Review of hess-2022-400 – *Stable water isotopes and tritium tracers tell the same tale: No evidence for underestimation of catchment transit times inferred by stable isotopes in SAS function models.*

The authors present an analysis of water ages obtained by different transit time (TT) models and two different tracers (i.e., stable water isotopes and tritium) in the Neckar basin (Germany). They test the common assertion that water ages above around five years cannot be well identified in TT models using stable water isotopes as model tracer, as opposed to tritium. They compare mean transit times (MTTs) as TT metric calculated via the sine-wave (SW) approach, convolution integral models (CO) and a process-based hydrological model using Storage Selection (SAS) functions. As MTT estimates substantially differ between the two tracers for SW and CO models, but are similar for the SAS model, the authors conclude that stable water isotopes are just as well fit for identifying old water ages as tritium and that differences in estimated water ages for the two tracers are an artifact of the SW and CO models.

General comments

The study fits the scope of HESS and makes a valuable contribution to the field of transit time modelling and tracer hydrology. Illustrating the capacity of stable water isotopes to quantify older water will open up new opportunities for TT modelling in catchments that are assumed to show comparably large MTTs. Hence, I support the general motivation and objectives of the study. However, I have some concerns related to the modelling approach and study basin used:

First, I am not sure whether a catchment (river basin) of 13,000 km² with at the same time limited availability of tracer data is the best choice for the study objectives. While individual controls on TTs remains largely elusive, it has been shown that TTs (or their metrics) vary widely depending on catchment characteristics such as elevation, topography or climate (e.g., Jasecko et al., 2016, Kumar et al., 2020). Modelling TTs in a river basin that shows a gradient of more than 800 mm yr^{-1} in annual precipitation, an elevation gradient of around 900 meters and varying land use types adds a lot of complexity that could have been avoided when using a much smaller and more homogeneous catchment. At the same time, the study relies on only one precipitation station for both stable water isotopes and tritium (within the basin) providing monthly composite samples. Hence, the tracer data are rather sparse both temporally and spatially, which adds another layer of uncertainty to the modelling. An alternative might be to compile data from previous TT modelling approaches that have been conducted in smaller catchments with more highly-resolved (space and/or time) stable water isotope and tritium data (e.g., Rodriguez et al., 2021 – reference already in manuscript).

Secondly, there is a remarkably great difference in model complexities between the individual TT modelling approaches. On the one hand, simple CO models with only one compartment, no temporal/seasonal variation and two pre-defined shape parameters for the TTs have been used, while on the other hand, the SAS model consists of three hydrological response units with multiple storage volumes each, has 11 calibration parameters and is also tested in a spatially distributed implementation. As the authors are clearly aware of, time-variant concepts of CO models (see Hrachowitz et al., 2010; and references cited therein) as well as multi-compartment models representing fast and slow flow routes have been used; using especially the latter is a common approach in CO modelling. Moreover, the SAS model with its comparably large number of parameters is calibrated simultaneously to discharge and at least one of the two tracers, while the CO models are calibrated to only one tracer. I am thus wondering to what extent results from these TT models can be compared at all. I understand that the objective of this paper is not to dismiss a specific model type, but rather to analyse the flexibility of stable water isotopes as TT model tracers. However, this requires to use model setups and data similar to those used in the papers that have demonstrated the truncation of TT distributions by calibration to stable water isotopes. To address this concern, one could think of (i) focussing on a smaller (or even headwater) catchment with preferably daily tracer data, (ii) using established CO models such as the more complex ones in Stewart et al. (2010), and (iii) using measured and modelled P, ET, storage and Q data as input for SAS modelling (potentially with non-random sampling) with one or a maximum of two SAS function compartments, as commonly done in more recent SAS modelling studies (e.g., Benettin et al., 2017; Harman, 2015; Nguyen et al., 2021).

Thirdly, the fact that spatial aggregation introduces bias in CO model-based MTTs, as stated also by the authors, raises the question to what extent comparison of MTT estimates is meaningful. I understand that the authors would like to test the validity of stable water isotopes in TT modelling particularly of *older* water ages, and that MTT has been a metric commonly reported for CO models. Nonetheless, according to Kirchner (2016 – reference already in manuscript), sine-wave fitting to seasonal isotope data does give robust estimates of the young water fraction F_{yw}. Hence, it might be more meaningful to compare F_{yw} estimates by the different TT model approaches, or, even better, to add this as further TT metric in the comparison.

Finally, I would highly appreciate if the authors could increase traceability of their results and provide the underlying tracer data as well as model codes. Traceability is one of the main criteria for HESS nowadays and given that the authors address such a fundamental claim in tracer hydrology and TT modelling, I find it necessary for the entire TT community to benefit from this study not only via the paper, but also in terms of data and code accessibility.

Overall, I would be glad to see how the authors address the above mentioned methodological issues. Please see below for some specific comments and suggestions.

Specific comments

- Lines 35—37: if this refers to the findings by Kirchner (2016), one could be more precise by specifying that the *MTT* (as commonly reported metric) derived from CO models is affected by spatial aggregation errors.
- Line 59: in what sense is there more coherence?
- Line 70: does Cl^- have a clear seasonal cycle? I assume both weathering and anthropogenic effects (e.g., application of road salt) govern its concentrations. Another possible distinction would be radioactive vs. conservative tracers.
- Lines 80—98: the focus on the amplitude ratio for the “traditional” TT approaches is fine for simple one-compartment gamma (and thus also exponential) models, but is this also relevant for multiple-compartment CO models and other pre-defined TT shapes such as the dispersion model? This suggests that CO models are exclusively based on the amplitude ratio and shift in seasonal isotope ratios.
- Lines 84—85: “practically” and “feasibly” twice?
- Lines 97: to what extent could a spatial aggregation bias also affect spatially lumped (one-compartment) SAS models?

- Lines 197: you used the CORINE dataset from 2018. To what extent has land use remained stable since 2001?
- Line 374: we do not necessarily see passive storage volumes in the most recent SAS model studies.
- Lines 398—414: I am wondering to what extent we can trust the spatially distributed implementation, given that there is only one calibration gauge at the outlet of the entire catchment. This also relates to my general comment about the considerable size and few data for the study basin.
- Line 411: could you specify what the distributed moisture accounting approach is?
- Lines 420—421: why have the authors not applied a multi-objective calibration to the CO models?
- Line 424: this is interesting but I think, as stated in my general comments, that TTs should be obtained from a SAS model with storage, input and output fluxes defined a priori (as if they were "real" data), rather than computing TTs from simultaneous calibration against flow and tracers. I think that this would be a more straightforward methodology given the scope of TT modelling and tracers. As presented here, we do not know to what extent simulated TTs are affected by equifinality in the hydrological model parameters.
- Lines 438—439: would normalisation of the errors help?
- Lines 553—555: not a complete sentence
- Line 571: not only, but also…?
- Lines 577—578: I think you could easily implement the multi-objective calibration for the CO models as well.
- Lines 619—620: so here one could at least test how time-variant/seasonal CO models perform
- Lines 642—644: could this not be an indication of the fact that there are too many degrees of freedom and the model always succeeds to fit the tracer data, regardless of whether it is spatially lumped or semi-distributed?
- Lines 656—657: see, e.g., Nguyen et al. (2022) who found substantial differences in SAS-based transport models between spatially lumped and semi-distributed setup.

References

Benettin, P., Soulsby, C., Birkel, C., Tetzlaff, D., Botter, G., and Rinaldo, A. (2017), Using SAS functions and high-resolution isotope data to unravel travel time distributions in headwater catchments, Water Resour. Res., 53, 1864– 1878, doi:10.1002/2016WR020117.

Harman, C. J. (2015), Time-variable transit time distributions and transport: Theory and application to storage-dependent transport of chloride in a watershed, Water Resour. Res., 51, 1– 30, doi:10.1002/2014WR015707.

Hrachowitz, M., Soulsby, C., Tetzlaff, D., Malcolm, I. A., and Schoups, G. (2010), Gamma distribution models for transit time estimation in catchments: Physical interpretation of parameters and implications for time-variant transit time assessment, Water Resour. Res., 46, W10536, doi:10.1029/2010WR009148.

Jasechko, S., Kirchner, J. W., Welker, J. M., & McDonnell, J. J. (2016). Substantial proportion of global streamflow less than three months old. Nature Geoscience, 9(2), 126–129. https://doi.org/10.1038/ngeo2636

Kumar, R., Heße, F., Rao, P.S.C. et al. Strong hydroclimatic controls on vulnerability to subsurface nitrate contamination across Europe. Nat Commun 11, 6302 (2020). https://doi.org/10.1038/s41467-020-19955-8

Nguyen, T. V., Kumar, R., Musolff, A., Lutz, S. R., Sarrazin, F., Attinger, S., & Fleckenstein, J. H. (2022). Disparate seasonal nitrate export from nested heterogeneous subcatchments revealed with StorAge Selection functions. Water Resources Research, 58, e2021WR030797. https://doi.org/10.1029/2021WR030797

Nguyen, T. V., Kumar, R., Lutz, S. R., Musolff, A., Yang, J., & Fleckenstein, J. H. (2021). Modeling nitrate export from a mesoscale catchment using storage selection functions. Water Resources Research, 57, e2020WR028490. https://doi.org/10.1029/2020WR028490