Comment on hess-2021-570

Anonymous Referee #1

Referee comment on "Technical Note: Flood frequency study using partial duration series coupled with entropy principle" by Sonali Swetapadma and Chandra Shekhar Prasad Ojha, Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-570-RC1, 2021

The Technical Note "Flood frequency study using partial duration series coupled with entropy principle" by Swetapadma and Ojha discusses methods to use partial duration series type of data to carry out flood frequency estimation. The topic is interesting and appropriate for the journal, but I somehow fail to see what the main contributions of the note, which I think does not provide a clear overview of the new developments, significant advances, and novel aspects of experimental and theoretical methods and techniques which are relevant for scientific investigations within the journal scope (this is a quote from the description of HESS technical notes).

The paper is fairly well organised and the references mostly suitable, giving an overview of what is the current understanding of the question. It presents the modelling framework using a case study in New Zealand.

My understanding is that the novel contribution proposed by the authors is to use entropy as a way to choice the PDS threshold, but I am not entirely sure this innovation is presented in a clear and convincing way. In particular there are a few points that I find quite unclear or that I believe undermine the strength of the authors' argument: I'll try to outline them below.

* I feel there note is somewhat lacking a discussion of the consequences connected to the many choices which are done in the modelling pipe-line: the more obvious one to me is the choice of estimating the distribution parameters with L-moments rather than with other methods. Would the threshold/distribution choice be different if we used standard moments or maximum likelihood to estimate the parameters?

* the choice of distributions used to model the number and magnitude of exceedances could be better motivated. The "traditional" framework uses the Poisson and the Generalised Pareto distribution respectively: these are motivated by some well known theoretical results. The Negative binomial extends the Poisson distribution, allowing for overdispersion. I do not quite understand how the Binomial distribution is instead fitted here, as we would need to have a k value of exceedances over N "trials" but the N value should be different from year to year since we only focus on independent peaks. Is this what the authors do? Further the use of the GEV, P3 and LP3 surprised me here as these are typically employed to describe annual maxima and have little theoretical or practical

justification in the context of threshold exceedances: they can of course be used, but I'd mention the fact that the GP has a somewhat stronger theoretical grounding.
* The definition of AIC and BIC is not correct in Table 3: the definition is, for AIC, n*log-lik(model) + 2k . For the Gaussian case it can be shown that the log-lik of the model reduces to the RSS, but that is a special case of a more general definition. In the caption of the table o_i and p_i should be written using capital letters for consistency with the table content
* Although he case study is quite interesting I find it is fairly hard to generalise anything from this. How do we know that this approach to PDS modelling is any more suitable than the other currently employed approaches? How could we evaluate that? How does this work in other places? How does this perform under different scenarios of true underlying processes? Overall I think the study does not give enough details about how generalisable the findings are (and actually it is not very clear what the main findings are). The note presents a modelling framework and applies it, but I feel it fails to convince the reader that this modelling approach is somehow better or worth adding to the currently available modelling tools. In particular I feel the modelling approach as presented is still very much needing the analyst to make some a-priori choices: something that is one of the main issues which make the widespread use of PDS harder to implement.

Some other small minor points in the presentation:

Line 22: interference -> inference
Line 29: the average number of events can be hardly be larger than the total number of annual maxima. It is often the case that the total number of PDS observations are more than the AMS observations, but this depends on the threshold: a very high threshold might result in PDS which have less observations than AMS.
Line 52: "gave the best results": in what sense? using what metrics? (This is a fundamental question which might also be addressed in the note: how do we evaluate what methods work well?)
Line 174: "To justify" sounds like an odd wording, maybe "to verify"? Further I would provide some more description of the test (very briefly) specifying the null and alternative hypothesis being tested and how to interpret the result (since these are not really commented on in the text around Figure 2)
Section 5: I would expect somewhere a plot showing the data series
Line 379: the threshold is much higher than 2.47 or 3.22: the threshold which is exceeded on average between 2.47 and 3.22 times per year.