STATISTICAL MODELING OF RAINFALL DISTRIBUTION IN

JOS, PLATEAU STATE

Abstract

Rainfall is a form of precipitation that occurs when water vapor in the atmosphere condenses into droplets which can no longer be suspended in the air. Flood is as a result of rainfall overflowing onto land. Flood is associated with a lot of negative consequences on the activities of man and animals in general and even threatens their existences. In this study, rainfall data obtained on rainfall occurrence in Jos, Plateau State were described. Five theoretical distributions; Weibull, Log-normal, Gamma, Extreme value type 1 and Log-Pearson type III distributions were fitted to the data. Kolmogorov-Smirnov and Anderson-Darling tests of goodness of fit were used to identify the most appropriate distribution. The Maximum likelihood estimator, Bayesian estimator and Principle of maximum entropy were used to estimate the parameters of the identified distribution. Alkaike information criterion (AIC) was used to compare the estimates of the parameters from the different estimators. The probability of the returning periods and the future amount of rainfall were then predicted. The results obtained show that the distribution is skewed to the left, with median annual rainfall of **2383.5mm** and inter-quartile range of **432.4mm**.The results obtained also show that the Log-Pearson Type III distribution best fitted the rainfall data with the Bayesian method as the best estimator of the parameters of the distribution. Results on the probability of the returning period show that it decreases as the number of years increases while the amount of rainfall increases as the number of year's increases.

Key Words: Rainfall, Probability distribution, Returning intervals, Estimation, Parameters

1. Introduction

Precipitation is the falling of water in various forms on the earth from the clouds. The usual forms of precipitation are rain, snow, sleet, glaze, hail, dew and frost [1]. The common form of precipitation in North central, Nigeria is rainfall. Rainfall is a form of precipitation that occurs when water vapor in the atmosphere condenses into droplets which can no longer be suspended in the air. The occurrence of rainfall is dependent upon several factors such as prevailing wind directions, ground elevation and location within a continental mass, and location with respect to mountain ranges, all have a major impact on the possibility of precipitation

1.1 Probability Distributions of Rainfall Data and Estimation of the Parameters

[2] dealt with the spatial pattern of mean annual rainfall that spans from 744.36 to 2189.47 mm/year, with rainfall increasing from the low plains to the high-elevation zones in the western region. MSWEP data showed more rainfall than ground data, it has been observed that the catchments (highly elevated areas) receive more rainfall, which might increase dam inflow and outflow and these regions that are located on the western side of the basin have largely deep forests and receive rainfall throughout the year.

[3] investigated the best fit probability distribution for two different climates in Italy by comparing light-tailed and heavy-tailed distributions. The Kolmogorov-Smirnov and ratio mean square error tests were used to assess the performances of the distributions. It was concluded from the study that heavy-tailed distributions give a better description of empirical data than light-tailed distributions.

[4] investigated the annual, seasonal and monthly rainfall trends of the northern region of Sri Lanka from 1970 to 2019. The results show that the historical annual rainfall of northern region of Sri Lanka has increased from 18.76 mm/decades to 37.68 mm/decade from 1970 to 2019.

[5] used the Markov model to formulate a four-state model in continuous time to study the annual rainfall data with respect to the manual rainfall distribution for crop production in Minna. He observed that if there is low rainfall in a given year, it will take at most 25%, 33% and 27% of the time to make a transition to moderate rainfall, high rainfall and moderate rainfall respectively in the far future.

2. Materials and Methods

2.1 Data collection

The data used for this study are secondary data collected on annual rainfall in Jos, Plateau State.

2.2 Weibull Distribution

Let X denote a random variable, a two- parameter Weibull density function is given by [6] as;

$$f(x, \propto, \beta) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{\alpha}}; \quad \alpha > 0, \beta > 0 \ and \ x \geq 0 \qquad (1)$$

where α is the shape parameter and β is the scale parameter with mean $\Gamma(\frac{1}{\alpha} + 1)$ and variance $\Gamma\left(\frac{2}{\alpha} + 1\right) - \Gamma^2(\frac{1}{\alpha} + 1)$

2.3 Lognormal Distribution

A random variable X is log normally distributed if $\ln(X)$ is normally distributed. The probability density function is given by [7] as;

$$f(x;\mu,\sigma^2) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(lnx-\mu)^2}{\sigma^2}} \quad ; x > 0, -\infty < \mu < \infty, \sigma^2 > 0 \qquad (2)$$

where μ is the location parameter as well the mean of the distribution and σ is the scale parameter as well the standard deviation of the distribution with mean, $\exp(\mu + \frac{\sigma^2}{2})$ and variance,$[\exp(\sigma^2) - 1]\exp(2\mu + \sigma^2)$

2.4 Gamma Distribution

Let X denote a random variable. A two parameter gamma density function with parameters α and β is given by [7] as;

$$f(x;\propto,\beta) = \frac{1}{\Gamma\alpha\beta^{\propto}} x^{\propto-1} e^{-\frac{x}{\beta}}; \qquad x > 0, \alpha > 0, \beta > 0 \qquad (3)$$

Where α is the shape parameter and β is the scale parameter with mean $\alpha\beta$ and variance $\alpha\beta^2$

2.5 **Log-Pearson Type III Distribution**

Let $y = lnx \ where \ x$ is a positive random variable. If y has a Pearson type III distribution, then x will have a LPT III distribution with pdf given by [8] as;

$$f(x) = \frac{1}{ax\Gamma b} (\frac{lnx-c}{a})^{b-1} \exp\left[-\left(\frac{lnx-c}{a}\right)\right] \qquad (4)$$

where,$a > 0, b > 0, and \ 0 < c < lnx, \Gamma(.) = gamma \ distributio$

2.6 Gumbel Distribution

A random variable x is said to have a Gumbel distribution if its probability density function is given by [9] as;

$$f(x) = a \exp(-a(x-b) - e^{-a(x-b)}) \qquad (5)$$

where, $a > 0, -\infty < b < lnx \ and \ x > 0$

2.7 The Test Statistics

In order to verify the goodness of fit of the models for the rainfall data, the Kolmogorov – Smirnov (K-S) and Anderson – Darling (A-D) tests were used. The lower the values of this statistics, the closer the fitted distribution appears to match the data. The hypothesis for the test is given as ;

H_0: The hypothesized distribution fits the data

Versus

H_1: The hypothesized distribution does not fit the data

Decision Rule: reject H_0 if p-value is less than the level of significance ($\alpha = 0.05$).

Given "N" ordered data points $X_1, X_2, \ldots, X_N$ the test statistic for the Kolmogorov – Smirnov test is given by [10] as;

$$D \equiv \max\left(F(X_i) - \frac{i-1}{N}, \frac{i}{N} - F(X_i)\right), 1 \leq i \leq n \qquad (6)$$

The test statistic for Anderson-Darling is given by [11] as;

$$A^2 = -N - S \qquad (7)$$

where,

$$S = \sum_{i=1}^{n} \frac{(2i-1)}{N}[lnF(X_i) + \ln(1 - F(X_{N+1} - i))] \qquad (8)$$

F (.) is the cdf of the continuous distribution and X_i is the ordered data and N is the population size.

2.8 Hydrological Analysis

In carrying out hydrological analysis such as flood and rainfall frequency analysis using Log-Pearson Type III distribution, the following steps suggested by [12] were adopted:

(i) the annual rainfall series X_i were assembled

(ii) the logarithm of the annual rainfall series were calculated as

$$y_i = logX_i \quad (3.66)$$

(iii) the mean $\bar{y}$, standard deviation σ_y and skew coefficient of the logarithm y_i were calculated

(iv) the logarithm of the rainfall i.e. $logX_i$ for each of the several chosen probability level P_i were calculated using the following frequency formula

$$logX_i = \bar{y} + k_j\sigma_y \quad (3.67)$$

where k_j is the frequency factor and function of probability P_i and skewness coefficient.

3. Applications

In this section, data on the rainfall data in Jos are presented and summarized, appropriate distributions were fitted to the data.

Table 1. Yearly Rainfall Data in Jos (1990 – 2022)

Year	Annual rainfall(mm)
1990	1557.9
1991	2153.2
1992	2396.1
1993	2482.9
1994	2075.5
1995	2563.7
1996	2581.5
1997	2961.3
1998	2567.4
1999	2424.1
2000	2182.8
2001	262.0
2002	2622.3
2003	2705.5
2004	2891.4
2005	1640.1
2006	2515.4
2007	2337.2
2008	3209.1
2009	2053.7
2010	2327.8
2011	1762.3
2012	2236.6
2013	2337.2
2014	2361.6
2015	2470.2
2016	2092.8
2017	2370.9
2018	2434.34
2019	2747.4
2020	2689.01
2021	2817.0
2022	2073.9

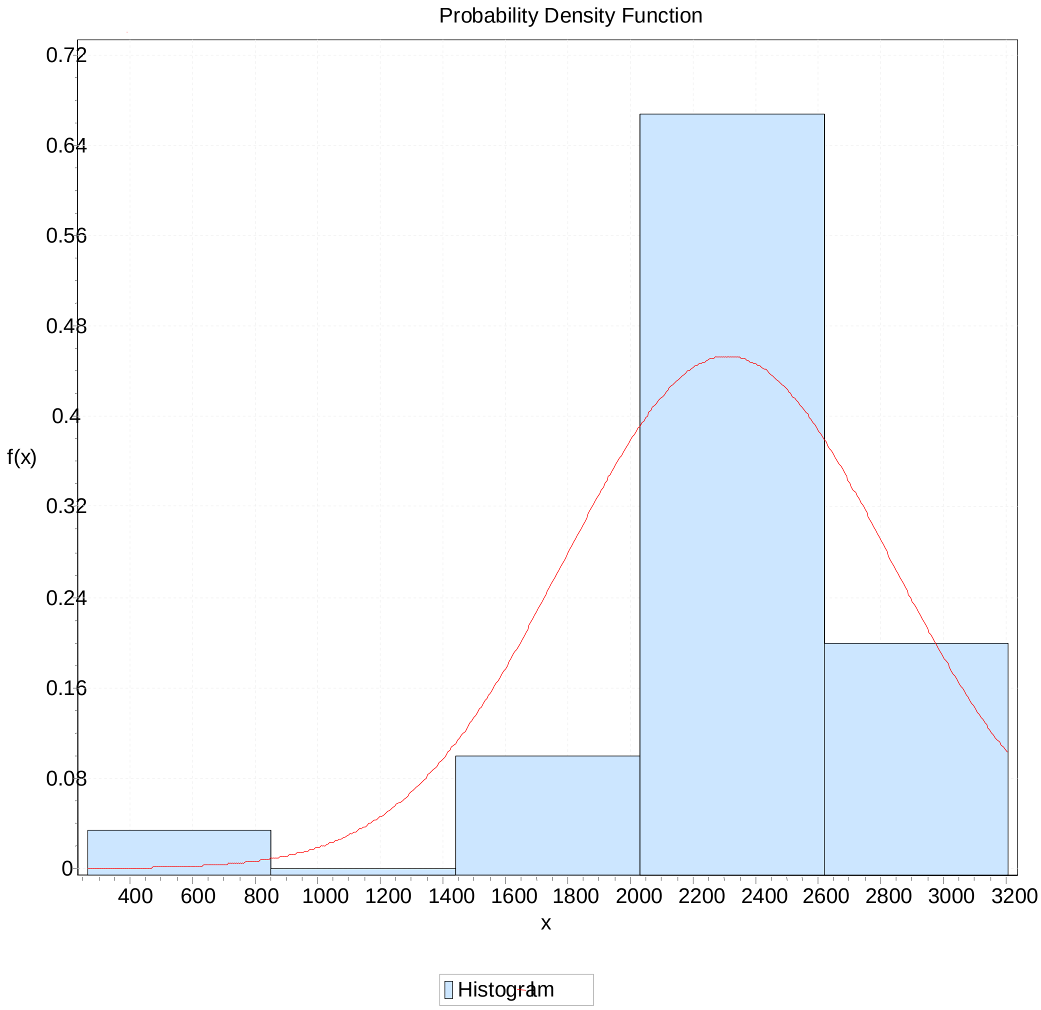

Fig. 1: Histogram of the Densities of the Annual Rainfal in Jos, Plateau State

It can be deduced from Figure above that the information on rainfall data (as collected) is negatively skewed. Figure 1 also shows that the most frequently occurring yearly amount of rainfall lies between 2000 and 2600mm.

In order to obtain more insights from the rainfall data, some basic descriptive statistics were calculated from the data. The results are shown in Table 2.

Table 2: Descriptive Statistics for Logarithm of Rainfall in Jos

Variable	Mean	St.Dev	Minimum	Maximum	Skewness	Kurtosis	median	Inter-quartile range
$\log X_i = y_i$	3.3380	0.1865	2.4183	3.5064	-4.39	21.85	3.772	2.6357

The results in Table 2 show that the skewness coefficient confirms that the distribution is negatively skewed while the kurtosis shows departure from normal distribution. The results also show that when the log of the original values were taken, the standard deviation was greatly reduced, the coefficient of skewness increased approximately by a factor of 2 when compared with the original value while the kurtosis increased by a factor of 3 approximately when compared with the original value. Thus, the following relationship could be deduced from the original data and its logarithm:

(i) skewness of log $X = 2$ skewness of X

(ii) kurtosis of log $X = 3$ kurtosis of X,

where X denotes the original rainfall data.

In order to choose the 'best' probability distribution to describe Rainfall data in Jos, Plateau State for the period studied, Kolmogorov Smirnov and Anderson-Darling goodness of fit tests were employed in Table 3 below.

Table 3: Results of Goodness of Fit Test

Distribution	Kolmogorov-Smirnov	Anderson Darling	p-values
Gamma	0.1758	0.56575	0.141564
Lognormal	0.18877	0.4143	0.3355
LPIII	**0.10841***	**0.29044***	**0.4724***
Weibull	0.1497	0.3822	0.3989
EV1	0.1342	0.5234	0.182702

NOTE: * denotes the best fit

Results in Table 3 show that the most appropriate distribution that described the rainfall data in Jos, Plateau State during the period under study is the log-Pearson Type III distribution since it had the highest p-value greater than the 0.05 level of significance and the least values of Kolmogorov-Smirnov and Anderson Darling test statistics among other distributions.

Table 4: Parameter Estimates of the fitted Probability Distribution

Distribution	***MLE***	**AIC**	***POME***	**AIC**	***B.E***	**AIC**
LPIII	$\hat{a}$=4.919 $\hat{b}$=-0.0092 $\hat{c}$=1.26	-80.675	$\hat{a}$=5.819 $\hat{b}$=-0.0072 $\hat{c}$=1.30	-80.615	$\hat{a}^*$=5.999 $\hat{b}^*$=-0.0084 $\hat{c}^*$=2.51	**-81.572**

The results in Table 4 show that for log-Pearson Type III distribution, the AIC yielded by Bayesian estimator was less than that of maximum likelihood estimator and principle of maximum entropy.

Table 5: Application of Log-Pearson Type III Distribution to the observed Data

Return Period T(years)	***Probabilities P(%)***	***Frequency Factor K(g=-2.06)***	$y_i = \overline{y} + k_j\sigma_y$	$X_i = antilog(y_i)(mm)$
2	50	0.307	3.3952	2484.3
5	20	0.777	3.4827	3038.8
10	10	0.895	3.5046	3196.0
25	4	0.959	3.5166	3285.5
50	2	0.980	3.5205	3315.1
100	1	0.990	3.5223	3328.9
200	0.5	0.995	3.5233	3336.6

The results in Table 5 show that the probability that 2484.3(mm) rainfall or greater will occur on the average of 2 years is 50%, the probability that 3038.8mm rainfall or greater will occur on the average of

5 years is 20%, the probability that 3196.0mm rainfall or greater will occur on the average of 10 years is 2%, the probability that 3285.5 mm rainfall or greater will occur on the average of 25 years is 4%, the probability that 3315.1 mm rainfall or grater will occur on the average of 50 years is 2%, the probability that 3328.9 mm rainfall or greater will occur on the average of 100 years is 1%, and the probability that 3336.6 mm rainfall or greater will occur on the average of 200 years is 0.5% in Jos Plateau State.

4. Conclusion

In this study, the probability distribution of the annual rainfall in Jos Plateau State was examined using Kolmogrov-Smirnov test of goodness of fit and Anderson-darling statistic. The most appropriate probability distribution that fitted the data was identified through the magnitudes of the Kolmogrov-Smirnov and Anderson-Darling statistics alongside their p-values. The estimates of the parameters of the identified distribution was obtained through the maximum likelihood estimation, principle of maximum entropy and Bayesian Information criterion. The estimates were compared with the magnitudes of their Akaike information criterion.

The results obtained show that the Log-Pearson Type III distribution is the most appropriate distribution that fitted the rainfall data and the Bayesian estimator is the best estimator of the parameters of the distribution. The probabilities of the returning intervals of rainfall in Jos, Plateau State were then calculated which shows that the probability of the returning period decreases as the number of years increases.

References

[1] Arora, KR. Irrigation, Water Power and Water Resources Engineering. Standard Publishers Distributors, Delhi, India; (2007).

[2] Hrudya, PH., Varikoden, H.; Vishnu, R. A review on the Indian summer monsoon rainfall, variability and its association with ENSO and IOD. Meteorol. Atmos. Phys; 2021, 133, 1–14

[3] Moccia, B, Mineo C, Ridolfi E, Russo, F, and Napolitano, F. Probability distributions of daily rainfall extremes in Lazio and Sicily, Italy, and design rainfall inferences. Journal of Hydrology: Regional Studies; (2021), Vol. 33.

[4] Piratheeparajah N, Chan NW and Tan ML Trend analysis of rainfall in the northern region of Sri Lanka from 1970 to 2019. GEOGRAFI; 2021 Vol, 9: 85-107, https://doi.org/10.37134/geografi.vol9.1.5.2021

[5] Abubakar, UY, Lawal, A and Muhammed, A.The Use of Markov Model in Continuous Time for Prediction of Rainfall for Crop Production. Journal of Mathematics; (2013) 7(1): 38-45.

[6] Lei. Y. Evaluation of three methods for estimating the Weibull distribution parameters of Chinese pine. .Journal of Forest Science; (2008), 54 (12): 566-571.

[7] Brenda, FG. Parameter Estimation for the lognormal distribution. Master's Thesis, Brigham Young University, Provo, United States*;* (2009)

[8] Forbes, C, Evans, M., Hastings, N., and Peacock, B. Statistical Distributions 4thEdition, John Wiley & Sons, Inc. Hoboken, New Jersey (2011).

[9] Singh, VP. and Guo, H. Parameter estimation for 3-parameter generalized Pareto distribution by the principle of maximum entropy (POME). Hydrol Sci. J.; (2010), 40: 165–181

[10] Koutsoyiannis, D. On the Appropriateness of the Gumbel Distribution in Modeling Extreme Rainfall. Proceedings of the ESF LESC Exploratory Workshop, Italy; (2003)

[11] Lu, HC, and Fang, GC. Predicting the Exceedance of a critical PM_{10} Concentration - A Case Study in Taiwan Atmospheric Environment; (2003): 37 (4), 3491 -3499.

[12] Jagadesh TR. and Jayaram MA. Design of Bridge Structures. PHI Publishers; (2009), 2nd edition.

.