



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

G.-Y. Niu^{1,2}, D. Pasetto^{3,4}, C. Scudeler³, C. Paniconi³, M. Putti⁴, P. A. Troch^{1,2,**}, S. B. DeLong^{1,*}, K. Dontsova¹, L. Pangle¹, D. D. Breshears^{5,**}, J. Chorover^{6,**}, T. E. Huxman^{7,**}, J. Pelletier^{8,**}, S. R. Saleska^{9,**}, and X. Zeng^{10,**}

¹Biosphere 2, University of Arizona, Tucson, AZ, USA

²Department of Hydrology and Water Resources, the University of Arizona, Tucson, AZ, USA

³Institut National de la Recherche Scientifique, Centre Eau Terre Environnement (INRS-ETE), Université du Québec, Quebec City, Canada

⁴Department of Mathematics, University of Padova, Padova, Italy

⁵School of Natural Resources and the Environment, University of Arizona, Tucson, AZ, USA

⁶Department of Soil, Water and Environmental Science, University of Arizona, Tucson, AZ, USA

⁷Ecology and Evolutionary Biology, University of California, Irvine, CA, USA

⁸Department of Geosciences University of Arizona, Tucson, AZ, USA

⁹Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA

¹⁰Department of Atmospheric Sciences, University of Arizona, Tucson, AZ, USA

* now at: United State Geological Survey, Menlo Park, CA, USA

**LEO Research and Design Team, Tucson, AZ, USA

Correspondence to: G.-Y. Niu (niu@email.arizona.edu)

Received: 20 September 2013 – Published in Hydrol. Earth Syst. Sci. Discuss.: 18 October 2013

Revised: 7 March 2014 – Accepted: 9 April 2014 – Published: 22 May 2014

Abstract. Evolution of landscape heterogeneity is controlled by coupled Earth system dynamics, and the resulting process complexity is a major hurdle to cross towards a unified theory of catchment hydrology. The Biosphere 2 Landscape Evolution Observatory (LEO), a 334.5 m² artificial hillslope built with homogeneous soil, may have evolved into heterogeneous soil during the first experiment driven by an intense rainfall event. The experiment produced predominantly seepage face water outflow, but also generated overland flow, causing superficial erosion and the 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 by comparing the modeling results from a three-dimensional physically based hydrological model with measurements of total mass change and seepage face flow. Our null hypothesis is that the soil is hydraulically homogeneous, while the alternative hypothesis is that LEO developed downstream heterogeneity from transport of fine sediments

driven by saturated subsurface flow. The heterogeneous case is modeled by assigning saturated hydraulic conductivity at the LEO seepage face ($K_{\text{sat,sf}}$) different from that of the rest (K_{sat}). A range of values for K_{sat} , $K_{\text{sat,sf}}$, soil porosity, and pore size distribution is used to account for uncertainties in estimating these parameters, 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. 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, highlighting the need of a coupled modeling system that integrates across disciplinary processes.

1 Introduction

Landscape heterogeneity is ubiquitous at various spatial scales, it may evolve over time, and it induces process complexity that still needs to be properly addressed in catchment hydrology. As such, predictions of the Earth system response to natural and anthropogenic forcing is 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) yet will allow prediction of the important co-evolution of these processes that will help predict future Earth system states.

To improve predictive understanding of the coupled physical, chemical, biological, and geological processes at Earth’s surface in changing climates, the University of Arizona has constructed a large-scale and community-oriented research infrastructure, the Biosphere 2 Landscape Evolution Observatory (LEO), near Tucson, Arizona, USA. The infrastructure is designed to facilitate investigation of emergent structural heterogeneity that results from coupled Earth surface processes (Hopp et al., 2009). Feedbacks and interactions between different Earth surface processes are studied through iterations of experimental measurement and development of coupled, physically based numerical models (Huxman et al., 2009). The controlled environment of LEO constitutes an ideal platform for validating and improving models and in turn models can help interpret measured data, corroborate and characterize the formation of soil and ecosystem heterogeneity, and design subsequent experiments.

LEO consists of three identical, 30 m long and 11.15 m wide, convergent landscapes. These landscapes are being studied in replicate as “bare soil” for an initial period of 2–3 years. During this time, investigations will focus on hydrological processes, surface modification by rainsplash and overland flow, hillslope-scale water transit times, evolution of moisture state distribution, rates and patterns of geochemical processes, emergent microbial ecology, and carbon and energy cycle dynamics within the shallow subsurface. The scale of environmental control and measurement capabilities, combined with a long-timescale focus for experimentation allow for substantial data–model coupling in a research environment. For example, detailed hydrogeochemical modeling predicted that within 3 years of rainfall initiation, the basalt parent material will develop significant changes in subsurface structure, including pore size and particle size changes that could potentially affect hydrologic flow pathways and system-scale hydrologic features (Dontsova et al., 2009). These physical system changes can be evaluated experimentally, along with the accelerated co-evolution of the

physical and biological systems expected following introduction of plant communities (e.g., Invanov, et al., 2010).

The Biosphere 2 LEO facility has been constructed after a period of community-based scientific planning (Hopp et al., 2009; Dontsova et al., 2009; Ivanov et al., 2010). The first hillslope of LEO (LEO-1) was commissioned at the end of 2012, while the second and third hillslopes were completed in the fall of 2013. From 2014 on, all three hillslopes will be monitored simultaneously while experiencing a climate that held representative transitions between wet and dry conditions, in both warm and cool growing seasons (e.g., a feature of many bioclimatic settings, including the semiarid southwest of the US). Monitoring will include rain amounts and intensity, soil moisture and soil water potential spatiotemporal distributions, perched groundwater dynamics, seepage flow, surface runoff and associated solute and sediment transport out of the hillslope, and total mass storage changes. Geochemical analysis of rain, soil, seepage, and surface runoff water and CO₂ analysis of soil air samples using embedded automatic sensors will complete routine monitoring procedures.

Between LEO-1 commissioning and the completion of the entire LEO (December 2012–December 2013), a series of stand-alone rainfall–runoff experiments with LEO-1 were scheduled. These experiments were designed to reveal internal hydrologic and geochemical dynamics, to test sensor and sampler infrastructure across a wide range of wetness conditions, and to fine-tune data acquisition and processing software and hardware. The amount of water used during these experiments will be applied to the other two hillslopes to provide similar geochemical conditions before the parallel, continuous long-term experiment starts in 2014. Simulations with uncoupled three-dimensional (3-D) hydrologic and solute transport models were run prior to the experiments to obtain preliminary predictions of the hydrologic and water particle response.

The objective of the first experiment, which started at 10:00 LT (local time) on 18 February 2013, was to bring the hillslope to a hydrologic steady state using a continuous and constant rain rate, and to observe how the hillslope outflows and internal states respond to this atmospheric forcing. Numerical simulations prior to the experiment had predicted that the hillslope would reach hydrologic steady state after 24 h (scenario M2 in Sect. 2.4). The rain was scheduled to be turned off to allow the hillslope to drain for a week after reaching steady state, and then another continuous and constant rain event labeled with deuterium was planned. Automatic sampling of rain and seepage water outflow was programmed at every 15 min, while manual sampling from a subset of the soil suction lysimeter array was attempted every 3 h. In the actual experiment, the hillslope never reached the predicted steady state but instead developed saturation excess overland flow, which transported 0.7 m³ of soil and generated a shallow gully in the central trough of the hillslope.

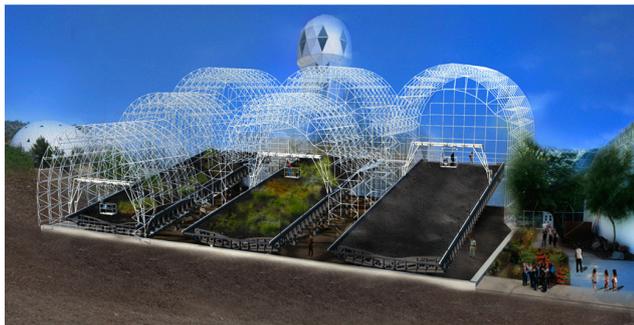


Figure 1. Diagram showing the three identical convergent landscapes (30 m long and 11.15 m wide) of the Biosphere 2 LEO constructed inside an environmentally controlled greenhouse facility.

LEO was built with homogeneous loamy sand (Sect. 2.1), carefully compacted at layering intervals of 25 cm thickness. The preexperiment numerical simulations thus assumed the LEO soil to be hydraulically homogeneous (M1 and M2 scenarios in Sect. 2.4), and predicted no overland flow. It is possible that heterogeneity in LEO developed during the intense rainfall event due to transport of fine sediments driven by subsurface saturated flow in the downstream direction, as evidenced in a lab experiment with the same soil (Hernandez and Schaap, 2012), and that this process played an important role in overland flow generation. We conducted a partial investigation of the fine soil particles to a depth of 72 cm at the seepage face shortly after the experiment and observed an accumulation of fines at the seepage face, but we were unable to test its effect on the hydraulic conductivity. A thorough probing of soil particles at the seepage face or upstream is infeasible as this would significantly disturb the soil structure and long-term experimental integrity.

In this work we pursue a modeling study to analyze the hydrologic response from the first LEO experiment and to answer the question: *why did the observed hydrological response differ so significantly from the preexperiment predicted response?* The analysis is based on simulation results using a 3-D physically based hydrological model. The investigation focuses on how overland flow was generated and on the important role of localized incipient heterogeneity in overland flow generation, a key general phenomena driving the design of the LEO apparatus. Heterogeneity is represented in the model through spatially varying saturated hydraulic conductivity (K_{sat}). We also consider uncertainty in K_{sat} and other soil parameters, namely soil porosity and the pore size distribution parameter (n) in the water retention characteristics. The modeling results are compared to detailed measurements of total mass change and seepage face flow collected during the experiment. Analysis of the modeling errors (or accuracies) over a wide parameter space with respect to homogeneous and heterogeneous soils allows us to make a probability assessment of the incipient heterogeneity hypothesis.

2 Methodology

2.1 Biosphere 2 LEO

LEO consists of three identical, sloping (10° on average), 334.5 m^2 convergent landscapes inside a 5000 m^2 environmentally controlled facility (Fig. 1). These engineered landscapes have a dry weight of $\sim 600\,000 \text{ kg}$ each and contain a 1 m depth of basaltic tephra ground to homogeneous loamy sand that is expected to evolve into structured soil over time. Each landscape was designed with a seepage face at its lower end to facilitate downslope flow; the seepage face consists of a 0.5 m wide gravel section held in place by a plastic plate perforated with 2 mm holes. Seepage face flow was recorded through six tipping buckets and six flow meters installed at six sections of the seepage face. Each landscape contains a spatially dense sensor and sampler network capable of resolving meter-scale lateral heterogeneity and submeter-scale vertical heterogeneity in moisture, energy, and carbon states and fluxes. The density of sensors and frequency at which they can be polled allows for measurements that are impossible to take in natural field settings. Embedded soil water solution and soil gas samplers allow for quantification of biogeochemical processes and facilitate the use of chemical tracers to study water movement at very dense spatial scales. The landscapes have load cells embedded into their structure to measure changes in total system mass weight with 0.05 % full-scale repeatability (equivalent to less than 1 cm of precipitation). Each landscape has an engineered rain system that allows application of precipitation at rates between 3 and 45 mm h^{-1} in spatially homogeneous or heterogeneous patterns and with enough capacity to produce hillslope-scale hydrological steady-state conditions or to run complex hyetograph simulations. The precipitation water supply storage system is flexibly designed to facilitate addition of tracers in constant or time-varying rates to any of the three hillslopes.

2.2 The hydrological model

We use the CATHY (CATchment HYdrology) model (Camporese et al., 2010) to simulate the partitioning of rainfall between runoff and infiltration, the subsurface redistribution of soil moisture and groundwater, and the discharge through the LEO seepage face. The subsurface flow module in CATHY solves the 3-D Richards equation describing flow in variably saturated porous media (Paniconi and Putti, 1994), while the surface flow module solves the diffusion wave equation describing surface flow propagation over hillslopes and in stream channels identified using terrain topography and the hydraulic geometry concept (Orlandini and Rosso, 1998). Surface–subsurface coupling is based on a boundary condition switching procedure that automatically partitions potential fluxes (rainfall and evapotranspiration) into actual fluxes across the land surface and calculates changes in surface storage.

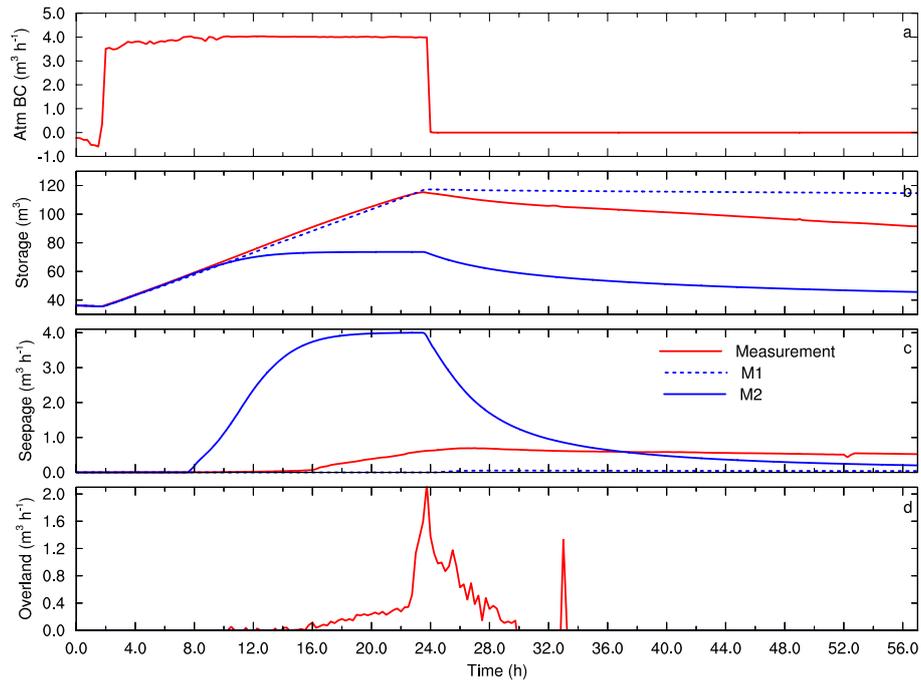


Figure 2. Comparison between the measured data and the modeling results from scenarios M1 and M2. From the upper panel to the lower panel are the atmospheric boundary conditions ($\text{m}^3 \text{h}^{-1}$), total water storage (m^3), seepage face flow ($\text{m}^3 \text{h}^{-1}$), and overland flow ($\text{m}^3 \text{h}^{-1}$).

2.3 The first LEO experiment

The first hydrological experiment on LEO-1 started at 10:00 LT, 18 February 2013 and ended at 08:00 LT, 19 February 2013. With the hillslope prewetted to an initial water storage of 108 mm (36.13 m^3 in Fig. 2b), rainfall over the entire hillslope at $\sim 12 \text{ mm h}^{-1}$ ($4.01 \text{ m}^3 \text{ h}^{-1}$ in Fig. 2a) for a duration of 22 h produced an input of $\sim 264 \text{ mm}$ into the 1.0 m deep soil of LEO. This experiment was designed to (1) test the functionality of all sensors, (2) investigate LEO's hydrological response under a heavy rainfall, and (3) generate a steady state of soil moisture for further tracer experiments. Prior to the experiment and based on laboratory and other analyses to assign parameter values, we used CATHY to estimate the time for LEO to reach an equilibrium state under a constant precipitation rate. In this simulation the seepage face outflow equaled the imposed precipitation rate after 1.5 d and no overland flow was predicted (see Sect. 2.4 for model configuration M2 and Sect. 3 for the results). The actual response of LEO to the imposed precipitation differed drastically from the preexperiment analysis. Overland flow occurred 15 h after the start of rainfall, resulting in erosion of the superficial soil layers and the formation of a surface channel. 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.

Total mass change, total seepage flow, and soil moisture at 496 locations were recorded every 15 min during the experiment. An estimation of the overland flow and soil evaporation rates was achieved from the closure of the water balance and from volumetric flow measurements. Figure 2 shows the hydrological data collected during the experiment. Time “0” corresponds to 08:00 LT 18 February (i.e., 2 h before the start of rainfall). Overland flow (Fig. 2d) reached a peak of about $1.8 \text{ m}^3 \text{ h}^{-1}$ around 08:00 LT 19 February when the rain system was turned off. The maximum seepage face flow occurred about 1 h later, with a magnitude of about $0.7 \text{ m}^3 \text{ h}^{-1}$.

2.4 Model setup

We discretized the $30 \text{ m} \times 11.15 \text{ m} \times 1 \text{ m}$ LEO soil into a grid of 60 cells \times 24 cells (61 nodes \times 25 nodes) in the lateral direction and 8 layers (9 nodes) in the vertical direction (Fig. 3), assigning a higher resolution (0.05 m) to the surface and bottom layers to better resolve infiltration at the soil surface and seepage flow at the bottom nodes of the seepage face. We set up a seepage face boundary condition at the 25×8 nodes of the downslope boundary (red dots in Fig. 3). The 25 nodes along the top edge of this lateral boundary were excluded; these nodes, together with all other nodes on the LEO surface, were assigned atmospheric boundary

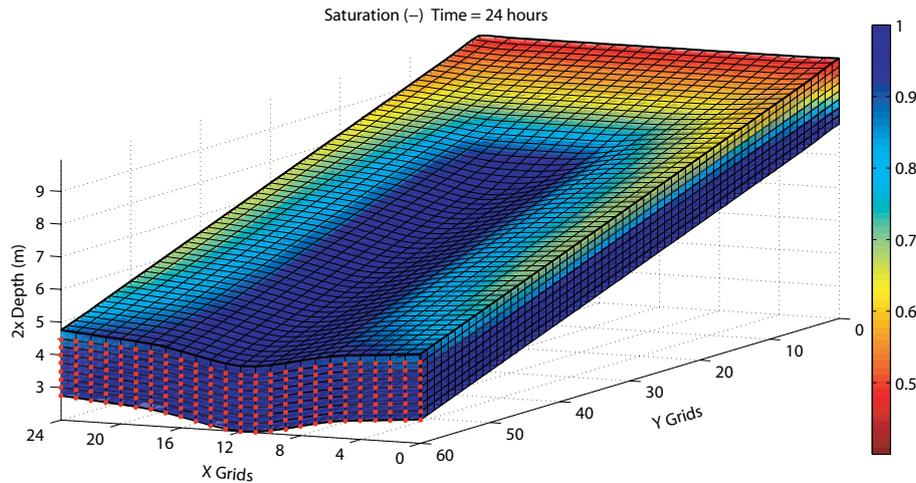


Figure 3. Discretization of the LEO soil with a grid of $60 \times 24 \times 8$ cells and $61 \times 25 \times 9$ nodes (the vertical depth of soil is exaggerated by a factor of 2). The red dots at the lower end of LEO represent the 25×8 nodes seepage face. Color indicates the modeled degree of saturation at time 24 h of the best realization ($n = 2.26$ in Fig. 9).

conditions. Aside from the seepage face and the land surface, all other LEO boundaries were set to a zero flux condition.

Because of the lack of a direct measurement of soil surface evaporation (E), the atmospheric boundary condition (Q_{atm}) of the model was estimated separately for three phases. During the daytime period from 08:00 to 20:00 LT of 18 February (time 0–12 h in Fig. 2a), E is not negligible. Q_{atm} was therefore estimated as the rate of change in total water storage (dS/dt) as measured by the load cell because the mass balance can be expressed as $dS/dt = P - E$, where P is rainfall, prior to the occurrence of major seepage face and overland flow. During the nighttime until the next morning when the rainfall was stopped (time 12–24 h in Fig. 2a), E was assumed to be negligible, and Q_{atm} was thus set to the sprinkler rainfall rate (12 mm h^{-1}). During the final phase after time 24 h with no rain, Q_{atm} was estimated at -2 mm d^{-1} , where 2 mm d^{-1} is the average evaporation rate from a wet surface for a winter month in Arizona.

Time stepping in the CATHY model is adaptive (based on the convergence of the iterative scheme used to linearize Richards' equation) and was set such that time step sizes ranged from 0.1 to 180 s. The convergence criterion on soil water pressure was set for a model accuracy of $1.0 \times 10^{-3} \text{ m}$.

We designed six scenarios of numerical simulations taking into account different configurations of model parameters characterizing the soil properties, including the van Genuchten curve fitting parameter (n), the porosity (θ_{sat}), and the saturated hydraulic conductivity (K_{sat}). The scenarios and corresponding model parameter values are summarized in Table 1.

Scenarios M1 and M2 assume that the soil is homogeneous and correspond to the numerical simulations performed before the physical experiment. M1 uses soil property parameters from an analysis of soil particle size distribution

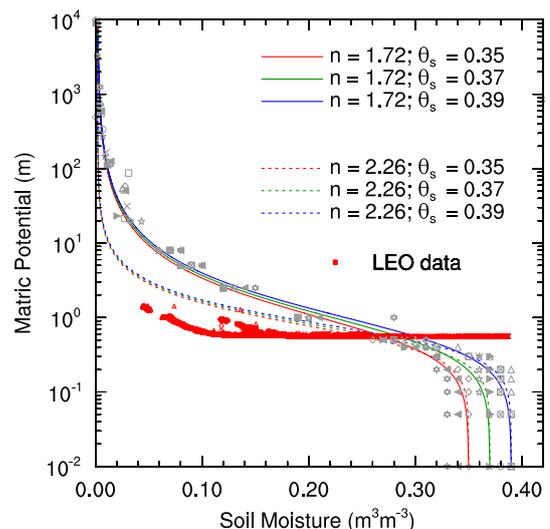


Figure 4. The relationship between soil moisture and matric potential from laboratory experiments (the grey markers represent different sampling depths) and from the van Genuchten fitting curves for different porosities. The solid curves attempt to match the laboratory data with $n = 1.72$ while the dashed curves are from a particle size distribution analysis and match better the in situ LEO data (red symbols) with $n = 2.26$.

($n = 2.26$, $\theta_{\text{sat}} = 0.39$, and $K_{\text{sat}} = 7.8 \times 10^{-6} \text{ m s}^{-1}$). M2 uses the same parameters except a greater K_{sat} ($3.8 \times 10^{-3} \text{ m s}^{-1}$) 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.

To generate overland flow it was found that the numerical model of LEO requires a lower soil porosity than the one used in M1 and M2 and/or a heterogeneous distribution

Table 1. Model scenarios and associated parameter values.

	M1	M2	M3_Homo	M3_Hetero	M4_Homo	M4_Hetero
van Genuchten n (–)	2.26	2.26	1.72	1.72	2.26	2.26
Saturated matric potential ψ_{sat} (m)	–0.48	–0.48	–0.6	–0.6	–0.6	–0.6
Residual moisture θ_r ($\text{m}^3 \text{m}^{-3}$)	0.035	0.035	0.002	0.002	0.002	0.002
Specific storage S_s (–)	5.0×10^{-4}	5.0×10^{-4}	5.0×10^{-4}	5.0×10^{-4}	5.0×10^{-4}	5.0×10^{-4}
Porosity θ_{sat} ($\text{m}^3 \text{m}^{-3}$)	0.39	0.39		21 values from 0.33 \rightarrow 0.38		
Saturated hydraulic conductivity K_{sat} (10^{-5}m s^{-1})	0.78	380		30 values from 1 \rightarrow 30		
K_{sat} at the seepage face $K_{\text{sat,sf}}$ (10^{-5}m s^{-1})	0.78	380	K_{sat}	18 values 1.4 \rightarrow 3.1	K_{sat}	18 values 1.4 \rightarrow 3.1
Total number of simulations	1	1	21 \times 30 = 630	21 \times 30 \times 18 = 11 340	21 \times 30 = 630	21 \times 30 \times 18 = 11 340

of the hydraulic conductivity that slows down the seepage face outflow. Scenarios M3 and M4 are designed to assess the probability that K_{sat} at LEO's seepage face, $K_{\text{sat,sf}}$, may be lower than that of the rest of the LEO soil. M3 consists of two groups of experiments, one under the hypothesis of homogeneous soil (M3_Homo) and the other assuming that $K_{\text{sat,sf}}$ is generally less than the K_{sat} of the rest of the LEO soil (M3_Hetero). The values of the van Genuchten parameters n , ψ_{sat} , and θ_r used in scenario M3 were obtained by fitting the soil water retention data from laboratory experiments on the LEO soil samples (Fig. 4). In particular, for scenario M3 the value of n is 1.72. M3_Homo simulations were conducted with combinations of 21 values of θ_{sat} ranging from 0.33 to 0.38 at a step of 0.0025 and 30 values of K_{sat} ranging from 1 to $30 \times 10^{-5} \text{m s}^{-1}$ at a step of $1 \times 10^{-5} \text{m s}^{-1}$, for a total of 630 simulations. M3_Hetero further combines 18 values of $K_{\text{sat,sf}}$ ranging from 1.4×10^{-5} to $3.1 \times 10^{-5} \text{m s}^{-1}$ at a step of $1 \times 10^{-6} \text{m s}^{-1}$, for a total of $630 \times 18 = 11\,340$ simulations. This restricted range of $K_{\text{sat,sf}}$ was determined through sensitivity tests using a wider range of parameter values that are not reported here. Scenario M4 is analogous to scenario M3 except that $n = 2.26$, the same as in M1 and M2. This larger n value, estimated from a preexperiment analysis of particle size distribution, tends to better match the in situ LEO data (Fig. 4). This higher value may be justified for the large volume of LEO (334.5m^3) compared to the small volume of the cores in the laboratory. In situ measurements of volumetric water content (with 5TM Decagon probes) and pore water pressure (with MPS-2 Decagon probes) indicate that the higher n values are not unrealistic for the LEO soil (Fig. 4). However, there is significant uncertainty in these measurements due to sensor inaccuracy. The pore water pressure sensors became saturated at levels above -6kPa , making them ineffective for wet conditions. For this reason, the wetter part ($> 0.2 \text{m}^3 \text{m}^{-3}$) of the retention characteristics derived from the in situ data (red dots in Fig. 4) may not be very reliable.

To evaluate what set of parameter values allows us to best approximate the observed response amongst these several thousand model simulations, we computed the mean relative

error between the measured and simulated data. For instance, let $\Delta S_{\text{m}}(t)$ and $\Delta S_{\text{s}}(t)$ be the measured and simulated variation of water storage at time t . We define the relative error $e_{\Delta S}$ as

$$e_{\Delta S} = \frac{\int_0^T |\Delta S_{\text{m}} - \Delta S_{\text{s}}| dt}{\int_0^T \Delta S_{\text{m}} dt}. \quad (1)$$

The relative error for the seepage face flow (e_{Q_s}) is computed in the same way. The mean relative error is then defined as an average of the two:

$$e = \frac{1}{2} (e_{\Delta S} + e_{Q_s}). \quad (2)$$

We did not include the relative error of overland flow in the averaged error above because the observed response for this variable was derived from mass balance calculations based on other measured variables. Its derivation also involves estimation of surface evaporation at later stages. Total water storage, however, was measured directly by means of 10 load cells, and seepage flow was measured by means of tipping bucket rain gauges and electromagnetic flow meters.

3 Modeling results

Figure 2 compares the modeling results from M1 and M2 to the measured overland flow, seepage face flow, and water storage. Neither M1 nor M2 produce any overland flow. Compared to the measured seepage face flow, M1 with a smaller K_{sat} ($7.8 \times 10^{-6} \text{m s}^{-1}$) produces negligible outflow at the seepage face, and therefore the modeled water storage stays at a constant value after it reaches its peak value. M2 however, with a much higher K_{sat} ($3.8 \times 10^{-3} \text{m s}^{-1}$), produces much higher outflow at the seepage face and lower water storage than the measured values. The M2 results indicate that the calibration of K_{sat} against the starting time of seepage face flow from the LEO-1 test run (see Sect. 2.4) is

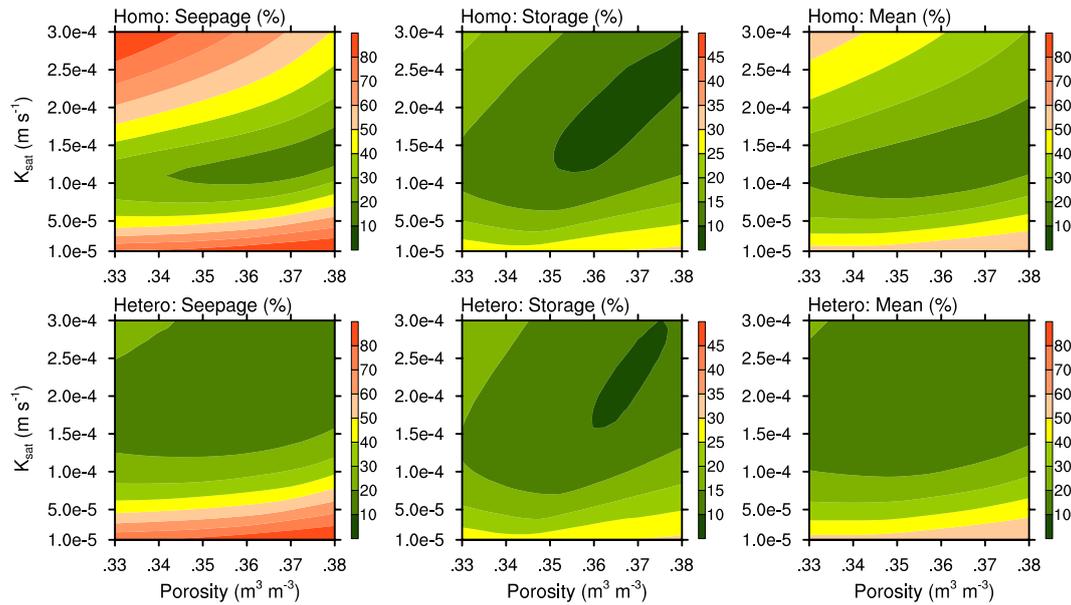


Figure 5. Relative model error (e) of seepage flow and water storage and the mean error for M3_Homo (upper panels) and M3_Hetero (lower panels); with $K_{\text{sat},\text{sf}} = 2.1 \times 10^{-5} \text{ m s}^{-1}$) over the parameter space of K_{sat} and porosity.

misleading, because the LEO soil at this early stage may not have been well compacted, resulting in faster outflows at the seepage face. M1 and M2 produce seepage face flow and water storage that are very different from the measurements, and at opposite extremes. Since the modeled overland flow is zero for both cases, changes in K_{sat} are insufficient to retrieve the observed overland flow. We therefore conducted several sensitivity simulations to reduce θ_{sat} and/or $K_{\text{sat},\text{sf}}$. These simulations helped produce overland flow and improved the simulation of seepage face flow and water storage, informing the design of the M3 and M4 experiments summarized in Figs. 5 and 6.

For scenario M3, Fig. 5 shows the relative model error across the parameter space of K_{sat} and θ_{s} for both the M3_Homo and M3_Hetero experiments. The results for M3_Hetero shown in this figure are obtained with $K_{\text{sat},\text{sf}} = 2.1 \times 10^{-5} \text{ m s}^{-1}$. M3_Hetero shows a relatively greater area of best simulations of seepage face flow (i.e., with relative errors that are smaller than 20 %) compared to M3_homo, for which the best results are concentrated along a narrow band around a K_{sat} value of $1.1 \times 10^{-4} \text{ m s}^{-1}$. This suggests that M3_Hetero has a greater number of best simulations than M3_Homo. However, M3_Hetero shows a smaller area of best simulations of water storage with relative errors smaller than 10 % than does M3_Homo. In terms of the mean relative error combining the two response variables, M3_Hetero yields a larger number or greater probability of best simulations than M3_Homo. To clarify this point, in Fig. 7 we show a frequency analysis of the mean relative errors obtained in M3_Homo and M3_Hetero. The frequencies are normalized by the total number of simulations (630),

so that the histograms are an approximation of the probability density functions (PDFs) of the mean errors. Taking a relative error smaller than 15 % as a marker, M3_Hetero has more than 40 % best simulations compared with only 6 % for M3_Homo.

Similar results are obtained for scenario M4, where a value of 2.26 instead of 1.72 was set for parameter n . Figure 6 shows the comparison of M4_Homo and M4_Hetero simulations in terms of the relative errors across the parameter space of K_{sat} and θ_{sat} . The results for M4_Hetero shown in this figure are obtained with $K_{\text{sat},\text{sf}} = 1.9 \times 10^{-5} \text{ m s}^{-1}$. M4_Hetero shows a larger area (or greater number) of best simulations than M4_Homo, more notably for seepage face flow. In terms of the mean relative error, M4_Hetero yields a greater area (or probability) of best simulations than M4_Homo. This is confirmed from the PDFs in Fig. 7, where M4_Hetero has about 16 % best simulations (taking relative error smaller than 10 % as a marker) while M4_Homo has only about 2 %. This implies that the assumption of $K_{\text{sat},\text{sf}} < K_{\text{sat}}$ produces a greater probability of best realizations than that of $K_{\text{sat},\text{sf}} = K_{\text{sat}}$, supporting the hypothesis of localized heterogeneity at the LEO hillslope.

Figure 7 also suggests that the overall performance of M4_Hetero is better than M3_Hetero. M4_Hetero produces 16 % best simulations with a mean relative error smaller than 10 % whereas M3_Hetero produces none, while at the 15 % relative error level M4_Hetero yields a 50 % probability of best realizations compared to 42 % for M3_Hetero. In addition, the best simulation of M4_Hetero produces a smaller error (7.38 %) than that of M3_Hetero (10.74 %) (Table 2).

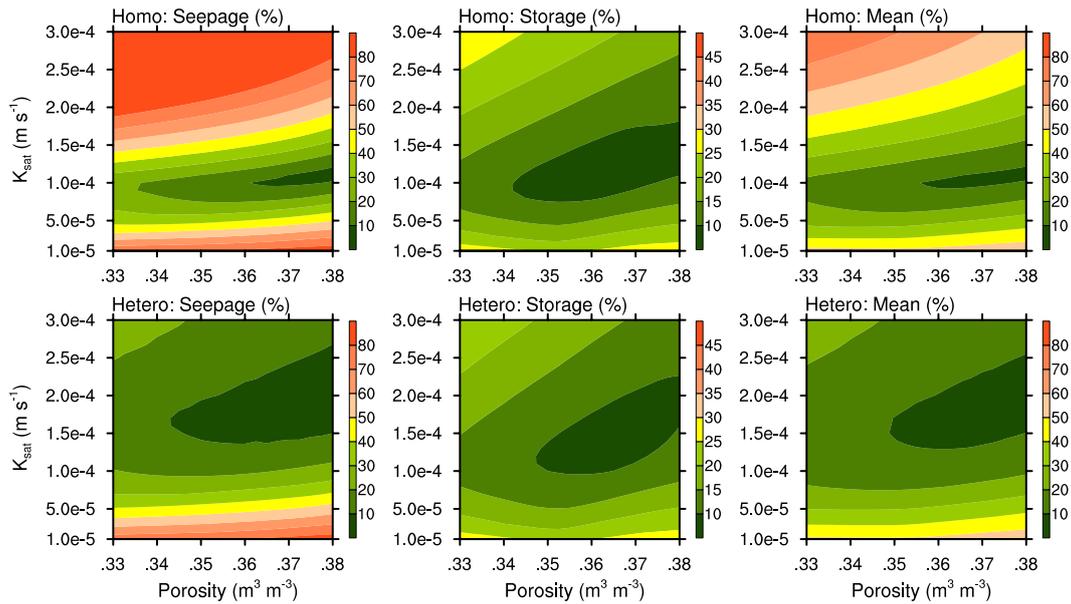


Figure 6. Relative model error (e) of seepage flow and water storage and the mean error for M4_Homo (upper panels) and M4_Hetero (lower panels; with $K_{\text{sat,sf}} = 1.9 \times 10^{-5} \text{ m s}^{-1}$) over the parameter space of K_{sat} and porosity.

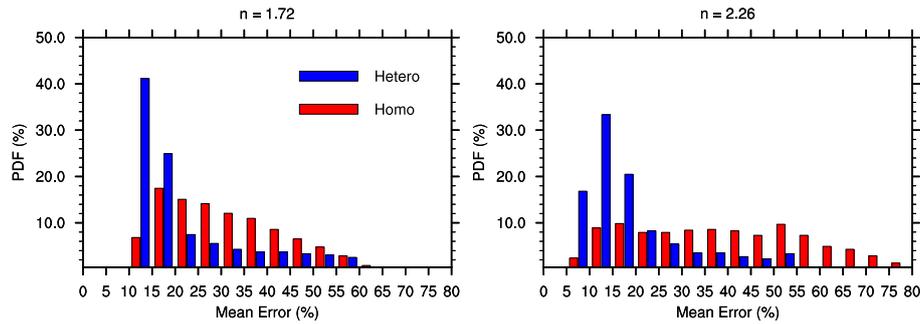


Figure 7. Probability density functions of the mean error for M3 ($n = 1.72$) and M4 ($n = 2.26$). The results for M3_Hetero and M4_Hetero are obtained with the optimized $K_{\text{sat,sf}}$ values used in Figs. 5 and 6, i.e., $2.1 \times 10^{-5} \text{ m s}^{-1}$ and $1.9 \times 10^{-5} \text{ m s}^{-1}$, respectively.

A further comparison between scenarios M3_Hetero and M4_Hetero is depicted in Fig. 8, which shows the PDFs of these two experiments across all 18 $K_{\text{sat,sf}}$ values and for 3 different values of mean relative error level (10, 15, and 20 %). When $K_{\text{sat,sf}} = 2.1 \times 10^{-5} \text{ m s}^{-1}$, M3_Hetero reaches the greatest probability of best simulations with mean a relative error of less than 15 %; and when $K_{\text{sat,sf}} = 1.9 \times 10^{-5} \text{ m s}^{-1}$, M4_Hetero reaches the greatest probability with a relative error of less than 10 %. M4_Hetero ($n = 2.26$) performs notably better than M3_Hetero ($n = 1.72$) over almost all the $K_{\text{sat,sf}}$ values (particularly at the 10 % error level).

The optimized $K_{\text{sat,sf}}$ values, corresponding to the best realizations out of the 11 340 simulations each of M3_Hetero and M4_Hetero, are, respectively, 2.3×10^{-5} and $2.2 \times 10^{-5} \text{ m s}^{-1}$ (Table 2) (slightly larger than those corresponding to their greatest probabilities). These

optimized values of $K_{\text{sat,sf}}$ coincidentally fall within the range of K_{sat} values obtained from the laboratory measurements (1.9×10^{-5} – $2.5 \times 10^{-5} \text{ m s}^{-1}$) with the same soil (Hernandez and Schaap, 2012). The optimized K_{sat} values for the upslope portion of the hillslope are about 6.4 times greater than $K_{\text{sat,sf}}$ for M4_Hetero and 7.4 times greater for M3_Hetero. These modeling results thus once again support the hypothesis of localized heterogeneity at the lower end of LEO.

The modeled time series of seepage flow (Fig. 9b) from the best simulations of M3_Hetero and M4_Hetero explains why the best results favor the higher n value. A higher n produces a faster early response of outflow at the seepage face and more sustainable flow during the recession period. The optimized n value (2.26) is also consistent with the larger optimized K_{sat} value ($1.4 \times 10^{-4} \text{ m s}^{-1}$), both suggesting a

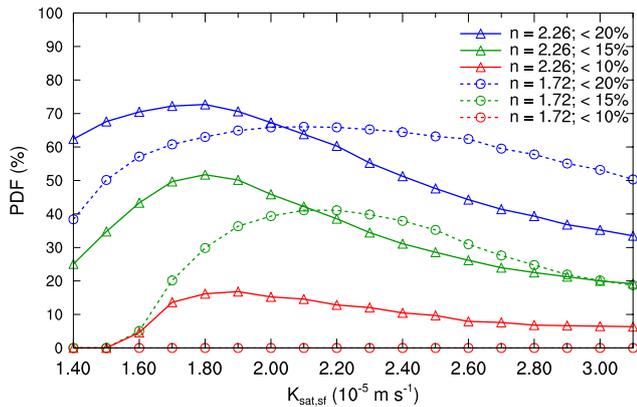


Figure 8. Probability density functions of best simulations for M3_Hetero ($n = 1.72$) and M4_Hetero ($n = 2.26$) simulations at various error levels across the 18 $K_{\text{sat,sf}}$ values considered.

greater permeability of the LEO soil than that of the same soil in the laboratory (and at the seepage face).

As a result of calibration against seepage face flow and water storage, the best realizations for both M3_Hetero and M4_Hetero also produce a reasonable overland flow hydrograph, in phase with the hydrograph estimated from mass balance calculations though with a longer tail during the recession period (Fig. 9c). The modeled longer tail of overland flow may be induced by the uncertainty in the soil surface evaporation estimate (2 mm d^{-1}) used as the upper boundary condition during this period. With the large conductivity of the LEO soil (e.g., $K_{\text{sat}} = 1.4 \times 10^{-4} \text{ m s}^{-1}$ upslope of the seepage face for the optimal M4_Hetero simulation), the overland flow generation mechanism is saturation excess (Gevaert et al., 2014), and therefore calibration of θ_{sat} and $K_{\text{sat,sf}}$ is critical for accurately reproducing this response. Figure 3 shows the degree of saturation of LEO when overland flow reaches its peak value. The water table first builds up at the lower end of LEO and then propagates upslope, with overland flow being triggered when the water table reaches the surface. This saturation-excess runoff generation process was confirmed by a detailed analysis of the 496 soil moisture sensors (Gevaert et al., 2014).

4 Discussion

A grand challenge in science is understanding the coupled evolution of Earth system processes (Anderson et al., 2008). Experimental facilities that can tackle the evolution of structure and function in physical and biological systems, along with their emergent processes at scale, will be extremely useful for understanding future Earth system states and the significant deviation from stationarity seen in our current climate system. Unlike other artificial laboratories such as the Hydrohill in China (Kendall et al., 2001) and the Chicken Creek in Germany (Gerwin et al., 2009; Hofer et al., 2011),

Table 2. Optimized parameter values for K_{sat} , $K_{\text{sat,sf}}$, and θ_{sat} and mean relative errors (e ; %).

	M3_Homo	M3_Hetero	M4_Homo	M4_Hetero
n (–)	1.72	1.72	2.26	2.26
θ_{sat} ($\text{m}^3 \text{ m}^{-3}$)	0.3625	0.3625	0.370	0.3675
K_{sat} (m s^{-1})	1.2×10^{-4}	1.7×10^{-4}	1.0×10^{-4}	1.4×10^{-4}
$K_{\text{sat,sf}}$ (m s^{-1})	1.2×10^{-4}	2.3×10^{-5}	1.0×10^{-4}	2.2×10^{-5}
e (%)	12.99	10.74	8.40	7.38

LEO was built with homogeneous soil and with a focus on evolving heterogeneity from a “time-zero” homogeneous condition through co-evolution of the soil–water–biota system over a timescale 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 unanticipated 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 preexperiment 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 probability assessment of the incipient heterogeneity hypothesis while considering also 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 believe macropore-related processes are 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 (not vertically) and are static during the modeling period. Evolution of heterogeneity due to coupled water and sediment transport processes, which may

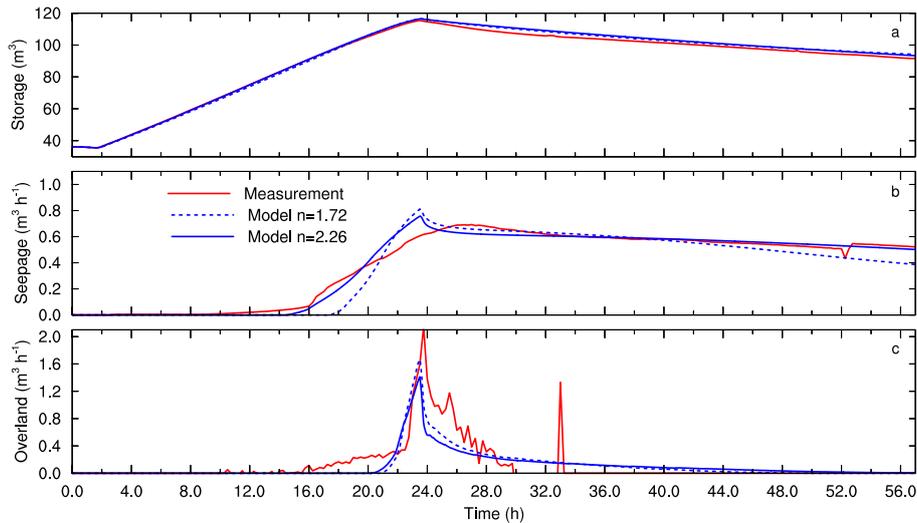


Figure 9. Comparison between the measured total water storage (a), seepage face flow (b), and overland flow (c) and the simulated results obtained with the optimized parameter values for M3_Hetero ($n = 1.72$) and M4_Hetero ($n = 2.26$).

occur particularly 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 apparently inadequate to this task. What is presented by experimental examples of unique events in LEO is the opportunity to develop new coupled models with sufficient data-rich descriptions to push our learning forward. Clearly this is a goal of subsequent efforts of our research group.

5 Conclusions

LEO was constructed anticipating challenging contemporary models with measurable, coupled Earth system dynamics. Ironically, the first rainfall experiment on LEO-1 was designed to test the functionality of subsurface sensors and to generate a hydrologic steady state for system dynamics characterization and further tracer experiments, yet provided a first example of generalized physical system evolution research. The design of this experiment in terms of rainfall intensity and duration was informed by hydrologic model simulations based on estimates of soil hydraulic properties. These preexperiment model simulations predicted that the hillslope would reach steady state in a reasonable amount of time (about 24 h) and that no overland flow through saturation excess would occur. The actual experiment resulted in saturated soils in the central trough of the hillslope that caused saturation excess overland flow and gully erosion. This study has explored possible reasons for the mismatch between model prediction and observations by performing numerous post-experiment model simulations over a much wider parameter space that allows probability assessment of

alternate hypotheses and consideration of parameter uncertainty, informed by a data-rich setting that exceeds the capacity of existing experimental and field settings.

Model simulations under homogeneous soil conditions, using soil parameters estimated from an analysis of particle size distribution (e.g., porosity $\theta_{\text{sat}} = 0.39 \text{ m}^3 \text{ m}^{-3}$) and a range of K_{sat} values, did not produce any overland flow. When θ_{sat} or the value of K_{sat} at the seepage face ($K_{\text{sat,sf}}$) were reduced, it was possible to produce overland flow, and this result informed the design of sensitivity experiments to test two hypotheses: that the soil is homogeneous, and that the soil has developed some heterogeneity in the downstream direction due to saturated soil compaction near the seepage face. We then performed over 20 000 simulations to assess these hypotheses, while considering the uncertainties in $K_{\text{sat,sf}}$, K_{sat} in the upslope soil, and θ_{sat} . We also considered two values of the pore size distribution parameter (n), obtained from a particle size distribution analysis ($n = 2.26$) and by laboratory fitting of the van Genuchten relationship for the LEO soil ($n = 1.72$). The optimized values for n (2.26) and for the upslope K_{sat} ($1.4 \times 10^{-4} \text{ m s}^{-1}$) are higher than the values measured in the laboratory ($n = 1.72$ and $K_{\text{sat}} \sim 1.9\text{--}2.5 \times 10^{-5} \text{ m s}^{-1}$). For both n values, we obtained that (1) simulations with $K_{\text{sat,sf}} < K_{\text{sat}}$ (incipient heterogeneity hypothesis) produced a higher probability of best realizations than those with $K_{\text{sat,sf}} = K_{\text{sat}}$ (homogeneity hypothesis) and (2) the best realizations with the heterogeneous soil yielded smaller errors than those with the homogeneous soil. The modeling results thus support the hypothesis of localized heterogeneity due to downslope compaction of the LEO soil. A possible mechanism for the compaction may be fine sediments transported during subsurface saturated flow prior to the onset of overland flow. 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. Additionally, these results highlight the need to consider the boundary processes that couple our disciplinary modeling frameworks and assumptions of space- and timescales that affect processes within a coupled system. We anticipate robust opportunities for similar model challenges in a number of disciplines over the next several years.

Acknowledgements. We thank the the Phileology Foundation for financial support and Michael Sibayan, John Adams, Nate Abramson for technical support. We also thank the many participants, from different institutions across the USA, in the design and planning workshops that led to the implementation of LEO at Biosphere 2. This study was also partly supported by projects of DOE (DE-SC0006773) and NSF (EF 1065790). We also thank the three anonymous reviewers for their constructive comments, which greatly helped improve the quality of the paper.

Edited by: N. Romano

References

- Anderson, S. P., Bales, R. C., and Duffy, C. J.: Critical Zone Observatories: Building a network to advance interdisciplinary study of Earth surface processes, *Mineralog. Mag.*, 72, 7–10, 2008.
- Beven, K. J., Wood, E. F., and Sivapalan, M.: On hydrological heterogeneity – catchment morphology and catchment response, *J. Hydrol.*, 100, 353–375, 1988.
- 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, W02512, doi:10.1029/2008WR007536, 2010.
- Dontsova, K., Steefel, C. I., Desilets, S., Thompson, A., and Chorover, J.: Solid phase evolution in the Biosphere 2 hillslope experiment as predicted by modeling of hydrologic and geochemical fluxes, *Hydrol. Earth Syst. Sci.*, 13, 2273–2286, doi:10.5194/hess-13-2273-2009, 2009.
- Gerwin, W., Raab, T., Biemelt, D., Bens, O., and Hüttl, R. F.: The artificial water catchment “Chicken Creek” as an observatory for critical zone processes and structures, *Hydrol. Earth Syst. Sci. Discuss.*, 6, 1769–1795, doi:10.5194/hessd-6-1769-2009, 2009.
- Gevaert, A. I., Teuling, A. J., Uijlenhoet, R., and Troch, P. A.: Hillslope experiment demonstrates role of convergence during two-step saturation, *Hydrol. Earth Syst. Sci. Discuss.*, 11, 2211–2232, doi:10.5194/hessd-11-2211-2014, 2014.
- Hernandez, C. M. and Schaap, M. G.: Hydrology simulations on basalt soil for the Landscape Evolution Observatory (LEO), Abstract No. H531-1644, AGU Fall Meeting, 3–7 December 2012, San Francisco, 2012.
- Hofer, M., Lehmann, P., Biemelt, D., Stähli, M., and Krafczyk, M.: Modelling subsurface drainage pathways in an artificial catchment, *Phys. Chem. Earth A/B/C*, 36, 101–112, doi:10.1016/j.pce.2010.04.020, 2011.
- Hofer, M., Lehmann, P., Stahli, 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, W01531, doi:10.1029/2011wr010719, 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 co-evolution at Biosphere 2, *Hydrol. Earth Syst. Sci.*, 13, 2105–2118, doi:10.5194/hess-13-2105-2009, 2009.
- 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*, 90, p. 120, 2009.
- Ivanov, V. Y., Fatichi, S., Jenerette, G. D., Espeleta, J. F., Troch, P. A., and Huxman, T. E.: Hysteresis of soil moisture spatial heterogeneity and the “homogenizing” effect of vegetation, *Water Resour. Res.*, 46, W09521, doi:10.1029/2009WR008611, 2010.
- Kendall, C., McDonnell, J. J., and Gu, W. Z.: A look inside ‘black box’ hydrograph separation models: a study at the Hydrohill catchment, *Hydrol. Process.*, 15, 1877–1902, 2001.
- Lehmann, P., Hinz, C., McGrath, G., Tromp-van Meerveld, H. J., and McDonnell, J. J.: Rainfall threshold for hillslope outflow: an emergent property of flow pathway connectivity, *Hydrol. Earth Syst. Sci.*, 11, 1047–1063, doi:10.5194/hess-11-1047-2007, 2007.
- McDonnell, J. J., Sivapalan, M., Vache, K., Dunn, S., Grant, G., Haggerty, R., Hinz, C., Hooper, R., Kirchner, J., Roderick, M. L., Selker, J., and Weiler, M.: Moving beyond Water Resour. Res., 43, W07301, doi:10.1029/2006WR005467, 2007.
- Orlandini, S. and Rosso, R.: Parameterization of stream channel geometry in the distributed modeling of catchment dynamics, *Water Resour. Res.*, 34, 1971–1985, 1988.
- Paniconi, C. and Putti, M.: A comparison of Picard and Newton iteration in the numerical solution of multidimensional variably saturated flow problems, *Water Resour. Res.*, 30, 3357–3374, 1994.
- Sivapalan, M.: Pattern, process and function: Elements of a unified theory of hydrology at the catchment scale, in: *Encyclopedia of Hydrological Sciences*, vol. 1, part 1, chapter 13, edited by: Anderson, M. G., John Wiley & Sons, Hoboken, N.J., 193–219, 2005.
- Troch, P. A., Carrillo, G. A., Heidbüchel, I., Rajagopal, S., Switanek, M., Volkmann, T. H. M., and Yaeger, M.: Dealing with landscape heterogeneity in watershed hydrology: a review of recent progress toward new hydrological theory, *Geogr. Compass*, 3/1, 375–392, doi:10.1111/j.1749-8198.2008.00186.x, 2009.