Let me start my comment in an unusual way: Imagine yourself in the middle of the difficult process of choosing one or maximum two metrics for model calibration within a wider model intercomparison project, involving several modelling groups using different models. The project wants to renew national-scale climate change impact simulations because the existing ones show deficiencies with respect to low flow simulations. You have your own preference for calibration metrics ("I only use NSE because at least I know very well what it does and it has a clear link to formal parameter estimation literature"), but you have to discuss with colleagues who have very different preferences ("KGE-like metrics are the new standard", "we should use the new metric by xxx"). After several discussion rounds, you discover a paper that investigates exactly the question that you have been ruminating for several weeks: Is there a modern hydrological metric that outperforms all others across a typical range of signatures for typical hydrological regimes? What a chance that someone took the time to analyze this question! Many thanks to the authors of this study for having provided a detailed analysis of a specific modelling question that many of us are facing.

Thank you for sharing your perspective. We are very happy to hear that our study provides some useful input for your own discussions and research questions.

The design of the study emphasizes an aspect that is often omitted: the variability arising from different model structures, which is often lacking from available studies, in particular in times of large-ensemble hydrology. We have become so excited about using hundreds of catchments for our studies that we tend to omit model ensembles.

We agree that it is important to consider model structural uncertainty and we hope this is sufficiently clear from the introduction.

As pointed out by other comments in this discussion, the study does not explicitly address data or parameter uncertainty: every data point included in the analysis (a signature value obtained for a given model calibrated with a given objective function for a given catchment) could in principle be augmented by analyzing e.g. the 20 best calibrations or by analyzing different input data sets or by analyzing different data portions. But this would be a different study.

Thank you for making this point. As outlined in our other responses, we agree with the general point that accounting for all sources of uncertainty at the same time would lead to the most robust conclusions but argue that in a practical sense this is difficult. In this work we tried to find a useful balance between various components (models, catchments, objective functions) and simplified other parts. We will clarify this further in the discussion as outlined in our response to Prof Beven's comments.

The study does also not put a lot of emphasis on connecting the results to dominant hydrological processes or controls. In particular, it would be nice to see some discussion of what signatures are (by nature!) dominated by the input data. I could think of rising limb density, which can arguably only be well simulated if the rainfall input has good properties. An attempt is made in Table 5 where the signatures are put into categories, but the category “streamflow” is perhaps a bit too generic. How do these signatures depend on input data quality and the climate, hillslope-scale partionning functions or storage-release functions ? The ones that strongly depend on "functions" are probably more strongly influenced by objective function choice. Any thoughts on this could perhaps complement the discussion.

Thank you for bringing this up. We will reconsider the classification of Table 5 and add a section on influences on signature values to our discussion section.

Two detailed comments:

- The term « hydrological regime » is used at instances e.g. in 83 where I believe it should read the “hydrological response” or “hydrological behaviour” since the signatures do not only capture the regime
- Sentence at line 463 ("The claim that KGE resolves (..)") seems incomplete.

We agree and will fix this.