We appreciate this review; we believe that it will enhance the apparent relevance of our manuscript via incorporating additional contexts and perspectives.

Responses to general comments:

-We agree that the manuscript misses an opportunity to compare numerical errors to other sources of error (e.g. observational, structural). Therefore, we will examine relevant literature to determine what kinds of errors in precipitation and discharge observations can be expected, especially with respect to changing precipitation amounts. An extra subsection will be added to the discussion which compares our results in numerical error to observational errors, further explaining the relevance of numerical error. This subsection will also likely include discussion of heteroschedastic error in hydrology (as suggested by Prof. Vrugt).

-We examine the daily time step because it is commonly used in hydrological models. However, as the reviewer indicates, simulations are often performed with a smaller time step. If this experiment was repeated with a smaller time step, we would expect to see the same trends in evolution of numerical error with respect to precipitation extremeness, simply with lower magnitudes of error. This is supported by the theory described in Section 2, where numerical error is described as proportional to the time step to some integer power. Therefore, we can expect lower magnitudes of error for all methods, and especially for those methods which are proportional to the time step to the second power, when a smaller time step is used, in the context of this experiment. We intend to describe this in a new paragraph in the discussion.

Responses to specific comments:

-Abstract: we agree that we examine conceptual models only and that this information should be included in the abstract.

-Introduction: as the results of our literature review tentatively indicate, the eight numerical methods used by FUSE are the most common methods used in conceptual hydrological models in general. In other words, of our surveyed modeling codes, none used a numerical method that is completely different from those that FUSE can offer. We will emphasize that the methods used by FUSE represent popular, current choices.

-Methods, P 9: We agree with the reviewer's comment that the relevant flood modeling period depends on catchment size. Therefore we will incorporate the suggested edit, indicating that 5, 10, or 20 days are relevant time scales for modeling floods in larger catchments.

-Methods, Fig. 3: This manuscript does not consider world records resulting from different climatologies, rather we scale precipitation data from individual world records in order to have a baseline of rainfall intensities that are possible at a given duration.

Therefore, no climate-specific conclusions about numerical error can be drawn, rather we simply use precipitation datasets that span a large number of different climates. However, we can indicate the location of a few notable events, including hurricane Katrina, in order to make the figure more concrete.

-Results, Fig. 9: We appreciate that this figure could be somewhat better explained, but we don't believe that using a stretched color palette would be appropriate, considering that all p-values for numerics groupings are below 1e-20, and p-values for groupings based on structure are typically greater than 0.05. Since the purpose of this figure is to demonstrate that numerical method is vastly more impactful than controlling numerical error than structure is, we will include the information that p-values for groupings based on numerical method choice are extremely low, and p-values for groupings based on structural choice are not, in the caption of this figure.

-Results, Fig. 10: We will attach this information as a supplement.

-Discussion, P 23: In a distributed model, numerical error is proportional to the chosen spatial discretization raised to some power, as well as to the time step raised to some power. This theoretically implies that numerical error could potentially increase in terms of maximum value in the context of distributed models. However, a fully distributed model (with distributed forcing data) could have its local extremes in space smoothed out given the choice of spatial discretization, while this option is not available for lumped conceptual models. More generally, it is possible for numerical errors due to spatial and temporal discretizations to interact. Therefore, we believe that this is an interesting question which is not straightforward to answer. We will indicate in the discussion that our results are not directly applicable to distributed models. How large the net numerical error is in case of a spatially explicit model, and what controls the magnitude of the errors involved, are interesting subjects for future research.

-Discussion, P 24: In the reference suggested by the reviewer, it is found that the impact of differing rainfall signals is to some extent diminished by the lumped rainfall-runoff model structure. This could partially explain why we find that trends in numerical error with respect to precipitation extremeness might not be sensitive to rainfall signal geometry, i.e. distributed models may be more sensitive to the impact of rainfall signal geometry. We can therefore easily include the suggested reference.

-Appendix: We will incorporate the suggested edit: TOPMODEL -> dynamic TOPMODEL.

Technical corrections

-P 8: We will make this change.

-P 12: The reviewer is correct; we will change the right hand label to "Precipitation".

Finally, we would like to thank the reviewer for the thoroughness of the review, the comment that the manuscript is mostly well written, and the compliment on the title.