Review of hess-2023-234

This manuscript critics the use of non-asymptotic (NA) distributions of block maxima. The authors bring two main arguments to support their critic. The first is that NA require knowledge of the parent distribution and that, when this is known, there would be no need for deriving distributions of block maxima. The second is that the presence of serial correlation in the observations would decrease the potential advantage of NA. The manuscript then follows with some targeted comments to specific studies.

The paper addresses a relevant topic, but the manuscript falls short at supporting its main conclusions. This is not due to technical errors, rather to a narrow (mono-disciplinary) vision of the problems at hand (see main comments) and to an erroneous generalization of case-specific objections. Since all statistical models present advantages and disadvantages, which depend on how much the underlying assumptions are met/not met (or, to quote the statistician George Box: “all models are wrong, but some are useful”), I believe that the presented critics cannot be generalized to NA methods as a whole. Rather, they should be targeted to highlight specific aspects of NA methods that need attention and/or the specific NA approaches that need attention.

Considering the main comments below, I believe the manuscript should be deeply revised before reconsideration. I am not sure this can be handled within a major revision, because the take-home messages would need important adjustments. The title should be revised to be more pertinent with the actual outcomes. This also pertains the title of some sections/subsections which border disrespectfulness (e.g., section 3, 4, 5). The manuscript is long and contains numerous repetitions. It includes unclear and/or incorrect reasoning in some sections, which prevent from fully understanding some parts (see comments 6 and 7 below). It could be halved in length without changing the message.

All the references in this document can be found in the preprint of the manuscript.

Given my lack of specific expertise, comment 6 was written with the help of a colleague expert in Bayesian statistics.

I hope my comments are helpful.
Kind regards,
Francesco Marra

Main comments

1. My main comment turns out to be a citation from the manuscript itself (line 694): “*models and methods should be thought and used in the right context and for suitable*

purposes". This sentence in the conclusions contradicts several of the arguments presented in the manuscript. The authors proceed for 30 pages (pages 1-30) repeatedly claiming that NA methods are 'superfluous'. Only in lines 685-694 they contradict this argument stating that what they report in the manuscript "*does not mean that they are not useful at all*". They then proceed listing what, in their view, are potentially useful applications of NA approaches and finish off with the citation I started with. I may add that the reported list of 'useful' applications is limited by the imagination of the authors (as shown by the sole presence of self-references in here) and, mostly, by the perspective they adopt. In this, it seems they forget the gap that there exists between theory and practice, between advancements in theory and practical use of extreme value distributions of any kind by hydrologists, risk modelers and end users in general. In fact, several applications of NA methods follow the directions accepted as 'useful' by the authors, and other applications of NA methods follow directions that are useful, although not within the directions imagined by the authors. One notable example is the connection between physical processes and statistics, which can only exist in a NA model, given that real world physics is not asymptotic. It is my believe that physical processes should direct the statistics we use. The physics of the processes we are dealing with is not asymptotic. At this concern I must cite again the authors (lines 623-624): "*Historically, the main scientific progresses occurred when some one called into question widely accepted mainstream theories using arguments more solid than those of the superseded theories*". It almost seems we think alike on this point, although with different concepts for 'mainstream'.

2. One argument is that NA methods require the knowledge of the parent distribution and that when this is known, there would be no need for deriving a block maxima (BM) distribution. This is technically true but seems to neglect situations in which a BM distribution is helpful (even though, I agree, not technically essential). Some examples: empirical comparison with observations of BM only; fair comparison with estimates from EVT distributions; providing information that practitioners can use without changing habits. Doing these directly from the compound parent, although possible, would be troublesome and possibly confusing for non-experts.

3. The issue with serial correlation is important and could affect some applications of NA methods. I believe future NA applications (either for block maxima or directly from the parent distribution) should keep this in mind. In this, the paper is a relevant addition to the literature. Still, it falls short at supporting the adjective 'superfluous' that accompanies the reader. The importance of serial correlation depends on the type of variable one wants to examine and on how the variable is used in the model. It cannot

be generalized to the application of NA methods as a whole. Incidentally, serial correlation also negates the assumptions of extreme value theory (EVT), with the effect of making the convergence much slower. Slow convergence actually suggests that NA methods should be used, making the reasoning circular and thus highlighting once again the complexity of the problem.

4. Relatively large portions of the manuscript are dedicated to commenting specific works (sections 4.2, 5.3; 10 pages in total). Given that previous work, also by some of the authors (Serinaldi et al., 2020), and this work itself confirm that NA methods are formally correct, it is not fully clear how objections to specific works should affect NA methods in general.

Specific comments

5. In section 3, the authors treat EVT as if it was the truth. Statistically it is, provided that the underlying assumptions are met. Among these the asymptotic assumption. In some relevant cases, convergence to the asymptote is (very) slow, such as the case of the powered exponential family of distributions. Notably, this is the case most relevant for precipitation, and precipitation is the main variable on which NA methods are confidently used (due to the relatively simpler relation with the underlying physics). In fact, in the case of precipitation tails from EVT are too heavy. This becomes clear when one tries to generate stochastic time series from a EVT distribution, and led to the development of a family of powered-exponential distributions for the generation of stochastic precipitation series (Papalexiou, 2022). These tails explain well the statistics of observed extremes, as shown by Marra et al. (2023) (more on this later). Overall, in that paper, we showed that GP tails from EVT and powered exponential tails from NA models can be indistinguishable, with the difference that the former are asymptotic distributions fitted to NA data. The message is once again that no model is perfect, and that different models may lead to similar answers, thus advancing our understanding of nature.

6. Some concepts in Section 5.3.1 are misused. Montecarlo simulations (both in its standard term and in its Markov Chain variant) are numerical methods to compute integrals and expectation, sampling from a target distribution numerically and approximating the expectations via empirical average. However, the description provided by the authors is confused and, at least for what concern the different approaches to statistical inference (here the frequentist or classical paradigm and the

Bayesian one), wrong.
Montecarlo simulations in frequentist inference and Markov Chain Montecarlo (MCMC) in Bayesian inference target totally different objects. The authors correctly assess the role of Montecarlo simulations under a frequentist approach to statistics. Under this point of view there exists a true population's characteristic (or statistics, using the authors term) that is estimated (intrinsically with some uncertainty) from a finite sample. The variability of the estimator (and not of the parameter that the estimator is targeting) can be assessed in many ways, e.g. exploiting Montecarlo sampling to mimic the repeated sampling principle thus allowing to construct frequentist confidence sets. In Bayesian inference, instead, do not exist a 'true' parameter of the population as this is consider a random variable itself. Consistently with this, the posterior distribution of any unknown, which is often approximated via MCMC sampling is the target of inference. While posterior summaries like the posterior mean are common, they represent fundamentally different entities from frequentist estimators. In Bayesian inference, MCMC draws are used to construct credible sets, intrinsically different from the notion of frequentist confidence sets. The uncertainty that the posterior is describing is not the same uncertainty that the estimator variance in frequentist inference (obtained in any way, including Montecarlo sampling) is describing. Additionally, Bayesian model averaging is a well-known and successful concept that is not related to the summarization of the (MCMC approximated) posterior distribution of any kind.
Despite stemming from confusing arguments about basic concept of frequentist and Bayesian inference, the discussion starting from eq. (10) to the end of Section 5.3.1 is correct. However, it is a mere consequence of eq. (17) and deserves less space. Perhaps lines 453-485 can be removed and the subsequent text rearranged.

7. Section 5.3.2 is not clear. Specifically, I could not grasp whether the objection concerns (a) the average from the synthetic timeseries of the Montecarlo samples, or (b) the average in the MEV formulation. Are the authors claiming that the figure in Marani and Ignaccolo draws something different from what is claimed, or that the MEV framework is incorrect? The suggested changes in Fig. 7 indicate that we are in case (a). Should this be the case, the entire section 5.3 would be a direct comment to Marani and Ignaccolo (2015) that not necessarily pertain NA methods in general, but only the Montecarlo sampling in here. Should (b) be the case, it is not clear why section 5.3.1 is there and why all the distributions (not only MEV) change in figure 7. Even in this case, the comment would not pertain NA methods in general.

8. In section 6, the authors briefly comment on a paper of mine in which NA (Weibull) and asymptotic (GP) tails are compared for the case of precipitation. They quickly dismiss our study claiming that we used a low threshold “*out of its range of validity*”. We reported results for threshold equal to the 95-th percentile for consistence with the Weibull model, but we clearly stated that “*Results derived from higher thresholds such as the 98-th percentile used by Serinaldi and Kilsby (2014) are qualitatively analogous but characterized by larger uncertainties*” (Marra et al., 2023). For reference, I report here the same as figure 3 in Marra et al. (2023) as it was obtained using a threshold equal to the 98-th percentile (Figure 1 below). As it can be seen, the instances in which GP provide too heavy or too light tails are even increased when using the 98-th percentile with respect to the 95-th percentile case (please refer to Marra et al., 2023). This is because in addition to theoretical convergence issues (what the authors focus on), there are important (practical) issues with stochastic (sampling) uncertainty.

9. The manuscript presents numerous self-citations.

10. Incidentally, as a user of NA methods, I never claimed they are ‘superlative’. They are as good as other models are: they offer advantages in some situations and disadvantages in others.

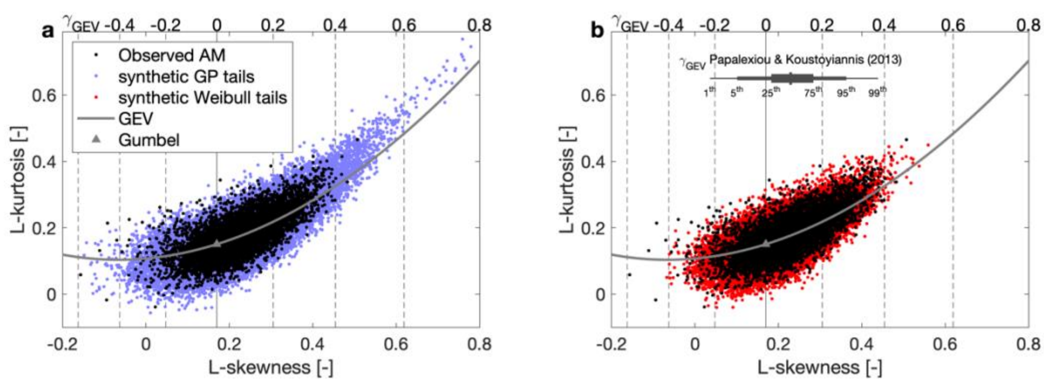

Figure 1: Same as figure 3a from Marra et al. (2023), but obtained using GP tails estimated with a threshold equal to the 98-th percentile