Author Comment: Accounting for Hydroclimatic Properties in Flood Frequency Analysis Procedures

Joeri B. Reinders and Samuel E. Munoz

Reviewer 2

Thank you for the opportunity to review the subject-manuscript. In this study, the authors evaluate the effects of accounting for hydroclimatic properties in flood frequency procedures. While the authors evaluate the importance hydroclimatic information for various flood frequency distributions in the hydrologically diverse United States, the general approach of climate regions may be too broad, resulting in an overgeneralized description of the best candidate distributions across the U.S. Additional comments are also provided to improve the overall study and findings.

We thank Reviewer 2 for the extensive comments on our manuscript. We agree that climate regions (and precipitation) are broad categories for flood distributions. However, in practice procedures for flood frequency analysis often apply only a single parametric distribution – and this work is to show that generalities like climate and precipitation can be used to provide higher accuracy. We do agree with many of the comments from Reviewer 2 about a more nuanced description of our results, especially with regard to trend in the skew. We will adopt these changes to better represent the debate on the usability of broad categories within flood frequency analyses.

Specific/technical comments:

In the abstract and Introduction sections, why do the authors describe evaluating the performance of 500-year flood events? Does this seem reasonable given the more common lengths of annual peak discharge records that are commonly less than 75 years of record? Furthermore, why do the authors state that we can estimate a 1,000 or 10,000-year flood event based on these records and draw conclusions about candidate distributions this far in the tails when we do not know the true distribution? These descriptions seem to counter what the author's used for their analysis: "The annual maxima records in the final dataset have an average length of 78 years and a range from 30 to 118 years." Therefore, recommend revising the manuscript to describe drawing inference for the performance of the candidate distributions among the heavy tails closer to the 100-[at most] the 500-year annual exceedance probabilities.

We agree with the important point that Reviewer 2 makes about the usability of relatively short observational records for the analysis of extreme flood probabilities. We agree that the records used in this analysis would be insufficient to accurately estimate for example 1,000 or 10,000-year floods. Yet, frequency analyses are used to compute flood probabilities of these large return periods – whether it be for infrastructure design purposes or to estimate the return period of an extreme event. We agree the abstract could be rewritten to better reflect the large uncertainties associated with these type of analyses – something we try to address with this research.

As shown in Figure 2b, for skews ~ -0.1 to 0.2, the LP3 performs well just as the LN3. However, the LP3 performs poorly for negative skews <-0.1. This is why B17C proposes the LP3 distribution, but with the expected moments algorithm (EMA) to estimate the moments. It has long been recognized that there are statistically low outliers in a block maxima approach (e.g., selecting one annual peak per year) for a flood frequency analysis. To handle low outliers, B17C, recommends using the multiple Grubbs-Beck test (MGBT) to detect these statistical low outliers (e.g., potentially influence low floods (PILFs)) that may have undue influence on the upper right-hand tail (e.g., low exceedance probabilities). If PILFs are detected, those flows are recoded as censored flows and the EMA method is used to estimate the parameters of the LP3 distribution. If a more robust statistical test for the identification and treatment of PILFs is not employed, the LP3 distribution will preform poorly, especially for those very negatively skewed distributions which are found across CONUS and not just in the arid southwest U.S., for example. Given this major caveat when using the LP3 distribution, recommend this study take that into account when describing the results of this study. And recognize the current approach for the LP3 in this study does not honor the updates to the Federal guidelines to better fit the agreed-upon use of the LP3 in the U.S.

We agree with Reviewer 2 – this is a valuable point that should be added to the discussion on the use of the LP3 distribution.

Also, because we do not know the true distribution of the annual maxima series, these caveats need to be address and discussed.

Do not think there are any consistent better distributions throughout the range of skews in the arid region. Even in the log transformed space there are no clear results of better fits. This is most likely attributed to the additional diversity among the arid region. There are regional differences between central and southern CA, the southern four corners region vs. the Midcontinent region west of the 100th meridian. Those two regions are not commonly grouped together.

We agree with Reviewer 2 that accounting for the diversity within climate zones would result in more accurate fitted distributions. However, the purpose of this paper was to explore whether broad divisions like climate and precipitation characteristics would result in specific conclusions regarding the use of one distribution over another. We agree however that the results could be adjusted to account for the nuance made by Reviewer 2.

We could run a similar analysis, for example, based on the regions used in the National Climate Assessment, which would better account for some of the diversity described by Reviewer 2.

Disagree with this description: "This shift is not observed in log-space, as log-transformed records in arid climates exhibit relatively small L-Ck compared to records from continental and temperate climates and are thus better represented by the GEV distribution (Fig. 3b)." Again, it depends on what range of skews among the arid group positive, near zero and negative.

We agree with Reviewer 2 for the same reasons as explained in the previous comments.

Disagree with this description: "Continental and temperate regions are both well represented by the LP3 and LN3 distribution (Figs. 3d and 3f)." Recommend splitting this discussion. Temperate is best represented by X distribution over different negative/positive skews. And the same type of description for the continental region. If these recommended descriptions are used, likely differences will become more apparent.

We agree with Reviewer 2 and will adjust this accordingly.

Recommend adding more thorough descriptions about positive/negative skews and the best fit of the distributions.

We agree with Reviewer 2 and will adjust this accordingly.

Do not necessarily agree that the LN3 is the best distribution for the highest Psc (figure 4e). More commonly the LN3 is higher than the upper confidence bounds but is overall closer than the P3. Recommend revising the description.

We agree with the nuance described by Reviewer 2 and will adjust this accordingly.

Disagree with this description: "The weighted moving average of the intermediate group falls in between 205 the GEV and LN3 distribution lines, indicating a shift from the GEV to the LN3 distribution as PSC values become higher (Fig. 4b)." Should this description be for figure 4c? Also, as values increase the better fits goes from the LN3 (0.2-0.45) then to the GEV (~>0.45).

This description describes the change seen from figure 4a to 4c. We agree with Reviewer 2 that a nuance on the skew parameter is in place.

Disagree with this description: "We found no strong contrast between the groups when records are log-transformed (Figs. 4b, 4d, and 4f); the weighted moving averages of all three groups follow the LN3 distribution line." Only the middle Psc values exhibit this description. The WMA for the Lowest Psc are substantially higher than the three distributions (4b), especially for skew >-0.1. And while the LN2 performs better (4f), after ~ >0.1, all distributions appear higher.

We agree with Reviewer 2 that the description of both 4b and 4f could be made more specific. It is true that 4b only follows the LN3 line for the lowest skew values. In 4f we would argue that the LN3 reflects the distributions well up till a skew of 0.5 – but we agree that it is important to make this point.

Recommend rewording the overgeneralized descriptions for figures 3 and 4.

We agree with Reviewer 2 and will make the adjustments accordingly.

Disagree with this description: "In general, our results point to the LN3 distribution as the best distribution to characterize annual maxima data from across the United States, but also demonstrate how regional hydroclimatic differences explain part of the variance among individual flood distributions." Again, this is too generalized and doesn't reflect a combination of results based on different skews.

We agree with Reviewer 2 as described in the previous comments. These changes in the results will be further translated into the discussion.

"While our work does not prove a causal relation between Köppen climate region, precipitation intensity, and flood frequency distribution shape, we do demonstrate that hydroclimatic factors explain part of the L-moment sample variance and flood frequency distribution shapes across the United States." Recommend that the author's attempt to provide some causal mechanisms, especially because the climate regions seem too broad. Recommend at least finding supporting literature of other cluster methods that have found similar regions.

We agree with Reviewer 2 and will make the adjustments accordingly.

Disagree with this description: "In our study, we demonstrate that the average statistical properties of annual hydrologic maxima across the United States are most closely represented by the LN3 distribution — although sample variance remains high (Fig. 2)." Recommend the author's further address the sample variances and how they relate to skew ranges and the corresponding better performing distributions.

We agree with Reviewer 2 and will make these changes.

“Our findings demonstrate that the distribution family that best characterized hydrologic maxima shifts from the GEV towards the LN3 distribution as we move from colder and wetter climates (Köppen group D) to arid and drier climates (Köppen group B) (Fig. 3).” Too vague and disagree with the Arid regions overall description. In Figure 3a (the Arid climate region B), the WMA is between LN3 and the P3 for skew values < 0.45. Again, the arid region is likely too broad and PILFs (low outliers) were not accounted for.

As discussed above we agree with Reviewer 2 on that these results/ discussion points can be more precisely formulated to take into account differences along the skew values.

“Our results also support findings of Pitlick (1994) who showed that local flood frequency distributions in mountainous areas of the Western United States are shaped by regional precipitation intensity.” This finding has not been talked about nor have any results been discussed specifically related to the mountainous region in the western United States. Often high-elevation sites in the western U.S. have mixed populations that strongly deviate from one particular distribution (e.g., annual peaks generated from rainfall, snowmelt and rain-on-snow. These sites can have s-shaped hooks in the right-hand tail, for example. See B17C and other references for a come complete description of mixed populations in the western U.S. These high elevation sites are not broadly related to regional precipitation intensity.

Here we did not mean to say that our results explain specifically distribution variations in mountains areas, but only that the precipitation intensity explains some of the variance in distribution shape as was also shown by Pitlick (1994) for mountainous watersheds in the Western United States. We agree with Reviewer 2 that these results can be formulated more precisely and that we should take into account the broader research on high elevation sites. We will adapt the discussion accordingly.

Disagree with this description: “Climate classification schemes provide another benefit as they encompass large contiguous areas...Encountering this problem becomes less likely with Köppen regions which often cover entire watersheds.” Large regions, such as the Arid region in this study, is likely too broad and oversimplifies various seasonal controlling factors on precipitation and flood generating attributions, for example.

We agree with Reviewer 2 that there is a lot of variation within climate regions. The aim of this work is to identify whether these broad classifications can help to make flood frequency analysis within national procedures more adaptable. That is why we did not account for these variations, but we do agree we could describe how such an analysis could account for further detail

Reviewer 1

I totally agree with the importance of the objective of this paper trying to shed some light on the distribution selection for flood frequency analysis, and with the main conclusion that "probability model selection can be improved when it is based on the hydroclimatic properties of the basin". It is a concise paper and I have enjoyed reading it. Some questions from my side:

We thank Reviewer 1 for the supportive comments.

In Fig 2 and its derivations in terms of Kopen index and Psc, the WMA clearly helps to see the potential similarities with the theoretical L-moments of the 3 distributions, but what do the WMA confidence limits really provide?

In all figures (or in one representative) I am missing some "better uncertainty", such as that of the sample L-moments, in order to explain their dispersion. However, they are a function of the sample size, so it will be difficult to deal with it. What is your opinion?

We thank Reviewer 1 for these comments on the uncertainty measure in our work and combined the two points in one reply. The limits provide the 95%CI for the WMA – so not an uncertainty that relates to the spread of the sample L-moments – but rather the spread of the L-Ck's per bin in the WMA analysis. A better indicator of uncertain over the sample L-moments itself could be provided by adding the standard deviation over sections of L-sk. This would also be in line with helpful comments made by Reviewer 2 who suggested taking into account the variation along the skew axis.

In section 4.2 authors define the following steps in their research, but it would not be very complex to advance part of them. I particular

- Why only these 3 distributions with 3 parameters? Excluding the 4-parameter ones and mixed distributions (I would mention also the TCEV), I am curious about the potentiality of the 3-parameter Generalized Pareto.

We think these would be very interesting additions and can add these in an updated manuscript.

- You have used as explanatory variables the Kopen and the Psc and I totally agree with the conclusions in section 3. Did you try the catchment area, as mentioned its importance in L50 and 69?

We did try catchment area in a preliminary study however this did not give results that were significantly different from those in figure 2. However, we can add these in an appendix.

Very minor comments:

L17. (log-)Pearson 3 (P3) or Pearson 3 (P3)? I.e., (log-) is confusing.

Thank you for pointing this out, we will adjust this accordingly.

L25 and 26 are not needed. Too often authors (myself included) try to give a general framework that is too general and unnecessary.

Thank you for pointing this out, we will adjust this accordingly.