Non-asymptotic distributions of water extremes: Superlative or superfluous?

By F. Serinaldi, F. Lombardo, C.G. Kilsby

Submitted to *HESS*

MS-NR: hess-2023-234

Reply Reply on RC1 (Anonymous Reviewer #1's report)

(Note: In the text below, Referees' comments were copied verbatim in **black**.)

Response We thank the Reviewer for the interesting feedback. Please find below our response to raised concerns.

A classic choice in the statistical modelling of extremes is between (a) constructing a detailed model of an entire process, from which its extremal properties can be estimated, either analytically or more usually by numerical methods, or (b) direct modelling of the extremes themselves. If adequate reliable data are available and the investigator has sufficient time, then approach (a) allows information from other sources (such as physical models) to be included at the modelling stage and has the benefit that estimates of all quantities, including extremes, stem from a single overall model and therefore are consistent. However this approach is demanding of data and of time, and makes the implicit assumption that the details of the underlying process are relevant to the extremes. Approach (b) is less demanding of data and avoids detailed modelling by applying the classical theorems of extreme-value theory (EVT) to block maxima or threshold exceedances for the phenomenon of interest. Although originally developed for independent and identically distributed observations, these theorems have been shown to be robust to plausible types of dependence in the underlying data, and have been widely and generally successfully applied in environmental settings. They can be regarded as semiparametric models, in the sense that they do not depend heavily on the underlying process. A major concern is that they rely on limiting approximations (the GEV and GPD) that may fit data at observed levels satisfactorily but extrapolate poorly to unobserved levels. Such models provide a simple and direct empirical approach to modelling extremes of the underlying phenomenon but it may be a struggle to incorporate physical constraints or other background knowledge into them.

The paper under discussion can be viewed as a critique of a particular type (a) approach, namely metastatistical extreme-value (MEV) modelling, from the viewpoint of a classical type (b) approach, namely the fitting of GEV and GPD models to block maxima and threshold exceedances.

Response We are grateful to the Reviewer for the thorough overview of statistical modelling of extremes. Anyway, we must clarify that in our work we do not compare asymptotic and non-asymptotic (NA) models of block maxima (BM) from the point of view of asymptotic models (the latter are indeed not even mentioned in the abstract). We contrast models of BM with the corresponding compound parent distributions. As clearly stated in the paper the key message is:

"*We discussed their redundancy and practical uselessness in real-world analysis. This apparently bold statement relies on very basic facts: (i) the distribution F_Z of a process Z provides all information about any quantile or summary statistics (extreme or not); (ii) extreme value distributions F_Y of BM corresponding to the parent process Z are just approximations of the tails of the distribution F_Z, and they have a role only if F_Z is unknown; and (iii) NA distributions require the preliminary knowledge/estimation of F_Z; however, once F_Z is known or fitted to data, NA distributions of BM are no longer needed, and their derivation is superfluous as F_Z already provides all information. In this context, the use of asymptotic extreme value models is justified by the fact that they do not require the preliminary knowledge or estimate of F_Z (under suitable conditions).*"

Asymptotic models are only mentioned in the paper to stress that their merit is to disconnect inference about the tails of the parent distribution from an accurate knowledge of the parent distribution itself (as mentioned by the Reviewer). Instead, NA models of BM require the preliminary inference of components (conditional marginals) that are sufficient to build the compound parent distribution. Therefore, NA models become useless as they provide just an approximation of the tails of a compound parent that is already completely defined. This is the message conveyed by Figs. 1 and 3 and Section 4.

Since we aim at conveying this message, we think that targeting "*much of the general text in the earlier sections for cuts*" is not a good idea, as these sections contain the epistemological justification of our criticism. This is also needed because we strongly believe that conceptual reasoning is crucial even in a modern era focused on massive data analysis and technicalities, because they must be supported by preliminary epistemological reasoning.

There are two main criticisms:

that papers proposing MEV have done so by application to and illustration on `real data', in which the true data-generating mechanism is unknown, which implies that it is impossible to compare the behaviour of different approaches under ideal conditions (when the target of inference is known);

that in any case the comparisons are incorrect, because of confusion over the target of inference (see Figure 7). Here the point is more subtle, but it is summarised in equation (16) of the paper. The point here is that if one is estimating a quantile function $Q_\theta(p)=F^{-1}(p;\theta)$ that depends on an unknown parameter θ and one will estimate θ from a single sample using an estimator $\hat\theta$, then the estimator of $Q_\theta(p)$ is $Q_{\hat\theta}(p)$, whose properties should be assessed over repeated sampling using independent replicates $Q_{\hat\theta_1}(p), \ldots, Q_{\hat\theta_S}(p)$ based on S samples leading to estimates $\hat\theta_1,\ldots, \hat\theta_S$. The average of these estimates would be $S^{-1} \sum_{s=1}^S Q_{\hat\theta_s}(p)$, i.e., the right-hand side of (16), rather than $Q_{\bar \hat\theta}(p)$ (the left-hand side of (16)), where $\bar\hat\theta$ is the average of the parameter estimates for the S samples. The paper under discussion illustrates the difference via the left- and right-hand panels of Figure 7.

Response Thanks again for this interesting comment. However, as mentioned above, our main concern comes from introducing NA models (not only MEV, but also our own models in Eq. 4) neglecting the epistemological justification of EVT (asymptotic) models.

In real-life problems, we mainly need the probability of values taken by processes such as discharge, precipitation intensity/depth, water level, etc. We do not need or look for the probability BM or POT of such processes. We focus on BM or POT for convenience, as these sub-processes can conveniently be described by a couple of distributions that do not require the knowledge of the parent process (under suitable conditions).

Since NA models of BM do not exhibit this disconnection, they automatically lose the only practical advantage of using “sub-optimal” BM and POT processes.

All the technical inconsistencies discussed throughout the paper are just a consequence of missing these epistemological concepts, reducing the development of new models to a mechanistic exercise focusing on “how” to do that, but missing the fundamental preliminary step, i.e. “why” we develop new models! This is the meaning of (the title of) Section 3: “Modeling extreme values: asking ‘why’ before looking for ‘how’”.

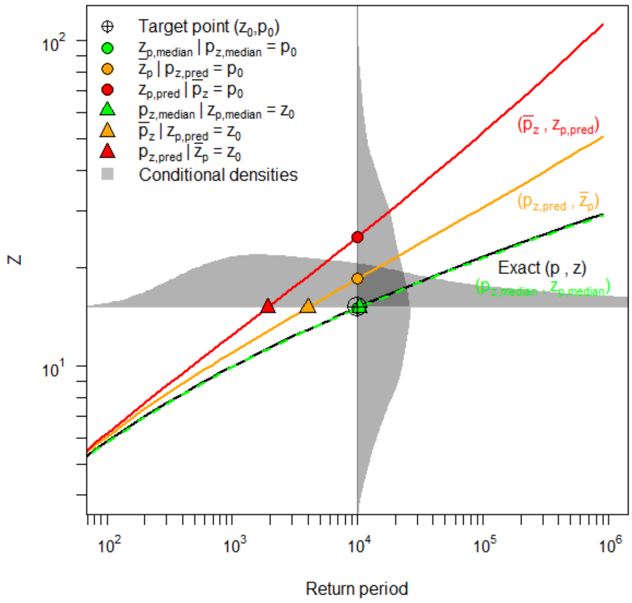
The literature on NA models reveals that these concepts are widely overlooked and, perhaps, need to be recalled as we did in the manuscript.

(Though the discussion at lines 532-535 leaves it unclear how the `median GEV/Gumbel’ curves are computed — the median for each p, giving a result that would not corresponding to any single quantile function, or what? And if the median, why not the mean?).

Response

Median GEV/Gumbel probability and quantile functions are obtained by Eq. 14 and 15.

We state “*Therefore, to be consistent with EVT, the ensemble of GEV and Gumbel distributions should be summarized using a transformation, such as the median, that retains the expected GEV/Gumbel shape.*” In this respect, we devote the preliminary sub-section 5.3.1 to explain why the mean provides biased summary of MC ensembles, whereas rank-based summary statistics, such as the median, are insensitive to non-linear transformations linking parameters, F_Z, and Z, and are therefore more appropriate to preserve the shape of the distribution. To further explore such concepts, a graphical representation is shown in figure below (with self-explanatory labels); it refers to GEV distribution with shape parameter equal to 0.1, sample size equal to 50, and 3000 Monte Carlo replications to assess sampling uncertainty.



Moreover, we introduced Section 5.3.1 by explicitly stating that "*We anticipate that the foregoing discrepancies depend on the misuse of methods used to summarize multi-model ensembles. Thus, before describing Monte Carlo experiments and their outcome, we need to recall some theoretical concepts that are required to correctly interpret numerical results.*"

Therefore, the text already cares about the reader's understanding of the discussion. As mentioned above, specific sentences in the text are consistent with the premises reported in the introducing sections.

Both of these criticisms seem to me to be correct, and they should in my view embarrass the reviewers of the original MEV papers and the journals that published them.

Response We believe that the objective statement by an independent Reviewer about the correctness of our criticisms is fitting to the task for the Authors of this paper, who have honed their research over the years to the very aims of mapping out and understanding theoretical consistency in analysis of geophysical data, without getting much consensus for that. Therefore, the Reviewer has our genuine thanks for providing us with such an important comment in acknowledgement of our work. Other than that, we agree with the Reviewer that "new" methods could sometimes be uncritically accepted in scientific journals and then routinely applied by the scientific community without double-checking their theoretical basis, taking for granted that they are conceptually/formally correct just because they are published once somewhere.

I found the paper to be quite poorly written, to the point of unclarity in numerous places, including lines 405 (`the spreader …'?), 429 (` …, or better, …'?) or 524 (what is a predictive quantile function of a predictive quantile function?), and with many minor errors. Examples of the latter are that (i) the Beirlant et al book cited at line 18 was published in 2004, not 2006, and (ii) stating on line 476 that the distribution of an order statistic is beta is incorrect — the beta distribution represents variables on a finite interval, and clearly this does not apply to order statistics from, say, a Gaussian sample (did the authors mean that the distribution of an order statistic can be represented _using_ that of a beta random variable?), (iii) equation (17), the left-hand side of which is a function of z, while the right-hand side is a number (as the expectation of Z is a constant), and (iv) at line 461, where results from a simulation study are `eventually used to build confidence intervals' — but in a simulation study the truth is known, so confidence intervals are not needed — as a confidence interval is based on a single sample, we have to guess that the authors mean that their S return level estimates will be used to compute quantiles of a distribution. The paper is full of inaccuracies of this sort, so the reader is continually wondering `is that correct?' and concluding `not quite'; this does not give confidence in the main results. It is the role of the authors to produce a well-crafted article, not that of a reviewer, so I will not give more examples (it would take many pages to list them all), but generally I found the writing to be unclear, long-winded, and in need of a careful review by a native English-speaker (see, e.g., line 514). Reducing the paper radically by revising and trimming the text throughout would improve it. I would also target much of the general text in the earlier sections for cuts, since it is mostly not germane to the criticism of the MEV work. A 15-page paper in the current format would make the main points more clearly and should be more readable.

Response We will double-check language and presentation. Concerning the specific points:

- "*predictive quantile function*": the full sentence states that "*the predictive quantile functions (<u>over S samples</u>) of the predictive quantile functions (<u>over n_Y samples</u>) <u>associated to MEV structure</u>*".

  Predictive quantile functions are just ensemble averages: Figure 7 just shows the ensemble averages over S samples of the ensemble averages over the n_Y quantile functions contributing to the MEV quantile function. In other words, there are two levels of compounding (let us say hierarchy): the first one is related to the derivation of MEV (which is itself a predictive distribution integrating the inter-block variability of parameters), while the second one is related to the derivation of the predictive version of MEV, integrating (averaging) over the S MEV functions.

  Thus, the curves in Figure 7 are ensembles of ensembles, where the average is taken firstly over n_Y parent quantile functions and therefore over S MEV quantile functions. This is formalized in L. 525, where we state "*In more detail*, $z_{p,\bar{Q}} \cong E_S[E_{\Omega_{\theta_S}}[{F_{WEI}^{l}}^{-1}(p|\theta_S)]]$…"

(i) We will fix the typo in the BibTex entry of Beirlant et al., thanks.

(ii) We understand Reviewer's concern and we will clarify this point in the revised text as follows. Under i.i.d. assumption, order statistics have a binomial distribution (which is equivalent to a beta) in the sense described by Equation 1 (see also David and Nagarajah 2003, pp.9-10). In other words, the distribution of the order statistics is a beta distribution of the variable $F_z(Z)$, or equivalently a so-called beta-extended distribution of Z (Eugene et al. 2002), which is also known as generalized beta-G distribution, where “G” denotes generalized classes, such as exponentiated-G or Kumaraswamy-G (e.g., Tahir and Cordeiro, 2016).

Therefore, from Eq. 1 and the expression reported in L.476, it should be clear that the range of the distribution is not bounded.

(iii) Thanks, we will double-check notation and fix these typos.

(iv) We believe a sentence cannot be extrapolated from the context. We state: “*Monte Carlo simulations are usually used to study the uncertainty affecting estimates based on finite-size samples (that provide incomplete information about the underlying process) or to approximate population distributions (or statistics) when mathematical closed-form expressions are not available. Examples of these applications are the experiments reported in Sections 5.1 and 5.2… In all cases, the primary output of Monte Carlo simulations is a set of parameters identifying a set of models (multi-model ensemble) that is then used to estimate the target statistics of interest. For example, simulations of S finite-size samples in Sections 5.1 and 5.2 are used to fit a set of S GEV distributions. These are then used to calculate a set of S 100-year return levels, which are eventually used to build confidence intervals*”

The sentence refers to the use of MC to build confidence intervals describing the uncertainty of estimates from finite-size samples.

To summarize, we will fix the above issues by double-checking notation and typos and adding some minor details. Anyway, we also think that these issues are far from being sufficient to call into question the overall content of the paper.

Concerning the length of the paper, we believe the message of the first part of the paper is very important despite its simplicity and iteration throughout the text. We do believe that in a shorter version such message could be easily neglected or misinterpreted.

We also highlight the content from Section 4.1.1 to 5.2 (i.e., L. 178-437), where we discuss key points such as the interpretation of the relationship between parent models and models of BM, the effect of serial correlation, and provide examples by simulations and data re-analysis. The content of the paper goes far beyond Sections 3 and 5.3, we will stress this in the revised manuscript.

References

Eugene N., Lee C., Famoye F. (2002) Beta-Normal distribution and its applications, Communications in Statistics - Theory and Methods, 31:4, 497-512

Tahir, M.H., Cordeiro, G.M. Compounding of distributions: a survey and new generalized classes. J Stat Distrib App 3, 13 (2016)