSE = \left(\frac{\sum (R_o - R_f)^2}{(n-m)} \right)^{\frac{1}{2}}$
2	Relative Root Mean square error	RRMSE	$RRMSE = \left(\frac{\sum \left(\frac{R_{oi} - R_{fi}}{R_o} \right)^2}{(n-m)} \right)^{\frac{1}{2}}$
3	Mean absolute deviation index	MADI	$MADI = \frac{1}{N} \sum_{i=1}^{N} \left\| \frac{R_o - R_f}{R_o} \right\|$
4	Maximum absolute error	MAE	$MAE = \max\left(\left\| R_o - R_f \right\| \right)$
	Probability plot correlation coefficient	PPCC	$PPCC = \frac{\sum \left\{ (R_{0i} - R_m)(R_{fi} - \overline{R_{fm}}) \right\}}{\left[\sum (R_{0i} - \overline{R}_m)^2 \sum (R_{fi} - \overline{R_{fm}})^2 \right]^{\frac{1}{2}}}$

2.5 Application of Best Distribution /Parameter Estimation Procedure for Rainfall Return Levels at Stations

The selected best fit distribution/ parameter estimation procedure at each station determined by scoring and ranking procedure will be used to predict rainfall return levels (R_T) at each station. A return level of rainfall (R_T) with return period T years is the level exceeded on average once in every T year [5]. Estimates of R_T through the formula presented in Table 5, were useful in expressing the degree of hazard related to extreme precipitation at a station. The return period or recurrence interval of interest for the scope of this study are T = 2, 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 75, 100 and 200 years and they would be computed by applying the best fitting procedure at each station.

Table 5: Formulae for Estimating the Rainfall Return Levels (R_T)

Distributions	R_T
EV1	$R_T = \xi - \alpha \ln(-\ln F)$
GEV	$R_T = \xi + (-\ln F)^{-k} - 1$ $F = \xi + \frac{\alpha}{k}(-\ln F)^{-k} - 1$
GLO	$R_T = \xi + \frac{\alpha}{K}\left[1 - \left(\frac{\{1-F\}}{F}\right)^K\right]$
GPA	$R_T = \mu + \frac{\alpha}{K}\left[1 - \left(\frac{\{1-F\}}{F}\right)^K\right]$

2.6 Development of Rainfall Frequency Curves

Rainfall Frequency Curves (RFCs) which is a function of the annual 1-day maximum rainfall and return periods or recurrence intervals was developed using Microsoft EXCEL and was an output of this study. This was achieved by plotting a graph of the forecasted rainfall intensity from the best rainfall distribution model, against return periods of 2, 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 75, 100 and 200 years at each station. By linking rainfall intensity with the likelihood of occurrence over different time frames, these curves provide valuable insights into precipitation patterns and can help in assessing flood risks and designing appropriate infrastructure [11].

2.7 Model Validation Using Kolmogorov-Smirnov (K-S) Test

The best rainfall distribution model for the ten substations was validated using Kolmogorov-Smirnov (K-S) test with the aid of *RAINBOW software* [12]. The Kolmogorov-Smirnov test is a non-parametric test used to compare the empirical cumulative distribution function (ECDF) of a sample against a given theoretical distribution or to compare two samples to determine if they are drawn from the same distribution. The procedure to carry out this test include arranging the data in ascending order. The Empirical Cumulative Distribution Function was then calculated by equation 15.

$$ECDF = \frac{Rank}{Numbers\ of\ observation} \quad (15)$$

The expected CDF values using the NORM.DIST function was then calculated using the Excel spreadsheet. The mean and the standard deviation of the data determined. The absolute differences between the ECDF and the theoretical CDF were calculated. The K-S statistic (D) is the maximum value. The critical value for the K-S test depends on the significance level and the sample size was then compared with the maximum value. The critical value is given by equation 16.

$$D_{critical} \approx \frac{1.36}{\sqrt{n}} \quad (16)$$

If the D-statistic is greater than the critical value, the null hypothesis that the data follows the specified distribution will be rejected.

3.0 Results and Discussion

3.1 Descriptive Statistics of Rainfall Data

Table 6 provides an overview of the descriptive statistics for daily rainfall data collected from rainfall gauging stations in Akure, Calabar, Ikeja, Owerri and Port Harcourt (PH) respectively, with data sourced from the Nigeria Meteorological Agency (NIMET).

Table 6: Descriptive Statistics of the NIMET Rainfall (1965 – 2012)

	Mean	St. Dev	Variance	Min.	Max.	Range	Kurtosis	Skewness
Ikeja	97.97	58.18	3385.38	0.00	237.30	237.30	0.19	0.26
Owerri	91.89	55.37	3065.96	0.00	206.60	206.60	-0.61	-0.41
PH	99.06	43.06	1854.19	0.00	185.30	185.30	0.84	-0.65

The descriptive statistics results reveals that the mean rainfall varies, with Port Harcourt recording the highest average of 99.06 and Owerri the lowest of 91.89. The standard deviation ranges from 43.06 to 58.18, signifying variability in the data. Ikeja has the highest maximum rainfall of 237.30 and the greatest variance of 3385.38. The skewness of the rainfall distribution is generally negative, indicating left skewness, except in Ikeja. Kurtosis values also vary, with Port Harcourt having the highest of 0.84, suggesting a moderately peak, and Owerri the lowest of -0.61, indicating a flatter distribution.

A normal distribution typically has skewness close to 0 and kurtosis around 3 [15]. The rainfall data from Ikeja exhibit skewness and kurtosis values that suggest they approximate a normal distribution fairly well, with only minor deviations. Meanwhile, the rainfall data from Owerri, and Port Harcourt show moderate deviations, characterized by noticeable skewness and varying degrees of peakedness and flatness.

3.2 Homogeneity Test

The homogeneity of the annual maximum rainfall data at each station was tested using the Rainbow software.
The Figure 2 below shows that the annual maximum rainfall data from Ikeja station is all from the same population at 99% probability level only.

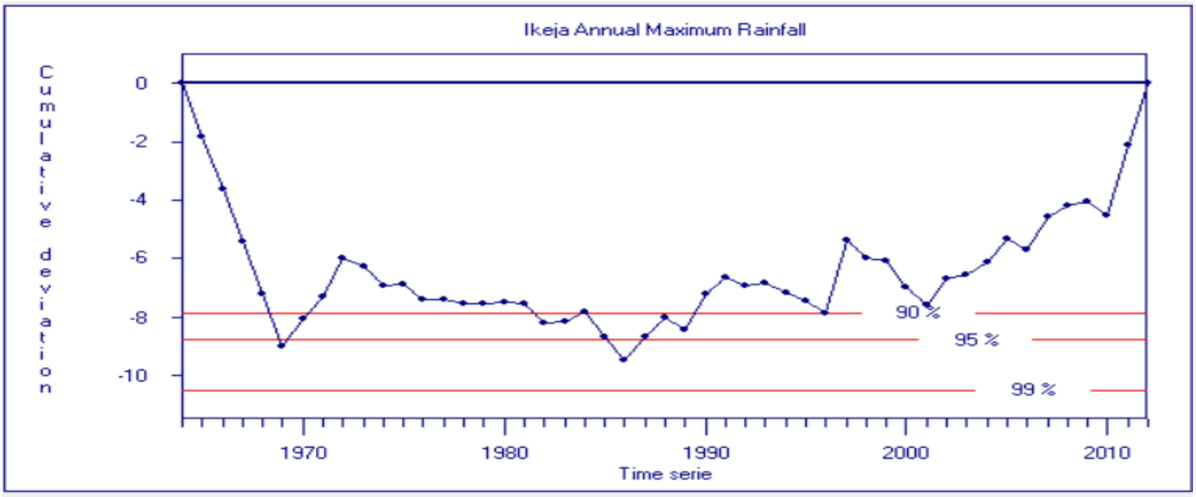

Figure 2: The Homogeneity Statistics of Ikeja Rainfall Data

Figure 3 below shows that the range of cumulative deviation and maximum cumulative deviation evaluated, could not be rejected at any of the (90%, 95%, and 99%) probability levels, demonstrating that the annual maximum rainfall data from Owerri station is all from the same population.

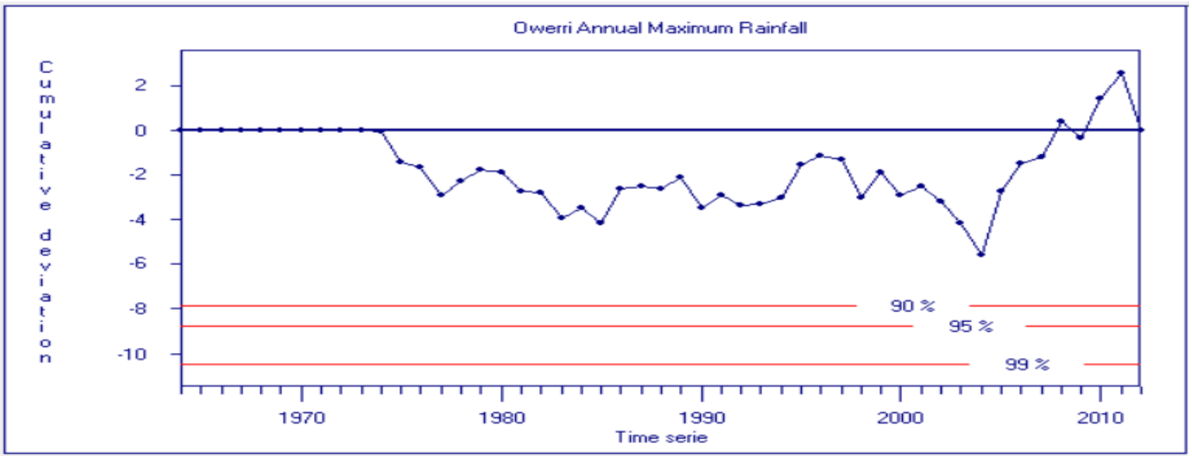

Figure 3: The Homogeneity Statistics of Owerri Rainfall Data

The Figure 4 below shows that the annual maximum rainfall data from Port Harcourt station is all from the same population at 99% probability level only.

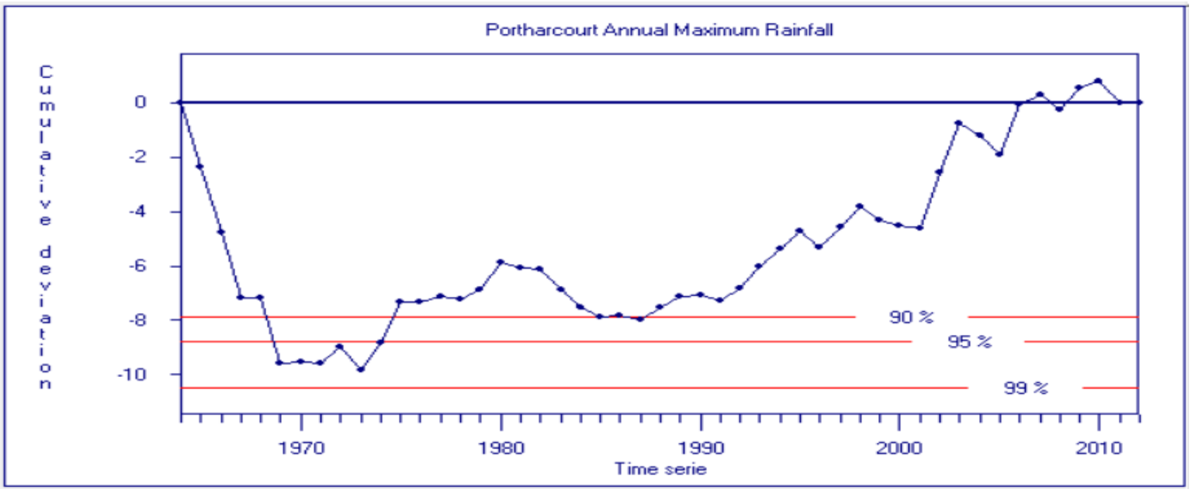

Figure 4: The Homogeneity Statistics of Port Harcourt Rainfall Data

3.3 Probability Distribution by MOM and L-Moments

To evaluate the L-Moments, the probability weighted moments, b, were first estimated for the given data set at the various sub stations as presented on Table 7. Afterwards, the sample L-Moments and ratios were estimated as presented on Table 8.

Table 7: Sample Probability Weighted Moments of the Study Stations

Station	Sample Probability Weighted Moments			
	b_0	b_1	b_2	b_3
Ikeja	102.23	70.65	55.85	47.38
Owerri	113.09	93.69	86.49	85.96
Port-Harcourt	101.16	63.85	47.24	38.11

Table 8: L-Moments and L-Moment Ratios of the Study Stations Data

Station	Sample L-Moments				L-Moment Ratios		
	L_1	L_2	L_3	L_4	L-C_v (τ_2)	L-C_s (τ_3)	L-C_k(τ_4)
Ikeja	102.23	39.08	13.39	17.89	0.38	0.34	0.46
Owerri	113.09	74.29	69.87	135.76	0.66	0.94	1.83
Port-Harcourt	101.16	26.54	1.51	10.01	0.26	0.06	0.38

From Table 8, it was observed that the L-coefficient of variation ranged between 0 and 1 for all stations. Additionally, the L-coefficient of skewness was less than 1 for most stations. The L-coefficient of kurtosis ratios for Ikeja and Port Harcourt were less than 1, while Owerri were greater than 1. These results imply that the majority of the stations conform to the findings proposed by Hosking [16], who stated that the L-coefficient of skewness and kurtosis are typically less than 1, and the L-coefficient of variation lies between 0 and 1.
Afterwards, the parameters of the distribution either by method of moments (MOM) and/ or by L-Moment's method were estimated as presented in Table 9.

Table 9: Probability Distribution Parameters

		IKEJA	OWERRI	PH
EV1/MOM	**α**	42.87	28.08	31.58
	ε	77.48	96.89	82.93
EV1/LMO	**α**	129.81	246.80	88.17
	ε	27.30	-29.36	50.27
GEV/LMO	**z**	-0.03	-0.12	0.02
	k	-0.25	-0.92	0.19
	α	1.34	0.52	0.68
	ε	101.02	106.51	100.88
GPA/LMO	**k**	-0.02	-0.94	0.79
	α	75.73	4.84	131.98
	ε	24.88	34.24	27.24
GLO/LMO	**k**	-0.34	-0.94	-0.06
	α	36.08	9.09	26.49
	ε	93.48	43.76	100.29

3.4 Statistical Goodness of Fit Measures on the Probability Distribution Model

Table 10 show the performance or accuracy of the probability distribution model/ parameter estimation procedure at each station using statistical goodness of fit measures namely: root mean square error (RMSE), relative root mean square error (RRMSE), probability plot correlation coefficient (PPCC), maximum absolute error (MAE), and mean absolute deviation index (MADI) tests.

Table 10: Results of goodness of fit tests

Station	Dist. Model	RMSE	RRMSE	MADI	MAE	PPCC
Ikeja	EV1 MOM	11.177	0.311609	9.75126	19.18864	0.286
	EV1 LMO	33.8425	0.311609	29.5256	58.10068	0.2583

Station	Dist. Model	RMSE	RRMSE	MADI	MAE	PPCC
	GEV LMO	2.5071	0.311299	2.20081	4.0435	0.33
	GPA LMO	48.3419	0.31179	42.2384	82.1588	0.2074
	GLO LMO	77.2223	0.310981	67.8126	127.595	0.1533
Owerri	EV1 MOM	7.3195	0.311609	6.38552	12.56612	0.3058
	EV1 LMO	64.3404	0.311609	56.13318	110.4594	0.2174
	GEV LMO	21.8673	0.308432	18.8826	42.26272	0.2147
	GPA LMO	99.6846	0.30835	86.0052	193.337	0.04699
	GLO LMO	188.6477	0.30833	162.752	365.93	0.0754
Port Harcourt	EV1 MOM	8.234129	0.311608	7.18379	14.13645	0.2977
	EV1 LMO	22.98019	0.311609	20.05408	39.46258	0.2542
	GEV LMO	0.179438	0.311525	0.155492	0.32364	0.34914
	GPA LMO	4.81682	0.31026	4.1322	9.93757	0.32726
	GLO LMO	19.82348	0.311402	17.3179	33.6498	0.26173

In Table 10, the analysis of goodness-of-fit for various distribution models applied to data from different weather stations reveals distinct performance metrics.
Examining results from the Ikeja station, the GEV LMO model demonstrated the best fit, with the minimum RMSE, MADI, MAE and highest PPCC values at 2.5071, 2.20081, 4.0435 and 0.33 respectively. The GLO LMO model yielded the least RRMSE at 0.310981.
While at Owerri station, the EV1 MOM model demonstrated the best fit, with the minimum RMSE, MADI, MAE and highest PPCC values at 7.3195, 6.38552, 12.56612 and 0.3058 respectively. The GLO LMO model yielded the least RRMSE at 0.30833.
Moving on to the Port Harcourt station, the GEV LMO model showcased the most favourable performance, with the lowest RMSE, MADI, MAE and highest PPCC values at 0.179438, 0.155492, 0.32364, and 0.34914 respectively. The GPA LMO model had the minimum RRMSE of 0.31026.
Table 11 presents the ranking of the models based on the scoring of the results of the different models.

Table 12: Ranking of Distribution Models at Stations

Station	Dist. Model	Total	Rank
Ikeja	EV1 MOM	19	2nd
	EV1 LMO	15	3rd
	GEV LMO	24	1st
	GPA LMO	10	4th
	GLO LMO	9	5th
Owerri	EV1 MOM	22	1st
	EV1 LMO	15	3rd
	GEV LMO	18	2nd
	GPA LMO	11	4th
	GLO LMO	10	5th
Port Harcourt	EV1 MOM	14	3rd
	EV1 LMO	5	5th

Station	Dist. Model	Total	Rank
	GEV LMO	23	1st
	GPA LMO	21	2nd
	GLO LMO	12	4th

Table 12: Summary of Best Fit Probability Distribution-Parameter Estimation Procedure at Stations

Station	Best fit Probability Distribution-Parameter Estimation Procedure
Ikeja	GEV - LMO
Owerri	EV1 - MOM
Port Harcourt	GEV - LMO

Table 12 shows that GEV LMO is the best-fit probability distribution model for the stations at Ikeja and Port Harcourt. While EV1 – MOM model fit best the rainfall data at Owerri station. This result corresponds to that of [6], who found that the EV1 and GEV distribution models are appropriate for the three selected stations in Southern Nigeria. Also, [17] and [18] found that the GEV distribution is most appropriate distribution of the monthly rainfall data in Bangladesh.

3.5 Validation of the Best Distribution model for Each Stations

The best rainfall distribution model for the three substations was validated using Kolmogorov-Smirnov (K-S) test, and this is as seen in Table 13.

Table 13: Model validation Result with K-S Goodness of fit Statistics

Station	Best Fit Distribution Model	Test Performed	Calculated D' Value	Critical D' value at 5% significance level	Decision
Ikeja	GEV LMO	K-S test	0.1232	0.4301	Accept Ho
Owerri	GEV LMO	K-S test	0.1579	0.4301	Accept Ho
PH	GEV LMO	K-S test	0.0918	0.4301	Accept Ho

The Kolmogorov-Smirnov test is a non-parametric test used to compare the empirical cumulative distribution function (ECDF) of a sample against a given theoretical distribution or to compare two samples to determine if they are drawn from the same distribution. The K-S test in Table 4.10 were calculated using the excel spreadsheet and the Critical (D) values at 5% significance level were approximated. The calculated D values range from 0.0918 to 0.1579. These values represent the maximum vertical distance between the ECDFs of the observed data and the theoretical distribution assumed by the models being tested. Since all calculated D values are smaller than the critical D value of 0.4301 at a 5% significance level, this indicates that there is no significant difference between the observed data and the theoretical distribution assumed by the models. In other words, the models GEV-LMO provide a good fit to the data at the 5% significance level. The critical D value of 0.4301 serves as the threshold for determining whether the models being tested are a good fit to the data. Since the calculated D values are below this critical value, it suggests that the models are statistically consistent with the observed data based on the Kolmogorov-Smirnov test at the 5% significance level.

3.6 Development of Rainfall Frequency Curves

The best fit probability distribution was used to predict the rainfall intensity per the return period of 2, 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 75, 100 and 200 as presented in Figure 6.

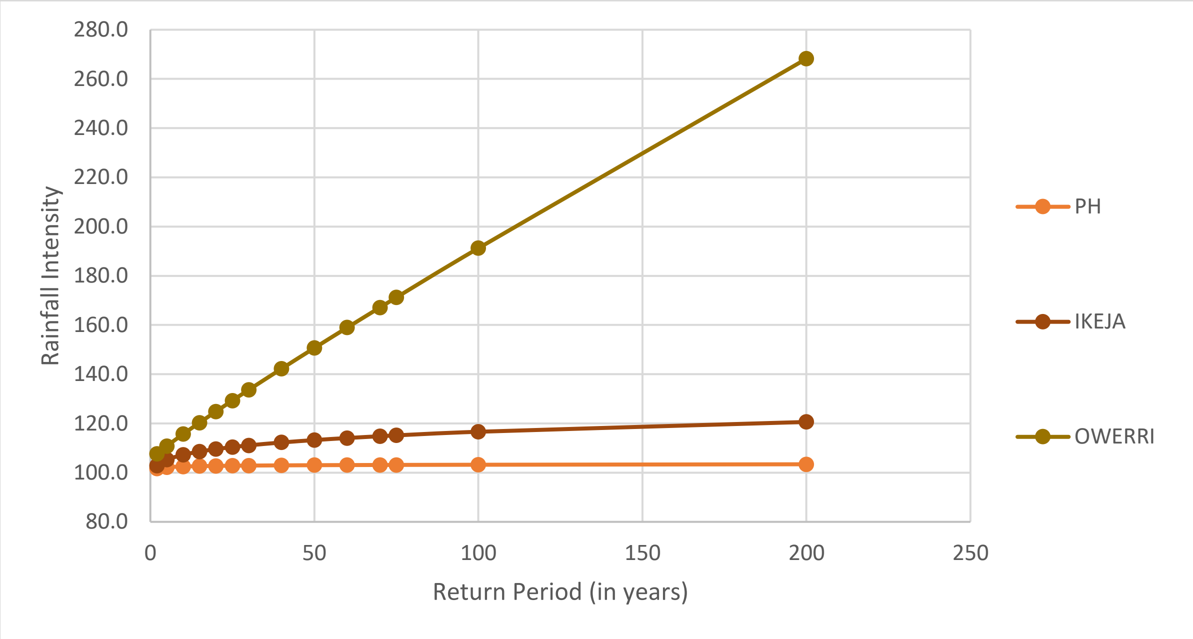

Figure 5: Rainfall frequency curves for the ten substations.

4. Conclusions

The probability weighted moments analysis for various distribution models applied to data from different weather stations reveals distinct performance metrics. Notably, at Ikeja and Port Harcourt stations, the GEV LMO model demonstrated the best fit with the lowest RMSE, RRMSE, MADI, MAE, and highest PPCC values. While EV1 MOM demonstrated best at Owerri station. The Kolmogorov-Smirnov (K-S) test was used to validate the rainfall distribution model by comparing the empirical cumulative distribution function (ECDF) of a sample against a given theoretical distribution. The calculated K-S test values ranged from 0.09182 to 0.1579, represent the maximum vertical distance between the ECDFs of the observed data and the theoretical distribution assumed by the models being tested. Since all calculated D values are smaller than the critical D value of 0.4301 at a 5% significance level, this indicated that there is no significant difference between the observed data and the theoretical distribution assumed by the models, suggesting that the GEV-LMO and EV1 MOM models provide a good fit to the data at the 5% significance level.

References

[1] Bell, J. E. et al., (2018). Changes in extreme events and the potential impacts on human health. *Journal of the Air & Waste Management Association*, 68(4), pp. 265-287

[2] Curriero, F. C., Patz, J. A., Rose, J. B. and Lele, S., (2001). The Association Between Extreme Precipitation and Waterborne Disease Outbreaks in the United States, 1948–1994. *American Journal of Public Health* , 91(8), pp. 1194-1199.

[3] Salarpour, M., Yusop, Z. and Yusof, F., (2012). Modelling the Distributions of Flood Characteristics for a Tropical River Basin. *Journal of Environmental Science and Technology*, 5(6), pp. 419-429.

[4] Douben, K.-J. (2006). Characteristics of river floods and flooding: a global overview, 1985–2003. *Irrigation and Drainage, 55*(S1), S9-S21. Retrieved from https://doi.org/10.1002/ird.239

[5] Rao, A. R. and Hamed, K., (2019). *Flood Frequency Analysis.* Boca Raton, CRC Press, p. 376.

[6] Agbonaye, A. I. and Izinyon, O. C., (2022). Evaluation of Best- Fit Probability Distribution Models for Prediction of Rainfall in Southern Nigeria. *Arid Zone Journal Of Engineering, Technology & Environment*, 18(1), pp. 143-158.

[7] Olofintoye, O., Sule, B. F. and Salami, A., (2009). Best–fit Probability distribution model for peak daily rainfall of selected Cities in Nigeria. *New York Science Journal*, 2(3), pp. 1-12.

[8] Machekposhti, K. H., H. S., Telvari, A. and Babazadeh, H., (2016). Determination of Suitable Probability Distribution for Annual Discharges Estimation (Case Study: Karkheh River at Iran). *International Journal of Probability and Statistics* , 5(3), pp. 73-81.

[9] Vivekanandan, N., (2022). Effect of Data Length on Estimation of Rainfall using Six Probability Distributions. *Water and Energy International,* 64r(11), pp. 13-19.

[10] Donat, M. G. et al., (2016). More Extreme Precipitation in the World's Dry and Wet Regions. *Nature Climate Change,* Volume 6, pp. 508-513.

[11] Alabi, T. A. and Olajide, B., (2023). Who Wants to Go Where? Regional Variations in Emigration Intention in Nigeria. *African Human Mobility Review,* 9(1), pp. 77-101.

[12] Raes, D., Willems, P. and GBaguidi, F., (2006). *RAINBOW – a software package for analyzing data and testing the homogeneity of historical data sets.* Islamabad, Pakistan, 4th International Workshop on 'Sustainable management of marginal drylands', pp. 27-31.

[13] Chadwick, A., Morfett, J. and Borthwick, M., (2004). *Hydraulics in Civil and Environmental Engineering.* 4th ed. s.l.:CRC Press.

[14] Heo, J.-H.et al., (2008). Regression equations of probability plot correlation coefficient test statistics from several probability distributions. *Journal of Hydrology,* 355(1-4), pp. 1-15.

[15] Hatem, G., Zeidan, J., Goossens, M. and Moreira, C., (2022). Normality Testing Methods and the Importance of Skewness and kurtosis in Statistical Analysis. *BAU Journal - Science and Technology,* 3(2).

[16] Hosking, J., (2007). Some theory and practical uses of trimmed L-moments. *Journal of Statistical Planning and Inference,* 137(9), pp. 3024-3039.

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

[18] Alam, J., Muzzammil, M. and Khan, M. K., (2016). Regional flood frequency analysis: comparison of L-moment and conventional approaches for an Indian catchment. *ISH Journal of Hydraulic Engineering,* 22(3), pp. 247-253.