

Editor-in-Chief
Nat. Hazards Earth Syst. Sci.

Ms. Ref. No.: nhess-2014-197; D. Maposa et al., doi:10.5194/nhessd-2-5401-2014.

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

Dear Editor-in-Chief

We would like to thank you for your patience. We extend our thanks and appreciation to the anonymous reviewers and J.L. Salinas for helpful comments and suggestions which have resulted in much improvement of our work. We appreciate that the reviewers have seen the paper in a different light and offered very helpful comments and suggestions. We have tried to address all the issues raised by the reviewers. We have made the following corrections (in italics) in response to the suggestions made by the reviewers. The reviewer's comments are retained for easy reference.

COMMENTS ON THE MANUSCRIPT OF REVIEWER 1: Anonymous (published 31 August 2014)

• **General comments and responses**

As a conclusion, the scientific contribution of this paper is limited and the approach used also need to be improved. Thus the current manuscript is not suggested to publish in NHESS.

The approach used will be improved; in particular, a brief subsection on the prior distribution methodology used in the manuscript will be added. This is in line with the suggestions by the other two reviewers. A description of how the priors were obtained will likely improve the scientific contribution of the paper. It is our hope that the manuscript will be ready for publication in the NHESS once these ambiguities are cleared.

• **Specific comments and responses:**

Technical approach: Authors argued that Bayesian approach improves the results because it offers higher flood height estimates compared with MLE. This is problematic since it is hard to judge whether the Bayesian approach over-estimate the flood height or the MLE under-estimate the flood height.

In line with comments from other reviewers, a separate nearby site will be used to check the consistency of the results of the two approaches.

Possible improvement on analyzing extreme flood risks:

- Include the information of human activity, climate variability/change etc.

The reviewer would agree with us that information such as human activity is very difficult to include in an extreme value model, the only closest way to cater for such information is the use of Bayesian inference which is what we did in this paper (see Gaioni et al., 2010). We will shade more light on how the prior distribution was obtained.

- To provide a more reliable estimation on the extreme quantiles, authors could investigate more on the shape parameter, which controls the tail behavior of the distribution. (see e.g. Martins and

Stedinger (2000), DOI: 10.1029/2001WR000367). Otherwise, authors could also use some expert priors for the Bayesian approach to improve the estimation of shape parameters.

Our additional subsection, as suggested by the other reviewers, on the prior distribution procedure used will help to respond to the suggestion raised above.

- The uncertainty of single-site analysis is usually large, which leads to a great limitation on estimating the extreme events. Thus it could be better to consider a multi-site regional model to reduce the uncertainty (e.g. Renard, B., et al. (2006)).

Renard, B., et al. (2006). "An application of Bayesian analysis and Markov chain Monte Carlo methods to the estimation of a regional trend in annual maxima." *Water Resources Research* 42(12): W12422.

*We thank the reviewer for providing us with useful references such as Renard et al. (2006) for further reading to improve on the contribution of our paper, and possible citation. We are also aware of the endless debate that exists between single-site analysis and regional analysis, with no winner between the two approaches (see also Smithers J.C., 2012, *Water SA* Vol. 38 No. 4, <http://dx.doi.org.10.4314/wsa.v38i4.19>). Our thrust in this paper is on single-site analysis for similar reasons given in Gaioni et al. (2010) and that we have sufficiently long records of data at a single site. To check for consistency in the two approaches the findings will be compared to the findings from a nearby site. We believe this will significantly improve the scientific contribution of the paper particularly in most economically challenged African countries where data records in the neighbourhood of a particular site may be scarce or simply unavailable.*

COMMENTS ON THE MANUSCRIPT OF REVIEWER 2: J.L. Salinas (published 18 September 2014).

• General comments and responses

.... Namely one can not argue that the Bayesian approach improves the estimates of extreme water levels, as compared with MLE approach only because they are larger. ... C2115 line 3

Smith (1987) shows that estimation of extreme value distribution parameters with the maximum likelihood method is a non-regular problem when the shape parameter (extreme value index) is less than $-\frac{1}{2}$. This means that certain regularity conditions have to be met. It is therefore more attractive to use the Bayesian approach which does not depend on these regularity conditions (Beirlant et al., 2004; Coles, 2004; Ardia and Hoogerheide, 2010; among others). Since our study involves tail estimation based on data sets with little information available, the Bayes approach can be used to capture and take into account all the available information including additional information through prior elicitation (Beirlant et al., 2004; Coles, 2004; among others).

Bayarri and Berger (2004) discuss the interplay of Bayesian and frequentist analysis, and identify the most common areas of important and useful connections between the two as the scenarios when no external information, except data and the model itself, is to be introduced into the model. These same authors (Bayarri and Berger) indicated that there are many areas of frequentist methodology that can be replaced by Bayesian methodology that provide superior answers. This clearly reveals that our findings are not strange, although we agree that more explanations on the prior distribution methodology (an additional subsection) are needed to add to the current manuscript. A clear discussion of the prior distribution used (trivariate normal prior or conjugate prior) will be given in the final improved manuscript.

...The authors are only applying two of the already existent (and well established) methods in parameter estimation for the GEV distribution.....

The thrust in this paper is not to prove a new theorem, but to extend the application of the two existing methods to new circumstances to estimate and develop flood frequency curves in developing and economically challenged countries like Mozambique where availability of river discharges/volumes and rainfall data records is limited. The main quality data records the country has is flood height (hydrometric) data. The application of spatial extreme value theory in such countries is also limited due to small number of gauging sites in the neighbourhood of the single-site with long records of water level (flood height) data sets. It was interesting to find that the findings in this paper which is based on flood height data in Mozambique using at-site approach are comparable to the regional analysis findings in more developed South Africa based on river discharges particularly with regard to the return period of the year 2000 flood level.

The reviewer, Jose Salinas, also wrote "...if the authors want to publish this "difference" in estimates, they should e.g. find the reason for this difference. Is it mainly due to the outlier of year 2000? Does this also happen at a regional scale (check if in streamgauges nearby Bayesian systematically gives higher estimates)?..."

To answer the questions raised in the above paragraph, a comparison of the estimates will be made with a nearby downstream site in the same river at Sicacate hydrometric station.

- **Specific issues:**

Last sentence of Abstract is trivial [, which implies....]

Avoid the expression "had a return period in excess of XXX years", use "had a return of XXX years"

page 5402 Line 21 [The hydrology...] to page 5403 line 8 should go to section 2, new subsection defining the case study.

page 5405, lines 4 to 8, irrelevant to list all the distributions.

page 5405, line 19 to page 5406 line 3, trivial to 99.99% of the readers.

I would finish the introduction with page 5405, line 15-18, so I would move page 5406 lines 4-26 before this (and also explain here shortly the MCMC approach).

Page 5407, sect 2.2 trivial, I would substitute this section with the basics of the MLE approach, which is missing in the manuscript

Page 5409, a mention to the prior distribution used is needed here.

Plots - I would like to see the two different flood frequency curves in the same plot, to ease the comparison

- **Response to all the specific issues of Reviewer 2: Jose L. Salinas**

We will improve our writing style and a few typos in the current version of the manuscript will also be corrected.

We thank the reviewer, Jose L. Salinas for making suggestions to us to read the articles by Viglione, Merz, Salinas and Blöschl (2013), Martins and Stedinger (2000), Martins and Stedinger

(2001), and Gaume et al. (2010). These articles are good and have improved our understanding, and some will be cited in our final manuscript.

COMMENTS ON THE MANUSCRIPT OF REVIEWER 3: Anonymous (published 30 September 2014).

- **General comments and responses**

... Firstly, I must admit, that I have a dilemma how to assess the paper. On one hand the paper is pretty well written and important for the hydrology of South Africa region, but on the other hand, I have not spotted a single new idea in the manuscript! Both methods Authors present are very well known and fossilised in hydrological sciences. ...

Like our previous response to Jose Salinas, the thrust in this paper is not to prove a new theorem, but to extend the application of the two existing methods to new circumstances to estimate and develop flood frequency curves in developing and economically challenged countries like Mozambique where availability of river discharges/volumes and rainfall data records is limited. The main quality data records the country has is flood height (hydrometric) data. The application of spatial extreme value theory in such countries is also limited due to small number of gauging sites in the neighbourhood of a single-site with long records of water level (flood height) data sets. It was interesting to find that the findings in this paper which is based on flood height data in Mozambique using at-site approach are comparable to the regional analysis findings in more developed South Africa based on river discharges particularly with regard to the return period of the year 2000 flood level.

While the main thrust of the paper is on the application, the Markov chain Monte Carlo method algorithm and the prior distribution used will be explained in the revised/improved version of the manuscript. This, in our view, will improve the scientific contribution of the manuscript particularly in developing countries such as most countries in Southern Africa.

... I would suggest to rephrase the article to underline and stress the novelty of the research. ... In my opinion the Introduction is too wordy – the whole paper would benefit, if Authors considered to cut the long story short and limit this chapter only to the most relevant issues. The important issues, such as goal of the paper and motivation of the Authors drown in myriads of unimportant or easy-to-check facts. I would suggest to divide the Introduction into several shorter chapters. ... In fact, the recommendation to shorten a paper concerns all the manuscript.

These suggestions coincide with those by Jose Salinas. Indeed we are going to rephrase most parts of the article mainly by following the suggestions given by Jose Salinas (which are closer to our original paper) and incorporating the ideas suggested by the Anonymous Reviewer in the paragraph above.

... I also wonder what distribution functions and what parameter the Authors used as the priors?

The trivariate normal prior (conjugate prior) used and Markov chain Monte Carlo algorithms used will be explained in the final improved paper.

... Besides it is worth mentioning that the ML method concentrates on main mass probability and therefore it is not quite recommended for estimation of high quantiles. ... The Authors select the heavy-tailed GEV distribution function as the underlying model, but I am not convinced by the arguments supporting this choice. What could be expected ...

These concerns raised above are problematic because the Anonymous Reviewer#3 here seems to be in contradiction with most revered literature and recently re-affirmed (see also Dombry, 2013

<http://arxiv.org/pdf/1301.5611.pdf> ; Ferreira and de Haan, 2013
<http://arxiv.org/pdf/1310.3222.pdf>, and many others). Also see Gaioni et al. (2010), Coles (2001), Beirlant et al. (2004) and many others for both the GEV and ML method. Also a point of correction, the GEV in this paper is not heavy-tailed but short-tailed since $\xi < 0$. Our previous paper which is also cited in the manuscript proved that the GEV distribution is consistent in the river basin.

...The chapter 4 seems to be a bit of a surprise. It was not announced earlier and I am not sure if it brings extra information to the manuscript. ...

We are not sure what the Anonymous Reviewer refers to as unannounced chapter 4 here! In the last paragraph of the Introduction section chapter 4 is well announced. However, we would agree that this style of writing may have come as an anomaly to the usual style the Reviewer may be accustomed to. This chapter was motivated by recommendations of the International Disaster and Risk Conference (IDRC 2014) held in Davos, Switzerland, 24-28 August. If the Reviewer attended the conference, they will be aware that the thrust for this year in all the journal articles that deal with Disaster Risk Reduction was to include this chapter 4 on Added value for the post 2015 framework for disaster risk reduction. Let me take this opportunity to announce that this paper was also presented at the IDRC 2014 Davos conference.

...The last chapter, Conclusions, repeats what was written earlier but in a concise way. In my opinion it is not what Conclusions look like. ...

The views on the Anonymous Reviewer on the differences in the writing style of the Conclusions are noted, and some adjustments will be made in the final revised manuscript.

- **Specific comments**

Tables 1 and 2 could be merged – it will be easier to compare the results.

In our view, merging the two tables for ML and Bayes might lead to confusing readability of the new table by some readers. So the current form the tables are appearing is more preferred (this is also supported by the other Reviewers). We can suggest that the Editors put the two tables on the same page as they were in the original manuscript.

Maybe a common diagram made of graphs of lower left corner from Fig. 3 and Fig. 5 would do.

This suggestion is in agreement with suggestion by Jose Salinas. We will make an effort to combine the flood frequency curves in our revised final manuscript.

In conclusion we want to thank the Anonymous Reviewer for providing us with more references to increase our understanding on the topic and improve the readability of our manuscript. We will indeed include some of the references in our improved version of the manuscript.

OVERALL COMBINED COMMENTS

Overall remarks: It is clear that all the reviewers agree on the significance of the manuscript particularly in Africa, agree on the need to re-structure the introduction section, agree on the need to include a subsection in Section 2 explaining the priors used and the MCMC algorithms used in the paper, and again the reviewers coincide on the need to combine the flood frequency curves. These issues will be given priority in the final revised version of the manuscript.

Thank you

