Response to Reviewer comments for manuscript HESS-2016-228, “Multiresponse modeling of an unsaturated zone isotope tracer experiment at the Landscape Evolution Observatory”, by Carlotta Scudeler, Luke Pangle, Damiano Pasetto, Guo-Yue Niu, Till Volkmann, Claudio Paniconi, Mario Putti, and Peter A. Troch

SUMMARY:
This manuscript deals with the modeling of an unsaturated flow and isotope tracer experiment. The experiment, conducted at the Landscape Evolution Observatory (LEO), involved successive injections of water and deuterium-enriched water into an initially very dry hillslope. Multivariate observations were presented for flow and transport: soil moisture, water and tracer outflow, breakthrough curves and total water storage. Simulations were performed with the physically-based distributed numerical model CATHY that solves the 3D Richards and advection-dispersion equations and includes coupling with surface routing equations. The modeling approach succeeded in simulating the integrated flow and transport responses. However, with the same parameterization it failed to restitute the point measurements of the water contents and the tracer concentrations.

OVERALL QUALITY:

This manuscript is clear, well structured, and pleasant to read. The experimental results are new. However, they should be better described and discussed. It is surprising to see that these well calibrated experiments are so difficult to model. Some of the numerous parameterizations added in the successive simulations look arbitrary and their choice should be better justified. Furthermore, the cumulative mass balance error of tracer in the CATHY simulations is relatively large (~2% with respect to the total mass injected) and this fact should therefore be discussed. The conclusions of the manuscript would be more convincing if more than one numerical code were used. But this task could be further accomplished in a future publication. Very surely, these experiments and their first simulations could serve as a nice benchmark for physically-based distributed numerical models provided the full dataset is rendered available. In my opinion, the experimental results and the corresponding simulations are very interesting and deserve to be published in HESS. However, some corrections and/or clarifications should be accomplished prior to publication. The authors will find below some remarks to correct or complete their manuscript.

1) We wish to thank the Reviewer for the attention to our work and the very detailed and constructive comments. The main issues raised above are taken up individually below and we will respond to each point raised. We agree that the data from the LEO experiments would make nice modeling benchmarks; the dataset from this paper is indeed available, as noted in the acknowledgements.

MAJOR COMMENTS:

(1) The experimental results are new and interesting. However, the description and discussion of the water contents and concentrations measured should be improved. You will find below some examples of questions that arise about the experimental results.

(1a) Page 4, Figure 1: Please comment the peak of d^2H *during the first irrigation event.*

2) Please note that we have made a mistake in that plot. The curve should be shifted by 23.5 h (see Figures 4, 6, 8, and 13). Thus the peak appears with the second pulse of rain, in the early seepage face flow. This is probably due to the fact that the residual soil water in the landscape prior to irrigation had become somewhat enriched in deuterium (compared to the irrigation water) during evaporation. In fact, during evaporation, hydrogen will preferentially go into the vapor phase

compared to deuterium, so that the liquid phase remaining in the soil easily becomes slightly enriched in deuterium. The delta d^2H values in that early seepage flow may reflect some mixing of the new irrigation water with the evaporatively-enriched residual soil moisture. This slight enrichment may disappear in the seepage flow at later times just due to dilution of that residual soil moisture by the newly infiltrating irrigation water. We are going to add this information to the section of the paper that describes the experiment.

(1b) Page 18, line 7: You are using the soil water content at 4 different depths averaged over 496 sensors. Can you quantify the soil heterogeneity from a statistical analysis of these 496 experimental vertical profiles?

3) The detailed information generated from the LEO experiments constitutes a valuable dataset for analyses such as the one suggested by the reviewer. It should be possible, albeit not within the scope of the present work, to perform inverse modeling to retrieve conductivity distributions based on this moisture information (note: in reality the 496 sensors do not correspond to 496 vertical profiles, since at each sampling position there are 2 to 4 sensors, at different depths). In addition to soil heterogeneity, there are other factors that can affect the soil moisture response at the different locations (heights and positions) of the hillslope. In the two figures below we plot the standard deviation for both the observed and modeled profiles. For the modeled response we have done this for both the homogeneous n_{vg} and heterogeneous n_{vg} cases. From this analysis we can see that the deviations from the average profiles for the observed and modeled responses are similar (apart from the results at 85 cm depth), suggesting that the model parameterization is quite adequate. We propose to replace Figure 10 in the original manuscript with these two more detailed figures, and to revise the description of these results accordingly.

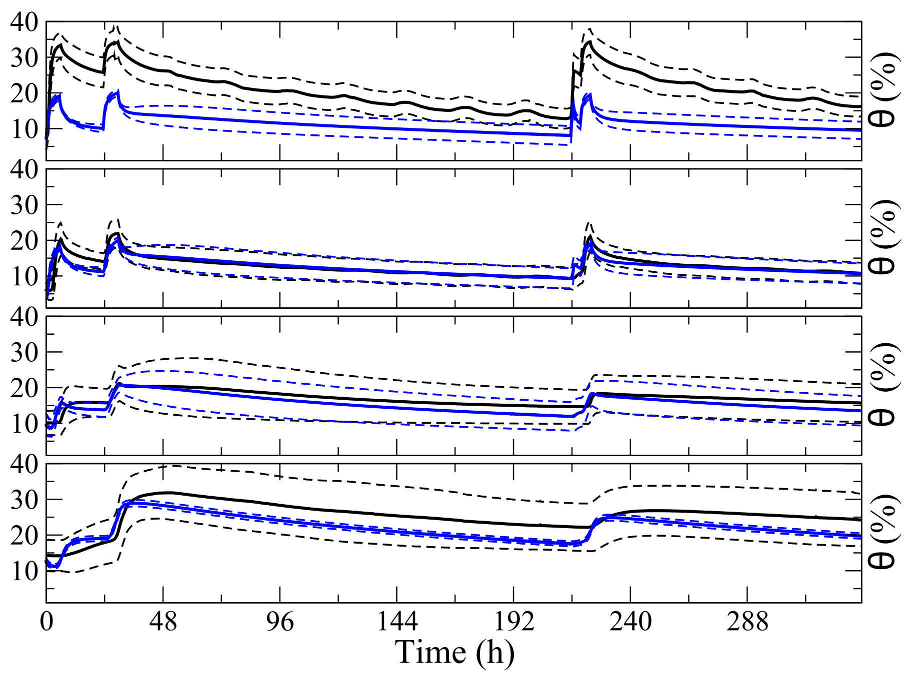

Figure 1. Averaged soil water content (θ) profiles at 5, 20, 50, and 85 cm (top to bottom) depth from the surface: observed (solid black curves) and calculated (solid blue curves, simulation f). In each graph the deviation from the mean (one standard deviation above and below) is shown as dashed lines (blue for the model and black for the measurements).

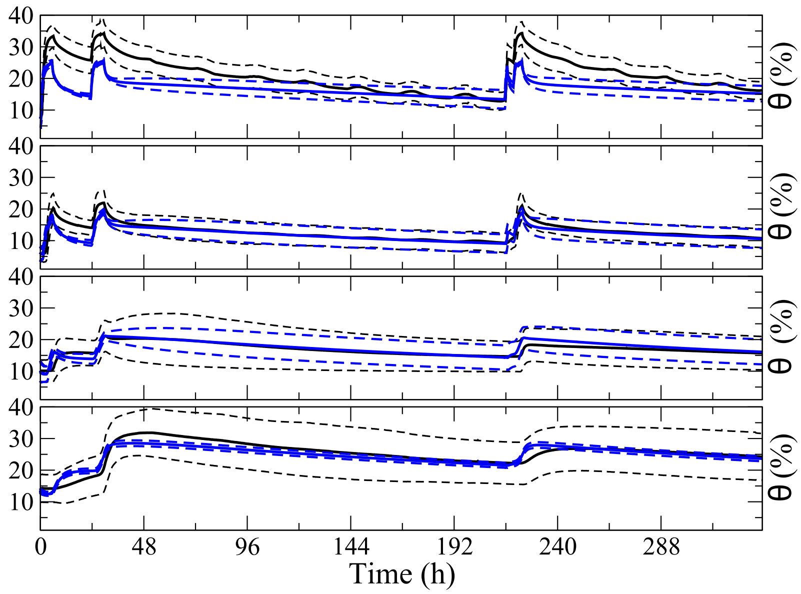

Figure 2. Averaged soil water content (θ) profiles for the variable n_{vg} simulation (simulation l) at 5, 20, 50, and 85 cm (top to bottom) depth from the surface: observed (solid black curves) and calculated (solid blue curves). In each graph the deviation from the mean (one standard deviation above and below) is shown as dashed lines (blue for the model and black for the measurements).

The landscape geometry is symmetric. All parameter heterogeneities included in the simulations are symmetrical as well. Does one observe this symmetry also in the experimental results? For example, are the theta vertical profiles measured along two vertical lines that are located at the same distance from the seepage face but on either side of the landscape similar? Is the variability of the profiles correlated with the rainfall variability?

4) In Figures 3 and 4 below we provide a comparison between the soil water content (θ) response at 5, 20, 50, and 85 cm depth from surface measured at 6 m from the seepage face and 4 m at the right and left of the central axis (Figure 3) and at 22 m from the seepage face and 2 m at the right and left of the central axis (Figure 4). It can be seen that the left and right responses are reasonably symmetric even if they are not identical. Although the rainfall is not uniform, as reported in the paper, there is no axial asymmetry in the engineered rain system at LEO. Thus we do not expect rainfall variability to be correlated to the variability of the soil moisture profiles in this sense.

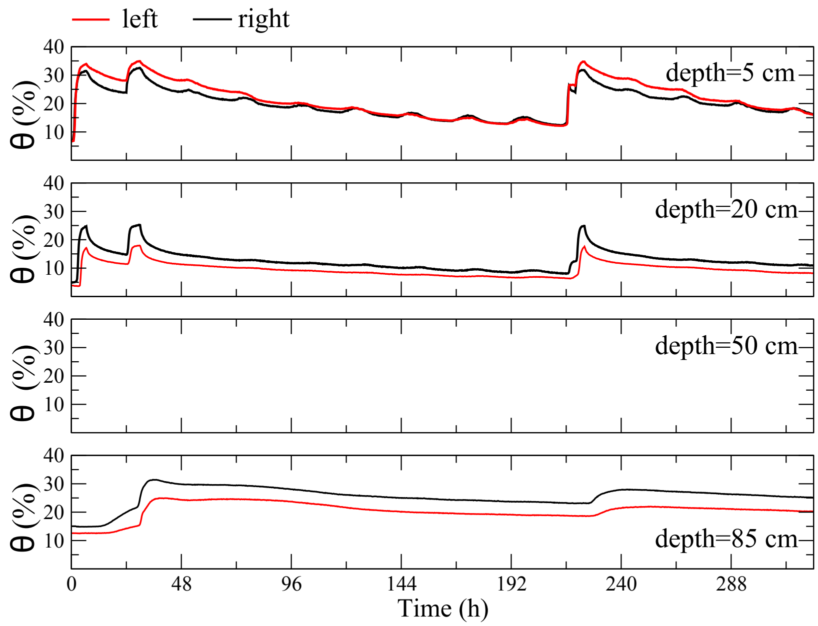

Figure 3. Soil water content (θ) observed response for the profiles at 5, 20, and 85 cm (top to bottom) depth from surface for two points located 6 m from the seepage face and at the left (red curves) and right (black curves) of the central axis.

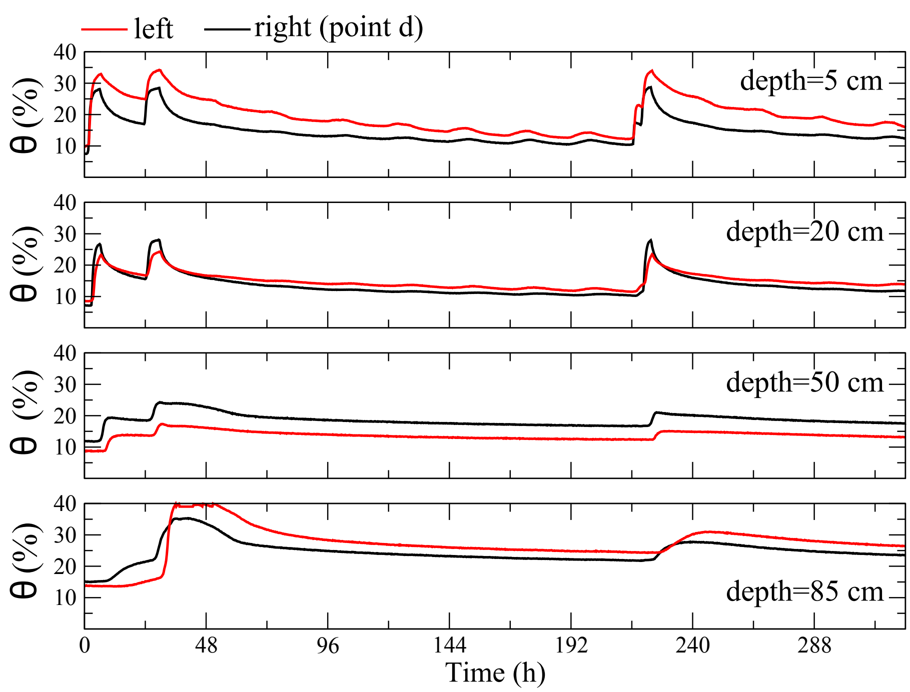

Figure 4. Soil water content (θ) observed response for the profiles at 5, 20, and 85 cm (top to bottom) depth from surface for two points located 22 m from the seepage face and at the left (red curves) and right (black curves) of the central axis. The profiles at the right are relative to point d shown in the paper.

(1c) Page 20, Figure 11: The time evolutions of theta measured at the points located at the center of the hillslope (points a, b, and c) clearly show that the bottom of the hillslope has become water saturated after the 2nd rainfall event. This point should be discussed in the paper. CATHY has clearly failed to simulate this saturated zone. More generally, practically all theta values obtained from the simulation, presented in Fig.11 are lower than the corresponding values obtained from the measurements. Do you have an explanation for this lack of water in the CATHY simulations? I think it would be better to calibrate some parameters (e.g., n_{VG}) from distributed theta profiles instead of calibrating parameters from averaged theta profiles. As a matter of fact, the vertical evolution of the wetting front varies depending on whether it is observed at the top of the landscape or in the zone of flow convergence.

5) We agree with the Reviewer that the proposed model is not able to simulate the water content dynamics for some sensor locations. This is already highlighted in the manuscript at page 20, lines 3-4. However, also with these limitations, the proposed model is partially able to simulate the saturated zone, since the simulated water table level is only a few centimeters lower than the 85 cm sensors. We do not think that there is a lack of water in the CATHY simulations: in fact the CATHY mass balance (Figure 7) is consistent with the measured data and, moreover, the measured and modeled averaged profiles of water content depicted in Figure 10 show a good agreement. The CATHY underestimation of water content along the central transect must be compensated by a general overestimation of water content along the lateral slopes. This means that the model is failing to reproduce the water convergence toward the center. This might also explain why the deviation for the averaged θ modeled profiles is very small. As stated at page 10, lines 8-10, to obtain a good calibration with respect to the θ profiles (non-averaged), it is necessary to increase the complexity of the parameter spatial distribution, for example as done in Pasetto et al. [2015] for a synthetic test at LEO, or to perform inverse modeling. However, as stated in the response point 3 above, this goes beyond the scope of the work. Finally, the physical model driving the system already provides different responses for the simulation of wetting fronts at the top and at the bottom of the landscape, without the need to introduce additional complexity in the parameter distribution.

(2) Some hypotheses and some results of the modeling approach require further argumentation and discussion.

(2a) Page 8, line 16: Several parameterizations in the simulations are arbitrary and not justified. For example, why did you choose a depth of 38 cm? Did you perform a calibration? Evaporation is often assumed to be active only over the first few centimeters.

6) We agree that our implementation to model fractionation is somewhat empirical. In adopting a sink term representation for the isotope (transport model) behavior during evaporation, our aim was to represent the range of possible responses between no solute, some solute, and maximum solute leaving the system with the evaporating water. Thus, the scope is not to calibrate the set of empirical parameters related to the phenomenon empirically described. The parameterizations were chosen in order to qualitatively reproduce the experimental results obtained by Barnes and Allison [1988], where it is shown that, for isotope profiles in unsaturated soil and under evaporation, the maximum concentration can also occur at 50 cm from the surface. Above this point the isotope concentration decreases rapidly towards the surface due to the diffusion of water vapor to the soil surface. In our model we assume that the region dominated by water vapor diffusion is also the one characterized by evaporation, and selected 38 cm for the threshold. We will describe this better in the revised manuscript.

(2b) Page 8, line 21: Same as remark (2a): Why did you choose λ=1 m-1?

7) See the response to the previous point.

(2c) Page 8, line 20: There is no moisture content dependence term in the sink term given by Eq. (15). What happens if there is not enough water for evaporation in the upper 38 cm of soil?

8) The moisture content cannot be lower than the moisture deficit threshold. Thus, when one element reaches the moisture deficit threshold, parameterized by its corresponding pressure head level (given as input), the evaporation process becomes soil limited. When this occurs, the actual sink term function will be automatically smaller than the imposed value. However, this did not occur in our experiment. We will add a note on this in the revised manuscript.

(2d) Page 9, lines 3-5: Same as remarks (2a) and (2b): the choice of f_c looks very arbitrary. Please justify it.

9) Water vapor diffusion increases with evaporation close to the surface and, consequently, more solute evaporates. This explains why f_c decreases towards the surface. We will add a sentence in the revised paper to better explain this.

(2e) Page 11, Table 3: Same as remarks (2a), (2b), and (2d): how did you choose the k values? Did you perform a calibration?

10) A systematic modeling analysis on the soil parameters was performed for the first LEO experiment [Niu et al., 2014]. In this first step of our analysis (integrated flow response to which Table 3 refers), we started from the results obtained by Niu et al., 2014. For example, the clogging hypothesis (with the same k values used in our work) was already invoked by Niu et al., 2014, as also the vertical k value is the one obtained for the first LEO experiment. We introduced a slight anisotropy as a result of a modeling analysis that we performed by running different simulations for different horizontal conductivity within a reasonable range. The six simulations are reported to show that a homogeneous parameterization is not representative of the LEO hillslope and that the clogging hypothesis at the seepage face applies also for our case, and also to investigate the effects of the initial conditions and rainfall distribution.

(2f) Page 11, lines 1-6: The heterogeneity and anisotropy of k_s are justified by invoking the processes of clogging and compaction. Such modifications should induce a modification of the topography. Did you observe topographic changes caused by diffusive geomorphic processes such as rain splash during the rain events that lasted several hours? Pangle et al. (2015) affirm that "digital elevation models will be constructed at regular intervals and following all events with the potential to modify the topography", with a model surface precision of 0.002 m. Have you performed such measurements? If yes, did you observe some changes in the topography? Have you observed the formation of some crusts at the soil surface? The properties of the soil, e.g., its permeability, must be changed with time if crusts are forming.

11) For the clogging phenomenon, we doubt that subsurface translocation of fine particles would have a measurable effect on topography. The discussion of the clogging hypothesis and consequent reduction in hydraulic conductivity at the seepage face can be found in Niu et al. (2014), where it is explained that during the experiment movement of some fine material into the seepage face was observed and that shortly after the experiment the gravel at the seepage face was removed to a depth of 72 cm and a 2% fraction of fines per volume of gravel was measured. The vertical compaction hypothesis was introduced in the present study to accommodate anisotropy but no new surveys of the surface topography were taken for this experiment, which is in any case of quite short duration so

that eventual alterations of topography over time can be safely neglected in this study. For the longer term experiments planned for LEO, involving co-evolution of physical, geochemical, and ecological aspects of the hillslopes, the topography will indeed be closely monitored.

(2g) Page 14, line 28: The best numerical results are obtained with the smallest value of the dispersivity. Can you discuss this result? Is it a proof that the soil is very homogeneous? It would be interesting to measure α_L for example from transport experiments in a column filled with the same porous media.

12) Since dispersion depends on the scale of measurement and can also be quite different for saturated versus unsaturated processes, a soil column measurement would probably not be useful for the hillslope model. We cannot say for sure that having obtained the best results with the smallest value of dispersivity is a proof that the soil is very homogeneous. In our case, dispersion is also related to the grid size in the model, with Peclet constraints dictating the smallest value that we were able to test (0.001 m for longitudinal dispersivity). Various issues surrounding dispersion are discussed in the concluding sections (5 and 6) of the paper.

(2h) Page 16, line 5: Can you explain why the cumulative mass balance error is so large (~2%)?

13) The 2% mass balance error clearly arises from the jump that occurs with the third pulse of rain (see Figures 5, 7, and 9 in the manuscript). This is probably due to the discontinuities in the time derivative of concentration and the water saturation close to the surface (being the soil very dry at this level and after the long evaporation period) as a consequence of the discontinuity in the atmospheric boundary condition. Moreover, with a finite element based model for advection-dispersion we always expect a non-zero mass balance error. To have a near perfect mass balance, the advective fluxes, governed by a hyperbolic conservation law, should be resolved by means of a numerical technique that mimics a mass balance within each cell of the computational domain or control volume (e.g., finite volume, discontinuous Galerkin) by using as input a mass-conservative velocity field, as we have recently implemented. We will add in the revised paper a comment on this.

MINOR COMMENTS:

(3) Page 4, line 3: Please clearly indicate the location of the seepage face. Is it the 11 m^2 boundary at the downslope end of the landscape?

14) Yes, where we also set a seepage face boundary condition in the model setup. We will clarify this in the revised manuscript.

(4) Page 4, line 9: The estimated evaporation rates are two times and ten times larger than the rates reported in Niu et al. (2014) and Pasetto et al. (2015), respectively, although the soil is drier. Can you explain this difference?

15) The first experiment (Niu et al., 2014) was performed in February 2013, two months before ours. The month of April is characterized by higher temperature and solar radiation compared to February, which explains the higher evaporation rate (the LEO hillslopes are housed in a transparent, greenhouse-like structure). The simulations reported in Pasetto et al. (2015) are based on synthetic conditions.

(5) Page 4, Figure 1: The irrigation rate is equal to 12 mm/h ~ $1.1x10^{-3}$ m^3/s. Please correct the y-scale for Qr in Fig. 1.

16) You are right. We will correct this mistake.

(6) Page 4, Figure 1: Does the size of the symbols for $d^2H(t)$ correspond to the 0.5‰ analytical precision?

17) We are not sure we understand this point. What do you mean by "size which corresponds to 0.5‰ analytical precision"?

(7) Page 5, Equation (1): The CATHY model solves the coupling between surface and subsurface flows. Why do you not quote the surface flow equation?

18) Yes, CATHY is a model for integrated surface/subsurface numerical simulations. We are not showing the surface equations (for both flow and transport) because the experiment is characterized by subsurface processes alone.

(8) Page 5-6, section 3.2: How are the nonlinear terms in the equations being solved? Is it based on an iterative scheme with Picard iterations?

19) The nonlinear system arising from the numerical discretization of Richards' equation is solved by means of a mixed Newton/Picard iteration with time step adaptation. The method applies Newton linearization to the term involving $\theta(\psi)$ and Picard linearization to all the remaining nonlinear terms (see Scudeler et al., 2016 cited in the manuscript). The system arising from the discretization of the transport equation (9) is linear and does not require an iterative procedure.

(9) Page 6, line 17: Please remove Eq.(6) because the effective saturation has already been defined in line 20.

20) We will do it.

(10) Page 6, line 19: Please replace the exponent in Eq.(8) with "-m" and add the definition of m: m = 1 - 1/nVG.

21) We will do it.

(11) Page 6-7, Eqs. (9)-(13): I am not convinced of the interest to present Equations (9)-(13). What is new in comparison with the schemes already described by Putti et al. (1998) or by Weill et al. (2011)?

22) In these prior papers the advection-dispersion equation is solved by means of a time-splitting technique that combines finite volumes for advection and finite elements for dispersion, and the boundary conditions are implemented in a different way since in the version of the model for our paper a finite element discretization is used for both advection and dispersion. Equations (9)-(13) are thus important for documenting how the boundary conditions have been implemented.

(12) Page 7, section 3.3: How did you choose the horizontal and vertical discretizations? Did you verify the spatial convergence of the numerical simulations?

23) The surface mesh is the same as the one used in Niu et al., 2014. The horizontal discretization was chosen in order to have the nodes of the computational mesh aligned with the sensor and sampler locations, thereby allowing us to directly compare simulated and measured distributed

responses. This same principle was used to guide the vertical discretization (the interface between two layers is set at the sensor and sampler heights). In addition, for the vertical discretization mesh refinement was required where strong velocities occurred (close to the surface), in accordance with Peclet constraints and to properly resolve the infiltration dynamics.

(13) Page 8, line 23: Please correct the values for the evaporation: 5 mm/d ≡ 5.8x10-8 m/s and 3.9 mm/ d ≡ 4.5x10-8 m/s

24) We will correct them.

(14) Page 9, Caption of Table 1: zi is the depth of the middle of the ith layer.

25) We will change the definition in the caption of Table 1 as well as in the text.

(15) Page 9, Table 2: Please verify the values given in Table 2. For example, for layer 5, fc1i=1.91x10^{-8}c, for layer 7, fc1i=2.36x10^{-8}c and fc2i=1.41x10^{-8}c, for layer 11, fc1i=4.77x10^{-8}c, for layer 12, fc1i=6.75x10^{-8}c.

26) The values in Tables 1 and 2 were calculated with equation (11) and using F_{ev}=-5.7 m/s and -3.4 m/s, the two values in line 23, page 8 that will be corrected (see response 24 above). In accordance, we will also update the values in the two tables.

(16) Page 11, Table 3: Please correct the name of the last simulation: "f" instead of "e".

27) We will do it.

(17) Page 11, line 19: Please provide the definition of the coefficient of efficiency CE.

28) We will provide the definitions of *CE* and *RMSE* in the revised manuscript. These are calculated as:

$$CE = 1 - \frac{\sum_{i=1}^{n}\left(Q_i - \hat{Q}_i\right)^2}{\sum_{i=1}^{n}(Q_i - \bar{Q}_i)^2} \qquad RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^2}{n}}$$

where n is the total number of observed data available at the different times, Q_i and $\hat{Q}_i$ are the i^{th} modeled and observed values, respectively, and $\bar{Q}_i$ is the average of the observed values.

(18) Page 13, Figure 3: Figure 3 would be clearer if the time evolutions of the seepage face flow and of the total water storage for a given case were reported side by side instead of one above the other. Furthermore, the superposition of two simulated test cases in each figure is unnecessary.

29) We agree with these suggestions. The revised figure and caption to be included in the manuscript is shown below.

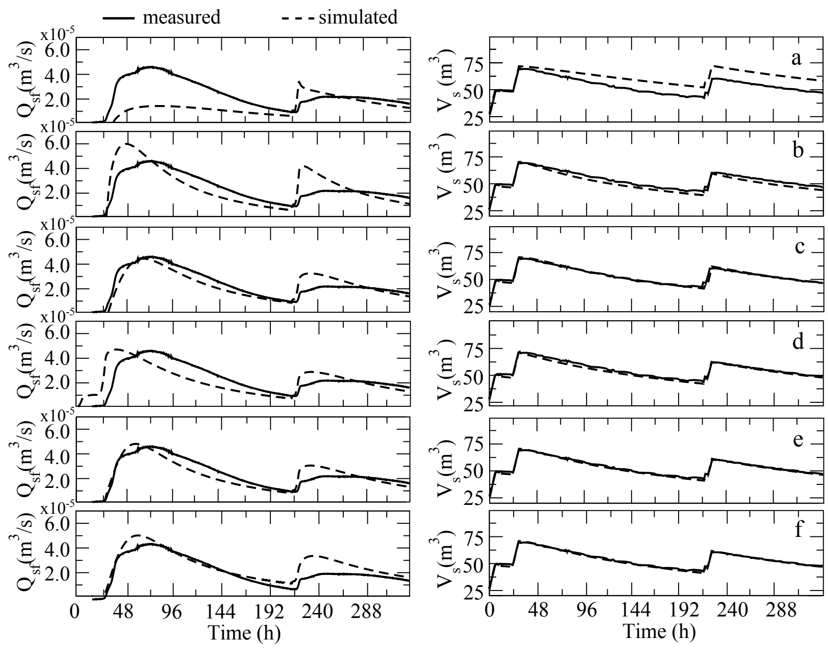

Figure 5. Results for the 6 simulations (a to f, from top to bottom, as defined in Table 3) of the integrated flow response analysis. For each case the seepage face flow Q_{sf} (left) and total water storage V_s (right) are reported. The solid lines correspond to the measured responses and the dashed lines to the simulated responses.

(19) Page 14, lines 19-20 and line 26: Finally, which simulation (e or f) is used for the subsequent simulations? Please correct the text accordingly.

30) It is the sixth (f) for the integrated transport analysis and, again, simulation f for the homogeneous n_{VG} parameterization in the distributed analysis, as already indicated in the text. We will insert "(simulation f)" after "sixth flow simulation".

(20) Page 14, lines 30-32: I do not understand what you mean. In my opinion, in Fig.4, 2H-labeled water appears in the measured outflow discharge and also in all simulated outflow discharges after the second pulse.

31) It is true that we have non-zero solute concentration at the seepage face also after the second pulse but the values are not as high as after the third pulse. We will change the sentence to "At the highest value, significant levels of ^2H-labeled water appeared in the outflow discharge after the second pulse, whereas in the measured data and in the model results for the smaller dispersivity values the levels were much lower."

(21) Page 16, line 11: You cannot claim that a ~50% increase of the seepage face concentration after the third event is a slight increase.

32) The term "slight" here for the change in seepage face concentration from 4% to 8% is intended in contradistinction to the change in the amount of tracer mass remaining in storage (90% to 40% – also a ~50% change, but of much more significant magnitude).

(22) Page 16, line 17: The definition of c implies: $0<c<1$. What do you mean by a tracer concentration as high as 15? It would be interesting to show some vertical profiles of the water content and concentration.

33) The concentration is normalized with respect to the maximum value (0 deficit). Higher values can occur 1) in the presence of localized injection (adding a source with concentration different from 0) or 2) when only water evaporates, as the same amount of mass that remains in the system would become more concentrated, which is what happens in our case. In the revised manuscript we will remove the sentence: “Further investigation is needed to understand whether this phenomenon is physically realistic or a numerical artifact.” We agree that it would be interesting to see some vertical profiles of water content and concentration but as this is not a main point of the article we prefer to avoid introducing new figures.

(23) Page 18, line 2: In the first simulation, a part of the isotope tracer may evaporate but it is not all lost by evaporation.

34) Here we wanted to say that for the previous simulation all mass in solution with the evaporating water was lost. We agree that it is a little bit confusing and we will rephrase the sentence.

(24) Page 20, line 2, and Page 21, line 4: Please add the name of the simulation: simulation l from Table 4. More generally, indicate in all figures the name of the simulations as specified in Table 4.

35) We will do it. In particular, in the caption of each figure we will indicate the name of the simulations as reported in Table 4.