Abstract
Climate change has brought about unprecedented new weather patterns, one of which is changes in extreme rainfall. In Kenya, heavy rains and severe flash floods have left people dead and displaced hundreds from their settlements. In order to build a resilient society and achieve sustainable development, it is paramount that adequate inference about extreme rainfall be made. To this end, this research modelled and predicted extreme rainfall events in Kenya using Extreme Value Theory for rainfall data from 1901-2016. Maximum Likelihood Estimation was used to estimate the model parameters and block maxima approach was used to fit the Generalized Extreme Value Distribution (GEVD) while the Peak Over Threshold method was used to fit the Generalized Pareto Distribution (GPD). The Gumbel distribution was found to be the optimal model from the GEVD while the Exponential distribution gave the optimal model over the threshold value. Furthermore, prediction for the return periods of 10, 20, 50 and 100 years were made using the return level estimates and their corresponding confidence intervals were presented. It was found that increase in return periods leads to a corresponding increase in return levels. However, the GPD gave higher return levels for 10 and 20 years compared to GEVD. While, for higher return periods 50 and 100 years, the GEVD gave higher return levels compared to the GPD. Model diagnostics using probability, density, quantile and return level plots indicated that the models provided were a good fit for the data.

Keywords: block maxima, extreme rainfall, extreme value theory, generalized extreme value distribution, generalized Pareto distribution, maximum likelihood estimation, peak over threshold, sustainable development

1. Introduction
The East African community is prone to climate and weather extremes with a highly variable climate, and has relatively high levels of population exposure and vulnerability. Specifically, Kenya is not new to extreme rainfall and the study of Parry, Echeverria, Dekens, and Maitima (2012) found out that Kenya’s exposure to climate risk is high, experiencing major droughts about every 10 years and moderate droughts or floods every 3 to 4 years, and as such regarded as one of the most disaster-prone countries in the world. The study further showed that, floods and droughts in Kenya accounts for a liability of about 2.0 to 2.4% of the Gross Domestic Product (GDP) each year. Sadly, as global climate change increase it has been shown that there will be a concomitant increase, in climate risks on Kenya’s economy. In addition, Huho and Kosonei (2014) found that there is an inverse relationship between economic development and extremes in rainfall that causes disasters. This has been seen in the Long-Rains during March - May of 2018 which was regarded as one of the wettest seasons (Kilavi et al, 2018).

It has been observed that heavy rainfall has some advantages as noted by the Kenya Food Security Steering Group (2018), in a post-season assessment of the 2018 March-May rains which indicated that the heavy rainfall had reversed some of the pre-existing food security impacts of preceding poor rainfall seasons, with many counties moving back from ‘Crises’ to ‘Stressed’ status. However, there are also negative effects of heavy rainfall as noted by The World Weather Attribution (2018), the heavy rains in 2018 displaced nearly 300,000 people and killed 158. Also, floods caused by heavy rainfall affects the economy by bringing businesses to a halt, decreasing agricultural production, reducing prices for agricultural produces and livestocks, destruction of the immediate environment which in turn exacerbate the impacts and leaves displaced people with limited humanitarian support due to a lack of access to the area (Huho & Kosonei, 2014).

Therefore, the importance of modelling and predicting rainfall performance in Kenya cannot be overemphasised since
Kenya is an agric-based economy and the effect of climate change induced rainfall pose great challenges and opportunities for Kenya. It goes without saying that, economic planners, climatologists, meteorologists, and policy makers in Kenya need to understand extreme rainfall patterns and future behaviours for effective decision making, planning and mitigation purposes. Extreme Value Theory (EVT) furnishes us with pertinent tools for modelling and predicting extreme rainfall in Kenya and this is the focus of this article.

The rest of the paper is structured accordingly. Section 2 discusses some relevant related literature to the present study. Section 3 presents the research methods and materials that guides the study while section 4 contains the results and discussion. Finally, the conclusion is presented in section 5.

2. Literature Review

Extreme Value Theory has profound applications in different areas such as flooding, rainfall, storms, precipitation, insurance claims, price fluctuations, to mention a few. Several researchers around the world have applied EVT in different countries to come up with several estimates about extreme events. The study of Lazoglou and Anagnostopoulou (2017) applied EVT in the Mediterranean region on ten stations data. Block Maxima and Peak Over Threshold (POT) with Maximum Likelihood, L-Moments, and Bayesian methods were employed in calculating the parameters of the extreme distributions. Results from the study indicated that GPD provides a better theoretical justification compared to Generalized Extreme Value (GEV) distribution in the prediction of extreme precipitation. Bayesian method gave the most accurate estimate parameters for the highest precipitation levels in most of the stations. GEV distribution with Bayesian estimator was found to give the highest return levels for the western stations while GPD with Bayesian estimator gave the highest return levels for the eastern regions for 50, 150 and 300 return periods.

Similarly, García-Cueto and Santillán-Soto (2012) applied EVT to two climatic datasets in Mexico, namely, temperature and rainfall. For the maximum temperature modelling, the study applied block maxima approach and the findings revealed that the Weibull distribution gave the most adequate estimates for the summer maximum temperature. POT method was used to fit the GPD model on the daily maximum rainfall. The work of Chifurira and Chikobvu (2014) fitted GEV distribution on the mean annual rainfall to describe rainfall extremes in Zimbabwe. Probabilities of meteorological floods were estimated by EVT. GEV distribution was fitted by the maxima distribution. The GEV model of the Gumbel distribution was found to fit the data well using the Anderson-Darling goodness of fit test. GEV distribution with constant shape and scale parameters, but with varying location parameter over time was inadequate to model Zimbabwe’s extreme maximum rainfall. The study indicated that a high mean annual rainfall of 1193 mm is expected in approximately 300 years.

In the same vein, Uwimana and Joseph (2018) applied EVT to Kigali monthly rainfall data, the best model was used to forecast, and this was compared to observed data to check if the estimated results were in agreement with reality. Gumbel distribution was also found to be the best distribution model. Estimates from return level showed that extreme rainfall will happen in 11 years. Also, Iyamuremye, Wanyonyi, and Mbate (2019) described the change in extreme rainfall using rainfall indices related to extremes for Dodoma, Tanzania. EVT techniques, block maxima and POT, were used for univariate sequences for independent and identically distributed rainfall data. The study did not find any linear trend in the extreme rainfall for 77 years in the Dodoma rainfall data. The Gumbel distribution of the GEV gave the best fit for the data. Finally, the researchers used GPD to fit the excesses under the assumption that the rainfall data was stationary.

Meanwhile, Mondal and Mujumdar (2015) used EVT to analyse characteristic changes in extreme rainfall in India using a high-resolution daily gridded dataset. Non-stationary distributions with varying parameters for physical covariates like ENSO-index, global average temperature and local mean temperatures were used to model intensity, duration and frequency of extreme rainfall over a high threshold. Intensity, duration and frequency were found to be non-stationary and no spatially uniform pattern was found in their changes across India. Duration of extreme rainfall was found to be stationary in most of the locations in India, while associations between frequency, intensity and local changes in temperature were found to be non-stationary. Wang et al. (2015) used change point detection and adjustment procedures to homonize long-term Macau observations, and then used EVT to study the changes in extreme meteorological events including summer heat waves and extreme rainfall. Extreme rainfall was modelled using Poisson-generalized Pareto model, heat waves were modelled by an extended Poisson-generalized Pareto model. In the observational data for Macau, the frequency of extreme rainfall increases significantly, while the positive trend is non-significant.

3. Research Methods and Materials

3.1 Area of the Study

The area of the study is Kenya which is a country in East Africa, bordering the Indian Ocean in south east, neighbouring countries are Ethiopia, Somalia, Tanzania, Uganda and South Sudan. With an area of 580,000 km², Kenya’s landscape varies from low plains near its coast at the Indian Ocean, to a fertile plateau in west. Nairobi is the largest city as well as the capital city. Figure 1 belows shows the map of Kenya.
3.2 Data Used in the Study

The data used in the study was obtained from the World Bank Group, Climate Change Knowledge Portal https://climateknowledgeportal.worldbank.org/ on monthly rainfall in Kenya from 1901 to 2016.

3.3 Techniques for Selecting Extremes

In order to model extreme events, it is paramount that one selects extreme values from the data. There are two methods for identifying extremes in a dataset, namely, Peak Over Threshold method and Block Maxima method.

3.3.1 Peak Over Threshold

The POT approach models excess values of a sample over a given (high) threshold within a time period, and estimates the tail behavior using the conditional distribution of these exceedances.

Given a threshold \( u \), the distribution function of extreme values of \( X \) over \( u \) is,
\[
F_u(x) = P(X - u \leq x | X > u) = \frac{F(x + u) - F(u)}{1 - F(u)}
\]  

(1)

Where \( F_u(x) \) is called the conditional excess distribution which represents the probability that the value of \( X \) exceeds the threshold \( u \) by most amount where \( x = X - u \).

Figure 3. Peak Over Threshold Method for Selecting Extreme Values

Figure 3 shows the POT method for selecting extreme values. From the diagram, the values above the threshold \( u \) namely, \( x_2, x_4, x_6 \) and \( x_7 \) are the extreme values.

3.3.2 Block Maxima

The block maxima approach consists of dividing the observation period into non-overlapping blocks of equal size and restricts attention to the maximum observation in each block. Larger blocks mean fewer maxima and larger variance in estimation, while smaller blocks may not fit the GEV distribution and lead to bias. Often block size is determined from the pragmatic perspective or data availability, leading to data such as annual, monthly, or daily maxima.

Figure 4. Block Maxima Method for Selecting Extreme Values
From Figure 4, the data is divided into three equal blocks and the maximum from each block namely, \( x_2 \), \( x_6 \) and \( x_7 \) forms the extreme values. Comparing the Block Maxima Method to the POT method, it can be seen that using the Block Maxima method some values that may be extreme but not the maximum in the block would not be included in the extreme values. Thus, the POT method gives a more efficient usage of the extreme value information.

### 3.4 Extreme Value Theory

Extreme Value Theory is a well developed theory in the field of probability that studies the distribution of extreme realizations of a given distribution function, or of a stochastic process, satisfying suitable assumptions. The foundations of the theory were set by Fisher and Tippett (1928) and Gnedenko (1943). A modern and large account of univariate Extreme Value Theory can be found in, Galambos (1985), Ferreira and de Haan (2006), Beirlant, Goegebeur, Segers, and Teugels (2004), Embrechts, Klüppelberg, and Mikosch (1997), and recently Lo, Ngom, Kpanzou, and Diallo (2018).

#### 3.4.1 Generalized Extreme Value Distribution

Consider that \( X_1, X_2, \ldots, X_n \), is a sequence of independent and identically distributed \((i.i.d.)\) random variables from distribution function \( H \). Let \( M_n = \max\{X_1, X_2, \ldots, X_n\} \). The exact distribution of \( M_n \) is \( H^n \). Suppose that there exists sequences of constants \( b_n > 0 \) and \( a_n \) such that:

\[
P\left( \frac{M_n - a_n}{b_n} \leq x \right) = H^n(b_n x + a_n) \rightarrow G(x)
\]

where \( G \) is non-degenerate distribution function, then \( G \) belongs to the Gumbel, Fréchet and Weibull families. The cumulative distribution function of these three distributions can be summarized by the GEVD given by:

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

where \( x \) are the extreme values from the blocks, \( \mu \) a location parameter; \( \sigma \) a scale parameter; \( \xi \) a shape parameter. The condition for a distribution to belong to any of the extreme value distributions is given as:

- (a) \( \xi = 0 \), Gumbel distribution
- (b) \( \xi > 0 \), Fréchet distribution
- (c) \( \xi < 0 \), Weibul distribution

Estimating the unknown parameters of the GEV above, leads to finding likelihood of GEV which can be written as:

\[
L_\ell(\mu, \sigma, \xi) = -q \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{q} \log \left[1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)^{-1}\right] = \sum_{i=1}^{q} \left[1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)^{-1}\right]^{-1}
\]

where \( 1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)^{-1} > 0 \) for \( i = 1, \ldots, q \).

#### 3.4.2 Generalized Pareto Distribution

In the heart of the threshold exceedances approach there is the GPD. The data which exceed the threshold are modelled according to the GPD.

The cumulative distribution function for the GPD is:

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

The GPD includes three types of distribution:

- (a) \( \xi = 0 \), Exponential distribution.
- (b) \( \xi > 0 \), Ordinary Pareto distribution.
- (c) \( \xi < 0 \), Pareto-II type distribution.
3.5 Return Periods

Let $z_{1-p}$ be the extreme quantile with order $(1 - p)$ of generalized extreme value distribution fundamental to the random variable $X$, with $p$ sufficiently small. The return level is obtained as:

$$
z_{1-p} = \begin{cases} 
\mu + \frac{\sigma}{\zeta} [(- \log(1 - p))^{-\zeta} - 1], & \zeta \neq 0 \\
\mu - \sigma \log (- \log(1 - p)), & \zeta = 0
\end{cases}
$$  \hfill (6)

For the GPD model, the return level is given by $x_q$ which defines the extreme level that exceeded the average once every $q$ observations. Thus the return level for the GPD is given as:

$$
x_q = \begin{cases} 
u + \frac{\sigma}{\zeta} [\zeta_q - 1], & \zeta \neq 0 \\
u + \sigma \log(q\zeta), & \zeta = 0
\end{cases}
$$  \hfill (7)

4. Results and Discussion

This section entails the presentation of the results after analysis using EVT and GPD packages in R. Also, discussion about the obtained results is presented.

4.1 Generalized Extreme Value Modelling

This section shows the results and discussion for modelling extreme rainfall in Kenya using GEV.

Figure 5. Annual Maximum Rainfall in Kenya (1901-2016)

Figure 5 does not display any trend on annual maximum data in Kenya. The highest maximum rainfall occurred in November 1961 with a recorded magnitude of 311.4080mm. According to Conway (2002), this extreme rainfall event in November 1961 caused widespread flooding, prolonged and rapid increase in the levels of many lakes in East Africa and brought about great disruption in the Kenyan economy. Similarly, the second highest extreme rainfall occurred in November 1997 with a recorded magnitude of 247.0810mm. As shown by Conway (2002), it brought about serious flooding, loss of life, damage to housing and infrastructure. Also, there was a 33% decrease in crop yields and an outbreak of Rift Valley Fever.
Table 1. Generalized Extreme Value Parameter Estimates

<table>
<thead>
<tr>
<th>Kenya</th>
<th>Parameter Estimates</th>
<th>Standard Errors</th>
<th>Confidence Interval (CI) (95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \xi )</td>
<td>( \sigma )</td>
<td>( \mu )</td>
</tr>
<tr>
<td>Parameter Estimates</td>
<td>0.237</td>
<td>21.101</td>
<td>190.736</td>
</tr>
<tr>
<td>Parameter Standard Errors</td>
<td>0.251</td>
<td>5.645</td>
<td>6.955</td>
</tr>
<tr>
<td>Confidence Interval (CI) (95%)</td>
<td>(-0.256, 0.730)</td>
<td>(10.038, 32.164)</td>
<td>(177.104, 204.369)</td>
</tr>
</tbody>
</table>

Table 1 indicates the results of the GEV modelling on extreme rainfall data in Kenya using Block Maxima approach. The GEV parameters were estimated using Maximum Likelihood Estimation (MLE). Since the confidence intervals of \( \xi \) contains 0, it implies that the Gumbel distribution is the optimal model for the GEV family.

Table 2. Generalized Extreme Value Return Level

<table>
<thead>
<tr>
<th>Return Periods</th>
<th>10 Years</th>
<th>20 Years</th>
<th>50 Years</th>
<th>100 Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Return Levels</td>
<td>253.499</td>
<td>281.758</td>
<td>326.293</td>
<td>366.753</td>
</tr>
<tr>
<td>CI (95%)</td>
<td>(-222.671, 423.878)</td>
<td>(236.450, 639.547)</td>
<td>(254.967, 815.733)</td>
<td>(268.153, 825.194)</td>
</tr>
</tbody>
</table>

Table 2 shows the predicted maximum rainfall return level (in mm) for the return periods of 10, 20, 50 and 100 (in years) along with their respective 95% confidence intervals. It can be seen from Table 2 the results that increase in return periods leads to a corresponding increase in return levels.

Figure 6. GEV Diagnostic Plots

Figure 6 shows the diagnostic plots for the GEV. The points of the probability plot and quantile plot lies close to the unit diagonal. This implies that, generalized extreme value distribution function provides a good fit.

The return level plot shows that the empirical return levels match well with those from the fitted distribution function. Finally, the density plot also shows good agreement between the fitted GEV distribution function and the empirical density.
4.2 Generalized Pareto Distribution Modelling

The hill plot in Figure 7 was used to determine the threshold. The threshold is selected from the plot where the tail index is fairly stable. At the order statistics of 780 the tail index is seen to be stable, which gives a threshold of 36.9865.

![Hill Plot](image)

Table 3 shows the parameter estimates of Generalized Pareto Distribution using maximum likelihood estimation. The confidence interval of $\zeta$ contains zero, thus, the exponential distribution fits the data well.

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
<th>$\sigma$</th>
<th>$\zeta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence Interval (CI) (95%)</td>
<td>48.829</td>
<td>-0.114</td>
</tr>
<tr>
<td>Parameter Standard Errors</td>
<td>2.328</td>
<td>0.032</td>
</tr>
</tbody>
</table>

Table 4 shows the predicted maximum rainfall return level (in mm) for the return periods of 10, 20, 50 and 100 (in years) along with their respective 95% confidence interval. It can be seen from the results that increase in return periods leads to a corresponding increase in return levels.

<table>
<thead>
<tr>
<th>Return Periods</th>
<th>10 Years</th>
<th>20 Years</th>
<th>50 Years</th>
<th>100 Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Return Levels</td>
<td>286.048</td>
<td>299.742</td>
<td>316.267</td>
<td>327.673</td>
</tr>
<tr>
<td>CI (95%)</td>
<td>(249.488, 322.608)</td>
<td>(257.5116, 341.9725)</td>
<td>(266.354, 366.179)</td>
<td>(271.877, 383.470)</td>
</tr>
</tbody>
</table>

Comparing the return levels for the GEVD and GPD we observe that the GPD gives higher return levels for 10 and 20 years compared to GEVD. However, for higher return periods 50 and 100 years it can be seen that the GEVD gives higher return levels compared to the GPD.
The model diagnostic plot for GPD looks reasonable for rainfall data in Kenya as it can be seen from Figure 8. This means that the GPD is a good fit to the block maxima.

Table 5. Mann Kendall Test for Trends

<table>
<thead>
<tr>
<th></th>
<th>Kendall (τ)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kenya</td>
<td>-0.0819</td>
<td>0.19346</td>
</tr>
</tbody>
</table>

Trends in the annual maximum rainfall data for Kenya was performed using the non-parametric Mann-Kendall test. The Kendall value (τ) is a correlation coefficient which lies in the interval, 0 ≤ τ ≤ 1 and denotes the relative strength of the trend. T-test is performed to show the probability of the trend occurring by chance, given as (p-value). If the p-value is less than 0.05, a trend is considered significant at 5% level of significance. Since, p-value = 0.19346 > 0.05, it can be concluded that the trend is not significant.

5. Conclusion

This research demonstrates the application and significance of EVT at describing extreme rainfall events in Kenya. The GEVD and GPD models are considered for maximum annual rainfall data in Kenya from 1901 to 2016. The model parameters were estimated using the Maximum Likelihood Estimation, block maxima approach was used to fit the Generalized Extreme Value Distribution while the Peak Over Threshold method was used to fit the Generalized Pareto Distribution.

Our findings reveal that the Gumbel distribution is the optimal model from the GEV family for the annual maximums of monthly rainfall data while the Exponential distribution gave the optimal model for the monthly rainfall data over the threshold value of 36.9865. It was found that increase in return periods leads to a corresponding increase in return levels. When comparing the return levels for the GEVD and GPD, our results show that the GPD gives higher return levels for 10 and 20 years compared to GEVD. However, for higher return periods 50 and 100 years, the GEVD gives higher return levels compared to the GPD. The model diagnostics showed that the models were reasonable for modelling the rainfall data. Performing the Mann Kendall test for trends indicates that the trend in annual maxima over the period of 1901 to 2016 in Kenya has been insignificant.
This study will help decision makers in Kenya with knowledge about extreme rainfall events in the return periods considered, to enable them make appropriate decisions to reduce damage to crops, infrastructure and lives that is caused by extreme rainfall. As climate change persists, continuous preparedness and adaptation measures are essential for the Kenyan communities resilience. Thus, this research will be useful in coming up with flood risk early warning, management, preparedness, response and mitigation. Although Kenya is heading in the right direction in terms of creating an enabling environment to respond to climate change, there is still much that needs to be done. In line with Kenya’s vision 2030, there is need to fully implement the National Climate Change Action Plan. However, future studies can model and predict extreme rainfall in Kenya with respect to specific regions in the country. Also, modelling both extreme rainfall and temperature in Kenya is a possible research direction.

References

Copyrights
Copyright for this article is retained by the author(s), with first publication rights granted to the journal.
This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).