Nogueira et al presented a modeling study for estimating the water sources and mixing spots in a riparian zone. The authors developed a SW-GW coupling model and applied it in a 4th-order stream, then compared against hydrochemical monitoring data. With the modeling quantification, the authors show that discharge events increase the mixing and present that their tool can serve as a complementary approach for hotspot identification. I feel this paper is novel and well-written so I encourage publications after some of my comments are resolved.

We thank the reviewer for the constructive comments on our work, which have helped to improve our manuscript. We also recognize that some points were not clear and may have been confusing. We have revised the manuscript to clarify those points based on the comments provided by both reviewers. Below we provide our detailed responses to the reviewer's comments.

L92: The authors mentioned other options of models such as ParFlow, and others. To my knowledge I think HydroGeoSphere is a commercial model and is not open-source - Why would you use it? I noticed that the code development paper is in press on WRR, and I'm wondering how much reproducibility of this work from a commercial model? For example, did the authors reuse codes or functions from HydroGeoSphere? Is this work of code development going to be publicly available?

The reviewer is correct, HGS is a commercial software, which initially started as a research code. Hence the routines used in the code are scientifically established and have been widely documented in the scientific literature (e.g., Therrien and Sudicky 1996, Jones et al. 2008, Unger et al. 2008, Park et al 2009, Brunner and Simmons 2011). We chose HGS as a code due our long-term experience with the code, which dates back to earlier, pre-commercial times of initial model development, the numerical efficiency of the code and its ease of use as well as the codes direct interface to the HMC model. For a detailed implementation of the HMC code in the HGS, we refer to Partington et al. (2011). Although HMC can be easily used in HGS the use of this method is not exclusive to HGS. It could be used together with any other numerical code that provides an explicit water balance at every cell in the model domain (Partington et al., 2011), such as MODFLOW for example. With our study we seek to demonstrate the applicability and usefulness of the HMC method in mapping water sources in a riparian zone, regardless of the software employed for the flow simulations. Our proposed methodology involving the HMC routines could be easily adapted to and employed in other non-commercial groundwater flow software.

L94: I think here should follow up with a few particle tracking paper introduction to say that particle tracking is the method without using the solute transport.

We appreciate this suggestion. We have added some references in the respective sentence to better exemplify this concept.

L247: Is this equation 4) defined by the authors or from other literatures? If the former, a more detailed explanation may be needed. If the latter, a citation will be needed.

We apologize for the confusion. The equation is based on the equation presented by Berezowski et al., 2019 (cited on L242). We have modified the sentence to clarify this. We have also added a figure to the supplementary material to facilitate a better understanding of the concept (see now Fig.S2 – also attached at the end of this document).

In brief, for a three end-member mixing case, any combination of three different source water fractions can be represented as a point *d* in a 3D coordinate space, in which the maximum Euclidian distance between point d and the point of equal mixing (equal fractions of all mixing members) within the mixing space is the radius of a circle (centred at [1/3, 1/3, 1/3]) escribed on an equilateral triangle (side length of $\sqrt{2}$). For a two end-member mixing case, the maximum segment is the diameter of a circle (centred at [0.5, 0.5]) with side length of $\sqrt{2}/2$. For a four (or more) end-member mixing case a spatial representation is not possible but equation (4) would equally apply. Therefore we would like to keep it in the text.

L297: Is Schmadel et al (2016) a study at the same research region? Would the discharge event affect the calculation? I feel it is not a direct comparison between this modeling result with other studies, unless more justification is needed. Also, what is "average losses" and "simulated losses" being defined?

Yes, Schmadel et al. (2016) performed their study at the same reach that we have performed our simulations at. With "average losses" we refer to the net water losses computed from discharge measurements at two different points along the reach (e.g., upstream and downstream in the domain), as similarly carried out by Schmadel et al. (2016).

We believe discharge events could affect this calculation because the reach could present slightly different net losses during some events in comparison to baseflow conditions, for instance, due to different hydraulic gradients between SW-GW, which result in different SW-GW exchanges. By using the water balance resulting from our simulations, we can compare our simulated water losses with their calculated value. We apologize for the confusion; we have now rewritten the sentence and changed the terms to water losses in order to better explain what we are referring to. It now reads:

> *The stream reach was characterized by predominantly losing conditions with average net water losses to the subsurface of around 40-50% of total discharge. This is higher than the 25% measured in the field by Schmadel et al. (2016) during a small discharge event in July 2014; however, our simulated net water losses for the same period of their analyses were around 30%, indicating a good match to observed reach conditions during the discharge event.*

L350: I'm more or less confused about the GW initial component in this plot. If GW initial term indicate the water budget from initial condition, does the F_GW term should be 0?

The reviewer is correct. At the very beginning of the simulations, the initial GW fraction (f_{GWi}) is equal to 1 in all model cells, while all the other water fractions are equal to 0. Therefore, a "spin-up" period is carried-out to flush out this virtual initial water and have a more representative

distribution of the other water fractions of interest throughout the domain. In the plots, we do not present this initial, "pre-spin-up" condition since our analyses focused on the remaining three water fractions. We have rewritten the figure caption to try and make this clearer. We have also added a sentence to clarify this in the methods section. It now reads (end of section 2.2):

> *We ran the model for a spin-up period at the beginning of the simulations in order to establish a more realistic distribution of the three water fractions over the domain at the beginning of our analyses. The spin-up period consisted of a two-year simulation period using constant average BC values. Following this period, the f_{GWi} fraction was virtually zero, whereas the three remaining water fractions were the only fractions observed throughout the domain. Thus, in the remaining analyses we mainly consider the three remaining water fractions for our calculations.*

Fig.5: Still, I'm trying to understand the fraction response to a rainfall event, for example, summer 2013. I see that fSW actually decreases when this new rainfall event occurs - could you explain why and is this something you are expecting?

Thank you for bringing up this point for discussion. In this case, we could first observe an increase of f_{SW} at the peak of the discharge event resulting from the large volume of stream water pushed into the riparian aquifer due to the increase of SW depth in the stream channel. This is then followed by a general increase of f_{FW} (during the falling limb of the event) as pointed out by the reviewer. This due to larger portions of stream water overflowing the banks during the discharge event, which then re-infiltrates into the subsurface during the recession of the flow event. This reinfiltrating water is considered to be part of the f_{FW} fraction within the simulation. Therefore, even though discharge is great during this event, there is a relatively larger increase in the f_{FW} in comparison to f_{SW} following the event.

This is something we expect in case of large discharge events due to flooding of the near stream area (Fig.S4, for example). For smaller discharge events, in which overbank flow does not occur, we expect an increase of f_{SW} only, as it can be observed in other periods of the simulation.

Fig.6: I think the figure caption does not explain why multiple profiles are shown in each plot? This is confusing to me - what each slice represents?

We apologize for the confusion. In the plots, the slices are cuts throughout the 3D domain (instead of showing an opaque volume or a fully translucent object), which facilitates the visualization of the subsurface properties in our opinion. In each subplot, the slices represent the same properties, however in different segments of the subsurface. We have rewritten the captions of the figure (and of Fig.7 as well) to better explain this idea. For instance Fig.6 now reads:

> *Fig.6: slices throughout the simulated 3D domain showing the minimum, maximum, median values, as well as standard deviations (σ_{HMC}) of stream water (f_{SW}) **(a-d)**, groundwater (f_{GW}) **(e-h)**, and floodplain water (f_{FW}) **(i-l)** fractions for the entire simulation period in different segments of the domain. The black line **(a-c, e-g, i-k)** indicates the HMC fractions of 0.5. Note the vertical exaggeration of the 3D plots (20x).*

L526: Why is Spearman's rank correlation used instead of Pearson's correlation, which I think is more widely used?

We prefer to use the Spearman's rank correlation instead of the more commonly used Pearson's correlation, because with the later we primarily want to characterize and quantify the monotonic relationships between the parameters independent of the magnitudes of the changes in parameter values. As we do not expect linear relationships between the analysed parameters we considered the Spearman's rank correlation to be the more straight forward and robust metric to characterize the monotonic relationships. However, we attest that by using the Pearson's correlation coefficient, one would likely obtain somewhat similar values and ranges in the final correlation coefficients.

L635: I know a solute transport model is not yet developed so it is difficult to estimate how much computational saving by using this method. This can be an open questions I guess, but how do you justify the computational saving?

Thank you for bringing up this point for discussion. Indeed, particularly for the modelling area, this could be seen as an open question at the moment. However, an explicit solute transport model at the same site was used in Nogueira et al. (2021) with substantially longer computation times for comparable conditions. However, we agree that this is not a direct comparison of computation times. In order to avoid further misunderstanding, we have rewritten the sentence. It now reads:

> *We intentionally did not conduct explicit simulations of reactive transport in this study since our main goal here was to explore the HMC method (coupled to a flow model) in order to assess the development of mixing spots on the riparian zone and their relation to hydrological variations. We additionally showed the importance of such macroscopic mixing spots for groundwater-borne NO_3^- turnover by comparing the quantitative mixing results of the HMC method with previous biogeochemical assessments carried out in the study area. For a direct quantification of nitrate removal rates, however, the use of reactive-transport models or other data-drive analyses would be indispensable. Such simulations would have allowed a comparison of observed and simulated concentration values and their dynamics for a more rigorous evaluation of model performance (Nogueira et al., 2021b). However, the additional computational effort to numerically solve the transport equations would likely also increase computational costs.*

References

Brunner, P., Simmons, C.T.: HydroGeoSphere: A Fully Integrated, Physically Based Hydrological Model. Ground Water 50, 170–176. https://doi.org/10.1111/j.1745-6584.2011.00882.x, 2011.

Jones JP, Sudicky EA, McLaren RG.: Application of a fully-integrated surface–subsurface flow model at the watershed-scale: a case study. Water Resour Res., 44:W03407, DOI:10.1029/WR005603, 2008

Nogueira, G. E. H., Schmidt, C., Brunner, P., Graeber, D. and Fleckenstein, J. H.: Transit-time and temperature control the spatial patterns of aerobic respiration and denitrification in the riparian zone, Water Resour. Res. 57, https://doi.org/10.1029/2021WR030117, 2021.

Park, Y.J., E.A. Sudicky, R.G. McLaren, and J.F. Sykes: Analysis of hydraulic and tracer response tests within mod- erately fractured rock based on a transition probability geo-statistical approach. Water Resources Research 40:W12404. DOI: 10.1029/2004WR003188, 2004.

Partington, D., Brunner, P., Simmons, C.T., Therrien, R., Werner, A.D., Dandy, G.C., Maier, H.R.: A hydraulic mixing-cell method to quantify the groundwater component of streamflow within spatially distributed fully integrated surface water-groundwater flow models. Environ. Model. Softw. 26, 886–898. https://doi.org/10.1016/j.envsoft.2011.02.007, 2011.

Schmadel, N.M., Ward, A.S., Kurz, M.J., Fleckenstein, J.H., Zarnetske, J.P., Hannah, D.M., Blume, T., Vieweg, M., Blaen, P.J., Schmidt, C., Knapp, J.L.A., Klaar, M.J., Romeijn, P., Datry, T., Keller, T., Folegot, S., Arricibita, A.I.M., Krause, S., War, A.S., Kurz, M.J., Fleckenstein, J.H., Zarnetske, J.P., Hannah, D.M., Blume, T., Vieweg, M., Blaen, P.J., Schmidt, C., Knapp, J.L.A., Klaar, M.J., Romeijn, P., Datry, T., Keller, T., Folegot, S., Marruedo Arricibita, A.I., Krause, S.: Stream solute tracer timescales changing with discharge and reach length confound process interpretation. Water Resour. Res. 52, 3227–3245. https://doi.org/10.1002/2015WR018062, 2016.

Therrien, R., and E.A. Sudicky: Three-dimensional analysis of variably-saturated flow and solute transport in discretely-fractured porous media. Journal of Contaminant Hydrology 23, no. 1–2: 1–44, 1996.

Unger, A.J.A., Q. Li, E.A. Sudicky, D. Kassenaar, E.J. Wexler, and S. Shikaze: Simulating the multi-seasonal response of a large-scale watershed with a 3D physically- based hydrologic model. Journal of Hydrology 357, no. 3–4: 317–336, 2008.

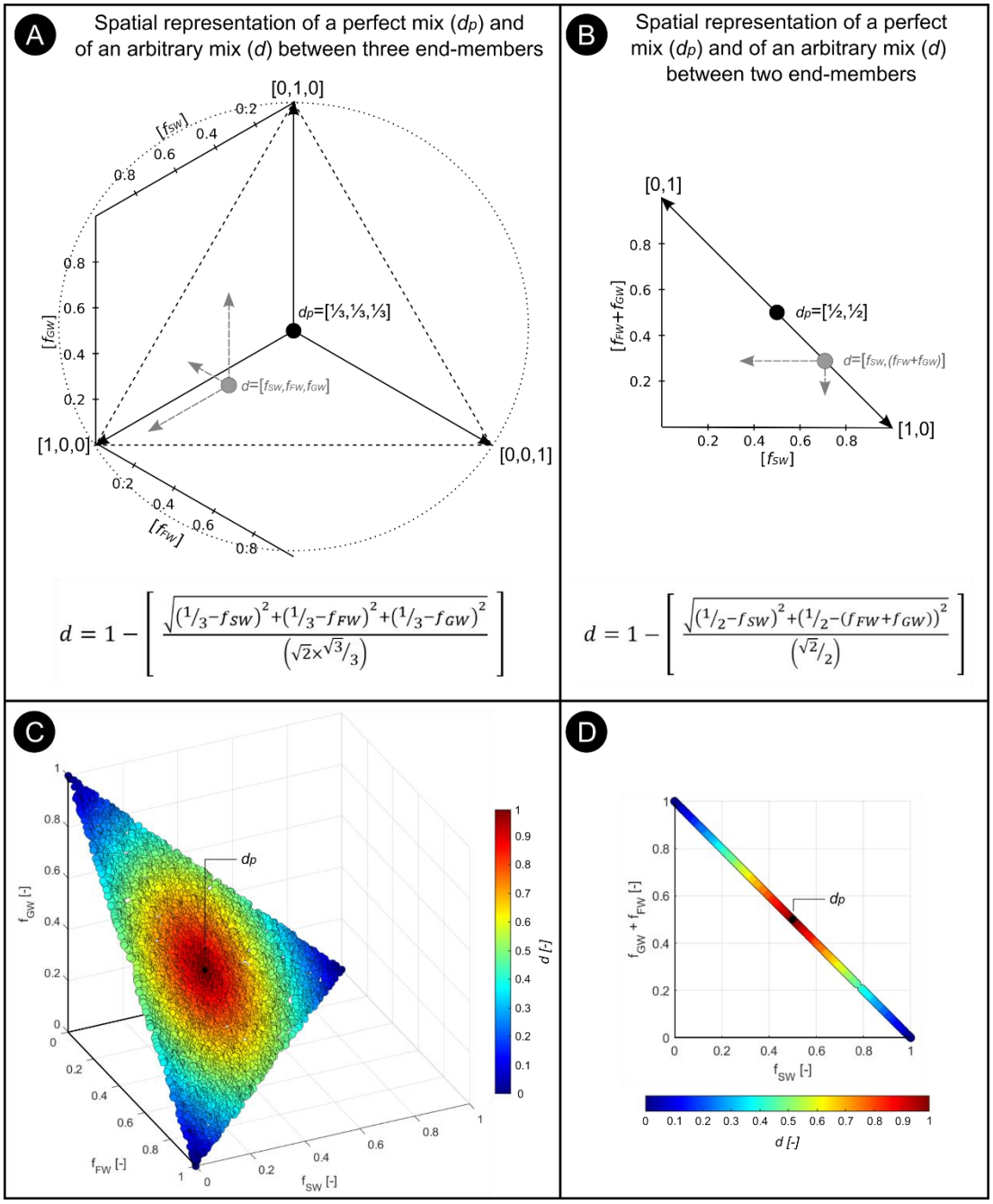

Fig.S2: Spatial representation of a perfect mixing (d_p) and of and arbitrary mixing (d) for the cases of three **(a)** and two **(b)** end-members mixing. The final mixing d can be calculated as the Euclidean distance between points d_p and d. For a three end-members mixing (3D case), any combination of fractions can be represented as a point d in a 3D coordinate space, in which the maximum distance is a radius of a circle (centred at [1/3, 1/3, 1/3]) escribed on an equilateral triangle (side length of $\sqrt{2}$). Thus, the maximum distance between d_p and d is $(\sqrt{2} \times \sqrt{3}/3)$ For a two end-members mixing (2D case), the maximum segment is the diameter of a circle (centred at [0.5, 0.5]), whereas the maximum distance between dp and d is $(\sqrt{2}/2)$. The long-dashed lines in (a) delimit the solution space for any possible mixing d where fractions sum up to 1. In (b) final mixing d values would fall over the solid line passing through d_p. Example of theoretical mixings between three **(c)** and two **(d)** end-members coloured according to computed d values (warmer colours indicate a more homogenous mixing); d_p is indicated as a black circle. The theoretical mixings were generated with 10000 random combinations of HMC fractions that sum up to 1. For a four (or more) end-members mixing a spatial representation is not possible but the general Eq.4 would equally work.