International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2016): 79.57 | Impact Factor (2017): 7.296

Modelling Extreme Maximum Rainfall Using Generalized Extreme Value Distribution: Case Study Kigali City

Uwimana Oliver[1], Joseph K. Mung'atu

[1]Jomo Kenyatta University of Agriculture and Technology, Kenya
[2]Jomo Kenyatta University of Agriculture and Technology, Kenya

Abstract: ***Extreme high Rainfall is a global phenomenon that occurs almost in all landscapes causing significant damage such as flood that can destroy infrastructure, interrupt economic activities and retard development. Early detection of Extreme High Rainfall helps to implement strategies and measures, before they occur. This research used statistical techniques to build models that can be used to predict Extreme high rainfall in Rwanda. The Methodology of EVT (Extreme Value Theory) was applied to model monthly rainfall, the forecasted results using the best model were compared with the observed data to check whether the obtained results show reasonably good agreement with the reality, this will be done by comparing difference between empiric distribution function and fitted distribution function. The model Quantile plot and return revel plot will be used to test the goodness of the model.***

Keywords: Extreme Value Theory, Generalized Extreme Value Distribution, Maximum Likelihood Estimation, Kigali Rainfalls, likelihood ration test

1. Introduction

Extreme High rainfall has been dangerous issue in Rwanda especial in Kigali city, because it leads to floods and landslide. The aim of this work is to use Mathematical techniques to model extreme high rainfall in Rwanda, specifically in Kigali city.

Kigali city is capital of Rwanda and it is located at Rwanda's geographical heart, Kigali city has become Rwanda's most important business center also the main port of entry. Kigali city has experienced heavy rainfall events caused rapid surges in the flow of rivers and drainage systems leading to flood downstream. Although flooding has been experienced since 1960s in Kigali city, but its frequency has significantly increased since 2000s, and its impact have been great on Human development, properties, infrastructures as well as environment (Nsengiyumva, 2012). In 2006,27 percent of buildings in Kigali city were in flood –prone zones within the Nyabugogo River floodplain where vulnerable populations, infrastructures, and various economic activities were exposed to flash floods. The prevention from this devastation in Kigali city will depends on monthly information on maximum rainfall in Kigali. In this study we will investigate the change of maximum rainfall in order to forecast extreme high rainfall that can occur in Kigali city.

2. Problem Statement

Extreme high rainfall is dangerous hazard of the nature because extreme high rainfall can lead to flood and landslides, which can threaten human life, disrupt transport. Extreme high rainfall can severely affects both environment and Human lives. Kigali city has experienced extreme high rainfall that hit its several parts, some of the victims were drowned in flood water, others died after houses collapsed under the heavy rain. This can generally affect economic growth of the country due to the disturbance in agriculture sector that depend on the change in climate. This study aims to use Mathematical and statistical techniques to make model that can be used to forecast the extreme high rainfall in Kigali city

3. Justification

Kigali city has experienced extreme high rainfall, many households; infrastructures have remained vulnerable to extreme rainfall such as flood. Many researches on rain water harvesting, flood in Kigali city have been extensively covered. These have been produced controversial results. Therefore, this study has scope which is different from the previous studies because it attempts to focus in details on the extreme high rainfall model by using Extreme value theory (EVT). In addition to this extreme high rainfall model will be used to know what will happen in future and help to implement measure to prevent extreme high rainfall (flood).

4. Objectives of the Study

4.1 General Objective

The main objective of this research was to develop Generalized extreme value model that can be used to forecast the occurrence of extreme maximum rainfall in Kigali city.

4.2 Specifics Objectives

The specific objectives of this research are:

1. To determine Generalized Extreme Value Distribution (GEVD) that can be corresponded to forecasting extreme high rainfall in Kigali city
2. To forecast extreme high rainfall in Kigali city using generalized extreme value distribution (GEVD)
3. To test the Stationarity of time series data (Monthly rainfall)

5. Research Hypothesis

It hypothesized that:

1. Statistical analysis cannot help to understand better the Occurrence of Extreme high rainfall in Kigali city.
2. The forecasting model cannot be used effectively to forecast the extreme high rainfall in Kigali city Using data from meteorological agency.

6. Data and Research Methodology

Monthly maximum rainfall in Kigali city for the period 1996 to December 2017 was obtained from Rwanda Meteorological Agency; the data set contains 264 values of monthly maximum rainfall. Extreme values analysis was performed on this study by fitting the generalized extreme value distribution to the sample using method of maximum likelihood estimates (MLE)

We consider generalized extreme value distribution having the following families.

I. $G(x) = \text{Expo}\left\{-\exp\left[-\left(\frac{z-b}{a}\right)\right]\right\} \quad -\infty < z < \infty$

$$II. = \begin{cases} 0 & z \le b \\ G(z) = \exp\left\{\left(\frac{Z-b}{a}\right)^{\varepsilon}\right\}, & z>b \end{cases}$$

$$III = \begin{cases} G(z) = \text{Expo}\left\{-\left[-\left(\frac{z-b}{a}\right)^{\alpha}\right]\right\}, & z<b \\ 1 & z > b \end{cases}$$

The above three distributions are referred as extreme value distribution, with types II, I. and III known as Gumbel, Frechet and Weibull Families respectively.

6.1 Gumbel Distribution

Gumbel distribution or extreme value type I has two forms, one is based on the smallest extreme and other is based on the largest extreme those extreme are called minimum and maximum cases respectively. It has been used to predict earthquakes, floods and other natural disasters, as well as modelling operational risk in risk management and the life of products that quickly wear out after after certain age. Gumbel distribution has the following characteristics:

- The shape of the gumbel distribution is skewed to the left. The gumbel pdf has no shape parameter. It means that the gumbel pdf has only one shape, which does not change.
- The gumbel pdf has location parameter μ which is equal to the mode but differs from median and the mean. This is because the Gumbel distribution is not symmetrical about μ
- As μ decrease, the pdf is shifted to the left. As μ increases, the pdf is shifted to the right.
- As σ increases, the pdf spreads out and become shallower. As σ decreases, the pdf become taller and narrower. The general form of the pdf of the maximum Gumbel is almost the same (the difference is the change in sign for the first exponent). The change in sign means that the shape of the PDF is identical mirror image of the minimum Gumbel.

The following is the formula for probability density function for the Gumbel(minimum)

$$f(x,\mu,\sigma) = \frac{1}{\sigma}\exp\frac{x-\mu}{\sigma}\exp - \exp\frac{x-\mu}{\sigma}$$

6.2 Weibull Distribution

The Weibull distribution is also known as the extreme value type III distribution, it was originally developed to address the problems arising in Material sciences, it is widely used in many other areas thanks to its flexibility. Weibull model relates to minima (Smallest extreme value) it is characterized by two parameters, one is the shape parameter ε (dimensionless) and the other is the scale parameter σ

Probability density function of Weibull Distribution.

Let X be a random variable and let $\varepsilon, \sigma > 0$ then X is said to have a Weibull Distribution with parameters ε and σ if the PDF of X is:

$$f(x,\varepsilon,\sigma) = \left\{\frac{\sigma}{\varepsilon}\left(\frac{x}{\varepsilon}\right)^{\sigma-1}\right\}\exp\left(-\left(\frac{x}{\varepsilon}\right)^{\sigma}\right) \quad \text{where} \quad x \ge 0$$

and 0 else where

6.3 Frechet Distribution

Frechet Distribution or extreme value type II, it is slowly converges to 1 and has Three parameter: shape parameter ε, scale parameter σ, and location parameter μ. It is defined on interval μ $\propto$ in other words ,it is bounded(restricted) on the lower side.it has been applied to Extreme events such as Hydrology example for annually maximum one-day rainfalls and river discharges. Flood Analysis, human lifespans.

It has the following probability density function.

$$f(x) = \frac{\varepsilon}{\sigma}\left(\frac{\sigma}{x-\mu}\right)^{\varepsilon+1}\exp\left(-\left(\frac{\sigma}{x-\mu}\right)^{\varepsilon}\right)$$

With ε : shape parameter

μ : location parameter

σ : scalar parameter

6.4 Graphic Model Checking

In order to check whether GEVD will be a good fit for our data, the assessment can be done with reference to the observed data by using the following approach:

6.4.1 Probability Plot

A probability plot is a comparison of the empirical and fitted distribution functions.

The empirical distribution function evaluated in i-th ordered block maximum **$Z_{(i)}$,is** $\tilde{G}\left(z_{(i)}\right) = i/(1+m)$ and the fitted distribution function is the same point is $\hat{G}\left(z_{(i)}\right) = \exp\left\{-\left[1+\hat{\varepsilon}\left(\frac{z(i)-\hat{\mu}}{\hat{\sigma}}\right)\right]^{\frac{-1}{\varepsilon}}\right\}$ in order to have a good model it is necessary to have this Condition: $\tilde{G}\left(z_{(i)}\right) = \hat{G}\left(z_{(i)}\right)$ in practice plot of $(\tilde{G}\left(z_{(i)}\right), \hat{G}\left(z_{(i)}\right))$ i=1…..m should lie close to the first diagonal .but because both functions are bounded to approach 1 as the values of Z increase, the plot is least informative in this region. The following graph avoid this deficiency

6.4.2 Quantile Plot

The quantile plot is a representation of the point.

$$\left(\hat{G}^{-1}\left(\hat{i}/(1+m)\right), Z(i)\right) \quad i=1……..m$$

$$G^{-1}\left(\hat{i}/(1+m)\right) =$$

$$\hat{\mu} - \frac{\hat{\sigma}}{\hat{\varepsilon}}\left[1-\left(-\log\frac{i}{1+m}\right)^{-\varepsilon}\right] \quad i= 1……..m$$

In the ideal situation the plot should show a linear. Departure from linearity in quantile plot also indicate model failure.

6.4.3 Return Level Plot

The return level plot represents the point $\left(\log y_p, z_p\right)$ where 0<p<1

Confidence intervals are usually added to this plot to increase its information. The importance of return periods in engineering is due to the fact that the return period is used as a design criterion. Furthermore, to use this plot as a model diagnostic one, the empirical estimates of the return level function are also added. For suitable models the model based curve and empirical estimates should be in agreement.

7. Data Analysis

7.1 Introduction

By referring to what have seen in methodology, this study deal with the formulation of GEVD model that will fit the collected maximum monthly rainfall to forecast the occurrences of flood in Kigali city. To build GEVD model requires first check the stationarity of our data, after checking the stationarity we test to see if GEVD is a good fit for our data this procedure is called Discussion about Different Diagnostics Plots for GEV The main focus in this study is to determine the model that will fit our data knowing that this model will come from the generalized extreme value families of distribution that best fits our data this will be done by fitting our data in to each of this class distribution and observe the one that has the best fit with our data. We select the best distribution model on basis of likelihood ration test which is described below: Null Hypothesis is defined by ε =0 vs alternative hypothesis $\varepsilon \neq 0$ When the likelihood ration test is less than chi square critical value we fail to reject null hypothesis with means that Gumbel family will be good fit for our data. When the likelihood ration test is greater than chi square critical value we reject null hypothesis with means that our data belong to either Weibull distribution or freshet according to the sign of the shape parameter we select if our model fit in Weibull when shape parameter is equal to negative and to freshet when shape parameter is positive.

7.2 Stationarity Checking

It is important to study the stationarity of rainfall data, the sequence of random variables represents the series of Maximum rainfall, it must have constant properties through time. Random variables must be independent and identically distributed satisfying stationarity assumption in order to fit GEV Distribution to accomplish the assumption of stationarity of GEV family distribution, it necessary to test whether they exist trend in our data. We have two type of stationarity test: 1. Augmented Dickey-Fuller (ADF) test 2. Kwiatkowski phillips, Schmidt and shin(KPSS)test The aim of those test is to check if data (monthly rainfall) in time periods is stationary, to use ADF test we take Null Hypothesis H0 states that Data are not Stationary while H1 state that they are stationary. It is the opposite for KPSS (Hasan, Al..2012)

7.3 Model Checking

For our data in order to check whether our model will be a good fit ta, the assessment can be done with reference to the observed data by using the following approach:

Quantile plot: the plot should show a linear (almost all point should lie to the line). Departure from linearity in quantile plot also indicates model failure.

Probability plot: A probability plot is a comparison of the empirical and fitted distribution functions should lies closer to the first diagonal.

7.4 Statistical Description of the Data

The analysis is based on the data available in Rwanda Meteorological Services. the data have been recorded at Headquarter from 1996 up 2017.by considering available monthly rainfall data. the table below shows the statistical description of data.

Table 1: Statistical Description of Monthly Rainfall in Kigali from 1996 to 2017

Minimum	Max	First Quartile	Means	Median	3 Quartile
0	324.3	35.4	82.4	71.8	117.8

7.5 Parameter Estimation

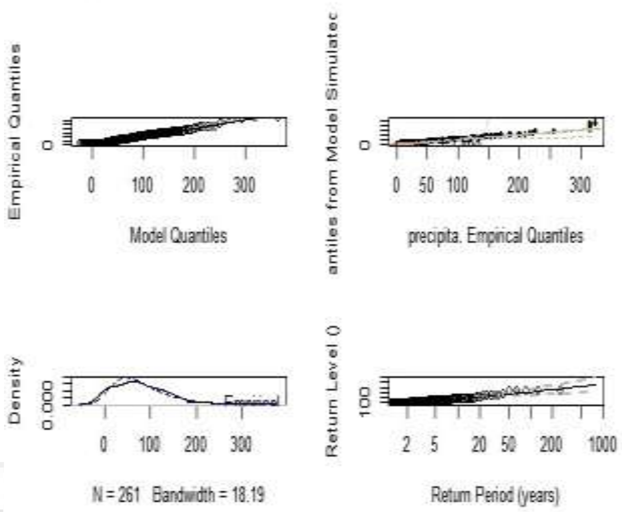
Parameter	Symbol	Parameter estimates	Standard error estimates
Location	μ	**51.47757800**	**3.37623182**
Scale	σ	**2.56438424**	**2.56438424**
Shape	ε	**0.06410368**	**0.05553106**

According to the results from tables above the shape parameter is not negative. From three families of GEV distribution, this is the case of either gumbel distribution or frechet.

7.6 Model Selection

The purpose of the study, is to show the best model from three families of GEV distribution. in this model selection, goodness of fit criteria is taken into account by using graphical Method Monthly rainfall data are plotted. Figures below are sample of diagnostics plots of Gumbel distribution

Diagnostic plot of Maximum rainfall data to fit Gumbel distribution



By analysis of Empirical Quantile plot almost all point lies to the line that means that our model is adequate. By looking at Density plot, it shows that the distribution of our data is normal because data are symmetry with respect to the mean. Return level plot is Convenient for model presentation and Validation that show that our model is adequate for our data.

7.7 Likelihood Ration Test

It is good to use numerical methods in statistical test than using any other method. This method of likelihood ration test has Chi-squared distribution of one degree of freedom for 0.05 level of significance. The purpose is to Compare GEV distribution with its particular Gumbel distribution.

Period	Likelihood ratio	Chi-square	P-value	Significance level
Month	0	0.01	1	0.05

The alternative hypothesis says that GEV distribution is better than Gumbel distribution. Since p-value =1 is greater than alpha =0.05 similarly as the Chi-square critical value=0.01 is greater than likelihood ration test 0 we fail to reject the null hypothesis. This means that the Gumbel distribution provides an acceptable model for the data (The Gumbel model is adequate for our data.

Table 5: 95% Confidence interval of shape parameter estimates

Period	Lower bound(mm)	Upper bound (mm)
Month	-0.0291	0.1497

The table above shows that Zero is included in the Confidence interval, Then the Gumbel distribution is a good fit for Maximum monthly rainfall.

7.8 Return Level Estimate

The maximum rainfall from monthly data which was recorded at Rwanda Meteorological service specifically at Kigali city is about 324.30 mm. Generalized extreme value distribution is used to estimate return levels with their corresponding return periods. By using GEV distribution, return level which is expected to exceed the maximum of available data is calculated for different selected return period

Table 6: Return levels with corresponding return periods for Monthly rainfall

Return period	**5**	**10**	**15**	**20**	**50**	**140**
Return level	**125.4**	**165.1**	**188.4**	**205.1**	**259.1**	**324.36**

From Tables above, return levels are changing with respect to the change in return periods. Changing in increasing return periods, return levels of extreme maximum rainfall also increase. the current maximum Monthly rainfall is 324.36 mm and the extreme maximum rainfall is expected to be 324.365mm in a return period of 140 Months or 11 years or more. The estimated return level lies in 95% confidence interval of (252.663, 396.0664) The obtained confidence interval shows that the true value of the return level of extreme maximum monthly rainfall is 100% within the interval

8. Summary

Flood have been the most common and serious disaster in most countries including Rwanda, over the past few decades the major cause of that disasters is the occurrence of extreme maximum rainfall a research on extreme high rainfall can contribute to the understanding of environment and climate change in general. The aim of this study was to analyze the monthly rainfall data obtained from Rwanda Meteorological services. The data was collected om Monthly basis from January 1996 to December 2017 in Kigali city.

This research used the GEV distribution in modelling extreme maximum rainfall. Dickey-Fuller (ADF) and Kwiatkowski-phillips-schmidt-shin (KPSS) statistical test were used to test the stationarity assumption. The likelihood ration test was used to select the good model among GEVD families, the parameter estimate and corresponding standard error estimates were estimated using maximum likelihood estimation method, according to value of parameter estimates, the shape parameter of GEV Distribution has positive sign which means that our data follow either gumbel distribution or frechet. Analysis gave gumbel distribution as the best-fit of modelling the extreme maximum rainfall data of Kigali city. Return level of maximum rainfall in this study was the amount of extreme maximum rainfall which is rare to happen; it was expected to exceed the maximum possible value of data randomly in a given return period of time, the corresponding return period was around 140 months or 11 years for monthly maximum rainfall data of the same station.

8.1 Conclusion

Generalized extreme value distribution has been found as the best distribution model. In this study the extreme maximum rainfall was modeled using Gumbel distribution which is one of three families of GEV distribution.

The stationarity of our data were approved in order to make forecasting of extreme high rainfall. The forecasts of Extreme high rainfall in Kigali city were estimated. This research with the purpose of estimating the return level and predict its coming period, can contribute in the heritage of future generation. It can also help decision maker from different institution which depend on climate like RAB (Rwanda Agriculture Board), it helps also for easy planning and implementation of some strategies of mitigating flood with effect of extreme rainfall. For further studies, the results prove that maximum rainfall is quite increasing with respect to time.

The future researchers can start to think of extreme maximum rainfall in the whole country

8.2 Recommendation

The Generalized extreme value Distribution has proved to be powerful tool to analyze data of extreme event.

As we have seen in this project, Gumbel distribution has proved to be a good model for modelling extreme high rainfall. Return level estimate was shown that after 11 years they will be a high rainfall. Rwandan Government should take measure of prevention of flood like planting trees, to relocate people's house in high risk zone.

The researcher must investigate the use of Generalized pareto Distribution (GPD) to model extreme maximum rainfall in Kigali city. The researcher may also consider to model extreme high rainfall in the whole country.

References

[1] Herve V. HABONIMAN, Jean Pierre BIZIMANA, Ernest UWAYEZU, Joseph TUYISHIMIRE, John MUGISHA Integrated flood modeling for flood hazard assessment in Kigali City, Rwanda. 2014.
[2] The Fidelis Group, LLC The Weibull Distribution. [March 6, 2014].
[3] Husna Hasan, Norfatin Salam, and Mohd Bakri Adam Modelling Extreme Temperature in Malaysia Using Generalized Extreme Value Distribution
[4] Triphonia Jacob Ngailo, Joachim Reuder, Edwin Rutalebwa, Shaban Nyimvua, Michaels's. Mesquita Modelling of Extreme Maximum Rainfall using Extreme Value Theory for Tanzania. 2016.
[5] N. Ayuketang, E. Joseph MODELLING EXTREME TEMPERATURE IN CAMEROON USING GENERALIZED EXTREME VALUE DISTRIBUTION March 21,2014
[6] Husna Hasan, Norfatin Salam, and Mohd Bakri Adam Modelling Extreme Temperature in Malaysia Using Generalized Extreme Value Distribution. 2013.
[7] Helena Penalva1, Manuela Neves2 and Sandra Nunes3 Topics in Data Analysis Using R in Extreme Value Theory. 2013.
[8] Innocent Nzeyimana1, Kwitonda Philliper2 1Minagri, Kigali, Rwanda, innonzey@gmail.com 2RMA, Kigali,

Rwanda Drought conditions and management strategies in Rwanda

[9] T. A. BUISHAND Royal Netherlands Meteorological Institute (KNM), PO Box 201, 3730 AE De Bilt, The Netherlands Extreme rainfall estimation by combining data from several sites* .2017

[10] Sheri Markose and Amadeo Alentorn the Generalized Extreme Value (GEV) Distribution, Implied Tail Index and Option Pricing. December 2010

[11] Michael McAleer Lan-Fen Chu Ching-Chung Chang Statistical Modelling of Extreme Rainfall in Taiwan. December 2012

[12] National Institute of Statistics of Rwanda DEMOGRAPHIC AND HEALTH SURVEY RDHS 2014-2015

[13] Rwanda Environment Management Authority Kigali State of Environment and Outlook Report 2013