General comments to the editor

We greatly appreciate the constructive and insightful comments on our manuscript. In response to your comments and those of the reviewers, we have made many improvements to the original manuscript. First, we will give an overview of the major changes that have been made to the manuscript based on these comments and other changes we feel improved the manuscript. Then we give detailed responses to each of your comments as well as those made by the reviewers in a point-by-point fashion. Finally, we report other relevant changes that were made to further improve the manuscript, largely based on comments given by the co-authors that have been added to the revised version, which are all members of the Landscape Evolution Observatory (LEO) Research and Design Team. The LEO project is in an early phase and it has recently been decided to include these primary investigators as co-authors of all publications, which was not the case initially. The added co-authors have also provided many additional constructive comments that we feel improved the manuscript.

The most important changes in the revised manuscript are summarized as follows:

- Piezometer data have been added by means of two new figures: Figure 6 and Figure 8. Despite the temperature sensitivity that led us to exclude the data from the original manuscript, we believe that the new figures support the observations made and conclusions drawn in this study.
- We have added the results of simple column storage calculations to support the importance of convergent subsurface flow (page 8, lines 13-19).
- The results of our experiment have been related to existing knowledge in conceptual models like the Sacramento model (page 12, lines 5-24).
- Soil hydraulic theory that is relevant to observations made during the two-step saturation process and groundwater ridging has been presented in more detail (page 13, lines 11-12; page 15, lines 2-14).
- A discussion of the overland flow generation processes has been added in a separate subsection, including reasons why Hortonian overland flow can be excluded (page 16, lines 5-14).
- The relation between this experiment and the LEO project as well as future experiments at LEO has been made clearer (page 3, line 26 to page 4, line 5; page 18, lines 8-14).

Responses to Editor comments

1) Explain the relation between the two-step saturation concept and some existing knowledge including the two-storage concept in Sacramento Model and Xinanjiang Model, and soil hydraulic theory.

We agree that the two-step saturation concept can in theory be simulated with a tension water reservoir and a free water reservoir, as is the case in the Sacramento and Xinanjiang models. However, these models do not allow lateral redistribution of water and therefore cannot simulate convergent flow. Without lateral subsurface flow from the upslope to convergent areas, the hillslope would not have saturated to the surface and overland flow would not have occurred during this experiment. In addition, in models with multiple soil layers, such as the Sacramento model, the rate of percolation from the upper to the lower soil layer depends on the storage deficit of the lower soil layer. Keeping the constant rainfall rate in mind, this means that the storage of the upper soil layer will increase while the lower soil reservoir is being filled (lowering the deficit). In our experiment, after the infiltration has passed, the moisture content only changes in response to the upward propagation of the saturation front. In other words, storage in the upper soil layer only changes once the lower soil layer is filled completely. A discussion of conceptual models and the two-step saturation has been added to the revised manuscript (page 12, lines 5-24).

The discussion of observations that are not in line with existing knowledge of soil hydraulic theory has also been extended in the revised manuscript. One of these observations is the constant specific infiltration flux despite decreasing water content. Air entrapment was offered as an explanation by one of the reviewers. In our response to the reviewer, we describe why we do not consider this to be the case: no ponding was observed and the rainfall rate was not completely homogeneous. Other possible explanations we have included in the revised version are changes in soil water retention characteristics or saturated hydraulic conductivity with depth, but the true explanation cannot be determined based on the collected data. These points have been included in the Discussion section of the revised manuscript (page 13, lines 11-23). Also, the discussion of groundwater ridging has been extended based on the comments and references provided by George Waswa (page 15, lines 1-14 and 21-28).

2) Provide more data to support the conclusion drawn in the manuscript

Motivated by the reviewers' comments we have analyzed the additional groundwater data. Piezometer data have now been added to the revised version to support the soil moisture data and the drawn conclusions by means of two figures (Figures 6 and 8). Although the data cannot be used quantitatively due to temperature sensitivity, the data can indicate the timing of water table rise and differences between the upslope and convergent areas. The first figure (Figure 6) shows that the timing of groundwater table rise is similar to the timing of step 2 in the soil moisture data (Figure 9). Also, the data suggest that the groundwater table reached the surface along the central trough. In the upslope area, groundwater tables remained lower. The second figure shows the development of a cross-slope groundwater ridge (Figure 8). This figure is in response to a comment made by George Waswa, pointing out that soil can be tension saturated and thus that piezometric or tensiometric data is required to justify the presence of a groundwater ridge. Despite the fact that the piezometers do not give us independent and reliable groundwater levels, the values listed can be trusted relative to each other and thus the development of a ridge is very clear.

The results of simple column storage calculations have also been added to support the importance of subsurface flow from the upslope to the convergent area (page 8, lines 13-19). These calculations show that the convergent area saturated much sooner than would be expected based on the initial soil moisture conditions and rainfall rate alone. Similarly, storage in the upslope area was lower than would be expected.

Response to reviewer Erwin Zehe

We would like to thank Dr. Zehe for his positive review and constructive comments, which we address below.

Main points

1) The reviewer points out that the results of the hillslope are dependent on the boundary conditions of the hillslope. Particularly due to the impermeable lateral boundaries the reviewer questions whether the results are typical of hillslopes or of small confluent catchments.

We agree that the impermeable lateral boundaries suggest that the dynamics are in fact typical of small confluent catchments or zero order basins. However, we expect that this system would behave similarly without the lateral impermeable boundaries, as is the case for natural hillslopes. The topography of the surface as well as the shape of the impermeable 'bedrock' as shown in Figure 1 suggest that even water falling along the outer edges of the hillslope will not flow along or over the lateral boundaries, but at a slight diagonal towards the trough as in a fishbone pattern. Once the groundwater table extends to the sides of the experimental hillslope, the impermeable lateral boundaries become relevant. In natural hillslopes, water would move across these boundaries to adjacent hillslopes. However, in natural rain events or storms, adjacent slopes will receive similar amounts of rainfall and thus the groundwater table may rise in a similar fashion, sustaining the no-flow boundary condition at the topographic divide. This discussion has been included in the revised manuscript (page 10, lines 8-19).

We can further consider the case where a planar hillslope is placed adjacent to the confluent B2-LEO hillslope and is exposed to the same forcing. We would expect groundwater tables in the planar hillslope, where there is no significant lateral flow, to be higher than along the edges of the confluent hillslope, where there is lateral flow towards the central trough. This difference in groundwater level would result in flow over the boundary from the planar hillslope to the confluent hillslope. While the flow over the lateral boundary would change the magnitude and timing of the response, the overall dynamics in the confluent slope would be similar to what we observed during the experiment. This illustrates the relevance of the dynamics we observed to hillslopes.

2) The reviewer mentions that it would be interesting to benchmark TOPMODEL with the experimental data, which would improve the manuscript and underpin the potential of the B2 (LEO) hillslopes.

We agree that the hillslope setup has the potential to test the assumptions and concepts in hydrological models. This experiment has already been simulated by physically based models, based on the 3D Richards' equation (Niu et al. 2014, HESS). Other colleagues are currently testing the hillslope-storage Boussinesq model with data from our experiment. However, so far no work has been done to benchmark TOPMODEL, and we believe this interesting idea is outside the scope of the current work. Even so, we agree that the hillslopes are very suitable for this and it would be interesting to work on this in the future.

3) In the last main point, the reviewer asks for an indication of how likely the experimental conditions are to occur in reality, such as providing the return period for the rain event in Tucson and other climates.

Though the B2 facility is located in Arizona, the experiment was not designed to be in line with the local extreme rainfall characteristics, but rather the aim was to bring the hillslope in a hydrologic steady state. The rainfall rate was chosen based on its relatively even spatial distribution and the irrigation was stopped when the (unplanned) overland flow was observed. The resulting event is comparable, at least in the magnitude of the 24h precipitation sum, to events that trigger (flash) floods and/or landslides around the world. However, the main focus of the research is not to reflect ambient conditions in certain natural hillslopes, but to observe underlying hillslope hydrological processes in great detail and under the simplified (but controlled) conditions of the artificial hillslope compared to natural hillslopes. This has been made clearer in the revised manuscript (page 3, line 24 to page 4, line 2).

Technical points

- Figure 5: Might be instructive to plot cumulated storage against cumulative rainfall?
This figure was included in the response on the online discussion forum. Due to the constant rainfall rate, the figure is very similar to Figure 5b of the online manuscript. At the beginning of the event, the storage closely follows the 1:1 line because runoff has not yet started. At the end of the event, the storage data increasingly deviates from this line as runoff increases. In the revised version, we have added a line representing cumulative rainfall to the revised version of Figure 5b.

- Please specify the error margins of your measurements.
The error margins of the soil moisture sensors and load cells have been added to the manuscript (page 5, line 28; page 6 line 5). Piezometer errors are dominated by temperature sensitivity. Under normal applications these piezometers hang in deep wells where temperature is more or less constant. In our application, these piezometers are mounted from below the hillslopes and are subject to rather large diurnal temperature fluctuations. It was impossible to find reliable correction methods as the impact of T fluctuations on the piezometer readings kept changing between days. The piezometer errors are discussed on page 7, lines 14-22 of the revised manuscript. We have since replaced the vibrating wire piezometers with Campbell pressure transducers that were tested in great detail, so we have restored the capability to measure water table dynamics in LEO.

- You explain the overshoot of the soil moisture observations by the influence of the capillary fringe of the ground water table. Can you specify how this should work for a TDR or and FDR sensor with respect to the measurement principles?

The overshoot is due to limitations of our calibration curves. We are in the process of recalibrating the sensors. Meanwhile we know that capping the sensors at an average porosity of 39% yields good results compared to the load cell readings, which we consider to be reliable (see Fig. 5b of the online manuscript).

- How did you measure the retention curves?

Retention curves were measured in the laboratory. Soil cores were taken from several depths of a barrel that had been filled with the same material as the hillslope and also compacted similarly. In the laboratory, the retention characteristic was established for these cores using Tempe cells and a WP4-T Dewpoint Potentiometer for the wet and dry ends, respectively. This information has been included in the revised version of the manuscript (page 5, lines 4-9).

- Please specify the hydraulic conductivity curve of the material. Do you expect ks to be anisotropic (now and in the long term future)?

Based on the water retention characteristics, the K(theta) curve is described by the following parameters, which have been added to the retention curve in Figure 2 of the revised manuscript:

theta sat = 39
theta res = 0.08
alpha = 1.86
n = 1.76
m = 0.43

We are also planning to determine the curve experimentally. We do not expect ks to be anisotropic now because the material and compaction are homogeneous and because this was the first experiment performed on the hillslope. However, we expect anisotropy to play a more important role in the future as hydrological pathways develop, especially after vegetation is planted (page 11, line 25-27).

- Subsurface hydrological dynamics at chicken creek (a large artificial hillslope) turned out to be pretty much contaminated by artificial structures (capillary barriers between cones when the site was filled). Do you expect B2-Leo to be free from this? If so I would expect symmetric patterns of saturation in Cross section B. This is not the case for the early stage of the experiment. Where does this come from - fingering?

During construction of the LEO hillslopes, great care was taken to fill and compact the material homogeneously. Instead of flattening and filling soil cones as was done in Chicken Creek, loose material was spread over a cross-slope strip of the hillslope to a certain depth. Then the material was compacted to another specified depth. This process was repeated for several (vertical) layers and (horizontal) cross-slope strips moving from the toe of the slope to the upper end. This method is described in the revised version of the manuscript (page 5, lines 1-4). Also, great care was taken in choosing and preparing the source material for the hillslopes to ensure a homogeneous texture, whereas the material at Chicken Creek was from a natural source and therefore heterogeneous.

We agree that the early stage of Cross section B is not entirely symmetrical, as would be expected. However, this is likely due to small-scale variation due to the indicated time and location of the cross section shown in Figure 7 of the revised manuscript. The evolution of the early phase, with the propagation of the infiltration front, is in fact quite similar at hillslope scale. This is evident from the small error bars on the timing of the first step as shown in Figure 9 of the revised manuscript. This point has been added to the discussion on page 11, lines 15-20.

Response to Reviewer #2

We thank Anonymous Referee #2 for the constructive comments. We address these below.

Main points

1) The reviewer states that despite the high density of sensors, many state values are unknown. These factors could affect the results. Given the differences in volumetric water content and hydraulic conductivity as measured in the laboratory and observed in the experiment as well as the time that has passed since the hillslope was filled, the reviewer questions whether the hillslope can be regarded as homogeneous.

Even though the soil was packed on the hillslope in such a way as to create homogenous hydraulic properties, the scale of the hillslope prevents true homogenous conditions. We expect the subsurface structure to constantly change as more experiments will be executed. Actually the observation of this evolving subsurface structure and its effect on hydrology is one of the main objectives of the long-term experiment of LEO. This has been made clearer in the project description (page 3, line 26 to page 4, line 2).

The measured water content values exceeding the maximum porosity are a sensor issue. We tested the volumetric water content sensors and found that the sensors measure values exceeding the maximum porosity when under the influence of a capillary fringe or groundwater table. The overshoot is due to limitations of the calibration curves. In the experimental data, this is supported by the similarity of the storage estimate based on load cell data and the estimate based on water content data capped at 39% (see Fig. 5b of the online manuscript). The soil moisture sensor issue is discussed in depth in the Results section (page 7, lines 1-13).

2) The reviewer points out that some traditional models conceptualize the observed two-step saturation process by a tension water reservoir and a free water reservoir.

As suggested by the reviewer, conceptual models that have a tension and free water reservoir, which interact as described, can represent the two-step saturation process. However, lumped conceptual models cannot simulate the effect of convergence and lateral flow that played such an important role during the experiment. We thank the reviewer for the references and have added this point and the references to the Discussion section (page 12, lines 5-24).

3) Finally, the reviewer suggests that more data be collected on soil structures and hydraulic characteristics, for example by performing dye tracer experiments.

We are performing additional soil characterizations in the lab to determine the K(theta) relationship. In the revised version of the paper we have added piezometric data (Figures 6 and 8) to support our observations based on soil moisture data regarding soil saturation (see response to George Waswa below).

Other concerns

- P2L11-14L: In some conceptual models, similar two-step soil saturation has been considered like the Sacramento model and the Xinanjiang model.

This point and the references have been included as discussed above. Thank you.

- P3L12-14: Can the filled homogeneous material represent the highly heterogeneous soils of natural hillsopes?

The aim of the Landscape Evolution Observatory is not to reflect conditions in certain natural hillslopes, but rather to study the underlying hydrological processes in great detail. This is facilitated by the controlled conditions and simplified design. This has been made clearer in the Introduction section (page 3, line 26 to page 4, line 7). We expect subsurface heterogeneity to develop over the course of the 10-year experiment, which has been made clearer in the manuscript as well (page 11, lines 25-27).

- P4L5: Since 2009, have the overall shape and relief of the hillslope been changed under sprinkler tests? Do you re-shape the micro-terrain every time after a rainfall experiment?

As mentioned previously, the hillslope was completed in 2012. After completion, a small number of sprinkler tests were performed to characterize the rainfall distribution. After these tests, the surface was reshaped once to restore the initial condition of the surface before performing the experiment. The surface is not reshaped after every rainfall event because the aim of the Landscape Evolution Observatory is to observe the evolution of the hillslope through time.

- P4L15: Could the large difference between measured saturated hydraulic conductivity and the effective value be due to preferential pathways like soil cracks within the soil material? If you could provide more soil hydraulic experiment, e.g., a 1D soil column test, it will help readers understand your observations.

We do not expect the difference in hydraulic conductivity estimates to be due to preferential pathways because of the care that was taken in filling and compacting the hillslope. Also, soil cracking is unlikely because the soil does not contain clay minerals that are required to create soil cracks due to swelling and shrinking when minerals absorb and release water. The reported clay fraction of 3% refers to particle size only, not mineral composition. In Fig. 7b (Figure 9b in the revised manuscript), the confidence interval of the median as shown by the error bars is very narrow, showing that the propagation of the infiltration front was uniform over the hillslope. Finally, the subsurface runoff starts after the infiltration front reaches the bottom of the hillslope in the convergent area. If preferential flow paths would be an important factor, we would expect some of the deepest sensors to saturate before the arrival of the infiltration front as shown in the figure and/or the subsurface runoff to start before the infiltration front reached the bottom of the hillslope. We have expanded the discussion of preferential pathways in the revised manuscript to include soil cracking (page 11, line 23-25).

- P5L9: Did you only carry out one rainfall experiment?

Yes, only one rainfall experiment was performed for this study.

- P5L24: In Figure 5a, the volumetric water contents in Phase 3 exceed the maximum porosity. Could it be the reason that there are some macropores which lead to higher hydraulic conductivity and water contents?

The overshoot in water content values is a sensor issue, as discussed in response to the first main point.

- P6L27: In figure 7, it is better to add the rainfall line in the legend.

We have included the rainfall line in the legend as suggested. Thank you.

- P8L14-21: The discussion that lateral distribution of water was a major contributor to groundwater table and overland flow generation in this part is not sufficiently supported by present data. Because justly in L10-13, the authors admit that the volume water content will not exceed the maximum soil porosity at the bottom of the soil profile at the toe of the slope, where should mostly tend to be saturated due to lateral subsurface flow.

The importance of lateral redistribution of water can be seen by the differences in the response of the convergent and upslope areas as illustrated in Figures 6 and 7 (Figures 7 and 9 in the revised manuscript). In addition, simple column storage calculations show that soil columns in the central trough saturated sooner than would be expected by initial conditions and rainfall rate alone. In upslope areas the storage was lower than would be expected (page 8, lines 13-19).

Although Figure 6a (Figure 7a in the revised manuscript) suggests that the soil is not saturated at the toe of the hillslope, we expect that this is not what really happened. The measured water contents reached 36-37%, within a few percent of the maximum porosity, and the sensors have an error of ±2%. Therefore we expect that in reality these locations were saturated. These points have been added to the Discussion section (page 11, lines 9-14).

- P9L5-8: Could you really exclude macropore flow? Soil macropores can be shaped through vegetation roots, worm holes as mentioned by the authors. While other factors like the processes from wetting to drying, freeze thawing… also cause soil cracks. So I suggest a dye tracer experiment may help to verify it.

We do not believe macropore flow to be significant at this stage (recall that the construction of the hillslope was completed less than two years ago). The absence of clay minerals makes soil cracking unlikely, as mentioned previously under the comment of preferential flow paths. The hillslope is bare soil and was never vegetated and no worms have been observed. These points have been made clearer in the Discussion section (page 11, lines 20-25). A dye tracer experiment is not practical because this would disturb the hillslope.

Response to Reviewer #3

We thank Anonymous Referee #3 for the constructive comments, which we address below.

1) The reviewer states that the results of the study support, rather than challenge, existing hydrological theory.

We agree that the results support existing theory in which vertically infiltrating water reaches an impermeable boundary, forms a groundwater table and subsequently moves downslope. However, there are also observations that are not easily explained. For example, the decreasing water content with depth after the passage of the wetting front begs an explanation. We made the distinction between observations that are in line with hydrological theories and observations that are not easily explained clearer in the revised version (page 13, line 6).

2) The reviewer recommends to define the groundwater ridge more clearly and to make it clear that the topographic effect is due to the bedrock topography rather than the soil surface topography. Also, the reviewer suggests that the groundwater in the trough is divergent rather than convergent in the lateral direction, contradicting the authors analysis that the side slopes contribute to the ridge.

The groundwater ridge is present in the lateral cross-section rather than the longitudinal direction, as suggested by the reviewer. We also agree that the topography of the impermeable 'bedrock' rather than the shape of the soil surface causes convergence. This has been made clearer (for example, page 10, line 13-14). However, we still argue that lateral flow from the side slopes contributes to the ridge formation. Simple column storage calculations show that the convergent area saturates sooner than would be expected due to rainfall alone, and vice versa for the upslope areas. This has also been added to the revised version (page 8, line 13-19).

3) The reviewer suggests we test the volumetric water content sensors in controlled cases to learn more about the overshoot sensors show under saturated conditions.

Controlled testing of the volumetric water content sensors showed that the measured water content exceeds the porosity when under influence of a water table. However, the measured values under saturated conditions were not compared to pressure head. At present, the sensors are being recalibrated to prevent these issues from happening in future experiments.

4) The reviewer indicates it would be helpful to discuss the relevance of the experimental rainfall to natural conditions.

The initial intention of the rainfall-runoff experiment was to bring the hillslope response to a hydrologic steady state; it was not foreseen that the applied rainfall intensity would trigger surface runoff and erosion. The rainfall rate was chosen due to its relatively even distribution over the hillslope and the irrigation was stopped when the (unplanned) overland flow was observed. The 24h precipitation sum of the resulting event is comparable to events that trigger floods and/or landslides around the world. However, the main focus of the research is not to mimic natural conditions, but to observe underlying hillslope hydrological processes in great detail and under the simplified (but controlled) conditions of the artificial hillslope, compared to natural hillslopes. This has been made clearer where we describe the LEO project in the revised Introduction section (page 3, line 26 to page 4, line 7).

Response to short comment by George Waswa

We also use this opportunity to respond to the short comment posted by George Waswa. We are grateful for his constructive comments, which have greatly helped improve our manuscript.

1) Subsurface flow and the formation of a groundwater ridge: The reviewer suggests that the results show that upward saturation was a result of accumulation of vertically infiltrating water from above and not from subsurface flow from the side slopes. In fact, it may be possible that the groundwater ridge in the convergent area supplied some water to the side slopes.

We agree that once the infiltration front reached the impermeable boundary at the bottom of the hillslope, further accumulation of infiltrating water started to form a groundwater table. Part of the reason that this started earlier in the convergent area could be, as the reviewer suggests, the wetter initial conditions at the bottom of the central trough. However, based on simple column storage calculations we found that soil columns in the central trough saturated sooner than would be expected based on the initial conditions and rainfall rate alone. Similarly, storage in columns in the upslope area is less than would be expected, indicating that there was lateral flow between the upslope and convergent areas. We have added a few sentences discussing these calculations in the Results section (page 8, lines 13-19).

2) Contribution of subsurface flow to overland flow: The reviewer points out that saturation reached the ground surface at 19 hours after the start of the rainfall event, while overland flow had started 5 hours earlier (Fig. 7b), which makes Hortonian overland flow more likely than saturation excess overland flow. Also, the contribution of subsurface flow to overland flow can only be verified if saturation in the convergent area is of phreatic water rather than saturation due to water held in tension.

The estimation of overland flow based on the water balance indicates that overland flow may have started as early as 14 hours into the experiment, but the error bars show that we can only be expect significant overland flow after 20 hours. Before this time, error bars extend to 0 (Fig. 9) due to uncertainties in the water balance analysis used to estimate overland flow before the end of the rainfall event. We do not expect that Hortonian overland flow occurred for several reasons. Firstly, the constant rain rate is lower than the saturated hydraulic conductivity, both as determined in the laboratory and as determined by model calibration. Also, overland flow did not start until after many hours of rainfall and once it started, overland flow was limited to the central trough. In subsequent experiments with the same rainfall rate, but shorter duration, no ponding or overland flow was observed. Moreover, model simulations based on the 3D Richards' equation do not show overland flow due to infiltration excess (Niu et al., 2014, HESS), but confirm overland flow due to saturation excess in the central trough. We have added a discussion on the runoff generation process and listed the reasons why Hortonian overland flow can be excluded in the revised manuscript (page 16, lines 2-14).

While at some moment it is possible that tension saturation at the surface caused saturation excess overland flow to occur, the continuation of overland flow for almost a day after the end of the rainfall event (Figure 9 and page 9, lines 28-29) indicates that it was mainly exfiltrating subsurface flow that contributed to overland flow. Though piezometric data would help justify this interpretation, we chose not to include the data in the initial version of the manuscript due to the temperature sensitivity of the vibrating wire piezometers installed in the hillslope. Despite the temperature sensitivity, the measurements can give us an idea about the timing of water table formation and when the water table reached the surface. While the retrieved values should not be interpreted as accurate representations of actual groundwater levels (the values suggest groundwater levels extending below the hull of the hillslope as well as above the soil surface), the data in Figure 6 support the timing of water table rise as seen in Figure 9 of the revised manuscript. Also, the data indicate that the water table reached the surface in the central trough. A summary of this discussion has been included in the revised manuscript (page 19, lines 15-20).

3) Groundwater ridging and water table: For effective discussion of a groundwater ridge, it might be necessary to include piezometric data to distinguish between the phreatic surface and the saturation level.

We agree that the inclusion of piezometric and tensiometric data would be of great value. However, the MPS-2 sensors installed in the hillslope saturate at pressures of -6 kPa and therefore do not provide pore water pressure data under wetter soil conditions. The piezometric data, as mentioned earlier, are temperature sensitive and in future studies, the vibrating wire piezometers will be replaced by pressure transducers. However, the piezometric data support the presence of a groundwater ridge in the central trough and we have included a new figure in the revised manuscript to show this (Figure 8). The development of the ridge as shown in this figure is similar to what was observed in the soil moisture data (Figure 7).

4) Physical processes involved in groundwater ridging: The reviewer references several representative studies in environments suitable for groundwater ridging that do not clearly show rapid groundwater table rise and/or groundwater table rise to the surface. From these studies, it appears that the rainfall intensity plays a more significant role than just filling the capillary meniscus.

We agree that the discussion of groundwater ridging in the online version of the manuscript is limited. In this comment, the reviewer provides some very useful information and references to take into consideration. We have enhanced the discussion of groundwater ridging in the revised manuscript (page 15, lines 1-14).

The reviewer continues to state that there could be various interpretations of the results based on the limited data supplied, such as air entrapment ahead of a wetting front. This might account for some observations, such as the reduced wetting front velocity, decreased soil water content in the infiltration profile and the upward saturation front.

We considered air entrapment as an explanation for the decreasing water content after the passage of the infiltration front and other observations. However, there are a few reasons why we deem it unlikely, such as the scale of the hillslope and the fact that we did not observe any ponding. Furthermore, we observe the decrease in moisture content with depth during times when none of the overlying soil is saturated, leaving enough empty pores for pressurized air to escape. Also, the rainfall intensity distribution is not completely uniform over the hillslope, making air entrapment unlikely. At this time we think the decreasing water content in the second phase, with the infiltration rate remaining constant, can be explained by changes in water retention characteristics with depth.

The discussion of air entrapment and other explanations has been added to the manuscript (page 13, lines 12-22).

The reviewer concludes by questioning whether the laboratory samples could have been more compact than soils in the hillslope. This could explain the significant difference between the effective hydraulic conductivity and the laboratory measured hydraulic conductivity as well as the observations that showed water contents exceeding the maximum porosity.

The compaction of laboratory samples was similar to that of the hillslope: large barrels were filled with the soil material in layers, and after the addition of each layer the soil was compacted to a specific level (added to manuscript on page 5, lines 1-5). After this was done, samples were taken at different depths. Extra testing of the 5TM sensors showed that sensors overestimated soil moisture once the sensors were influenced by the capillary fringe or a groundwater table. The manufacturer (Decagon) confirmed that values above total porosity are due to calibration issues. Finally, when a cutoff value equal to the porosity is used for soil moisture, the storage estimates based on these values are almost identical to the mass accumulation measured by the load cells (Figure 5). The discussion of the overshoot observed in soil moisture data has been made clearer in the revised version (page 7, lines 1-13).

Other relevant changes

- In the title, ‘Hillslope’ was replaced by ‘Hillslope-scale’ to emphasize the unique size of the artificial hillslope used in our study.
- An opening problem statement was added to the abstract, as well as a closing statement linking this first experiment to future work at the Landscape Evolution Observatory.
- Page 4, line 8: The rainfall event is characterized as ‘extreme’ rather than ‘intense’. The rain rate is not particularly high, but the duration is long.
- The structure of the discussion section has been changed due to the number of added discussion points.
- Page 14, lines 12-16: We added a few sentences discussing the power of the calculated rates of groundwater table rise as shown in Figure 9. We point out that the power of these rates of change is reduced by the size of the error bars and the limited number of measurements at 35 cm depth in the upslope area. This could be a reason that the observed rates of groundwater table rise at hillslope scale are not entirely as expected. For example, though we do not observe accelerated groundwater table rise as we would expect due to the decreasing available pore space (Figure 4b), the size of the error bars indicate it is possible such an acceleration did occur.
- A figure was added to the supplementary material (Figure S1) to demonstrate the relevance of the two-step saturation process at hillslope scale (in the text, the figure is described on page 6, lines 24-26 and discussed on page 11, lines 1-7).

Reference

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 generation – insight from a modeling study of the first experiment at the Biosphere 2 Landscape Evolution Observatory, Hydrol. Earth Syst. Sci., 18, 1873-1883, 2014.