MJFAS
Malaysian Journal of Fundamental and Applied Sciences

RESEARCH ARTICLE

Stochastic Model of the Annual Maximum Rainfall Series Using Probability Distributions

Nurul Azizah Muzakir, Nurtiti Sunusi*, Sri Astuti Thamrin

Department of Statistics, Faculty of Mathematics and Natural Sciences, Hasanuddin University, 90245 Makassar, Sulawesi Selatan, Indonesia

Abstract Rainfall is a natural process that is often characterized by significant variability and uncertainty. Stochastic models of rainfall typically involve the use of probability distributions to describe the likelihood of different outcomes occurring. This study aimed to model the annual maximum of daily rainfall in Makassar City, Indonesia for the period 1980–2022, specifically focusing on the rainy season (November to April) using probability distributions to estimate return periods. The study used the Generalized Extreme Value (GEVD) and Gumbel distributions. The Kolmogorov-Smirnov test was used to determine the suitability of each distribution, and the likelihood ratio test was employed to determine the best-fit model. The Mann-Kendall test was used to detect any trends in the data. The results indicated that the Gumbel distribution was the best-fit model for data in November, December, January, March, and April, while GEV was appropriate for February. No trends were observed in any of the months. The study then estimated the maximum rainfall for various return periods. January produced the highest maximum rainfall estimates for the 2, 3, and 5-year return periods, while February produced the highest maximum rainfall estimates for the 10 and 20-year return periods. Information about maximum rainfall can be valuable for the government and other stakeholders in developing flood prevention strategies and mitigating the effects of heavy rainfall, particularly during the peak months of the rainy season in Makassar City, which are December, January, and February.

Keywords: Maximum rainfall, GEV distribution, Gumbel distribution, return period, rainy season.

Introduction

Rainfall is a natural phenomenon that occurs randomly, making it a perfect example of a stochastic process. Stochastic processes are mathematical models that describe the behavior of random variables over time [1]. Indonesia is a tropical country with a diverse climate, and rainfall patterns vary significantly across the country. Indonesia experiences two separate seasons, which are the dry and rainy seasons. The rainy season typically begins in November and lasts until April, with the highest rainfall occurring between December and February. During this time, extreme rainfall and thunderstorms are common in many parts of the country, particularly in the western and central regions. Heavy precipitation occurrences can greatly affect both the environment and society. It can lead to flooding, landslides, and other hazards that can damage infrastructure, crops, and other property [2,3]. A stochastic model is a mathematical representation of a random process. The stochastic model is used to describe and simulate the behavior of extreme rainfall events, and it can be calibrated and validated using the available maximum rainfall data [4].Stochastic models of extreme rainfall typically involve the use of probability distributions to describe the likelihood of different outcomes occurring and to estimate the associated uncertainty [5,6].

Several probability distributions have been suggested to represent the annual hydrological extremes in a particular area. Among the most frequent distributions utilized for extreme rainfall are Beta-K, Beta-P, Gamma, Generalized Extreme Value (GEVD), Generalized Pareto (GPA), Generalized Logistic (GLO), Pearson Type III, Lognormal, and Wakeby [7]. Different countries have their own national guidelines that suggest using various distributions. In countries like Austria, Australia, Germany, Italy, and Spain, the GEV distribution is the preferred option. Meanwhile, the GPA and GLO distributions are recommended in Belgium and the United Kingdom, respectively [8,9]. Several researchers in Indonesia have made an attempt to identify the most suitable probability distribution for representing the maximum rainfall data.

***For correspondence:**
nurtitisunusi@unhas.ac.id

Received: 23 March 2023
Accepted: 21 Sept. 2023

For instance, Sanusi *et al*. [10] evaluated different distributions, including GEVD, GPA, GLO, and Pearson Type III, to monitor maximum precipitation in Makassar City. Their findings revealed that among these distributions, the GEVD demonstrated superior performance. Another study conducted by Sanusi *et al*. [11] discovered that the annual maximum rainfall data in various South Sulawesi Province districts, such as Pinrang and Takalar, was best represented by the GEVD. Susanti *et al*. [12] concluded that the GEVD performed better than the GPA distribution for maximum rainfall data in Pekanbaru City, Riau.

In stochastic modeling, the Generalized Extreme Value distribution (GEVD) can be used to model the distribution of extreme values for a variety of random variables, including rainfall, wind speeds, river flows, and earthquake magnitudes. The GEVD is particularly useful when the distribution of the data is unknown or cannot be well-modeled by other distributions, such as the normal distribution.The GEVD is commonly referred to as a family of distributions as it includes three extreme value distributions, namely Gumbel, Frechet, and Weibull [13]. It was first introduced to meteorology by Jenkinson in 1955 [14], and it has been widely used for extreme rainfall analysis [15]. The GEVD has been utilized by researchers from various parts of the globe, including Park *et al*. [16] and Nadarajah and Choi [17] for studies conducted in South Korea, Nashwan *et al*. [18] for studies conducted in Malaysia, Kumar *et al*. [19] and Babar and Ramesh [20] for studies conducted in India, and Abbas *et al*. [21] and Aurangzeb *et al*. [22] for studies conducted in Pakistan.

The return period helps us understand how often we can anticipate a certain amount of rainfall occurring, on average, within a given timeframe. The return period of extreme rainfall refers to the average interval of time over which a particular level of rainfall intensity is expected to be equaled or exceeded. Return period is often associated with return level. Various studies have implemented the GEVD to estimate the maximum rainfall return level. As an example, Boudrissa *et al*. [23] employed the GEVD to simulate the annual maximum rainfall data in northern Algeria, and calculated the return levels of the event. Their outcomes revealed that the return levels of maximum rainfall in diverse locations in Algeria, such as Miliana, Algiers, and Oran, showed an increase as the return period increased. The same results were also obtained by Min and Halim [24] for the return levels of maximum rainfall in Malaysia.

The aim of this study is to model the annual maximum of daily rainfall series in Makassar City, located in the central part of Indonesia, using probability distributions, and to estimate maximum rainfall for varous return periods. The Generalized Extreme Value distribution (GEVD) and its nested model, Gumbel, are considered for this purpose. To analyze daily rainfall series, it is common to first examine rainy and non-rainy months separately. This research exclusively focuses on the rainy season in Makassar City (November to April), as heavy rainfall during this period increases the likelihood of natural disasters such as floods and landslides. The study also predicts the probability of extreme events occurring in Makassar City, which can be used by stakeholders to develop effective mitigation measures and early warning systems for hydrometeorological disasters.

Materials and Methods

Data and Location

This research was conducted in Makassar, the capital city of Indonesian province of South Sulawesi. It is the biggest city in Eastern Indonesia and the country's fifth-largest metropolitan center. Makassar city is located at coordinates 5.133° S and 199.417° E. The city has a total area of 175.77 km^2. The highest elevation in Makassar City is 25 meters above sea level. Makassar is a tropical area with a tropical monsoon climate type. This is shown by the contrast of the rainy and dry seasons. The rainy season in the city occurs from November to April, with the heaviest rainfall occurring from December to February. During this period, the city experiences high humidity levels and frequent thunderstorms. In contrast, the dry season in the city lasts from May to October. Although the city still receives some precipitation during this time, it is generally sporadic and in the form of brief showers. The months of August and September are typically the driest months [10,25].

The daily rainfall data (mm/day) recorded at the Hasanuddin rain gauge station from 1980 to 2022 were used. The data is provided online by the Indonesian Meteorology, Climatology and Geophysics Agency (abbreviated as BMKG) through the following website: https://dataonline.bmkg.go.id/home. The Hasanuddin rain gauge station was chosen because of its data completeness and it had the longest period of data variability. The data was grouped by month, and then the maximum rainfall in each year of observation was determined, which we call the monthly annual maximum rainfall data. Because this study only focused on rainfall data in the rainy season, we only analyzed the maximum rainfall from November to April. Table 1 shows the data structure.

Table 1. Structure of monthly annual maximum rainfall data

Year	Maximum rainfall (mm/day)					
	November	December	January	February	March	April
1980	37	128	163	70	49	36
1981	76	85	79	87	63	50
1982	30	74	84	95	78	44
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2022	136.1	138.4	134.5	166.3	109.6	29.6

Generalized Extreme Value Distribution

The Generalized Extreme Value distribution (GEVD) is a continuous probability distribution used in statistics to model the extreme values in a dataset. It is a three-parameter distribution that generalizes the Gumbel, Frechet, and Weibull distributions. The GEVD is defined by the following cumulative distribution function (CDF) [26]:

$$F(z;\mu,\sigma,\xi)=e^{-\left(1+\xi\left(\frac{z-\mu}{\sigma}\right)\right)^{-1/\xi}} \tag{1}$$

defined on the set $\left\{z:1+\xi\left(\frac{z-\mu}{\sigma}\right)>0\right\}$. The location parameter, $-\infty<\mu<\infty$, determines the location of the distribution center, the scale parameter, $\sigma>0$, determines the spread or scale of the distribution, and the shape parameter, $-\infty<\xi<\infty$, determines the type of extreme value distribution (EVD). The GEVD leads to the Frechet or type II EVD for $\xi>0$. If $\xi<0$, then it is the type III EVD or reversed Weibull. The Gumbel distribution is a special case of the GEVD, with a shape parameter of $\xi=0$. The Gumbel distribution is also known as the type I EVD with the CDF as in Eq. (2) below.

$$F(z;\mu,\sigma)=e^{-e^{\left(-\frac{z-\mu}{\sigma}\right)}} \tag{2}$$

The probability density function (PDF) for the GEVD is obtained by taking the first derivative of the CDF with respect to z as shown in Equation (3) for $\xi\neq 0$ and Equation (4) for $\xi=0$.

$$f(z;\mu,\sigma,\xi)=\frac{1}{\sigma}\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-\left(\frac{1}{\xi}\right)-1}e^{-\left(1+\xi\left(\frac{z-\mu}{\sigma}\right)\right)^{-1/\xi}} \tag{3}$$

$$f(z;\mu,\sigma)=\frac{1}{\sigma}e^{\left(-\frac{z-\mu}{\sigma}\right)}e^{-e^{\left(-\frac{z-\mu}{\sigma}\right)}} \tag{4}$$

Parameter Estimation

The Maximum Likelihood Estimation (MLE) method is one way to estimate unknown population parameters. This method is a classic and is most widely used because of its simplicity [27]. In the process, this method seeks to find estimator for parameters that can maximize the likelihood function. Suppose $z_1,z_2,\dots,z_n$ are GEV-distributed random samples. The likelihood function from the PDF of the GEVD as in Equation (5) for $\xi\neq 0$ and Equation (6) for $\xi=0$.

$$L(\mu,\sigma,\xi|z_1,\dots,z_n)=\prod_{i=1}^{n}f(z_i|\mu,\sigma,\xi)=\prod_{i=1}^{n}\left[\frac{1}{\sigma}\left[1+\xi\left(\frac{z_i-\mu}{\sigma}\right)\right]^{-\left(\frac{1}{\xi}\right)-1}e^{-\left(1+\xi\left(\frac{z_i-\mu}{\sigma}\right)\right)^{-1/\xi}}\right] \tag{5}$$

$$L(\mu,\sigma|z_1,\dots,z_n)=\prod_{i=1}^{n}f(z_i|\mu,\sigma)=\prod_{i=1}^{n}\left[\frac{1}{\sigma}e^{\left(-\frac{z_i-\mu}{\sigma}\right)}e^{-e^{\left(-\frac{z_i-\mu}{\sigma}\right)}}\right] \tag{6}$$

When attempting to maximize the likelihood function in Equations (5) and (6), one can also aim to maximize the logarithm of the likelihood function. As a result, for $\xi\neq 0$, the ln likelihood function can be expressed as follows:

$$\begin{aligned}\ln L(\mu,\sigma,\xi|x_i)&=\ln\prod_{i=1}^{n}f(z_i|\mu,\sigma,\xi)=\sum_{i=1}^{n}\ln f(z_i|\mu,\sigma,\xi)\\&=\sum_{i=1}^{n}\ln\left[\frac{1}{\sigma}\left[1+\xi\left(\frac{z_i-\mu}{\sigma}\right)\right]^{-\left(\frac{1}{\xi}\right)-1}e^{-\left(1+\xi\left(\frac{z_i-\mu}{\sigma}\right)\right)^{-1/\xi}}\right]\\&=-n\ln(\sigma)+\left(-\frac{1}{\xi}-1\right)\sum_{i=1}^{n}\ln\left(1+\xi\left(\frac{z_i-\mu}{\sigma}\right)\right)-\sum_{i=1}^{n}\left(1+\xi\left(\frac{z_i-\mu}{\sigma}\right)\right)^{-1/\xi}\end{aligned} \tag{7}$$

While the ln likelihood function for $\xi=0$ as in Equation (8) below.

$$\ln L(\mu,\sigma|z_i) = \ln \prod_{i=1}^{n} f(z_i|\mu,\sigma) = \sum_{i=1}^{n} \ln f(z_i|\mu,\sigma)$$
$$= \sum_{i=1}^{n} \ln\left[\frac{1}{\sigma} e^{\left(-\frac{z_i-\mu}{\sigma}\right)} e^{-e^{\left(-\frac{z_i-\mu}{\sigma}\right)}}\right]$$
$$= -n\ln(\sigma) - \sum_{i=1}^{n}\left(\frac{z_i-\mu}{\sigma}\right) - \sum_{i=1}^{n} e^{\left(-\frac{z_i-\mu}{\sigma}\right)} \qquad (8)$$

Once the ln likelihood function is obtained, the next step is to determine the first-order partial derivative of the ln likelihood function for each parameter and set it equal to zero. If no analytical solution is possible, then parameter estimation is performed using numerical methods, one of which is the BFGS quasi-Newton [28].

Goodness of Fit Test

Goodness of fit tests establish how well sample data fits what is expected of a population. The Kolmogorov-Smirnov (K-S) test is a one-sample test used to determine how well a given set of data fits a theoretical distribution [29]. Assume that X is a random variable that comes from a population that follows a certain distribution. $F_T(x)$ represents the CDF of the reference distribution, and $F_S(x)$ is the empirical CDF of the sample. The hypotheses utilized in the K-S test are as follows:

$H_0 : F_S(x) = F_T(x)$
$H_1 : F_S(x) \neq F_T(x)$
The statistical test is given by

$$D = max|F_S(x) - F_T(x)| \qquad (9)$$

Hypothesis null is rejected if $D > D_\alpha$ in the one sample Kolmogorov-Smirnov table with a significance level of α. Alternatively, the null hypothesis is rejected if the $p - value < \alpha$.

Likelihood Ratio Test

The likelihood ratio (LR) test is a statistical method used to compare the goodness of fit of two nested models, where one model is a simplified version of the other. It is commonly used in hypothesis testing to determine if a more complex model provides a better fit to the data than a simpler model. Assume M_0 is the GEVD model with parameter of $\theta_0 = (\mu,\sigma,\xi)$ and M_1 is its nested model, that is the Gumbel distribution with parameter of $\theta_1 = (\mu,\sigma)$. The hypotheses are as follows:

H_0 : the simpler model (Gumbel) provides a good fit to the annual maximum rainfall data
H_1 : the more complex model (GEVD) provides a better fit to the annual maximum rainfall data than the simpler model (Gumbel)

The test statistic, which is the difference in log-likelihoods between the two models, can be expressed as follows [30]:

$$LRT = 2\left[l(\hat{\theta}_0) - l(\hat{\theta}_1)\right] \qquad (10)$$

where $l(\hat{\theta}_0)$ represents the log-likelihood for the GEVD model and $l(\hat{\theta}_1)$ represents the log-likelihood for the Gumbel model. The critical value of the LR test is calculated from the distribution of the test statistic, which is approximated by a chi-square distribution with degrees of freedom equal to 1 in this case at the significance level of α. If the calculated LR is greater than the critical value, then there is sufficient evidence to reject the null hypothesis and conclude that the GEVD model provides a better fit to the data. Alternatively, the null hypothesis is rejected if the $p - value < \alpha$.

Mann-Kendall Trend Test

The Mann-Kendall (M-K) test is applied to the data to check for trends in order to satisfy the GEV distribution's stationary assumption. One benefit of utilizing the M-K method is that it can identify patterns in temporal data, including both linear and nonlinear trends [31]. If the time series data is known with data length $n\ (x_1, x_{2,...,}x_n)$, then the test statistic is given by:

$$S = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a_{ij} \qquad (11)$$

where

$$a_{ij} = sign(x_j - x_i) = sign(R_j - R_i) = \begin{cases} 1 \; if \; x_j - x_i \; > \; 0 \\ 0 \; \text{if} \; x_j - x_i \; = \; 0 \\ -1 \; if \; x_j - x_i \; < \; 0 \end{cases} \quad (12)$$

For large sample sizes ($n \geq 8$), the normal distribution approach can be used for the distribution of S with the mean and variance as follows.

$$E(S) = 0 \quad (13)$$

$$V(S) = \frac{n(n-1)(2n+5)}{18} \quad (14)$$

Hypothesis testing using the Z test as in Equation (15) below.

$$Z = \begin{cases} \dfrac{S-1}{\sqrt{V(S)}} \; if \; S > 0 \\ 0 \; if \; S = 0 \\ \dfrac{S+1}{\sqrt{V(S)}} \; if \; S < 0 \end{cases} \quad (15)$$

The Z value is the standard normal distribution. The trend is considered significant if $|Z| > Z_{\alpha/2}$ where α denotes the significance level.

Return Level

A return level represents the expected value of a variable that is exceeded, on average, once in a specified number of years. The p-year return level is defined as the high quantile with a probability of $1/p$ that the annual maximum exceeds this quantile. The p-year return level is associated with a return period of p years. The return level is the same for all years under the assumption of stationary [32]. Using the GEV distribution, The p-year return level (for $p > 1$) is given by [33]:

$$\hat{z}_p = \begin{cases} \hat{\mu} + \dfrac{\hat{\sigma}}{\hat{\xi}}\left[\left(-\log(1-p^{-1})\right)^{-\hat{\xi}} - 1\right], & \hat{\xi} \neq 0 \\ \hat{\mu} - \hat{\sigma}\log\left(-\log(1-p^{-1})\right), & \hat{\xi} = 0 \end{cases} \quad (16)$$

Results and Discussion

Data Description

A data description was conducted to determine the characteristics of annual maximum rainfall in Makassar City during the rainy months from 1980 to 2022. Figure 1 provides a summary of the data's statistics. The results indicate that the Hasanuddin rain gauge station reported the highest maximum rainfall of 385 mm/day in February, followed by 270 mm/day in January, and 243.3 mm/day in December. Rainfall intensity generally increased from December to February, which corresponds to Makassar City's peak rainy season. The average maximum rainfall was highest in January at 113 mm/day and lowest in April at 60.35 mm/day. The maximum rainfall in November had the smallest standard deviation, while February had the largest. This suggests that maximum rainfall in February varies greatly. Figure 2 shows the time series plot of the annual maximum rainfall in Makassar City during the rainy season. From the plot, the maximum rainfall series fluctuated around its average in each month of observation, and visually, there was no upward or downward trend in the series.

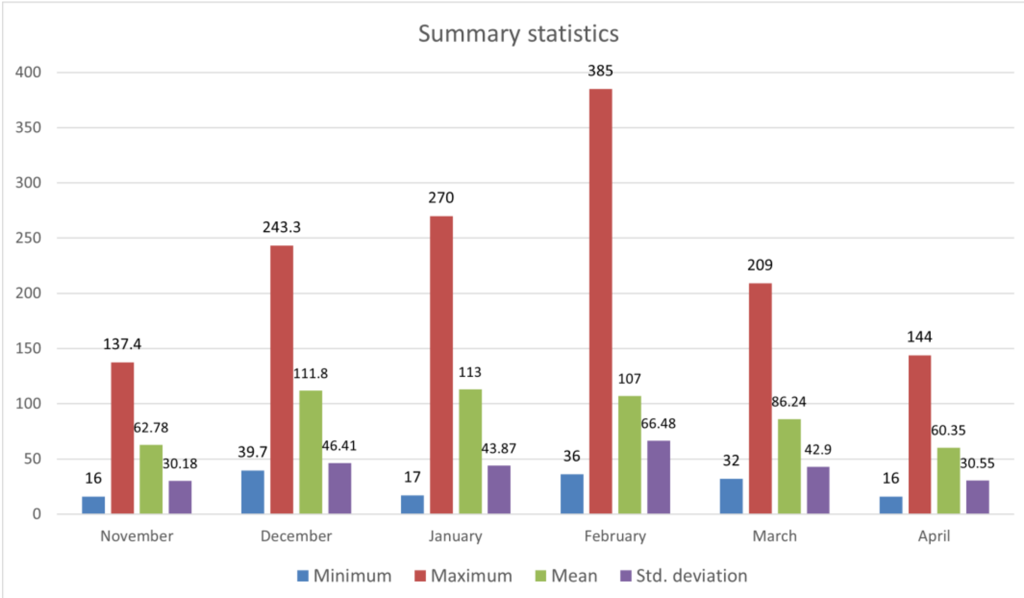

Figure 1. Summary statistics of annual maximum rainfall series during the rainy season from 1980–2022 in Makassar City.

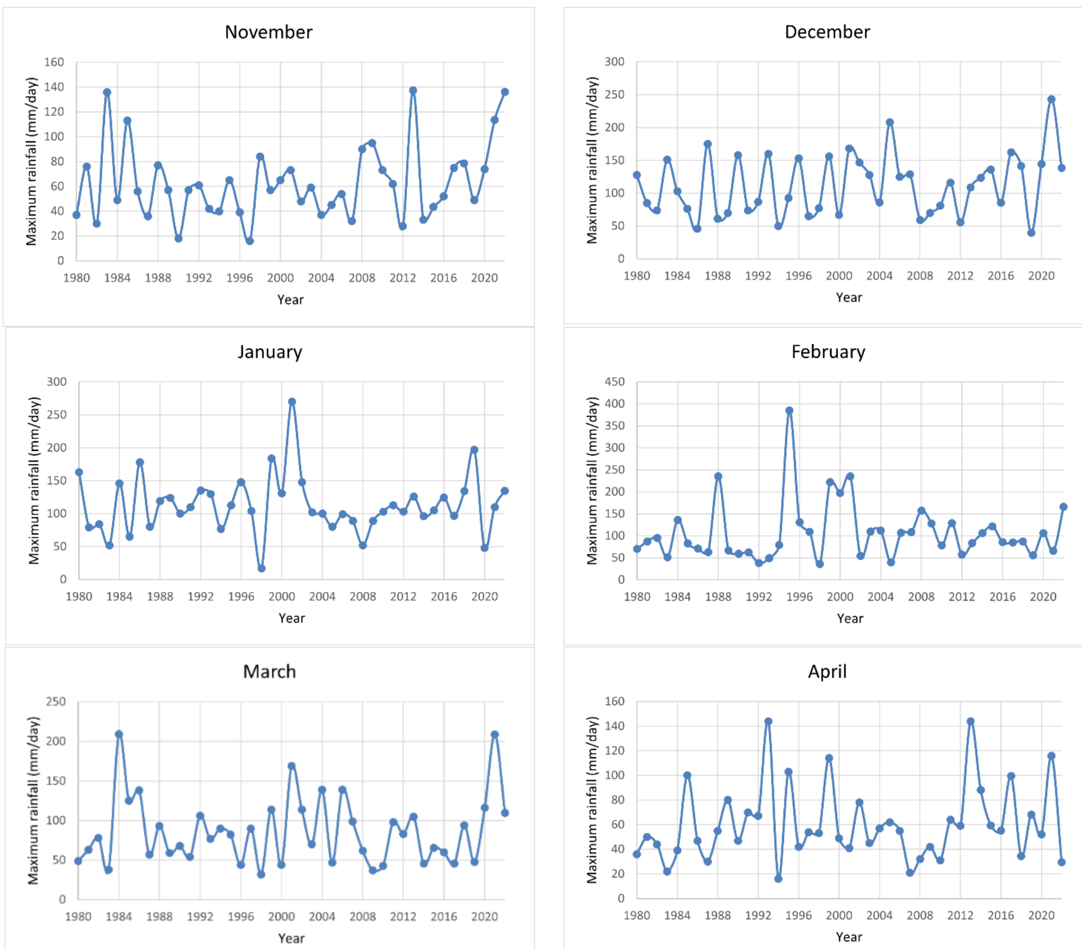

Figure 2. Time series plot of annual maximum rainfall in Makassar City over the period 1980–2022 in the rainy season

Parameter Estimation and Model Diagnostics

The annual maximum rainfall series was modeled with the GEVD. The results of the point parameter estimates and 95% confidence intervals using the MLE method are presented in Table 2. The shape parameter determined the type of GEVD. The point estimate of the shape parameter for November was positive, meaning that the distribution of the maximum rainfall series in November leads to the GEVD type II (Frechet distribution). However, the 95% confidence interval for the shape parameter contained zero, indicating insufficient evidence to support this conclusion. Therefore, it may be appropriate to consider its nested model, that is Gumbel distribution (also known as GEVD type I) in the analysis. Although the maximum likelihood estimates for ξ were negative in December and January, indicating a distribution with finite tails, the 95% confidence interval extended above zero. Hence, it was insufficient to conclude that the distribution's tails were finite. As a result, the Gumbel distribution could be a candidate. For February, the GEVD type was Frechet because the shape parameter was positive, which was supported by the fact that the lower and upper bounds of the 95% confidence intervals were both positive. In March and April, the GEVD type was Frechet, but the Gumbel distribution can also be considered since zero was within the 95% confidence interval.

Table 2. Parameter estimates of the GEVD and Gumbel for annual maximum rainfall series using the MLE method

Month	Distribution model	Location parameter		Scale parameter		Shape parameter	
		$\hat{\mu}$	Confidence Interval	$\hat{\sigma}$	Confidence Interval	$\hat{\xi}$	Confidence Interval
November	GEV	48.93	[41.29, 56.56]	22.55	[16.93, 28.18]	0.03	[-0.20, 0.27]
	Gumbel	49.34	[42.17, 56.51]	22.81	[17.41, 28.20]	-	-
December	GEV	91.14	[77.62, 104.67]	38.36	[28.17, 48.55]	-0.05	[-0.34, 0.24]
	Gumbel	90.09	[78.24, 101.94]	37.59	[28.66, 46.52]	-	-
January	GEV	94.59	[82.17, 107.01]	38.30	[29.97, 46.62]	-0.10	[-0.23, 0.04]
	Gumbel	92.95	[80.86, 105.05]	38.28	[30.02, 46.52]	-	-
February	GEV	75.49	[63.72, 87.25]	34.07	[24.21, 43.92]	0.28	[0.003, 0.56]
	Gumbel	81.04	[68.63, 93.44]	39.86	[29.89, 49.82]	-	-
March	GEV	64.17	[54.18, 74.15]	27.34	[19.16, 35.52]	0.21	[-0.13, 0.56]
	Gumbel	67.52	[57.98, 77.06]	30.43	[22.95, 37.90]	-	-
April	GEV	45.92	[38.76, 53.08]	21.10	[15.70, 26.51]	0.10	[-0.14, 0.34]
	Gumbel	47.05	[40.17, 53.93]	21.94	[16.67, 27.22]	-	-

Table 3. Kolmogorov-Smirnov, likelihood ratio, and Mann-Kendall tests for annual maximum rainfall series in Makassar City

Month	Number of observations	Distribution model	P-value K-S test	P-value LR test	P-value M-K test
November	43	GEV	0.99	0.78	0.15
		Gumbel	0.99		
December	43	GEV	0.55	0.75	0.28
		Gumbel	0.47		
January	43	GEV	0.76	0.21	0.74
		Gumbel	0.59		
February	43	GEV	0.95	0.01	0.58
		Gumbel	0.66		
March	43	GEV	0.81	0.17	0.72
		Gumbel	0.82		
April	43	GEV	0.98	0.40	0.27
		Gumbel	0.90		

The goodness of fit of the GEVD and its nested model, the Gumbel, was assessed using the K-S test with a significance level of 0.05. Table 3 indicates that both models were suitable for the annual maximum rainfall series in all rainy months since the p-values of the K-S test were greater than 0.05. Additionally, the M-K test results with a significance level of 0.05 showed that there was no upward or downward trend from year to year in all months of observation, indicating that the assumption of stationary was met.

The LR test was performed to determine the best-fit model between GEVD and Gumbel. The model with fewer parameters (in this case, Gumbel) was selected if the p-value of the LR test was greater than the significance level of 0.05. Based on Table 3, the Gumbel distribution was the best-fit model for the annual maximum rainfall series in November, December, January, March, and April, while the GEVD was the most appropriate model for February. These findings are in agreement with the outcomes of a study conducted by Park *et al*. [16], which demonstrated that the stationary GEVD and its nested model

(Gumbel) were effective in describing the maximum rainfall behavior during the summer rainy season in South Korea.

Figures 3 and 4 show diagnostic plots used to evaluate the accuracy of the model fit to the annual maximum rainfall series. The Q-Q plot and density plot reveal that each plotted observation point was approximately linear and that the empirical distribution curves follow the best-fit model distribution well, respectively. These findings suggest the validity of the best-fit model for each month.

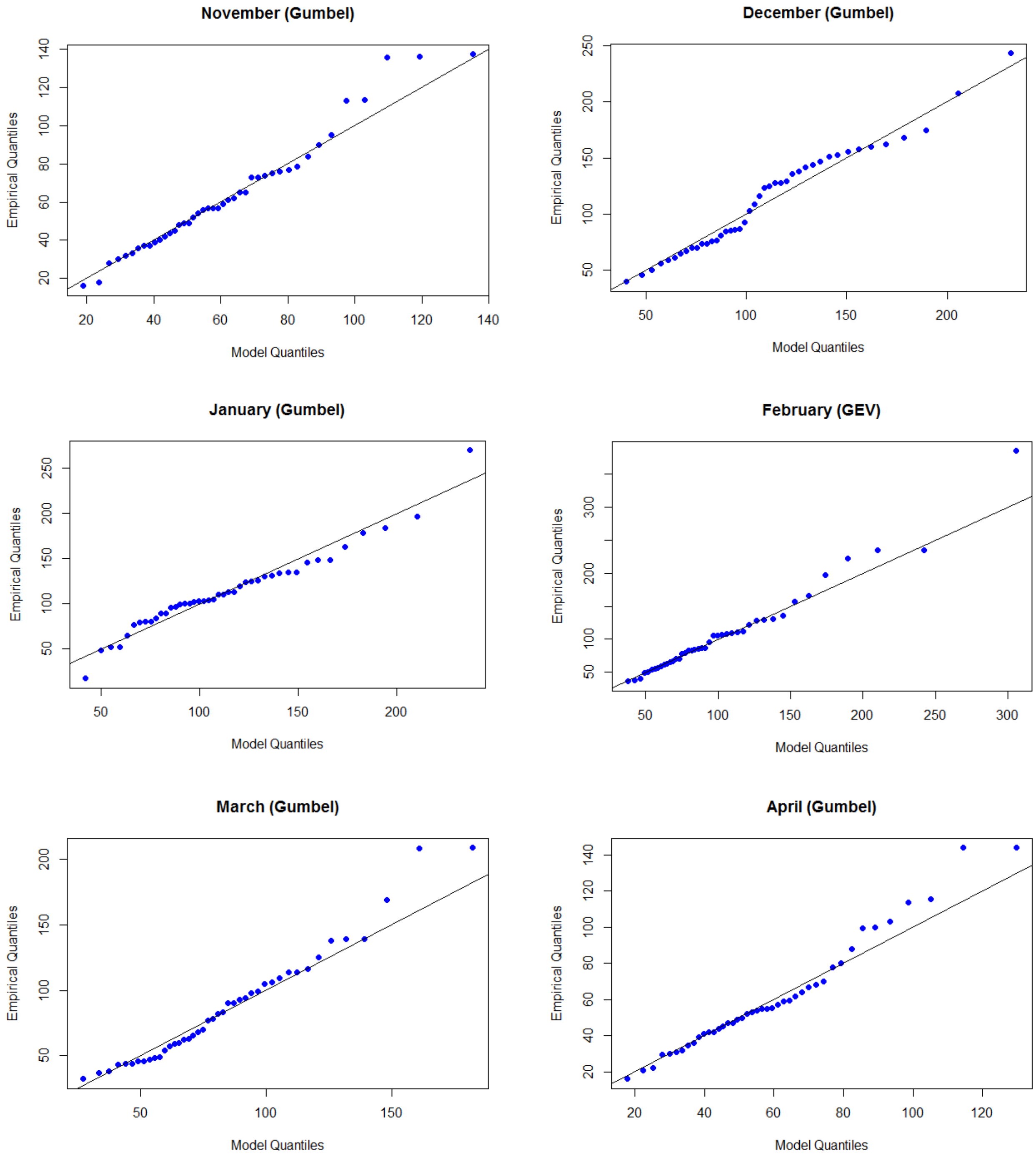

Figure 3. Q-Q plots of the best-fit model for the annual maximum rainfall series in the rainy season in Makassar City

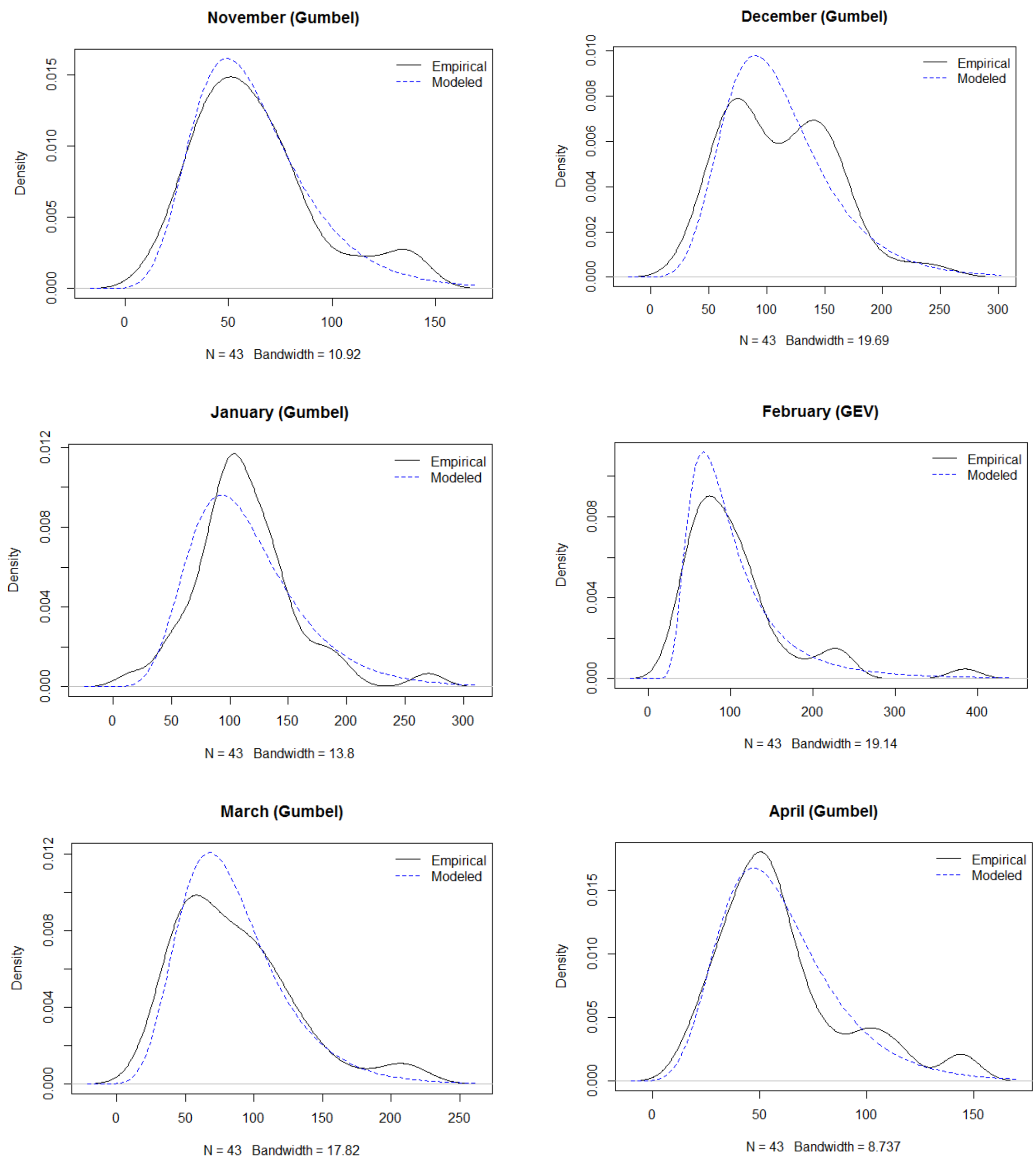

Figure 4. Empirical density vs best-fit model distribution of annual maximum rainfall in the rainy season in Makassar City

Return Levels and Probability of Exceedances

The return level of maximum rainfall is a crucial parameter that plays a significant role in designing and evaluating infrastructure and facilities that are sensitive to extreme weather events, such as flood control systems, dams, and levees. In order to estimate the return levels of annual maximum rainfall (mm/day) in Makassar City during the raining season for different return periods of 2, 3, 5, 10, and 20 years, the GEVD and Gumbel models were employed. The maximum rainfall return level estimates based on the GEVD and Gumbel, as well as return level plots along with 95% confidence intervals for the best model in each month of observation are presented in Table 4 and Figure 5, respectively.

Table 4 shows that the return levels in all months increased as the selected return period increased for both the GEVD and Gumbel models. Using the Gumbel distribution, the maximum rainfall in November is expected to exceed the level of 57.70 mm/day once every 2 years (known as a 2-year event) with a probability of occurrence of 0.5. Then, once every 3 years, the maximum rainfall in November above 69.93 mm/day is expected with a probability of 0.33. Finally, with a probability of 0.05, the maximum rainfall in November above 117.08 mm/day is expected once every 20 years. Based on the Gumbel distribution, the maximum rainfall is expected to exceed 103.87 mm/day in December and 106.98 mm/day in January once every two years, with a probability of occurrence of 0.5. For longer return periods, the maximum rainfall is expected to exceed 174.69 mm/day in December and 179.09 mm/day in January once every 10 years with a probability of exceedance of 0.1.

The maximum rainfall in February is expected to exceed 88.65 mm/day once every two years, with a probability of occurrence of 0.5 using the GEV distribution. Then, once every 5 years, the maximum rainfall in February is expected to exceed 139.22 mm/day with a probability of 0.2. Using the Gumbel distribution, the maximum rainfall is expected to exceed 94.99 mm/day in March and 66.86 mm/day in April once every 3 years. For longer return periods, the maximum rainfall will exceed 157.91 mm/day in March and 112.23 mm/day in April once every 20 years with a probability of exceedance of 0.05.

Table 4. Return levels of maximum rainfall (mm/day) in the rainy season by the probability distributions in Makassar City

Month	Distribution model	Return period				
		2-year	3-year	5-year	10-year	20-year
November	GEV	57.25	69.60	83.61	101.62	119.32
	Gumbel	57.70	69.93	83.55	100.66	117.08
December	GEV	105.08	124.99	146.57	172.76	196.98
	Gumbel	103.87	124.03	146.48	174.69	201.74
January	GEV	108.38	127.66	147.97	171.84	193.13
	Gumbel	106.98	127.50	150.36	179.09	206.64
February	GEV	88.65	110.55	139.22	182.90	234.53
	Gumbel	95.64	117.02	140.82	170.73	199.42
March	GEV	74.59	91.39	112.50	143.15	177.54
	Gumbel	78.67	94.99	113.17	136.00	157.91
April	GEV	53.80	65.84	80.02	99.04	118.65
	Gumbel	55.10	66.86	79.97	96.43	112.23

According to Table 4, January showed the highest return levels for the 2, 3, and 5-year return periods compared to other months. While for the 10-year and 20-year return periods, February showed the highest return levels. The high estimate of maximum rainfall in January and February is due to the peak of the rainy season occurring in these months. As a result, floods are frequent during this period. It was recorded that in January 2019, a flood disaster hit the city of Makassar and inundated 1658 houses, affecting 9328 residents [34]. Heavy rains with an intensity of 166.8 mm/day that occurred on February 13, 2023, caused flooding in most parts of Makassar [35]. The results of the 20-year return level predicted higher rainfall intensity in all rainy months, but with a low probability of occurrence. In contrast, the predicted rainfall intensity was lower but the probability of occurrence was higher for the 2-year return period results. This is consistent with the research findings of Alam *et al* [36].

The annual maximum rainfall series during all rainy months exhibited stationarity, indicating the absence of discernible long-term trends or patterns in the data. As a result, its statistical properties can be considered consistent over time. Consequently, the best-fit model can be utilized to predict the maximum rainfall for the upcoming year. As shown in Figure 6, the maximum rainfall in 2023 is predicted to exceed 57.70 mm/day in November, 103.87 mm/day in December, 106.98 mm/day in January, 88.65 mm/day in February, 78.67 mm/day in March, and 55.10 mm/day in April, with an occurrence probability of 0.5. The phenomenon of rainfall that occurs in the tropics is quite dynamic, with different places tending to have different rainfall intensities. This study used rainfall data from a single location, namely the Hasananuddin rain gauge station. According to Sunusi and Giarno [37], the maximum rainfall event at one location in Makassar City may not necessarily occur at other locations. As a result, the use of rainfall data from multiple rain gauge stations must be considered in order to improve the accuracy of maximum rainfall estimates.

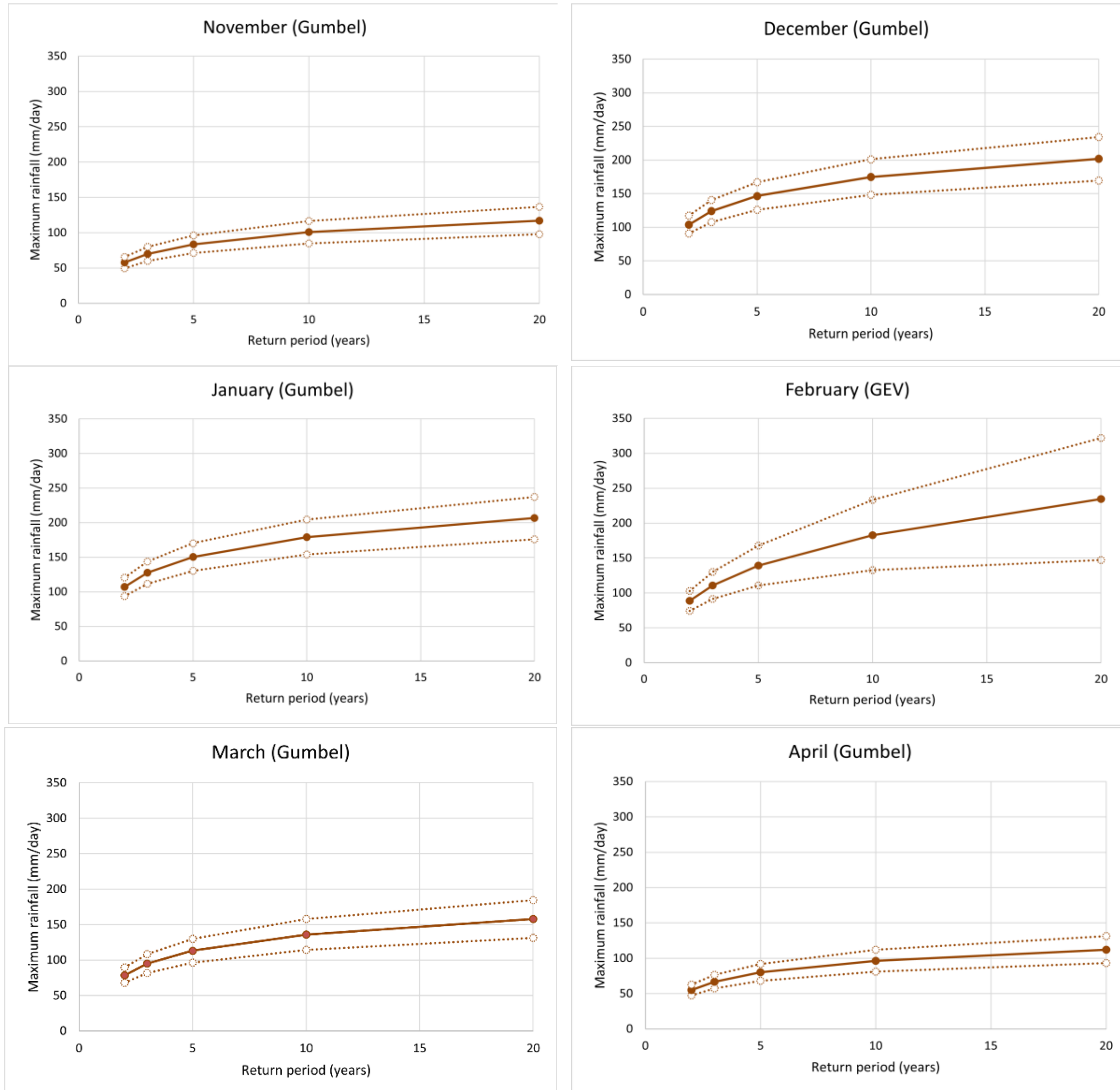

Figure 5. Return levels of annual maximum rainfall in Makassar City with a 95% confidence interval using the best-fit model

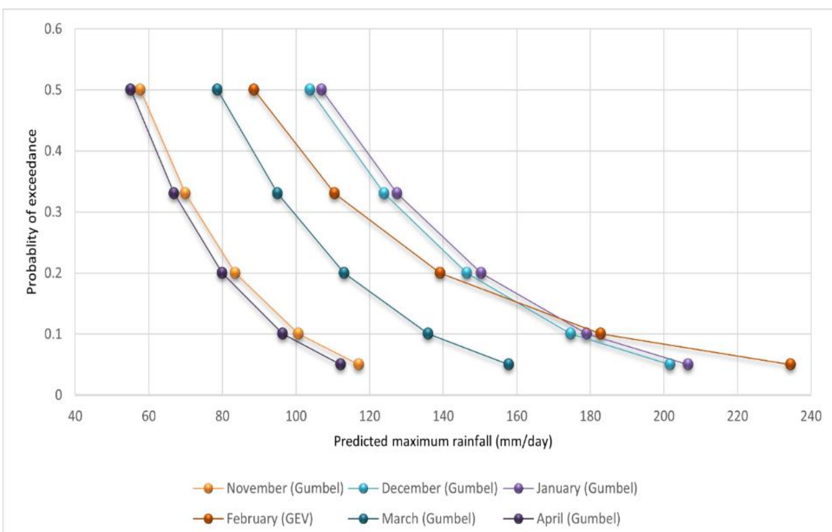

Figure 6. Predicted maximum rainfall in Makassar City in the next year (2023, 2024, and so on) with their probability of exceedances

Conclusions

This study used stochastic models based on probability distributions to analyze the patterns of annual maximum rainfall in Makassar City from 1980 to 2022 during the rainy season (November to April). The GEVD and its nested model (Gumbel) were considered, and the parameters of the GEVD and Gumbel were estimated using the MLE method. To validate the model, the K-S and LR tests were employed. The GEVD and Gumbel were found to be suitable for the data in all rainy months. According to the LR test results, the Gumbel distribution was the best-fit model for the distribution of the annual maximum rainfall series in November, December, January, March, and April, while the GEV distribution was appropriate for February. The authors also explored possible trends in the data using the Mann-Kendall test but found no evidence of trends, either upward or downward, in all months of observation.

Furthermore, the return levels of maximum rainfall for return periods of 2, 3, 5, 10, and 20 years in each month were estimated using quantiles from the best-fit distribution. January showed the highest return levels for the 2, 3, and 5-year return periods when compared to the other months, while February produced the highest return level estimates for the 10-year and 20-year return periods. Finally, the authors discovered that the longer the return period, the higher the estimated maximum rainfall return level, but with a low probability of that event occurring, and vice versa. Stochastic models of maximum rainfall using the GEVD and Gumbel distributions are useful tools for predicting the likelihood of maximum rainfall and can help inform decision-making in fields such as flood management. Further work needs to include non-rainy months to get a better insight into the maximum rainfall in each month in Makassar City. Moreover, the use of rainfall data from multiple rain gauge stations must be considered in order to improve the accuracy of maximum rainfall estimates.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgment

All authors are grateful to the Indonesian Meteorology, Climatology, and Geophysical Agency (BMKG) for supplying rainfall dataset. Besides, the authors would also like to thank the editor and reviewers for their contributions to the manuscript's improvement.

References

[1] De Luca, D. L., and Capparelli, G. (2022). Rainfall nowcasting model for early warning systems applied to a case over Central Italy. *Nat. Hazards*, *112*, 501-520. https://doi.org/10.1007/s11069-021-05191-w.

[2] Harmel, R. D., Richardson, C. W., and King, K. W. (2000). Hydrologic response of a small watershed model to generated precipitation. *Trans. ASAE*, *43*(6), 1483-1488. https://doi.org/10.13031/2013.3047.

[3] Trenberth, K. (2011). Changes in precipitation with climate change. *Clim. Res.*, *47*(1), 123-138. https://doi.org/10.3354/cr00953.

[4] Chapman, T. (1998). Stochastic modelling of daily rainfall: the impact of adjoining wet days on the distribution of rainfall amounts. *Environ. Model. Softw.*, *13*(3-4), 317-324. https://doi.org/10.1016/S1364-8152(98)00036-X.

[5] Du, H., Xia, J., Zeng, S., She, D., and Liu, J. (2014). Variations and statistical probability characteristic analysis of extreme precipitation events under climate change in Haihe River Basin, China. *Hydrol. Process.*, *28*(3), 913-925. https://doi.org/10.1002/hyp.9606.

[6] Li, Z., Brissette, F., and Chen, J. (2014). Assessing the applicability of six precipitation probability distribution models on the Loess Plateau of China. *Int. J. Climatol.*, *34*(2), 462-471. https://doi.org/10.1002/joc.3699.

[7] WMO. (2009). *Guide to hydrological practices, volume II: Management of water resources and application of hydrological practices.* 6th edition. Geneva, Switzerland: World Meteorological Organization.

[8] Ball, J., *et al.* (2016). *Australian rainfall and runoff: A guide to flood estimation, Commonwealth of Australia*.

[9] Salinas, J. L., Castellarin, A., Viglione, A., Kohnová, S., and Kjeldsen, T. R. (2014). Regional parent flood frequency distributions in Europe – Part 1: Is the GEV model suitable as a pan-European parent? *Hydrol. Earth Syst. Sci.*, *18*(11), 4381-4389. https://doi.org/10.5194/hess-18-4381-2014.

[10] Sanusi, W., Mulbar, U., Jaya, H., Purnamawati, and Side, S. (2017). Modeling of rainfall characteristics for monitoring of the extreme rainfall event in Makassar City. *Am. J. Appl. Sci.*, *14*(4), 456-461. https://doi.org/10.3844/ajassp.2017.456.461.

[11] Sanusi, W., Chaerunnisa, S., Annas, S., Side, S., and Abdy, M. (2022). Estimated parameters of rain flow distribution using L-Moment method in South Sulawesi, Indonesia. *J. Appl. Math. Comput.*, *6*(1), 30-40. https://doi.org/10.26855/jamc.2022.03.006.

[12] Susanti, W., Adnan, A., Yendra, R., and Muhaijir, M. N. (2018). The analysis of extreme rainfall events in Pekanbaru city using three-parameter generalized extreme value and generalized Pareto distribution. *Appl. Math. Sci.*, *12*(2), 69-80. https://doi.org/10.12988/ams.2018.712355.

[13] Fisher, R. A., and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Math. Proc. Camb. Philos. Soc.*, *24*(2), 180-190. https://doi.org/10.1017/S0305004100015681.

[14] Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Q. J. R. Meteorol. Soc.*, *81*(348), 158-171. https://doi.org/10.1002/qj.49708134804.

[15] Lana, X., Martínez, M. D., Burgueño, A., Serra, C., Martín-Vide, J., and Gómez, L. (2006). Distributions of long dry spells in the iberian peninsula, years 1951–1990. *Int. J. Climatol.*, *26*(14), 1999-2021. https://doi.org/10.1002/joc.1354.

[16] Park, J.-S., Kang, H.-S., Lee, Y. S., and Kim, M.-K. (2011). Changes in the extreme daily rainfall in South Korea. *Int. J. Climatol.*, *31*(15), 2290-2299. https://doi.org/10.1002/joc.2236.

[17] Nadarajah, S., and Choi, D. (2007). Maximum daily rainfall in South Korea. *J. Earth Syst. Sci.*, *116*(4), 311-320. https://doi.org/10.1007/s12040-007-0028-0.

[18] Nashwan, M. S., Ismail, T., and Ahmed, K. (2019). Non-stationary analysis of extreme rainfall in Peninsular Malaysia. *J. Sustain. Sci. Manag.*, *14*(3), 17-34.

[19] Kumar, V., Shanu, and Jahangeer. (2017). Statistical distribution of rainfall in Uttarakhand, India. *Appl. Water Sci.*, *7*(8), 4765-4776. https://doi.org/10.1007/s13201-017-0586-5.

[20] Babar, S., and Ramesh, H. (2014). Analysis of extreme rainfall events over Nethravathi basin. *ISH J. Hydraul. Eng.*, *20*(2), 212-221. https://doi.org/10.1080/09715010.2013.872353.

[21] Abbas, K., Alamgir, Khan, S. A., Khan, D. M., Ali, A., and Khalil, U. (2012). Modeling the distribution of annual maximum rainfall in Pakistan. *Eur. J. Sci. Res.*, *79*(3), 418-429.

[22] Aurangzeb, A., Abbas, K., Iqbal, A., Altaf, M., Nadeem, M. T., and Shahzad, F. (2022). Statistical analysis of rainfall trends in Balochistan and Sindh. *Tech. J. Univ. Eng. Technol. UET Taxila*, *27*(1), 8-21.

[23] Boudrissa, N., Cheraitia, H., and Halimi, L. (2017). Modelling maximum daily yearly rainfall in northern Algeria using generalized extreme value distributions from 1936 to 2009. *Meteorol. Appl.*, *24*(1), 114-119. https://doi.org/10.1002/met.1610.

[24] Min, J. L. J., and Halim, S. A. (2020). Rainfall modelling using generalized extreme value distribution with cyclic covariate. *Math. Stat.*, *8*(6), 762-772. https://doi.org/10.13189/ms.2020.080617.

[25] Malino, C. R., Arsyad, M., and Palloan, P. (2021). Analysis of rainfall and air temperature parameters in Makassar City related to climate change phenomena. *J. Sains Dan Pendidik. Fis.*, *17*(2), 139-145. https://doi.org/10.35580/jspf.v17i2.22167.

[26] Coles, S. (2001). *An introduction to statistical modeling of extreme values*. London, New York: Springer.

[27] Coles, S. G., and Dixon, M. J. (1999). Likelihood-based inference for extreme value models. *Extremes*, *2*(1), 5-23.

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

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

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

[31] Wu, H., Soh, L.-K., Samal, A., and Chen, X.-H. (2008). Trend analysis of streamflow drought events in Nebraska. *Water Resour. Manag.*, *22*(2), 145-164. https://doi.org/10.1007/s11269-006-9148-6.

[32] Cooley, D. (2013). Return periods and return levels under climate change in *Extremes in a Changing Climate.* A. AghaKouchak, D. Easterling, K. Hsu, S. Schubert, and S. Sorooshian. Eds. Dordrecht: Springer Netherlands. pp. 97–114. https://doi.org/10.1007/978-94-007-4479-0_4.

[33] Caires, S. (2016). A comparative simulation study of the annual maxima and the peaks-over-threshold methods. *J. Offshore Mech. Arct. Eng.*, *138*(5). https://doi.org/10.1115/1.4033563.

[34] Thoban, M. I., and Hizbaron, D. R. (2020). Urban esilience to floods in parts of Makassar, Indonesia. *E3S Web Conf.*, *200*. https://doi.org/10.1051/e3sconf/202020001007.

[35] VOI. (2023). *BMKG: The rob flood in Makassar City was triggered by sea waves rising to a height of 4 meters.* Retrieved from https://voi.id/en/news/253896.

[36] Alam, M., Emura, K., Farnham, C., and Yuan, J. (2018). Best-Fit probability distributions and return periods for maximum monthly rainfall in Bangladesh. *Climate*, *6*(1), 1-16. https://doi.org/10.3390/cli6010009.

[37] Sunusi, N., and 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 Environ. Sci*, *9*(3), 260-276. https://doi.org/10.3934/environsci.2022018.