

Modelling of Extreme maximum Rainfall using Extreme Value Theory for Tanzania

Triphonia Jacob Ngailo

Department of Mathematics
University of Dar es Salaam, Dar es Salaam, Tanzania
triphongailo@gmail.com

Joachim Reuder

Geophysical Institute,
University of Bergen, Bergen, Norway
Joachim.Reuder@uib.no

Edwin Rutalebwa

General Studies Department
Dar es Salaam Institute of Technology
erutalebwa@gmail.com

Shaban Nyimvua

Department of Mathematics
University of Dar es Salaam, Dar es Salaam, Tanzania
shabanmbare@gmail.com

Michel d.S. Mesquita

Uni Research Climate, Bjerknes Centre for Climate Research, University of Bergen
Bergen, Norway
michel.mesquita@uni.no

Abstract: A series of rainfall data over 31 years in the period 1961-2014 and 1984-2014 recorded at fourteen different stations in Tanzania is modelled using Extreme Value Theory. In order to reduce destruction and loss of life and property, it is necessary to make proper inference about extreme rainfall. The main goal of this study was to determine the best fitting distribution to the extreme daily rainfall based for each station, while considering both stationary and non-stationary processes. The model parameters were estimated and predicted the extreme rainfall return periods and their confidence intervals. The evidences of non-stationary for Dar es Salaam region and stationary for the remaining stations were found. The model fit suggest that, the Gumbel distribution provides the most appropriate model for the annual maximums of daily rainfall and the Exponential distribution gives the reasonable model for the daily rainfall data over the threshold value of 99% for all stations in Tanzania.

Finally the return level estimates and their corresponding confidence intervals for the return periods of 10, 20, 50 and 100 years shows that extreme rainfall has less than 20 years return period in most stations in Tanzania. This indicates that the frequency of occurrence is increasing.

Keywords: Extreme Value Theory, Annual maximum, Threshold value, Generalized Extreme Value distribution, Generalized Pareto distribution, Maximum likelihood estimation, Extreme rainfall, Tanzania.

1. INTRODUCTION

Extreme rainfall events cause significant damage to agriculture, ecology and infrastructure, disruption of human activities, injuries and loss of lives. In Tanzania many areas are affected by heavy rainfall and the associated floods and lands lides. Some examples of the loss caused by floods in the region are the damage both to life and property experienced throughout the country during the 1997/1998 El Niño associated with floods, and the 2011 floods that wrecked the coastal city of Dar es Salaam.

Data related to human and economic losses from disasters that have occurred between 1980 and 2010 in Tanzania show that floods rank second among the top ten natural disasters threatening the country's economy, only outnumbered by epidemics (Mboera et al., 2012). Some towns in Tanzania, for example Dar es Salaam and Mwanza, experience floods almost every rainy season and causes significant damage (Casmiri,2008).Some notable examples are 2010,2011,2012 and 2013 Dar es Salaam floods (Kebede and Nicholls, 2012). In December 2003, March 2006 and December 2007 floods occurred along the low lands of northern Kilimanjaro which caused damage to infrastructure and discontinuity of traffic on the main road and railroad due to overtopping (Mul et al.,2008).Furthermore in 2010, floods occurred in Kilosa (Morogoro), Mpwapwa and Kondoa (Dodoma) more than 50,000 people were affected, 5,100 hectares of crops were destroyed and agricultural land were covered with soil, mud and sand, public facilities were destroyed. Such an indication of increasing frequency and intensity of rainfall means that Tanzania will face a higher probability of huge damages from extreme rainfall in the future. Thus, understanding the patterns of extreme rainfall and their future behaviour is of increasing importance to policy makers in Tanzania.

Several published papers have analysed extreme rainfall using either generalized extreme value (GEV) or generalized pareto (GPD) distributions which provides evidence of the importance of modelling rainfall from different regions of the world: Europe ((Buishand et al., 2008));France (Carreau et al., 2013);Taiwan (Chu et al., 2013); China (Ender and Ma, 2014) ;(Li-Ge et al., 2013) ; Australia (Li et al., 2005) ; Mozambique (Maposa et al., 2014) ; Bangladesh(Shahid, 2011) and Colombo(Varathan et al., 2010) . The observational and statistical modeling results of the above mentioned studies have shown that there are remarkable increases in intensity of precipitation extremes. However, there has been little or no published research that has attempted to detect extreme rainfall by using GEV or GPD in Tanzania. Therefore, this paper would seem to be the first application of the GEV and GPD distributions for extreme rainfall in Tanzania.

In this study, we first find the best fitting distribution to extreme daily rainfall by using all available past data from 1961-2014 and 1984-2014 in fourteen rainfall stations in Tanzania. Secondly we determine the likely future behaviour of these extremes. We use two techniques to define extreme events: one is block maxima that is considering the annual maximum of daily rainfall (one year block) and the second is peak over threshold selecting exceedances over a specific threshold value. The Generalized Extreme Value distribution (GEV) and the Generalized Pareto Distribution (GPD) are used to find the best fitting distribution for the above two techniques respectively. Each of these models has an asymptotic justification, and it is common to use the data to determine which of the two models is most appropriate. The parameters are estimated by maximum likelihood method.(Coles, 2001) considered Maximum Likelihood Estimate (MLE) as the best method because of its all-round utility and adaptability to model change. This means that, the underlying methodology is essentially unchanged even though the estimating equation is modified. Moreover, the confidence intervals for predicted extreme rainfalls are also considered in our study. Details for the GEV and GPD approach are given in Section 2. We elaborate for both models the return levels of extreme events for 10, 20, 50, and 100 years.

According to Kijazi and Reason (2005) the economies of Tanzania is mainly depend on rain fed agriculture, which is highly vulnerable to the amounts and distribution of rainfall. Efforts to achieve food security, reliable hydroelectric power in most parts of the country have been hampered by floods and droughts that are natural events and if not managed in a timely and effectively can be disastrous. The study of extreme rainfall in Tanzania is of paramount importance. Therefore the study will inform decision makers, planners, climatologists and meteorologists about the behaviour of extreme rain- fall. Appropriate decisions and plans can be made based on the results of this study to prepare the general public for changes brought on by extreme rainfall. The paper is organized as follows. Section 2, describes study area, data and methodology for fitting probability models, and identifies best fitted procedures. Finally, the results of the best fitted probability distribution models and their implications are discussed in section 3.

2. METHODOLOGY

2.1 Study Area

The study area covers fourteen stations in Tanzania that lie within $1 - 12^{\circ}\text{S}$ and longitudes $29 - 40^{\circ}\text{E}$ between the great East African lakes, namely; lake Victoria in the north, Tanganyika to the west and Nyasa to the south and to the east, lies the Indian ocean. The map and location of rainfall stations are displayed in Fig. 1.

Map of Tanzania



Figure 1. A map of Tanzania showing Rainfall Stations

Tanzania has a tropical equatorial type of climate. However its climate has a great diversity due to the country's diversity in topography. The country is characterized by two rainfall regimes: namely unimodal and bimodal rainfall regime. The seasonal rains over the unimodal regime occur between October and May (Msimu) over the Southern, South-western, central, and Western areas of the country. The bimodal rainfall regime has two rain seasons, the long rain season (Masika) experienced between March and May (MAM) and the short rain season (Vuli) occurring between October and December (OND) over the Northern coast, North-eastern highlands, Lake Victoria basin, and the Islands of Zanzibar (Unguja and Pemba). These rains are normally associated with the northwards and southwards seasonal movement of the Inter-tropical Convergence Zone (ITCZ) (Kondowe and Aniskina, 2014).

2.2 Data

The daily rainfall data recorded from fourteen weather stations collected from the Tanzania Meteorological Agency (TMA) between 1984-2014 and 1961-2014 respectively. Each station signifies one region in Tanzania. The stations are randomly distributed with different rainfall patterns and geographical location, we analyze the data station by station, and therefore homogeneity test was not applicable in this study. The stations with shorter records or high missing data were not selected and the outliers are considered as extremes.

Daily Rainfall in Dar es Salaam

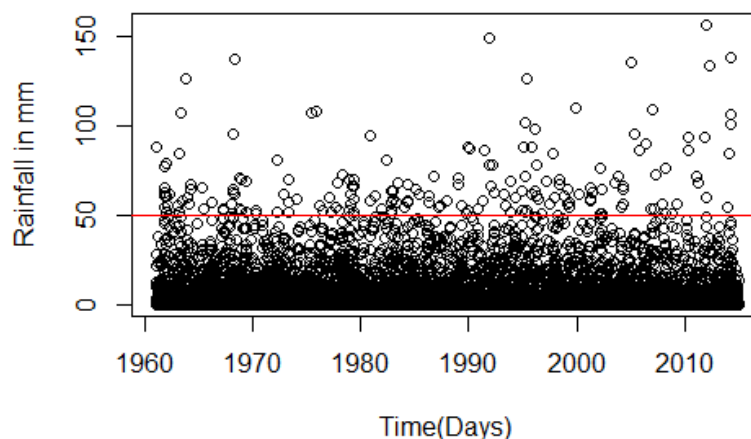


Figure 2. Original Daily rainfall data for Dar es Salam

Figure 2 represents the daily rainfall at Dar es Salaam over the period 1961-2014 and we see some extreme rainfall over the periods (rainfall >50mm). Similarly, we have the same comment from other stations rainfall plots which are not presented here for simplicity.

2.3 Extreme Value Theory (EVT)

EVT provides a tool for modeling the asymptotic distribution of the maximum of a sequence and the sample size based on (Fisher and Tippett, 1928) theorem. In order to describe the behavior of extreme rainfall at a particular area, it is necessary to identify the distribution (s) which best fit the data. In this study, we fit and compare the performance of two different distributions GEV and GPD distributions. We used the R Package for extremes (Gilleland et al., 2013) that is able to perform parametric inferential analysis of the GEV and GPD distributions for each station.

2.3.1 Generalized Extreme Value Distribution (GEV)

In this study we use the Generalized Extreme Value (GEV) distribution, which has all the flexibility of the three extreme value distributions; Gumbel, Fréchet and negative Weibull. This model is appropriate when the maximum observations of each period or block with a predefined and fixed length are assembled from a large number of identically and independently distributed (iid) variables (Coles, 2001). The cumulative distribution function of these three distributions can be summarized by the GEV (Jenkinson, 1955) and is given by;

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

Where x are the extreme values from the blocks, μ a location parameter; σ a scale parameter; and ξ a shape parameter. For $\xi > 0$ we get the Fréchet distribution, $\xi = 0$, the Gumbel distribution and $\xi < 0$ the weibull distribution. The estimates of unknown parameters of GEV was obtained by maximizing the log likelihood with respect to parameter vectors (μ, σ, ξ) . The log likelihood of GEV can be written as;

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

Provided that $\left[1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right] > 0$ for $i = 1 \dots m$.

2.3.2 Generalized Pareto Distribution (GPD) and Peak over Threshold (POT)

In this technique the data are collected over some specific threshold (cut-off) value. Modelling the extremes under this method enables a more efficient usage of extreme value information than that given by an analysis of annual maxima data, which excludes from inference many extreme events that did not happen to be the largest annual event. As a statistical modeling technique this procedure was popularized by (Davison and Smith, 1990). Given a threshold u , the distribution function of extreme values of X over u is,

$$F_u(y) = \Pr(X - u \leq y | X > u) = \frac{F(y + u) - F(u)}{1 - F(u)} \quad (3)$$

$F_u(y)$ represents the probability that the value of X exceeds the threshold u by most amount where $y = X - u$. Balkema and De Haan (1974) and Pickands III (1975) showed that the distribution $F_u(y)$ converges to GPD when the threshold is sufficiently high. In this study, data were fit to the GPD distribution several times with different thresholds. The estimated

GPD parameters were then evaluated to find a stable range of parameter values, the 99% quantile of the data provided reliable parameter estimates. The cumulative distribution function for the GPD is:

$$GPD(x, \xi, \sigma, \mu) = \begin{cases} \left(1 - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right), & \xi \neq 0 \\ \left(1 - \exp\left(-\frac{x - \mu}{\sigma} \right) \right), & \xi = 0 \end{cases} \tag{4}$$

The parameters of GPD were estimated by MLE.

2.3.3 Return Periods for GEV and GPD

Return period (T): Once the best probability model for the data has been determined, the interest is in deriving the return levels of rainfall. The T year return level, say X_T , is the level exceeded on average only once in T years ($x_T = 1 - 1/T$) (Ender and Ma, 2014). The return level for GEV with return period 1/p is obtained by;

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi} [1 - (-\log(1-p))^\xi] \\ \mu - \sigma \log[1 - (-\log(1-p))] \end{cases} \tag{5}$$

For GPD model, the return level is explained by x_m that defines the extreme level that is exceeded on average once every m observations.

$$x_m = \begin{cases} u + \frac{\sigma}{\xi} [(\zeta_u) - 1], & \xi \neq 0 \\ u + \sigma \log(m\zeta_u), & \xi = 0 \end{cases} \tag{6}$$

2.3.4 Stationary test

Mann-Kendall stationarity test is performed on the data to test for stationarity. The test determines the existence of either an increasing or decreasing tendency in daily annual maximum extreme rainfall (Chu et al., 2013). The null hypothesis of the p-value test is that there is no trend while the alternative hypothesis is that there is a trend in the data.

3. RESULTS AND DISCUSSION

This section gives statistical results from the analysis. Table 1 gives annual maximum daily rainfall from all stations in Tanzania. The coastal regions include Dar es Salaam, Tanga, Mtwara, Zanzibar and Pemba, all these regions has high amount of rainfall. The mainland regions, Bukoba in the western part have the highest amount of rainfall followed by Mahenge in the south and Arusha on the north while Sumbawanga in the western- south has moderate rainfall.

Table 1. Descriptive statistics of the annual maximum of daily rainfall data

Location	Years of Data	Min	1st Qu	Median	Mean	3rd Qu	Max
Dar es Salaam	1961-2014	35.80	63.85	71.10	81.71	92.08	156.40
Tanga	1961-2014	46.00	78.98	97.20	107.20	122.50	218.00
Mtwara	1961-2014	42.00	67.35	87.50	94.09	113.20	232.00
Arusha	1984-2014	39.40	49.60	76.60	73.02	82.20	188.70
Morogoro	1984-2014	31.70	51.35	62.20	62.58	76.00	89.80
Dodoma	1984-2014	47.90	63.80	72.10	76.25	89.95	113.00
Bukoba	1984-2014	51.40	70.80	78.50	87.76	92.45	248.00
Kigoma	1984-2014	42.40	64.05	67.00	70.90	78.85	141.00
Pemba	1984-2014	46.60	103.10	129.80	131.5	152.10	227.50

Zanzibar	1984-2014	69.50	96.25	112.20	122.00	140.00	234.00
Songea	1984-2014	43.80	59.95	69.20	73.83	985.60	132.00
Iringa	1984-2014	28.60	44.45	53.50	54.28	60.35	106.90
Mahenge	1993-2014	69.20	94.48	116.00	121.60	152.10	207.10
Sumbawanga	1989-2014	37.10	43.02	50.55	52.90	64.68	75.60

3.1 Modelling using GEV

The annual maximum data are shown in Figure 3. No obvious trend is observed. The annual maxima of daily rainfall accumulations (mm) at different locations in Tanzania, from 1961 to 2014 at coastal regions and 1984 to 2014 at mainland regions.

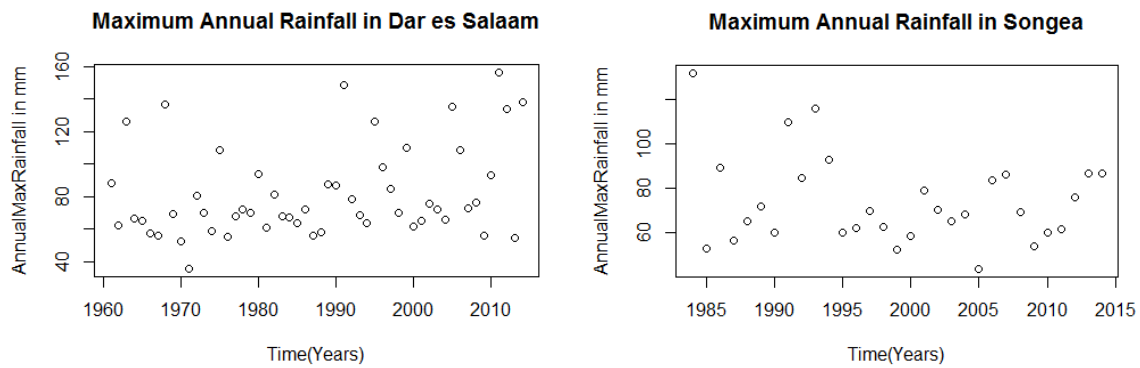


Figure 3. Scatter plots for annual maxima rainfall

Fig. 3 represents maximum annual daily rainfall at Dar es Salaam over the period 1961-2014 and Songea over the period of 1984-2014, showing extreme rainfall over the region using block maxima method. The outstanding feature is that the highest magnitude of 156.4mm occurred in the year 2011 in Dar es Salaam. While this high magnitude appears as an outlier, there is historical evidence that it did occur in three days causing flooding in all districts of Dar es Salaam City, which led significant impacts to lives and economy of the city (Bucchignani et al., 2015).

Table 2. Parameter estimates for GEV

	μ	σ	ξ
Dar es Salaam	68.53 (2.81) 95% CI (63.02,74.05)	18.71(2.11) (14.57,22.85)	0.11(0.09) (-0.07,0.29)
Tanga	87.81(4.84) (78.32,97.30)	31.15(3.60) (24.09,38.21)	0.04(0.11) (-0.176,0.256)
Arusha	59.44 (3.99) (51.62,67.26)	18.79(3.13) (12.65,24.92)	0.14(0.17) (-0.19,0.47)
Morogoro	57.76(3.23) (51.43,64.09)	15.63(2.50) (10.73,20.53)	-0.36(0.17) (-0.69,-0.027)
Dodoma	68.56 (3.42) (61.86,75.26)	16.05(2.55) (11.05,21.05)	-0.13(0.19) (-0.5,0.24)
Bukoba	73.89 (3.21) (67.60,80.18)	16.15(2.50) (11.25,21.05)	0.20(0.12) (-0.04,0.44)
Kigoma	63.25 (2.65) (58.06,68.44)	13.45(1.87) (9.78,17.12)	0.002(0.10) (-0.194,0.198)
Mtwara	75.93 (4.42) (67.27,84.59)	27.67(3.97) (19.89,35.45)	0.07(0.13) (-0.25,0.26)
Pemba	114.99 (7.79) (99.73,130.27)	38.95(5.44) (28.28,49.61)	-0.18(0.12) (-0.415,0.055)
zanzibar	102.89 (5.71) (91.70,114.08)	27.66(4.36) (19.11,36.21)	0.11(0.16) (-0.20,0.42)

Modelling of Extreme maximum Rainfall using Extreme Value Theory for Tanzania

Songea	64.63 (2.82) (59.10,70.16)	13.91(2.11) (9.77,18.05)	0.08(0.14) (-0.19,0.35)
Iringa	47.96 (2.26) (43.53,52.39)	11.45(1.58) (8.25,14.55)	-0.02(0.10) (-0.216,0.176)
Mahenge	104.68 (7.49) (90.00,119.36)	29.47(5.67) (18.36,40.58)	-0.013(0.22) (-0.44,0.42)
Sumbawanga	48.40 (2.84) (46.75,53.96)	11.84(2.05) (7.82,15.86)	-0.25(0.17) (-0.58,0.08)

For GEV estimation, the Block Maxima of daily annual maximum rainfall are extracted for all fourteen stations. The blocks $n = 365$ days have been chosen to be reasonably large, so we fitted our Model to the $N = 31$ and $N = 54$ annual maxima using maximum likelihood estimation. Table 2 gives the parameters of GEV as results of MLE fitted to daily annual maxima. The estimates are as follows; The first line of each parameter indicates the value and the standard errors (s.e). The corresponding 95% confidence intervals (CI) are included in the second line, the results shows that the confidence interval of ζ contains 0, which means the Gumbel Distribution could be a more accurate model in the entire GEV family. A further support is provided by the profile confidence interval for ζ displayed in table 2 the value $\zeta = 0$ is well contained within the confidence interval in all stations.

Table 4 gives the return level estimates at different return periods for daily annual maxima. The 95% CI are included in brackets.

The estimated return levels and CI increase with the increase of the return period. Compared with the lower bound of the confidence interval, the upper bound is likely to be further away from the predicted return level when the return period is longer. There is an increasing pattern for Tanga, Mtwara, Arusha, Dodoma, Bukoba, Kigoma, Pemba, Zanzibar, Songea, Iringa, Mahenge and Sumbawanga for return period of 10 to 20 years. According to Tanzania Meteorological Agency (2011), 24-hour rainfall amount exceeding 50mm is considered as extreme rainfall that can cause floods. For all stations in Tanzania the extreme event of floods occurs much more as 50mm is included in the CI of $T = 10$ and $T = 20$. Hence, floods might occur once in every 10 to 20 years in Tanzania.

Table 3. GEV return level estimates

Period	10-years	20-years	50-years	100-years
Dar es Salaam	116(103, 138)	134(116,171)	160(133, 225)	181(145, 276)
Tanga	161(143, 195)	186(161,243)	220(183, 324)6	246(199, 403)
Mtwara	144(125,179)	168(144,231)	201(166, 325)	228(180, 428.)
Arusha	109 (93, 157)	128 (106, 228)	156(121, 325)	180 (132, 422)
Morogoro	82(76,91)	86(81, 101)	90 (85, 115)	93(87, 125)
Dodoma	100(91, 120)	108(98, 141)	118(105, 177)	125(109, 211)
Bukoba	120(104, 156)	140(117, 204)	170(133, 303)	196(146, 405)
Kigoma	94(85, 111)	103(92, 131)	116(102, 163)	125(108, 195)
Pemba	187(168, 221)	204(183, 255)	224(199, 303)	236(209, 344)
Zanzibar	173(150, 230)	199(168, 303)	236(190, 447)	267(203, 597)
Songea	99(88, 123)	111(96,153)	128(106, 209)	142(113, 265)
Iringa	73(66, 87)	81(72,103)	91(79, 129)	98(84, 151)
Mahenge	170(147, 239)	191(162, 321)	217(179, 477)	236(189, 648)
Sumbawanga	69(63, 81)	73(67, 91)	77(71, 108)	80(73, 121)

The model diagnostic plot looks reasonable for rainfall at Dar es Salaam for GEV (figure 3). Similar way we made diagnostic plot for other stations that also looks good, are not presented in this study. All of the four diagnostic plots for all the stations; probability plot, quantile plot, return level plot and density plot, provides evidence that GEV is a good fit to the block maxima.

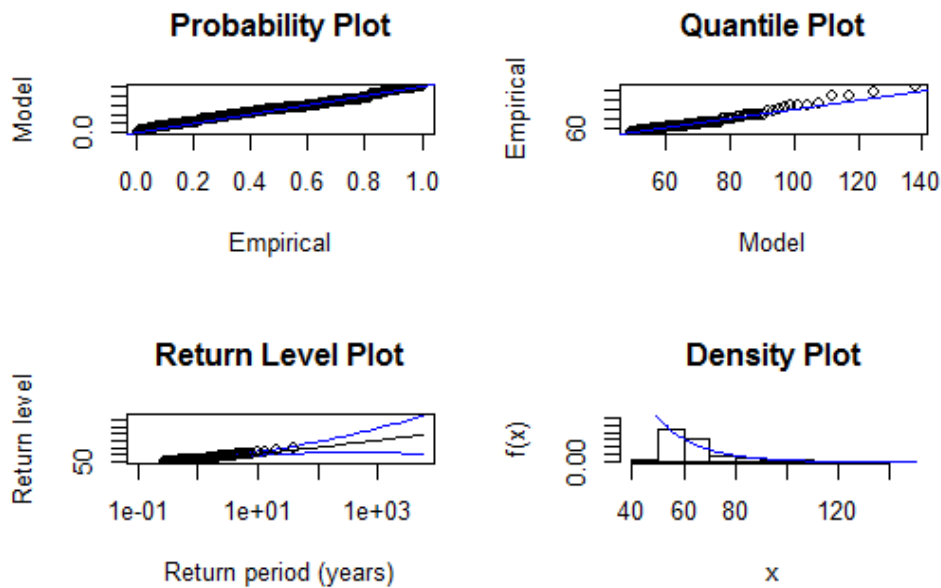


Figure 4. Diagnostic plot for GEV

3.2 Modeling using GPD

The issue of how to choose the threshold is similar to that of selecting the size of a block in the sense that both imply a balance between bias and variance. A low level leads to failure in the asymptotic approximation of the model and a high level provides few observations and then high variance. The choice of appropriate thresholds in this study is based on mean excess plots methods presented in figure 5. We select one constant threshold for every region. The mean excess plots is plotted for Dar es Salaam to represent other regions. The possible threshold is the point when the mean excess plot shows linearity, and estimated parameters look stable at different thresholds (Chu et al., 2013) as an example, we can observe an increasing linear tendency between 40 and 75 for Dar es Salaam, this increasing tendency indicates a positive value for parameter ζ . We choose a threshold u the value 48, as presented Fig. 5, which represents the 99 percentile.

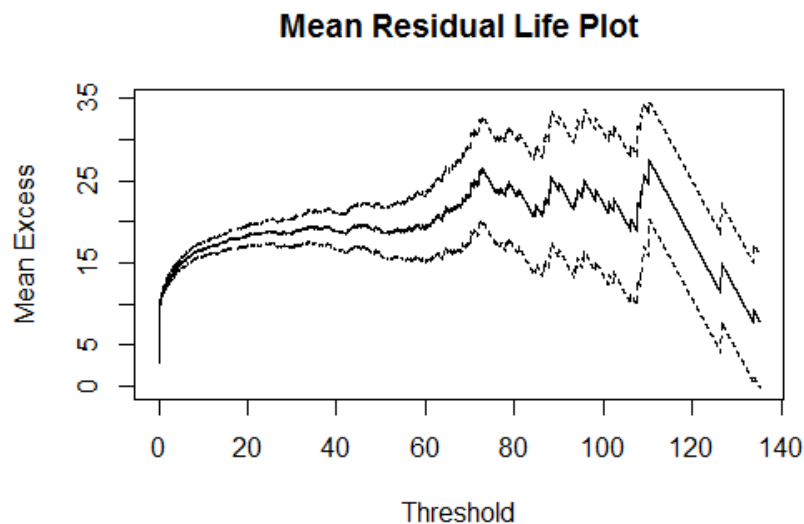


Figure 5. Mean Residual plot for Dar es Salaam

The same has been done for other regions, the plots are not shown in this work. After threshold selection, the parameters for GPD as given in equation (4) are estimated using MLE. The results are represented in Table 5. In all regions CI include the zero for the shape parameter that is, the data fits the Exponential distribution.

Table4. GPD Parameter estimates

	Threshold	σ	ξ
Dar es Salaam	48 n=201	17.75(1.87) (14.08,21.42)	0.07(0.079) (-0.08,0.22)
Tanga	37 n=194	26.85(2.97) (21.02,32.67)	0.07(0.08) (-0.087,0.227)
Arusha	39 n=113	16.10(2.34) (11.51,20.68)	0.12(0.11) (-0.096,0.34)
Morogoro	37 n=113	13.56(2.18) (9.29,17.83)	-0.03(0.13) (-0.28,0.22)
Dodoma	39 n=114	22.96(3.06) (16.96,28.95)	-0.18(0.096) (-0.368,0.0082)
Bukoba	56 n=114	13.98(1.76) (10.51,17.43)	0.13(0.08) (-0.03,0.29)
Kigoma	39 n=112	18.36(2.10) (14.24,22.48)	-0.085(0.065) (-0.21,0.042)
Mtwara	48 n=198	23.84(2.53) (18.88,28.80)	0.09(0.08) (-0.07,0.25)
Pemba	67 113	31.35(4.61) (22.31,40.39)	0.03(0.08) (-0.13,0.19)
zanzibar	63 n=114	29.78(4.34) (21.27,38.28)	0.07(0.11) (-0.15,0.28)
Songea	45 n=112	15.55(2.14) (11.36,19.74)	0.002(0.0996) (-0.193,0.197)
Iringa	29 n= 113	15.96(1.70) (12.63,19.29)	-0.11(0.07) (-0.247,0.027)
Mahenge	69 n= 81	36.44(5.80) (25.07,47.81)	0.13(0.11) (-0.09,0.35)
Sumbawanga	29 n= 116	15.96(1.90) (12.24,19.68)	-0.11(0.07) (-0.23,0.03)

Model checking QQ-plots of the empirical versus fitted models are reasonably linear and the density plots indicate that the assumptions for using the GPD and GEV are fulfilled for all stations.

Table5. GPD return level estimates

Period	10-years	20-years	50-years	100-years
Dar es Salaam	112(110,143)	137(122,170)	160(135,216)	178(148,255)
Tanga	164(147,198)	188(163,242)	223(184,318)	250(200,385)
Mtwara	148(132,178)	172(148,222)	205(169,288)	231(183,351)
Arusha	111(96,148)	129(107,191)	155(122,268)	177(134,348)
Morogoro	84(75,105)	92(81,127)	103(87,161)	111(90,195)
Dodoma	100(92,116)	108(98,132)	117(104,151)	122(108,167)
Bukoba	120(107,149)	136(117,182)	160(134,242)	180(146,300)
Kigoma	96(87,110)	105(97,124)	116(104,148)	124(112,264)
Pemba	187(163,242)	212(180,298)	245(198,397)	272(210,492)
Zanzibar	185(160,242)	212(180,310)	250(198,418)	280(210,492)
Songea	101(90,122)	111(98,145)	126(108,182)	136(114,216)
Iringa	76(70,90)	84(76,103)	92(80,121)	98(87,137)
Mahenge	174(157,216)	189(167,249)	207(179,300)	219(188,352)
Sumbawanga	77(76,100)	85(81,116)	96(87,141)	103(92,161)

Comparing return levels estimated by GEV with the results of GPD, there are not significant differences for all periods and all stations (T=10,20,50 and 100). From the Tables 3 and 5 we can notice, the predicted return values and the confidence levels are very close in both sampling techniques. Mann-Kendall trend analysis was performed at each station to test the trends of 95 percentile annual maximal extreme rainfall. The results are tabulated in Table 6. Only one station out of fourteen stations under study shows significant change the p-value is less than 0.005 in annual maxima rainfall. The trends for other stations were insignificant.

Table 6. Results of statistical tests for trends at weather stations in Tanzania

station	Data span (years)	Kendall τ	P-value
Dar es Salaam	54	0.199	0.035
Tanga	54	-0.118	0.2100
Mtwara	54	-0.135	0.1520
Arusha	31	-0.012	0.952
Morogoro	31	-0.058	0.659
Dodoma	31	0.047	0.722
Bukoba	31	-0.069	0.598
Kigoma	31	-0.083	0.532
Pemba	31	-0.204	0.1101
Zanzibar	31	0.018	0.9458
Songea	31	-0.0301	0.825
Iringa	31	0.204	0.133
Mahenge	22	0.0043	1
Sumbawanga	25	-0.229	0.143

The model diagnostic plot for GPD looks reasonable for rainfall at Dar es Salaam and Kigoma figure 6. Similar way we made diagnostic plot for other stations that also looks good, are not presented in this study. Thus the model chosen is valid.

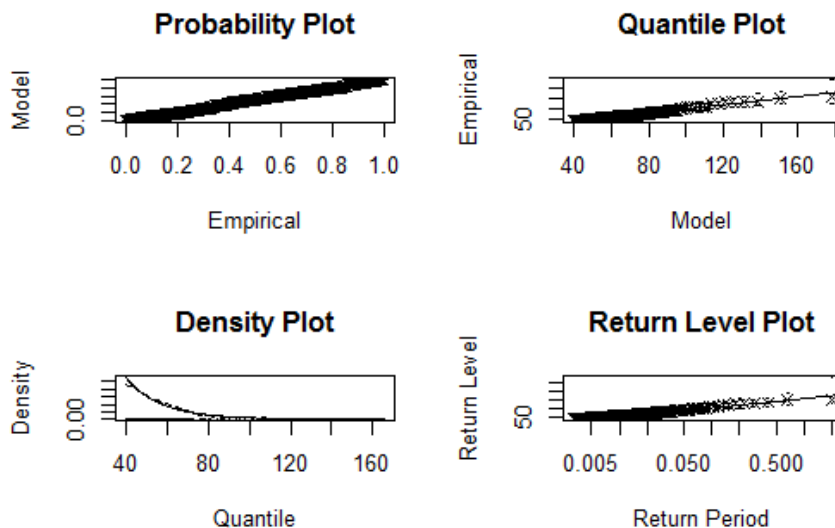


Figure 6. Diagnostic plots for Dar es Salaam

4. CONCLUSION

Throughout this study, how Extreme Value Theory used as a significant tool in describing extreme events is explained. Two models are considered for daily rainfall data as well as annual maximum rainfall data, both model are well fitted and interrelated. We estimated return levels of extreme rainfall using stationary and non-stationary models for 1961-2014 and 1984-2014 extreme rainfall data for fourteen stations. The purpose is to support decision makers in these regions with

statistical knowledge about extreme rainfall that they can choose appropriate risk mitigating measures to reduce the damage caused by floods.

The empirical results indicated that there have been significant increases in annual maxima over the period 1961-2014 in Dar es Salaam confirming trends over the region, other stations trend were insignificant. Among these stations, have a greater value for a location parameter such as Tanga and Bukoba, Mtwara, Pemba, Zanzibar and Mahenge stations showing that there is an increase in extreme rainfall. Given 50-year return level (for the year 2064), the return levels of daily extreme rainfall of all the cities except Iringa and Sumbawanga stations were all greater than 100mm, reaching a warning line of extreme high rainfall, as defined by the Tanzania Meteorological Agency in Tanzania.

Increasing extreme rainfall events are observed in Tanzania with the greater annual mean as seen in table 1 for some regions which might have number of negative implications. An interesting result of this study, which requires more in-depth examination, is the frequency of extreme rainfall occurrences. Therefore there is a need to extend this work to study frequency of extreme rainfall occurrences. This study would play as advantageous in understanding extreme rainfall in Tanzania. This lends strong support to take proper caution against the natural disasters due to extreme rainfall and save the life and property.

ACKNOWLEDGEMENTS

I thank Tanzania Meteorological Agency for providing their data; WIMEA-ICT project through NORAD and Dar es Salaam Institute of Technology for their financial supports.

REFERENCES

- [1] Balkema, A. A. and De Haan, L., Residual life time at great age. *The Annals of probability*, pages 792–804, (1974).
- [2] Bucchignani, A. G.-a. E., Palazzi, E., Onofrio, D. D., Gasparini, P., Marzocchi, W., Garcia-Aristizabal, A., Bucchignani, E., Palazzi, E., DOnofrio, D., Gasparini, P., and Marzocchi, W., Analysis of non-stationary climate-related extreme events considering climate change scenarios: an application for multi-hazard Assessment in the Dar es Salaam region, Tanzania. *Natural Hazards*, 75(1):289–320, (2015).
- [3] Buishand, T. a., de Haan, L., and Zhou, C., On spatial extremes: With application to a rainfall problem. *Annals of Applied Statistics*, 2(2):624–642, (2008).
- [4] Carreau, J., Neppel, L., Arnaud, P., and Cantet, P., Extreme Rainfall Analysis at Ungauged Sites in the South of France: Comparison of Three Approaches. 154 (2):119–138, (2013).
- [5] Casmiri, D., Vulnerability of Dar es Salaam City to Impacts of Climate Change. *Emps*, (2008).
- [6] Chu, L.-f., McAleer, M., and Chang, C.-C., Statistical Modelling of Extreme Rainfall in Taiwan. *Tinbergen Institute, Tinbergen Institute Discussion Papers: 13-006/III*, 2013, 15(December 2012):6536, (2013)
- [7] Coles, S., An introduction to statistical modeling of extreme values. Springer-Verlag, (2001)
- [8] Davison, A. C. and Smith, R. L., Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 393–442, (1990).
- [9] Ender, M. and Ma, T., Extreme Value Modeling of Precipitation in Case Studies for. *International Journal of Scientific and Innovative Mathematical Research*, 2(1):23–36, (2004).
- [10] Fisher, R. A. and Tippett, L. H. C., Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 180–190. Cambridge Unit Press, (1928).
- [11] Gilleland, E., Ribatet, M., and Stephenson, A. G., A software review for extreme value analysis. *Extremes*, 16(1):103–119, (2013).
- [12] Jenkinson, A. F., The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81(348):158–171, (1955).

- [13] Kebede, A. S. and Nicholls, R. J., Exposure and vulnerability to climate extremes: Population and asset exposure to coastal flooding in Dar es Salaam, Tanzania. *Regional Environmental Change*, 12(1):81–94, (2012).
- [14] Kijazi, A. and Reason, C., Relationships between intraseasonal rainfall variability of coastal Tanzania and ENSO. *Theoretical and applied climatology*, 82(3-4):153–176, (2005).
- [15] Kondowe, A. L. and Aniskina, O. G., Impacts of grid spacing and cumulus schemes on the quality of rainfall forecasts over Tanzania. (678), (2014).
- [16] Li, Y., Cai, W., and Campbell, E., Statistical modeling of extreme rainfall in southwest Western Australia. *Journal of Climate*, 18(1999):852–863, (2005).
- [17] Li-Ge, C., Jun, Z., Bu-Da, S., Jian-Qing, Z., and Gemmer, M., Probability Distribution and Projected Trends of Daily Precipitation in China. *Advances in Climate Change Research*, 4(3):153–159, (2013).
- [18] Maposa, D., Cochran, J. J., Lesaoana, M., and Sigauke, C., Estimating high quantiles of extreme flood heights in the lower Limpopo River basin of Mozambique using model based Bayesian approach. *Natural Hazards and Earth System Sciences Discussions*, 2(8):5401–5425, (2014).
- [19] Mboera, L. E., Mayala, B. K., Kweka, E. J., and Mazigo, H. D., Impact of climate change on human health and health systems in Tanzania: a review. *Tanzania journal of health research*, 13(5), (2012).
- [20] Mul, M. L., Savenije, H. H. G., and Uhlenbrook, S., Spatial rainfall variability and runoff response during an extreme event in a semi-arid catchment in the South Pare Mountains, Tanzania. *Hydrology and Earth System Sciences Discussions*, 5:2657–2685, (2008).
- [21] Pickands III, J., Statistical inference using extreme order statistics. *The Annals of Statistics*, pages 119–131, (1975).
- [22] Shahid, S., Trends in extreme rainfall events of Bangladesh. *Theoretical and Applied Climatology*, 104:489–499, (2011).
- [23] Varathan, N., Perera, K., and Wikramanayake, N., Statistical Modelling of Daily Extreme Rainfall in Colombo. *International Conference on Sustainable Built Environments 2010 (ICSBE 2010)*, (December):13–14, (2010).

AUTHOR'S BIOGRAPHY



Triphonia Jacob Ngailo, PhD student at University of Dar es Salaam, My research interest is on mathematical modeling of weather extremes and numerical Weather predictions. Works as an Assistant Lecturer (Mathematics) at Dar es Salaam Institute of Technology.