Hydrol. Earth Syst. Sci. Discuss., 12, C4210–C4217, 2015
www.hydrol-earth-syst-sci-discuss.net/12/C4210/2015/

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Examination for robustness of parametric estimators for flood statistics in the context of extraordinary extreme events" *by* S. Fischer et al.

Anonymous Referee #2

Received and published: 13 October 2015

General comments

The authors attempt to estimate the 99% and 99.9% quantiles of extreme value distributions using different estimations at synthetically generated data. The authors list the analytical expressions of different estimators and base their assumptions on cited literature. This study could be a simple exercise, and I am not convinced that it constitutes so important an advance to be published in HESS. Also, I have concerns on both its physical basis and on the authors' understanding on the performed experiments.

Specific comments

In our current knowledge of statistical hydrology a study that focuses in distribution fitting cannot merely rely on point estimates, especially when dealing with higher order statistics and significantly small samples (e.g., of size 30). The variability of estimations needs to be characterized by quantitative means (e.g, confidence intervals (c.i.)). The authors need to characterize the most important possible sources of bias. Quantifying the variation of skewness in synthetically generated small samples (i.e., simulations by sampling from the extreme value distributions) is crucial for understanding to what extent the results presented here (which are based on sampling from a 3 parameter distribution, and on different estimators) are valid. The structure of the presented experiments allows for quantifying both a) parameter estimation c.i. for different estimators, and the b) uncertainty of sampling from GEV in samples of relatively small sizes. Therefore, it is surprising that the associated uncertainty is not quantified and included in this study (see comments below).

Moreover, the definition of "extraordinary extreme events" raises questions. The authors characterize as "extraordinary extreme events" the maxima exceeding a specified threshold. This definition is based on a threshold that strictly depends on a specific sample, with no reference to the population, whatsoever. The value of this threshold is by definition sample-dependent and strictly subjective. Thus, discussions on robust estimation with respect to "extraordinary extreme events" do not have a rigorous basis, in my opinion (see comments below).

A potential source of bias on sample statistics is setting 2% of the sample equal to the 99.9% quantile. This is a non-linear modification of the theoretical distribution and the authors make no reference to this potential source of bias on higher order statistics and on the corresponding estimates of distribution parameters. To quantify this bias the authors should estimate the statistics of the modified GEV distribution with the modified tail. Moreover, I suspect that the variability from perturbing the part (2%) of the sample that is set to the selected quantile, leading to different non-linear alterations of the distribution tail, would suffice to explain different behavior of the results in Tables

2,4, and 6.

Also I am not sure the authors fully understand the roles of bias and RMSE (see comments below). The authors attempt to interpret the results by systematically confusing the nature and causation of the two performance indicators. Higher bias does not lead to higher RMSE. The results in Table 1 illustrate this, however, in general this is a clear misinterpretation of the fundamental tools used to support the validity of the conclusions. Moreover, I am not sure whether the authors fully understand the theoretical basis behind the numerical alterations they perform to the theoretical distributions. The authors numerically modify the theoretical distributions and they then use estimators derived from the original, unmodified distributions. The authors do not characterize nor do they discuss the potential sources of bias from these manipulations. Considering the above, I am not confident that the authors provide the sufficient justification and discussion on different behaviors emerging in the results of this study.

Below the most important comment are listed, followed by technical corrections.

P8554L10:What does "when extraordinary extreme events are known to appear in the sample" means? This requires knowledge of a threshold a priori to characterize the population such that maxima are classified to extreme and to "extraordinary extreme" events, which can only be subjectively defined (as per the suggestions in P8566L9-14). This statement seems wrong as it is. Please rephrase.

P8555L22: what does "mispecification of the underlying model" means? Please explain. Does this refer to the exercise of P8566L1-9, where the authors truncated the theoretical distribution and then attempted to apply estimators derived from the untruncated theoretical one? That comparison is not valid (see corresponding comment). Unless there are errors in the use of English here (syntax), "Robustness against mispecification of the underlying model" has no meaning with respect to the theoretical basis of estimation.

P8557L20: How do the authors justify this "2%"? How does perturbing this degree of

freedom affect the overall results? The authors can illustrate a sensitivity analysis on the length of the sample subset that is being replaced by the 99.9% quantile, on the fits in Tables 2,4, and 6. This introduced non-linearity in the theoretical distribution is a potential source of bias that is not discussed nor is it accounted for throughout the manuscript.

P8563L10: I am not sure to what extent this statement is valid. Theoretically, the presence of what the authors call "an unnecessary third parameter" in GEV, which is a generalization of Gumbel may not affect the performance of robust estimators. An important source of bias in Table 1 is the sample size. As one would expect, as n decreases the RMSE of GEV-fitting significantly increases. This can be attributed to attempting to incorporate skewness estimates in distribution fitting from a small sample of e.g., 30 values. Including confidence intervals for skewness estimation for all sample sizes would shed light on the validity of this discussion. Also, reproducing skewness in small sample simulations can also be significantly biased.

P8563L12: "..fitting a Gumbel distribution causes a substantial negative bias, which dominates the RMSE." Bias and RMSE can be both performance indicators, but they are not correlated nor are they theoretically connected in the manner implied here. Additional comment (not proof): for instance, if anything in Table 1 the results for Gumbel fitting with high RMSE correspond to lower bias. These relations are merely statistical occurrences among different samples. Concerns are raised at this point on the authors' intuitive interpretation, with regard to the indicated causations that drive the results.

P8563L15: I am not sure if this is a striking result. This can be largely attributed to the significant bias of reproducing and estimating higher order statistics in this experiment, for small sample sizes. Again, depending on the tail of the distribution, attempts to estimate higher order statistics such as skewness, can be significantly biased. In a small sample of e.g., size of 30: i) sampling from the theoretical distribution, and ii) attempting to estimate skewness using any point estimator can be significantly biased with respect to the theoretical skewness. Fluctuations in results and different behaviors

can be fully justified by the large potential sources of bias associated with higher order statistics and small sample sizes.

P8563L28- P8564L1: In my opinion, this is not a "very striking result". As discussed above, the results for small samples are very likely to be significantly biased. Attempting to generate GEV distributed random variables for small samples (e.g., n=30) may lead to skewness values that significantly deviate from the theroretical ones (i.e., the ones estimated from the selected location, scale, and shape parameters). Accounting also for the associated estimation bias, which can be significant for small sample sizes, may justify the high values of bias and RMSE of Table 5 for n=30.

P8563L29: "large positive bias resulting in a large RMSE" this causation is not valid (see previous relevant comment)

P8564L14: please see previous comments on the "relation" (or lack thereof) between RMSE and bias discussed by the authors

P8564L13-14: The high GEV bias and RMSE a can be given to estimation bias at small samples. Attempting to estimate parameters that are based on higher order statistics (skewness) may be a significant source of bias, as discussed above.

P8564L16-23: All this clearly depends on the selected "2%" which is being replaced by the 99.9% quantile (see comment above). This is a free parameter, the value of which is not sufficiently justified, in my opinion.

P8565L8-10: I have concerns about this conclusion, which are discussed in previous comments on small sample sizes

P8565L11: the frequency of rare events is predefined in these experiments. Please see previous comments on modifying the sample distribution in a non-linear manner.

P8566L1-8: This is not a valid experiment in my opinion. Cutting off the tail is a non-linear alteration leading to a truncated distribution. Replacing the omitted values with different ones leads to an entirely different distribution. Yet it seems that the authors

used the estimators corresponding to the unmodified theoretical distributions discussed earlier in this manuscript. Concerns are raised at this point on to what extent the authors have complete understanding of the performed manipulations, implications to potential sources of bias, and on the physical basis of the very comparison they suggest.

P8566L9-14: this discussion is necessary (even though the classification of extreme events by no means has an objective character)

P8566L13-14: In my opinion indicating effects of the likelihood of "extraordinary events" on parameter estimation falls within the scope of this work and it is very likely to provide a stronger basis for the conclusions (see previous comments on sensitivity analysis on the frequency of assumed extreme events)

Technical corrections:

P8555L13: the occurrence of extreme events is not a problem. Estimation of statistics can be challenging. Please correct syntax.

P8555L21:" with robustness meaning robustness": please replace with something like "robustness referring to"

P8555L22: I am not sure I understand what "single extraordinary events" means

P8556L4: this sentence is vague

P8556L10: is n<50 the criterion for small sample sizes suggested by DWA? Please specify

P8556L14: "fair evaluation": please rewrite using more appropriate terms

P8558L20: "being similar to ML" is not "an advantage". please rephrase

P8558L23: please identify which these situations are. This is vague

P8558L24-5: mean and stand. dev are not "characteristic values of a sample"; they

are "statistical characteristics of a sample". please rephrase

P8558L25- P8559L1: I am not convinced that what precedes "Therefore" sufficiently justifies this statement. Can you please rephrase or strengthen this argument with cited literature?

P8559L5-6: the syntax can be confusing; "respectively" perhaps can be used more clearly

P8562L8-11: this is a correct remark, but it can be restated to be more clear (also please avoid "do not seem to be suitable in this hydrological context")

P8562L20: please define i.i.d in P8554

P8563L8-9: what do you mean by "approximately valid models"? Please explain with examples.

P8563L15-16: I strongly advise against the use of "much" or other adverbs implying statistical significance (potentially important for this study) which has not been performed in this study.

P8563L19: Can you please specify what you mean by "due to the large variability"? Variability of what?

P8564L12: "L-Moments behaving" I would rephrase this

P8564L18: please replace "situation" with "case"

P8565L8-10: I am not sure I understand this sentence. Please see previous comments on effects of small samples.

P8565L7-8: can you please rephrase this? (syntax)

P8565L22: typo "the"

P8566L1: Please rephrase this; I am not sure this sentence makes sense.

P8566L2: "an uncertainty"- please correct grammar

I hope these comments help the authors improve the quality of this study. Thank you for the opportunity to review.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 12, 8553, 2015.