General comments:

The technical note “Flood frequency study using partial duration series coupled with entropy principle” is interesting. The subject is practical in flood frequency analysis. However, I could not determine the tangible advantage of the applied method. From my point of view, this paper is more like a research paper than a technical note. In general, it follows the scopes of the HESS journal. The advantage and novelty of the paper have to be highlighted in the manuscript. Therefore, the manuscript has more room for improvement.

Specific comments:

L7 & L14: In the text, you mentioned “quality discharge” several times; what does this term mean?

The authors have mentioned the importance of quality discharge measurement and frequency analysis for effective design flood estimation. The quality discharge signifies the availability of river discharge measurements delivered in real-time. Stream stages and the related discharge determine the hazard level during any flood event. So the availability of the good quality of measured discharge or streamflow data plays a vital role in flood estimation and risk management. The discharge records represent the ground-truth data for developing and continuously improving the hydrologic model’s accuracy for forecasting stream flows. So acquiring quality discharge data for streams is critically essential for design flood estimation.

L17-19: In the abstract, more focus on results is needed. Why “POME” is an effective tool?

The authors have modified the abstract by adding the following results from the study.

“For all the four candidate distributions, the average number of peaks per year at the optimum threshold was between 2.47 to 3.22. The PDS sample with λ of 3.2 with Log Pearson type 3 and Poisson model performed better. Also, the results obtained from the proposed method were by the standardized procedure applied using eight model selection criteria.”

The principle of maximum entropy given by Jaynes (1957) states that while making inferences from limited available data, the probability distribution with the maximum entropy is the best to represent the data. Such a probability distribution is the "largest one"; it will ignore no possibility, being the most uniform one, subject to the given constraints. This minimally biased distribution will be more probable or less predictable than other distributions with lower entropy values. Therefore, while characterizing unknown events or limited data with any statistical model, one should prefer the maximum entropy distribution. The following lines are added in the abstract to include the effectiveness of the POME tool.

“The POME allows choosing the distribution with the maximum entropy from the set of all probability distributions compatible with one or more mean values of one or more random variables. Initially introduced for solving a problem in statistical mechanics, POME has become a widely applied tool for constructing the probability distribution in statistical inference. Because here the information is generally expressed by mean values of some random variables with a need of a suitable probability distribution which ignores no possibility subject to the relevant constraints.”

In the manuscript and especially in the introduction, you used 34 references that are old (before 2000). It is better to employ recent research. However, it is not a critical point.

Some new pieces of literature are added in the modified manuscript, such as Rosbjerg and Madsen (2004), Ben-Zvi (2009), Deidda (2010), Bhunya et al. (2012), Bhunya et al. (2013), Shinyie and Ismail (2012), Caballero-Megido et al. (2018), Pan and Rahman (2021).

L31-33: Could you please elaborate more? How is it possible to thoroughly evaluate the flood generating processes?!

A Partial Duration Series (PDS) compromises traditional time series analysis and Annual Maximum Series (AMS) modeling. It represents more information about any flood event as it involves dual modeling, i.e., the magnitude and time of arrival of peaks above a threshold. Generally, in AMS, we only focus on modeling the maximum value of each year, while in PDS, maximum values higher than a threshold and their rate of occurrence are considered. So it provides a better way to thoroughly evaluate the flood generating process as it incorporates the time of arrival of peaks.

L37: What does "better performance of PDS" mean?

Here the better performance of PDS as compared to AMS signifies less sampling variance of T year return period estimations Q (T) than AMS. For example, Cunnane (1973) observed that the PDS estimate of Q (T) for the same range of return periods has a smaller sampling variance than the AMS estimate only if the PDS has a λ of 1.65.

L39: Please explain "Poisson arrival of peaks" before mentioning it in the text.

Poisson arrival of peaks means when the rate of occurrence of peaks above a threshold is suitably modeled using Poisson distribution. The same has been described in the revised manuscript.

L 41: What do you mean by "Poisson process"? Readers demand to have clear fundamental literature in the introduction.

A Poisson process is a model for a series of discrete events where the average time between events is known, but the exact timing of events is random. The arrival of an event is independent of the event before (waiting time between events is memoryless). It is usually used in scenarios where we count the occurrences of certain events that appear to happen at a specific rate but entirely at random. In the case of a Poisson process, events are independent of each other, i.e., the occurrence of one event does not affect the probability another event will occur; the average rate (events per time period) is constant, and two events cannot happen at the same time. The same has been included in the introduction section before presenting the literature on the Poisson model.

Please reflect and indicate your method advantage in the introduction. By having a wide variety of "λ", what is entropy-based models' preference?

In the traditional statistical approach, the degree of fitness of various probability distributions to model the magnitude of exceedances is compared using some GOF metrics. Based on this, the threshold with better performance of error metrics is selected. The present study is the first of its kind where entropy is applied to locate the optimum threshold in PDS modeling of FFA. Instead of considering only the degree of fitness of magnitude of exceedances, the proposed methodology includes entropy of both the models of PDS, i.e., the arrival rate of peaks and their magnitude, to find the optimum threshold and the respective distributions. The advantage of the proposed method has been included in the introduction section of the revised manuscript. By having a wide variety of 'λ,' the entropy-based model suggests selecting the value of λ where the combined entropy of both the models is the maximum from a region where an increase in threshold causes a decrease in λ value along with linearity in the mean residual life plot.

You mentioned several times “probability dist.” And “fitting dist”, what are your purposes to point them in the introduction part? I understand what you did, but is it not vivid in your manuscript.

In the introduction, the authors mentioned “fitting distributions”, i.e., fitting probability distributions to the magnitude of exceedance in PDS. In PDS, one model is used for the arrival rate of peaks above a threshold and the other for their magnitude.
The authors agree with the reviewer’s suggestion, and the same has been taken care of in the revised manuscript.

L66-72: This paragraph must be rewritten to address the purposed method's necessity and novelty. Now, I did not get any points.

Even though several methods exist for threshold identification in PDS modeling of FFA, there is no universal guideline for the same. The present study is the first of its kind, where entropy is applied in PDS modeling. Instead of considering only the degree of fitness of magnitude of exceedances like in the existing standardized statistical approaches, the proposed methodology includes entropy of both the models of PDS, i.e., the arrival rate of peaks and their magnitude to find the optimum threshold and the respective distributions. The proposed methodology is applied to the daily discharge data of the Waimakariri River at the Old highway bridge site in New Zealand. Similar changes are made in the revised manuscript.

L75: In this sentence, what do you mean by “dual”?

Here ‘dual’ means the two components of a PDS model, i.e., (i) to model the arrival rate of peaks above a threshold and (ii) to model the magnitude of these peaks.

Table 1: What is “Γ”? I did not find its definition in the text.

“Γ” represents the gamma function. It’s added in the footnote of Table 1.

L100: What is the benefit of the “negative binomial dist.” in your context?

Negative binomial (NB) distribution is an alternate choice of discrete probability distribution to model the arrival rate of peaks in PDS apart from Poisson and Binomial distribution based on the value of dispersion index. From Figure 5, it’s clear that NB is suitable for some low range of thresholds. For these values of thresholds, the average number of peaks per year increases with the threshold (Figure 3). So these are excluded for further entropy analysis. However, it might be possible for any other study area to calculate model 1 using NB distribution as it extends the Poisson distribution, allowing for over-dispersion. For such cases, the expression of NB distribution given in Table 1 can be applied in the entropy expression.

L119: Why, in this paper, “e” has to be in a logarithm base?

The logarithm base in the expression entropy defines the unit of entropy. Here, “e” is used as the base in the entire work; however, some other units can also be used as they won’t affect the core of the methodology proposed in the study.

L129-130: “therefore, while … (Lee et al., 2011). I do not understand this sentence.

Here the importance of the principle of maximum entropy (POME) theory is described as observed by Lee et al. (2011). For a statistical model with limited available data, the distribution with the maximum entropy should be chosen as it will be more probable or less predictable than other distributions with smaller entropy values.

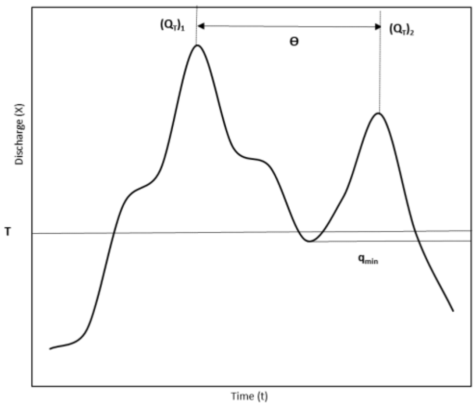
L131-136: irrelevant to previous sentences; it has to be somewhere else.

This section represents some literature on applying the POME theory, which has now been shifted to the introduction section in the revised manuscript.

L163-165: Rewrite the sentences. It is not understandable.

“The peak discharge in a PDS also depends upon various catchment dynamics with respect to space and time, such as catchment area, the frequency of rainfall and their magnitude, etc. So the independence criteria of peaks can depend not only on statistical phenomenon as proposed in some previous literature.”

L169-171: Could you please graphically explain this condition? then “intermediate” discharge can be intelligible in L173.



This graph is added to the revised manuscript.

L171: Do you mean mathematically and logically using OR in Eq. 13? Because it has to be in this form.

θ < 5days + ln (A) **or** $q_{min} > (3/4) \min [q_1, q_2]$; this is the right expression.

Table 3: ADC has to be mentioned after AD, not in the end.

The changes are made in the revised manuscript.

L200-201: Does not have vivid meaning.

As described previously, the PDS extracted at any threshold comprises two models; a discrete distribution is used to model the arrival of peaks per year (M1). A continuous probability distribution fits the magnitude of these peak values (M_2). The present study suggests selecting a threshold where both models' combined entropy is the maximum as per the principle of maximum entropy theory.

L206: What is conventional statistics in this research? And what do you want to point by comparing these models?

The conventional statistics used in this research are listed in Table 3. The standardized statiscal procedure found in literature evaluates the degree of fitting of various probability distribution models to the magnitude of exceedances applying one or more such statistical measures. So the authors have compared the result obtained from the proposed methodology (overall degree of fitting of candidate distributions at various thresholds) with this conventional standardized procedure to analyze whether this entropy-based analysis can be applied as an alternate for threshold selection.

Page 9: What does this method work if λ=2? Two independent events per year. Is there any way to calculate the threshold by assuming two events per year?

λ = n/N; where 'n' is the total number of peaks above the threshold and N is the number of years available data. Here, N = 49, so the threshold at which there are 98 peaks (n) above it corresponds to a λ = 2. So for Figure 4, the independence criteria are applied to extract the PDS at each threshold starting from the minimum daily discharge (22.033m^3/s). This analysis shows that at t = 920.368 m^3/s, there are 98 peaks above it, referring to λ = 2.

Figure 1: It is good instruction; however, you need to elaborate more on the second box "Extract PDS at …", explain the third box in the text, before this figure, and answer the question of "what if for non-linear approach" for the fourth box.

The second box: Extract PDS at all the thresholds starting from the minimum daily discharge and apply USWRC independence criteria; check for independence using Modified Mann Kendal's Tau test and autocorrelation plots

An explanation for the third box (before Figure 2): The gradual variation of the average number of peaks per year divides the entire range of thresholds into four domains, as described by Lang et al. (1999). For a PDS, an extremely low threshold makes the whole series lie above it in domain 1. Then with an increase in threshold, more peaks are identified and retained in domain 2. Further rise in threshold makes a decrease in the average number of peaks per year in domain 3, and finally, when the threshold reaches the time series maximum, no peaks are retained in domain 4.

Fourth Box: Linearity of MRLP gives us a rough idea about a range of thresholds where the optimum threshold might lie. It also implies choosing a threshold to maximize the stability of the distribution parameter estimates for the PDS. If for some cases, no region of linear variation is found in the MRLP, the independent thresholds from Domain 3 (Figure 4) can be considered further. However, the MRLP and parameter stability plot should be analyzed from the bootstrap sampling to ensure the same.

L 225-226: remove it. It is not relevant. "Flood management …"

The lines are removed in the revised manuscript.

L228: What does excellence mean in FFA? Does it mean long-term?

Here the excellence of the data set means the long–term available data and its quality (with very few missing values).

Table 4: The mean and std of maximum daily flow are the same. I do not have your data. Do you think, is it correct?!

There was a mistake in the calculation. The values have been modified in the revised manuscript.

L238: What are the applied thresholds? Please write them in this part.

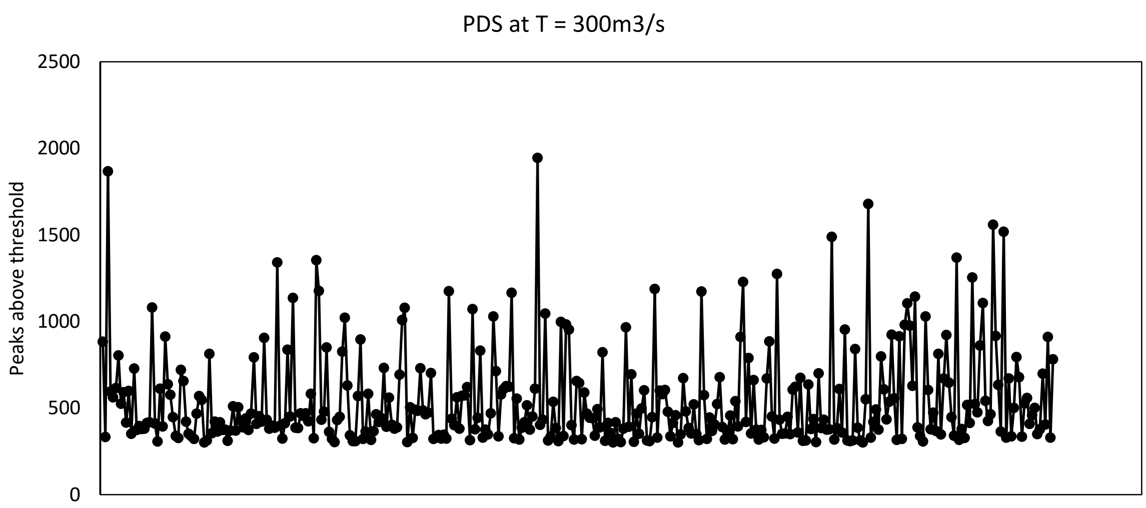
Initially, each data point of the daily data series starting from the minimum daily discharge (22.033m^3/s) is considered a threshold. USWRC independence criteria removed the dependent peaks (Figure 3). After identifying a suitable range for the threshold from the graphical tests proposed by (Lang et al., 1999), i.e., in Figure 3, arbitrary thresholds are selected within that range at an interval of 10 m^3/s such as 220 m^3/s,230 m^3/s,240 m^3/s …….1150 m^3/s. Similar explanations have been included in the revised manuscript.

Figure 2: in a, did you omit the values upper the critical dash line? What is the interpretation of the negative values in Kendal tau (y-axis)?

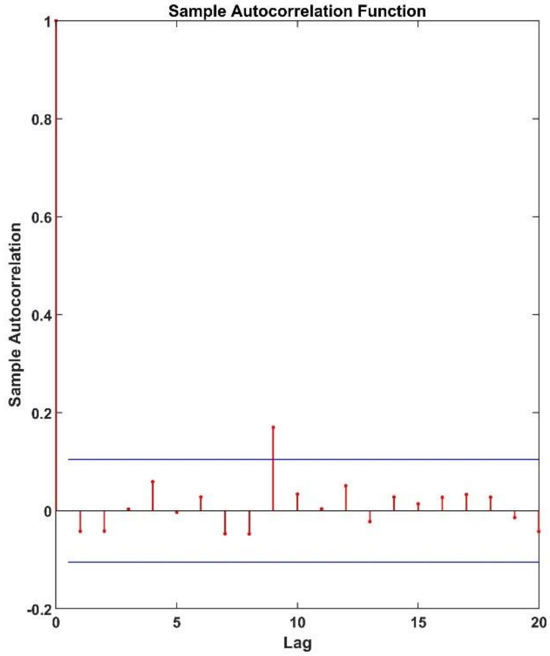
The thresholds at which tau is greater than tau critical (upper dashed line) are omitted from further analysis as it violates the independence assumption of peaks. The negative Kendal tau value represents a negative correlation among the variables.

I am curious to know the reason for the higher correlation in 9 step time lag in b.

Figure 2b represents the autocorrelation plot for the PDS extracted at a threshold of 300 m^3/s. The partial duration series is shown below.



The authors have used MATLAB software to plot the autocorrelation graph using autorcorr() function, and the result is shown below.



The PDS graph shows that except for two higher discharges, the peaks extracted at a threshold (T) of 300 m^3/s do not have any particular pattern, i.e., they tend to fluctuate randomly. The presence of these two higher values at 9 step time lag might be the reason for a higher correlation. But since most of the spikes are not statistically significant, this implies that the peaks are mainly independent of each other.

L244: How many peaks did you select in the designated threshold?

Only independent peaks are selected at each threshold, those retained after applying USWRC criteria.

Figure3: is an excellent figure, but a question rase up, why did you consider values below λ=1? What is the benefit of showing, for example, eight peaks per year in FFA? Because they are not "flood" anymore.

Here to analyze the graphical test proposed by Lang et al. (1999), we have considered all the peaks starting from the lowest daily discharge value. The number of independent peaks in the PDS and the respective λ values are calculated at each point. So it also includes the values below λ = 1 to identify all the four domains of the plot. Eight peaks per year are not shown for any particular reason; it's just the part of the entire plot of t vs. λ as λ attains a maximum of 8.22. This represents the lower limit of domain 3.

L278-279: How do you recognize the linear behavior in the figure? This is an entirely ocular and non-mathematical diagnosis. Where does the plot start to shift?! How do you consider linearity if the threshold in Figure 4b? By changing the y-axis, it is not linear anymore!

The authors agree with the reviewer that it's based on visual observation. It's observed from the calculated values that for a threshold varying between 220 m^3/s to 1150 m^3/s, the mean excess varied between 216 to 360 m^3/s. So the authors have identified a region where the value of mean excess doesn't show much variation with change in the threshold, i.e., within a threshold of 550 to 1000 m^3/s, mean excess showed a slight variation around 300 m^3/s. The Y-axis of Figure 4b is made the same as Figure 4a in the revised manuscript.

L283: explain more.

A similar explanation given in the previous question is added in the revised manuscript.

Page 14: What is the range of dispersion index? What does 1 in your study mean? What is/are the reasons for having high DI at low thresholds? and reflect it in the manuscript.

Dispersion index can take any positive value greater than one or less than or equal to 1. For Poisson's process, mean equals to the variance lead to DI of 1. More the line comes close to the line DI = 1, Poisson's hypothesis becomes more applicable. In the present study, DI values are plotted to identify suitable probability distribution to model the arrival rate of peaks. It's unnecessary to highlight the DI = 1 line in our context. So the required changes are made in the revised manuscript. At low threshold values, more peaks above the thresholds cause over-dispersion, leading to high DI values.

L319: "The average number of peaks per … 2.5 to 3.2", It is a wide range for long-term high-resolution time series. What do you think about that? Could you suggest $\lambda=3$ as an average value for your case study area? Or, it is still sensitive to this range.

Here, the authors expressed the average number of peaks per year of the four candidate distributions at their respective optimum thresholds (Table 5, column 3). For P3/PD, the optimum threshold has a λ of 2.47 while GEV/PD and GP/PD λ equals 3.22. "The average number of peaks per year varied between 2.5 to 3.2": this line was not the correct representation. So it's been modified in the revised manuscript.

Table 5: It is not needed to write the λ column here. It is not in the continuation of the entropy section.

The same has been removed from Table 5.

L320-324: Do not need to mention here.

These lines are removed from here as they have already been discussed in the methodology section.

L324: Rewrite the sentence "A similar analysis …"

"The degree of fitting of the four candidate distributions were assessed and ranked according to the test statistics (Table 3). The final rank was computed at the selected thresholds as described in the methodology section."

L339-340: Did you have any other expectation for having a higher design flood for the considerable return period in GEV?!

The authors did not find any apparent reasons for having a higher design flood for a considerable return period in GEV.

Figure 9: What is the reason for the abrupt jump around 950? I know, what is LL & UL but please mention their abbr.

The PDS extracted at $950 m^3/s$ contains some dependent peaks (tau=0.122, tau critical = 0.1082). So the presence of these peaks might have caused an abrupt jump around 950. LL and UL represent the lower and upper limit at 95% confidence level for the bootstrap sampling. Similar changes are made in the revised manuscript.

L359: Why and how do you select 730?

According to the guidelines proposed by (Lang et al., 1999), (1) identify an interval of threshold values which gives good results for tests nos. 2 and 3 (Figure 4 and Figure 5 in the present study); (2) select within this interval the largest threshold with $\lambda > 2$ or 3 (test no. 1) (Figure 3). Based on this, 730 is selected.

Page 17: Still, I do not understand the advantages of your method!. Is it faster? Is it hydrologically more reasonable? Is it prevent to do some additional steps?

Instead of applying several statistical measures to assess the degree of fitness of models to the value of exceedances, only this combined entropy of both the models proposed in the study can lead to similar results of the optimum threshold. It's hydrologically more reasonable as it considers the entropy of both the models of a PDS, i.e., the uncertainty involved in both the models while finalizing the threshold. The authors have explored the application of entropy in PDS modeling and proposed this POME-based approach as an alternate for other threshold identification techniques.

Conclusion question: Is it possible to have no peaks per year? I mean, the average peak per year is 3.2, and theoretically, it is possible to have several independent peaks in a year and no peak in another year. Did you have such a drought year or period?

Yes, it's possible to have a dry year, i.e., no peaks are above the selected threshold in a particular year. Yes, it's possible for any value of λ. Such drought years were observed at some threshold in our study also.

Technical corrections: The authors would like to thank the reviewer for suggesting these technical corrections, and the required changes are done in the revised manuscript.

L36: You already mentioned "λ" in the text.

L46: Please define EDF abbr.

EDF stands for Empirical Distribution Function.

L93-94: Please write this part in equation format.

L93-94 is written in equation format in the revised manuscript.

Table 1: Please cross-check the L moment expressions. I believe it has a mistake in the typing.

The expression is corrected.
$C = 2/(3+t_3) - (\ln 2/\ln 3)$

L98: GPA is wrong. It is GP all over the manuscript.

The GPA is changed to GP in L98.

L 109: Can be merged to the above equation.

It’s merged with the equation written above.

Page 5: “y” and others are not the same format as other parts of the paper. i.e., “y” à y

L 114: by (Shannon, 1948) is the wrong citation form.

L 141: Eq. (4), while in line 124, it is written Eqn. So it has to be the same in all parts of the text.

L 149: the “Generalized extreme value” should be “Generalized Extreme Value” or “generalized extreme value”.

“Generalized Extreme Value”

L154: Why eq. 11 is bold?

L177: The repetition of (PDS) is not needed anymore.

L185: When you mention the “Schwarz bayesian criterion” instead of the “Bayesian information criterion” term, you should write SBC, SIC, or SBIC, not BIC.

The corrections are done in the revised manuscript.

L203: “The Dispersion ind.” Should be “The dispersion ind.”.

It’s been changed to “The dispersion index”.

L228: Extra parenthesis

It is removed in the revised manuscript.

L239: in Sect. 2.4 ---à in Sect. 2.4.

Figure 3: Please fix the place of the arrow for the domain3.

The same has been corrected.

Figure 4: It is better to have the same x-axis (200-400). Also, y-ais in b is not appropriate.

The required change in Figure 4 is done.

L284-285: by Cunnane is mentioned several times.

L286: DI, did you mention this abbr in the text before?

Please take care of using abbr in the text. Sometimes, it seems that they are written too much!!

DI stands for Dispersion Index. All such abbreviation-related mistakes are modified in the revised manuscript.

Figure6: Different y-axis makes it difficult to compare total entropies. You can at least use the same minor grid with two decimals.

Using colors may be better to show the result. Sometimes it is not easy to recognize the exact points.

Figure 6 is modified with colors and the same range of the Y-axis.

L325: KS and AD statistics

Figure 7&8: Surely use colors. Legends are not readable for me.

Figure 7 and Figure 8 are modified accordingly.

L373: Different abbr. at OH. Sometimes it is OBH.

The same abbr. OHB (for the Old Highway Bridge site) has been updated in the revised manuscript.

References

Ben-Zvi, A.: Rainfall intensity-duration-frequency relationships derived from large partial duration series, J. Hydrol., 367(1–2), 104–114, doi:10.1016/j.jhydrol.2009.01.007, 2009.

Bhunya, P. K., Singh, R. D., Berndtsson, R. and Panda, S. N.: Flood analysis using generalized logistic models in partial duration series, J. Hydrol., 420–421, 59–71, doi:https://doi.org/10.1016/j.jhydrol.2011.11.037, 2012.

Bhunya, P. K., Berndtsson, R., Jain, S. K. and Kumar, R.: Flood analysis using negative binomial and Generalized Pareto models in partial duration series (PDS), J. Hydrol., 497, 121–132, doi:https://doi.org/10.1016/j.jhydrol.2013.05.047, 2013.

Caballero-Megido, C., Hillier, J., Wyncoll, D., Bosher, L. and Gouldby, B.: Technical note: comparison of methods for threshold selection for extreme sea levels, J. Flood Risk Manag., 11, 127–140, doi:https://doi.org/10.1111/jfr3.12296, 2018.

Cunnane, C.: A particular comparison of annual maxima and partial duration series methods of flood frequency prediction, J. Hydrol., 18(3–4), 257–271, doi:10.1016/0022-1694(73)90051-6, 1973.

Deidda, R.: A multiple threshold method for fitting the generalized Pareto distribution to rainfall time series, Hydrol. Earth Syst. Sci., 14(12), 2559–2575, doi:10.5194/hess-14-2559-2010, 2010.

Jaynes, E. T.: Information theory and Statitical mechanics, Phys. Rev., 106(4), 620-630, 1957**.**

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

Lee, S., Vonta, I. and Karagrigoriou, A.: A maximum entropy type test of fit, Comput. Stat. Data Anal., 55(9), 2635–2643, doi:10.1016/j.csda.2011.03.012, 2011.

Pan, X. and Rahman, A.: Comparison of annual maximum and peaks-over-threshold methods with automated threshold selection in flood frequency analysis: a case study for Australia, Nat. Hazards, doi:https://doi.org/10.1007/s11069-021-05092-y, 2021.

Rosbjerg, D. and Madsen, H.: Advanced approaches in PDS/POT modeling of extreme hydrological events, in

Hydrology: Science and Practice for 21st Century, Proceedings of the British Hydrological Society International Conference, edited by B. Webb, N. Arnell, C. Onof, N. MacIntyre, R. Gurney, and C. Kirby, pp. 217–220, Imperial College, London, U. K., 2004.

Shinyie, W. L. and Ismail, N.: Analysis of T-year return level for partial duration rainfall series, Sains Malaysiana, 41(11), 1389–1401, 2012.