Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-501-RC2, 2019

Hydrology and Earth System Sciences
Discussions
Open Access

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.

Anonymous Referee #2

Received and published: 10 December 2019

1 General comments

The manuscript tested the hypothesis that ^{3}H tracer provides information over longer transit times than ^{2}H. The authors calibrated the StorAge Selection (SAS) function model for each tracer and examined information gain using the posterior distributions of the model parameters. They rejected the hypothesis based on their results. Nevertheless, they concluded that ^{3}H tracer is more informative and cost-efficient compared to ^{2}H.

e topic is timely and very interesting. However, the manuscript needs substantial revision. First, I do not think that the results presented in this study support most of their conclusions. Their SAS function-based model performed poorly even with 12 parameters, and it is not clear how much we can learn from the poorly performing and not well-constrained model. Second, I have several issues with their analysis and the hypothesis test. These points are described in more detail in what follows.

2 Major comments

2.1 Not enough dataset and the poorly performing complex model

The model has an unusually large number of parameters (12 parameters; e.g., Line 249) compared to the previous SAS function-based modeling studies. I believe that the authors illustrated the need for more parameters well in their previous study, which is now published in WRR. However, the model does not perform well even with the 12 parameters (with the maximum NSE 0.24 for ^{2}H), and I am not sure what we can learn from the poorly-performed model. The large number of parameters also causes several issues described below. Also, the dataset is very limited, and it is not clear if the limited number of samples and the limited sampling period support their conclusions.

First, it is not clear if the ^{3}H dataset is enough. The number of samples is too limited to constraint 12 parameters. I can easily guess that the parameters are not well-constrained. Thus, it is obscure how much information we can extract from the time series, the posterior distributions of those parameters, the TTDs, and the SAS functions, which were used to test the hypothesis and examine if those tracers contain non-redundant information to each other. For example, the authors stated that “stable and radioactive isotopes have information in common about young water” in Lines 472-475. However, the argument cannot be supported by those 24 samples. Furthermore, how much information we can extract from the 2-years of ^{2}H dataset? Can we talk about transit time longer than 2 years (at the maximum) based on the model results?

Second, I think that their Latin hypercube sampling (Line 262) suffers the curse of dimensionality. They sampled 12,096 parameter sets from the 12-dimensional parameter space. It can be easily guessed that those samples are very sparsely distributed in the 12-dimensional parameter space (i.e., $12^4 > 12,096$), and the sparse sampling can potentially limit their ability to construct well-constrained posterior distributions of those parameters.

Lastly, the poor performance of the model leads me to think that maybe their model structure is not adequate, and any discussion based on the model results should be conducted more carefully. It is clear that the model fails to reproduce short time-scale dynamics. Figure 4 shows that their ^{2}H-based model cannot capture the observed large fluctuation. It seems that the large fluctuation is, in part, due to the high correlation between $C_{p,2}$ and $C_{Q,2}$ especially when the system is dry, and It implies that short time-scale dynamics are not captured by the model (as they mentioned in Lines 512-513). The fluctuation seems much more pronounced in the ^{2}H time series. Thus, if we have a better model that captures the short time-scale dynamics, it may contradict the authors' argument in Line 472: "stable and radioactive isotopes have information in common about young water."

2.2 Analysis and Hypothesis test

The use of the Kullback-Leibler Divergence D_{KL} in the hypothesis test seems inappropriate. Throughout the manuscript, the authors stated that using both tracers together is valuable since $D_{KL} > 0$ (e.g., in Lines 435-436 and Lines 468-470). However, the criterion $D_{KL} > 0$ cannot determine whether the criterion is met because multiple tracers are used or because there is just any additional information. For example, D_{KL} between the model constrained by, let's say, 100 ^{2}H data and the model constrained by the rest of the ^{2}H data will be greater than zero.

Moreover, different performance measures were used for their models (Lines 265-270), and it makes the use of D_{KL} even more inappropriate. The authors used the NSE for the ^{2}H-based model and used the MAE for the ^{3}H-based model. Thus, the difference between the posterior distributions estimated by those behavioral models can be, in part, explained by the choice of performance measure. For example, if the authors estimate the posterior distributions using the ^{2}H dataset based on the MAE, the posterior distributions would differ from those estimated based on the NSE. Then, D_{KL} would be greater than zero. Thus, it is not hard to follow their argument that using both tracers together is valuable (e.g., in Lines 331-333, Lines 435-436, Lines 478-470, and Lines 580-581).

Furthermore, I disagree with their cost analysis (in Lines 445-451), which led them to conclude that ^{3}H tracer is more cost-effective (e.g., Line 17). As described in Lines 462-463, “The amount of information learned from the isotopic data probably scales non-linearly and probably reaches a plateau as the number of observation points grows.” However, they assumed “linearity” in their cost analysis. Thus, the analysis is not valid.

Lastly, it seems that the ET SAS functions are very important in this study but rarely explained. One of its parameters, μ_{ET} is the most valuable parameter in terms of the information gain in this study (see Table 2). However, no explanation is provided why it is the most valuable and how it affects their interpretation of the results. For example, Figure 5 is one of the most important figures that clearly illustrates the difference between the ^{2}H-based model and the ^{3}H-based model. The simulated ^{3}H concentration using the ^{2}H-based model, in general, is higher than that simulated using the ^{3}H-based model. It means that tracer mass partitioned into discharge is smaller in the ^{3}H-based model during the period. Since there is no explanation on the difference, I had to guess that either more ^{3}H tracer mass is stored in the system in the ^{3}H-based model or more ^{3}H tracer was partitioned into evapotranspiration in the model. Overall, it seems that the partitioning is one of the most important differences between the two models. Thus, the partitioning should be explained in more detail.

3 Minor comments

1. The posterior distribution of the parameters should be presented in this manuscript. It is hard to grasp most of the authors' arguments without those distributions.

2. Line 375: Typo in “[0,∞[“

3. Line 224: It is stated that $\lambda_1(t)$ is the smallest weight. However, it is not clear how that was constrained in the model calibration.

4. Lines 236-237: S_{ref} is chosen not calibrated, so probably introducing the chosen value here would be better, rather than introducing it in the next section, 2.6 Model calibration.

5. The initial condition for the SAS function model is not described. If there was a spin-up for the SAS function model (like the storage estimation), what tracer time series were used?

6. Lines 404-405: How this comparison between 2016 finding and 2017 finding helps readers to understand the higher age estimated using the ^{3}H-based model?

7. Lines 437-439: Those parameters are not independent. Thus, those were not independently constrained.

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