This manuscript presents a series of catchment scale numerical experiments intended to explore how the choice of tracer data and model type controls the mean transit times and TTDs. The authors applied the O18 and H3 tracers with relatively long time series. They mainly found that the use of O18 tracer and H3 tracer could somehow lead to similar MTT once the right model type was chosen (IM-SAS model). The science questions and approach would appeal to the HESS journal audience and make a nice contribution to understanding the effect of data, model types and in-stream concentrations.

However, I have a major concern that need to be addressed prior to publication:

1) I suggest to provide more context / justification / details about the calibration procedure – for example, how do you make sure your calibrated best-fits were not local best-fits but globe ones. The best-fit results of different implementations (such as IM-sas-L and IM-sas-D) were similar, but that does not mean the modeled results such as MTT was true. This generally requires an analysis of the potential uncertainty. While I understand a full uncertainty analysis may be unfeasible, the impact of operational choices done in the calibration exercise need to be better discussed.

2) I do agree with the authors that the H3 and δ18O tracers both are informative for the flow systems, what is needed is just a model good enough to resolve such information in a meaningful way. Especially for the catchments with strong seasonality. However, I am not sure if the model has to use combined date sets of hydrological and tracer as the author argued that "only the combined information using hydrological and tracer data and the consideration of transient flow conditions gives similar MTT, independent of the used tracer". I think the important thing is that the flow model can represent the reality in a good way, such that the tracer transport can be well reproduced. Using hydrological data in calibration may not a key control for that.

Other comments with line number:

Line 160: What are Ep and P?

Line 368: perhaps say that the storage component is just locally full-mixed and those local full mixtures do not lead to an overall fully mixed system…

Line 373: I don't think that to reduce computational time and computer memory requirements is good reason for using uniform sas functions rather than other shapes of sas function. I think the right way should be describing the model of reduced complexity (parameters) was already enough for your modelling targets.

Line 378: could you explain in more detail how was the tracer sampled from the passive and active volumes? Also random sampling from $S_{s,tot}$ ?

Line 393-395: maybe simply say the lumped implementation used a single HUR to represent the entire basin. Is that what you mean? In this case the precipitation zones were not used any more, right? Maybe clarify this.

Equation 14: what are $E_{mse,Q,n}$ and $E_{mse,tracer,m}$?

Line 473: it looks like that when using all the data, the lumped model (scenario 9) was even better than the distributed model (scenario 12) that has more parameters, does that mean the high model complexity is not essential for a better model performance in your case, could you clarify that

Line 508: Table 3?