

Authors' responses to the comments of reviewer#1

We appreciate your review and comments on our manuscript, “Deriving Subsurface Stormflow Mechanisms with Isotopes and ERT”. Your feedback is valuable to us, and we will make the recommended revisions accordingly. We provide detailed responses to each of your comments below.

This is an interesting paper with a very nice experimental setup. Combining a sprinkling experiment, deuterium tracer, trench flow, groundwater wells, soil cores, and ERT gives the authors a strong data set. The main idea is valuable, and I think the paper can make a useful contribution to understanding subsurface stormflow generation.

However, the manuscript needs more work in organization and interpretation. In several places, the authors present interpretations as if they were directly observed. One important example is the distinction between “unlabeled water” and “pre-event water.” Because the first half of the sprinkling experiment used unlabeled stream water, the authors need to show clearly how different this water was from pre-event trench water and whether the two could be separated isotopically. If their isotope compositions were similar, then the isotope data would mainly distinguish labeled water from unlabeled water, rather than event water from pre-event water.

This does not weaken the experiment, but it does mean that the wording around “pre-event water” should be more careful. The paper would also benefit from a clearer separation between results and interpretation. Some mechanisms, such as piston displacement, preferential flow, and matrix storage, are plausible but they are still interpretations. The authors should make this clearer throughout the abstract, discussion, and conclusions.

Line by line Comments

Introduction

The introduction needs some organization. It is hard to follow some of the ideas because many concepts are introduced one after another without enough transition. Reworking the introduction could greatly improve the manuscript and help the reader understand why this experiment is important.

L36-42: The discussion of isotopes and the old-water paradox is important, but the transition from preferential flow to old water feels abrupt. The authors should add one or two sentences explaining why preferential flow and old-water dominance can happen at the same time.

We will add a better transition and explain the co-occurrence of preferential flow and old water discharge.

L43-50: This paragraph about difficulties with natural rainfall experiments needs a better connection with the paragraph above. It currently reads as a new topic. The authors could first say that one reason the mechanism remains difficult to identify is that natural rainfall inputs are hard to constrain.

Thank you for the suggested transition, we will change it to this in the revision.

L56: The transition into ERT feels disconnected. The paper moves from controlled irrigation to direct observation of flow pathways very quickly. I suggest adding a sentence explaining why isotope data alone are insufficient.

We will add a smoother transition from sprinkling to ERT

L68-74: The limitations of ERT may fit better in the Methods or Discussion. The introduction is already quite long, and this part slows down the motivation of the paper.

We will move this to discussion.

L80: The main knowledge gap should appear earlier. The mechanism by which event water triggers release of stored water is the central motivation, and it should be introduced near the beginning of the introduction.

We will state this knowledge gap earlier when introducing the flow pathways of SSF.

L82: The third research question seems partly answered already in the introduction, where the authors explain that isotopes and ERT are complementary. I suggest reformulating Q3. Maybe something like: “To what extent can the combination of isotopic tracers and ERT improve the interpretation of SSF generation mechanisms?”

Thank you for this suggestion, we will change the third question.

Methods

L100-110: The trenches are described here, and then section 2.2.1 describes the trenches again. Consider merging these parts or reducing repetition.

We will merge the descriptions.

The trench setup is very important for the paper, but it is still difficult to understand exactly what each trench captures. A more detailed schematic would help. Figure 1 is currently somewhat rough (it literally says “rough sketch”). The authors should replace this with a cleaner figure showing soil horizons, trench layers, bedrock/refusal depth, spring location at Trench 1, and the collection depths.

We will add a schematic of the Trench setup to the paper and mark the location of the spring in Trench 1. The current Fig 1 was only intended for soil horizon visualization.

The sprinkling experiment needs more information about spatial uniformity. Was rainfall distribution measured with collectors? How variable was sprinkling intensity across the plot? This matters because the authors later interpret spatial differences in ERT and soil tracer storage.

Sprinkling uniformity was assessed in another paper (<https://doi.org/10.22541/essoar.15006118/v1>), however only in a lab setting. The lab test yielded a Christiansen's uniformity coefficient of 88%, indicating good spatial uniformity of the applied water. In real-world conditions, branches and tree stumps will have lowered uniformity. We will add this uncertainty to the discussion.

The isotope design should be explained more carefully. Because the first 1.5 hours of sprinkling used unlabeled stream water, the isotope data clearly distinguish labeled water from unlabeled water, but may not fully distinguish pre-event water from the initial unlabeled irrigation water if their isotope compositions were similar. The authors should state the isotope composition of the initial sprinkling water and pre-event trench water, and clarify how this affects the interpretation of “pre-event” fractions.

Since both waters had a similar isotopic composition, we can only differentiate between labelled and unlabeled water, and we will make this clearer in the revision.

L180-185: The isotope mixing model section needs more detail. What isotope values were used for each end member? Were source uncertainties included? What priors and MCMC settings were used in simmr? How was convergence checked?

We will include the source end member values used. We did not assign any uncertainties. For d2H we fit a two-component (labelled/unlabelled) mixing model in simmr under the mixing relation $C_s = f_E C_{\text{label}} + (1 - f_E) C_{\text{unlabeled}}$, and report the posterior f_E ; simmr serves as a single-tracer two-source solver, not as a multivariate model. We will clarify this in the revision.

L187-192: The storage estimate in the upper 20 cm is based on strong assumptions, e.g., full saturation, homogeneous porosity. However, the authors collected very few soil cores. This should be presented as an approximate estimate, not as a precise value.

We will revise this and add an uncertainty assessment of different saturation and porosities, and also state the values as rough estimates

L193-204: The ERT inversion description is useful, but the choice of regularization parameters needs more justification. Saying that values were selected after “several tests” is not enough for reproducibility.

The regularization parameters were chosen given that the misfit chi2 were close to 1. More details on how they were applied will be added.

L210-211: The 3 %, 5 %, and 10 % thresholds need a clearer explanation. Why is 3 % considered meaningful given measurement and inversion uncertainty? A noise or repeatability test would help.

In the chapter of Methodological uncertainties, we mention that the interpreted waterfront with a 3 % threshold was selected because using values of 5 % and 10 % would only show small, localized changes in resistivity after 24 h and only the 3 % threshold captured a spatially consistent signal in Trench 1. Other authors had used 5% and 10% thresholds, however such a value would have masked the decrease observed the day following the experiment, since most of the resistivity changes fell below the 10 % level. We therefore justified a lower threshold in our case since it preserved sensitivity to the response we sought to characterize, whereas a more conservative cut-off would have suppressed a physically meaningful feature of our data. Additionally, in Trench 1 we have one ERT profile measured 13.9 h before the baseline measurement, and resistivity variations during this period reached maximum values of almost 1% and mean values of around 0.3% in the first 50 cm, decreasing with depth. Which shows that the natural variations were quite small.

Results

The results are generally interesting, but some parts already include interpretation. I suggest keeping the Results more descriptive and saving mechanism language for the Discussion.

L214-228: The description of the SSF response is good. However, “event runoff coefficient” should be used carefully. The trenches may not capture all water, and later the authors discuss possible deep or lateral bypass. Maybe call this “trench-recovered runoff coefficient” or define the limitation clearly.

We will clarify the calculation method of the coefficient in the revision

L249-263: The tracer breakthrough is one of the strongest results. However, the authors report tracer arrival within 7-18 minutes, while samples were collected every 20 minutes. Please clarify how these arrival times were calculated.

We collected samples every 20 Minutes, but measured SSF every 10 minutes, so we linearly interpolated Isotope concentrations to 10 minutes. That's why time differences are always in 10 Minute intervals. Additionally, we calculated the length and volume of the hoses connecting the water pools to the sprinklers. They were still filled with unlabeled water at time of pool switching. It then took for instance 3 Minutes for the labelled water to reach the sprinklers, which is why uneven time numbers can occur. We will clarify this in the revision.

L265-272: The sample size for the soil cores is small. The conclusion that tracer was only stored in the upper 20 cm should be limited to the sampled locations. The authors should avoid implying that the entire plot was characterized by only three cores.

We will tone down our statement and state more uncertainty.

L271: The estimate that 45 % of labeled water was stored in the topsoil seems too precise. It depends heavily on porosity, saturation, and spatial representativeness. Please add uncertainty or present this as a rough mass-balance estimate.

As above, we will add that this is a rough mass balance.

L274-323: The ERT results are detailed, but the section is hard to follow. There are many percentages, locations, and times. A summary table or simplified interpretation figure would help a lot.

The ERT results will be summarized.

Discussion

The wording on the dual-domain interpretation is too strong in places. The data support rapid movement of labeled water through preferential pathways and storage of some labeled water in the shallow matrix. However, the statement that most SSF was pre-event water is less certain because unlabeled irrigation water cannot be separated from pre-event water.

We will tone down this statement in the revision.

L349-356: This is a strong paragraph and one of the best parts of the paper. The contrast between tracer in deep trench outflow and no tracer in deeper soil cores is convincing evidence for bypass flow. However, the authors should still acknowledge that the soil cores are spatially limited.

Yes, we will acknowledge this limitation in the revision.

L357-364: The manuscript says the tracer budget shows preferential flow carried only a small fraction of applied water. This is probably true, but recovery depends on trench capture efficiency and storage estimates. Please make the uncertainty more visible.

We will state this uncertainty clearer.

L365-370: The piston-flow interpretation is plausible, but ERT non-response in the deeper saturated zone does not by itself prove piston displacement. It could also reflect low ERT

sensitivity under saturated conditions or low contrast between waters. The authors discuss this later, but the interpretation here should be softer.

We will soften the interpretation and state that this is limited due to low contrast.

L399-420: The “unaccounted labeled water” section is useful and should maybe be emphasized more. It shows that the mass balance is still uncertain. This uncertainty should be connected back to the main conclusions.

With the revision stating clearer the uncertainties associated with the mass balance, we will also update this section.

L421-463: The methodological uncertainty section is good. I think some of this should appear earlier because it affects how the results are interpreted. The issue of unlabeled irrigation water versus pre-event water is not a small limitation, it is central to the main claim.

As stated above, we will incorporate them above.

Conclusions

The conclusions should be rewritten with more careful wording. The study shows that labeled event water reached the trenches quickly and that much of the labeled water was retained or delayed. It also supports a dual-domain interpretation. But the authors should avoid saying that the majority of SSF was definitively pre-event water unless they can better account for the unlabeled irrigation water.

Like above, we will adapt this.

The final suggestion about using multiple tracers, including a conductive tracer detectable by ERT, is very good. This is one of the clearest take-home messages from the study and could be highlighted more.

Yes, we are currently doing this experiment and see clear improvements in the detection of flow in the “saturated” zone.

Minor comments

L9-10: The statement that 45 % of labeled water remained stored in the upper 20 cm should include uncertainty or be softened.

We will change this.

L17: “confirming” is too strong. Maybe use “suggesting” or “supporting.”

We will adapt in the revision.

L84-85: “gives insights of where” should be “provides insight into where.”

We will fix this.

Figure 5: Add more explanation of how the reference profile is used and whether the upper 5 cm were discarded.

We will make this clearer in the methods and adapt the graphic

Figures 6-7: Consider adding depth-of-investigation or sensitivity masks for ERT results.

It is going to be considered. Most lines, have a zoom-in on the infiltration area, and it is within the DOI.

L399: "According to the retrieved tracer data" could be clearer. Maybe "Based on the tracer recovery and storage estimates..."

We will change this.

L412-419: The discussion of bedrock uptake and lateral flow beneath the trench is important, but it also means the trench water balance is incomplete. This should be stated more explicitly.

We will adapt in the revision.

Data availability: The authors mention data and supplementary ERT data, but code for isotope mixing and ERT processing could also be made available if possible.

Yes, we will make this available.

Questions for the authors

How different were the isotope values of the initial stream water, pre-event trench water, and labeled water?

We will include this in the revision. Stream and trench were extremely similar, but the label strongly differed from the natural conditions (220 permille)

How were breakthrough times shorter than the sampling interval estimated?

See answer above.

Was sprinkler uniformity measured across the 200 m² plots?

See answer above.

How sensitive is the 45 % storage estimate to porosity and saturation assumptions?

We will include an assessment in the revision.

Could the first unlabeled sprinkling phase have changed the flow pathways before the labeled water was applied?

The first phase probably only wetted up the soil so the labelled water could infiltrate and flow, without being directly bound to the matrix.

How much confidence do the authors have that the lower trench discharge represents water from within the sprinkled area rather than lateral inflow from outside it?

Since we saw a strong flow increase related to baseline, and had no rain during the experiment, so it must stem from the sprinkled area.

Overall, I think this is a strong and interesting experiment, but the interpretation needs to be tightened. The paper will be much more convincing if the authors are clearer about what is directly observed, what is inferred, and where uncertainty remains.

Thank you!