

## Research Article

# On Compound Distributions for Natural Disaster Modelling in Kenya

**Antony Rono, Carolyne Ogutu , and Patrick Weke**

*School of Mathematics, University of Nairobi, Nairobi, Kenya*

Correspondence should be addressed to Carolyne Ogutu; [cogutu@uonbi.ac.ke](mailto:cogutu@uonbi.ac.ke)

Received 10 February 2020; Accepted 4 April 2020; Published 25 April 2020

Academic Editor: Hernando Quevedo

Copyright © 2020 Antony Rono et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Kenyan communities are exposed to natural disasters by an amalgamation of factors such as poverty, aridity, and settlements in areas susceptible to natural disasters or in areas with poor infrastructure. This is expected to increase due to the effects of climate change. In an attempt to explain some of these variabilities, we model the extreme damages from natural disasters in Kenya by developing a compound distribution that takes into account both the frequency and the severity of the extreme events. The resulting distribution is based on a threshold model and compound extreme value distribution. For frequency of events exceeding a threshold of 150,000, we found that it follows a negative binomial distribution, while severity of exceedance follows a generalized Pareto distribution. This distribution fits the data well and is found to be a better model for natural disasters in Kenya than the traditional extreme value threshold model.

## 1. Introduction

Kenya has continued to face an increasing vulnerability to natural disaster risk. The communities are exposed to natural disasters by an amalgamation of factors such as poverty, aridity, and settlements in areas susceptible to natural disasters or in areas with poor infrastructure. These factors coupled with naturally occurring hazards, which are currently being propelled by climate change, pose an extreme threat to the Kenyan society. As of 2018, a total of 113 natural disaster events had been recorded in the last six decades, affecting approximately 62 million people and resulting to 6,900 deaths. The total damage from these occurrences is estimated to be 609 million US dollars. Most of the natural disaster events are weather-related, with almost 70% of the landmass being affected by drought and a total of 55 flooding events recorded in various parts of the country.

As with most parts of the world, natural disasters are expected to increase in the future due to climate change. World Bank projects the number of drought days in many parts of the world to increase by more than 20% by 2080, and the number of people exposed to drought could increase by 9 – 17% in 2030 and 50 – 90% in 2080. The number of

people exposed to river floods could also increase by 4 – 15% in 2030 and 12 – 29% in 2080 [1]. However, the effects would be felt mostly in the less-developed countries like Kenya. A study conducted by the Center of Research and Epidemiology of Disasters (CRED) found that people living in the poorer nations are six times more likely to be injured, to lose their homes, be displaced, or require emergency assistance than those in the wealthier nations. They are also seven times more likely to die from a natural disaster than those in richer nations. Therefore, with the increasing frequency and potential damage of natural disasters in Kenya, there is an urgent need to understand the characteristics of such events in the country.

In modelling extremal and rare events, extreme value theorem (EVT) emerges as a vital tool to model such risks. There are two main approaches in EVT: classical EVT (block-maxima) and excess over threshold (EOT) [2]. Several studies have been conducted on EVT and its application to model natural disasters including Engeland et al. [3] who used the block maxima method to model hydrological floods and droughts in the USA and Jindrova and Pacakova [4] who used EOT to model historical natural catastrophe losses in the USA. In both studies, they found

the EVT distributions to be a good fit for the data. However, such models only consider the loss severity distribution without explicitly considering that of the frequency of occurrence. As a result, Tebfu and Fengshi proposed the compound extreme value distribution (CEVD) and successfully used it to model typhoon in South China. They assume that the frequency of occurrence is Poisson distributed while that of the severity follows an extreme value distribution. Other compound extreme value distributions have since been developed, including Poisson-Weibull CEVD [5], Poisson-generalized Pareto CEVD [6], and geometric-Gumbel CEVD [7].

In this paper, we seek to model the extreme damages from natural disasters in Kenya by developing a compound distribution that takes into account both frequency and severity of the extreme events. The rest of the paper is structured as follows: Section 2 is the methodology where we will discuss extreme value theory and compound distributions, which we will use to develop a model to model natural disasters in Kenya. The data analysis results and discussion are then presented in Section 3, and the study concludes in Section 4.

## 2. Methodology

**2.1. Extreme Value Theory.** The cornerstone of extreme value theory is the study of the stochastic behaviour of the maximum (or minimum) of a sequence of random variables. Define

$$M_n = \max\{Y_1, \dots, Y_n\}, \tag{1}$$

where  $\{Y_1, \dots, Y_n\}$  is a sequence of independent random variables with a common distribution function  $F$  and  $M_n$  represents the maxima (minima) of the observed process over  $n$  blocks or time units. If  $F$  is known, the distribution of  $M_n$  is

$$P\{M_n \leq y\} = P\{Y_1 \leq y, \dots, Y_n \leq y\} = \{F(y)\}^n. \tag{2}$$

However,  $F$  is usually unknown in practice and will have to be estimated from the data. This poses a problem since a small error in the estimation of  $F$  can lead to large disparities for  $F^n\{y\}$ . An alternate approach is to model  $F^n\{y\}$  through asymptotic theory of  $M_n$ , where we study the behaviour of  $F^n\{y\}$  as  $n \rightarrow \infty$ . Since  $F(y) < 1$  for  $y < y_{\text{sup}}$ , where  $y_{\text{sup}}$  is the upper end-point of  $F$ , we have  $F^n\{y\} \rightarrow 0$  as  $n \rightarrow \infty$ . We can remove the degeneracy problem by allowing some linear renormalization of  $M_n$ . Consider a linear renormalization:

$$\widehat{M}_n = \frac{M_n - d_n}{c_n}, \tag{3}$$

where  $\{c_n\}$  and  $\{d_n\}$  are sequences of constants with  $c_n > 0$ . Under a suitable choice of  $c_n$  and  $d_n$ , the distribution of  $M_n$  can be stabilised leading to “extremal types theorem” [2]:

**Theorem 1** (extremal types theorem). *For a nondegenerate distribution function,  $G$ , if there exist sequences of constants  $\{c_n > 0\}$  and  $\{d_n\}$ , as  $n \rightarrow \infty$ , such that*

$$P\left(\frac{M_n - d_n}{c_n} \leq y\right) \rightarrow G(y), \tag{4}$$

then  $G$  belongs to one of the following families:

$$\begin{aligned} \text{I Gumbel: } G(y) &= \exp\left\{-\exp\left[-\left(\frac{y-d}{c}\right)\right]\right\}, \quad -\infty < y < \infty, \\ \text{II Frechet: } G(y) &= \begin{cases} 0, & y \leq d, \\ \exp\left\{-\left(\frac{y-d}{c}\right)^{-\kappa}\right\}, & y > d, \end{cases} \\ \text{III Weibull: } G(y) &= \begin{cases} \exp\left\{-\left[-\left(\frac{y-d}{c}\right)\right]^\kappa\right\}, & y < d, \\ 1, & y \geq d, \end{cases} \end{aligned} \tag{5}$$

For  $c > 0$  and  $d \in \mathbb{R}$ .

The three classes of distributions are called extreme value distributions, with type I (Gumbel), type II (Frechet), and type III (Weibull), respectively. von Mises [8] and Jenkinson [9] combined the three types of extreme value distributions leading to the generalized extreme value distribution (GEV).

**Theorem 2.** *If there exist sequences of constants  $c_n$  and  $d_n$  such that*

$$P\left(\frac{M_n - d_n}{c_n} \leq y\right) \xrightarrow{n \rightarrow \infty} G(y), \tag{6}$$

where  $G$  is a nondegenerate distribution function, then  $G$  is a member of the GEVD family:

$$G(y) = \exp\left\{-\left[1 + \zeta\left(\frac{y-\nu}{\sigma}\right)\right]^{-1/\zeta}\right\}, \tag{7}$$

defined on  $y$  such that  $1 + (\zeta((y - \nu)/\sigma)) > 0$  and with parameters: scale  $\sigma > 0$ , location  $\nu \in \mathbb{R}$ , and scale  $\zeta \in \mathbb{R}$ .

Theorem 2 suggests that regardless of the population distribution of  $M_n$ , if a nondegenerate limit can be obtained by linear renormalization, then the limit distribution will be the GEV distribution. This approach is, however, inefficient in terms of data usage since only the maximum within each time period is used for modelling. An alternative approach is excess over threshold, where all the data above, some sufficiently high threshold is used for modelling.

**Theorem 3.** *Given a set of independent and identically distributed random variables  $Z_1, \dots, Z_n$ , with a common distribution function,  $F$ , the conditional excess distribution function,  $F_\nu = P(Z - \nu | Z > \nu)$ , of a random variable  $Z$  above a high threshold  $\nu$  can be approximated by*

$$F_\nu(z) \approx \begin{cases} 1 - \left(1 + \frac{\zeta(z-\nu)}{\sigma}\right)^{-1/\zeta}, & \text{for } \zeta \neq 0, \\ 1 - \exp\left(-\frac{z-\nu}{\sigma}\right), & \text{for } \zeta = 0, \end{cases} \tag{8}$$

for  $z > \nu$  and  $(1 + (\zeta(z - \nu) / \sigma)) > 0$  and parameters scale  $\sigma > 0$  and shape  $\zeta \in \mathbb{R}$ .

Equation (8) is the generalized Pareto distribution. So, for  $Z > v$ , the unconditional distribution function of the exceedance,  $z - v$ , is

$$G(z) = P(Z - v \leq z) = 1 - a \left( 1 + \left( \frac{\zeta(z-v)}{\sigma} \right) \right)^{-(1/\zeta)}, \quad \text{for } \zeta \neq 0, \tag{9}$$

where  $a = P(X > v)$  is treated as a parameter to be estimated. The density function of the exceedances can be proved to be

$$g(x) = \frac{d}{dx} G(x) = \frac{a}{\sigma} \left( 1 + \frac{\zeta(x-v)}{\sigma} \right)^{-(1+1/\zeta)}, \quad \text{for } \zeta \neq 0. \tag{10}$$

**2.2. Compound Distributions.** Let  $\{X_1, \dots, X_N\}$  be a sequence of independent and identically distributed random variables with a common distribution function  $Q$ . Also, let  $N$  be a counting random variable with probability function  $P_n = P(N = n)$  and independent of  $X_i$ . A compound distribution is the distribution of the random sum

$$S_N = X_1 + \dots + X_N = \sum_{i=1}^N X_i. \tag{11}$$

The distribution function of  $S_N$  is given by

$$F(x) = P(S_N \leq x) = \sum_{n=1}^N Q^{n*}(x) P_n, \tag{12}$$

where  $Q^{n*}(x)$  is the  $n$ -th fold convolution power of  $Q$ .

**Definition 1** (convolution). The convolution of two density functions  $Q_X(\cdot)$  and  $Q_Z(\cdot)$  on the positive real line is

$$Q_X(x) * Q_Z(s) = Q_{X+Z}(s) = \int_{-\infty}^{\infty} Q_Z(s-x) q_X(x) dx, \tag{13}$$

where  $q_X(x) = (d/dx)Q_X(x)$  and  $S = X + Z$ .

**Remark 1.** The distribution function of a sum of independent and identically distributed random variables with common cdf  $Q$  is the  $n$ -th fold convolution power of  $Q$ :

$$Q(X_1 + X_2 + \dots + X_n) = Q * Q * \dots * Q = Q^{*n}. \tag{14}$$

Compound distributions are used to model total losses in a portfolio or a group of insurance policies. In this context,  $S$  denotes the total losses,  $N$  is the number of losses, and  $X_i$  is the size of the  $i$ -th independent loss. However, in extreme value analysis, we are interested in the tails of distributions, i.e., the distribution of the largest losses, and how they affect the total losses. A natural class of large loss distribution is given by the subexponential family, which is a subclass of the heavy-tailed distribution. By definition, heavy-tailed distributions have heavier tails than the exponential distribution, and their tails decay like a power function. All commonly used heavy-tailed distributions are subexponential.

**Definition 2** (subexponential distribution). A distribution function  $Q$  with  $Q(x) = 0$  for  $x < 0$  is called subexponential if

$$\lim_{x \rightarrow \infty} \frac{1 - Q^{*n}(x)}{1 - Q(x)} = n, \quad \text{for } n \geq 2. \tag{15}$$

This can be interpreted as for a sum of  $n$  independent random variables,  $X_1, \dots, X_n$ , with common distribution  $Q$ :

$$P\{X_1 + \dots + X_n > x\} \approx P\{\max(X_1, \dots, X_n) > x\}, \text{ as } x \rightarrow \infty. \tag{16}$$

Equation (16) is usually referred to as catastrophe principle. We can also express it as

$$\lim_{x \rightarrow \infty} \frac{P\{M_n > x\}}{P\{S_n > x\}} = 1, \tag{17}$$

where  $M_n = \max(X_1, \dots, X_n)$  and  $S_n = \sum_{i=1}^n X_i$ . This relation implies that the total losses are directly dependent on the largest losses. In other words, the sum of  $n$  losses gets large if and only if its maximum gets large.

Therefore, assuming that the underlying distribution is subexponential (heavy-tailed), we can write the distribution function of the compound distribution in terms of the maximum values:

$$F(x) = 1 - P(S_n > x) \approx 1 - P(M_n > x) = \sum_{n=1}^N P_n G^n(x). \tag{18}$$

For  $n = 0, 1, 2, \dots$ ,

$$F(x) = \sum_{n=0}^{\infty} P_n G^n(x), \tag{19}$$

where  $G(x)$  is the distribution of the maxima.

Equation (19) is called a compound extreme value distribution (CEVD). Unlike EVT models, which do not consider the distribution of the frequency of occurrence in detail, CEVD assumes that the frequencies of extreme events are random variables. Theorem 4 presents the CEVD as proposed by Liu and Ma [10].

**Theorem 4.** Let  $Y$  and  $Z$  be random variables with cumulative distribution functions  $G(x)$  and  $T(x)$ , respectively. Let  $N$  be another random variable independent of  $Y$  and  $Z$ , with probability function

$$P(N = n) = P_n, n = 0, 1, 2, \dots, \tag{20}$$

$$\sum_{n=0}^{\infty} P_n = 1.$$

Define a random variable  $X$  as

$$X = \begin{cases} Z, & N = 0, \\ \max_{1 \leq i \leq N} Y_i, & N \geq 1, \end{cases} \tag{21}$$

where  $Y_i$  is the  $i$ -th independent observation of  $Y$ .

Then, the distribution function of  $X$  is

TABLE 1: Descriptive statistics of the occurrence and severity of natural disaster data.

Statistic	Annual occurrence	Annual severity
N	55	55
Total	112	62,160,910
Mean	2	1,130,198
Std. dev	2.26	3,494,302
Minimum	0	0
1st quantile	0	0
Median	1	15,000
3rd quantile	4	252,718
Maximum	9	23,331,469
Interquantile range	4	252,718

$$F(x) = \sum_{n=0}^{\infty} P_n [G(x)^n] - P_o [1 - T(x)]. \quad (22)$$

We can express equation (22) as

$$F(x) = F_0(x) - \varepsilon(x), \quad (23)$$

where  $F_0(x) = \sum_{n=0}^{\infty} P_n [G(x)]^n$  and  $\varepsilon(x) = P_o [1 - T(x)]$ .

Since we are interested in the upper limits of the distribution, we can take  $F_0(x)$ , ignoring  $\varepsilon(x)$ , to be the value of  $F(x)$ . The full principle of proof is discussed by Liu and Ma [10]. It can be shown that  $F_0(x)$  is monotone nondecreasing and right continuous [6]. It satisfies  $F_0(\infty) = 1$  and  $F_0(-\infty) = P_o$ . It is, however, worth noting that  $F_0(x)$  is not a distribution function when  $P_o > 0$ . We can modify the probability  $P_o$  to make  $F_0(x)(-\infty) = P_o = 0$ , but since we are interested in the upper limits of  $F_0(x)$ , we will not consider the details of the modification.

As a result of the above discussion, we can formally define CEVD as follows.

*Definition 3.* Given a random variable  $N$  with probability mass function  $P(N = n) = P_n$  for  $n = 0, 1, \dots$  and a set of independent and identically distributed random variables  $Y_1, \dots, Y_N$  with a common distribution function  $G(x)$  and assumed to be independent of  $N$ , the compound extreme value distribution comprising of  $N$  and the running maximum of  $Y$ ,  $X = \max_{1 \leq i \leq N} Y_i$ , is defined as

$$F_0(x) = \sum_{n=0}^{\infty} P_n [G(x)]^n, \quad \text{for } x \in \mathbb{R}. \quad (24)$$

Using Definition 3, we can now develop a distribution to model natural disasters in Kenya. Let  $\{X_1, \dots, X_N\}$  be a sequence of independent and identically distributed random variables. For a sufficiently high threshold,  $v$ , the observations that exceed  $v$ ,  $x - v$ , are called exceedances. Denote the number of exceedances by  $N_v$ . Assume the distribution of  $N_v$  is a negative-binomial distribution with parameters  $\kappa > 0$  and  $0 < \rho < 1$  such that

$$P(N_v = n) = P_n = \binom{\kappa + n - 1}{n} \rho^n (1 - \rho)^\kappa, \quad \text{for } n = 0, 1, 2, \dots \quad (25)$$

Replacing equation (25) into CEVD formula (24),

$$F(x) = \sum_{n=0}^{\infty} \binom{\kappa + n - 1}{n} \rho^n (1 - \rho)^\kappa (G(x))^n = \left\{ \frac{1 - \rho}{1 - \rho G(x)} \right\}^\kappa, \quad x > 0, \quad (26)$$

where  $G(x)$  is the cumulative distribution function of the exceedances given in equation (9). Equation (26) is then the distribution function of what we will call the negative binomial-generalized Pareto compound extreme value distribution (NB-GP CEVD).

To estimate the parameters, we will use maximum likelihood estimation (MLE). From distribution function (26), we get the density function:

$$f(x) = \frac{d}{dx} F(x) = \frac{a\kappa\rho(1-\rho)^\kappa}{\sigma} \cdot \left( 1 - \rho \left[ 1 - a \left( 1 + \frac{\zeta(x-v)}{\sigma} \right)^{-(1/\zeta)} \right] \right)^{-(\kappa+1)} \cdot \left( 1 + \frac{\zeta(x-v)}{\sigma} \right)^{-(1+(1/\zeta))}. \quad (27)$$

The log-likelihood is then

$$\ln L = \sum_{i=1}^n \ln f(x_i). \quad (28)$$

We can compute the MLE estimates  $\hat{a}$ ,  $\hat{\kappa}$ ,  $\hat{\rho}$ ,  $\hat{\sigma}$ , and  $\hat{\zeta}$  by equating  $(\partial \ln L / \partial a)$ ,  $(\partial \ln L / \partial \kappa)$ ,  $(\partial \ln L / \partial \rho)$ ,  $(\partial \ln L / \partial \sigma)$ ,  $(\partial \ln L / \partial \zeta)$  all to zero.

### 3. Data Analysis, Results, and Discussion

We use data for all the natural disasters recorded in Kenya in the period 1964 – 2018 that were obtained from the CRED database. The severity of natural disasters is quantified in terms of the total number of people affected on an annual basis, which we deemed to be more reliable than the total damage in monetary terms. Table 1 shows the descriptive statistics for both the annual occurrence and the impacts. In summary, the minimum number of disaster occurrence and the resulting severity are zero, which corresponds to those years where no natural disaster event was recorded. On the contrary, the maximum number is 9, and the number of people affected is 23,331,469. The mean is 2 for the number of occurrences and 1,130,198 for the

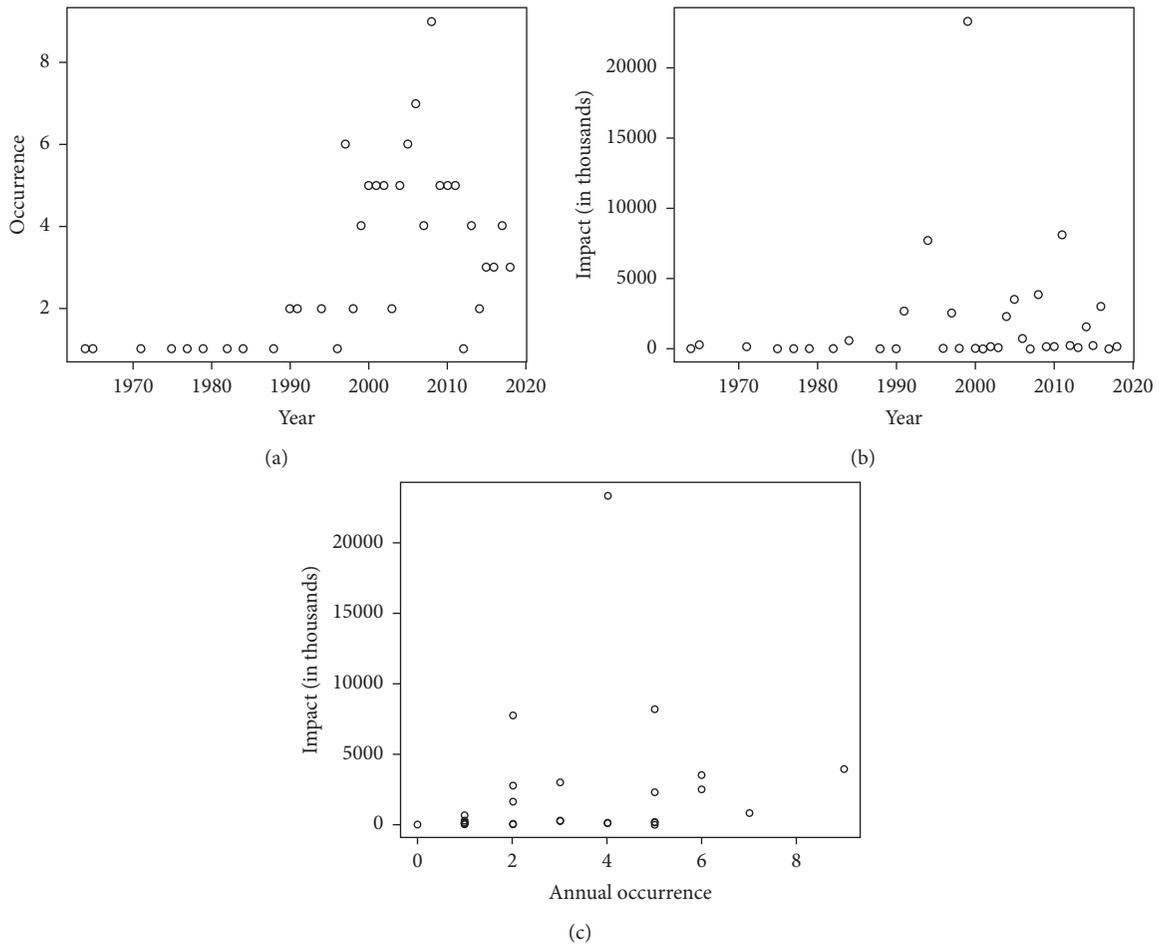


FIGURE 1: Scatter plots to check for independence. (a) Annual occurrence against the corresponding years. (b) Annual severity against the corresponding years. (c) Annual occurrence against annual severity.

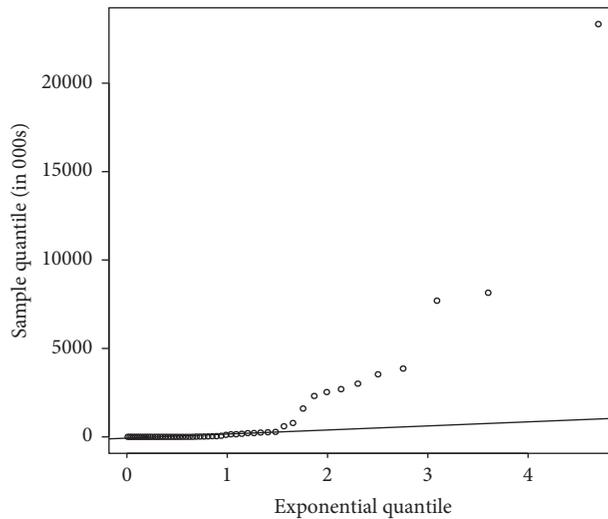


FIGURE 2: Exponential Q-Q plot.

severity. We can also observe that the mean is greater than the median for both variables, indicating that the data are right-skewed.

We will start with exploratory analysis of the natural disaster data. The scatterplots in Figure 1 show that there are no serious violations of the independence assumptions

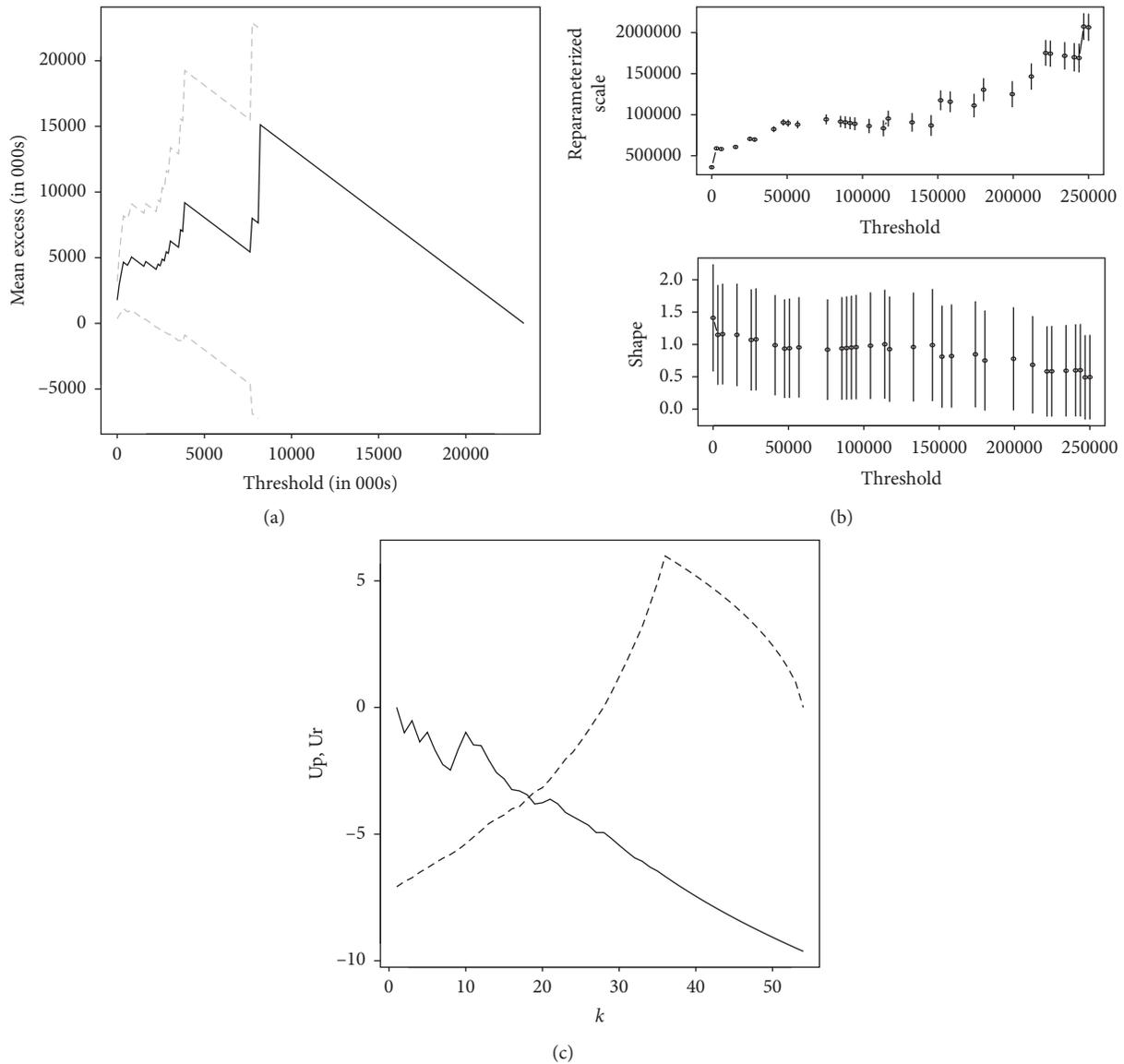


FIGURE 3: Threshold selection tools. (a) Mean residual life plot. (b) Parameter stability plot. (c) Gertensgarbe plot. Up and Ur are the series of differences from start to end and end to start, respectively, and  $k$  is the observation number in the order statistics.

among the two variables. The exponential Q-Q plot in Figure 2 displays a convex departure from the straight line, indicating that the theoretical quantiles grow slower than the empirical quantiles. This suggests that the severity data are heavy-tailed.

**3.1. Threshold Selection.** We use three graphical tools to select an appropriate threshold. First, we plot the mean excesses for each value of 200 different thresholds across the whole dataset, against their corresponding thresholds, with a significance level of 5%. Figure 3(a) shows that the graph becomes linear right from the beginning, until around 8,000,000. This suggests a threshold of 0.

Next, we plot the maximum likelihood estimates of the GPD parameters at 80 different thresholds against their corresponding thresholds, together with 95% confidence

TABLE 2: Mann–Kendall test results.

Statistic	Value
$k_0$	19
$p$ value	0.00014
Threshold	150000
Tail index	0.48507

intervals. Figure 3(b) shows that the scale parameters become stable at around 50,000, while the reparameterized shape parameter becomes constant right from the start. The threshold is then 0 and 50,000.

Finally, we plot a Gertensgarbe plot, which involves plotting the series of differences  $\Delta_r = z_{(r)} - z_{(r+1)}$ ,  $r = 2, 3, \dots, n$ , of the order statistics,  $z_{(1)} \leq z_{(2)} \leq \dots \leq z_{(n)}$ , from the start to end and from the end to start. The cross

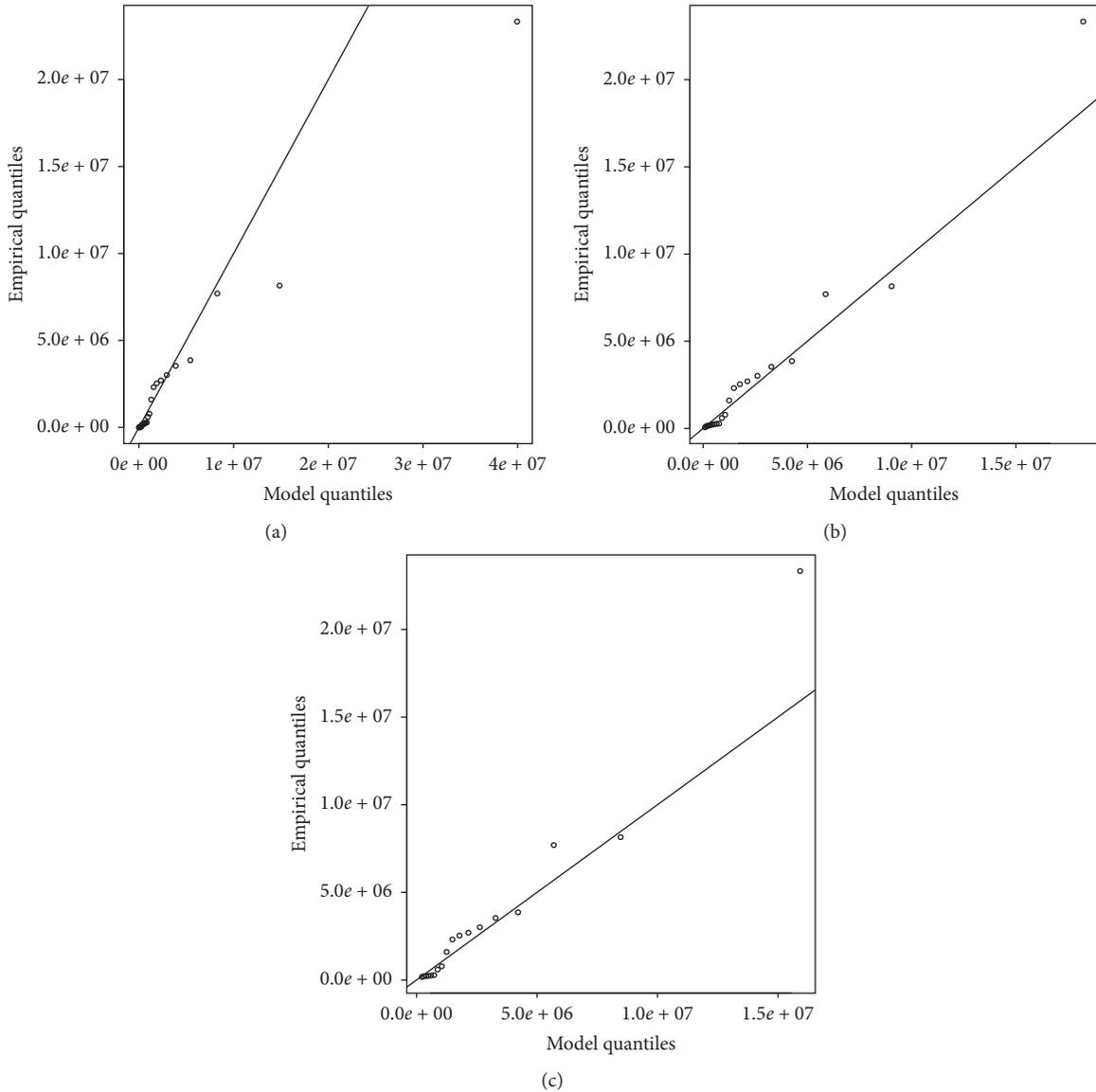


FIGURE 4: Q-Q plot of GPD at different thresholds. (a) Threshold = 0. (b) Threshold = 50,000. (c) Threshold = 150,000.

point of the two series (Figure 3(c)) is at the observation numbered  $k = 19$ , which corresponds to a threshold of 150,000. We also carry out the sequential version of Mann-Kendall test to contrast whether this point is the starting point of the extreme region. The results are indicated in Table 2. The null hypothesis that there is no change in the series of differences is rejected with a  $p$  value less than 0.001.

We will now investigate the goodness of fit of the GPD in each of the three threshold values. Figure 4 shows that GPD fits the data best at the 50,000 and 150,000 thresholds. To avoid violating the asymptotic arguments underlying the GPD, we choose the threshold to be 150,000.

3.2. *Fitting the Negative Binomial-Generalized Pareto Compound Extreme Value Distribution.* Given a threshold of 150,000, we will first investigate the fit of the negative

binomial distribution to the number of exceedances and the GPD to the exceedances. Tables 3 and 4 show that the  $p$  value is greater than 0.01 in both cases, indicating that the distributions are a good fit to their respective variables.

We can then fit the NB-GP CEVD to the data of natural disasters in Kenya. Table 5 shows the parameter estimates, and Table 6 shows the fit of the distribution.

The maximized value of the log-likelihood function is found to be  $-192.0693$ .

As observed in Table 6, the  $p$  values in both tests are greater than 0.01. Thus, we fail to reject the null hypothesis that natural disasters in Kenya follow a NB-GP CEVD at 1% level of significance. We can therefore conclude that the proposed distribution is a good fit for the data.

To assess the improvement achieved by using NB-GP CEVD instead of GPD, we investigate the quality of the proposed model relative to that of the GPD. This is done

TABLE 3: Goodness of fit of negative binomial to the distribution of the number of exceedances.

Name	Value
Chi-squared statistic	0.52
Degree of freedom	2.00
Chi-squared $p$ value	0.77

TABLE 4: Goodness of fit of GPD to the exceedances.

Hypothesis	$p$ value	$R$ -statistic
$H_0: X$ has a GPD	0.2142	0.9652

TABLE 5: Maximum likelihood estimates.

Parameter	Estimate	Standard error
$\hat{\kappa}$	9,000,000	—
$\hat{\rho}$	0.0000035	0.000006
$\hat{\sigma}$	786.42	—
$\hat{\zeta}$	2.1567	0.010162
$\hat{a}$	0.3455	0.018874

TABLE 6: Goodness-of-fit test results.

Test	Statistic	$p$ value
Kolmogorov–Smirnov	0.2387	0.2300
Anderson–Darling	2.3430	0.0607

TABLE 7: Comparison of the quality of models.

Distribution	BIC	AIC
Negative binomial-generalized Pareto	398.86	395.87
Generalized Pareto	574.53	572.75

using Bayesian information criterion (BIC) and alkaline information criterion (AIC).

Table 7 shows that both measures are smaller for the NB-GP CEVD as compared to those of GPD, suggesting that the former is a better model for natural disasters in Kenya.

#### 4. Conclusion

A compound distribution is developed to model the extreme damages from natural disasters in Kenya. Unlike the traditional extreme value theory models that only consider the severity of extreme events, the distribution proposed here captures both frequency and severity. The distribution is based on a threshold model and compound extreme value distribution, where the frequency of events exceeding a threshold of 150,000 is found to follow a negative binomial distribution, while the severity of the exceedance follows a generalized Pareto distribution. The exceedances are assumed to be independent, and the number of exceedances is also assumed to be independent of the severity. The distribution is shown to fit the data well and is found to be a

better model for natural disasters in Kenya than the traditional extreme value threshold model.

The proposed distribution can be an important tool to understand the risks associated with natural disasters in Kenya. This can be particularly useful to the country's disaster management bodies and other stakeholders to improve the existing disaster preparedness strategies, which will in turn reduce the negative economic and social impacts of such events.

#### Data Availability

The data used in this study is open source and available at [https://www.emdat.be/emdat\\_db/](https://www.emdat.be/emdat_db/).

#### Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

#### References

- [1] S. Hallegatte, M. Bangalore, L. Bonzanigo et al., *Shock Waves: Managing the Impacts of Climate Change on Poverty*, The World Bank, Washington, DC, USA, 2015.
- [2] S. Coles, J. Bawa, L. Trenner, and P. Dorazio, *An Introduction to Statistical Modeling of Extreme Values*, Vol. 208, Springer, Berlin, Germany, 2001.
- [3] K. r. Engeland, H. Hisdal, and A. Frigessi, "Practical extreme value modelling of hydrological floods and droughts: a case study," *Extremes*, vol. 7, no. 1, pp. 5–30, 2004.
- [4] P. Jindrová and V. Pacáková, "Modelling of extreme losses in natural disasters," *International Journal of Mathematical Models and Methods in Applied Sciences*, vol. 10, 2016.
- [5] T. F. Liu, "Long-term distribution of hurricane characteristics," in *Proceedings of the Offshore Technology Conference*, Houston, TX, USA, May 1982.
- [6] J. Liu, D. Shi, and X. Wu, "Estimation of Poisson-generalized pareto compound extreme value distribution by probability-weighted moments and empirical analysis," *Transactions of Tianjin University*, vol. 14, no. 1, pp. 50–54, 2008.
- [7] G. Qingsong and P. Wei, "Parameter estimation for geometric-gumbel compound extreme-value distribution based on the  $\pi$ -th quantiles of samples," in *Proceedings of the 2015 Conference on Informatization in Education, Management and Business (IEMB-15)*, September 2015.
- [8] R. von Mises, "La distribution de la plus grande de  $n$  valeurs," pp. 271–294, American Mathematical Society, Providence, RI, USA, 1954.
- [9] A. F. Jenkinson, "The frequency distribution of the annual maximum (or minimum) values of meteorological elements," *Quarterly Journal of the Royal Meteorological Society*, vol. 81, no. 348, pp. 158–171, 1955.
- [10] T. F. Liu and F. S. Ma, "Prediction of extreme wave heights and wind velocities," *Journal of the Waterway, Port, Coastal and Ocean Division*, vol. 106, no. 4, pp. 469–479, 1980.