General comments and recommendation

The manuscript submitted by La Follette et al. presents an interesting analysis of the potential numerical errors of lumped conceptual hydrological models emerging with increasing extremeness of precipitation. The authors conduct their analysis considering multiple model structures, covering the whole parameter space of these, and take also into account different initial conditions, generally justifying all their experimental set-up choices and critically questioning most of these, e.g. testing the robustness of their main findings or performing an additional check.

The first part of the manuscript is a nice introduction and/or recapitulative overview of some of the main numerical methods applied in hydrology, as well as many other fields dealing with the problem of solving ODEs. The authors highlight –as other hydrologists did before them- a problem that is usually suppressed in most hydrological studies, even those specifically dealing with different sources of uncertainties in a hydrological set up. I am actually quite convinced many users of hydrological model don't even know which numerical scheme is applied in the model they are using, and this information is often also not exactly the most straight-forward to find in the documentation of a hydrological model.

The manuscript as a whole is well structured, the methods are generally thoroughly described or supported by relevant sources and/or equations, the discussion is rather exhaustive, and actually covers some of the critical points that popped to my mind while reading the results. However, dealing the paper with a source of error and its relevance in hydrological modelling, specifically focusing on extreme precipitation, I miss putting this source of error into context, as compared with all the other sources of error a modeler would expect, and specifically the ones the modeler would expect when dealing with extreme precipitation and possibly with floods. Per se we are dealing with an ill-posed problem most of the time in hydrology, without even the need to go into other interrelated dynamics (see Di Baldassarre et al. 2016).. but anyway, I think it would be important to mention the uncertainty and errors the authors would expect to be related for instance to discharge measurements in a modelling application (see e.g. Westberger et al. 2020) – being this the most common variable used for calibrating and validating hydrological models-, or errors in the input data themselves, and in particular in precipitation, being this the main input the authors are focusing on. Errors in precipitation estimates can be considerable – if not deviating- and can occur in the original precipitation measurement itself, and further in the extrapolation over a larger area.
Another aspect which might be worth to spend a few (more) words on is the time scale: why do you only look at the daily time scale, and what would you expect to be difference by going down to the hourly time scale – which is the time scale at which e.g. flood forecasting is performed ? While the daily time scale is more relevant and common in climate change application, this might not be the case for present applications, which however also deal with the difficulties of making intense precipitation and hydrological models getting along, and for which a correct representation of the process is of primary importance.

I am not a native speaker, and as such not the best judge, but I dare to say the paper is well written, and it mostly reads fine, just sometimes it becomes a bit' cumbersome to read. I think this is the case mainly because of the close repetition of words in some paragraphs (e.g. P15 L 328 The numerical techniques were sorted into one of the three ranked groups based on their rank,…).
The manuscript features high-quality and interesting figures, accompanied by clear and good descriptions in the relative captions.

Finally, I would like to compliment the authors also for the nice and catchy title.

Because of these considerations, I think the manuscript requires only few corrections and some complementary or additional justifications/comments before it can be recommended for publication.

Please find my specific and technical comments here following.

Specific comments

- Abstract:
  - While the authors and probably many hydrologists are familiar with FUSE, it might be advisable and useful to specify you are looking at conceptual models only in the abstract of your paper (as Clark & Kavetski 2010 and Kavetski & Clark 2010 do in the abstract, but also actually in the title of their two papers).

- Introduction:
  - P3-L62: You choose to use the same numerical methods Clark & Kavetski (2010) and Kavetski & Clark (2010) applied: why didn't you consider expanding the numerical methods to include more of these?

- Methods:
  - P9-L218-219: Being a swiss hydrologist and mainly familiar with European rivers and hydrology I must say I do not agree with the assumption 5,10 and 20 days are the time frames usually used for modeling floods. This is really highly depending on the catchments' scale and other catchments' features resp. catchments' type. For mesoscale catchments 3-4 days are already relevant event durations (see also Froidevaux et al.2015), for some smaller alpine or quickly reacting catchments 1-2 days can also be enough. So here I would specify you consider event durations that are relevant for larger catchments, such as the Rhine for instance?
  - Fig.3: what is the meaning from the climatological perspective and where would Katrina be in this graph?

- Results:
  - Fig.9: It is just a suggestion, but wouldn't it be more interesting to show the figure using stretched colours resp. a colours' palette, using hatches of something else to indicate if 0.05 is exceeded? p=0.05 is commonly used in literature but it is still a subjective threshold.
  - Fig.10: it is OK to show the grouping of the reviewed codes as a mere visual, but it be more interesting for the readers to actually see a table with the exact numerical methods found resp. applied for the different codes? You could also attach it as supplement, if you consider it too much or not that informative, but as a matter of transparency and as you already did the job of extracting that piece of information, I think it would be a pity to not show it somewhere.

- Discussion:
  - P 23-L488-492: what would you expect to change if you used a (semi-)distributed model? What do you think is the role of numerical error by allowing water transport in space?
  - P24-L496-501: You might want to reinforce your findings with some literature? See e.g. Müller-Thomy & Sikorska-Senoner 2019

- Appendix:
  - P27-L609: maybe instead of TOPMODEL => (dynamic) TOPMODEL would be more correct? Metcalfe et al.2015 applied the dynamic TOPMODEL – even though I am aware that here the main difference between the two models is how water is distributed, what you are not applying here.

Technical corrections

- P8-L171: typos error=>Appendex should be Appendix
- P12 Fig.4: the y-axis on the right side shouldn't be precipitation?

References

Di Baldassarre, G., Brandimarte, L., Beven, K. (2016), The seventh facet of uncertainty: wrong assumptions, unknowns and surprises in the dynamics of human–water systems, *Hydrological Sciences Journal*, 61:9, 1748-1758, DOI: 10.1080/02626667.2015.1091460

Froidevaux, P., Schwanbeck, J., Weingartner, R., Chevalier, C., and Martius, O. (2015), Flood triggering in Switzerland: the role of daily to monthly preceding precipitation, *Hydrol. Earth Syst. Sci.*, 19, 3903–3924, https://doi.org/10.5194/hess-19-3903-2015

Müller-Thomy, H., Sikorska-Senoner, A. E. (2019), Does the complexity in temporal precipitation disaggregation matter for a lumped hydrological model? *Hydrological Sciences Journal*, 64:12, 1453-1471, DOI: 10.1080/02626667.2019.1638926

Westerberg, I.K., A.E. Sikorska-Senoner, D. Viviroli, D., M. Vis, and J. Seibert (2020), Hydrological model calibration with uncertain discharge data. *Hydrological Sciences Journal*, https://doi.org/10.1080/02626667.2020.1735638