

Reply to Reviewer #1

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

The paper is well structured while the content is very dense and properly concise. The paper gives an important contribute to the analysis of factors affecting the FF curve by performing a deep analysis of the effects produced by models structure, model parameters, interaction between model parameters and model structure, initial conditions (in terms of water content) and precipitation events. In my opinion the paper can be accepted subject to minor revision.

We thank the reviewer for their positive assessment of the manuscript and their helpful specific comments highlighting areas for further improvement. Please see our point-by-point replies below.

I suggest the authors prepare a flux diagram describing their workflow, i.e. all the steps of their procedure. Indeed, there is a lot of attention on how the available data are used and this, in the end, slightly obscures the logic and the sequence of the steps. In this (or these) diagram(s) the authors should highlight the deep meaning of each step independently of the way the available data are used for their quantification.

We have added a workflow diagram as new Figure 3. Sections 3.1, 3.4, and 3.6 have been extensively edited, and additional clarification has been added to Section 3.7.

It is not completely clear to me how the rainfall events are generated starting from the precipitation frequency curve. I understand that simulations are performed with a time step of one day and thus precipitations are generated with this time step. However, in the case of Island Park, a two-day precipitation event is generated from the frequency curve and I do not understand if the total of precipitation in two days is generated or if the event is scanned at daily level.

This is clarified in Figure 3 and Section 3.1.

Overall, I prefer continuous simulations than event simulations: in fact, in this way the "natural" combination between rainfall periods and flood periods is obtained without any artificial combination between initial conditions and rainfall events. Instead of using the regional analysis, why don't the authors have set up a rainfall model (Neyman-Scott or Poisson model), performed long simulations and extracted annual maxima from them?

We agree that continuous simulations over many hundreds to thousands of years is another way to perform FF estimation We have added discussion of this point in Section 3.1.

Incidentally, Hashimi -> Hashemi

Thank you for catching this typo, we have corrected all instances of it.

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 have added brief mention in the introduction on lines 97-99.

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

We have corrected this sentence.

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.

We have rewritten most of section 3.1 (lines 175-200) and added a detailed workflow diagram as new Figure 3. We have clarified that we are using a watershed model in this study on lines 183 and 256.

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.

We have added discussion of this point in section 3.4 on lines 309-316.

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 have clarified that we are using a watershed (lumped) model in this study on lines 183 and 256.

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

We have shortened and clarified Section 3.6.

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.

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 track down the relevant details they might have missed or forgotten. This section referencing would help in a number of other places too.

We have improved our wording – e.g. "KGE interval metric-based calibration" is now annual peak flow calibration, and added references to specific sections throughout the paper.

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 have added new Figure 4 to show the precipitation frequency curves.

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.

We added section referencing here.

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

We have modified the discussion of the meteorology used during the event simulations in the re-written Section 3.6 and throughout the discussion in Section 5.

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

We have added to this.

L428: consider replacing “across” with “between”

Changed.

#### 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 Space 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>.

## Reply to Reviewer #3

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

### General comments:

This paper presents a comprehensive study of the uncertainties of a Flood Frequency analysis method based on stochastic simulation. The different sources of uncertainty are distributed between the structure of the model, the estimation of the parameters, the initial conditions and the inputs (rainfall). Overall, this paper is well written and presents significant and comprehensive scientific results. As the other reviewers I will recommend minor corrections, mainly to clarify some points about the tools used, described only by publications. Clarifications on the methodology would allow a better understanding of certain points detailed below.

We thank the reviewer for their positive assessment of the manuscript, and their helpful specific comments highlighting areas for further improvement. Please see our point-by-point replies below.

### Specific comments:

Section 3.1: As the other reviewers, I think that a diagram presenting the workflow would help in understanding the different steps and tools used.

Thank you for the comment, see our replies to reviewers #1 and #2 regarding improving clarification and the new Figure 3.

L134: specify if the modeling is lumped or distributed, knowing that the input data for the calibration looks distributed (in Newman et al, 2015). Clarify how you calibrate the hydrological models from an ensemble of historical meteorology (is the ensemble related to ground data interpolation uncertainties?). If the modeling is global, are the ensemble really very different sets (spatial mean)?

For this study the model is lumped. We have clarified this on lines 183 and 256.

L137: Clarify if "two event sequence possibilities" are in fact two periods/seasons? And which ones?

We have clarified Section 3.6 and reworded our description of our two different meteorology inputs for our two sets of event-based modeling.

L141: We ask ourselves the question of 11 parameters sets: how are they obtained? This is not from the sets of historical meteorology since there are 100 used. (ok, the answer is in line 328... but only 10 sets for the Altus basin)

We have clarified this on lines 214-216.

L220: why not take the whole distribution of initial conditions (IC) and only the strongest initial conditions. This does not allow to associate dry CI with heavy rainfall, and it reduces the impact of the uncertainties related to IC (for the current frequencies at least).

We have added discussion of this in Section 3.4.

L250: Rainfall events are in fact total rainfall generated from a regional probability law. Is it a limitation of the method, to simulate two-day events to generate extreme flood flows? As the time step of the hydrological modeling is daily, how are the rains distributed over the two days simulated? Moreover, can you explain how the problem of changing from a point rainfall to a basin rainfall is solved (how the areal-reduction factor is taken into account).

The precipitation frequency distribution is a two-day total only for Island Park, for Altus it is a single day event. For time splitting in Island Park, we randomly split the events across two days so that the FUSE models receive the precipitation over two days. These points are clarified in Figure 3 and Section 3.1.

We have added additional clarification regarding how we applied the areal-reduction factor (ARF) for these specific basins in Section 3.5.

L250-252 and 260-262: it should be better explained how you use historical precipitation events that are equal to the basin-average magnitudes sampled from the frequency curve, especially for extremes events.

We have clarified how precipitation events are generated from the frequency curves in Section 3.5 and clarified how we are using them in event simulations in Sections 3.6 and 3.7.

L309: How is the calibration performance assessed. There do not appear to be any validation procedures to verify this (as LOO procedure for instance). The graphs in Figure 4 presented calibration results. It is therefore difficult to judge the performance of the different models, which also depends on their robustness. It is better to show validation results.

We have used all data for calibration and have not performed an out-of-sample evaluation of the model as we are not concerned with true forecast performance and need the largest possible sample sizes for calibration. This is clarified on lines 295-297.

Specific comments:

Figure 1: as it is a theoretical explanation diagram, put theoretical curves (straight lines for example?) because we have the impression that it is a result.

Thank you for the figure suggestion. We have chosen to leave Figure 1 as is.

### References:

Bell, F.C. 1976. The areal reduction factors in rainfall-frequency estimation. Natural Environmental Research Council, Report 35. Institute of Hydrology, Wallingford, United Kingdom.

Clark, M. P., and A. Slater 2006: Probabilistic quantitative precipitation estimation in complex terrain. *J. Hydrometeor.*, 7, 3–22, doi:10.1175/JHM474.1.

Hansen, E.M., Schreiner, L.C., and Miller, J.F. (1982) Application of Probable Maximum Precipitation Estimates, United States East of the 105th Meridian. Hydrometeorological Report No. 52, National Weather Service, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, Silver Spring, MD, 168 p.

Hansen, E.M., Fenn, D.D., Schreiner, L.C., Stodt, R.W., and Miller, J.F. (1988) Probable Maximum Precipitation Estimates-United States between the Continental Divide and the 103rd Meridian. Hydrometeorological Report No. 55A, National Weather Service, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, Silver Spring, MD, 242 p.

Hansen, E.M., Fenn, D.D., Corrigan, P., Vogel, J.L., Schreiner, L.C. and Stodt, R.W. (1994) Probable Maximum Precipitation-Pacific Northwest States, Columbia River (including portions of Canada), Snake River and Pacific Coastal Drainages. Hydrometeorological Report No. 57, National Weather Service, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, Silver Spring, MD, 338 p.

Newman, A. J., Clark, M. P., Craig, J., Nijssen, B., Wood, A., Gutmann, E., Mizukami, N., Brekke, L. D., and Arnold, J. R. (2015). Gridded ensemble precipitation and temperature estimates for the contiguous United States. *Journal of Hydrometeorology*, 16(6), 2481-2500.

Schreiner, L.C. and Riedel, J.T. (1978) Probable Maximum Precipitation Estimates, United States East of the 105th Meridian. Hydrometeorological Report No. 51, National Weather Service, National Oceanic and Atmospheric Administration, U.S. Department of Commerce, Silver Spring, MD, 87 p.