Check for updates

RESEARCH ARTICLE

Parametric modelling of rainfall return periods in south-western Nigeria: Survival analysis approach [version 1; peer review: 1 approved, 1 approved with reservations]

Phillip Awodutire ¹, Blessing Sasanya ², Olohita Ufuoma ¹,
Oluwafemi Samson Balogun ³

¹Department of Mathematics and Computer Sciences, University of Africa Toru Orua, Toru Orua, Bayelsa, Nigeria

²Department of Crop and Soil Science, University of Port Harcourt, Port Harcourt, Rivers, Nigeria

³School of Computing, University of East Finland, Kuopio, Finland

V1 First published: 24 Jan 2022, 11:83
<https://doi.org/10.12688/f1000research.75722.1>
 Latest published: 24 Jan 2022, 11:83
<https://doi.org/10.12688/f1000research.75722.1>

Abstract

Background: Rainfall is the main source of water on the earth's surface. It infiltrates and percolates deep into the soil for groundwater recharge. Rainfall patterns, amounts, durations, and intensities can vary daily, monthly, annually, and spatially. It is therefore important to accurately estimate rainfall return periods, which can be employed in hydraulic design and flood control measures.

Methods: This research considered the survival analysis approach for the prediction of rainfall return periods including intensity, and months during which these would occur in south-western Nigeria. Twenty years' of annual rainfall data were obtained from three metrological stations and these were subjected to nine different probability plotting position methods. Results from the plotting positions was further subjected to four survival models using five years of censor time. The Akaike Information Criterion (AIC) was used to determine the best-fitting model for the dataset.

Results: The Laplace probability plotting position in conjunction with the log-logistic distribution best describes the datasets, since it gave the lowest AIC value of 22.53. The log-logistic distribution is also suitable for the prediction of return period from the Weibull probability plotting position since the AIC values were 6.934 and -4.332 respectively. The Hirsh plotting position in conjunction with the Weibull distribution is also suitable for the description of the dataset.

Conclusion: The established parametric models are suitable for the accurate prediction of return periods of peak rainfall events during any month of the year.

Keywords

return periods, peak rainfall intensities, survival analysis, parametric modelling, probability plotting position

Open Peer Review

Approval Status

	1	2
version 1		
24 Jan 2022	view	view
1. Thomas Xavier , Kannur University, Kannur, India		
2. Sohail Chand , University of the Punjab, Lahore, Pakistan		

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Phillip Awodutire (phillip.awodutire@gmail.com)

Author roles: **Awodutire P:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Writing – Original Draft Preparation; **Sasanya B:** Conceptualization, Data Curation, Investigation, Methodology, Writing – Original Draft Preparation; **Ufuoma O:** Methodology, Writing – Review & Editing; **Balogun OS:** Funding Acquisition, Supervision

Competing interests: No competing interests were disclosed.

Grant information: This manuscript was funded by Digiteknologian TKI-ymparisto project A74338 (ERDF, Regional Council of Pohjois-Savo).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2022 Awodutire P *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Awodutire P, Sasanya B, Ufuoma O and Balogun OS. **Parametric modelling of rainfall return periods in south-western Nigeria: Survival analysis approach [version 1; peer review: 1 approved, 1 approved with reservations]** F1000Research 2022, 11:83 <https://doi.org/10.12688/f1000research.75722.1>

First published: 24 Jan 2022, 11:83 <https://doi.org/10.12688/f1000research.75722.1>

Introduction

The concept of survival analysis as a statistical tool in handling time to an event related problem was initially designed for medical related studies. Example of such studies include time taken for a patient to recover from diseases, time to die from a disease, and the likes. Several studies have been conducted using parametric models to analyse health related issues (Awodutire *et al.*, 2018; Naseri, *et al.*, 2018). In recent times, survival analysis has been extended to finances (Laitinen, 2005; Witzany *et al.*, 2012; Lee, 2014) engineering (reliability analysis, Awodutire *et al.*, 2021), politics (Jonathan, 2014) and similar. It may be of interest to model the relationship of time to event to other covariates (sometimes called predictors). For this research, the event of interest was rainfall intensities and its return period (i.e. the time taken to experience rainfall).

Nigeria is a country with many rivers, lakes, ponds (natural and manmade) and tides of these rivers are controlled majorly by seasons. There are different climatic seasons in Nigeria namely dry and rainy seasons. During the rainy seasons, rainfall amounts, and intensities are usually very high. Nigeria's economy is linked to climate sensitive activities (Houessou-Doussou *et al.*, 2019). Precipitation (rainfall) is formed when saturated air is heated, rises by a mountain, convectional current or frontal action (Ogungbenro and Morakinyo, 2014). Rainfall is the main source of water on the earth's surface which percolates through the soil for groundwater recharge and runs off the soil surface forming surface water such as streams and rivers (Salase *et al.*, 2015). It also forms depression storages, it is trapped by vegetative interception and it is then transferred from the land to the atmosphere by evapotranspiration. The importance of rainfall in water resources management, water supply, agricultural activities and food production cannot be over emphasized; however, rainfall varies spatially and temporally. Precipitation is the main driver of variability in the water balance over time and space (Davie, 2008).

Rainfall can be of immense benefit to the environment, but it can also be accompanied by adverse effects at certain times. Excessive rainfall can lead to undesirable disasters such as flooding and landslides. Rainfall patterns, amount, duration and intensity can vary daily, monthly, annually and from one place to another. This makes it subject to uncertainties and probabilities, which are partly explained by the concept of rainfall return periods. Return period is an essential tool in hydrology which estimates the time interval between events of similar sizes or intensities (Laura and Richard, 2015). An understanding of historical and current rainfall trends is essential in determining the return periods in a particular area. Factors such as missing data and unknown probability distribution function of annual peaks makes the estimation of return period of real events a tedious task; frequency analysis is thus used to estimate the return periods of specific events (Houessou-Doussou *et al.*, 2019). Consequently, it is important to estimate rainfall return period by modelling its components with real world behaviours and attributes. Varying precipitation can influence hydrology, water resources as well as extreme events such as flood and drought. Statistical estimates are therefore used for forecasting, prediction, correlation, collation and analysis of daily, monthly or annual rainfall rates and duration data (Ybanez, 2013). Analysis of past rainfall data provides estimates of recurrence interval, which can also be used to predict into the future (Olatunde and Adejoh, 2017). According to Olatunde and Adejoh (2017), stochastic analysis of rainfall is of high importance in the design and development of civil engineering structures such as buildings, bridges, water storage structures (reservoirs, detention basins, rainwater tanks). These structures are needed to maintain continual usage, under specified reliability, environmental or agricultural conditions. Probability analysis of past rainfall records are useful in that regard for the determination and prediction of highest rainfall months and years (Ewemoje and Ewemoje, 2011). These are also important for preparation against disaster.

Generally, stochastic analysis involves characterization of the probability distribution of the variable (rainfall in this case) and its associated predictors, so that conditional probability distribution can be derived. For rainfall, the characterization is more specific to the time scale being modelled, which could be annual, monthly, daily or sub daily time scales (Olatunde and Adejoh, 2017).

Several studies have been done around rainfall prediction from past records using probability functions. Ewemoje and Ewemoje (2011) researched the best plotting position and distribution for flood estimation of the Ona river using 18 years' peak flood data from the Ogun Oshun River Basin, Nigeria. Three probability distributions and plotting position methods were compared. The suitability of the Hazen, Weibull and California probability position methods were compared, as well as those of the normal, log-normal and log Pearson type (III) distributions. The Hazen plotting position and log Pearson type (III) distribution performed best. Hurford *et al.* (2012) studied the validation of return period of rainfall thresholds used for extreme rainfall alerts using links with rainfall intensities and observed surface flood events. The research hinged on investigating if return period is adequate for the warning of surface water flooding by examining the intensity and return period of rainfall associated with observed surface water flood events. Rainfall amount recorded by rain gauges and flood events were analysed which showed that most surface flood water events were associated with rainfall intensities of less than a 1-in-10-year return period. It was concluded that better understanding of the relationship

between flood magnitude and rainfall intensity could be enhanced through the improvement of data recording on flood magnitude and duration for informed comparison between surface water flood warning thresholds. Agbede and Aiyelokun (2016) established the most suitable stochastic model for flood management in the Yewa sub-basin, south-western Nigeria. The peak floods were fitted into normal, gamma, gumbel and Weibull distributions using 13 years' peak flood data, with return periods obtained from the Hazen plotting position method. The Weibull distribution was reported to be the most suitable distribution for predictions of flood in the Yewa sub-basin. In the same vein, Aiyelokun *et al.* (2017) fitted 31 years' hydrologic data from the gauged Opeki river to various probability distributions using return periods founded by the Hazen method. The researchers employed normal, log normal, log Pearson type (III), exponential, extreme value type (I), extreme value type (II) and the three-parameter burr distribution. The exponential and normal distributions were reported incapable of predicting flood flows from the Opeki river. It was further reported that the log Pearson type (III) distribution was the most suitable for the estimation of peak flood from the Opeki river. Santos *et al.* (2015) analysed seasonal return periods for maximum daily precipitation in the Brazilian Amazon. The extreme value theory was adopted using the non-parametric generalized extreme value (GEV) distribution and the generalized Pareto distribution (GPD). The GEV and GPD goodness of fit were evaluated by applying the Kolmogorov–Smirnov (KS) test, which compares the cumulative empirical distributions with theoretical ones. The KS test indicated that the tested distributions had a good fit, particularly the GEV distribution. They were thus adequate for the study of seasonal maximum daily precipitation. Furthermore, Yahaya (2012) and Ogungbenro and Morakinyo (2014) used statistical methods to justify the changes observed in monthly and annual rainfall trends over some years. Obot (2010) used the non-parametric Mann-Kendall test to check for significant trends in rainfall in Nigeria in some randomly selected locations.

Having reviewed these studies, none used the survival analysis approach to determine possible return period of maximum or peak rainfall events. Several studies have predicted return periods of rainfall from complete datasets, but according to Houessou-Doussou (2019), there may be missing or censored data. Therefore, this study aims to develop models which can adequately predict return periods from peak rainfall intensities and the months of occurrence of such peaks in south-western Nigeria in cases of missing or censored information. The return periods were achieved by plotting probability positions for obtained rainfall data using nine different methods. The plotting position methods were compared and subjected to statistical analysis using parametric modelling which compared four survival models. The parametric approach has been proven over time to be the best method of analysing time data events. This has resulted from its ability to handle datasets with minimal sample sizes and its efficient and consistent estimations. Studies applying survival analysis to rainfall data are rare, but this study takes this approach for the analysis of this time data event.

Methods

The return periods of annual peak rainfall intensities were studied. Monthly rainfall data were obtained from three meteorological stations in south-western Nigeria. The peak intensities from 2009–2018 were considered and their corresponding months of occurrence recorded (Awodutire *et al.*, 2021a). The annual maximums of the monthly rainfall intensities were arranged in descending order of magnitude. These were subjected to probability plotting positions by comparing the California, Weibull, Hazen, Adamowski, Blom, Chegodagev, Gringoten, Hirsh, and Laplace methods. These are given by equations (1) to (9).

$$\text{California method } T = \frac{n}{m} \quad (1)$$

$$\text{Weibull method } T = \frac{n+1}{m} \quad (2)$$

$$\text{Hazen method } T = \frac{n}{m-0.5} \quad (3)$$

$$\text{Admowski method } T = \frac{n+0.5}{m-0.25} \quad (4)$$

$$\text{Blom method } T = \frac{n+0.25}{m-0.375} \quad (5)$$

$$\text{Chegodagev method } T = \frac{n+0.4}{m-0.3} \quad (6)$$

$$\text{Gringorten method } T = \frac{n+0.12}{m-0.44} \quad (7)$$

$$\text{Hirsh method } T = \frac{n+1}{m+0.5} \quad (8)$$

$$\text{Laplace method } T = \frac{n+2}{m+1} \quad (9)$$

Where m = rank order of the rainfall intensities, n = number of years of record, T = return period (years). For this research, the return period is censored at five years. The parametric survival model (also known as accelerated failure time model (AFT)) is of the form:

$$\ln T = \tau + d' \gamma + s \quad (10)$$

where s is said to follow a particular distribution, γ is the covariates, d is the coefficient of the covariates, τ is the intercept of the model and T is the time taken for the event to happen. The covariates under study are rainfall intensity and months while the time T is the return period.

For this study, four different parametric survival models were employed for both return periods generated from the probability plotting equations. These are Exponential, Weibull, log-normal and log-logistic parametric models described by the equations 11 to 14.

$$\text{Exponential: } f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}} \quad x > 0, \theta > 0 \quad (11)$$

$$\text{Log - Logistic: } f(x) = \frac{x^{\alpha-1} \mu \alpha}{(1 + \mu x^{\alpha})^2} \quad x > 0, \alpha > 0, \mu > 0 \quad (12)$$

$$\text{Weibull: } f(x) = \alpha \mu x^{\alpha-1} e^{-\mu x^{\alpha}} \quad x > 0, \alpha > 0, \mu > 0 \quad (13)$$

$$\text{Log - Normal: } f(x) = \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} \quad x > 0, \mu > 0, \sigma > 0 \quad (14)$$

The Akaike Information Criterion (AIC) was used for the comparative studies of the resulting models. The AIC is given as equation 15:

$$\text{AIC} = 2k - 2 \ln(L) \quad (15)$$

Where L is the likelihood value of the model.

The model with the lowest AIC performed the best. The significance of the independent variables in the model (contribution to the dependent variable) were assessed at 0.05 significance level with hypothesis as $H_0: \rho_i = 0$ vs $H_0: \rho_i \neq 0$. The H_0 was rejected at $p < 0.05$. Data analysis was conducted using SPSS version 20.0 (IBM Corp. Released, 2020) (RRID:SCR_019096) and R 4.10 programming (R core team, 2017) (RRID:SCR_001905). The R code for data analysis is presented in Awodutire *et al.* (2021b).

Results and discussion

The average highest rainfall intensity from the data obtained from the weather stations was 453.1 mm/month experienced in August 2015. while the least rainfall intensity was 201.1 mm/month in July 2017. The return periods of the highest rainfall intensity were 20, 21, 40, 27.33, 32.4, 29.14, 35.92, 16.8 and 11 years respectively for the California, Weibull, Hazen, Adamowski, Blom, Chegodayev, Gringorten, Hirsh and Laplace probability plotting position methods. The return periods of the lowest rainfall intensity were 1.00, 1.00, 1.03, 1.03, 1.03, 1.03, 1.03, 1.04 and 1.05 respectively for the California, Weibull, Hazen, Adamowski, Blom, Chegodayev, Gringorten, Hirsh and Laplace probability plotting position methods. Tables 1–9 show the various parametric models relating rainfall return period, monthly intensities and month of the year during which the peak rainfall intensity was experienced. Parametric models capable of adequately

Table 1. Parametric models of return period California probability plotting positions.

Distribution	Parameter	Estimate	S.E	p-value	LogLik	AIC	P
Lognormal	Intercept	−2.752	0.351	0.00	−9.20	26.385	0.00
	x_1	0.082	0.079	0.29			
	x_2	0.013	0.183	0.00			
Exponential	Intercept	−3.992	1.831	0.029	−28.30	62.564	0.00
	x_1	0.025	0.373	0.947			
	x_2	0.019	0.005	0.000			
Log-Logistic	Intercept	−2.637	0.260	0.000	−7.30	22.527	0.00
	x_1	0.0247	0.065	0.700			
	x_2	0.0125	0.223	0.000			
Weibull	Intercept	−2.637	0.259	0.000	−7.30	33.115	0.00
	x_1	−0.025	0.065	0.700			
	x_2	0.013	0.000	0.000			

Table 2. Parametric models of return period Weibull probability plotting positions.

Distribution	Parameter	Estimate	S.E	p-value	LogLik	AIC	P
Lognormal	Intercept	−2.460	0.202	0.000	−0.200	8.341	0.00
	x_1	−0.020	0.045	0.660			
	x_2	0.012	0.000	0.000			
Exponential	Intercept	−3.680	1.840	0.045	−27.700	61.401	0.00
	x_1	−0.1567	0.388	0.686			
	x_2	0.019	0.005	0.000			
Log-Logistic	Intercept	−2.523	0.181	0.000	0.500	6.934	0.00
	x_1	−0.001	0.041	0.980			
	x_2	0.012	0.000	0.000			
Weibull	Intercept	−2.467	0.192	0.000	0.200	7.595	0.00
	x_1	−0.028	0.036	0.430			
	x_2	0.013	0.000	0.000			

Table 3. Parametric models of return period Hazen probability plotting positions.

Distribution	Parameter	Estimate	S.E	p-value	LogLik	AIC	P
Lognormal	Intercept	−2.865	0.271	0.000	−5.300	18.521	0.000
	x_1	−0.0453	0.062	0.460			
	x_2	0.014	0.001	0.000			
Exponential	Intercept	−3.972	1.831	0.030	−28.100	62.153	0.000
	x_1	−0.165	0.390	0.670			
	x_2	0.021	0.005	0.000			
Log-Logistic	Intercept	−2.926	0.258	0.000	−4.700	17.418	0.000
	x_1	−0.031	0.057	0.580			
	x_2	0.014	0.000	0.000			
Weibull	Intercept	−2.886	0.216	0.000	−4.100	16.179	0.000
	x_1	−0.034	0.048	0.470			
	x_2	0.014	0.000	0.000			

Table 4. Parametric models of return period Adamowski probability plotting positions.

Distribution	Parameter	Estimate	S.E	p-value	LogLik	AIC	P
Lognormal	Intercept	−2.635	0.223	0.000	−2.100	12.19	0.000
	x_1	−0.025	0.050	0.620			
	x_2	0.013	0.001	0.000			
Exponential	Intercept	−3.820	1.841	0.038	−27.900	61.87	0.000
	x_1	−0.158	0.390	0.686			
	x_2	0.020	0.005	0.000			
Log-Logistic	Intercept	−2.700	0.201	0.000	−1.300	10.57	0.000
	x_1	−0.009	0.046	0.840			
	x_2	0.013	0.000	0.000			
Weibull	Intercept	−2.700	0.201	0.000	−0.600	9.186	0.000
	x_1	−0.009	0.004	0.840			
	x_2	0.013	0.000	0.000			

Table 5. Parametric models of return period Blom probability plotting positions.

Distribution	Parameter	Estimate	S.E	p-value	LogLik	AIC	P
Lognormal	Intercept	−2.731	0.242	0.000	−3.500	14.983	0.000
	x_1	−0.034	0.054	0.530			
	x_2	0.013	0.000	0.000			
Exponential	Intercept	−3.881	1.835	0.024	−28.000	62.029	0.000
	x_1	−0.161	0.389	0.679			
	x_2	0.020	0.005	0.000			
Log-Logistic	Intercept	−2794	0.225	0.000	−2.800	13.678	0.000
	x_1	−0.020	0.050	0.690			
	x_2	0.014	0.209	0.000			
Weibull	Intercept	−2.717	0.194	0.000	−2.100	12.160	0.000
	x_1	−0.030	0.041	0.460			
	x_2	0.014	0.000	0.000			

Table 6. Parametric models of return period Chegodayev probability plotting positions.

Distribution	Parameter	Estimate	S.E	p-value	LogLik	AIC	P
Lognormal	Intercept	−2.670	0.230	0.000	−2.600	13.250	0.000
	x_1	−0.028	0.052	0.590			
	x_2	0.013	0.000	0.000			
Exponential	Intercept	−3.842	1.839	0.037	−28.000	61.930	0.000
	x_1	−0.159	0.390	0.683			
	x_2	0.020	0.005	0.000			
Log-Logistic	Intercept	−2.735	0.209	0.000	−1.900	11.720	0.000
	x_1	−0.013	0.047	0.780			
	x_2	0.013	0.000	0.000			
Weibull	Intercept	−2.649	0.189	0.000	−1.100	10.204	0.000
	x_1	−0.020	0.386	0.450			
	x_2	0.013	0.000	0.000			

Table 7. Parametric models of return period Gringorten probability plotting positions.

Distribution	Parameter	Estimate	S.E	p-value	LogLik	AIC	P
Lognormal	Intercept	−2.804	0.256	0.000	−4.400	16.779	0.000
	x_1	−0.038	0.058	0.510			
	x_2	0.014	0.000	0.000			
Exponential	Intercept	−3.933	1.833	0.032	−28.000	62.070	0.000
	x_1	−0.162	0.389	0.677			
	x_2	0.020	0.004	0.000			
Log-Logistic	Intercept	−2.870	0.241	0.000	−3.800	15.554	0.000
	x_1	−0.024	0.053	0.650			
	x_2	0.014	0.209	0.000			
Weibull	Intercept	−2.802	0.205	0.000	−3.100	14.142	0.000
	x_1	−0.031	0.044	0.480			
	x_2	0.014	0.004	0.000			

Table 8. Parametric models of return period Hirsh probability plotting positions.

Distribution	Parameter	Estimate	S.E	p-value	LogLik	AIC	P
Lognormal	Intercept	−2.361	0.182	0.000	2.500	2.964	0.000
	x_1	−0.001	0.040	0.970			
	x_2	0.011	0.000	0.000			
Exponential	Intercept	−3.616	1.756	0.040	−29.500	64.921	0.000
	x_1	−0.041	0.384	0.914			
	x_2	0.017	0.004	0.000			
Log-Logistic	Intercept	−2.420	0.167	0.000	3.200	1.563	0.000
	x_1	0.012	0.037	0.720			
	x_2	0.011	0.000	0.000			
Weibull	Intercept	−2.382	0.169	0.000	4.100	−0.228	0.000
	x_1	−0.018	0.032	0.570			
	x_2	0.012	0.000	0.000			

Table 9. Parametric models of return period Laplace probability plotting positions.

Distribution	Parameter	Estimate	S.E	p-value	LogLik	AIC	P
Lognormal	Intercept	−2.128	0.168	0.000	5.400	−2.754	0.000
	x_1	0.003	0.036	0.940			
	x_2	0.010	0.000	0.000			
Exponential	Intercept	−3.680	1.840	0.046	−27.700	61.401	0.000
	x_1	−0.157	0.388	0.686			
	x_2	0.019	0.005	0.000			
Log-Logistic	Intercept	−2.195	0.151	0.000	6.200	−4.332	0.000
	x_1	0.015	0.032	0.650			
	x_2	0.011	0.000	0.000			
Weibull	Intercept	−2.178	0.150	0.000	7.200	6.440	0.000
	x_1	−0.013	0.027	0.630			
	x_2	0.011	0.000	0.000			

predicting rainfall return periods from rainfall intensities and annual calendar months were obtained from the log-normal, exponential, log-logistic and Weibull distributions. They are of the form shown on [equation 12](#):

$$\ln T = C + Bx_1 + Ax_2 \quad (12)$$

T is the return period, C is the intercept of the equation, B is the coefficient of x_1 , which is the rainfall intensity, A is the coefficient of the month variable x_2 .

Parametric models were derived by comparing four probability distributions using return periods from the nine probability plotting position methods. Right censoring was considered. The rainfall return period was censored at five years. The p-values in [Tables 1–9](#) for each survival model under consideration indicate that the models fit the data very well ([Awodutire et al., 2021c](#)). The AIC values were used to compare the models from each of the distributions and the probability plotting position methods.

[Table 1](#) compares the probability distributions employed for the derivation of the suitable parametric model from the California probability plotting positions. The log-logistic distribution proved to be the most suitable distribution for the plotting position since it has the lowest AIC value of 22.53. In the same vein, the log-logistic distribution is most suitable for the prediction of return period from the Weibull and Laplace probability plotting position methods since the AIC values were 6.934 and -4.332 respectively. The exponential distribution however had the highest AIC values for both the California and Weibull plotting position methods. The AIC values were 62.56 and 61.40 respectively. The exponential distributions proved to be most unsuitable for all the plotting position methods. The AIC values of the exponential distribution consistently ranged between 61 and 65. The Weibull distribution, however, was the best fit to the return period data from the Hazen, Adamoski, Blom, Chevodayev, Gringoten, and Hirsh probability plotting positions. The AIC values attributed to the Weibull distribution were 16.18, 9.19, 12.16, 10.20, 14.14 and -0.23 for the respective plotting position methods. From the aforementioned results, it was inferred that the distributions which best fitted the models were the log-logistic and the Weibull probability distributions. The Laplace probability plotting position in conjunction with the log-logistic distribution best described the datasets. This is in conformity with the model equation derived by [Hurford et al. \(2012\)](#). The Hirsh probability plotting position in conjunction with the Weibull probability distribution was also suitable for the description of the dataset and prediction of return period from the rainfall intensities and month of occurrence of such intensities. This is described by the model [equation 18](#)). It must, however, be noted that the censor period for this study is five years. This means that any rainfall return period predicted from the equations exceeding five years must be censored according to the theory of survival analysis.

$$\ln T = 0.012x_2 - 0.018x_1 - 2.382 \quad (17)$$

$$\ln T = 0.011x_2 + 0.015x_1 - 2.195 \quad (18)$$

These findings are slightly contrary to the report of [Obot \(2010\)](#), which reported that the Weibull probability position in conjunction with the normal distribution gave the highest fit for the Apoje sub-basin of Osun River. The combinations were reported to result in an R^2 value of 0.9950 and root mean square error (RMSE) value of 35.09 m³/s. This is contrary to the findings of [Agbede and Aiyelokun \(2016\)](#) who compared the Hazen, Weibull and California probability plotting positions using the normal, log-normal and log Pearson type III distribution for the prediction of flows in Ona river. The Hazen plotting position method was reported to perform best since it gave a higher regression coefficient (R^2) and minimal RMSE value. The log-Pearson (III) distribution gave the least absolute difference for all the plotting positions compared for the study. [Adeboye and Alatise \(2007\)](#) compared seventeen probability plotting positions using the Gumbel distribution. The study reported that the Hazen plotting position was the best for sample sizes ranging from 10 to 20.

Conclusions

The results obtained from the analysis of rainfall intensity and return period revealed that parametric models are essential tools for the estimation of time intervals between extreme and peak rainfall events in different months of the year. The combination of the Laplace plotting position and log-logistic distribution or the Hirsh plotting position and Weibull distribution fitted the datasets best. The established parametric models were suitable for the accurate prediction of return periods of peak rainfall events during any month of the year. The accelerated failure time approach was found to be suitable for the analysis of rainfall data by determining the best of several parametric models. In this research, the parametric models employed showed the relationships between rainfall return periods, rainfall intensity and month of the year.

Data availability

Underlying data

Zenodo: Underlying data for 'Parametric modelling of rainfall return periods in south-western Nigeria: Survival analysis approach'.

This project contains the following underlying data:

- Data file: rain.csv <https://doi.org/10.5281/zenodo.5797868> (Awodutire *et al.*, 2021a)
- Table 1: Parametric models of return period California probability plotting positions.
- Table 2: Parametric models of return period Weibull probability plotting positions.
- Table 3: Parametric models of return period Hazen probability plotting positions.
- Table 4: Parametric models of return period Adamowski probability plotting positions.
- Table 5: Parametric models of return period Blom probability plotting positions.
- Table 6: Parametric models of return period Chegodayev probability plotting positions.
- Table 7: Parametric models of return period Gringorten probability plotting positions.
- Table 8: Parametric models of return period Hirsh probability plotting positions.
- Table 9: Parametric models of return period Laplace probability plotting positions.

<https://doi.org/10.5281/zenodo.5799930> (Awodutire *et al.*, 2021c)

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](#) (CC-BY 4.0).

Software availability

Archived source code at time of publication: <https://doi.org/10.5281/zenodo.5800018> (Awodutire *et al.*, 2021b)

License: [Creative Commons Attribution 4.0 International](#)

References

- Adeboye OB, Alatise MO: **Performance of Probability Distributions and Plotting Positions in Estimating the Flood of River Osun at Apoje Sub-basin, Nigeria.** *Agricultural Engineering International: The CIGR E Journal*. 2007; **IX**. July, 2007.
- Agbede OA, Aiyelokun OO: **Establishment of a stochastic model for sustainable economic flood management in Yewa sub-basin, Southwest Nigeria.** *Civil Eng. Journal* 2016; **2**(12): 646–655.
[Publisher Full Text](#)
- Aiyelokun OO, Ojelabi A, Malomo S, *et al.*: **Efficient flood forecasting for the operation of hydraulic structures in a typical river basin.** *Int. J. Sci. Eng. Res.* 2017; **8**(11).
- Awodutire PO, Olapade AK, Oladapo OAI, *et al.*: **Assessing Survival Times of Breast Cancer Patients Using Type I Generalized Half Logistic Survival Model.** *JAMMR* 2018; **25**: 1–7.
[Publisher Full Text](#)
- Awodutire PO, Balogun OS, Olapade AK, *et al.*: **The modified beta transmuted family of distributions with application using exponential distribution.** *PLoS One* 2021; **16**(11): e0258512.
[Publisher Full Text](#)
- Awodutire PO, Sasanya BF, Ufuoma OG, *et al.*: **Dataset on Modelling Rainfall Return Periods and Intensity: Survival Analysis Approach** [Data set]. *Zenodo*. 2021a.
[Publisher Full Text](#)
- Awodutire PO, Sasanya BF, Ufuoma OG, *et al.*: **Tables showing the results of the models using the different plotting positions.** *Zenodo*. 2021b.
[Publisher Full Text](#)
- Awodutire PO, Sasanya BF, Ufuoma OG, *et al.*: **R Code on Modelling Rainfall Return Periods and Intensity: Survival Analysis Approach.** *Zenodo*. 2021c.
[Publisher Full Text](#)
- Davie T: *Fundamentals of hydrology*. Routledge Fundamental of Physical Geography. Ed. By John Gerrard. 2008.
- Ewemoje TA, Ewemoje AS: **Best distribution and probability positions for daily maximum flood estimation at Ona River in Ogun-Oshun River basin, Nigeria.** *Agric. Eng. Int. CIGR J.* 2011; **2011**: 3.
- Houessou-Doussou EAY, Gathanya JM, Njuguna M, *et al.*: **Flood Frequency Analysis Participatory GIS and Rainfall Data For Two Stations in Narok Town, Kenya.** *Hydrology* 2019; **6**: 90.
[Publisher Full Text](#)
- Hurford AP, Parker DJP, Priest SJ, *et al.*: **Validating the return period of rainfall thresholds used for extreme rainfall alerts by linking rainfall intensities with observed surface water flood events.** *J. Flood Risk*

Management 2012; **5**(5): 134–142.

[Publisher Full Text](#)

IBM Corp. Released: *IBM SPSS Statistics for Windows, Version 27.0* Armonk, NY: IBM Corp; 2020.

[Reference Source](#)

Jonathan G: *Survival Analysis and European Union Decision Making. European Union Poilitics*. SAGE Publications; 2014; vol. **8**: 155–179.

Laitinen EK: **Survival Analysis and Financial Distress Prediction: Finnish Evidence Review of Accounting and Finance**. 2005.

[Publisher Full Text](#)

Laura KR, Richard MV: **Reliability, return periods and risk under nonstationarity**. *Water Resour. Res.* 2015; **51**(8).

[Publisher Full Text](#)

Lee M: **Business Bankruptcy Prediction Based on Survival Analysis Approach**. *International Journal of Computer Science and Information Technology*. 2014; **6**: 103–119.

[Publisher Full Text](#)

Naseri P, Baghestani AR, Momenyan N, *et al.*: **Application of a Mixture Cure Fraction Model Based on the Generalized Modified Weibull Distribution for Analyzing Survival of Patients with Breast Cancer**. *Int. J. Cancer Manag.* 2018; **11**: 2018.

[Publisher Full Text](#)

Obot N, Chendo M, Udo S, *et al.*: **Evaluation of rainfall trends in Nigeria for 30 years (1978-2007)**. *Int. J. Phys. Sci.* 2010; **5**.

Ogungbenro SB, Morakinyo TE: **Rainfall Distribution and Change Detection Across Climatic Zones in Nigeria**. *Weather Clim. Extremes*. 2014; **5-6**: 1–6.

[Publisher Full Text](#)

Olatunde AF, Adejoh I: **Annual exceedance probability and return periods of rainstorms in Lokoja**. *Int. J. Soc. Sci.* 2017; **11**.

R Core Team: *R: A language and environment for statistical computing* Vienna, Austria: R Foundation for Statistical Computing; 2017.

[Reference Source](#)

Salase AE, Agyimpomaa DEE, Selasi DD, *et al.*: **Precipitation and rainfall types with their characteristic features**. *J. Nat. Sci. Res.* 2015; **5**(20).

Santos EB, Lucio PS, Silva CMSE: **Seasonal Analysis of Return Periods for Maximum Daily Precipitation in the Brazilian Amazon**. *J. Hydrometrol.* 2015; **16**: 973–984.

[Publisher Full Text](#)

Witzany Y, Rychnovsky M, Charamza P: **Survival Analysis in LGD modelling**. *European Financial and Accounting Journal* 2012; **7**: 6–27.

[Publisher Full Text](#)

Ybanez R: **Understanding Rainfall Return Periods** *Project NOAA Open-File Reports* 2013; **1**: pp 3–4.2362 7409.

Yahaya AS, Nor NM, Jali NRM, *et al.*: **Determination of probability plotting position for type 1 extreme value distribution**. *J. Appl. Sci.* 2012; **12**(14): 1501–1506.

[Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 11 April 2022

<https://doi.org/10.5256/f1000research.79632.r129245>

© 2022 Chand S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Sohail Chand

College of Statistical and Actuarial Sciences, University of the Punjab, Lahore, Pakistan

In this paper, the authors have fitted a survival model to rain return periods. They have considered nine different methods for the computation of the return period and four survival models. Here are a few suggestions for the improvement of the paper.

1. A statistical summary of data should be provided. For this purpose, descriptive statistics and appropriate graphs can be used. Moreover, it will help readers to look at the statistical behavior of rain overall data.
2. Precisely define x_1 and x_2 predictors especially their types i.e. nominal or ordinal or scale. It is important to discuss how the categorical type variables are considered in the model and how their coefficients will be interpreted.
3. Graphical presentation e.g. scatter diagrams can be helpful to visualize the relationship between response and predictors.
4. Models' out-of-sample performance should be evaluated e.g. using some cross-validation techniques.
5. It would be worth considering, if possible, also some other performance measures in addition to Akaike Information Criterion.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Statistical Modeling

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 08 February 2022

<https://doi.org/10.5256/f1000research.79632.r120739>

© 2022 Xavier T. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Thomas Xavier

Department of Statistical Sciences, Kannur University, Kannur, Kerala, India

The authors have used survival analysis for the prediction of rainfall return periods. Twenty years of annual rainfall data were obtained from three meteorological stations for the analysis. Akaike Information Criterion was used to determine the best-fitting model. This work is useful to predict the return periods of peak rainfall events during any month of the year that would occur in south-western Nigeria.

The authors have considered nine different probability plotting procedures and the best model out of these has been obtained. Four different survival models namely, exponential, log-logistic, Weibull, and log-normal are considered. The paper should be really helpful to understand how survival analysis can be used to predict the rainfall return period.

It would have been great if the authors could show plots or explain the distribution of the datasets. Also, they don't discuss if there are any missing observations or not. It is something that can be considered for further studies.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Statistical distributions, survival analysis.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research