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 organized 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.

1. I feel their 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?

In the present study, parameters of the distributions are estimated using L moments, and the reason for the same is described in the manuscript. However, other estimation methods, such as maximum likehood, probability-weighted moments, or method of moments, may be used. Different parameter estimation methods might lead to the selection of different thresholds and distribution choices as the value of entropy may vary accordingly. A similar analysis was carried out by Bezak et al. (2014) where they observed the better performance for the method of L-moments (ML) when compared with the conventional moments and maximum likelihood estimation.

But, all these choices available in the modelling pipe-line will not affect the core of the methodology described in section 3, i.e., the optimum threshold is selected as the one where the total entropy of a PDS model is the maximum. So for any particular parameter estimation technique, the entropy-based methodology proposed in the study will work in the same way, leading to selecting an optimum threshold and the respective distribution models.

2. 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 Generalized Pareto distribution respectively: these are motivated by some well-known theoretical results. The Negative binomial extends the Poisson distribution, allowing for over dispersion. 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 Poisson assumption for modeling the number of exceedances above a threshold is the traditional one; however other studies have proposed suitability of Binomial and Negative binomial distributions for the same (Lang et al., 1999; Önöz and Bayazit, 2001; Nagy et al., 2017). As per the Dispersion index (DI) test proposed by Cunnane(1979) if the value of dispersion index (i) falls within the lower and upper critical DI value ($I_{\alpha/2.}$ $I_{1-\alpha/2}$), at a particular significance level α, Poisson process is accepted. If $i < I_{\alpha/2}$, Binomial could be an alternative, and for $i > I_{1-\alpha/2}$, Negative binomial could be used. In the present study, DI at a 5% significance level was calculated at different thresholds and plotted in Figure 5 of the main manuscript. From the plot, it is clear that different distribution models are suitable for different thresholds based on DI value. For the range of thresholds selected from test 1 and test 2 proposed by Lang et al. (1999) (as shown in Figure 3 and Figure 4 of the main manuscript), suitable distribution models such as Poisson and Binomial are chosen from the DI plot. The authors agree with the reviewer that the negative binomial (NB) extends the Poisson distribution, allowing for over-dispersion, and for the study area at some lower value of thresholds, NB was the suitable candidate. However, this range of thresholds was dropped in the further analysis based on Figure 3.

GEV, LP3, P3, and GP distributions are applied to model the magnitude of exceedances in the present study, whereas GP has a more theoretical background. Nagy et al. (2017) carried out PDS modeling of the same study area using similar distributions. So for a better comparison of threshold and distribution models obtained from the present study with their findings, probability distributions from extreme value and Pearson family are also considered in the study along with traditional GP distribution.

3. 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.

For a statistical model with 'k' parameters, AIC developed by (Akaike, 1974) is applied to find suitable probability distribution. It represents the model's lack of fit and unreliability due to the number of parameters. AIC is expressed as, AIC = -2(log maximum likehood for the model) + 2(number of fitted parameters) = -2lnL + 2k

For 'n' data points assuming the error to be independent identical normally distributed, AIC = 2k + n ln(RSS/n); where RSS is the residual sum of squares. BIC can be expressed as BIC = n ln(RSS/n) + k ln(n) with the same assumption of errors obeying Gaussian distribution.

These expressions of metrics have been applied in some previous studies, such as (Karmakar and Simonovic, 2008; Karmakar and Simonovic, 2009; Zhang and Singh, 2007). Based on this literature survey, these definitions of AIC and BIC (Table 3) of the manuscript are applied in the present study.

The necessary corrections are made in the caption of the table.

4. Although the case study is quite interesting I find it is fairly hard to generalize 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 generalizable 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.

The present research attempts to propose an alternate method for selecting the optimum threshold in PDS modeling of flood frequency analysis. This method is applied to the daily discharge for the Waimakariri River at the Old Highway Bridge site. The reason being, (i) this is one of the frequently flooded watersheds, and (ii) a recent FFA on this area using the same data series was conducted by Nagy et al. (2017) which provides a base for comparing the results. However, the proposed methodology can also be explored at other sites, which is out of the scope of the recent work. From the literature survey, it's observed that the significant uncertainty in the application of the PDS model in FFA lies in the fact that there is no single method that performs the best for threshold detection in PDS. So the applicability of this proposed entropy-based approach is analyzed only by comparing the value of optimum threshold obtained from some existing guidelines and previous studies.

The authors represent entropy as an alternate tool for threshold selection in the PDS model and found that the threshold obtained from this is close to the other techniques. This is the first study of its kind, where the concept of entropy is applied in PDS modeling of FFA, and also it helps to locate the optimum threshold and the dual models appropriate to model the respective PDS.

Some other small minor points in the presentation:

Line 22: interference -> inference

The correction is applied in the revised manuscript.

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.

The required changes are made in line 29.

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?)

The authors have discussed in detail in the main manuscript (Page number 2) based on which metrics best results were obtained by (Nagy et al., 2017). They analyzed the degree of fitting of PDS samples to the magnitude of exceedances at various thresholds using three statistics such as Chi-square, Kolmogorov-Smirnov (KS), and Filliben Correlation (FCC). They compared the results to find the value of threshold at which PDS has lower χ^2 and KS with higher FCC value.

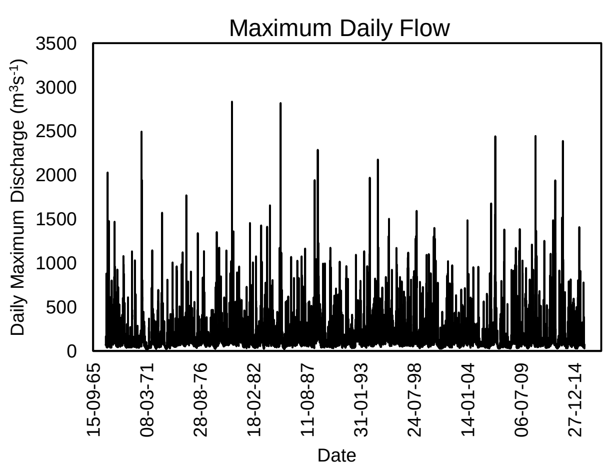
In the recent work, total entropy at each threshold is compared to find the maximum value. The various error statistics listed in Table 3 are used to evaluate the degree of fitting of the magnitude of exceedances. However, different methods existing in the literature for threshold identification in PDS are compared based on the value of optimum threshold only.

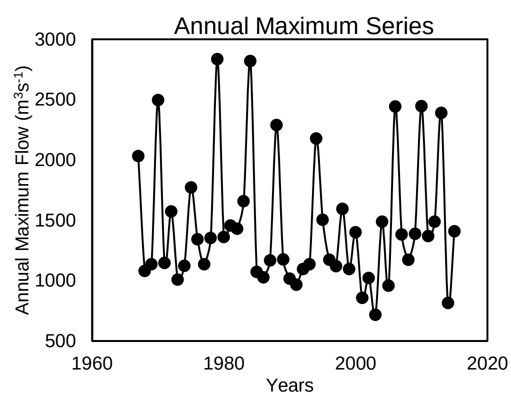
Line 174: "To justify" sounds like an odd wording, maybe "to verify"? Further I would provide some more description of the test (very briefly) specifying the null and alternative hypothesis being tested and how to interpret the result (since these are not really commented on in the text around Figure 2)

The details of these statistical tests are included in Appendix A of the revised manuscript.

Section 5: I would expect somewhere a plot showing the data series

Figure 2 of the revised manuscript represents the daily and annual maximum data series of the study area.





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.

The corresponding changes are made in the revised manuscript.

References

Akaike, H.: A New Look at the Statistical Model Identification, IEEE Trans. Automat. Contr., 19(6), 716–723, doi:10.1109/TAC.1974.1100705, 1974.

Bezak, N., Brilly, M. and Šraj, M.: Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis, Hydrol. Sci. J., 59(5), 959–977, doi:10.1080/02626667.2013.831174, 2014.

Cunnane, C.: A note on the Poisson assumption in partial duration series models, Water Resour. Res., 15(2), 489–494, doi:10.1029/WR015i002p00489, 1979.

Karmakar, S. and Simonovic, S. P.: Bivariate flood frequency analysis: Part 1. Determination of marginals by parametric and nonparametric techniques, J. Flood Risk Manag., 1(4), 190–200, doi:10.1111/j.1753-318x.2008.00022.x, 2008.

Karmakar, S. and Simonovic, S. P.: Bivariate flood frequency analysis. Part 2: A copula-based approach with mixed marginal distributions, J. Flood Risk Manag., 2(1), 32–44, doi:10.1111/j.1753-318X.2009.01020.x, 2009.

Lang, M., Ouarda, T. B. M. J. and Bobe´e, B: Towards operational guidelines for over-threshold modeling, J, Hydrol, 225, 103–117, 1999.

Nagy, B. K., Mohssen, M. and Hughey, K. F. D.: Flood frequency analysis for a braided river catchment in New Zealand: Comparing annual maximum and partial duration series with varying record lengths, J. Hydrol., 547, 365–374, doi:10.1016/j.jhydrol.2017.02.001, 2017.

Önöz, B. and Bayazit, M.: Effect of the occurrence process of the peaks over threshold on the flood estimates, J. Hydrol., 244(1–2), 86–96, doi:10.1016/S0022-1694(01)00330-4, 2001.

Zhang, L. and Singh, V. P.: Trivariate Flood Frequency Analysis Using the Gumbel–Hougaard Copula, J. Hydrol. Eng., 12(4), 431–439, doi:10.1061/(asce)1084-0699(2007)12:4(431), 2007.