



Analysis and mathematical modeling for flood surveillance from rainfall by Extreme Values Theory for agriculture in Phitsanulok Province, Thailand

Natthinee Deetae

Department of Statistics, Faculty of Science and Technology, Pibulsongkram Rajabhat University, Phitsanulok, Thailand

Abstract. Attempts to use the generalized extreme value distribution and generalized Pareto distribution with the maximum likelihood estimates on the extreme rainfall data at one weather station over Phitsanulok province. This paper gathered the rainfall data from January 1987 to December 2021. The estimated return level is 10, 20, 50, and 100 years. The result displays the modeling of the generalized extreme value distribution (GEV); the Gumbel distribution was a fitting proposal for extreme annual rainfall in Phitsanulok province, with $\mu = 24.95(1.39)$, $\sigma = 22.59(1.06)$, and $\xi = 0.03(0.05)$. In addition, the modeling of the generalized Pareto distribution (GPD) with daily rainfall data in Phitsanulok province had to fit on $\sigma = 49.08(2.99)$, and $\xi = -0.28(0.03)$. Moreover, we obtained return levels with an upward trend in the change in return periods.

2020 Mathematics Subject Classifications: 60G70, 62G32

Key Words and Phrases: Extreme values theory, Generalized extreme values distribution, Generalized Pareto distribution, Return level, Maximum likelihood estimation, Rainfall

1. Introduction

The problem of natural disasters caused by water is a long-standing problem for agriculture in Thailand, whether the amount of water is so large that it results in flooding conditions or the amount of water is not enough until water scarcity causes drought, all directly affect agricultural productivity and farmers. Phitsanulok province is an area where rice is grown in large quantities and is experiencing natural disasters caused by water. Annually occurring natural disasters affect rural communities and are accepted as being part of the way of life in the community. A solution is expressway drainage, a form of drainage on roads and road junctions where rainfall over the junction is temporarily collected in a holding pond where it is held until it can be pumped safely to the river after the water level in the river has lowered. Large dams are also a potential solution to accommodate the flooding and can be used to store reserve water in times of heavy rainfall, and available

DOI: <https://doi.org/10.29020/nybg.ejpam.v15i4.4558>

Email address: Natthinee@psru.ac.th (N. Deetae)

for release in the event of drought. A technical, analytical solution is also possible, where natural disaster analysis models are created for specific areas, rather than attempting to create long-term solutions for broad geographic areas. However, in water management, there are still uncontrollable factors, such as global climate change affecting flooding [4, 6–9]. Therefore, water management solutions are necessary and can be effectively modeled for greater accuracy in specific situations, such as installing additional drainage systems or constructing reservoirs to meet water hazards. Information relevant to the field of hydrology includes rainfall measurements, wind speed and frequency, humidity, and temperature changes. Of particular relevance is where these data show the extreme values from which hydrological forecast models can be created to analyze outliers; abnormal highs or abnormally lows, to obtain the highest or lowest value for the forecasting model. These models can provide forward-looking surveillance, projections and forecasts, leading to the creation of a good management policy.

Extreme Value Theory is a model popular for forecast modeling that can be applied to excessive or low rainfall. Several researchers have studied practical applications of the generalized extreme theory on rainfall data from different parts of the world: for example, P. Bussababodin and A. Kaewman [2] studied extreme statistics, especially in hydrology, applied extreme value theory in various disciplines, and discussed the concept and development of the theory of extreme values and summarized the inference statistics of the extreme values in 2015. Chikobvu and Chifurira [3] created models of extreme minimum rainfall for Zimbabwe using generalized extreme value distribution. Subsequently, in 2000, C. S. Withers and S. Nadarajah [12] studied evidence of a trend in return levels for daily rainfall in New Zealand by using annual maxima of daily rainfall over time with the generalized extreme value distribution. In 2007, El Adlouni et al. [1] presented generalized maximum likelihood estimators for the nonstationary generalized extreme value model. More recently, in 2020, Samuel et al. [11] proposed modeling extreme rainfall using the generalized extreme value distribution from monthly rainfall data from the Nigeria Meteorological Agency in Kaduna, fitted the data to the generalized extreme value distribution, the generalized Pareto distribution, and a return level. The research cited here is only part of the research into the use of the extremes value theory to predict rainfall, and could essentially support the design of urban flood control structures for appropriate flood risk management.

The objective of this study is to quantify and describe the behavior of maximum rainfall for flood surveillance by the extreme values theory with generalized extreme values (GEV) distribution and generalized Pareto distribution (GPD). The results will provide an estimator of parameters and return level.

2. Methodology

In the extreme values theory, we used generalized the extreme values distribution and the Pareto distribution; the details of which are as follows.

2.1. Generalized Extreme Values (GEV) Distribution

Let X_1, X_2, \dots, X_n are a sequence of independent random variables having common distribution function F . Suppose that M_n is a maximum of distribution for all values of n by

$$M_n = \max(X_1, X_2, \dots, X_n).$$

If n is the number of observations in a year then M_n represents the annual maximum and the distribution of M_n can be derived exactly for all values of n as follows:

$$\begin{aligned} Pr\{M_n \leq z\} &= Pr\{X_1 \leq z, X_2 \leq z, \dots, X_n \leq z\} \\ &= Pr\{x_1 \leq z\} \times Pr\{x_2 \leq z\} \times \dots \times Pr\{x_n \leq z\}. \end{aligned}$$

It comes that $Pr\{M_n \leq z\} = \{F(z)^n\}$.

The finite maxima M_n converges to the upper endpoint of $\{F(z)\}^n$ when $n \rightarrow \infty$. Now, let's look at the behavior of F^n as $n \rightarrow \infty$. Suppose that there exists sequences of normalizing constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that:

$$P\left(\frac{M_n - b_n}{a_n} \leq z\right) = F^n(a_n z + b_n) \rightarrow G(z);$$

where G is a non-degenerate distribution function, then G as follows

1. Gumbel distribution

$$G(z) = \exp\left(-\exp\left(-\frac{z-b}{a}\right)\right), \quad -\infty < z < \infty,$$

2. Fréchet distribution

$$G(z) = \begin{cases} 0, & z \leq b \\ \exp\left(-\left(\frac{z-b}{a}\right)^{-\alpha}\right), & z > b, \alpha > 0, \end{cases}$$

3. Weibull distribution

$$G(z) = \begin{cases} \exp\left(-\left(-\frac{z-b}{a}\right)^\alpha\right), & z > b, \alpha > 0 \\ 1, & z \leq b. \end{cases}$$

For some parameters $a > 0, b \in \mathbb{R}$ and $\alpha > 0$, then Gumbel, Fréchet and Weibull can be combined into a single family of models having the distribution function of the form:

$$G(z) = \exp\left(-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{\frac{1}{\xi}}\right);$$

where z are the extreme values from the blocks, μ is a location parameter; σ is a scale parameter; ξ is a shape parameter. The condition for a distribution to belong to any of the extreme values distributions is given as:

- (i) $\xi = 0$, Gumbel distribution
- (ii) $\xi > 0$, Fréchet distribution
- (iii) $\xi < 0$, Weibul distribution.

Estimating the unknown parameters of the GEV above, leads to finding the maximum likelihood estimation (MLE) of GEV which can be written as:

$$l(\mu, \sigma, \xi) = -n \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \log \left(1 + \xi \left(\frac{z_i - \mu}{\sigma}\right)\right) - \sum_{i=1}^n \log \left(1 + \xi \left(\frac{z_i - \mu}{\sigma}\right)\right)^{\frac{1}{\xi}};$$

where $1 + \xi \left(\frac{z_i - \mu}{\sigma}\right) > 0$ and $\xi \neq 0$.

Return Periods can calculate the return level z_T^{GEV} at T of *GEV* as

$$z_T^{GEV} = \mu - \frac{\sigma}{\xi} \left(1 - \left(-\log \left(1 - \frac{1}{T}\right)\right)^{-\xi}\right);$$

for $T = N \times n_y$ when n_y is one observation per year and N is a number of years.

2.2. Generalized Pareto Distribution (GPD)

The GPD uses the threshold exceedances approach where data exceeds the appropriate threshold. Therefore, the distribution function for the GPD is:

$$H(z) = 1 - \left(1 + \frac{\xi z}{\sigma}\right)^{\frac{1}{\xi}}.$$

where σ is a scale parameter and ξ is a shape parameter. Parallely to the GEV distribution, the GPD includes three types of distribution [10].

- (i) $\xi = 0$, Exponential distribution.
- (ii) $\xi > 0$, Ordinary Pareto distribution.
- (iii) $\xi < 0$, Pareto-II type distribution.

To calculate the maximum likelihood estimation of GPD, we have

$$l(\mu, \sigma) = -n \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \log \left(1 + \xi \left(\frac{z_i - \mu}{\sigma}\right)\right).$$

For the return level of GPD, u corresponds to a suitable threshold value, and $X > u$ is given as:

$$Pr(X > u) = \zeta_u \left(1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right)^{\frac{-1}{\xi}}.$$

When $\zeta_u = Pr(X > u)$. The return level z_T^{GPD} at T of GPD as

$$z_T^{GPD} = u + \frac{\sigma}{\xi} (T\zeta_u)^\xi - 1, \quad \xi \neq 0.$$

For $T = N \times n_y$ when n_y is one observation per year and N is a number of years.

3. Data and Research Methodology

The study area domain included Phitsanulok province's agricultural area, which has 20% of the total area of the province, which is located in the upper central region or the lower northern region of Thailand, with an area of 10,815.8 square kilometers (6,759,909 rai) or 6.4% of the northern area, and is 2.1% of the country's total area. It is bordered on the north by Uttaradit province, on the south by Phichit province, on the east by Phetchabun province, and on the west by Kamphaeng Phet and Sukhothai provinces. Climate and weather conditions throughout the year in Phitsanulok include the rainy season, which is hot, windless, and frequently overcast. The rainy season starts around May - October, with an average rainfall of 1,375 millimeters per year. The dry season is hot and humid, partly cloudy. Generally, the weather is hot all year; with the temperature varying in the range from 19°C to 37°C and very rarely below 15°C or above 39°C.

The data used in this study were the monthly highest rainfall data of Phitsanulok province from 1987 – 2021 from the weather station 378201-Phitsanulok. The dataset contains 368 values of maximum monthly rainfall. By using the data courtesy of the Thailand Meteorological Department, we performed extreme values analysis by fitting the generalized extreme value distribution and the generalized Pareto distribution to the sample data using the method of maximum likelihood estimates. We considered the generalized extreme value distribution as having the following families of distribution.

We used the R package for extremes by Gilleland et al.[5] that can perform parametric inferential analysis of the GEV distribution and GPD for the 378201-Phitsanulok station.

The research results were divided into two parts: analysis of the generalized extreme value distribution and the generalized Pareto distribution. The details are as follows.

3.1. Analysis of the generalized extreme value distribution

The Augmented Dickey-Fuller test (ADF) in Table 1 indicating only the significance of p-value statistics. Hence, the premise of stationary for rainfall data of Phitsanulok province from 378201-Phitsanulok station, and therefore we conclude the stationarity of rainfall data.

Table 1: Stationarity test

Test	Test statistic	p-value
Augmented Dickey Fuller	-13.828	0.01

The analysis is based on the data available in Phitsanulok province. The data have been recorded by the Thailand Meteorological Department from 1987 to 2021, and is the best available monthly rainfall data. Table 2 shows the statistical description of the data, and the p-value of the Jarque–Bera test is 4.344e-13. Hence, the data comes from a normal distribution.

Table 2: Summary statistics of normality tests of the monthly highest rainfall data of Phitsanulok province from 1987 – 2021

<i>N</i>	Maximum	Mean	Standard deviation	Coefficient of skewness	Jarque-Bera Statistic
368	167.1	38.8133	29.3043	0.8584	56.929 (4.344e-13)

Table 3 indicates the results of the GEV modeling on the extreme rainfall data in Phitsanulok province using the Block Maxima approach. The GEV parameters were estimated using maximum likelihood estimation. The estimated results are $(\mu, \sigma, \xi) = (24.9578, 22.5914, 0.0324)$, with standard errors (1.3935, 1.0627, 0.0515). The approximate 95% confidence intervals (CI) for the parameters are thus (22.2233,27.6856) for μ , (20.5079, 24.6740) for σ , and (-0.0685,0.1335) for ξ . Since the confidence intervals of ξ contain 0, the Gumbel distribution is the optimal model for the GEV family. The estimated parameter covariance matrix of location scale and shape are shown in Table 4.

Table 3: Maximum likelihood estimates (standard errors) of the GEV distribution parameters

	Location (μ)	Scale (σ)	Shape (ξ)
Estimation	24.9578	22.5914	0.0324
Standard errors	1.3935	1.0627	0.0515
CI 95%	(22.2233,27.6856)	(20.5079,24.6740)	(-0.0685,0.1335)

Table 4: Estimated parameter covariance matrix

	Location (μ)	Scale (σ)	Shape (ξ)
Location (μ)	1.9417	0.7178	-0.0321
Scale (σ)	0.7178	1.1295	-0.0231
Shape (ξ)	-0.0321	-0.0231	0.0026

Table 5: Kolmogorov–Smirnov tests to determine whether annual rainfall data for Phitsanulok province for 1987–2021 follow a GEV distribution

Test	Test statistic	p-value
Kolmogorov–Smirnov test	0.0539	0.2348

The quality of convergence of the rainfall extremes is accessed using the Kolmogorov–Smirnov (K-S) goodness-of-fit tests. The K-S test, relying on the empirical study of the cumulative distribution function, is used to determine if the sample is from the hypothesized continuous distribution. The K-S approach is less sensitive for normal distribution. The assumption is that the data is from a population independent of identical distribution (i.i.d). The alternative hypothesis is a two-tail test assuming that the data follow a monotonic trend. Therefore, from the results obtained, the p-value of the K-S test is 0.2348, $\sigma=22.5914(1.0627)$, and $\xi=0.0324(0.0515)$ is appropriate for this dataset.

Figure 1 shows the diagnostic plots for the GEV. The probability plot and quantile plot points lie close to the unit diagonal. This implies that the generalized extreme value distribution function provides a good fit. The return level plot shows that the empirical return levels match well with those from the fitted distribution function. Finally, the density plot shows good agreement between the fitted GEV distribution function and the empirical density.

Table 6: Return level estimates (mm) at selected return intervals (T) determined using the GEV distribution

Return Periods	T=10	T=20	T=50	T=100
Return Levels	77.6981	95.4001	118.9373	137.0483
CI (95%)	(71.29,84.11)	(85.54,105.26)	(102.24,135.64)	(113.46,160.64)

Table 6 shows the estimated return periods of maximum rainfall likely to occur over the next 10, 20, 50, or even 100 years fitted by GEV distribution. Return levels are concerning the change in return periods. Changes in increasing return periods and return levels of extreme maximum rainfall also increase. The current maximum Monthly rainfall is 137.0483 mm, and the extreme maximum rainfall is expected to be 137.0483 mm in a return period of 100 years or more. The estimated return level lies in 95% confidence at the interval of (113.46,160.64). The obtained confidence interval shows that the actual value of the return level of extreme maximum monthly rainfall is 100% within the interval.

3.2. Analysis of the generalized Pareto distribution

The estimate of generalized Pareto distribution in Table 7 indicates the results of the GPD modeling on extreme rainfall data in Phitsanulok province using the threshold approach. The GPD parameters were estimated using maximum likelihood estimation. The estimated results are $(\sigma, \xi) = (49.0750, -0.2819)$, with standard errors (2.9901, 0.0276).

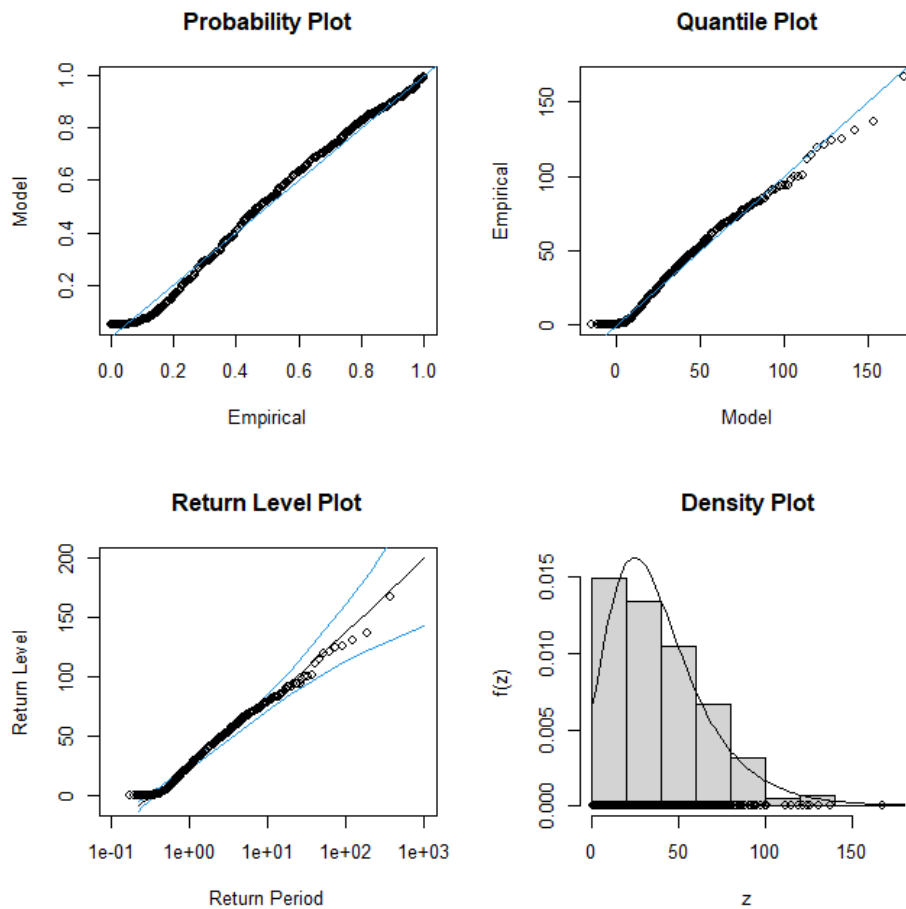


Figure 1: The normal probability density function of maximum annual rainfall for Phitsanulok province for the period 1987–2021 for the GEV.

The approximate 95% confidence intervals for the parameters are thus (43.2021, 54.92173) for σ , and (−0.3359, −0.2277) for ξ . The estimated parameter covariance matrix of location scale and shape are shown in Table 8.

Table 7: Maximum likelihood estimates (standard errors) of the GPD parameters

	Scale (σ)	Shape (ξ)
Estimation	49.0750	−0.2819
Standard errors	2.9901	0.0276
CI 95%	(43.2021,54.9173)	(−0.3359, −0.2277)

The K-S test in Table 9 indicates that the p-value is 0.1003. Hence, the GPD with $\sigma=49.0750(2.9901)$ and $\xi=-0.2819(0.0276)$ is appropriate for this dataset.

Table 8: Estimated parameter covariance matrix

	Scale (σ)	Shape (ξ)
Scale (σ)	8.9319	-0.0691
Shape (ξ)	-0.0691	0.0007

Table 9: Kolmogorov–Smirnov tests to determine whether annual rainfall data for Phitsanulok province for 1987–2021 follow the GPD

Test	Test statistic	p-value
Kolmogorov–Smirnov test	0.0637	0.1003

Table 10: Return level estimates (mm) at selected return intervals (T) determined using the GPD

Return Periods	T=10	T=20	T=50	T=100
Return Levels	161.5650	164.8923	168.4024	170.5163
CI (95%)	(133.31,189.82)	(131.24,198.54)	(125.79,211.01)	(119.39,221.64)

Table 10 shows the estimated return periods of maximum rainfall likely to occur over the next 10, 20, 50, or even 100 years fitted by GPD. Return levels are concerning the change in return periods. Changes in increasing return periods, return levels of extreme maximum rainfall also increase. The current maximum monthly rainfall is 170.5163 mm, and the extreme maximum rainfall is expected to be 170.5163 mm in a return period of 100 years or more. The estimated return level lies in 95% confidence at the interval of (119.3940, 221.6386). The obtained confidence interval shows that the actual value of the return level of extreme maximum monthly rainfall is 100% within the interval.

4. Conclusions

The study will inform meteorologists about the behavior of high rainfall with an extreme value theory. Furthermore, government or non-governmental organizations can make appropriate decisions and plans based on the results of this study to prepare the general public for changes brought on by the highest rainfall. This paper’s results were based on the monthly highest rainfall data of Phitsanulok province retrieved from the Weather Station 378201-Phitsanulok from 1987 to 2021 and analyzed by GEV distribution and GPD. As a result, we get the estimation of parameters from GEV for the return level for rainfall values for 10, 20, 50, and 100 years, which are 77.6981 mm, 95.4001 mm, 118.9373 mm, and 137.0483 mm, respectively. On the other hand, the estimation of parameters from GPD for the return level for rainfall values for 10, 20, 50, and 100 years are 161.5650 mm, 164.8923 mm, 168.4024 mm, and 170.5163 mm, respectively. Decision-makers benefits include early warning systems, food security, poverty alleviation, disaster

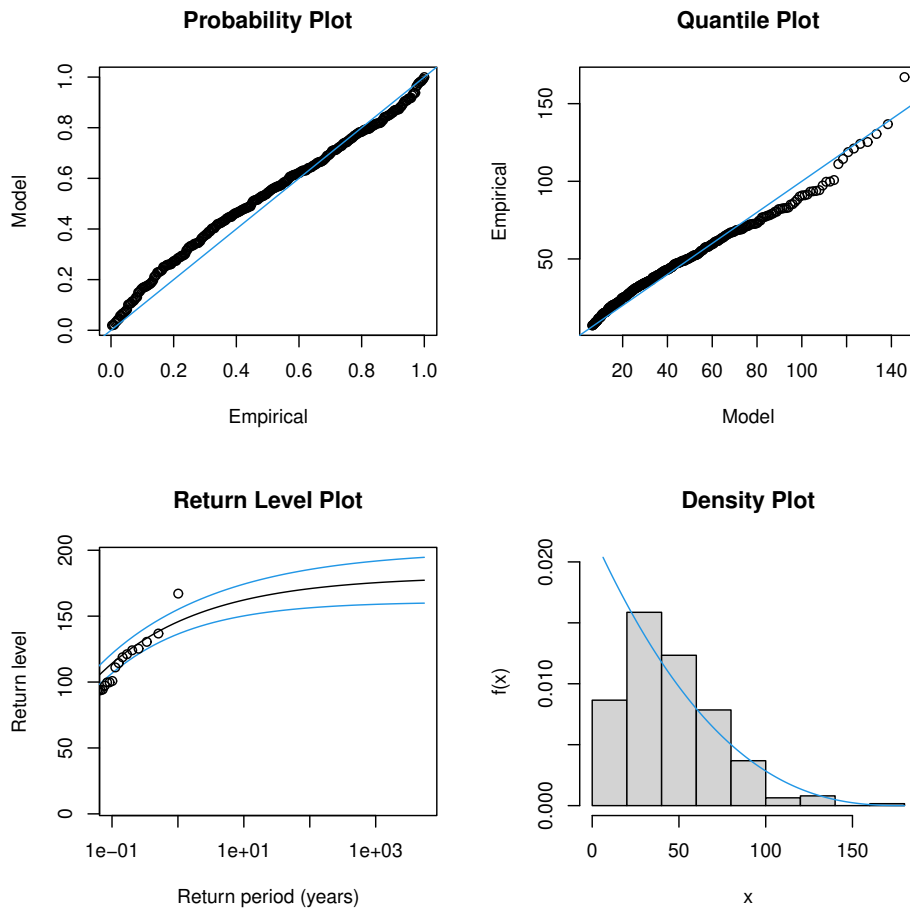


Figure 2: The normal probability density function of maximum annual rainfall for Phitsanulok province for the period 1987–2021 for the GPD.

management, risk management, etc. Other than that, we can develop an extreme value theory to enhance knowledge and other theories to be more efficient.

Acknowledgements

This work is partially supported by Pibulsongkram Rajabhat University, funded Research and innovation for the development of the Bio-economy, Circular Economy, and Green Economy (BCG) Fiscal Year 2022 Capital Code RDI-2-65-52. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

References

- [1] S. El Adlouni, T. B. M. J. Ouarda, X. Zhang, R. Roy, and B. Bobée. Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resources Research*, 43:1–13, 2007.
- [2] P. Busababodhin and A. Kaewmun. Extreme values statistics. *The Journal of King Mongkut's University of Technology North Bangkok*, 25(2):55–65, 2015.
- [3] D. Chikobvu and R. Chifurira. Modelling of extreme minimum rainfall using generalised extreme value distribution for zimbabwe. *South African Journal of Science*, 111:1–8, 2015.
- [4] L. Gao, L. Zhang, and X. Xu. A global data set for economic losses of extreme hydrological events during 1960–2014. *Water Resources Research*, 5:5165–5175, 2019.
- [5] E. Gilleland, M. Ribatet, and A. G. Stephenson. A software review for extreme value analysis. *Extremes*, 16(1):103–119, 2013.
- [6] E. A. Ijigah and T. A. Akinyemi. Flood disaster: An empirical survey of causative factors and preventive measures in kaduna, nigeria. *International Journal of Environment and Pollution*, 3:53–56, 2015.
- [7] C. Jones, D. E. Waliser, K. M. Lau, and W. Stern. Global occurrences of extreme precipitation and the madden–julian oscillation: Observations and predictability. *Journal of Climate*, 17:4575–4589, 2004.
- [8] J.-L. Martel, M. Alain, and B. Francois. Global and regional projected changes in 100-yr subdaily , daily , and multiday precipitation extremes estimated from three large ensembles of climate simulations. *Journal of Climate*, 33(3):1089–1103, 2020.
- [9] M. M. Q. Mirza. Global warming and changes in the probability of occurrence of floods in bangladesh and implications. *Global Environmental Change*, 12:127–138, 2002.
- [10] F. C. Onwuegbuche, A. B. Kenyatta, S. B. Affognon, E. P. Enock, and M. O. Akinade. Application of extreme value theory in predicting climate change induced extreme rainfall in kenya. *International Journal of Probability and Statistics*, 8(4):85–94, 2019.
- [11] S. B. Sunday, N. S. Agog, P. Magdalene, M. Abdullateef, and K. A. Gladys. Modeling extreme rainfall in kaduna using the generalised extreme value distribution. *Scientific World Journal*, 15(3):73–77, 2020.
- [12] C. S. Withers and S. Nadarajah. Evidence of trend in return levels for daily rainfall in new zealand. *Journal of Hydrology (New Zealand)*, 39:155–166, 2000.