Hydrology and
Earth System
Sciences

Form and function in hillslope hydrology: in situ imaging and characterization of flow-relevant structures

Conrad Jackisch[1], **Lisa Angermann**[2,3], **Niklas Allroggen**[3], **Matthias Sprenger**[4,5], **Theresa Blume**[2], **Jens Tronicke**[3], and **Erwin Zehe**[1]

[1]Karlsruhe Institute of Technology (KIT), Institute for Water and River Basin Management,
Chair of Hydrology, Karlsruhe, Germany
[2]Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Section Hydrology, Potsdam, Germany
[3]University of Potsdam, Institute of Earth and Environmental Science, Potsdam, Germany
[4]University of Freiburg, Institute of Geo- and Environmental Natural Sciences, Chair of Hydrology, Freiburg, Germany
[5]University of Aberdeen, School of Geosciences, Geography & Environment, Aberdeen, Scotland, UK

Correspondence to: Conrad Jackisch (jackisch@kit.edu)

Abstract. The study deals with the identification and characterization of rapid subsurface flow structures through pedo- and geo-physical measurements and irrigation experiments at the point, plot and hillslope scale. Our investigation of flow-relevant structures and hydrological responses refers to the general interplay of form and function, respectively. To obtain a holistic picture of the subsurface, a large set of different laboratory, exploratory and experimental methods was used at the different scales. For exploration these methods included drilled soil core profiles, in situ measurements of infiltration capacity and saturated hydraulic conductivity, and laboratory analyses of soil water retention and saturated hydraulic conductivity. The irrigation experiments at the plot scale were monitored through a combination of dye tracer, salt tracer, soil moisture dynamics, and 3-D time-lapse ground penetrating radar (GPR) methods. At the hillslope scale the subsurface was explored by a 3-D GPR survey. A natural storm event and an irrigation experiment were monitored by a dense network of soil moisture observations and a cascade of 2-D time-lapse GPR "trenches". We show that the shift between activated and non-activated state of the flow paths is needed to distinguish structures from overall heterogeneity. Pedo-physical analyses of point-scale samples are the basis for sub-scale structure inference. At the plot and hillslope scale 3-D and 2-D time-lapse GPR applications are successfully employed as non-invasive means to image subsurface response patterns and to identify flow-relevant paths.

Tracer recovery and soil water responses from irrigation experiments deliver a consistent estimate of response velocities. The combined observation of form and function under active conditions provides the means to localize and characterize the structures (this study) and the hydrological processes (companion study Angermann et al., 2017, this issue).

1 Introduction

1.1 Form–function relationship in hydrological sciences

From a general perspective the interplay of processes and spatial structures (Grayson and Blöschl, 2001) manifests itself as patterns in dynamics (Sivapalan, 2005) and self-organization (Zehe et al., 2013). This interplay can be expressed as a form–function relationship, which is addressed in many disciplines. Especially in systems biology, the form–function relations are deeply rooted (e.g., Aristotle in Blits, 1999; Thompson, 1917) and under debate until today (e.g., Mugler et al., 2011). In abstract terms the relation of form and function is fundamental for the concept that we can predict the behavior of a system under different forcings by knowing its constructive properties. In this respect we understand form as the shape and material property of the soil domain, whereas function refers to the dynamic behavior of wa-

ter within the same. Although the existence and importance of form–function relationships are generally agreed upon, it is not clear to what extent form follows or reveals function and vice versa. In a soil-hydrological context of soil–water interactions the retention curve relates the pore size distributions and their covariance structure to storage of water against gravity and root water uptake. The hydraulic conductivity curve relates the pore size distribution and the interconnectedness of the pores to the conductance/release function of water depending on the wetting state. These are classic examples of form–function relations at the Darcy scale. However, the established relation does not directly translate to water displacement and contact angles at the actual pore scale (Armstrong et al., 2016). At larger scales, accepted form–function relations turn out to be incomplete when preferential flow paths become important, as observed at plots of different soil types (Flury et al., 1994) and in most catchments (Uhlenbrook, 2006). Form–function relations for plots and hillslopes should reflect how macropore density and connectivity in conjunction with the rainfall forcing and initial state control initiation and interaction of macropore flow with the soil matrix and thus ultimately export and redistribution of water from or within the control volume. In either case determining topology and connectivity (form) and understanding their implication for soil water transport (function) is seen as the "forefront of multiphase flow research" (Armstrong et al., 2016).

It is a long-standing vision in eco-hydrology to observe and characterize the form and function of all possible different flow paths in the subsurface. However, this is hindered by a lack of observation techniques which are capable of measuring and visualizing flow paths across the relevant range of scales in a continuous manner. In this study, we address the challenge of in situ observation, identification and characterization of flow-relevant structures through a series of complementary methods at the point, plot and hillslope scale.

1.2 Identification and characterization of flow-relevant structures in the subsurface

While heterogeneity is seen as a purely random variation of soil properties, organized heterogeneity implies a spatial covariance of these properties and connected flow paths. As such we define structure based on their functional implication in line with Gerke (2012) and others. While such structures can be classical macropores like earthworm burrows (Palm et al., 2012; Blouin et al., 2013; van Schaik et al., 2014), decayed root channels (Nadezhdina et al., 2010) or cracks and geogenic structures like voids in periglacial cover beds (Heller, 2012), we also attribute connected inter-aggregate pores to structure. They have in common that gravity induced preferential subsurface flow is facilitated through the directed drainage paths, partially bypassing large sections of the soil. Beven and Germann (1982) initiated a discussion about macropores and preferential flow and more recently

repeated that the topic is still not given the attention appropriate to its significance in all areas of soil and catchment hydrology (Beven and Germann, 2013).

Despite observation of fast responses through such macroporous networks, e.g., as tracer breakthroughs (Schotanus et al., 2012; Klaus et al., 2013) or in multi-modal reactions (Martínez-Carreras et al., 2016), it was shown that quick responses of catchments are often fed by pre-event water (Neal and Rosier, 1990; Jones et al., 2006) which is known as the "old water paradox" (Kirchner, 2003).

Due to limited direct observability of subsurface flow, most evidence is either inferred from integral responses or derived from model applications: in the field, a large spectrum of methods is applied to investigate subsurface connectivity (Bishop et al., 2015; Blume and van Meerveld, 2015) and to quantify preferential flow (Allaire et al., 2009). Dye staining has evolved as common practice since its first applications (presumably Bouma and Dekker, 1978) for a retrospective imaging of preferential flow paths. Even though Anderson et al. (2009) extended dye staining to the hillslope scale, the technique is usually limited to plot-scale applications. Another drawback is the requirement to excavate and thereby destroy the system, which prohibits analyses of function under variable forcing. Application of salt tracers in the vadose zone adds a quantitative measure, but at lower spatial resolution than dye staining. It also suffers from the a posteriori inference about the retention of the solutes.

Furthermore, breakthrough curves of precipitation or irrigation events at trenches or springs are commonly used (e.g., McDonnell et al., 1996; Tromp-van Meerveld and McDonnell, 2006; Bachmair and Weiler, 2014). In combination with fluorescent, salt and natural tracers they can provide quantitative information over the course of rapid flow events at this scale (e.g., Wienhöfer et al., 2009). However, such measurements can only capture spatially integral signals and require one to infer the form by the observed function.

So far, relatively few studies managed to actually in situ image spatially distributed subsurface flow paths at larger scales. On the one hand, applicability is also often technically limited to very small scales: Schlüter et al. (2016) examined multiphase flow with time-lapse X-ray microtomography in a sample of 4.2 mL. Koestel and Larsbo (2014) presented an X-ray tomography study with a sample of 258 mL undisturbed soil. Gerke (2012) analyzed the pore fractions in two samples of 785 mL undisturbed soil through a medical CT X-ray scanner. Wehrer and Slater (2015) report findings from tracer breakthrough experiments in laboratory lysimeters accompanied by 3-D time-lapse electrical resistivity tomography (ERT). Guo et al. (2014) conducted a multi-2-D time-lapse ground penetrating radar (GPR) survey to identify preferential flow structures in situ in a $2 \, \text{m}^2$ section of a hillslope. On the other hand, the lack of a unified theory of advective and diffusive soil water redistribution, mixing, storage and release (Beven and Germann, 2013) adds to unclearness about appropriate observation strategies.

Hydrological "standard approaches" attempt to explore parameters like soil layer depth, porosity and hydraulic conductivity based on distributed point-scale measurements. Also, state and flux monitoring most often consists of a set of point observations, e.g., of hydro-meteorological conditions and soil moisture. An appropriate sampling design is substantial for the statistical inference (e.g., de Gruijter et al., 2006). Thus there is also a conceptual issue arising from the fact that such samples necessarily integrate over sub-scale structures, such as inter-aggregate pore networks. At the same time such an integral may not necessarily allow inference about structures at the larger scale exceeding the support of the observation. When the respective sampled set and the subsurface setting is basically unknown, spatial scaling of soil moisture (Western and Blöschl, 1999) and other observed variables becomes problematic. In a "Special Section" on preferential flow, Gerke et al. (2010) highlighted that further analyses need to focus on the quantification of flow-relevant structures. They continue that experimentally non-invasive and imaging techniques are needed for research and model testing. We will take up these issues in the discussion section.

1.3 Hypotheses and overall aims of the study

The rationale of this study is to analyze insights into flow-relevant subsurface structures based on qualitative and quantitative measurements at the point, plot and hillslope scale. Specifically, we hypothesize that a combination of quantitative field methods and in situ imaging of subsurface response patterns with dye staining and time-lapse GPR provides the missing link between form of the flow structures and how their interactions determine rapid subsurface flow and thus function.

We test this hypothesis by addressing three main research questions.

Q1 What kind of information on sub-scale flow-relevant structures, their characteristics and their distribution can be inferred from a large set of direct point-scale measurements of soil hydraulic properties?

Q2 How do salt tracer data, dye tracer stains, soil moisture response patterns, and 3-D time-lapse GPR compare with respect to inference on vertical flow channels and apparent flow velocities at the plot scale?

Q3 How do methods and identified structures convey to the hillslope scale?

The study approaches the identification and characterization of flow-relevant subsurface structures as the aspect of *form*. The alternative starting point towards hillslope process understanding is taken in the companion study by Angermann et al. (2017, this issue) with the aspect of *function*.

2 Experimental approaches and study methods

The study at hand approached the topic on three complementary scales with a range of different methods: as a standard reference, results from auger exploration and in situ measurements of hydraulic conductivity and infiltration capacity were collected. They were extended with pedo-physical laboratory examination of 250 mL undisturbed ring samples for bulk density, porosity, texture, soil water retention characteristics, and saturated hydraulic conductivity. We then broadened the perspective to the plot scale with irrigation experiments accompanied by TDR (time domain reflectometry) measurements of soil moisture dynamics in a 1-D profile, 3-D time-lapse GPR imaging, and tracer recovery of dye, salt and stable isotopes. At the hillslope scale, 3-D GPR was used to identify flow-relevant structures in a static survey. For dynamic investigation, an irrigation experiment specifically designed to identify lateral flow structures was observed by a dense network of TDR soil moisture profiles and a series of trench-like 2-D time-lapse GPR transects.

2.1 Study site description

The study is situated in the headwaters of the Colpach River, a tributary of the Attert which has been investigated by several studies before (Pfister and Hoffmann, 2002; Hellebrand et al., 2011; Jackisch, 2015). Located at the southern edge of the schistose Ardennes Massif, the soils are characterized by eolian loess deposits and weathered schist debris. The hydrological setting of quick catchment reaction to precipitation especially during the non-vegetated season has been subject to some process hypotheses related to the periglacial deposit layers and flow at the bedrock interface (van den Bos et al., 2006; Fenicia et al., 2014; Wrede et al., 2015; Loritz et al., 2017). Our measurements and experiments focus on two forested hillslopes (mostly managed stands of beech, *Fagus sylvatica*, with mixed shrubs; some measurements took place in stands of spruce, *Picea abies*). The agriculturally used plateaus at the hilltops are not examined here. Figure 1 presents a map of the area and the location of the respective measurements and experiments.

2.2 Pedo-physical exploration

The soil physical exploration addressed our research question Q1 using an intentionally large set of hydrological and geophysical methods to survey the subsurface. The sampling is guided by a network of hydro-meteorological monitoring stations measuring all relevant fluxes and states in the atmospheric boundary layer and the subsurface (research project "Catchments As Organized Systems", Zehe et al., 2014).

2.2.1 Sampling design

Aligned with the monitoring stations, infiltration capacity and saturated hydraulic conductivity were measured. In or-
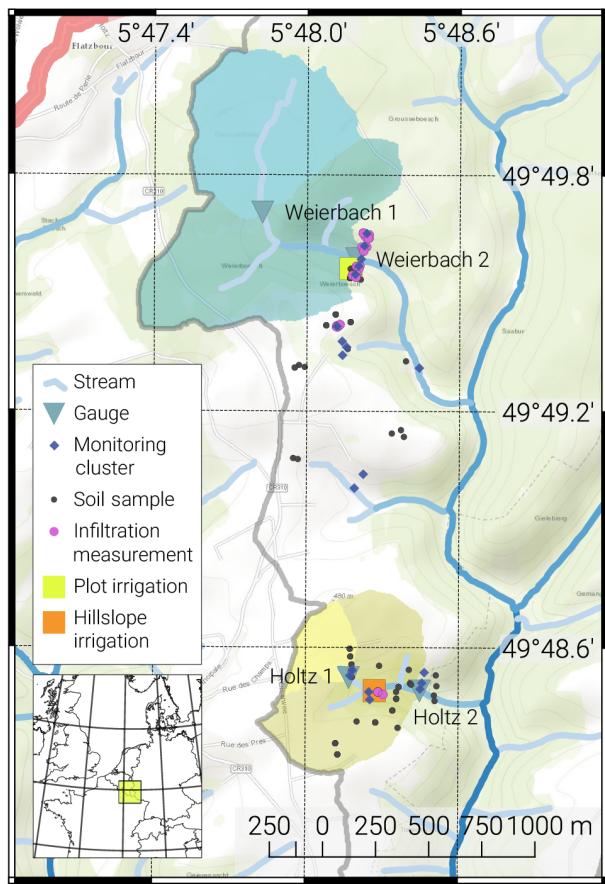
Figure 1. Map of the study sites in the upper Attert basin, Luxembourg.

der to address plot-scale (few meters) and hillslope-scale (a few hundred meters) heterogeneity, the design consisted of clustered sets of point measurements along two catenas plus one set at the site of the hillslope irrigation experiment presented in Sect. 2.4. A detailed map is included as Fig. B1 in the Appendix.

The distance between the clustered sets was 80–200 m. In each, three nested sets with a lag distance of 10–20 m along and perpendicular to the contour line were defined. In such a nested set at least one measurement of infiltration capacity and two profiles (laterally spaced 1 m) of saturated hydraulic conductivity in different depth levels were conducted. To complete the scale triplet (Bloschl and Sivapalan, 1995), the respective support is given in the description of each technique.

In addition to the point measurements, a series of percussion drilled profiles (drill head diameter of 4 cm) as 1-D profiles were drawn and 250 mL ring samples were taken within the top 0.6 m for laboratory analyses.

2.2.2 Exploration techniques

Infiltration capacity was measured at 40 points with a Hood Tension-Infiltrometer (IL-2700, UGT GmbH). It employs a tension chamber (12.4 cm radius) as infiltration water supply. Inside the chamber, a defined low negative pressure head is established, which allows a precise measurement of infiltration capacity at different tensions. Three to five tension levels between the 0 and 5.5 cm water column were applied at each spot.

In addition to infiltration capacity at the surface, we used a Compact Constant Head Permeameter (CHP, Ksat Inc.) for determination of saturated hydraulic conductivity in 32 borehole profiles with 3–7 depth levels of about 20 cm increments, with the lowest level at a depth where further hand-drilling was inhibited by stones. The permeameter establishes a constant water level (10.5 cm in our cases) above the bottom of a borehole (here 5 cm in diameter). The outflow is measured to calculate saturated hydraulic conductivity (Amoozegar, 1989).

The 63 undisturbed soil ring samples were analyzed for bulk density, porosity (assumed to be equal to saturated soil water content), soil water retention properties (Hyprop, UMS GmbH and WP4C Decagon Devices Inc.), saturated hydraulic conductivity (Ksat, UMS GmbH), and soil texture (ISO 11277, wet sieving and pipette method sedimentation).

2.3 Imaging and quantification of rapid flow in plot-scale irrigation experiments

In order to explore the network of flow-relevant structures and patterns of rapid subsurface flow, we conducted three plot-scale irrigation experiments. This relates to our second research question Q2. The general setup is very similar to the one described by Allroggen et al. (2015b), van Schaik (2009), Öhrström et al. (2004) and Kasteel et al. (2002). Marked on the map in Fig. 1, the three plots are located on a forested mid slope near gauge Weierbach 2 (see also Fig. B1).

2.3.1 Experimental design and multi-method approach

Three plots of $1\,\text{m}^2$ size were irrigated each for 1 h with an intensity of 50, 30, and $50\,\text{mm}\,\text{h}^{-1}$ on 30 October, 1 November, and 2 November 2013, respectively. The relatively high rates were chosen to activate the potential flow paths, thereby establishing connectivity. A layout of the experiment is presented in Fig. 2.

The irrigation was accomplished by spray irrigation (fullcone nozzle Spraying Systems Co.) using a wind-protection tent. Brilliant Blue dye tracer $(4\,\text{g}\,\text{L}^{-1})$ and bromide salt $(5\,\text{g}\,\text{L}^{-1}$ potassium bromide) were used for qualitative and quantitative reference, respectively.

In addition, temporal dynamics of soil moisture along a selected profile was monitored throughout the experiments
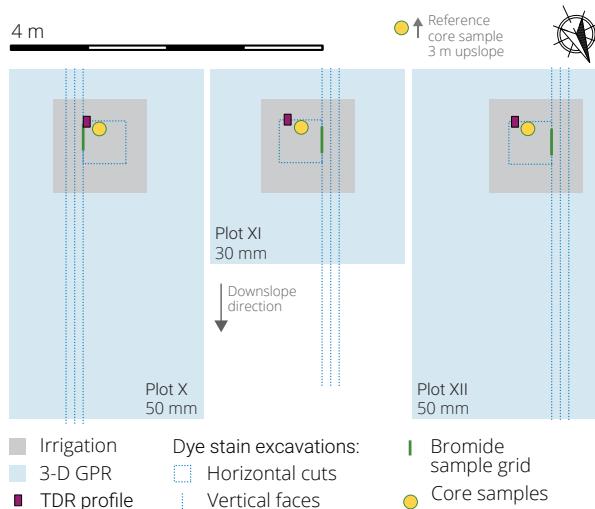
Figure 2. Plan view layout of the plot-scale irrigation experiments. Three irrigation plots (1 m², gray squares) are monitored by 3-D time-lapse GPR (blue rectangles) and TDR (soil moisture tube probe, red box). The plots are sampled for tracer recovery by percussion drilled core samples (yellow dot) and in a grid on the last of three vertical faces (dashed blue line). Moreover, dye stains are excavated at horizontal cuts in the center of the irrigation area (dashed blue square). A pre-irrigation reference for porewater stable isotope composition is sampled as a fourth core 3 m upslope.

through continuous TDR measurements in an access tube (Pico IPH, IMKO GmbH) down to 1.5 m depth and with a diameter of 4.2 cm. This technique is chosen to minimize the impact of sensor installation (percussion drilling and installation of the tubes from the surface) and to avoid interference with the GPR (sensor probe was removed during GPR measurements). The sensor measured an integral of about 1 L (depth increment of 18 cm, mean signal penetration of 5.5 cm). It was manually lowered in the tube to the respective depth for each reading. Each measurement took about 10 s. Hence the whole procedure added up to 4–10 min per profile record. The procedure was continuously repeated until 1.5 h after irrigation onset in line with the findings of Germann and al Hagrey (2008) and Germann and Karlen (2016). They propose that film flow in soil structures disperses into the matrix after 1.5 times the duration of a constant plot irrigation.

Two hours after the end of each irrigation, a percussion drilled soil core was taken (drill head diameter of 8 cm) and sampled in 5 cm depth increments down to 1 m. The plot was excavated 24 h after irrigation for vertical and horizontal recovery of Brilliant Blue stains. This was done by successive digging of three vertical faces into the plot (aligned with the slope line, 0.1 m distance starting from the lateral edge) and five to seven horizontal cuts in different depth levels down to the first deposit layer (0.5 × 0.5 m² in the center of the plot). On the third vertical face in the center of the plot core samples of 66 mL soil were taken in a 5 cm grid with 5 columns

and 14–21 rows. In order to minimize time lags in the 70–105 individual samples, a quick-sampler (see Appendix A) was developed, allowing for precise and nearly undisturbed sampling.

2.3.2 Bromide recovery and stable isotope analysis

All samples were analyzed for bromide (Br⁻). This was done by oven drying the samples and consecutively suspending them in 150 mL de-ionized water (72 h in an overhead shaker at nine rotations per minute). The samples were then left 4 days for sedimentation to exfiltrate the excess through (a) filtration paper (5–13 μm) and (b) 0.45 μm PP microfilter. The extracts were analyzed in an ion chromatograph (Metrohm 790 Personal IC) with an anion separation column (Metrosep A Supp 4 – 250/4.0) for Br⁻ concentration.

A recovery coefficient (RC) is calculated as proportion of recovered mass of Br⁻ in the soil samples scaled to the total irrigated area times the depth of the lowest sample. Through this we neglect lateral flow from the irrigation spot and further percolation in the calculation. We also assume the samples to be representative for the whole affected soil volume.

Prior to the bromide analysis, the percussion drilled soil core samples were also analyzed for their stable isotopic composition (δ^{18}O and δ^2H) of the porewater. See Appendix D for details and results, which are given in comparison to the bromide recovery.

2.3.3 Calculation of apparent vertical flow velocity

The quantitative measurements allow one to infer apparent vertical flow velocity along the profiles. For bromide we employ a cumulative curve method (Leibundgut et al., 2011). The distribution of the advective velocity v_{advect} is set to the depth distribution of the tracer concentration at the time of fixation t_{fix}. For the profile we assume apparent velocities:

$$v = z/t_{fix}. \tag{1}$$

Relating to our third research question Q3, they are projected to the recovered distribution of tracer concentration:

$$\Phi(v_{advect,z}) = c_{tracer,z} / \sum_{z=0}^{z_{max}} c_{tracer}, \tag{2}$$

where z is depth and Φ is the cumulative distribution function. Obviously, the estimated travel velocity distribution depends strongly on the selection of t_{fix} somewhere between irrigation and excavation. This can scale v several orders of magnitude. Again, the reference of 1.5 times the irrigation duration is chosen (Germann and Karlen, 2016). For Br⁻ in the sampled grids, each column was treated as an individual 1-D profile. The calculation further assumes full tracer recovery.

2.3.4 Analysis of soil moisture responses

The individual TDR soil moisture measurements (θ) were projected to a regular grid of 0.1 m depth increments and 10 min time increments for visualization of changes compared to the initial records. As an alternative and independent estimate of vertical response velocities (research question Q3), we calculated the distribution of first exceedance of soil moisture by $\geq 2\,\%$ vol in each depth level z:

$$v_{\text{response}} = z/t_{\Delta\theta \geq 0.02}. \tag{3}$$

For this the un-interpolated measurements were used.

2.3.5 3-D time-lapse GPR

GPR is known as geophysical imaging technique with high spatial resolution (Huisman et al., 2003; Binley et al., 2015). Applied at the shallow subsurface it has been proven as potential means to locate and characterize soil layers and subsurface structures (Holden, 2004; Gormally et al., 2011; Steelman et al., 2012; Klenk et al., 2015). GPR is also capable to monitor subsurface fluid migration in time-lapse approaches (Birken and Versteeg, 2000; Trinks et al., 2001). Our experiments were monitored by 3-D time-lapse GPR measurements as described by Allroggen et al. (2015b). We employed a PulseEKKO Pro GPR system (Sensors and Software Inc.) equipped with 500 MHz shielded antennas with constant offset of 0.18 m. Sampling interval was set to 0.1 ns, recording a total trace length of 100 ns in 8 internal stacks. Since precise positioning and accurate repeatability are key requirements, we used a kinematic survey approach relying on an automatic-tracking total station (Leica Geosystems AG, providing sub-centimeter coordinates) in combination with a portable measuring platform (Allroggen et al., 2015b).

Using this setup, we acquired one 3-D GPR data cube before irrigation, one directly after the end of irrigation, and a last one about 20 h after irrigation for each plot. One survey took about 45 min. Allroggen and Tronicke (2016) have shown that a pixel-to-pixel comparison of the radar amplitudes (A) is not suitable for analyzing time-lapse GPR data in the presence of limited repeatability and noisy data. They propose a structural similarity attribute inspired by (Wang et al., 2004) calculated in a moving window. It normalizes the cross-correlation $c_{x,y}$ of the residuals ($A - \mu_A$) of two different acquisition times (x, y) by the product of their standard deviations (σ_A). They further introduced a as 10% of the maximum amplitude to avoid numerical instabilities with near-zero σ values:

$$S_{\text{struct}}(x, y) = \frac{c_{x,y} + a}{\sigma_{A_x} \sigma_{A_y} + a}. \tag{4}$$

In our study we calculate the structural similarity attribute S_{struct} of the pre-irrigation reference and the two post-irrigation records using a local Gaussian window of 2.5 ns along the vertical axis and 0.1 m along the horizontal axis.

The attribute ranges between 1 and -1, with 1 being highly similar and -1 referring to most dissimilar. Points of low similarity indicate deviations that arise from changes in dielectric permittivity which likely reflect changes in local soil water content.

As an additional estimate of vertical response velocities, the same approach as for the soil moisture responses (Sect. 2.3.4) was employed with a threshold of the similarity attribute of zero between pre- and post-irrigation records.

2.4 Lateral subsurface flow paths in the hillslope

In order to examine the characteristics of flow-relevant structures and the periglacial deposit layers at the hillslope scale, we conducted an experiment on 21 June 2013 at a close-by hillslope. The experiment was specifically designed to explore the response in lateral preferential flow paths and to replicate the plot-scale experiments without tracer application. The site had to be chosen for facilitation reasons (permissions, accessibility, collaboration within the CAOS research project). With reference to its hydrological responses (companion paper Angermann et al., 2017, this issue), vegetation, slope, soils and hydraulic properties, we consider the hillslopes to be very similar.

2.4.1 3-D GPR survey of the hillslope

As an additional reference to the soil core profiles, a 3-D GPR survey of the hillslope was conducted prior to the natural event and the irrigation. The GPR data processing relies on a standard processing scheme including bandpass filtering, zero time correction, envelope-based automatic scaling, gridding to a regular 0.03 m by 0.1 m grid, inline fk-filtering and a 3-D topographic migration approach as presented by Allroggen et al. (2015a), using an appropriate constant velocity of 0.07 m ns^{-1}.

For structural analysis, the processed data are imported into the OpenDtect software (dGB Earth Sciences). Under heterogenous soil conditions the derived data cube is dominated by complex reflection patterns which prohibit a classical structural analysis based on picking reflectors (as done in a study with a similar cope but different setting by Gormally et al., 2011). Therefore, we support our interpretation of the 3-D GPR data and picking of potential flow-relevant horizons by a dip-corrected semblance attribute. The attribute calculates the spatial coherency and highlights areas of coherent reflections (Marfurt et al., 1998). Low semblance indicates more complex reflection patterns caused by high internal heterogeneity, possibly influencing the subsurface flow regime.

2.4.2 Experimental design

The experimental site is located at the lower part of a north facing hillslope. Vegetation is dominated by mixed beech forest. However, the experimental site is placed in an area with no major trees. Except for few young trees at the downhill

monitoring area, all shrubs were carefully removed from the experimental site to accomplish GPR measurements and allow for undisturbed and homogeneous irrigation. The topographic gradient is about 14°.
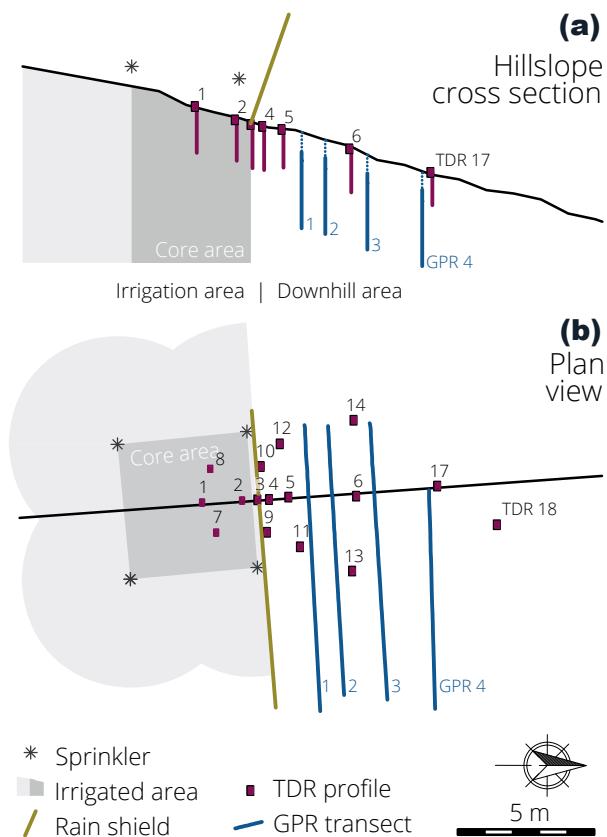

Figure 3. Layout of the hillslope-scale irrigation experiment as vertical view **(a)** and plan view **(b)**. The hillslope is divided into an irrigation area and a downhill area by a rain shield. Sixteen access tubes for TDR measurements of soil moisture profiles are arranged in three diverting transects. Parallel to the contour lines, four transects of 2-D time-lapse GPR are recorded.

The experimental layout is given in Fig. 3. Irrigation intensity, the duration of the experiment and the spacing of the observation profiles have been decided based on a priori modeling scenarios as described in Appendix C. The experiment was preceded by two strong storm events of 43 mm in total on 20 June. The events ended 20 h before irrigation onset. The irrigation of 141 mm in 4.5 h was fed from stream water and was realized by four circular sprinklers (Wobbler, Senninger Irrigation Inc.) arranged to overlap at a 5 m by 5 m core area with relatively homogeneous intensity. While boundary effects were mitigated by an irrigated buffer zone of about 4 m at the uphill and lateral borders of the core area, the downhill boundary was defined by a rain shield. This established a sharp transition to the non-irrigated area below. Water collected by the rain shield was routed off the experimental site. Irrigation was monitored by a flow meter to measure the ab-

solute water input, one tipping bucket to monitor the temporal variability, and 42 mini rain collectors evenly distributed across the core area to check spatial heterogeneity of the intensity.

Moreover, a surface runoff collector was installed across 2 m of the lower boundary of the core area. It was built from a plastic sheet installed approximately 1 cm below the interface between litter layer and Ah horizon of the soil profile. At the downhill end of the sheet, the water was captured by a buried and covered gutter. An in-ground tube was attached to the deepest point of the gutter to conduct the water to a tipping bucket downhill of the investigated area. The tube had been filled with water prior to the experiment to ensure an immediate reaction to the occurrence of surface runoff.

We monitored soil moisture dynamics in a setup of 16 access tubes with 3 manual TDR probes like in the plot-scale experiments (Imko GmbH, two with 12 cm integration depth and one with 18 cm). Measurements required manual positioning of the sensor probes for each reading. We continuously recorded the states in all tubes in 10 cm depth increments, realizing revisiting intervals of 5–20 min. The tubes were installed to reach to a depth of about 1.7 m. The layout consisted of three diverging transects with four TDR profiles in the lower half of the core area, the highest density of profiles just downhill from the rain shield, and the furthest profile about 9 m downhill.

Four 2-D time-lapse GPR transects were treated as *GPR-inferred, non-invasive trenches* parallel to the contour lines located 2, 3, 5, and 7 m downhill from the rain shield. Here, the GPR acquisition unit was equipped with shielded 250 MHz antennas. The data were recorded using a constant offset of 0.38 m, a sampling interval of 0.2 ns and a time window of 250 ns. Wooden guides and the automatic tracking total station guaranteed accurate and repeatable positioning.

2.4.3 Analysis of TDR data

In order to synchronize the almost 5000 individual TDR soil moisture records to a regular grid in time and depth, interpolation and resampling were required. To do so, we generated an intermediate grid of high data density onto which linearly interpolated versions of the time series of each profile were projected. We then resampled from this intermediate grid to derive a synchronized version of the records at 0.1 m depth and 15 min time increments. With this the spatial aggregation remains below the integration length of the TDR probes.

The temporal resampling and the therefore necessary linear interpolation is close to the acquisition timing of one profile (4–10 min each). Since the correlation length of distributed soil moisture observations is rather short and because we explicitly aim to analyze the responses of preferential flow structures, the issue of interpolation needs special attention and will be discussed in Sect. 4.2.1.

All soil moisture measurements are converted to changes in soil moisture referenced to the state previous to irrigation

onset to identify activated flow paths. Lateral interpolation between different TDR profiles over distances of about 1 m and above is unfeasible. Soil moisture as extensive state variable is discontinuous at interfaces. The found subsurface setting does not exhibit any isotropic continuum required for such interpolations.

As in the plot irrigation experiments, vertical response velocities are calculated for the TDR profiles in the core area. The calculation of lateral response velocities is given in the companion study (Angermann et al., 2017, this issue).

2.4.4 GPR transects and structural similarity attribute interpretation

The 2-D time-lapse GPR data are derived from nine repeated recordings along the four vertical GPR transects. Each record is processed after a standard processing scheme of bandpass filtering, zero time correction, exponential amplitude preserving scaling, inline fk-filtering, topographic migration with constant velocity ($0.07 \, \mathrm{m \, ns^{-1}}$), and consecutive gridding to a 2-D transect with regular trace-spacing of 0.02 m.

Most time-lapse GPR data analyses are based on calculating trace-to-trace differences (Birken and Versteeg, 2000; Trinks et al., 2001) or picking and comparison of selected reflection events in the individual time-lapse transects (Allroggen et al., 2015b; Haarder et al., 2011; Truss et al., 2007). Like in the 3-D time-lapse GPR applications, the radargrams in the young, highly heterogeneous soils do not exhibit explicit reflectors as suitable references. In addition, the limited repeatability of the measurements and the desired identification of lateral flow structures require an alternative approach.

Like for the plot-scale experiments, we use the time-lapse structural similarity attribute (Allroggen and Tronicke, 2016, and Sect. 2.3.5). It is calculated using a local Gaussian window of 2 ns along the vertical axis and 0.06 m along the horizontal axis.

Due to the presence of remaining event water from the preceding storm event (Angermann et al., 2017, this issue), all measurements are referenced to the last acquisition time approximately 23 h after irrigation start and about 19 h after irrigation. The resulting structural similarity attribute images are used as a qualitative indicator of relative deviations from the reference state.

2.4.5 Discriminating the natural storm event and the irrigation experiment

The experiment was preceded by two strong storm events of 43 mm in total on 20 June 2013. In reference to the gauge reaction the experiment was conducted shortly before the second peak of the runoff reaction to the preceding storm events (see Fig. 5 in Angermann et al., 2017, for details). Accordingly, the structural similarity attributes, which compare the distributed states at the respective acquisition time to the last record, identify responses to both drivers, the natural storm

event and the irrigation experiment. To discriminate the signals we analyze the dynamics of each pixel in the GPR transects over time. The first two structural similarity attribute transects 7.5 h before and directly at irrigation start are attributed to the natural event and show an increasing structural similarity (towards accordance with the reference state). Once the attribute value of a pixel decreases again (increasing deviation from the reference state) it is attributed to the irrigation. For stability reasons, a threshold of 0.15 was introduced for the attribute to be exceeded to detect changes. This is discussed in Appendix F.

3 Results

3.1 Soil physical exploration

3.1.1 Point samples show high heterogeneity

The in situ point measurements of infiltration capacity and saturated hydraulic conductivity showed high variability without clear relationships with simple morphological descriptors like depth, hillslope position or topographic flow gradient (details given in Fig. B1). Infiltration capacity ranged from 5×10^{-5} to $5 \times 10^{-3} \, \mathrm{m \, s^{-1}}$. The values for saturated hydraulic conductivity ranged from 1×10^{-8} to $1 \times 10^{-3} \, \mathrm{m \, s^{-1}}$ and even exceeded the measuring range of the constant head permeameter. Only at the site of the hillslope-scale experiment was a pattern of elevated conductivity at about 0.6 m depth found. The strong heterogeneity and large spread of values was also depicted in the analyses of undisturbed soil samples (Fig. 4).

On average the area is dominated by silty soils (see also Juilleret et al., 2011). This was corroborated by texture analyses and the mean retention characteristics. However, the measurements of saturated hydraulic conductivity are on average 2 orders of magnitude larger than what might be expected for these soils given their texture (Schaap et al., 2001). Also, porosity clearly exceeds the expected values, while bulk density is smaller than expected (compare Fig. 4, middle row of the panels). All measurements exhibited a large spread of values which does not correlate well with simple morphological variables like depth or hillslope position (Fig. 4, bottom row). The high hydraulic conductivity and large porosity is maybe explained by aggregation of fine silty material in conjunction with a network of rapidly draining inter-aggregate pores.

3.1.2 Soil core profile snapshots

The soil core profiles (Fig. 5) generally confirmed the presence of the periglacial slope deposits by gravel bands but also showed a high degree of heterogeneity. The thickness of the horizons was variable, with a humidified mineral A-horizon of up to 0.3 m. The gravel content gradually increased over depth in the Bw-horizon and further increased in the C-horizon, starting between 0.4 and 1.1 m depth. Be-
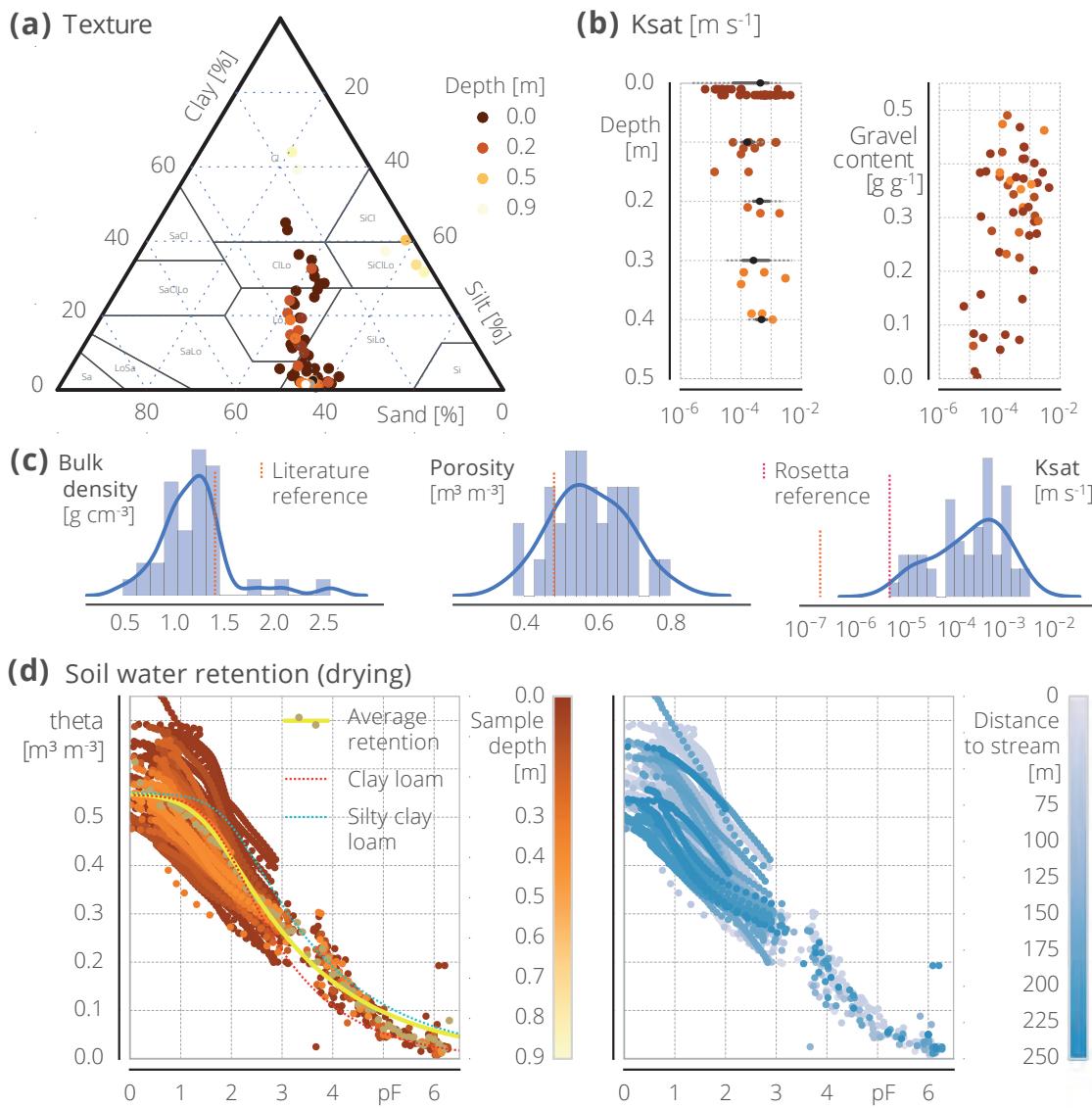
Figure 4. Results of laboratory analyses of 63 undisturbed 250 mL ring samples. **(a)** Soil texture analyzed with wet sieving and sedimentation (pipette method). Saturated hydraulic conductivity (Ksat) measured with the Ksat apparatus and plotted against sample depth and gravel content. Black box plots of the respective depth increment. **(b)** Histograms and kernel density estimate of measured bulk density, porosity and Ksat. Reference as the mean silty loam value from the literature (Hillel, 1980; Rawls et al., 1982; Carsel and Parrish, 1988). Rosetta (Schaap et al., 2001) reference based on mean values of samples (15.7, 47.9, 36.4 % sand, silt, clay and BD 1.1 g cm^{-3}). **(c)** Soil water retention relation measured with the Hyprop and the WP4C apparatus with the average retention estimate (respective mean of each 0.05 pF-bin and fitted van Genuchten model) and literature references (Carsel and Parrish, 1988) scaled to measured average θ_s and θ_r.

low the depth of 0.5 m, scattered layers of weathered rock with usually horizontal orientation were found in some soil cores. Percussion drilling was often inhibited at a depth between 1.5 and 2.0 m (lower end of the bars in Fig. 5) due to even higher stone content with a more and more vertical orientation of the weathered rocks. In core 7, concretions of iron and manganese oxides were found at depths between 1.6 and 1.9 m below ground, indicating hydromorphic conditions.

Based on these standard techniques the overall setting of a heterogeneous silty soil deviating from expected low hy-

draulic conductivity was revealed. So far gained insight is limited to the general existence of periglacial deposit layers (high gravel content in soil profiles), rapid flow paths (hydraulic conductivity several orders of magnitude above literature references), and some integral retention properties. However, details about its spatial organization and the detection of specific and potentially continuous structures remained obscured by high heterogeneity.
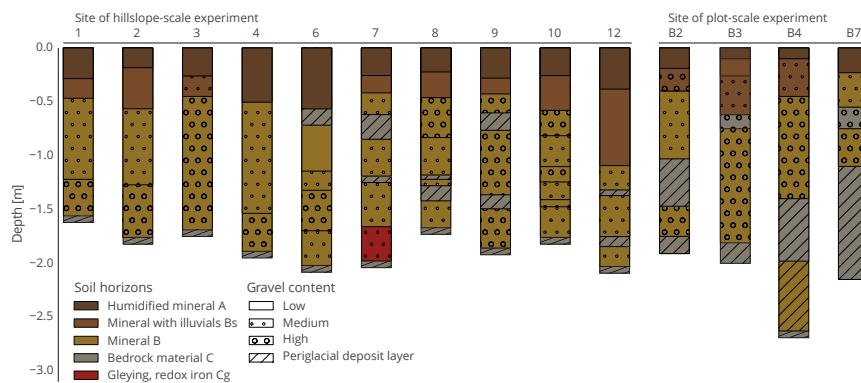
Figure 5. Soil core profiles from the upper Colpach River basin. See Figs. 3 and B1 for positions. Bar depth is the maximum drilling depth of the cores restricted by large stones or bedrock.

3.2 Plot-scale flow path activation and vertical velocities

3.2.1 Irregular patterns of dye stains

In the plot-scale tracer experiments the Brilliant Blue dye stains identified patchy infiltration patterns partially bypassing large sections of the soil without clear traces of the actual flow path (Fig. 6). For all experiments stained patches were found down to the periglacial deposit layer at 0.6–0.8 m depth. During the excavation apparently isolated dye traces were recovered even several meters downhill from the irrigation spot (4 m downslope, 1 m deep). The stains did not reveal a network of large macropores but an irregular mesh of connected inter-aggregate voids. This is in line with the observed hydraulic capacity (Fig. 4).

3.2.2 Bromide breakthrough to the periglacial deposit layer

The connectedness and large transport capacity of this network of inter-aggregate pores is corroborated by the distributions of bromide tracer recovery (Fig. 7, top row). All plots suggested a relatively strong response at a depth of approximately 0.6 m. This depth correlates with the upper boundary of the first layer of periglacial deposits found in core profile B3 (Fig. 5) and in the excavated soil profiles. This response is contrasted by low bromide concentration at a shallower depth. Even plot XI, where only 30 mm were applied, showed the same pattern with a clear breakthrough to the periglacial deposit layer.

At plot XII we found a stronger interaction with the soil matrix, which led to more dye staining and a higher bromide recovery. Overall, tracer recovery was incomplete (0.45, 0.38, and 0.83 for plots X to XII, respectively) and even declined when including the core samples (0.24, 0.3, 0.63) once more, pointing to strongly irregular soil water redistribution.
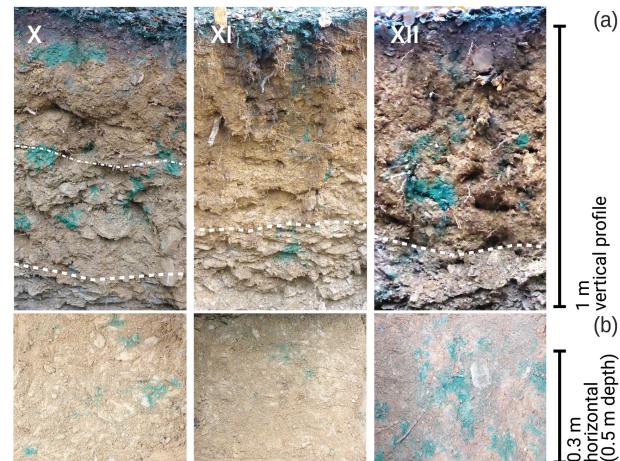

Figure 6. Recovered dye patterns in plot irrigation experiments. **(a)** Excavated vertical faces, **(b)** horizontal cuts in 0.5 m depth. Dashed lines indicate level of periglacial deposit layer.

3.2.3 Quick soil moisture response in greater depth

The observed soil moisture changes (Fig. 7, bottom row) corroborated the results from the tracer data. Especially at plot X and XII we found a relatively quick and strong response at 0.7 and 0.5 m depth, respectively. This even preceded soil moisture changes in shallower layers in plot X. Hence the records highlighted an important characteristic of the identified flow-relevant structures, although the signal had a much lower spatial resolution than the tracer results. In contrast to the recovered tracers, we did not observe significant changes in soil moisture in plot XI. This can be explained by its position offset from the main flow field (Fig. D2 in Appendix D).

3.2.4 3-D view on soil water redistribution

The structural similarity attribute of the 3-D time-lapse GPR measurements provided qualitative information of changes
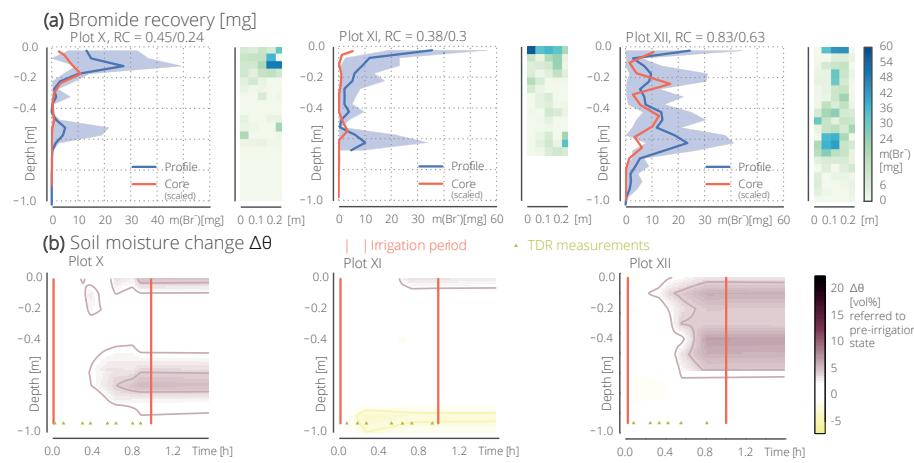
Figure 7. Results from plot-scale irrigation experiments with 50, 30, and 50 mm spray irrigation for 1 h. **(a)** Recovered bromide mass profiles and grids (5 × 5 cm). Blue line as mean and shaded area between min/max for each depth of the sampling grid. Orange line is the mass recovered in drilled profile samples (scaled to the same volume reference). Recovery coefficient (RC) calculated for the profile samples (first value) and the profile and core samples (second value). **(b)** Observed soil moisture change referenced to the first measurement shortly before onset of the irrigation. Individual measurement times marked with triangles.

in soil moisture in a spatial context. At all plots the response patterns of low structural similarity pointed out quick vertical flow to a depth of 80 ns or about 1.4 m within 1.5 h after irrigation start (Fig. 8, and Figs. D2 and D3 in Appendix D). Also here, strongest deviations were recorded in the mid horizon between 40 and 60 ns two way travel time (TWT) corresponding to approximately 0.7 to 1 m depth. The top horizon between 20 and 40 ns (0.35–0.7 m) had comparably high similarity. Measurements above that depth were technically not possible. Patches of low structural similarity until 20.5 h after irrigation start suggested further lateral redistribution in the later course of the experiment at plot X. At plot XI with 30 mm irrigation further vertical transport with a slight lateral component was recorded. Plot XII had a very high similarity between the first and third acquisitions. This is a sign of stronger macropore–matrix interaction and dispersive redistribution.

The contrasting attribute distributions over time and comparing plots X and XII not only revealed diverse patterns. It also highlighted the qualitative nature of the analytical method of the GPR data. Although visual interpretation of the radargrams (top rows in Figs. 8, D2 and D3) is very difficult, they show how the structural similarity attribute highlighted areas where radar patterns changed. Due to the complex reflection energy patterns it is not suitable to trace individual reflectors. This prevents a quantitative interpretation as shown by Allroggen et al. (2015b).

For the identification of structures, the results did not exhibit specific macropores like the dye stains, but areas of response to the irrigation. Nevertheless, the patchy characteristic of the found response patterns was very similar to that of Brilliant Blue.

3.2.5 Derivation of vertical flow velocities

Based on all applied techniques, hydraulic conductivity and apparent vertical flow velocities were calculated (kernel density estimates plotted in Fig. 9). The many point-scale measurements (left panel based on 63 ring samples, 40 infiltrometer points, 102 individual permeameter measurements) resulted in disagreeing distributions stretching across a large spectrum of flow velocities. The reason for this spread stems from the fact that the measurements consist of matrix flow and flow in structures. The response-related methods of the irrigation experiments were in much better accordance because they all relate to the same processes. They revealed an apparent vertical velocity of $1 \times 10^{-3.5}$ m s^{-1} (Fig. 9b, based on bromide recovery with an estimated time of fixation (t_{fix}) after 1.5 h, first excess of TDR recorded soil moisture $\geq 2\,\%$ vol, and GPR structural similarity attributes below zero between pre-irrigation and the first post-irrigation records).

All results ranged several orders of magnitude above the literature reference of 2.5×10^{-7} m s^{-1} (mean of reported values for silt and silty loam – Hillel, 1980; Rawls et al., 1982; Carsel and Parrish, 1988) and the Rosetta value of 6.2×10^{-6} m s^{-1} (Schaap et al., 2001). The role of inter-aggregate pores facilitating this quick redistribution even through comparably small voids without noticeable dye staining was corroborated.

3.3 Hillslope-scale detection of lateral flow paths

3.3.1 3-D GPR survey suggests a fragmentary layer

The 3-D GPR survey at the site of the hillslope experiment identified fragmented structures at about 1.5 m depth
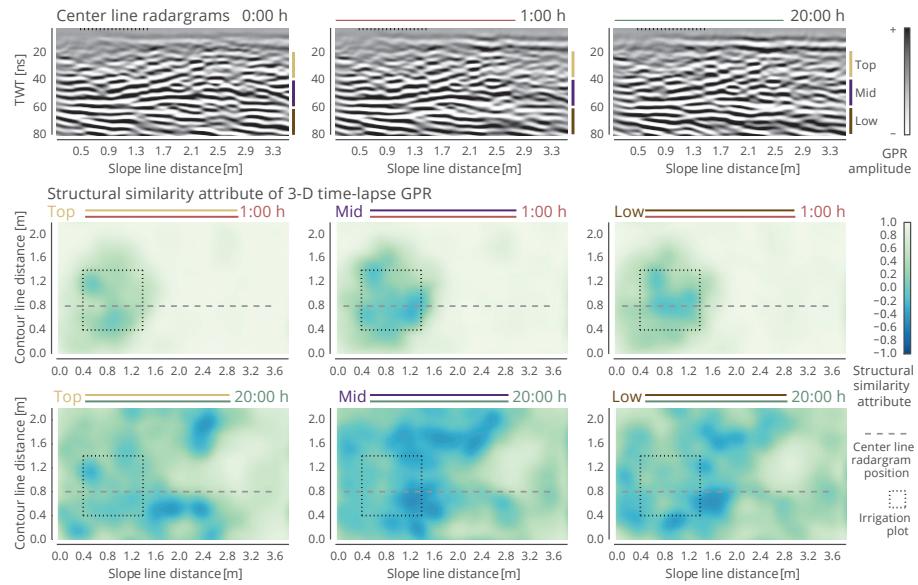
Figure 8. Time-lapse 3-D GPR of irrigation experiment at plot X. Center line radargrams at the marked transect (gray dashed line in lower panels) for the three acquisition times (before 0:00 h, directly after irrigation 1:00 h, 20:00 h after irrigation) are given in the top row. Two way travel time (TWT) is given as original depth reference. The structural similarity attribute of the 3-D data cube is given in three different depth layers (top 20–40 ns, mid 40–60 ns, low 60–80 ns) in the lower panels. The irrigation plot is marked by a black dashed box/line. Slope line distance is increasing downslope.
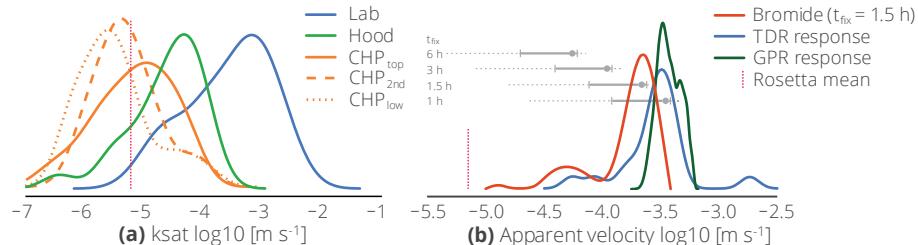
Figure 9. Saturated hydraulic conductivity and apparent vertical flow velocity kernel density estimates. **(a)** Point-scale measurement results (Lab: Ksat apparatus; Hood: Hood Tension-Infiltrometer; CHP: Constant Head Permeameter in different depth levels); **(b)** results from plot-scale irrigation experiments. Vertical gray lines are box plots of velocity distribution based on different t_{fix} estimates for bromide. Rosetta (Schaap et al., 2001) reference based on mean values of ring samples (15.7, 47.9, 36.4 % sand, silt, clay and BD 1.1 g cm^{-3}).

(Fig. 10). This is in accordance with the soil core profile depth (Fig. 5). Especially profile 7 suggested an impermeable layer just below that depth. Although a potential structure can be identified, it remains unclear to which degree this area of high spatial inhomogeneity in terms of radar reflection characteristics is flow-relevant, unless a reaction to an event is observed.

3.3.2 Hillslope responses

The results of the hillslope-scale irrigation experiment can be distinguished into the core area observations with TDR profiles only and observations at the downhill monitoring area, including TDR profiles as well as 2-D GPR transects. The change in soil moisture at the core area (TDR 2 and 8

in Fig. 11) was very much in line with the findings from the plot-scale experiments. Given sufficient irrigation, both experiments showed a quick and clear response at greater depth, even before intermediate layers responded. While the patterns were similar, the signal was much stronger during the hillslope-scale experiment, which is due to the higher irrigation amount and duration. The calculated apparent vertical response velocities (lower right panel in Fig. 11) had a wider spread towards the faster end but ranged around the values identified in the plot irrigation experiments. The downhill profiles (e.g., TDR 9 and 11 in Fig. 11) showed a more diverse response. With greater distance to the core area the reaction was more and more limited to single depth levels. But since the depth levels and responses were highly diverse, it remains rather ambiguous to determine their lateral con-
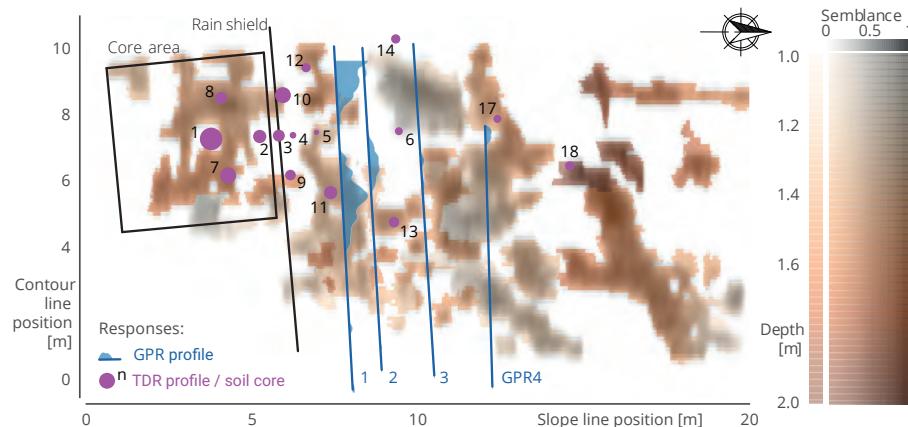
Figure 10. Potential subsurface structures from 3-D GPR survey and setup of hillslope experiment. Structure identification guided by the dip corrected semblance attribute. Depth estimated based on mean measured effective radar velocity in soil of $0.07\,\mathrm{m\,ns^{-1}}$. Summary of the hillslope experiment given by locations of TDR profile tubes (purple, also location of respective soil cores in Fig. 5) and GPR transects (blue). Dot size of TDR scaled to maximum of observed change in soil moisture. Along GPR transects lateral marginals of the structural similarity attribute as proxy for recorded advection. Note that the picked potential subsurface structures are located in different depth (white to black) and that variations in spatial contrast can be seen in the semblance attribute (white to orange). Where more than one horizon has been identified the top one is plotted.

nection. Overall changes in soil moisture as a maximum at each TDR profile did not corroborate the potential subsurface structures identified in the 3-D GPR survey (compare identified potential structures with dot sizes in Fig. 10). The full set of profiles is reported in the companion study (Anger-mann et al., 2017, this issue).

The four successive GPR transects across the downhill monitoring area provided spatially distributed images of hillslope-scale flow patterns and boundary fluxes. The structural similarity attribute of storm event water (green) and irrigation water (blue) revealed distinct, heterogeneously distributed patterns (Fig. 12) pointing to discrete connected flow paths instead of an irregular network of inter-aggregate pores. The measurements suggested that lateral flow takes place in a very diverse network with very low similarity between the transects. Moreover, the responses to the irrigation decayed with distance to the core area.

The patches which reacted to the storm event are mostly different ones than the structures used to drain the irrigation water. Apparently the irrigation experiment initiated flow in more shallow structures (compare transect 1 irrigation reaction with transect 3 storm water in Fig. 12). Areas of high temporal dynamics of the similarity attribute were identified as regions of such flow-relevant structures (Fig. 12, bottom row). Note that the last recorded difference 18 to 23 h after irrigation start (not shown) exhibited high similarity in all profiles. The mean of the attribute was between 0.93 and 0.96 and standard deviation between 0.076 and 0.048 for GPR transects 1 and 3, respectively. Apparently, the system had reached a steady state without much further change in soil moisture (see Appendix F for more details).

The patchy structures at the transects highlighted the irregularly distributed nature of lateral preferential flow paths which was similarly observed in the plot experiments. Although some areas exert a higher density of reacting flow paths than others, no continuous patterns could be specified throughout the hillslope. We also saw a decay of the signal strength and areal share with distance from the core area. As the patterns from transect 1 did not simply propagate further downslope, the flow paths must be tortuous and leaky. Hence inferring the configuration of the connection between the four transects in the downhill direction is not feasible. A comparison of the suggested structures of the 3-D GPR survey to the overall response to irrigation recorded at the GPR transects did not correlate well (compare identified potential structures with the reaction summary at GPR transects in Fig. 10).

4 Discussion

4.1 Identification of flow-relevant structures across scales

Our results have shown that the silty soils coincide with high porosities and high hydraulic conductivity at the point scale. Such a coincidence is not what is expected for cohesive, fine textured soils and can be explained by a setting of aggregated fine material in conjunction with a network of inter-aggregate pores. With respect to our research question Q1, the pedo-physical analyses initiated the recognition of these sub-scale structures. However, neither their position nor their general setup can be identified based on point observations
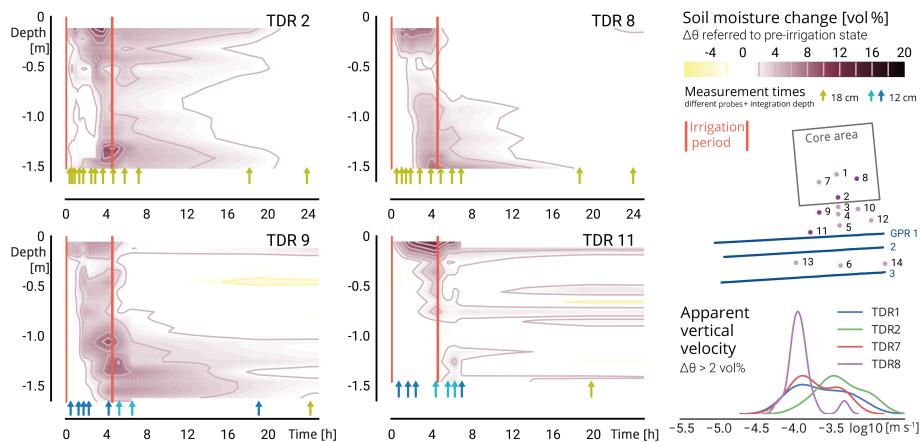
Figure 11. Development of soil moisture in TDR profiles during and after the hillslope irrigation experiment. Exemplary transect with changes referenced to pre-irrigation conditions and attributed to irrigation water. Time is given in hours after irrigation start. Individual measurements and probe reference marked with triangles. More data and explanation in Angermann et al. (2017), this issue. Right bottom: apparent vertical flow velocity as first excess of $\Delta\theta \geq 2\,\%$ vol at core area profiles.

because of its scale below the support of the measurements. Vice versa, methods at the next scale do not provide information about porosity and bulk density.

Irrigation experiments at the plot scale visualized that a network of these inter-aggregate voids connects the surface to the periglacial deposit layer and is responsible for highly diverse soil water redistribution. These structures are different from what we usually expect (cracks, worm burrows, roots channels) at this scale. This could be depicted from dye tracer stains (Fig. 6), which still have the highest spatial resolution on the cost of a lack of temporal insight. It requires strong assumptions about macropore–matrix interaction, time of fixation and dye supply, retention and recoverability. Despite all uncertainty about what process caused staining, the technique allows to identify the structures activated by irrigation and to infer much about their setting where dye has been retained. Although dye stains are closely related to actual flow and thus function, they only reveal the potential pathways and thus form as the actual processes and timing remain unknown. When irrigation intensity and irrigation amount ranges near the hydraulic capacity of the macropore network while still avoiding ponding or macropore clogging, the entire network of flow-relevant structures is marked. 3-D time-lapse GPR has proven to be capable to detect similar response patterns. However, the spatial and temporal resolution of the method is still insufficient to detect the flow-relevant inter-aggregate voids marked by dye stains. Some of the structures have not even been traced with dye, nor could GPR identify them. Nevertheless, the overall characteristics of the structures as patchy responses are depicted well and in a non-invasive, spatially continuous manner. Thus most of the point-sampling related issues (Sects. 3.1.1 and 4.4) are resolved. Regarding research question Q2, the visualization of flow structures based on re-

sponses to irrigation succeeded at the plot scale. They are in good coherence with the quantitative findings from salt tracers, stable isotopes and soil moisture dynamics. Interestingly, the found vertical response velocity distributions correspond well to the saturated hydraulic conductivity measurements in soil samples, although their distribution is much more tight.

At the hillslope scale (Q3), applications of 3-D time-lapse GPR are technically impossible due to the long acquisition times. Consequently we altered the setup to four trench-like 2-D time-lapse GPR profiles to facilitate the required high temporal resolution. The responses suggest structures similar to but less diverse than the found inter-aggregate voids at the plot scale. They are spatially persistent and leaky and apparently feed from diverse sources. As such the irrigation experiment caused a similar response in different structures than the previous storm event. Moreover, the relatively high input rates have proven adequately chosen to identify lateral subsurface flow paths. At this scale the capability of point-based methods for structure identification is even more limited as the dense network of soil moisture profile observations did not allow the derivation of a conclusive picture.

4.2 Event response patterns reveal flow-relevant structures

Interestingly, static methods failed to unravel structures from overall heterogeneity. This corroborates our idea that responses to an event are required for the identification of flow-relevant structures. Furthermore, it confirms that a combined assessment of form and function is needed to mutually reduce ambiguity. This is also shown in the companion study with a focus on function and processes at the hillslope scale (Angermann et al., 2017, this issue).
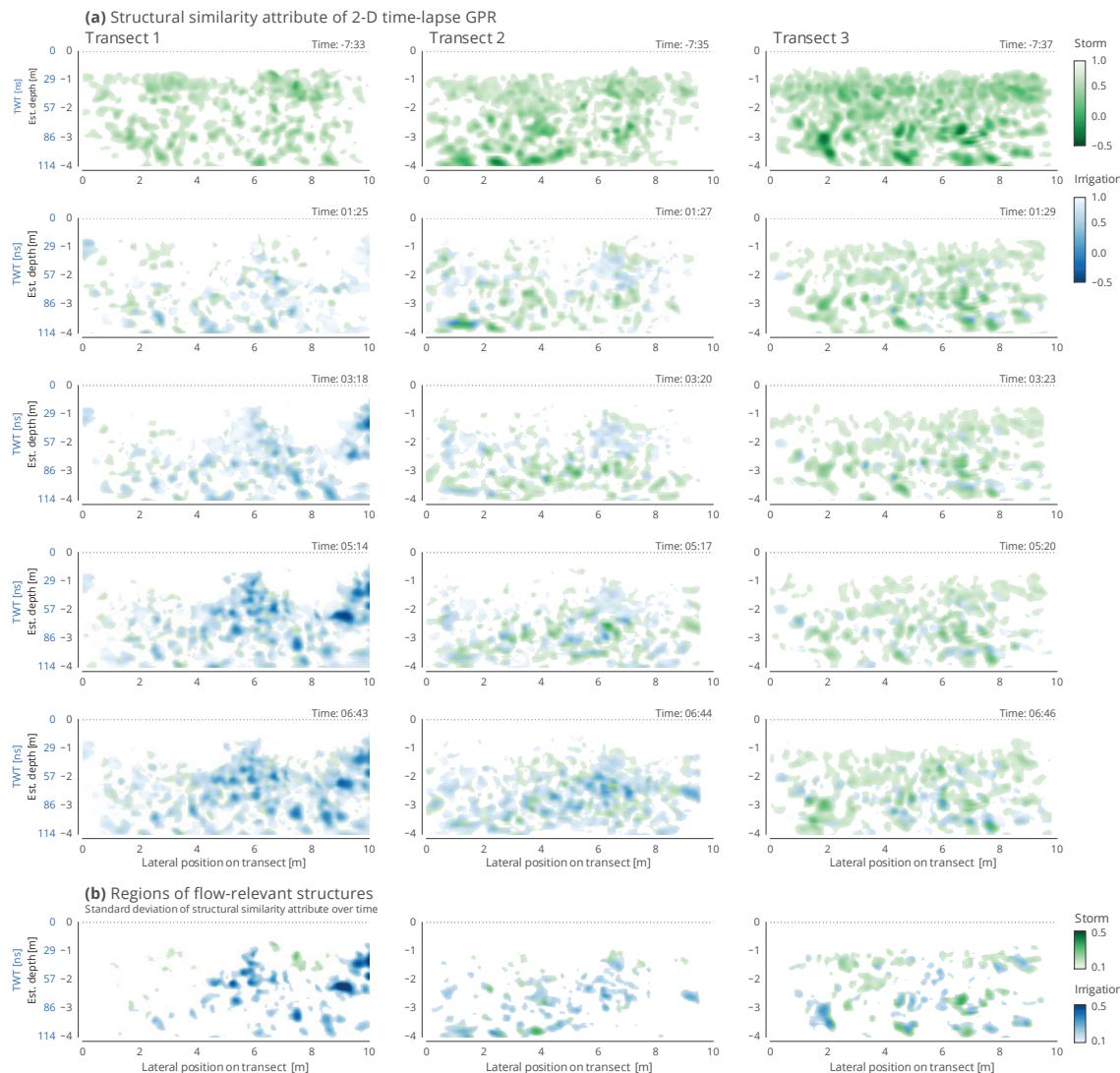
Figure 12. Structural similarity attribute in time-lapse 2-D GPR transects. Blue: irrigation event water; green: storm event water. Columns: time series in one transect; rows: different transects at the same time. **(b)** Identified regions of rapid subsurface flow based on the standard deviation of all structural similarity attributes at one transect over time. Note: the structural similarity attribute calculates similarity between the radargram at the respective time to the last record 23 h past irrigation. A threshold of 0.15 is applied to identify significant changes. It is a qualitative measure based on the assumption that the last record is in steady state and that all differences are induced by soil water redistribution.

4.2.1 Soil moisture responses

In our case TDR measurements through access tubes were employed as low-impact means to monitor soil water dynamics in order to detect areas of quick and strong response. Structures in general and the inter-aggregate voids in our case cover only a very small fraction of the measured volume. We may underestimate detected flow paths when they do not alter the total volumetric soil water content much (bypassing). This can explain the observed patterns of low response in the topsoil and changes in regions where the fast flow is decelerated at some kind of bottleneck. Referring to the theoretical

integration volume of 1 L, it would require a macropore of about 1 cm diameter within the support of the sensor to be filled to just reach a threshold of 2 % vol. Adding this 20 mL of water diffusively would result in the same measurement. This shows that soil moisture measurements exhibit a conceptual bias towards the diffusive fraction of the soil water.

The quantification of advective water from the recorded changes in soil moisture has been proven as not feasible. Given the insight of the discretely structured flow domain and the high lateral response velocities identified in the companion study (Angermann et al., 2017, this issue), the soil moisture measurements leave us with many questions. Com-

paring the identified regions of flow structures (Fig. 12) with the support of a TDR sensor quickly reveals that even a large number of point observations remains highly uncertain if the overall spatial context is unknown. This is especially the case at the hillslope scale. At the plot scale, the issue is less pronounced, as we have shown with the good correspondence between tracers, GPR and soil moisture reaction at plot X (Figs. 7 and 8). However, at plot XI with less intense irrigation, the soil moisture profile did not react despite the evidence of quick vertical redistribution in all the other methods. Apparently, the TDR records were simply not close enough to the relatively few activated flow paths (Fig. D2).

4.2.2 Time-lapse GPR patterns

The potential horizons identified by the static 3-D GPR survey do not coincide with the observed responses (Fig. 10). This is another example for the requirement of a shift between active and non-active state to identify flow-relevant structures. The structural similarity attributes derived from time-lapse GPR reveal the patterns of soil water redistribution. The continuous 2-D and 3-D data allow to relate temporal changes in space as images of the subsurface as proposed by Gerke et al. (2010); Beven and Germann (2013) and others.

The comparison of radargrams in time needs further attention: In other time-lapse GPR applications for soil water dynamics in structured domains (Truss et al., 2007; Haarder et al., 2011; Allroggen et al., 2015b; Klenk et al., 2015) analysis is guided by reference to a reflector and its apparent displacement can be used to calculate changes in soil moisture. Alternatively, a wetting front could generate a reflector (Léger et al., 2014). In our case none of these existed.

On the one hand, we minimized methodological problems concerning the noise arising from the imperfect positioning of repeated GPR measurements by using a measuring platform at the plot scale, transect guides at the hillslope scale, and an automatic-tracking total station (Allroggen et al., 2015b). On the other hand, we base our analysis on the structural similarity attribute instead of pixel-to-pixel comparison or picked reflectors. The disadvantage is that this allows only for a qualitative measure. The advantage is that the spatial organization of areas with changing reflection and transmission properties (which are attributed to changes in soil moisture) can be revealed even in heterogeneous soils. The 3-D applications at the plot scale avoid strong assumptions about the continuity of preferential flow paths when inferring 3-D networks from 2-D measurements (Gormally et al., 2011; Guo et al., 2014). Especially in the soils under study, the found response patterns (Fig. 8) and the excavated stained soil profiles (Fig. 6) show highly tortuous flow paths. Thus we refrain from interpolations between the multi-2-D transects at the hillslope scale.

Although the 3-D time-lapse attribute data of the plot irrigation experiments are of low spatial resolution (blur due to similarity attribute method and long duration of one acquisition) and limited temporal resolution (few acquisition times), they are suitable to identify regions of flow-relevant structures and their characteristics. In the multi-2-D transects resolution was enhanced (short duration of one acquisition and many repeated measurements) which depicted the structures much better. Hence, time-lapse GPR can especially be improved by enhancing the acquisition time and frequency.

The observation of changes during activation of flow-relevant structures generated the required contrast to overall heterogeneity. For large structures, this ledto precise identification and localization. Smaller flow paths cannot be fully resolved. Nevertheless, the continuous 2-D and 3-D images of the subsurface response patterns provide means to non-invasively study the form–function relationship in situ and to overcome some of the restrictions of retrospective and destructive tracer methods. However, quantitative interpretation of time-lapse GPR data remains challenging.

4.3 Methodological assessment

In contrary to our first expectation, the value of pedophysical analyses of soil core samples has been relatively high even for characteristics of flow facilitated by the revealed paths at larger scales. Structure identification is not only obscured in heterogeneity as one would expect, but properties deviating from the standard situation (fine texture, low bulk density and high porosity) gave rise to the identification of the inter-aggregate flow paths. However, the spatial organization of structures below and above the support of the samples cannot be revealed. This is also the reason for the relatively low information which could be drawn from the in situ infiltration measurements: The observed flow rates are largely affected by the capacity of the connected flow paths draining the measurement point. This adds to the critical assumption of homogeneity (Langhans et al., 2011).

Besides the high information gain through the state shift of flow-relevant structures in irrigation experiments, the employed methods at the plot scale have very specific advantages and disadvantages: Especially the laborious and costly analysis of salt tracers and stable isotopes is contrasted by relatively little additional information. Moreover, the lack of a temporal information about when the solute or water molecule was retained in a certain depth is seen problematic. Soil moisture profile dynamics and time-lapse GPR do not suffer this drawback. Both can be employed with very low or even no impact on the subsurface from the surface. While GPR requires to be operated in higher temporal resolution (see Sect. 4.5), soil moisture profiles lack the desired spatial discretization. Dye staining delivers the highest spatial resolution to reveal subsurface structures on the cost of blindness about the temporal dynamics. Furthermore, a tomographic excavation of the stains has proven very difficult.

Under strongly structured conditions as at the hillslope under study, point observations remain a needle in a haystack.

Unlike for vertical structures at the plot scale, the dense network of soil moisture profiles could not depict the lateral flow paths well. Here, the *GPR-inferred trenches* have shown to be a valuable surrogate for massive staining like in the study by Anderson et al. (2009). In addition, the temporal dynamics of the hillslope reaction could be observed.

With regard to our a priori model application, the combination of vertical and lateral flow paths (identified in the irrigation experiments) with layers of low permeability just below the structures (observed in the soil core profiles) could refine the domain towards more lateral soil water transport. The mean retention properties (derived from pedo-physical analyses) are adequate. Hence, the combination of data from all scales can contribute to a refinement of the model.

4.4 Heterogeneity versus structure

Based on numerous point-scale measurements, the overall layering and mean property of a heterogeneous soil with periglacial deposit layers were described in Sect. 3.1. Given the large effort to conduct such an exploration, this result appears rather unsatisfactory and could have been achieved with much simpler means (e.g., compare Heller and Kleber, 2016, in a similar setting). However, they have been key to the identification of the sub-scale inter-aggregate structures which convey to vertical drainage paths and a lateral network in the subsurface. Without high supply rates from the point scale, subsurface storm flow in the lateral structures could not be sustained.

At the hillslope scale, the attribute supported picking of potential structures in the 3-D GPR data cube also had high discrepancies from the actual relevant structures (see the differences between potential subsurface structures and recorded reaction in the TDR and GPR profiles in Fig. 10), which is in contrast to similar GPR applications by Gormally et al. (2011) and explains the large spread in the results of the hydrological measurements.

It has proven particularly difficult to distinguish heterogeneity and structure. This has conceptual implications: as introduced, we regard statistical heterogeneity as random small-scale changes in hydrological soil properties (de Marsily et al., 2005) and structure as spatially organized flow paths (Gerke, 2012) and their connectedness (Tetzlaff et al., 2010) or persistent spatial covariance of high conductivity values. Hence the structures require multivariate or topological characterization. To infer on the directed flow in subsurface structures, a spatially continuous observation of the reaction to an event is required. This has proven especially challenging as the spatial scale of flow-relevant structures (5×10^{-3} to 5×10^{-2} m) is several orders of magnitude below the support of standard soil samples and hydrological measurements (1×10^{-4} to 1×10^{-3} m^3).

In more general terms, heterogeneity can be seen as deviation of the found reality from the concept of quasi-homogeneous elementary volumes. If this deviation concerns only the apparent parameters of the same physical process, more samples are adequate to determine their distribution. In cases (like here) where this deviation also means a shift in the physical processes, heterogeneity may introduce bias as it becomes a scale problem: any measurement will consist of an unknown subset of connected or non-connected flow paths. This makes it impossible to unravel the properties of the different flow domains without knowing the composition of the explored ensemble of each measurement. Hence the point-based techniques cannot determine the super-scale organization outside the support of the measurement. Without detection of organization and thus flow-relevant structures, they can only recover heterogeneity independently of the number of samples.

4.5 Outlook on structure identification with time-lapse GPR

In the context of preferential flow studies in watersheds around the globe and in many different models, our results open new ways to visualize subsurface flow and to facilitate more field studies to understand stormflow generation (as recently found in a meta-analysis by Barthold and Woods, 2015). Although we cannot fully resolve macropores as needed in spatially explicit representation of macropores as vertically and laterally connected flow paths (Vogel et al., 2006; Sander and Gerke, 2009; Klaus and Zehe, 2011), our findings provide the experimental basis to further develop such models. More implicit approaches like stochastic stream tubes (Jury and Roth, 1990), the scale way idea (Vogel and Roth, 2003), or dual porosity and permeability approaches (Gerke, 2006) could be extended by providing spatial and temporal context, which is one of their assumptions. Also, more integrating concepts like the representative watersheds (REW, Reggiani et al., 1998; Tian et al., 2006; Lee et al., 2007) could define zones for quick drainage based on repeated response observations in vertical and lateral structures.

In the form and function framework one implication of the study is that a disjunct analysis of the two is a source of unnecessary ambiguity and susceptibility to bias. Although the conjugated nature of form and function is very much in line with the general findings and perception of early studies (Aristotle in Blits, 1999; Thompson, 1917; Wittgenstein, 1922, and others), it contradicts a general notion in hydrological surveying and modeling to separate the two. In most models different flow paths are defined in a lumped manner using effective parameters after all. These domains and their parameters could be determined based on irrigation experiments and time-lapse GPR measurements.

While models require specific parameters about the site under study which are coherent with their conceptual assumptions or modeler's perception (Holländer et al., 2014), the experiments are also strongly shaped by our perceptual model about the processes. With this, the matter of model

adequacy is not restricted to numerical aspects alone (Gupta et al., 2012). Methodologically, the in situ imaging of sub-surface flow processes can be used to reduce ambiguity of measurements and to constrain the process conceptualization in heterogeneous and structured soils. In our case, the a priori model overestimated deep percolation and underestimated the velocity of lateral soil water redistribution through subsurface flow paths. Based on field information about the overall distribution of flow paths or quickly reacting areas, sampling and monitoring could be guided. This would reduce the limitations of point-scale methods with relatively little effort.

5 Conclusions

In the hillslopes under study, silty, cohesive soils coincide with high porosity and high flow velocities at the Darcy scale. This motivated in depth investigation of flow-relevant structures explaining this. We have shown that subsurface heterogeneity and the mismatch of observation and process scales obscured the identification of flow-relevant structures under static conditions without a shift between active and non-active states. The pedo-physical analyses initiated the recognition of these sub-scale inter-aggregate structures. The point-scale exploratory methods could quantify the general characteristics of the subsurface only within a wide spectrum of the respective target properties. However, they failed to identify flow-relevant structures in terms of position, distribution and capacity at larger scales. Measurements of infiltration capacity and hydraulic conductivity require special attention, because they integrate over an unknown set of advective and diffusive flow paths. The discrepancy between results from the soil core profiles and a 3-D GPR survey on the one hand and the time-lapse approaches on the other hand indicates that structures identified from inhomogeneities are not necessarily flow-relevant pathways.

Joint application of tracers and time-lapse GPR during irrigation experiments revealed details about the structures and their activation by flow. At the plot scale a network of inter-aggregate pores enables fast soil water redistribution in a less directed manner and at much finer scales than usually expected in macropores like cracks, worm burrows or root channels. This facilitates high apparent vertical flow velocities ranging around $10^{-3.5}\,\mathrm{m\,s^{-1}}$, while operating in fine pores at scales very difficult to identify even with dye staining. The combination of tracer and time-lapse GPR methods enabled the more holistic view into the subsurface which was further applied to the hillslope scale. There persistent lateral pathways connecting along the hillslope have been identified through *GPR-inferred trenches*.

Our findings show that form and function in hydrological systems operate in conjugated pairs. This implies that it is very difficult to observe them separately and that their projections are inherently non-unique and scale-dependent. Besides the fine scale of the inter-aggregate voids, form needs to be addressed in its context to reveal information about its structure and characteristics, but addressing function also needs details about the spatial circumstances to be conclusive. Overly strong assumptions about structures or processes can be avoided by the presented non-invasive time-lapse GPR method, which can visualize and localize response patterns at the plot and hillslope scale. They compare well with soil moisture dynamics and tracer recovery. As such the localization of responses provides the missing link to relate form to function (taken up in the companion study by Angermann et al., 2017, this issue) and to guide more specific investigation, monitoring and modeling of subsurface processes.

Appendix A: Quick sampler for fast undisturbed core sampling on excavated profiles

In addition to the dye tracer stain records, quantitative analysis of salt tracer recovery distribution in the excavated profiles underneath the irrigation plots was done. One challenge to address was the required time to collect an adequate array of such soil samples with known volumetric reference. We developed a re-loadable core sampler with a calibrated sample volume of 66 mL.

The sampler is applied like a ring sample with an attached hammering adaptor. In order to minimize time and impact on the profile we enabled a pull-withdrawal of the sample. For this, the sampler is about 15 mm longer than the desired sample. The irregular open edge is scraped off by a calibration twist drill. The prepared and accurate to volume sample is finally pushed out by a piston from the sampler into a sealable brown glass bottle for further treatment in the laboratory.

Appendix B: Detailed results of hydrological measurements

In situ measurements of infiltration capacity and saturated hydraulic conductivity had a highly heterogeneous distribution. To detail the respective records and found profiles, Fig. B1 shows them in a spatial context.

Appendix C: A priori model reference

Based on the findings of the pedo-physical exploration, we setup the 2-D process model CATFLOW (Zehe et al., 2001) as representative hillslope for hypothetical a priori simulation of the experiment in order to determine the required irrigation intensity, the spatial extent of the observation network, the temporal resolution, and the duration of the monitoring. The model domain was set up based on the soil property estimates from the soil physical exploration assuming a fractured periglacial deposit layer as conductive layer in the hillslope (Fig. C1). In a series of scenarios, the one with $30\,\mathrm{mm\,h^{-1}}$ irrigation for 4 h turned out to be well balanced with respect to anticipated hillslope reaction given a limited source area. Fast soil water redistribution was modeled to last for 12 h.

Comparing the results from the model with the experiment shows strong deviation in terms of the activation of a conductive layer. However, this could be improved by adding a layer of low permeability below, since the modeled reaction on the bedrock interface is quite similar but slower than the observed dynamics.

Appendix D: Porewater stable isotope analysis in plot irrigation experiments

In addition to bromide as a conservative salt tracer, the same percussion drilled core samples were analyzed for their stable isotopic composition (δ^{18}O and δ^2H) of the porewater. This was done with the direct equilibration method as proposed by Wassenaar et al. (2008) and described in detail by Sprenger et al. (2015) using a wavelength-scanned cavity ring-down spectrometer (Picarro Inc.). The precision for the method is reported to be $0.31\,‰$ for δ^{18}O and $1.16\,‰$ for δ^2H (Sprenger et al., 2016). The measured isotopic signal is given relative to the Vienna Standard Mean Ocean Water. As a pre-experiment reference, a fourth reference core was sampled prior to the experiments about 3 m upslope.

We calculate the volumetric event water portion (–) in the soil water as

$$\frac{\Theta_{\mathrm{event}}}{\Theta_{2\,\mathrm{h}}} = \frac{\Theta_{2\,\mathrm{h}} \cdot \delta^2 \mathrm{H}_{2\,\mathrm{h}} - \Theta_{\mathrm{pre}} \cdot \delta^2 \mathrm{H}_{\mathrm{pre}}}{\Theta_{2\,\mathrm{h}} \cdot \delta^2 \mathrm{H}_{\mathrm{event}}}, \tag{D1}$$

with δ^2H as deuterium composition (‰) in the pre-event reference sample ("pre"), in the core sample 2 h after irrigation start (2 h), and in the irrigation water ("event"). The amount of soil water is given as Θ (mm).

Figure D1a–c show the depth profile of irrigation water as a portion of total water content, calculated from the deviation in δ^2H concentration between the reference and 2 h past irrigation core samples. The results are also compared to the bromide concentrations in the soil water phase of the same samples, showing slight correlation. However, the values are rather noisy due to low difference of the isotopic composition of the soil water and the non-enriched irrigation water. Figure D1d–e highlight the very weak soil moisture signal and low deviation between the respective soil cores close to the method's precision. Especially interpretation of the peak at about 0.5 m depth and signals below may be erroneous, because the signature of the reference core coincides with the irrigation water there.

In line with the findings of Klaus et al. (2013) the isotopic signal of non-enriched water required strong assumptions for its interpretation. In our case this specifically applies to the plot-scale core samples where we calculated the difference to the pre-experiment core regardless of the fact that soil water and irrigation water deviated only slightly ($\geq 15\,‰$) and even had the same values at 0.5 m depth. Moreover, the reference core was required to be at a different location. Hence, flow paths and thus the initial isotope profile are not necessarily the same as at the respective plots. However, as an assumably ideal tracer, the stable isotope data allowed for an additional and coherent measurement. With respect to the overall findings of rapid flow in discrete structures the assumption is justified.
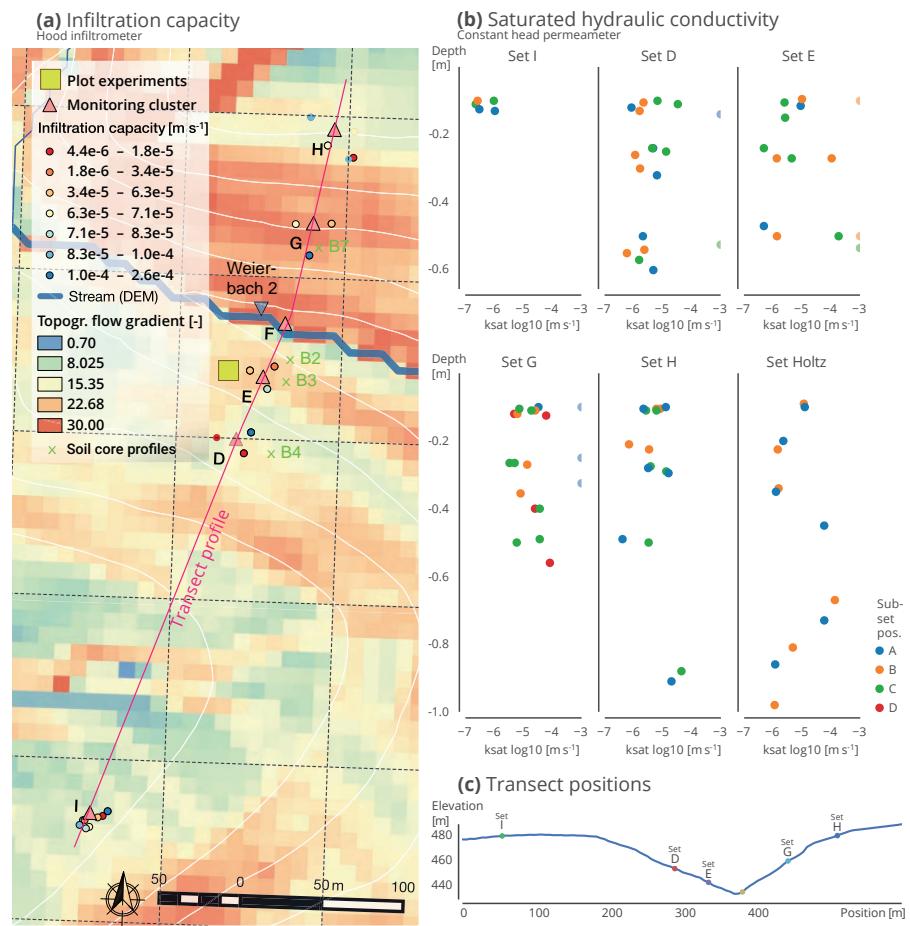
Figure B1. Hydrological exploration results. **(a)** Infiltration capacity (values color coded) measured with hood infiltrometer. Base map with topographic flow gradient $G = \arctan(\partial z_{rel}/\partial d_{rel})$, topography contour lines and positions of hydro-meteorological monitoring clusters. **(b)** Saturated hydraulic conductivity measured with a constant head permeameter. Individual profile position in the respective nested set color coded (two measurement holes with dist = 1 m). Values exceeding the device capacity set to 10^{-3} m s^{-1}. **(c)** Elevation profile of the transect as a characteristic landscape feature in the sub-basin.
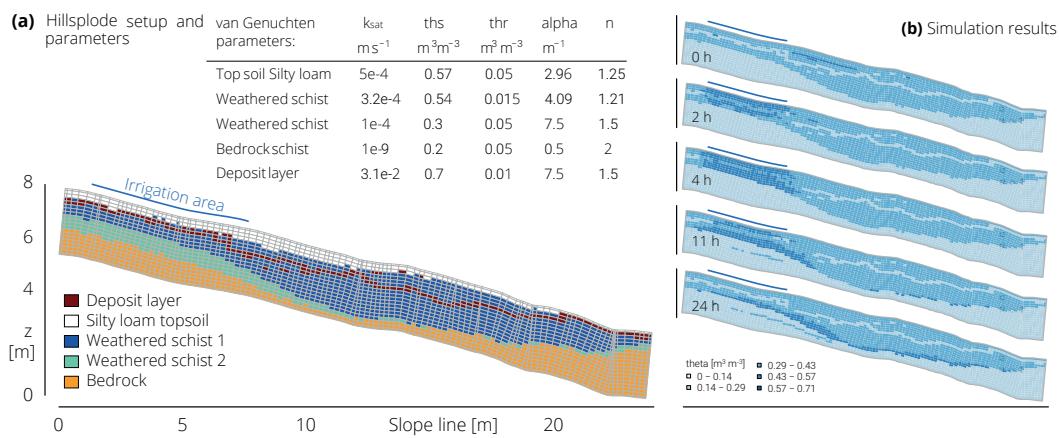
Figure C1. CATFLOW model reference of the hillslope experiment. **(a)** Table of used soil definitions and hypothetical hillslope setup. **(b)** Simulated soil moisture with 30 mm h^{-1} irrigation for 4 h.
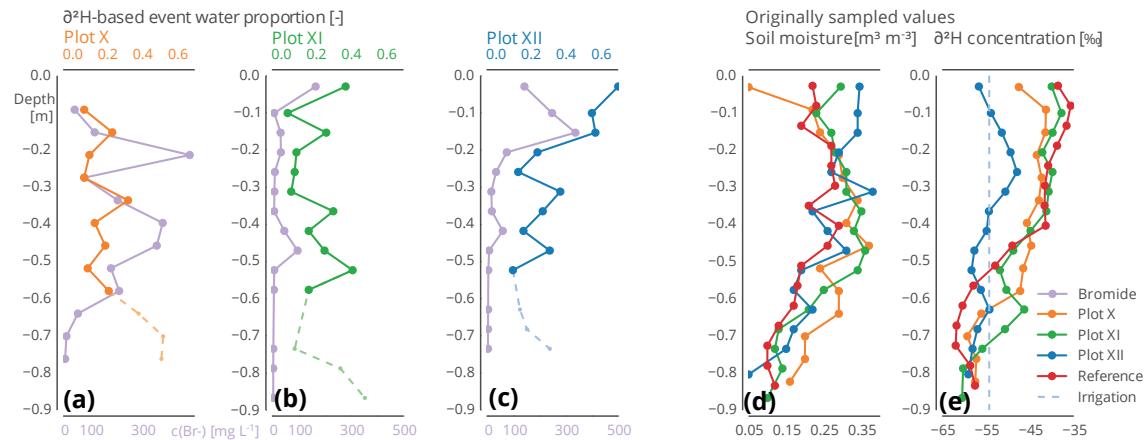
Figure D1. Plot-scale irrigation experiments. Proportion of event water derived from deviations in concentrations of deuterium and bromide in the soil water of sampled cores to the reference **(a–c)**. Absolute measured soil water content **(d)**. Deuterium concentration in samples and signature of irrigation water **(e)**. Note that the δ^2H concentration profiles and the signature of the irrigation water show low deviation and even coincide at 0.5 m depth. Thus the interpretation needs precaution, especially below that level (dashed lines, **a–c**).
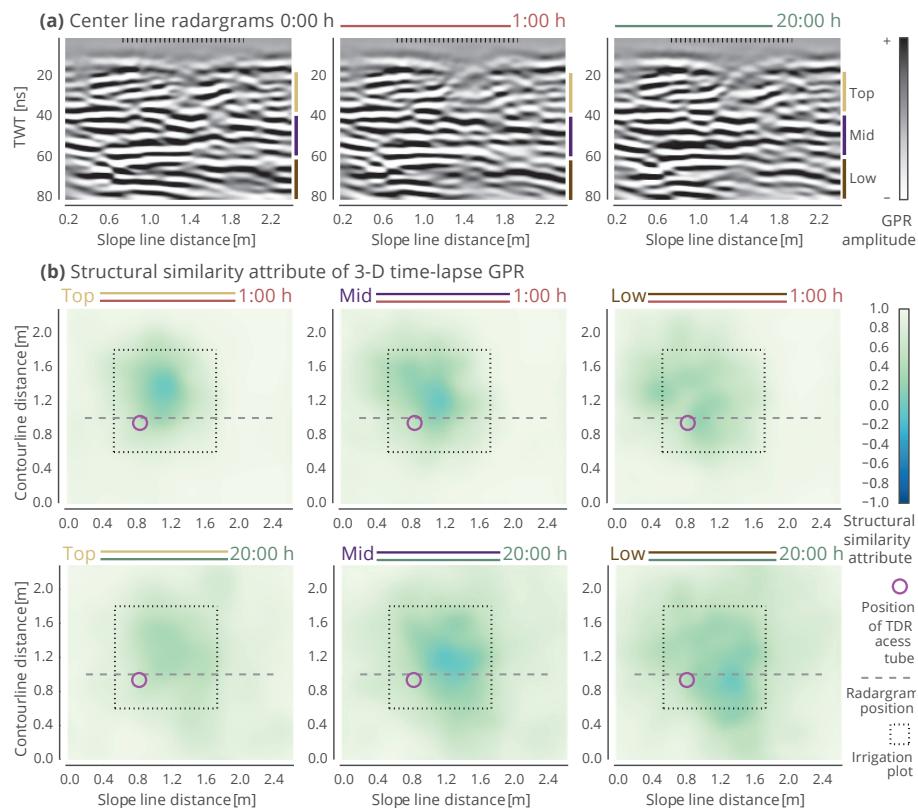
Figure D2. Time-lapse 3-D GPR of the irrigation experiment at plot XI. Center line radargrams at the marked transect (gray dashed line in the lower panels) for the three acquisition times (before 0:00 h, directly after irrigation 1:00 h, 20:00 h after irrigation) are given in the top row. The structural similarity attribute of the 3-D data cube is given in three different depth layers (top 20–40 ns, mid 40–60 ns, low 60–80 ns) in the lower panels. The irrigation plot is marked by a black dashed box/line. Slope line distance is increasing downslope.
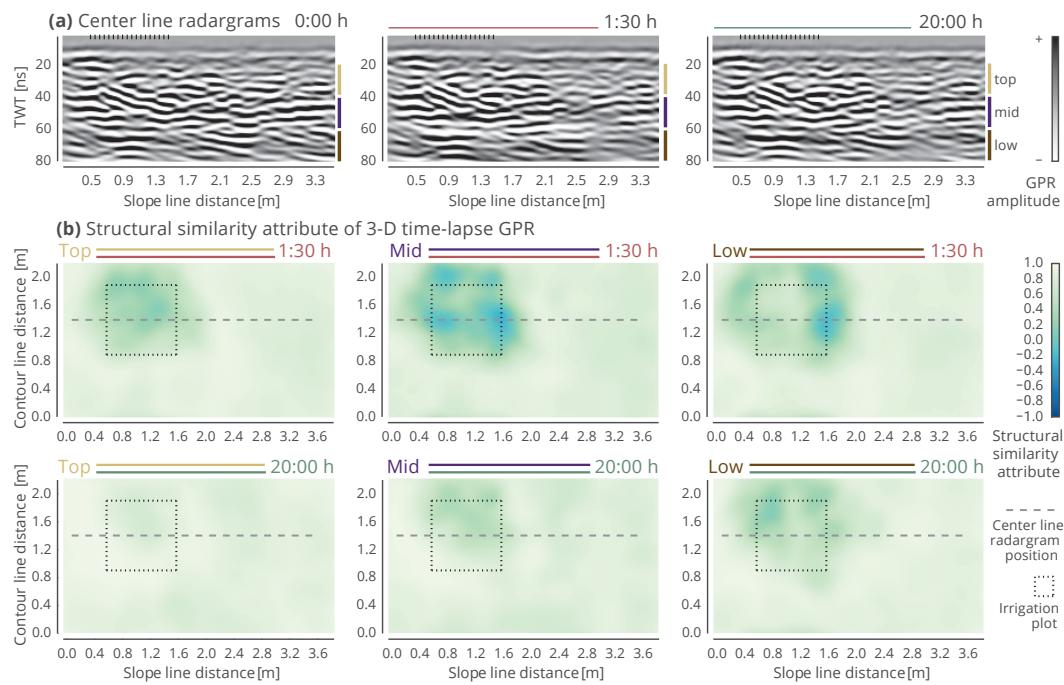
Figure D3. Time-lapse 3-D GPR of the irrigation experiment at plot XII. Center line radargrams at the marked transect (gray dashed line in the lower panels) for the three acquisition times (before 00:00 h, directly after irrigation 01:30 h, 20:00 h after irrigation) are given in the top row. The structural similarity attribute of the 3-D data cube is given in three different depth layers (top 20–40 ns, mid 40–60 ns, low 60–80 ns) in the lower panels. The irrigation plot is marked by a black dashed box/line. Slope line distance is increasing downslope.

Appendix E: Results for the 3-D time-lapse GPR at plots XI and XII

In addition to the results in Sect. 3.2, here the radargrams and structural similarity attributes for the other two plot-scale experiments are given in Figs. D2 and D3. In plot XI with less intense irrigation the lateral spread of water is less pronounced. As found by the tracer methods, interaction with the soil matrix was elevated in plot XII. Moreover, the acquisition of the GPR data took longer at this plot.

Appendix F: Technical concerns of the time-lapse GPR and the structural similarity attribute

The demands on the precision of the repeated acquisition with spatial determination and antenna contact to the ground are very high and are assumed to be nearly perfect within our experiments. Under field conditions precision is limited due to numerous effects like micro-topography, topsoil conditions, signal attenuation and even weather. The lack of distinguished reflectors also inhibited any estimation of quantitative values. Further, the referenced depths in Fig. 12 are only estimates based on a constant mean GPR velocity which can also vary in time and space depending on the initial conditions.
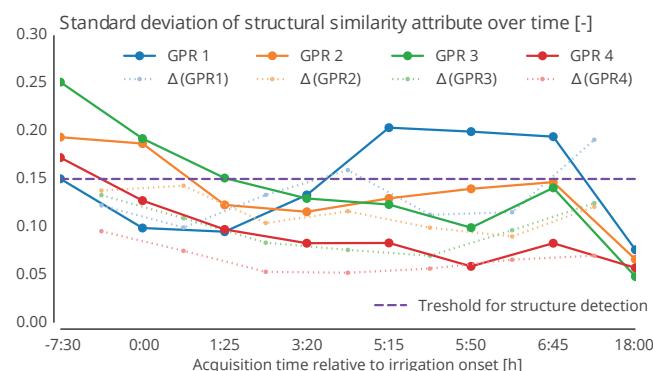

Figure E1. Standard deviation of the structural similarity attribute at the different GPR transects in the hillslope experiment over time (solid lines) and standard deviation of the differences of two successive attribute distributions (dotted lines). The used threshold for the detection of flow-relevant structures is marked as the dashed purple line.

The highlighted assumptions clearly frame the limits of the technique. The overall sensitivity of the approach can be judged from the structural similarity attribute of the last pairs of records in the hillslope experiment when we assume the soil water to be in equilibrium again. Figure E1 presents the development of the standard deviations of the structural similarity attribute of the respective transects over time. In

dotted lines we plotted the standard deviations of the step-wise attribute differences. The standard deviations of the attribute for the last pairs of records is 0.06. Using this value as methodological noise reference, it implies that weak responses and local effects must not be over interpreted. Hence, the introduced threshold of 0.15 for irrigation signal detection appears to be a reasonable choice for qualitative interpretation in our case.

Another limit is the interpretability of changes in the radargrams, as water can have different effects under different situations. A wetted well-defined surface may quickly become a reflector which is easy to detect. However, tortuous flow paths may not be as ideal. Small structures might be well below the limits of detectability in the complex reflection pattern. As such the structural similarity attribute can only detect zones of significant changes which can be induced by many lumped small structures, one big flow path, or even a favorably oriented stone which gets wetted.

References

Allaire, S. E., Roulier, S., and Cessna, A. J.: Quantifying preferential flow in soils: A review of different techniques, J. Hydrol., 378, 179–204, https://doi.org/10.1016/j.jhydrol.2009.08.013, 2009.

Allroggen, N. and Tronicke, J.: Attribute-based analysis of time-lapse ground-penetrating radar data, Geophysics, 81, H1–H8, https://doi.org/10.1190/geo2015-0171.1, 2016.

Allroggen, N., Tronicke, J., Delock, M., and Böniger, U.: Topographic migration of 2D and 3D ground-penetrating radar data considering variable velocities, Near Surf. Geophys., 13, 253–259, https://doi.org/10.3997/1873-0604.2014037, 2015a.

Allroggen, N., van Schaik, L., and Tronicke, J.: 4D ground-penetrating radar during a plot scale dye tracer experiment, J. Appl. Geophys., 118, 139–144, https://doi.org/10.1016/j.jappgeo.2015.04.016, 2015b.

Amoozegar, A.: Comparison of the Glover Solution with the Simultaneous-Equations Approach for Measuring Hydraulic Conductivity, Soil Sci. Soc. Am. J., 53, 1362, https://doi.org/10.2136/sssaj1989.03615995005300050010x, 1989.

Anderson, A. E., Weiler, M., Alila, Y., and Hudson, R. O.: Dye staining and excavation of a lateral preferential flow network, Hydrol. Earth Syst. Sci., 13, 935–944, https://doi.org/10.5194/hess-13-935-2009, 2009.

Angermann, L., Jackisch, C., Allroggen, N., Sprenger, M., Zehe, E., Tronicke, J., Weiler, M., and Blume, T.: Form and function in hillslope hydrology: characterization of subsurface flow based on response observations, Hydrol. Earth Syst. Sci., 21, 3727–3748, https://doi.org/10.5194/hess-21-3727-2017, 2017.

Armstrong, R. T., McClure, J. E., Berrill, M. A., Rücker, M., Schlüter, S., and Berg, S.: Beyond Darcy's law: The role of phase topology and ganglion dynamics for two-fluid flow, Phys. Rev. E, 94, 043113, https://doi.org/10.1103/PhysRevE.94.043113, 2016.

Bachmair, S. and Weiler, M.: Interactions and connectivity between runoff generation processes of different spatial scales, Hydrol. Proc., 28, 1916–1930, https://doi.org/10.1002/hyp.9705, 2014.

Barthold, F. K. and Woods, R. A.: Stormflow generation: A meta-analysis of field evidence from small, forested catchments, Water Resour. Res., 51, 3730–3753, https://doi.org/10.1002/2014WR016221, 2015.

Beven, K. and Germann, P.: Macropores and water flow in soils, Water Resour. Res., 18, 1311–1325, https://doi.org/10.1029/WR018i005p01311, 1982.

Beven, K. and Germann, P.: Macropores and water flow in soils revisited, Water Resour. Res., 49, 3071–3092, https://doi.org/10.1002/wrcr.20156, 2013.

Binley, A., Hubbard, S. S., Huisman, J. A., Revil, A., Robinson, D. A., Singha, K., and Slater, L. D.: The emergence of hydrogeophysics for improved understanding of subsurface processes over multiple scales, Water Resour. Res., 51, 3837, https://doi.org/10.1002/2015WR017016, 2015.

Birken, R. and Versteeg, R.: Use of four-dimensional ground penetrating radar and advanced visualization methods to determine subsurface fluid migration, J. Appl. Geophys., 43, 215–226, https://doi.org/10.1016/S0926-9851(99)00060-9, 2000.

Bishop, J. M., Callaghan, M. V., Cey, E. E., and Bentley, L. R.: Measurement and simulation of subsurface tracer migration to tile drains in low permeability, macroporous soil, Water Resour. Res., 51, 3956–3981, https://doi.org/10.1002/2014WR016310, 2015.

Blits, K. C.: Aristotle: Form, function, and comparative anatomy, The Anatomical Record, 257, 58–63, https://doi.org/10.1002/(SICI)1097-0185(19990415)257:2<58::AID-AR6>3.0.CO;2-I, 1999.

Bloschl, G. and Sivapalan, M.: Scale issues in hydrological modelling: A review, Hydrol. Proc., 9, 251–290, https://doi.org/10.1002/hyp.3360090305, 1995.

Blouin, M., Hodson, M. E., Delgado, E. A., Baker, G., Brussaard, L., Butt, K. R., Dai, J., Dendooven, L., Peres, G., Tondoh, J. E., Cluzeau, D., and Brun, J. J.: A review of earthworm impact on soil function and ecosystem services, Eur. J. Soil Sci., 64, 161–182, https://doi.org/10.1111/ejss.12025, 2013.

Blume, T. and van Meerveld, H. J. I.: From hillslope to stream: methods to investigate subsurface connectivity, Wiley Interdisciplinary Reviews: Water, 2, 177–198, https://doi.org/10.1002/wat2.1071, 2015.

Bouma, J. and Dekker, L. W.: A case study on infiltration into dry clay soil I. Morphological observations, GEODERMA, 20, 27–40, https://doi.org/10.1016/0016-7061(78)90047-2, 1978.

Carsel, R. F. and Parrish, R. S.: Developing joint probability distributions of soil water retention characteristics, Water Resour. Res., 24, 755–769, https://doi.org/10.1029/WR024i005p00755, 1988.

de Gruijter, J., Brus, D. J., Bierkens, M. F. P., and Knotters, M.: Sampling for Natural Resource Monitoring, Springer Science & Business Media, 2006.

de Marsily, G., Delay, F., Gonçalvès, J., Renard, P., Teles, V., and Violette, S.: Dealing with spatial heterogeneity, Hydrogeol. J., 13, 161–183, 2005.

Fenicia, F., Kavetski, D., Savenije, H. H. G., Clark, M. P., Schoups, G., Pfister, L., and Freer, J.: Catchment properties, function, and conceptual model representation: is there a correspondence?, Hydrol. Proc., 28, 2451–2467, https://doi.org/10.1002/hyp.9726, 2014.

Flury, M., Flühler, H., Jury, W. A., and Leuenberger, J.: Susceptibility of soils to preferential flow of water: A field study, Water Resour. Res., 30, 1945–1954, https://doi.org/10.1029/94WR00871, 1994.

Gerke, H. H.: Preferential flow descriptions for structured soils, J. Plant Nutr. Soil Sc., 169, 382–400, https://doi.org/10.1002/jpln.200521955, 2006.

Gerke, H. H.: Macroscopic Representation of the Interface between Flow Domains in Structured Soil, Vadose Zone J., 11, https://doi.org/10.2136/vzj2011.0125, 2012.

Gerke, H. H., Germann, P., and Nieber, J.: Preferential and Unstable Flow: From the Pore to the Catchment Scale, Vadose Zone J., 9, 207–212, https://doi.org/10.2136/vzj2010.0059, 2010.

Germann, P. and al Hagrey, S.: Gravity-Driven and Viscosity-Dominated Infiltration into a Full-Scale Sand Model, Vadose Zone J., 7, 1160–1169, https://doi.org/10.2136/vzj2007.0172, 2008.

Germann, P. F.: Preferential Flow, Geographica Bernensia, Institute of Geography, University of Bern, 2014.

Germann, P. F. and Karlen, M.: Viscous-Flow Approach to In Situ Infiltration and In Vitro Saturated Hydraulic Conductivity Determination, Vadose Zone J., 15, https://doi.org/10.2136/vzj2015.05.0065, 2016.

Gormally, K. H., McIntosh, M. S., Mucciardi, A. N., and McCarty, G. W.: Ground-Penetrating Radar Detection and Three-Dimensional Mapping of Lateral Macropores: II. Riparian Application, Soil Sci. Soc. Am. J., 75, 1236, https://doi.org/10.2136/sssaj2010.0342, 2011.

Grayson, R. and Blöschl, G., eds.: Spatial Patterns in Catchment Hydrology: Observations and Modelling, Cambridge University Press, 2001.

Guo, L., Chen, J., and Lin, H.: Subsurface lateral preferential flow network revealed by time-lapse ground-penetrating radar in a hillslope, Water Resour. Res., 50, 9127–9147, https://doi.org/10.1002/2013WR014603, 2014.

Gupta, H., Clark, M. P., Vrugt, J. A., Abramowitz, G., and Ye, M.: Towards a Comprehensive Assessment of Model Structural Adequacy, Water Resour. Res., 48, 1–40, https://doi.org/10.1029/2011WR011044, 2012.

Haarder, E. B., Looms, M. C., Jensen, K. H., and Nielsen, L.: Visualizing unsaturated flow phenomena using high-resolution reflection ground penetrating radar, Vadose Zone J., 10, 84–97, https://doi.org/10.2136/vzj2009.0188, 2011.

Hellebrand, H., Müller, C., Matgen, P., Fenicia, F., and Savenije, H.: A process proof test for model concepts: Modelling the meso-scale, Phys. Chem. Earth, 36, 42–53, https://doi.org/10.1016/j.pce.2010.07.019, 2011.

Heller, K.: Einfluss periglazialer Deckschichten auf die oberflächennahen Fließwege am Hang – eine Prozessstudie im Osterzgebirge, Sachsen, PhD thesis, Technical University Dresden, 2012.

Heller, K. and Kleber, A.: Hillslope runoff generation influenced by layered subsurface in a headwater catchment in Ore Mountains, Germany, Environ. Earth Sci., 75, 943–15, https://doi.org/10.1007/s12665-016-5750-y, 2016.

Hillel, D.: Fundamentals of Soil Physics, Academic Press, 1980.

Holden, J.: Hydrological connectivity of soil pipes determined by ground-penetrating radar tracer detection, Earth Surf. Proc. Land., 29, 437–442, https://doi.org/10.1002/esp.1039, 2004.

Holländer, H. M., Bormann, H., Blume, T., Buytaert, W., Chirico, G. B., Exbrayat, J.-F., Gustafsson, D., Hölzel, H., Krauße, T., Kraft, P., Stoll, S., Blöschl, G., and Flühler, H.: Impact of modellers' decisions on hydrological a priori predictions, Hydrol. Earth Syst. Sci., 18, 2065–2085, https://doi.org/10.5194/hess-18-2065-2014, 2014.

Huisman, J. A., Hubbard, S. S., Redman, J. D., and Annan, A. P.: Measuring Soil Water Content with Ground Penetrating Radar: A Review, Vadose Zone J., 2, 476–491, https://doi.org/10.2113/2.4.476, 2003.

Jackisch, C.: Linking structure and functioning of hydrological systems – How to achieve necessary experimental and model complexity with adequate effort, PhD thesis, KIT Karlsruhe Institute of Technology, https://doi.org/10.5445/IR/1000051494, 2015.

Jones, J. P., Sudicky, E. A., Brookfield, A. E., and Park, Y.-J.: An assessment of the tracer-based approach to quantifying groundwater contributions to streamflow, Water Resour. Res., 42, W02407, https://doi.org/10.1029/2005WR004130, 2006.

Juilleret, J., Iffly, J. F., Pfister, L., and Hissler, C.: Remarkable Pleistocene periglacial slope deposits in Luxem-bourg (Oesling): pedological implication and geosite potential, Bulletin de Société des Naturalistes Luxembourgeois, 2011.

Jury, W. A. and Roth, K.: Transfer functions and solute movement through soil, theory and applications, Birkhauser, 1990.

Kasteel, R., Vogel, H.-J., and Roth, K.: Effect of non-linear adsorption on the transport behaviour of Brilliant Blue in a field soil, Eur. J. Soil Sci., 53, 231–240, https://doi.org/10.1046/j.1365-2389.2002.00437.x, 2002.

Kirchner, J. W.: A double paradox in catchment hydrology and geochemistry, Hydrol. Proc., 17, 871–874, https://doi.org/10.1002/(ISSN)1099-1085, 2003.

Klaus, J. and Zehe, E.: A novel explicit approach to model bromide and pesticide transport in connected soil structures, Hydrol. Earth Syst. Sci., 15, 2127–2144, https://doi.org/10.5194/hess-15-2127-2011, 2011.

Klaus, J., Zehe, E., Elsner, M., Külls, C., and McDonnell, J. J.: Macropore flow of old water revisited: experimental insights from a tile-drained hillslope, Hydrol. Earth Syst. Sci., 17, 103–118, https://doi.org/10.5194/hess-17-103-2013, 2013.

Klenk, P., Jaumann, S., and Roth, K.: Quantitative high-resolution observations of soil water dynamics in a complicated architecture using time-lapse ground-penetrating radar, Hydrol. Earth Syst. Sci., 19, 1125–1139, https://doi.org/10.5194/hess-19-1125-2015, 2015.

Koestel, J. and Larsbo, M.: Imaging and quantification of preferential solute transport in soil macropores, Water Resour. Res., 50, 4357–4378, https://doi.org/10.1002/2014WR015351, 2014.

Langhans, C., Govers, G., Diels, J., Leys, A., Clymans, W., Putte, A., and Valckx, J.: Experimental rainfall–runoff data: Reconsidering the concept of infiltration capacity, J. Hydrol., 399, 255–262, https://doi.org/10.1016/j.jhydrol.2011.01.005, 2011.

Lee, H., Zehe, E., and Sivapalan, M.: Predictions of rainfall-runoff response and soil moisture dynamics in a microscale catchment using the CREW model, Hydrol. Earth Syst. Sci., 11, 819–849, https://doi.org/10.5194/hess-11-819-2007, 2007

Léger, E., Saintenoy, A., and Coquet, Y.: Hydrodynamic parameters of a sandy soil determined by ground-penetrating radar inside a single ring infiltrometer, Water Resour. Res., 50, 5459–5474, https://doi.org/10.1002/2013WR014226, 2014.

Leibundgut, C., Maloszewski, P., and Külls, C.: Tracers in Hydrology, Wiley, 2011.

Loritz, R., Hassler, S. K., Jackisch, C., Allroggen, N., van Schaik, L., Wienhöfer, J., and Zehe, E.: Picturing and modeling catchments by representative hillslopes, Hydrol. Earth Syst. Sci., 21, 1225–1249, https://doi.org/10.5194/hess-21-1225-2017, 2017.

Marfurt, K. J., Kirlin, R. L., Farmer, S. L., and Bahorich, M. S.: 3-D seismic attributes using a semblance-based coherency algorithm, Geophysics, 63, 1150–1165, 1998.

Martínez-Carreras, N., Hissler, C., Gourdol, L., Klaus, J., Juilleret, J., François Iffly, J., and Pfister, L.: Storage controls on the generation of double peak hydrographs in a forested headwater catchment, J. Hydrol., 543, 255–269, https://doi.org/10.1016/j.jhydrol.2016.10.004, 2016.

McDonnell, J. J., Freer, J., Hooper, R., Kendall, C., Burns, D., Beven, K., and Peters, J.: New method developed for studying flow on hillslopes, EOS T. Am. Geophys. Un., 77, 465–472, https://doi.org/10.1029/96EO00306, 1996.

Mugler, A., Grinshpun, B., Franks, R., and Wiggins, C. H.: Statistical method for revealing form-function relations in biological networks, P. Natl. Acad. Sci. USA, 108, 446–451, https://doi.org/10.1073/pnas.1008898108, 2011.

Nadezhdina, N., David, T. S., David, J. S., Ferreira, M. I., Dohnal, M., Tesar, M., Gartner, K., Leitgeb, E., Nadezhdin, V., Cermak, J., Jimenez, M. S., and Morales, D.: Trees never rest: the multiple facets of hydraulic redistribution, Ecohydrology, 3, 431–444, https://doi.org/10.1002/eco.148, 2010.

Neal, C. and Rosier, P. T. W.: Chemical studies of chloride and stable oxygen isotopes in two conifer afforested and moorland sites in the British uplands, J. Hydrol., 115, 269–283, https://doi.org/10.1016/0022-1694(90)90209-G, 1990.

Öhrström, P., Hamed, Y., Persson, M., and Berndtsson, R.: Characterizing unsaturated solute transport by simultaneous use of dye and bromide, J. Hydrol., 289, 23–35, https://doi.org/10.1016/j.jhydrol.2003.10.014, 2004.

Palm, J., van Schaik, L., and Schröder, B.: Modelling distribution patterns of anecic, epigeic and endogeic earthworms at catchment-scale in agro-ecosystems, Pedobiologia, 56, 23–31, https://doi.org/10.1016/j.pedobi.2012.08.007, 2012.

Pfister, L. and Hoffmann, L.: Experimental hydro-climatological atlas of the Alzette river basin, Centre de Recherche Public – Gabriel Lippmann, 2002.

Rawls, W. J., Brakensiek, D. L., and Saxtonn, K. E.: Estimation of Soil Water Properties, T. ASAE, 25, 1316–1320, https://doi.org/10.13031/2013.33720, 1982.

Reggiani, P., Sivapalan, M., and Hassanizadeh, S. M.: A unifying framework for watershed thermodynamics: balance equa-

tions for mass, momentum, energy and entropy, and the second law of thermodynamics, Adv. Water Resour., 22, 4, 367–398, https://doi.org/10.1016/S0309-1708(98)00012-8, 1998.

Roth, K.: Steady State Flow in an Unsaturated, Two-Dimensional, Macroscopically Homogeneous, Miller-Similar Medium, Water Resour. Res., 31, 2127–2140, https://doi.org/10.1029/95WR00946, 1995.

Sander, T. and Gerke, H. H.: Modelling field-data of preferential flow in paddy soil induced by earthworm burrows, J. Contam. Hydrol., 104, 126–136, https://doi.org/10.1016/j.jconhyd.2008.11.003, 2009.

Schaap, M. G., Leij, F. J., and Van Genuchten, M. T.: Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions, J. Hydrol., 251, 163–176, 2001.

Schlüter, S., Berg, S., Rücker, M., Armstrong, R. T., Vogel, H.-J., Hilfer, R., and Wildenschild, D.: Pore-scale displacement mechanisms as a source of hysteresis for two-phase flow in porous media, Water Resour. Res., 52, 2194–2205, https://doi.org/10.1002/2015WR018254, 2016.

Schotanus, D., van der Ploeg, M. J., and van der Zee, S. E. A. T. M.: Quantifying heterogeneous transport of a tracer and a degradable contaminant in the field, with snowmelt and irrigation, Hydrol. Earth Syst. Sci., 16, 2871–2882, https://doi.org/10.5194/hess-16-2871-2012, 2012.

Sivapalan, M.: Pattern, Process and Function: Elements of a Unified Theory of Hydrology at the Catchment Scale, Enc. Hydrol. Sci., 2005.

Sprenger, M., Volkmann, T. H. M., Blume, T., and Weiler, M.: Estimating flow and transport parameters in the unsaturated zone with pore water stable isotopes, Hydrol. Earth Syst. Sci., 19, 2617–2635, https://doi.org/10.5194/hess-19-2617-2015, 2015.

Sprenger, M., Seeger, S., Blume, T., and Weiler, M.: Travel times in the vadose zone: Variability in space and time, Water Resour. Res., 52, 5727–5754, https://doi.org/10.1002/2015WR018077, 2016.

Steelman, C. M., Endres, A. L., and Jones, J. P.: High-resolution ground-penetrating radar monitoring of soil moisture dynamics: Field results, interpretation, and comparison with unsaturated flow model, Water Resour. Res., 48, W09538, https://doi.org/10.1029/2011WR011414, 2012.

Tetzlaff, D., Carey, S., Laudon, H., and McGuire, K.: Catchment processes and heterogeneity at multiple scales – benchmarking observations, conceptualization and prediction, Hydrol. Proc., 24, 2203–2208, https://doi.org/10.1002/hyp.7784, 2010.

Thompson, D. W.: On growth and form, Cambridge University Press, 1917.

Tian, F., Hu, H., Lei, Z., and Sivapalan, M.: Extension of the Representative Elementary Watershed approach for cold regions via explicit treatment of energy related processes, Hydrol. Earth Syst. Sci., 10, 619–644, https://doi.org/10.5194/hess-10-619-2006, 2006.

Trinks, I., Stümpel, H., and Wachsmuth, D.: Monitoring water flow in the unsaturated zone using georadar, First Break, 19, 679–684, 2001.

Tromp-van Meerveld, H. and McDonnell, J. J.: Threshold relations in subsurface stormflow: 1. A 147-storm analysis of the Panola hillslope, Water Resour. Res., 42, W02410, https://doi.org/10.1029/2004WR003778, 2006.

Truss, S., Grasmueck, M., Vega, S., and Viggiano, D. A.: Imaging rainfall drainage within the Miami oolitic limestone using high-resolution time-lapse ground-penetrating radar, Water Resour. Res., 43, 1–15, https://doi.org/10.1029/2005WR004395, 2007.

Uhlenbrook, S.: Catchment hydrology – a science in which all processes are preferential, Hydrol. Proc., 20, 3581–3585, https://doi.org/10.1002/hyp.6564, 2006.

van den Bos, R., Hoffmann, L., Juilleret, J., Matgen, P., and Pfister, L.: Regional runoff prediction through aggregation of first-order hydrological process knowledge: a case study, Hydrol. Sci. J., 51, 1021–1038, https://doi.org/10.1623/hysj.51.6.1021, 2006.

van Schaik, L.: Spatial variability of infiltration patterns related to site characteristics in a semi-arid watershed, CATENA, 78, 36–47, https://doi.org/10.1016/j.catena.2009.02.017, 2009.

van Schaik, L., Palm, J., Klaus, J., Zehe, E., and Schröder, B.: Linking spatial earthworm distribution to macropore numbers and hydrological effectiveness, Ecohydrology, 7, 401–408, https://doi.org/10.1002/eco.1358, 2014.

Vogel, H.-J. and Roth, K.: Moving through scales of flow and transport in soil, J. Hydrol., 272, 95–106, 2003.

Vogel, H.-J., Cousin, I., Ippisch, O., and Bastian, P.: The dominant role of structure for solute transport in soil: experimental evidence and modelling of structure and transport in a field experiment, Hydrol. Earth Syst. Sci., 10, 495–506, https://doi.org/10.5194/hess-10-495-2006, 2006.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P.: Image Quality Assessment: From Error Visibility to Structural Similarity, IEEE T. Image Process., 13, 600–612, https://doi.org/10.1109/TIP.2003.819861, 2004.

Wassenaar, L. I., Hendry, M. J., Chostner, V. L., and Lis, G. P.: High Resolution Pore Water $\delta 2$H and $\delta 18$O Measurements by H$_2$O (liquid) – H$_2$O (vapor) Equilibration Laser Spectroscopy, Environ. Sci. Technol., 42, 9262–9267, https://doi.org/10.1021/es802065s, 2008.

Wehrer, M. and Slater, L. D.: Characterization of water content dynamics and tracer breakthrough by 3-D electrical resistivity tomography (ERT) under transient unsaturated conditions, Water Resour. Res., https://doi.org/10.1002/2014WR016131, 2015.

Western, A. W. and Blöschl, G.: On the spatial scaling of soil moisture, J. Hydrol., 217, 203–224, https://doi.org/10.1016/S0022-1694(98)00232-7, 1999.

Wienhöfer, J., Germer, K., Lindenmaier, F., Färber, A., and Zehe, E.: Applied tracers for the observation of subsurface stormflow at the hillslope scale, Hydrol. Earth Syst. Sci., 13, 1145–1161, https://doi.org/10.5194/hess-13-1145-2009, 2009

Wittgenstein, L.: Tractatus logico-philosophicus, Kegan Paul, Trench, Trubner & Co. Ltd., 1922.

Wrede, S., Fenicia, F., Martínez-Carreras, N., Juilleret, J., Hissler, C., Krein, A., Savenije, H. H. G., Uhlenbrook, S., Kavetski, D., and Pfister, L.: Towards more systematic perceptual model development: a case study using 3 Luxembourgish catchments, Hydrol. Proc., 29, 2731–2750, https://doi.org/10.1002/hyp.10393, 2015.

Zehe, E., Maurer, T., Ihringer, J., and Plate, E.: Modeling water flow and mass transport in a loess catchment, Phys. Chem. Earth Pt. B, 26, 487–507, https://doi.org/10.1016/S1464-1909(01)00041-7, 2001.

Zehe, E., Ehret, U., Blume, T., Kleidon, A., Scherer, U., and Westhoff, M.: A thermodynamic approach to link self-organization, preferential flow and rainfall-runoff behaviour, Hydrol. Earth Syst. Sci., 17, 4297–4322, https://doi.org/10.5194/hess-17-4297-2013, 2013.

Zehe, E., Ehret, U., Pfister, L., Blume, T., Schröder, B., Westhoff, M., Jackisch, C., Schymanski, S. J., Weiler, M., Schulz, K., Allroggen, N., Tronicke, J., van Schaik, L., Dietrich, P., Scherer, U., Eccard, J., Wulfmeyer, V., and Kleidon, A.: HESS Opinions: From response units to functional units: a thermodynamic reinterpretation of the HRU concept to link spatial organization and functioning of intermediate scale catchments, Hydrol. Earth Syst. Sci., 18, 4635–4655, https://doi.org/10.5194/hess-18-4635-2014, 2014.