Richards equation at the hillslope scale: Can we resolve the heterogeneity of soil hydraulic material properties?

Hannes H. Bauser[1,2], **Minseok Kim**[1], **Wei-Ren Ng**[1], **Aaron Bugaj**[1], **Peter A. Troch**[1,2]

[1]Biosphere 2, University of Arizona, Tucson, AZ, USA
[2]Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, AZ, USA

Key Points:

- Gravity flow experiment at the Landscape Evolution Observatory hillslopes to observe heterogeneity of soil hydraulic material properties
- The heterogeneity shows large variations between neighboring sensor locations and can't be resolved with the hillsopes' dense sensor network
- Detailed process-based modeling of soil water movement remains unfeasible at the hillslope scale and requires effective material properties

Corresponding author: Hannes H. Bauser, hbauser@arizona.edu

Abstract

Process-based modeling of soil water movement with the Richards equation requires the description of soil hydraulic material properties, which are highly uncertain and heterogeneous at all scales. This limits the applicability of Richards equation at larger scales beyond the patch scale. The experimental capabilities of the three hillslopes of the Landscape Evolution Observatory (LEO) at Biosphere 2 provide a unique opportunity to observe the heterogeneity of hydraulic material properties at the hillslope scale. We performed a gravity flow experiment where through constant irrigation the water content increases until the hydraulic conductivity matches the irrigation flux above. The dense water content sensor network at LEO then allows to map the heterogeneity of hydraulic conductivity at a meter scale resolution. The experiment revealed spatial structures within the hillslopes, mainly a vertical trend with the lowest hydraulic conductivity close to the surface. However, the variation between neighbouring sensors is high, showing that the heterogeneity cannot be fully resolved even at LEO. By representing the heterogeneity in models through Miller scaling we showed the impact on hillslope discharge. For the hillslope with the smallest heterogeneity, representing the dominant structures was sufficient. However, for the two hillslopes with the larger overall heterogeneity, adding further details of the local heterogeneity did impact the discharge further. This highlights the limitations of Richards equation, which requires the heterogeneous field of material properties, at the hillslope scale and shows the relevance to improve our understanding of effective parameters to be able to apply the process-based model to larger scales.

1 Introduction

Soil water movement is a key process in Earth's critical zone. However, accurate model predictions remain challenging and become more and more difficult with increasing scales, since soil water flow is dominated by the soil's pore structure and its multiscale heterogeneity.

From the core to the patch scale, process-based modeling of soil water movement is typically based on the Richards equation. The Richards equation itself can be derived from a stationary approximation of the Navier-Stokes equation and upscaling the description of soil water movement from to pore to the core scale. The assumption of stationarity leads to the requirement of a local equilibrium of the hydraulic state variables on the scale of a representative elementary volume as well as on the resolution of numerical discretizations (Roth, 2008). This upscaling further leads to a description lacking of hysteresis and preferential flow in the Richards equation (Vogel, 2019). Nevertheless, at the pedon scale the Richards equation can be successfully applied in situations when its assumptions hold (e.g., Bauser et al., 2016).

At larger scales from hillslopes to catchments no upscaled process-based description of the water dynamics in the unsaturated zone is available. Consequently, applications of the Richards equation to the hillslope and catchment scale have been attempted (e.g., Bittelli et al., 2010; Koch et al., 2016; Camporese et al., 2010, 2019). However, these applications then face the additional challenge of unknown soil hydraulic material properties and their spatial distribution (Vogel et al., 2018). At the hillslope and catchment scale, the heterogeneity often cannot be resolved beyond soil layers, even in experimental hillslopes (e.g., Botto et al., 2018). Within the soil layers, material properties are assumed to be homogeneous and are described through effective properties. However, due to the nonlinear dynamics, simple averaging or first-order upscaling is unable to fully reproduce upscaled fluxes. Moreover, the effective material properties do not only depend on the heterogeneous distribution but also on boundary conditions, making the use of Richards equation at these larger scales rather a pragmatic choice than based on a sound physical understanding (Vereecken et al., 2007).

With sufficient local measurements, the heterogeneity of soil hydraulic material properties may be resolved, which then allows, potentially, the appropriate application of Richards equation. In a synthetic study, Chaudhuri et al. (2018) inversely estimated three-dimensional heterogeneous fields of soil hydraulic material properties based on local water content measurements that resolved the heterogeneity. Recently, Yu et al. (2021) showed in a one-dimensional approach that the spatial measurement interval for water content measurements should be set to spatial correlation length of the heterogeneity to be able to resolve the heterogeneity in a data assimilation framework. However, in real-world applications correlation lengths of the heterogeneity are typically unknown and are not resolved by local measurements, challenging the use of Richards equation.

The Landscape Evolution Observatory (LEO) at Biosphere 2 (B2), The University of Arizona, consists of three artificial hillslopes and features unique environmental control and measurement capabilities. One of the key foci of LEO is understanding spatial variability within hillslopes (Huxman et al., 2009). Each hillslope features a network of 496 water content sensors which provide an opportunity to resolve the heterogeneity of soil hydraulic material properties and apply Richards equation at the hillslope scale. The LEO hillslopes were designed homogeneously, but the introduction of heterogeneity during construction was expected. Further heterogeneity is likely to form through weathering (Dontsova et al., 2009) and rearrangement of fine sediments. For a first hydrologic experiment at LEO, Gevaert et al. (2014) analyzed water content measurements in detail and found indications for heterogeneity within the hillslopes: a slightly asymmetric infiltration front and decreasing water contents with depth during gravity flow. For the same experiment Niu et al. (2014) estimated soil hydraulic material properties and found an improved performance when including some heterogeneity through a reduced hydraulic conductivity at the seepage face. Scudeler et al. (2016) modeled a tracer experiment at LEO and found improved performance when including vertical heterogeneity within the hillslopes by introducing soil layers in the model. In a synthetic study using data assimilation Pasetto et al. (2015) investigated the ability to estimate heterogeneous fields of hydraulic conductivity with the sensors network at LEO and the impact of a reduced number of sensors due to expected sensor failure over time. They found that total water content could be estimated reliably while the ability to determine the three-dimensional field of hydraulic conductivity improved with increasing correlation length of the heterogeneity as well as increasing number of active sensors.

In this study we leveraged the capabilities at LEO to explore the heterogeneity of soil hydraulic material properties at LEO. The sensor network allows to map the heterogeneity at a unique resolution at the hillslope scale to then investigate the possibility to achieve accurate process-based modeling of soil water movement with the Richards equation beyond the patch scale. For this we performed a gravity flow experiment, where extended constant rainfall is applied to the hillslopes until a gravity flow regime is reached where flow is dominantly driven by gravity and local water measurements reveal information about the local hydraulic conductivity. In the next step we then incorporated different representations of the heterogeneity in soil hydrologic models to explore the impact on discharge as an integrated hillslope response.

2 Materials and Methods

2.1 Soil water movement

Assuming single phase flow and local equilibrium, water movement in soils can be described by the Richards equation:

$$\frac{\partial \theta}{\partial t} - \nabla \cdot \left[K(\theta) \left[\nabla h_m(\theta) - \mathbf{e}_z \right] \right] = 0 \,, \tag{1}$$

with the water content θ $(-)$, time t (T), isotropic hydraulic conductivity K (LT^{-1}), matric head h_m (L), and unit vector in direction of gravity $\mathbf{e}_z$ $(-)$.

Soil hydraulic material properties describe the sub-scale physics not explicitly represented in equation (1). These are the soil hydraulic conductivity function $K(\theta)$ and the soil water characteristics $h_m(\theta)$. The material properties are typically parameterized using simplified functions. We employ the Mualem-van Genuchten parameterization:

$$K(\Theta) = K_s \Theta^{\tau} \left[1 - \left[1 - \Theta^{n/[n-1]}\right]^{1-1/n}\right]^2 , \tag{2}$$

$$h_m(\Theta) = \frac{1}{\alpha}\left[\Theta^{-n/[n-1]} - 1\right]^{1-1/n} , \tag{3}$$

with saturation

$$\Theta = \frac{\theta - \theta_r}{\theta_s - \theta_r} . \tag{4}$$

The parameters θ_s $(-)$ and θ_r $(-)$ are the saturated and residual water content. Parameter α (L^{-1}) is a scaling factor, which can be associated with an inverse air entry value and n $(-)$ can be associated with the width of the pore size distribution. Parameter K_s (LT^{-1}) is the saturated hydraulic conductivity and τ $(-)$ can be associated with the increasing flow path tortuosity for decreasing saturation.

In the case of a constant infiltration with a flux $j < K_s$ into a homogeneous soil the water content approaches a constant given by $K(\theta) = j$. In this limit $\nabla h_m(\theta)$ vanishes and gravity becomes the only driving force in equation (1). This is referred to as gravity flow. In practice the establishment of gravity flow is typically limited due to, e.g., nonconstant infiltration, the presence of a groundwater table, and the impact of soil layers.

Soil hydraulic material properties are heterogeneous at all scales. A simplified way to represent heterogeneity is Miller scaling. Based on a reference material, the material properties at each location are scaled by a scalar field. By assuming geometric similarity the resulting material properties at each location $\mathbf{x}$ are:

$$h_m(\theta, \mathbf{x}) = h_m^*(\theta)\frac{1}{\xi(\mathbf{x})} , \tag{5}$$

$$K(\theta, \mathbf{x}) = K^*(\theta)\,\xi(\mathbf{x})^2 , \tag{6}$$

with * denoting the reference material, and ξ is the scaling factor. Small scaling factors $(\xi < 1)$ correspond to a finer material compared to the reference material, while larger scaling factors $(\xi > 1)$ correspond to coarser texture. A limitation of Miller scaling is that it assumes a homogeneous porosity.

2.2 Landscape Evolution Observatory

LEO consists of three artificial hillslopes and offers unique measurement capabilities. The three hillslopes are constructed identically and represent zero-order basins with an average slope of 10°. Each hillslope is 11 m wide, 30 m long, and filled with 1 m of crushed basalt. The texture is primarily loamy sand. Only for the most downslope 0.5 m of each hillslope the texture of the basalt is gravel. The downslope end is a seepage-face boundary.

To distinguish the three hillslopes in this manuscript they are labeled as 'West', 'Center', and 'East', according to their geographical location within the B2 facility. Locations within each hillslope are given based on their cross-slope location from -5.5 m to 5.5 m, their up-slope location from 0 m (downslope) to 30 m (upslope), and their depth within the soil from 0 m to 1 m. For a specific location at the surface, we use the notation (cross-slope location (m), up-slope location (m)), e.g., $(-4, 26)$. For a specific location within the hillslopes, we use the notation (cross-slope location (m), up-slope location (m), depth (m)), e.g., $(-4, 26, 0.05)$.

LEO is equipped with an intensive sensor and sampler network. In the following only the sensors used in this study are discussed. A more complete description of LEO can be found in Pangle et al. (2015). All measurements used in this study were recorded in 0.25 h intervals. The discharge at the seepage-face boundary at the downslope end of the hillslopes is measured spatially distributed in six sections through two different sensors in series: (i) magnetic flow meters (SeaMetrics, PE102) for high discharge and (ii) tipping buckets (NovaLynx, 260-2501-A tipping buckets) for low flow. The center hillslope has a leak, presumably close to the seepage face. The leaking water is collected and measured through an additional flow meter and tipping bucket. All tipping buckets were calibrated prior to the experiment in this study using a calibrator (NovaLynx, 260-2595 Rain Gauge Calibrator). A custom-engineered irrigation system allows to apply rainfall rates ranging from 3 to 40 mm h^{-1} through different combinations of five irrigation circuits. For the experiment one of these irrigation circuits was used ('Multi 1'), delivering approximately 13 mm h^{-1}.

Within the soil of each hillslope, 496 water content sensors (Decagon 5TM) are co-located with matric potential sensors (Decagon MPS-2). The sensors are installed in a 1 m x 2 m grid in different depths (5, 20, 35, 50, and 85 cm), depending on location. Due to a limited life span the number of active sensors is lower. During the experiment on average 56 % of the water content sensors and 75 % of matric potential sensors were operational. This still leaves over 270 active water content and over 330 matric potential sensors in each of the hillslopes.

The measured bulk relative dielectric permittivity ϵ_r by the water content sensors is related to water content through a petrophysical relationship. Following Roth et al. (1990), we use the complex refractive index model (CRIM), for an isotropic bulk medium:

$$\sqrt{\epsilon_r} = \theta\sqrt{\epsilon_{r,w}} + (\phi - \theta)\sqrt{\epsilon_{r,a}} + (1 - \phi)\sqrt{\epsilon_{r,s}} \,, \tag{7}$$

with the relative permittivity of air $\epsilon_{r,a} = 1$, and assuming a relative permittivity of the soil matrix of $\epsilon_{r,s} = 6.5$. The relative permittivity of water, $\epsilon_{r,w}$, depends on the temperature and is $\epsilon_{r,w} = 80$ at 20 °C. Based on the temperature measured by the water content sensors we employ the temperature correction by Kaatze (1989). Following Niu et al. (2014) and van den Heuvel et al. (2018) the porosity ϕ for all three hillslopes is assumed to be $\phi = 0.3675$.

2.3 Gravity flow experiment

To determine the heterogeneity of hydraulic conductivity in the three LEO hillslopes we performed a gravity flow experiment. In a first step the spatial rainfall distribution was determined. In a second step an extended temporally constant rainfall was applied to reach a gravity flow regime. The experiment was performed consecutively on each of the three hillslopes (Center: 9-10 July 2020, East: 15-16 July 2020, and West: 22-23 July 2020). A parallel realization of the experiment was not possible, due to the required long rainfall, which cannot be supported on all three hillslopes at the same time.

2.3.1 Rainfall distribution

To measure the rainfall distribution, 154 cups were placed on the hillslopes in a 1 m x 2 m grid from location $(-5, 2)$ to $(5, 30)$ matching the locations of the below-ground water content sensors (Figure 1). On day one of the experiment the hillslopes were irrigated for 1 h (8:30-9:30 local time). During the irrigation all air handlers within the LEO space were turned off to reduce the impact of air movement on the rainfall distribution. Afterwards, cups were collected using a gantry system to avoid disturbance of the soil. By comparing the weight to previously recorded dry-weight of the cups, the local irrigation rate was determined.

Figure 1. Center hillslope with cups to measure the rainfall distribution. The picture is taken at the upslope end looking downslope.

During the collection, three cups tipped over and no data is available for these locations (West slope $(5, 6)$ and Center slope $(-5, 10)$ and $(2, 26)$). Values for the locations were approximated by interpolating the surrounding measurements.

To reduce the impact of evaporation from the cups, they were immediately sealed with plastic wrap when collected. Although small, evaporation from the cups between the end of the 1 h irrigation and the collection time was accounted for. Five water filled cups were placed around the hillslope and weighted regularly. Evaporation was then determined through a linear interpolation. The evaporation for each individual cup was accounted for based on this linear interpolation and the collection time of the corresponding cup. For the Center and East hillslopes evaporation was less then 0.4 mm for the cup collected last. For the West hillslope, an issue with the gantry system caused an interruption during the collection leading to evaporation of up to 1.9 mm for the last collected cup.

2.3.2 Constant infiltration

On the second day, each hillslope was irrigated for an extended period of 12 h (8:30-20:30 local time). For consistency with the calibration on the previous day, again all air handlers within the LEO space were turned off. The constant infiltration enables to reach a gravity flow regime in parts of the hillslope. Limitations for establishing the gravity flow regime are (i) the impact of the rising groundwater table during the experiment, and (ii) a restricted maximum irrigation time to avoid potential overland flow. To minimize the impact of the groundwater table, the hillslopes were not irrigated for several weeks prior to the experiment to create drier initial conditions. The maximum irrigation length of 12 h was set through prior model simulations.

To determine if gravity flow was approached sufficiently, the temporal development of the water content of each sensor was first characterized qualitatively whether water content had reached a constant value without an impact of groundwater. Additionally, sensors had to fulfill the condition that water content did not change more than 0.01 in the last hour of the rainfall. For locations fulfilling both we assume that gravity flow was sufficiently reached and that the matric head gradient was negligible compared to gravity. Data recording for the West and Center slopes was interrupted few times during the irrigation event, however the last hour was uninterrupted.

Table 1. Calibrated effective soil hydraulic Mualem-van Genuchten parameters for the three LEO hillslopes as reported by van den Heuvel et al. (2018) for the West and Center hillslopes and by Niu et al. (2014) for the East hillslope.

Parameter	West	Center	East[a]
θ_s (-)	0.3675	0.3675	0.3675
θ_r (-)	0.0	0.0	0.002
α (m^{-1})	-2.252	-5.076	-1.667
n (-)	2.25	1.88	2.26
K_s ($\mathrm{m\,s}^{-1}$)	1.19×10^{-4}	1.79×10^{-4}	1.4×10^{-4}
τ (-)	0.5	0.5	0.5

[a]For the East hillslope Niu et al. (2014) used a lower K_s for the gravel layer at the seepage face of 2.2×10^{-5}.

Discharge at the seepage face was recorded in six separate sections and the additional leak for the Center slope throughout the experiment. For the Center slope the piping from one section leading to the flow meters was leaking at the beginning of the experiment. This was fixed shortly after the onset of the discharge on 10 July 2020 at 19:40.

2.3.3 Surface crust

Surface crusts can lead to reduced infiltration capacity and surface runoff (Eldridge et al., 2020). We observed a newly forming thin surface crust on the LEO hillslopes during the experiment. This crust can also be seen in Figure 1. When collecting cups for the rainfall calibration each location was qualitatively classified as *crust present*, *no crust present*, or *inconclusive* when no clear classification was possible due to both surface covers close to the location. During the following extended rainfall, no surface runoff due to the crust was observed. A more detailed analysis of the results with respect to the presence of the crust also showed no significant impact (Appendix A). Consequently, the crust is not considered in the remaining manuscript.

2.4 Local heterogeneity

At locations where gravity flow was established, the hydraulic conductivity for the measured water content matches the irrigation flux above. This provides for each sensor location a locally measured point on the hydraulic conductivity function. Assuming Miller scaling and reference material properties, Miller scaling factors can be determined for each measurement location using equation (6).

Effective soil hydraulic material properties are available from previous studies at LEO and have been calibrated to seepage-face outflow and total storage (Table 1). These are used as the reference material for each hillslope. These reference parameters have been calibrated assuming homogeneous rainfall distributions for each hillslope. The calibrations also assumed homogeneous material properties, except for the East hillslope, where a lower saturated hydraulic conductivity was used for the gravel layer at the seepage face. The reference material properties incorporate and compensate for any model structural errors, for example, the leak in the Center slope, which was not represented explicitly.

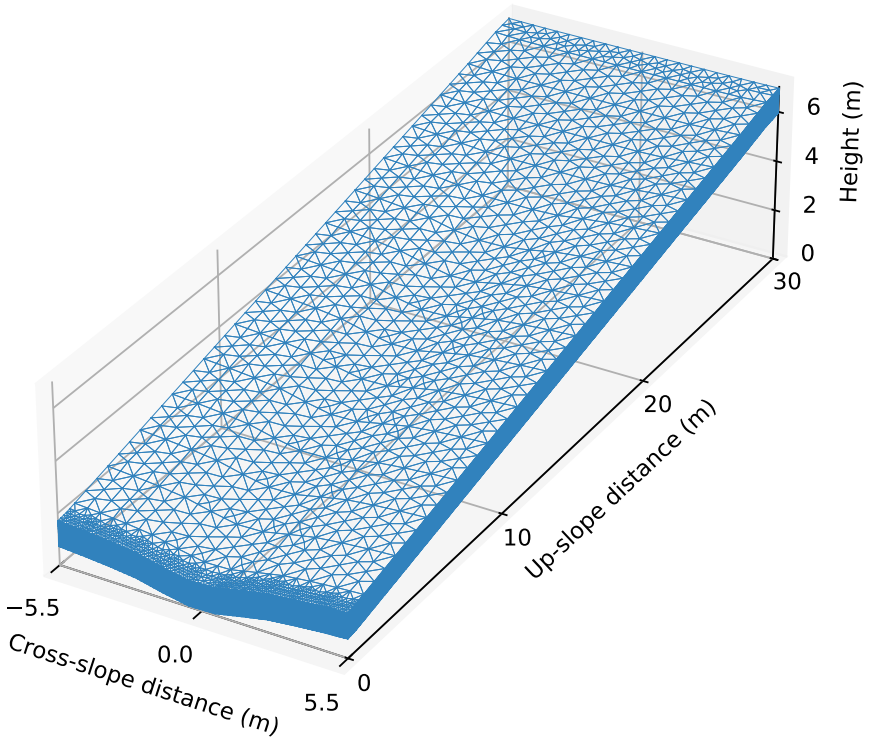

Figure 2. Mesh generated with GMSH for model simulations with DORiE.

2.5 Modeling

To solve Richards equation we employ the recently developed DORiE software framework which uses a discontinuous Galerkin method (Riedel et al., 2020). All model runs use the same grid, which was created using GMSH (Geuzaine & Remacle, 2009). The hillslopes are resolved vertically through 30 Layers, where the 10 top and 10 bottom layers have a resolution of 2.5 cm and the middle layers a resolution of 5 cm. The horizontal grid is based on 1 m × 2 m grids around each measurement location, wherein a mesh is generated with a set mesh resolution size of 0.9 m (Figure 2). While this is a rather coarse resolution for numerically solving Richards equation, this mesh results already in 1 143 720 degrees of freedom for each time step in DORiE. With this resolution each simulation run took about two weeks on a single 3 GHz Intel Core i5 processor.

The simulations shown in this study begin with the start of the 12 h irrigation and cover 14 d. The initial condition for these simulations are created for each model run through a spin-up run starting 8 d earlier. The initial condition for the spin-up period is determined by interpolating matric head measurements and extrapolating the values beyond the covered volume assuming static equilibrium. We chose a rather short spin-up period since the measurement range of the water potential sensors is limited to dry conditions below a head of −0.5 m. The hillslopes were not irrigated for several weeks prior to the experiment, leading to the driest conditions at the start of the experiment. A shorter spin-up period then does not only save computation costs, but also corresponds to a drier and consequently improved initial condition for the spin-up phase itself.

The upper boundary condition at the surface during the spin-up phase is realized through a Dirichlet boundary condition from interpolating the water potential sensors closest to the surface and extrapolating to the surface assuming static equilibrium. During the rain event the measured rainfall distribution is applied as a Neumann boundary

condition. After the rainfall an accurate Dirichlet boundary condition was not feasible any more to describe evaporation due to the limited measurement range of the matric potential sensors. While an average evaporation flux can be determined for the LEO hillslopes from mass measurements, applying this average flux as a Neuman boundary condition is not feasible, since this leads to non-physical situations in heterogeneous model representations of the hillslope. Since the current version of DORiE does not provide an evaporation boundary condition, that could switch to a Dirichlet boundary condition in case of a non-physical flux, the upper boundary condition is set to a no-flux boundary condition after the rainfall for the spin-up phase as well as the remaining simulation. The seepage-face boundary condition is realized through an outflow boundary condition, which is realized through a Dirichlet boundary condition of 0 m, but only if it leads to a flux out of the domain. Otherwise it is a no-flux boundary condition. All other boundary conditions are set to no-flux boundary conditions as well. This means that the leak in the Center hillslope is not represented in the model.

To investigate the impact of the observed heterogeneity several model scenarios with and without representation of heterogeneity were simulated. Miller scaling factors were determined locally at measurement locations where gravity flow was reached, but need to be inter- and extrapolated for a continuous representation in the model. For all scenarios, we assumed that the local Miller scaling factors do not contain information that can directly be extrapolated to the gravel layer (the most downslope 0.5 m) and consequently we did not apply Miller scaling there. The different scenarios are:

1. *Homogeneous rain and soil*. In this scenario no heterogeneity is assumed for the soil. The reference material properties of each hillslope are used. Additionally, as the only scenario, the rainfall distribution is averaged and assumed to be spatially homogeneous as well.
2. *Homogeneous soil*. No heterogeneity is assumed for the soil and the reference material properties are used. In contrast to scenario 1 (*homogeneous rain and soil*), the spatially resolved heterogeneous rainfall distribution is used.
3. *Vertical heterogeneity*. The logarithm of the Miller scaling factors are averaged horizontally in the different sensor depths of 5, 20, 35, and 50 cm. At the depth of 85 cm no gravity flow was reached. At this depth a Miller scaling factor of 1 (i.e., no scaling) is assumed to avoid that the local Miller scaling factors at a depth of 50 cm are extrapolated for the entire lower 50 cm of the domain. The horizontally averaged Miller scaling factors are then linearly interpolated vertically, and extrapolated constantly to the upper and lower boundaries of the domain. Again, the heterogeneous rainfall distribution is used.
4. *Kriged interpolation*. In this scenario, kriging is employed to interpolate the logarithm of the Miller scaling factors with the GeoStatTools python package (Müller & Schüler, 2020). In a first step, the trend of the vertical heterogeneity as in scenario 3 is removed with Miller scaling factors at a depth of 85 cm set to 1. A single horizontal variogram is calculated using all four depths with estimated Miller scaling factors. A Gaussian variogram model is fitted to the empirical variogram to estimate the range. For the vertical heterogeneity insufficient information is available to estimate a range. We set this range to 15 cm. Based on these variograms, the interpolated kriging field is calculated.
5. *Linear interpolation*. The logarithm of the local Miller scaling factors is interpolated linearly in three dimensions and extrapolated constantly towards the domain boundaries. Again, at a depth of 85 cm and below, Miller scaling factors are assumed to be 1. The heterogeneous rainfall distribution is used.

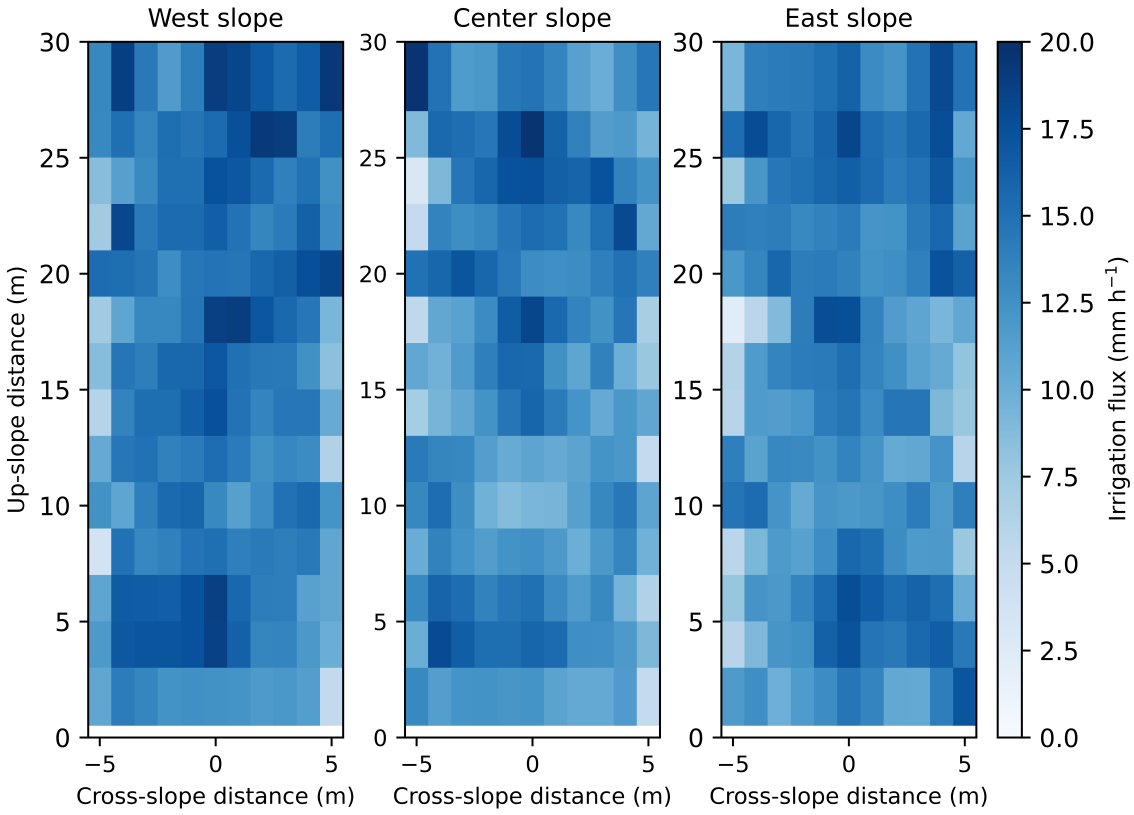

Figure 3. Rainfall distribution as estimated from 154 cups on each hillslope. The gravel at the most downslope 0.5 m does not receive rainfall since it is covered by a tarp.

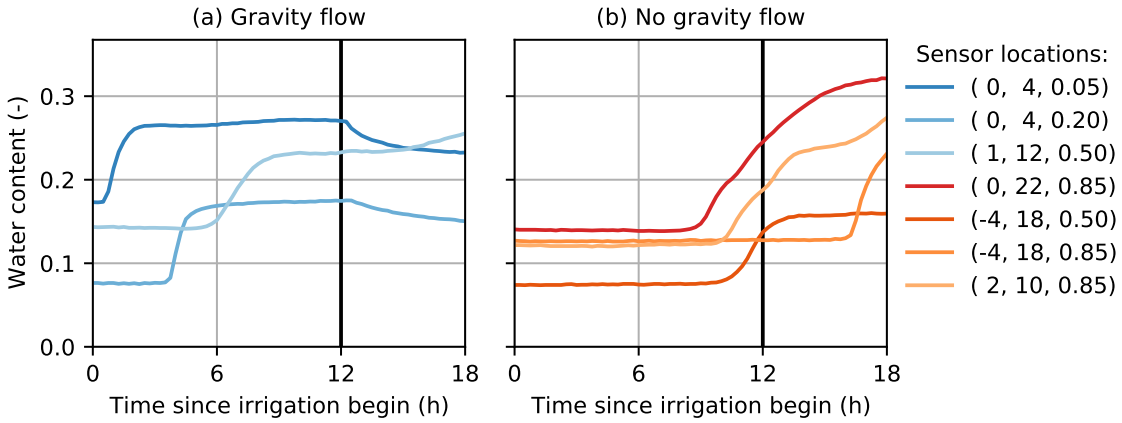

Figure 4. Examples for sensors on the East hillslope where a gravity flow regime was (a) established or (b) not established by the end of the 12 h rain event (vertical black line). More shallow sensors at $(0, 4, 0.05)$ and $(0, 4, 0.20)$ show a decline of water content soon after the irrigation stops. Deeper sensors do not show this response as fast and may show a further increase in the water content due to an increasing groundwater table, e.g., at $(1, 12, 0.50)$ and $(2, 10, 0.85)$.

3 Results

3.1 Rainfall distribution

The observed rainfall distribution is shown in Figure 3. Overall, the distribution shows slightly higher fluxes in the cross-slope middle of the hillslopes as well as towards the up-slope end of the hillslopes. Due to the design of the irrigation system with sprinklers at 7 different locations at each side of the hillslopes, the variation of fluxes becomes larger at the sides of the hillslopes as well. These overall patterns are consistent with previous calibrations (Pangle et al., 2015).

The average rainfall differs slightly between the hillslopes. It is $14.1\,\mathrm{mm\,h^{-1}}$ for the West slope, $12.7\,\mathrm{mm\,h^{-1}}$ for the Center slope, and $13.0\,\mathrm{mm\,h^{-1}}$ for the East slope.

3.2 Gravity flow

During the 12 h rain event gravity flow was established down to the sensors at 50 cm depth, except for few locations with low irrigation rate above. The rain event was not long enough to establish gravity flow at the sensors at the depth of 85 cm. Figure 4 shows few example locations for the East slope where gravity flow was established (Figure 4a) and not established (Figure 4b) by the end of the 12 h rain event. The figure also shows two examples for sensors at a depth of 50 cm. At $(1, 12, 0.50)$ a gravity flow regime is reached. The rainfall rate above is $12.2\,\mathrm{mm\,h^{-1}}$ and only slightly lower than the average. In contrast, the rainfall rate above the sensor at $(-4, 18, 0.50)$ is much lower with $5.6\,\mathrm{mm\,h^{-1}}$ and no gravity flow is reached there.

Figure 5 shows the water content for locations where gravity flow was established together with the irrigation flux above the respective sensor. In a homogeneous soil, this would show the hydraulic conductivity function. However, for all three hillslopes rather a point cloud is visible indicating the heterogeneity. Several differences and features are already apparent. For all hillslopes, and most distinguished for the East hillslope, the topmost sensors at a depth of 5 cm show a higher water content compared to the deeper sensors. Comparing the the three hillslopes, the West hillslope shows a smaller spread of water contents compared to the Center and East hillslopes.

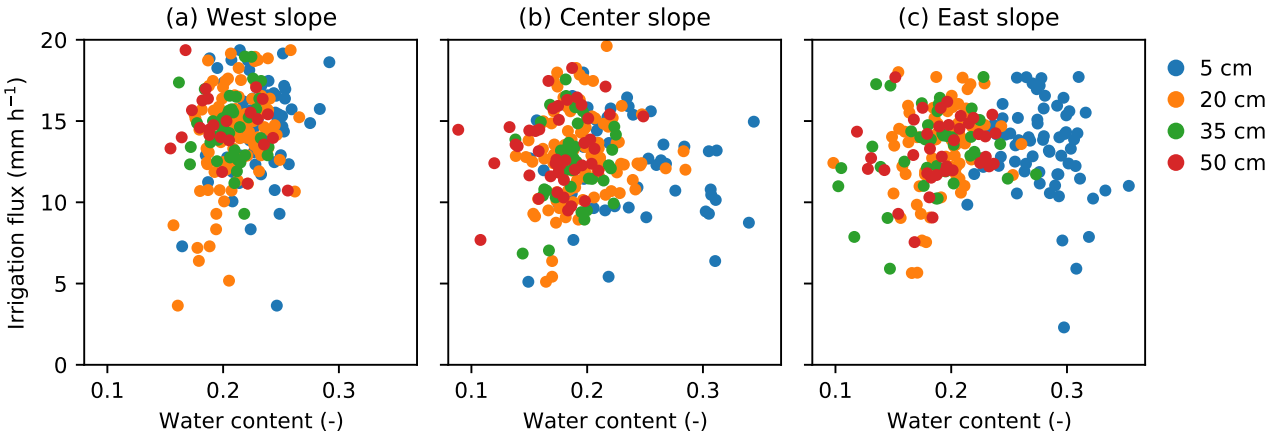

Figure 5. Observed water content under gravity flow conditions versus irrigation flux above the corresponding sensors for all three hillslopes. The different colors indicate the different depths of the sensors.

3.3 Observed heterogeneity

Assuming Miller scaling and reference material properties, the data pairs of irrigation flux and water content can be used to calculate a local Miller scaling factor (see section 2.4). Since this requires that a gravity flow regime was established at the corresponding sensor locations, no Miller scaling factors could be determined for the sensors at a depth of 85 cm. For a better comparison of the three hillslopes, with different reference material properties, we do not directly show the Miller scaling factors, but rather calculate the hydraulic conductivity at a saturation of $\Theta = 0.6$ (water content of $\theta = 0.22$), which is roughly in the middle of the observed water content range during gravity flow conditions. The relevance of this choice can be illustrated by comparing the hydraulic conductivities of the reference material properties. The saturated hydraulic conductivity of the Center slope is the highest of the three hillslopes. However, due to a smaller value of the parameter n in the Mualem-van Genuchten parameterization the hydraulic conductivity drops faster with decreasing saturation and is the smallest conductivity compared to the other two hillslopes at $\Theta = 0.6$.

Figure 6 shows a histogram of the calculated hydraulic conductivity for $\Theta = 0.6$. It reveals relevant heterogeneity within the hillslopes, where the West hillslope exhibits the least heterogeneity. The range of the hydraulic conductivity covers about one order of magnitude for the West hillslope and about two orders of magnitude for the Center and East hillslope. This is consistent with other studies at LEO. For example, Kim et al. (2022) reported higher differences in the soil water characteristics for the East hillslope compared to the West hillslope. Within the hillslopes the sensors closest to the surface at a depth of 5 cm show the lowest conductivity. This effect is particularly pronounced for the East hillslope. Trends for deeper depths are less pronounced. There are further differences between the hillslopes. The Center hillslope overall features a higher hydraulic conductivity, while the West hillslopes has the lowest conductivity for all depths except for the depth of 5 cm, where the East hillslope has the lowest conductivity.

Figure 7 depicts the distribution of hydraulic conductivity at a depth of 20 cm for each hillslope. This is the depth with the most active sensors. It shows that differences between neighboring locations are high and dominate the observed heterogeneity. Only little to no larger structures can be distinguished. The most relevant structure is in the Center hillslope showing lower hydraulic conductivity closer to the seepage face. This

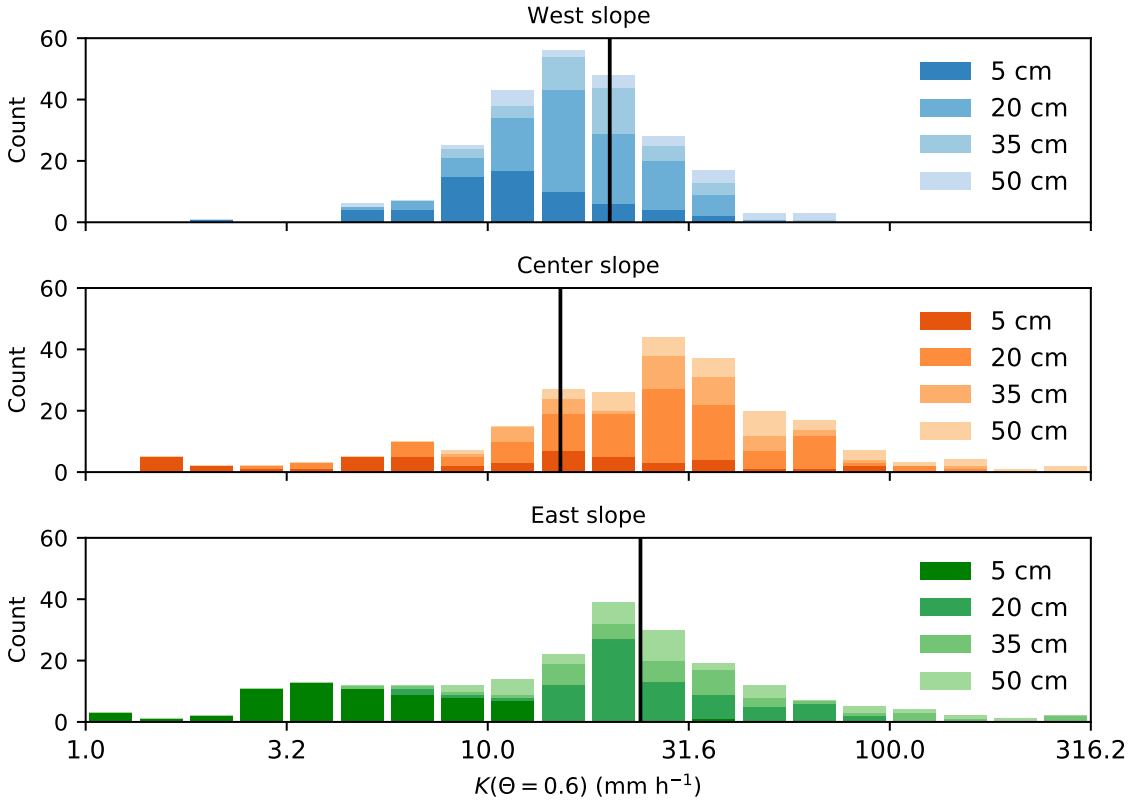

Figure 6. Comparison of hydraulic conductivity at a saturation of $\Theta = 0.6$ (water content of $\theta = 0.22$) under the assumption of Miller scaling and reference material properties as given in Table 1. The different color tones indicate the different depths of the sensors. The vertical black line marks the hydraulic conductivity of the corresponding reference material properties. Note the logarithmic scale for the hydraulic conductivity.

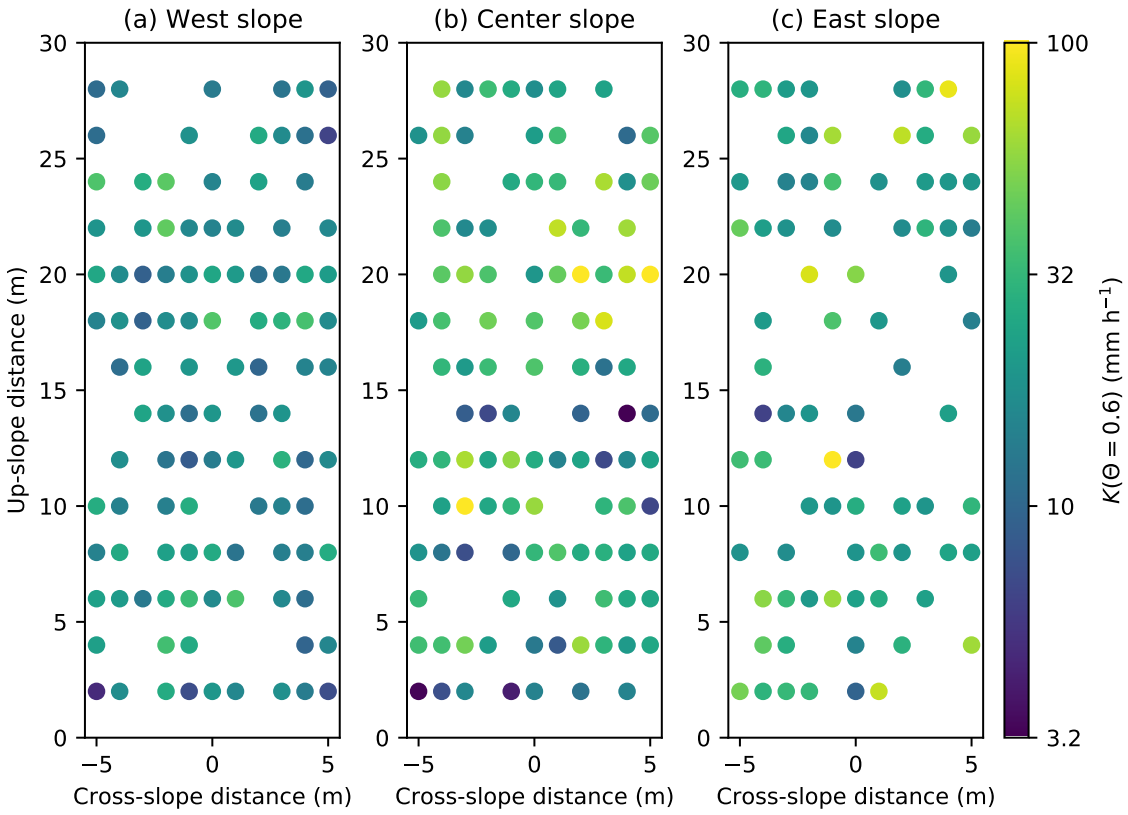

Figure 7. Spatial distribution of the hydraulic conductivity at a depth of 0.2 m. The hydraulic conductivity is shown at a saturation of $\Theta = 0.6$ (water content of $\theta = 0.22$) under the assumption of Miller scaling and reference material properties as given in Table 1.

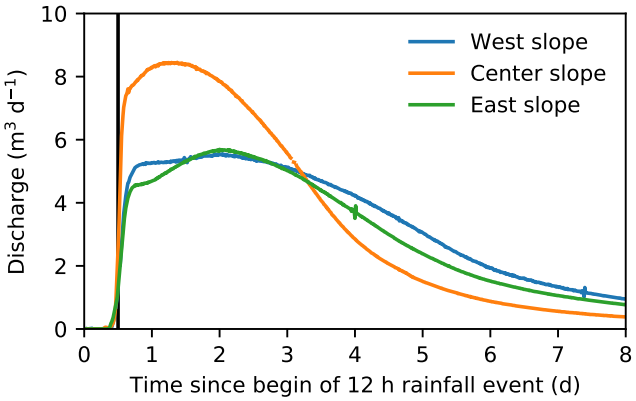

Figure 8. Observed discharge for the three hillslopes resulting from the 12 h rainfall event. The vertical black line marks the end of the irrigation.

is also confirmed by the calculated variogram, which yields a length scale of 7 m for the center hillslope. However, the nugget (i.e., the short range variance) still remains almost a factor of 1.8 larger than the variance of the structure. For the East hillslope the variogram has a length scale < 3 m, which is smaller than structures that could be resolved by the density of the sensor network, and for the West hillslope the variogram did not flatten off and no physically reasonable range was determined. This indicates that the horizontal heterogeneity is dominantly small scale heterogeneity, which cannot be resolved, even by the rather dense sensor network at LEO.

3.4 Observed Discharge

The discharge at the seepage face resulting from the 12 h irrigation event started approximately 9 h after the begin of the irrigation and differs between the three hillslopes (Figure 8). Initially the discharge of the Center hillslope is higher than the discharge of the West and East hillslope. After about 80 h the situation reverses and the discharge of the Center hillslope is smaller than of the other two hillslopes. The Center hillslope also reaches its maximum discharge the earliest. This is consistent with the higher hydraulic conductivity observed for the Center hillslope. The discharge behavior of the West and East hillslopes is more similar and shows a distinguished feature. Initially the discharge increases sharply, followed by a time with almost constant discharge, before discharge increases again slightly to reach the maximum discharge at about 48 h. Overall the discharge of the West slope is higher than the East hillslope (except for a short time period at around 48 h) which is consistent with the higher irrigation rate of the West hillslope.

The discharge for all three hillslopes continuously decreased further and 14 d after the begin of the rain event discharge is less than $0.2\,m^3\,d^{-1}$ for the West and East hillslopes, and less than $0.1\,m^3\,d^{-1}$ for the Center hillslope. Total discharge of all three hillslopes within these 14 d is comparable with $29\,m^3$, $28\,m^3$, and $27\,m^3$ discharge for the West, Center, and East hillslopes. Total evaporation was significant with $10\,m^3$ for the West and $11\,m^3$ for the Center and East hillslopes.

3.5 Modeled discharge

For all three hillslopes, five different model scenarios with different representation of heterogeneity were performed (section 2.5). This also leads to different initial conditions. However, differences in initial total storage are small. For the West hillslope ini-

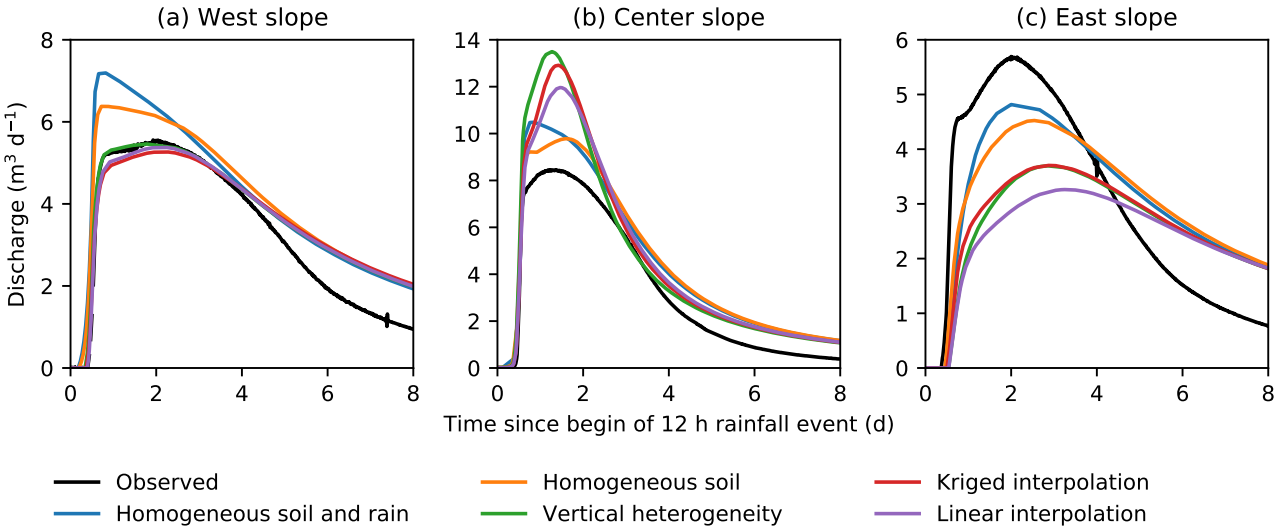

Figure 9. Observed and modeled discharge for the five different model scenarios described in section 2.5. Note the different axes for the three hillslopes. For comparison of the observed discharge see Figure 8.

tial storage is 31 m^3 for the two scenarios with homogeneous soil and 30 m^3 for the scenarios with heterogeneous representation of the soil. For the East slope all model scenarios have an initial storage of 32 m^3, except for the *linear interpolation* scenario with an initial storage of 33 m^3. Only for the Center hillslope, differences are slightly larger with 31 m^3 for the two scenarios with homogeneous soils, 28 m^3 for the *vertical heterogeneity* and the *kriged interpolation* scenarios and 29 m^3 for the *linear interpolation* scenario. Common to all the scenarios is that evaporation is not represented. Due to this, we expect that the models overestimate discharge. The total discharge during the 14 d following the rain event confirms this. For the West and Center slope modeled discharge for all model scenarios is higher than the observed discharge and ranges from 37 m^3 to 41 m^3 for the West hillslope and from 39 m^3 to 41 m^3 for the Center hillslope. Only for the East hillslope the modeled total discharge is more comparable to the observed total discharge and ranges from 27 m^3 to 33 m^3. However, we also expect that the impact of evaporation on the discharge is mainly relevant towards later times and less relevant in the first few days. Initially evaporation only impacts the topmost few centimeters at the surface.

The modeled discharge during the first 8 d for all hillslopes and scenarios is shown together with the observed discharge in Figure 9. As expected, the modeled discharge exceeds the observed discharge for all hillslopes and scenarios towards later times. During the first few days, the situation is more differentiated.

The scenario *homogeneous soil and rain* uses homogeneous rainfall distributions and the homogeneous reference material properties as calibrated in previous studies. Nevertheless, significant deviations from the observed discharge occur for all three hillslopes. These material properties were calibrated on specific, different experiments and consequently yielded effective material properties optimally describing those situations. Since effective material properties do not only depend on the heterogeneous distribution of soil hydraulic material properties, but also on the boundary conditions (Vereecken et al., 2007), deviations are expected. More specifically, for the West and Center slope the modeled peak discharge is too high and occurs too early, while for the East hillslope the peak discharge is too small.

The scenario *homogeneous soil* includes the heterogeneous rainfall distribution but still the homogeneous reference material properties. The cumulative rainfall remains the same as in the *homogeneous soil and rain* scenario. This has a significant impact on the modeled discharge. For all three hillslopes, the peak discharge is reduced. For the Center and East hillslope, the peak discharge occurs later. This is consistent with the slightly higher irrigation flux for the upslope part of the hillslopes. It also changes the qualitative shape of the curve, particularly for the Center slope.

The scenarios representing heterogeneity of the soil hydraulic material properties further decrease the peak discharge for the West and East hillslope, while the peak discharge is increased for the Center hillslope. This is consistent with the overall difference of observed hydraulic conductivity in relation to the hydraulic conductivity of the reference material properties. The observed conductivity is lower for the West and East slope and higher for the Center hillslope (Figure 6). This also indicates that the differences in the initial water content only have a comparatively small impact on the discharge. A higher initial water content is expected to lead to a higher peak discharge, however, for the Center hillslope, where the differences in the initial condition are largest, the opposite is the case.

The different scenarios with heterogeneous soil represent different features of the heterogeneity. The *vertical heterogeneity* scenario considers the vertical structure of the horizontally averaged heterogeneity. Most prominent here is the lower hydraulic conductivity close to the surface for all three hillslopes. The *kriged interpolation* scenario does also consider the vertical trend but additionally includes horizontal structures. Only the center hillsope shows a clear horizontal structure with lower hydraulic conductivity downslope. The *linear interpolation* scenario linearly interpolates between measurement locations. This way some local information at the measurement locations is included to the extent it could be resolved by the measurement network at LEO. However, the correlation length of the variations between neighbouring sensor locations remains unknown. Consequently, the scenario is not an accurate representation of the heterogeneity of hydraulic conductivity in the hillslopes and could even overemphasize the impact of the heterogeneity due to the interpolation.

For the West hillslope, including the *vertical heterogeneity* only, improves the modeled discharge and it very closely matches the observed discharge. Additionally including the horizontal structures in the *kriged interpolation* scenario or further details of the heterogeneity in the *linear interpolation* scenario does not relevantly change the discharge further. For the Center hillslope, including the *vertical heterogeneity* only increases the residual between observed and modeled discharge, but improves the timing of the maximum discharge. Additionally including the horizontal structures in the *kriged interpolation* as well as further details of the heterogeneity in the *linear interpolation* scenario both change the discharge further. For the East hillslope, including the *vertical heterogeneity* only, also increases the residual between observed and modeled discharge, but additionally including the horizontal structures in the *kriged interpolation* does not impact the discharge strongly. But including additional details from the local heterogeneity in the *linear interpolation* scenario further changes the modeled discharge.

The impact of the different scenarios on the discharge are consistent with the features of the observed heterogeneity. The West hillslope has the lowest heterogeneity. Only a clear structure for the vertical heterogeneity was be observed. Consequently, the different scenarios yield very similar results for the modeled discharge. In contrast, the Center hillslope shows clear differences between the three scenarios. The Center hillslope has a larger overall heterogeneity, and does not only feature a vertical structure, but also a horizontal structure. Consequently, adding the horizontal structure of the heterogeneity in the *kriged interpolation* scenario leads to changes in the modeled discharge compared to the *vertical heterogeneity* scenario. Due to the higher overall heterogeneity and the larger differences between individual locations, the *linear interpolation* scenario changes

the discharge further. The East hillslope features a large overall heterogeneity, but only a vertical structure. No clear horizontal structure of the heterogeneity was observed. Consequently, the *kriged interpolation* scenario and the *vertical heterogeneity* scenario show a similar discharge behavior. But again, due to the large overall heterogeneity between individual locations, the *linear interpolation* scenario alters the discharge behavior.

It is not expected that adding the heterogeneous features of rainfall and hydraulic material properties improves the predicted discharge. The effective material properties were calibrated to best describe the discharge of the respective experiment. The effective material properties already compensate the missing representation of heterogeneity as well as any other model errors. Consequently, we cannot expect that adding these heterogeneous features improves the modeled discharge. However, we still see the qualitative impact of the features on the discharge behavior. The quantitative close match of predicted discharge for the scenarios with heterogeneous material properties and observed discharge for the West hillslope are most likely due to chance.

The model results show that representing the heterogeneous rainfall distribution impacts the discharge, including the shape of the hydrograph and the time of maximum discharge. Process-based modeling of soil water movement at LEO should consequently include this distribution in the model, but will be limited in practice due to labor intensive calibration. The model results also show that representing the heterogeneity of soil hydraulic material properties in the topmost 50 cm of the LEO hillslopes has a clear impact on the predicted discharge behavior. Including this heterogeneity changes the outflow behavior not only quantitatively but also alters the shape of the hydrograph. It is noteworthy, that the groundwater table mostly stayed below 50 cm, which means that the impact on the discharge is mainly through the heterogeneity in the vadose zone above.

The three hillslopes represent different situations on how the heterogeneity needs to be represented. For the West hillslope the heterogeneity is the smallest. In this case it was sufficient to represent the dominant structure of the heterogeneity, which is the vertical heterogeneity. For the West hillslope including additional local heterogeneity did not have an impact on the discharge. For the East hillslope the dominant structure of the heterogeneity is again the vertical heterogeneity. However, overall the differences in material properties are larger. Including local heterogeneity through linear interpolation impacts the discharge. However, the sensor network is not able to fully resolve this heterogeneity. The chosen linear interpolation is only a simplified representation as well. For the Center hillslope the situation is similar to the East hillslope, except that the dominant structures that have an impact on the discharge are not only the vertical heterogeneity, but also the horizontal trend.

3.6 Infiltration front

To analyze the vertical dynamics across the hillslopes in more detail we calculated the arrival time of the infiltration front at each sensor location. Following Gevaert et al. (2014) we define the arrival time as the time when the water content increases for the first time. To avoid detecting an increase due to measurement noise, we set a threshold of 0.01 in the water content. Figure 10 shows the observed arrival time over the irrigation flux above the corresponding sensors location for all three hillslopes. As expected, the infiltration front reaches sensors in deeper depths later. Sensors at locations with higher irrigation flux above show an earlier arrival time of the infiltration front. The heterogeneity of soil hydraulic material properties, as well as different initial conditions cause additional variation in the arrival time. Modeled arrival times for all hillslopes and scenarios are earlier than the observed arrival times as illustrated in Figure 10 for the *linear interpolation* scenario.

For the West hillslope, where the observed discharge is represented best in the model scenarios, we exemplary show the observed and modeled water content during the irri-

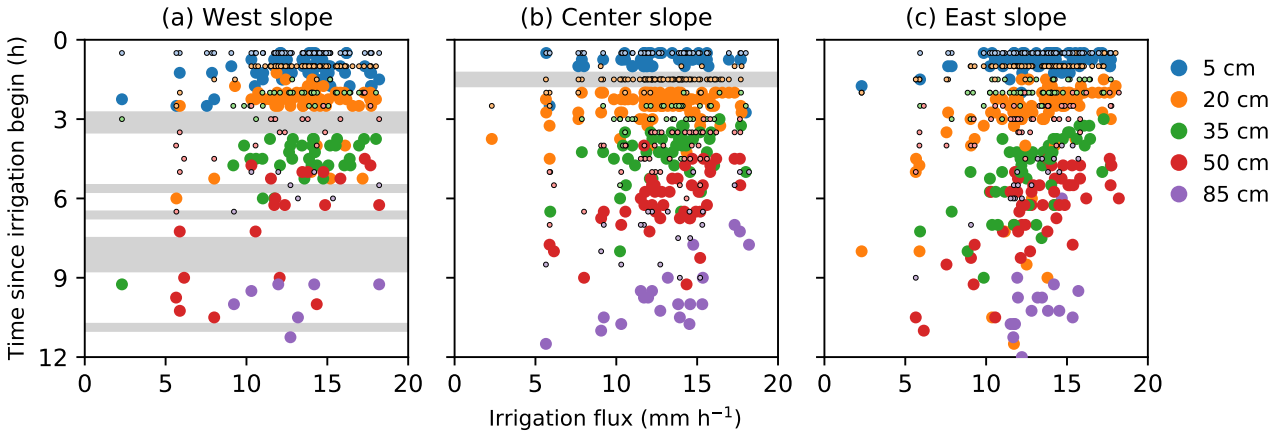

Figure 10. Arrival time of the infiltration front as determined from water content sensors versus irrigation flux above the corresponding sensors for all three hillslopes. Large circles are the arrival times determined by the observations. Small circles correspond to the modeled arrival times at the same sensor locations for the *linear interpolation* scenario. The grey bars for the West and Center slope indicate times when data recording was interrupted and no arrival time can be determined from the observations. Note that the measurement interval for the water content is 15 minutes, while model output was only written every 30 minutes. The different colors indicate the different depths of the sensors.

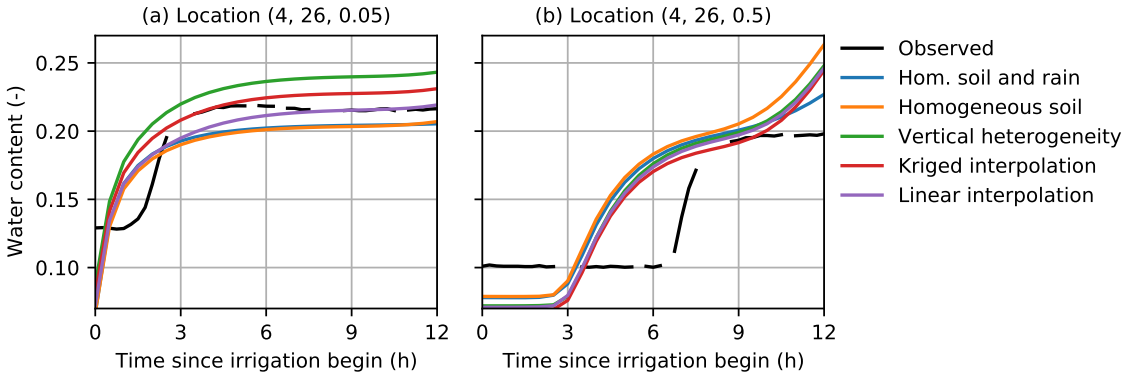

Figure 11. Examples for observed and modeled water content in the West hillslope. The two sensors are at different depths at the same location.

gation event for two sensors at the depths of 5 cm and 50 cm at the same location (Figure 11). The arrival time of the irrigation front differs strongly from the observed arrival time for all model scenarios. It is much earlier. This can be seen for both depths.

The observed water content reaches a gravity flow regime for both depths by the end of the 12 h rain event. Despite the same irrigation flux above the sensors, the water content differs between the two depths due to the local heterogeneity of the material properties. For all modeled scenarios, a gravity flow regime is sufficiently reached at the end of the 12 h rain event at a depth of 5 cm. However, despite the earlier arrival time of the infiltration front, the gravity flow regime, is not reached earlier than for the observed water content. The water content under gravity flow conditions for scenarios *homogeneous soil and rain* and *homogeneous soil* are very close, since the local irriga-

tion of 14.0 mm h^{-1} is close to the average irrigation rate of 14.1 mm h^{-1}. The water content in the other modeled scenarios differs due to the different representation of the heterogeneity. The scenario *linear interpolation* matches the observed water content, since in this scenario the locally determined Miller scaling factor is used, which ensures the matching hydraulic conductivity for the applied irrigation flux. In contrast, none of the model scenarios reach a gravity flow regime at the depth of 50 cm, since the water content increases further due to a developing groundwater table below.

This example shows that, despite the good match in discharge for the model scenarios that include a representation of the soil heterogeneity, the internal dynamics is not represented accurately. This highlights that a model calibration only based on discharge, even when relevant structures of the heterogeneity are included cannot be expected to yield physically reasonable material properties. Consequently, further model results like water content profiles, evapotranspiration, or solute transport are unreliable even during the calibrated time period.

3.7 Consistent representation of hillslope dynamics

All the performed model simulations are based on previously calibrated reference material properties, that effectively describe the overall hillslope discharge behavior. The calibrations were based on seepage-face discharge and total storage of the LEO hillslopes. In this manuscript we represent heterogeneity by modifying these reference material properties through Miller scaling based on a single point on the hydraulic conductivity function determined through the established gravity flow regime. Consequently, the heterogeneity in the model scenarios mainly represents the structure of heterogeneity based on the reference material properties. Our approach cannot estimate the full material properties (hydraulic conductivity function and soil water characteristic) and their heterogeneous distribution. This leads to the observed discrepancies for the local dynamics (Figure 11) despite a good description of discharge.

This underlying issue is well known in hydrologic modeling as equifinality (e.g., Beven, 2006) where the measurements used to calibrate the model parameters do not contain sufficient information leading to multiple acceptable solutions. This difficulty in model calibration due to equifinality has already been shown for the LEO hillslopes (Kim, 2018). It is noteworthy that this issue arises even at LEO, where model structural errors including the parameterizations, uncertainties in initial and boundary condition, as well as limitations on representing the heterogeneity are present, but are excepted to be smaller than for most real-world hillslopes.

To achieve a more consistent representation that describes the hillslope dynamics more accurately, additional information is necessary. By including not only discharge and storage information, but also water table measurements, Kim (2018) improved the model prediction of the breakthrough curve of a passive solute tracer at LEO, which indicates an improved representation of the internal dynamics. Similarly, including tracer information in the calibration can improve the consistency of the representation as well (e.g., McDonnell & Beven, 2014).

The large number of local water content measurements at LEO further provide information about the internal dynamics. Including this information can help finding more consistent representations. For example, Scudeler et al. (2016) used horizontally averaged water contents to improve the estimation of soil hydraulic material properties at LEO. However, the observed small scale heterogeneity limits the applicability of such averaging approaches due to the nonlinearity of the system and makes it challenging to fully harvest the information of the local measurements. To directly include the local measurements in the estimation of material properties, the local heterogeneity has to be included as well. In a synthetic study using data assimilation, Bauser et al. (2020) recently estimated effective reference material properties in heterogeneous materials based on lo-

cal water content measurements by including the local heterogeneity through Miller scaling. This approach showed high accuracy for predicting water contents and fluxes. Such an approach can be applied to the LEO hillslopes as well to effectively represent the unresolved small scale heterogeneity but needs to be extended by the observed dominant structure of the scaling factors as observed in this study. A limitation of our study as well as of Bauser et al. (2020) is the assumption of Miller scaling, which assumes a constant porosity for the entire domain. This is unlikely for LEO. However, Miller scaling can be extended by a second scaling factor that also scales the porosity. A preliminary application of the approach to individual soil profiles at LEO showed promising results. We are confident that such an approach can lead to an accurate description of both local dynamics and hillslope discharge for the West hillslope, since there the heterogeneity is small enough that the representation of the local heterogeneity does not have a strong impact on the discharge behavior. For the Center and East hillslope, representing the local heterogeneity does impact the modeled discharge and is consequently expected to impact the estimation of effective material properties. This may lead to a less consistent representation of the hillslope dynamics impacting further processes such as solute transport that rely on an accurate description of the water dynamics. A remaining limitation is that the observed heterogeneity in this study is constrained to the topmost 50 cm. Below that depth no information about the structure of the heterogeneity is available from the performed experiment.

While the observed heterogeneity at LEO is in an artificial system, we are convinced that theses findings are equally relevant for real-world systems. We expect local heterogeneity to be present in real-world soils as well. This can be seen in data sets that feature several sensors within individual soil layers (e.g., Bauser et al., 2016; Martini et al., 2021). While information from local measurements should be included for a more consistent estimation of material properties, the likely local heterogeneity surrounding individual sensors must be accounted for. This is particularly important when the available number of measurements is small and could lead to biased material properties if local heterogeneity is not accounted for.

4 Conclusions

This study used the unique measurement capabilities of the LEO hillslopes to investigate the heterogeneity of soil hydraulic material properties at the hillslope scale. The hillslopes were constructed homogeneously and typically would be represented by a single homogeneous layer in models. The measurement network of over 270 active water content sensors in each hillslope in combination with controllable boundary conditions allows to map the heterogeneity within the three hillslopes through a gravity flow experiment. This gives the unique opportunity to explore the possibility to represent heterogeneity at the hillslope scale and investigate the stringent application of small-scale process understanding of the Richards equation to larger scales.

The observed heterogeneity ranges 1-2 orders of magnitude for the hydraulic conductivity, where one of the hillslopes (West) has a smaller heterogeneity than the two other hillslopes (Center and East). The heterogeneity is dominantly local with large differences between neighbouring measurement locations and cannot be resolved by the measurement network. The correlation length of this local heterogeneity remains unknown. This makes it consequently also impossible to resolve the water content in detail. However, the experiment did reveal also larger structures. All three hillslopes feature a vertical structure, with the lowest conductivity observed by the most shallow sensors. One of the three hillslopes (Center) additionally features a horizontal structure, with lower hydraulic conductivity observed downslope.

We investigated the impact of the heterogeneity on the hillslope discharge as an integrated hillslope response. For this the observed heterogeneity was represented assum-

ing Miller scaling in a Richards equation solver. Resolving the structures of the heterogeneity impacted the modeled discharge. Representing some of the local heterogeneity through linearly interpolating the observed heterogeneity further did not change the modeled discharge for the West hillslope, where the heterogeneity is smallest, but did for the Center and East hillslopes, where the heterogeneity is larger. At LEO the rainfall distribution is also heterogeneous and alters the discharge behavior. Consequently, detailed modeling of soil water movement at LEO should ideally include the observed structure of the heterogeneity as well as the heterogeneous rainfall distribution. However, since the small scale heterogeneity cannot be resolved, this heterogeneity needs to be represented through effective material properties, even if the larger structures are specifically prescribed.

The misrepresentation of the local dynamics in the model, even when hillslope discharge is predicted accurately confirms the need of additional information when calibrating hillslope scale process-based models to achieve a consistent representation. This is particularly needed if further processes that depend on the accurate description of the soil water dynamics are included. At LEO this will become more important with the introduction of plants in the near future and the increasing relevance of root water uptake. To achieve a more consistent description of soil water movement additional information is needed, for example, through tracer data or through local measurements. When incorporating local measurements this needs to be done by accounting for the local heterogeneity at each sensor location which differs from the effective material properties describing the effective behavior of the heterogeneous soil. This becomes especially relevant in real-world applications if only few local measurements are available. If local heterogeneity is not considered this can lead to biases and may be even detrimental.

The failure to sufficiently resolve the heterogeneity of soil hydraulic material properties at LEO, one of the world's hydrologic flagship experiments, highlights the importance of new strategies for further upscaling of process-based modeling, especially in real-world applications. Since small scale heterogeneity cannot be resolved, the use of effective material properties becomes inevitable and fundamentally restricts the applicability of Richards equation, leading to reduced predictive capabilities. To still achieve high model accuracy and employ the process-based small scale representations we need to better understand and predict the uncertainties resulting from effective properties as well as their dependency on the specific hydraulic situation, for example, boundary conditions.

Appendix A Surface crust

During the experiment we observed a recently formed surface crust on the hillslopes. The extent of this crust was largest on the West slope. Surface crusts can lead to surface runoff, but no surface runoff was observed during the experiment. Additionally, we explored whether the observed heterogeneity indicates a reduced infiltration below locations with observed crust cover. If a reduced infiltration occurred, this would lead to a reduced water content at sensors locations below the crust during the gravity flow conditions. The lower water content in combination with the assumed unaffected irrigation flux measured above the surface would then lead to the estimation of a too large Miller scaling factor, which translates into a larger hydraulic conductivity. Figure A1 shows a histogram of the hydraulic conductivity at sensors locations with established gravity flow under the different surface covers above the sensor. Differences between the sensors below the different surface covers are small. The data do not show higher conductivity below the surface crust.

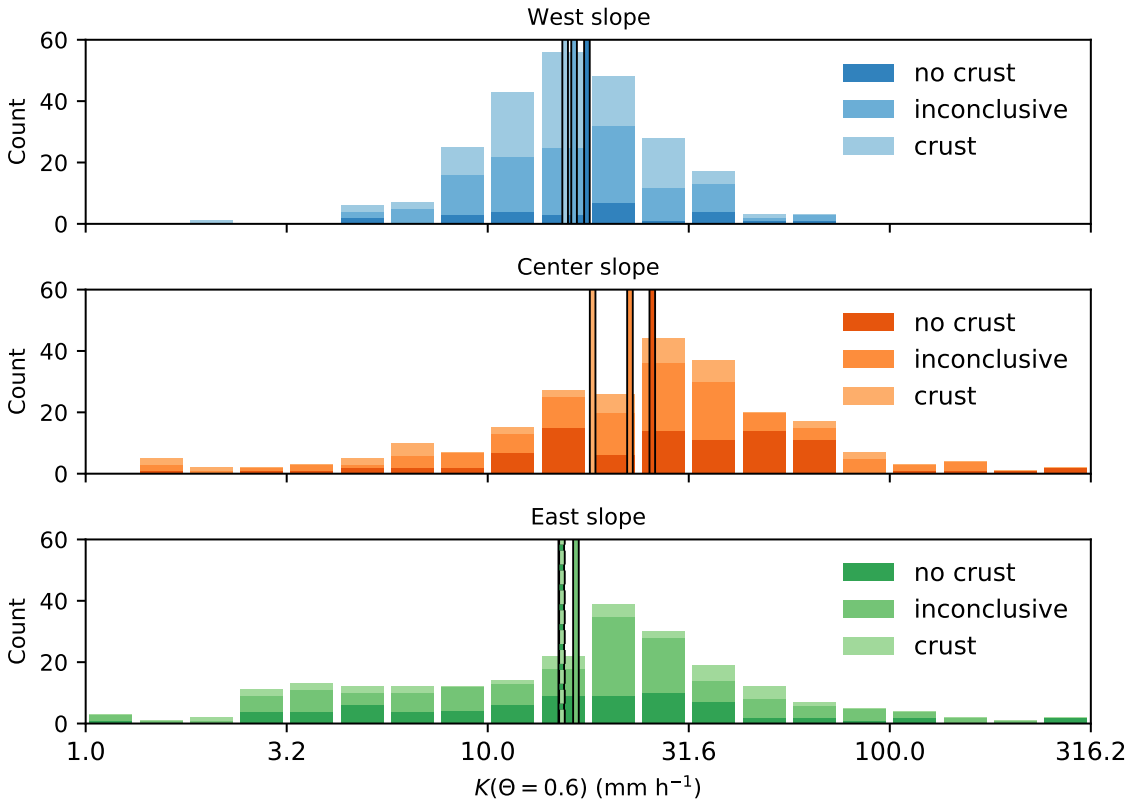

Figure A1. Comparison of hydraulic conductivity at a saturation of $\Theta = 0.6$ (water content of $\theta = 0.22$) under the assumption of Miller scaling and reference material properties as given in Table 1. The different color tones indicate the different crust covers above the corresponding sensors. The vertical lines mark the averaged logarithm of the hydraulic conductivity.

Acknowledgments

DORiE is publicly available (Riedel et al., 2020). The underlying data is available upon contacting Biosphere 2 (https://biosphere2.org/research/systems-data).

H.H. Bauser was funded by the Deutsche Forschungsgemeinschaft (DFG) through Project BA 6635/1-1 and by the National Science Foundation (NSF) through grant EAR-2120113. The authors gratefully acknowledge support from the Philecology Foundation of Fort Worth Texas. Additional funding support was provided by the Office of the Vice President of Research at the University of Arizona and by the Technology and Research Initiative Fund (TRIF) Water, Environmental, and Energy Solutions (WEES) initiative at the University of Arizona (Shared Equipment Enhancement Funds).

We thank Lukas Riedel and Santiago Ospina De Los Ríos for their support with DORiE.

References

Bauser, H. H., Jaumann, S., Berg, D., & Roth, K. (2016). EnKF with closed-eye period – towards a consistent aggregation of information in soil hydrology. *Hydrology and Earth System Sciences*, *20*(12), 4999–5014. doi: 10.5194/hess-20-4999-2016

Bauser, H. H., Riedel, L., Berg, D., Troch, P. A., & Roth, K. (2020). Challenges with effective representations of heterogeneity in soil hydrology based on local water content measurements. *Vadose Zone Journal*, *19*(1), e20040. doi: 10.1002/vzj2.20040

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, *320*(1), 18-36. doi: 10.1016/j.jhydrol.2005.07.007

Bittelli, M., Tomei, F., Pistocchi, A., Flury, M., Boll, J., Brooks, E. S., & Antolini, G. (2010). Development and testing of a physically based, three-dimensional model of surface and subsurface hydrology. *Advances in Water Resources*, *33*(1), 106-122. doi: 10.1016/j.advwatres.2009.10.013

Botto, A., Belluco, E., & Camporese, M. (2018). Multi-source data assimilation for physically based hydrological modeling of an experimental hillslope. *Hydrology and Earth System Sciences*, *22*(8), 4251–4266. doi: 10.5194/hess-22-4251-2018

Camporese, M., Paniconi, C., Putti, M., & McDonnell, J. J. (2019). Fill and spill hillslope runoff representation with a richards equation-based model. *Water Resources Research*, *55*(11), 8445-8462. doi: 10.1029/2019WR025726

Camporese, M., Paniconi, C., Putti, M., & Orlandini, S. (2010). Surface-subsurface flow modeling with path-based runoff routing, boundary condition-based coupling, and assimilation of multisource observation data. *Water Resources Research*, *46*(2). doi: 10.1029/2008WR007536

Chaudhuri, A., Hendricks Franssen, H.-J., & Sekhar, M. (2018). Iterative filter based estimation of fully 3D heterogeneous fields of permeability and Mualem-van Genuchten parameters. *Advances in Water Resources*, *122*, 340-354. doi: 10.1016/j.advwatres.2018.10.023

Dontsova, K., Steefel, C. I., Desilets, S., Thompson, A., & Chorover, J. (2009). Solid phase evolution in the Biosphere 2 hillslope experiment as predicted by modeling of hydrologic and geochemical fluxes. *Hydrology and Earth System Sciences*, *13*(12), 2273–2286. doi: 10.5194/hess-13-2273-2009

Eldridge, D. J., Reed, S., Travers, S. K., Bowker, M. A., Maestre, F. T., Ding, J., ... Zhao, Y. (2020). The pervasive and multifaceted influence of biocrusts on water in the world's drylands. *Global Change Biology*, *26*(10), 6003-6014. doi: 10.1111/gcb.15232

Geuzaine, C., & Remacle, J.-F. (2009). Gmsh: A 3-D finite element mesh generator with built-in pre- and post-processing facilities. *International Journal for Nu-*

merical Methods in Engineering, *79*(11), 1309-1331. doi: doi.org/10.1002/nme.2579

Gevaert, A. I., Teuling, A. J., Uijlenhoet, R., DeLong, S. B., Huxman, T. E., Pangle, L. A., ... Troch, P. A. (2014). Hillslope-scale experiment demonstrates the role of convergence during two-step saturation. *Hydrology and Earth System Sciences*, *18*(9), 3681–3692. doi: 10.5194/hess-18-3681-2014

Huxman, T., Troch, P. A., Chorover, J., Breshears, D. D., Saleska, S., Pelletier, J., & Zeng, X. (2009). The hills are alive: Earth science in a controlled environment. *Eos, Transactions American Geophysical Union*, *90*(14), 120-120. doi: 10.1029/2009EO140003

Kaatze, U. (1989). Complex permittivity of water as a function of frequency and temperature. *Journal of Chemical & Engineering Data*, *34*(4), 371–374. doi: 10.1021/je00058a001

Kim, M. (2018). *Processes and landscape structure underlying system scale hydrologic transport: Theory, experiment, and modeling* (Doctoral dissertation, Johns Hopkins University, Baltimore, Maryland). Retrieved from `http://jhir.library.jhu.edu/handle/1774.2/60123`

Kim, M., Volkmann, T. H. M., Wang, Y., Meira Neto, A. A., Matos, K., Harman, C. J., & Troch, P. A. (2022). Direct observation of hillslope scale StorAge Selection functions in experimental hydrologic systems: Geomorphologic structure and preferential discharge of old water. *Water Resources Research*, *n/a*(n/a). (e2020WR028959) doi: 10.1029/2020WR028959

Koch, J., Cornelissen, T., Fang, Z., Bogena, H., Diekkrüger, B., Kollet, S., & Stisen, S. (2016). Inter-comparison of three distributed hydrological models with respect to seasonal variability of soil moisture patterns at a small forested catchment. *Journal of Hydrology*, *533*, 234-249. doi: 10.1016/j.jhydrol.2015.12.002

Martini, E., Bauckholt, M., Kögler, S., Kreck, M., Roth, K., Werban, U., ... Zacharias, S. (2021). *STH-net:* a soil monitoring network for process-based hydrological modelling from the pedon to the hillslope scale. *Earth System Science Data*, *13*(6), 2529–2539. doi: 10.5194/essd-13-2529-2021

McDonnell, J. J., & Beven, K. (2014). Debates—the future of hydrological sciences: A (common) path forward? A call to action aimed at understanding velocities, celerities and residence time distributions of the headwater hydrograph. *Water Resources Research*, *50*(6), 5342-5350. doi: doi.org/10.1002/2013WR015141

Müller, S., & Schüler, L. (2020, April). *Geostat-framework/gstools: Volatile violet v1.2.1.* Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.3751743` doi: 10.5281/zenodo.3751743

Niu, G.-Y., Pasetto, D., Scudeler, C., Paniconi, C., Putti, M., Troch, P. A., ... Zeng, X. (2014). 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. *Hydrology and Earth System Sciences*, *18*(5), 1873–1883. doi: 10.5194/hess-18-1873-2014

Pangle, L. A., DeLong, S. B., Abramson, N., Adams, J., Barron-Gafford, G. A., Breshears, D. D., ... Zeng, X. (2015). The Landscape Evolution Observatory: A large-scale controllable infrastructure to study coupled earth-surface processes. *Geomorphology*, *244*, 190–203. doi: 10.1016/j.geomorph.2015.01.020

Pasetto, D., Niu, G.-Y., Pangle, L., Paniconi, C., Putti, M., & Troch, P. A. (2015). Impact of sensor failure on the observability of flow dynamics at the Biosphere 2 LEO hillslopes. *Advances in Water Resources*, *86*, 327-339. (Data assimilation for improved predictions of integrated terrestrial systems) doi: 10.1016/j.advwatres.2015.04.014

Riedel, L., Ríos, S. O. L. D., Häfner, D., & Klein, O. (2020). DORiE: A discontinuous Galerkin solver for soil water flow and passive solute transport based on DUNE. *Journal of Open Source Software*, *5*(52), 2313. doi: 10.21105/joss.02313

Roth, K. (2008). Scaling of water flow through porous media and soils. *European Journal of Soil Science*, *59*(1), 125-130. doi: 10.1111/j.1365-2389.2007.00986.x

Roth, K., Schulin, R., Flühler, H., & Attinger, W. (1990). Calibration of time domain reflectometry for water content measurement using a composite dielectric approach. *Water Resources Research*, *26*(10), 2267–2273. doi: 10.1029/WR026i010p02267

Scudeler, C., Pangle, L., Pasetto, D., Niu, G.-Y., Volkmann, T., Paniconi, C., ... Troch, P. (2016). Multiresponse modeling of variably saturated flow and isotope tracer transport for a hillslope experiment at the Landscape Evolution Observatory. *Hydrology and Earth System Sciences*, *20*(10), 4061–4078. doi: 10.5194/hess-20-4061-2016

van den Heuvel, D. B., Troch, P. A., Booij, M. J., Niu, G.-Y., Volkmann, T. H., & Pangle, L. A. (2018). Effects of differential hillslope-scale water retention characteristics on rainfall–runoff response at the Landscape Evolution Observatory. *Hydrological Processes*, *32*(13), 2118-2127. doi: https://doi.org/10.1002/hyp.13148

Vereecken, H., Kasteel, R., Vanderborght, J., & Harter, T. (2007). Upscaling hydraulic properties and soil water flow processes in heterogeneous soils: A review. *Vadose Zone Journal*, *6*(1), 1-28. doi: 10.2136/vzj2006.0055

Vogel, H.-J. (2019). Scale issues in soil hydrology. *Vadose Zone Journal*, *18*(1), 190001. doi: 10.2136/vzj2019.01.0001

Vogel, H.-J., Bartke, S., Daedlow, K., Helming, K., Kögel-Knabner, I., Lang, B., ... Wollschläger, U. (2018). A systemic approach for modeling soil functions. *SOIL*, *4*(1), 83–92. doi: 10.5194/soil-4-83-2018

Yu, D., Zha, Y., Shi, L., Bolotov, A., & Tso, C.-H. M. (2021). Spatiotemporal sampling strategy for characterization of hydraulic properties in heterogeneous soils. *Stochastic Environmental Research and Risk Assessment*, *35*(3), 737–757. doi: 10.1007/s00477-020-01882-1