Reply to comments by referee 1 (Prof. Dr. Birkel) on the manuscript "Incorporating experimentally derived streamflow contributions into model parameterization to improve discharge prediction" by Hartmann et al. The referee's comments are provided in *ITALIC*, our response in regular style.

I had the pleasure to read the paper "Incorporating experimentally derived streamflow contributions into model parameterization to improve discharge prediction" published in HESSD under https://doi.org/10.5194/hess-2021-179 by Hartmann et al. The paper is generally well-written and the quest for reducing model uncertainty and increasing model realism is of interest to the HESS readership. I found the combination of a relatively simple conceptual rainfall-runoff model constrained by observation-based water source contribution estimates worthwhile and adequate to respond the research questions asked.

- ➔ Thanks a lot for your positive assessment

Having said that, I have some comments and suggestions that I put forward for the authors to consider:

- While I appreciate the tracer sampling effort and new data from a lesser known study site, the paper structure does not reflect this at all. The mixing model and some mentioning of the tracer data used appears in the study site description even with the results of the water source contributions. I strongly suggest to separately put these into the methods and later into results assuming this data and analysis was not previously published (no reference suggests this!). Furthermore, it would be instructive to actually see some of this data to get a notion of the space-time variability, e.g. in form of a bi-variate plot since the water source estimates are crucial for the analysis. I also wonder why no throughfall end-member was included in the mixing model and if the 20% margin for FHZ and FGW used to accept/reject models is based on an uncertainty estimate of the mixing model results (none are presented in Table 1)?

- ➔ Thanks for this suggestion (which was also given by referee #2). Following this advice, we will re-structure the manuscript to provide a complete section of the experimental work including fieldwork, sampling, hydrochemical analysis and derivation of streamflow contributions including respective figures. After that, we will present the revised modeling part. The new structure will be set as follows (or very similar):

1. Introduction (as it is now)
2. Description of the experimental work
2.1. Test catchment: location, topography, soils, vegetation; meteorology
2.2. Hydrometric measurements (discharge)
2.3. Sampling different waters: streamflow, hillslope soil water, riparian zone soil water
2.4. Analysis of water chemistry: parameters, methods
2.5. Deriving contributions to streamflow
2.5.1. Introduction to hydrograph separation
2.5.2. Selection and characterization of end members: show data; explain that the time-invariant behavior was not given; explain that we therefore delineated four periods for which we calculated mean end member concentrations
2.5.3. Tracer time series in streamflow over sampling period

2.5.4. Results of hydrograph separation
3. Modeling work - Methods
3.1. The model
3.2. Step-wise parameter estimation and quantification of information content of observations
3.3. Quantification of uncertainty of simulated model internal fluxes and discharge
4. Results
4.1. Step-wise parameter estimation and quantification of information content of observations
4.2. Quantification of uncertainty of simulated model internal fluxes and discharge
5. Discussion (structured as is right now)

- ➔ A throughfall end-member was not applied because on the daily field visits during the sampling period, overland flow was never observed. We therefore did not expect throughfall, i.e. new water, to contribute directly to streamflow. As per our observations, the generation of streamflow was a result of subsurface flow processes and, thus, we only considered subsurface water sources as end members.
- ➔ A margin for F_{HS} and F_{GW} was chosen because of previous studies that highlighted the uncertainties going along with hydrograph separation (e.g., Genereux, 1998;see P8, L11-14 in the original manuscript). But since particular estimates of uncertainties of previous hydrographs separations at other sites are difficult to transfer, we found the final value of 20% through a trial-and-error procedure within realistic ranges based on the uncertainties found in previous studies. We will clarify this and discuss our value of 20% in relation to the uncertainty of the separated streamflow components, which we will state in the expanded exponential section, in the revised manuscript.

- The fact that throughfall was sampled made me wonder about the importance of interception at this forested catchment and the effect it might have on the modelling since this is not included in the model structure. I also have a bit of an issue with the model structure itself and how the storage outflows are used to represent water sources: The model is essentially lumped with a vertical two-storage cascade fed by a soil reservoir that re-distributes the water for runoff generation. Now, I was thinking conceptually that all the hillslope outflow must feed into the riparian zone and from there runoff is generated that together with the groundwater flow constitutes streamflow (two end-members only). However, the latter would require a minimal semi-distributed model structure with two-storages in parallel (hillslopes draining into the riparian zone) and a third groundwater reservoir. In contrast, your HZ and GW is coming from the same source (storage V1). I would definitely appreciate some more explanations here.

- ➔ Interception will certainly play a role in this forested catchment. However, during the monsoon period, which is our main focus and our main sampling period, the events are characterized by high intensities (daily rainfall during the monsoon period on the days with rainfall mostly between 20-125 mm). Under these circumstances, the relevance of interception in relation to the rainfall reaching the ground surface is expected to be small. Therefore, we did not represent the process of interception in the model.
- ➔ The model structure implemented in this study is based on the conceptual understanding of the mutual dynamics of the hillslope, the riparian zone and the groundwater. Fast flow in the hillslope is triggered when groundwater levels increase exceed heights that correspond to an effective groundwater storage of FC (maximum storage in the hillslope [mm], see Table 2 in the original

manuscript). This model process represents conceptually the impact of rising groundwater levels on lateral transmissivities that allow fast-saturated flow down the hillslope towards the riparian zone. We will clarify this in the revised version of the manuscript.

- *My main concern however, is that the paper falls a little short in terms of the analysis related to many assumptions that are currently not sufficiently justified. For example, the choice of the KGE statistic that clearly influences the validation of mostly low flows in 2014, which is almost unfair as there is visibly no information content in these measurements. Without tracer measurements for 2014 it almost bags the question of why 2014 was included in the first place. The threshold of KGE>0.8 to accept models seems arbitrary. All three recession constants have the same initial parameter limits, but you would certainly accept a slower response of the groundwater reservoir outflow. Or you could think of fixing the groundwater recession constant based on a Master recession curve such as suggested in work by Hrachowitz et al. On the importance of the modeler's choices. As a matter of fact there is more literature on previous work (you could potentially cite) that attempted to reduce parameter uncertainty through constraining parameters with additional information such as tracers that did not necessarily included the need for more model complexity in terms of number of parameters. I would therefore suggest to try and test different statistics to see how they perform and apply the different criteria for model parameter selection also to the full 2 million parameter sets for a more comprehensive assessment of information content. Furthermore, throughout the paper you suggest quantitative assessments of information content, uncertainty in the context of a likelihood-weighted uncertainty estimate (GLUE), parameter identifiability and sensitivity, but this was not really done. Here, I would suggest to consistently use terminology and maybe provide some quantitative analysis such as e.g. a Shannon criterion for information content and/or a sensitivity metric such as Sobol and/or a measure of the width of the likelihood-weighted uncertainty bound used for prediction. With that you more comprehensively support your interpretations and allow the reader to really assess your statements in the discussion and conclusion.*

- ➔ Thank you for these very helpful remarks. Unfortunately, only two years of observations were available with 2013 the only year, for which streamflow contributions could be calculated. Hence, we decided to use the 2014 monsoon for evaluation only. Since it is much dryer that 2013, we consider the evaluation more rigorous (at least for the simulated discharge) and still consider the rather low values of the validation KGE_Q (0.02 ± 0.39) still acceptable allowing to conclude that including the streamflow contribution oin the calibration year provides more stable predictions coapred to using discharge for calibration only (which lead to a validation KGE of only -0.98 ± 1.54). We will clarify this in the revised version of the manuscript.
- ➔ Concerning our selection of KGE > 0.8, we relied on previous work (e.g, Hartmann et al., 2017). In order to evaluate the sensitivity of our results on this threshold, we will vary this threshold systematically, and use error measures different to KGE as also recommended by the referee. In the revised version, we will present and discuss the impact of this analysis.
- ➔ We will provide a more complete literature review on studies that attempted to reduce parameter uncertainty through constraining parameters with additional information such as tracers in the revised version of the manuscript.
- ➔ We will also double check and correct usage of terms like "information content", "sensitivity analysis", "uncertainty analysis" to provide a clear and consistent terminology throughout the paper.

- *Figure 5 is quite hard to interpret and I suggest to use a log-scale for streamflow visualization.*

  ➔ Ok, we will use log-scale in the revised paper.

- *There are some occasions in the paper where you wrote "be", but I think it should be "by".*

  ➔ We will double-check the manuscript for this type (and others).

For the above reasons, I would recommend major revisions before potential publication of this paper.

Sincerely,
Christian Birkel

References

Genereux, D.: Quantifying uncertainty in tracer-based hydrograph separations, Water Resour. Res., 34(4), 915–919, doi:10.1029/98WR00010, 1998.

Hartmann, A., Antonio Barberá, J. and Andreo, B.: On the value of water quality data and informative flow states in karst modelling, Hydrol. Earth Syst. Sci., 21(12), doi:10.5194/hess-21-5971-2017, 2017.