Nat. Hazards Earth Syst. Sci. Discuss., 2, 5401–5425, 2014
www.nat-hazards-earth-syst-sci-discuss.net/2/5401/2014/
doi:10.5194/nhessd-2-5401-2014

Natural Hazards and Earth System Sciences Discussions

This discussion paper is/has been under review for the journal Natural Hazards and Earth System Sciences (NHESS). Please refer to the corresponding final paper in NHESS if available.

Estimating high quantiles of extreme flood heights in the lower Limpopo River basin of Mozambique using model based Bayesian approach

D. Maposa[1,*], J. J. Cochran[2], M. Lesaoana[1], and C. Sigauke[3]

[1]University of Limpopo, Sovenga, South Africa
[2]University of Alabama, Tuscaloosa, Alabama, USA
[3]University of Witwatersrand, Johannesburg, South Africa
[*]Ph.D. Candidate: University of Limpopo, Sovenga, South Africa

Received: 13 July 2014 – Accepted: 18 July 2014 – Published: 20 August 2014

Correspondence to: D. Maposa (danmaposa@gmail.com)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

In this paper we discuss a comparative analysis of the maximum likelihood (ML) and Bayesian parameter estimates of the generalised extreme value (GEV) distribution. We use a Markov Chain Monte Carlo (MCMC) Bayesian method to estimate the parameters of the GEV distribution in order to estimate extreme flood heights and their return periods in the lower Limpopo River basin of Mozambique. The return periods of extreme flood heights based on the Bayesian approach show an improvement over the frequentist approach based on the maximum likelihood estimation (MLE) method. However, both approaches indicate that the 13 m extreme flood height that occurred at Chokwe in the year 2000 due to cyclone Eline and Gloria had a return period in excess of 200 years, which implies that this event has a very small likelihood of being equalled or exceeded at least once in 200 years.

1 Introduction

Floods have increasingly become a common natural disaster in Southern Africa. Mozambique is one of the most affected countries in the region mainly due to its geographical position with nine transboundary rivers (Mabaso and Manyena, 2013; Spaliviero et al., 2014). Among the transboundary rivers in Mozambique, Zambezi River is the largest river in the territory followed by Limpopo River which is the second largest African river that drains to the Indian Ocean. Unlike the Zambezi River which is characterised by very large dams such as Kariba and Cohora Bassa, the Limpopo River has no large dams implying that the flow is not regulated. The hydrology of the Limpopo River basin is characterised by one cycle of rainfall that extends from October of the previous year to April of the following year with peak monthly totals in February, while the dry season runs from May to September (World Meteorological Organization [WMO], 2012). The Limpopo River is well pronounced by extreme natural hazards;

alternating between extreme floods and severe droughts (WMO, 2012; Spaliviero et al., 2014).

In this paper we address extreme floods in the lower Limpopo River basin of Mozambique. The Limpopo River separates South Africa from Botswana and Zimbabwe, and then flows eastwards through Mozambique to the Indian Ocean. Its catchment area distribution among Botswana, Mozambique, South Africa and Zimbabwe is 20, 20, 45 and 15 % respectively, with most of the catchment lying under semi-arid conditions (WMO, 2012; South Africa Limpopo, 2013; Spaliviero et al., 2014).

Gohil and Chowdhary (2013) define flood as a rare high event of a river usually as a result of extremely high rainfall triggered by unusual meteorological conditions. The Limpopo River basin in Mozambique experienced a series of disastrous floods over the past two decades; the most catastrophic and expensive of these floods were the year 2000 floods which killed more than 700 people, drowned more than 20 000 cattle and caused economic damages estimated at USD 500 million (Mabaso and Manyena, 2013; Maposa et al., 2013; Mondlane et al., 2013). In just over a decade after the disastrous floods of the year 2000, extreme floods revisited the Limpopo River basin in 2013 in the Chokwe district of Mozambique completely flooding the town of Chokwe including the General Hospital and other major facilities, and forcing two women to give birth on rooftops (Jackson, 2013; Maposa et al., 2013; Musiya, 2013). These events pose serious problems to the engineering structure by destroying bridges, roads and other major infrastructure in the towns and cities along the floodplains of the Limpopo River basin, and also cause major problems to the insurance sector and developed nations that help fund these insurance programmes. Maposa et al. (2013) reports that aid money can buy four times as much humanitarian impact if used before a disaster rather than on post disaster relief operations. Each year a disaster occurs, a substantial amount of money that has been originally designated for development in Mozambique gets diverted to relief and rehabilitation assistance.

The motivation for studying these extreme floods in the Limpopo River basin is to reduce the associated risk and mitigate the deleterious impacts of these floods on

humans and property. Several authors have provided results that portend that the frequency, magnitude and intensity of extreme weather events such as floods and temperature are on the rise (Mabaso and Manyena, 2013; Mondlane et al., 2013; WMO, 2013; Spaliviero et al., 2014). According to a unique survey of 139 National Meteorological and Hydrological Services carried out by WMO (2013) floods were the most frequently experienced extreme events over the course of the decade 2001–2010 worldwide including Africa. Munich Re (2013) affirms that natural catastrophic statistics for the year 2013 was dominated by floods that caused billions of American dollars in losses.

The Limpopo River basin has recently attracted researchers from various fields of specialisation ranging from academics to non-governmental organisations. Asante et al. (2007) studied and developed flood monitoring system from remotely sensed data for the Limpopo River basin in Mozambique. WMO (2012) proposes to improve the flood forecasting and early warning systems in the basin. Mujere (2011) argues that despite the accurate short term flood forecasts provided by meteorological forecasts, the shortage of time allowed for disaster preparedness and the incidence of false alarms lead people not to take the short term forecasts seriously. These short comings in flood forecasting justify the need to use statistical methods. Mabaso and Manyena (2013) advocate for contingency planning in Southern Africa to be considered as an event rather than a process in disaster preparedness and response planning in an effort to reduce disaster risk. Most recently Spaliviero et al. (2014) give a detailed account of flood risk analysis in the Limpopo River basin from a geosciences point of view based on the river's past evolution and geomorphological characteristics.

Mondlane et al. (2013) perform a comparative analysis of extreme flood frequency distribution models based on 20 years rainfall data recorded at Xai-Xai precipitation station in lower Limpopo River basin using the Gumbel, Fréchet, Pareto and Weibull distributions. The histogram of the collected data in Mondlane et al. (2013) showed a multi-modal distribution and the Gumbel Max distribution appeared to approximate its skewness better compared to other distributions in the paper while the two-parameter Weibull and Gumbel Min fitted the negatively skewed and unimodal distribution of the randomly simulated data. In a paper presented at the Extreme Value Analysis (2013) conference, Maposa et al. (2013) compare ten candidate distributions for their goodness of fit at Chokwe and Sicacate hydrometric stations based on over 50 years of annual daily maximum river flows (flood heights) using the generalised extreme value (GEV), generalised gamma (GG), Gumbel, two-parameter gamma (Gamma 2P), three-parameter gamma (Gamma 3P), two-parameter lognormal (LN2), three-parameter lognormal (LN3), log-Pearson type 3 (LP3), two-parameter Weibull and three-parameter Weibull distributions. Maposa et al. (2013) conclude that the GEV is more consistent at the two sites, with Gamma 3P, Gumbel Min and LN3 providing alternative models for the basin. Smithers (2012) details a comprehensive literature of advances that have been made to model extreme floods. However, Smithers (2012) stresses that the demand for reliable and improved estimates of flood frequency in terms of flood peaks and return periods have not been met and still poses a challenge in hydrology despite the improved understanding of the fundamental hydrological processes.

The purpose of this paper is to perform a comparative analysis of maximum likelihood and Bayesian estimates of the GEV distribution. We use Markov Chain Monte Carlo (MCMC) Bayesian inference to improve on the results achieved through fitting a GEV distribution with parameters estimated by the maximum likelihood estimation (MLE) method. Gaioni et al. (2010) allude that the GEV distribution arises naturally when modelling the maxima over a sequence of observations. Recently Ferreira and de Haan (2013) demonstrate conditions under which the block maxima method may prevail over the peaks-over-threshold method and derive the theoretical proofs based on probability weighted moment (PMW) estimators. Ferreira and de Haan (2013, p. 1) define block maxima approach in extreme value theory (EVT) as a method that "*consists of dividing the observation period into non-overlapping periods of equal size and restricts attention to the maximum observation in each period*". In hydrology a block is usually a year and the probability distribution of the new observations formed is assumed to approximately follow an extreme value distribution under extreme value conditions. In peaks-over-threshold approach observations that exceed a certain predetermined high

threshold are selected and the probability distribution of the selected observations of excesses over a pre-determined threshold is assumed to approximately follow a generalised Pareto distribution (GPD). The present paper concentrates on block maxima.

The MLE method is one of the most commonly used methods for estimating parameters when working with block maxima (Dombry, 2013; Ferreira and de Haan, 2013). The consistency of maximum likelihood estimators has recently been proved (Dombry, 2013). The present paper uses MLE and Bayesian methods to estimate the parameters of the GEV distribution. Gaioni et al. (2010) propose a model based Bayesian approach for direct quantile elicitation which translates into prior distribution assessment. Gaioni et al. (2010) argue that although the proposed approach is quite general, it is deemed particularly useful in river data cases in which direct assessments on the prior distribution are extremely difficult. Most recently Vidal (2014) emphasises the novelty of model based Bayesian inference approach when he used Bayesian analysis of the Gumbel distribution to analyse extreme rainfall data in Chile. Vidal (2014) leaves the use of GEV distribution in Bayesian analysis to further research. This provides some evidence that the application of this method in hydrology is still relatively new and thus supports its application in least developed countries such as Mozambique. In this study we propose to use the GEV distribution as the likelihood function and estimate the parameters using the MLE method and Bayesian parameter estimation method in order to use the prior to develop the predictive distribution which provides the basis for future expectations regarding the behaviour of the lower Limpopo River. The advantages of using Bayesian models are explained in Gaioni et al. (2010), and for further reading on Bayesian inference models we refer the reader to Beirlant et al. (2004) and Reiss and Thomas (2007). Some of the advantages of using Bayesian parameter estimation methods are the use of prior knowledge and that the modeller is able to capture uncertainty of the parameter estimates.

The rest of the paper is such that Sect. 2 explains the data and gives the theoretical framework of the statistical models used in the paper, Sect. 3 presents and discusses the results, Sect. 4 outlines the contributions of this paper to disaster risk reduction, and Sect. 5 gives the concluding remarks.

2 Materials and methods

In this section we present the data; probability framework of block maxima and the theoretical models applied later in this paper.

2.1 The data

The data used in this study were obtained from the Mozambique National Directorate of Water, which is the authority responsible for water management in Mozambique under the Ministry of Public Works. The data are hydrometric and measured in metres. Annual daily maximum flood heights recorded at Chokwe hydrometric station over the period 1951–2010 were used for this study. The block maxima approach was achieved by taking sequential steps to select the highest peak flood height in each hydrological block or year (Dombry, 2013; Ferreira and de Haan, 2013).

2.2 Probability framework of block maxima

Recent advances in block maxima are derived in Ferreira and de Haan (2013) and Dombry (2013). In hydrology it is natural that the observations are blocked by years, in particular, if the sample size is large enough.

The probability framework of block maxima is derived as follows:

Let $X_i = (X_1, X_2, \ldots, X_n) \overset{iid\ rv}{\sim} F$, each X_i representing the annual instantaneous daily observed flood heights (water levels).

Now let $M_n = \max(X_1, X_2, \ldots, X_n)$, then $P(M_n \leq x) = P(X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x) = P(X_1 \leq x) \times P(X_2 \leq x) \times \ldots \times P(X_n \leq x) = F^n(x)$, but F is unknown, so we approximate F^n by limit distributions as $n \to \infty$ (see proof in Ferreira and de Haan, 2013, p. 3). The

distribution of F is assumed to be in the domain of attraction of some extreme value distribution (Fisher and Tippett, 1928; Dombry, 2013; Ferreira and de Haan, 2013):

$$G_\xi(x) = \exp(-(1+\xi x)^{-1/\xi}), \xi \in \Re, 1+\xi x > 0, \tag{1}$$

where ξ is the extreme value index. The closest approximation of the extreme value distribution of F is the GEV distribution.

2.3 The GEV distribution model

The GEV cumulative distribution function, G, is given as:

$$G(x) = \begin{cases} \exp\left(-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right), \ 1+\xi\frac{x-\mu}{\sigma} > 0, \ \xi \neq 0, \\ \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right), \ x \in \Re, \ \xi = 0, \end{cases} \tag{2}$$

where μ, σ and ξ are the location, scale and shape parameters, respectively, and are estimated in this paper by MLE and MCMC Bayesian method (Gaioni et al., 2010; Dombry, 2013; Ferreira and de Haan, 2013; Vidal, 2014). In Eq. (2), we have the heavy-tailed Fréchet class of distributions if $\xi > 0$, the short-tailed Weibull class of distributions if $\xi < 0$, and the light-tailed Gumbel class of distributions if $\xi = 0$. The short-tailed Weibull class of distributions is bounded above by $\mu - \frac{\sigma}{\xi}$.

The estimates of extreme quantiles of the GEV are obtained from:

$$X_p = G^{-1}(x) = \begin{cases} \mu + \frac{\sigma}{\xi}\left[(-\ln(1-p_i))^{-\xi} - 1\right], \ \xi \neq 0, \\ \mu - \sigma \ln(-\ln(1-p_i)), \ \xi = 0, \end{cases} \tag{3}$$

where $p = P(X > x) = 1 - G(x)$ is the probability of exceedance. In Eq. (3) as $p \to 0$ and $\xi < 0$, we get $X_p = \mu - \frac{\sigma}{\xi}$.

2.4 Bayesian flood frequency model

The Bayesian estimation of a parameter vector $\boldsymbol{\theta} = (\mu, \sigma, \xi)$, in this paper, is given as:

$$f(\boldsymbol{\theta}_i|x) = \frac{f(\boldsymbol{\theta}_i)f(x|\boldsymbol{\theta}_i)}{f(x)} = \frac{f(\boldsymbol{\theta}_i)f(x|\boldsymbol{\theta}_i)}{\sum_j f(\boldsymbol{\theta}_j)f(x|\boldsymbol{\theta}_j)}, \text{ for a discrete parameter vector } \boldsymbol{\theta}, \text{ and}$$

$$f(\boldsymbol{\theta}|x) = \frac{f(\boldsymbol{\theta})f(x|\boldsymbol{\theta})}{f(x)} = \frac{f(\boldsymbol{\theta})f(x|\boldsymbol{\theta})}{\int f(\boldsymbol{\theta})f(x|\boldsymbol{\theta})d\boldsymbol{\theta}}, \text{ for a continuous parameter vector } \boldsymbol{\theta}, \tag{4}$$

where $f(\boldsymbol{\theta}), f(\boldsymbol{\theta}|x), f(x|\boldsymbol{\theta})$ and $f(x)$ is prior, posterior, likelihood and normalisation constant, respectively. It can be said that the posterior is proportional to the prior times the likelihood. In other words the posterior information is the combined sum of the prior and sample information (Beirlant et al., 2004; Reiss and Thomas, 2007; Vidal, 2014). Bayesian methods have the objective of computing the posterior distribution of the desired variables, in our case, the parameters of the annual daily maximum flood height distribution. Equation (4) can also be written in the form (Kwon et al., 2008):

$$P(\boldsymbol{\theta}|x) = \frac{P(\boldsymbol{\theta}) \times P(x|\boldsymbol{\theta})}{P(x)} = \frac{P(\boldsymbol{\theta}) \times P(x|\boldsymbol{\theta})}{\int_{\Phi} P(\boldsymbol{\theta}) \times P(x|\boldsymbol{\theta})d\boldsymbol{\theta}} \propto P(\boldsymbol{\theta}) \times P(x|\boldsymbol{\theta}), \tag{5}$$

where $\boldsymbol{\theta}$ is vector of the parameters of the distribution to be fitted (in this case GEV distribution), $P(\boldsymbol{\theta}|x)$ is the posterior distribution, x is the vector of observations, Φ is the space parameter, $P(\boldsymbol{\theta})$ is the prior distribution, and $P(x|\boldsymbol{\theta})$ is the likelihood function. The $100(1-\alpha)\%$ Bayesian credible set C (or in particular credible interval) is a subset of the space parameter Φ such that: $\int_C P(\boldsymbol{\theta}|x)\mathrm{d}\boldsymbol{\theta} = 1-\alpha$, where the sum replaces the integral if the space parameter Φ is discrete. The quantile-based credible intervals are such that if $\boldsymbol{\theta}_L^*$ is the $\alpha/2$ posterior quantile for $\boldsymbol{\theta}$, and $\boldsymbol{\theta}_U^*$ is the $1-\alpha/2$ posterior quantile for $\boldsymbol{\theta}$, then $\left(\boldsymbol{\theta}_L^*, \boldsymbol{\theta}_U^*\right)$ is the $100(1-\alpha)\%$ credible interval for $\boldsymbol{\theta}$. In this paper R programming package is used to produce these quantiles and to plot the return level of the posterior distribution.

3 Results and discussion

In this section we present, interpret and discuss the results of our analysis. All the results in this section were obtained using R statistical programming package and R Studio (R Core Team, 2013). Excel software was used to produce the time series plot of annual daily maximum flood heights at Chokwe hydrometric station.

3.1 Time series plot of the data

The results in Fig. 1 show that the highest peak flood height of magnitude 13 m occurred in the year 2000. In general, however, the time series plot of annual daily maximum flood heights does not indicate an upward trend of floods. In fact the flood heights at Chokwe hydrometric station for the period 1951–2010 are considerably random as indicated by the trend line R^2 of 0.26 % in Fig. 1.

3.2 Maximum likelihood estimation (MLE) approach model results

The GEV distribution was fitted using the R package Ismev (Heffernan and Stephenson, 2012). The parameters of the GEV distribution were estimated by the MLE method. The results obtained are presented in Table 1 and Fig. 3. Figure 2 presents results for the empirical distribution of the annual daily maximum flood heights at Chokwe hydrometric station. The boxplot, density plot and other plots in Fig. 2 show that the distribution of annual daily maximum flood heights at Chokwe is positively skewed and the 13 m flood height is an outlier. Table 1 presents the MLE estimates of the GEV parameters, the associated standard errors and the associated 95 % confidence intervals of the parameters μ, σ and ξ. The confidence interval for the population mean, μ, reveals that we can be 95 % certain that the true population mean of annual daily maximum flood heights at Chokwe lies between 3.76 m and 4.77 m high. The diagnostic plots in Fig. 3 indicate that the GEV model is a good fit as indicated by the probability plot and the probability density plot. The probability plot shows that the points are very close to the reference line (line best fit) and the probability density plot shows that the GEV mimics the shape of the empirical distribution represented by the histogram in Fig. 3. Both the return levels plot in Fig. 3 and results in Table 3 reveal that the 13 m flood height of the year 2000 has a return period in excess of 500 years based on the MLE method, implying that it was a very rare event.

Since $\xi < 0$, we can use the short-tailed Weibull class of distributions to model the annual daily maximum flood heights at Chokwe hydrometric station. The upper bound of the annual daily maximum flood heights at Chokwe based on the MLE parameters in Table 1 is $X_p = \mu - \frac{\sigma}{\xi} = 4.26452 - \frac{1.78893}{-0.08351} = 25.69$ m. This reveals that the 13 m flood height that occurred at Chokwe hydrometric station in the year 2000 is about half the upper bound annual daily maximum flood height expected at the site, although the 13 m flood height is way higher than rest of the flood heights for the period 1951–2010 as shown by the boxplot in Fig. 2. Results in Table 3 show that the 100 year flood height at Chokwe based on MLE approach is estimated to be 11.10 m. Any flood height above the 100 year flood level is considered to be very extreme (Coles, 2001; Beirlant et al., 2004; Reiss and Thomas, 2007).

3.3 Bayesian approach model results

The MLE estimates of the GEV parameters μ, σ and ξ were then used to run the Markov Chain Monte Carlo (MCMC) simulations to produce the traces and marginal posterior densities shown in Fig. 4. The trace plots in Fig. 4 show fast convergence and the results of the marginal posterior densities show that it is highly unlikely that the posterior estimate of μ will be below 3.5 m and highly unlikely that it will be above 5.0 m. Table 2 presents results for the Bayesian posterior parameter estimates of the GEV distribution, the associated naïve standard errors and the associated 95 % credible intervals of the parameters μ, σ and ξ. Credible intervals are the Bayesian analogue of confidence intervals. However, the interpretation differs and Gaioni et al. (2010) argues that one of the main advantages of Bayesian methods compared to frequentist methods is that credible intervals give more assurance than confidence intervals. For instance,

a 95 % credible interval for μ from Table 2 is interpreted as: the true population mean of the annual daily maximum flood heights at Chokwe lies between 3.76 m and 4.78 m with probability 0.95. It can also be noted that in this study the 95 % credible intervals for all the three parameters μ, σ and ξ in Table 2 are wider than the 95 % confidence intervals for the same parameters in Table 1, and all the upper limits of the 95 % credible intervals are greater than the corresponding upper limits for 95 % confidence intervals. These differences indicate that the Bayesian approach gives higher quantile estimates in the upper tail than the MLE frequentist approach as presented in Table 3. Modelling the upper tail is very important in flood frequency analysis as it results in reducing the impact of a flood event through reliable forecasting and disaster preparedness.

The return levels of the posterior distribution in Fig. 5 and results in Table 3 reveal that the 13 m annual daily maximum flood height event of the year 2000 caused by cyclone Eline and cyclone Gloria has a return period in excess of 200 years. The 95 % credible intervals in Fig. 5 gives a confirmation of our earlier discussion that the Bayesian approach gives higher quantile estimates in the upper tails than the frequentist approach shown in Fig. 3. Both the 95 % credible intervals in Table 2 and the 95 % confidence intervals in Table 1 for ξ change sign from negative to positive revealing the existence of all the three families of distributions encompassed in the GEV distribution. This means that using, for example, the Weibull distribution alone will not model all the characteristics of the distribution of annual daily maximum flood heights at the site.

Again since $\xi < 0$, the upper bound of the annual daily maximum flood heights at Chokwe hydrometric station based on the Bayesian approach parameter estimates in Table 2 is $X_p = \mu - \frac{\sigma}{\xi} = 4.27235 - \frac{1.90141}{-0.06824} = 32.14$ m. The upper bound based on Bayesian approach is about two-and-half times the 13 m flood height of the year 2000. The results in Table 3 show that the 100 year flood height based on Bayesian parameter estimation approach is 11.78 m which implies that the 13 m flood height of the year 2000 was indeed an extreme event (see boxplot in Fig. 2) despite the fact that it is almost one-third lower than the expected upper bound at the site based on Bayesian approach. Table 3 also reveals that the 13 m flood height has a return period in excess of 200 years based on Bayesian approach which is a shorter return period compared to the MLE method that is in excess of 500 years for the same flood height. The Bayesian estimates of annual daily maximum flood heights and their associated return periods (see Table 3) in this paper seem to be closer to reality as compared to the MLE approach. This coincides with the findings by Reis and Stedinger (2005).

4 Added value for the post 2015 framework for disaster risk Reduction

The work in this paper supports the implementation of the Hyogo Framework for Action (World Conference on Disaster Reduction, 2005) through addressing the gaps and challenges in reducing underlying risk factors; risk identification, monitoring and early warning; and preparedness for effective response and recovery. Accurate forecasts of the return periods of extreme floods can be used to reduce uncertainties associated with these natural hazards, thereby reducing the underlying risk factors and enabling people to better prepare for and respond to these rare events. Mozambique is one of the developing (or least developed) flood-prone countries in Southern Africa and therefore warrants particular attention because of its vulnerability and risk levels which far exceed its capacity to respond to and recover from flood disasters (World Conference on Disaster Reduction, 2005). Knowledge of the distribution of maximum flood heights helps in substantial reduction of flood disaster-related losses in lives, in the social, economic and environmental assets of Mozambique.

The work in this paper also contributes in the sharing of research findings, lessons learned and best practices which are some of the aspects needed to enhance international and regional cooperation and assistance in disaster risk reduction.

Disaster Risk Management in the Post 2015 Framework for Disaster Risk Reduction should continue to focus on knowledge sharing, international and regional cooperation and assistance in disaster risk reduction. The search for improved and reliable statistical techniques in long-term flood frequency forecasting should be considered as an ongoing process in disaster risk reduction.

5 Conclusions

In this paper we have considered the block maxima approach of extreme value theory, also known as at-site approach. The frequentist approach considered fitting the generalised extreme value distribution using the maximum likelihood estimation method. The choice of the GEV distribution over other alternative distributions for the lower Limpopo River at Chokwe hydrometric station was inspired by previous studies done by the authors at the same station and other stations in the lower Limpopo River basin. Diagnostic tests showed consistency of the goodness of fit of the GEV distribution at the site. The importance of the GEV distribution in modelling the lower Limpopo River annual daily maximum flood heights was revealed in this study through the use both credible and confidence intervals.

In an attempt to improve on the results achieved using the frequentist approach, a model based Markov Chain Monte Carlo Bayesian (MCMC) approach was applied to the data. The maximum likelihood estimates of parameters of the GEV distribution were used to develop the MCMC simulations in order to develop Bayesian posterior distribution estimates of the parameters of the GEV distribution. This study has revealed that Bayesian approach estimates of the return periods are shorter than those of the MLE approach. For instance, the 13 m flood height of the year 2000 has a return period in excess of 200 years based on Bayesian approach and in excess of 500 years based on the MLE approach. The study has also revealed that for a particular return period (or high quantile), Bayesian approach offers higher flood height estimates as compared to MLE approach when the GEV distribution is fitted to the Limpopo River data. Our findings suggest that the Bayesian approach improves on results achieved through the frequentist approach when using the GEV distribution as the likelihood function.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

Authors Contribution

D. Maposa contacted the research and did paper write-up as part of contributions towards his Ph.D. thesis. J. J. Cochran and M. Lesaoana, as Ph.D. supervisors for D. Maposa, suggested the use of Bayesian estimation parameter method to model the maximum flood heights, and the two supervisors also reviewed the draft manuscript to reach journal requirements and publishable standards. C. Sigauke helped with R programming in collaboration with D. Maposa.

Acknowledgements. The authors are indebted to the Mozambique National Directorate of Water (NAM), which is the authority responsible for water management under the Ministry of Public Works in Mozambique, for providing data used in this study. Special thanks go to Isac Filimone of NAM who went all his way to provide us with all the necessary data used in this study. We are also grateful to United Nations Office for the Coordination of Humanitarian Affairs-Southern Africa (OCHA) for providing us with weekly updates reports of floods in Southern Africa, particularly for the Limpopo River basin in Mozambique.

References

Asante, K. O., Macuacua, R. D., Artan, G. A., Lietzow, R. W., and Verdin, J. P.: Developing a flood monitoring system from remotely sensed data for the Limpopo basin, IEEE T. Geosci. Remote, 45, 1709–1714, 2007.

Beirlant, J., Goegebeur, Y., Segers, J., Waal, D. D., and Ferro, C.: Statistics of Extremes: Theory and Applications, John Wiley & Sons Ltd, Chichester, West Sussex, 2004.

Coles, S.: An Introduction to Statistical Modelling of Extreme Values, Springer-Verlag, London, 2001.

Dombry, C.: Maximum likelihood estimators for the extreme value index based on the block maxima method,ArXiv: 1301.5611v1 [math.PR], available at: http://arxiv.org/pdf/1301.5611.pdf (last access: 17 August 2014), 2013.

Ferreira, A. and de Haan, L.: On the block maxima method in extreme value theory, ArXiv: 1310.3222v1 [math.ST], available at: http://arxiv.org/pdf/1310.3222.pdf (last access: 17 August 2014), 2013.

Fisher, R. A. and Tippett, L. H. C.: Limiting forms of frequency distribution of the largest or smallest member of a sample, Cambridge Philosophical Soc., 24, 180–190, 1928.

Gaioni, E., Dey, D., and Ruggeri, F.: Bayesian modelling of flash floods using generalised extreme value distribution with prior elicitation, Chilean J. Statistics, 1, 75–90, 2010.

Gohil, R. B. and Chowdhary, D. R.: Study of flood frequency for Tan River at station Amba, Gujarat, Paripex, Indian J. Res., 2, 132–133, 2013.

Heffernan, J. E. and Stephenson, A. G.: Ismev: An Introduction to Statistical Modeling of Extreme values. R package version 1.39, available at: http://CRAN.R-project.org/package=ismev (last access: 17 August 2014), 2012.

Jackson, J.: Mozambique Floods Spur Roof Births, Ruin and Diarrhoea, Agence France-Presse (AFP), available at: http://reliefweb.int/report/mozambique/mozambique-floods-spur-roof-births-ruin-and-diarrhoea (last access: 17 August 2014), 2013.

Kwon, H.-H., Brown, C., and Lall, U.: Climate informed flood frequency analysis and prediction in Montana using hierarchical Bayesian modelling, Geophys. Res. Lett., 35, L05404, doi:10.1029/2007GL032220, 2008.

Mabaso, E. and Manyena, S. B.: Contingency planning in Southern Africa: events rather than processes?, Jamba, J. Disaster Risk Stud., 5, 95, doi:10.4102/jamba.v5i1.95, 2013.

Maposa, D., Cochran, J. J., and Lesaoana, M.: Investigating the goodness-of-fit of ten candidate distributions and estimating high quantiles of extreme floods in the lower Limpopo River basin, Mozambique, Presented at the 8th Conference on Extreme Value Analysis (EVA2013), 8–12 July 2013, Fundan University, Shanghai, China, p. 60, available at: http://people.math.gatech.edu/~peng/EVAFuDan/doc/EVA2013-ProgramAbstract.pdf (last access: 16 August 2014). 2013.

Mondlane, A. V., Hansson, K., and Popov, O.: Analysis of mathematical models and their application to extreme events, World Academy of Science, Engineering and Technology, International Journal of Mathematical, Computational Sci. Eng., 7, 378–387, 2013.

Mujere, N.: Flood frequency analysis using the Gumbel distribution, Int. J. Comput. Sci. Eng., 3, 2774–2778, 2011.

Munich Re.: Floods Dominate Natural Catastrophe Statistics in First Half of 2013, Press Release: 9 July 2013, Munich Reinsurance, Munich, Germany, 2013.

Musiya, T.: UN: Mozambique floods displace 150 k, leave 38 dead, Associated Press CHOKWE, Mozambique (AP), available at: http://article.wn.com/view/2013/01/29/UN_

Mozambique_floods_displace_150k_leave_38_dead_0/ (last access: 17 August 2014), 2013.

R Core Team.: R: a Language and Environment For Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, available at: http://www.R-project.org/ (last access: 16 August 2014), 2013.

Reis, D. S. and Stedinger, J. R.: Bayesian MCMC flood frequency analysis with historical information, J. Hydrol., 313, 97–116, 2005.

Reiss, R.-D. and Thomas, M.: Statistical Analysis of Extreme Values With Application to Insurance, Finance, Hydrology and Other Fields, 3rd edn., Birkhäuser Verlag, Basel, 2007.

Smithers, J. C.: Methods for design flood estimation in South Africa, Water SA, 38, 633–646, 2011.

South Africa Limpopo: The Limpopo River: 2013 Limpopo Tourism and Parks Board, available at: http://www.golimpopo.com/fact-details_limpopo-river_15.html (last access: 17 August 2014), 2013.

Spaliviero, M., De Dapper, M., and Maló, S.: Flood analysis of the Limpopo River basin through past evolution reconstruction and a geomorphological approach, Nat. Hazards Earth Syst. Sci., 14, 2027–2039, doi:10.5194/nhess-14-2027-2014, 2014.

Vidal, I.: A Bayesian analysis of the Gumbel distribution: an application to extreme rainfall data, Stoch. Env. Res. Risk A., 28, 571–582, doi:10.1007/s00477-013-0773-3, 2014.

WMO: Limpopo River Basin – a Proposal to Improve the Flood Forecasting and Early Warning System, World Meteorological Organization, with the support of Limpopo Water Course Secretariat and the Riparian States of Botswana, Mozambique, South Africa and Zimbabwe, 2012.

WMO: Global Climate 2001–2010: a Decade of Climate Extremes – Summary Report, WMO no. 119, World Meteorological Organization, Geneva, Switzerland, 2013.

World Conference on Disaster Reduction: Hyogo Framework for Action 2005–2015: Building the Resilience of Nations and Communities to Disasters, 18–22 January 2005, Kobe, Hyogo, Japan, available at: http://www.unisdr.org/wcdr (last access: 16 August 2014), 2005.

Table 1. Maximum likelihood parameter estimates of the GEV distribution.

Parameter	Estimate	Standard error (SE)	95 % *CI
μ	4.26452	0.25374	(3.7570, 4.7723)
σ	1.78893	0.17725	(1.4343, 2.1436)
ξ	−0.08351	0.07273	(−0.2290, 0.0620)

*CI means Confidence Interval.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

Table 2. Bayesian posterior parameter estimates of the GEV distribution.

Posterior parameter	Estimate	Naive SE	95 % *CI
μ	4.27235	0.00602	(3.7636, 4.7872)
σ	1.90141	0.00484	(1.5565, 2.4002)
ξ	−0.06824	0.00184	(−0.2027, 0.1046)

*CI means Credible Interval.

Table 3. In-sample evaluation of estimated tail quantiles at different probabilities.

Quantiles	Exceedance probability (p)	Return period	ML estimate (Exceedances)*	Bayesian estimate (Exceedances)*
95th	0.05	20 years	8.97 m (1)	9.38 m (1)
98th	0.02	50 years	10.22 m (1)	10.79 m (1)
99th	0.01	100 years	11.10 m (1)	11.78 m (1)
99.5th	0.005	200 years	11.92 m (1)	12.72 m (1)
99.6th	0.004	250 years	12.18 m (1)	13.02 m (0)
99.8th	0.002	500 years	12.94 m (1)	13.90 m (0)
99.9th	0.001	1000 years	13.65 m (0)	14.74 m (0)
99.99th	0.0001	10 000 years	15.76 m (0)	17.27 m (0)

(Exceedances)* refer to the number of sample observations above the estimated quantile (flood) level.

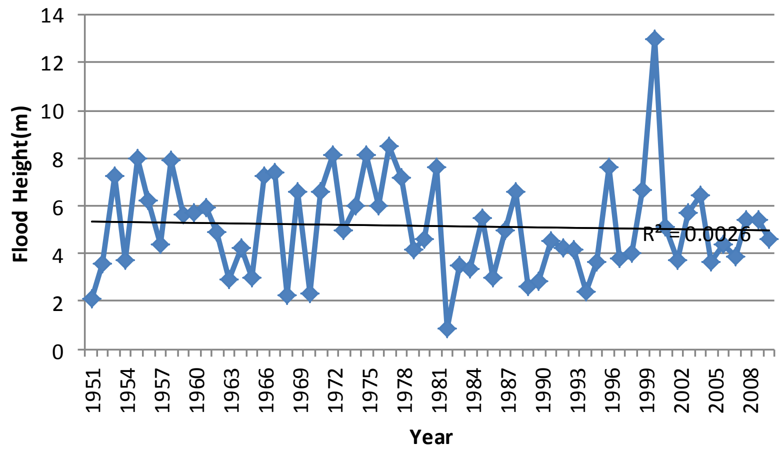

Figure 1. Time series plot of annual daily maximum flood heights at Chokwe hydrometric station, 1951–2010.

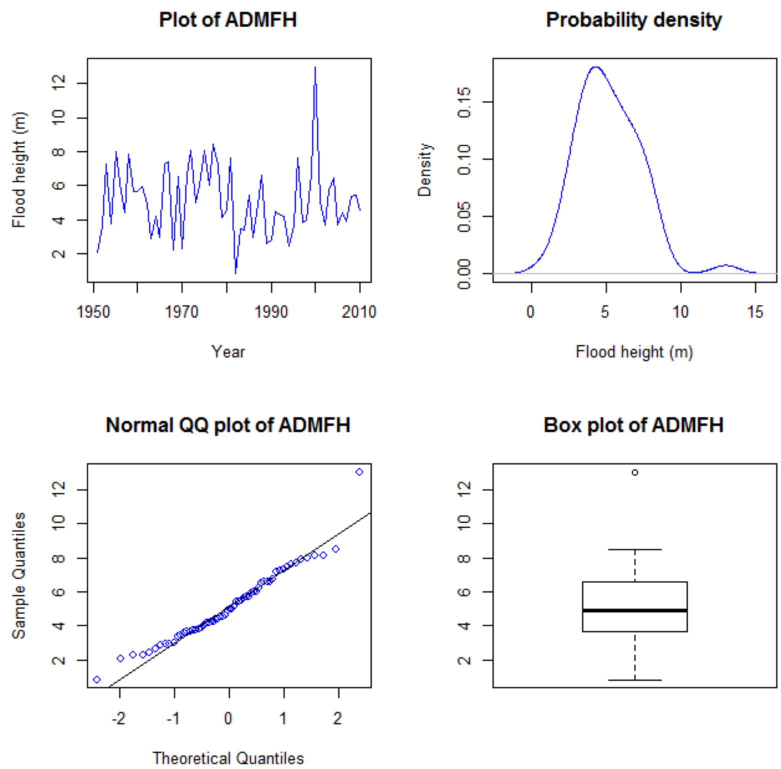

Figure 2. Plots of the empirical distribution of the annual daily maximum flood heights (ADMFH) at Chokwe hydrometric station.

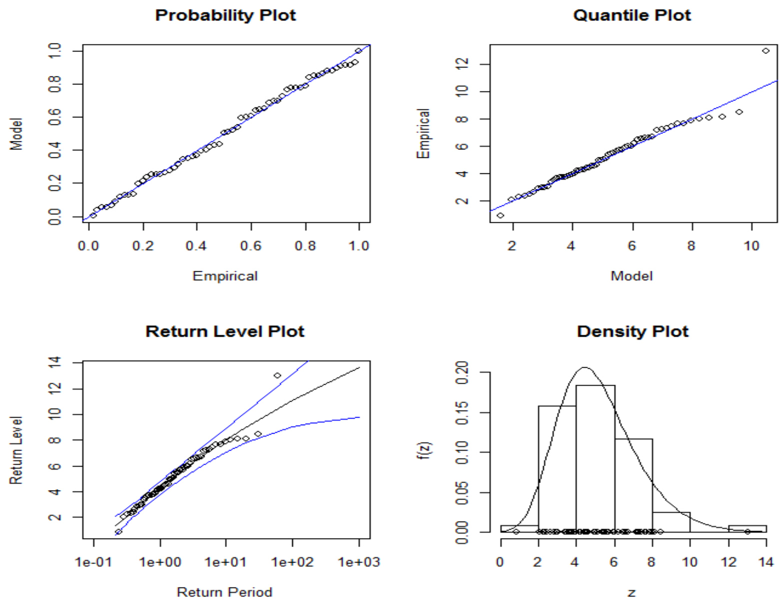

Figure 3. Diagnostic plots of the MLE approach GEV fit at Chokwe hydrometric station.

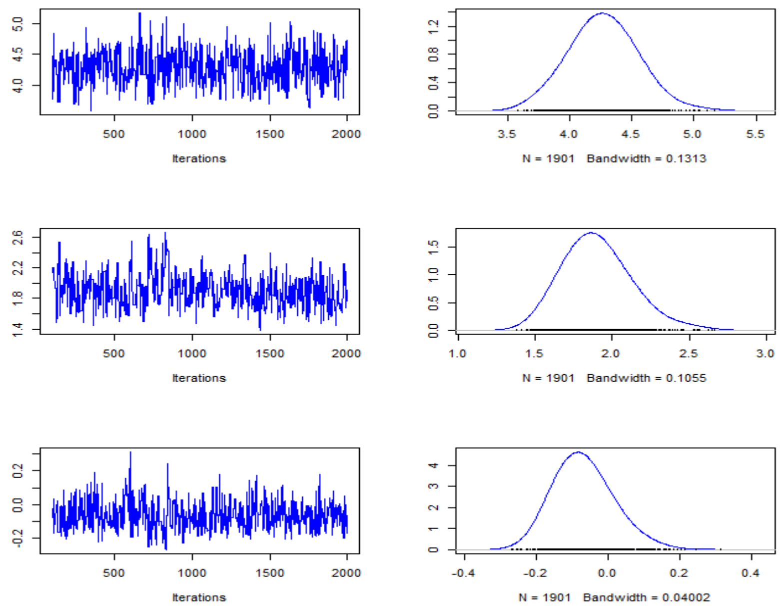

Figure 4. (a) Top panel: trace and marginal posterior density of the parameter μ, **(b)** middle panel: trace and marginal posterior density of the parameter σ and, **(c)** bottom panel: trace and marginal posterior density of the parameter ξ.

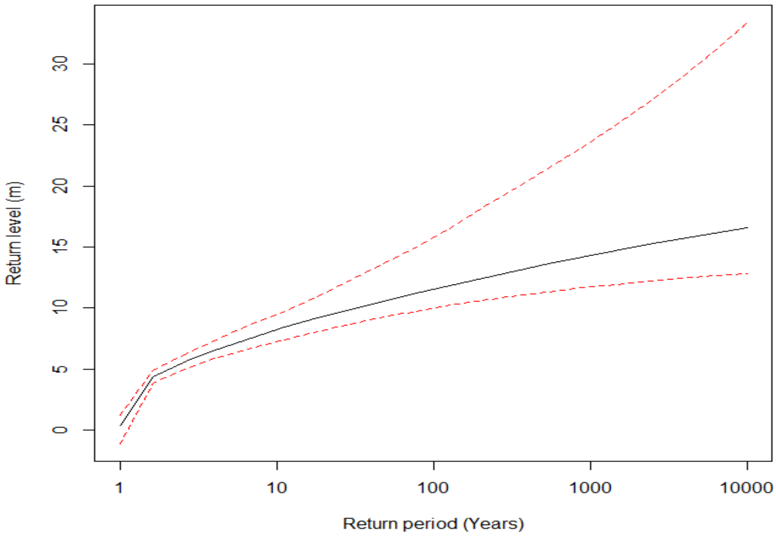

Figure 5. Return level plot of posterior distribution with 95 % Bayesian credible intervals (dashed lines) at Chokwe hydrometric station.