Construction of flood frequency curves in the lower Limpopo River basin of Mozambique using Bayesian and Markov chain Monte Carlo methods

Daniel Maposa*
University of Limpopo, Polokwane, South Africa – danmaposa@gmail.com
James J. Cochran
University of Alabama, Tuscaloosa, Alabama, USA– jcochran@cba.ua.edu
'Maseka Lesaoana
University of Limpopo, Polokwane, South Africa – lesaoana.maseka@ul.ac.za

Abstract

In this paper we discuss the disastrous extreme floods that have recently become a common feature in the economically challenged Southern African Development Community (SADC) member state of Mozambique. The year 2000 floods and recent 2013 floods killed a combined total of over 800 people, caused economic damages worth over USD550 million and led at least 3 women to give birth on rooftops and treetops in Chokwe district in lower Limpopo River basin of the country. We use the Markov chain Monte Carlo (MCMC) Bayesian method to estimate the parameters of the generalised extreme value distribution at a regional level in an attempt to develop the long term forecasts of the return periods of extreme floods in the vulnerable basin. We construct the flood frequency curves for the region which can be used to predict the return periods and their corresponding return levels in the region. It is hoped that these long term forecasts will complement the short term flood forecasting and early warning systems in the basin in reducing the associated risk and mitigating the deleterious impacts of these floods on humans and property.

Keywords: Bayesian inference; economic damages; roof births; flood frequency curves.

1. Introduction

Since the introduction of the first book on the applications of statistics of extremes mainly in hydrology and climatology by Emil Gumbel in 1958, statistics of extremes has played an important role in engineering practice for water resources design and management (Katz et al. 2002). Emil Gumbel noted that the oldest problems of extreme values arose from floods, however he was confronted with resistance from engineers and other practitioners in his attempt to link physical practice and extreme value theory hence he was compelled to make the following statement:

"*It seems the rivers know the theory. It only remains to convince the engineers of the validity of this analysis.*" (Alves and Neves, p.2, 2010).

The other opposition to the Gumbel distribution (named after Emil Gumbel) was its unboundedness which was questioned by many opponents of statistics of extremes who argued that a distribution of largest values could not be unbounded and Gumbel was again compelled to state that: "The exploration of how unlimited distributions behave at infinity is just part of the common general effort of mathematics and science to transgress the finite, as calculus has done since Newton's time for the infinite, and nuclear physics is doing for the infinitesimal." (Gumbel, p.2, 1958; Katz et al., p.1288, 2002).

In a paper he published in 1941 on the estimation of the return period of flood flow, the statistician and pioneer in the application of statistics of extremes, Emil Gumbel, warned that: "In order to apply any theory we have to suppose that the data are homogenous, i.e., no systematical change of climate and important change in the basin have occurred within the observation period and that no such change will take place in the period for which such extrapolations are made." (Katz et al., p.188, 2002).

Since the time of Emil Gumbel's quotes statistics of extremes has undergone extensive further development, including various distributions and approaches that incorporate nonstationarity, spatial extremes and spatial dependence (Coles, 2001, Katz, 2010; Ferreira and de Haan, 2013). Among the distributions developed after the Gumbel distribution, is the generalised extreme value (GEV) distribution which arises naturally when modelling the maxima or minima over a sequence of observations (Gaioni et al., 2010; Maposa et al., 2014b).

Smithers (2012) asserted that flood frequency analysis remains a field of great importance mainly because of its economic and environmental impact. In respect of advances in statistics of extremes methods for flood frequency analysis (FFA), South Africa is the most advanced while Mozambique is among the least developed in the Southern African Development Community (SADC) region which encompasses 15 member states including Botswana, Mozambique, South Africa and Zimbabwe. Literature on statistics of extremes is scarce in Mozambique, particularly in the lower Limpopo River basin (LLRB), with the exception of a few recent papers including Mondlane et al. (2013) and Maposa et al. (2014a and 2014b). Mondlane et al. (2013) studied the distribution of extreme precipitation using 20 years of rainfall data records at Xai-Xai in the LLRB while Maposa et al. (2014a and 2014b) used hydrometric flood heights to model the distribution of floods at Chokwe and Sicacate in the LLRB of Mozambique.

A lot of work and implementation has been done in the LLRB of Mozambique regarding meteorological forecasts that provide short term forecasts since the end of disastrous civil war in 1992 that lasted almost two decades (World Meteorological Organisation [WMO], 2012). The World Bank project installed 19 real time rain gauge stations in LLRB. Some flood forecasting and early warning models were also installed in Mozambique using Mike 11 'Flood Watch' and its upgrades and geo-spatial streamflow forecasting modelling, but less than 14% of the installed real time stations are still working and the flood forecasting and early warning system is struggling to be operational mainly due to inadequate project design, technical issues and maintenance problems (WMO, 2012). Some of the donors who provide financial and technical assistance to disaster risk management in the LLRB are the World Bank, British Department for International Development (DFID), United States Agency for International Development (USAID), Germany's Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) and African Development Bank. However, little, if any, funding has been channelled towards advancement of statistics of extremes in the basin.

The main purpose of this paper is to construct flood frequency curves for the LLRB of Mozambique. We also wish to suggest improvements in the modelling of extreme floods in the basin. The interest in the LLRB is mainly driven by the low lying nature of the region across the coastal floodplain that makes it susceptible during periods of high flows, posing a major problem in the lower part of the basin. The basin is characterised by extreme natural hazards, alternating between extreme floods and severe droughts. The high incidence of flooding is mainly attributed to tropical cyclones that form in the Indian Ocean coast of Mozambique. According to historical records on natural disasters over the past 58 years (1956-2013), Mozambique experienced 10 droughts, 23 flood events, 14 tropical cyclones, 18 epidemics and one earthquake (Mozambique Floods 2013). During the 2013 flood it was reported that two women gave birth on rooftops in Chokwe district and over 100 people were killed (Jackson, 2013). The most catastrophic and expensive of these natural disasters were the flood events of the year 2000 which killed more than 700 people and caused economic damages estimated at USD500 million in the LLRB of Mozambique alone and one woman was reported to give birth on a tree top. The Limpopo River has a history of worst floods in the country. The flood that occurred in February 1977 in the river was reported to be the worst that occurred in Mozambique until that date, but it was surpassed by the year 2000 flood which reached a maximum height of 13 m in Chokwe district, completely flooding the town of Chokwe and parts of Xai-Xai city in Mozambique and shared water with Incomati River which had never happened before (WMO, 2012).

2. Materials and Methods

2.1 Available data

The data used in this study is hydrometric data for the LLRB at Combomune, Chokwe, Pafuri and Sicacate hydrometric stations obtained from the Mozambique National Directorate of Water (DNA) which is the authority responsible for water management in Mozambique. The uninterrupted flood water level data series for Chokwe (E35) was from 1951 to 2010, Sicacate (E36) from 1952 to 2010, Combomune (E33) from 1966 to 2010, Pafuri (E31) from 1953-2010, generating 60, 59, 45 and 58 years of annual maximum series, respectively. The hydrometric data, in its original form, was daily flood heights measured in metres. Sequential steps were taken to obtain the annual daily maximum flood heights.

2.2 Study area

The Limpopo River rises at the confluence of Marico and Crocodile rivers at the limit of North West Province in South Africa and flows northeast along the border of South Africa with Botswana. The name Limpopo is a modified version of the original Sepedi name *diphororo tsâ meetse* meaning 'gushing strong waterfalls' (Chilundo et al., 2008). The river separates South Africa from Botswana and Zimbabwe by creating an international border of nearly 900 km, and then flows eastwards through Mozambique to the Indian Ocean. 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 (WMO, 2012).

The LLRB is divided into three main geographic regions: the upper Limpopo (which stretches from the headwaters down to the Shashe River confluence at the border of Botswana, South Africa and Zimbabwe), the middle Limpopo (which is located between the Shashe River confluence and Pafuri) and the lower Limpopo (which stretches downstream from Pafuri to the mouth of the river on the Indian Ocean), (Chilundo et al., 2008).

The LLRB, where the study takes place, is further divided into three major climatic zones:

- The coastal Xai-Xai zone, which is situated in the lower portion of the region between Xai-Xai and the Limpopo River's connection to the sea.
- The lower Limpopo Valley (including the Chokwe and Sicacate hydrometric stations), which is situated in the area between Xai-Xai and Macarretane Dam.
- The upper Limpopo Valley (including Combomune and Pafuri hydrometric stations), which is the area from Macarretane Dam up to Mozambican border at Pafuri.

2.3 Research methodology

Design flood estimation procedures are categorised into two broad areas known as methods based on analysis of observed flood (or streamflow) data and rainfall-based methods (Smithers, 2012). Since our data is hydrometric flood heights we use methods based on the analysis of floods. Smithers (2012) pointed out that FFA of observed flows methods are probabilistic in nature and hence are suitable for design floods estimation. FFA may be performed at a single site (commonly known as at-site) when long records of streamflow data are available at a site or at regional level which utilises data from several sites to estimate the flood frequency distribution of observed data at each site. Previous studies have argued that the regional approach is more preferred in practice than the at-site approach. This paper considers the regional approach. The streamflow data records are sufficiently long enough for the 4 sites implying that the regional approach results should be closer to the at-site findings. Regionalisation of groups (stations) into homogeneous regions was done by subjective judgement using prior information of the frequency distributions at the sites and geographic information (Kachroo et al., 2000; Haile, 2011). It must be conceded that in our follow up paper, more detailed regionalisation approaches are going to be considered for identification of homogenous regions.

The GEV distribution has been identified to fit the annual maxima data in the basin (Maposa et al., 2014a and 2014b). Let $Y_1,\ldots,Y_m$ be independent and identically distributed (i.i.d.) daily flood heights. Let X_i denote the annual maxima for year i (Charpentier and Sibaï, 2008). We consider the i.i.d. random variables $(X_i)_{i\geq 1}$ with common distribution $F \in D(G_{\xi_0})$ and corresponding normalisation sequences $(a_m > 0)$ and b_m such that $\lim_{m\to\infty} F^m(a_m + b_m) = F_\xi(x),\ x \in \Re.$ [1]

The distribution function F in Eqn. [1] satisfies the extreme value condition with index ξ or equivalently F belongs to the domain of attraction of G_ξ (Fisher and Tippett, 1928; Coles, 2001; Maposa et al., 2014b). The closest approximation of the extreme value distribution of F is the max-stable GEV distribution with parameters a_m, b_m and ξ_0 commonly estimated by the maximum likelihood method (Coles, 2001; Maposa et al., 2014b). The GEV cumulative distribution function, G is given in Eqn. [2] as:

$$G_{(\mu,\sigma,\xi)}(x) = \begin{cases} \exp\left(-\left(1+\xi\dfrac{x-\mu}{\sigma}\right)^{-1/\xi}\right), & 1+\xi\dfrac{x-\mu}{\sigma} > 0,\ \xi \neq 0, \\ \exp\left(-\exp\left(-\dfrac{x-\mu}{\sigma}\right)\right), & x \in \Re,\ \xi = 0. \end{cases} \qquad [2]$$

where μ, σ and ξ in Eqn. [2] are the location, scale and shape parameters, respectively, as in Maposa et al. (2014b). The parameters of the GEV in this study are to be estimated by the maximum likelihood (ML) method and Markov chain Monte Carlo (MCMC) Bayesian method. In Eqn. [2] we have the Fréchet class of distributions if $\xi > 0$, the Weibull class of distributions if $\xi < 0$, and the Gumbel class of distributions if $\xi = 0$.

We use the trivariate normal prior (conjugate prior) discussed in Coles and Powell (1996), Stephenson and Ribatet (2006) and Sigauke (2014). Let the prior distribution be denoted by $\pi(\boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\mu,\sigma,\xi)$ is a vector of the parameters to be estimated. Let $\boldsymbol{\theta}' = (\mu, \log\sigma, \xi)$, then the prior distribution on $\boldsymbol{\theta}$ is defined as (Stephenson and Ribatet, 2006; Sigauke, 2014):

$$\pi(\boldsymbol{\theta}') \propto \frac{1}{\sigma}\exp\left\{-\frac{1}{2}(\boldsymbol{\theta}'-\mathbf{v})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}'-\mathbf{v})\right\} \qquad [3]$$

where $\mathbf{v}$ is a mean vector and $\boldsymbol{\Sigma}^{-1}$ is the covariance matrix. The likelihood function is given by:

$$\pi(x/\boldsymbol{\theta}) = \prod_{i=1}^{n}\frac{1}{\sigma}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-1/\xi-1}\exp\left\{-\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \qquad [4]$$

The joint posterior density in Eqn. [5] below is:

$$\pi(\boldsymbol{\theta}/x) \propto \frac{1}{\sigma^{n+1}}\exp\left\{-\frac{1}{2}(\boldsymbol{\theta}'-\mathbf{v})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}'-\mathbf{v}) - \sum_{i=1}^{n}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \times \prod_{i=1}^{n}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-1/\xi-1}$$

The initial values of the Markov chain Monte Carlo (MCMC) are denoted by $\boldsymbol{\theta}_0 = (\mu_0, \sigma_0, \xi_0)$ and $s = (s_\mu, s_\sigma, s_\xi)$. We use ML estimates for these initial values (Stephenson and Ribatet, 2006; Sigauke, 2014).

3. Results and discussion

Results obtained from the methods proposed in Section 2 are presented in this section. The regional analysis classified the 4 sites into two homogeneous groups (Group1: Chokwe and Sicacate, and Group2: Combomune and Pafuri).

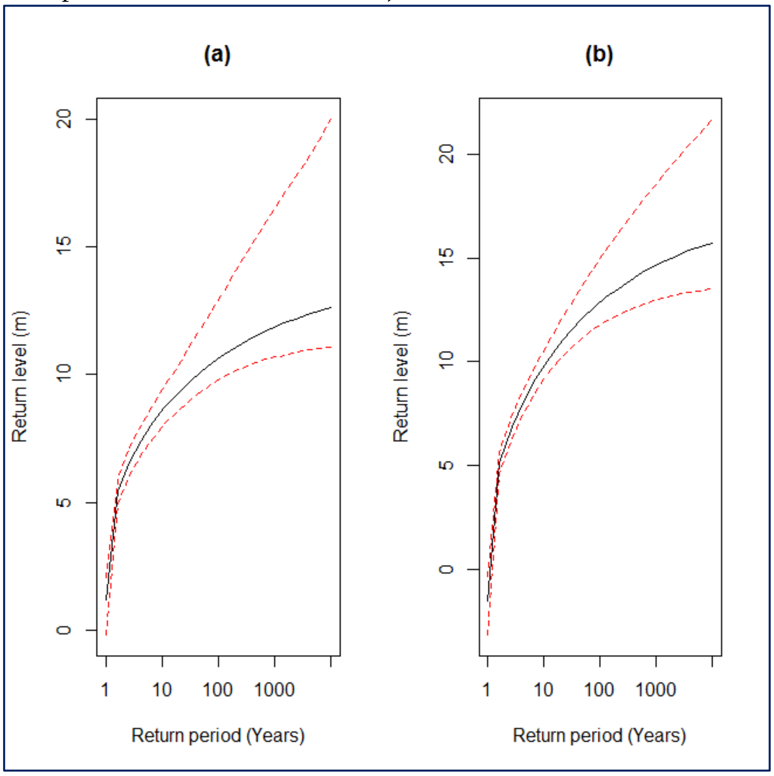

Figure 1: Flood frequency curves of posterior distribution with 95% Bayesian credible intervals (dashed lines) at (a): upper Limpopo Valley (Combomune & Pafuri) (b): lower Limpopo Valley (Chokwe & Sicacate hydrometric stations).

The results of the flood frequency curves in Figure 1 reveal that the 100-year flood height at the upper Limpopo Valley (Combomune & Pafuri) is approximately10.39 m which is less than the highest flood height of 10.97 m that occurred in the valley at Combomune in 2000. The 100-year flood height at lower Limpopo Valley (Chokwe & Sicacate) is approximately 12.72 m which is also below the 13 m flood height that occurred in the year 2000 (with a return period of approximately in excess of 135 years). These findings imply that engineering structures such as bridges, roads, factories, houses, spillways, dams and cities that were built based on the standard 100-year flood level recommended by engineers were destroyed and this indeed explains the high value of the economic damages and deaths that occurred in the LLRB mainly in Chokwe district with reported births on treetops.

5. Conclusions

The principal aim of this study was to develop flood frequency curves in the lower Limpopo River basin with the purpose of making contributions in the long term forecasts of floods in the basin to complement the much established and sponsored short term forecasts in the region. The regional flood frequency curves were developed for the lower Limpopo River basin and these curves appear to be consistent with the occurrence of maximum daily flood heights in the homogeneous regions identified. It can be concluded that the findings in this study are promising for this vulnerable part of the basin.

Further research in the lower Limpopo River basin will advance this study to consider the application of Bayesian inference and Markov chain Monte Carlo methods to the nonstationary models in location and scale parameters, regional trend analysis and spatial extremes.

References

Alves, I.F. and Neves, C. (2010). Extreme value distributions. Available at: http://docentes.deio.fc.ul.pt/fragaalves/fraga_alves_lexicon.pdf. (last access: 29 January 2015).

Charpentier, A. and Sibaï, D. (2008). Dynamic flooding modelling combining Hurst and Gumbel's approach. *Environmetrics*, doi:10.1002/env.909.

Chilundo, M., Kelderman, P. and 'Keeffe, J.H.O. (2008). Design of a water quality monitoring network for the Limpopo River basin in Mozambique. *Physics and Chemistry of the Earth, 33(8-13), 655-665.*

Coles, S. (2001). An introduction to statistical modelling of extreme values. Springer-Verlag, London.

Coles, S.G. and Powell, E.A. (1996). Bayesian methods in extreme value modelling: A review and new developments. *International Statistical Review*, 64, 119-136.

Ferreira, A. and de Haan, L. (2013). 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 September 2014).

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

Gaioni, E., Dey, D. and Ruggeri, F. (2010). Bayesian modelling of flash floods using generalised extreme value distribution with prior elicitation. *Chilean Journal of Statistics*, 1(1), 75-90.

Gumbel, E.J. (1958). Statistics of extremes. New York, Columbia University Press.

Haile, A.T. (2011). *Regional flood frequency analysis in Southern Africa*. MSc Thesis, Department of Geosciences, Faculty of Mathematics and Natural Sciences, University of Oslo.

Jackson, J. (2013). Mozambique floods spur roof births, ruin and diarrhoea. Agence France-Presse (AFP). http://reliefweb.int/report/mozambique/mozambique-floods-spur-roof-births-ruin-and-diarrhoea (last access: 19 January 2015).

Kachroo, R.K, Mkhandi, S.H. and Parida, B.P. (2000). Flood frequency analysis of Southern Africa: I. delineation of homogeneous regions. *Hydrological Sciences Journal – Journal Des Sciences Hydrologiques*, 45(3), 437-447.

Katz, W.K. (2010). Statistics of extremes in climate change. *Climatic Change*, 100, 71-76.

Katz, W.K., Parlange, M. B. and Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, 25, 1287-1304.

Maposa, D., Cochran, J.J. and Lesaoana, M. (2014a). Investigating the goodness-of-fit of ten candidate distributions and estimating high quantiles of extreme floods in the lower Limpopo River basin, Mozambique. *Journal of Statistics and Management Systems*, 17(3), 265-283, doi:10.1080/09720510.2014.927602.

Maposa, D., Cochran, J.J., Lesaoana, M. and Sigauke, C. (2014b). Estimating high quantiles of extreme floods in the lower Limpopo River of Mozambique using model based Bayesian approach. *Natural Hazards and Earth System Sciences Discussions*, 2, 5401-5425, doi:10.5194/nhessd-2-5401-2014.

Mondlane, A.V., Hansson, K. and Popov, O. (2013). Analysis of mathematical models and their application to extreme events. *World Academy of Science, Engineering and Technology; International Journal of Mathematical, Computational Science and Engineering*, 7(4), 378-387.

Mozambique Floods 2013. Consolidated early recovery strategy. Humanitarian Country Team, Maputo, 25 April 2013

Sigauke, C. (2014). Modelling electricity demand in South Africa. PhD Thesis, University of Free State.

Smithers, J.C. (2012). Methods for design flood estimation in Southern Africa. *Water SA*, 38(4), 633-646.

Stephenson, A.G. and Ribatet, M.A. (2006). A User's Guide to the evdbayes package ,Version 1.1. Available at: http://cran.r-project.org/ (last access: 29 January 105).

WMO. (2012). 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.