Please find below our responses to the comments of reviewer 1. Our responses are given in blue.

1. The authors have calibrated one-zone stream transient storage models (TSMs) using two different tracers: salt (NaCl) and Radon. The authors state that pairing salt tracers with Radon increases the window of detection (WoD) by up to 21 days, which is much longer than that of salt. The authors argue that including Radon improves the ability to identify the true mean value of transient storage parameters. This paper will make a nice contribution to the literature by providing new findings into measurement and modeling of stream tracer breakthrough curves.
   The presentation is good but there were a few places where this reviewed got confused. To benefit the reader through an improved presentation, below are a few general comments and several specific comments for the authors to consider.

   The authors state that jointly calibrating to two tracers is an improvement, which is exciting and novel; however, it is not clear to this reviewer where the joint calibration results are presented. In the figures and captions, it appears results are based on induvial tracers. It would help the reader to specifically state which results are the joint calibration and which are individual.

   - *Thank you for these important points. We agree that we have not explicitly mentioned what we define as "joint calibration". We will clarify this aspect in the manuscript. The revised version of the text will read as:*

     "Since we tested the same combinations of parameter values in the TSM for both tracers, the intersection of the behavioral parameter sets from both tracers reflects the parameter sets obtained when the model is calibrated with both tracers together. This indicates that when the behavioral parameter sets for both tracers are identical, the choice of tracer does not affect estimates of solute transport. We refer to this intersecting set as the 'joint calibration'. In contrast, other parameter sets were included in the behavioral set for only one tracer but not the other. These sets represent parameterizations that result in acceptable model performance for a single tracer, but are not robust in simulating both tracers simultaneously. We refer to these as the 'individual calibration'. The behavioral parameter sets were used for all subsequent analysis and calculations."

   - *Furthermore, we will state clearly in the figure and table captions where we refer to the joint and the individual calibration. For example, in Figure 3 in Table 2, we will specify which of the presented values of the Shannon entropy H and Kullback-Leibler divergence correspond to the joint or individual calibration. The revised captions will read as follows:*

     "Figure 3: Cumulative distribution plots (upper row) and posterior distribution plots (lower row). Each plot shows A_{TS} results using the best 1% and best 10% parameter sets from the individual calibration. (…)"

     "Figure 4: Cumulative distribution plots (upper row) and posterior distribution plots (lower row). Each plot shows groundwater inflow (q_l) results based on the best 1% and best 10% parameter sets from the individual calibration. (…)"

     "Figure 6: Distributions of behavioral A_{TS} values from model setups with varying groundwater inflow locations, each calibrated with chloride only. (…)"

     "Table 2: Shannon entropy H and Kullback-Leibler divergence D_{KL} for the prior and posterior distributions of model parameters (D, A_{TS}, α and the groundwater inflow q_l)

resulting from joint ($H(X|radon \cap chloride)$; $D_{KL}(X|(radon \cap chloride),X)$ and individual ($H(X|chloride)$; $H(X|radon)$; $D_{KL}(X|chloride)$ calibration; $D_{KL}(X|radon)$) of the TSM with chloride and radon. (…).”

2. If Radon is relatively steady and therefore does not capture the entire BTC, why would one assume that a TSM could be calibrated using only Radon? Pairing salt with Radon creates a more representative BTC. In simple terms, better calibration data, a better chance to identify parameters. That type of statement would benefit the reader.

   - *We agree that clearer explanation will be needed regarding the assumption of using radon alone to calibrate the TSM. Conceptually, the underlying assumption is that exchange processes occur regardless of whether tracer input is steady (e.g., from natural tracers like radon) or from a defined injection (e.g., chloride slug or constant-rate injections). From a physical point of view, these processes should label the same flow paths within a reach, leading to similar model parameters independent of the tracer used. This principle has previously been supported by studies simulating transit time distributions at the catchment scale (Rodriguez et al. 2021). We will rewrite part of the introduction to clarify this aspect:*

     “To answer these questions, we apply a coherent mathematical framework to radon and slug injections of sodium chloride (NaCl). This approach is motivated by previous catchment-scale studies showing that applying the same model framework to different tracers yields comparable insights into hydrological transport processes (e.g., Rodriguez et al., 2021; Wang et al., 2023). To ensure a coherent mathematical framework for both NaCl and radon, we adapt the transient storage model OTIS (‘One-Dimensional Transport with Inflow and Storage model,’ Runkel, 1998) by incorporating radon-specific processes such as degassing. We then jointly and individually calibrate the model with slug tracer (chloride) and radon data. We perform a global sensitivity analysis approach to assess parameter identifiability. Finally, we apply information theory to quantify the information gained from joint and individual calibration of the TSM with these tracers.”

   - *Thank you also for pointing out the need for more accessible language throughout the manuscript. We will revise the text accordingly to improve clarity and readability.*

3. Parameter “certainty” is difficult for this reviewer to understand and follow because “sensitivity” and “identifiability” seem to already cover that concept. How is certainty different than identifiability? If clearly a different metric, that makes sense. If certainty means something like identifiability, introducing a new term “certainty” just adds confusion.

- *Thank you for pointing this out. We agree that the distinction between identifiability, sensitivity, and certainty requires clarification. In our study, we are specifically interested in parameter certainty, which we define as the ability to estimate a parameter with low uncertainty. A key reason for parameter uncertainty is insensitivity, meaning that the model outcome does not change substantially with the parameter value. In such cases, a wide range of parameter values can produce similarly low nRMSE values. To address this, we first analyze sensitivity using nRMSE-versus-parameter plots and cumulative distribution function (CDF) plots. However, insensitivity is not the only cause of parameter uncertainty. Parameter interactions can also reduce identifiability, which we examine using scatter plots. These analyses allow us to characterize the main drivers of potential uncertainty. Finally, to quantify certainty itself, we analyze posterior distributions and compute Shannon entropy. A parameter is identifiable when it is sensitive, certain, and does not show strong interactions with other parameters. We will clarify this aspect in the text:*

  "The parameter identifiability analyses include the visual inspection of I) nRMSE vs. parameter plots (Wagener et al., 2003), II) cumulative parameter distribution plots (Kelleher et al., 2019), III) posterior distribution plots (Wagener et al., 2002; Ward et al., 2017), IV) scatter plots of the behavioral parameters, and V) calculation of the Shannon entropy for the posterior distribution of model parameters as metric for certainty (sensu Rodriguez et al. 2021, section 2.6)."

  "We used the information content to evaluate parameter certainty by calculating the Shannon entropy of their posterior distributions (Cover and Thomas, 2005; Loritz et al., 2018). Rodriguez et al. (2021) applied this approach in a catchment-scale study. The posterior distribution is the probability density function of the behavioral parameters sets."

4. If the background was subtracted for salt, why is groundwater influx also needed for the salt TSM? If QLHS is estimated from discharge, it would help the reader that conditions are still steady state. So groundwater inflow was assumed fixed for all reaches? And then estimated by gage data (dQ/dx). And also estimated directly via calibration. Correct?

   - *Subtracting background concentration is common in slug tracer tests, and justified by the peak concentrations significantly exceeding background concentration in the stream. However, inflowing water with less chloride concentration than those measured in the BTC may still have a diluting effect on tracer concentration. Therefore, groundwater influx is also needed for calibrating a TSM with chloride. We will highlight this aspect in 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."

   - *Discharge was calibrated in the Q_{LHS} approach rather than kept fixed. We will revise the section describing the calibration procedure to highlight this point. The revised version of the section will read as:*

     "In the second calibration approach (Q_{LHS}), discharge was calibrated as model parameter, like D, α, and A_{TS}, before calculating groundwater inflow. For each reach,

discharge was sampled from a normal distribution, because Schmadel et al. (2010) reported that discharge measurement errors follow a normal distribution. (…)"

5. What new information is presented in Figure 9? The red flow path does not seem technically correct.

   – *The novelty of Figure 9 is indeed the flow path "C," which is why it appears in red.*

   should state that this figure has been modified from Payne et al. 2009.

   – *We will clarify that this figure was modified from Payn et al. (2009).*

   Flow path C does not seem technically correct. You do not know if that represents tracer that bypasses measurement site. The WoD is your measurement window, not the real window in which flow paths exist. And why is C red? That adds confusion. This reviewer does not find any new information added by this figure, and it does not seem technically correct. Suggest to revise to clarify which arrow is specifically improved in this study.

   – *Flow path "C" is correctly labeled; it is identified only by radon, not chloride. We will explain the improvements in measuring this flow path and the novel insights provided by the figure throughout the text and in the figure caption:*

     "Figure 9: Conceptual figure of flow paths in streams. The box shows the reach of investigation. A is a flow path that is labelled by the tracer at the upstream location and returns within the WoD. B is a flow path with timescales longer than the WoD, where transit times exceed the duration of the tracer experiment, as indicated by the dashed arrow. C is a subsurface flow path entering the hyporheic zone upstream of the reach. This flow path is characterized by subsurface transit times shorter than 21 days and radon activity that has not yet reached secular equilibrium. This flow path can be identified by radon but not by chloride concentration, which explains the red coloration. D is a flow path that bypasses the downstream end of the reach. E is a groundwater flow path with transit times longer than 21 days and radon activity at secular equilibrium. Flow path D is conceptually excluded from the TSM, which is why it is highlighted with a red dashed arrow. (Figure adapted from Payn et al. (2009))"

     "Spatially variable gross water gains exceeding net discharge from TSM calibration with radon suggest that the water balance of Oak Creek is affected by stream–subsurface water exchange that is occurring at multiple spatial scales. Previous studies have shown that large spatial-scale subsurface flow paths, that are subsurface flow paths originating from further upstream of the stream reach, play a critical role in explaining water mass balances in streams (e.g., Payn et al., 2009; Stanford and Ward, 1993; Ward et al., 2023). Unlike chloride, radon uniquely labels these flow paths (Fig. 9, arrow C), which are otherwise undetectable. This unique labeling underpins the superior sensitivity of radon in revealing subsurface flow paths and highlights its novel contribution to understanding their role in stream water balances. Conceptually, this labeling capacity of radon is reflected in measured steady state activity at both the upstream and downstream ends of the reach before the slug tracer experiment began. Steady state activity indicates that the measured radon activity includes subsurface water parcels pre-labelled with radon prior to the experiment. Thus, large spatial-scale subsurface flow paths from further upstream also contribute to measured radon activity at the downstream end of the reach (Fig. 9, arrow C), along with groundwater inflow and temporally shorter flow paths (Fig. 9, arrows A, B, and E; Cook et al., 2013; McCallum et al., 2012). In contrast, chloride is injected as a distinct input signal (Dirac injection) at the upstream end of the reach and thus only reveals flow paths between locations of injection and measurement (Fig. 9, arrows A and B). This inherent bias in chloride injections toward labeling

timescales within the WoD complements the bias of radon toward large spatial-scale subsurface flow paths, underscoring the value of using both tracers together. However, despite these novel insights on large spatial-scale flow paths from calibrating TSM with radon, such paths are not explicitly represented in the TSM. Instead, they are indirectly accounted for through the calibration of groundwater inflow. This indirect consideration may lead to an overestimation of groundwater inflow values during calibration. Such overestimation reduces certainty in the A_{TS} and α parameters and causes interactions among groundwater inflow, A_{TS}, and α. Although radon's sensitivity to large spatial-scale subsurface flow paths provides valuable insights, it therefore also introduces bias in constraining transient storage parameters in TSMs."

6. This reviewer got a bit lost in the Discussion. The alternating representation and descriptions of groundwater and groundwater locations caused a tough read, and the key point seemed buried as a result. Radon helps constrain the GW inflow. Great. The salt is needed for advection and dispersion. Try to write in more simple terms where possible.

   – *Thank you for pointing this out. We will carefully revise the discussion, remove unnecessary repetition of results, and highlight our key findings more clearly. For example, the first section of the discussion will read as:*

     "Calibrating the TSM with chloride provides more information on solute transport than calibrating the model with radon because chloride is particularly informative on D (Table 2). Previous studies have shown that D mainly affects the rising limb of BTCs (e.g., Kelleher et al., 2013; Scott et al., 2003; Wlostowski et al., 2013). This highlights that tracers with a distinct rising concentration limb, such as chloride, are necessary to identify D. Radon, by contrast, provides more information on groundwater inflow and A_{TS} than chloride (Table 2) due to its higher activity in groundwater compared to surface water. Because chloride concentrations are smaller in groundwater compared to surface water, groundwater inflow simply dilutes chloride concentrations in the stream without adding additional information on inflow. In summary, the most information on solute transport is obtained by using both tracers jointly because each tracer contributions uniquely. Joint calibration with radon and chloride improves certainty in solute transport estimates compared to calibrating with either tracer alone. We therefore recommend calibrating TSMs with multiple tracers to improve estimates of solute transport in future studies.
     This recommendation aligns with recent calls for joint calibration of hydrological models with multiple tracers. For example, Neilson et al. (2010b) demonstrated that calibrating a TSM with both temperature and slug tracer data provides more insights into solute transport and exchange compared to calibrating the TSM with temperature alone. At the catchment scale, Rodriguez et al. (2021) demonstrated that jointly calibrating a storage selection function with deuterium and tritium reduced uncertainty in model parameters compared to calibrating the model with either tracer alone. Notably, the authors used a different quality criterion for behavioral parameter selection than we did. Rodriguez et al. (2021) used distinct threshold values for each tracer to obtain a comparable number of behavioral parameter sets due to differences in sampling frequency and thus the dataset length between deuterium and tritium. In contrast, we selected the best 1% and 10% of parameter sets as behavioral to remain consistent with previous solute transport studies (e.g., Kelleher et al., 2019; Wagener et al., 2002; Ward et al., 2013, 2017; Wlostowski et al., 2013). This selection resulted in lower nRMSE for radon compared to chloride (Fig. 2). Thus, TSM calibration with radon carried more weight in the joint calibration. Therefore, the conclusion that joint calibration of the TSM with radon and chloride yields more information than calibration with either tracer alone is partly influenced by the quality criterion for

parameter sets and, ultimately, by subjective modeling decisions - a well-known challenge in hydrology (e.g., Beven and Binley, 1992).”

7. This reviewer feels that some sort of concise recommendation from the authors would be helpful for the reader. How would this transfer to other streams or watersheds? Low versus high background Radon? This reviewer assumes this tracer approach is limited by river size; perhaps it only really applies to smaller headwater type streams? What are the limitations if streamflow is dynamic? Would this study need to be repeated for other streamflow conditions? A reader would benefit from a concise statement.

   – *Thank you for this important point. We will revise the last part of our discussion section and add a subheader titled “Implications.” In this new section, we will explicitly explain how to apply these findings to other streams. We will also provide detailed recommendations for future research. Please see our answer to comment no. 33.*

Citations look mostly complete and are well thought out. Equations appear correct. Nice work overall.

- *Thank you!*

Specific comments:

8. 14: How do you know a longer estimated timescale would lead to a larger volume? Longer timescale does not necessarily mean more volume.

   – *We will delete this aspect.*

9. 35: consider Schmadel et al. https://doi.org/10.1002/hyp.9854 for supporting definition of WoD.

   - *Thank you for this suggestion.*

10. 58: "slug", suggest to define as a near instantaneous injection of mass.

    - *Thank you for this suggestion. We will define the term "slug":*

      "This limitation is critical, as a growing body of research highlights the presence of flow paths that exceed the duration of experiments using instantaneous tracer injections (hereafter 'slug tracer experiments'; e.g., Ward et al., 2023)."

11. 62: Do not follow what "they" are. Exchange fluxes?

    - *We will replace the term "they" by "these subsurface flow paths" to clarify this aspect.*

12. 67: Sentence needs another look. Redundant information, consider delete.

    - *There is a small grammatical error in the sentence, which we will correct. However, we think this sentence is important here, as it provides a logical bridge between the preceding and following paragraphs.*

13. 69: switching to "duration" adds confusion for reader. You specifically mean the WoD, correct? Suggest to state the WoD for consistence as that is clearly defined. Duration of any slug is infinity; the WoD defines which portion of the slud returns real information.

    - *We agree and will change the term "duration of slug tracer experiments" to WoD.*

14. 84: The goal is…there are several places where qualifying words like "overarching" add to the word count and not needed.

    - *Thank you for this important comment. We will carefully revise the entire manuscript and removed qualifying words from the text.*

15. 150" "derived"? Confusing. Solute transport is simply parameterized. You are not necessarily deriving anything new.

    - *Thank you, we will revise the sentence and write:*

      "Solute transport parameters are commonly determined by calibrating TSMs against measured BTCs."

16. 153: well or a "completely mixed" transient storage zone.

    - *We agree and will revise as suggested.*

17. 162: fine, but mention of the ADE if no storage is not necessary for the reader.

- *Thank you for this suggestion. We will delete the sentence "In case of a lack of exchange where α = 0, eq. 1 reduces to the advection-dispersion equation."*

18. 163: the model formulation was further modified to account for gas exchange. "Not suited" reads awkwardly.

- *We will remove the term "not suited for" and change the sentence to "The model formulation above does not account for key processes affecting radon activity,…"*

19. 253: "certainty" adds confusion for this reviewer. H is entropy, but you are also calling that "certainty." This reviewer has referred to parameter sensitivity and identifiability, but not certainty.

– *We will revise the definition of certainty in the text (see response to comment no. 3).*

20. 300: Runkel and Chapra 1993 should probably be cited here. Also, important to point out for the reader that this is steady state, meaning flow is considered steady for every slug injection. If dynamic, dA/dt + dQ/dx = qi = qout. Also, it is not clear what was needed for the TSM. You only need this equation to estimate qi and plug into the TSM. And what was the Radon concentration assumed in GW? Does qi not matter for salt because the background was subtracted?

– *Thank you for raising this concern. We will state that this applies under steady-state conditions. The revised sentence will read as follows:*

"This is because OTIS, and by extension OTIS-R, accounts for water mass balance under steady state by parameterizing groundwater inflow using the following equation: (…)"

– *Furthermore, we will report the radon activity used for lateral groundwater inflow:*

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

21. 320: Reads as flat. A reader is left wondering what the key result is. Perhaps say that tracer results were ideal for testing this approach because there was a clear difference between surface water and groundwater concentrations. "tracers revealed spatial variability" is not necessarily novel. Suggest to add a concise statement to help the reader regarding what you see in the tracer data. For example, "pairing two tracers allowed for improved identification of parameters because surface water and groundwater are distant in this watershed." Something like that.

- *Thank you for pointing this out. We will rewrite the section as follows:*

"BTCs (Fig. S1) showed a distinct peak concentration at both the upstream and the downstream locations of the study reaches, thereby meeting a key requirement for the calibration of the TSM. Radon activity in groundwater was 23 times higher than in surface water, providing the necessary contrast for quantifying groundwater inflows into the stream. Stream radon activity ranged from 285 (± 22) Bq m^{-3} to 337 (± 26) Bq m^{-3} with the highest activity was observed at reach #2."

22. 334: Increased certainty? You mean improved parameter identifiability? It is not clear how certainty and identifiability are different or same. Same comment for line 348.

   - *We will explain in detail what we mean by certainty (see our response to the comment no. 3)*

23. 360: In the caption, it would help the reader to specify these are two separated calibrations with an “only salt” and “only radon”

   - *We will revise as suggested.*

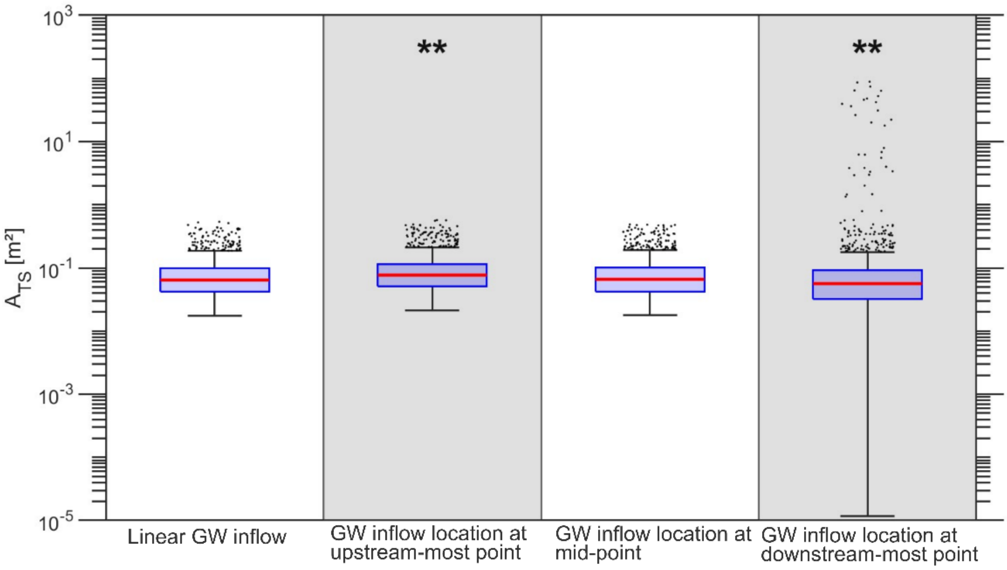
24. 365: unclear what Qfix, QLHS, Qout are directly from the table. So ATS is highly sensitive to Radon tracer?

   - *We will clarify these aspects by revising the table caption. The caption will define the terms 'high', 'moderate', 'poor', and 'non-sensitive' for clarity. For consistency, we will also align the terminology used in the table with that introduced in the Methods section. The revised caption will read as follows:*

     “Table 3: Overview of the sensitivity based on the K-S test for all model parameters (D, α, A_{TS}, and the groundwater inflow (q_l)) from the individual calibration of the TSM with chloride and radon. The terms 'high', 'moderate', 'poor', and 'non-sensitive' refer to the classification of parameter sensitivity based on the K-value and p-value (see Section 2.6), following the methodology of Ouyang et al. (2014). Results from all three calibration approaches are shown here, which differ in how groundwater inflow was calibrated (Q_{fix}, Q_{LHS}, and Q_{out}). Results are presented for all reaches, but only for the 1% behavioral parameters. For simplicity, we show the klow model setup for calibrating the TSM with radon only.”

25. 410: This reviewer does not follow the main point or reason for this Figure 6, and finds this figure confusing. The GW categories are not the same as previous discuss in text. Qfix, QLHS, Qout. What is GW inflow downstream? And why is that shaded grey. The point is a few outliers are caused? Medians and percentiles are nearly the same.

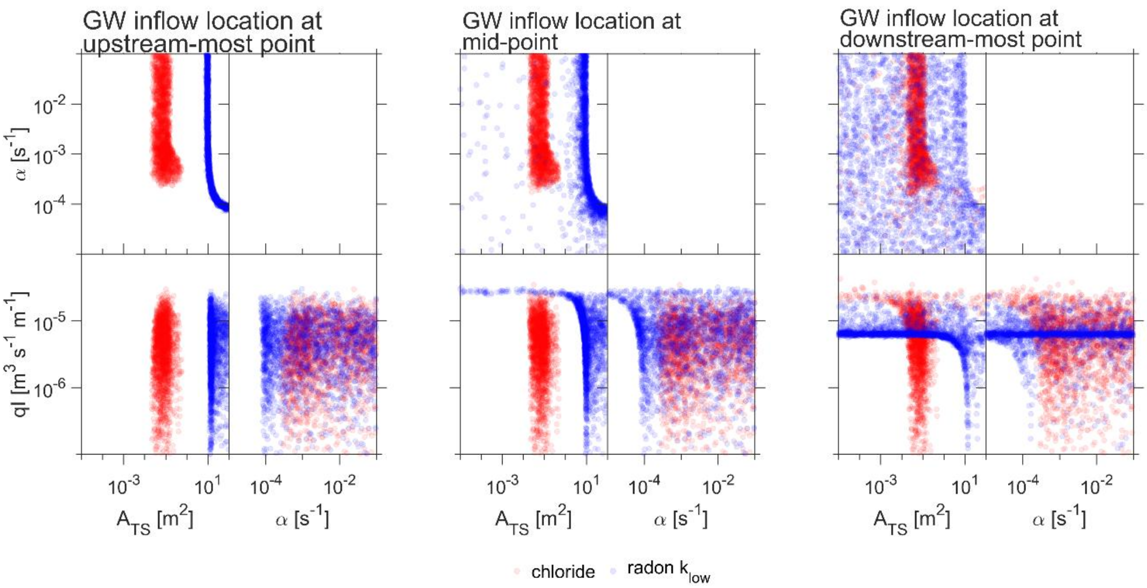
   – *Thank you for this important point. We will use the same terminology as in the previous text for consistency. We will also clarify the purpose of the shaded grey area.*



"Figure 6: Distributions of behavioral ATS values from model setups with varying groundwater inflow locations, each calibrated with chloride only. The model setup labelled 'linear groundwater (GW) inflow' refers to the calibration approach where the groundwater inflow was calculated from the calibrated discharge (QLHS). The red line is the median of the distributions, while black dots highlight outliers. Asterisks and the grey areas show a significant difference between the variance of the parameter distributions compared to the setup with linear groundwater inflow. Results are shown for the best 1% behavioral model parameters and reach #1 only."

26. 415: Figure 7, it is not clear how this is different from Figure 5. New GW terminology adds confusion.

   – *We agree and will change the titles of the sub figures accordingly.*



"Figure 7: Scatter plots of the best 1% behavioral model parameters (A_{TS}, α and the groundwater inflow ql) from calibrating the TSM with radon and chloride alone. Three

different model setups are presented that differ in the location of the groundwater inflow along the reach (upstream-most point, mid-point, downstream-most point model setups). For simplicity, results from calibrating the TSM with radon in the k_{low} model setup are shown only."

27. 435: "Calibrating TSMs with multiple tracers better constrains model parameters." Where are the actual results of calibrating using two tracers simultaneously? You calibrated using an individual tracer, and then calibrate to one set of combine tracer data, correct?

    - *We will explain in detail what we mean by joint and individual calibration (see response to comment no. 1).*

28. 456: "TSM is jointly calibrated with radon and chloride." Did you actually do a joint calibration? To suggest two tracers were used, a multi-objective function should be considered similar to Neilson with solute and temperature. This reviewer does not clearly follow where the joint results are.

    - *Please refer to our previous response, as well as the detailed explanation provided in response to comment no. 1.*

29. 501: "critical for contextualizing" in future radon studies? How? QLHS and Qout results look very similar to this reviewer. It just provides a longer WoD.

    - *Thank you for highlighting this point. We will revise the section and remove the term "critical for contextualizing." Furthermore, we will specify which results we explicitly refer to here to avoid confusion, clarifying that we do not refer to the Q_{LHS}/Q_{out} results. The revised text will read as:*

      "Therefore, our findings emphasize that the spatial variability of groundwater inflow location should be explicitly accounted for in future radon-based studies. This consideration may challenge the common perception that radon mass balances can be fully closed solely by including exchange with subsurface transient storage zones."

30. 505: Statement is not clear from Figure 6. Central tendency is all the same regardless of GW location. A reach-by-reach water balance seems critical. "Location selected for GW" is arbitrary.

    - *Although the behavioral parameter distributions appear similar regardless of groundwater inflow location, our results show that these distributions differ significantly. This point will be emphasized in the main text:*

      "We found significantly different distributions of the behavioral parameters depended on the different locations of groundwater inflow (upstream-most point, mid-point, downstream-most point model setups) independently of which tracer was used for calibration (Fig. 6, Fig. 7)."

31. 519: "Previous studies have shown that large spatial-scale subsurface flow paths play a critical role in explaining water mass balances in streams." Redundant information, and reads too vague and may add confusion for the reader. Why are the large-spatial-scale so important? It is only important for estimating an accurate WoD. Right? Meaning, your paper is about the WoD, not regional GW processes. The point the help the reader is that local groundwater fluxes control the water balance and should be quantified as part of the in-stream BTC. That is explicit and easy to understand. Suggest to write in more explicit terms specific to your study.

- *We will clarify that "large spatial scale subsurface flow paths" do not refer to regional groundwater flow. Rather, they indicate subsurface flow paths originating further upstream in the reach. The revised text will read as:*

  "Previous studies have shown that large spatial-scale subsurface flow paths, that are subsurface flow paths originating from further upstream of the stream reach, play a critical role in explaining water mass balances in streams (e.g., Payn et al., 2009; Stanford and Ward, 1993; Ward et al., 2023)."

32. 540: Figure 9, should state that this figure has been modified from Payne et al. 2009. Flow path C does not seem technically correct. You do not know if that represents tracer that bypasses measurement site. The WoD is your measurement window, not the real window in which flow paths exist. And why is C red. That adds confusion. This reviewer does not find any new information added by this figure, and it does not seem technically correct.

    - *Thanks for pointing this out. We will rewrite the section with an emphasis on the novelty of labelling this flow path with radon in our study. Please see our answer to comment no. 33.*

33. 549: Just say break the transient storage zone into two zones, each with its own exchange flux. This is important for temperature and reactive solutes. It might be critical? Too vague. And critical how? How would this help with Radon? Discussion related to two-zones is a little weak; sure, denitrification representation might improve, but how is that relevant for radon and chloride?

    - *We will revise the discussion on the two-storage zone TSM and relocate it to a new section titled "Implications." The revised text will read as:*

      "The improved parameter identifiability through joint TSM calibration with radon and chloride may be a critical step toward enhancing the physical realism of TSMs and providing more reliable estimates of solute transport. This is important because past studies found that the relationships between TSM parameters and hydrologic drivers are often contradictory (Ward and Packman, 2019; Bonanno et al., 2022). We envision that future studies will derive model parameters by jointly calibrating TSMs across diverse environments and hydrologic conditions. This approach can help clarify how model parameters vary with hydrologic drivers. To advance this goal, studies should compare streams across varying scales and geological settings to determine how these differences affect calibration outcomes when radon is included. Through such comparative studies, research could test our expectation that streams with higher radon activity, typically smaller-order streams with granite-rich geology, will yield greater certainty for groundwater inflow q_l and improved model performance when TSMs are jointly calibrated with radon and chloride. This greater certainty for q_l might result from the smaller greater difference in radon concentration activity between groundwater and stream water. Furthermore, future research could jointly calibrate TSMs with radon and tracer data from constant rate injections. This approach would test whether radon captures longer flow paths than those detected by constant rate injections but not by slug tracer injections.
      For future research with radon on gross exchange fluxes, we recommend constraining the transient storage parameter by considering surface (e.g., pools) and subsurface transient storage zones (e.g., hyporheic zones) in TSM evaluations (e.g., Choi et al., 2000). This distinction might be critical because only surface storage contributes to radon degassing. In contrast, radon activity increases mainly in the subsurface storage and, to a lesser extent, in surface storage. These storage-specific processes are not fully captured by using a single storage zone, as implemented in

OTIS-R, instead of two. 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.
(…)"

34. 563: "The goal was.."

- *We will remove the term "overarching", thanks for pointing out.*

35. 564: Radon is a solute too. Be specific here regarding which solute, which tracer.

- *We will change to:*

  "The goal of this study was to quantify flow paths of different timescales at the reach scale using measurements of chloride concentration and naturally occurring radon."

36. 571: Not sure this reviewer agrees that the recommendation should always be to include Radon and chloride jointly. What is missed if Radon is excluded? Can we estimate qi from groundwater first, and then use salt once that qi parameter is set? What is Radon is non-detect? Are you assuming Radon is measurable in all streams? When longer flow paths need to be identified, sure, Radon makes sense, but there is still large uncertainty in the parameter value. Plus, a one zone may be much more of a limitation than excluding Radon. Suggest to state in simple terms that Radon can extend the WoD with paired with salt, plain and simple.

- *Thank you for your suggestions. We will add a section titled "Implications and future research" to the discussion to address your recommendations. Please refer to our detailed response to comment no. 33 for more information.*