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, we agree with the need for further clarification through development of at least one workflow diagram and additional text describing our methods and tools.

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 will clarify this around line 134. We combine all grid cells that intersect the basin polygon using their fractional areas to create the basin mean meteorology. We calibrate each ensemble member to arrive at 100 model parameter sets per basin.

The ensemble estimates interpolation and instrument uncertainty as described in Clark and Slater (2006) and Newman et al. (2015). We will add a brief description of this methodology and dataset to the revised paper.

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

We will clarify this phrase with the following: The two event sequences are 1) Randomly selected historical precipitation and temperature for the days following the specified event precipitation; and 2) the same historical temperature time series but with zero post-event precipitation.

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)

Thank you for catching this omission, we will add an explanation of the model parameter set selection earlier in the paper.

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).

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 as mentioned by Reviewer #2.

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, especially in more arid environments where conditions such as surface sealing and rock-mantled slopes may actually result in more severe flooding under intense short-duration thunderstorms. While the basins tested here did not consider those conditions, we agree that in the appropriate hydrologic regime, users should consider a wider distribution for such initial conditions which may increase the importance of initial conditions in flood response.

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

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. We will be sure to clarify this point. 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.

We will also add additional clarification regarding how we applied the areal-reduction factor (ARF) for these specific basins. In general, the process of converting a point precipitation-frequency curve (which is generally considered valid over 10 mi^2) to a basin-average precipitation-frequency curve can be accomplished through the application of an ARF. According to Reclamation (2012), the authors developed a linear relationship between point (x-axis) and basin-average (y-axis) storm totals from 12 different storms that impacted the Altus Dam region identified in HMR 51(Schreiner and Riedel, 1978), HMR 52 (Hansen et al., 1982), and HMR 55A (Hansen et al., 1988).

Although some published studies in the literature demonstrate that ARFs vary as a function of annual exceedance probability for common to extreme events (e.g., Bell, 1976), authors of the Island Park study applied a constant ARF for all AEPs. More specifically, Reclamation (2015) multiplied the point-specific precipitation-frequency curve by a constant ARF of 0.85, which they estimated using historical point and basin-average precipitation depths available in HMR 55A(Hansen et al. 1988) and HMR 57 (Hansen et. al.1994).

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.

Thank you for this comment. To clarify, the sampled and simulated precipitation events that are equal to the basin-average magnitude sampled from the frequency curve would be scaled by a factor of 1 (i.e., no change). Sampled and simulated precipitation events that are greater than the sampled basin-average magnitude sampled from the frequency curve would be scaled by a factor less than 1 (i.e., decreased). Finally, sampled and simulated precipitation events that are less than the sampled basin-average magnitude sampled from the frequency curve would be scaled by a factor greater than 1 (i.e., increased).

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.

Thank you for bringing this point up. We agree that out of sample validation results, particularly for a forecast application, are the appropriate way to evaluate model performance. In this case we had two factors to consider for how to calibrate and validate our models. First, we had limited data available in both catchments for calibration, and second we were not performing a true 'forecast' application in the sense of worrying about absolute values for floods. Thus, we decided to use all available data for calibration and select models using the in-sample performance. Going forward this would be an area for exploration to identify split sample calibration-validation strategies for floods to minimize model behavioral changes, such as changes in the model sensitivity explored here in the validation phase.

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 will modify Figure 1 by simplifying the curves to be much smoother, such as exponential or square root functions, but may not move to straight lines.

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.