

## ***Interactive comment on “Testing the truncation of travel times with StorAge Selection functions using deuterium and tritium as tracers” by Nicolas B. Rodriguez et al.***

**Francesc Gallart (Referee)**

francesc.gallart@idaea.csic.es

Received and published: 29 November 2019

The authors propose a very interesting piece of work that may shed light on the future joint use of deuterium and tritium isotopes on water age studies. The volume of the original analytical information is outstanding, the text is a little verbose but clear, the graphs are explicative and the rationale and methods are well explained although a relevant part is not described as it is under review in another journal.

Nevertheless, there are a few methodological issues that should be fixed or justified before the manuscript is acceptable for publication.

C1

General comments.

The procedure used by the authors to test the “truncation” hypothesis “that streamflow TTDs calculated using only deuterium (2H) or only tritium (3H) are different” does not follow the established methods for hypothesis testing. As a rule, for rejecting a null hypothesis it is necessary to verify that its probability is lower than a prefixed assumable error risk, typically  $p < 0.01$ . High uncertainty of the results is not sufficient for rejecting a null hypothesis.

The authors found that “differences between the various statistics of the TTDs were smaller than the uncertainties of the calculations when comparing the results obtained with 2H alone and with 3H alone”. But the authors also state that “even though the uncertainties are sufficient to account for the differences between 2H - and 3H -derived age and storage measures, it is worth noticing that 3H systematically gave higher estimates”. Therefore, even if the authors did not estimate the probability of the null (truncation) hypothesis, this last sentence suggests that its probability was not sufficiently low for rejecting it, so the result of this work is that the authors cannot reject the “truncation” hypothesis.

Furthermore, this hypothesis testing exercise had other issues. Indeed, although the authors “treated 2H and 3H equally by calculating TTDs using a coherent mathematical framework for both tracers (i.e. same method and same functional form of TTD)” they did not treat these isotopes with similar sampling strategies. Indeed, nearly 30 stream samples of 3H collected during highly varying flow conditions cannot be compared with the 1088 stream samples of 2H collected every 15 hours on average, even if the period was the same. Among the diverse causes that can explain the modest differences found between the results obtained with deuterium and tritium, the potential role of the different sampling strategy must be taken into account (differences respect to the sample number and flow representativeness, as also suggested by the authors in the discussion). The test performed by the authors compares the results and potentials of both isotopes when used under the current state of the art but not their own potentials.

C2

A rigorous test for comparing the own potentials of both isotopes would need to use an equal number of samples taken simultaneously for both.

This leads to another relevant issue on the sample treatment. The authors, as commonly made, weighted the isotopic signal of rainfall waters with the respective rainfall depths. But nothing is said on the weighting of stream samples, as regrettably also recurrently made. So the reader has to assume that the raw (unweighted) isotopic signals of stream samples were used for constraining the model.

My point is that this approach, if actually used, will provide a set of model parameters adequate to describe the isotopic signal of the samples as they are in the record, but not to simulate the isotopic mass balances, i.e. the main rationale of the model. If the isotopic mass balances are sought, it is necessary to weight the isotopic signal of every sample with the associated flow (time span X discharge). Furthermore, looking to Figure 4, it seems that the most highly scattered 2H samples were taken during low flows, so it could happen that the, really low, efficiency of the model would improve by flow-weighting the stream samples.

Another associated question is the representativeness of the stream samples of the diverse flow ranges in the catchment. In the discussion, the authors sensibly wonder if “tritium ... may still be biased towards hydrological recessions” and “how many measurements are enough and when to sample isotopes for maximum information gain on water ages”. If the stream samples must represent the mass flow of water and tracers and a detailed flow record is available, it is possible to compare the distribution functions of both flow records (only measured versus measured and sampled) for assessing the degree of representativeness of the sampling designs. This kind of analysis should be customary in all catchment environmental tracing studies, particularly for small catchments where the flow duration curve is usually highly skewed.

Detailed comments:

Lines 12-13: The truncation (null) hypothesis cannot be rejected from the work results.

C3

Line 122: “phyllade” is a French geological term. The closest English term, as far as I know, is “phyllite”

Line 330: ... This is not the case for d3H...

Line 380: The truncation (null) hypothesis cannot be rejected from the work results.

The model calibration method that consists of using a range of parameter sets instead of an ‘optimal’ parameter set was developed by Beven & Binley (1992). I suggest to cite this work also because it, as far as I know, was the first using the Shannon entropy for analysing the value of additional data in the calibration of a model.

Beven, K., & Binley, A. (1992). The future of distributed models: model calibration and uncertainty prediction. *Hydrological processes*, 6(3), 279-298.

---

Interactive comment on *Hydrol. Earth Syst. Sci. Discuss.*, <https://doi.org/10.5194/hess-2019-501>, 2019.

C4