Journal of Geoscience and Environment Protection, 2025, 13(12), 151-168
https://www.scirp.org/journal/gep
ISSN Online: 2327-4344
ISSN Print: 2327-4336

Best Model Selection for Extreme Values Analysis Using the Peaks-Over-Threshold Method and the Annual Maximum Method: Case Study of Bamako, Mali

Alikalifa Sanogo[1,2*], Roland Songotu Kabange[3], Prince Appiah Owusu[3], Ousmane Mohamed Maïga[4], Lassina Dit Papa Koné[5]

[1]Graduate Research Program on Climate Change and Water Resources, West African Science Service Center on Climate Change and Adapted Land Use (WASCAL), University of Abomey-Calavi, Abomey-Calavi, Benin
[2]Laboratory of Applied Hydrology, University of Abomey-Calavi, Abomey-Calavi, Benin
[3]Department of Civil Engineering, Faculty of Engineering and Technology, Kumasi Technical University, Kumasi, Ghana
[4]Graduate Research Program on Climate Change and Human Habitat, West African Science Service Center on Climate Change and Adapted Land Use (WASCAL), Federal University of Technology, Minna, Nigeria
[5]Graduate Research Program in West African Climate Systems, West African Science Service Center on Climate Change and Adapted Land Use (WASCAL), Federal University of Technology Akure, Akure, Nigeria
Email: *AlikalifaSanogo@gmail.com, sanogo.a@edu.wascal.org

How to cite this paper: Sanogo, A., Kabange, R. S., Owusu, P. A., Maïga, O. M., & Koné, L. D. P. (2025). Best Model Selection for Extreme Values Analysis Using the Peaks-Over-Threshold Method and the Annual Maximum Method: Case Study of Bamako, Mali. *Journal of Geoscience and Environment Protection, 13,* 151-168.
https://doi.org/10.4236/gep.2025.1312009

Received: October 10, 2025
Accepted: December 2, 2025
Published: December 5, 2025

Abstract

Heavy rains and floods are long considered critical societal concerns due to adverse effects on society, environment, and economies. The best appropriate identification for rainfall distribution is equally of significant concern to society and planners due to its application in the hydrological and water resources management sectors, and agricultural planning. Two methods for extreme values theory namely the annual maximum (AM) series method and the peaks-over-threshold (POT) method are generally employed for extreme values analyses. This study therefore compared the results of both methods for ARC2 daily rainfall data at Bamako-Senou station for the period 1991 - 2021. Five (05) commonly used distribution functions, namely the Normal, Log-Normal (LN), Gumbel type I, Gamma and Pearson type 3 (P3) distributions were used to fit the AM data. The method of moments (MOM) and the method of maximum likelihood estimation (MLE) were employed for parameters estimation in AM analyses. The generalized Pareto (GP) distribution was also used to fit the peaks over threshold (POT) method. The results indicated that the P3 distribution gave better result than other distributions when parameters were estimated with the MLE. The LN distribution was also best fit distribution to AM series when parameters were estimated with the MOM. The P3 distribution gave higher quantile estimates than other distributions. The POT method gave better results for quantiles estimation than the AM series. It is recommended that further study should include various truncation levels and tests for the choice of an optimal threshold in the POT method.

Keywords

Annual Maximum (AM) Series, Best Distribution Models, Extreme Values Analysis and Peaks-Over-Threshold (POT) Method

1. Introduction

One of the major concerns for society is torrential rain and floods as their impacts on the environment, society and economics are devastating (Panthou et al., 2012). The design of reliable and effective hydraulic structures for managing flood can be associated primary with a perfect understanding of future extreme conditions. The estimation of return levels has considerable engineering importance especially in urban drainage system's area. Many statistical methodologies for such estimation have been developed in recent years. Generally, two main methods are used to define extreme values (Thompson et al., 2009). First, the time period over which the data are collected is divided into block, the most extreme value defined in each block (monthly or daily maxima) is then used for future analysis (Thompson et al., 2009). The second method is based on exceedances over a threshold or excess over a threshold (Thompson et al., 2009).

Extreme events are those that have low occurrence in nature but have high severity or impact (Sharma, 2016). Modelling such events occurrence and magnitude is defined by a statistical method called extreme value theory (EVT). For instance, floodwalls are not designed for common occurrences, but rather for rare and disastrous occurrences (Sharma, 2016). EVT is based on asymptotic results, so, the used data to model events is often a relatively small subset of the whole dataset generally above the 90th or 95th quantile (Sharma, 2016). Fisher and Tippett (1928) indicated that the maximal behavior can be defined for a single process by three asymptotic limits describing extremes distributions assuming the variables are independent. Gumbel (1958) also codified this theory in a statistics book of extremes, comprising the Gumbel distributions. An annual maximum (AM) sample is defined by the maximum peak flow of each year (Bezak et al., 2014). However, defining the sample in this manner can result in loss of information (Langbein, 1949; Lang et al., 1999; Bačová-Mitková & Onderka, 2010). For instance, some peaks that were not AMs, but are nonetheless relatively high, are not taken into account in AM analyses. Another limitation of the AM series is that the sample size is typically quite small (Bezak et al., 2014).

Probability distribution functions are defined to modeling magnitudes of exceedances in AM rainfalls at a single site (Kite, 1977; Hosking & Wallis, 1997;

Stedinger et al., 1993). Though, more than one distribution may well fit the data, the best model selection can be difficult (Salas et al., 2013). However, the model selection remains one of the most challenges in engineering practice, since there is no one-size-fits-all approach on which distribution (s) should be employed for frequency analysis of extreme rainfalls (Bezak et al., 2014). Guidelines have been developed by countries suggesting which distribution to use (Bezak et al., 2014). However, finding the best method remains a challenge to engineers (Bezak et al., 2014). The EVT is based on three (3) extreme event types of distribution often used to model extreme values, the Gumbel, Frechet and Weibull models, sometimes referred to as type I, type II and type III distributions respectively. These three (3) distributions can be linked by a single distribution called the Generalized extreme values distribution (GEV). For extreme values distribution above a specific threshold, the peaks over threshold (POT) method is generally applied to extreme random variables, and the method, which fits that distribution is generally referred to as the Generalized Pareto (GP) distribution. Under certain conditions, however, the GP distribution family has important connections with the GEV family. However, the shape parameter is expected to be asymptotically the same as the threshold $u \to \infty$ and other parameters can be found through theoretical relations. For hydrological extreme events modelling, previous studies found that fitting the GP distribution on a reasonable number of exceedances over a proper threshold leads to more accurate extreme quantile estimates than fitting the GEV distribution on annual maxima (Cunnane, 1973; Madsen et al., 1997).

Methods developed to estimate parameters for improved efficiency and robustness of distributions include Maximum Likelihood Estimation (MLE), and the Method of Moment (MOM), Probability Weighted Moments (PWM), and Linear Moments (LM). Sankarasubramanian and Srinivasan (1999) compared the MOM and the LM for parameter estimation of the generalized normal (GN), generalized extreme value (GEV), Pearson type 3 (P3) and generalized Pareto (GP) distributions. The performance of the methods depended on the sample size and data skewness. Ahilan et al. (2012) analyzed data from 172 gauging stations in Ireland and investigated which of the EV distributions was most appropriate for flood frequency analysis (FFA). The Gumbel distribution was preferred over two (2) other types of EV distributions (type II or Frechet, and type III or Weibull). Again, Seckin et al. (2011) made FFA on the data from 543 gauging stations in Turkey, and the LM was used for the distribution parameters estimation. The GEV distribution gave better goodness-of-fit values than the other distributions, such as the generalized logistic (GL) and LP3 distributions. Rosbjerg et al. (1992) applied the MOM and PWM for parameter estimation for partial duration series or the POT method. When the parameters were estimated with MOM, they found that the mean square error (MSE) was smaller than in the case of PWM.

In this study, the maximum likelihood estimation (MLE) was used to estimate the GP distribution parameters. The method of moments (MOM) and the MLE were used to estimate parameters in annual maximum (AM) series. The objectives of this study were: (a) to compare statistical distribution functions and explore the best fit distribution to AM series; (b) to compare the POT and the AM series method.

2. Materials and Methods

2.1. Data sources

The main data source for this study was the National Meteorological Agency of Mali (MALI-METEO), which provided monthly rainfall data from Bamako-Senou rain gauge station for a period 1991 - 2020. As extreme rainfalls are generally of daily rainfall series, African Rainfall Climatology version-2 (ARC2) data was used to model extreme rainfalls in this study. Monthly rainfall data from MALI-METEO was used to validate the gridded ARC2 data, the validation of ARC2 with gauge data is given by (Sanogo et al., 2023). The location of Bamako-Senou's rain gauge station is provided in (**Table 1**).

Table 1. Bamako-Senou's rain gauge characteristics.

Station	ID	Longitude	Latitude	Elevation (m)
Bamako-Senou	61,291	7.95 ˚W	12.53 ˚N	380

Missing values presented in daily rainfalls data is of 68 out of 11 323 days. A missing rate of 5% or less in the datasets is inconsequential (Schafer & Graham, 2002). The summary statistics of ARC2 data is presented in **Table 2** below.

Table 2. Summary statistics of ARC2 data from 1991 to 2021.

Minimum (mm)	1st Quantile (mm)	Median (mm)	Mean (mm)	3rd Quantile (mm)	Maximum (mm)	NA's
0.00	0.00	0.00	2.621	0.00	178.8	68

2.2. Study area

As the capital and administrative center of the Republic of Mali located at 8˚0'10" W and 12˚38'21" N on both sides on the bank of Niger river (**Figure 1**), Bamako has a 245 km^2 land area a population of about 1.3 million, and a population density of 5300 people per square kilometer (Keita et al., 2020). Bamako has one rainy season which usually lasts about six (6) months, beginning mid-May to October, peaking usually in August. However, the dry season occurs between February and the first two (2) weeks of May. The Intertropical Convergence Zone (ITCZ) largely controls Bamako's rainfall characteristics by oscillations across the North and South of the African continent, which usually brings rainfall to South Mali between June and October each year (World Bank Group, 2011).

The average monthly rainfall in Bamako ranges from 54.1 mm in May to 290.2 mm in August, with a mean annual rainfall of 991.3 mm (Keita et al., 2020). The warmest and coolest months are therefore April and May respectively, and December to January representing the months without rain (Keita et al., 2020). Daily maximum and minimum temperature of 40˚C and 17˚C respectively are recorded in April and December, with an average annual temperature of 29˚C usually recorded in March. Floods are common in Bamako in the rainy season, as the city can usually receive more than 600 mm of rainfall. Bamako is located in a valley covered with sandstone deposits and has two (2) types of soil: the soil formed by rock formation, and soil caused by lateralization and alluvial formation occupying the primary and secondary riverbeds and tributaries (Keita et al., 2020). Bamako's vegetation is largely characterized by savannah forest and rivers (Keita et al., 2020).

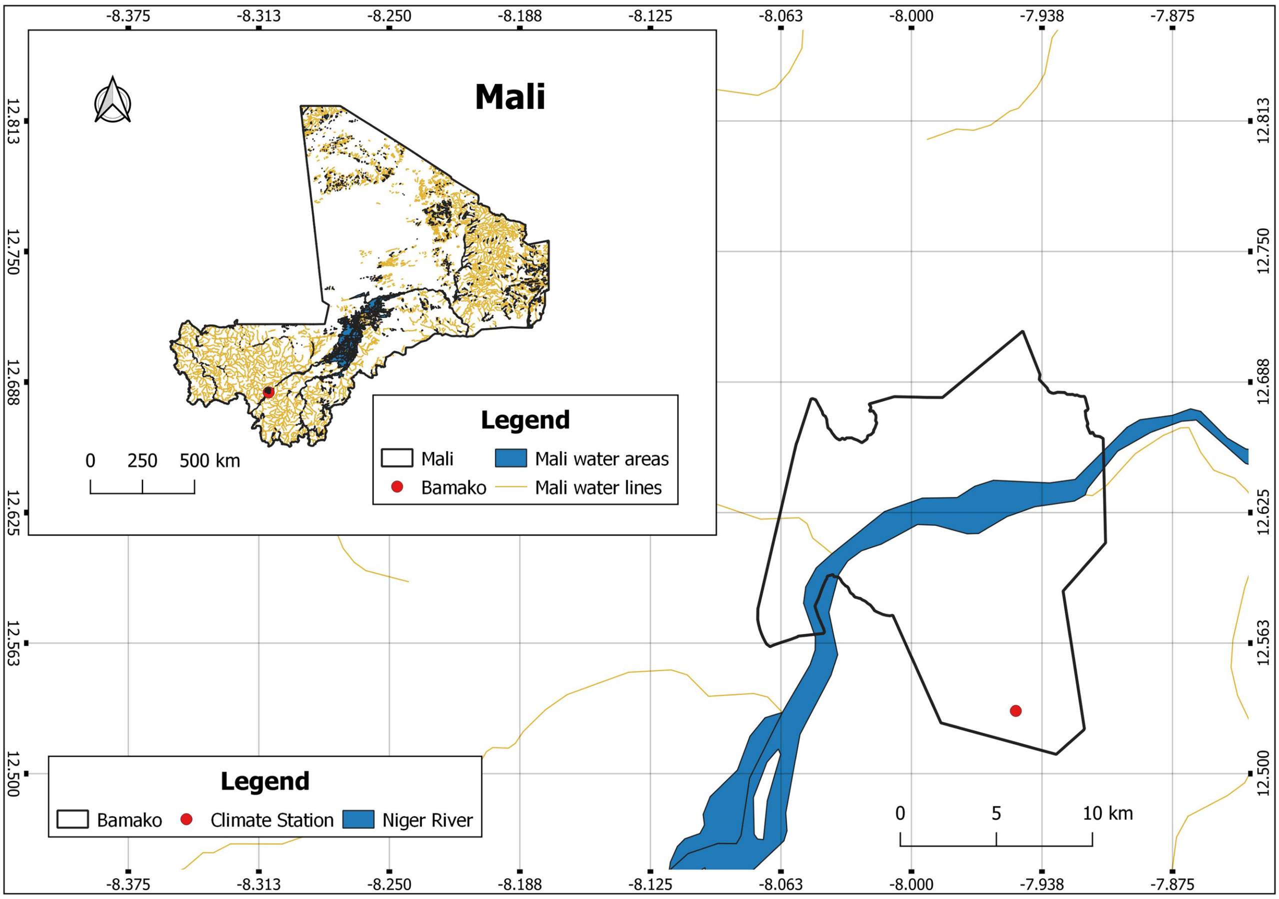

Figure 1. Location of the Bamako area with climate station.

2.3. Methods

2.3.1. Quantile Method for Threshold Selection

The quantile of a distribution is a quantity Q_P such that a proportion p of the population must be less than or equal to Q_P. For instance, the quantile Q_{99} is the value such that 99% of the population is less than or equal to Q_{99}.

Consider a population of discrete values or a continuous population density, the $k-thq$-quantile for a variable X can be written as:

$$P[X < x] \le k/q \quad \text{or,} \quad P[X \ge x] \ge 1 - k/q \text{ and } P[X \le x] \ge k/q \tag{1}$$

The “$p-quantile$” is based on a real number p with $0 < p < 1$ then p replaces k/q in the above formulas.

2.3.2. Generalized Pareto (GP) Distribution Method

The peaks over threshold (POT) was used to model the tail of the distribution, the approach consisted in the analysis of the data that are considered as extreme observations, the data that surpass a threshold level u. Let us consider $X_1,\ldots,X_n$ a sequence of independent and identically distributed random variables, having distribution function F, the focus was on conditional probability such as:

$$F_u(y) = P(X \le u + y | X > u) \quad \text{then,} \quad F_u(y) = \frac{F(u+y) - F(u)}{1 - F(u)},$$

$0 \le y \le x_F - u$.

With x_F is the endpoint of *F*, the following result gives an approximation to this probability for high values of the threshold u. The distribution function of $(X-u), conditioned\ to\ X > u,$ is approximately by the generalized Pareto (GP) distribution given as follows:

$$F_u(y) \approx G_{\xi,\sigma}(y), u \to \infty$$

With

$$G_{\xi,\sigma}(y) = \begin{cases} 1 - \left(1 + \dfrac{\xi y}{\sigma}\right)^{-\frac{1}{\xi}} & for\ \xi \ne 0 \\ 1 - exp\left(-\dfrac{y}{\sigma}\right) & for\ \xi = 0 \end{cases} \tag{2}$$

where $y \ge 0\ for\ \xi \ge 0\ and\ 0 \le y \le \frac{\sigma}{\xi}\ for\ \xi < 0$, $\xi\ and\ \sigma$ are the shape and scale of the GP distribution.

Or

$$F_u(y) = \frac{F(u+y) - F(u)}{1 - F(u)} \Rightarrow F_u(y) = \frac{F(x) - F(u)}{1 - F(u)} and\ F(x) = [1 - F(u)]F_u(y) + F(u)$$

Denoting by $\delta_u = P(X > u)$ and from the conditional distribution the expression of

$$P\{X > x | X > u\} = \left\{1 + \xi\left(\frac{x-u}{\sigma}\right)\right\}^{-1/\xi} \ and\ therefore, P(X > x) = \delta_u \left\{1 + \xi\left(\frac{x-u}{\sigma}\right)\right\}^{-1/\xi} \tag{3}$$

The levels x_m that is exceeded on average once every m observations is the solution of

$$\frac{1}{m} = \delta_u \left\{1 + \xi\left(\frac{x_m - u}{\sigma}\right)\right\}^{-1/\xi} \ then, x_m = u + \frac{\sigma}{\xi}\left((m\delta_u)^{\xi} - 1\right) \tag{4}$$

2.3.3. Fitting Probability Distributions to Annual Maximum (AM) Series

For fitting probability distribution to AM series, the Normal, Lognormal (LN), Gamma, Pearson 3 and Gumbel distributions were fitted to AM series variables

(Table 3).

Table 3. Summary table of distributions used to fit AM series in Bamako.

Distribution	Pdf, equation for MOM and log-likelihood function	Range
Normal	$f(x)=\frac{1}{\sigma_x\sqrt{2\pi}}exp\left[-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right]$ μ_x and $\sigma_x^2;\gamma_x=0$ $lnL(\mu_x,\sigma_x)=-\frac{n}{2}ln(2\pi)-nln\sigma_x-\frac{1}{2\sigma_x^2}\sum_{i=1}^{n}(x_i-\mu_x)^2$	$-\infty<x<\infty$
LN	$f(x)=\frac{1}{x\sigma_y\sqrt{2\pi}}exp\left[-\frac{1}{2}\left(\frac{lnx-\mu_y}{\sigma_y}\right)^2\right]$ $\mu_x=exp\left(\mu_y+\frac{\sigma_y^2}{2}\right)$, $\sigma_x^2=\mu_x^2\left[exp(\sigma_y^2)-1\right]$; $\gamma_x=3CV_x+CV_x^3$ $lnL(\mu_y,\sigma_y)=-nln(2\pi)-nln\sigma_y-\sum_{i=1}^{n}lnx_i-\frac{1}{2\sigma_y^2}\sum_{i=1}^{n}(lnx_i-\mu_y)^2$	$x>0$
P3	$f(x)=\frac{1}{\beta\Gamma(\alpha)}\left(\frac{x-c}{\beta}\right)^{\alpha-1}exp\left(-\frac{x-c}{\beta}\right)$ $\alpha=\left(\frac{2}{\gamma_x}\right)^2,\beta=\frac{\sigma_x\gamma_x}{2}and\ c=\mu_x-\alpha\beta$ $lnL(\alpha,\beta,c)=-nln(\beta)-nln(\Gamma(\alpha))-n(\alpha-1)ln(\beta)+(\alpha-1)\sum_{i=1}^{n}ln(x_i-c)-\frac{1}{\beta}\sum_{i=1}^{n}(x_i-c)$	$\beta>0$
Gamma	$f(x)=\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}exp(-\beta x)$ $\alpha=\left(\frac{\mu_x}{\sigma_x}\right)^2,\beta=\frac{\mu_x}{\sigma_x^2}$ $lnL(\alpha,\beta)=n\alpha ln(\beta)-nln(\Gamma(\alpha))+(\alpha-1)\sum_{i=1}^{n}lnx_i-\beta\sum_{i=1}^{n}x_i$	$\alpha>0;\beta>0$
Gumbel	$f(x)=\frac{1}{\alpha}exp\left[-\frac{x-\xi}{\alpha}-exp\left(-\frac{x-\xi}{\alpha}\right)\right]$ $\mu_x=\xi+0.5772\alpha$ $\sigma_x^2=\frac{\pi^2\alpha^2}{6}\approx1.645\alpha^2;\gamma_x=1.1396$ $lnL(\xi,\alpha)=-nln(\alpha)-\frac{1}{\alpha}\sum_{i=1}^{n}(x_i-\xi)-\sum_{i=1}^{n}exp\left(-\frac{x-\xi}{\alpha}\right)$	$-\infty<x<\infty$

where μ_x, σ_x^2 and γ_x are the mean, standard deviation and skewness of x, respectively. $Y=lnx$; parameters $\alpha,\beta\ and\ c$ represent the shape, scale and location respectively of the P3 distribution, and Γ is the gamma function.

One of the purposes of this article was to explore which distribution functions give the best fit with the AM series sample. The Chi-square test and the Kolmogorov-Smirnov test were used to check the adequacy of the tested distribution functions at 0.05 significance level. The root mean square error (RMSE), and the Akaike information criterion (AIC) were used to define the best fit distribution.

Table 4 below gives the test statistics for model selection.

Table 4. Test statistics for model selection.

Statistics	Formula	Best Value
Root Mean Square Error	$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(O_i - P_i)^2}$	0
AIC	$AIC = nlog\left(RMSE^2\right) + 2k$	*Minimum*
K-S	$D = \max\left\|F(x) - \hat{F}(x)\right\|$	-
Chi-square	$\chi^2 = \sum_{i=1}^{k}\frac{\left(f_i - \hat{f}_i\right)^2}{\hat{f}_i}$	-

where F is the actual cumulative relative frequency and $\hat{F}$ is expected cumulative relative frequency. The estimated value for the test statistic should then be compared with the critical value selected based on the significance level and the number of data points. f_i is the actual frequency and $\hat{f}_i$ is the expected frequency, i is the number of classes and k is the total number of categories. This estimation is close enough for testing the goodness-of-fit whenever the expected frequency in any category is equal or greater than 5. χ^2 is chi-square distributed with $k-1$ degree of freedom. This form of AIC presented in Table 4 was used by Karmakar and Smonovic (2008, 2009).

3. Results and Discussion

3.1. Threshold Estimation for Maximum Daily Rainfalls

The threshold estimation was done by the method of quantile, the mean excess plot using the peak over threshold (POT) method was used to set the optimal threshold. The percentage of missing values out of the number of rainy days was 0.6%. A missing rate of 5% or less in the datasets is inconsequential (Schafer & Graham, 2002). The mean excess plot should be linear, and the parameter estimates should be stable above the threshold at which the GP model becomes valid. Confidence intervals were added to take account the effects of estimation uncertainty in the evaluation of optimal threshold. From Figure 2(a), a threshold above 42.9 mm is not appropriate because the confidence bounds increase significantly. This is visible in Figure 2(b) that there is not enough data above the threshold 42.9 mm. Meanwhile by taking a threshold below 33.8 mm, the variances in Figure 2(b) are too small, which means that the number of observations is too large, and the asymptotic approximation will be violated in GP distribution. Therefore, it can be approximated that the threshold should be around 33.8 - 42.9 mm. In this case, the Poisson distribution may help define the optimal threshold. As the variance of the data is greater than the mean of the observed values, the Poisson distribution (in which the variance is equal to the mean) will not be a good candidate leading to low p-values. In addition, as the mean residual plot must be approximately linear with the threshold, the linearity still exists between 33.8 - 42.9 mm. Therefore, the threshold which corresponds to 99th percentile of the data was chosen (40.8 mm) as an optimal threshold. The number of exceedances for the threshold 40.8 mm was 112 exceedances corresponding to on average 3.6 peaks per year. When the quantile Q_{95} was chosen, the number of exceedances above the threshold was 536, corresponding to on average 17.3 peaks per year. Moreover, when the quantile Q_{90} was chosen the number of peaks per year on average was 36.3 peaks per year. However, various studies suggested the threshold value should be high so that the model assumptions are not violated (Lang et al., 1999). Jenkinson (1955) and Bernardara et al. (2014) showed some strategies to select the best threshold. The best threshold should be at least greater than the mean but cannot be too high (Henrique & Castro, 2019). **Figure 2(c)** and **Figure 2(d)** show the number of peaks-over-threshold 40.8 mm and the generalized Pareto (GP) distribution diagram of the threshold 40.8 mm.

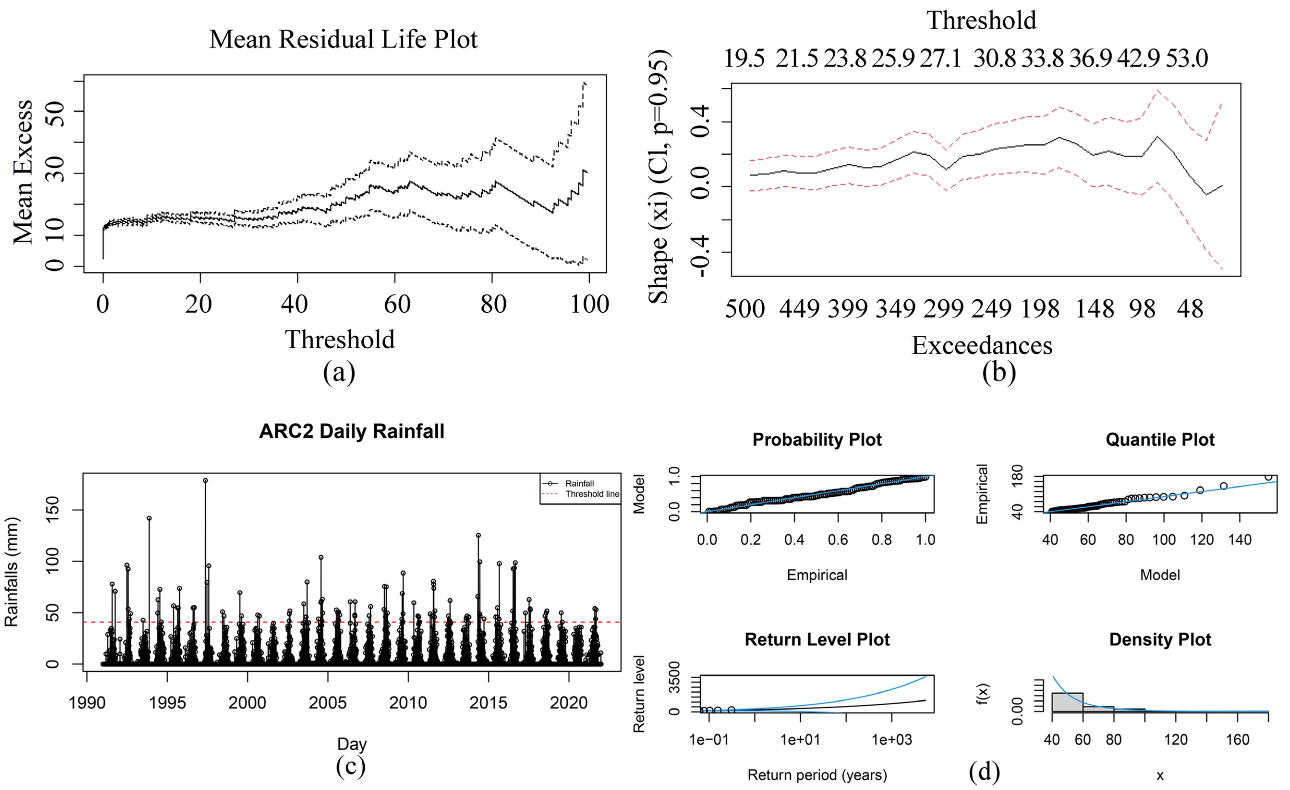

Figure 2. (a) The mean excess plot; (b) the shape parameter of the GP distribution; (c) the peaks over the threshold $u = 40.8mm$ for the period 1991 - 202 and (d) the GP distribution diagram.

Fofana et al. (2022) on urban Flash flood and extreme values analysis in Bamako revealed that the threshold at which flood may occur in Bamako was about 47 mm. Moreover, Panthou et al. (2014) showed that extreme thresholds in Sahel regions were in most cases between 20 to 50 mm. Another study on the debilitating floods in Sahel countries, principally the flash flood which triggered in most Sahel countries, predicted that most floods which occurred in Sahel countries were thresholds of 99th percentile (Sarr & Lona, 2018). **Table 5** below gives parameters estimation of the GP distribution and the goodness-of-fit AIC value.

Table 5. GP distribution parameters estimation using the method of Maximum Likelihood and the AIC value.

GP distribution parameters Estimate using the POT method	
Threshold u	40.8
Number of Exceedances	112
Scale	15.1269
Shape	0.1852
AIC	877.9746

It is well documented that the specific choice of the threshold value would affect the magnitude of exceedances beyond the selected threshold and the arrival time of the extreme events (Zawiah et al., 2009). The threshold $u = 40.8mm$ should be large enough to ensure that the observations are independent without loss of appropriate information. Furthermore, choosing a low threshold value could introduce serial dependence of both occurrence times and magnitudes, thereby violating the assumption of independence (Zawiah et al., 2009).

3.2. Best Fit Probability Distribution for Annual Maximum (AM) Series

Several studies have been conducted over the year for the best model selection to describe extreme rainfalls in various areas. To the best of our knowledge, we have not been across any specific study online to describe the best model selection for extreme rainfalls in Bamako area. Many probability distribution models have been suggested for representing the distribution of annual hydrologic extremes at a single site (Chow, 1964; Kite, 1977; Stedinger et al., 1993; Hosking & Wallis, 1997; Rao & Hamed, 2000; Salinas et al., 2014a; Salinas et al., 2014b). Distributions such as exponential, Gamma, Weibull, log-normal, logistic, Gumbel, GP and GEV distributions have been commonly employed in hydrological studies as referenced by (Sharma & Singh, 2010; Hussain et al., 2022; Agbonaye & Izinyon, 2017; Oseni & Ayoola, 2013; Alam et al., 2018; Coronado-Hernández et al., 2020). For instance, in an extreme rainfalls study in Bangladesh (Mahtab, 2024) found that the Generalized Extreme Value (GEV), Generalized Normal (GNO) and P3 distributions are the best models that can be used for describing accurately the distributions of annual maximum rainfalls in Bangladesh. The national guidelines of different countries suggest the use of different distributions. For instance, Log-Pearson 3 has been recommended in the US in Bulletin 17B (Griffis & Stedinger, 2007). The generalized extreme value (GEV) distribution is recommended in various countries in Europe, including Australia, Germany, Italy, and Spain (Salinas et al., 2014a). The GEV distribution and LP3 are also recommended in Australia (Ball et al., 2016). However, many other distributions have also been employed popularly in various regions including the generalized Pareto (GP) distribution in Belgium, Gumbel distribution in Finland and Spain and the generalized logistic (GLO) distribution in the UK (Salinas et al., 2014a). In Canada, the choice of a specific distribution is not mandatory, however, LP3, Log-normal three parameters (LN3), GEV, Gumbel have been used popularly (Chow & Watt, 1992; Alila, 1999; Hansen, 2015). Recent study for instance in Kenya found that when applying the Normal, Lognormal, Weibull, GEV, Exponential, Gamma, Logistic, Gumbel, Uniform and the GP distributions to modelling maximum temperature and total rainfall, the result showed that the GEV, Gamma, and Log-Normal distributions were well-suited for both maximum temperature and total rainfall datasets (Otieno et al., 2025). In this study, five (5) commonly used probability distribution functions namely, the Normal, Log-normal (LN), Gamma, Gumbel type 1 and Pearson 3 (P3) distribution were used to fit the AM series. **Figure 3** below shows the empirical and density plots of various distributions using the MLE and MOM as parameters estimations.

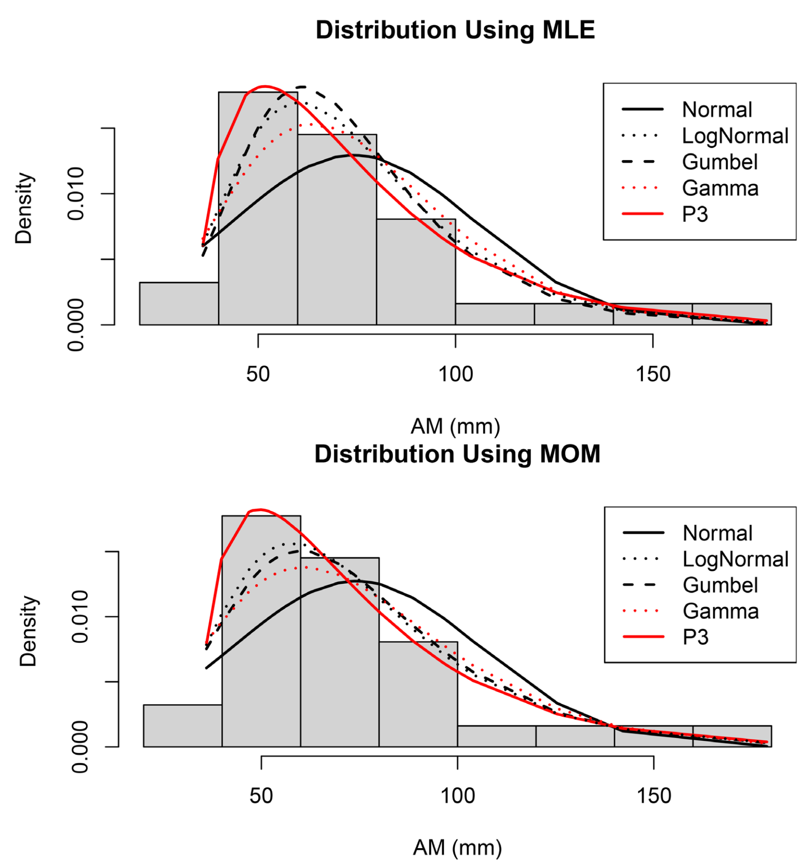

Figure 3. Empirical and density plots of distributions using the MLE as parameters estimation in the left and the MOM in the right.

The K-S, Chi-square test, AIC, and the RMSE tests were applied to the AM series data for the period 1991 - 2021. The results of the tests for the case when the parameters were estimated with the MLE and the MOM are shown in Table 6 below. All the distributions except the Normal distribution using both the MLE and the MOM for parameters estimation were passed for the Chi-square test at 0.05 significance level. When parameters were estimated with the MLE, P3 gave better result than other distributions followed the Gumbel distribution and the LN distribution as third best fitted distribution, the gamma and the normal distributions gave weak results compared to other distributions. When the MOM was used for parameters estimation, the LN distribution comes first as the best fitted distribution followed by the P3 distribution as second and the Gumbel distribution as third, the Gamma and Normal distribution gave weak results compared to other distributions. All the distributions passed the K-S test at 0.05 significance level, the critical value for the K-S test when the number of categorical intervals is 9 was 0.43001. The P3 distribution using the MLE and the LN distribution using the MOM can be used for AM series frequency analysis in the study area.

Table 6. Tests results for the AM method for the period 1991 - 2021 and parameters estimated with the MLE and the MOM, the significance level for the test was chosen as 5%.

Distribution/Test	RMSE	AIC	K-S	χ^2	χ^2 P-values
		Maximum Likelihood Estimation			
Normal	34.16578	222.9359	0.1742684	21.683	0.005539
LN	32.59133	220.0109	0.0702064	5.9374	0.6542
Gumbel	32.54013	219.9134	0.07495211	7.6818	0.4652
Gamma	33.19853	221.1554	0.08578309	9.0084	0.3416
P3	31.5673	220.0316	0.04129497	3.5636	0.8942
		Method of Moments Estimation			
Normal	34.10679	222.8288	0.1779364	20.175	0.009694
LN	31.81689	218.5198	0.05945942	4.399	0.8195
Gumbel	32.08416	219.0385	0.07297036	5.1768	0.7385
Gamma	32.23384	219.327	0.09177828	6.6278	0.5773
P3	31.81689	220.5198	0.04275357	3.0146	0.9334

In comparing the peaks-over-threshold method and the annual maximum method for flood frequency analysis, Bezak et al. (2014) found that when the distributions parameters were estimated with the MOM and L-moment, the LP3 distribution gave the best results in most of the applied tests. However, when the distributions parameters were estimated using the MLE, the best results were obtained using the P3 distribution, while the lognormal (LN) and LP3 distributions were second and third respectively.

3.3. Comparison between the POT and AM Series Method

The comparison between the POT and the AM series was based on ARC2 daily rainfall values for the period 1991 - 2021. The POT sample was 3.6 times higher than the AM sample. The truncation level selection is one of the disadvantages of the POT method. Tavares and da Silva (1983) found that a POT sample with an average of two peaks per year gave better results than the AM method. Cunnane (1973) reported that an average of at least 1.65 peaks per year should be selected in the POT method to gain advantage over the AM method. Moreover, Madsen (1996) found that more than 0.91 peaks per year should be selected for the POT model to have advantage over the AM method when the parameters are estimated with the MOM. From **Figure 4**, for smaller return periods the POT method and the AM method gave approximately similar results when parameters are estimated with the MLE and the MOM. However, for higher return periods the POT gave higher quantile rainfalls than the AM series method and described better than the AM series. The P3 distribution when parameters are estimated with the MLE and the MOM gave best and higher quantiles than other distributions in AM method. The POT method using the GP distribution also gave best result for quantiles estimation than the P3 distribution in AM series. In conclusion, therefore, the POT method gave best result than the AM series for estimating quantiles. For flood frequency analyses (FFA) in the region, the POT method could be used to inform the return periods of upcoming flash floods in Bamako. The design of most urban water infrastructures mostly depends on return periods of extreme events; this could also be an alternative using the POT method to design those infrastructures in Bamako.

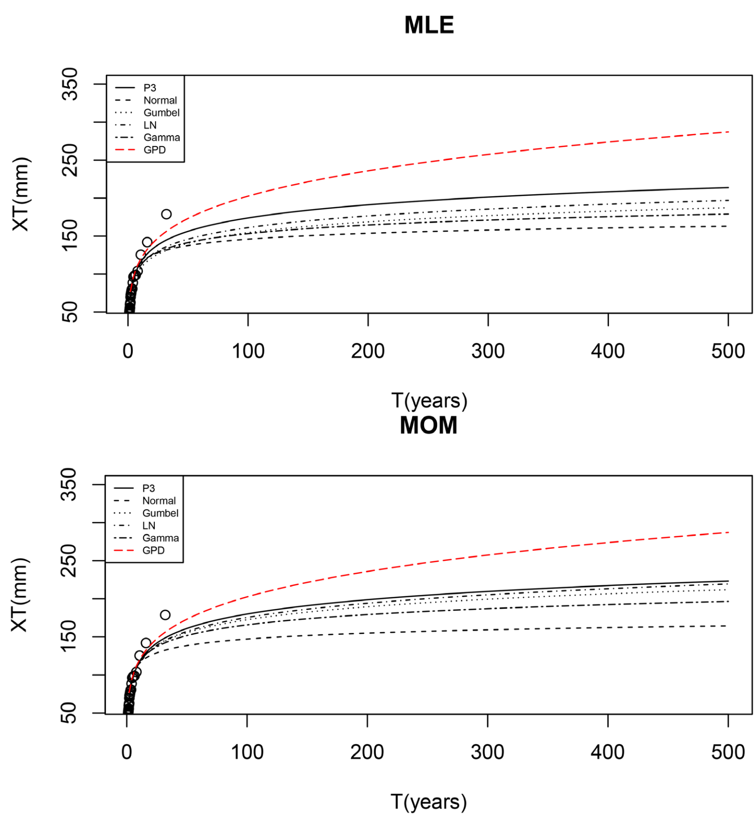

Figure 4. Results of the POT and the AM series quantiles.

4. Conclusion

Torrential rains lead to floods which are long known societal challenges due to the adverse devastating effects on societies, economies, and the environment. Extreme rainfall modelling is therefore critical in forecasting periods of extreme rainfall to inform hydrological and water resources management policies direction. In this study, extreme rainfall analyses were carried out with the AM and partial duration series (POT). The threshold value selection was the main disadvantage of the POT method. The quantile method was applied to ARC2 daily rainfall values and the mean excess plot method was used to set the optimal threshold value. However, the optimal threshold was approximated to be around 33.8 - 42.9 mm. A value which corresponded to 99th percentile was chosen as optimal threshold. 40.8 mm was found as optimal threshold and corresponds to on average 3.6 peaks per year. The GP distribution was used to fit the POT method and the goodness-of-fit test results were presented in **Table 5**. Different parameter estimation techniques were used and five (05) commonly probability distribution functions were applied to the AM data. The P3 distribution gave the best fit to AM data than other distributions when parameters were estimated with the MLE. The LN distribution also gave the best fit to AM data when parameters were estimated with MOM. The Gumbel distribution was the second best fit and the LN distribution was the third best fit to AM data when parameters were estimated with MLE. The POT method gave higher quantile values and described better extreme rainfalls than in the case of AM series method. The P3 distribution in the case of AM series method gave higher quantile estimates than other distributions using MLE and MOM as parameters estimations. For higher extreme rainfall frequency analyses quantiles estimation in Bamako-Senou station area, we suggest using the MOM MLE as parameter estimation technique and the P3 and Gumbel as distribution functions. The LN can also be used as distribution function using the MOM as parameter estimation technique in Bamako area. The POT method could also be used for more extreme rainfall quantiles using the GP as distribution function. Different truncation levels and tests should be applied to the data to find the optimal threshold. We further suggest that, for AM method, the Log-Pearson type 3 (LP3) distribution and the generalized logistic (GL) distribution be used for extreme rainfalls analyses and added the L-moment method for parameters estimation. Furthermore, for more adequacy extreme rainfall analyses, similar analyses should also be performed using different satellite rainfall data in various parts of the Bamako area.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Agbonaye, A. I., & Izinyon, O. C. (2017). Best-Fit Probability Distribution Model for Rainfall Frequency Analysis of Three Cities in South Eastern Nigeria. *Nigerian Journal of Environmental Sciences and Technology, 1,* 34-42. https://doi.org/10.36263/nijest.2017.01.0024

Ahilan, S., O'Sullivan, J. J., & Bruen, M. (2012). Influences on Flood Frequency Distributions in Irish River Catchments. *Hydrology and Earth System Sciences, 16,* 1137-1150. https://doi.org/10.5194/hess-16-1137-2012

Alam, M. A., Farnham, C., & Emura, K. (2018). Best-Fit Probability Models for Maximum Monthly Rainfall in Bangladesh Using Gaussian Mixture Distributions. *Geosciences, 8,* Article 138. https://doi.org/10.3390/geosciences8040138

Alila, Y. (1999). A Hierarchical Approach for the Regionalization of Precipitation Annual Maxima in Canada. *Journal of Geophysical Research, 104,* 31645-31655.

Bačová-Mitková, V., & Onderka, M. (2010). Analysis of Extreme Hydrological Events on the Danube Using the Peak over Threshold Method. *Journal of Hydrology and Hydromechanics, 58,* 88-101. https://doi.org/10.2478/v10098-010-0009-x

Ball, J., Babister, M., Nathan, R., Weeks, W., Weinmann, E., Retallick, M., & Testoni, I. (2016). *Australian Rainfall and Runoff: A Guide to Flood Estimation.* Geoscience Australia.

Bernardara, P., Mazas, F., Kergadallan, X., & Hamm, L. (2014). A Two-Step Framework for Over-Threshold Modelling of Environmental Extremes. *Natural Hazards and Earth System Sciences, 14,* 635-647. https://doi.org/10.5194/nhess-14-635-2014

Bezak, N., Brilly, M., & Šraj, M. (2014). Comparison between the Peaks-Over-Threshold Method and the Annual Maximum Method for Flood Frequency Analysis. *Hydrological Sciences Journal, 59,* 959-977. https://doi.org/10.1080/02626667.2013.831174

Chow, K. C. A., & Watt, W. E. (1992). Use of Akaike Information Criterion for Selection of Flood Frequency Distribution. *Canadian Journal of Civil Engineering, 19,* 616-626. https://doi.org/10.1139/l92-071

Chow, V. T. (1964). *Handbook of Applied Hydrology.* McGraw-Hill Book Company.

Coronado-Hernández, Ó. E., Merlano-Sabalza, E., Díaz-Vergara, Z., & Coronado-Hernández, J. R. (2020). Selection of Hydrological Probability Distributions for Extreme Rainfall Events in the Regions of Colombia. *Water, 12,* Article 1397. https://doi.org/10.3390/w12051397

Cunnane, C. (1973). A Particular Comparison of Annual Maxima and Partial Duration Series Methods of Flood Frequency Prediction. *Journal of Hydrology, 18,* 257-271. https://doi.org/10.1016/0022-1694(73)90051-6

Fisher, R. A., & Tippett, L. H. C. (1928). Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample. *Mathematical Proceedings of the Cambridge Philosophical Society, 24,* 180-190. https://doi.org/10.1017/s0305004100015681

Fofana, M., Adounkpe, J., Larbi, I., Hounkpe, J., Djan'na Koubodana, H., Toure, A. et al. (2022). Urban Flash Flood and Extreme Rainfall Events Trend Analysis in Bamako, Mali. *Environmental Challenges, 6,* Article 100449. https://doi.org/10.1016/j.envc.2022.100449

Griffis, V. W., & Stedinger, J. R. (2007). Evolution of Flood Frequency Analysis with Bulletin 17. *Journal of Hydrologic Engineering, 12,* 283-297. https://doi.org/10.1061/(asce)1084-0699(2007)12:3(283)

Gumbel, E. J. (1958). *Statistics of Extremes.* Columbia University Press.

Hansen, C. R. (2015). Comparison of Regional and At-Site Frequency Analysis Methods for the Estimation of Southern Alberta Extreme Rainfall. *Canadian Water Resources*

Journal / Revue Canadienne des Ressources Hydriques, 40, 325-342. https://doi.org/10.1080/07011784.2015.1060871

Henrique, D., & Castro, M. (2019). What Is the Return Period of Intense Rainfall Events in the Capital Cities of the Northeast Region of Brazil? *Atmospheric Science Letters, 20,* 1-11.

Hosking, J. R. M., & Wallis, J. R. (1997). *Regional Frequency Analysis: An Approach Based on L-Moments.* Cambridge University Press.

Hussain, B., Qureshi, N. A., Buriro, R. A., Qureshi, S. S., Pirzado, A. A., & Saleh, T. A. (2022). Interdependence between Temperature and Precipitation: Modeling Using Copula Method toward Climate Protection. *Modeling Earth Systems and Environment, 8,* 2753-2766. https://doi.org/10.1007/s40808-021-01256-8

Jenkinson, A. F. (1955). The Frequency Distribution of the Annual Maximum (or Minimum) Values of Meteorological Elements. *Quarterly Journal of the Royal Meteorological Society, 81,* 158-171. https://doi.org/10.1002/qj.49708134804

Karmakar, S., & Simonovic, S. P. (2008). Bivariate Flood Frequency Analysis: Part 1. Determination of Marginals by Parametric and Nonparametric Techniques. *Journal of Flood Risk Management, 1,* 190-200. https://doi.org/10.1111/j.1753-318x.2008.00022.x

Karmakar, S., & Simonovic, S. P. (2009). Bivariate Flood Frequency Analysis. Part 2: A Copula-based Approach with Mixed Marginal Distributions. *Journal of Flood Risk Management, 2,* 32-44. https://doi.org/10.1111/j.1753-318x.2009.01020.x

Keita, M. A., Ruan, R., & An, R. (2020). Spatiotemporal Change of Urban Sprawl Patterns in Bamako District in Mali Based on Time Series Analysis. *Urban Science, 5,* Article 4. https://doi.org/10.3390/urbansci5010004

Kite, G. W. (1977). *Frequency and Risk Analyses in Hydrology.* Water Resources Publications.

Lang, M., Ouarda, T. B. M. J., & Bobée, B. (1999). Towards Operational Guidelines for Over-Threshold Modeling. *Journal of Hydrology, 225,* 103-117. https://doi.org/10.1016/s0022-1694(99)00167-5

Langbein, W. B. (1949). Annual Floods and the Partial-Duration Flood Series. *Transactions, American Geophysical Union, 30,* 879-881.

Madsen, H. (1996). *At-Site and Regional Modelling of Extreme Hydrologic Events.* Thesis (PhD), Technical University of Denmark.

Madsen, H., Rasmussen, P. F., & Rosbjerg, D. (1997). Comparison of Annual Maximum Series and Partial Duration Series Methods for Modeling Extreme Hydrologic Events: 1. At-Site Modeling. *Water Resources Research, 33,* 747-757. https://doi.org/10.1029/96wr03848

Mahtab, S. B. (2024). Selection of Best Probability Distribution for Annual Maximum Rainfall Series in Bangladesh. In Q. Xu, W. Xu, S. Yao, & X. Ding (Eds.), *Abstract Book of the 24th IAHR APD Congress*. IAHR. https://www.iahr.org/library/regional?pid=557

Oseni, B. A., & Ayoola, F. J. (2013). Fitting the Statistical Distribution for Daily Rainfall in Ibadan, Based on Chi-Square and Kolmogorov-Smirnov Goodness-of-Fit Tests. *West African Journal of industrial and Academic Research, 7,* 93-100.

Otieno, K., Chaba, L., Odhiambo, C., & Omolo, B. (2025). A Systematic Approach to Modeling Monthly Maximum Temperature and Total Rainfall in Kenya. *Scientific Reports, 15,* Article No. 31758. https://doi.org/10.1038/s41598-025-12810-0

Panthou, G., Vischel, T., & Lebel, T. (2014). Recent Trends in the Regime of Extreme Rainfall in the Central Sahel. *International Journal of Climatology, 34,* 3998-4006.

https://doi.org/10.1002/joc.3984

Panthou, G., Vischel, T., Lebel, T., Blanchet, J., Quantin, G., & Ali, A. (2012). Extreme Rainfall in West Africa: A Regional Modeling. *Water Resources Research, 48,* 1-19. https://doi.org/10.1029/2012wr012052

Rao, A. R., & Hamed, K. H. (2000). *Flood Frequency Analysis.* CRC Press.

Rosbjerg, D., Madsen, H., & Rasmussen, P. F. (1992). Prediction in Partial Duration Series with Generalized Pareto-Distributed Exceedances. *Water Resources Research, 28,* 3001-3010. https://doi.org/10.1029/92wr01750

Salas, J. D., Heo, J. H., Lee, D. J., & Burlando, P. (2013). Quantifying the Uncertainty of Return Period and Risk in Hydrologic Design. *Journal of Hydrologic Engineering, 18,* 518-526.

Salinas, J. L., Castellarin, A., Kohnová, S., & Kjeldsen, T. R. (2014a). Regional Parent Flood Frequency Distributions in Europe—Part 2: Climate and Scale Controls. *Hydrology and Earth System Sciences, 18,* 4391-4401. https://doi.org/10.5194/hess-18-4391-2014

Salinas, J. L., Castellarin, A., Viglione, A., Kohnová, S., & Kjeldsen, T. R. (2014b). Regional Parent Flood Frequency Distributions in Europe—Part 1: Is the GEV Model Suitable as a Pan-European Parent? *Hydrology and Earth System Sciences, 18,* 4381-4389. https://doi.org/10.5194/hess-18-4381-2014

Sankarasubramanian, A., & Srinivasan, K. (1999). Investigation and Comparison of Sampling Properties of L-Moments and Conventional Moments. *Journal of Hydrology, 218,* 13-34. https://doi.org/10.1016/s0022-1694(99)00018-9

Sanogo, A., Kabange, R. S., Owusu, P. A., Djire, B. I., Donkoh, R. F., & Dia, N. (2023). Investigation into Recent Temperature and Rainfall Trends in Mali Using Mann-Kendall Trend Test: Case Study of Bamako. *Journal of Geoscience and Environment Protection, 11,* 155-172. https://doi.org/10.4236/gep.2023.113011

Sarr, B., & Lona, I. (2018). Les fortes pluies enregistrées au Sahel au cours de l'hivernage 2007: Variabilité et/ou changement climatique. Centre Régional Agrhymet (CRA), Département Formation et Recherche.

Schafer, J. L., & Graham, J. W. (2002). Missing Data: Our View of the State of the Art. *Psychological Methods, 7,* 147-177. https://doi.org/10.1037/1082-989x.7.2.147

Seckin, N., Haktanir, T., & Yurtal, R. (2011). Flood Frequency Analysis of Turkey Using L-Moments Method. *Hydrological Processes, 25,* 3499-3505. https://doi.org/10.1002/hyp.8077

Sharma, K. (2016). An Application of Extreme Values Theory to Learning Analytics: Predicting Collaboration Quality from Eye-Tracking Data. *SSRN Electronic Journal,* 22 p. https://doi.org/10.2139/ssrn.2831445

Sharma, M. A., & Singh, J. B. (2010). Use of Probability Distribution in Rainfall Analysis. *New York Science Journal, 3,* 40-49.

Stedinger, J. R., Vogel, R. M., & Foufoula-Georgiou, E. (1993). Frequency Analysis of Extreme Events. In D. R. Maidment (Ed.), *Handbook of Hydrology* (Chapter 18). McGraw-Hill.

Tavares, L. V., & Da Silva, J. E. (1983). Partial Duration Series Method Revisited. *Journal of Hydrology, 64,* 1-14. https://doi.org/10.1016/0022-1694(83)90056-2

Thompson, P., Cai, Y., Reeve, D., & Stander, J. (2009). Automated Threshold Selection Methods for Extreme Wave Analysis. *Coastal Engineering, 56,* 1013-1021. https://doi.org/10.1016/j.coastaleng.2009.06.003

World Bank Group (2011). *Vulnerability, Risk Reduction, and Adaptation to Climate*

Change. World Bank Group.

Zawiah, W. Z. W., Jemain, A. A., Kamarulzaman, I., Suhaila, J., & Sayang, D. M. (2009). A Comparative Study of Extreme Rainfall in Peninsular Malaysia with Reference to Partial Duration and Annual Extreme Series. *Sains Malaysiana, 38,* 751-760.