Mohamed A. Ismail

Statistical modelling of extreme rainfall in Khartoum, Sudan

Mohamed A. Ismail, PhD, Assistant Professor of Applied Statistics, Institute of Public Administration, Riyadh, Saudi Arabia
Email: mohamedismail4183208@gmail.com

The objective of this study was to analyse the extreme rainfall data of Khartoum, the capital city of Sudan. The study's data consist of monthly rainfall data obtained from the World Bank Climate Change Knowledge Portal for the years 1901 to 2021. The generalized extreme value (GEV) distribution was used to model the annual maximum rainfall using the block maxima method. The model parameters were estimated using the maximum likelihood estimation (MLE) method. We found evidence to suggest that the Gumbel model better fits the maximum rainfall data based on the likelihood ratio test of the GEV and Gumbel models. The return levels of maximum monthly rainfall in Khartoum and their 95 percent confidence intervals were computed for selected return periods: 2, 10, 20, 25, 30, 50, 100, 200, and 500 years.

Keywords: rainfall, generalized extreme value, Gumbel distribution, block maxima, return level, Khartoum, Sudan

Heavy rainfall and floods have wide-range effects, particularly in developing countries *(Dewan, 2015)*. The main negative effects of extreme rainfall and flood events are death, disease, damage to the environment, destruction of basic infrastructure, population displacement, contamination of drinking water, loss of crops and livestock, and landslides *(Pregnolato et al., 2017; Walsh et al., 1994)*.

Sudan is prone to different types of natural hazards, including frequent occurrences of floods and droughts. In Sudan, there are two kinds of floods: run-off floods and flash floods brought on by extremely heavy rains, as well as floods brought on by the Nile and its tributaries overflowing. During the past decades, there have been torrential rains causing devastating floods in 1962–1965, 1978–1979, 1988, 1994, 1998, 1999, 2003, 2006–2007, 2009 and 2020 in different parts of the country, including Khartoum (*Osman–Sevinc, 2019; El Tohami, 2019; Siddig–Ahmed, 2016; NASA Earth Observatory, 2022)*. Among the past floods,

the 1946 and 1988 floods were the worst to hit Sudan. Likewise, Sudan has suffered from low rainfall, causing extreme droughts in 1906, 1984–1985, 1989, 1990, 1997, 2000, and 2011, which led to severe famines in 1906, 1984 and 1985 and food shortages in other years *(Hamid–Eltayeb, 2022)*.

More than eight million people in Sudan were affected by heavy rains, and rain-induced floods occurred during the period 1983–2020 *(World Bank Climate Change Knowledge Portal, 2022)*, which shows high levels of vulnerability of the population to natural hazards. The intensity and frequency of rainfall, poor infrastructure, poor government planning, and mismanagement of disasters play a major role in the devastating floods that occur in Sudan. It is thus of paramount importance to analyse extreme rainfall events to mitigate the effect of recurrent rainfall-induced floods in Khartoum city, the most populous city in the country. The results of such analysis may help policy-makers, executives, climatologists, urban planners, and insurance companies, among others, understand extreme rainfall patterns for effective decision-making.

Hence, the aim of this study is to provide a holistic analysis of extreme rainfall in Khartoum, the capital city of Sudan. Specifically, we aim to build the best probability distribution to predict extreme rainfall and to determine extreme rainfall return levels for given return periods.

1. The study data

Secondary data comprising records of monthly rainfall in Khartoum measured in millimetres were obtained from the Climate Change Knowledge Portal *(World Bank Group, Climate Change Knowledge Portal, 2022)*. The rainfall data for this study covered the period from 1901 to 2021.

2. Rainfall patterns in Khartoum

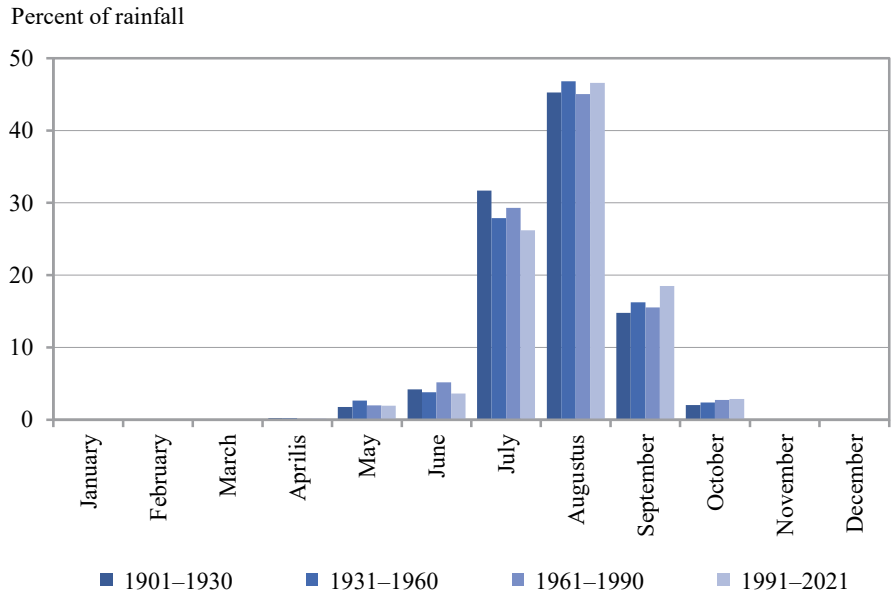
Greater Khartoum is the capital and the largest city of Sudan; it is sited approximately in the centre of the country, located between latitudes 15 to 16 °N and longitudes 31.5 to 34. Greater Khartoum consists of three towns: Khartoum at the confluence of the White and Blue Niles, Khartoum North (Bahri), which lies north of the confluence and east of the main Nile, and Omdurman, located west of the Nile River *(Zerboni et al., 2021)*. The three towns host

approximately 8.7 million people in 2021, accounting for 19% of the country's total population *(Bank of Sudan, 2022)*.

Khartoum has a hot arid desert climate *(Elagib–Mansell, 2000)*, with a rainy season that lasts from May to October, and traces of rain occurring before and after this period (Figure 1). Moreover, the figure shows that the monthly distribution of rainfall has slightly changed in the past three decades. While the rainy season has not shifted in its start and end, the total rainfall fell in July and August and decreased from 76.9 percent during 1901–1930 to 72.8 percent during 1991–2021, whereas the percent of rainfall in September–October increased from 16.8 percent to 21.4 percent during the 1901–1930 and 1991–2021 periods, respectively.

Figure 1

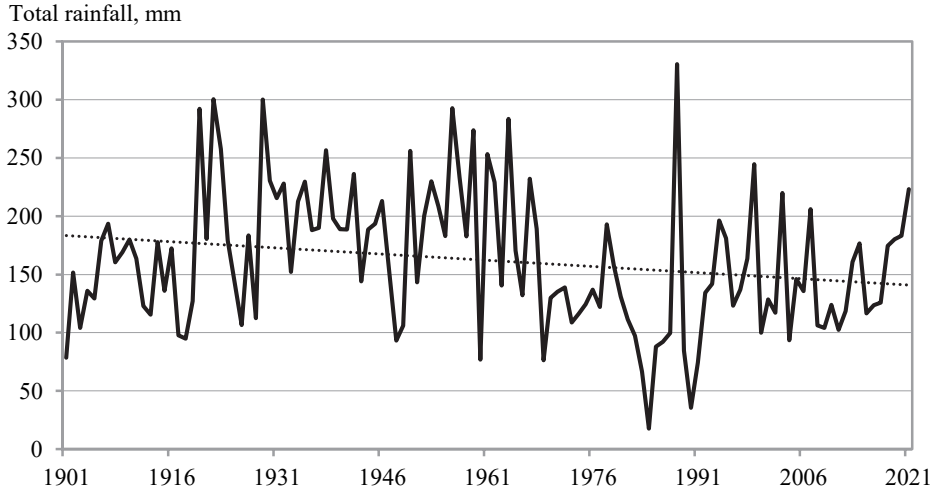
Percent distribution of monthly rainfall in Khartoum



The short rainy season in Khartoum is characterized by high temporal variations with a slight decreasing trend (Figure 2). The lowest total rainfall occurred in 1984, with an unprecedented level of 17.7 mm, which is lower than the long-term mean level by 89%. In 1988, Khartoum received the highest rainfall level of more than 300 mm, resulting in its worst flooding. Moreover, the mean annual rainfall of 165 mm during the study period is associated with a coefficient of variation of 36%, which shows high interannual variability (Table 1).

Figure 2

Annual total rainfall in Khartoum



The descriptive statistics shown in Table 1 also explain the monthly rainfall pattern, which reveals that the maximum amount of rainfall occurs in monsoon months from April to October in the study area. The values of the monthly coefficient of variation are remarkably high, particularly during March and November. This pattern shows that there was high variability in the rainfall within months over the years.

Table 1

Descriptive statistics of monthly rainfall (mm) in Khartoum, 1901–2021

Month	Minimum	Maximum	Range	Mean	Std. Deviation	Coefficient of Variation
January	0.0	0.0	0.0	0.0	0.0	–
February	0.0	0.0	0.0	0.0	0.0	–
March	0.0	0.1	0.1	0.0	0.0	343.6
April	0.0	2.3	2.3	0.3	0.4	121.0
May	0.0	14.4	14.4	3.5	3.2	91.9
June	0.0	30.1	30.1	6.7	5.9	86.8
July	1.4	143.7	142.3	46.6	29.6	63.5
August	1.3	194.4	193.1	74.5	38.1	51.1
September	2.2	70.7	68.4	26.3	14.6	55.4
October	0.0	12.7	12.7	4.1	3.3	80.6
November	0.0	0.1	0.1	0.0	0.0	196.4
December	0.0	0.0	0.0	0.0	0.0	–
Total	20.7	333.4	312.7	164.6	59.4	36.1

3. Methodology

3.1 Extreme value theory

Extreme value models have been widely used to analyse the behaviour of extreme events in different fields, including climate extremes. The main objective of these models is to estimate the uncertainty of extreme levels of a process *(Coles, 2001)*. Broadly, classical extreme value models can be classified into generalized extreme value (GEV) models for fitting the block maxima or block minima (BM) method and generalized Pareto (GP) models for fitting high threshold exceedances using the peak over threshold (POT) method *(Carmona, 2013; Coles, 2001; Gilleland–Katz, 2016; de Haan–Ferreira, 2006)*. The GEV and GP models have long been used to model extreme climate variables in different countries and regions *(Singirankabo–Iyamuremye, 2022; Nadarajah–Choi, 2007)*.

In the BM method, the study period is divided into a sequence of nonoverlapping periods of equal size, and then the maximum (minimum) value is extracted from each subperiod (block) to form the series of extremal observations. Thus, the number of maximum observations depends on the block size; as the block size increases, the number of extremal observations decreases.

3.2 Generalized extreme value distributions

Let $X_1, X_2, \ldots X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables from a distribution function F ($F(x) = \Pr(X \leq x)$). Suppose $M_n = \max(X_1, X_2, \ldots X_n)$ is the maxima of n observations and the exact distribution of M_n is $F^n(x)$. Moreover, assume that there exist sequences of constants $b_n > 0$ and a_n such that:

$$\Pr\left[\frac{M_n - a_n}{b_n} \leq x\right] = F^n\left(b_n + a_n\right) \to G\left(x\right)$$

where G is a nondegenerate distribution function. As n approaches infinity ($n \to \infty$), G belongs to the Gumbel, Frechet and Weibull families. The GEV distribution is a parametrization of these three distributions into a single-family model with the following cumulative function *(Coles, 2001)*:

$$GEV\ (x, \xi, \sigma, \mu) = \begin{cases} \exp\left(-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right), & \xi \neq 0 \\ \exp\left(-\exp\left(\frac{x-\mu}{\sigma}\right)\right), & \xi = 0 \end{cases}, \quad (1)$$

where x are the maximum (minimum) observations of the blocks, μ ($-\infty < \mu < \infty$) is a location parameter, σ ($\sigma > 0$) is a scale parameter, and ξ ($-\infty < \xi < \infty$) is a shape parameter or the tail index. The GEV comprises three distributions depending on the value of the shape parameter ξ; for $\xi = 0$, the distribution is Gumbel, and for $\xi > 0$ and $\xi < 0$, the distributions are Weibull and Fréchet, respectively.

Assuming the independence of m extreme observations, the maximum likelihood estimation of the unknown parameters of the GEV can be determined by maximizing the likelihood function, given by the following form:

$$\ell(\mu,\sigma,\xi) = -m\,\log\sigma - \left(1+\tfrac{1}{\xi}\right)\sum_{i=1}^{m}\log\left[1+\xi\left(\tfrac{x_i-\mu}{\sigma}\right)\right] - \sum_{i=1}^{m}\left[1+\xi\left(\tfrac{x_i-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right] \quad (2)$$

where $\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right] > 0$ for $i = 1,2,\ldots,m$.

3.3 Return periods

One of the main objectives of modelling extreme events is to estimate extreme quantiles of the annual maxima (minima), as they give an estimate of the level the process is expected to exceed once, on average, in a given number of years.

The return level q_p associated with the 1/p-year return period for the GEV distribution is given as follows:

$$q_p = \begin{cases} \mu - \frac{\sigma}{\xi}\left[1-\left\{-\log\left(1-p\right)\right\}^{-\xi}\right], & \xi \neq 0 \\ \mu - \sigma\log\left(-\log\left(1-p\right)\right), & \xi = 0 \end{cases} \quad (3)$$

q_p is the level exceeded by the annual maxima in any year with probability p.

4. Comparison of models

The likelihood-ratio (LR) test has been widely used to compare nested models by determining whether additional parameters significantly improve model fit. The LR test is applied to compare the full model with a reduced model. Assume that ℓ_R is the log-likelihood for the reduced model M_R and ℓ_F the full model M_F. Then, the LR statistic is given as follows:

$$\lambda = 2\left(\ell_1 - \ell_0\right) \tag{4}$$

where λ is approximately distributed as a chi-squared distribution ($\lambda \sim \chi_d^2$), where d is the difference in the number of parameters between the two models. We reject M_R if $\lambda > C$, where C is the $(1 - \alpha)$ quantile of the χ_d^2 distribution.

5. Model diagnostics

The most commonly used model diagnostic plots to assess the goodness of fit of the fitted model are quantile plots, probability plots, return level plots, and density function plots. The probability plot (PP) is a comparison of the empirical and fitted distribution functions. The empirical distribution function is given by:

$$\tilde{G}\left(x_{(i)}\right) = \frac{i}{m+1}$$

where $x_{(i)}$ is the ordered maxima data $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(m)}$. The corresponding fitted distribution is given by:

$$\hat{G}\left(x_{(i)}\right) = \exp\left\{-\left[1+\hat{\xi}\left(\frac{x_{(i)}-\hat{\mu}}{\hat{\sigma}}\right)\right]^{-1/\hat{\xi}}\right\}.$$

For the best GEV fit model, the correlation between the empirical and fitted distribution points is positive and very high, and most points lie close to the unit diagonal.

The quantile plot (QP) compares the ranks of the fitted distribution with those of the data. The QP is a plot between $\hat{G}^{-1}\left(\frac{i}{m+1}\right)$ and $x_{(1)}$ against $\hat{G}^{-1}\left(\frac{i}{m+1}\right) = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}\left[1-\left\{-\log\left(\frac{i}{m+1}\right)\right\}^{-\hat{\xi}}\right]$. For the best-fit model, most points of the two series lie close to the unit diagonal.

The return level plot is the plot of $\hat{q}_p$, the maximum likelihood estimate of q_p given by (3), against $y_p = -\log(1-p)$ on a logarithmic scale. The plot is linear when the shape parameter is equal to zero, convex with an asymptotic limit of $p \to 0$ at $\mu - \sigma/\xi$ if the shape parameter is less than zero, and concave with no finite bound if the shape parameter is greater than zero. For an adequate GEV model, the model-based curve and empirical estimates should agree.

The fourth diagnostic plot is the density function plot, which compares the probability density function of a fitted model with a histogram of the data.

6. Results

6.1 Stationarity of annual rainfall maxima

The data for GEV models represent the maxima (minima) of blocks within a specified period. In this study, the data are blocked into yearly lengths, resulting in 121 annual rainfall maxima (1901–2021).

Checking for stationarity, Figure 3 shows that the pattern of the maximum rainfall over the study period (1901–2021) has not changed. Moreover, the respective p values of the Augmented Dickey Fuller (ADF) and Phillips Perron (PP) tests are less than 0.05 (Table 2). Thus, we can conclude that there is stationarity in the annual maximum rainfall in Khartoum.

Figure 3

Scatter plot of annual rainfall maxima

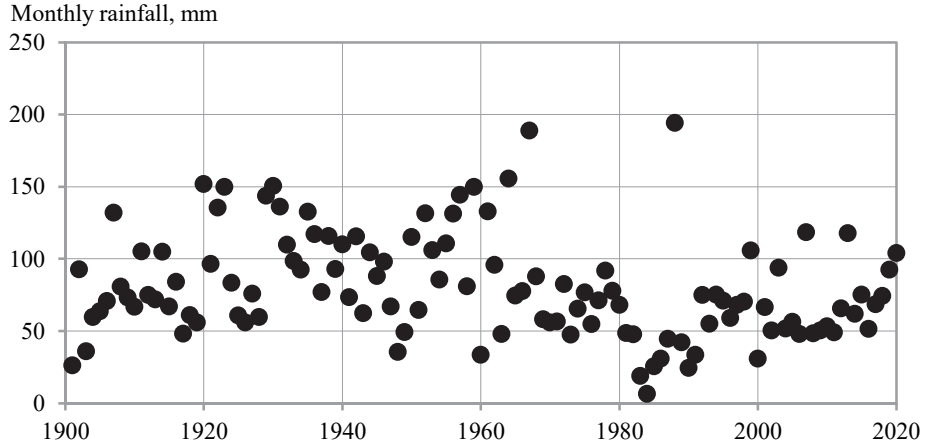


Table 2

Unit root test for maximum rainfall in Khartoum (n=121)

Denomination	t-statistic	Critical value ($\alpha = 0.05$)	P-value
ADF test			
Constant	–8.9	–2.9	0.000
Constant and linear trend	–9.4	–3.4	0.000
PP test			
Constant	–9.1	–2.9	0.000
Constant and linear trend	–9.4	–3.4	0.000

6.2 Results of GEV models

Based on the results of stationarity tests, three-parameter GEV and Gumbel models have been developed using the maximum likelihood estimation (MLE) method (Table 3). First, the classical three-parameter GEV model was fitted to the rainfall data. The results presented in Table 3 show that the estimated shape parameter in the fitted model is negative $\left(\hat{\xi} = -0.067\right)$, which is inconsistent with the a priori positive sign ($\xi > 0$) for rainfall data. *Nadarajah and Choi (2007)* noted that for rainfall data, the shape parameter usually takes values greater than zero $\left(\hat{\xi} > 0\right)$; however, the Gumbel distribution fits the data better than the three-parameter GEV distribution. Moreover, the 95% confidence interval of the shape parameter obtained from the profile likelihood includes 0 $\left(0 \in 95\%\ \mathrm{CI}\right)$ (Figure 4), implying that there is not enough evidence to reject the null hypothesis (H_0: $\xi = 0$).

Moreover, the goodness-of-fit was evaluated by Anderson–Darling and Kolmogorov−Smirnov tests. Since the p values of the respective two tests are greater than 0.05, we conclude that both the GEV and Gumbel distributions fit the maximum rainfall data (Table 4).

Therefore, a Gumbel model was fitted, and then the LR test was performed to determine which of the two models fit the data better. The null and alternative hypotheses of the LR test are as follows:

H_0: Gumbel model ($\xi = 0$) vs $H_{1:}$: GEV model ($\xi \neq 0$).

Since the p value of the likelihood-ratio test (0.308) is greater than 0.05, we conclude that there is not enough evidence to reject the null hypothesis. Hence, the Gumbel model is suitable for the annual rainfall maxima data in Khartoum.

Table 3

GEV and Gumbel model parameter estimates

Parameter	GEV model	Gumbel model
Location μ	65.32 (59.37, 71.26)	64.25 (58.73, 69.78)
Scale σ	29.89 (25.66, 34.11)	29.40 (25.39, 33.41)
Shape ξ	–0.067 (–0.19, 0.057)	–
Negative Log-Likelihood	597.521	598.0405
AIC	1201.041	1200.081
BIC	1209.429	1205.673

Note: Numbers in parentheses are 95% upper and lower confidence intervals.

Figure 4

95% profile likelihood confidence interval (dashed vertical lines) for the shape parameter of the GEV model to annual maxima rainfall

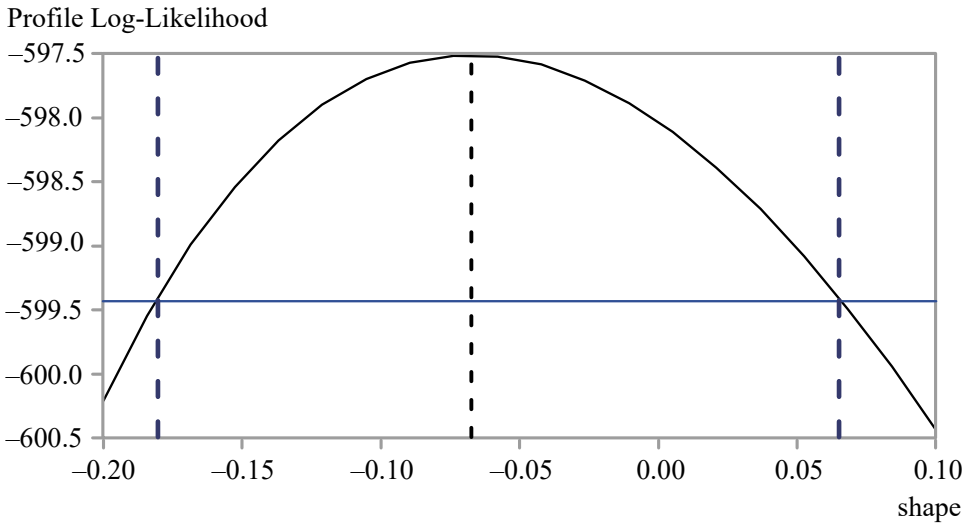


Table 4

Anderson–Darling and Kolmogorov−Smirnov tests to determine whether the annual maxima of monthly rainfall follow GEV and Gumbel distributions

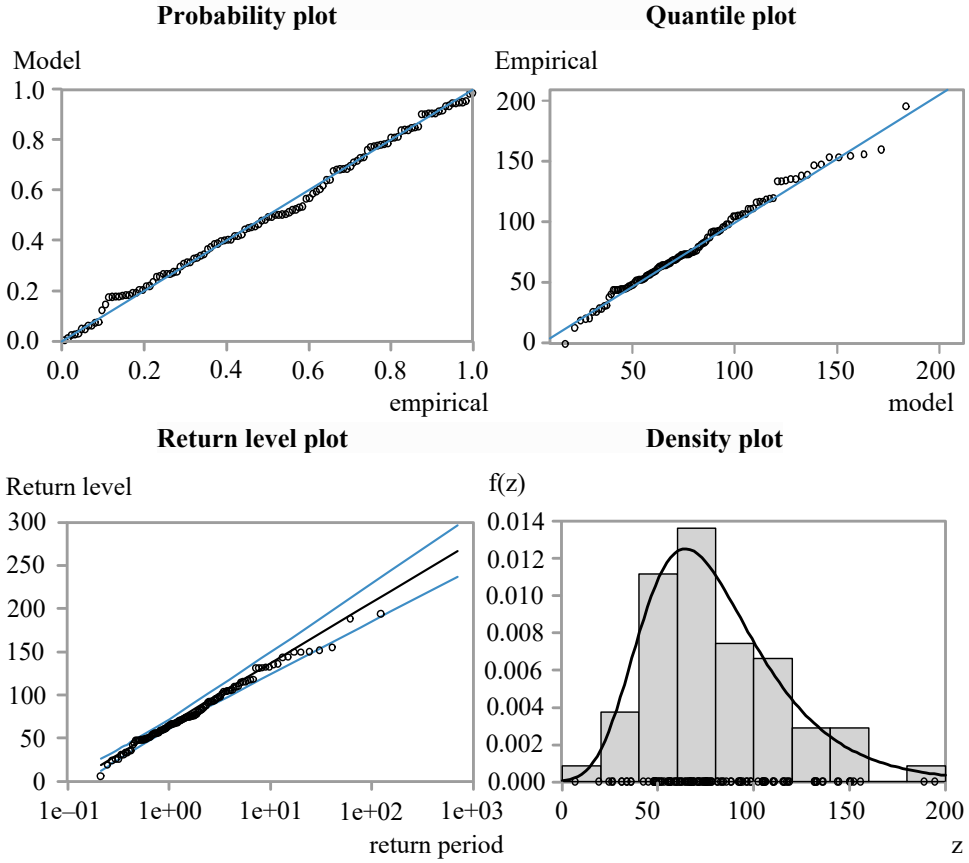
Model	Anderson–Darling test		Kolmogorov−Smirnov test	
	test statistic	p-value	test statistic	p-value
GEV	0.3543	0.457	0.065	0.688
Gumbel	0.3079	0.556	0.064	0.704

7. Model diagnostics

For model diagnostics, we used four plots: probability, quantile, return-level and density function plots. For both probability and quantile plots, Figure 5 shows that most of the data points fall roughly on a straight line. The return level plot shows that almost all the points lie within the 95% confidence interval. Moreover, the density plot shows that there is a good match between the empirical and fitted Gumbel distributions. Therefore, the diagnostic plots indicate that the Gumbel distribution provides a very good fit to the annual maximum rainfall in Khartoum.

Figure 5

Diagnostics of the Gumbel model fitted to the Khartoum maximum rainfall. From upper left to lower right: probability, quantile, return level, and histogram with fitted Gumbel density



8. Return levels

Table 5 shows the estimated maximum monthly rainfall return levels (mm) for the return periods of 2, 10, 20, 25, 30, 50, 100, 200, and 500 (in years) along with their respective 95% confidence intervals and exceedance probabilities. Most of the maximum rainfall events from 1901 to 2021 were found to fall within a 1–10 year return period (Figure 6), indicating that extremely high rainfall events (return

periods of 50–100) rarely occurred in this period. The results also show that the increase in return periods leads to a corresponding increase in return levels and lower exceedance probabilities.

For example, a maximum monthly rainfall of 179.0±18.2 mm is predicted to occur in Khartoum once every 50 years during the rainy season (June–October). A high rainfall level of 246.9±27.2 mm is estimated to occur in a month in the rainy season once every 500 years, which is very unlikely to be experienced in the city. Other return levels are interpreted the same way.

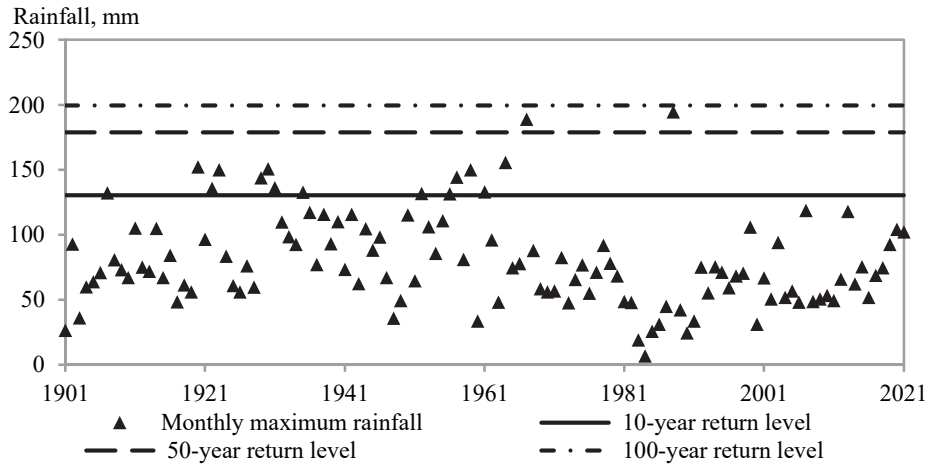
Table 5

Estimated return levels and 95% confidence intervals for several return periods for the Gumbel model fitted to the rainfall maxima in Khartoum

Time	Rainfall returns level (mm/month)			Exceedance probability
	point estimate	95% lower CI	95% upper CI	
2-year	75.0	68.9	81.2	0.500
5-year	108.3	99.0	117.7	0.200
10-year	130.4	118.4	142.4	0.100
20-year	151.6	136.9	166.2	0.050
25-year	158.3	142.8	173.8	0.040
30-year	163.7	147.6	179.9	0.033
50-year	179.0	160.8	197.1	0.020
100-year	199.5	178.6	220.4	0.010
200-year	219.9	196.4	243.5	0.005
500-year	246.9	219.8	274.1	0.002

Figure 6

Annual rainfall maxima in Khartoum



9. Conclusion

In this study, we modelled the maximum monthly rainfall in Khartoum for the period 1901–2021. The generalized extreme value (GEV) distribution was used to model the annual rainfall maxima using the block maxima method. First, the classical three-parameter GEV model was fitted to the rainfall data. The results showed that the estimated shape parameter in the fitted model is negative $\left(\hat{\xi} = -0.067\right)$, which is inconsistent with the a priori positive sign ($\xi > 0$) for rainfall data. Therefore, a Gumbel model was fitted, and then the LR test was performed to determine which of the two models fit the data better. The results of the LR test indicate that the Gumbel model fits the rainfall maxima data in Khartoum better than the three-parameter GEV model. Model diagnostics, which include probability, quantile, return-level, and density function plots, show that the Gumbel model fits the data well. Return level estimates, which are the return level expected to be exceeded in a certain period of time T in years, were computed for Khartoum rainfall. A high rainfall level of 246.9±27.2 mm is estimated to occur in a month in the rainy season once every 500 years, which is very unlikely to be experienced in Khartoum.

The results of this study will help decision-makers with the extreme rainfall levels and their respective return periods considered to enable them to make suitable decisions to protect the population and the basic infrastructures of the city. Thus, this study will be useful in developing rain-induced flood risks for early warning, management, preparedness, response and mitigation. However, future studies can model and predict extreme rainfall in Sudan using daily rainfall data rather than monthly data for Khartoum and the other cities of the country.

References

Bank of Sudan (2022): 2021 Annual Report. https://cbos.gov.sd/

Carmona, R. (2013): Statistical Analysis of Financial Data in R. Springer-Verlag New York Inc.

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

de Haan, L. – Ferreira, A. (2006): Extreme Value Theory: An Introduction. Springer, New York.

Dewan, T. H. (2015): Societal impacts and vulnerability to floods in Bangladesh and Nepal. *Weather and Climate Extremes*. Vol. 7. pp. 36–42. https://doi.org/10.1016/j.wace.2014.11.001

El Tohami, A. A. (2019): Seasonal flood in Sudan and its environmental, health and socioeconomic impacts on the livelihoods: Review of a five-year period. *American Journal of Biomedical Science & Research*. Vol. 5. No. 4. pp. 296–301.

Elagib, A. N. – Mansell, M. G. (2000): Recent trends and anomalies in mean seasonal and annual temperatures over Sudan. *Journal of Arid Environments*. Vol. 45. pp. 3263–3288.

Gilleland, E. – Katz, R. W. (2016): extRemes. 2.0: An Extreme Value Analysis Package in R. *Journal of Statistical Software.* Vol. 72. No. 8. pp. 1–39.

Hamid, A. A. – Eltayeb, Y. H. (2022): Space borne technology for drought monitoring in Sudan. https://www.un-spider.org/sites/default/files/Hamid.pdf (downloaded: September 2022)

Nadarajah, S. – Choi, D. (2007): Maximum daily rainfall in South Korea. *Journal of Earth System Science*. Vol. 116. pp. 311–320.

NASA Earth Observatory (2022): Record flooding in Sudan. https://earthobservatory.nasa.gov/images/147288/record-flooding-in-sudan (downloaded: September 2022)

Osman, M. M. – Sevinc, H. (2019): Adaptation of Climate-Responsive Building Design Strategies and Resilience to Climate Change in the Hot/Arid Region of Khartoum, Sudan. *Sustainable Cities and Society*. Vol. 47. No. 101429. https://doi.org/10.1016/j.scs.2019.101429

Pregnolato, M. – Ford, A. – Wilkinson, S. – Dawson, R. (2017): The impact of flooding on road transport: A depth-disruption function. *Transportation Research Part D: Transport and Environment.* Vol. 55. pp. 67–81. https://doi.org/10.1016/j.trd.2017.06.020

Siddig, D. – Ahmed, A. (2016): Floods as an example of extreme events in Khartoum. Paper presented in a workshop on climate change adaptation in the Economic development sector using integrated water resources management tools, Amman, Jordan. *Economic and Social Commission for Western Asia* (ESCWA).

Singirankabo, E. – Iyamuremye, E. (2022): Modelling extreme rainfall events in Kigali city using generalized Pareto distribution. *Meteorological Applications.* Vol. 29. No. 4. https://doi.org/10.1002/met.2076

Walsh, R. P. D. – Davies, H. R. J. – Musa, S. B. (1994): Flood frequency and impacts at Khartoum since the early nineteenth century. *The Geographical Journal*. Vol. 160. No. 3. pp. 266–279.

World Bank Group, Climate Change Knowledge Portal (2022): Sudan's historical and projected climate data. https://climateknowledgeportal.worldbank.org/country/sudan (downloaded: September 2022)

Zerboni, A. – Brandolini, F. – Mariani, G. S. – Perego, A. – Salvatori, S. – Usai, D. – Pelfini, M. – Williams, M. A. J. (2020): The Khartoum–Omdurman conurbation: A growing megacity at the confluence of the Blue and White Nile rivers. *Journal of Maps*. Vol. 17. No. 4. pp. 227–240. https://doi.org/10.1080/17445647.2020.1758810