Reply to Dr. Daniel Wright (Reviewer #2)

We have replied to all the reviewer's comments in red.

Review of HESS-2021-49: "Identifying Sensitivities in Flood Frequency Analyses using a Stochastic Hydrologic Modeling System" by Newman et al.

The authors present a sensitivity study examining the relatively contributions to flood quantiles from precipitation, initial conditions, model structure and parameters, and meteorological sequencing for two watersheds in the western US. The results are interesting and useful, while the manuscript is well-written. Like reviewer #1, I think that only minor revisions are needed. Similarly, like reviewer #1, I think occasionally found it difficult to understand exactly what was done or why. I'll point out those issues that I noticed, but I agree with reviewer #1 that an overall workflow diagram might be helpful if done with care.

Thank you Dan for the positive review and very helpful comments. We agree that the manuscript needs further clarification and an additional well formulated workflow figure. Please see our point by point replies below.

Specific comments:

I agree with the authors' discussion of AEP equality assumptions, but the potential problems don't end there. Even the assumption that precipitation annual maxima—which are the values used here and in most studies—are the drivers of streamflow annual maxima is not really correct. In Yu et al. (2019; specifically, Table 3 in that paper), we found that you need to get into 200+ year return periods before that assumption is really reliable, at least for the midsized midwestern watershed we looked at. Clearly, this is less of an issue for really big floods.

Thanks for the additional discussion related to AEP neutrality assumptions. This is an interesting point and we will certainly add some of this discussion and associated references into our introduction.

L68-70: It seems like some element is missing from this sentence. "higher sensitivity…" higher than what?

Thanks for catching this, we mean that some parameters related in the hydrologic model influence FF estimates more than others in a relative sense.

Section 3.1: I think you need to provide more explanation on how you used the total probability theorem. I *think* I understand what you did, but the reader shouldn't have to guess. Out of curiosity, I'm wondering if that approach would be valid when using distributed models. With lumped models (which I assume the authors are using here, but I'm not actually sure; see below), a bigger rainfall event combined with a higher IC will result in a higher peak than a smaller rainfall combined with a drier IC. But with distributed models, that is only true in general but not universally due to routing effects—I've seen cases where this isn't.

This general point was also highlighted by the other two reviewers. Thus, we will improve our discussion regarding our use of the total probability theorem in section 3.1, and possibly add more detail to other methodological points in this section.

Also, thanks for the ideas regarding lumped versus distributed models. We are indeed using lumped models in this study. We will consider adding a few sentences in the introduction and possibly in this section regarding the nuances between lumped and distributed models and how that may impact an FF sensitivity analysis like ours.

Section 3.4: It would have been nice to know how important this assumption of picking a few (high) ICs is, as opposed to letting the ICs vary more widely. My particular concern is that to some degree or another, your rainfall quantiles are probably based in part on some events (probably some big ones!) that are outside of this Feb-July (Altus) and Apr-June (Island Park) periods. The Colorado 2013 floods are a good example of this. I suspect that there is some degree of misrepresentation of the relative importance of ICs and precipitation for this reason.

This is a good question, and one we will add some general discussion about in the manuscript. For the current study, we chose to focus on wetter initial conditions (ICs) to focus on floods with large return periods (e.g. 10,000+ years) and following general Reclamation FF estimation methodologies. Admittedly, we are showing results across even frequent return periods and the reliance on only wet ICs may influence the importance of IC uncertainty for these more frequent return periods, also mentioned by reviewer #3.

In general Reclamation focuses on larger events and wetter periods of the observed record and generally uses the range of conditions for those larger events to inform the distribution of initial conditions sampling from those events. This assumption may not be valid in all hydrologic regimes or for all events, especially in more arid environments and for unique events like the Colorado 2013 floods.

We plan to add the above clarification and discussion and emphasize the point that there may be an underestimation of the sensitivity of ICs, particularly for frequent return periods.

Section 3.5: I found the explanation of spatial precipitation structure to be unclear-both how it was done, and why it was done. In the latter case, my confusion stems from the lack of description of the models' spatial discretization (or maybe I missed that somewhere).

We will clarify that the model is lumped, therefore there is no spatial structure to the precipitation inputs. Using the gridded distributed precipitation inputs, we combine all grid cells that intersect the basin polygon using their fractional areas to create the basin mean precipitation input.

Section 3.6: I found this section difficult to follow, and didn't totally understand what was being done.

We agree this section is poorly worded and may be better served by including an example figure of the two different precipitation sequences. Essentially, we are trying to distinguish between using the standard Reclamation approach of event modeling where they have the event precipitation input

followed by zero precipitation for the rest of the simulation, and a more realistic approach of specifying the event precipitation and then continuing with historical precipitation after the specified event.

Section 3.7: While neither Peleg et al. (2017) for I (in Zhu et al. 2018) examined model structure, we did use ANOVA (in Zhu et al.) or something like ANOVA (in Peleg et al.) to examine the roles of other things (ICs, for one) in FFA. I won't be offended if you don't, but you may consider whether those prior studies' findings provide relevant contrasts with your work.

Thank you for these citations. We will examine these two studies and include relevant discussion of them in the introduction, this section, and throughout the results.

L304: Usage of "overrepresentation" is unclear.

Thanks for catching this, we meant to say 'overestimation'.

L304 and more generally: given all the moving parts here, some section referencing would help, as well as a bit more precision with terminology. For example, "KGE interval metric-based calibration"-it took me a minute to figure out what you were talking about. You mean calibration based on peak flows, right? Furthermore, referring back to Section 3.3 (e.g. "(see Section 3.3)") would help the reader the track down the relevant details they might have missed or forgotten. This section referencing would help in a number of other places too.

Thanks for the helpful suggestion. We agree that section referencing may help improve detail tracking for readers. We will identify places in the text where we can do this and use the subsequent comments from your review as well. We will also work on our terminology to improve precision and readability.

You are correct that 'KGE interval metric-based calibration' refers to annual peak flow calibration using KGE.

Section 5: It would be nice to know if the "shapes" of the flood frequency curves are driven by the shapes of the precip frequency curves, which aren't shown.

We will add a figure highlighting the precipitation frequency distributions. In general the shapes of the flood frequency curves do roughly follow the precipitation frequency curve shapes.

L365 and around there: I struggled with this paragraph, in part because I didn't understand the descriptions in Section 3.6. Also, this is another good place to refer back to earlier sections/descriptions.

Thanks for the suggestion. We will modify this paragraph and add section referencing.

L376: "Dry to historical meteorological sequence"-I found this wording confusing

We will modify this phase within the sentence.

L390: You could refer back to the first mention that you're analyzing different streamflow timescales

We will add reference to the correct previous section.

L428: consider replacing "across" with "between"

We will change 'across' to 'between'.

References:

Yu, G., D. B. Wright, Z. Zhu, C. Smith, and K. D. Holman. "Process-Based Flood Frequency Analysis in an Agricultural Watershed Exhibiting Nonstationary Flood Seasonality." Hydrol. Earth Syst. Sci. 23, no. 5 (May 7, 2019): 2225–43. https://doi.org/10.5194/hess-23-2225-2019.

Zhu, Zhihua, Daniel B. Wright, and Guo Yu. "The Impact of Rainfall Spaceâ Time Structure in Flood Frequency Analysis." Water Resources Research 54, no. 11 (2018): 8983–98. https://doi.org/10.1029/2018WR023550.

Peleg, N., F. Blumensaat, P. Molnar, S. Fatichi, and P. Burlando. "Partitioning the Impacts of Spatial and Climatological Rainfall Variability in Urban Drainage Modeling." Hydrol. Earth Syst. Sci. 21, no. 3 (March 14, 2017): 1559–72. https://doi.org/10.5194/hess-21-1559-2017.