

Hydrol. Earth Syst. Sci. Discuss., referee comment RC3
<https://doi.org/10.5194/hess-2022-149-RC3>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on hess-2022-149

Anonymous Referee #3

Referee comment on "Exploring tracer information in a small stream to improve parameter identifiability and enhance the process interpretation in transient storage models" by Enrico Bonanno et al., Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2022-149-RC3>, 2022

General

The manuscript highlights an important aspect of inverse modeling based on stream tracer test – the uncertainty and identifiability of the estimated model parameters and the connected metrics used to assess solute transport processes in streams. Although I think that the text is generally well written, the structure of the paper (e.g. order of sections) could be streamlined and improved for clarity. As it is right now, I found myself going back and forth between the different sections in order to understand the iterative and rather complex structure. Overall, I appreciate the approach and the general topic, but I believe the manuscript would benefit from addressing a number of issues highlighted below.

Major comments:

- I think that the terms "identifiability" and "sensitivity" needs to be defined early on in the introduction so that there is no doubt what is meant by these key terms in the context of the study. As it is now the words are used already from the start of the paper, but the definitions are "hidden" in Appendix A. I suggest moving the definitions (lines 902-905) and placing them in the main paper where they are first introduced.
- There are several threshold values used to select behavioral parameter sets and subsequently to define identifiability. However, these thresholds are not sufficiently motivated and discussed, which makes it difficult for the reader to assess the results. This holds for the definitions of both the global identifiability (e.g. the top 0.1-10% of the models when assessing the CDF deviation from the 1:1 line having, the grouping

based on the K-S results) and the dynamic identifiability (e.g. information content > 0.66). This leads to questions about the subjectivity of these thresholds when assessing the identifiability and how sensitive the overall assessment of the model parameters are to the chosen thresholds (e.g. in Figure 8).

- The objective function used to assess the model performance, i.e. the RMSE, is based on the difference between observed and modelled BTC over a given time scale/number of observations. What are the effects on the RMSE for concentration values that may differ more than an order of magnitude over assessed time window, i.e. for high values (peak of the BTC corresponding to 50 mg/l) compared to the low values (tail of the BTC corresponding to <1 mg/l)? How does this affect (i) the global identifiability analysis and (ii) the dynamic identifiability? Previously alternative objective functions have been suggested, including RMSE with a logarithmic (e.g. Ward et al., 2018) or mixed scale (e.g. Bottacin-Busolin et al., 2011; Riml et al., 2013), to account for the different magnitudes of the concentrations across the BTC. I suggest that the authors assess this and discuss the implications. Moreover, I miss a visual comparison of the observed and simulated BTC as a complement to the presented RMSE values, preferable using log-transformed concentrations to highlight how well the model captures the tail of BTC that is argued to be of importance for transient storage processes (e.g. lines 55-57).
- The fact that using erroneous model parameter estimates (obtained either from the literature, from a simplified model (ADE) or from a Monte-Carlo simulation with too wide parameter ranges and/or not sufficient iterations) leads to uncertainty/errors when estimating the transport metrics (Eqn 5-8) is rather intuitive. Firstly, I find (the rather long) discussion in Section 4.1 as well as the conclusion (line 24-26) and abstract (lines 21-26) about misinterpretations/uncertainty when comparing the different models "unfair". The conditions for the OTIS-MCAT simulations seem to be equivalent to the first iteration of the proposed methodology. Thus the conditions when OTIS-MCAT was used differ substantially from the 3-4 iterations with a successive refinement of the parameter ranges of the proposed methodology. I understand that the authors would like to make a point and compare the results against an existing model framework, but I think that manuscript would benefit from significantly downplaying the role of OTIS-MCAT. I would prefer a stronger focus on how the refinement of parameter ranges using the existing model framework resulted in a reduced uncertainty/increased identifiability of the model parameters. Secondly, although the authors successfully reduced the uncertainty of the model parameters by an iterative and smart sampling of the parameter space, it was surprising to see that the results from the OTIS-P outperformed (2 out of 3 experiments) the proposed methodology when using the objective function preferred by the authors (RMSE, table 2). This is something that is not sufficiently discussed in the paper and, in my view, opens for questions when the iterative sampling procedure is needed. Could e.g. the DYNIA approach be combined with OTIS-P using a given confidence interval as input for the parameter ranges to assess the identifiability in parameter estimates from OTIS-P? I guess that similarity in performance between parameters obtained by OTIS-P and the proposed sampling procedure might connect to the objective function used to evaluate the performance (see major comment #3), the limited number of experiments (in a single reach) and how initial values (and the possibility of finding a local minimum when optimizing) were defined in the OTIS-P model.
- I believe that the paper lacks a thorough discussion regarding different model representations of transient storage. Eq. 3 assumes an exponential residence time distribution (RTD) in the transient storage zone as originally defined in the TSM model (e.g. Bencala and Walters 1983). Subsequently other type of exchange models and RTDs in the transient storage zone have been introduced (e.g. Wörman et al., 2002; Haggerty et al., 2002). I see a great advantage of the proposed model framework – compared to the OTIS-MCAT and OTIS-P – to explore alternative model formulations (including multiple transient storage zones) and alternative RTDs. This flexibility, when discussed properly, could levitate the readers understanding of the usefulness of the

model framework.

Detailed comments:

Line 26-29: This is unclear, what is meant by "clear potential"?

Line 53: "The numerous contradictory outcomes", in what context? Please clarify

Line 109-111: Unclear. "to keep constant a rather identifiable parameter". Please rephrase.

Eq. 3: I believe that C_S should be replaced by C_{TS} in the bottom equation

Line 216-: I miss information of the width of the window in the DYNIA.

Line 233-234 "The best 1% of the results were used to define its parameter space in the successive TSM iteration". Not in agreement with Figure 1 that says "New parameter range defined from the top 10% of the results". Please revise.

Line 245-246: v_{peak} is not clearly defined. Is this the time from injection to the BTC peak divided by the stream length (i.e. 55 m)? Please clarify.

Line 249: How was the set up of the OTIS-P simulations in terms of initial parameter values? Does "multiple OTIS-P iterations" mean that several initial conditions were tested to reduce the risk of ending up with in local minimum in the optimization? Please clarify.

Line 255-258: What parameter ranges were used and how many iterations were performed with the OTIS-MCAT? It is difficult to compare the results if the simulation conditions are not provided.

Line 293-295: From Figure 3, it is unclear how $A_{TS} < 5.356 \text{ m}^2$ has information content > 0.66 , Figure 3 i,j). Previously (line 223-224) it is stated that the information content is expressed as one minus the width of the 90 confidence interval, which I assume uses the entire parameter distribution. Please clarify why there is no lower bound on the confidence interval. Moreover, although I realize that this is the first iteration, to have a transient storage area several orders of magnitude larger than the cross-sectional area of the stream makes little sense.

Figure 3. I suggest to show the y-axis of the alpha plot (Figure 3 g) using a log scale, due to the small values.

Line 312-314: How much of this result can be derived to the used parameter ranges and the number of iterations? If the MC analysis would been set up differently (smaller parameter ranges, larger number of iterations), how would the result differ?

Line 338: "orange boxplots" is labeled "red boxplots" in Figure 7. Please revise.

References:

Bencala, K. E., and Walters, R. A. (1983), Simulation of solute transport in a mountain pool-and-riffle stream: A transient storage model, *Water Resour. Res.*, 19(3), 718– 724,

Bottacin-Busolin, A., Singer, G., Zaramella, M., Battin, T.J., and Andrea Marion, A., (2009) Effects of Streambed Morphology and Biofilm Growth on the Transient Storage of Solutes *Environmental Science & Technology* 43 (19), 7337-7342

Haggerty, R., Wondzell, S. M., and Johnson, M. A., (2002) Power-law residence time distribution in the hyporheic zone of a 2nd-order mountain stream, *Geophys. Res. Lett.*, 29(13),

Riml, J., Wörman, A., Kunkel, U., Radke, M. (2013) Evaluating the fate of six common pharmaceuticals using a reactive transport model: Insights from a stream tracer test, *Science of The Total Environment*, 458-460, 344-354

Ward, A. S., Kelleher, C. A., Mason, S. J. K., Wagener, T., McIntyre, N., McGlynn, B., et al. (2017). A software tool to assess uncertainty in transient-storage model parameters using Monte Carlo simulations. *Freshwater Science*, 36(1), 195– 217

Wörman, A., Packman, A. I., Johansson, H., & Jonsson, K. (2002). Effect of flow-induced exchange in hyporheic zones on longitudinal transport of solutes in streams and rivers. *Water Resources Research*, 38(6).