Science World Journal Vol. 15(No 3) 2020
www.scienceworldjournal.org
ISSN 1597-6343
Published by Faculty of Science, Kaduna State University
https://doi.org/10.47514/swj/15.03.2020.010

Full Length Research Article

MODELING EXTREME RAINFALL IN KADUNA USING THE GENERALISED EXTREME VALUE DISTRIBUTION

Bako Sunday Samuel*, Agog Nathan Samuel*, Magdalene Peter*, Mubarak Abdullateef* and Anyam Gladys Kwaghdoo*

* Department of Mathematical Sciences, Kaduna State University, Nigeria.

*Corresponding Author's Email Address: bakosundays@gmail.com or bakosundays@kasu.edu.ng

ABSTRACT
An important statistical distribution use in modeling such extreme events is the generalized extreme value distribution while the generalized Pareto distribution is suitable in modeling threshold excesses of extreme values. In this study, monthly rainfall data from the Nigeria Meteorological Agency in Kaduna are fitted to the generalized extreme value distribution and for a suitable threshold of 251mm, threshold excesses were fitted to the generalized Pareto distribution and a return level computed for 25, 50 and 100 years return period respectively. The threshold excesses follow the Weibull distribution and are bounded above implying that there is a finite value which the maximum above the threshold cannot exceed. For the 25, 50 and 100 years return period, a return level of 350mm, 390mm and 490mm with probability of exceedances of 0.04, 0.02 and 0.01 respectively were observed. The result further show that with the increasing level of rainfall as return period increases, there is a high likelihood of monthly maximum rainfall increasing steadily over the years and this has great consequences on the environment. If this trend continues unchecked as a result of global warming, residents will continue to experience flood unless the government build more drainages and ensure that existing drainages are free from dirt to enhance proper channeling and free flow of water in the event of rainfall.

Keywords: Extreme Value Theory (EVT), Generalized Pareto Distribution (GPD), Peak over Threshold, Return level, Extreme rainfall

1. INTRODUCTION

Rainfall is among the most devastating weather phenomena and have significant environmental consequences as they are frequently followed by flash floods and sometimes accompanied by severe weather conditions such as lightning, strong surface winds, and intense vertical wind shear. An extreme event is a phenomenon with a very high or low value which has a low probability of occurrence but with devastating consequences (Mirza, 2002). The variability of rainfall and the pattern of extreme high or low precipitation are very important for disaster planning, as well as the economy of any State. Extreme rainfall are events of greater concern and one of the most devastating weather phenomena which have adverse effect on lives, property and economy of a nation. These events are frequently followed by floods and sometimes accompanied by severe weather such as hail, lightning, intense vertical wind shear and strong surface winds (Jones et al. 2004). In North-western part of Nigeria, Kaduna state in particular have witness some extreme high rainfall in recent time. The state has severally experienced incidence of flooding as a result of extreme downpour almost annually. Most notably, in 2003, flash flood caused inundation of huge areas which sacked over 5000 people out of their homes and submerged more than 30,000 houses (Ijigah and Akinyemi, 2015). It is therefore important to study and model these extremes periodical rainfall for the purpose of making accurate prediction of return level and to make future planning considering the disastrous impact these events may have.

Extreme value analysis is one of the most important statistical disciplines and has been applied in various fields; from financing to hydrology to atmospheric chemistry to climatology to financial econometrics. The important feature of extreme value analysis is the objective of assessing the external behavior of random variables and quantifying the stochastic behavior of maxima and minima of independent and identically distributed random variables (Coles 2001). Extreme value theory is unique as a statistical discipline in that it develops techniques and models for describing the unusual rather than the usual. Extreme Value Theory (EVT) is a unique statistical discipline for developing techniques and models for describing the extreme behavior (the tail of the distribution) rather than the average behavior of a system**.** The distinguishing feature of an extreme value analysis is that it provides a framework in which anticipated forces could be estimated using historical data (Coles, 2001). It seeks to asses from a given ordered sample the probability of events that are more extreme than any previously observed.

There are two commonly used approaches of identifying extreme behavior of a stationary dataset. The first method considers the maximum of the variable taking in successive periods called block maxima, for example months or years. These selected observations constitute the extreme events. The second method focuses on the values exceeding a given threshold. The block maxima approach is the traditional method used to analyze data with seasonality, for example rainfall data. (Coles, 2001). In the block maxima approach, historical datasets are divided into blocks of equal width and the maximum data value within each block is extracted for analysis. However, in the approach of threshold methods, data are used more efficiently and all extremes values are utilized. This method is the most adopted approach in recent applications of EVT given the aforementioned reasons.(Gilli and Kellezi, 2006; Jones et al., 2014). Both approaches can be characterized in terms of a Poisson process, which allows for simultaneously fitting of parameters concerning both the frequency and intensity of extreme events (Eric and Richard, 2016).

Several studies have been conducted using extreme value theory to model extreme rainfall (maxima), these has led to different approach in predicting the extreme events. Sun-hee (2016) proposed a general procedure for accessing the descriptive and predictive abilities of ten probabilities distribution model that have been used in extreme rainfall frequencies analyses. L-moment

Science World Journal Vol. 15(No 3) 2020
www.scienceworldjournal.org
ISSN 1597-6343
Published by Faculty of Science, Kaduna State University
https://doi.org/10.47514/swj/15.03.2020.010

and maximum likelihood method were used for parameter estimation for the distributions. It was concluded that three of the models were found to be the best choice for the selected daily, sub-daily and annual maximum rainfall which include GNO, PE3 and GEV models. Supari (2012) also conducted a study of extreme rainfall events and trend assessment in Java Island, Indonesia using data from 1981 – 2010 in 84 stations. A pre-analyses test was conducted on the datasets such as duplicated data check, spatial outliers check, missing value check and homogeneity test. A threshold of 100mm with 5 years return period was defined using generalized extreme value distribution. A study conducted by Maya (2014), carried out an analysis of summer (June, July, August) extreme rainfall events in five regions in England and Wales using daily rainfall data of 100 years (1900-2000). The GPD method was adopted in this study, a threshold in excess of 60mm daily rainfall was considered. Their results show uncertainty in return values for return periods c 50 years to be very high, which reduces confidence in the r 15 for higher return periods.

This study analyze extreme rainfall events within the past few years recorded by the Nigeria Meteorological Agency in Kaduna using the Generalized Extreme Value (GEV) distribution and generalized Pareto distribution. The objectives are; to fit an appropriate distribution for the tails of the distribution of rainfall using the generalized extreme value distribution, to model threshold exceedances using the Generalized Pareto Distribution (GPD) and to obtain the return level and exceedances probability of rainfall for given return periods.

The structure of this article is as follows; In Section 2, we explain the theoretical background of the generalized extreme value distribution, the extremal types theorem and the generalized Pareto distribution. In Section 3, we present the results and discussions of the analysis for the rainfall data while Section 4 presents the conclusions derived from the study.

2. Materials and Methods

For this analysis of precipitation time series, the extreme value analysis technique is used to analyses the data with the main focus of using generalized Pareto distribution approach to fit the excess threshold. To achieve the objectives of this research which include identifying extreme indices and excess threshold using the peak-over-threshold (POT) data selection method extracted from sample data series to produce the series of extreme values above a chosen high threshold and used them with the generalized Pareto probability distribution. The maximum likelihood (MLE) method is used for parameter estimation for the model inference. A return level estimation is further used to check model goodness of fit.

2.1 Extreme Value Theory

The theory which governs the extreme value theory (EVT) stems from the study of block maxima given a sample of independent random variables$X_1, X_2 \dots, X_n$with common distribution F. To understand the extreme behavior, then value of the maximum of the sample $M_n = \max\{X_1, \dots, X_n\}$is certainly required. The behavior of the mean of this observation has often been the focus of classical statistical theory. The discussion of the classical statistical theory is what underpins the central limit theorem. Let

$$M_n = Max\,\{X_1, X_2 \dots, X_n\} \qquad \text{......(1)}$$

Deriving $M_n$under the theory of distribution for all values of n taking z as the upper point of F, we have

$$\Pr\{M_n\} = Pr\,\{X_1 \le z, X_2 \le z \dots, X_n \le z\}$$
$$\Pr\{M_n\} = Pr\,\{X_1 \le z\} \times \Pr\{X_2 \le z\} \dots \times \Pr\{X_n \le z\}$$
$$\Pr\{M_n\} = Pr\,\{F(z)\}^n \qquad \text{........ (2)}$$

2.2 Extremal Types Theorem

From equation (2), the finite maxima M_n converges to the upper endpoint of $\{F(z)\}^n$. Now, let's look at the behavior of F^nas$n \to \infty$.First, F is said to be sum stable if there exist a renormalizing constant $\{a_n > 0\}$ and$\{b_n\}$, such that $P\left(\frac{M_n - b_n}{a_n} \le z\right) = P(X \le z)$. Given a non-degenerate distribution function G, if$X_i$is a distribution on a max stable G,then $P(M_n \le z) = G^n(z)$, therefore a distribution F is max stable if

$$G^n(a_n z + b_n) = G(z) \qquad \text{.............. (3)}$$

Now, for a distribution of a normalized $M_n$of a continuous distribution given that F converges to G as $n \to \infty$, then

$$P\left(\frac{M_n - b_n}{a_n} \le z\right) = F^n(a_n z + b_n) \to G(z) \qquad \text{............. (4)}$$

Therefore, if equation (4) holds for the normalizing sequence$a_n and b_n$, then we can say the distribution of the maximum M_n of a finite random variables $X_1, X_2 \dots, X_n$, (provided n is large) converges to a max stable distribution $G(z)$.

This non-degenerate distribution function $G(z)$. belongs to one of three families:

Gumbel: $G(z) = exp\left\{-\exp\left[-\left(\frac{z-b}{a}\right)\right]\right\}, \quad -\infty < z < \infty$

Frėchet: $G(z) = \begin{cases} 0, & z \le b; \\ exp\left\{-\left(\frac{z-b}{a}\right)^{\propto}\right\}, & z > b; \end{cases}$

(Weibull):$G(z) = \begin{cases} exp\left\{-\left[-\left(\frac{z-b}{a}\right)^{\propto}\right]\right\}, & z < b; \\ 0, & z \ge b; \end{cases}$

With *a> 0* and *α > 0* where *a, b* and *α* is the scale, location and shape parameter respectively. Collectively, the three classes of distribution are referred to as the extreme value distributions (Coles 2001).

2.3 Generalized Extreme Value Distribution

Extreme value theory is based on Extremal Types theorem which state that the limiting distribution of maxima is converging to one of the three distributions called Gumbel, Frechet and Weibull. Then we can say that, GEV distribution is a generalization of these three distributions. The three families of distribution, Gumbel, Frechet and Weibull having cumulative distribution function given by

$$F(x) = \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-\frac{1}{\varepsilon}}\right\}$$

defined for $\left\{ z : 1 + \frac{(z-\mu)}{\sigma} > 0 \right\}$, $-\infty < \mu < \infty, \sigma > 0, and -\infty < \xi < \infty$, where μ is the location parameter, σ is the scale parameter and ξ is the shape parameter. The ξ value determines the class of GEV distribution such that $\xi > 0$ corresponds to Frechet distribution, $\xi < 0$ correspond to Weibull distribution and $\xi = 0$ (as$n \to \infty$) correspond to Gumbel distribution. (Coles, 2001)

2.4 Generalized Pareto Distribution

The principal drawback to the classical GEV method is that in the selection of the extreme value only one value is selected per block size. This reduces the observations available for analysis. To increase the number of data for analysis, an alternative approach is to obtain "Peak-Over-Threshold (POT)" maxima, to produce a series of extreme values extracted from sample data series above a chosen (high) threshold, which is then modeled with the generalized Pareto distribution (GPD). This common approach to extreme value analysis is based on the exceedances over high thresholds. If $x_1, . . . x_n$are independent and identical (iid) random variables and *x* is the differences between the observations over the threshold and the threshold itself. The cumulative distribution function of the GPD distribution defined on $x > 0$ and $\left(1 + \frac{\xi x}{\tilde{\sigma}}\right) > 0$ is given by

$$H(x) = 1 - \left(1 + \frac{\xi x}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}}$$

where $\tilde{\sigma}$ and ξ are the two parameters of the GPD. $\tilde{\sigma}$ is the scale parameter, ξ is the shape parameter and $\tilde{\sigma} = \sigma + \xi(u - \mu)$ (u correspond to a suitable threshold value 'not a parameter). The parameters μ and σ are the scale and shape parameter just like in the case of GEV model. The value of ξ determine the behavior of the GPD just as in GEV. For $\xi < 0$ the distribution of exceedances has an upper boundary of $u - \frac{\tilde{\sigma}}{\xi}$, the distribution of excesses has no upper limit and $\xi = 0$ gives an exponential distribution with parameter $1/\tilde{\sigma}$. Although the values of ξ in GPD and GEV should be the same, however, it is difficult to get this equal value in the two methods, since GPD computes the exceedances values from a threshold and GEV takes into account the maxima values in blocks, therefore, this makes the number of observations for the extreme values for both approaches different. If the number of observations used for both approaches are almost similar then the value of ξ will get closer. In determining extreme events using GPD model, although, there is no universal choice of threshold for any given dataset, hence the threshold value, *u* must be set appropriately given the nature of the datasets. If the choice of threshold is too low, it will affect the asymptotic fundamental of the model, and on the other hand if it is too high, it will only generate a few excesses which will result to high standard error and variance of the estimated model parameter. To check the validity of a GPD model, graphical goodness-of-fit checks such as probability, quantile, return level and density plots are approaches used.

3. Results and Discussion

This section presents the results of analysis for monthly rainfall data sourced from the Nigeria Meteorological Agency in Kaduna. The extreme value analysis was carried out using the Peak-over excess threshold approach of the monthly rainfall data of Kaduna and discussion of the result obtained are based on the methodologies described above. The package, *texmex,* in R statistical software developed by Harry and Jane (2013) is use to conduct this analysis.

3.1 Extremal Index Estimation and Declustering

Firstly, we begin by looking at the threshold parameter stability plot for the extremal index and also the parameters of the GPD fitted to cluster maxima.

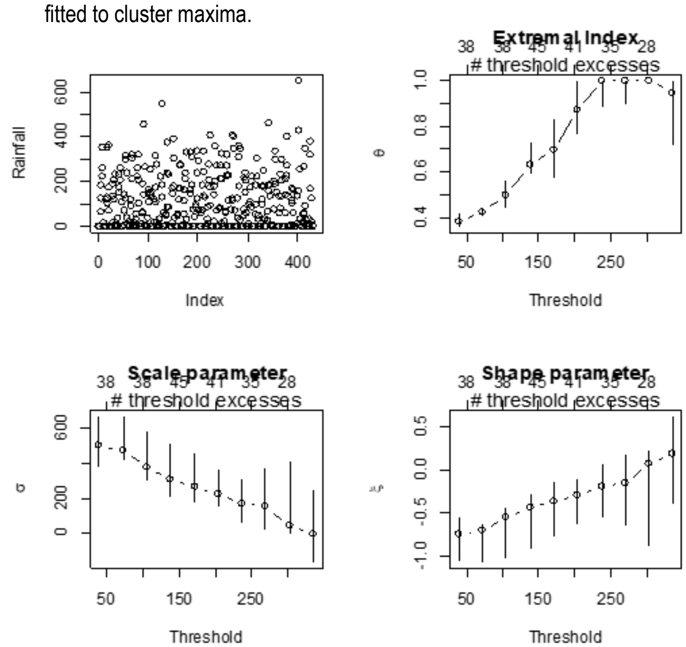

Fig 1.1: Parameter Stability plot and Extremal Index

It's evident from Fig 1.1 that while the estimated parameters of the GPD show that the parameter estimates (shape and scale) do not appear to vary considerably for even small values. This implies that they are relatively stable to the choice of threshold used for declustering. However, the estimates of the extremal index θ are not. The extremal index on the right tails tends to the maximum boundary (0≤ θ ≤1). Since we are interested in finding the clustering tendencies of the very highest values, we choose the value of threshold for which the estimates of all three

Science World Journal Vol. 15(No 3) 2020
www.scienceworldjournal.org
ISSN 1597-6343
Published by Faculty of Science, Kaduna State University
https://doi.org/10.47514/swj/15.03.2020.010

parameters are approximately constant. This leads to a threshold choice of 251 mm or above. A check of the goodness of fit of the choice of threshold for GPD model for cluster occurrence by using the Q-Q plot is described in below.

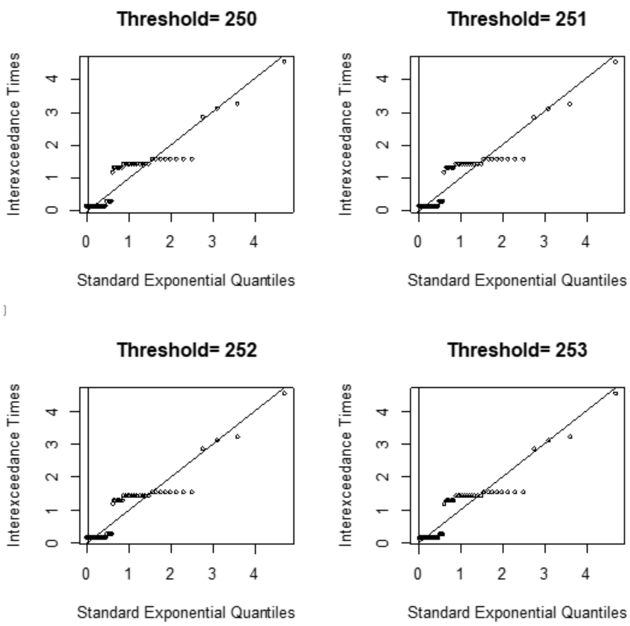

Fig 1.2: Q-Q plots of choices of threshold

We can be reassured by Fig 1.2that a threshold of 251 mm gives a good fit of the model to the rainfall data. For this choice of threshold, the estimate of the extremal index is 0.98, so that the average clusters size is$^{1}/_{0.98} = 1.02$. This is telling us that rainfall tends to be heavy on consecutive days but very rainy spells tend not to last longer than a day. We next decluster the sequence by using the automatic declustering method.

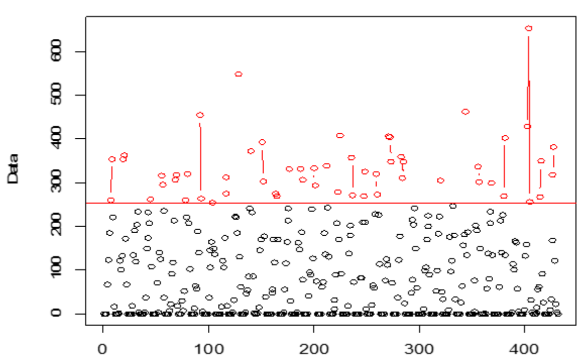

Fig 1.3: Plot Indicating Exceedances of the threshold

From Fig 1.3 above, given the original observation which is of length 432, there are 52 exceedances of the threshold 251 mm. However, we have identified serial dependence in the data, so the threshold excesses are not independent and in fact correspond to 37 approximately dependent clusters.

3.2 Fitting the Generalized Pareto distribution to cluster maxima

Fitting the GPD to the cluster maxima we estimate the parameters of the generalized Pareto distribution used to describe the conditional distribution of a cluster maximum given that it exceeds the threshold used for declustering. Table 1.1 shows the values for parameter estimate for the fitted cluster maxima model.

Table 1.1: Parameter Estimate for Cluster Maxima model

Parameter	*Estimate*	*Standard Error*	*AIC*
Scale (σ)	4.7971	0.2085	396.4056
Shape (ε)	-0.1915	0.1262	

Negative log-likelihood value of GPD -196.2028. Excess rates 0.0810

To further confirm that the threshold selected is good to use in fitting the GPD, diagnostic plots were plotted based on the cluster maximum. The below Figure 1.5 is the diagnostic plots to check the fit of the GPD.

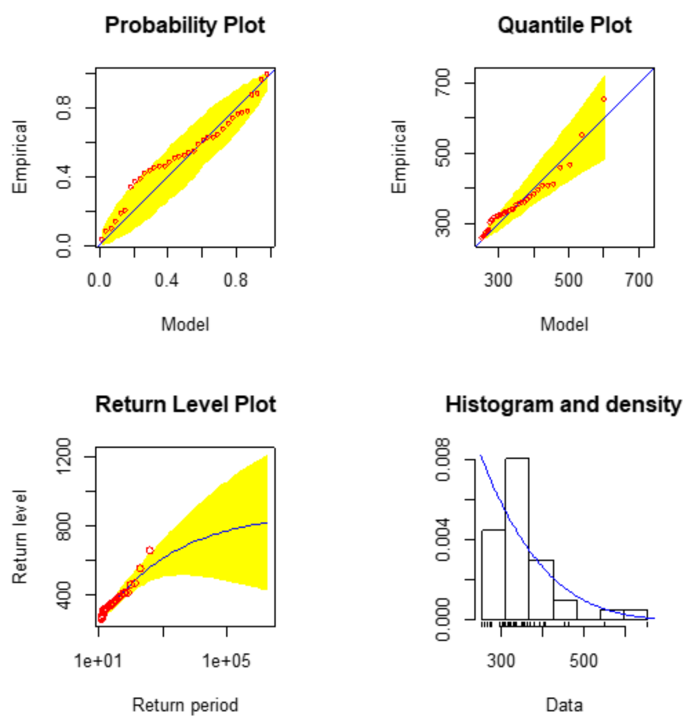

Fig 1.4: GPD Diagnostic plots

Fig 1.4 gives a good fit of the cluster maxima. The QQ-plot shows that all the points are approximately linearly distributed along the unit diagonal showing a good fit of the GPD. The plot gives access to the adequacy and validity of the fitted generalized Pareto distribution. Both the probability plot and the quantile plot depict a reasonable extreme value fit because, the probability plot is linear. The empirical density plot also affirms how adequate the GPD is in terms of modeling the data. Hence the diagnostic plots do not raise any alarm on the adequacy and validity of the generalized Pareto fitting of the maxima cluster. The distribution

Science World Journal Vol. 15(No 3) 2020
www.scienceworldjournal.org
ISSN 1597-6343
Published by Faculty of Science, Kaduna State University
https://doi.org/10.47514/swj/15.03.2020.010

of the rainfall data is Frechet distributed with heavy tailed that decays polynomially such that higher values of the maximum are obtained with greater probability than would be the case with a lighter tail.

3.3 Estimation of return levels

Application of EVT in rainfall studies are concerned about how well the mathematical theory can be applied to further answer questions relating to the probability that extreme rainfall will exceed a certain level in a given period which is referred to as the return level. Return levels in extremes explain the value of extreme events that occur on average once in a given period. The return level plot of the cluster maxima is given in Fig 1.5.

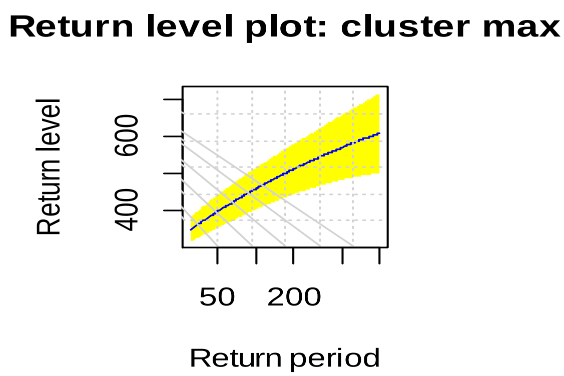

Fig 1.5: Parameter Estimate for Cluster Maxima model

Return levels can be computed from the GPD fitted to cluster maxima by using the estimated parameters of the GPD model and the chosen threshold of 251mm. Since estimated Shape (ε) parameter ε < 0, the generalized Pareto distribution belongs to the Weibull domain of attraction. The model can be express as

$$H(y) = 1 - \left(1 + \frac{-0.1915y}{4.7971}\right)^{\frac{1}{0.1915}}$$

4. Conclusion

The study uses the Generalized Extreme Value (GEV) and the Generalized Pareto Distribution (GPD) to model total monthly rainfall. The Parameter estimates for both the GEV and GPD were estimated using the maximum likelihood method. The threshold stability and extremal index plot was used to determine the appropriate threshold exceedances level, which was observed to be suitable at 251 mm. This choice of threshold was further reassured using the quantile-quantile plot. The distribution of the monthly total rainfall is Frechet distributed with heavy tailed that decays polynomially so that higher values of the maximum are obtained with greater probability than would be the case with a lighter tail. The threshold excesses follow the Weibull distribution and are bounded above implying that there is a finite value which the maximum above the threshold cannot exceed. For the 25, 50 and 100 years return period, a return level of 350mm, 390mm and 490mm with probability of exceedances of 0.04, 0.02 and 0.01 respectively were observed. If this trend continues unchecked as a result of global warming and increasing rainfall, residents will continue to experience flood unless the government build more drainages and ensure that existing drainages are free from dirt to enhance proper channeling and free flow of water in the event of extreme rainfall.

REFERENCES

Coles Stuart. An Introduction to Statistical Modelling of Extreme Values - London: Springer: Springer Series in Statistics, 2001. - ISBN 1-85233-459-2.

Eric G and Richard W.K. (2016). extRemes 2.0: An Extreme Value Analysis Package in R. *Journal of Statistical Software.* 72(8): pp 1-39

Gilli, M. and Kellezi, E. (2006). An Application of Extreme Value Theory for Measuring Financial Risk. *Computational Economics, 27*(1), pp 1-23.

Ijigah, E.A. and Akinyemi, T.A. (2015). Flood Disaster: An Empirical Survey of Causative Factors and Preventive Measures in Kaduna, Nigeria. *International Journal of Environment and Pollution Research*. 3: pp. 53-56.

Jones M. R., Blenkinsop S., Fowler H. J., and Kilsby C. G. (2014). Objective classification of extreme rainfall regions for the UK and updated estimates of trends in regional extreme rainfall. *International Journal of Climatology*, *34*(3), pp 751–765.

Jones, C., Waliser, D. E., Lau, K. M and Stern, W. (2004), Global Occurrences of Extreme Precipitation and the Madden–Julian Oscillation: Observations and Predictability. *Journal of Climate*, 17: Pp. 4575–4589.

Maya B. (2014). Extreme Rainfall and the Relationship to Summer Flooding Events in the UK. Available on http://www.met.reading.ac.uk/~sgs02rpa/REPORTS/Maya_Dissertation.pdf

Mirza, M. M. Q., (2002). Global warming and changes in the probability of occurrence of floods in Bangladesh and implications. *Global Environmental Change*. 12: pp127-138.

Sun-hee Lim (2016). Statistical Modelling of Extreme Rainfall Processes in British Colombia. Available on http://digitool.library.mcgill.ca/webclient/StreamGate?folder_id=0&dvs=1544385239611~330

Supari S, Janneke E, Edvin A and Sudibyakto S. (2012). Spatiotemporal Characteristics of Extreme Rainfall Events over Java Island, Indonesia. *The Indonesia Journal of Geography.* 44(1): pp 62-86.