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

Conrad Jackisch[1], Lisa Angermann[2], 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 3D time-lapse ground penetrating radar (GPR) methods. At the hillslope scale the subsurface was explored by 3D GPR survey. A natural storm event and an irrigation experiment were monitored by a dense network of soil moisture observations and a cascade of 2D 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. 2D and 3D 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., 2016, 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) lies at the core of our understanding, exploration and modeling. It manifests itself as patterns in dynamics (Sivapalan, 2005) and self-organization (Zehe et al., 2013). This interplay can be expressed as form-function relationship, which is addressed in many disciplines. Especially in systems biology the form-function relations are deeply rooted (e.g. Thompson, 1917) and under debate until today (e.g. Mugler et al., 2011). In logical and philosophical terms the separation of form and function is fundamental for the concept that we can predict the behavior of a system under different forcing by knowing its constructive properties. Already Wittgenstein (1922) highlighted that it is form and content which make up the substance configuring the world and that form exhibits the possibility of structure. Although it is generally agreed on the existence and importance of form-function relationships, 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 pores size distribution and their co-variance structure to storage of water. The hydraulic conductivity curve relates the pores size distribution and the interconnectedness of the pores to conductance/release function of water. These are classic examples of form-function relations at the Darcy-scale. However, this requires revision at the actual pore-scale (Armstrong et al., 2016). When we scale-up, preferential flow paths become important as has been observed in most catchments (Flury et al., 1994; Uhlenbrook, 2006). Also here the Darcy-scale form-function relations become of limited use when macropore density, connectivity, initiation and interaction with the soil matrix are main controls of flow and redistribution. In either case determining topology and connectivity (form) and understanding their implication on soil water transport (function) is seen as "forefront of multiphase flow research" (Armstrong et al., 2016).

It is a long-standing vision in eco-hydrology to observe and characterize 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 to measure and visualize 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 pedon-, plot- and hillslope-scale.

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

Preferential or rapid subsurface flow takes place through structures that add directed drainage paths to the already often highly heterogeneous setting of the soil. Channelled through biogenic structures, such as earthworm burrows (Palm et al., 2012; Blouin et al., 2013; van Schaik et al., 2014) and plant roots (Nadezhdina et al., 2010), and geogenic structures like voids in periglacial cover beds (Heller, 2012), it is partially bypassing large sections of the soil. Beven and Germann (1982) initiated a discussion about macropores and preferential flow and recently resumed 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 breakthrough (Schotanus et al., 2012; Klaus et al., 2013) or in double peak 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 "old water paradox" (Kirchner, 2003). Analyses of the integral responses inspired a multitude of studies in experimental basins around the globe and studies based on models.

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. This technique provides valuable information but requires strong assumptions about macropore-matrix interaction, time of fixation and recoverability. 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.

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 and salt tracers they can also provide quantitative information over the course of rapid flow events at this scale (e.g. Wienhöfer et al., 2009). However, also such measurements can only capture integral signals and require to infer the form by the observed function.

So far, relatively few studies managed to actually observe the internal, spatially distributed flow paths. On the one hand, applicability is also often limited to very small scales: Schlüter et al. (2016) examined multiphase flow with time-lapse X-ray microtomography in a sample of $4.2\,\text{ml}$. Koestel and Larsbo (2014) presented an X-ray tomography study with a sample of $258\,\text{ml}$ undisturbed soil. Gerke (2012) analyzed the pore fractions in two samples of $785\,\text{ml}$ undisturbed soil through a medical CT X-ray scanner. Wehrer and Slater (2015) report findings from tracer breakthrough experiments in laboratory lysimeters accompanied by 3D time-lapse electrical resistivity tomography (ERT). Guo et al. (2014) conducted a multi-2D 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 structured entities. 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) highlight that further analyses need to focus on the quantification of flow-relevant structures. They continue that exper-

imentally 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 infer flow-relevant subsurface structures in a qualitative and quantitative sense. Specifically, we hypothesize that a combination of quantitative field methods and in situ imaging of subsurface response patterns with time-lapse GPR at three scale levels provides the missing link between form and function of flow paths and rapid subsurface flow.

We test this hypothesis by addressing three main research questions:

Q1 What kind of information on flow-relevant structures, their characteristics and their distribution can be inferred from direct measurements using a large set of soil core profiles, soil core samples, permeameter measurements and a GPR survey?

Q2 How do salt tracer data, dye tracer patterns, and soil moisture responses from plot- and hillslope-scale irrigation experiments compare with localized imaging with 2D and 3D time-lapse GPR?

Q3 How do identified flow-relevant structures and estimates of vertical response velocities compare between the different methods?

The study is approaching the topic from the aspect of *form* while the alternative starting point *function* is taken in the companion study by (Angermann et al., 2016, this issue).

4

2 Experimental approaches and study methods

The study at hand approached the topic on three complementary scales with a large number of methods: As standard reference, results from auger exploration and in-situ measurements of hydraulic conductivity and infiltration capacity were collected. They were extended with laboratory examination of 250 ml undisturbed ring samples for bulk density, porosity, texture, soil

5   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 1D profile, 3D time-lapse GPR imaging, and tracer recovery of dye, salt and stable isotopes. At the hillslope-scale, 3D 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

10   trench-like 2D time-lapse GPR transects.

2.1 Study site description

The study is situated in the headwaters of the Colpach river which is 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

15   the schistose Ardennes Massif the soils are characterised 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). Our measurements and experiments focus on two forested hillslopes (mostly managed stands of *Fagus sylvatica* with mixed shrubs, some measurements took place in stands of *Picea abies*). The agriculturally used plateaus

20   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 at three hillslopes. The sampling is guided by a network of hydro-meteorological

25   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 sensor clusters infiltration capacity and saturated hydraulic conductivity was measured. In order to address plot-scale (few meters) and hillslope-scale (few 100 meters) heterogeneity, the design consisted of clustered sets of point

30   measurements along two catenas plus one set at the site of the hillslope irrigation experiment presented in section 2.4. A detailed map is included in appendix figure 13.
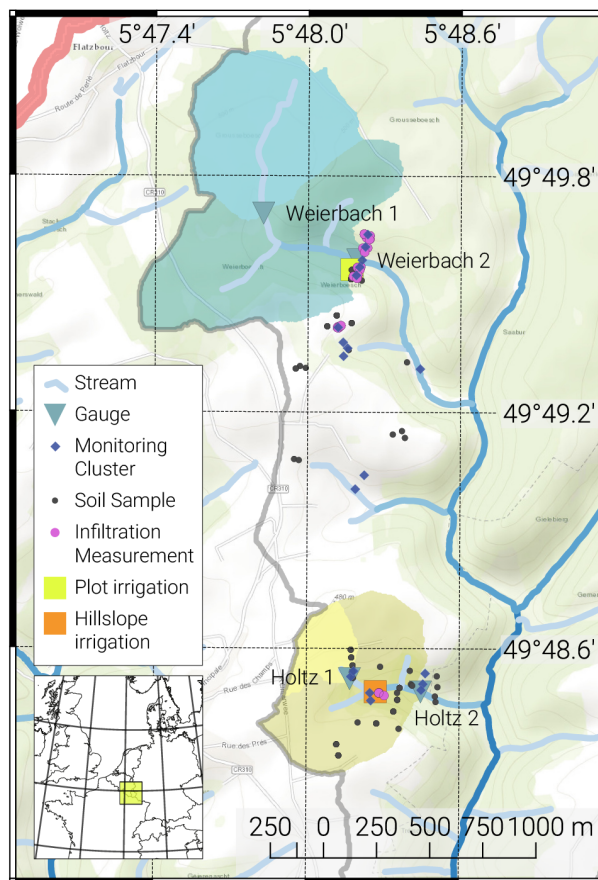
Figure 1. Map of the study sites in the upper Attert basin, Luxembourg.

The distance between the clustered sets was $80\,\mathrm{m}$ to $200\,\mathrm{m}$. In each three nested sets with a lag distance of $10\,\mathrm{m}$ to $20\,\mathrm{m}$ 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 $1\,\mathrm{m}$) 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.

5   In addition to the point measurements a series of percussion drilled profiles (drill head diameter of $4\,\mathrm{cm}$) as 1D profiles were drawn and $250\,\mathrm{mL}$ ring samples were taken within the top $0.6\,\mathrm{m}$ 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 ($12.4\,\mathrm{cm}$ radius) as infiltration water supply. Inside the chamber, a defined low negative pressure head is established,

10   which allows a precise measurement of infiltration capacity at different tensions. 3 to 5 tension levels between 0 and $5.5\,\mathrm{cm}$ water column were applied at each spot.

6

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 to 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 $5\,cm$ diameter). The outflow is measured to calculate

5   saturated hydraulic conductivity (k_{sat}) (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

10   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 figure 1 the three plots are located on a forested mid slope in direct vicinity of observation station E (see also appendix figure 13).

2.3.1 Experimental design and multi-method approach

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

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

20   reference, respectively.

In addition, temporal dynamics of soil moisture along a selected profile was monitored throughout the experiments 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

25   an integral of about 1.05L (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 seconds. Hence the whole procedure added up to $4\,min$ to $10\,min$ 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) that film flow in soil structures equilibrates gravity and thus the main fraction of advective flow is dispersed after 1.5 times the duration of a constant input rate.

30   $2\,h$ 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 3 vertical faces into the plot (aligned with the slope line, $0.1\,m$ distance starting from the lateral edge) and 5-7 horizontal cuts in different depth levels down to the first deposit layer ($0.5 \times 0.5\,m^2$ in the center
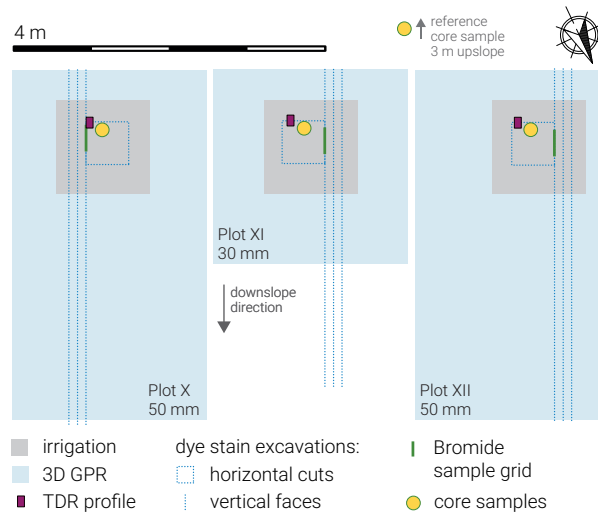
Figure 2. Plan view layout of the plot-scale irrigation experiments. Three irrigation plots ($1\,\text{m}^2$, gray squares) are monitored by 3D time-lapse GPR (blue rectangles) and TDR (soil moisture tube probe, red box). The plots are sampled for tracer recovery by percussion drilled core samples (yellow 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 pore water stable isotope composition is sampled as fourth core $3\,\text{m}$ upslope.

of the plot). On the third vertical face in the center of the plot mini core samples of $66\,\text{mL}$ soil were taken in a $5\,\text{cm}$ grid with 5 columns and 14 to 21 rows. In order to minimize time lags in the 70 to 105 individual samples a quick sampler was developed allowing for precise and nearly undisturbed core sampling.

2.3.2 Bromide recovery and stable isotope analysis

5   All samples were analyzed for Bromide (Br^-). This was done by oven drying the samples and consecutively suspending them in $150\,\text{mL}$ de-ionised water ($72\,\text{h}$ in overhead shaker at 9 rotations per minute). The samples were then left 4 days for sedimentation to exfiltrate the excess through a) filtration paper ($5\,\mu\text{m}$ to $13\,\mu\text{m}$) and b) $0.45\,\mu\text{m}$ PP micro-filter. 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.

10   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 ($\delta^{18}O$ and δ^2H) of the pore water. See appendix B for details and results.

2.3.3 Calculation of apparent vertical flow velocity

The quantitative measurements allow 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 individual 1D 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:

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

For this the un-interpolated measurements were used.

2.3.5 3D 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 3D 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 3D 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 alone is not suitable for analyzing time-lapse GPR data in the presence of limited repeatability and noisy data. They propose a structural similarity attribute, which basically calculates the crosscorrelation of the residuals normalized by the product of the standard deviations of two moving windows at the same positions and different acquisition times. In our study we calculate the structural similarity attribute 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 axes. The attribute ranges between 1 and -1 with 1 being highly similar and -1 referring to most dissimilar.

As additional estimate of vertical response velocities the same approach as for the soil moisture responses (section 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 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. With reference to its hydrological responses (companion paper Angermann et al., 2016, this issue), vegetation, slope, soils and hydraulic properties we consider the hillslopes to be very similar.

2.4.1 3D GPR survey of the hillslope

As additional reference to the soil core profiles a 3D GPR survey of the hillslope was conducted prior to 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 3D topographic migration approach as presented by Allroggen et al. (2015a), using an appropriate constant velocity of $0.07 \, \text{m ns}^{-1}$.

For structural analysis, the processed data are imported into the OpenDtect software (dGB Earth Sciences). Under heterogeneous 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 3D 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 A priori model reference

Based on the findings of the pedo-physical exploration, we setup the 2D 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 assuming periglacial deposit layers as conductive layers in the hillslope. In a series of scenarios, the one with $30\,\mathrm{mm\,h^{-1}}$ irrigation for $4\,\mathrm{h}$ 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 $12\,\mathrm{h}$.

5 2.4.3 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 removed from the experimental site to accomplish GPR measurements and allow for undisturbed and homogeneous irrigation. The topographic gradient is about $14°$.
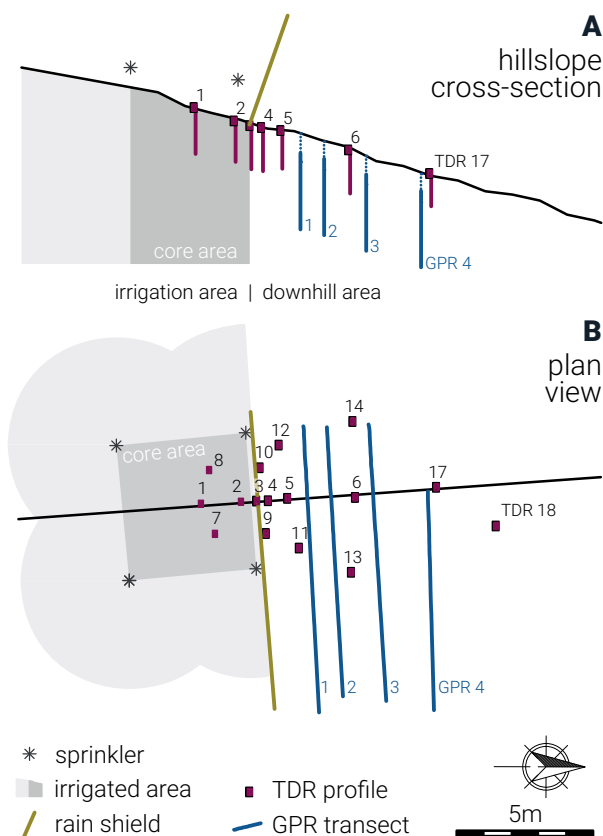
Figure 3. Layout of hillslope-scale irrigation experiment as vertical view (A) and plan view (B). The hillslope is divided into an irrigation area and a downhill area by a rain shield. 16 access tubes for TDR measurements of soil moisture profiles are arranged in 3 diverting transects. Parallel to the contour lines 4 transects of 2D time-lapse GPR are recorded.

The experimental layout is given in figure 3. The irrigation of $141\,\mathrm{mm}$ in $4.5\,\mathrm{h}$ was fed from stream water and was realized by four circular sprinklers (Wobbler, Senninger Irrigation Inc.) arranged to overlap at a $5\,\mathrm{m}$ by $5\,\mathrm{m}$ core area with relatively homogeneous intensity. While boundary effects were mitigated by an irrigated buffer zone of about $4\,\mathrm{m}$ 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 absolute 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 $2\,\mathrm{m}$ of the lower boundary of the core area. It was built from a plastic sheet installed approximately $1\,\mathrm{cm}$ 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 TDR sensors like in the plot scale experiments (Imko GmbH, two with $12\,\mathrm{cm}$ integration depth and one with $18\,\mathrm{cm}$). Measurements required manual positioning of the sensor probes for each reading. We continuously recorded the states in all tubes in $10\,\mathrm{cm}$ depth increments realizing revisiting intervals of $5\,\mathrm{min}$ to $20\,\mathrm{min}$. The tubes were installed to reach to a depth of about $1.7\,\mathrm{m}$. 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 downslope of the rain shield, and the furthest profile about $9\,\mathrm{m}$ downhill.

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

2.4.4 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 was 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 synchronised version of the records in $0.1\,\mathrm{m}$ depth and $15\,\mathrm{min}$ 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 ($4\,\mathrm{min}$ to $10\,\mathrm{min}$ 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 section 4.4.1.

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 $1\,\mathrm{m}$ 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.

5   As in the plot irrigation experiments, vertical response velocities are calculated for the TDR profiles at the core area. The calculation of lateral response velocities is given in the companion study (Angermann et al., 2016, this issue).

2.4.5   GPR transects and structural similarity attribute interpretation

The 2D time-lapse GPR data is derived from 9 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,

10   inline fk-filtering, topographic migration with constant velocity ($0.07\,\mathrm{m\,ns^{-1}}$), and consecutive gridding to a 2D transect with regular trace-spacing of $0.02\,\mathrm{m}$.

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 3D time-lapse GPR applications, the radargrams in the young,

15   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 presented by Allroggen and Tronicke (2016). It is calculated using a local Gaussian window of $2\,\mathrm{ns}$ along the vertical axis and $0.06\,\mathrm{m}$ along the horizontal axis.

Due to the presence of remaining event water from the preceding storm event (Angermann et al., 2016, this issue), all

20   measurements are referenced to the last acquisition time approximately $23\,\mathrm{h}$ after irrigation start and about $19\,\mathrm{h}$ after irrigation. The resulting structural similarity attribute images are used as a qualitative indicator for relative deviations from the reference state. Points of low similarity indicate deviations that arise from changes in dielectric permittivity which likely reflect changes in local soil water content.

2.4.6   Discriminating the natural storm event and the irrigation experiment

25   The experiment was preceded by two strong storm events of $43\,\mathrm{mm}$ in total on June 20, 2013. In reference to the gauge reaction the experiment was conducted shortly before the major peak of the resulting runoff reaction (Please see figure 5 in Angermann et al., 2016, for details). Accordingly, the structural similarity attributes 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 $7.5\,\mathrm{h}$ before and directly at irrigation start are attributed to the

30   natural event and show an increasing structural similarity (accordance with the reference state). Once the attribute value of a pixel decreases again (lower structural similarity) it is attributed to the irrigation. For stability reasons, a threshold of 0.15 was introduced, which is discussed in appendix D.

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 to simple morphological descriptors like depth, hillslope position or topographic flow gradient (details given
in the appendix figure 13). Infiltration capacity ranged between $5 \times 10^{-5}\,\mathrm{m\,s^{-1}}$ to $5 \times 10^{-3}\,\mathrm{m\,s^{-1}}$. The values for saturated
hydraulic conductivity ranged from $1 \times 10^{-8}\,\mathrm{m\,s^{-1}}$ 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 a pattern of elevated conductivity in about $0.6\,\mathrm{m}$ depth
was found. The strong heterogeneity and large spread of values was also depicted in the analyses of undistorted soil samples
(figure 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 showed a strong positive
bias with respect to reported literature and text book values (compare figure 4 middle row of panels). Accordingly, the range
of porosity was positively biased and the range of bulk density lies below literature references. All measurements exhibited a
large spread of values which does not correlate well with simple morphological variables like depth or hillslope position.

3.1.2 Soil core profile snapshots

The soil core profiles (figure 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 $0.3\,\mathrm{m}$. The gravel content gradually increased over depth in the Bw-horizon and further increased in the C-horizon, starting
between $0.4\,\mathrm{m}$ and $1.1\,\mathrm{m}$ depth. Below the depth of $0.5\,\mathrm{m}$ scattered layers of weathered rock with usually horizontal orientation
were found in some soil cores. Percussion drilling was often inhibited at a depth between $1.5\,\mathrm{m}$ and $2.0\,\mathrm{m}$ (lower end of the
bars in figure 5) due to even higher stone content with a more and more vertical orientation of the weathered rocks. In core
T7, concretions of iron and manganese oxides were found in the depth between $1.6\,\mathrm{m}$ and $1.9\,\mathrm{m}$ below ground, indicating
hydromorphic conditions.

Based on these standard techniques the overall setting of a heterogeneous silty soil deviating from expected low hydraulic
conductivity was revealed. So far 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 prop-
erties could be derived. However, details about its spatial organization and the detection of specific and potentially continuous
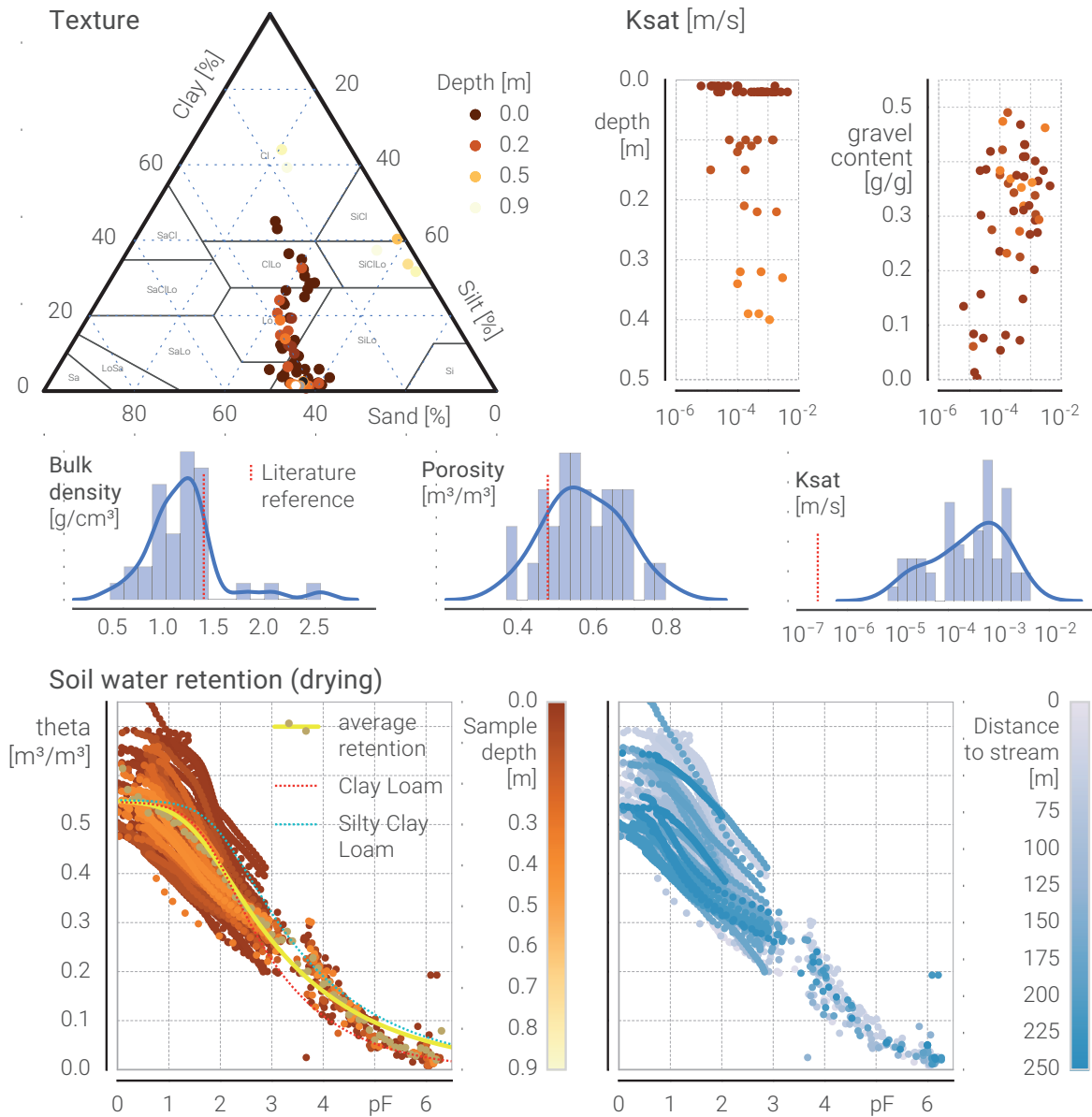structures remained obscured by high heterogeneity.

Figure 4. Results of laboratory analyses of 63 undisturbed 250 ml ring samples. **Top row:** Soil texture analysed with wet sieving and sedimentation (pipette method). Saturated hydraulic conductivity (Ksat) measured with the Ksat apparatus and plotted against sample depth and gravel content. **Mid row:** Histograms and kernel density estimate of measured bulk density, porosity and Ksat. Reference as mean silty loam value from literature (Hillel, 1980; Rawls et al., 1982; Carsel and Parrish, 1988; Schaap et al., 2001). **Bottom row:** Soil water retention relation measured with the Hyprop and the WP4C apparatus with average retention estimate (respective mean of each 0.05pF-bin and fitted van Genuchten model) and literature references (Carsel and Parrish, 1988) scaled to measured average θ_s and θ_r.
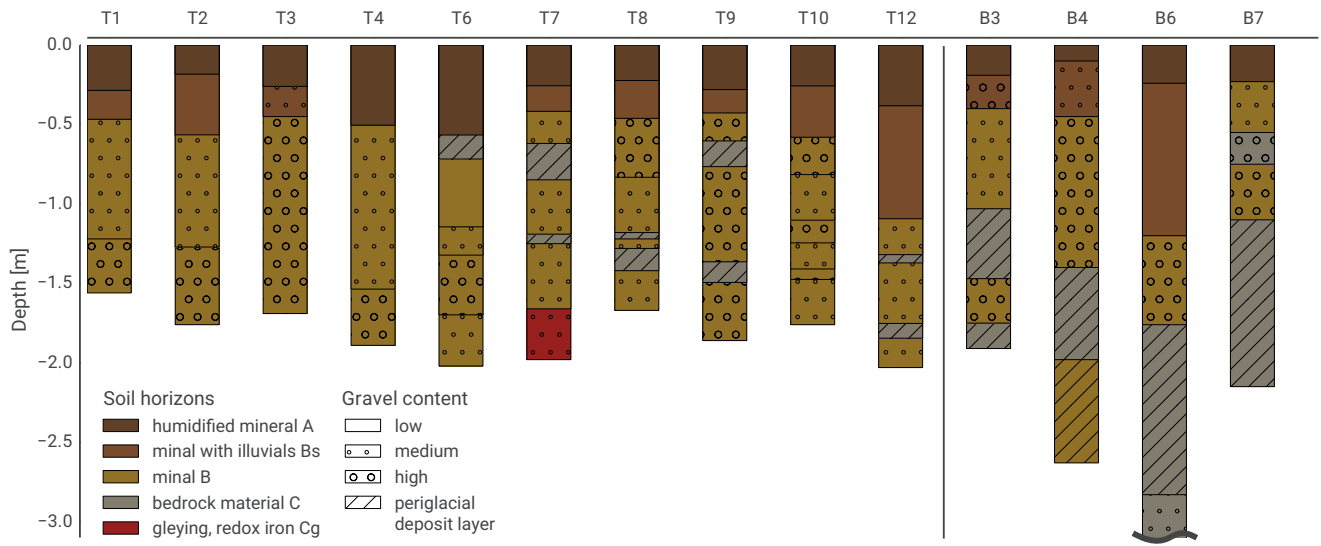
Figure 5. Soil core profiles from the upper Colpach River basin. The T-nomenclature indicates cores from the installation of the TDR access tubes at the Holtz hillslope experiment. B corresponds to examples at the catena of the plot experiments. Bar depth is the maximum drilling depth of the cores restricted 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 (figure 6). For all experiments stained patches were found down to
5   the periglacial deposit layer in $0.6\,\mathrm{m}$ to $0.8\,\mathrm{m}$ depth. During the excavation apparently isolated dye traces were recovered even several meters downhill the irrigation spot ($4\,\mathrm{m}$ downslope, $1\,\mathrm{m}$ deep). The stains did not reveal a network of large macropores but an irregular mesh of connected inter-aggregate voids. This explains the positive bias in the soil physical analyses.

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
10   of Bromide tracer recovery (figure 7, top row). All plots suggested a relatively strong response in the depth of approximately $0.6\,\mathrm{m}$. This depth correlates with the upper boundary of the first layer of periglacial deposits found in core profile B7 (figure 5) and in the excavated soil profiles. This response is contrasted by low Bromide concentration in shallower depth. Even plot XI, where only $30\,\mathrm{mm}$ 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 very low and even decreased when including the core samples once more pointing to heterogeneity.
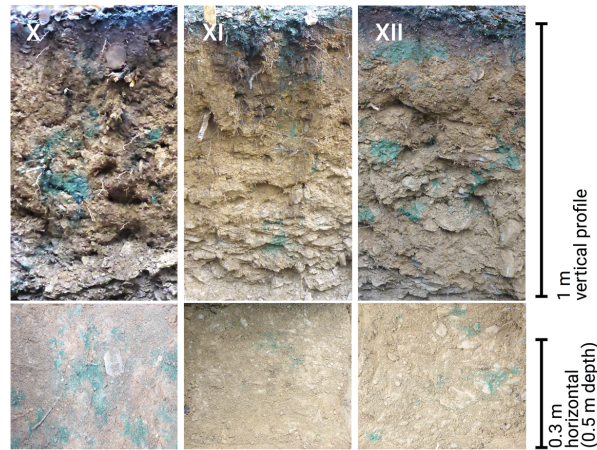

Figure 6. Recovered dye patterns in plot irrigation experiments. Top row excavated vertical faces, bottom row horizontal cuts in 0.5 m depth.

3.2.3 Quick soil moisture response in greater depth

The observed soil moisture changes (figure 7, bottom row) corroborated the results from the tracers. Especially at plot X and XII we found a relatively quick and strong response in 0.7 m 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 off set from the main flow field (figure 15 in appendix C).

3.2.4 3D view on soil water redistribution

The structural similarity attribute of the 3D time-lapse GPR measurements provided qualitative information of changes 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 (figure 8, and figure 15 and 16 in appendix C). Also here, strongest deviations were recorded in the mid horizon between 40 and 60 ns corresponding to approximately 0.7 to 1 m depth. The top horizon between 20 and 40 ns (0.35 to 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
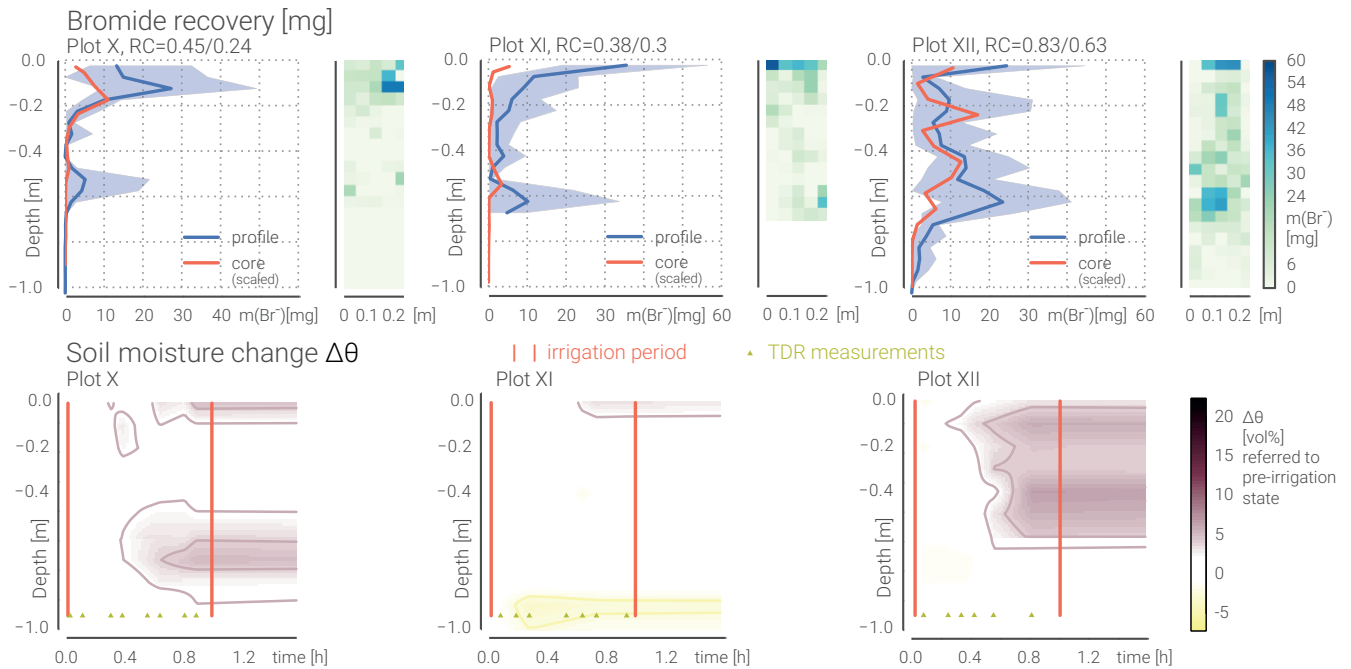
Figure 7. Results from plot-scale irrigation experiments with 50 mm, 30 mm and 50 mm spray irrigation for 1 h. **Top:** Recovered Bromide mass profiles and grids (5x5cm). Blue line as mean and shaded area between min/max for each depth of the sampling grid. Orange line is mass recovered in drilled profile samples (scaled to same volume reference). Recovery coefficient (RC) calculated for the profile samples (first value) and the profile and core samples (second value). **Bottom:** Observed soil moisture change referenced to the first measurement shortly before onset of the irrigation. Individual measurement times marked with triangles.

a slight lateral component was recorded. Plot XII had very high similarity between first and third acquisition. Thus stronger macropore-matrix interaction and dispersive redistribution is corroborated.

The contrasting attribute distributions over time and comparing plots X and XII did not only reveal heterogeneity. It also highlighted the qualitative nature of the analytical method of the GPR data. Although visual interpretation of the radargrams (top row in figures 8, 15 and 16) 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 figure 9). The many point-scale measurements (left panel based on 63 ring samples, 40 infiltrometer points,
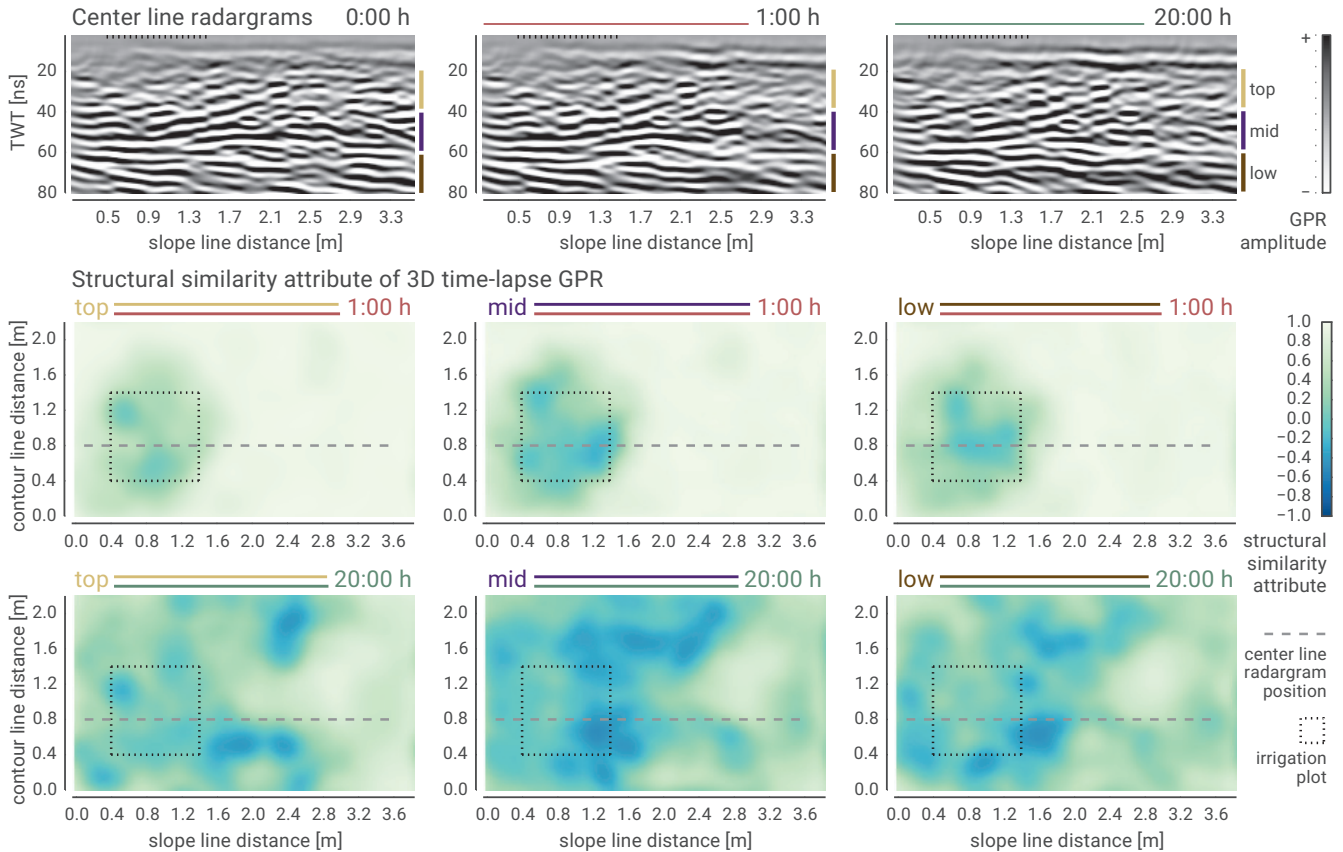
18

Figure 8. Time-lapse 3D GPR of irrigation experiment at plot X. Center line radargrams at the marked transect (grey dashed line in lower panels) for the three acquisition times (before 0:00 h, directly after irrigation 1:00h, 20:00 h after irrigation) are given in the top row. The structural similarity attribute of the 3D data cube is given in three different depth layers (top 20 ns to 40 ns, mid 40 ns to 60 ns, low 60 ns to 80 ns) in the lower panels. The irrigation plot is marked by a black dashed box/line. Slope line distance is increasing downslope.

102 individual permeameter measurements) resulted in rather inconsistent 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 $1 \times 10^{-3.5}\,\mathrm{m\,s^{-1}}$ (figure 9 right panel based on Bromide recovery with an estimated time of fixation (t_{fix}) after 1.5 h, first excess of TDR recorded soil moisture $\geq 2\mathrm{vol}\%$, and GPR structural similarity attributes below zero between pre-irrigation and first post-irrigation records).

All results ranged several orders of magnitude above the literature reference of $2.5 \times 10^{-7}\,\mathrm{m\,s^{-1}}$ (mean of reported values for silt and silty loam Hillel, 1980; Rawls et al., 1982; Carsel and Parrish, 1988; 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.
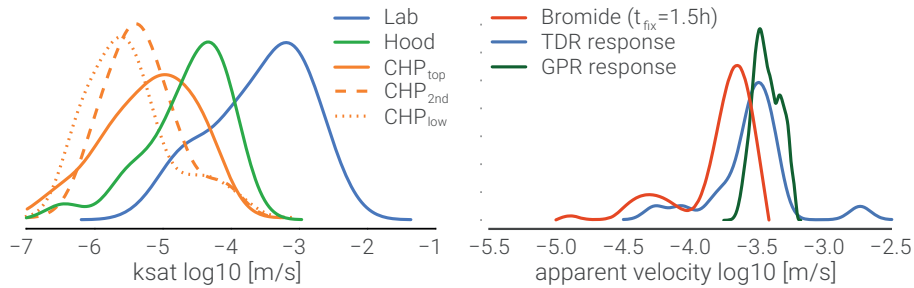
Figure 9. Saturated hydraulic conductivity and apparent vertical flow velocity kernel density estimates. Left panel: Point-scale measurement results; Right panel: Results from plot-scale irrigation experiments.

3.3 Hillslope scale detection of lateral flow paths

3.3.1 3D GPR survey suggests a fragmentary layer

The 3D GPR survey at the site of the hillslope experiment identified fragmented structures in about $1.5\,\mathrm{m}$ depth (figure 10). This is in accordance with the soil core profile depth (figure 5, T1..T12). Especially profile T7 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.

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 2D GPR transects. The change of soil moisture at the core area (TDR 2 and 8 in figure 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 in 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 figure 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 figure 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 divers, it remains rather ambiguous to determine their lateral connection. Overall changes in soil moisture as maximum at each TDR profile did not corroborate the potential subsurface structures identified in the 3D GPR survey (figure 10). The full set of profiles are reported in the companion study (Angermann et al., 2016).

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
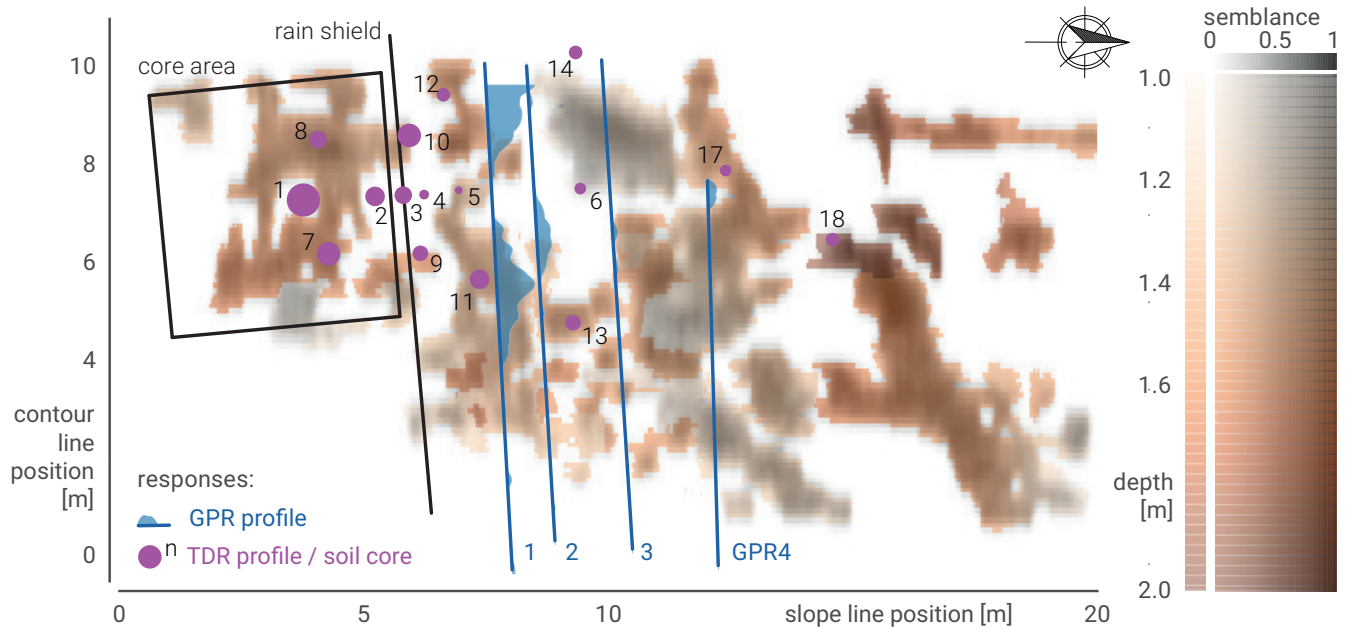
Figure 10. Potential subsurface structures from 3D GPR survey and setup of hillslope experiment. Structure identification guided by the dip corrected semblance attribute. Depth estimated based on mean measured effective radar velocity in soil of $0.07\,\mathrm{m\,ns^{-1}}$.

Summary of the hillslope experiment given by locations of TDR profile tubes (purple, also location of respective soil cores in figure 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.

(blue) revealed distinct, heterogeneously distributed patters (figure 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. Areas of high temporal dynamics of the similarity attribute were identified as regions of such flow-relevant structures (figure 12, bottom row). Note that the last recorded difference 18 h 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 transect 1 and 3 respectively. Apparently, the system had reached a steady state without much further change in soil moisture.

The patchy structures at the transects highlighted the irregularly distributed nature of preferential flow paths which was already 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 downhill direction is
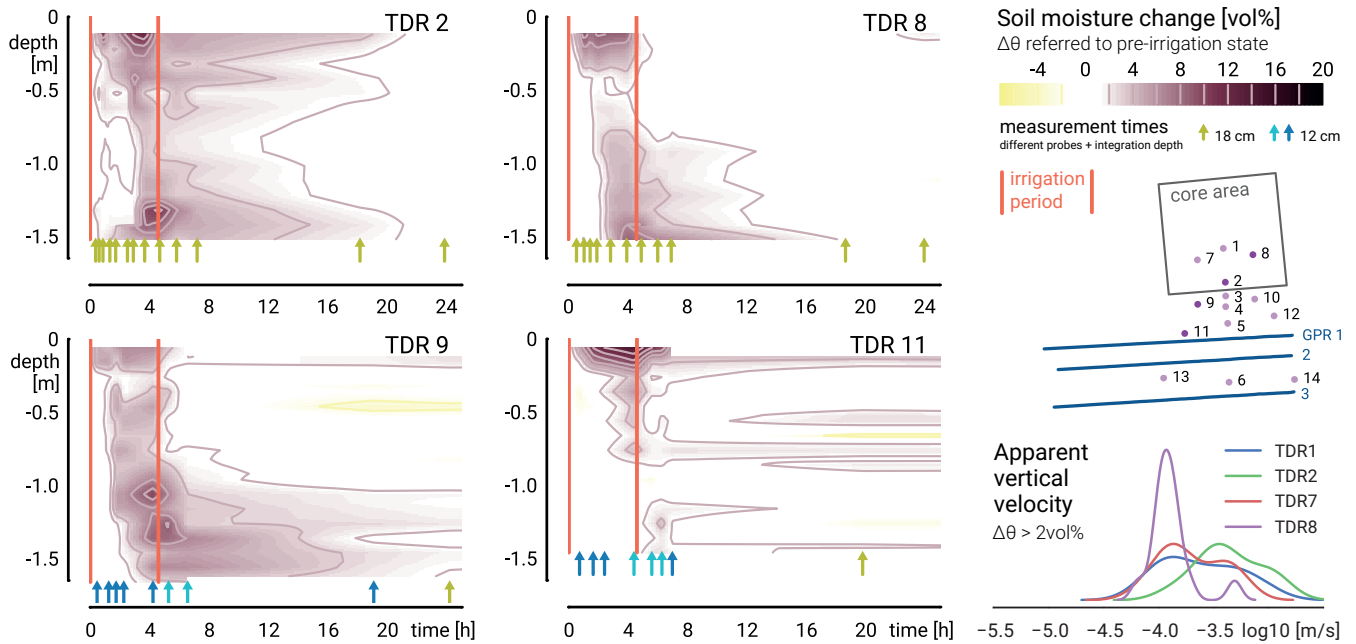
21

Figure 11. Development of soil moisture in TDR profiles during and after hillslope irrigation experiment. Exemplary transect with changes referenced to pre-irrigation conditions and attributed to irrigation water. Time is given in [h] after irrigation start. Individual measurements and probe reference marked with triangles. More data and explanation in Angermann et al. (2016), this issue. Right bottom: Apparent vertical flow velocity as first excess of $\Delta\theta \geq 2$vol% at core area profiles.

not feasible. A comparison of the suggested structures of the 3D GPR survey to the overall response to irrigation recorded at the GPR transects did not correlate well (figure 10).
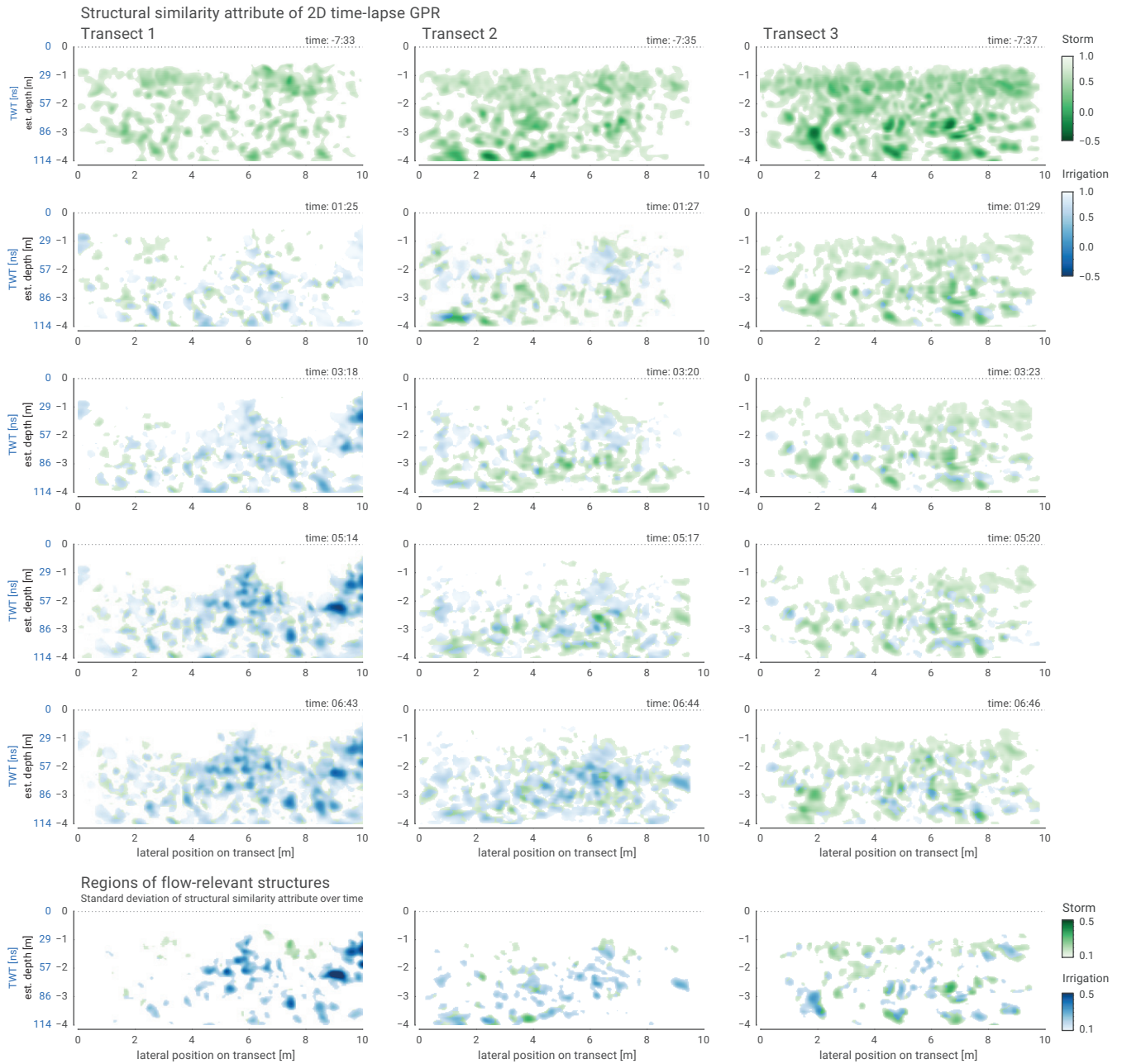
Figure 12. Structural similarity attribute in time-lapse 2D GPR transects. Blue: irrigation event water, Green: Storm event water. Columns: time series in one transect, Rows: different transects at the same time. Bottom row: Identified regions of rapid subsurface flow based on the standard deviation of all structural similarity attributes at one transect over time.

Notice: The structural similarity attribute calculates similarity between the radargram at the respective time to the last record 23 h past irrigation. A threshold or 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   Discussion

4.1   Identification of flow-relevant structures

Our results have shown that the silty soils coincide with high porosities and high hydraulic conductivity at the Darcy-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 questions, pedo-physical approaches have only limited capability to identify flow-relevant structures (Q1). Although the measurements hint to their existence, neither position nor general setup can be derived from point observations. Using irrigation experiments has been far more insightful, especially in combination with time-lapse GPR.

Irrigation experiments at the plot scale suggest that inter-aggregate voids are responsible for highly heterogenous soil water redistribution. These structures are different from what we usually expect (cracks, worm burrows, roots) and operate at much finer scales. This could be depicted from dye tracer stains, which still have the highest spatial resolution. Despite all uncertainty about what process caused staining (high supply of dye water, high sorption into the matrix, long contact times, etc.) the technique allows to identify the structures affected by irrigation and to infer much about their setting. Although dye stains are closely related to actual flow and thus function, they reveal form as the temporal dimension remains unknown. When irrigation intensity and irrigation amount ranges near the hydraulic capacity of the macropore network while still avoiding ponding or macropore clogging the potential network of flow-relevant structures is marked. 3D time-lapse GPR has proven to be capable to detect similar response patterns. However, the spatial and temporal resolution of the method is still of concern. Some of the flow-relevant inter-aggregate voids have not even been traced with dye, nor could GPR identify these specific structures. Notwithstanding, the overall characteristics of the structures as patchy responses is depicted well and in a spatially continuous manner. Thus most of the point-sampling related issues are resolved. To answer Q2 activating flow structures is absolutely necessary to identify them.

At the hillslope scale 2D time-lapse GPR has facilitated high temporal resolution while the structures are also less diverse. The responses suggest similar inter-aggregate voids which 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. At this scale the capability of point-based methods for structure identification is very limited. Also static methods cannot unravel structures from overall heterogeneity here.

4.2   Structure characterization and apparent vertical flow velocities

The value of pedo-physical analyses is much higher when it comes to the characteristics of flow facilitated by the revealed paths. Referring to our third research question (Q3) we have derived hydraulic conductivity and apparent vertical response velocity for comparison at all scales in figures 9 and 11. Although none of the methods is specific to the velocities achieved in the structures, the distributions clearly corroborate that the network of inter-aggregate pores elevates the velocities about three to four orders of magnitude compared to reported matrix capacity.

The results also put a critical perspective on hydrological field measurements of infiltration capacity and saturated hydraulic conductivity. Both methods rely on concepts which cannot account for anisotropy. Thus, the resulting values have to be taken with some precaution if one seeks to derive quantitative estimates. Because the inter-aggregate flow network operates at much finer scales than cracks or worm burrows, the measurements of saturated hydraulic conductivity based on 250 ml ring samples depict the overall picture. However, a very large number of samples will be necessary to converge. Yet, exactly this motivated the presented experiments.

This is much more concisely derived from the responses to irrigation basically irrespective to the applied method. The calculation based on salt tracers point to rather short times of fixation. This estimate is avoided when using time-lapse methods. However, TDR-based responses have a high uncertainty concerning the representativity of the point records. They also are highly sensitive to the selected threshold. GPR-based estimates so far lack temporal resolution and require to infer the depth of the identified signal through estimates of the radar velocity.

4.3 Heterogeneity versus structure

Based on a relatively large number of pedo-physical measurements the overall layering and mean property of a heterogeneous soil with periglacial deposit layers was described in section 3.1. Given the large effort to conduct such exploration, this result is rather dissatisfactory and could have been achieved with much simpler means (e.g. compare Heller and Kleber, 2016, in a similar setting). The attribute supported picking of potential structures in the 3D GPR data cube also had high discrepancies to the actual relevant structures (figure 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.

Also the low representativity of the observations of soil moisture dynamics in many TDR profiles compared to the 3D and multi-2D time-lapse GPR records in sections 3.2.4 and 3.3.2 underlines the methodological drawbacks of point observations. Moreover, low recovery of Bromide in the more than 120 samples per plot points out that the sampling of only 0.5 % of the total affected soil volume cannot fully represent the highly heterogeneous distribution of irrigation water.

It has proven particularly difficult to distinguish heterogeneity and structure. This has conceptual implications: We may regard statistical heterogeneity as random small scale changes in hydrological soil properties (de Marsily et al., 2005) and structure as spatially organized flow paths and their connectedness (Tetzlaff et al., 2010). The spatial scale of flow-relevant structures (5×10^{-3} m to 5×10^{-2} m) is several orders of magnitude below the support of standard soil samples and hydrological measurements (1×10^{-4} m^3 to 1×10^{-3} m^3). At the same time we find the flow process in such structures several orders of magnitude faster than in the soil matrix (also Germann, 2014). Continuum theory assumes homogeneity within the representative elementary volume (REV). The inter-aggregate flow path network enabled the same high flow velocities as classic macropores and the lateral flow paths at the hillslope scale experiment. Nevertheless, their responses can be quantified in a larger REV. Hydrological standard approaches tend to operate at a scale between the actual processes and an appropriate REV.

In more general terms, heterogeneity can be seen as deviation of the found reality from the concept of quasi-homogeneous entities. If this deviation concerns only the apparent values 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 not 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 sub-scale organization within the support of the measurement. Without detection of organization and thus flow-relevant structures they can only recover heterogeneity independent of the number of samples.

4.4 Event response patterns reveal flow-relevant structures

In order to distinguish flow-relevant structures from heterogeneity, approaches focussing on the dynamics of response patterns have proven to be adequate. This corroborates our hypothesis and confirms that a combined assessment of form and function is needed to mutually reduce ambiguity.

4.4.1 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 $1\,\mathrm{L}$ it would require a macropore of about $1\,\mathrm{cm}$ diameter within the support of the sensor to be filled to just reach a threshold of 2 vol%. Adding this $20\,\mathrm{ml}$ 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 proven not feasible. Given the insight of the discretely structured flow domain and the high response velocities identified in the companion study (Angermann et al., 2016, this issue), the soil moisture measurements leave us with many questions. Comparing the identified regions of flow structures (Figure 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 (which is the case in virtually all soil moisture measurement studies). This was also shown in the experiment at plot XI (figure 7 and 15): While tracer and GPR exhibit quick drainage of the irrigation water to the periglacial deposit layer, the TDR records are simply not close enough to the comparably few activated flow paths due to their limited support. As for the local exploration, a precise location of such monitoring at specific points in the domain (e.g. directly at responsive structures and at less reactive places) may enhance the interpretability of monitoring data.

4.4.2 Time-lapse GPR patterns

The finding that potential horizons identified by the static 3D GPR survey do not coincide with the observed responses (figure 10) 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 2D and 3D fields 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 and the detected local changes in the complex reflection patterns allowed only for qualitative interpretation.

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 and an automatic-tracking total station (Allroggen et al., 2015b). On the other hand we base our analysis on the structural similarity attribute in stead 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. It also enables 3D applications avoiding strong assumptions about the continuity of preferential flow paths when inferring 3D networks from 2D measurements (Gormally et al., 2011; Guo et al., 2014). Especially in heterogeneous soils the found response patterns (figure 8) and the excavated stained soil profiles (figure 6) question a direct linear connection across several decimeters. Because of the low spatial continuity we found there, we refrain from interpolations between the multi-2D transects at the hillslope experiment.

Although the 3D 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-2D 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 enhanced the contrast to overall heterogeneity. For large structures, this lead to precise identification and localization. Smaller flow paths cannot be fully resolved. Nevertheless, the continuous 2D and 3D 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.

4.5 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 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.

In the form and function framework one implication of the study is that a disjunct analysis of the two is source of unnecessary ambiguity and susceptibility to bias. Although this is very much in line with the general findings of 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), also the experiments are 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 subsurface flow processes can be used to reduce ambiguity of measurements and to constrain the process conceptualization in heterogeneous and structured soils. Based on 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. We have shown that subsurface heterogeneity and the mismatch of observation and process scales obscured the identification of flow-relevant structures in the subsurface under static conditions without a shift between active and non-active state. The exploratory methods could discover and quantify the general characteristics of the subsurface only within a wide spectrum of the respective target properties. Moreover, they failed to identify flow-relevant structures in terms of position, distribution and capacity. 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 3D GPR survey on the one hand and the time-lapse approaches on the other hand points out 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.

Our findings raise strong concerns about approaches attempting to determine function from form alone. Besides the fine scale of the small inter-aggregate voids we come back to Wittgenstein (1922): Form requires to be addressed in its context to reveal information about structure and its characteristics. But also addressing function needs details about the spatial circumstances to be conclusive.

In hydrological systems form and function are mutually paired. 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. They compare well with soil moisture dynamics and tracer recovery. As such the localization of responses provides the missing link to relate form with function and to guide more specific investigation, monitoring and modeling of subsurface processes.

Appendix A: Detailed results of hydrological measurements

In situ measurements of infiltration capacity and saturated hydraulic conductivity had a highly heterogeneous distribution. To detail on the respective records and found profiles figure 13 shows them in spatial context.

Appendix B: Pore water stable isotope analysis in plot irrigation experiments

In addition to Bromide as conservative salt tracer the same percussion drilled core samples were analyzed for their stable isotopic composition (δ^{18}O and δ^2H) of the pore water. 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 pre-experiment reference, a fourth reference core has been sampled prior to the experiments about 3 m upslope.

We calculate the volumetric event water portion [-] in the soil water as:

$$\frac{\Theta_{event}}{\Theta_{2h}} = \frac{\Theta_{2h} \cdot \delta^2 \mathrm{H}_{2h} - \Theta_{pre} \cdot \delta^2 \mathrm{H}_{pre}}{\Theta_{2h} \cdot \delta^2 \mathrm{H}_{event}} \tag{B1}$$

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

Figures 14A-C show the depth profile of irrigation water as portion of total water content, calculated from the deviation in δ^2H concentration between reference and $2h$ 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 not-enriched irrigation water. Figure 14D-E highlights 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 in about 0.5 m depth and signals below may be erroneous, because the signature of the reference core coincides with the irrigation water there.
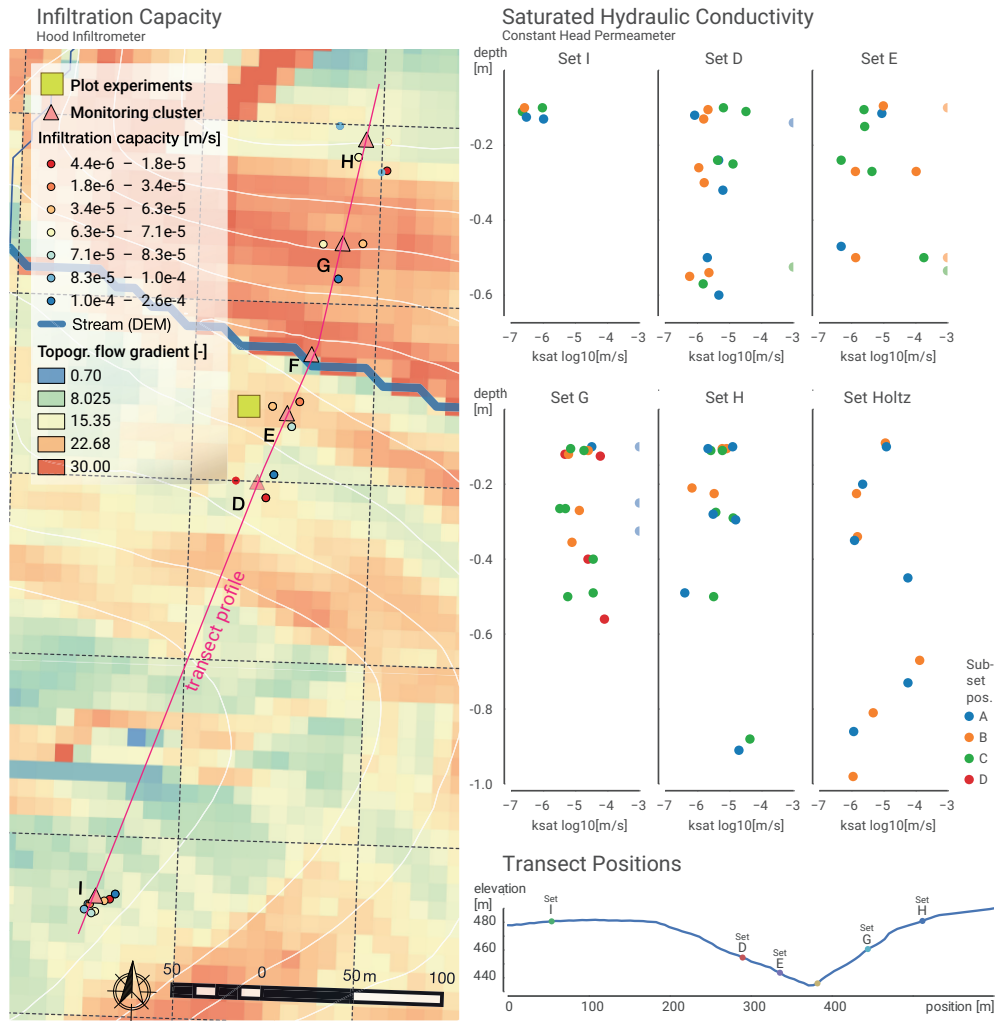
Figure 13. Hydrological exploration results. Left: Infiltration capacity (values color coded) measured with hood infiltrometer. Basemap with topographic flow gradient $G = arctan(\partial z_{rel}/\partial d_{rel})$, topography contour lines and positions of hydro-meteorological monitoring clusters. Right top: Saturated hydraulic conductivity measured with a constant head permeameter. Individual profile position in the respective nested set color coded (2 measurement holes with dist=1 m). Values exceeding the device capacity set to $10^{-3}\,\mathrm{m\,s^{-1}}$. Right bottom: Elevation profile of transect as characteristic landscape feature in the sub-basin.

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 the fact, that soil water and irrigation water deviated only slightly ($\geq 15\,\%_o$) and even had the same values in $0.5\,\mathrm{m}$ 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 assumably ideal tracer the stable
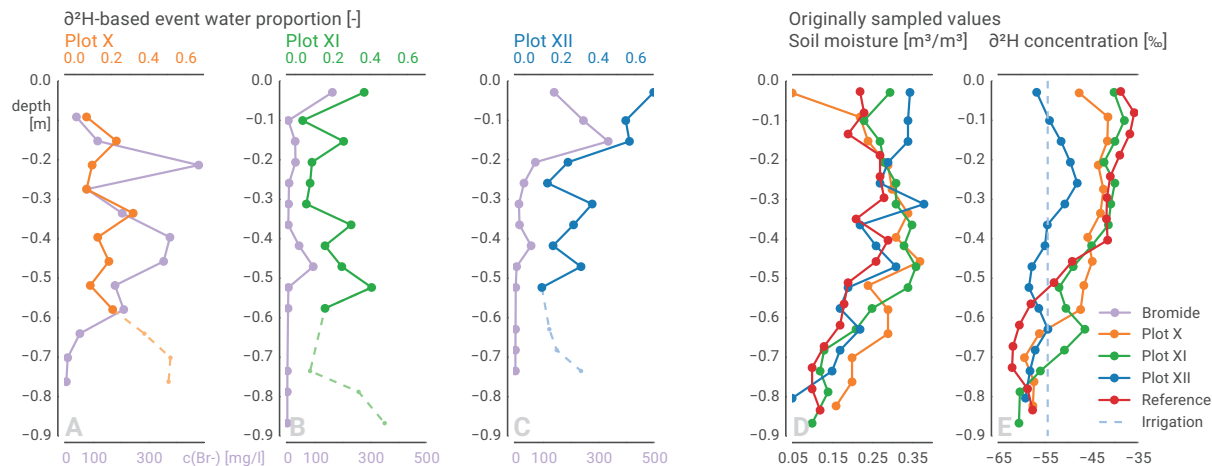
Figure 14. Plot-scale irrigation experiments. Proportion of event water derived from deviations in concentration of Deuterium and Bromide in soil water of sampled cores to reference (A-C). Absolute measured soil water content (D). Deuterium concentration in samples and signature of irrigation water (E). Note that the $\delta^2 H$ concentration profiles and the signature of the irrigation water show low deviation and even coincide in $0.5\,\mathrm{m}$ depth. Thus the interpretation needs precaution, especially below that level (dashed lines A-C).

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.

Appendix C: Results 3D time-lapse GPR at Plot XI and XII

In addition to the results in section 3.2 here the radargrams and structural similarity attributes for the other two plot-scale
5 experiments are given in figure 16 and 15. 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 D: Technical concerns of time-lapse GPR and the structural similarity attribute

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

    The highlighted assumptions clearly frame the limits of the technique. The overall sensitivity of the approach can be judged
15 from the structural similarity attribute of the last pairs of records in the hillslope experiment when we assume the soil water to
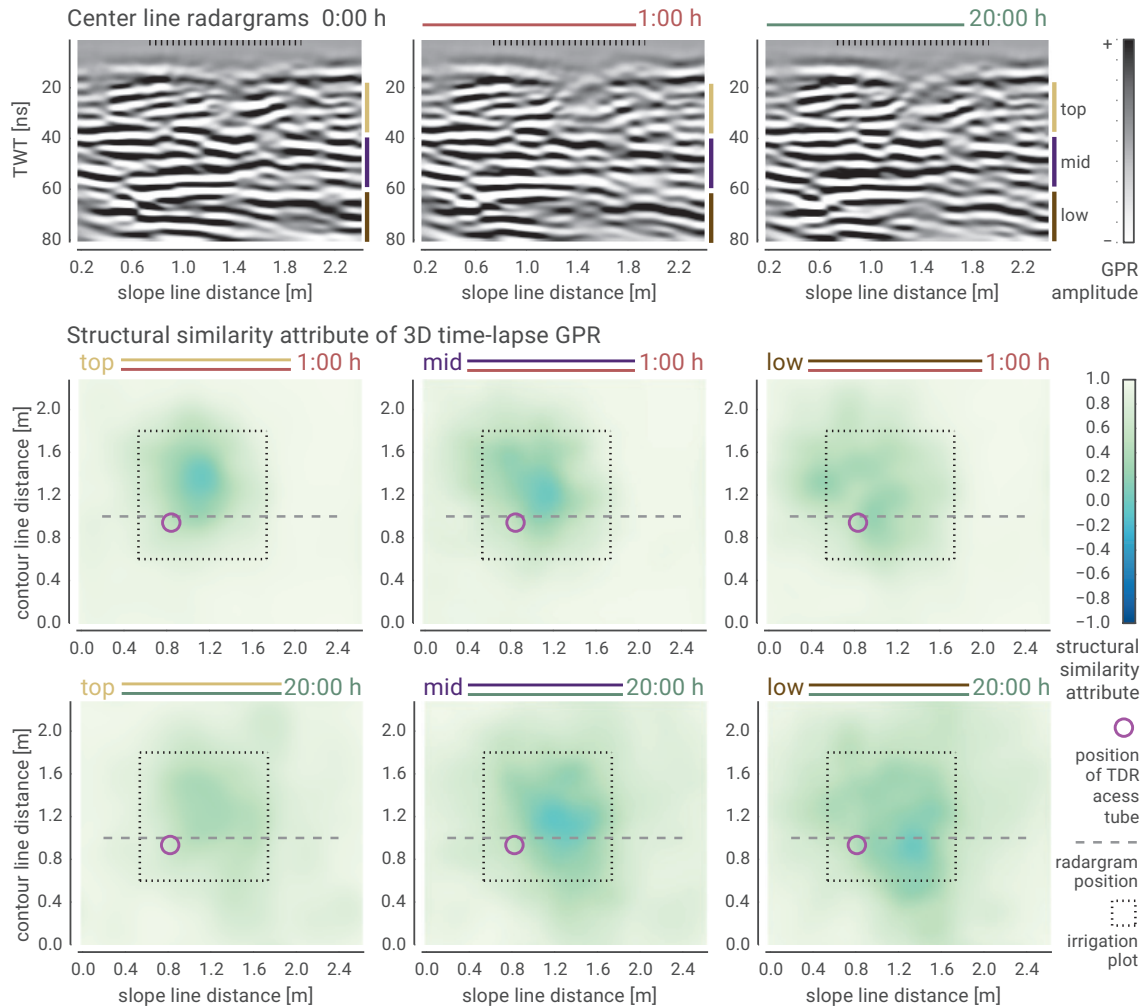
Figure 15. Time-lapse 3D GPR of irrigation experiment at plot XI. Center line radargrams at the marked transect (grey dashed line in lower panels) for the three acquisition times (before 0:00 h, directly after irrigation 1:00h, 20:00 h after irrigation) are given in the top row. The structural similarity attribute of the 3D data cube is given in three different depth layers (top 20 ns to 40 ns, mid 40 ns to 60 ns, low 60 ns to 80 ns) in the lower panels. The irrigation plot is marked by a black dashed box/line. Slope line distance is increasing downslope.

be in equilibrium again. Figure 17 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 stepwise 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

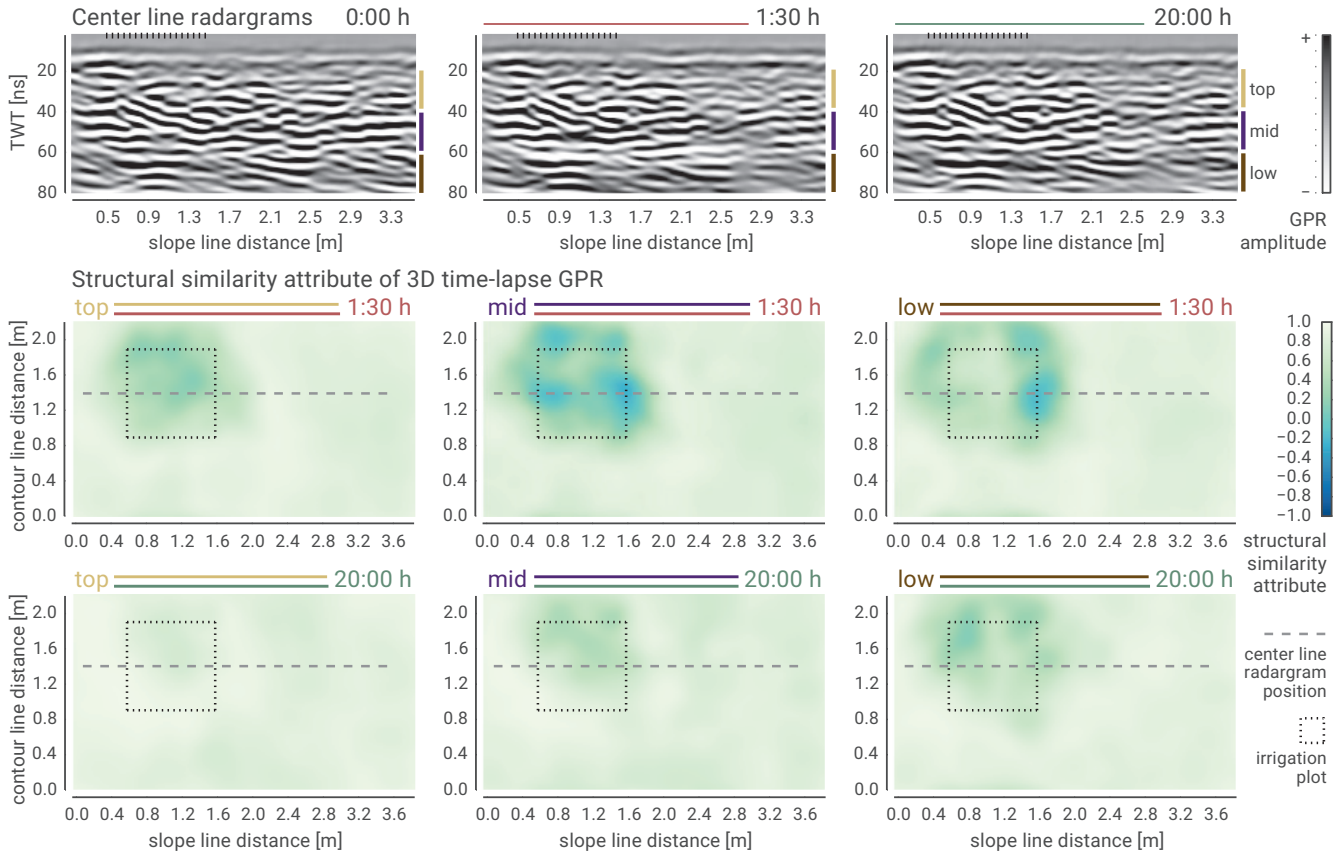5   signal detection appears to be a reasonable choice for qualitative interpretation in our case.

Figure 16. Time-lapse 3D GPR of irrigation experiment at plot XII. Center line radargrams at the marked transect (grey dashed line in lower panels) for the three acquisition times (before 0:00 h, directly after irrigation 1:30h, 20:00 h after irrigation) are given in the top row. The structural similarity attribute of the 3D data cube is given in three different depth layers (top 20 ns to 40 ns, mid 40 ns to 60 ns, low 60 ns to 80 ns) in the lower panels. The irrigation plot is marked by a black dashed box/line. Slope line distance is increasing downslope.

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.
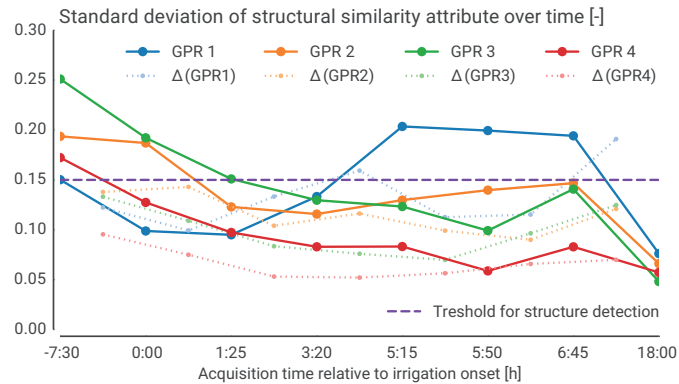
Figure 17. Standard deviation of structural similarity attribute at the different GPR transects in the hillslope experiment over time (solid lines) and standard deviation of the differences of two successive attribute distributions (dotted lines). The used threshold for the detection of flow-relevant structures is marked as dashed purple line.

References

Allaire, S. E., Roulier, S., and Cessna, A. J.: Quantifying preferential flow in soils: A review of different techniques, Journal of Hydrology, 378, 179–204, doi: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, doi: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 Surface Geophysics, 13, 253–259, doi: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, Journal of Applied Geophysics, 118, 139–144, doi: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 Science Society of America Journal, 53, 1362, doi: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, Hydrology and Earth System Sciences, 13, 935–944, doi: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, Hydrology and Earth System Sciences, this issue, 2016.

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, Physical Review E, 94, 043 113, doi:10.1103/PhysRevE.94.043113, 2016.

Bachmair, S. and Weiler, M.: Interactions and connectivity between runoff generation processes of different spatial scales, Hydrological Processes, 28, 1916–1930, doi: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 Resources Research, 51, 3730–3753, doi:10.1002/2014WR016221, 2015.

Beven, K. and Germann, P.: Macropores and water flow in soils, Water Resources Research, 18, 1311–1325, doi:10.1029/WR018i005p01311, 1982.

Beven, K. and Germann, P.: Macropores and water flow in soils revisited, Water Resources Research, 49, 3071–3092, doi: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 Resources Research, 51, 3837, doi: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, Journal of Applied Geophysics, 43, 215–226, doi: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 Resources Research, 51, 3956–3981, doi:10.1002/2014WR016310, 2015.

Bloschl, G. and Sivapalan, M.: Scale issues in hydrological modelling: A review, Hydrological Processes, 9, 251–290, doi: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, European Journal of Soil Science, 64, 161–182, doi: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, doi: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, doi: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 Resources Research, 24, 755–769, doi: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, Hydrogeology Journal, 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?, Hydrological Processes, 28, 2451–2467, doi: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 Resources Research, 30, 1945–1954, doi:10.1029/94WR00871, 1994.

Gerke, H. H.: Preferential flow descriptions for structured soils, Journal of Plant Nutrition and Soil Science, 169, 382–400, doi:10.1002/jpln.200521955, 2006.

Gerke, H. H.: Macroscopic Representation of the Interface between Flow Domains in Structured Soil, Vadose Zone Journal, 11, 0, doi: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 Journal, 9, 207–212, doi: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 Journal, 7, 1160–1169, 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 Journal, 15, 0, doi: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 Science Society of America Journal, 75, 1236, doi: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 Resources Research, 50, 9127–9147, doi: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 Resources Research, 48, 1–40, doi: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 Journal, 10, 84–97, doi: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, Physics and Chemistry of the Earth, Parts A/B/C, 36, 42–53, doi:doi: 10.1016/j.pce.2010.07.019, 2011.

Heller, K.: Einfluss periglazialer Deckschichten auf die oberflächennahen Fließwege am Hang - eine Prozessstudie im Osterzgebirge, Sachsen, Ph.D. 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, Environmental Earth Sciences, 75, 943–15, doi: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 Surface Processes and Landforms, 29, 437–442, doi: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., Bloschl, G., and Flühler, H.: Impact of modellers' decisions on hydrological a priori predictions, Hydrology and Earth System Sciences, 18, 2065–2085, doi: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 Journal, 2, 476–491, doi: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, Ph.D. thesis, KIT Karlsruhe Institute of Technology, doi: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 Resources Research, 42, W02 407, doi: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, European Journal of Soil Science, 53, 231–240, doi:10.1046/j.1365-2389.2002.00437.x, 2002.

Kirchner, J. W.: A double paradox in catchment hydrology and geochemistry, Hydrological Processes, 17, 871–874, doi: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, Hydrology and Earth System Sciences, 15, 2127–2144, doi: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, Hydrology and Earth System Sciences, 17, 103–118, doi:10.5194/hess-17-103-2013, 2013.

5 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, Hydrology and Earth System Sciences, 19, 1125–1139, doi:10.5194/hess-19-1125-2015, 2015.

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

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

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

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
15 double peak hydrographs in a forested headwater catchment, Journal of Hydrology, doi: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, Transactions American Geophysical Union, 77, 465–472, doi:10.1029/96EO00306, 1996.

Mugler, A., Grinshpun, B., Franks, R., and Wiggins, C. H.: Statistical method for revealing form-function relations in biological networks, Proceedings of the National Academy of Sciences, 108, 446–451, doi:10.1073/pnas.1008898108, 2011.

20 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, doi: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, Journal of Hydrology, 115, 269–283, doi:10.1016/0022-1694(90)90209-G, 1990.

25 Öhrström, P., Hamed, Y., Persson, M., and Berndtsson, R.: Characterizing unsaturated solute transport by simultaneous use of dye and bromide, Journal of Hydrology, 289, 23–35, doi: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, doi: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
30 Lippmann, 2002.

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

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

35 Schaap, M. G., Leij, F. J., and Van Genuchten, M. T.: Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions, Journal of Hydrology, 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 Resources Research, 52, 2194–2205, doi: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, Hydrology and Earth System Sciences, 16, 2871–2882, doi:10.5194/hess-16-2871-2012, 2012.

Sivapalan, M.: Pattern, Process and Function: Elements of a Unified Theory of Hydrology at the Catchment Scale, Encyclopedia of Hydrological Sciences, 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, Hydrology and Earth System Sciences, 19, 2617–2635, doi: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 Resources Research, 52, 5727–5754, doi: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 Resources Research, 48, W09 538, doi: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, Hydrological Processes, 24, 2203–2208, doi:10.1002/hyp.7784, 2010.

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

Trinks, I., St u mpel, 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 Resources Research, 42, W02 410, doi: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 Resources Research, 43, 1–15, doi:10.1029/2005WR004395, 2007.

Uhlenbrook, S.: Catchment hydrology - a science in which all processes are preferential, Hydrological Processes, 20, 3581–3585, doi: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, Hydrological Sciences Journal, 51, 1021–1038, doi:doi: 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, doi: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, doi:10.1002/eco.1358, 2014.

Vogel, H.-J. and Roth, K.: Moving through scales of flow and transport in soil, Journal of Hydrology, 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, Hydrology and Earth System Sciences, 10, 495–506, doi:10.5194/hess-10-495-2006, 2006.

Wassenaar, L. I., Hendry, M. J., Chostner, V. L., and Lis, G. P.: High Resolution Pore Water $\delta 2H$ and $\delta 18O$ Measurements by H2O (liquid) - H2O (vapor) Equilibration Laser Spectroscopy, Environmental Science & Technology, 42, 9262–9267, doi: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 Resources Research, doi:10.1002/2014WR016131, 2015.

Western, A. W. and Blöschl, G.: On the spatial scaling of soil moisture, Journal of Hydrology, 217, 203–224, doi:doi: 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, Hydrology and Earth System Sciences, 13, 1145–1161, doi: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, Hydrological Processes, 29, 2731–2750, doi:10.1002/hyp.10393, 2015.

Zehe, E., Maurer, T., Ihringer, J., and Plate, E.: Modeling water flow and mass transport in a loess catchment, Physics and Chemistry of the Earth Part B - Hydrology Oceans and Atmosphere, 26, 487–507, doi: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, Hydrology and Earth System Sciences, 17, 4297–4322, doi:10.5194/hess-17-4297-2013, 2013.

Zehe, E., Ehret, U., Pfister, L., Blume, T., Schroder, 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, Hydrology and Earth System Sciences, 18, 4635–4655, doi:10.5194/hess-18-4635-2014, 2014.