Modeling Extreme Rainfall Events in the Tropical Monsoon Region: A Generalized Pareto Distribution Approach

Nurul Azizah Muzakir[a], Nurtiti Sunusi[b*],Sri Astuti Thamrin[c]

[a]namuzakir@gmail.com, [b]nurtitisunusi@unhas.ac.id, [c]tuti@unhas.ac.id
[a,b,c] Department of Statistics, Faculty of Mathematics and Natural Sciences, Hasanuddin University, Makassar, Indonesia

Abstract

Extreme weather events have become more frequent due to climate change, and extreme rainfall is a common occurrence in the tropical monsoon areas. Makassar City was chosen as a representative area of the tropical monsoon climate. The objective of this research is to utilize the generalized Pareto distribution (GPD), along with its nested model (exponential), to predict the monthly return levels of extreme rainfall in the designated geographical region. The study utilized daily precipitation data from January 1980 to December 2022. To determine the monthly daily rainfall data that exceeds a certain threshold, the peaks over threshold (POT) method was applied. The result discovered that the exponential distribution is the most appropriate for extreme rainfall series in most months, except for February and July, where the GPD is more appropriate. No trends or seasonal patterns were identified in any of the months. The calculated return levels of extreme rainfall for each month at the 2, 3, 5, and 10-year return periods indicate that February has the highest rainfall return level for all selected return periods compared to other months with December and January following closely behind. These findings are expected to assist the government in developing flood prevention strategies and mitigating their effects, particularly during the rainy season's peak months in the city, which are December to February.

Keywords: extreme rainfall; generalized Pareto; peaks over threshold; return levels

1. Introduction

Global climate change is an issue that is currently a concern for many researchers. The consequences of global climate change are multifaceted, affecting both the quality of human life and the environment. As a result of climate change, extreme weather, such as heavy rain and high temperatures will become more common (Jentsch & Beierkuhnlein, 2008). Climate change affects various regions across the world, including tropical areas (Loo et al., 2015). The tropical monsoon climate is a tropical climate subtype. Areas with a tropical monsoon climate are coded Am according to the Koppen climate classification (McKnight & Darrel, 2000). The tropical monsoon climate is characterized by a contrast between the wet and dry seasons. In such regions, flooding for the time of the wet season and drought for the time of the dry season are common, which can be attributed to the extreme levels of precipitation and temperatures (Kusumastuti & Weesakul, 2014). Existing definitions of extreme rainfall events typically focus on the amount of rain that falls over a given duration (Yin et al., 2022). Kusumastuti & Weesakul (2014) conducted research on the extreme rainfall indices for Southeast Asian monsoon countries. According to their findings, precipitation ≧ 60 mm/day is considered heavy rainfall, precipitation ≧ 80 mm/day is considered very heavy rainfall, and precipitation ≧ 100 mm/day is considered extreme rainfall.

The probability distribution is one method for predicting changes in extreme rainfall. This method is extremely useful for determining the frequency of extreme rainfall. Frequency analysis aims to ascertain the probability of specific occurrences. The distributions of Weibull, generalized extreme value (GEVD), generalized Pareto (GPD), generalized logistic, gamma, lognormal, Gumbel, Burr, and several other probability distributions are widely used for extreme rainfall investigations (Benyahya et al., 2014; Li et al., 2014; Rahman et al., 2013). Each probability distribution has its own applications and constraints. Choosing the appropriate probability distribution is therefore critical in analyzing the frequency of extreme events (Li et al., 2015). Several studies have proven that the GEVD or GPD is the optimal model for representing precipitation data. For instance, Zalina et al. (2002) compared eight candidate distributions to provide an accurate forecast of maximum rainfall for Malaysia. They considered the gamma, generalized normal, GPD, GEVD, Gumbel, Log Pearson Type III, Pearson Type III, and Wakeby distributions. According to their findings, the GEVD model was deemed the most suitable for characterizing Malaysia's yearly precipitation data. In a research conducted by Shaharudin et al. (2020), they used several probability distributions, such as GPD, Lognormal, and Gamma to model the daily rainfall data in Malaysia. The study discovered that the GPD was appropriate for describing daily rainfall data. In a third study, Rahman et al. (2013) considered fifteen different probability distributions in the annual maximum flood data in Australia. The Log Pearson Type III, GEVD, and GPD were identified as the three most suitable distributions. The suitability of GEVD and GPD in modeling extreme rainfall series has led to their common usage in frequency analysis of extreme rainfall. Various researchers worwilde have utilized either the GEVD or the GPD for rainfall analysis. For instance, Boudrissa et al. (2017) used the GEVD to analyze the yearly maximum daily precipitation in northern Algeria, Martins et al. (2020) used the GPD to examine maximum precipitation in Uruguaiana, Brazil, while Singirankabo & Iyamuremye (2022) applied the GPD to investigate extreme precipitation occurences in Kigali City.

The GPD is a powerful tool for modeling extreme value data, especially when the underlying distribution is unknown or difficult to determine. It is a flexible distribution that can be used to accurately model a wide range of extreme phenomena, including heavy-tailed and skewed data. Regarding precipitation in Makassar, the capital of South Sulawesi province in Indonesia, the distribution of rainfall is uneven throughout the year. Some months experience very high rainfall, while other months have relatively low rainfall. Due to this variability, the GPD is a suitable distribution for modeling rainfall in Makassar City. Consequently, Makassar City was chosen as a representative area for the study of tropical monsoon climates. To our knowledge, there remains a lack of literature about the frequency analysis of extreme precipitation in the tropical monsoon area, particularly in the study area utilizing the GPD. Previous studies conducted by Rahim et al. (2019) and Sanusi et al. (2017) have examined the characteristics of extreme precipitation in Makassar City using the GEVD. However, the use of the GPD for analyzing extreme precipitation in this specific tropical monsoon region remains underexplored in the existing literature. Consequently, the authors attempted to analyze the extreme rainfall in the same study area using an alternative probability distribution, namely the GPD. The aim of this study is to utilize the GPD, along with its nested model (exponential), to predict the monthly return levels of extreme precipitation in Makassar City. Return levels offer valuable insight into the likelihood of extreme rainfall during a particular period, allowing local government to devise effective flood mitigation strategies.

2. Methodology

2.1. Data and Location

This research was carried out in Makassar, which is the administrative capital of Indonesia's South Sulawesi province. Makassar is the most populous city in the eastern part of the country and is recognized as the fifth-largest metropolitan area in Indonesia. The city’s geographic coordinates are 5.133° S and 199.417° E, and it covers a total area of 175.77 km^2. The highest elevation in Makassar City is 32 meters above sea level, and its temperature range between 23°C to 32°C (Sunusi & Giarno, 2022). Makassar has a monsoon climate (Köppen Am) characterized by distinct wet and dry seasons (Figure 1) (The Ministry of Foreign Affairs of Iran, 2020). The rainfall in Makassar City is affected by the monsoon, which is caused by the alternating high-pressure and low-pressure systems in the Asian and Australian continents.

The daily rainfall data (mm/day) recorded at the Sultan Hasanuddin meteorological station from January 1980 to December 2022 was used. It is a weather observation station located at Sultan Hasanuddin International Airport (the coordinates are 05°03'71" S and 119°33'65" E). The station is operated by the Meteorological, Climatological, and Geophysical Agency of Indonesia, known as BMKG. The daily rainfall data was grouped into monthly periods, and POT method was employed to identify monthly daily rainfall that exceeded a sufficiently high threshold. The GPD was then used to analyze the identified values.

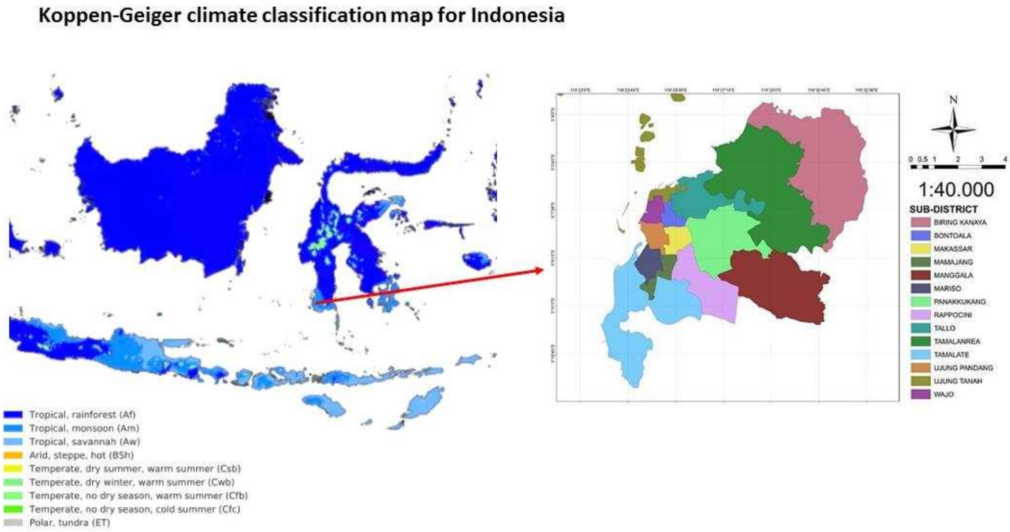

Figure 1. Map of the study area

2.2. Generalized Pareto Distribution

The GPD was utilized to model the distribution of occurrences surpassing a particular threshold. If the threshold is represented by u and a random variable by Z, the Cumulative Distribution Function (CDF) of the GPD can be expressed in the following equation (Pickands, 1975):

$$F(z-u;\sigma,\xi) = 1 - \left(1 + \xi\left(\frac{z-u}{\sigma}\right)\right)^{-\frac{1}{\xi}}. \quad (1)$$

The scale parameter, σ, is responsible for determining the distributon's scale and can only take on positive values. The shape parameter, ξ, determines the distribution's tail behavior and can take on any real value. The support of Z is $z \geq u$ when $\xi \geq 0$, and $u \leq z \leq u - \sigma/\xi$ when $\xi < 0$. In the event that the shape parameter $\xi = 0$, the distribution is simplified to an exponential distribution, which is characterized by its CDF:

$$F(z-u;\sigma) = 1 - e^{\left(-\frac{z-u}{\sigma}\right)}. \quad (2)$$

When ξ is positive, the distribution exhibits a right-skewed tail, resulting in a Pareto distribution. Conversely, when ξ is negative, the distribution leads to a beta distribution (Caires, 2016).

2.3. Threshold Selection

Selecting an appropriate threshold is crucial in the POT method. To ensure the stability of the GPD parameters, a sufficiently high threshold value must be maintained. However, choosing a threshold that is too high reduces the number of observations and increases the variance (Wu & Qiu, 2018). Graphical approaches, such as the Mean Residual Life (MRL) plot and the threshold stability plot, are commonly used to determine the threshold. Davison & Smith (1990) proposed a graphical method for selecting a threshold called the MRL plot. This method is based on the expectation of the GPD excesses,

$$E(Z - u_0 | Z > u_0) = \frac{\sigma_{u_0}}{1 - \xi}, \tag{3}$$

where the shape parameter must be less than one to guarantee the existence of the mean. The GPD's threshold stability property states that if it effectively models the tail behavior for a threshold u_0, it also accurately describes the tail behavior for any higher threshold $u > u_0$. Hence, for $u > u_0$,

$$E(Z - u | Z > u) = \frac{\sigma_u}{1 - \xi} = \frac{\sigma_u + \xi u}{1 - \xi}. \tag{4}$$

Thus, $E(Z - u | Z > u)$ is a linear function of u for all $u > u_0$. Figure 2 shows the example of MRL plot with a confidence interval of 95%.

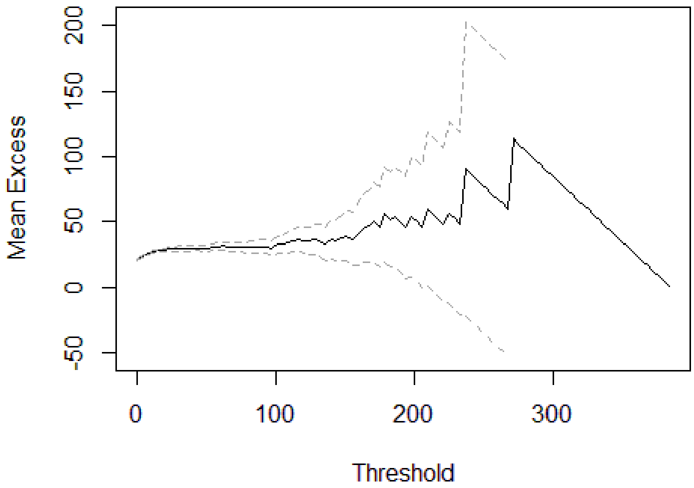

Figure 2. The graph of the mean residual life

The basic idea behind the MRL plot is to choose a threshold that, when plotted, shows a linear trend as the threshold value increases, indicating that the GP distribution model is valid.

2.4. Parameter Estimation

The parameters of the GPD were estimated through the utilization of the maximum likelihood method. This method is a classic and is most widely used because of its simplicity (Coles & Dixon, 1999). Suppose $z_1, z_2, \dots z_m$ are GP-distributed random samples. The likelihood function for $\xi = 0$ as in Eq. (5).

$$L(\sigma|z_1,\dots,z_m) = \prod_{i=1}^{m} f(z_i|\sigma) = \prod_{i=1}^{m}\left[\frac{1}{\sigma}e^{\left(-\frac{z_i-u}{\sigma}\right)}\right] \tag{5}$$

For $\xi \neq 0$, the likelihood function as in Eq. (6):

$$L(\sigma,\xi|z_1,\dots,z_m) = \prod_{i=1}^{m} f(z_i|\sigma,\xi) = \prod_{i=1}^{m}\left[\frac{1}{\sigma}\left(1+\xi\left(\frac{z_i-u}{\sigma}\right)\right)^{-\frac{1}{\xi}-1}\right]. \tag{6}$$

When attempting to maximize the likelihood function in Eq. (5) and (6), one can also aim to maximize the logarithm of the likelihood function. As a result, for $\xi = 0$, the ln likelihood function as in Eq. (7)

$$\begin{aligned}\ln L(\sigma|z_1,\dots,z_m) &= \ln \textstyle\prod_{i=1}^{m} f(z_i|\sigma) = \sum_{i=1}^{m} \ln f(z_i|\sigma)\\ &= \textstyle\sum_{i=1}^{m} \ln\left[\frac{1}{\sigma}e^{\left(-\frac{z_i-u}{\sigma}\right)}\right]\\ &= -m\ln(\sigma) - \textstyle\sum_{i=1}^{m}\left(\frac{z_i-u}{\sigma}\right).\end{aligned} \tag{7}$$

The ln likelihood for $\xi \neq 0$ as in Eq. (8)

$$\begin{aligned}\ln L(\sigma,\xi|z_1,\dots,z_m) &= \ln \textstyle\prod_{i=1}^{m} f(z_i|\sigma,\xi) = \sum_{i=1}^{m} \ln f(z_i|\sigma,\xi)\\ &= \textstyle\sum_{i=1}^{m} \ln\left[\frac{1}{\sigma}\left(1+\xi\left(\frac{z_i-u}{\sigma}\right)\right)^{-\frac{1}{\xi}-1}\right]\\ &= -m\ln(\sigma) + \left(-\frac{1}{\xi}-1\right)\textstyle\sum_{i=1}^{m} \ln\left(1+\xi\left(\frac{z_i-u}{\sigma}\right)\right)^{-\frac{1}{\xi}-1}.\end{aligned} \tag{8}$$

Once the ln likelihood function is obtained, the subsequent procedure is to determine the first-order partial derivative of the ln likelihood function with respect to each parameter and and equate it to zero. If no analytical solution is possible, then parameter estimation is performed using numerical methods, one of which is the BFGS quasi-Newton (Gilli et al., 2019).

2.5. Hypothesis Testing

After obtaining the parameters of the GPD, the model fit was evaluated through the utilization of the Kolmogorov-Smirnov (K-S) test. This one-sample test compares the CDF of the excesses over a specific threshold to the CDF of the GPD. The null hypothesis holds that the excesses over a threshold follow the GPD. The null hypothesis is deemed to be rejected when the p-value $< \alpha$. (Berger & Zhou, 2014). The Mann-Kendall (M-K) test was utilized to ascertain the existence of a trend over time. The M-K statistic sums the sign of all differences between pairs of observations. The null hypothesis assumes that the data does not exhibit any trend (Wu et al., 2008).

The likelihood ratio (LR) test was employed to assess and compare the goodness of fit of two models, where one model is a simplified version of the other. Assume M_0 is the GPD with parameter of $\theta_0 = (\sigma, \xi)$ and M_1 is its nested model, that is the exponential distribution with parameter of $\theta_1 = (\sigma)$. The null hypothesis assumes that the simpler model (exponential) provides a good fit to the extreme rainfall series. The test statistic can be expressed as follows (Reiss & Thomas, 2007):

$$LRT = 2\left[l(\hat{\theta}_0) - l(\hat{\theta}_1)\right] \tag{9}$$

where $l(\hat{\theta}_0)$ represents the log-likelihood for the GPD and $l(\hat{\theta}_1)$ represents the log-likelihood for the exponential. The critical value of the LR test is calculated from the distribution of the test statistic, which is estimated by a chi-square distribution with one degree of freedom in this case at the significance level of α. If the LR value computed is greater than the critical value, it is indicative of adequate evidence to reject the null hypothesis and infer that the GPD model offers a superior fit to the given data. Alternatively, the null hypothesis is rejected if the $p - value < \alpha$.

2.6. Return Levels

To simplify the understanding of extreme value models, they are often presented in the form of quantiles or return levels. These return levels are typically expressed on an annual scale and represent the level that is expected to be exceeded once every P years. In the case where there are n_x observations per year, this is equivalent to the m-observation return level, where $m = P \times n_x$. Hence, the P-year return level can be expressed using the following equation (Coles, 2001):

$$Z_p = \begin{cases} \hat{u} + \dfrac{\hat{\sigma}}{\hat{\xi}}\left[\left(Pn_x\hat{\lambda}_u\right)^{\hat{\xi}} - 1\right], & \xi \neq 0, \\ \hat{u} + \hat{\sigma}\log\left(Pn_x\hat{\lambda}_u\right), & \xi = 0. \end{cases} \tag{10}$$

For the estimates of return levels, parameter estimation of GPD is required. However, it is also necessary to know the estimate of $\hat{\lambda}_u$, the probability of a single observation exceeding the threshold u. It can be estimated by using the following equation

$$\hat{\lambda}_u = \frac{k}{n}. \tag{11}$$

3. Results and Discussion

Figure 3 illustrates the rainfall pattern in Makassar City from January 1980 to December 2022. The city experiences a rainy season from November to April, with the highest amount of rainfall occurring from December to February, which is influenced by the west monsoon, also known as the Asian monsoon. During this period, the city encounters high humidity levels and frequent thunderstorms. Conversely, the dry season spans from May to October, and it is impacted by the east monsoon winds, commonly referred to as the Australian monsoon. These winds blow from the Australian continent to the Asian continent, and since the sun is in the northern hemisphere during this time, rain is uncommon, particularly in August and September.

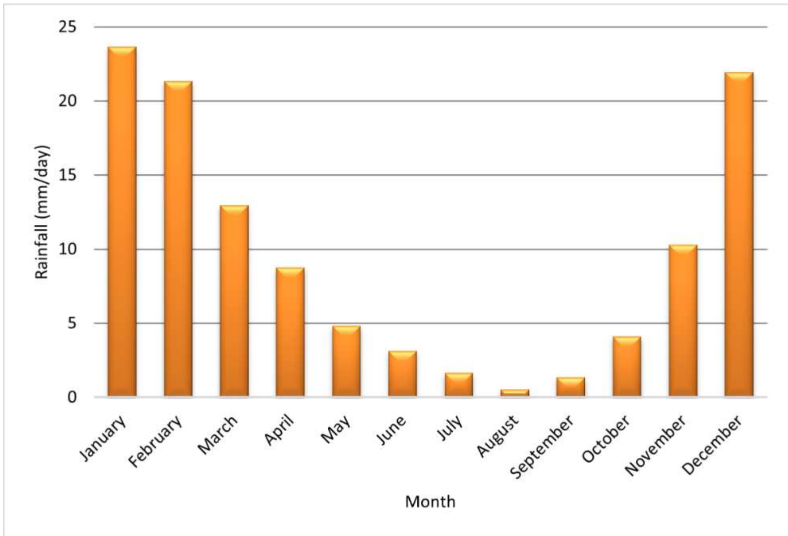

Figure 3. Distribution of the average rainfall in Makassar City for the period 1980–2022

The extreme rainfall threshold for each month was determined by creating a plot of the MRL with a 95% confidence interval, as outlined in Section 2.3. Table 1 shows that the extreme rainfall threshold in Makassar City is higher during the peak rainy season months (December to February) compared to other months. Specifically, the highest extreme rainfall threshold of 90 mm/day is observed in December. Hence, daily rainfall exceeding 90 mm/day in December is considered extreme. From March to August, the rainfall threshold decreases, while it increases from September to November. August has the lowest rainfall threshold of 5 mm/day, reflecting the peak of the dry season.

Table 1. Threshold selection for daily rainfall data and Mann-Kendall trend test

Month	Threshold	Number of events	P-value M-K test
January	86	60	0.650
February	80	57	0.553
March	60	57	0.289
April	45	58	0.909
May	28	59	0.521
June	20	55	0.127
July	9	62	0.104
August	5	34	0.265
September	10	42	0.508
October	21	68	0.494
November	48	62	1.00
December	90	63	0.887

To determine the presence of trends or seasonal patterns in extreme rainfall data obtained through the POT method, a visual inspection of the time series plot was conducted. As shown in Figure 4, the occurrences of extreme rainfall exhibit stochastic fluctuations in all months, indicating the absence of any identifiable patterns or recurrent cycles in the series. Additionally, there is no visible upward or downward trend in the series. The statistical analysis performed using the M-K test with a confidence level of 95% confirmed the absence of any trend in the data for all months.

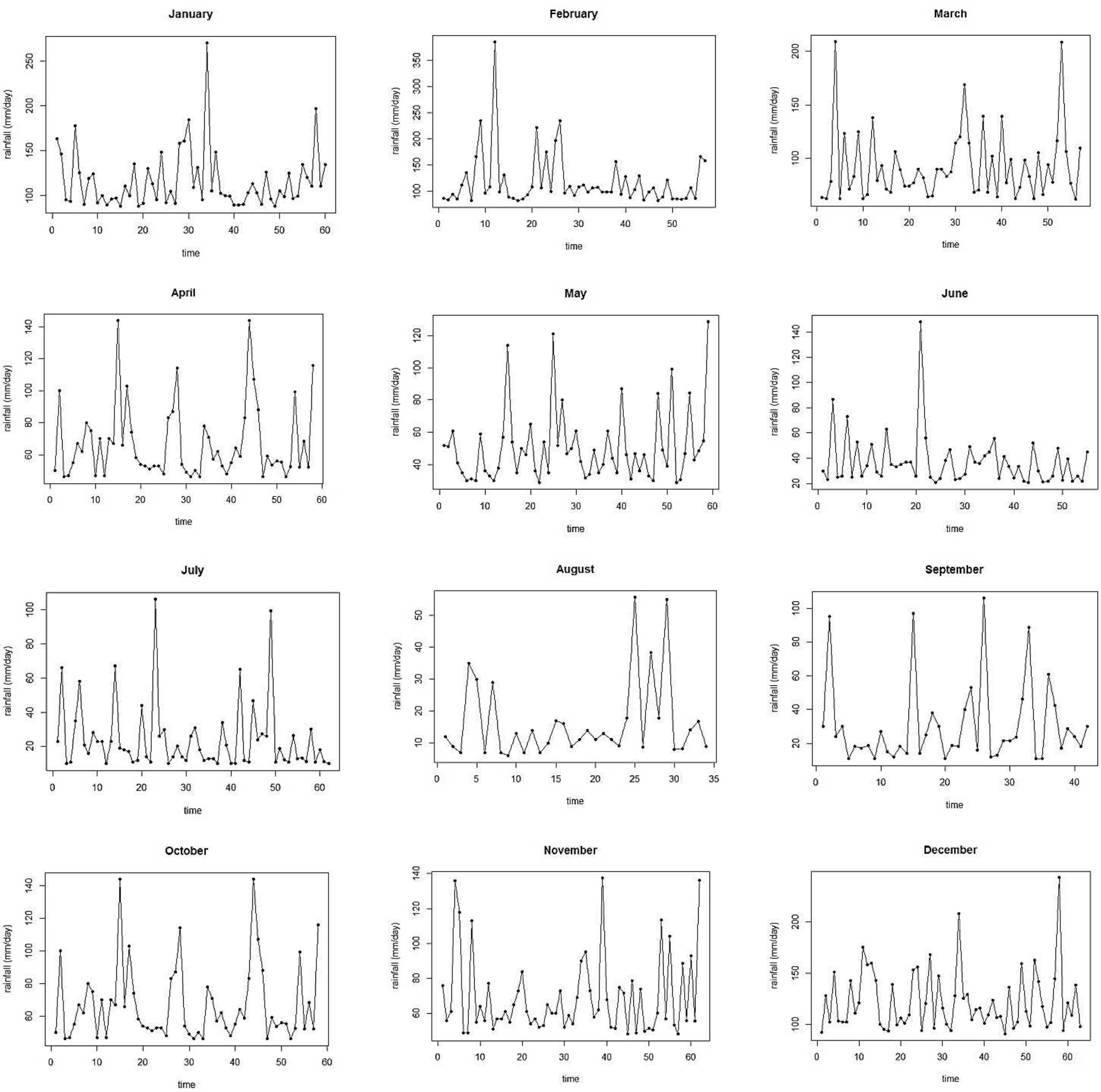

Figure 4. Time series graph of extreme precipitation events according to the POT approach

After verifying the assumptions of the extreme value theory model, the GPD was applied to model the extreme rainfall series for each month. Table 2 provides the point estimates of parameters and their corresponding 95% confidence intervals, which were obtained using the maximum likelihood method. Furthermore, a goodness of fit test was performed using the K-S test at a confidence level of 95%. The results demonstrate that the GPD is appropriate for modeling rainfall that exceeds the predetermined threshold for all months. This is supported by the fact that the p-value of the K-S test is greater than 5%.

Table 2. Parameter estimation and Kolmogorov-Smirnov test results for extreme rainfall series in Makassar City

Month	Scale parameter		Shape parameter		P-value K-S test
	$\hat{\sigma}$	CI	$\hat{\xi}$	CI	
January	27.529	(17.050, 38.008)	0.088	(-0.197, 0.373)	0.976
February	27.976	(16.553, 39.399)	0.280	(-0.043, 0.603)	0.525
March	32.026	(19.733, 44.319)	0.011	(-0.271, 0.294)	0.999
April	20.525	(11.934, 29.115)	0.074	(-0.256, 0.405)	0.858
May	22.270	(13.859, 30.682)	0.008	(-0.270, 0.287)	0.859
June	14.854	(9.178, 20.531)	0.135	(-0.143, 0.413)	0.728
July	10.191	(5.830, 14.552)	0.349	(-0.010, 0.708)	0.671
August	9.893	(4.861, 14.924)	0.125	(-0.260, 0.510)	0.415
September	15.436	(7.647, 23.226)	0.256	(-0.155, 0.668)	0.908
October	20.340	(13.237, 27.443)	0.045	(-0.210, 0.301)	0.863
November	18.802	(11.117, 26.489)	0.110	(-0.215, 0.435)	0.844
December	35.490	(23.746, 47.235)	-0.096	(-0.318, 0.125)	0.983

The positive shape parameter for January indicates a polynomial decrease in the tail of the distribution, suggesting a possible Pareto distribution. However, since the 95% confidence interval for the shape parameter encompassess zero, this is inadequate evidence to confirm that the extreme rainfall distribution type for January is Pareto. Consequently, it may be appropriate to consider the exponential distribution in the analysis. The shape parameters for February to November are positive, indicating a heavy right tail, and suggesting that the distribution conforms to the Pareto distribution. However, the exponential distribution can also be considered since zero is within the 95% confidence interval for each month. In December, the maximum likelihood estimate for the shape parameter is negative, indicating that the extreme rainfall distribution type is beta. However, the 95% confidence interval extended above zero, which is insufficient evidence to conclude that the distribution type for December is beta. Thus, the exponential distribution could be considered as a potential candidate for further analysis.

Table 3. Likelihood ratio test to select the optimal fit model for extreme precipitation series in Makassar City

Month	Likelihood ratio test	Chi-square critical value	P-value	Decision
January	0.435	3.841	0.509	Exponential
February	4.748	3.841	0.029	Generalized Pareto
March	0.006	3.841	0.939	Exponential
April	0.210	3.841	0.646	Exponential
May	0.003	3.841	0.955	Exponential
June	1.293	3.841	0.255	Exponential
July	6.088	3.841	0.014	Generalized Pareto
August	0.459	3.841	0.498	Exponential
September	2.031	3.841	0.154	Exponential
October	0.132	3.841	0.717	Exponential
November	0.496	3.841	0.481	Exponential
December	0.579	3.841	0.447	Exponential

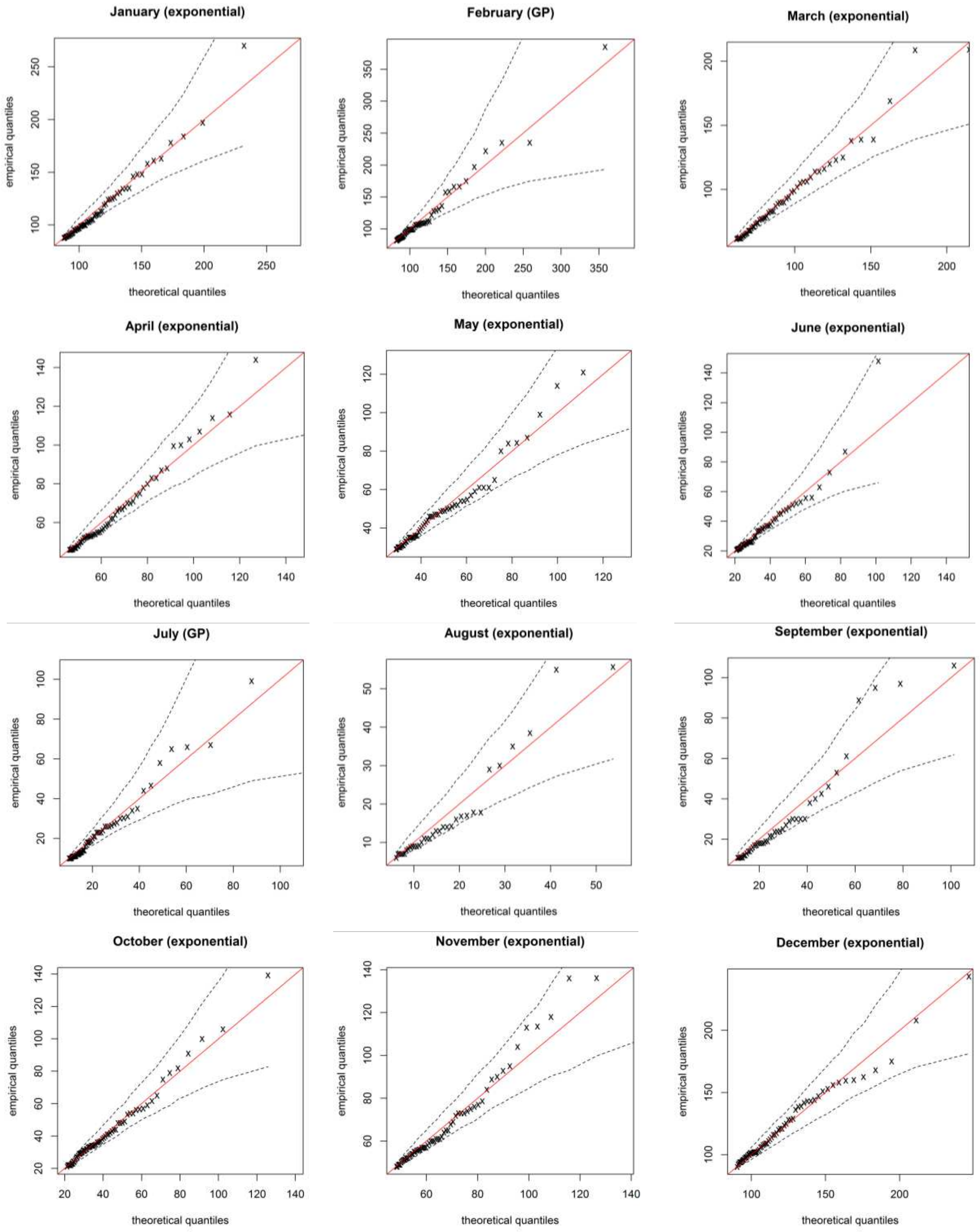

Figure 5. Q-Q plots of the optimal distribution model for the extreme precipitation series

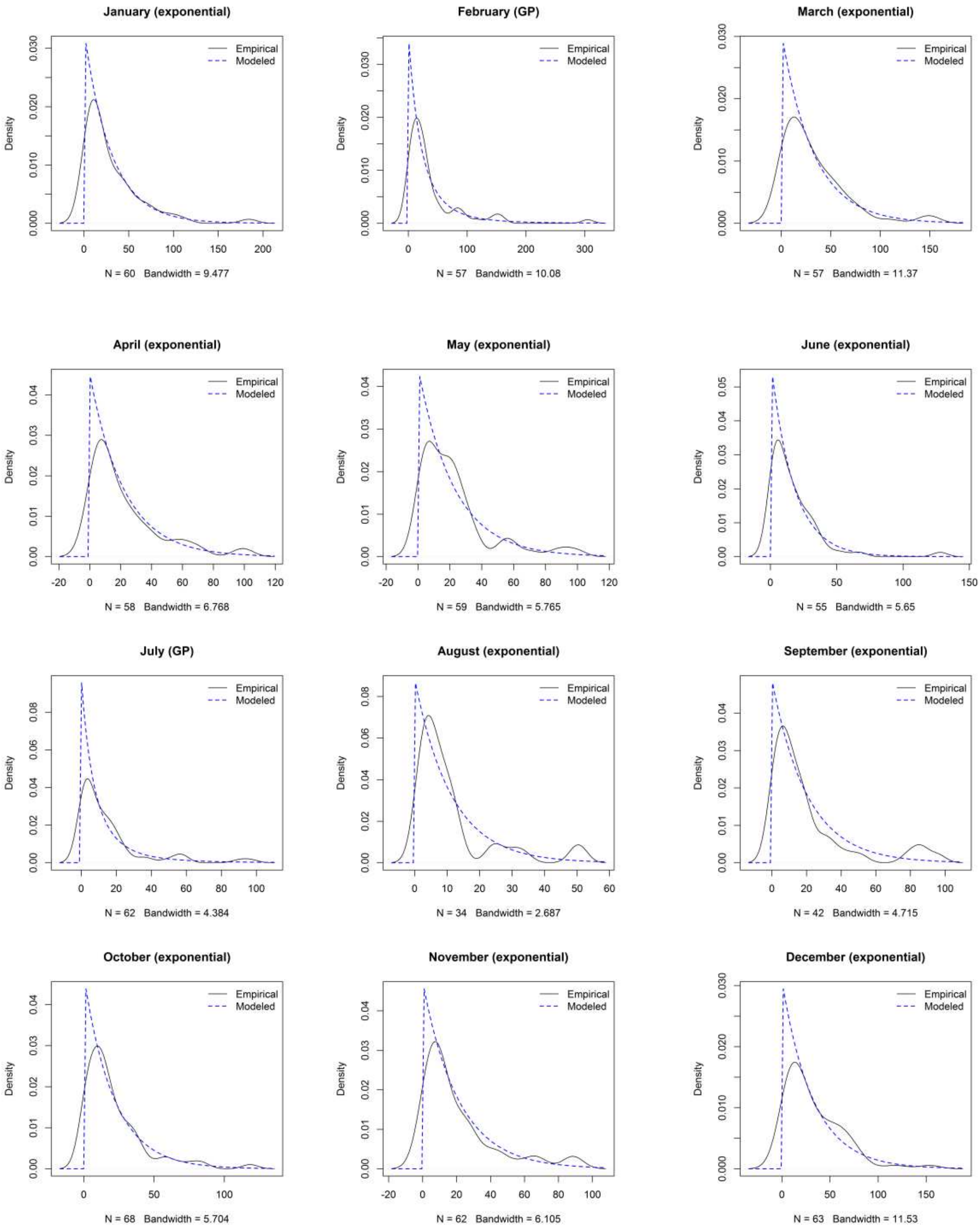

Figure 6. Empirical density vs best suitable distribution of extreme rainfall precipitation in Makassar City

The likelihood ratio test with a confidence level of 95% was performed to determine the best fit distribution between GPD and its nested model, the exponential. The model with fewer parameters (in this case, exponential) was selected if the likelihood ratio test yielded a smaller value than the critical threshold or if the p-value exceeded 0.05. According to Table 4, the exponential distribution is deemed the most appropriate model for extreme rainfall series in January, March, April, May, June, August, September, October, November, and December. While the GPD is most appropriate for February and July. To validate the chosen model, the techniques of Q-Q plot and density plot were employed. Figure 5 illustrates the empirical and theoretical quantiles, and the dashed lines represent the 95% confidence interval. The Q-Q plot indicates that the extreme rainfall series in each month conforms to its best fit distribution, as evidenced by the straight line. The density plot in Figure 6 demonstrates that the empirical distribution curves follow the best-fit distribution well. These findings suggest the validity of the best fit model for each month.

Table 4. Return levels of extreme rainfall for various return periods in Makassar City

Month	Distribution	Return period (year)			
		2	3	5	10
January	Exponential	193.1	205.3	220.8	241.7
	GP	200.7	216.2	236.6	265.7
February	Exponential	219.2	234.8	254.5	281.3
	GP	254.3	287.3	334.6	410.5
March	Exponential	174.7	187.8	204.4	226.9
	GP	175.7	189.2	206.4	229.8
April	Exponential	124.5	133.5	144.8	160.2
	GP	129.5	140.5	154.9	175.3
May	Exponential	108.2	117.3	128.8	144.3
	GP	108.7	118.0	129.8	145.9
June	Exponential	80.6	87.5	96.3	108.3
	GP	87.0	96.9	110.2	129.9
July	Exponential	63.7	69.8	77.6	88.1
	GP	82.5	98.1	121.1	159.8
August	Exponential	38.5	43.0	48.8	56.6
	GP	40.4	46.4	54.3	66.0
September	Exponential	75.9	84.2	94.6	108.8
	GP	87.5	102.6	123.9	157.8
October	Exponential	99.5	108.1	119.0	133.8
	GP	102.6	112.4	125.0	142.7
November	Exponential	124.9	133.4	144.2	158.7
	GP	132.4	144.0	159.5	181.8
December	Exponential	206.7	219.9	236.4	258.8
	GP	198.1	208.0	220.05	235.4

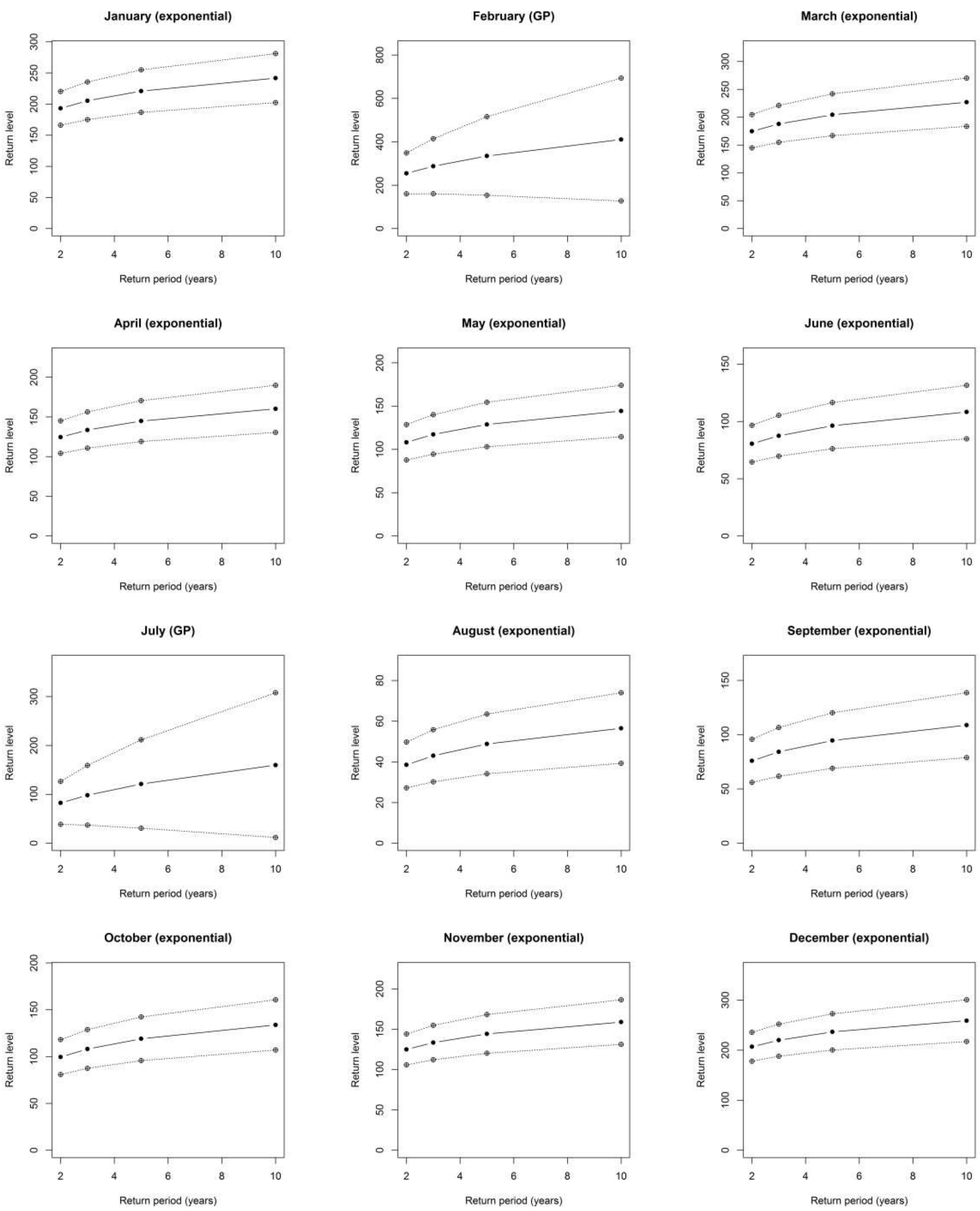

Figure 7. Return level of extreme rainfall in Makassar City with a 95% confidence interval

The GPD and exponential models were utilized to estimate the return for extreme precipitation events (mm/day) in Makassar City for various return periods, including 2, 3, 5, and 10 years. Table 4 displays the point estimates of the return levels obtained from the GP and exponential distributions. Additionally, Figure 7 showcases the return level plots along with their corresponding 95% confidence intervals for the most appropriate distribution. According to Table 5, February has the highest rainfall return level for all selected return periods compared to other months. December and January follow closely behind. By utilizing the GPD, the maximum amount of rainfall in February is expected to exceed the level of 254.3 mm/day once every 2 years with a probability of 0.851. However, there is some uncertainty around this estimate, and it is likely that the true maximum of rainfall could fall within the range of 160.3 to 348.4 mm/day with a 95% level of confidence, as shown in Figure 7. Additionally, it is predicted that the extreme rainfall in February above 287.3 mm/day will occur once every 3 years with a probability of 0.848, but the true maximum rainfall may lie within the range of 160.5 to 414.2 mm/day.

According to the exponential distribution, the probability of extreme rainfall in December exceeding 206.7 mm/day is 0.875, which corresponds to an expected return period of once every 2 years. This estimate comes with a 95% confidence interval ranging from 177.9 to 235.6 mm/day. For a 5-year return period, the return level point estimate of rainfall is 236.4 mm/day, meaning that this is the expected level of rainfall to occur in December. However, the actual return level of rainfall could fall within the range of 200.3 to 272.6 with a 95% level of confidence. For a longer return period, the extreme rainfall in January is expected to exceed 241.7 mm/day once every 10 years with a probability of 0.862 based on the exponential distribution. However, there is a possibility that the actual extreme rainfall could fall between 202.3 to 281.1 mm/day with a 95% level of confidence.

Makassar City is highly susceptible to flooding due to its topography, which consists of lowlands near the coast and the mouths of two large rivers, the Jeneberang and Tallo rivers (Musliadi et al., 2021). The peak of flooding in Makassar usually occurs from December to February, when the high intensity of rainfall leads to inundation. For instance, in mid-December 2017, suburbs of Makassar City experienced flooding with a height of one meter (Halim et al., 2019). In January 2019, a flood disaster hit the city and inundated 1658 houses, affecting 9328 residents (Thoban & Hizbaron, 2020). Even recently, on February 13, 2023, heavy rains with an intensity of 166.8 mm/day resulted in flooding in most parts of Makassar (VOI, 2023). Rudiyanto (2015) discovered that 32.9% of the city of Makassar is an area prone to inundation in his research on critical flood inundation zones in the city. Several areas in the Tamalate, Tamalanrea, and Biringkanaya sub-districts are at risk of flooding. The study's findings are expected to help the government develop flood prevention strategies and mitigate their effects. The government can use the estimated return levels to prepare for potential extreme rainfall events and reduce the negative impact of flooding on people's lives, infrastructure, and the economy.

4. Conclusions

In this paper, we propose the GPD with its nested model (exponential) to analyze the extreme rainfall events in Makassar City. To determine monthly daily rainfall that exceeded a certain threshold, the POT method was utilized. The extreme rainfall threshold in the city is higher during the peak rainy season months (December to February). The highest rainfall threshold of 90 mm/day is observed in December and there were 63 extreme rainfall events in Makassar City above 90 mm/day during the period 1980-2022. In contrast, August records the lowest rainfall threshold of 5 mm/day. We also explored possible trends and seasonal patterns but found no evidence.

According to the LR test results, the exponential distribution is the most appropriate model for extreme precipitation series in most months, except for February and July, where the GPD is more appropriate. The return levels of extreme rainfall from January to December were calculated for 2, 3, 5, and 10-year return periods. February has the highest rainfall return level for all selected return periods compared to other months. December and January follow closely behind. By utilizing the GPD, the maximum amount of rainfall in February is expected to exceed the level of 254.3 mm/day once every 2 years. For a 5-year return period, the return level point estimate of rainfall in December is 236.4 mm/day. For a longer return period, the extreme rainfall in January is expected to exceed 241.7 mm/day once every 10 years based on the exponential distribution. The GPD and exponential models can be useful in predicting the likelihood of extreme rainfall and informing decision-making in fields such as flood management. We recommend further work on applying spatial extreme value analysis using the GPD to understand and predict extreme events over a spatial region.

5. Acknowledgement

We are grateful to the Indonesian Meteorology, Climatology, and Geophysical Agency (BMKG) for supplying data. Besides, we wish to extend our appreciation to the editor and reviewers for their valuable contributions towards enhancing the quality of the manuscript.

6. References

Benyahya, L., Gachon, P., St-Hilaire, A., & Laprise, R. (2014). Frequency analysis of seasonal extreme precipitation in southern Quebec (Canada): An evaluation of regional climate model simulation with respect to two gridded datasets. Hydrology Research, 45(1), 115–133. https://doi.org/10.2166/nh.2013.066

Berger, V. W., & Zhou, Y. (2014). Kolmogorov–Smirnov Test: Overview. In N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri, & J. L. Teugels (Eds.), Wiley StatsRef: Statistics Reference Online (1st ed.). Wiley. https://doi.org/10.1002/9781118445112.stat06558

Boudrissa, N., Cheraitia, H., & Halimi, L. (2017). Modelling maximum daily yearly rainfall in northern Algeria using generalized extreme value distributions from 1936 to 2009: Modelling maximum rainfall in northern Algeria. Meteorological Applications, 24(1), 114–119. https://doi.org/10.1002/met.1610

Caires, S. (2016). A Comparative Simulation Study of the Annual Maxima and the Peaks-Over-Threshold Methods. Journal of Offshore Mechanics and Arctic Engineering, 138(5). https://doi.org/10.1115/1.4033563

Coles, S. (2001). An introduction to statistical modeling of extreme values. Springer.

Coles, S. G., & Dixon, M. J. (1999). Likelihood-Based Inference for Extreme Value Models. 2(1), 5–23.

Davison, A. C., & Smith, R. L. (1990). Models for Exceedances Over High Thresholds. Journal of the Royal Statistical Society: Series B (Methodological), 52(3), 393–425. https://doi.org/10.1111/j.2517-6161.1990.tb01796.x

Gilli, M., Maringer, D., & Schumann, E. (2019). Chapter 11-Basic Methods. In Numerical Methods and Optimization in Finance (2nd edition, pp. 229–271). Elsevier Inc. https://doi.org/10.1016/B978-0-12-815065-8.00023-6

Halim, H., Arifin, A., Nonci, N., Zainuddin, R., Anriani, H. B., & Kamaruddin, S. A. (2019). Flood disaster and risk anticipation strategy. IOP Conference Series: Earth and Environmental Science, 235, 012032. https://doi.org/10.1088/1755-1315/235/1/012032

Jentsch, A., & Beierkuhnlein, C. (2008). Research frontiers in climate change: Effects of extreme meteorological events on ecosystems. Comptes Rendus Geoscience, 340(9–10), 621–628. https://doi.org/10.1016/j.crte.2008.07.002

Kusumastuti, C., & Weesakul, S. (2014). Extreme Rainfall Indices for Tropical Monsoon Countries in Southeast Asia. Civil Engineering Dimension, 16(2), 112–116. https://doi.org/10.9744/ced.16.2.112-116

Li, Z., Brissette, F., & Chen, J. (2014). Assessing the applicability of six precipitation probability distribution models on the Loess Plateau of China: APPLICABILITY OF SIX PRECIPITATION PROBABILITY DISTRIBUTION MODELS. International Journal of Climatology, 34(2), 462–471. https://doi.org/10.1002/joc.3699

Li, Z., Li, Z., Zhao, W., & Wang, Y. (2015). Probability Modeling of Precipitation Extremes over Two River Basins in Northwest of China. Advances in Meteorology, 2015, 1–13. https://doi.org/10.1155/2015/374127

Loo, Y. Y., Billa, L., & Singh, A. (2015). Effect of climate change on seasonal monsoon in Asia and its impact on the variability of monsoon rainfall in Southeast Asia. Geoscience Frontiers, 6(6), 817–823. https://doi.org/10.1016/j.gsf.2014.02.009

Martins, A. L. A., Liska, G. R., Beijo, L. A., Menezes, F. S. de, & Cirillo, M. Â. (2020). Generalized Pareto distribution applied to the analysis of maximum rainfall events in Uruguaiana, RS, Brazil. SN Applied Sciences, 2(9), 1479. https://doi.org/10.1007/s42452-020-03199-8

McKnight, T. L., & Darrel, H. (2000). Physical Geography: A Landscape Appreciation. Upper Saddle River, N.J. : Prentice Hall.

Musliadi, Chaerul, M., & Gusty, S. (2021). Study on Flood Simulation of Tallo Watershed, Makassar City, South Sulawesi Province. Journal of Physics: Conference Series, 1899(1), 012063. https://doi.org/10.1088/1742-6596/1899/1/012063

Pickands, J. (1975). Statistical Inference Using Extreme Order Statistic. The Annals of Statistics, 3(1), 119–131. https://doi.org/10.1214/aos/1176343003

Rahim, A., Bakri, B., Anisa, Mutholib, A., & Haerunnisa, A. (2019). Characteristic of extreme rainfall in Makassar, Province of South Sulawesi. Journal of Physics: Conference Series, 1341(9), 092017. https://doi.org/10.1088/1742-6596/1341/9/092017

Rahman, A. S., Rahman, A., Zaman, M. A., Haddad, K., Ahsan, A., & Imteaz, M. (2013). A study on selection of probability distributions for at-site flood frequency analysis in Australia. Natural Hazards, 69(3), 1803–1813. https://doi.org/10.1007/s11069-013-0775-y

Reiss, R.-D., & Thomas, M. (2007). Statistical Analysis of Extreme Values: With applications to insurance, finance, hydrology and other fields (3rd edition). Birkhäuser.

Rudiyanto, R. (2015). Identification of the Makassar City Flood Inundation Zone Based on GIS [Undergraduated Thesis of Civil]. Hasanuddin University.

Sanusi, W., Mulbar, U., Jaya, H., Purnamawati, & Side, S. (2017). Modeling of Rainfall Characteristics for Monitoring of the Extreme Rainfall Event in Makassar City. American Journal of Applied Sciences, 14(4), 456–461. https://doi.org/10.3844/ajassp.2017.456.461

Shaharudin, S. M., Ahmad, N., Mohamed, N. S., & Mahdin, H. (2020). Fitting statistical distribution of extreme rainfall data for the purpose of simulation. Indonesian Journal of Electrical Engineering and Computer Science, 18(3), 1367. https://doi.org/10.11591/ijeecs.v18.i3.pp1367-1374

Singirankabo, E., & Iyamuremye, E. (2022). Modelling extreme rainfall events in Kigali city using generalized Pareto distribution. Meteorological Applications, 29(4). https://doi.org/10.1002/met.2076

Sunusi, N. & Giarno. (2022). Comparison of some schemes for determining the optimal number of rain gauges in a specific area: A case study in an urban area of South Sulawesi, Indonesia. AIMS Environmental Science, 9(3), 260–276. https://doi.org/10.3934/environsci.2022018

The Ministry of Foreign Affairs of Iran. (2020). *Indonesia's climate and precipitation*. https://indonesia.mfa.gov.ir/en/generalcategoryservices/13009/indonesias-climate-and-precipitation

Thoban, M. I., & Hizbaron, D. R. (2020). Urban resilience to floods in parts of Makassar, Indonesia. E3S Web of Conferences, 200, 01007. https://doi.org/10.1051/e3sconf/202020001007

VOI. (2023). BMKG: The Rob Flood In Makassar City Was Triggered By Sea Waves Rising To A Height Of 4 Meters. https://voi.id/en/news/253896

Wu, G., & Qiu, W. (2018). Threshold Selection for POT Framework in the Extreme Vehicle Loads Analysis

Based on Multiple Criteria. Shock and Vibration, 2018, 1–9. https://doi.org/10.1155/2018/4654659

Wu, H., Soh, L.-K., Samal, A., & Chen, X.-H. (2008). Trend Analysis of Streamflow Drought Events in Nebraska. Water Resources Management, 22(2), 145–164. https://doi.org/10.1007/s11269-006-9148-6

Yin, C., Wang, J., Yu, X., Li, Y., Yan, D., & Jian, S. (2022). Definition of Extreme Rainfall Events and Design of Rainfall Based on the Copula Function. Water Resources Management, 36, 3759–3778. https://doi.org/10.1007/s11269-022-03229-7

Zalina, M. D., Desa, M. N. M., Nguyen, V.-T.-A., & Kassim, A. H. M. (2002). Selecting a probability distribution for extreme rainfall series in Malaysia. Water Science and Technology, 45(2), 63–68. https://doi.org/10.2166/wst.2002.0028