This document contains a point-by-point response to the three reviews, where the relevant changes made in the manuscript are highlighted in blue (for Reviewer 1), red (for Reviewer 2), and in pink (for Reviewer 3) and refer to the marked-up manuscript version reported below.

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

Reviewer #1 (Claude Mugler)

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 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) The peak 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 d²H 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 have add this information to the section of the paper that describes the experiment. (See page 5, lines 8-13)

(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 have replaced Figure 10 in the original manuscript with these two more detailed figures, and we have revised the description of these results accordingly. (See Figures 11 and 12; page 19 lines 3-4 and 14-15)



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 2^{nd} 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 have described this better in the revised manuscript. (See page 10-11, lines 14-1)

(2b) Page 8, line 21: Same as remark (2a): Why did you choose $\lambda = 1 \text{ m-} 1$?

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 have added a note on this in the revised manuscript. (See pages 9-10, lines 21-1)

(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 have added a sentence in the revised paper to better explain this. (See page 11, lines 3-4)

(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 ($\sim 2\%$)?

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 have added in the revised paper a comment on this. (See page 17, lines 3-5)

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 have clarified this in the revised manuscript. (See page 5, line 12)

(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 $12 \text{ mm/h} \sim 1.1 \times 10^{-3} \text{ m}^3/\text{s}$. Please correct the y-scale for Qr in Fig. 1.

16) You are right. We have corrected 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 have done 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 have done it. (See page 7, lines 16-19)

(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: $5 \text{ mm/d} \equiv 5.8 \times 10^{-8} \text{ m/s}$ and $3.9 \text{ mm/d} \equiv 4.5 \times 10^{-8} \text{ m/s}$

24) We have corrected them. (See page 9, line 20)

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

25) We have changed the definition only in the text (See page 9, line 19) since in the revised manuscript Table 1 is not present anymore (see response 10 to Reviewer 3 specific comment)

(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 of the first mauscript that have been corrected (see response 24 above). In accordance, the values of the two tables have been updated. However, Tables 1 and 2 are not present anymore in the revised manuscript (see response 10 to Reviewer 3 specific comment) but with the corrected values Figure 3 has been built.

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

27) We have done it.

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

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

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

where *n* is the total number of observed data available at the different times, Q_i and \hat{Q}_i are the *i*th observed and modeled values, respectively, and \bar{Q}_i is the average of the observed values. (See page 12, lines 26-29)

(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. (See Figure 4)



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 have inserted "(simulation f)" after "sixth flow simulation". (See page 15, line 21)

(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 have changed the sentence to "At the highest value, significant levels of ${}^{2}H$ -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." (See page 16, lines 5-7)

(21) Page 16, line 11: You cannot claim that a \sim 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: $0 \le c \le 1$. 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 have removed 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. (See page 17, lines 17-18)

(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 have rephrased the sentence. (See page 18, line 3)

(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 have done it. In particular, in the caption of each figure we have indicated the name of the simulations as reported in Table 4. (See Figures 5, 6, 7, 8, 11, 12, and 13)

Reviewer # 2

Overview

The manuscript describes in detail efforts to fit a variably-saturated flow and convection-dispersion transport models to a very highly controlled field-scale experiment in a 1:1 physical analog of a hillslope. The physical models and numerical approximations are described in mathematical terminology (e.g. zero Neumann for no flow boundary condition etc.) yet readable for a wide community of hydrologists. The beauty of the paper is in the clear description of the need to increase the complexity of the model in the process of fitting first the integrated flow response (in which unique transient observations of seepage face flow-rate, and total storage of water are available in this experimental system) than the integrated transport response, than further complexity is needed to fit point observations of water content and concentrations. The model does not go very far with complexity, it starts in uniform hydraulic properties, moving to different properties near the seepage face and layered porous medium but does not go further to variability within layers, or mobile immobile formulations etc.. Hence, the well-known, good fit of the macro phenomenon relatively to poorer fits to point observations is described very clearly.

Recommendation

I am not sure there is completely new modeling knowledge here, nevertheless, the paper has "educational quality" for hydrological modelers as well as very unique experimental data (although not in focus in the manuscript), and therefore, I warmly recommend publication in HESS, following the authors pay their attention to the comments herein.

We wish to thank the Reviewer for the attention to our work. The major and specific comments raised by the Reviewer are addressed below.

Major Comments

1) Title – The hillslope problem as well as the model used here and the results of the experiment, are variably saturated rather than unsaturated (saturation at 85 cm for significant duration of the experiment in most locations, Figure 11). Suggest to change to: Multiresponse modeling of variably saturated flow and isotope tracer transport in a hillslope experiment at the Landscape Evolution Observatory

1) We like the title proposed by the Reviewer and have adopted it, changing only "in a hillslope" to "for a hillslope". (See new title)

2) Discussion - In line with the previous comment. I don't understand why the authors do not discuss the more specific setup of a hillslope that was studied here, rather than concentrating on general unsaturated flow. The hillslope case has significant differences than the general unsaturated zone (variably saturated, lateral flow component dominant, relations with evaporation and runoff etc.). Many simulation studies of hillslopes can be discussed (e.g., Fiori and Russo, 2008 WRR).

2) In the Discussion section we wished to draw attention to the numerous challenges in modeling solute transport (rather than flow) phenomena in the unsaturated zone, given that this is the area that in our opinion raised the most difficulties - physical and numerical - in our modeling of the LEO experiment, in particular with regards to capturing point-scale responses. We did not make a specific distinction between field, hillslope, and (small) catchment-scale studies, as we feel that the issues and proposed explanatory hypotheses (last paragraph of the Discussion) apply across the board. Nonetheless, taking the cue from the Reviewer's concern about a hillslope focus and the mention of the Fiori and Russo paper, we have added a mention of transit time distribution research at the hillslope scale, as this is an excellent example of the need for better modeling of flow and transport dynamics and a very active area of current research. The first two sentences of the Discussion section have thus become three and read: "Mass transport in unsaturated soils is extremely important in the context of biosphere, critical zone, and Earth systems research because of exchanges of water and solutes that occur across the land surface interface. The study of hillslope transit time distributions (e.g., Fiori and Russo, 2008; Botter et al., 2010; Heidbüchel et al., 2013; Tetzlaff et al., 2014) is a good example of the need for a better understanding of such water and solute exchanges and the consequent subsurface flowpaths. The simulation of unsaturated zone mass transport phenomena is however known to be a particularly complex problem, ...". (See page 22, lines 9-12)

3) List of symbols – There are many symbols in equations and within the text. For example, it took me too long to find what does the nee in line 27 page 7 stands for. I suggest adding a list of symbols at the beginning of the paper.

3) We do not think a list of symbols is warranted for a standard-length paper. We have however examined all the places in the paper where symbols are used.

4) Use of the term heterogeneity – is misleading. Changing a homogenous model deterministically to have lower Ks near the seepage face, or different hydraulic properties at different layers doesn't make it a heterogeneous model (a term now used for a medium in which the properties vary from pixel to pixel randomly usually constrained to a PDF and a spatial correlation function). I suggest describing this type of additional complexity with different (more explicit) terms (e.g. low Ks at seepage face, layered n(vg) etc.), throughout the text, tables and figures.

4) We agree with this comment and have made the suggested changes throughout the paper (including the figures and tables). (See page 1, lines 12 and 13; page 12, lines 4 and 5; page 14, line 6; Table 3; page 15, line 5; page 25, line 10; Figures 13 and 14)

5) Fractionation? - in water isotopes during evaporation. The term fractionation is brought up late in the methods section (page 8) as if it is totally trivial. I suggest to add a paragraph on fractionation of water isotopes during evaporation in the introduction to introduce the topic before jumping into the details of dealing with modeling it in the methods section.

5) Although we totally agree that fractionation is not a trivial topic, it is nonetheless not a main topic of the paper. We therefore prefer not to emphasize it too strongly in the Introduction, as this would entail having to also describe the other configurations that were tested (this is all done in Section 3.3). We have thus added, rather than adding an entire paragraph on fractionation, the following sentence at the very end of the Introduction: "The boundary condition configurations, for instance, includes a sink-based treatment of isotope fractionation to allow only a portion of the tracer to evaporate with the water." (See page 3, lines 14-16)

6) Van Genuchten (1985) - should be van Genuchten (1980). It would have been a specific comment for any other paper in hydrology (p. 6, l. 15 and in reference list).

6) Thanks for catching this. The error maybe crept in because we often cite the van Genuchten and Nielsen 1985 paper for these constitutive relationships. We have corrected "1985" to "1980".

Specific Comments

1) P. 4, Figure 1. a) Lowest pane (delta2H) – zoom into the interval of interests in the vertical axis (< 53); b) say something on the high readings at the beginning before tracer introduction, and just before the third rain pulse. Or looking at Figure 4 there seems to be a shift of the data to the left? Solve the problem, explain.

1) Indeed there is an error in Figure 1, thanks for making us realize this. The delta2H graph should be shifted by 23.5 h, as is evident from Figures 4, 6, 8, and 13. The corrected figure is shown below. The high readings straight after the beginning of the tracer introduction is a point that was also raised by the Reviewer 1. The peak in the early seepage face flow 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 and reflects some mixing of the new irrigation water with the evaporatively-enriched residual soil moisture. We have added this information to the section of the paper that describes the experiment. Note that in the corrected graph there is no longer a high reading before the third rain pulse. As suggested we have also rescaled the y-axis in this graph. (See Figure 1 in the revised paper)



2) P. 5, l. 14, Eq. 2. I suggest to add the sink\source term -f(c) to the 2H transport equation here as well, rather than only elaborating on it in table 2 and related text.

2) Thanks for the suggestion. We have added two terms in the transport equation: $qc^*[M/TL^3]$, with $c^{*}=c$ if q (the source/sink term in the flow equation 1) is a sink term, otherwise an imposed source concentration, and f_c , which is a generic source/sink term (our correction term used to model fractionation). (See Equation 2; page 6, lines 2-4; page 8, lines 22-23)

3) P. 7, l. 27. Shouldn't the left hand side of the equation be n (or theta) v^{nee} , rather than only v^{nee} (porous medium approximation of ratio of flux and velocity).

3) The boundary term arising from the finite element P1 Galerkin discretization of equation 1 (at the left hand side of the = sign) is:

$$-\int_{\Gamma_f} K_r(\psi) K_s(\nabla \psi + \eta z) \cdot \nu \,\mathrm{d}\Gamma_f$$

where Γ_f is the Neumann boundary of the domain for the flow. Thus, the equation in line 27 is correct. You can find more details on the flow equation discretization in Scudeler et al. (2016), cited in the paper.

4) P. 10 l. 5-8. Excellent lines – don't touch, makes it so much easier to follow the long descriptions after.

4) Thanks.

5) *P. 11, l. 11. The evaporation rates – were they calculated from the water balance and the load cell data? Or how? Please elaborate.*

5) The evaporation rates were calculated from the seepage face measurement and the load cell data. In particular, being V_{sf} the cumulative volume flowing out from the seepage face between two events or after the last event until the end of the experiment (information directly obtained from the flow meter measurements), being dV the change in water volume between two events or after the last event until the end of the experiment (information directly obtained from the load cell data), being dT the time interval between two events or after the third event until the end of the experiment, the average evaporation rate was calculated as $(-V_{sf}+dV)/dT$. We have added this explanation on page 4 after line 10. (See page 5, lines 18-20)

6) P. 15 Figure 4. A) Say something on the early breakthrough during the heavy isotope injection. B) Elaborate in the text why was the high dispersivity simulation so much biased upwards in the mass of tracer exiting the system (earlier arrival times are expected in high dispersivity but also late ones. What were the left-in-storage or evaporated components of the mass balance in the high dispersivity run?

6) A) Do you mean for the measured response? If so we have answered this point above (see response to specific comment 1) and we are going to add that information in the revised manuscript. It is also true that the model response presents an early breakthrough after the injection of isotope as we have nonzero solute concentration at the seepage face also after the second pulse but the values are not as high as after the third pulse. These early breakthroughs may be due to dispersion effects. B) We show below the mass balance results for the simulation relative to the 0.1 m longitudinal dispersivity case. At the end of the simulation almost 40% of the mass injected has flown out from the seepage face, 16% has evaporated, and 42% has remained in storage (compared to, respectively, ~4%, 52%, and 42% of the α_{i} =0.001 m case). 0.1 m is 1/10 of the depth of the hillslope, compared to 0.001 m and 0.01 which are only the 1/1000 and 1/100, respectively. The effect of the high dispersivity makes the solute percolate down quickly to then flow out of the domain from the seepage face boundary. As a consequence it is also less exposed to evaporation. We would expect another breakthrough with a fourth pulse of rain. Unfortunately, no measurements were taken after 336 h. We have not added the figure below to the paper but we discuss a little bit more in detail the results for the 0.1 m case. (See page 16, lines 3-5)



Simulated mass balance results for the $\alpha_l=0.1$ m case.

7) P. 16, Figure 6 and related text: solute is not a proper term for ${}^{2}H$.

7) We agree, hydrogen isotopes are not solutes. In the revised manuscript we no longer refer to deuterium as a solute. (See page 1, line 5; Table 3; Figures 7 and 9)

Reviewer # 3

Summary:

This manuscript provides detailed comparisons of multiple hydrologic response variables using a sophisticated integrated hydrology model and highly controlled experiment at the Landscape Evolution Observatory. The authors experiment with different levels of complexity within the model and demonstrate the importance of model heterogeneity if the goal of the model is to match spatially distributed points as opposed to integrated responses. Results also indicate the importance of considering more than just integrated hydrologic response variables when determining model parameters.

Recommendation:

Overall I find the paper to be well written. I think it provides an interesting comparison of a state of the art experiment with state of the art modeling that will be interesting to the hydrologic community and should be published in HESS. I find their scientific approach to be sound; however, I do think that some changes to the manuscript to better outline all of the test cases and highlight differences would make the discussion easier to follow. I also think that the manuscript would be of broader interest if the authors would devote some discussion the relevance of these findings to other commonly used or similar modeling approaches. I have provided detailed suggestions to this effect below.

We wish to thank the Reviewer for the attention to our work. The comments raised by the Reviewer are addressed below.

Major Comments:

1. The introduction is focused on the need for multi objective parameter optimization. This is a good motivator for this work, but also the study is not really presenting advances for parameter optimization. Rather it's evaluating the impact of different parameterizations on model response. Therefore, I think it would be helpful to provide more background on heterogeneity and variably saturated flow processes and the state of the practices for both modeling and observations. I think this would provide a better context for where both the modeling and observations used here compare to previous work.

1) We agree with the reviewer that the paper is not really about parameter optimization. In the same sense, neither is it really about heterogeneity, as we do not conduct a systematic analysis based on complex configurations and innumerable realizations. But parameter estimation and heterogeneity are certainly underlying themes of the paper, and we agree that a mention of heterogeneity in the context of variably saturated flow processes is warranted. In the Introduction we have added a sentence on this and include additional citations. Lines 7-9 of page 2 have become: "... to more

complex models. Traditional challenges, on both experimental and modeling sides, are associated with soil heterogeneity, variability in parameters, and variably saturated conditions (e.g., Binley et al., 1989; Woolhiser et al., 1996; Neuweiler and Cirpka, 2005; see Reference below). An added source of complexity arises when passing from flow modeling to flow and transport modeling (e.g., Ghanbarian-Alavijeh et al., 2012; Russo et al., 2014)." (See page, lines 12-17) See also reply to specific comment 1 below.

2. I would appreciate more details on why the observational experiments were setup the way they were. For example, how were the rainfall rates and timing determined?

2) We have added in section 2 just after the sentence ending "... precipitation at rates between 2 and 40 mm/h." the following additional description of the rain system: "Each landscape at LEO has 5 independent plumbing circuits, each including a different array of sprinkler heads, and therefore generating a different irrigation flux." And in section 3.1 we have added after the first paragraph the following new paragraph providing additional details on the rainfall rates and timing: "At the time of this experiment we consistently used one plumbing circuit because the spatial distribution of irrigation produced by this circuit had been well characterized by in situ testing. This allowed us to examine the possible influence of spatially heterogeneous irrigation patterns on flow and transport. The purpose of the first irrigation application was to increase the average moisture content of the landscape, which had received no irrigation for more than 40 days prior. The second irrigation application was used to introduce the deuterium tracer. No additional irrigation was applied for multiple days so that the tracer transport within, and out of the landscape, would be affected by soilmoisture redistribution and evaporation. The third and final irrigation application was applied with the intention of forcing additional tracer mass beyond the seepage face boundary, to reveal additional detail in the measured breakthrough curve. In retrospect, and following laboratory analysis that spanned several weeks, we only observed the initiation of the tracer breakthrough curve at the seepage face." (See pages 3-4, lines 32-1; page 5, lines 1-9)

3. It can be hard to keep all of the different simulations setups straight throughout the paper. I think this could be addressed by expanding on Figure 2 to better label different aspects of the domain that are discussed in the model setup and creating a new table or conceptual model that summarizes all of the runs in one place.

3) In Figure 2 we now indicate the seepage face and the atmospheric forcing boundary. The revised figure is shown below. The only thing that differs in the model setup amongst the different simulations is the treatment of the atmospheric boundary. With the introduction of a new table containing the boundary condition setup for each simulation it should be easier to follow the setup of the different simulations (see response to specific comment 7 below). (See Figure 2 in the revised manuscript)



Figure 2. 3D numerical grid for the LEO landscape. Points a, b, c, and d are the locations where samples were extracted during the experiment for subsequent laboratory analysis.

4. The discussion of differences between basins is mostly qualitative. I think some additional figures that plot differences between scenarios for key metrics and discussion points would strengthen the conclusions.

4) We are not sure we understand what is meant by 'differences between basins'. Does the reviewer mean the 3 different LEO hillslopes? The experiment being analyzed in this paper was conducted on just one of the 3 hillslopes. This is mentioned in the first sentence of section 3.1 ("...performed at the LEO-1 hillslope ...") but we agree that this is not at all very clear. We have added the following clarification after the first sentence of the last paragraph of the Introduction: "Both of these experiments were performed on the first of the three hillslopes at LEO to be commissioned, hereafter referred to as LEO-1." We believe that the revisions to the main text and figures that we are bringing to the paper, in response also to the other two reviewers, help strengthen the conclusions (see, for example, response 3 to Reviewer 1 and response 6 to the specific comments of Reviewer 2). (See page 3, lines 6-7)

5. This study uses the CATHY model, but it is focuses on addressing larger questions in model uncertainty and parameterizations. Given this goal I think some additional discussion on the degree to which these results are specific to the model you are using or would be universal to other integrated flow and transport models would be quite helpful.

5) We agree with this remark and have added the following new paragraph at the end of the Discussion section: "The broad results of our study should be quite universal, particularly to deterministic numerical models based on the 3D Richards and advection-dispersion equations. However, any model has its specific features and differs, for example, in the way equations are coded (e.g., choice of numerical solvers) or interface conditions are implemented (e.g., free-surface vs boundary condition switching). For insights on the impact of specific model differences in the performance of CATHY-like models, see the intercomparison studies of Sulis et al. (2010; see References below) and Maxwell et al. (2014; already cited in the paper). These intercomparison studies have thus far focused only on flow processes, and there is an urgent need to extend the analyses to solute transport phenomena, in order to properly guide our assessment of the physical and numerical correctness of competing models as these models continue to increase in complexity. For instance for this study there are aspects of the CATHY model related to how we implemented evaporation and fractionation that might be expected to negatively impact the generality of our findings, although in terms of isotope tracer mass exiting the seepage face the impact was quite small. But the implementation here was somewhat ad hoc, and more study is needed on the

importance and proper representation of fractionation in solute transport models, especially under strongly unsaturated conditions." (See page 26, lines 17-28)

Specific Comments:

1. Page 2, line 8: Please expand on this point. What do you mean by 'an important example of this complexity'? Are you saying that parameter estimation has been particularly challenging for mass transport?

1) We are alluding here to the added complexity (more equations, more parameters, etc) when passing from flow to flow and transport modeling. We have clarified this sentence to: "An added source of complexity arises when passing from flow modeling to flow and transport modeling (e.g., Ghanbarian-Alavijeh et al., 2012; Russo et al., 2014)." This sentence is also made clearer with the new sentence added just before this one (see reply to major comment 1 above).

2. Page 3, line 6: Clarify, "infrastructure" for what?

2) We have changed this to "research infrastructure". (See page 3, line 18)

3. Page 3, line 10: From this description it sounds like a simple sloping slab but from Figure 2 it appears that it is actually a tilted v sloping to the center of the domain. Please clarify. Also you could annotate the slopes on Figure 2 to make this even more clear.

3) In fact LEO consists of three v-shaped hillslopes. The average slope of each landscape is 10° , as stated in the paper, while the local slope varies from upslope positions to the convergence zones, with maximum slope of 17° near the convergence zone. Since it is difficult to incorporate this information graphically in Figure 2, we have added this information to the text, in section 2. (See page 3, lines 24-25)

4. Page 4, line 2: You should clarify that you are talking about just the rain from the first event here not 'all the rain water'.

4) We are talking about all rain water. The confusion is perhaps due to the phrase "... and generated seepage face outflow that started after 5 h" at the end of this sentence. We have split this sentence in two: "All the rain water applied infiltrated into the soil. Seepage face flow started 5 h after the beginning of the experiment." (See page 5, lines 11-13)

5. Page 4, line 2: Also here you switch from using the term 'irrigation' to 'rain'. It will be easier to follow if you pick one term and stay consistent.

5) We agree that this is inconsistent and may cause confusion. Since we use the terms "rainfall" / "rain" / "precipitation" more than "irrigation" in the paper, we now use these former terms exclusively. (See page 4, lines 6-12; page 8, line 18)

6. Page 7, line 14: Please expand here to clarify how you decided on this lateral resolution.

6) In the revised manuscript we now explain why we have chosen this discretization (See page 8, lines 9-12). See also our response 23 to Reviewer 1, comment 12.

7. Page 7, line 25: This is a very dense and long sentence. In my opinion it would easier to follow and refer

back to if this information were provided in the form of a table. Also, if you keep this in paragraph form you should tie the three numbered experiments listed to simulations a-f in Table 3?

7) We have added the suggested table (it became Table 1 in the revised manuscript), shown below, and we have revised (shorten and simplify) the paragraph in question. (See Table 1; pages 8-9, lines 25-1)

~								
Simulation	Rain with ² H-en	riched water	Rain with no ² H	-enriched water	Evaporation (between			
(see Tables	(second pulse)		(first and third p	ulses)	rain pulses and after the			
4 and 5)					third pulse)			
a-f, g-i	Flow Transport		Flow	Transport	Flow	Transport		
	q_n^{f} =-12 mm/h	$q_{c}^{t} = v \bullet vc^{*}, c^{*} = 1$	q_n^{f} =-12 mm/h	$q_c^t = 0$	$q_n^{f} = 5$ or	$q_n^t = 0$		
					3.9 mm/h			
j	q_n^{f} =-12 mm/h	$q_{c}^{t} = v \bullet vc^{*}, c^{*} = 1$	q_n^{f} =-12 mm/h	$q_c^t=0$	Sink q	Source f_c		
		-			(Table 1)	(Table 2)		
k	q_n^{f} =-12 mm/h	$q_{c}^{t} = v \bullet vc^{*}, c^{*} = 1$	q_n^{f} =-12 mm/h	$q_c^t = 0$	Sink q	Source f_c		
					(Table 1)	with $f_c = -qc$		

Table 1. Treatment of boundary conditions at the land surface during the rainfall and evaporation periods for the flow and transport models.

8. Page 8 line 16: How did you determine the 38cm depth for evaporation? This seems arbitrary.

8) This is a point that has also been raised by another reviewer (see our response 6 to comment 2a of Reviewer 1). 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 have described this better in the revised manuscript.

9. Page 9 line 6: It would be helpful to have visual on your model figure for where the seepage face is occurring.

9) In the revised manuscript we show graphically, in Figure 2, where the seepage face is set. The revised figure and caption are shown above (see response to major comment 3 above).

10. In my opinion the source sink terms listed in Tables 1 and 2 would be more easily interpreted graphically. Alternatively, I'm not sure that this information is necessary for the interpretation of the results as long as you describe how you got these terms so potentially these tables could also be deleted.

10) We agree and have replaced Tables 1 and 2 with the figure shown below. (See Figure 3 in the revised manuscript; page 11, line 3)



Figure. Sink term q and source term fc over depth z added to the flow and transport equation, respectively. q1 and fc1 are applied between rain pulses 1, 2, and 3, while q2 and fc2 are applied after rain pulse 3.

11. Table 3: Why is simulation e repeated twice in this table

11) This is a mistake, the last one should be "f".

12. Table 4 is difficult to follow. I think you need a separate table describing the setup of runs g-l and then report only the output metrics in this table. Also, it might help to just focus on runs g-l here and add the information for simulations a-f to Table 3.

12) We think it is important to keep Table 4 since it summarizes all simulations performed and makes it easier to follow the text in both the Simulations performed subsection and in the Results section.

13. Figure 3: Please describe what 'simulated, preceding case' means in the caption.

13) The "preceding case" results have been removed from this figure (see our response 29 to Reviewer 1, comment 18, that also includes the new figure and caption).

14. Figures 4, 6 and 8: I think the diamonds for the measured values should be smaller so that they are not overlapping each other or the axes so much.

14) We agree. We show below the new Figure 4 with smaller diamonds for the measured values. We have done this also for Figures 1, 6, 8, 12, and 13. (See Figures 1, 5, 7, 9, and 15 in the revised manuscript)



References

Binley, A., J. Elgy, and K. Beven (1989), A physically based model of heterogeneous hillslopes: 1. Runoff production, Water Resour. Res., 25(6), 1219–1226, doi:10.1029/WR025i006p01219

Woolhiser, D. A., R. E. Smith, and J.-V. Giraldez (1996), Effects of spatial variability of saturated hydraulic conductivity on Hortonian overland flow, Water Resour. Res., 32(3), 671–678, doi:10.1029/95WR03108.

Neuweiler, I., and O. A. Cirpka (2005), Homogenization of Richards equation in permeability fields with different connectivities, Water Resour. Res., 41, W02009, doi:10.1029/2004WR003329.

Sulis, M., S. B. Meyerhoff, C. Paniconi, R. M. Maxwell, M. Putti, and S. J. Kollet (2010), A comparison of two physics-based numerical models for simulating surface water–groundwater interactions, Adv. Water Resour., 33 (4), 456–467, doi:10.1016/j.advwatres.2010.01.010.

Multiresponse modeling of an unsaturated zone isotope tracer experiment^{R2} variably saturated flow and isotope tracer transport for a hillslope experiment^{R2} at the Landscape Evolution Observatory

Carlotta Scudeler^{1,2}, Luke Pangle³, Damiano Pasetto⁴, Guo-Yue Niu^{5,6}, Till Volkmann⁶, Claudio Paniconi¹, Mario Putti², and Peter Troch^{5,6}

¹Institut National de la Recherche Scientifique, Centre Eau Terre Environnement (INRS-ETE), Université du Québec, Quebec City, Canada

²Department of Mathematics, University of Padova, Padova, Italy

³Department of Geosciences, Georgia State University, Atlanta, USA

⁴Laboratory of Ecohydrology, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

⁵Biosphere 2 - Earth Science, University of Arizona, Tucson, USA

⁶Department of Hydrology and Water Resources, University of Arizona, Tucson, USA

Correspondence to: C. Scudeler (carlotta.scudeler@ete.inrs.ca)

Abstract. This paper explores the challenges of model parameterization and process representation when simulating multiple hydrologic responses from a highly controlled unsaturated flow and transport experiment with a physically-based model. The experiment, conducted at the Landscape Evolution Observatory (LEO), involved alternate injections of water and deuteriumenriched water into an initially very dry hillslope. The multivariate observations included point measures of water content and

- 5 tracer concentration in the soil, total storage within the hillslope, and integrated fluxes of water and solute^{R2} tracer^{R2} through the seepage face. The simulations were performed with a three-dimensional finite element model that solves the Richards and advection-dispersion equations. Integrated flow, integrated transport, distributed flow, and distributed transport responses were successively analyzed, with parameterization choices at each step supported by standard model performance metrics. In the first steps of our analysis, where seepage face flow, water storage, and average concentration at the seepage face were the
- 10 target responses, an adequate match between measured and simulated variables was obtained using a simple parameterization consistent with that from a prior flow-only experiment at LEO. When passing to the distributed responses, it was necessary to introduce heterogeneity^{R2} complexity^{R2} to additional soil hydraulic parameters to obtain an adequate match for the point-scale flow response. Augmented heterogeneity^{R2} This^{R2} also improved the match against point measures of tracer concentration, although model performance here was considerably poorer. This suggests that still greater complexity is needed in the model
- 15 parameterization, or that there may be gaps in process representation for simulating solute transport phenomena in very dry soils.

1 Introduction

Simulation models of water and solute interaction and migration through complex geologic media are essential tools for addressing fundamental and practical problems, ranging from basic scientific understanding of critical zone processes (Brooks et al., 2015) to improving the management of our freshwater resources (Gorelick and Zheng, 2015). Physically based dis-

- 5 tributed numerical models require a careful definition of spatially variable parameters and time variable boundary conditions, and can produce information for numerous response variables at different levels of spatio-temporal aggregation. It is increasingly acknowledged that proper implementation and verification of these models, in terms of both process representation and parameter identification, requires detailed, multiresponse field or laboratory data, in contrast to traditional model evaluation based on a single, integrated response variable such as total discharge (Paniconi and Putti, 2015). However, multiobjective pa-
- 10 rameter estimation for nonlinear or coupled models with a high number of degrees of freedom is very challenging (Anderman and Hill, 1999; Keating et al., 2010), since classical techniques developed for simpler hydrological models (e.g., Gupta et al., 1998; Fenicia et al., 2007) are not readily extendable, in terms of robustness and efficiency, to more complex models. Traditional challenges, on both experimental and modeling sides, are associated with soil heterogeneity, variability in parameters, and variably saturated conditions (e.g., Binley et al., 1989; Woolhiser et al., 1996; Neuweiler and Cirpka, 2005). An added
- 15 source of complexity arises when passing from flow modeling to flow and transport modeling (e.g., Ghanbarian-Alavijeh et al., 2012; Russo et al., 2014).^{R3} An important example of this complexity arises in the modeling of mass transport phenomena in unsaturated soils (e.g., Ghanbarian-Alavijeh et al., 2012; , 2014).^{R3}

While many hydrologic model assessment studies have reported good agreement between simulated and observed data when performance is measured against a single response variable, there are comparatively few studies that have made use of

- 20 observation data from multiple response variables. Brunner et al. (2012), for instance, examined the performance of a onedimensional (1D) unsaturated zone flow model when water table measurements were supplemented by evapotranspiration and soil moisture observations. Sprenger et al. (2015) assessed the performance of three inverse modeling strategies based on the use of soil moisture and pore water isotope concentration data for a 1D unsaturated flow and transport model. Kampf and Burges (2007) obtained encouraging results for a 2D Richards equation flow model using integrated (subsurface outflow) and
- 25 internal (piezometric water level and volumetric water content) measurements from a hillslope-scale experiment. Kumar et al. (2013) used multiple discharge measurements to calibrate and apply a distributed hydrologic model to 45 subcatchments of a river basin in Germany. Investigations based on hypothetical experiments are more common. Mishra and Parker (1989), for example, obtained smaller errors for simultaneous estimation of flow and transport parameters than for sequential estimation based on synthetically-generated observations of water content, pressure head, and concentration.
- 30 In this study we perform a modeling analysis of the experimental data collected from an intensively-measured hillslope at the Landscape Evolution Observatory (LEO) of the Biosphere 2 facility (Hopp et al., 2009). The simulations were conducted with the CATHY (CATchment HYdrology) model (Camporese et al., 2010; Weill et al., 2011), a physics-based numerical code that solves the 3D Richards and advection-dispersion equations and includes coupling with surface routing equations. The availability of extensive observational datasets from detailed multidisciplinary experiments (recent examples in addition

to LEO include the TERENO network of experimental catchments (Zacharias et al., 2011) and the Chicken Creek artificial catchment (Hofer et al., 2012)) can contribute vitally to testing and improving the current generation of integrated (surface-subsurface) hydrological models (Sebben et al., 2013; Maxwell et al., 2014).

Two experiments have been conducted to date at LEO, a rainfall and drainage test in February 2013 (Gevaert et al., 2014; Niu 5 et al., 2014), which featured both subsurface and overland flow, and an isotope tracer test in April 2013 (Pangle et al., 2015), run under drier soil conditions and with reduced rainfall rates to avoid occurrence of surface runoff. Both of these experiments were performed on the first of the three hillslopes at LEO to be commissioned, hereafter referred to as LEO-1.^{R3}

Using both integrated (load cell and seepage face) and distributed (point-scale soil moisture and concentration) data collected during the tracer experiment, the objective of this study is to explore the challenges of multiresponse performance assessment

- 10 for a 3D variably saturated flow and solute transport model. In a first step we consider only integrated flow responses, and the CATHY model is initially parameterized according to analyses of the February 2013 experiment. As integrated transport and point-scale flow and transport observations are progressively introduced in the analysis, the impact of different configurations (spatially uniform versus spatially variable parameters, treatment of initial and boundary conditions, etc) on the model's ability to capture the expanding and increasingly detailed response variables is examined. The boundary condition configurations,
- 15 for instance, include a sink-based treatment of isotope fractionation to allow only a portion of the tracer to evaporate with the water.^{R2}

2 Study site: Biosphere 2 Landscape Evolution Observatory

LEO is a large-scale community-oriented research^{R3} infrastructure managed by the University of Arizona at the Biosphere 2, Oracle, U.S.A. (Hopp et al., 2009; Huxman et al., 2009; Pangle et al., 2015). It consists of three identical convergent artificial landscapes (or hillslopes) constructed with the aim of advancing our predictive understanding of the coupled physical, chemical, biological, and geological processes at Earth's surface in changing climates. For the first years of LEO operation, vegetation is not present and the research is focused on the characterization of the hydrological response of the hillslopes in terms of water transit times, generation of seepage and overland flow, internal dynamics of soil moisture, and evaporation. The three hillslopes are of 10° average slope and^{R3} are 30 m long and 11 m wide and of 10° average slope. The local slope varies from upslope positions to the convergence zone, with a maximum slope of 17° near the convergence zone.^{R3} The landscapes are filled with 1 m of basaltic tephra ground to homogeneous loamy sand, chosen mainly for its primary elemental composition that includes critical nutrients for plant growth. The three landscapes are housed in a 2000 m² environmentally controlled facility. Each landscape contains a sensor and sampler network capable of resolving meter-scale lateral heterogeneity and

30 at which they can be polled allows for a monitoring intensity that is impossible to achieve in natural field settings. Additionally, each landscape has 10 load cells embedded into the structure that allow measurement of changes in total system mass and an engineered rain system that allows application of precipitation at rates between 2 and 40 mm/h. Each landscape at LEO has 5 independent plumbing circuits, each including a different array of sprinkler heads, and therefore generating a different rain

submeter-scale vertical heterogeneity in water, energy, and carbon states and fluxes. The density of sensors and the frequency



Figure 1. Hydrological response to the tracer experiment at the LEO-1 hillslope. From top: measured rain input pulses Q_r (the red pulse is deuterium-enriched); seepage face flow Q_{sf} ; total water storage V_s ; and mean $\delta^2 H$ values at the seepage face. Time 0 corresponds to 9:30 am, 13 April 2013. The vertical dashed lines indicate the timing of the three pulses of rain (red when the water is deuterium-enriched and blue when it is not).

flux.^{R3} Tracers can be introduced into the system via the rainfall simulator at a constant or time-varying rate. The embedded soil water solution and soil gas samplers facilitate the use of these tracers to study water and solute movement within the hillslopes at a very dense spatial scale.

3 Methodology

5 3.1 Isotope tracer experiment

The first tracer experiment performed at the LEO-1 hillslope began at 9:30 am on April 13, 2013. The experiment consisted of three irrigation^{R3} rainfall^{R3} events that were applied over 10 days (Fig. 1). During each event the irrigation^{R3} rainfall^{R3} was applied at a rate of 12 mm/h for durations respectively of 5.5 h, 6 h, and 5.25 h. Irrigation^{R3} Rainfall^{R3} was interrupted for 2.75 h during the third event (1.25 h from the start) due to necessary equipment maintenance, then restarted. During the

10 second event deuterium-enriched water was introduced into the irrigation^{R3} rain^{R3} system. The enriched water had a hydrogen isotopic composition (expressed using the delta-notation as $\delta^2 H$) of approximately 0%, which corresponds to an enrichment of approximately 60% compared to typical (non-enriched) irrigation^{R3} rainfall^{R3} source water. At the time of this experiment we consistently used one plumbing circuit because the spatial distribution of rainfall produced by this circuit had been well characterized by in situ testing. This allowed us to examine the possible influence of spatially heterogeneous rain patterns on flow and transport. The purpose of the first rain application was to increase the average moisture content of the landscape, which had received no rain for more than 40 days prior. The second rainfall application was used to

- 5 introduce the deuterium tracer. No additional rain was applied for multiple days so that the tracer transport within, and out of the landscape, would be affected by soil moisture redistribution and evaporation. The third and final rainfall application was applied with the intention of forcing additional tracer mass beyond the seepage face boundary, to reveal additional detail in the measured breakthrough curve. In retrospect, and following laboratory analysis that spanned several weeks, we only observed the initiation of the tracer breakthrough curve at the seepage face.^{R3}
- The initial conditions of the system were very dry. The estimated total initial volume of water was about 26 m³ (the total water storage capacity of the hillslope is approximately 130 m³). All the rain water applied infiltrated into the soil.^{R3} and generated seepage^{R3} Seepage^{R3} face outflow at the downslope vertical plane^{R1} that^{R3} started after^{R3} 5 h after the beginning of the experiment^{R3}. Two outflow peaks were observed: the first one after the second pulse of rain, with a peak of 4.5×10^{-5} m³/s, and the second one after the final pulse, with a peak of 2.1×10^{-5} m³/s. Temporal changes in total soil water storage were
- 15 monitored via the load cell measurements, flow from the seepage face boundary was measured with electronic flow meters and tipping bucket gauges, and matric potential and water content were measured at 496 locations with, respectively, MPS-2 and 5TM Decagon sensors installed at depths 5 cm, 20 cm, 50 cm, and 85 cm from the landscape surface. Cumulative fluxes and instantaneous state variables were recorded at 15-min intervals. The estimated evaporation rate, derived based on water balance calculations^{R2} from the seepage face measurements and load cell data and calculated as the difference between the change in
- water volume and the cumulative volume flowing out from the seepage face over the selected time interval,^{R2} was, on average, 1.9×10^{-5} m³/s (5.0 mm/d) between rain pulses and 1.5×10^{-5} m³/s (3.9 mm/d) after the third rain pulse.

The movement of the deuterium-enriched water within and out of the landscape was monitored through manual sampling and subsequent laboratory analysis. Prenart quartz water sampling devices were used to extract soil water samples periodically throughout the experiment. Data reported in this manuscript include samples collected at 5, 20, 50, and 85 cm depth from

- surface at the four locations shown in Fig. 2. Flow from the seepage face boundary was collected with a custom autosampler (sampling cylinders of 5 cm length and 3 cm circumference). The deuterium concentration within all water samples was measured via laser spectroscopy (LGR LWIA Model DLT-100) at the University of Arizona. Analytical precision was better than 0.5% for $\delta^2 H$. All isotopic data are expressed relative to the international reference VSMOW or VSMOW-SLAP scale. The seepage face isotopic data indicate that the residual soil water in the landscape prior to the experiment had become enriched
- 30 in deuterium (compared to the rainfall water) during evaporation. In fact, during evaporation, hydrogen preferentially goes into the vapor phase compared to deuterium, so that the liquid phase remaining in the soil easily becomes deuterium-enriched. Thus, the $\delta^2 H$ values in the early seepage face flow may reflect some mixing of the new rain water with the evaporatively-enriched water. This slight enrichment disappears in the seepage flow at later times because of the dilution by the newly infiltrating water.^{R1}



Figure 2. 3D numerical grid for the LEO landscape. Points a, b, c, and d are the locations where samples were extracted during the experiment for subsequent laboratory analysis.

3.2 Hydrological model

5

The CATHY (CATchment HYdrology) model (Camporese et al., 2010) used to simulate the isotope tracer experiment has been previously implemented for LEO to study coupled surface and subsurface flow (Niu et al., 2014) and sensor performance (Pasetto et al., 2015). The description here will thus be limited to aspects pertaining particularly to the implementation for LEO of the solute transport component of the model. The numerical solver for the advection-dispersion transport equation is described in detail in Putti and Paniconi (1995), and, like the flow solver, is based on a three-dimensional finite element discretization in space and a weighted finite difference discretization in time. The velocity field and nodal saturation values computed by the flow solver are passed as input at given time steps to the transport solver. The governing equations for the flow and transport solvers are:

10
$$S_w S_s \frac{\partial \psi}{\partial t} + n \frac{\partial S_w}{\partial t} = \nabla \cdot [K_r(\psi) K_s(\nabla \psi + \eta_z)] + q$$
(1)
$$\frac{\partial (n S_w c)}{\partial t} = \nabla \cdot [K_r(\psi) K_s(\nabla \psi + \eta_z)] + q$$
(1)

$$\frac{\partial(nS_wc)}{\partial t} = \nabla \cdot (D\nabla c) - \nabla \cdot (vc) + qc^* + f_c \tag{2}$$

where $S_w = \theta/\theta_s$ is the water saturation [-], θ is the volumetric moisture content [-], θ_s is the saturated moisture content [-](generally equal to the porosity n [-]), S_s is the aquifer specific storage coefficient [1/L], ψ is the pressure head [L], t is the time [T], ∇ is the gradient operator [1/L], $K_r(\psi)$ is the relative hydraulic conductivity function [-], K_s is the hydraulic con-

15 ductivity tensor [L/T] (considered to be diagonal, with k_s the saturated hydraulic conductivity parameter for the isotropic case and k_v and k_h , respectively, the vertical and horizontal hydraulic conductivity parameters for the anisotropic case), $\eta_z = (0,0,1)^T$, z is the vertical coordinate directed upward [L], q is a source (when positive) or sink (when negative) term [1/T], c is the solute concentration $[M/L^3]$, D is the dispersion tensor $[L^2/T]$, and $^{R_2}v = (v_1, v_2, v_3)^T$ is the Darcy velocity vector $[L/T]^{R_2}$, $c^* = c$ if q is a sink, otherwise it is an imposed source concentration $[M/L^3]$, and f_c is a correction term $[M/TL^3]$ used in the treatment of the surface boundary condition for the transport equation during evaporation.^{R_2} The velocity vector is obtained from the flow equation as $v = -K_r K_s (\nabla \psi + \eta_z)$ while the dispersion tensor can be expressed as:

$$D_{ij} = nS_w \tilde{D}_{ij} = \alpha_t |v| \delta_{ij} + (\alpha_l - \alpha_t) \frac{v_i v_j}{|v|} + nS_w D_o \tau \delta_{ij} \quad i, j = 1, 2, 3$$
(3)

where $|v| = \sqrt{v_1^2 + v_2^2 + v_3^2}$, α_l is the longitudinal dispersivity [L], α_t is the transverse dispersivity [L], δ_{ij} is the Kronecker delta [-], D_o is the molecular diffusion coefficient $[L^2/T]$, and τ is the tortuosity (we assume $\tau=1$) [-]. The evaluation of integrals arising in finite element discretization of the dispersion fluxes is performed using a rotated reference system spanned by the unit vectors (x_1, x_2, x_3) that are aligned with the principal directions of anisotropy of D, whereby $x_1 = v/|v|$. Within

this reference system, D becomes diagonal, with the three components defined as:

5

10

$$D_{11} = \alpha_l |v| + n S_w D_o \tau \tag{4}$$

$$D_{22} = D_{33} = \alpha_t |v| + n S_w D_o \tau$$
⁽⁵⁾

The soil moisture–pressure head and relative conductivity–pressure head dependencies are described by the van Genuchten 15 (1980) relationship:

$$S_e = \left[1 + \left(\frac{|\psi|}{|\psi_{eat}|}\right)^{n_{VG}}\right]^{-m} \tag{6}$$

$$K_r(\psi) = S_e^{0.5} \left[1 - (1 - S_e^{\frac{1}{m}})^m \right]^2$$
(7)

where $S_e = (S_w - S_{wr})/(1 - S_{wr})$ is the effective saturation [-], $R^1 S_{wr}$ is the residual water saturation [-], $m = (1 - 1/n_{VG})$, $R^1 n_{VG}$ is a fitting parameter ranging between 1.25 and 6 [-], and ψ_{sat} is related to the air entry suction [L].

The transport equation (2) is solved in its conservative form, i.e., without applying the chain rule to the advective and storage terms. Using Euler time stepping, the resulting discretized system is:

$$([A+B]^{k+1} + \frac{1}{\Delta t_k}M^{k+1})\hat{c}^{k+1} = \frac{1}{\Delta t_k}M^k\hat{c}^k - b^{t,k+1}$$
(8)

where k is the time counter, \hat{c} is the vector of the numerical approximation of c at each node of the grid, and the coefficients of the, respectively, dispersion, advection, and mass matrices are:

$$25 \quad a_{ij} = \int D\nabla \phi_i \nabla \phi_j d\Omega \tag{9}$$

$$b_{ij} = \int_{\Omega}^{\Omega} \nabla(v\phi_j)\phi_i d\Omega \tag{10}$$

$$m_{ij} = \int_{\Omega}^{\Omega} nS_w \phi_i \phi_j d\Omega \tag{11}$$

where i, j = 1, ..., N with N the number of nodes, Ω is the discretized domain, and ϕ are the basis functions of the Galerkin finite element scheme. The boundary condition vector for the discretized transport equation is:

$$b_i^t = \int_{\Gamma^t} (-D\nabla c) \cdot \nu \phi_i d\Gamma^t = \int_{\Gamma^t} q_n^t \phi_i d\Gamma^t$$
(12)

where Γ^t is the boundary of the domain Ω , $q_n^t [M/(L^2T)]$ is the Neumann (dispersive) flux, and ν is the outward normal 5 vector to the boundary. Cauchy, or mixed, boundary conditions can be easily implemented as variations of Eq. (12), involving an additional term in the system matrix implementing the advective component of the Cauchy condition.

3.3 Model setup for the LEO tracer experiment

We discretized the 30 m x 11 m x 1 m LEO hillslope into 60×22 grid cells in the lateral direction and 30 layers in the vertical direction (Fig. 2). The resulting surface mesh consists of 1403 nodes and 2640 triangular elements. This horizontal

- 10 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).^{R3} The surface mesh was projected vertically to form a 3D tetrahedral mesh with parallel layers of varying thickness, with the thinnest layers assigned to the surface and bottom layers. This allows the numerical model to accurately capture infiltration/evaporation processes at
- 15 the surface and the formation of base flow at the bottom of the domain. From top to bottom the thickness of the 30 layers is: 0.01 m for the first five layers, 0.025 m from layer 6 to layer 9, 0.05 m for layer 10, 0.06 m from layer 11 to layer 20, 0.05 m for layer 21, 0.025 m from layer 22 to layer 25, and 0.01 m from layer 26 to layer 30.

Measurements showed that the average $\delta^2 H$ of the irrigation^{R3} rain^{R3} source water at LEO was -60%. For the transport model, we used a normalized concentration defined as:

$$c = \frac{\delta^2 H_{ref} - \delta^2 H}{\delta^2 H_{ref}} \tag{13}$$

where $\delta^2 H_{ref}$ =-60% and $\delta^2 H$ is the actual value. Thus the initial conditions, as well as the concentrations of the first and third pulses, were *c*=0, while the second pulse had an imposed concentration of *c*=1. Note that, with this transformation, the dimension of the term f_c of Eq. (2) becomes 1/T.^{R2}

- A careful treatment of boundary conditions was essential to modeling the isotope tracer experiment, in particular at the 25 land surface where three different cases needed to be considered. These cases are schematically summarized in Table 1, in relation to the simulations performed, and further noted here^{R3}: 1) Rain with ²*H*-enriched water (second pulse)^{R3}, handled as a Neumann prescribed flux condition for flow $(q_n^f = -K_s K_r(\psi)(\nabla \psi + \eta_z) \cdot \nu = v \cdot \nu \text{ with } q_n^f = -12 \text{ mm/h}^{R3})$ and a Cauchy prescribed advective flux condition for transport $(q_c^t = (vc - D\nabla c) \cdot \nu = v \cdot \nu c^* \text{ with } e^* = 1^{R3})$; 2) Rain with no ²*H*-enriched water (first and third pulses)^{R3}, handled with the same Neumann condition as case 1 for flow $(q_n^f = -12 \text{ mm/h})^{R3}$ and a zero
- 30 Cauchy prescribed total flux condition for transport ($q_c^t = (vc D\nabla c) \cdot \nu = 0$ with the concentration values at the surface nodes computed by the model); 3) Evaporation (between rain pulses and after the third pulse)^{R3}, handled with the same Neumann

Table 1. Treatment of boundary conditions at the land surface during the rainfall and evaporation periods for the flow and transport models^{R3}

Simulation	Rain with ^{2}H -enriched water		Rain with no ${}^{2}H$ -enriched water		Evaporation		
(see Tables 2 and 3)	(second pulse)		(first and third pulses)		(between rain pulses and		
					after the third pulse)		
	Flow	Transport	Flow	Transport	Flow	Transport	
a-f, g-i	q_n^f =-12 mm/h	$q_c^t = v \cdot \nu c^*, c^* = 1$	q_n^f =-12 mm/h	$q_c^t=0$	q_n^f =5 or 3.9 mm/h	$q_c^t=0$	
j	q_n^f =-12 mm/h	$q_c^t = v \cdot \nu c^*, c^* = 1$	q_n^f =-12 mm/h	$q_c^t=0$	Sink q (Fig. 3)	Source f_c (Fig. 3)	
k	q_n^f =-12 mm/h	$q_c^t = v \cdot \nu c^*, c^* = 1$	q_n^f =-12 mm/h	$q_c^t=0$	Sink q (Fig. 3)	Source f_c , $f_c = -qc$	

condition as case 1 for flow but with $q_n^f = 5.0 \text{ mm/d}$ between the rain pulses and $q_n^f = 3.9 \text{ mm/d}$ after the third pulse, R3 and a zero Neumann prescribed dispersive flux condition for transport $(q_n^t = -D\nabla c \cdot \nu = 0$ with the concentration values at the surface nodes computed by the model). With the zero dispersive flux condition of case 3, all the isotopic mass in solution with the evaporating water leaves the domain by advection.

- 5 In addition to this "base case" treatment of rainfall and evaporation, we also introduced some variations on the surface boundary conditions. For rainfall (cases 1 and 2 above), we tested both uniform and variable spatial distributions. For the latter, a rainfall pattern with slightly higher rates towards the center of the landscape was used, as indicated by measurements taken during testing of the engineered rain system. This pattern was generated in such a way that the mean rainfall rate and the total volume of water injected were preserved. For evaporation, since there were no measurements of soil evaporation isotopic
- 10

15

composition at the LEO landscape, we tested two other hypotheses — that none or only a portion (fractionation) of the isotope tracer evaporated — in addition to the zero dispersive flux condition of case 3.

To prevent isotope tracer from leaving the system through the landscape surface, we treated the evaporation as a sink term in the flow model, distributed exponentially from the surface to a depth of 38 cm, rather than as a Neumann boundary condition. The reasoning here is that evaporation involves not just the surface but also deeper soil layers.^{R1} In generating the sink term, we ensured that the total volume of water evaporated was the same as in the Neumann boundary condition treatment. The sink term function q in Eq. (1) applied to each layer i (i=1,...,13 for a total depth of 38 cm) is:

$$q_i = \frac{F_{ev}}{\sum_{i=1}^{13} (e^{-\lambda z_i} \Delta z_i)} e^{-\lambda z_i}$$
(14)

where q_i is applied to each tetrahedron of layer i, $\lambda [1/L]$ is a parameter set to 1 m⁻¹ in this case, z_i is the depth from surface to the center^{R1} of layer i, Δz_i is the thickness of layer i, and F_{ev} [L/T] is the homogeneous evaporative flux used in the Neumann boundary condition case (with rates $-5.7^{R1}-5.8^{R1}\times10^{-8}$ m/s between rain pulses and $-3.4^{R1}-4.5^{R1}\times10^{-8}$ m/s after

20 the third pulse). The applied sink fluxes are reported in Table 1^{R3} shown schematically in Fig. 3.^{R3} Note that if the element reaches the residual water saturation, parameterized by its corresponding pressure head level, the evaporation process becomes



Figure 3. Sink term q and correction source term f_c over depth z added to the flow and transport equations, respectively. q_1 and f_{c1} are applied between rain pulses 1, 2, and 3, while q_2 and f_{c2} are applied after rain pulse 3.

soil-limited. When this occurs, the actual sink term function is automatically smaller than the imposed value.^{R1} To ensure that all the tracer mass stays in the system, for the transport model we added a correction term f_c in Eq. (2) and set it^{R2} we set the correction term f_c in Eq. (2)^{R2} equal to -qc. In this way we inject back into the system the same amount of tracer mass that has exited with the sink term q.

- Most land surface hydrological models still neglect fractionation, even though it can significantly influence the mass exchange at the land surface and the concentration profiles in the soil. Barnes and Allison (1988) examined isotope transport phenomena under both saturated and unsaturated conditions. In the latter case they experimentally observed that at steady state the maximum concentration of the heavier isotope species (e.g., ${}^{2}H$) occurs a short distance below the surface and decreases rapidly beyond that depth. The resulting profile can be explained as the result of vapor diffusion and isotopic exchange domi-
- 10 nating the zone above the drying front and the balance between capillary and diffusive liquid water transport below the drying front (Craig and Gordon, 1965; Clark and Fritz, 1997; Horita et al., 2008). Alternative conceptualizations of the fractionation process have also been recently developed (e.g., Braud et al., 2009; Haverd and Cuntz, 2010). In this work the fractionation process was incorporated into the CATHY model using the sink term approach described above, setting 38 cm as the soil depth at which the maximum ${}^{2}H$ tracer concentration occurs (thereby assuming that the soil above is dominated by water

Table 2.	Configurations	for the 6	simulations of	f the integrated	flow analysis.

Simulation	Saturated hydraulic conductivity (m/s)			Initial conditions	Rainfall
	Horizontal, k_h	Vertical, k_v	Seepage face, k_{sf}		
а	1.4×10^{-4}	1.4×10^{-4}	1.4×10^{-4}	Uniform	Spatially uniform
b	6×10^{-4}	1.4×10^{-4}	1.4×10^{-4}	Uniform	Spatially uniform
c	6×10^{-4}	1.4×10^{-4}	2.2×10^{-5}	Uniform	Spatially uniform
d	6×10^{-4}	1.4×10^{-4}	2.2×10^{-5}	Steady state	Spatially uniform
e	6×10^{-4}	1.4×10^{-4}	2.2×10^{-5}	Interpolated soil moisture measurements	Spatially uniform
f ^{R1}	6×10^{-4}	1.4×10^{-4}	2.2×10^{-5}	Interpolated soil moisture measurements	Spatially variable

vapor diffusion due to evaporation).^{R1} The correction source^{R2} injection^{R2} term f_c introduced into the transport equation is now modified such that there is no tracer mass re-injection in the first layer, and the amount re-injected progressively increases from qc/12 to qc between layers 2 and 13 (Table 2)^{R3} (Fig. 3)^{R3}. The reasoning here is that the rate at which tracer evaporates increases with evaporation and water vapor diffusion close to the surface.^{R1}

- Besides the surface boundary, we set up a seepage face condition at the 23 x 30 nodes that constitute the downslope lateral boundary. For the transport equation the seepage face nodes have a zero Neumann (dispersive) assigned flux so that ${}^{2}H$ is allowed to exit the domain through advection with the outflowing water. All other LEO boundaries (the three other lateral boundaries and the base of the hillslope) were set to a zero Neumann condition for both the flow and transport equations (with a zero water flux this implies that the advective flux for the transport equation is also zero).
- 10 The time stepping for the flow model is adaptive (based on convergence of the iterative scheme used to linearize Richards' equation (1)) and we set the time step range between 10^{-4} s and 90 s. The results in terms of velocity and saturation values were saved every 90 s or 900 s, respectively, during and between the rain events. These were linearly interpolated in time and read as input by the transport model, which was run with a fixed time step of 90 s for the entire simulation.

3.4 Simulations performed

15 The model simulations were used to interpret the integrated and point-scale flow and transport responses of the LEO hillslope. The guiding idea was to assess the need to increase the complexity of the model in progressing from first trying to reproduce the integrated flow response, then the integrated transport response, and finally the point-scale flow and transport responses. With the requirement that each new parameterization still had to satisfy the observation dataset from the previous level, the space of admissible solutions was progressively reduced. Initially the soil was assumed to be homogeneous and isotropic. The values of the van Genuchten parameters (n_{VG} =2.26, ψ_{sat} =-0.6 m, and θ_r =0.002), the porosity (n=0.39), the saturated hydraulic conductivity (k_s =1.4×10⁻⁴ m/s), and the specific storage (S_s =5×10⁻⁴ m⁻¹) were obtained from laboratory analyses and simulations of prior LEO experiments (Niu et al., 2014; Pasetto et al., 2015). From this base set of parameter values for the first simulations, anisotropy, heterogeneity,^{R2} and other variations were progressively introduced in the model.

- In the first step of this procedure (integrated flow response), we examined the influence of heterogeneity^{R2} spatial variability^{R2} and anisotropy in saturated hydraulic conductivity (different k_s at the seepage face and over the rest of the hillslope, on the basis of a clogging hypothesis from accumulation of fine particles (Niu et al., 2014); higher k_h than k_v , on the basis of a hypothesis of slight vertical compaction leading to enhanced flow in the horizontal direction), rainfall (spatially uniform; spatially variable), and initial conditions (uniform; generated from a steady state simulation under drainage and evaporation; matching
- 10 the soil moisture distribution at each sensor location). Six simulations were run in the first step. The configurations for each run are summarized in Table 2. For the initial conditions, in all three configurations (uniform for runs a through c, steady state for run d, and matching sensors for runs e and f), the same total initial water storage (26 m³ as reported earlier) was used. For the atmospheric forcing, the spatially uniform rainfall rate (runs a through e) was the mean measured rate reported earlier (12 mm/h), while the spatially variable case (run f) was handled as described earlier. The evaporation rate, on the other hand,
- 15 was kept spatially uniform for all 6 simulations and equal to the mean rate of 5.0 mm/d between the three pulses and 3.9 mm/d after the third pulse.

In the second step (integrated transport response), the effects of the dispersivity coefficients α_l and α_t and of isotope evaporation mechanisms on the amount of tracer at the seepage face outlet were explored. In the third step (flow point-scale data), the analysis focused on the soil moisture profiles obtained by averaging the observations and model results at specific depths (5, 20, 50, and 85 cm), and spatially variable (by layer) soil hydraulic properties (n_{VG}) were introduced. Finally, for the point-scale

transport we compared the results obtained from some of the different parameterizations used in the previous steps.

20

25

The simulations performed are summarized in Table 3. Model performance was assessed against available observations using the coefficient of efficiency (*CE*) on seepage face flow Q_{sf} for the integrated flow response and the root mean squared error (*RMSE*) on concentration *c* at the seepage face for the integrated transport response and on averaged θ profiles for the flow point-scale response. The *CE* and *RMSE* metrics, also reported in Table 3, are calculated as in Dawson et al. (2007):

$$CE = 1 - \frac{\sum_{i=1}^{n_o} (Q_i - \hat{Q}_i)^2}{\sum_{i=1}^{n_o} (Q_i - \bar{Q})^2}$$

$$\sqrt{\sum_{i=1}^{n_o} (Q_i - \hat{Q}_i)^2}$$
(15)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (Q_i - Q_i)^2}{n}}$$
(16)

^{R1} where n_o is the total number of observed data available at the different times, Q_i and \hat{Q}_i are the observed and modeled values, respectively, and \bar{Q} is the observed average value.^{R1}

Integrated flow analysis				Sim	Simulations			CE for Q_{sf}		
			а	base case	(Niu et	t al., 20	14)		-0.62	
			b	anisotropy			0.64			
Effect or	n seepage	Effect on total	с	c low k_s at seepage face ^{R2}					0.79	
face flo	ow Q_{sf}	water storage V_s	d	initia	al cond	itions			0.28	
			e initial c		al cond	l conditions			0.82	
			f	rainfall distribution			0.85			
In	Integrated transport analysis			α_l	evaporation				RMSE	
			g	0.1		all tra	cer ^{R2}		0.12	
Effect on	concentration	c at the seepage face	h	0.01	all tracer ^{R2}			0.037		
(flow	configuration	from simulation f)	i	0.001	all tracer ^{R2}			0.026		
			j	0.001	fractionation			0.03		
				0.001	no tracer ^{R2}				0.045	
									$RMSE$ for averaged θ	
	Point-scale								(at 5, 20, 50, 85 cm depth)	
Effect on	Effect on	Effect on c point-scale	f	depth (cm)	5	20	50	85	10 36 1 17 1 73 3 78	
averaged θ	point-scale	profiles (transport		n_{VG} (homogeneous)		2.26			10.50, 1.17, 1.75, 5.70	
profiles	θ profiles	configuration from	1	depth (cm)	5	20	50	85	5 61 1 43 0 95 1 72	
		simulation i)	1	$\frac{n_{VG}}{(\text{layered}^{R2})}$	1.8	.8 2.26 2.0 1.9		1.9	5.01, 1. 1 5, 0.95, 1.72	

Table 3. Simulation descriptions, parameter configurations, and performance metrics (coefficient of efficiency CE and root mean squared error RMSE) for the integrated flow, integrated transport, and point-scale analysis steps.



Figure 4. Results for the 6 simulations of the integrated flow response analysis (see Table 2). For each case the seepage face flow Q_{sf} (top^{R1} left^{R1}) and total water storage V_s (bottom^{R1} right^{R1}) are reported.

4 Results

4.1 Integrated flow response

In the first set of simulations we attempt to reproduce two integrated flow responses of the LEO hillslope, the measured seepage face flow and the measured total water storage. The results of the 6 simulations are presented in Fig. 4. The water

- 5 balance partitioning between seepage face flow and internal storage was found to be strongly affected by the introduction of anisotropy and variability^{R2} heterogeneity^{R2} in the hydraulic conductivity. We also found that the distribution of initial condition determines the timing of the first simulated seepage face peak and its shape. The spatial distribution of rain, on the other hand, was not found to have a significant impact on the model response. These general findings are described in more detail below.
- 10 In the first simulation (Fig. 4a), under the assumption of homogeneity, isotropy, uniform initial conditions, and spatially uniform rainfall and evaporation, the discrepancy between the simulated and observed response was large (a negative CE



Figure 5. Results for the integrated transport response analysis for different values of dispersivity (simulations g, h, and i of Table 3)^{R1}. The vertical dashed lines indicate the timing of the three pulses of rain (red when the water is ${}^{2}H$ -enriched and blue when it is not).

is reported in Table 3), with the first and second peaks of the discharge hydrograph, respectively, underestimated and overestimated by the model. In the second simulation, with the introduction of anisotropy (increasing the horizontal hydraulic conductivity k_h to 6×10^{-4} m/s), the overall model results for the seepage face flow improved notably compared to simulation a (*CE* passed from -0.62 to 0.64) and the match for the total water storage was improved significantly (Fig. 4b). Next, the

- 5 introduction of heterogeneity between the seepage face and the rest of the hillslope^{R2} low k_s at the seepage face^{R2} lowered the hydrograph peaks and smoothed out its overall shape (Fig. 4c), moving the simulated hydrograph closer to the measured one (and increasing *CE* to 0.79). The effect of using distributed instead of uniform initial conditions is seen in comparing Figs. 4c, 4d, and 4e. Under uniform starting conditions the response was delayed in time, compared to the steady state case (generated under a drainage and evaporation run from initially wet conditions), where the response to the first rain pulse was
- 10 faster. This faster response resulted in increased drainage due to longer recession periods, adversely affecting the match for the second pulse but improving the result for the third pulse. The simulation for Fig. 4e, with initial conditions closest to the initial state of the hillslope, resulted in a further increase in CE to 0.82. For this run, the good match for the first hydrograph peak from simulation c of Table 2 was recovered, whilst retaining the good match for the second peak from simulation d. The simulated total water storage dynamics was already very well captured by simulation c and was not greatly affected by the
- 15 initial conditions. The initial conditions from simulation e were used for all subsequent simulations discussed in this study. In the final simulation for the integrated flow response analysis, incorporating the spatial distribution of rainfall had a nominal impact on the results (Fig. 4f), with a slight increase in CE to 0.85. Thus the actual distribution of atmospheric forcing, so long as it is not highly variable (which was part of the experimental design for the LEO tracer experiment), is less important than capturing the correct mean rate and total volume of these hydrologic drivers.

20 4.2 Integrated transport response

The velocity field and saturation obtained from the sixth flow simulation (simulation f)^{R1} of the preceding section were used as input to the transport model. Fig. 5 and Table 3 show, respectively, the results for the average tracer concentration at the



Figure 6. Simulated mass balance results for $\alpha_l = 0.001$ m (simulation i of Table 3)^{R1}. From top to bottom: ²H mass that enters the system, M_{in} (normalized with respect to the total mass added to the system during the simulation); that exits through the seepage face, M_{sf} ; that exits through evaporation, M_{ev} ; and that remains in storage, M_{st} . The bottom graph shows the cumulative mass balance error $E_r = (M_{in} - M_{sf} - M_{ev} - M_{st})$. The vertical dashed lines indicate the timing of the three pulses of rain (red when the water is ²H-enriched and blue when it is not).

seepage face and the *RMSE* for different longitudinal dispersivity α_l values, namely 0.1 m, 0.01 m, and 0.001 m. The transverse dispersivity α_t was set one order of magnitude smaller than α_l . The three graphs and the *RMSE* values show that the discrepancy between the measured and simulated outflow concentration decreases with α_l . The results show that the effect of the high dispersivity makes the tracer percolate down quickly to then flow out of the domain from the seepage face boundary. In fact, at^{R2} At^{R2} the highest value, significant levels of^{R1 2}*H*-labeled water appeared in the outflow discharge after the second pulse, whereas this did not occur^{R1} in the measured data and in the model results for the <u>model dispersivities^{R1}</u> smaller dispersivity values the levels were much lower.^{R1} In all three cases the model reproduced the increase in tracer concentration after the last pulse, but whereas for α_l =0.1 m the values were four times higher than the observed ones, for α_l =0.01 m and α_l =0.001 m they decreased significantly. The simulation using the lowest value of dispersivity was able to reproduce reasonably well the integrated measure of tracer response for the LEO hillslope.

5

10

To assess model accuracy, we report in Fig. 6 the mass balance results for the α_l =0.001 m case, in terms of a balance between the cumulative mass of deuterium that entered the hillslope (with the second rainfall pulse), that exited the system (through seepage face outflow and evaporation), and that remained in storage. The results show that for α_l =0.001 m and α_t =0.0001 m,



Figure 7. Measured and modeled average tracer concentration at the seepage face for the cases in which no solute^{R2} tracer^{R2} and all solute^{R2} tracer^{R2} and all solute^{R2} tracer^{R2} leaves the system with evaporation (simulations i and k of Table 3)^{R1}. Both simulations are run for α_l =0.001 m and α_t =0.0001 m. The vertical dashed lines indicate the timing of the three pulses of rain (red when the water is ²*H*-enriched and blue when it is not).

at the end of the simulation (after 14 d), 52% of the mass of ${}^{2}H$ injected has been lost through evaporation, about 4% has seeped out, and the rest remained in storage, minus a cumulative mass balance error of about 2% with respect to the total mass injected. The sudden mass balance error jump which occurs at the beginning of the third pulse of rain is most probably due to discontinuities in the time derivative of concentration and water saturation close to the surface (since the soil is very dry at this

- 5 level and after the long evaporation period) as a consequence of the discontinuity in the atmospheric boundary condition.^{R1} The high evaporative component computed by the model is a direct outcome of the zero dispersive flux surface boundary condition for the transport equation, through which any tracer in solution with evaporating water is advected away with the water. We examine next the impact of the sink term treatment of solute^{R2} tracer^{R2} exchange across the land surface boundary, preventing any isotope tracer from evaporating.
- 10 The results of the sink term simulation in terms of average seepage face tracer concentration and mass balance are reported, respectively, in Fig. 7 and 8. As expected, the seepage face concentration has now increased, but only slightly, compared to the previous simulation. In mass terms, the seepage component has increased from 4% to 8% by the end of the simulation. With no tracer mass now exiting via the landscape surface, it is found instead that much more of the mass has remained in storage (about 90% compared to about 40% when the tracer was allowed to evaporate with the water). This result strongly
- 15 suggests that the tracer does not percolate far (deep) into the hillslope, perhaps as a result of the very dry conditions initially and during the whole experiment. A negative consequence of not allowing any tracer mass to evaporate, combined with low percolation, is an intense accumulation of the mass near the landscape surface, with tracer concentrations as high as 15. Further investigation is needed to understand whether this phenomenon is physically realistic or a numerical artifact.^{R1} A compromise between allowing zero or all isotope tracer to leave the system via evaporation is to introduce isotopic fractionation processes
 20 into the model.

The results of the isotope fractionation simulation are reported in Figs. 9 and 10, respectively, for the average tracer concentration at the seepage face and the model mass balance results. The curve for the average concentration in Fig. 9 justly lies



Figure 8. Simulated mass balance results for α_l =0.001 m when the sink term is used to perform evaporation and the correction term f_c added to the transport equation is used to force all the isotopic mass to stay in the system (simulation k of Table 3)^{R1}. From top to bottom: ²H mass that enters the system, M_{in} (normalized with respect to the total mass added to the system during the simulation); that exits through the seepage face, M_{sf} ; that exits through evaporation, M_{ev} ; and that remains in storage, M_{st} . The bottom graph shows the cumulative mass balance error $E_r = (M_{in} - M_{sf} - M_{ev} - M_{st})$. The vertical dashed lines indicate the timing of the three pulses of rain (red when the water is ²H-enriched and blue when it is not).

between the curves obtained by making all and no isotope evaporate with water. The mass balance shows that at the end of the simulation 6.5% of the total mass injected has gone out through the seepage face, this result also falling between the previous simulations where zero or all isotope tracer in solution with the evaporating water^{R1} was lost via evaporation. As expected, the evaporative mass loss is now significant (38%), but not as high as obtained when evaporation was treated as a land surface Neumann boundary condition (52%). The final mass balance error (0.75%) is lower than for the two previous simulations,

5 Neumann boundary condition (52%). The final mass balance error (0.75%) is lower than for the two previous simulations, and the accumulation of isotope mass just below the land surface that occurred in the preceding case was not observed in this simulation.

4.3 Distributed flow response

For the distributed flow response analysis we first examined the behavior in time of the averaged soil water content value at the

10 4 depths of the sensor network (5, 20, 50, and 85 cm). That is, we compared the average of all sensor measurements at a given depth to the average of all simulated nodal θ values at that depth. The graphs for the results of simulation f from Table 2 (the



Figure 9. Measured and modeled average tracer concentration at the seepage face for the cases in which no solute, all solute, and some solute^{R2} tracer, all tracer, and some tracer^{R2} (fractionation) leaves the system with evaporation (simulations i, j, and k of Table 3)^{R1}. The three simulations are run for α_l =0.001 m and α_t =0.0001 m. The vertical dashed lines indicate the timing of the three pulses of rain (red when the water is ²*H*-enriched and blue when it is not).

configuration that best retrieved the integrated flow response) are shown in Fig. 11^{R1} , while the *RMSE* values are reported in Table 3. The results show that at 50 cm there is a small underestimation by the model and that the model does not perform well at 5 cm and 85 cm compared to the profile at 20 cm. At 85 cm depth the observed and calculated deviation from the mean are also completely different (for the model it is almost 0).^{R1}

- To address this problem we increased the retention capacity of the soil by reducing, selectively, the n_{VG} parameter of the van Genuchten hydraulic functions as reported in Table 3. We subdivided the soil profile into 4 strata encompassing the 4 sensor depths, and we decreased n_{VG} for the strata closest to the surface (from 0 to 10 cm, n_{VG} =1.8), from 32 to 68 cm (n_{VG} =2.0), and from 68 cm to bottom (n_{VG} =1.9). For the second stratum (from 10 cm to 32 cm) the retention parameter was left unaltered from all previous simulations (n_{VG} =2.26) since the model already captured the observed response for the sensor at 20 cm
- 10 depth quite well. The highest retention capacity (lowest n_{VG} value) was set in the first stratum since the observation data show that the water content close to the landscape surface remains quite high, both during infiltration and drainage. The n_{VG} values for the 4 strata reported here are the combination, from many trials, that best retrieved the observed averaged θ profiles. The results of this simulation are also^{R1} shown in Fig. 12^{R1} and reported in Table 3. Compared to the results of the homogeneous n_{VG} case, the model response improves significantly for the average profile at 5, 50, and 85 cm, even if the deviations at 85 cm
- 15 are still very different.^{R1}

To take the distributed flow response analysis further, in Fig. 13 we show the water content time series at the four specific points shown in Fig. 2, at 5, 20, 50, and 85 cm depth from the surface. Sensor data at each of the 4 points and for each of the 4 soil depths is compared against both the homogeneous n_{VG} case (simulation f from Table 2) and the heterogeneous^{R2} layered^{R2} n_{VG} case (different value for each of the four strata). Once again the more detailed parameterization (simulation 1

from Table 3^{R1}, variable n_{VG}) gives better results, although for some of the soil depths (in particular at 50 cm and 85 cm) and for some of the points (in particular point c) the discrepancies between simulated and measured θ time series are quite marked.



Figure 10. Simulated mass balance results for α_l =0.001 m when the correction^{R2} source term f_c is added to the transport equation to perform isotopic fractionation (simulation k of Table 3)^{R1}. From top to bottom: ²H mass that enters the system, M_{in} (normalized with respect to the total mass added to the system during the simulation); that exits through the seepage face, M_{sf} ; that exits through evaporation, M_{ev} ; and that remains in storage, M_{st} . The bottom graph shows the cumulative mass balance error $E_r = (M_{in} - M_{sf} - M_{ev} - M_{st})$. The vertical dashed lines indicate the timing of the three pulses of rain (red when the water is ²H-enriched and blue when it is not).

It should be remarked that we did not run, as we did for the simulation summarized in Fig. 12^{R1} , repeated trials to find a best fit, so it may perhaps be possible to optimize the fits against both the averaged θ data (Fig. 10)^{R1} and the point data (Fig. 13) by manipulating the soil retention capacity for the 4 strata. However, it seems more likely that in going from a distributed but nonetheless averaged response variable to a distributed, point-scale response variable, additional model parameter complexity is needed to obtain an adequate response for all individual response variables.

4.4 Distributed transport response

For the distributed transport response analysis we compared, as we did in Fig. 13 for the internal state flow response, the model results at individual points (a, b, c, d from Fig. 2) and individual soil depths (5, 20, 50, and 85 cm) for simulations using uniform (corresponding to configuration f from Table 2) and spatially variable (4-strata^{R1} simulation 1 from Table 3^{R1}) soil retention capacity. The results are shown in Fig. 14, and it can be seen that the model does not perform well at several locations within

10

5



Figure 11. Averaged θ profiles at 5, 20, 50, and 85 cm depth from the surface. In each graph the deviation from the mean (one standard deviation above and below) is shown as dashed lines. The results are obtained for simulation f reported in Table 3.^{R1}

the hillslope (consistently at 20 cm depth, and at 5 cm depth for point b). Encouragingly, however, there is consistency with the previous distributed flow results, in that the variable n_{VG} run performs noticeably better than the spatially uniform case. For instance, with variable n_{VG} the results are improved at the bottom of the hillslope, at 50 cm (for points b and c the modeled response gets closer to the measurements particularly after the third flush), and slightly at 5 cm (for point a).

- 5 For the distributed transport analysis we did not examine averaged concentration profiles at each of the 4 sensor depths (as we did for soil water content in Fig. 12) due to insufficient data. The sampling time and laboratory analysis costs for exhaustive measurement of isotopic compositions were prohibitive, thus there are much less data available for the distributed transport analysis compared to the flow case. The data gaps are also evident in Fig. 14: there are no measurements for 3 of the graphs, and scarce data at 50 cm depth for points a and d. It is also important to note that no additional parameterization was attempted
 - 10 for the distributed transport analysis. The main explicit parameters in the transport equation are the dispersivity coefficients, and these were experimented with in the integrated transport analysis. The transport equation is of course strongly dependent on flow velocities, and thus implicitly on the conductivity and other soil hydraulic parameters that were assessed in the flow model analyses. These and other parameterization issues will be further discussed in the next section.

To complete the sequence of analyses from integrated flow and then transport to distributed flow and transport, we used the simulation results from the additional parameterization introduced for the distributed analyses (spatially variable soil retention



Figure 12. Averaged θ profiles at 5, 20, 50, and 85 cm depth from the surface. In each graph the deviation from the mean (one standard deviation above and below) is shown as dashed lines. The results are obtained for simulation 1 reported in Table 3.^{R1}

capacity) to assess model performance against the integrated flow and transport responses. The results (Fig. 15) show that while the match against tracer concentration at the seepage face has somewhat improved (compare with Fig. 9), the match against both of the integrated flow responses (seepage outflow and total water storage) has significantly deteriorated (compare with simulation f of Fig. 4). This is not a surprising result, given that no attempt was made to parameterize the model in tandem against both integrated and point-scale observations (nor against joint flow and transport data); the implications will be discussed below.

5 Discussion

5

Mass transport in unsaturated soils is extremely important in the context of biosphere, critical zone, and Earth systems research because of exchanges of water and solutes that occur across the land surface interface. The study of hillslope transit time distri-

10 butions (e.g., Fiori and Russo, 2008; Botter et al., 2010; Heidbüchel et al., 2013; Tetzlaff et al., 2014) is a good example of the need for a better understanding of such water and solute exchanges and the consequent subsurface flowpaths. The simulation of unsaturated zone mass transport phenomena is however^{R2} However, its simulation is also^{R2} known to be a particularly complex problem, compounded by any presence of heterogeneity. Wilson and Gelhar (1981), for instance, showed that spatial variations



Figure 13. Distributed (internal state) hydrological response for the θ profiles at 5, 20, 50, and 85 cm depth from the surface for four locations on the LEO-1 hillslope: point a (top left), point b (top right), point c (bottom left), and point d (bottom right) of Fig. 2. The results are obtained for simulations f and l reported in Table 3.^{R1}



Figure 14. Distributed (internal state) hydrological response for the tracer concentration breakthrough curves at 5, 20, 50, and 85 cm depth from the surface for four locations on the LEO-1 hillslope: point a (top left), point b (top right), point c (bottom left), and point d (bottom right) of Fig. 2. There were no tracer concentration measurements at 5 cm depth for point c and at 5 and 20 cm depth for point d. The transport model is run for α_l =0.001 m and α_t =0.0001 m. The vertical dashed lines indicate the timing of the three pulses of rain (red when the water is ²*H*-enriched and blue when it is not).



Figure 15. Performance of the model against integrated flow and transport responses (seepage face flow Q_{sf} , total water storage V_s , and average tracer concentration c at the seepage face) using the additional parameterization from the distributed analyses (spatially variable soil retention capacity, simulation 1 of Table 3^{R1}). The vertical dashed lines indicate the timing of the three pulses of rain (red when the water is ²*H*-enriched and blue when it is not).

in moisture content affect solute plume spreading even without dispersive mixing, and that the rates of solute displacement are typically much smaller than the rates of moisture displacement. Birkholzer and Tsang (1997) demonstrated significant channeling effects (preferential solute pathways, with accompanying higher dispersion) at the extremes of very low saturation and full saturation. Studies that have combined comprehensive experimental observation with detailed subsurface simulation have

- 5 also documented some of the challenges faced in modeling solute transport under unsaturated and heterogeneous conditions (e.g., Haggerty et al., 2004; Zheng et al., 2011). In this context, for the tritium and bromide tracer experiments at the Las Cruces trench site, standard models gave good prediction of wetting front movement during infiltration but poor prediction of point soil water content and tracer transport (Hills et al., 1991; Wierenga et al., 1991). For the macrodispersion (MADE) experiment, Russo and Fiori (2009) found that heterogeneity further enhances solute spreading and breakthrough curve arrival times when
- 10 the unsaturated zone is relatively dry or deep. In the present study, the additional heterogeneity^{R2} complexity^{R2} introduced for the point-scale responses (namely spatially variable soil retention capacity) did not match as favorably the integrated (flow) observation dataset (Fig. 15). While this could perhaps be remedied using more rigorous or quantitative parameter estimation, the particular difficulties in capturing the point-scale concentration profiles, especially near the landscape surface, can be taken

as further evidence for flaws or gaps in theoretical understanding and model formulation (process representation) for simulating solute transport phenomena in very dry, heterogeneous soils.

Various hypotheses have been invoked to explain possible factors that affect the migration and distribution of solutes under unsaturated, heterogeneous conditions, including: turbulent mixing due to high rainfall (Havis et al., 1992); solute transfer

- 5 between mobile and immobile water (De Smedt and Wierenga, 1984); mobile-immobile exchange and hysteresis (Butters et al., 1989; Russo et al., 1989a, b, 2014); lateral mixing due to velocity fluctuations (Russo et al., 1998); isotope effects (Barnes and Allison, 1988; LaBolle et al., 2008; Zhang et al., 2009); variable, state-dependent anisotropy (McCord et al., 1991); non-Gaussian early-time mean tracer plume behavior (Naff, 1990); non-Fickian solute migration at low water contents (Padilla et al., 1999) and for macroscopically homogeneous sand (Bromly and Hinz, 2004); and saturation-dependent dispersivity (Raoof
- and Hassanizadeh, 2013). In addition, Konikow et al. (1997) and Parker and van Genuchten (1984) discuss the importance 10 of boundary condition treatment (e.g., water-solute injection, solute exchange between soil and atmosphere). Given the many open questions, for this first analysis of the LEO isotope tracer experiment the modeling was kept to the standard formulation of the Richards and advection-dispersion equations. Limitations encountered in the multiresponse performance assessment, from the standpoint of experimental procedure, model formulation, or numerical implementation, will inform follow-up studies at
- LEO. The simulation results from this tracer experiment, for instance, point to highly complex effects on plume migration of 15 spatially variable water content in the dry soils that characterized the experiment, especially at early times.

The broad results of our study should be quite universal, particularly to deterministic numerical models based on the 3D Richards and advection-dispersion equations. However, any model has its specific features and differs, for example, in the way equations are coded (e.g., choice of numerical solvers) or interface conditions are implemented (e.g., free-surface vs boundary condition switching). For insights on the impact of specific model differences in the performance of CATHY-like models, 20 see the intercomparison studies of Sulis et al. (2010) and Maxwell et al. (2014). These intercomparison studies have thus far focused only on flow processes, and there is an urgent need to extend the analyses to solute transport phenomena, in order

to properly guide our assessment of the physical and numerical correctness of competing models as these models continue to increase in complexity. For instance for this study there are aspects of the CATHY model related to how we implemented 25 evaporation and fractionation that might be expected to negatively impact the generality of our findings, although in terms of isotope tracer mass exiting the seepage face the impact was quite small. But the implementation here was somewhat ad hoc, and more study is needed on the importance and proper representation of fractionation in solute transport models, especially under strongly unsaturated conditions.^{R3}

6 Conclusions

30 In this study we have used multivariate observations (soil moisture, water and tracer outflow, breakthrough curves, and total water storage) culled from the first isotope tracer experiment at the LEO-1 hillslope of the Biosphere 2 facility to explore some of the challenges in modeling unsaturated flow and transport phenomena. Integrated (first flow and then transport) and distributed (again flow followed by transport) measurements were progressively introduced as response variables with which to assess the results from simulations with CATHY, a 3D numerical model for variably saturated flow and advective-dispersive solute migration. Compared to the first flow experiment at LEO that was successfully modeled with CATHY (Niu et al., 2014), the modeling task for the tracer experiment was significantly more complicated due to: joint simulation of both flow and transport processes; considerably drier initial conditions and reduced forcing; performance assessment against both system-

5 wide and point-scale measurements; and multiple periods of water/tracer injection compared to a single rainfall episode. In some sense the previous flow study looked at the first order response of the LEO hillslope, whereas the modeling exercise for the tracer experiment represents a first look at higher order responses of the Biosphere 2 landscapes.

There are several findings from this first set of simulations of a LEO isotope tracer experiment. At the start of the exercise, where integrated flow measurements were used, we were able to obtain good matches for two response variables (total water

10 storage and seepage face outflow) using parameter values and initial and boundary conditions that correspond quite closely to the actual experimental conditions and previous (flow experiment) model implementation (Niu et al., 2014). The same soil parameterization was successfully used to reproduce the integrated transport response. When passing to point-scale flow and finally point-scale transport, a refinement of the model setup (augmenting the degree of heterogeneity, mainly) was needed. Moreover, providing more information to the model (for example, the distribution of initial water storage rather than just the initial total volume) generally helped to improve the simulation results.

The effect of saturated hydraulic conductivity (heterogeneity and anisotropy) on the response of subsurface hydrologic models is well known, and was also borne out in this study. Also not surprisingly, the dispersivity parameter had a big impact on the transport simulations, with a clear trend to a better match against measured seepage face concentration as dispersivity was decreased. The spatial distribution of rainfall was not found to have a big impact on simulation results, and there was not much difference, in terms of isotope tracer mass exiting the seepage face, between the zero, partial, and no fractionation cases, suggesting that the injected tracer did not percolate very far into the hillslope, likely due to the very dry initial conditions.

20

We conclude with a few specific recommendations for alleviating some of the modeling and experimental limitations encountered during this study. On the modeling side, a more sophisticated treatment of solute transport phenomena beyond the standard advection-dispersion equation could start with incorporation of a mobile-immobile conceptualization and/or saturation-

- 25 dependent dispersivity. Other upgrades to the CATHY model (e.g., Scudeler et al., 2016) will mitigate the grid Peclet constraint and provide more reliable flow velocity calculations, essential to maintaining low mass balance errors and high accuracy in solute transport. On the experimental side, higher solute^{R2} tracer^{R2} concentrations (including labeled tracers), wetter initial conditions, and more intensive direct or indirect measures of total solute^{R2} tracer^{R2} mass could help address the high sensitivity of solute transport to small scale heterogeneity under dry soil conditions. Any experiments that provide spatially detailed
- 30 observations of both flow and transport response variables that are then jointly used in estimating, for instance, conductivity and other soil hydraulic parameters (traditionally identified based solely on flow responses), would be critical to advancing the present state of hydrologic model verification, given the high impact that Darcy velocities, which are directly dependent on such parameters, have on solute mixing processes. Finally, future LEO isotope tracer experiments that also generate some surface runoff would offer valuable benchmark data for improving integrated surface-subsurface models.

Acknowledgements. This work was supported in part by the National Science Foundation (NSF grants EAR-1344552 and EAR-1417097). The authors wish to thank Nate Abramson, Anouk Gevaert, Ed Hunt, Michael Pohlmann, Michael Sibayan, and Yvonne Smit for technical support, and Claude Mugler and two anonymous reviewers for their helpful comments. The Biosphere 2 facility and the capital required to conceive and construct LEO were provided through a charitable donation from the Philecology Foundation, and its founder, Mr. Edward Bass. We gratefully acknowledge this charitable donation. The data for this paper are available by contacting the corresponding author.

References

Anderman, E. R. and Hill, M. C.: A new multistage groundwater transport inverse method: presentation, evaluation, and implications, Water Resour. Res., 35, 1053–1063, doi:10.1029/1998WR900114, 1999.

Barnes, C. J. and Allison, G. B.: Tracing of water movement in the unsaturated zone using stable isotopes of hydrogen and oxygen, J. Hydrol.,

- 5 100, 143–176, doi:10.1016/0022-1694(88)90184-9, 1988.
- Binley, A., Elgy, J., and Beven, K.: A physically based model of heterogeneous hillslopes: 1. Runoff production, Water Resour. Res., 25, 1219–1226, doi:10.1029/WR025i006p01219, 1989.
 - Birkholzer, J. and Tsang, C.-F.: Solute channeling in unsaturated heterogeneous porous media, Water Resour. Res., 33, 2221–2238, doi:10.1029/97WR01209, 1997.
- 10 Botter, G., Bertuzzo, E., and Rinaldo, A.: Transport in the hydrologic response: Travel time distributions, soil moisture dynamics, and the old water paradox, Water Resour. Res., 46, doi:10.1029/2009WR008371, W03514, 2010.
 - Braud, I., Biron, P., Bariac, T., Richard, P., Canale, L., Gaudet, J. P., and Vauclin, M.: Isotopic composition of bare soil evaporated water vapor. Part I: RUBIC IV experimental setup and results, J. Hydrol., 369, 1–16, doi:10.1016/j.jhydrol.2009.01.034, 2009.
 - Bromly, M. and Hinz, C.: Non-Fickian transport in homogeneous unsaturated repacked sand, Water Resour. Res., 40,

15 doi:10.1029/2003WR002579, W07402, 2004.

Brooks, P. D., Chorover, J., Fan, Y., Godsey, S. E., Maxwell, R. M., McNamara, J. P., and Tague, C.: Hydrological partitioning in the critical zone: Recent advances and opportunities for developing transferable understanding of water cycle dynamics, Water Resour. Res., 51, 6973–6987, doi:10.1002/2015WR017039, 2015.

Brunner, P., Doherty, J., and Simmons, C. T.: Uncertainty assessment and implications for data acquisition in support of integrated hydrologic

20 models, Water Resour. Res., 48, doi:10.1029/2011WR011342, W07513, 2012.

- Butters, G. L., Jury, W. A., and Ernst, F. F.: Field scale transport of bromide in an unsaturated soil: 1. Experimental methodology and results, Water Resour. Res., 25, 1575–1581, doi:10.1029/WR025i007p01575, 1989.
 - Camporese, M., Paniconi, C., Putti, M., and Orlandini, S.: Surface-subsurface flow modeling with path-based runoff routing, boundary condition-based coupling, and assimilation of multisource observation data, Water Resour. Res., 46, doi:10.1029/2008WR007536,
- 25 W02512, 2010.

30

Clark, I. D. and Fritz, P.: Environmental Isotopes in Hydrogeology, CRC Press, 1997.

Craig, H. and Gordon, L. I.: Deuterium and oxygen 18 variations in the ocean and the marine atmosphere. In: Stable Isotopes in Oceanographic Studies and Paleotemperatures, E. Tongiorgi (Ed.), pp. 9–130, Laboratorio di geologia nucleare, Pisa, 1965.

Dawson, C. W., Abrahart, R. J., and See, L. M.: HydroTest: a web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts, Environ. Model. Softw., 22, 1034–1052, doi:10.1016/j.envsoft.2006.06.008, 2007.

- De Smedt, F. and Wierenga, P. J.: Solute Transfer through columns of glass beads, Water Resour. Res., 20, 225–232, doi:10.1029/WR020i002p00225, 1984.
 - Fenicia, F., Savenije, H. H. G., Matgen, P., and Pfister, L.: A comparison of alternative multiobjective calibration strategies for hydrological modeling, Water Resour. Res., 43, doi:10.1029/2006WR005098, W03434, 2007.
- 35 Fiori, A. and Russo, D.: Travel time distribution in a hillslope: Insight from numerical simulations, Water Resour. Res., 44, doi:10.1029/2008WR007135, W12426, 2008.

- Gevaert, A. I., Teuling, A. J., Uijlenhoet, R., DeLong, S. B., Huxman, T. E., Pangle, L., Breshears, D. D., Chorover, J., Pelletier, J. D., Saleska, S. R., Zeng, X., and Troch, P. A.: Hillslope-scale experiment demonstrates the role of convergence during two-step saturation, Hydrol. Earth Syst. Sci., 18, 3681–3692, doi:10.5194/hess-18-3681-2014, 2014.
- Ghanbarian-Alavijeh, B., Skinner, T. E., and Hunt, A. G.: Saturation dependence of dispersion in porous media, Phys. Rev. E, 86, doi:10.1103/PhysRevE.86.066316, 2012.
- Gorelick, S. M. and Zheng, C.: Global change and the groundwater management challenge, Water Resour. Res., 51, 3031–3051, doi:10.1002/2014WR016825, 2015.
 - Gupta, H. V., Sorooshian, S., and Yapo, O. P.: Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, Water Resour. Res., 34, 751–763, doi:10.1029/97WR03495, 1998.
- 10 Haggerty, R., Harvey, C. F., Freiherr von Schwerin, C., and Meigs, L. C.: What controls the apparent timescale of solute mass transfer in aquifers and soils? A comparison of experimental results, Water Resour. Res., 40, doi:10.1029/2002WR001716, W01510, 2004.
 - Haverd, V. and Cuntz, M.: Soil-Litter-Iso: A one-dimensional model for coupled transport of heat, water and stable isotopes in soil with a litter layer and root extraction, J. Hydrol., 388, 438–455, doi:10.1016/j.jhydrol.2010.05.029, 2010.
 - Havis, R. N., Smith, R. E., and Adrian, D. D.: Partitioning solute transport between infiltration and overland flow under rainfall, Water
- 15 Resour. Res., 28, 2569–2580, doi:10.1029/92WR01366, 1992.

5

30

- Heidbüchel, I., Troch, P. A., and Lyon, S. W.: Separating physical and meteorological controls of variable transit times in zero-order catchments, Water Resour. Res., 49, 7644–7657, doi:10.1002/2012WR013149, 2013.
 - Hills, R. G., Wierenga, P. J., Hudson, D. B., and Kirkland, M. R.: The second Las Cruces trench experiment: Experimental results and two-dimensional flow predictions, Water Resour. Res., 27, 2707–2718, doi:10.1029/91WR01538, 1991.
- 20 Hofer, M., Lehmann, P., Stähli, M., Seifert, S., and Krafczyk, M.: Two approaches to modeling the initiation and development of rills in a man-made catchment, Water Resour. Res., 48, doi:10.1029/2011WR010719, W01531, 2012.
 - Hopp, L., Harman, C., Desilets, S. L. E., Graham, C. B., McDonnell, J. J., and Troch, P. A.: Hillslope hydrology under glass: confronting fundamental questions of soil-water-biota coevolution at Biosphere 2, Hydrol. Earth Syst. Sci., 13, 2105-2118, doi:10.5194/hess-13-2105-2009, 2009.
- 25 Horita, J., Rozanski, K., and Cohen, S.: Isotope effects in the evaporation of water: a status report of the Craig-Gordon model, Isotopes Environ. Health Stud., 44, 23–49, doi:10.1080/10256010801887174, 2008.
 - Huxman, T., Troch, P. A., Chorover, J., Breshears, D. D., Saleska, S., Pelletier, J., Zeng, X., and Espeleta, J.: The hills are alive: interdisciplinary Earth science at Biosphere 2, EOS, Trans. Am. Geophys. Union, 90, 120, 2009.

Kampf, S. K. and Burges, S. J.: Parameter estimation for a physics-based distributed hydrologic model using measured outflow fluxes and internal moisture states, Water Resour. Res., 43, doi:10.1029/2006WR005605, W12414, 2007.

Keating, E. H., Doherty, J., Vrugt, J. A., and Kang, Q.: Optimization and uncertainty assessment of strongly nonlinear groundwater models with high parameter dimensionality, Water Resour. Res., 46, doi:10.1029/2009WR008584, W10517, 2010.

Konikow, L. F., Sanford, W. E., and Campbell, P. J.: Constant-concentration boundary condition: lessons from the HYDROCOIN variabledensity groundwater benchmark problem, Water Resour. Res., 33, 2253–2261, doi:10.1029/97WR01926, 1997.

- 35 Kumar, R., Samaniego, L., and Attinger, S.: Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations, Water Resour. Res., 49, 360–379, doi:10.1029/2012WR012195, 2013.
 - LaBolle, E. M., Fogg, G. E., Eweis, J. B., Gravner, J., and Leaist, D. G.: Isotopic fractionation by diffusion in groundwater, Water Resour. Res., 44, doi:10.1029/2006WR005264, W07405, 2008.

- Maxwell, R. M., Putti, M., Meyerhoff, S., Delfs, J.-O., Ferguson, I. M., Ivanov, V., Kim, J., Kolditz, O., Kollet, S. J., Kumar, M., Lopez, S., Niu, J., Paniconi, C., Park, Y.-J., Phanikumar, M. S., Shen, C., Sudicky, E. A., and Sulis, M.: Surface-subsurface model intercomparison: A first set of benchmark results to diagnose integrated hydrology and feedbacks, Water Resour. Res., 50, 1531–1549, doi:10.1002/2013WR013725, 2014.
- 5 McCord, J. T., Stephens, D. B., and Wilson, J. L.: Hysteresis and state-dependent anisotropy in modeling unsaturated hillslope hydrologic processes, Water Resour. Res., 27, 1501–1518, doi:10.1029/91WR00880, 1991.
 - Mishra, S. and Parker, J. C.: Parameter estimation for coupled unsaturated flow and transport, Water Resour. Res., 25, 385–396, doi:10.1029/WR025i003p00385, 1989.

Naff, R. L.: On the nature of the dispersive flux in saturated heterogeneous porous media, Water Resour. Res., 26, 1013–1026, doi:10.1029/WR026i005p01013, 1990.

- Neuweiler, I. and Cirpka, O. A.: Homogenization of Richards equation in permeability fields with different connectivities, Water Resour. Res., 41, doi:10.1029/2004WR003329, W02009, 2005.
- Niu, G.-Y., Pasetto, D., Scudeler, C., Paniconi, C., Putti, M., Troch, P. A., DeLong, S. B., Dontsova, K., Pangle, L., Breshears, D. D., Chorover, J., Huxman, T. E., Pelletier, J., Saleska, S. R., and Zeng, X.: Incipient subsurface heterogeneity and its effect on overland flow
- 15 generation–Insight from a modeling study of the first experiment at the Biosphere 2 Landscape Evolution Observatory, Hydrol. Earth Syst. Sci., 18, 1873–1883, doi:10.5194/hess-18-1873-2014, 2014.
 - Padilla, I. Y., Yeh, T.-C. J., and Conklin, M. H.: The effect of water content on solute transport in unsaturated porous media, Water Resour. Res., 35, 3303–3313, doi:10.1029/1999WR900171, 1999.
 - Pangle, L. A., DeLong, S. B., Abramson, N., Adams, J., Barron-Gafford, G., Breshears, D. D., Brooks, P., Chorover, J., Dietrich, W. E.,
- 20 Dontsova, K., Durcik, M., Espeleta, J., Ferre, T. P. A., Ferriere, R., Henderson, W., Hunt, E., Huxman, T. E., Millar, D., Murphy, B., Niu, G.-Y., Pavao-Zuckerman, M., Pelletier, J. D., Rasmussen, C., Ruiz, J., Saleska, S., Schaap, M., Sibayan, M., Troch, P., Tuller, M., van Haren, J., and Zeng, X.: Landscape Evolution Observatory: A large-scale controllable infrastructure to study coupled Earth-surface processes, Geomorphology, 244, 190–203, doi:10.1016/j.geomorph.2015.01.020, 2015.

Paniconi, C. and Putti, M.: Physically based modeling in catchment hydrology at 50: Survay and outlook, Water Resour. Res., 51, 7090–7129,

25 doi:10.1002/2015WR017780, 2015.

10

- Parker, J. C. and van Genuchten, M. T.: Flux-averaged and volume-averaged concentrations in continuum approaches to solute transport, Water Resour. Res., 20, 866–872, doi:10.1029/WR020i007p00866, 1984.
- Pasetto, D., Niu, G.-Y., Pangle, L., Paniconi, C., Putti, M., and Troch, P. A.: Impact of sensor failure on the observability of flow dynamics at the Biosphere 2 LEO hillslopes, Adv. Water Resour., 86, 327–339, doi:10.1016/j.advwatres.2015.04.014, 2015.
- 30 Putti, M. and Paniconi, C.: Picard and Newton linearization for the coupled model of saltwater intrusion in aquifers, Adv. Water Resour., 18, 159–170, doi:10.1016/0309-1708(95)00006-5, 1995.

Raoof, A. and Hassanizadeh, S. M.: Saturation-dependent solute dispersivity in porous media: Pore-scale processes, Water Resour. Res., 49, 1943–1951, doi:10.1002/wrcr.20152, 2013.

Russo, D. and Fiori, A.: Stochastic analysis of transport in a combined heterogeneous vadose zone-groundwater flow system, Water Resour.

35 Res., 45, doi:10.1029/2008WR007157, W03426, 2009.

Russo, D., Jury, W. A., and Butters, G. L.: Numerical analysis of solute transport during transient irrigation: 1. The effect of hysteresis and profile heterogeneity, Water Resour. Res., 25, 2109–2118, doi:10.1029/WR025i010p02109, 1989a.

- Russo, D., Jury, W. A., and Butters, G. L.: Numerical analysis of solute transport during transient irrigation: 2. The effect of immobile water, Water Resour. Res., 25, 2119–2127, doi:10.1029/WR025i010p02119, 1989b.
- Russo, D., Zaidel, J., and Laufer, A.: Numerical analysis of flow and transport in a three-dimensional partially saturated heterogeneous soil, Water Resour. Res., 34, 1451–1468, doi:10.1029/98WR00435, 1998.
- 5 Russo, D., Laufer, A., Gerstl, Z., Ronen, D., Weisbrod, N., and Zentner, E.: On the mechanism of field-scale solute transport: Insights from numerical simulations and field observations, Water Resour. Res., 50, 7484–7504, doi:10.1002/2014WR015514, 2014.
 - Scudeler, C., Putti, M., and Paniconi, C.: Mass-conservative reconstruction of Galerkin velocity fields for transport simulations, Adv. Water Resour., 94, 470–485, doi:10.1016/j.advwatres.2016.06.011, 2016.
- Sebben, M. L., Werner, A. D., Liggett, J. E., Partington, D., and Simmons, C. T.: On the testing of fully integrated surface-subsurface
 hvdrological models, Hvdrol, Proc., 27, 1276–1285, doi:10.1002/hvp.9630, 2013.
 - Sprenger, M., Volkmann, T. H. M., Blume, T., and Weiler, M.: Estimating flow and transport parameters in the unsaturated zone with pore water stable isotopes, Hydrol. Earth Syst. Sci., 19, 2617–2635, doi:10.5194/hess-19-2617-2015, 2015.
 - Sulis, M., Meyerhoff, S. B., Paniconi, C., Maxwell, R. M., Putti, M., and Kollet, S. J.: A comparison of two physics-based numerical models for simulating surface water–groundwater interactions, Adv. Water Resour., 33, 456–467, doi:10.1016/j.advwatres.2010.01.010, 2010.
- 15 Tetzlaff, D., Birkel, C., Dick, J., Geris, J., and Soulsby, C.: Storage dynamics in hydropedological units control hillslope connectivity, runoff generation, and the evolution of catchment transit time distributions, Water Resour. Res., 50, 969–985, doi:10.1002/2013WR014147, 2014.
 - van Genuchten, M. T.: A closed-form equation for predicting the hydraulic conductivity of unsaturated soils, Soil Sci. Soc. Amer. J., 44, 892–898, doi:10.2136/sssaj1980.03615995004400050002x, 1980.
- 20 Weill, S., Mazzia, A., Putti, M., and Paniconi, C.: Coupling water flow and solute transport into a physically-based surface-subsurface hydrological model, Adv. Water Resour., 34, 128–136, doi:10.1016/j.advwatres.2010.10.001, 2011.
 - Wierenga, P. J., Hills, R. G., and Hudson, D. B.: The Las Cruces trench site: Characterization, experimental results, and one-dimensional flow predictions, Water Resour. Res., 27, 2695–2705, doi:10.1029/91WR01537, 1991.

Wilson, J. L. and Gelhar, L. W.: Analysis of longitudinal dispersion in unsaturated flow: 1. The analytical method, Water Resour. Res., 17,

- 25 122–130, doi:10.1029/WR017i001p00122, 1981.
 - Woolhiser, D. A., Smith, R. E., and Giraldez, J.-V.: Effects of spatial variability of saturated hydraulic conductivity on Hortonian overland flow, Water Resour. Res., 32, 671–678, doi:10.1029/95WR03108, 1996.
 - Zacharias, S., Bogena, H., Samaniego, L., Mauder, M., Fuß, R., Pütz, T., Frenzel, M., Schwank, M., Baessler, C., Butterbach-Bahl, K., Bens,O., Borg, E., Brauer, A., Dietrich, P., Hajnsek, I., Helle, G., Kiese, R., Kunstmann, H., Klotz, S., Munch, J. C., Papen, H., Priesack, E.,
- 30 Schmid, H. P., Steinbrecher, R., Rosenbaum, U., Teutsch, G., and Vereecken, H.: A network of terrestrial environmental observatories in Germany, Vadose Zone J., 10, 955–973, doi:10.2136/vzj2010.0139, 2011.
 - Zhang, X.-P., Yang, Z.-L., Niu, G.-Y., and Wang, X.-Y.: Stable water isotope simulation in different reservoirs of Manaus, Brazil, by Community Land Model incorporating stable isotopic effect, Int. J. Climatol., 29, 619–628, doi:10.1002/joc.1740, 2009.
 - Zheng, C., Bianchi, M., and Gorelick, S. M.: Lessons learned from 25 years of research at the MADE site, Ground Water, 49, 649-662,
- 35 doi:10.1111/j.1745-6584.2010.00753.x, 2011.