Please find below our responses to the comments of reviewer 2. Our responses are given in blue.

1. The authors present an interesting application of the transient storage model (TSM) in which instream tracer data is supplemented by naturally occurring radon data. The paper is generally well written and the authors have some good points. The manuscript has a number of shortcomings, however, and major revisions will be needed to develop a manuscript that warrants publication.
   The authors raise two fundamental issues in the introduction that are recurring themes in the TSM literature. I haven't done much work in this arena for some time, but I have conducted other reviews where I've made similar points. Please keep in mind that many of the comments that follow are directed at the field in general rather than these particular authors.
   The first issue is the alleged inability of the tracer-modeling approach to quantify "long flow paths". I can certainly think of cases where this is important -- e.g. when the stream crosses a fracture zone and water leaves the channel and doesn't return before the observation point (path C in Fig 9; lines 59/60) - but in general I would say that this concern is overblown. The importance of this issue is dependent on the overall goal of the study. If the goal is to quantify the processes that affect constituents that are present in the water column, such as the case of an accidental spill into a waterway, the "failure" of the approach is not of consequence -- the tracer mimics the constituent of concern, so if these long flow paths don't affect the tracer, they don't affect the constituent (the amount of mass in long flow paths is trivial). If your interest is the fate of molecules in the riparian zone as they interact with the stream (e.g. diffuse agricultural pollution that enters laterally), the long flow paths may be of greater importance. But I would argue that the typical stream tracer approach isn't an appropriate means to address this latter scenario. Further, if one is concerned about long flow paths, why use slug additions that by definition have a short window of detection? Its important to note that 100% of the data sets used to establish the transient storage paradigm were based on continuous tracer injections wherein tracer concentrations are allowed to reach a steady-state plateau. Under this approach, plateau will be achieved when all flow paths that return to the stream have had sufficient time to do so (thereby eliminating the problem associated with path B). Unfortunately many contemporary investigators have utilized slug additions as a means of streamlining the field effort, and this simplification comes with a price.

   - *We would like to thank the reviewer for these comments. We agree that, ultimately, the importance of temporally longer flow paths depends on the aims of the study. In the provided example of an accidental spill into a waterway, a slug tracer injection effectively mimics the constituent of concern. The temporally longer flow paths influence the slug tracer in the same way as they do the constituent.*
   - *Despite the well-established knowledge on the WoD inherent in slug tracer injections, we maintain that studying alternative tracers, such as radon (as proposed in our study), is critical. Alternative tracers may extend the WoD and offer additional insights into solute transport. We agree that constant-rate injections may better capture temporally longer flow paths returning to the stream. Therefore, we will propose that the joint calibration of continuous tracer injections with radon represents a promising direction for future research:*

     <u>"Furthermore, future research could jointly calibrate TSMs with radon and tracer data from constant rate injections. This approach would test whether radon captures temporally longer flow paths than those detected by constant rate but not by slug tracer injections."</u>

2. The second fundamental issue is the much celebrated parameter identifiability/equifinality problem raised by numerous investigators. The point I'd like to make here is that the "failure" to develop unique estimates of the transient storage parameters does not

necessarily indicate a "problem" with the "traditional" tracer based modeling approach. The common explanation for parameter identifiability problems is that the approach is somehow inadequate. An equally plausible explanation in many cases is that transient storage simply isn't important in the reach of interest. In the extreme case, consider data from a straight, lined canal. If you applied the approach in this situation and were unable to identify the transient storage area and exchange rate, would you blame the approach, or conclude that transient storage is unimportant? For the case of a natural stream reach, there may be identification issues simply because these flow paths aren't relevant when viewed in terms of constituent mass.

As someone who has worked for years trying to help numerous researchers analyze problematic data sets, its my opinion that the transient storage approach is adequate when applied to a quality data set. Most problems arise when the data is sparse in critical parts of the breakthrough curve, the investigators have inaccurate estimates of streamflow, and/or the data is simply noisy (due to poor lab analyses, incomplete mixing, etc).

- *Thank you for this important point. We fully agree that the “failure” of parameter identifiability could also relate to transient storage mechanisms not playing a role. In such cases, the advection-dispersion equation might be a better choice. Furthermore, we agree that the quality of the BTC could explain the lack of parameter identifiability in some cases. However, we also think that not all BTCs are simply noisy. While data quality remains a significant issue, equifinality and parameter identifiability are critical aspects extensively discussed in hydrologic research (e.g., Beven and Freer 2001).*

- *In the revised manuscript, we will incorporate the suggested alternative explanations for the “failure” of parameter identifiability:*

  “In streams where exchange between advective flow and storage zones is not minimal, non-identifiability may arise when multiple parameter combinations yield equivalent model performance.”

specific comments:

----

3. 1) abstract "...the amount and location of groundwater inflow, which is not explicitly accounted for in TSM". This is incorrect, TSMs consider inflow (e.g. ql in equation 1).

   - *Thank you for this important point. We will revise the statement for improved clarity. We will change to:*

     “This non-identifiability arises because radon activity in streams remains at steady-state and is highly sensitive to the location and amount of groundwater inflow, as well as contributions of flow paths from subsurface transient storage zones. As a result, radon measurements are biased toward longer-timescale flow paths, limiting their applicability to uniquely constrain solute transport parameters in TSM calibration without complementary slug tracers.”

4. 2) line 32 "solute tracer experiments are biased towards faster flow paths". To reiterate, if it doesn't affect the tracer, it doesn't affect the constituent of interest (at the scale studied). The tracer is simply reproducing what constituent molecules would experience.

   - *Please refer to our detailed response to comment no. 1 above.*

5. 3) line 40 "TSMs assume a uniform, steady-state ..flow". This is incorrect -- the lateral inflow term can be used to implement non-uniform flow. Consideration of unsteady flow is rare, but possible:

   Runkel, R.L., McKnight, D.M., and Andrews, E.D., 1998. Analysis of transient storage subject to unsteady flow: Diel flow variation in an Antarctic stream, J North American Benthological Society, 17(2), 143–154, 10.2307/1467958 .

   - *Thank you for this suggestion. We will revise the text accordingly and incorporate the literature recommendation as follows:*

     “In their simplest forms, TSMs assume a uniform, steady-state, one-dimensional flow, modeled using the advection-dispersion equation (ADE), while also accounting for first-order mass transfer between the advective flow and a storage zone (Bencala and Walters, 1983; Gooseff et al., 2008). Extensions allow for implementing non-uniform groundwater inflow via lateral inflow terms (Runkel et al., 1998).”

   also, I don't agree with "effectively infinite dimensions" (nor does your text on line 154, "finite-size, well-mixed storage zone")

   - *We will remove this aspect from the text; please refer to our response to the previous comment for the improved text.*

6. 4) line 45, "The parameter values derived from TSMs provide a means of comparing solute transport within a single stream or across multiple streams". Could cite:

   Runkel, R.L., 2002. A new metric for determining the importance of transient storage, J. North American Benthological Society, 21(4), 529–543, 10.2307/1468428 .

- *We will follow the suggestion.*

7. 5) lines 50-55. Again, it depends on your objective; for most cases I'm not convinced adding the radon data really helps. I find the idea of supplementing the tracer data with other auxiliary data the most promising approach for separating surface and subsurface (hyporheic) storage. I think you allude to using radon data for this purpose later in the paper. For more of my thoughts (as if you haven't had enough :-), see:

    Runkel R.L., McKnight, D.M., and Rajaram, H., 2003. Modeling hyporheic zone processes, Advances in Water Resources, 26(9), 901–905, 10.1016/S0309-1708(03)00079-4

    – *Thanks for pointing this out. We will adapt parts of the discussion to highlight the research opportunity of using radon in a two-storage zone TSM.*

    "We selected the one-zone storage model because this setup aligns conceptually with most previously established radon models (e.g., Cook et al., 2006; Frei and Gilfedder, 2015) and many TSM calibration with slug tracers (e.g., Bonanno et al., 2023). To our knowledge, no prior radon study has considered two storage zones, highlighting a promising opportunity for future research."

8. 6) line 79-80. "When surface water exchanges with subsurface transient storage zones and contacts radium-bearing minerals in the streambed, radon activity increases as a function of the time spent in the hyporheic zone". Radon in the stream could come from this contact w/ streambed materials AND/OR groundwater inflow. I'm guessing the streambed part is modeled using the production term (gamma, equation 2) and the groundwater part is handled through the lateral inflow term. More description of how this is handled should be added, including what you're using to set the lateral inflow concentration.

    – *We will follow the suggestion The revised version will read as:*

    "We calculated the radon production term γ as the product of the decay constant (0.18 d^{-1}) and the measured equilibrium radon activity (Gilfedder et al. 2019). This approach assumes that radon production occurs only in the subsurface transient storage zone of the stream. However, radon may also increase when stream water interacts directly with the streambed surface."

    "For all three calibration approaches, the measured equilibrium radon activity was used as the activity of the groundwater inflow."

9. 7) line 108-110. "Each reach length was at least 20 times the Wetted Channel Width to control for expected variations in solute transport that occur as a function of reach selection". Similar metrics for reach length are often mentioned in regard to complete mixing of the tracer with depth and width, but I'm guessing you're alluding to the Dahmkohler number. You may want to clarify this.

    – *We do not explicitly refer to the Damköhler number in our study; however, we will revise the sentence to clarify the rationale behind the chosen reach length:*

    "Each reach length was at least 20 times the Wetted Channel Width. This was done to ensure thorough mixing of the tracer and solutes across both the width and depth of the channel, thereby minimizing spatial variability and potential biases due to reach selection"

10. 8) line 120. "Discharge was calculated for the resulting BTCs using dilution gaging". I assume this required some relationship between conductivity and chloride, which is fine. But you may want to try this method as a check:

    McCleskey, R. B., Runkel, R. L., Murphy, S. F., & Roth, D. A. (2025). Stream discharge determinations using slug additions and specific conductance. Water Resources Research, 61, e2024WR037771. https://doi.org/10.1029/2024WR037771 .

    I'm happy to help and provide an updated spreadsheet if interested.

    – *Thank you for highlighting this interesting manuscript. We agree that the method represents an important contribution to discharge estimation and will consider this approach in future research.*

11. 9) line 128. "Radon sampling sites were co-located with BTCs observations" - was there 1 sample per site? This paper is all about the radon yet we don't get to see the data - please add.

    – *We present the radon data in the results section, but only report the minimum and maximum values for the stream. A detailed dataset will be provided in the appendix.*

12. 10) line 187-188. Here you refer to D, alpha, Ats; Table 1 has D, A, Ats

    – *Thank you for this helpful suggestion. We will revise the table accordingly to improve clarity and accuracy.*

13. 11) line 188-189. Is it a uniform distribution or a logarithmic one?, I'm confused...

    – *We performed uniform sampling in logarithmic space. To clarify this, we will rewrite the sentence as follows:*

    "We sampled parameter values of D, α, and A_{TS} uniformly from a log10 transformed space to ensure approximately equal representation for each order of magnitude within the parameter space (Kelleher et al. 2013; Ward et al. 2017)."

14. 12) line 195, "intersection of behavioral parameter sets" - I like this approach...

    – *Thank you!*

15. 13) calibration approach described in section 2.4 and various flow approaches described later in the paper. I don't agree with the approach of fixing velocity and A, and estimating Q for several reasons. I highly recommend fixing Q and estimating A. A few subpoints:

    - why not use the Q from the slug additions? I think your three Q methods overly complicate things and these complications aren't relevant to what you're trying to show (the utility of adding radon data). I suggest using the Qfix approach and dropping the others. If you're worried about uncertainty in your slug estimates, develop a linear regression between the Q estimates and distance and use the regressed values at each site.

    – *The reason we apply three different calibration approaches is not to demonstrate the utility of adding radon. Instead, we aim to show why only using radon does not work.*

- *Only by using different calibration approaches can we demonstrate that:*
  - *The radon activity in the stream is highly sensitive to inflowing groundwater; even within the measurement uncertainty of discharge calculated from slug tracers, we observe differences in model performance (Q_{fix}).*
  - *A higher degree of freedom in the calibration method (Q_{LHS}) compared to Q_{fix} improves model performance, as indicated by a greater number of consistent behavioral parameter sets for both radon and chloride.*
  - *Furthermore, calibrating groundwater inflow allows us to infer water loss (Q_{out}) with radon calibration, an insight we do not obtain by calibrating the TSM to slug injections.*

- *We are not concerned about uncertainty in slug tracer estimates. Rather, our effort to apply three calibration approaches highlights the uncertainty inherent in radon. We will clarify this aspect in the manuscript:*

  "Radon activity in streams varies with the amount of inflowing groundwater, as radon activity differs significantly between groundwater and surface water (Cook, 2013). Small changes in the amount of inflowing groundwater may lead to differences in model performance. To account for this, we either calibrated or calculated groundwater inflow within three different calibration approaches, in addition to calibrating D, α, and A_{TS}."

  - fixing A does nothing to reduce "potential issues of equifinality" (line 205). The main channel area is by far the easiest parameter to estimate via simulation as it controls the velocity and thus the timing of the BTC. When estimating A, Ats, alpha and D using nonlinear regression, the A parameter always has the narrowest 95% confidence interval and is the parameter estimated with the most certainty. Equifinality problems usually arise when there's not enough data in certain parts of the BTC to uniquely identify D, Ats, and alpha. Wagner and/or Harvey have papers (book chapters?) which show the sensitivity of various parts of the BTC for various parameters and there's always ample info to estimate A. By fixing A (and/or velocity) you're ignoring this information and biasing your other parameter estimates. If you insist on fixing A/velocity, I suggest using the center of mass rather than the peak as this more truly represents the average reach velocity (see Runkel 2002 ref above), especially if there's an extended tail.

- *Thank you for mentioning this important point. We agree that different parts of the BTC provide distinct information on specific parameters, as demonstrated by previous studies using slug and constant-rate injections (e.g., Wagener et al. 2002; Wlostowski et al. 2013; Kelleher et al. 2013; Wagner and Harvey 1997). However, research does not consistently identify which part of the BTC corresponds to which information (see Figure 8 in Bonanno et al. (2022) for a visualization of these differences). Recently, Bonanno et al. (2022) systematically studied the effect of calibrating stream velocity on parameter identifiability in BTCs. They concluded that fixing the stream velocity is reasonable and does not affect the identifiability of the remaining parameters. We followed this approach since, to our knowledge, it is the most recent and state-of-the-art method, and the only systematic study that explicitly studied the calibration of stream velocity. Therefore, we consider fixing A/velocity to be reasonable for our study. A detailed explanation of this aspect will be provided in the manuscript:*

  "This choice was motivated by findings from Bonanno et al. (2022), who showed that A_{TS} and α are often not identifiable when v is calibrated instead of calculating v by dividing the stream length by the arrival time of the concentration peak of the downstream BTC."

16. 14) lines 215-220. Degassing is a function of turbulence which is effected by velocity and thus Q. Stream width (surface area) is also important. Why not use the value estimated at the most similar Q?

   – *We agree that the discharge at the most similar Q condition is likely to best reflect the experimental situation. However, Cargill et al. (2011) did not find a clear functional relationship (linear, exponential, or otherwise) between Q and k and investigated only three Q conditions. We therefore consider Cargill et al.'s (2011) results to be a reasonable approximation, but not sufficiently certain to be fixed at a single value. In addition, we explicitly test whether a potential misestimation of k would alter the conclusions. This ensures that our findings are not an artifact of choosing one particular k value, but remain valid across a plausible range.*

17. 15) line 334. "radon provided more information on groundwater inflow". With all the uncertainty involved (e.g. degassing rate, radon inflow concentration, radon analyses) why would these estimates of inflow be better than your dilution gaging? The Qfix approach is the way to go.

   – *We do not claim that radon provides more information on groundwater inflow than derived from dilution gauging. However, our quantitative results on information content show that TSM calibrations with radon offer more information than calibration with chloride concentration on groundwater inflow, which is not related to dilution gauging. This is likely because radon activity increases with inflowing groundwater, whereas chloride concentrations are diluted. We will clarify this aspect in the revised version of the manuscript:*

   "Although we expect groundwater inflow to primarily affect radon activity in streams, we also calibrated TSM parameters in three model setups that varied in groundwater locations using chloride concentrations. This was motivated by the assumption that chloride-free groundwater, as commonly assumed in TSMs, dilutes chloride concentrations in streams."

18. 16) line 402. "parameter interactions became evident when inflow was at the most downstream points". Back on line 310 you mention all of the inflow entering over 1 meter long area; it's my experience that these abrupt changes can cause numerical problems -- you may not 'see' these problems at sites farther away from the observation point (the Crank-Nicolson method usually 'recovers'), but this one at the downstream end could be a numerical artifact. I suggest spreading the inflow over several 1-m segments (maybe 10 m).

   – *Thank you for raising this important concern. As part of our calibration routine, we always check the numerical stability of the model run as a prerequisite to the actual calibration (see our publicly available calibration code in Bacher et al., 2025). All presented results therefore stem from numerically stable calibrations. Therefore, we can confirm that the calibration with groundwater inflow at the most downstream location is not a numerical artefact and we are confident that the numerical stability we observe is unrelated to any 'recovery' of the Crank–Nicolson method. The same data were used in all three setups, with only the groundwater inflow location varied.*
   – *However, we speculate that numerical instabilities may become more significant when higher lateral groundwater inflow is calculated or calibrated. In our simulations with three different inflow locations, we used data from reach #1 for calibration and discharge was calibrated within a parameter range defined by BTC measurements. The discharge in reach #1 did not change significantly (see Figure 8), which led to relatively low inflow values and may explain the numerical stability.*

19. 17) line 470. "This suggests that obtaining narrow, well-constrained estimates for groundwater inflow, ATS, or α from calibrating the TSM with radon will remain challenging unless at least one of these parameters is further constrained" -- this is easily done - fix Q and save the modeling approach for the more empirical/abstract parameters!!!

    – *Please refer to our response to Comment No. 15 for a detailed reply.*

Beven, Keith; Freer, Jim (2001): Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology 249 (1-4), pp. 11–29. DOI: 10.1016/S0022-1694(01)00421-8.

Bonanno, Enrico; Blöschl, Günter; Klaus, Julian (2022): Exploring tracer information in a small stream to improve parameter identifiability and enhance the process interpretation in transient storage models. Hydrol. Earth Syst. Sci. 26 (23), pp. 6003–6028. DOI: 10.5194/hess-26-6003-2022.

Kelleher, C.; Wagener, T.; McGlynn, B.; Ward, A. S.; Gooseff, M. N.; Payn, R. A. (2013): Identifiability of transient storage model parameters along a mountain stream. Water Resour. Res. 49 (9), pp. 5290–5306. DOI: 10.1002/wrcr.20413.

Rodriguez, Nicolas Björn; Pfister, Laurent; Zehe, Erwin; Klaus, Julian (2021): A comparison of catchment travel times and storage deduced from deuterium and tritium tracers using StorAge Selection functions. Hydrol. Earth Syst. Sci. 25 (1), pp. 401–428. DOI: 10.5194/hess-25-401-2021.

Wagener, T.; McIntyre, N.; Lees, M. J.; Wheater, H. S.; Gupta, H. V. (2003): Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. Hydrol. Process. 17 (2), pp. 455–476. DOI: 10.1002/hyp.1135.

Wagener, Thorsten; Camacho, Luis A.; Wheater, Howard S. (2002): Dynamic identifiability analysis of the transient storage model for solute transport in rivers. Journal of Hydroinformatics 4 (3), pp. 199–211. DOI: 10.2166/hydro.2002.0019.

Wagner, Brian J.; Harvey, Judson W. (1997): Experimental design for estimating parameters of rate-limited mass transfer: Analysis of stream tracer studies. Water Resour. Res. 33 (7), pp. 1731–1741. DOI: 10.1029/97WR01067.

Wlostowski, Adam N.; Gooseff, Michael N.; Wagener, Thorsten (2013): Influence of constant rate versus slug injection experiment type on parameter identifiability in a 1-D transient storage model for stream solute transport. Water Resour. Res. 49 (2), pp. 1184–1188. DOI: 10.1002/wrcr.20103.