Reply to the comments on “Analysis of an extreme rainfall-runoff event at the Landscape Evolution Observatory by means of a three-dimensional physically-based hydrologic model” by G.-Y. Niu et al.

Dear Editor,

All three reviewers think the paper is very interesting but suggested we improve the paper’s presentation and include analyses of the LEO soil moisture data.

We have revised the paper according to the three reviewers’ comments and suggestions. We followed the third reviewer’s structural comments, we addressed some common issues raised, and we made numerous other changes in response to specific suggestions. We revised each part of the paper including the title, abstract, introduction, and model setup, and we added a discussion section that addresses some of the main concerns of the reviewers. We also revised Figure 3 to show the exact locations of the seepage face nodes.

A detailed analysis of the soil moisture data has been included in another paper submitted to HESS-D, entitled “*Hillslope experiment demonstrates role of convergence during two-step saturation*” led by A. I. Gevaert. We cite this paper to confirm the saturation-runoff generation mechanism as simulated by our model.

Detailed replies to each of the questions of all three reviewers are attached here.

We hope that you will find our responses to the reviewer comments and our changes to the paper satisfactory, and we look forward to hearing from you.

Sincerely yours,

Guo-Yue Niu

Anonymous Referee #1
Received and published: 13 December 2013

General comments:
The paper presents numerical experiments conducted using data from the one hillslope of the Landscape Evolution Observatory (LEO). The concept of such experimental set up is very interesting as it allows testing functioning and modelling hypotheses under controlled conditions. The considered data set corresponds to rainfall simulation at a homogeneous rainfall rate, following by no rainfall. The land surface is bare soil. This first experiment was designed to test the functioning of the installation, but providing interesting data, all the more than the observed behaviour was completely different from the expected one, as given by previous numerical modelling. In particular overland flow and the formation of a small gully were observed and were not predicted by previous simulations. The objective of the numerical experiments is to investigate possible reasons for this mismatch. The question is of interest.

R: Thanks for the comments and for the time you have taken to review this paper.

However, only one general hypothesis, i.e. a possible heterogeneity of the soil hydraulic conductivity at the seepage face is considered, and the hillslope soil is still supposed to be homogeneous. Although the hillslope was artificially built, it is very likely that some soil heterogeneity is present in the soil and may also explain the unpredicted behaviour of the hydrological response. The authors could refer to interesting findings in the artificial Chicken Creek catchment built in Germany (e.g. Hofer et al., 2011, 2012; Hölzel et al., 2011 and more generally a special issue of Physics and Chemistry of the Earth, vol 36 (1-4), 2011). In the present paper, the authors have realised thousands of simulations with different homogeneous soils, but it would also have been possible to test the impact of possible heterogeneity in the soil properties (both horizontally and vertically).

R: LEO was built with homogenous sandy loam and carefully compacted at every 25 cm thickness to explore the co-evolution of the soil-water-plant system from an initially homogeneous soil. This is the reason we started our simulations with the homogeneity hypothesis.

The main purpose of this paper is to explore the possibility of heterogeneity development in the LEO soil and its effects on overland flow generation, while considering the uncertainties in soil property parameters.

We could include vertical heterogeneity in the simulations. However, this would significantly increase computational demands. It took about a week to finish the 20,000 runs presented in the paper. Combining the "vertical heterogeneity" (regarding the uncertainties in soil parameters at different layers) would take about "n" weeks, where n is the total number of combinations of varying parameter values in the vertical direction and can be easily about 1000. We are using a physically-based model based on the 3D Richards equation, which is much more time-consuming than the one developed in Hofer et al. (2012) for the Chicken Creek

catchment. Also, including vertical heterogeneity would not alter the final conclusion (that confirms incipient heterogeneity).

However, we do agree that vertical heterogeneity is important in the broader context and that it may have developed in LEO, in this way affecting the internal moisture states. This study was not aimed at investigating *where* the heterogeneity developed (although we did set different Ksat values at the seepage face) but rather to answer *what* happened in the LEO soil that produces a different hydrological response (especially the overland flow) from what was expected based on preliminary analyses.

We are currently running the CATHY model to simulate soil moisture at different layers and horizontal locations.

These points are now mentioned in the new Discussion section.

In addition, the authors mention the existence of lots of sensors measuring water pressure and water content. It could be interesting to analyse those data before building the hypotheses tested using the numerical model.

R: A detailed analysis of the soil moisture data of the LEO soil is included in another paper, "*Hillslope experiment demonstrates role of convergence during two-step saturation*" led by A. I. Gevaert, which has been submitted to HESS-D. We cite this paper to support our modeled mechanism of overland flow production.

The impact of possible macropores could also be analysed.

R: We have added the following paragraph to the Discussion section:
"*The model we used in this study solves the Richards equation based on Darcy-Buckingham theory, resolving matrix flow and not macropore flow. There are many modeling studies that use percolation theory and other approaches to simulate hydrologic connectivity of macropores to form preferential flow pathways and threshold-like hydrological responses (e.g., Lehmann et al., 2007; Hofer et al., 2011). At this early stage of LEO, with complete absence of organic matter and vegetation roots, we do not anticipate macropore-related processes to be dominant. Macropores might possibly exist around the sensors, although in this case subsurface flow would be enhanced and would very likely have prevented generation of overland flow*."

The feeling when reading the paper is that the authors try to get good simulations of discharge, but for the wrong reasons.

R: The main focus of our paper is not on getting good simulations but rather on hypothesis testing, as made clearer in the revised manuscript's the Discussion section:
"*A thorough investigation of the fine particles at the seepage face or upslope is not feasible as this would alter the soil structure of LEO-1. The physically-based*

hydrological model used in this study allowed us to make a probabilistic assessment of the incipient heterogeneity hypothesis while considering uncertainties in soil parameters. Under heterogeneous conditions the model produced better results for seepage flow and total water storage, as well as overland flow that is comparable to estimates from a water budget analysis. It was not our intention to improve the modeling accuracy through parameter calibration but to test the hypothesis of incipient heterogeneity development."

My point of view is that the publication of the results presented in the paper may be premature and that it could be more efficient to first analyse the data more in depth before possible publication of numerical simulations results.

R: See above responses that clarify the context and objectives of our study.

In addition, the paper does not detail enough some important part of the experimental design, the model used, his set up and this requires further attention.

R: We added two more references about the model, one on subsurface water and another on surface water routing.

We also revised the experimental design and model setup, and we added seepage face nodes in Figure 3.

The reference list is also very short and almost only limited to publications about the LEO. A comparison of the authors results with results from the literature would be welcome.

R: We added a Discussion section and a broader range of references.

More detailed comments are provided below.

Specific comments:

1) p.12620, section 2.2. The model description is very short and more information could be provided on the model functioning, numerical discretization, parameters required. Some points on how macropores and or are not taken into account could be useful also.

   R: We added more references on the model and some comments on macropores (see above points).

2) P.12620, section 2.3. More information about how the LEO hillslope was built could also be useful. Was it built to get a homogeneous soil and if yes how was this achieved?

   R: Additional information on the hillslope construction and constitution has

been added in the Introduction and in Section 2.1.

Is the rainfall applied over the whole hillslope or only at the top of the hillslope?

R: On the whole slope. The paper has been revised to make this information explicit (Section 2.3).

Where are the soil moisture sensors located? Do you have measurements at several depths?

R: This is not a focus of the paper. Detailed information on this is included in the HESSD paper mentioned previously.

Could you explain better how and where the seepage flow is measured? Is it measured at the bottom of the slope? A scheme with the experimental design could be useful.

R: We revised Figure 3 and added a sentence in Section 2.1:
"*Seepage face flow was recorded through six tipping buckets and six flow meters installed at six sections of the seepage face*."

3) P.12622, lines 1-9. The specification of the upper boundary conditions is quite rough? Did you made some sensitivity analysis of possible error on this boundary condition?

   R: Yes, we ran many experiments before the systematic runs.

   Are you sure that the imposed rainfall is homogeneous all along the slope (if it is applied over the whole slop, see also question in point 2).

   No, we are not sure. We compared the homogeneous rainfall pattern to an inhomogeneous pattern that was measured by hundreds of cups before the first experiment. The difference in the modeled seepage flows between the two cases is negligible. These results are not included in the paper.

4) P.12622, lines 19-24. Could you specify clearly that in senarii M1 and M2, the soil is assumed to be homogeneous? The Ksat value of simulation M2 is very large. Some comments about the realism of this value would be welcome.

   R: This is now stated more clearly.

   We already have a sentence that explains this value: "*M2 uses the same parameters except a greater K_{sat} (= 3.8×10^{-3} m s^{-1}) resulting from a calibration against the starting time of measured seepage face flow for a LEO-1 test run with 20 mm h^{-1} of rainfall applied for 5 h in November, 2012.*"

5) P.12623, liens 1-10. The authors mention that they consider heterogeneous configuration, but the way the heterogeneity is taken into account in the model is really not clear. Do you only modify the Ksat of the last layer of nodes (i.e with y value around 60 in Fig .3?). A figure showing how the heterogeneity is considered would help understanding what is really done.

   R: We revised Figure 3 to convey this better.

6) P.12623, lines 10-12. Why do you retain such a narrow range for Ksat and Ksat,sf, as compared to the range used in M1 and M2?

   R: We actually did more experiments with a wider range of parameter values than described in the paper. We focus our attention only in the narrow ranges because our model generates a feasible overland flow only within this range.

7) P.12628, lines 25-28. The authors mention the existence of soil moisture measurements scattered within the whole slope. It would be necessary to assess the relevance of the model simulations/hypotheses with these data, before concentrating on only one functioning hypothesis: soil heterogeneity at the seepage face, but without questioning the hypothesis that the remaining of the slope is homogeneous.

   R: As mentioned above this is included in another HESSD paper, and further analyses of LEO experiments using also internal data is ongoing.

8) Section 4. The discussion should be enhanced with reference/comparison with other studies.

   R: We added a Discussion section, citing more papers:

"*Unlike other artificial large-scale hillslopes such as Hydrohill in China (Kendall et al., 2001) and Chicken Creek in Germany (Gerwin et al., 2009; Hofer et al., 2011), LEO was built with homogenous soil and with a focus on evolving heterogeneity from a "time-zero" homogenous condition through co-evolution of the soil-water-biota system over a time scale of years (Hopp et al., 2009; Dontsova et al., 2009). Development of catchment morphology and soil catena driven by hydrological processes through soil erosion and deposition may be one of the major causes that induce heterogeneity and that in turn exert strong feedbacks on hydrological processes (e.g., Beven et al., 1988; Sivapalan, 2005; McDonnell et al., 2007; Troch et al., 2009). At LEO it was not expected that soil heterogeneity would develop in such a short time period during an intense rainfall event that induced significant subsurface saturated flow. This is one of the main reasons that our pre-experiment model predictions failed to produce overland flow.*

A thorough investigation of the fine particles at the seepage face or upslope is not feasible as this would alter the soil structure of LEO-1. The physically-based hydrological model used in this study allowed us to make a probabilistic assessment of the incipient heterogeneity hypothesis while considering uncertainties in soil parameters. Under heterogeneous conditions the model produced better results for seepage flow and total water storage, as well as overland flow that is comparable to estimates from a water budget analysis. It was not our intention to improve the modeling accuracy through parameter calibration but to test the hypothesis of incipient heterogeneity development.

The model we used in this study solves the Richards equation based on Darcy-Buckingham theory, resolving matrix flow and not macropore flow. There are many modeling studies that use percolation theory and other approaches to simulate hydrologic connectivity of macropores to form preferential flow pathways and threshold-like hydrological responses (e.g., Lehmann et al., 2007; Hofer et al., 2011). At this early stage of LEO, with complete absence of organic matter and vegetation roots, we do not anticipate macropore-related processes to be dominant. Macropores might possibly exist around the sensors, although in this case subsurface flow would be enhanced and would very likely have prevented generation of overland flow.

In this modeling study we assume that all soil parameter values vary horizontally and are static during the modeling period. Evolution of heterogeneity due to coupled water and sediment transport processes, which may occur in particular under intense rainfall conditions, is beyond the ability of state-of-the-art hydrological models and requires more attention in ongoing efforts to develop coupled Earth system models. Likewise, soil erosion models that consider only surface processes (e.g., Hofer et al., 2012) are also inadequate to this task."

9) P.12629, lines 1-10. This paragraph should come sooner in the discussion.

   R: Yes, we moved this paragraph to an earlier section.

10) Figure 4. Could the authors provide more information about the way red dots are obtained? Why are the data horizontal beyond 0.20 m3/m3. If this relates to the explanation given p.12629, lines 5-8, then the data should be removed from the analysis.

   R: We modified the text for Figure 4. We prefer to keep the data in the picture, because the curve is still indicative of the greater *n* value at least for the relatively dry range.

Anonymous Referee #2
Received and published: 30 December 2013

This is an interesting paper that points to the enormous potential of the artificial hillslope laboratory, LEO. The results of the first experiment conducted in LEO are already interesting, however the description of the results has the effect of guiding the reader to more traditional questions. This is unfortunate, and I hope the authors can fix this problem in a resubmission at the end of this discussion.

First of all, when viewed one way, the results show that, however well-conceived, one can never achieve perfect homogeneity in the real world. However, it is clear that this was already expected by the developers of LEO, in that the focus of the experiment on not reproducing the real world, but on exploring how heterogeneity evolves over time, indeed how hydrological variability and landscape heterogeneity co-evolve. The discussion of the results already indicate that this is already happening, in that the authors are ascribing differences between model predictions and actual observations to this emerging heterogeneity, explaining the compaction (even relative compaction) of the soils even as the experiment is happening.

R: Thanks for the comments and for your time in reviewing the paper.

Given all this, the focus on characterizing on the errors between outputs from various model configurations and actual observations, gives the impression that they are merely asking traditional questions, i.e., fitting a hydrograph, in this case for just one event. I am not against these details, as the modeler still must get the model to mimic the observations, and there is certain amount of equifinality in this fitting.

R: We revised the text everywhere in the paper to reflect that this is not a traditional calibration study. We wanted to address the impacts of soil property uncertainties on our conclusion about heterogeneity development, resulting in a probabilistic assessment of heterogeneity. We added a paragraph in the Discussion section:
"*A thorough investigation of the fine particles at the seepage face or upslope is not feasible as this would alter the soil structure of LEO-1. The physically-based hydrological model used in this study allowed us to make a probabilistic assessment of the incipient heterogeneity hypothesis while considering uncertainties in soil parameters. Under heterogeneous conditions the model produced better results for seepage flow and total water storage, as well as overland flow that is comparable to estimates from a water budget analysis. It was not our intention to improve the modeling accuracy through parameter calibration but to test the hypothesis of incipient heterogeneity development.*"

However, I would have found the results more informative, for this event and for the best parameter combinations of the model, some deeper insights into the internal dynamics that led to the hydrograph that was observed. For example, the dynamics of the groundwater table during the event, the soil moisture, and the saturation area etc. would provide more insights. Note that it is here that LEO is most innovative and helpful compared to real world field experiments, the ability to observe the space-time dynamics of water partitioning. Also, any additional information on change of structure and heterogeneity will also be insightful, and will shift the focus in the appropriate direction.

R: A detailed analysis of the soil moisture data has been included in another paper submitted to HESS-D, entitled "*Hillslope experiment demonstrates role of convergence during two-step saturation*" led by A. I. Gevaert. We cited the paper to confirm the saturation-runoff generation mechanism as simulated by our model.

Another comment on the presentation: from the beginning the authors framed the aim of the paper as hoping to explain the big difference between the observed and predicted hydrographs. This is the valid approach: however, towards the end the paper veers away somewhat from this goal. I was expecting a clear, conclusive statement on the causes of this difference, and I did not find it. They may want to make sure to go through the entire paper and ensure the main message is carried through to the end.

R: Thanks for the suggestions. We revised the paper significantly from the title all the way to the conclusion.

One final question/suggestion: the title has the word "extreme rainfall-runoff"> What is the motivation for this phrase?

R: We changed the title to "*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*"

Do the authors think that the event studied is extreme as to cause the erosion that happened? The 12 mm/hr intensity does not sound like too extreme to me.

R: We think this event is extreme by the total amount (264 mm) of this single event, generating significant subsurface flow and overland flow. Now we call it "an intense rainfall event" in the paper, though it is not very intense in intensity.

Another question/suggestion about the title: it might be better for the title to reflect the main message coming out of the paper. As it is now, the title is somewhat neutral, and does not attract attention to the main question/issue that is really highlighted in the paper.

R: Thanks for the suggestion. We changed it (see above).

Overall, I like this paper and would like this paper to be eventually published in HESS. I would prefer if the paper undergoes some (perhaps moderate) revisions to address the concerns raised above and attract sufficient attention to some really important issues in hydrologic process understanding and distributed modeling.

Anonymous Referee #3
Received and published: 11 January 2014

In this work, the physics based model Cathy is used to investigate the response of an artificial hillslope to the application of a uniform rainfall input. The artificial hillslope (named LEO) is built by using a homogeneous soil. Based on the supposedly well known soil properties and on earlier modelling applications, the rainfall was applied uniformly in space and time to bring the hillslope to a hydrologic steady-state. However, the hillslope never reached the predicted steady-state but instead developed saturation excess overland flow. The work aims to understand why did the observed hydrological response differ so significantly from the predicted response. The answer offered by the authors is that the experiment itself triggered some form of heterogeneity in the soil hydraulic conductivity of the seepage face. Indeed, when this heterogeneity is included in the model building, the model is shown to be able to accurately reproduce the hydrologic response. The topic is very interesting, in that it shows the hydrological implications of processes which introduce heterogeneity in a supposedly homogeneous environment. The objectives are of interest for the readers of HESS and the writing is good. Nevertheless, the paper lacks focus and a clear story line and suffers from structure.

R: Thanks for the comments and for your time in reviewing the paper.

Lack of focus and a clear story line: The title is misleading. The 'investigation of an extreme rainfall-runoff event' is evidently not the central focus of this paper. The introduction reserves too room for the general description of the LEO experiment, whereas too few is dedicated to establish a link between one main objective of LEO (examination of co-evolution of the physical and biological system) and the incipient heterogeneity which is tested by the field-numerical experiment. There are essentially no conclusions, in the sense that the implications of the obtained results are not even addressed.

R: We have changed the title, revised the introduction and conclusion, and added a discussion section.

Problems with structure: The main problem with the structure of the work is that it poses a very nice question assisted with a formidable experimental structure, and ends up with an answer which is just barely supported by the multiple monitoring means. The authors should use internal data (at least soil moisture data and soil hydraulic data at the seepage face) to add experimental foundation to the numerical simulations, and to reduce equifinality in the answers they are able to offer. Moreover, the authors should at least address what is the main implication of this field-numerical experiments. In my opinion, this can be stated as follows: during the observed event, the hydraulic properties of part of the hillslope evolved from one value to another. The numerical model cannot reproduce this behavior: it is based on use of static values for the hydraulic properties. This is interesting, because it is a process likely occurring in many intense events, and very often forgotten both by experimentalists and modellers. The discussion of the results should include consideration of this implication.

R: We have taken care to revise the paper throughout to make clearer the context and objectives of our study and the wider implications of our findings. On the issue of using soil moisture data, see our response to the previous reviewers.

Considering the general interesting topic I think that the work might be publishable after moderate revisions. In the following I will try to outline, where and how the manuscript can be improved.

R: We really appreciate the outline, which we followed closely.

Title: The title should focus on the main problem addressed by the work, which is not the investigation of an extreme rainfall-runoff event.

R: We changed the title to "*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*"

Abstract: The abstract should make clear the meaning of the 'saturated soil compactation near the seepage face'. This is apparently due to the transport of fine sediments during subsurface saturated flow prior the onset of overland flow. Moreover, the abstract should make clear how the heterogeneous model is built. The sentence starting with "We varied the saturated: : :" is central for this, but it is definitely hard to understand.

R: We significantly revised the abstract following the suggestions:

"Heterogeneity may have developed during the first experiment at one of Biosphere 2 Landscape Evolution Observatory (LEO) hillslopes. The LEO hillslopes are carefully compacted with homogeneous soil to a depth of 1m. The experiment, driven by an intense rainfall event, produced predominantly seepage face water outflow, but also generated overland flow that caused erosion of the superficial soil and formation of a small channel. In this paper, we explore the hypothesis of incipient heterogeneity development in LEO and its effect on overland flow generation using a three-dimensional physically-based hydrological model. The model simulations consider spatially varying saturated hydraulic conductivity (Ksat), as well as uncertainty in Ksat and other soil parameters. Our null hypothesis is that the soil is hydraulically homogeneous, while the alternative hypothesis is that the soil has developed some heterogeneity in the downstream direction due to transport of fine sediments driven by saturated subsurface flow. The heterogeneous case is modeled by assigning a different saturated conductivity at the LEO seepage face (Ksat,sf). A range of values is used for Ksat, Ksat,sf, soil porosity, and pore size distribution, resulting in more than 20,000 simulations. It is found that the best runs under the heterogeneous soil hypothesis produce smaller errors than those under the null hypothesis, and that the heterogeneous runs yield a higher probability of best model performance than the homogeneous runs. These results support the alternative hypothesis of localized incipient heterogeneity of the LEO soil, which facilitated generation of overland flow."

Introduction: After a short general description of LEO and of its aims, the introduction should describe only the facilities used for the described experiment and the relevant links to the general aims of LEO. This is not the case, and this is where the Introduction must be improved. For instance, the introduction describes the first experiment as a sequence of two artificial rain applications (P4 L18-20), with the second rain application being labeled with deuterium. It announces also that chemical analysis should inform about water transit times. Since the second rain event and the chemical analysis was never executed (or at least is not part of this work), this only adds confusion to the description of the experiment. Even more important: the reader cannot understand from the Introduction if the incipient heterogeneity tested with the field and numerical experiment is a documentation of an already started co-evolution of the system, or it is just due to an accident. Owing to lack of clarity on this, the reader cannot understand why it is important to understand the reasons for the mismatch between predicted and observed hillslope's behavior and which are the potential lessons to learn.

R: We revised the introduction following the suggestions, starting with heterogeneity issues in catchment hydrology:

"*Landscape heterogeneity is ubiquitous at various spatial scales, it may evolve over time, and it induces process complexity that still hasn't been properly addressed in catchment hydrology. As such, predictions of the Earth system response to natural and anthropogenic forcing are currently highly uncertain (Sivapalan 2005; McDonnell et al., 2007; Troch et al., 2009). To develop a unified theory of catchment hydrology, hydrologists should ask questions of "why" the heterogeneity exists rather than traditional questions of "what" heterogeneity exists (McDonnell et al., 2007). This requires an improved understanding of the intimately coupled processes of hydrology, geomorphology, ecology, pedology, and biogeochemistry (McDonnell et al., 2007; Troch et al., 2009).*"

The first LEO Experiment: This text doesn't include information on the hydraulic behavior of the seepage face. Part of this information is instead reported at P14, L18-27, almost at the end of the paper. This last text should be moved into the description of the first LEO Experiment, to provide ground to the choice to decrease the saturated hydraulic conductivity for the seepage elements of the computational mesh.

R: Agreed. We moved this paragraph to "The first LEO experiment" section:
"*Shortly after the experiment we removed the gravel to a depth of 72 cm and determined the fraction of fines per volume of gravel to be about 2%, which may or may not represent a significant reduction in hydraulic conductivity of the seepage face, considering also that precise measurements could not be made over the entire seepage face. In addition we observed some of the holes in the plate to be clogged with fines but were unable to test the effect of this clogging on the hydraulic conductivity of the seepage face.*"

Also, at P7, L8, it is stated that "total mass change, total seepage flow, and soil moisture at 496 locations were recorded every 15 min during the experiment". However, the authors only use total mass change and total seepage flow in the analysis. They should make clear why the use of soil moisture data is considered inessential to better clarify the hydrological functioning of the hillsope. Otherways, they should use these data to shed light on the comparison between observed behaviours and model results.

R: As mentioned in previous points, examination of the soil moisture data has been included in another paper submitted to HESS-D, and further analyses of LEO experiments data using internal data is ongoing.

Model setup: a figure with the description of the mesh organization should be reported to help the reader to understand how the Homogeneous and the Heterogeneous model simulations were built. This distinction is key to understand the model results; however it is left to three mere lines (P9 L 2-4) where the unclear term Ksat,sf is reported. This variable is never defined in the text. Also, it is difficult to locate the mesh grids where the conductivity was modified.

R: We revised Figure 3 to add the seepage face nodes where Ksat,sf was modified in the heterogeneous case, and *Ksat,sf* is now properly defined (including in the Abstract).

Modelling results: as reported above, one key implication of this work is that during the observed event, the hydraulic properties of part of the hillslope evolved from one value to another. The numerical model cannot reproduce this behavior: it is based on use of static values for the hydraulic properties. The discussion of the results should include consideration of this implication. In the current text, this is done only at P10 L16.

R: We do not of course know whether the heterogeneity developed gradually or suddenly, but very good model results were obtained with the hypothesis of localized heterogeneity. In the new Discussion section we provide the following impetus to further development of coupled Earth system models:
"*In this modeling study we assume that all soil parameter values vary horizontally and are static during the modeling period. Evolution of heterogeneity due to coupled water and sediment transport processes, which may occur in particular under intense rainfall conditions, is beyond the ability of state-of-the-art hydrological models and requires more attention in ongoing efforts to develop coupled Earth system models. Likewise, soil erosion models that consider only surface processes (e.g., Hofer et al., 2012) are also inadequate to this task.*"

Moreover, at P13 L2-4 the authors report: "With the large conductivity of the LEO soil (e.g., Ksat = 1.4×10^{-4} ms^{-1} upslope of the seepage face for the optimal M4_Hetero simulation), the overland flow generation mechanism is saturation-excess". This key statement should be supported by use of soil moisture data.

R: We added the sentence "*This saturation-excess runoff generation process was confirmed by a detailed analysis of the 496 soil moisture sensors (Gevaert et al., 2014)*."

Discussion and conclusion: This section falls short and fail to discuss the implications of this work. Here the authors really need to extend the discussion identifying pathways for future work. Why is the work relevant to the analysis of co-evolution? What moves it beyond the status quo in the analysis of events which are able to modify the constituent soil properties? Why should someone cite this work? I expect more from a HESS paper.

R: We have added a new Discussion section:
**"*Unlike other artificial large-scale hillslopes such as Hydrohill in China (Kendall et al., 2001) and Chicken Creek in Germany (Gerwin et al., 2009; Hofer et al., 2011), LEO was built with homogenous soil and with a focus on evolving heterogeneity from a "time-zero" homogenous condition through co-evolution of the soil-water-biota system over a time scale of years (Hopp et al., 2009; Dontsova et al., 2009). Development of catchment morphology and soil catena driven by hydrological processes through soil erosion and deposition may be one of the major causes that induce heterogeneity and that in turn exert strong feedbacks on hydrological processes (e.g., Beven et al., 1988; Sivapalan, 2005; McDonnell et al., 2007; Troch et al., 2009). At LEO it was not expected that soil heterogeneity would develop in such a short time period during an intense rainfall event that induced significant subsurface saturated flow. This is one of the main reasons that our pre-experiment model predictions failed to produce overland flow.*

A thorough investigation of the fine particles at the seepage face or upslope is not feasible as this would alter the soil structure of LEO-1. The physically-based hydrological model used in this study allowed us to make a probabilistic assessment of the incipient heterogeneity hypothesis while considering uncertainties in soil parameters. Under heterogeneous conditions the model produced better results for seepage flow and total water storage, as well as overland flow that is comparable to estimates from a water budget analysis. It was not our intention to improve the modeling accuracy through parameter calibration but to test the hypothesis of incipient heterogeneity development.

The model we used in this study solves the Richards equation based on Darcy-Buckingham theory, resolving matrix flow and not macropore flow. There are many modeling studies that use percolation theory and other approaches to simulate hydrologic connectivity of macropores to form preferential flow pathways and threshold-like hydrological responses (e.g., Lehmann et al., 2007; Hofer et al., 2011). At this early stage of LEO, with complete absence of organic matter and vegetation roots, we do not anticipate macropore-related processes to be dominant. Macropores might possibly exist around the sensors, although in this case subsurface flow would be enhanced and would very likely have prevented generation of overland flow.

In this modeling study we assume that all soil parameter values vary horizontally and are static during the modeling period. Evolution of heterogeneity due to coupled water and sediment transport processes, which may occur in particular under intense rainfall conditions, is beyond the ability of state-of-the-art hydrological models and requires

***more attention in ongoing efforts to develop coupled Earth system models. Likewise, soil erosion models that consider only surface processes (e.g., Hofer et al., 2012) are also inadequate to this task.*"**

And the paper now concludes with:

"*… This modeling study of the first LEO experiment suggests an important role of coupled water and sediment transport processes in the evolution of subsurface heterogeneity and on overland flow generation."*