Hydrol. Earth Syst. Sci. Discuss., 12, C5276–C5281, 2015
www.hydrol-earth-syst-sci-discuss.net/12/C5276/2015/

Hydrology and Earth System Sciences
Open Access
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.

S. Fischer et al.

svenja.fischer@rub.de

Received and published: 1 December 2015

Response to the referees' comments

We want to thank both referees for their stimulating comments and suggestions. Responses to the questions and suggestions of the referees are included below.

Review #1 (F. Serinaldi):

The first question raised by the referee concerns the role of robust estimation in the context of extreme value analysis. We agree that this is a difficult issue and consider

it therefore to be well worth investigation. Perhaps we should have added some more information on the foundations of robust estimation. Discarding or downweighting the most extreme and thus unusual observations used to be common practice when the interest was in the central location of the data. In the context of extreme value analysis, we agree that the most extreme observations may be the most interesting ones and should thus not be simply discarded subjectively, as explained by the reviewer. Since traditional estimators can strongly be influenced by extreme observations which occur occasionally in practice, particularly when observing many sites (gauges) in parallel, robust statistics has developed techniques like minimum distance or M-estimators which accommodate observations which are extreme in the light of the assumed model and the other observations. This can sometimes be understood as a kind of automatic downweighting, as opposed to manual downweighting or other subjective manipulations. The referee is also concerned about the danger of overlooking the information in the most extreme observations, since these can point at different physical mechanisms, etc. In fact, robust methods can even help to detect such different mechanisms when being applied with care. As pointed out by John W. Tukey (1979: Robust Techniques for the User. In: R.L. Launer, G.N. Wilkonson (eds.) *Robustness in Statistics*. Academic Press, New York, pp. 103-106), "It is perfectly proper to use both classical and robust/resistant methods routinely, and only worry when they differ enough to matter. BUT when they differ, you should think HARD." While data analysts should not apply a single method blindly, it is often a good strategy to apply a traditional and a robust estimator in parallel. If the results of both methods agree, one can be confident that these results are not just due to a single or a few extreme observations. Moreover, it indicates that the underlying model, e.g. the generalized extreme value distribution, fits the data well, if the most extreme observations are judged to be plausible given the model and the other, less extreme data. If, however, the chosen traditional and robust estimator disagree, one should think carefully about possible explanations for this discrepancy, like particular mechanisms driving the most extreme but not the other observations, for instance. Robust estimators can thus even help to detect model deficiencies or other problems.

The calculations of the referee confirm that observations at least as large as the 99.9%-quantile of the underlying distribution are not unlikely in samples of at least 30 observations. Indeed, it was our concern to identify methods which work well in the presence of such large but still plausible observations. We agree that more complicated mechanisms for generating very large observations as described by the referee are more realistic. Such mechanisms would lead to even more extreme observations and thus increase the difference between traditional and robust methods.

The referee outlines that the recommended estimator, TL(1,1)-Moments, is in none of the simulations the best performing estimator concerning bias and RMSE. That is correct, indeed the TL(1,1)-estimator is not the best estimator in the different simulations. It is nevertheless our recommendation, as it is robust and has good properties (concerning bias and RMSE) also when no extraordinary extreme events occur. Moreover, our focus was laid on the case where a GEV-distribution is fitted, since this is the most relevant practice. Under these considerations the TL(0,1)-Moments cannot be recommended because of their high RMSE and bias in the case where we have a GEV-distribution with positive shape parameter without inserted extreme events. The same remark is valid for the MD-estimator, which has even higher bias and RMSE, even in the case of extraordinary extreme events in small samples. Since in the presence of extraordinary extreme events the TL(1,1)-Moments behave similar to the TL(0,1)-Moments and in the case of a small sample size much better than the MD-estimator, we give the recommendation to use TL(1,1)-Moments. It is the robust estimator with small bias and RMSE in the case of unmodified data and small sample size. If we knew the data situation we are in beforehand we would agree with the referee concerning his recommendation to use L-Moments or ML-estimation in the case of no extraordinary extreme events, depending on whether the sample size is smaller or larger than 200. In the case of extraordinary extreme events and sample sizes smaller than 200 we would recommend the TL(0,1)-Moments and for a larger size the Minimum Dis-

tance estimator. We tried to make this distinction of an overall recommendation and a situation-dependent recommendation on p. 8565, lines 21-28, and hope this distinction is clarified by this explanation. The referee also mentions that fitting a GEV distribution to Gumbel samples is not truly wrong. This is correct, of course, but in the presence of extraordinary events, the estimation of a third parameter does not only lead to additional uncertainty but a non-robust estimation could lead to a very large estimated value of the shape parameter where in fact there are only one or two extraordinary events. A non-robust GEV-fitting would therefore result in a GEV-distribution very different from the Gumbel. This was the reason for us to discriminate between Gumbel- and GEV-fitting.

Review #2 (anonymous):

We agree with the reviewer that statistical analysis should provide information on the variability of the estimates, e.g. confidence intervals. Nevertheless, the first step is identification of a good point estimator which provides reliable information even in the presence of unusual extreme values. Once we have identified such an estimator, we can build confidence intervals based on it, but this is a next step thereafter. Concerning the skewness of real samples, we have analysed many data sets and chosen our parameter settings to represent typical data scenarios found in practice.

There seems to be a misunderstanding concerning the definition of extraordinary extreme events, since the definition applied in our paper does not depend on the specific sample but on the underlying distribution generating the data. Moreover, the referee fears that we have not fully understood the implications of our doing and points out that replacing 2% of the observations by a certain large value changes the tails of the distribution, so that we should actually estimate the quantiles of such a modified distribution. However, such an analysis was not our intention. We just regard the inserted very large observations as extreme observations from the same distribution as the other observations were generated from, and want to estimate high quantiles of this distribution with or without the occurrence of such extreme values. This type of analysis, where one assesses the performance of estimators under a distributional model in the presence of observations which are rare for this model, is common practice in robust statistics and we refer to the corresponding literature, e.g. the textbook "Robust Statistics: Theory and Methods" by Maronna, Martin and Yohai (2006) and the references cited therein. Even worse, the referee fears that we do not understand the roles of bias and RMSE and (repeatedly) claims that they are not correlated. However, looking into any statistical textbook dealing with estimation we find the equality that the MSE (mean square error) is the sum of the squared bias and the variance. The MSE can thus be splitted into a component describing the systematic deviation from the target (namely the squared bias) and another component measuring random variability (namely the variance). For a fixed variance, the MSE increases linearly with the squared bias, and so there is a strictly monotone relationship between the root of the MSE (RMSE) and the absolute bias if the variance is fixed. We prefer reporting the RMSE and the bias over the squared quantities, because the former are measured on the same scale as the parameter of interest, and since an absolute bias almost as large as the RMSE tells us that the random variation of the estimator, as measured by its standard deviation σ, is small compared to the bias, please compare the relationship

$$RMSE = \sqrt{\text{bias}^2 + \sigma^2}.$$

Has there perhaps been a confusion of MSE and the variance σ^2?

In the following we want to answer specific comments, which are not covered by the above comments.

P8555L22: The usage of the word "know" is indeed misleading in this context. We would change this passage to: "When very large events occur, which apparently do not fit to the remaining data, the additional usage of Trimmed L-Moments as extension to classical estimators are a recommendable choice for a robust estimation."

P8557L20: Replacing 2% of the observations by very large values for this kind of distribution is certainly subjective. Our choices correspond to situations which are still plausible in practice (if we ignore the presence of ties, but this is just a matter of taste), please compare the calculations of the other referee. Further situations with more or less replacements could be considered as well, of course.

P8563L10: If we understand the referee right, this is exactly what we want to express by fitting a GEV-distribution to Gumbel-data. The estimation of the third parameter causes additional uncertainty leading to a higher estimation error (p. 8563, l.10 and comment to Referee #1). We agree that such a situation might be detected when using confidence intervals for the shape parameter, but this was not our concern in this study of point estimators.

P8563L28-P8564L1: What is striking about this result is that the behaviour changes completely if the sample size is increased by only 20 data values. In other simulation cases this does not happen.

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