Non Homogeneous Poisson Process Modelling of Seasonal Extreme Rainfall Events in Tanzania

Triphonia Ngailo[1], Nyimvua Shaban[2], Joachim Reuder[3], Edwin Rutalebwa[4], Isaac Mugume[5]

[1]University of Dar es Salaam, Department of Mathematics, Dar–es–Salaam, Tanzania

[2]University of Dar es Salaam, Department of Mathematics, Dar–es–Salaam, Tanzania

[3] University of Bergen, Geophysical Institute, Bergen, Norway

[4] Dar es Salaam Institute of Technology, General Studies Department, Dar-es-Salaam, Tanzania

[5] Makerere University, Department of Geography, Geoinformatics and Climatic Sciences

Abstract: ***Extreme rainfall events due to heavy rainfall can vary greatly. This variability can be explained by different factors such as season of the year, temperature and local topography, among others. Statistical models using Extreme Value Theory have been used to model extreme weather events which assume stationarity of rainfall process. However, the stationarity requirement is not met in reality for rainfall data because rainfall time series usually exhibit seasonality. A stochastic model based on a non- homogeneous Poisson Process (NHPP) charactezised by a time-dependent intensity of rainfall occurrence, is employed in to study the seasonal and trend effects on extreme events modelling of daily rainfalls exceeding prefixed threshold value. Dataset from 14 Tanzania rainfall stations over the period 1981–2014 was used. The Akaike information criterion and likelihood ratio test methods were used to select NHPP model that best fits the data. The results showed a good fit for time–varying intensity of rainfall occurrence process by the first order harmonic Fourier law and improved analysis as well as modelling of extreme rainfall using NHPP intensity function.***

Keywords: Non homogeneous Poisson Process, maximum likelihood estimation, Seasonality, Extreme Rainfall Events, Intensity function

1. Introduction

Extreme rainfall events cause significant damage to agriculture, ecology and infrastructure, disruption of human activities injury and loss of life [2, 3, 7]. In recent years, floods have become more frequent in Tanzania which necessitate investigating their cause.

For example, in December 2009 and January 2010, Mkonda river banks burst, affected Kilosa town and led to the displacement of about 24,000 people. In Mpwapwa and Kongwa districts, an estimated 19,000 persons were displaced [1] and the cost of rehabilitation estimated at TShs 329 billion. The April 2011 floods in Kilombero valley demolished 663 houses in Morogoro region and submerged 2,942 making about 9,000 people homeless. The extreme rainfall in January 2008 led to floods which displaced hundreds of people and flooded mining pits in Mererani resulting in over 70 deaths [2]. In all these context, modeling these extreme rainfall events is of great interest to public safety alert, life insurance and protection, the design of civil infrastructures, town and regional planning, management and loss mitigation.

Different approaches have been used to define extreme rainfall events, with considerable discrepancies between the definitions of extreme. These definition includes; annual maximum [3], percentile based [4, 5] or duration of wettest 5 days of the year [6] and thresholds rainfall exceedances [7]. For our study, we define extreme rainfall event as rainfall amount exceeding the 99^{th} percentile of the distribution of seasonal rainfall.

Seasonality is one of the main feature of rainfall time series. The rainfall over Tanzania is driven mainly by the migration of the Inter–Tropical Convergence Zone (ITCZ) [8] which is over Tanzania during October–December (OND) and March–May (MAM) making Tanzania to have two rainy seasons. The southern, western and central parts of the country experience one wet season that starts from November lasting up to April or May (NDJFMA) [9].

Rainfall variability in Tanzania have been reported in previous studies. Kassile et al.[10] examined the evolution of rainfall over central Tanzania focusing on Dodoma region and did not find statistically significant trend in the amount of rainfall over the period 1981–2010. Hamisi [11] analyzed the monthly rainfall trends and variability over Tanzania from 1982 to 2012 and found significant decreasing trends for all stations except in Mwanza, Sumbawanga, and Dodoma. In a recent related study by Ngailo et al. [12] of extreme events over Tanzania found increasing intensity of extremes rainfall and decreasing return periods over different regions. Studies employing Non-homogeneous Poisson processes (NHPP) modeling of extreme rainfall over Tanzania are limited which is covered in this work.

NHPP models have become an important tool for modeling non stationary processes. It is considered as a more realistic method than the classic ones to model different day-to-day random phenomena. This is due, to the fact that they consider the intensity function as time-dependent [13]. The major feature of Non-Homogeneous Poisson Process (NHPP) is that the occurrence of event is allowed to change with time and also related to the location. Thus the occurrence rate of an event is function of time $\lambda(t)$. The NHPP modeling has been applied in many disciplines e.g. in meteorology for modeling inter- arrival times of rainfall events [14], estimating containership arrival rate in harbor operation and management [15], storms prediction [16], modeling of hot extreme events [17], and by Sirangelo et al. [18] to analyze occurrence of rainfalls.

The major objective of this work is to assess the possibility of trend in extreme rainfall events over Tanzania. The extreme rainfall events behavior are characterized by NHPP to represent the occurrence. We incorporated a linear transform function of seasonal and trend covariates in the NHPP model with the peak over threshold (POT) approach to the daily rainfall data. The rest of the paper is presented as, section 2 presenting study methods and data sources, section 3 discussing the results and conclusion in section 4.

2. Data Sources

In this study we used daily accumulated rainfall, as observed from a rain gauge network of Tanzania rainfall stations for a period from 1961 to 2014 obtained from Tanzania Meteorological Agency (TMA). These rain gauges provide good coverage of the area [19] and allow investigation of rainfall variability at the regional scale [19, 20]. Since different parts of Tanzania have different rainfall seasons (i.e. November–April (NDJFMA), October– December (OND) & March–May (MAM)), our analysis took the seasonality into account and allowed NHPP model parameters to be seasonally dependent.

3. Methods

3.1 Non Homogeneous Poisson Process Model(NHPP)

The Non-homogeneous Poisson Process (NHPP) model has been used to model the occurrence of events in time. The NHPP is the generalization of the Poisson process [21, 22, 23] which is characterized by a deterministic intensity function describing how the rate of the process changes in time [24]. Thus, it is reasonable to model the extreme rainfall occurrences by a NHPP, where points occur randomly in time, at a variable rate which depends on influential covariates such as seasonality and trend by incorporating these covariates in NHPP model. We model the intensity $\lambda(t)$, where t denotes time measured on a daily scale, as a deterministic function and consider seasonality and trend terms $(R_t 31)$ and $(R_t 15)$ of rainfall, corresponding to the MAM, OND and NDJFMA seasons of the harmonic functions describing the annual cycle as described by the model equation 1.

$$\lambda(t) = F(t) + G(t) \quad (1)$$

where

$$F(t) = \frac{1}{2}a_0 + \sum_{k=1}^{n} a_k \cos\left(\frac{2\pi kt}{p}\right) + b_k \sin\left(\frac{2\pi kt}{p}\right)$$

$$G(t) = c_1 R_t 31 + c_2 R_t 15$$

where n is the optimal number of harmonics for the season, a_0, a_k , b_k , c_1 and c_2 are unknown parameters,, R_x31 and R_x15 are trend terms , defined as a 31 and 15 days moving average centered in each day, which provide information on the local state of rainfall.

If t is time, we assume the counting process, $N(t)$ (i.e. counting the total number of events that have occurred up to time, t) is an NHPP with intensity function, $\lambda(t)$ and we require $N(t)$ to be considered a counting process for $t \geq 0$. If $t < s$ over the interval [t, s], such that $N(0) = 0$, then $N(s) - N(t)$ is the number of events that have occurred in the interval $[t, s]$. We define $N(t)$ as a Poisson process, [25] if;

$$p\{N(t+s) - N(s) = k\} = e^{-\lambda t}\frac{(\lambda t)^k}{k!} \quad (2)$$

and if the mean is

$$E[N(t)] = \lambda t$$

then $N(t)$ is a NHPP provided:

$$P\{N(t) = k\} = e^{-\Lambda(t)}\frac{(\Lambda t)^k}{k!} \quad (3)$$

The intensity function $\lambda(t)$, a parameter of interest in this study, describes how Poisson Process changes in time and $\Lambda(t)$ denotes the expected number of events of NHPP over time interval $[0,t]$.

For NHPP,

$$E[N(t)] = \Lambda(t) \quad (4)$$

where,

$$\Lambda(t) = \int_0^t \lambda(x)dx, t > 0$$

here for each,

$$0 < t_1, t_2, \ldots, < t_{n-1} < t_n$$

we have

$$N(t_1), N(t_2) - N(t_1), \ldots N(t_n) - N(t_{n-1})$$

independent random variables. The inter-arrival times (arriving time) are defined as

$$s_1 = t_1, s_2 = t_2 - t_1, \ldots, s_n = t_n - t_{n-1}$$

Therefore, the probability of having k arrivals or k events during the time $(t, t+x)$ is;

$$P_r(N(t+x) - N(x) = k) = \frac{e^{-[\Lambda(t+x)-\Lambda(x)]}[\Lambda(t+x) - \Lambda(x)]^k}{k!} \quad (5)$$

The parameter estimation is performed by maximum likelihood (section 3.2) and the selection of the variables to be included in the linear predictor is based on the likelihood ratio test (section 3.3.2).

3.2. Maximum Likelihood Estimation

Given a NHPP with intensity function $\lambda(t)$ in equation (1) is observed over a fixed interval (0, T). For each $0 < t_1 < t_2$, $\cdots$, $< t_{n-1} < t_n \leq T$. Our intention is to find the function $\lambda(t)$ that maximizes the Likelihood $L(\lambda)$, then likelihood function is given by:

$$L(\beta;(t_n)_{t=1}^{n}) = \exp\left[-\int_A \lambda(t;\beta)dt\right]\prod_{t=1}^{n}\lambda(t_i;\beta); \quad (6)$$

with $\lambda(t;\beta)$ is defined as:

$$\lambda(t;\beta) = \exp(\mathbf{X}^T(t)\beta)$$

where, $X^T(t)$ is the row vector of covariates at time t known as seasonality and trend, β is the vector of parameters which are (a_k, b_k, c_1 and c_2) and A is the space where the point process is defined. Assuming that $\lambda(t)$; β) is constant in each time unit, the expression of the log likelihood is:

$$l(\beta;(t_i)_{i=1}^{n}) = -\sum_{t=1}^{T}\lambda(t;\beta) + \sum_{t=1}^{n}\log\lambda(t;\beta) \quad (7)$$

where T is the length of the observation period. The maximum likelihood (ML) estimation of the $\hat{\beta}$ can be done using numerical optimization method in R.

3.3. Model Selection

In model selection, the aim is to find the smallest set of variables which provides an adequate description of the data. Several criteria are commonly used for model selection [26] e.g. the Aikaike Information Criterion, AIC (sec.3.3.1) and the Likelihood Ratio Tests, LRT (sec.3.3.2) have used in this work.

3.3. 1.Akaike Information Criteria

The AIC is a measure of goodness of fit that takes the number of fitted parameters into account and is an effective method of choosing between a given set of models [27]. A true model does not necessarily have to be in the set because the goal is to select the best approximating model of set [28]. It is widely used as a measure for selecting the best among competing models for a fixed data set e.g. in ecology [29]; wildlife [30] and many others. The chosen model is the one that minimizes the Kullback-Leibler distance between the model and the truth (that is a model that minimizes the loss of information) [31, 32]. The AIC is described by equation (8).

$$AIC = 2K - 2\ln(L) \quad (8)$$

where, L denotes the maximum log–probability of the estimated model and is the likelihood evaluated in the estimator; K is the number of estimated parameters in the approximated model [33]. The AIC scores are often shown as ΔAIC scores, or difference between the best model (smallest AIC) and each model (so the best model has a ΔAIC of zero). Therefore the model with the lowest AIC is the best model among all models specified for the data at hand.

3.3.2. Likelihood Ratio Test

Likelihood is the probability of the observed data given a selected model [32] and measures how well the data supports that particular value in the model. In this work we make comparison between the two adjustments with and without seasonal and trend covariates, the best parameter was identified by implementing the Likelihood–ratio Test (LRT). The LRT allows us to compare the models with and without covariates as explained by Drazek [13] and it is a statistical proof of the accuracy of the fitting between two models, where one fits better than the other [34]. The test statistics (eqn. 9) is:

$$\text{LRT} = 2(\ln L_1 - \ln L_0) \quad (9)$$

where, $\ln L_0$ is the maximum log–likelihood under the null model and $\ln L_1$ is the maximum log–likelihood under the alternative model. The null model has fewer parameters than the alternative model (without covariates). If the null model can be viewed as a special case of the alternative model, then statistical theory allows use of the χ^2 distribution to compute a p–value. LRT calculates for each covariate in the model the p-value of a likelihood ratio test comparing the original fitted NHPP with the model excluding that covariate from the linear predictor. The covariate with p–value < 0.05 is selected as the best covariate and is included in the model.

In this work, Statistical programming software, R was used to graphically and numerically described the data as well as estimating the intensity functions.

4. Results and Discussion

4.1 General characteristics of Extreme Rainfall occurrences

The rainfall time series for: Dar es Salaam on both seasons; Morogoro (MAM); Sumbawanga, and Dodoma (NDJFMA)) are presented using figure (1). For the period 1981–2014, results show that there was a changes in the intensity and frequency of extreme rainfall events. For Dar es Salaam on both seasons (MAM) and (OND) we can observe a more frequent occurrence of rainfall amount over threshold from the beginning of 1990's towards the end of the period.

International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

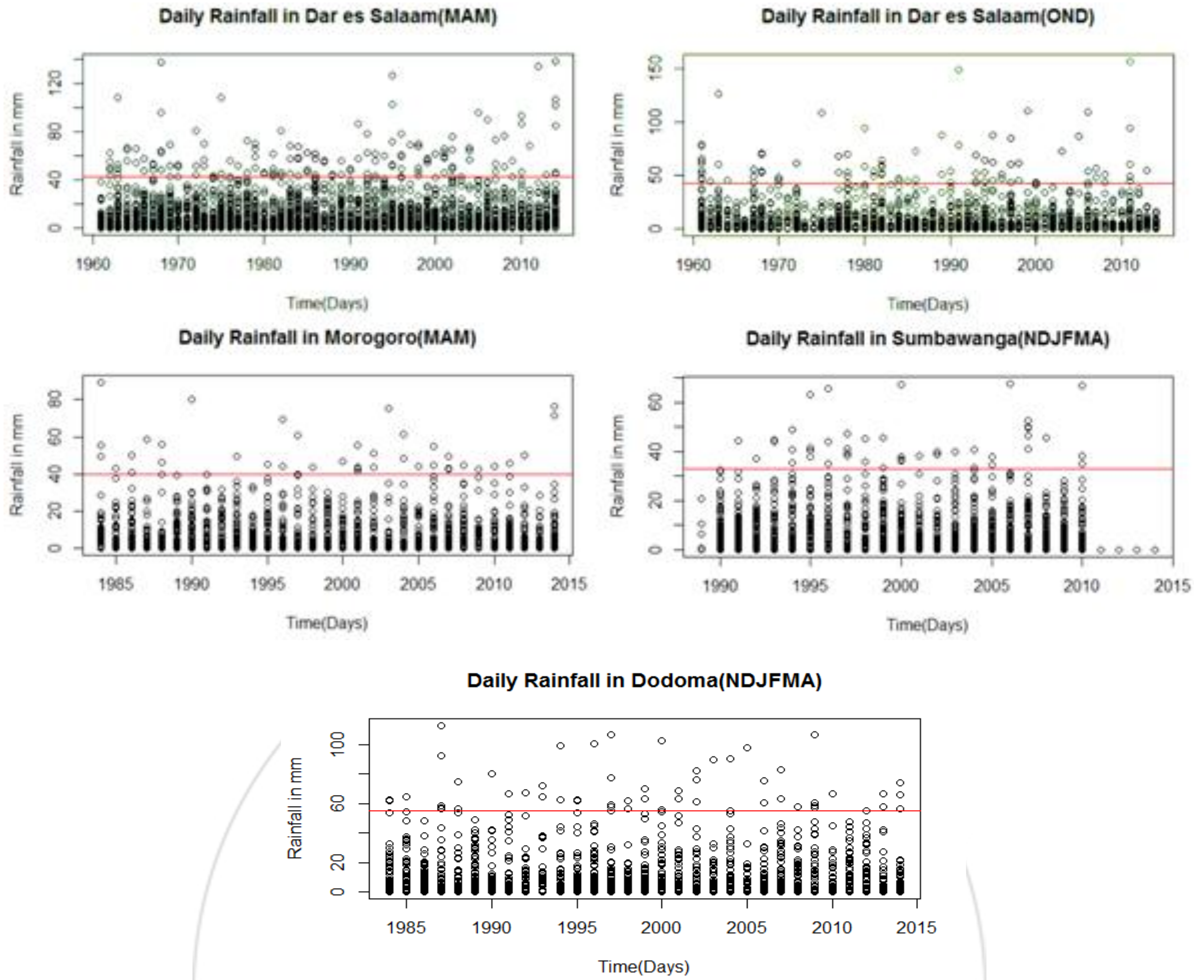


Figure 1: Daily rainfall series for MAM, OND and NDJFMA seasons

4.2. Trend Analysis on Rainfall Amounts

We used the Mann Kendall (MK) trend test as a non-parametric test to investigate the trend of extreme rainfall amounts over Tanzania seasonally. The trend analysis were split into seasons that are defined according to the Tanzania climatological criteria, from November to April (NDJFMA), from March to May (MAM) and from October to December (OND).Table 1 shows the P-values of trend test and tau in brackets of every station. However, the variability in the data in some regions based on MAM, OND and NDJFMA seasons indicates significant increasing trend. This can be confirmed by the p-values < 0.05 in table 1 and in some regions P-values are close to zero particularly in Sumbawanga and Arusha. In Bukoba, Morogoro, Zanzibar, Pemba, Iringa, Songea and Mahenge shows the trend was insignificant.

Table 1: MannKendall Trend Test

Region	(MAM)	(OND)	(NDJFMA)
Dar es Salaam	0.02(-0.05)	0.004(-0.04)	
Tanga	0.56(0.01)	0.01(-0.04)	
Mtwara	0.56(0.01)	0.00004(-0.04)	
Bukoba	0.93(0.00)	0.63(-0.006)	
Kigoma	0.38(-0.01)	0.49(-0.009)	
Morogoro	0.09(-0.02)	0.08(-0.03)	
Zanzibar	0.14(-0.02)	0.35(-0.01)	
Pemba	0.85(-0.03)	0.26(-0.02)	
Iringa	-	-	0.77(0.00)
Arusha	0.00(-0.10)	0.03(-0.03)	
Songea	-	-	0.18(-0.01)
Sumbawanga	-	-	0.00(-0.10)
Mahenge	-	-	0.94(0.00)
Dodom	-	-	1.04(-0.02)

4.3. Analyzing Extreme Rainfall Events

In order to fit a model based on a point process, an occurrence point must be associated to each event; we choose the day of the spell where the maximum R_t value is observed. Extreme Rainfall Event (ERE) is defined as a spell, of arbitrary length, of consecutive days with their R_t (Rainfall) values exceeding an extreme threshold, we selected a threshold of 99th percentile for the OND, MAM and NDJFMA daily rainfall data from 1961- 2014 and from 1983-2014.The occurrence point is defined as the point where maximum value of rainfall amount occurs within events. We have presented the number of events and number of exceedances over threshold in brackets are table 2 and 3 for MAM and NDJFMA seasons for Dar es Salaam, Tanga, Iringa, Songea and Dodoma to represent other regions.

Table 2: Extreme rainfall Events in Tanzania for MAM season

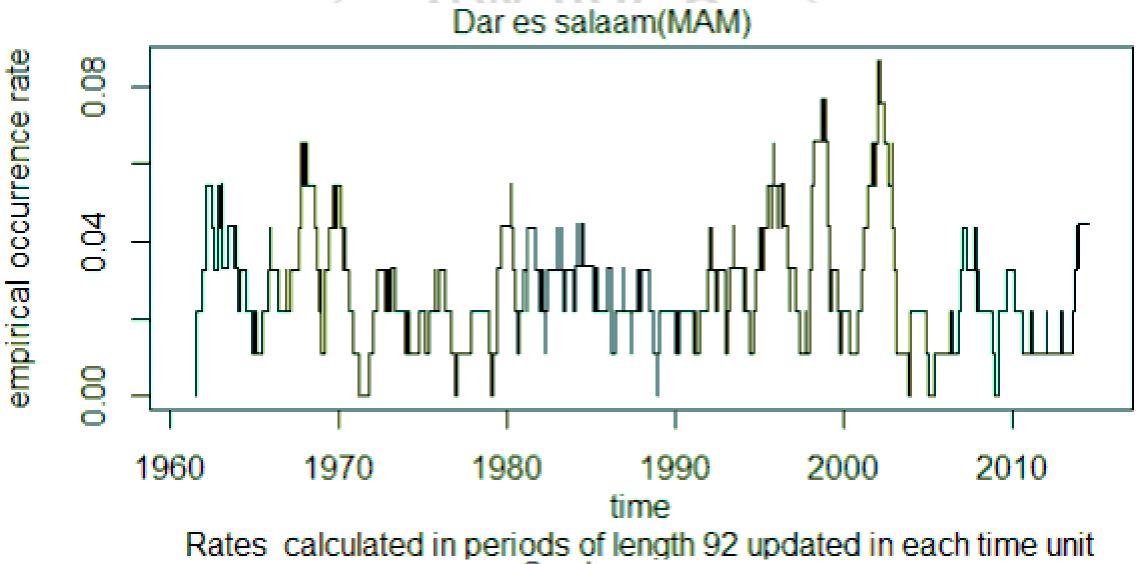
SN	*Decade*	*Events in Dar es Salaam*	*Events in Tanga*
1	1961-1970	30(31)	24(29)
2	1971-1980	18(20)	23(29)
3	1981-1990	25(29)	28(39)
4	1991-2000	29(32)	26(30)
5	2001-2010	25(27)	22(24)
6	2011-2014	7(9)	7(7)

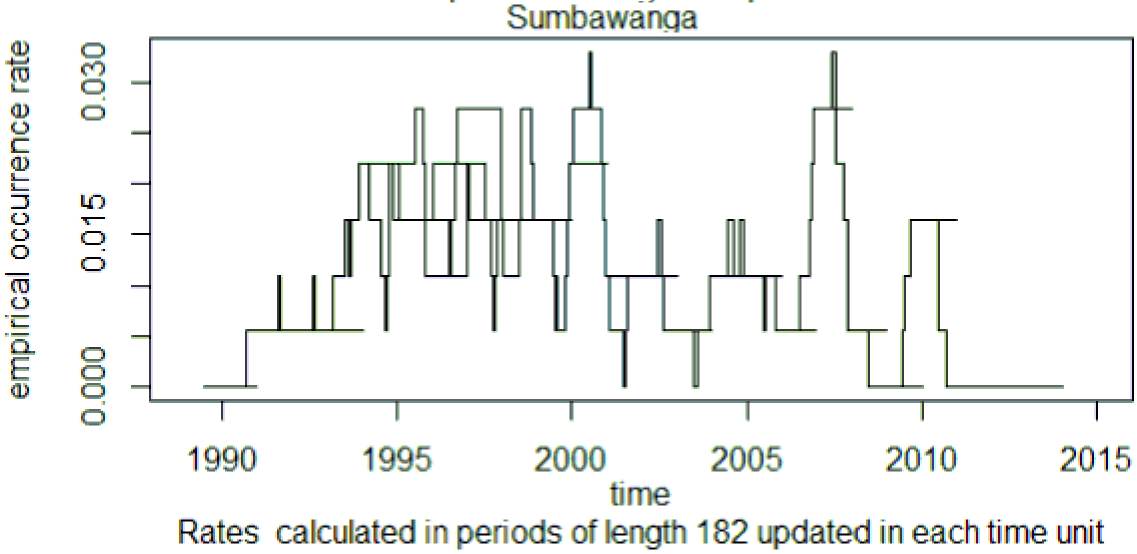
Table 3: Extreme Rainfall Events in Tanzania for NDJFMA season

SN	*Decade*	*Events in Iringa*	*Events in Dodoma*	*Events in Songea*
1	1984-1993	11(12)	25(25)	37(38)
2	1994-2003	23(24)	32(32)	38(38)
3	2004-2014	22(22)	24(24)	34(35)

The results from table 2 for Coastal regions shows that extreme rainfall events shows a slight increase in the 2^{nd} to the 4^{th} decades, while in table 3 shows a slight increase in extreme events in the 2^{nd} and 3^{rd} decades. Generally there are fluctuations in the number of events.

The empirical occurrence rates on overlapping and disjoint interval are also calculated. The occurrence rate is calculated as the number of points in the considered interval divided by its length, and each rate is assigned to the mean point of the interval. Overlapping intervals are defined by a constant length (L). In this study we use the length of 92 days for MAM and OND and 182 days for NDJFMA seasons. The disjoint intervals can be specified by the number of intervals or by a constant length. All the intervals have the same length except the last one, which is shorter because the length is not a multiple of constant length. A plot of the empirical rate over time were performed. The results show that, the occurrence rate is neither increasing nor decreasing (it's constant) as shown in figure 2. The maximum excess position defines the occurrence point in the point process.





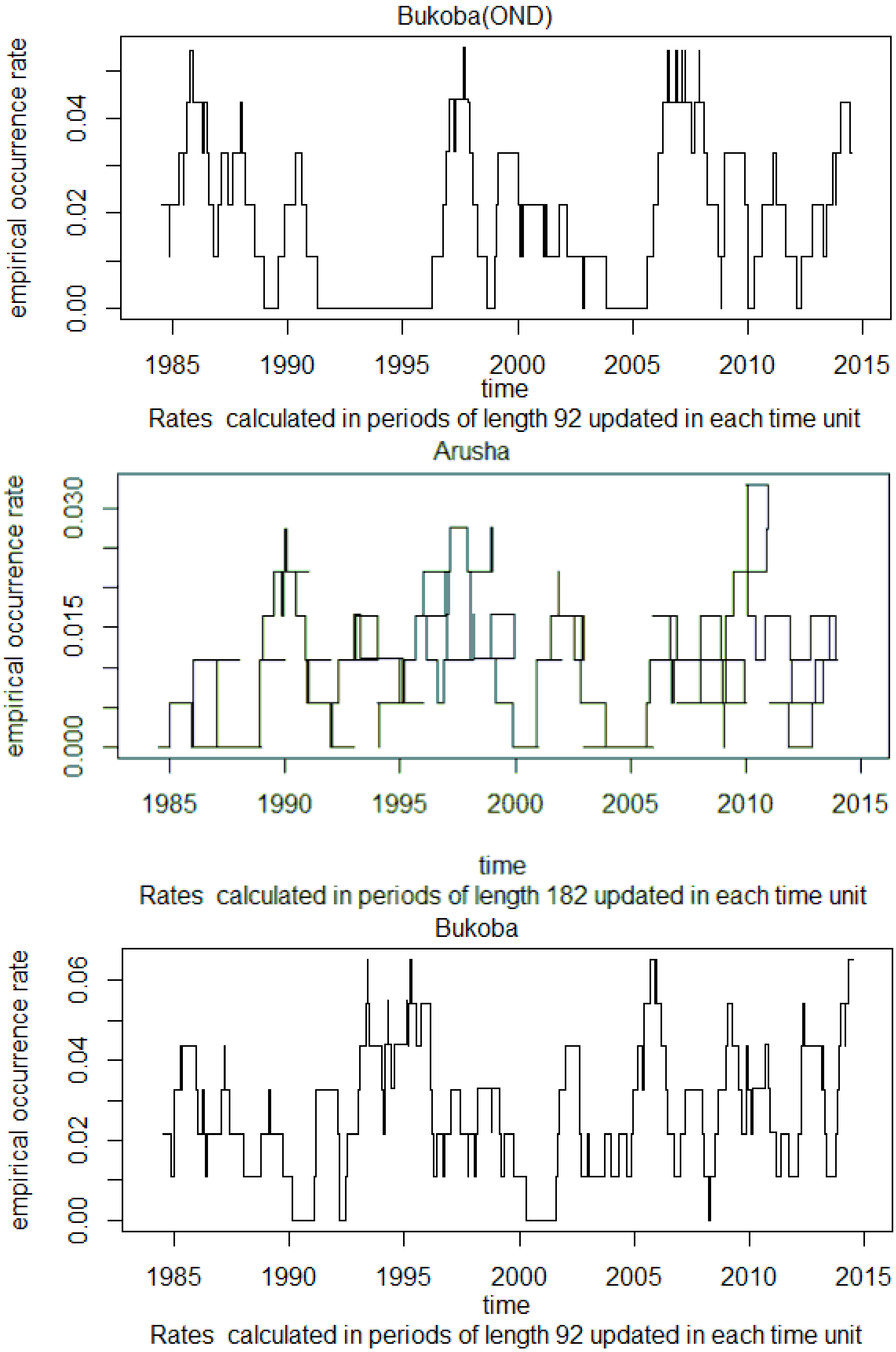


Figure 2: Empirical occurrence rates on overlapping and disjoint intervals

4.4 Fitting the model

Due to the characteristics of the ERE occurrence, seasonality behavior and trend was modeled by a Point Process with a non homogeneous intensity. The intensity function was modeled using equation (1), As an exploratory step to check which covariates are more influential, we carried out an automatic stepwise selection by AIC in both directions. First, the initial model which only includes the intercept was fitted for Dar es Salaam and Mahenge to present other regions are as shown in figure 3 which shows no variations in intensities.

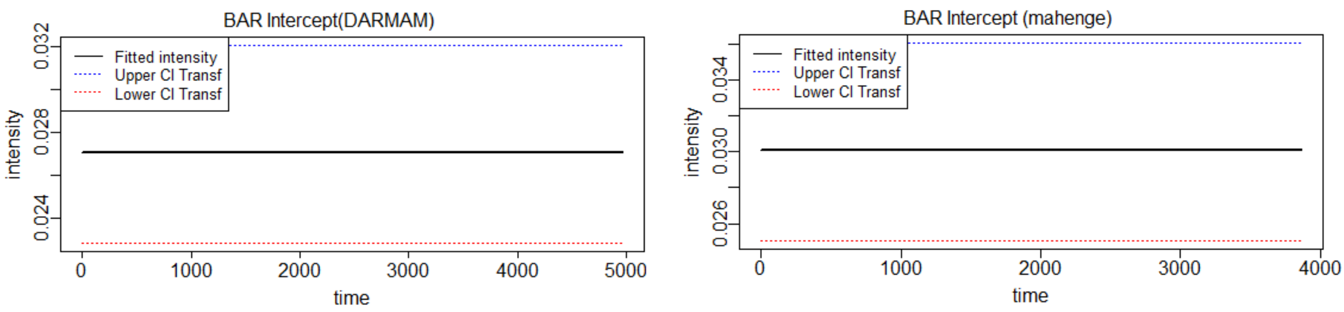


Figure 3: Bar graph plotted without covariates

The AIC values for MAM season was calculated using optimization method in R and result is given in table 4.

Table 4: Automatic stepwise selection by AIC

Region	Cos	Sin	R_t31	R_t15
Dar es Salaam	1239.40	1233.83	122.94	1234.97
Tanga	1190.37	1204.52	1207.98	1208.29
Mtwara	864.19	858.01	889.37	889.37

According to the AIC for the MAM season, in Dar es Salaam, the best model should only include the R_t31 covariate, for Tanga, the best model should only include the cosine covariate and for Mtwara, the best model should only include the sine covariate as shown in table 4. We combined also a selection based on the likelihood ratio test, using a forward stepwise approach controlled by the model for all stations, the results are shown in table 5.

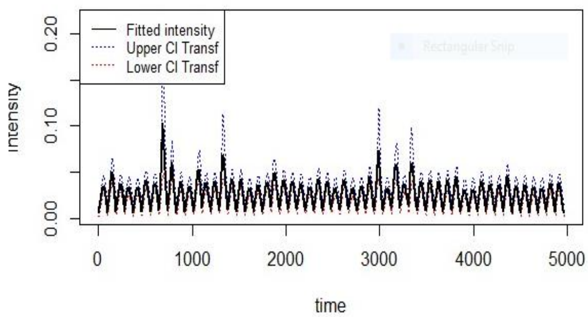
Table 5: Likelihood ratio test in both directions

Region	ML Ratio Test	P-Value
Dar es Salaam	21.39	0
Bukoba	14.36	0
Mtwara	32.92	0

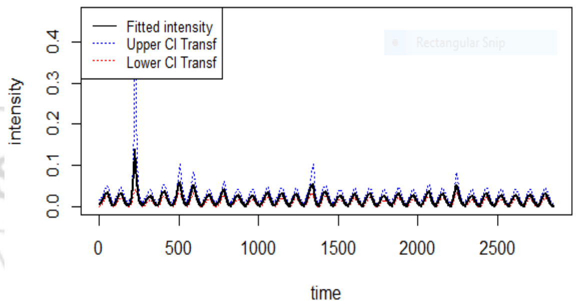
Since the p-value is 0, the first order harmonic is included and the inclusion of the second order harmonic is checked. After the inclusion of the second order harmonics for Dar es Salaam; ML ratio test statistics is 1.14 and P-value is 0.565.The p-value 0.565 rejects the inclusion of the second order harmonic. The final model resulting from this forward covariate selection process includes the first order harmonic term and trend term R_t31.

The model with selected covariates were fitted in R and the result of the fitted intensity were plotted as shown in figure 4.

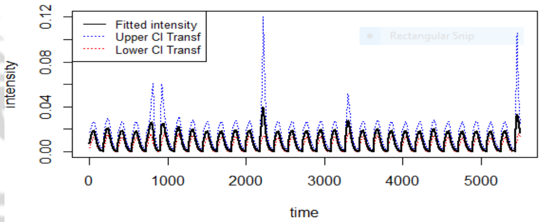
Dar es Salaam(MAM)

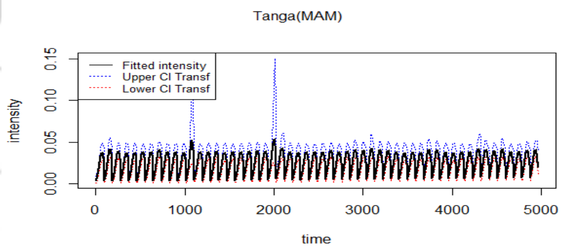


Arusha(MAM)

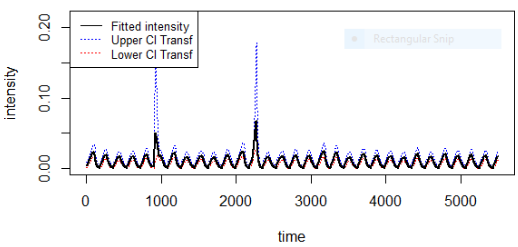


Iringa (NDJFMA)





Morogoro (MAM)



Sumbawanga (NDJDMA)

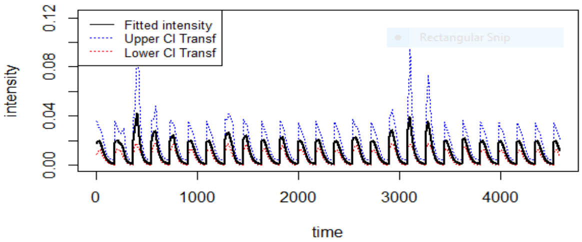


Figure 4: Poisson rate fitted for the observed time interval 1961 − 2014 and 1984 − 2014.

We can observe general evolution of the fitted rate of some regions showing slight increasing trend and majority of the regions showing a constant trend. This confirms with the values of extreme events in table 2 and 3 which shows slight increase in extreme events in the second and third decades and fluctuations behavior .

In comparing the original fitted NHPP with the model excluding the covariate from the linear predictor and the model with covariates the Likelihood Ratio test that calculates the p-values of a likelihood ratio test (LRT) for each covariate in the model were used. The p-values of the LRT comparing the initial model and the model without the covariates is given in table 6.

Table 6: LRT P-Values

p-Value	Dar es Salaam	Tanga	Songea	Bukoba
Cosine	0.00	0.00	0.00	0.00
Sine	0.00	0.00	0.01	0.01
R_t31	0.02	0.50	0.42	0.17

For all seasons and all regions the P-Value for first harmonic seasonality is statistically significant. In order to make comparable the empirical and the fitted occurrence rates, a cumulative fitted rate were used. This means, the fitted values are the sum of the intensities fitted by the model over the same interval where the empirical rates have been calculated. The cumulative curves shows statistically significant seasonal features. Thus, the temporal variation of rainfall occurrence intensity $\lambda(t)$ are well expressed through the Fourier series as a function of period $P = 1$ year, as shown in figure 5.

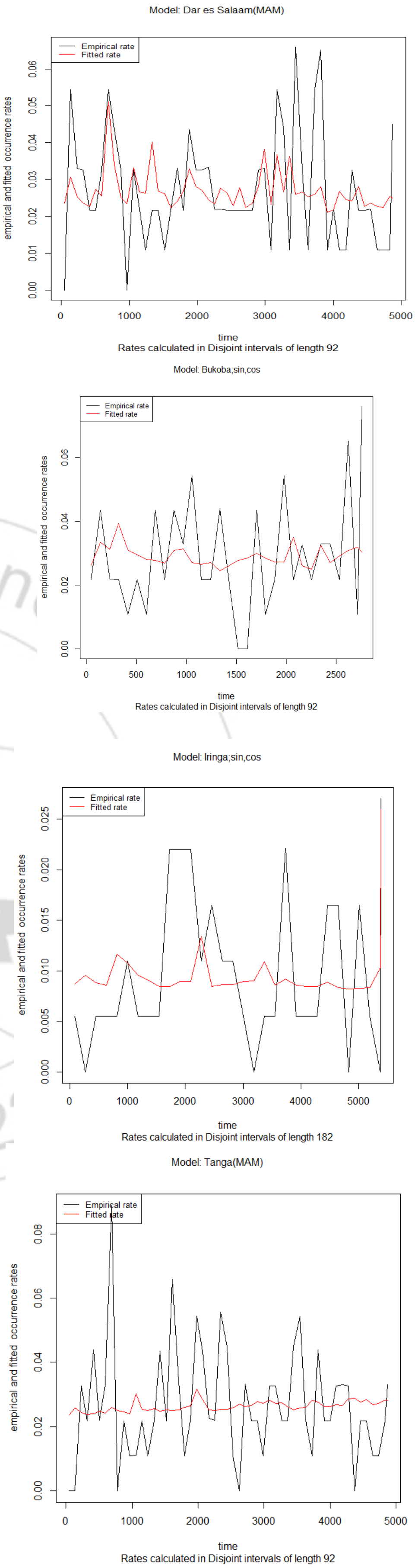


Figure 5: Empirical and Fitted intensities calculated on disjoint seasonal intervals

Fitting of our model requires to calculate the coefficients for intensity function $\lambda(t)$ (fitted parameters) and their corresponding confidence intervals. Table 7 contains the values of the fitted parameters which were estimated are with their standard errors in brackets.

Table 7: Parameter Estimation

Region	a_0	a_1	b_1	c_1
Dar es Salaam	-7.93(1.06)	-1.57(0.49)	4.13(1.09)	0.14(0.05)
Tanga	-6.94(1.05)	-2.26(0.54)	2.84(1.03)	0.06(0.08)
Mtwara	-7.20(1.35)	0.51(0.43)	3.67(1.42)	-0.02(0.07)
Bukoba	-7.81(1.39)	-2.07(0.64)	3.72(1.44)	0.09(0.06)
Iringa	-6.69(0.68)	1.72(0.42)	2.03(0.72)	0.20(0.16)
Songea	-6.31(0.57)	2.31(0.38)	2.17(0.58)	0.02(0.03)
Arusha	-12.04(2.35)	-1.53(0.75)	8.20(2.43)	0.28(0.12)
Mahenge	-4.91(0.35)	1.14(0.21)	1.05(0.43)	0.16(0.05)
Sumbawanga	-5.58(0.57)	1.54(0.40)	0.57(0.65)	0.12(0.09)

Confidence intervals for $\lambda(t)=\exp(v(t))$ was obtained by transforming the confidence intervals for $\lambda(t)=X^T(t)\beta$. The transformation approach applies an exponential transformation to the confidence interval of the linear predictor. The 95% confidence intervals of the parameters, based on the profile likelihood and on the properties of the ML estimators, was obtained, given in table 8.

Table 8: Confidence intervals for the intensity

Region	a_0	a_1	b_1	c_1
Dsm	-7.93 [10.12,5.93]	-1.57 [-2.51,0.70]	4.13 [2.07,6.34]	0.14 [0.03,0.2]
Tanga	-6.94 [-9.09,-4.96]	-2.26 [-3.40,-1.25]	2.84 [0.89,4.92]	0.06 [0.12,0.21]
Mtwara	-7.20 [-10.02,4.70]	0.51 [-0.35,1.32]	3.67 [1.03,6.62]	-0.02 [-0.18,0.10]
Bukoba	-7.81 [-10.52,5.09]	-2.07 [-3.33,-0.82]	3.72 [0.90,6.53]	0.09 [0.03,0.21]
Iringa	-6.69 [-8.20,-5.52]	1.72 [0.99,2.67]	2.03 [0.73,3.57]	0.20 [0.14,0.48]
Songea	-6.31 [-7.42,-5.20]	2.31 [1.57,3.06]	2.17 [1.08,3.26]	0.02 [0.03,0.08]
Arusha	-12.04 [-17.11,-7.82]	-1.53 [-3.11,-0.15]	8.20 [3.44,12.96]	0.28 [0.04,0.5]
Mahenge	-4.91 [-5.65,-4.26]	1.14 [0.76,1.59]	1.05 [0.19,1.90]	0.16 [0.06,0.2]
Sumba wanga	-5.58 [-6.86,-4.60]	1.54 [0.86,2.44]	0.57 [-0.64,1.95]	0.12 [0.06,0.28]

4.5. Model Validation

To validate the model ,we fitted NHPP with intensity $\lambda(t)$ using residuals. There are two types of residuals; uniform (or exponential) and raw residuals. Both of them are useful and provide complementary information. The validation analysis was done using the uniform residuals which consists Kolmogorov-Smirnov test and qqplots with a 95% confidence band based on a beta distribution. The residual plots were plotted which comprises the serial correlation based on the Pearson correlation coefficient, Ljung-Box tests and a lagged serial correlation . The plots for Bukoba for both seasons are plotted to represent other regions.

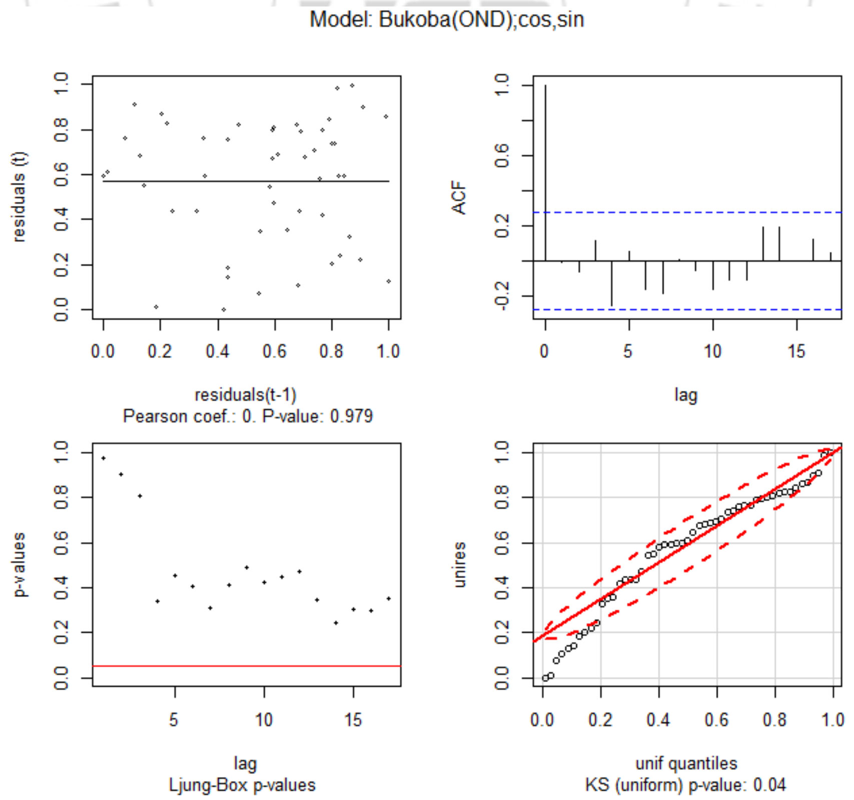


Figure 6: Validation plots of Uniform residuals of the final NHPP model

The results of the autocorrelation analysis and the uniform behaviour are satisfactory, as shown in figure 6. However the validation process was found to be satisfactory for the model fitted to Dar es Salaam, Bukoba and Tanga data. Concerning the NHPP checking, the uniform qq-plot, are shown, shows a linear behaviour inside the confidence band and the p-value of the corresponding Kolmogorov–Smirnov test is 0.263 for Dar es Salaam and 0.04 for Bukoba.

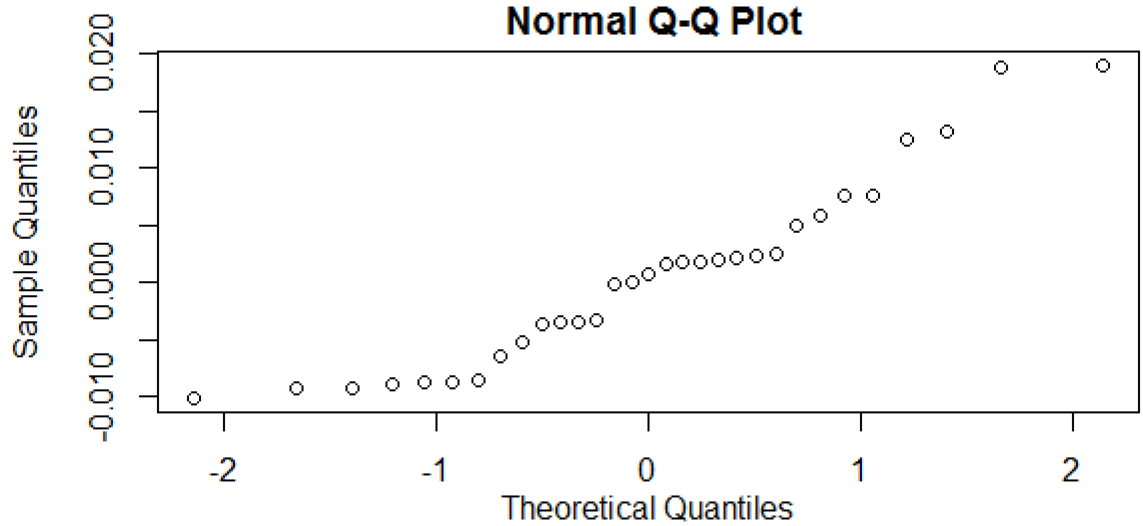


Figure 7: Uniform qqplot with beta confidence bands for Dar es Salaam

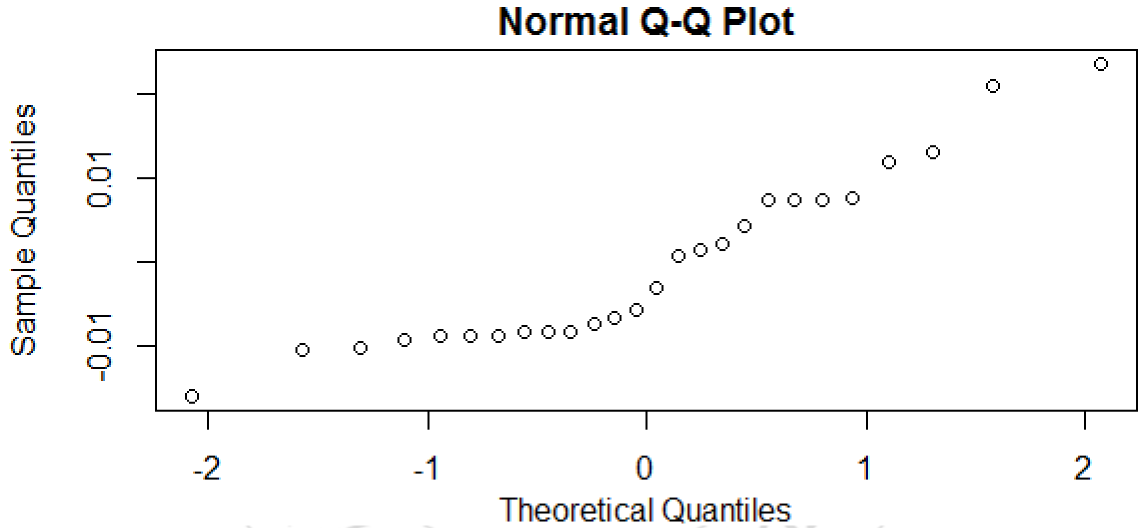


Figure 8: Uniform qqplot with beta confidence bands for Mahenge

5. Summary and Conclusion

In this paper a Non-homogeneous Poisson model has been used for the stochastic interpretation of the seasonal and trend variability of extreme rainfall occurrence process in 14 rain gauges selected data series in Tanzania. Analysis of individual stations reveal an increasing trend in accumulated rainfall amount. The procedure applied to the analysis time interval 1961- 2014 and 1983-2014 shows that a Fourier series with single harmonics represents a good fit for explaining the variability of the occurrence intensity function $\lambda(t)$ for all the rain gauges. The fitted model shows an important increase of the ERE occurrence rate from the 90's, and even greater from the late 2010's. This result is obtained by carrying out both a classical approach concerning model validation and a more robust technique related to the property that a non- homogeneous Poisson process was used.

The theoretical distribution so obtained has been adopted to verify possible changes of $\lambda(t)$ function for the validation period, by using the autocorrelation analysis and uniform residuals approach to generate synthetic series of rainfall occurrences. The results showed that the differences between the observed and the fitted $\lambda(t)$ behaviour were statistically significant.

Moreover, a statistical test based on Mann Kendal test on rainfall amount has shown that, there is a statistical significant increasing trend in most regions in Tanzania. However, there is no statistically significant evidence of extreme rainfall occurrence process changes for more recent periods in the analyzed regions. However, for each station, estimated fitted parameters and confidence intervals are statistically significant, since all the estimates are inside their 95% confidence interval.

Further applications of the non-homogeneous Poisson model is needed to model extreme rainfall intensities and the changes in variance arising from underlying seasonal behavior.

References

[1] IFRC, International federation of red cross and red crescent societies. Tanzania: Floods (2010).

[2] T. Z. A. Operations, Tanzania : Floods (February) (2010) pp. 1–6.

[3] A. Davison, Modelling of Extreme Rainfall in Space and Time (2011) pp.1–24.

[4] L.-f. Chu, M. McAleer, C.-C. Chang, Statistical Modelling of Extreme Rainfall in Taiwan, Tinbergen Institute, *Tinbergen Institute Discussion Papers*: 13-006/III, 2013 15 (December,2012),(2013pp.6536.doi:10.7763/IJESD.2013.V4.302.

[5] I. Mugume, S. Shuanghe, T. Sulin, G. Mujuni, Analysis of temperature variability over desert and urban areas of northern china, *J Climatol Weather Forecasting* 4 (162) (2016)2.

[6] Y. Liu, P. Kokic, Predictive Inference for Spatio-temporal Precipitation Data and Its Extremes (2014) 1–27arXiv:1411.4715.

[7] A .C. Davison, R. Huser, Statistics of Extremes, *Annual Review of Statistics and Its Application*-doi:10.1179/003962659792003612.

[8] A. Kijazi, C. Reason, Relationships between intraseasonal rainfall variability of coastal Tanzania and Enso, *Theoretical and applied climatology* 82 (3-4) (2005) 153–176.

[9] A. L. Kondowe, Impact of Convective Parameterization Schemes on the Quality of Rainfall Forecast over Tanzania Using WRF-Model (June) (2014) 691–699.

[10] Kassile, Trend Analysis of Monthly Rainfall Data in Central Zone, *Journal of Mathematics and Statistics* 9 (1) (2013) 1–11. doi:10.3844/jmssp.2013.1.11.

[11] J. Hamisi, Study of rainfall trends and variability over Tanzania (PhD thesis).

[12] T. J. Ngailo, J. Reuder, E. Rutalebwa, S. Nyimvua, M. d.S. Mesquita, Modelling of extreme maximum rainfall using extreme value theory for tanzania, *International Journal of Scientific and Innovative Mathematical Research* 4 (3) (2016) 34–45.

[13] L. C. Drazek, Intensity estimation for poisson processes, Ph.D. thesis, The University of Leeds (2013).

[14] C. Agnese, G. Baiamonte, C. Cammalleri, D. Cat Berro, S. Ferraris, L. Mercalli, Statistical analysis of inter-arrival times of rainfall events for Italian sub-alpine and mediterranean areas, *Advances in Science and Research* 8 (1) (2012) 171–177.

[15] Y.-H. T. Hsien-Lun Wong, Shang-Hsing Hsieh, Application of non-homogeneous poisson process modeling to containership arrival rate, 2009 *fourth international conference on innovative computing, in-formation and control.*

[16] G. B. Malika Chassan, Jean-Marc Aza¨ıs, N. Suard, proportional hazard model for storm occurrence risk.

[17] J. Asin, A. Cebrian, Modeling and forecasting extreme hot events in the central ebro valley, a continental-Mediterranean area, *global and planetary change* (57), pp 43-58.

[18] B. Sirangelo, E. Ferrari, D. De Luca, Occurrence analysis of daily rain- falls through non-homogeneous poissonian processes, *Natural Hazards and Earth System Sciences* 11 (6) (2011) 1657–1668.

[19] G. Villarini, W. F. Krajewski, Empirically-based modeling of spatial sampling uncertainties associated with rainfall measurements by rain gauges, *Advances in Water Resources* 31 (7) (2008) 1015–1023.

[20] D. I. F. Grimes, E. Pardo-Iguzquiza, R. Bonifacio, Optimal areal rainfall estimation using raingauges and satellite data, *Journal of hydrology* 222 (1) (1999) 93–108

[21] L. M. Leemis, Nonparametric estimation of the cumulative intensity function for a non homogeneous poisson process, *Management Science* 37 (7) (1991) 886–900.

[22] S. Lee, J. R. Wilson, M. M. Crawford, Modeling and simulation of a non- homogeneous poisson process having cyclic behavior, *Communications in Statistics-Simulation and Computation* 20 (2-3) (1991) 777–809.

[23] Y. Gao, A. Parameswaran, Finish them!: Pricing algorithms for human computation, *Proceedings of the VLDB Endowment* 7 (14) (2014) 1965–1976.

[24] J. H. Cha, M. Finkelstein, New shock models based on the generalized polya process, European Journal of Operational Research 251 (1) (2016) 135–141.

[25] M. Liao, Applied stochastic processes, CRC Press, 2013.

[26] S. I. Vrieze, Model selection and psychological theory: a discussion of the differences between the akaike information criterion (aic) and the bayesian information criterion (bic)., *Psychological methods* 17 (2) (2012) 228.

[27] H. Bozdogan, Model selection and akaike's information criterion (aic): The general theory and its analytical extensions*, Psychometrika* 52 (3) (1987) 345–370.