General comments

R2: "The text content is generally well written but the theme's complexity and the structuring makes honestly the reading rather difficult, which can force the reader going back and forth the chapters to follow a storyline. I found very logical how the text was structured in the 'Study site and Methods' section, and I would really try to stick to this same structure when presenting 'Results', in the 'Discussion' section and even in the 'Conclusion' paragraph. My personal taste would be:

- a.1) ADE parameters
- a.2) TSM parameters (i.e. hydrologic exchange parameters)
  - a.2.1. Identifiability of TSM parameters when v = vvariable
  - a.2.2. Identifiability of TSM parameters when v = vpeak
- a.3) Model iterations
  - a.3.1. TSM first iterations
  - a.3.2. TSM last iterations
  - a.3.3. DYNIA
  - a.3.4. Comparison with inverse modeling results (OTIS-P)
  - a.3.5. Comparison with random sampling approaches (OTIS-MCAT)
- a.4) Metrics and hydrologic interpretations of model results

According to this numbering, the 'Results' section starts with the a.3.1 jumps to a.3.3, then back to a.3.2, etc. The 'Discussion' section starts with a.4, then goes back to a.3.1 and a.3.2, follows with a.3.4, etc. This can make a difference between a quick and effective reading, and a tedious reading."

Authors: We thank the Reviewer 2 (R2) for the positive and constructive manuscript feedback. We understand the need for a revised manuscript structure. We will adopt the suggested structure for the revised results and discussion sections with a slight modification:

- a.1) ADE parameters
- a.2) TSM parameters
  - a.2.1 Identifiability of TSM parameters when v = vvariable
    - a.2.1.1 TSM first iterations
    - a.2.1.2 TSM last iterations
  - a.2.2 Identifiability of TSM parameters when v = vpeak
    - a.2.2.1 TSM first iterations
    - a.2.2.2 TSM last iterations
- a.3) DYNIA
  - a.3.1 DYNIA (first + last iteration - v = vvariable)
  - a.3.2 DYNIA (first + last iteration - v = vpeak)
- a.4) Comparison
  - a.4.1 Comparison with inverse modeling results (OTIS-P)
  - a.4.2 Comparison with random sampling approaches (OTIS-MCAT)
- a.5) Metrics and hydrologic interpretations of model results

We slightly modified the structure suggested by R2, because:

- For achieving identifiability of TSM parameters we need to address model iterations;
- This modified structure is also in agreement with the review of R1, who required the manuscript structure to follow the order of the research questions.

R2: "I would like to draw the attention about the use of the term 'parameters' in the manuscript. I personally have no experience with the identifiability of parameters in a TS model and it can bias my understanding of the term. Thus, this might be probably my wrong perception.
As a hydrologist, my understanding of parameters involved in the transport of solutes in a stream/river can be stream discharge, components of flow velocity, flow turbulence, grain size of streambed

sediments, groundwater-stream (GW-SW) water exchange fluxes, stream channel topography, existence/absence of riparian vegetation, etc. In the manuscript, these parameters are treated as such, but then it introduces the term TSM parameters and parameter sets. In this manuscript's TSM model, each iteration simulates 115,000 parameter sets. An unexperienced reader would tend to think that > 100,000 parameters as the ones I mentioned above are involved in the transport of solutes in the stream. This should be obviously clarified."
Authors: We agree with R2 that the term "parameters" and "parameter sets" have to be better clarified in the manuscript. In the main text we defined A, v, and D as "advection-dispersion parameters", A_TS and α as "transient storage parameters". The "advection-dispersion parameters" together with the "transient storage parameters" are referred to as "TSM parameters" (cf. lines 183-185). We agree that the similarity between "transient storage parameters" and "TSM parameters" can be confusing, thus we will modify "TSM parameters" as "model parameters" in the revised manuscript version. We will define the term "parameter set" as "a single combination of A, v, D, A_TS, and α from random-sampling" in the revised method section.

R2: "The manuscript's title states 'to reduce the uncertainty' but I have missed more uncertainty analysis throughout the text."
Authors: It is true that we have not used the term "uncertainty" in the manuscript, and we adopted in the main text the term "identifiability" or "non-identifiability" since we addressed parameter identifiability. Thus, we apologize for the lack of consistency in terms of used terminology. We will uniform and double-check the terminology in the revised manuscript. Namely, we will always use "identifiability" and "non-identifiability" and the title will be modified in: "Exploring tracer information in a small stream to improve the identifiability and enhance the process interpretation of transient storage models". We addressed parameter identifiability via global identifiability analysis based on Generalized Likelihood Uncertainty Estimation (Beven & Binley, 1992). This analysis includes parameter vs likelihood plots, parameter distribution plots, regional sensitivity analysis plots, and identifiability plots (Appendix A). Identifiability of a parameter is generally investigated via visual inspection of regional sensitivity analysis plots (e.g. Kelleher et al., 2019) or visual investigation of parameter distribution plots and parameter vs likelihood plots (e.g. Ward et al., 2017). We here addressed identifiability also via the two-sample Kolmogorov-Smirnov (K-S) test. In this way we could combine visual identifiability of TSM parameters with numerical thresholds defined by K-S test. In the revised manuscript we will emphasize the role of global identifiability analysis and K-S test to address identifiability of TSM parameters.

General comments

Abstract

R2: "The abstract is concise and clearly written. How to fill the research gap is properly addressed.

1. Line 14: I would consider adding 'adjacent groundwater bodies' as a exchanging agent.
2. Line 24: … TSM parameters, respectively. The severe differences…
3. Line 26: … at the study site → The article 'the' makes reference to a determined object, but the study site has not yet been introduced. Consider changing 'the' for 'our' or 'a study site in western Luxembourg', for example."

Authors: We thank R2 for the constructive suggestions. We will include the suggested changes in the abstract to improve its clarity.

Introduction

R2: "The introduction chapter needs to be revised in order to be fully understandable. The content is good, but an improved organization can elevate it."
Authors: We thank R2 for the supportive feedback. We will revise the introduction to improve its readability. In agreement with R1, we will re-organize the research gaps. We will also remove lines from 83 to 92, since we do not address a numerical modification of TSM.

R2: "Tend to list consecutive citations in chronological order (e.g. lines 72-73, 78-79, etc.)."
Authors: We will list consecutive citation in chronological order in the entire manuscript.

R2: "Line 49: Many readers can disagree with such a definition of 'hyporheic zone'. I don't agree with the idea of a fully saturated thickness. I usually consider the definition given by Cardenas & Wilson, 2007 (Cardenas, M. B., & Wilson, J. L. (2007). Exchange across a sediment–water interface with ambient groundwater discharge. Journal of Hydrology, 346, 69– 80), as a good approach."
Authors: We thank R2 for the suggested definition. The definition of the "hyporheic zone" is still controversial and not univocal. We will clarify as follows: "the saturated area that is physically influenced by water and solutes exchange between the stream channel and the adjacent groundwater" including the suggested paper in the citations.

R2: "Lines 79-82: This reads like a Research gap, and it is well stated, but it should fit at a later stage, before enumerating hypothesis and research questions."
Authors: We agree with R2 that this statement should come before addressing research questions. We will move it before the section that introduces the research question

R2: "Lines 103 and 113: I would not change paragraph in this sentence."
Authors: We will modify accordingly.

R2: "Line 105: 'stream velocity' → The stream itself does not move. Better using 'streamflow velocity'."
Authors: We will double-check the use of "streamflow velocity" in the entire manuscript.

R2: "Line 128: I would state the hypotheses in a new paragraph."
Authors: We will move the hypothesis in a new paragraph, together with the research gap moved from lines 79-82.

Study site and methods

R2: "Line 152: Study site and 'data' → Better specify which data."
Authors: We will clarify.

R2: "Study site and data → How long is the studied stream reach?"
Authors: The stream reach is 55 m long (cf. line 165 of the original manuscript).

R2: "Lines 161-162: How was Q calculated? (± Analytical Error?)"
Authors: Q was calculated from the tracer BTCs. As we did not perform repeated measurements, we cannot estimate the analytical error. However, the most common estimated error for discharge from dilution gauging method is ± 8% (Schmadel et al. 2010).

R2: "Line 164: How was EC calibrated according to NaCl concentrations? Was EC temperature compensated? (± Analytical Error?)"
Authors: EC-Cl conversion was done in laboratory using a known-volume sample of streamwater (taken before tracer injection) and adding known-quantities of a solution with a known concentration of Na-Cl. The regression line between EC-Cl had a $R2=0.9999$ (line 165-166). Temperature compensation was automatically considered via Nonlinear temperature compensation (nLF, according to EN 27 888). We will clarify this in the revised method section.

R2: "Was there any difference in background EC between the injection and the measurement locations? Or was it assumed to be equal? Or is it one of the testing parameters? These aspects are actually interesting."
Authors: The background concentration at the injection point is slightly lower from the background concentration at the measurement location. We account for this difference during the EC-Cl conversion by sampling the streamwater used for the EC-Cl conversion at the measurement location. We will clarify this detail in the revised version of the manuscript.

R2: "Line 181 (Equation 3): Why the second term of the second equation is negative?

If Cs < C, the concentration of a certain solute in stream tends to get diluted. But according to eq. 3, ∂C/∂t would become larger"

Authors: We double checked the equation 3, and we confirm that the second equation is negative. It could also be written as (Harvey, Wagner, & Bencala, 1996):

$\frac{dC}{dt} = \alpha \ \frac{A}{A_{TS}}(C - \ C_{TS}).$

Note that C_TS is not defined a-priori, as larger or smaller than C, but it is function of C itself and of the concentration C_TS at the time-step antecedent the one investigated. Crank-Nicolson approximation for the second equation of the system of equation 3 reads:

$$C_{TS}^{j+1} = \frac{(2-\gamma)C_{TS}^{j} + \gamma\left(C_i^j + C_i^{j+1}\right)}{2+\gamma}$$

Where: *γ=αAΔt/A_TS*; *i* denotes the central segment of the central-divided differences; *j* denotes the investigated initial timestep, while *j+1* denotes an advanced time. The full explanation of the Crank-Nicolson approximation for TSM resolution can be found in Runkel & Chapra (1993).

R2: "2.2: How the accuracy of the ADE and TSM results are assessed?"

Authors: Accuracy of both ADE and TSM results is assessed via Root Mean Squared Error objective function (lines 186-187).

R2: "Line 207: I assume CDF is the Cumulative Distribution Function, but it has not been defined as such yet."

Authors: R2 is right, we will clarify the abbreviation in the revised methods section.

R2: "2.3: How is the uncertainty of the model iterations' results, DYNIA's results and the comparison with OTIS-P and OTIS-MCAT assessed?"

Authors: We clarified uncertainty vs. identifiability above. Non-identifiability of TSM results and OTIS-MCAT was determined from global identifiability analysis and K-S test (lines 195-216). Identifiability in the dynamic identifiability analysis can be addresses via the metric called "Information content", defined as: "one minus the width of the 90 % confidence interval over the entire parameter range (Wagener et al., 2002)." Thus, a large 90 % confidence interval is associated to a low information content and uncertain/non-identifiable parameter (lines 221-229). Identifiability in OTIS-P (95% confidence limits) are derived in OTIS-P package via linear numerical approximation of the variance–covariance matrix within a local neighborhood of the solution (Runkel 1998; Ward et al., 2017). We will clarify how identifiability in OTIS-P is evaluated.

R2: "Line 270 (Equation 7): Introduce a tab in the equation's label, so it can be in line with the rest of the equations."

Authors: Nice catch! We will insert a tab prior to "Eq. 7".

R2: "2.4: I assume that including FMED reduced the systematic error, but is it quantifiable?"

Authors: Unfortunately, Fmed does not act on the reduction of error and non-identifiability of transient storage modelling. It is a non-dimensional metric used to interpret TSM results in comparison to other reaches or discharge conditions (Gooseff et al., 2013; Runkel, 2002, and section 2.4).

Results

R2: "This chapter is properly written but the sub-chapters could be re-organized to build-up a smoother storyline. See the last part of General comment (a)."

Authors: We thank R2 for the supportive suggestions for improvement. In light of the feedback of both R1 and R2 we will revise the structure of the results section to mirror the order of the research questions and the order of the revised method sections.

R2: "Lines 279-282: This reads more like in the 'Methods' section, right?"

Authors: We will remove this short paragraph to avoid redundancy between method section and discussion section.

R2: "Why was vpeak and no other velocity chosen for the fixed velocity scenario? Why not median velocity which could be more a representative velocity value?
What does vpeak mean?
- Peak velocity during tracer experiments?
- Mean velocity during the experiment E3 which had the largest discharge?
- Peak velocity during the entire monitoring period?"
Authors: We will clarify that v_peak is the velocity obtained via v_peak=L/tpeak. Where L is the length of the investigated reach and tpeak is the time for concentration peak to travel through study reach. We used v_peak as it is commonly used in many transient storage studies (Ward et al., 2013; Kelleher et al., 2013; Wlostowski et al., 2017; Ward et al., 2017; Ward et al., 2019a; Ward et al., 2019b). We will clarify this choice in the revised version of the method section.

R2: "Line 292: Figure 3c-f →Is Figure 3f informative in the tailing of the BTC?"
Authors: Yes, it is in our view. We will clarify in the revised manuscript that by "tail" we meant sections of the BTC from the peak to the end of the BTC.

R2: "Line 297: Add a comma after > 0.05),"
Authors: We will add a comma after "0.05".

R2: "8.8 l/s for E3 → I would not consider this discharge as a low qs, since you state that the exchanging flow is > 1/3 of the total Q of the stream."
Authors: We will remove subjective terms as "low", "high", "long" in the manuscript to avoid unprecise writing.

R2: "Line 343: 15 hrs for E3 → I would not consider this as a long residence time. The flux velocity is roughly 4.6 m/hr, which according to me can be slow for dead zones, but not for hyporheic flux, or GW-SW exchange."
Authors: We will remove subjective terms as "low", "high", "long" in the manuscript to avoid unprecise writing and subjectivity.

R2: "Lines 345-348: are these simulations physically possible at all, knowing the actual stream discharge? As for Tsto, do 0.8 (in E1) and 3.3 (in E3) m/s make sense?"
Authors: After the first TSM iteration, the parameters were not identifiable thus they may not make physically sense. We would recommend to only check for physical realism after the parameters became identifiable. However, in the following we will present the physical realism of the exchange rate (qs). We exemplified that for E1 (Q = 2.52 l/s).
Following the OTIS_MCAT results, the exchange rate over the stream length of 55 m was 23.1 l/s (cf. line 345). In this case the estimated storage volume in the stream channel was 1930 l. After the fourth TSM iteration the exchange rate over the stream length of 55 m was 2.7 l/s (cf. line 353). In this case the estimated storage volume in the stream channel was 2029 l. It is evident that both rates would be physically possible. This holds true for the other transport metrics. These results make clear that is important to reach parameter identifiability since non-identifiable results can reproduce physically possible behavior.

R2: "Lines 353-354: Again… double check the applicability of these results into the actual flow regime."
Authors: The results are physically possible. We have already outlined for E1 in the answer above. Here we summarize the same results for E2 and E3 after the last TSM iteration. Storage volume in the stream channel for E2 is ~2'850 l while the exchange rate over the stream length of 55 m was ~5 l/s. Storage in the stream channel for E3 is ~4'500 l while the exchange rate over the stream length of 55 m was ~23 l/s. Since all the evaluated storage metrics are interlinked by Alpha and A_TS the physically realism is preserved.

R2: "Comments 26, 27 and 28: Since you are doing a comparison between methods, rather than using terms like 'high' or 'low', it is better to use relative terms such as 'higher' or 'lower', or even 'distinct'."
Authors: We will account for this in the revised manuscript.

R2: "Line 349: 'whether' does not sound appropriate in this context. I would rather use 'regardless'?"
Authors: We will adapt accordingly.

Discussion

R2: "Same as in the previous chapter, the sub-chapters could be re-organized to build-up a smoother storyline. See the last part of General comment (a). Aren't 4.1 and 4.3 more related to each other, and 4.2 a unique sub-chapter. I think starting with 4.2 and then following with 4.1 and 4.3 would improve the chapter's sequence."
Authors: We will adapt by starting with 4.2, followed by 4.3 and 4.1. These modifications will allow us to respect the order of the research question, as suggested from R1.

R2: "4.1: Are the uncertainties specified?"
Authors: The uncertainties of the transport metrics are reported as boxplot limits (5-95 percentiles) of the top 100 model results for every TSM iteration in Figure 7.

R2: "Lines 402 and 406-407: Aren't these contradictory statements? Identifiability is contradictory under the same scenario (in both cases v = vpeak)."
Authors: In the revised manuscript we will clarify that non-identifiable alpha was found when velocity was assumed equal to v_peak.

R2: "Line 438: 100,000 instead of 100'000."
Authors: We will change accordingly.

R2: "Line 446-447: Seems like a very generalist closing statement."
Authors: We will revise to be more specific.

R2: "Lines 460-462: Aren't these contradictory statements? The assumption v = vpeak might not be representative of the advection role, but can encompass the effect of advection in the entire BTC."
Authors: As it is written, these statements could indicate that we suggest that vpeak can and cannot be representative of the advection role on the BTC. We will reformulate the sentences to make clear that we were comparing our findings to past studies.

R2: "Lines 480-503: Figure 8 is actually a very good review from other authors' findings. I quite like it. The piece of text relating to this figure is actually an important part of your discussion, but again it reads like a succession of sentences with vague organization, and it forces the reader to continuously go back and forth between the text and the figure. I propose the following. Instead of travelling through all the cited authors and comparing your results with their results, why can't it be presented in concordance with the different BTC's features, e.g.:
- Rising Limb
- Peak
- Falling Limb
- Tail

One can travel along the BT curve and compare your data with literature data. The reading would be more graphical and intuitive, and it could probably help to safe some word spacing.
By the way, it is interesting to know whether there is a method or a simple threshold value to distinguish between the end of the falling limb and the beginning of the tail?"
Authors: We will reformulate this section as suggested by R2. There is no univocal method or threshold value to distinguish between falling limb and the tail of the BTC. Therefore, we decided to consider as "tail" the section from the peak to the end of the BTC. We will include this information in the revised manuscript.

R2: "Lines 502-503: '… Kelleher et al., 2013 also indicate…" → whereas it has a strong influence based on your results, right?"
Authors: We will emphasize the difference between our results and Kelleher et al. (2013) results in the revised 4.3 section.

R2: "Lines 503-504: 'Different sensitivity…' → This reads like the staring of the next paragraph."
Authors: We will change accordingly.

Conclusion

R2: "The conclusion reads well, but again, I would try to expose the ideas trying to follow the same sequence of findings shown in the results."
Authors: We will adapt the conclusion to follow the revised structure of the methods, results, and discussion.

R2: "Line 523: '… that the BTC…' -→ Revise wording."
Authors: We will revise and clarify this sentence.

Tables and Figures

R2: "Line 739 (Table 2): *v = vpeak* → Try to keep the same notation as in the text (i.e. Italic and subscript)."
Authors: We will uniform the fonts of *vpeak* and others.

R2: "Line 777 (Figure 2): The meaning of colored dots for (m-o) plots not specified."
Authors: We will clarify that the indicated colors refer to plots from "k" to "o".

R2: "Line 778 (Figure 2): Try to keep the same notation as in the text (i.e. Italic and subscript)."
Authors: We will unify the font of *vpeak*.

R2: "Figure 3: Here you show the BTC in E1 when the streamflow is smaller and the identifiability is probably less dynamic. I would have liked to see differences in dynamism between the different tracer experiments. This does not mean that you need to re-do the figures for other tracer experiments, but probably comment a bit on it in the text."
Authors: The dynamic identifiability analysis showed that alpha and A_TS are highly identifiable (inf content > 0.66) for smaller section of the tail of the BTC with increasing discharge conditions. For example, the information content of A_TS is above 0.66 for 51% of the tail of the BTC for E1, for 23% for E2, and for 19% for E3. However, we decided to not discuss differences and similarities for E1, E2, E3 after the 1st TSM iteration, since the dynamic identifiability analysis after the 1st TSM iteration is conducted with non-identifiable TSM results (figure 2, green points). We decided to compare dynamic identifiability analysis results only for identifiable TSM results via the qualitative plots in Figure 8 to keep the manuscript at a readable length. However, we will briefly comment the observed difference in terms of identifiable alpha and A_TS on the tail of the BTC for the first TSM iteration in the revised result sections.

R2: "Figure 4: I would probably combine Figs. 3 and 4 in the same figure. Both are showing exactly the same info (Experiment E1 and v = vvariable) but only for different iteration states, and can be potentially confusing."
Authors: We will change accordingly. The revised figure 3 will include subplots from (a) to (t). Subplots (a)-(j) will report results of dynamic identifiability analysis for the 1st TSM (old figure 3). Subplots (k)-(t) will report results of dynamic identifiability analysis for the 4th TSM (old figure 4).

R2: "Figure 7: In the Y-axis of plots g-I and j-l, please use similar notation. Use either 0.01 or 10-2."
Authors: We will unify the numerical notion.

R2: "Lines 866-867 (Figure 7): 'm','n','o' seem to be leftovers."
Authors: We will remove these letters.

R2: "Line 868 (Figure 7): '… and equal to vpeak, **respectively**'."
Authors: We will adapt accordingly.

R2: "Figure 8: Plots g-l are presented first and explained later. They can both combine in the same piece of text."
Authors: We will present and explain the plots (g)-(l) in the same part of the figure description.