RESPONSE TO REFEREE COMMENTS 1

General comments

The paper presents flow-through experiments on sand packs, with inert solute and heat used as tracers. Experiments are complemented with micro-CT imaging and analyses of the samples' microstructure to extract measures of pore-scale heterogeneity. Key findings include: (1) more heterogeneous samples (sand mixture, SM) exhibit more anomalous signatures in the data, compared to homogeneous samples (S4); (2) the two tracers are not interchangeable, since the outcomes they produce differ quantitatively and qualitatively; and (3) better structural descriptors than those used herein (Euler number and smoothness parameter) are needed to link the micro and macro scales. Overall, the manuscript is clearly written and the experiments are valuable. That said, a few larger questions stand out that the authors might benefit from considering.

Authors' reply: We sincerely appreciate the referee's positive assessment of our manuscript and the insightful comments. The comments raise important questions regarding the interpretation of pore-scale processes and their representation at the Darcy scale. We will carefully address these points in the revised manuscript. Our point-by-point responses are provided below.

Specific comments

(1) The authors fit the ADE to measured breakthrough profiles, obtain the fitted tracer velocities, and find that these underestimate the Darcy velocities measured directly in the experiments. The retardation was also noted to be more pronounced for the heat tracer, and to increase nonlinearly with injection rate. I think this is expected, given the form of the ADE assumed, which contains no budget for tracer that might become immobile, or semi-mobile, by diffusing either into dead-end zones or into the grains themselves (in the case of heat). If such budgeting had been included in the governing Eq. 1, through a dual-medium representation, I suspect the fitted velocities would indeed match those measured in the experiments -- e.g., velocity could be constrained to the experimental value while the fit focuses on determining the mobile-immobile transfer coefficient. Such a coefficient would also explain the nonlinear increase in retardation for heat at high injection rates, which likely also breaks the model's local equilibrium assumption. As the authors correctly note, a good fit ensures consistency but not that the right model was chosen. I encourage discussion on why Eq. 1, rather than a dual-medium model, was chosen, whether a dual-medium model would make these observations less surprising, and whether that would not constitute the "upscaling" avenue they advocate.

Authors' reply: We thank the reviewer for this insightful comment. The observation that a dual-medium (mobile–immobile) formulation could absorb the apparent retardation into an exchange coefficient - and thereby recover the independently measured Darcy velocity - is well taken, and we agree that it represents a physically meaningful interpretive framework. We address the three specific questions raised below.

(1) Why was Eq. 1 (single-domain ADE) chosen rather than a dual-medium model? The single-domain ADE was deliberately selected because it is the standard continuum framework employed in virtually all field and laboratory heat- and solute-tracer interpretations (Anderson, 2005; Rau et al., 2014; Irvine et al., 2015). Our objective was not to identify the best-fitting model but to quantify the *systematic bias* that this conventional framework introduces when applied to media of increasing microstructural complexity. By

holding the model structure fixed and varying only the pore-network geometry (S4 → S5 → SM), we isolate how much of the observed misfit is attributable to unresolved microstructure rather than to model misspecification. Had we adopted a dual-medium model from the outset, the additional exchange parameter(s) would likely have absorbed much of the retardation signal, rendering the bias invisible - which is precisely the diagnostic failure we seek to expose. We will make this motivation explicit in the revised manuscript.

(2) Would a dual-medium model make the observations less surprising? Yes, in a parametric sense: introducing a mobile–immobile mass-transfer coefficient (or, for heat, a finite interphase heat-transfer coefficient) would allow the fitted front velocity to converge toward the measured Darcy flux, and the velocity dependence of the apparent retardation could be captured by a Péclet- or Damköhler-number scaling of the exchange term (e.g., Haggerty and Gorelick, 1995; Heinze and Hamidi, 2017). We fully acknowledge this and will state it clearly in the revised Discussion. However, we note two important caveats. First, the excellent goodness-of-fit already achieved by the single-domain ADE (mean $R^2 > 0.99$, NRMSE < 0.037) means that a dual-medium model would improve *physical interpretability* without substantially improving *statistical fit*; the bias we report is therefore not a fitting artefact but a consequence of forcing a structurally incomplete model to reproduce data it can mimic but not mechanistically represent. Second, similar Darcy-scale signatures (non-Gaussian breakthrough, velocity-dependent retardation) can arise from several distinct pore-scale processes - including exchange with dead-end pores, heterogeneous velocity distributions, preferential pathway formation, and diffusion-controlled trapping (van Genuchten and Wierenga, 1976; Haggerty and Gorelick, 1995; Dentz et al., 2023) - as well as from finite fluid–solid heat exchange and LTNE (Heinze and Hamidi, 2017; Lee et al., 2025). A dual-medium parameterisation would lump these mechanisms into a single effective coefficient without distinguishing among them. Recent work has further shown that heterogeneous velocity fields coupled with molecular diffusion can generate non-Fickian travel-time distributions without requiring an *a priori* dual-domain decomposition (Zehe et al., 2026). We will therefore present the dual-medium interpretation as one plausible mechanistic explanation - arguably the most parsimonious - while noting that it does not uniquely resolve the underlying pore-scale physics.

(3) Does the dual-medium framework constitute the "upscaling" avenue the authors advocate? We agree with the reviewer that a dual-medium (or, more generally, multi-continuum) formulation is a natural and promising candidate for the process-aware upscaling we call for. In the revised Discussion we will explicitly position the mobile–immobile exchange coefficient (and its thermal analogue, the interphase heat-transfer coefficient) as effective Darcy-scale parameters that could, in principle, be linked to measurable microstructural quantities such as pore-size distribution skewness, Euler number, and REV stability - precisely the metrics we characterise via micro-CT. In this sense, the present study provides the *diagnostic evidence* (i.e., where and how the single-domain ADE fails) that motivates and constrains such an upscaling effort, while the *prescriptive step* - deriving physically meaningful exchange coefficients from pore-scale topology - requires dedicated model-comparison, sensitivity, and parameter-identifiability analyses that are beyond the scope of the current experimental dataset. We will frame this distinction clearly and outline it as a concrete direction for future work.

We will revise Section 2.2 to clarify the rationale for adopting the single-domain ADE, expand **Section 4** to include a dedicated paragraph on the dual-medium interpretation and its relationship to the observed bias patterns, and add the suggested references. We thank the reviewer for prompting this important clarification.

References:

Anderson, M. P. (2005). Heat as a ground water tracer. *Groundwater*, *43*(6), 951-968.

Dentz, M., Hidalgo, J. J., & Lester, D. (2023). Mixing in Porous Media: Concepts and Approaches Across Scales: M. Dentz et al. *Transport in Porous Media*, *146*(1), 5-53.

Haggerty, R., & Gorelick, S. M. (1995). Multiple-rate mass transfer for modeling diffusion and surface reactions in media with pore-scale heterogeneity. *Water Resources Research*, *31*(10), 2383-2400.

Heinze, T., & Hamidi, S. (2017). Heat transfer and parameterization in local thermal non-equilibrium for dual porosity continua. *Applied Thermal Engineering*, *114*, 645-652.

Irvine, D. J., Simmons, C. T., Werner, A. D., & Graf, T. (2015). Heat and solute tracers: How do they compare in heterogeneous aquifers?. *Groundwater*, *53*(S1), 10-20.

Lee, H., Gebhardt, H., Blum, P., Bayer, P., & Rau, G. C. (2025). Numerical analysis of local thermal non-equilibrium experiments reveals conceptual regimes of grain-scale heat transport. *Water Resources Research*, *61*(11), e2025WR041260.

Rau, G. C., Andersen, M. S., McCallum, A. M., Roshan, H., & Acworth, R. I. (2014). Heat as a tracer to quantify water flow in near-surface sediments. *Earth-Science Reviews*, *129*, 40-58.

Van Genuchten, M. T., & Wierenga, P. J. (1976). Mass transfer studies in sorbing porous media I. Analytical solutions. *Soil science society of america journal*, *40*(4), 473-480.

Zehe, E., Pfister, L., Elhanati, D., & Berkowitz, B. (2026). On the relevance of molecular diffusion for travel time distributions inferred from different water isotopes. *Hydrology and Earth System Sciences*, *30*(7), 2093-2106.

(2) The metrics chosen for quantifying pore-scale heterogeneity must have been selected on the rationale that they bear the strongest relationship to the measurements made. That rationale was not articulated. Concretely: why the Euler number, and why the smoothness parameter? The Euler number is a topological measure of the porous medium, which tells one how connected the pore space is. But it says nothing about velocity variations (Line 524). Here is a thought experiment: pick a set of pores/nodes along a straight line as the porous medium. Now connect adjacent pores with curved pipes of identical radius and length that leave the originating pore at angle theta with respect to the pore-to-pore line, and enter the target pore at the same angle. In 3D, one can add an arbitrary number of pipes this way without changing the velocity variation (it remains a Dirac delta, since all pipes have the same resistance and the same pressure drop), while increasing connectivity (via coordination number) arbitrarily, driving the Euler number as negative as one pleases. Similarly, a real porous medium can be highly connected (e.g., a high-porosity fibrous medium) while its velocity variation is very small. The smoothness/roughness parameter was also a surprising choice of metric, which the authors eventually found does not exert primary control on the results (Line 595). Why was it expected to in the first place, over other factors such as the mobile/immobile conceptualization?

Authors' reply: We thank the reviewer for this penetrating critique. The thought experiment regarding the Euler number is mathematically elegant and correctly highlights that topological connectivity, in isolation, does not dictate velocity variance. We also agree that our initial rationale for selecting the Euler number and the Matérn smoothness parameter was not sufficiently articulated. We will substantially revise the manuscript to clarify our reasoning, temper our interpretations, and explicitly distinguish between static structural metrics and dynamic process conceptualizations. We address the specific points below.

(1) Rationale for the Euler number and the thought experiment: We selected the 3D Euler number as a topological metric because, in natural granular porous media, it serves as a primary indicator of pore-network

connectedness, percolation clusters, and the prevalence of dead-end pores (Vogel and Roth, 2001; Pires et al., 2020). The reviewer's thought experiment - where connectivity is increased via identical curved pipes without altering the velocity distribution - is entirely valid for an idealized, uniform network. However, in the natural, polydisperse granular media we investigated (specifically the Sand Mixture, SM), topological complexity and pore-size heterogeneity are intrinsically coupled. The highly negative Euler number in SM (−9,210 vs. −6,288 for S4) arises directly from the packing of diverse grain sizes, which simultaneously generates a right-skewed pore-size distribution (skewness = 0.781) and a broad distribution of pore-throat radii. It is this concomitant geometric heterogeneity - rather than the topological connectivity in isolation - that creates the preferential pathways and stagnant zones driving the observed velocity variations and non-Gaussian transport. We chose the Euler number as a computationally efficient, static proxy for this structural complexity, but we agree it is not a direct mathematical measure of velocity variance. We will revise Section 2.4 and the Discussion to explicitly state this limitation and avoid implying a direct causal link between the Euler number and local velocity variations.

(2) Rationale for the Matérn smoothness parameter: The Matérn smoothness parameter (ν) was selected based on recent geostatistical literature (e.g., Di Palma et al., 2017) suggesting that the short-range spatial continuity (or "roughness") of the pore-solid interface influences mechanical dispersion and boundary-layer dynamics. We hypothesized that a rougher interface (lower ν) might disrupt solid-phase conduction pathways and fluid-solid heat exchange, thereby amplifying Local Thermal Non-Equilibrium (LTNE) for the heat tracer. The fact that νfailed to distinguish between S4 and SM was a genuine, and somewhat surprising, empirical finding. Rather than omitting it, we reported it to demonstrate that standard geostatistical metrics, while useful for stochastic reconstruction, may lack the physical specificity required to predict tracer-specific biases like LTNE. We will clarify this initial hypothesis and the resulting empirical insight in the revised manuscript.

(3) Structural metrics vs. the mobile/immobile conceptualization: The reviewer rightly asks why we expected these structural metrics to exert control over the results, rather than factors like a mobile/immobile (MIM) conceptualization. This question strikes at the core of our upscaling objective. The MIM framework is a *dynamic, process-based parameterization* that requires fitting transport models to breakthrough curves or running expensive pore-scale flow simulations. Our objective was to determine whether *static, purely geometric metrics* derived directly from micro-CT images could serve as *a priori* predictors for when such dynamic models (like MIM or LTNE formulations) are actually necessary. We hypothesized that extreme topological complexity and interface roughness would act as structural "red flags" for non-Fickian transport. Our results show that while the Euler number and pore-size skewness successfully flagged the structural potential for complex transport, the Matérn smoothness parameter did not.

Revisions to the manuscript: To address these points, we propose to implement the following revisions:

- **Methods (Sect. 2.4):** We will explicitly articulate the rationale for choosing the Euler number (as a proxy for dead-end potential and connectedness in granular media) and the Matérn parameter (as a test of interface roughness effects on LTNE), while clearly stating their limitations regarding direct velocity variance.
- **Discussion (Sect. 4.4):** We will incorporate the logic of the reviewer's thought experiment to clarify that topological connectivity must be interpreted alongside pore-size distribution (skewness) to infer velocity variations in natural media.
- **Future Work (Conclusion):** We will expand the conclusion to emphasize that while static CT metrics can identify structural heterogeneity, bridging the gap to Darcy-scale transport ultimately requires process-based descriptors. We will explicitly incorporate the reviewer's suggested alternatives, such as pore-scale

velocity distributions, stagnant-domain fractions (Bijeljic et al., 2013), hydraulic tortuosity (Ghanbarian et al., 2013), and pore-throat conductance (Miao et al., 2017), as the necessary next steps for rigorous pore-to-Darcy upscaling.

References:

Bijeljic, B., Raeini, A., Mostaghimi, P., & Blunt, M. J. (2013). Predictions of non-Fickian solute transport in different classes of porous media using direct simulation on pore-scale images. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *87*(1), 013011.

Di Palma, P. R., Guyennon, N., Heße, F., & Romano, E. (2017). Porous media flux sensitivity to pore-scale geostatistics: a bottom-up approach. *Advances in Water Resources*, *102*, 99-110.

Ghanbarian, B., Hunt, A. G., Sahimi, M., Ewing, R. P., & Skinner, T. E. (2013). Percolation theory generates a physically based description of tortuosity in saturated and unsaturated porous media. *Soil Science Society of America Journal*, *77*(6), 1920-1929.

Miao, X., Gerke, K. M., & Sizonenko, T. O. (2017). A new way to parameterize hydraulic conductances of pore elements: A step towards creating pore-networks without pore shape simplifications. *Advances in water resources*, *105*, 162-172.

Pires, L. F., Auler, A. C., Roque, W. L., & Mooney, S. J. (2020). X-ray microtomography analysis of soil pore structure dynamics under wetting and drying cycles. *Geoderma*, *362*, 114103.

Vogel, H. J., & Roth, K. (2001). Quantitative morphology and network representation of soil pore structure. *Advances in water resources*, *24*(3-4), 233-242.

Technical corrections:

(1) I would rephrase "urgently" in Line 603. It somewhat overstates the need.

Authors' reply: We agree that the term "urgently" may overstate the intended message. We will revise the wording to a more neutral expression that better reflects the scope and implications of the present study.

(2) Some words are split across lines, e.g., "relationship" in Line 335.

Authors' reply: Thank you for pointing this out. We will carefully check the manuscript for unintended word splitting and formatting issues and correct them throughout the revised version.

(3) All figures are clear, but a bit busy and blurry. For example, the histograms in Fig. 7 might look cleaner as box plots, and Fig. 8 is a bit cluttered. Perhaps panels c-j could be moved to an appendix if they support, but do not directly serve, the key takeaway.

Authors' reply: We appreciate this helpful suggestion and agree that the readability of several figures can be improved. We will revise the figures to increase their resolution and reduce unnecessary visual complexity. For Fig. 7, we will reconsider the current histogram-based presentation and evaluate whether a box-plot representation provides a clearer summary of the data. For Fig. 8, we will simplify the layout and assess whether the supporting panels (c–j) can be moved to the Appendix or Supplement while retaining the panels most directly related to the main message in the manuscript.