Reply to Interactive comment by Anonymous Referee #2 "A sprinkling experiment to quantify celerity-velocity differences at the hillslope scale"

Willem J. van Verseveld, Holly R. Barnard, Chris B. Graham, Jeffrey J. McDonnell, J. Renée Brooks, Markus Weiler

First of all we would like to thank Referee #2 for his/her time, useful comments on this manuscript and asking relevant questions. His/her suggestions will for sure help to improve the quality of this manuscript. Our answers (in blue) to the suggestions are written below each suggestion (black).

1. Abstract. It is difficult to understand the significance of the study. The first sentence states that the difference between velocity and celerity is poorly understood. I would argue that it is very well understood. Perhaps the mechanisms explaining such differences are poorly understood. Many numbers are given in the abstract, but it is difficult to understand why such numbers would be interesting. I would stress in the abstract more the connection between experimenting and modelling, which I find the most interesting aspect of this work.

We fully agree with the reviewer that the significance of the study is not very clear from the abstract alone, and that indeed the connection between experimenting and modelling is an interesting aspect of our study. We will change the abstract to make the significance of our study clearer, also leaving out the many numbers. Additionally, we will also change the first sentence.

2. In the introduction, I would cite the paper "Velocity and celerity dynamics at plot scale inferred from artificial tracing experiments and time-lapse ERT", Journal of Hydrology, 2017, as it seems relevant for the study.

This paper seems indeed relevant for our study, and we will include it in the Introduction.

3. In section 3.1, before starting to describe what was done, it would be useful to illustrate why it was done it. What were the objectives of the experimental design? Which conditions did you want to recreate? Why? What about natural rainfall in addition to artificial rainfall? Did it happen? If not, what if it happened? Etc.

We think the objectives of this study are clearly mentioned in section 1, also linked to the experimental design. We agree that we did not include information on natural rainfall during the sprinkler experiment (also mentioned by another reviewer), and on what kind of conditions we did want to recreate (in comparison with natural conditions). We will add this information.

4. Equation 7 appears to be wrong – the integral of concentration is not equal to mass. What does Mout represent? If it is just the mass of tracer in the outflow, what about Evaporation? I guess the calculation of mean residence time should account also for this. . .

We agree that Equation 7 is not clearly explained, but it is correct, under steady-state water flow conditions (the exit distribution can then be calculated from concentrations alone). Mout is the mass of tracer at the exit boundaries (for evapotranspiration, recharge, seepage and lateral subsurface flow (captured at trench and total). We did calculate the exit time distribution for these exit boundaries.

Calculation of the mean residence time distribution was based on mass in the storage zone (unsaturated and saturated zone), Equation 8. Outgoing mass fluxes from evaporation and later subsurface flow are thus taken into account.

We will improve the description how we did calculate the exit time distributions and we will change Equation 7; we will only use the right part of the equation (mass), because of unsteady water flow conditions during our sprinkler experiment.

5. What was the recovery rate of water and tracers? Did it differ? Why?

The recovery rates of (simulated) water and tracers are presented in Table 5 and did differ, both at the trench and for total later subsurface flow, with higher recovery rates for water. We explain this difference in the paper by a dual porosity system; about 10% of deuterium remains in the unsaturated zone.

6. Model 1 and Model 2 is a misleading terminology. In fact it appears that the model structure is the same, just the evaluation criteria are different.

Yes, the model structure is the same. We think this depends on the definition, are two models with two different parameter sets, but the same model structure, different models? We think the models are different, and do not think Model 1 and 2 are misleading terms, as we need to make clear in the text to which model (or model with a specific evaluation criteria) we are referring to.

7. The criteria used to calculate the behavioural parameter sets for Model 1 and 2 make the comparison difficult as the criteria with which the models are evaluated are very different. I think it would make more sense to make the behavioural parameters of Model 2 a subset of the behavioural parameters for Model 1, by requiring them to satisfy some additional constraints based on the tracers.

We don’t think it makes more sense to constrain the parameter set of Model 2 by the behavioural parameters of Model 1. By doing so, one could a-priori exclude possible behavioural parameter sets for Model 2. Please also note that Model 1 is still part of the behavioral parameter sets of Model 2 (we did not reject Model 1 based on based on the objective criterion for the deuterium breakthrough).

8. Indeed, in Figure 6 the two distributions do not appear to differ significantly, and in my opinion, given how they are constructed, they are not even comparable (besides the fact that the caption is not clear, in Figure 6 there are 2 green colors. Which green color do you mean?)

Correct, the two distributions for most parameters of Model 1 and Model 2 are similar. However, parameters n, n0 and kb are more identifiable for Model 2 (from Figure 6 and lower parameter uncertainty (Table 3)). Please note that we use two different colors in Figure 6: green and blue. Green is transparent, so where the color is dark green, the two distributions overlap. We will make this clearer in the legend of Figure 6.