Picturing and modeling catchments by representative hillslopes

Ralf Loritz[1], Sibylle K. Hassler[1], Conrad Jackisch[1], Niklas Allroggen[2], Loes van Schaik[3], Jan Wienhöfer[1], and Erwin Zehe[1]

[1] Karlsruhe Institute of Technology (KIT), Institute of Water and River Basin Management, Karlsruhe, Germany
[2] University of Potsdam, Institute of Earth and Environmental Science, Potsdam, Germany
[3] Technical University Berlin, Institute of Ecology, Berlin, Germany

Correspondence to: Ralf Loritz (Ralf.Loritz@kit.edu)

Abstract: This study explores the suitability of a single hillslope as most parsimonious representation of a catchment in a physically-based model. We test this hypothesis by picturing two distinctly different catchments in perceptual models and translating these pictures into parametric setups of 2-D physically-based hillslope models. The model parametrizations are based on a comprehensive field data set, expert knowledge and process-based reasoning. Evaluation against stream flow data highlights that both models predicted the annual pattern of stream flow generation as well as the hydrographs acceptably. However, a look beyond performance measures revealed deficiencies in streamflow simulations during the summer season and during individual rainfall-runoff events as well as a mismatch between observed and simulated soil water dynamics. Some of these shortcomings can be related to our perception of the systems and to the chosen hydrological model, while others point to limitations of the representative hillslope concept itself. Nevertheless, our results corroborate that representative hillslope models are a suitable tool to assess the importance of different data sources as well as to challenge our perception of the dominant hydrological processes we want to represent therein. Consequently, these models are a promising step forward in the search of the optimal representation of catchments in physically-based models.

1 Introduction

The value of physically-based hydrological models has been doubted (e.g. Beven, 1989, Savenije and Hrachowitz, 2016) since their idea was introduced by Freeze and Harlan (1969). Physically-based models like MikeShe (Refsgaard and Storm, 1995) or CATHY (Camporese et al., 2010) typically rely on the Darcy-Richards concept for soil water dynamics, the Penman–Monteith equation for soil-vegetation-atmosphere exchange processes and hydraulic approaches for overland and stream flow. Each of these concepts is subject to limitations arising from our imperfect understanding of the related processes and is afflicted by the restricted transferability of process descriptions from idealized laboratory conditions to heterogeneous natural systems (Grayson et al., 1992; Gupta et al., 2012).

Nevertheless the usefulness of physically-based models as learning tool to explore how internal patterns and processes control the integral behavior of hydrological systems, has been corroborated in several studies. For example Pérez et al. (2011) used Hydrogeosphere (Brunner and Simmons, 2012) together with a regularization scheme for its calibration, to infer how changes in agricultural practices affect the stream flow generation in a catchment. Hopp and McDonnell (2009) explored the role of bedrock topography on the runoff generation using HYDRUS 3D (Simunek et al., 2006) at the Panola hillslope. Coenders-Gerrits et al. (2013) used the same model structure to examine the role of interception and slope on the subsurface runoff generation. Bishop et al. (2015), Wienhöfer and Zehe (2014) and Klaus and Zehe (2011) used physically-based models to investigate the influence of vertical and lateral preferential flow networks on subsurface water flow and solute transport, including the issue of equifinality and its reduction. These and other studies (e.g. Ebel et al., 2008, Scudeler et al., 2016) show that physically-based models can be set up using a mix of expert knowledge and observed parameters and may be tested against a variety of observations beyond stream flow – such as soil moisture observations, groundwater tables or tracer break-through curves. Such studies are, on the one hand an option to increase our limited understanding of the processes underlying physically-based models (Loague and VanderKwaak, 2004), and on the other hand reveal if a model allows consistent predictions of dynamics within the catchment and of its integral response behavior (Ebel and Loague, 2006).

Setting up a classical physically-based model in a heterogeneous environmental system is, however, a challenge as it requires an enormous amount of highly resolved spatial data, particularly on subsurface characteristics. Such data sets are rare and only available in rather homogeneous systems or in environmental system simulators as Biosphere 2 LEO (Hopp et al., 2009). Therefore, it has been a long standing vision to replace fully distributed physically-based models by aggregated but yet physically-based model concepts for instance the Hillslope Storage Boussinesq approach (HSB, Troch et al., 2003; Berne et al., 2005) or the REW approach (Representative Elementary Watershed e.g. Reggiani and Rientjes, 2005; Zhang and Savenije, 2005). The key challenge in applying these concepts to real catchments is the assessment of a closure relationship, which parameterizes a) hydrological fluxes (Beven, 2006a) and b) soil water characteristics in an aggregated effective manner (Lee et al., 2006; Zehe et al., 2006). Furthermore, is it not completely clear whether the entire range of variability in subsurface characteristics is relevant for hydrological simulations (Dooge, 1986; Zehe et al., 2014). There are, however, promising concepts emerging, for example the work of Hazenberg et al. (2016) who recently developed a hybrid model consisting of the HSB model in combination with a 1-D representation of the Richards equation for the unsaturated zone.

Regardless of whether one favors physically-based, hybrid or more statistical model approaches, a perfect representation of a hydrological system should balance the necessary complexity with greatest possible simplicity (Zehe et al., 2014). The former is necessary to avoid oversimplification. The latter attempts to avoid the drawbacks of over-parametrization (Schoups et al., 2008). In principle there are two ways how one can try to reach this optimum model structure. Either by starting with a complex system representation, for instance a full 3-D catchment model and simplify the model structure as much as possible or by starting at the other end of the spectrum, with the most parsimonious model structure and proceed towards higher complexity. In conceptual rainfall-runoff models which follow the HBV concept (Bergström and Forsman, 1973) the most parsimonious model structure for simulating the behavior of a catchment is a single reservoir. In the case of physically-based models there is more than one starting point. In flatland catchments without dominant lateral flow processes in the soil one might choose a single soil column. This “null model” could be refined into multiple parallel acting columns, to capture variability in vegetation and soil properties. This represents the first generation of land surface components in meteorological models (e.g. Niu et al., 2011) and the first generation of models for the catchment scale dynamics of nitrate (Refsgaard et al., 1999).

However, in hilly or mountainous terrain the smallest meaningful unit is a hillslope including the riparian zone, because rainfall and radiation input depend on slope and aspect, as well as on downslope gradients which cause lateral fluxes in the unsaturated zone (e.g. Bachmair and Weiler, 2011; Zehe and Sivapalan, 2007). This is the reason why hillslopes are often regarded as the key landscape elements controlling transformation of precipitation and radiation inputs into fluxes and stocks of water (e.g. Bronstert and Plate, 1997), energy (Zehe et al., 2010, 2013) and sediments (Mueller et al., 2010).

The most parsimonious representation of a small catchment in a physically-based model could thus be a single representative hillslope. However, the challenge of how to identify such a hillslope has rarely been addressed. This reflects the fact that the identifiability of a representative hillslope has been strongly questioned since the idea was born. For example, Beven (2006) argues that neither is the hillslope form uniquely defined nor is it clear whether it is the form that matters, the pattern of saturated areas (Dunne and Black, 1970) or the subsurface architecture. The enormous spatial variability of soil hydraulic properties and preferential flow paths in conjunction with process non-linearity are additional arguments against the identifiability of representative hillslope models (Beven and Germann, 2013). Nevertheless, hillslopes act as miniature catchments (Bachmair and Weiler, 2011), which made Zehe et al. (2014) postulate that structurally similar hillslopes act as functional units for the runoff generation and might thereby be a key unit for understanding catchments of organized complexity (Dooge, 1986). Complementarily, Robinson et al. (1995) showed that the behavior of catchments up to the lower mesoscale are strongly dominated by the hillslope behavior, and Kirkby (1976) highlighted that in catchments extending up to 50 km^2 random river networks had the same explanative power for runoff generation as the real river network. He concluded that as long as river networks are not dominant the characteristic areas of the catchment hold the key to understand its functioning.

In this context it is important to note that a representative hillslope cannot be a simple copy of a real hillslope in a catchment or a simple average of several hillslopes and their structural properties. A much more promising avenue is to set up the representative hillslope based on a perceptual model which is in turn a generalized and simplified picture of the catchment structure and functioning. This is because perceptual models provide a useful means to facilitate communication between field researchers and modelers (Seibert and McDonnell, 2002) and additionally often represent catchments as hillslope-like cross sections. The general idea to translate a perceptual model into a model structure is not new and has already been applied within a conceptual rainfall-runoff model framework even within the same area (Wrede et al., 2015). The scientific asset of using a physically-based model is that the perceptual model provides important information on typical ordinal differences in hydraulic conductivity of different subsurface strata and the nature and qualitative locations of the dominating preferential flow paths. This information can be implemented into hillslope models in a straightforward manner. The transformation of a qualitative model structure into a quantitative, parametrization of the model depends, however, strongly on the chosen hydrological model and the quality and amount of available data.

Objectives and approach

We hypothesize that a single hillslope in a physically-based model is the most parsimonious representation of a small hilly catchment. The objective of this study is to test this hypothesis in a two-step approach:

- First we derive a qualitative model structure of a representative hillslope from our perception of the dominant processes and the related dominant surface and subsurface characteristics in the catchment.
- In the second step we transform this qualitative model structure into a quantitative model structure without the use of an automatic parameter allocation.

The challenge in deriving a qualitative model structure lies in the separation of the important details from the idiosyncratic ones. This process is to a large extent independent of the chosen hydrological model and is strongly related to the available expert knowledge and quality of the data. The transformation of a qualitative to a quantitative model structure on the other hand depends on the chosen model and whether it is for example based on 2-D or 3-D hillslope module or how rapid flow paths are represented. For this reason the objective of our study is not to “sell” our particular model, but to share the way how we distilled the quantitative model setups in our target catchments from available data and to evaluate the ability of this parsimonious physically-based model to accurately simulate multiple state and flux variables. During the model setup we intendedly avoided using an optimization algorithm to fit the model to the data. In contrary, we relied on various available observations, process-based reasoning, and appropriate literature data for conceiving our perceptual models and parameterizing the representative hillslope models as their quantitative analogues. More specifically, we use geophysical images to constrain subsurface strata and bedrock topography and derived representative soil-water retention curves from a large data set of undisturbed soil samples. Furthermore, we use observations from soil pits, dye staining experiments and observed leaf area indices (LAI) for our model parametrization. Finally we benchmark the hillslope models against normalized double mass curves, the hydrograph as well as against distributed soil moisture and sap flow observations.

2. Study area, data basis and selected model

We focus our model efforts on two different catchments, the Colpach and the Wollefsbach, located in the Attert experimental basins in Luxembourg (Figure 1, Pfister et al., 2000). These sites offer comprehensive laboratory and field data collected by the CAOS (Catchments As Organized Systems) research unit (Zehe et al., 2014). Besides standard hydro-meteorological data the model setup is based on a) observed soil hydraulic properties of a large number of undisturbed soils cores, b) 2-D electric resistivity profiles in combination with soil pits and augering to infer on bedrock topography, and c) flow patterns from dye staining experiments and soil ecological mapping of earthworm burrows, to infer the nature and density of vertical preferential flow paths. The representative hillslopes for the two catchments were each set up as a single 2-D hillslope in the CATFLOW model (Zehe et al., 2001). The following subsections will provide detailed information on the perceptual models and on the water balance of both catchments. We will shortly refer to the key data and those parts of the model which are relevant for the quantitative model setup, while the appendix provides additional details on both.

2.1 The Attert experimental basin

The Attert basin is located in the mid-western part of the Grand-Duchy of Luxembourg and has a total area of 288 km². Mean monthly temperatures range from 18°C in July to a minimum of 0°C in January; mean annual precipitation in the catchment varies around 850 mm (1971–2000) (Pfister et al., 2000). The catchment covers three geological formations, the Devonian schists of the Ardennes massif in the northwest, Triassic sandy marls in the center and a small area of sandstone (Jurassic) in the southern part of the catchment (Martínez-Carreras et al., 2012). Our study areas are headwaters named Colpach in the schist area (19.4 km²) and Wollefsbach in the marl area (4.5 km²). As both catchments are located in distinctly different geologies and land use settings, they differ considerably with respect to runoff generation and the dominant controls (e.g. Bos et al.

1996, Martínez-Carreras et al. 2012, Fenicia et al. 2014, Wrede et al. 2015, Jackisch 2015).

2.1.1 Colpach catchment: perceptual model of structure and functioning

The Colpach catchment has a total area of 19.4 km² and elevation ranges from 265 to 512 m a.s.l. It is situated in the northern part of the Attert basin in the Devonian schists of the Ardennes massif. Around 65 % of the catchment are forested, mainly the steep hillslopes (Figure 2). In contrast, the plateaus at the hill tops are predominantly used for agriculture and pasture. Several geophysical experiments and drillings showed that bedrock and surface topography are distinctly different. The bedrock is undulating and rough with ridges, depressions and cracks (compare perceptual model Figure 3 A and ERT image in Figure 6 B). Depressions in the bedrock interface are filled with weathered, silty materials which may form local reservoirs with a high water holding capacity. These reservoirs are connected by a saprolite layer of weathered schist which forms a rapid lateral flow path on top of the consolidated bedrock. Rapid flow in this “bedrock interface” is the dominant runoff process (Wrede et al., 2015), and the specific bedrock topography is deemed to cause typical threshold-like runoff behavior similar to the fill-and-spill mechanism proposed by Tromp-Van Meerveld and McDonnell (2006). Further indication that fill-and-spill is a dominant process is given by the fact that the parent rock is reported as impermeable, which makes deep percolation through un-weathered schist layers into a large groundwater body unlikely (Juilleret et al., 2011). The lack of significant observations of base flow underpins this notion. Furthermore, surface runoff has rarely been observed in the catchment, except along forest roads, which suggests a high infiltrability of the prevailing soils (Bos et al., 1996). This is in line with distributed permeameter measurements and soil sampling performed by Jackisch (2015). Moreover, numerous irrigation and dye staining experiments highlight the important role of vertical structures for rapid infiltration and subsequent subsurface runoff formation (Jackisch 2015, Figure 2 B). These vertical preferential flow paths, the saprolite layer on top of the impermeable bedrock, the bedrock topography as well as the absence of a major groundwater body are regarded the dominant structures for the representative hillslope model (Figure 3 A and C).

2.1.2 Wollefsbach catchment: perceptual model of structure and functioning

The Wollefsbach catchment is located in the Triassic sandy marls formation of the Attert basin. It has a size of 4.5 km² and low topographic gradients, with elevation ranging from 245 to 306 m a.s.l. The catchment is intensively used for agriculture and pasture (Figure 2 C); only around 7 % are forested. Hillslopes are often tile-drained (compare perceptual model sketch in Figure 3 B). The heterogeneous marly soils range from sandy loams to thick clay lenses and are generally very silty with high water holding capacities. Similar to the Colpach catchment, vertical preferential flow paths play a major role for the runoff generation; their origin, however, is distinctly different between the seasons. Biogenic macropores are dominant in spring and autumn due to the high abundance of earthworms. Because earthworms are dormant during midsummer and winter, their burrows are partly disconnected by ploughing, shrinking and swelling of the soils (Figure 2 D, see also Figure 4). Soil cracks emerge during long dry spells in midsummer due to the considerable amount of smectite clay minerals in these soils, which drastically increase soil infiltrability in summer (Figure 4). The seasonally varying interaction of both types of preferential flow paths with a dense man-made subsurface drainage network is considered the reason for the flashy runoff regime of this catchment, where discharge rapidly drops to baseflow level when precipitation events end. This is the key feature that needs to be captured by the representative hillslope model. However, as the exact position of the subsurface drainage network and the worm burrows as well as the threshold for soil crack emergence are unknown, the specific influence of each structure on runoff generation in a hydrological model is difficult to estimate.

2.1.3 Water balance and seasonality

The water balance of the Colpach and Wollefsbach catchments for several hydrological years is presented in Figure 5 as normalized double mass curves. Normalized double mass curves relate cumulated runoff to cumulated precipitation, both divided by the sum of the annual precipitation (Pfister et al., 2002, Seibert et al. 2016). Annual runoff coefficients in the Colpach catchment vary around 0.51 ±0.06 among the four hydrological years (Figure 5 A). Annual runoff coefficients are smaller in the Wollefsbach catchment than in the Colpach catchment, and vary across a wider range, from 0.26 to 0.46 (Figure 5 B). In both catchments the winter period is characterized by step-like changes which reflect fast water release during rainfall events partly due to rapid subsurface flow. In contrast, the summer regime is characterized by a smooth and almost flat line when vegetation is active. Accumulated rainfall input is not transformed into additional runoff but is either stored in the system or released as evapotranspiration (Jackisch 2015). As suggested by Seibert et al. (2016) we used a temperature index model from Menzel et al. (2003) to detect the bud break of the vegetation and to separate the vegetation-controlled summer regime from the winter period in these curves.

2.2 Data basis

2.2.1 Surface topography and land use

Topographic analyses are based on a 5 m LIDAR digital elevation model which was aggregated and smoothed to 10 m resolution. Land use data from the "Occupation Biophysique du Sol" is based on CORINE land use classes analyzed by color infrared areal images published in 1999 by the Luxembourgian surveying administration "Administration du cadaster et de la Topographie" at a scale of 1:15000.

2.2.2 Subsurface structure and bedrock topography

We used hillslope-scale 2-D electrical resistivity tomography (ERT) in combination with augerings and soil pits to estimate bedrock topography in the schist area. Our auger profiles revealed, in line with Juilleret et al. (2011) and Wrede et al. (2015), that the vertical soil setup comprises a weathered silty soil layer with a downwards increasing fraction of rock fragments, which is underlain by a transition zone of weathered bedrock fragments and by non-weathered and impermeable bedrock. Based on a robust inversion scheme as implemented in Res2Dinv (Loke, 2003) and additional expert knowledge, the subsurface was subdivided into two main layers of unconsolidated material and solid bedrock. The bedrock interface was picked by the 1500 Ωm isoline, as explained in detail in the appendix. For our study we used seven ERT profiles from the Colpach area (example see Figure 6 B).

2.2.3 Soil hydraulic properties

We determined soil texture, saturated hydraulic conductivity and the soil water retention curve for 62 soil samples in the schist area and 25 in the marl area. Particularly for the soil hydraulic functions Jackisch (2015) and Jackisch et al. (2016) found large spatial variability, which was neither explained by slope position nor by the soil depth at which the sample was taken(Figure 7). As our objective was to assess the most parsimonious representative hillslope model, we neglected this variability but used effective soil water characteristics for both catchments instead. These were not obtained by averaging the parameter of the individual curves, but by grouping the observation points of all soil samples for each geological unit, and averaging them in steps of 0.05 pF. We then fitted a van Genuchten-Mualem model using a maximum likelihood method to these averaged values (Table 1 and Figure 7). The appendix provides additional details on measurement devices and on the dye staining experiments.

2.2.4 Meteorological forcing and discharge

Meteorological data are based on observations from two official meteorological stations (Useldange and Roodt) provided by the "Administration des services techniques de l'agriculture Luxembourg". Air temperature, relative humidity, wind speed and global radiation are provided with a temporal resolution of 1 h while precipitation data are recorded at an interval of 5 min. Precipitation was extensively quality checked against six disdrometers which are stationed within the Attert basin and by comparing several randomly selected rainfall events against rain radar observations, both using visual inspection. Discharge observations are provided by the Luxembourg Institute of Science and Technology (LIST).

2.2.5 Sap flow and soil moisture data

The Attert basin is instrumented with 45 automated sensor clusters. A single sensor cluster measures inter alia rainfall, and soil moisture in three profiles with sensors at various depths. In this study we use 38 soil moisture sensors located in the schist area and 28 sensors located in the marl area, at depths of 10 and 50 cm. Furthermore we use sap flow measurements from 28 trees at 11 of the sensor cluster sites. The measurement technique is based on the heat ratio method (Burgess et al., 2001), sensors are East30Sensors 3-needle sap flow sensors. As a proxy for sap flow we use the maximum sap velocity of the measurements from three xylem depths (5, 18 and 30 mm) as recorded by each sensor. To represent the daytime flux, we use 12-h daily means between 8am and 8pm.

2.3 The physically-based model CATFLOW

Model simulations were performed using the physically-based hydrological model CATFLOW (Maurer, 1997; Zehe et al., 2001). CATFLOW consists of a 2-D hillslope module which can optionally be combined with a river network to represent a catchment (with several hillslopes). The model employs the standard physically-based approaches to simulate soil water dynamics, optionally solute transport, overland and river flow and evapo-transpiration, which were already mentioned in the introduction and are described in more detail in the appendix. In the following we will only provide details implement of rapid flow paths in the model, as this aspect is differs greatly from model to model.

2.3.1 Generation of rapid vertical and lateral flow paths

Vertical and lateral preferential flow paths are represented as a porous medium with high hydraulic conductivity and very low retention. This approach has already been followed by others (Nieber and Warner 1991; Castiglione et al. 2003; Lamy et al. 2009; Nieber and Sidle 2010), and is one of many ways to account for rapid flow paths in physically-based models. However, it is import to note that such a macropore representation is obviously not an image of the real macropore configuration given the typical grid size of a few centimeters, but a conceptualization to explicitly represent parts of the subsurface with prominent flow paths and the adjacent soil matrix in an effective way. The approach includes the assumption that preserving the connectedness of the rapid flow network (Figure 3) is more important than separating rapid flow and matrix flow into different domains.

Implementations of this approach with CATFLOW were successfully used to predict hillslope scale preferential flow and tracer transport in the Weiherbach catchment, a tile-drained agricultural site in Germany (Klaus and Zehe, 2011), and at the Heumöser hillslope, a forested site with fine textured marly soils in Austria (Wienhöfer and Zehe, 2014). The locations of vertical macropores along the soil surface may either be selected based on a fixed distance or via a Poisson process based on the surface density of macropores. From these starting points the generator stepwise extends the vertical preferential pathways downwards to a selected depth, while allowing for a lateral step with a predefined probability of typically 0.05 to 0.1 to establish tortuosity. Lateral preferential flow paths to represent either pipes at the bedrock interface or the tile drains are generated in the same manner: starting at the interface to the stream and stepwise extending them upslope, again with a small probability for a vertical upward or downwards step to allow for tortuosity (Figure 3 C and D).

3. Parametrization of the representative hillslope models

3.1 Colpach catchment

3.1.1 Surface topography and spatial discretization

We extracted 241 hillslope profiles based on the available DEM in the Colpach catchment using Whitebox GIS (Lindsay J.B., 2014) following the LUMP approach (Landscape Unit Mapping Program, Francke et al., 2008). Based on these profiles (Figure 6 A) we derived a representative hillslope with a length of 350 m, a maximum elevation of 54 m above the stream, and a total area of 42600 m². The hillslope has a mean slope angle of 11.6° and is facing south (186°), similar to the average aspect of the Colpach catchment. The first step in generating the representative hillslope profile was to calculate the average distance to the river of all 241 extracted hillslope profiles as equal to 380 m. In the next step all elevation and width values of the profiles were binned into 1 m "distance classes" from the river ranging up to the average distance of 380 m. For each class the median values of the a) elevation above the stream and b) the hillslope width were derived and used for the representative hillslope profile (Figure 6 A). For numerical simulation the hillslope was discretized into 766 horizontal and 24 vertical elements with an overall hillslope thickness of 3 m. The vertical grid size was set to 0.128 m, with a reduced vertical grid size of the top node of 0.05 m. Grid size in downslope direction varied between 0.1 m within and close to the rapid flow path and 1 m within reaches without macropores (Figure 3 C). The hillslope thickness of 3 m was chosen to reflect the average of the deepest points of the available bedrock topographies extracted from ERT profiles, which was 2.7 m.

Boundary conditions were set to atmospheric boundary at the top and no flow boundary at the right margin. At the left boundary of the hillslope we selected seepage-boundary condition, where the stored water amount above field capacity leaves the system. A gravitational flow boundary condition was established for the lower boundary. We used spin-up runs with initial states of 70 % of saturation for the entire hydrological year of interest and used the resulting soil moisture pattern for model initialization. This initialization approach was also used for the Wollefsbach catchment.

3.1.2 Land use and vegetation parametrization

According to the land use maps, the hillslopes are mostly forested. As the hilltop plateaus account only for a very small part of the representative hillslope, the land use type for the entire hillslope is set to forest (Figure 2 A). Start and end of the vegetation period was defined using the temperature-degree model of Menzel et al. (2003), which allowed successful identification of the tipping point between the winter and vegetation season in the double mass curves of the Colpach and of the Wollefsbach (compare Figure 5). We further used observed leaf area indices (LAI) to parameterize the evapotranspiration routine. However, since only fourteen single measurements at different positions are available for the entire schist area and vegetation period, we use the median of all LAI observations from August as a constant value of 6.3 for the vegetation period. To account for the annual pattern of the vegetation phenology we interpolate the LAI for the first and last 30 days of the vegetation period linearly between zero and 6.3, respectively. The other evapotranspiration parameters are displayed in Table 2 and were taken from Breuer et al. (2003) or Schierholz et al. (2000).

3.1.3 Bedrock-topography, permeability and soil hydraulic functions

We used the shape of the bedrock contour line of the ERT image (Figure 6) to constrain the relative topography of the bedrock interface in the hillslope model as follows. We scaled the 100 m of bedrock topography to the hillslope length of 380 m. We then used the average depth to bedrock from all seven available ERT measurements (2.7 m) to scale the maximum depth to bedrock in our model. To this end we divided the average depth of 2.7 m by the deepest point of the bedrock in Figure 6 B (3.3 m) and used the resulting factor of 0.88 to reduce the bedrock depth of Figure 6 B relatively at all positions. As a result, the soil depths to the bedrock interface vary between 1 m to 2.7 m with local depressions that form water holding pools. Since no major groundwater body is suspected and no quantitative data on the rather impermeable schist bedrock in the Colpach is available, we use a relative impermeable bedrock parametrization suggested by Wienhöfer and Zehe (2014, Table 1). It is important to note that due to this bedrock parametrization water flow through the hillslope lower boundary tends to zero.

The silty soil above the bedrock was modeled with the representative hydraulic parameters obtained from field samples listed in Table 1. Since there was no systematic variation of hydraulic parameters of the individual soil samples with depth, soil hydraulic parameters were set constant over depth, except for porosity, which was reduced to a value of 0.35 (m^3m^{-3}) at 50 cm depth to account for the increasing skeleton fraction of around 40% in deeper soil layers.

3.1.4 Rapid subsurface flow paths

Macropore depths were drawn from a normal distribution with a mean of 1 m and a standard deviation of 0.3 m. These values are in agreement with the mean soil depth and correspond well with the results of dye staining experiments performed by Jackisch (2015) and Jackisch et al. (2016). Additionally, macropores were slightly tortuous with a probability for a lateral step of 5 %. Since no observations for the macropore density were available, we use a fixed macropore distance of 2 m. The macropore distance was chosen rather arbitrarily to reflect their relative density in the perceptual model and to establish a partly connected network of vertical and lateral rapid flow paths. The vertical flow paths were parametrized using an artificial porous medium with high hydraulic conductivity and low retention properties proposed by Wienhöfer and Zehe (2014, Table 1). Also the weathered periglacial saprolite layer which is represented by a 0.2 m thick layer above the bedrock was parameterized as a porous medium following Wienhöfer and Zehe, (2014). The estimated saturated hydraulic conductivity of $1*10^{-3}$ m s^{-1} corresponds well with the velocities described by Angermann et al. (2016). This ensures that the Reynolds number is smaller than 10, implying that flow can be considered laminar and the application of Darcy's law is still appropriate (Bear, 1972).

3.2 Wollefsbach catchment

3.2.1 Surface topography and spatial discretization

Since only eight relatively similar hillslope profiles were derived from the DEM in the Wollefsbach we randomly chose one of those with a length of 653 m, a maximal elevation above the river of 53 m and an area of 373600 m². The hillslope has a mean slope angle of 8.1° and is facing south (172°). The hillslope was discretized into 553 horizontal and 21 vertical elements with an overall hillslope thickness of 2 m (Figure 3 D). The vertical grid size was set to 0.1 m, with a reduced top and bottom node spacing of 0.05 m. Grid size in lateral direction varied between 0.2 m within and close to the rapid flow paths and 2 m within reaches without macropores (Figure 3 B and D).

3.2.2 Land use and vegetation parametrization

Land use was set to grassland within the steeper and lower part of the hillslope, and set to corn for larger distances to the creek (>325 m). Due to the absence of local vegetation data we used tabulated data characterizing grassland and corn from Breuer et al. (2003).

Start and end point of the vegetation period for the grassland and the start point for the corn cultivation were again identified by the temperature index model of Menzel et al. (2003). The vegetation period for the corn cultivation ends in the beginning of October since this is the typical period for harvesting. The intra-annual vegetation dynamics were taken from Schierholz et al. (2000).

3.2.3 Bedrock-topography, -permeability and soil hydraulic functions

Contrary to the Colpach, geophysical measurements and augerings revealed bedrock and surface to be more or less parallel. Soil depth was set to constant 1 m and the soil was parameterized using the representative soil retention curves shown in Figure 7. The bedrock was again parameterized according to values Wienhöfer & Zehe (2014) proposed for the impermeable bedrock at the Heumöser hillslope in Austria (Table 1), which is also in a marl geology.

3.2.4 Rapid subsurface flow paths

Based on the perceptual model (Figure 3 B and D) and the reported vertical and lateral drainage structures in the catchment we generated a network of fast flow paths. The depths of the vertical flow paths were drawn from a normal distribution with a mean of 0.8 m and a standard deviation of 0.1 m. The tile drain was generated at the standard depth of 0.8 m extending 400 m upslope from the hillslope creek interface. Due to the apparent changes in soil structure either by earthworm burrows or emergent soil cracks (Figure 4), we used different macropore setups for the winter and the vegetation season. For the winter setup we implemented vertical drainage structures every four meters. In the summer setup we added fast flow paths every two meters to account for additional cracks and earthworm burrows. The positions of the conceptual macropores were selected again arbitrarily to create an image of the perceptual model and to assure that the soil surface and the tile drain were well connected. Vertical flow paths and the tile drain were parametrized similar to the Colpach with the same artificial porous medium (Table 1). Boundary conditions of the hillslope, initialization and the spin up phase were the same as described for the Colpach model.

3.3 Model scenarios

Both hillslopes models were set up within a few test simulations to reproduce the normalized double mass curves in both catchments of the hydrological year 2014. We choose the normalized double mass curves as a fingerprint of the annual pattern of runoff generation since it is particular suitable for detecting differences in inter-annual and seasonal runoff dynamics of a catchment (Jackisch, 2015). Model performance was judged by visual inspection as well as by using the Kling-Gupta efficiency (KGE, Gupta et al. 2009).

In a second step we compared the simulated overland flow and subsurface storm flow across the left hillslope boundary to observed discharge. Water leaving the hillslope through the lower boundary was neglected from the analysis because in both setups the total amount was smaller than 1 % of the overall hillslope outflow. We compared the specific discharge of the hillslopes to the observed specific discharge of the two catchments in mm h^{-1} by dividing measured and simulated discharge by the area of the catchments and the hillslopes. Our goal was to test if our hillslope models represented the typical subsurface filter properties which are relevant for the runoff generation in both selected hydrological landscapes (schist and marl area in the Attert basin). We measured the model performance with respect to discharge again based on the KGE. Since it is advisable to calculate and display various measures of model performance (Schaefli and Gupta, 2007), we calculated the Nash-Sutcliffe efficiency (NSE; a measure of model performance with emphasis on high flows) and the logarithmic NSE (log NSE; a performance measure suited for low flows). As both catchments are characterized through a strong seasonality we further separated the simulation period in a winter and vegetation period and calculated the KGE, NSE as well as the logNSE separately for each of the seasons. In addition, we followed Klemeš (1986) and performed a proxy-basin test to check if the runoff simulation is transposable within the same hydrological landscape and conducted a split sampling to examine if the models also work in the hydrological year of 2013. Finally, we judged the model goodness visually for selected rainfall-runoff events.

In a third step we evaluated the model setups against available soil moisture observations. A natural starting point for a modeling study would be to classify the available soil moisture observation for instance by their landscape position. However, similar to the case of the soil water retention properties, the small scale variability of the soil properties seems to be too dominant, as grouping according to hillslope position was not conclusive (Jackisch, 2015; appendix A4). We therefore extracted simulated soil moisture at 20 virtual observation points at different downslope positions at the respective depth of the soil moisture observations (10 and 50 cm), and compared the median of the simulated virtual observations against the 12-hours-rolling median of the observed soil moisture using the KGE and the Spearman rank correlation. Finally, we analyzed simulated transpiration of the Colpach model by plotting it against the 12-hours-rolling median of the daily sap flow velocities observed in the Schist area of the Attert basin. As sap flow is a velocity and transpiration is a normalized flow they are not directly comparable. This is why we normalized both observed sap flow and simulated transpiration by dividing their values by their range and only discuss the correlation among the normalized values. The visual inspection shows additionally to which extent maximum and minimum values of both normalized time series coincide. This cannot be inferred from the correlation coefficient.

4. Results

4.1 Normalized double mass curves and discharge

The hillslope models reproduce the typical shape of the normalized double mass curves – the steep, almost linear increase in the winter period and the transition to the much flatter summer regime – in both catchments very well (Fig. 8 A, B).

The KGEs of 0.92 and 0.9 obtained for the Colpach and the Wollefsbach, respectively, corroborate that within the error ranges both double mass curves are explained well by the models. As a major groundwater body is unlikely in both landscapes, a large inter-annual change in storage is not suspected and we hence state that the hillslope models closely portray the seasonal patterns of the water balance of the catchments. This is further confirmed by the close accordance of simulated and observed annual runoff coefficients. We obtain 0.52 compared to the observed value of 0.55 in the Colpach and 0.39 compared to an observed value of 0.42 in the Wollefsbach.

In addition to the seasonal water balances, both models also match observed discharge time series in an acceptable manner (KGE 0.88 and 0.71; Table 3). A closer look at the simulated and observed runoff time series (Figure 9 and 10) reveals that the model performance differs in both catchments between the winter and the summer seasons. Generally we observe a better model accordance during the wet winter season, when around 80% of the overall annual runoff is generated in both catchments. In contrast, there are clear deficiencies during dry summer conditions. This is also highlighted by the different performance measures which are in both catchments higher during the winter period than during the vegetation period (Table 3).

The Colpach model misses especially the steep and flashy runoff events in June, July and August, and underestimates discharge in summer. It also misses the characteristic double peaks of the catchment as highlighted by runoff events 2 and 3 (Figure 9). Although the model simulates a second peak, it is either too fast (event 2) or the simulated runoff of the second peak is too small (event 3). This finding suggests that our perceptual model of the Colpach catchment needs to be revised, as further elaborated in the discussion. Another shortcoming is the missing snow routine of CATFLOW which can be inferred from event 1 (Figure 9 top left panel). While snow is normally not a major control of runoff generation in the rather maritime climate of the Colpach catchment, the runoff event 1 happened during temperatures below zero and was most likely influenced by snowfall and subsequent snow melt, which might explain the delay in observed rainfall-runoff response.

In the Wollefsbach model the ability to match the hydrograph also differed strongly between the different seasons (Table 3; Figure 10). The flashy runoff respone in summer is not always well captured by the model, as for example for a convective rainfall event with rainfall intensities of up to 18 mm (10 mins)$^{-1}$ in August (Figure 10, event 2).

On the contrary, runoff generation during winter is generally simulated acceptably (KGE = 0.74). Yet, the model strongly underestimates several runoff events in winter too (Figure 10, event 1). As temperatures during these events were close to zero, this might again be a result of snow accumulation, which cannot be simulated with CATFLOW due to the missing snow routine. It is of key importance to stress that we only achieve acceptable simulations of runoff production in the Wollefsbach when using two different macropore setups for the winter and the summer periods to account for the emergence of cracks (Figure 4) by using a denser 2m-spacing of macropores. When using a single macropore distance of either 2 m (summer setup) or 4 m (winter setup) in the entire simulation period the model shows clear deficits with a KGE of 0.61 and 0.53, respectively. Furthermore, we are able to improve the performance of the Wollefsbach model if we use velocities faster than $1*10^{-3}$ m/s for the drainage structures. However, this violates the laminar flow assumption and the application of Darcy's law becomes inappropriate.

4.2 Model sensitivities, split sampling and spatial proxy test

Sensitivity tests for the Colpach reveal that the model performance of matching the double mass curves is strongly influenced by the presence of connected rapid flow paths.

A complete removal of either the vertical macropores or the bedrock interface from the model domain decreases the model performance considerably (KGE 0.71 or 0.72, respectively). In contrast, reducing the density of vertical macropores from 2 m to 3 or 4 m only leads to a slight decrease in model performance (KGE 0.85 and 0.82, respectively). In an additional sensitivity test we changed the bedrock topography from the one inferred from the ERT data to a surface parallel one, which reduces model performance with respect to discharge (KGE < 0.6).

The temporal split-sampling reveals that the representative hillslope model of the Colpach also performs well in matching the hydrograph of the previous hydrological year 2012-13 (KGE = 0.82). Furthermore, the parameter setup was tested within uncalibrated simulations for the Weierbach catchment (0.45 km^2), a headwater of the Colpach in the same geological setting. This again leads to acceptable results (KGE = 0.81, NSE = 0.68). The same applies to the representative hillslope model of the Wollefsbach which also performs well in matching the hydrograph of the previous year (KGE = 0.7). Furthermore, the parameter setup was tested within an uncalibrated simulation for the Schwebich catchment (30 km^2), a headwater of the Attert basin in the same geological setting as the Wollefsbach, and again with acceptable results (KGE = 0.81, NSE = 0.7).

4.3 Simulated and observed soil moisture dynamics

We compare the ensemble of soil moisture time series from the virtual observation points to the ensemble of available observations (Figure 11). In the Colpach, soil moisture dynamics are matched well (Spearman rank correlation r_s = 0.83). This is further confirmed when comparing this value to the median Spearman rank correlation coefficient of all sensor pairs (r_s = 0.66). However, simulated soil moisture at 10 cm depth was systematically higher than the average of the observations. The predictive power in matching the observed average soil moisture dynamics was small (KGE = 0.43; Figure 11 A). Contrary to the positive bias, the total range of the simulated ensemble appears with 0.1 $m^3 m^{-3}$ much smaller than the huge spread in the observed time series (0.25 $m^3 m^{-3}$). In line with the model performance in simulating discharge, the model has deficiencies in capturing the strong declines in soil moisture in June and July. Simulated soil moisture at 50 cm depth exhibits a strong positive bias and again underestimates the spread in the observed time series. The predictive power is slightly better (KGE = 0.51), while simulated and observed average dynamics are in good accordance (r_s = 0.89).

Contrary to what we found for the Colpach, the ensemble of simulated soil moisture at 10 cm for the Wollefsbach falls into the state space spanned by the observations; it only slightly underestimates the rolling median of the observed soil moisture (Figure 11 C. The predictive power is higher (KGE = 0.67) than in the Colpach, while the match of the temporal dynamics is slightly lower (r_s = 0.81). Again the model fails to reproduce the strong decline in soil moisture between May and July. It is, however, interesting to note that the model is nearly unbiased during August and September. This is especially interesting since the Wollefsbach model does not perform too well in simulating discharge during this time period. Simulated soil moisture at 50 cm depth shows similar deficiencies as found for the Colpach, while the predictive power was slightly smaller (KGE = 0.44), and also the dynamics is matched slightly worse (r_s = 0.79).

When recalling the soil water retention curves (Figure 7), one can infer that a soil water content of 0.2 $m^3 m^{-3}$ corresponds to pF around 3.8 in the Colpach and to pF around 4.1 in the Wollefsbach. That in mind it is interesting to note that some observed soil moisture values are below this threshold throughout the entire year. This is particularly the case for soil moisture observation at 50 cm depth in the Colpach where almost 50 % of the sensors measure water contents close to the permanent wilting point throughout the wet winter period. This also holds true for 8 sensors at 10 cm depth.

4.4 Normalized simulated transpiration versus normalized sap flow velocities

As sap flow provides a proxy for transpiration, we compared normalized, averaged sap flow velocities of beech and oak trees to the normalized simulated transpiration of the reference hillslope model of the Colpach. The three-day-rolling-mean of sap flow data stays close to zero until the end of April and starts to rise after the bud break of the observed trees. The Colpach model is able to match the bud break of the vegetation well. Furthermore, simulations and observations are in good accordance during midsummer. In the period between August and October the simulations underestimate the observations, while in April and May the simulations are too high (Figure 12). Nevertheless, the model has some predictive power (KGE = 0.65), and is able to mimic the dynamics well (r_s = 0.75).

5 Discussion

The presented model results partly corroborate our hypothesis that single representative hillslopes might serve as the most parsimonious representations of two distinctly different lower meso-scale catchments in a physically-based model. The setups of the representative hillslopes were derived as close images of the available perceptual models and by drawing from a variety of field observations, literature data and expert knowledge. The hillslope models were afterwards tested against stream flow data, including a split sampling and a proxy basin test, and against soil moisture and partly against sap flow observations.

From the fact that stream flow simulations were acceptable in both catchments when being judged solely on model efficiency criteria, one could conclude that the hillslopes portray the dominant structures and processes which control the runoff generation in both catchments well. A look beyond streamflow-based performance measures revealed, however, clear deficiencies in stream flow simulations during the summer season and during individual rainfall-runoff events as well as a mismatch in simulated soil water dynamics. In the next sections we will hence discuss the strengths and the weaknesses of the representative hillslope model approach. More specifically, in section 5.1 we will focus on the role of soil heterogeneity, preferential flow paths and the added value of geophysical images. In section 5.2 we will discuss the consistency of both models with respect to their ability to reproduce soil moisture and transpiration dynamics. Finally in section 5.3, we discuss if the general idea to picture and model a catchment by a single 2-D representative hillslope is indeed appropriate to simulate the functioning of a lower-mesoscale catchment.

5.1.1 The role of soil heterogeneity for discharge simulations

By using an effective soil water retention curve, instead of accounting for the strong variability of soil hydraulic properties among different soil cores (section 2.2.3) we neglect the stochastic heterogeneity of the soil properties controlling storage and matrix flow. This simplification is a likely reason why the model underestimates the spatial variability in soil moisture time series (compare section 5.2.1). However, our approach does not perform too badly in simulating the normalized double mass curves as well as the runoff generation, at least to some extent, in both catchments. Especially during the winter, when around 80 % of the runoff is generated, runoff is reproduced acceptably well. As our models do not represent the full heterogeneity of the soil water characteristics but are still able to reproduce the runoff dynamics in winter, we reason in line with Ebel and Loague, (2006) that heterogeneity of soil water retention properties is not too important for reproducing the stream flow generation in catchments. In this context it is helpful to recall the fact that hydrological models with three to four parameters are often sufficient to reproduce the stream flow of a catchment. This corroborates that the dimensionality of stream flow is much smaller than one could expect given the huge heterogeneity of the retention properties. This finding has further implications for hydrological modelling approaches as it once more opens the question on the amount of information that is stored in discharge data and how much can be learned when we do hydrology backwards (Jakeman and Hornberger, 1993). Our conclusion should, however, not be misinterpreted that we claim the spatial variability of retention properties to be generally unimportant. The variability of the soil properties of course plays a key role as soon as the focus shifts from catchment-scale runoff generation to solute transport processes, infiltration patterns or to water availability for evapotranspiration.

5.1.2 The role of drainage structures and macropores for discharge simulations

By representing preferential flow paths as connected networks containing an artificial porous medium in the Richards domain, we assume that preserving the connectedness of the network is more important than the separation of rapid flow and matrix flow into different domains. The selected approach was successful in reproducing runoff generation and the water balance for the winter period in the Wollefsbach and Colpach catchments. Simulations with a disconnected network, where either the saprolite layer at the bedrock interface or the vertical macropores were removed, reduced the model performance in the Colpach model from KGE = 0.88 to KGE = 0.6 and KGE = 0.71, respectively. We hence argue that capturing the topology and connectedness of rapid flow paths is crucial for the simulation of stream flow release with representative hillslopes. We furthermore showed that a reduction in the spatial density of macropores from a 2 m to 4 m spacing did not strongly alter the quality of the discharge simulations. This insensitivity can partly be explained by the fact that several configurations of the rapid flow network may lead to a similar model performance. From this insensitivity and the equifinality of the network architecture (Klaus and Zehe, 2010; Wienhöfer and Zehe, 2014) we conclude that it is not the exact position or the exact extent of the macropores which is important for the runoff response but the bare existence of a connected rapid flow path (Jakeman and Hornberger, 1993).

However, our results also reveal limitations of the representation of rapid flow paths in CATFLOW. For instance model setups with higher saturated hydraulic conductivities ($>10^{-3}$) of the macropore medium clearly improved the model performance in the Wollefsbach but violated the fundamental assumption of Darcy’s law of pure laminar flow. This was likely one reason why capturing rapid flow was much more difficult with the selected approach for the Wollefsbach. Another reason was that the emergence of cracks, implying that the relative importance of rapid flow paths for runoff generation is not constant over the year, as highlighted by the findings of dye staining experiments (Figure 4). Given this non-stationary configuration of the macropore network it was indispensable to use a summer and winter configuration to achieve acceptable simulations. This indicates that besides the widely discussed limitations of the different approaches to simulate macropore flow, another challenge is how to deal with emergent behavior and related non-stationary in hydrological model parameters. This is in line with the work of Mendoza et al. (2015), who showed that the agility of hydrological models is often unnecessarily constrained by using static parametrizations. We are aware that the use of a separate model structure in the summer period is clearly only a quick fix, but it highlights the need for more dynamic approaches to account for varying morphological states of the soil structure during long-term simulations.

5.1.3 The role of bedrock topography and water flow through the bedrock

The Colpach model was able to simulate the double peak runoff events which are deemed as typical for this hydrological landscape. However, the model did not perform satisfactorily with regard to peak volume and timing. A major issue that hampers the simulation of these runoff events is that it the underlying hydrological processes are still under debate. While Martínez-Carreras et al. (2015) attributes the first peak to water from the riparian zone and the second to subsurface storm flow, other researchers (Angermann et al., 2016; Graeff et al., 2009) suggested that the first peak is caused by subsurface storm flow and the second one by release of groundwater. The representative hillslope model in its present form only allows simulation of overland flow and subsurface storm flow and not the release of groundwater because of the low permeability of the bedrock medium of 10^{-9} m s^{-1}. The deficiency of this model to reproduce double peak runoff events shows that neglecting water flow through the bedrock is possibly not appropriate

(Angermann et al. 2016) and that both the perceptual model and the setup of the representative hillslope for the Colpach need to be refined. We hence suggest that the representative hillslope approach provides an option for a hypothesis-driven refinement of perceptual models, within an iterative learning cycle, until the representative hillslope reproduces the key characteristics one regards as important.

The importance of bedrock topography for the interplay of water flow and storage close to the bedrock was further highlighted by the available 2-D electric resistivity profiles. A model with surface-parallel bedrock topographies performed considerably worse in matching stream flow in terms of the selected performance measures and particularly did not produce the double peak events. This underlines the value of subsurface imaging for process understanding, and is a hint that the Colpach is indeed a fill-and-spill system (Tromp-Van Meerveld and McDonnell, 2006). It also shows that 2-D electric resistivity profiles can be used to constrain bedrock topography in physically-based models (Graeff et al., 2009), which can be of key importance for simulating subsurface storm flow (Hopp and McDonnell, 2009; Lehmann et al., 2006). Although we used constrained bedrock topography only in a straightforward, relative manner in this study, our results corroborated the added value of ERT profiles for hydrological modelling in this kind of hydrological landscapes. Nevertheless, we know that a much more comprehensive study is needed to further detail this finding.

5.2 Integration and use of multi-response and state variables

5.2.1 Storage behavior and soil moisture observations

Both hillslope models reveal much clearer deficiencies with respect to soil moisture observations. While average simulated and observed soil moisture dynamics are partly in good accordance, both models are biased except for the Wollefsbach model at 10 cm depth. In the Wollefsbach catchment this might be explained by the fact that we use an uniform soil porosity for the entire soil profile, although porosity is most likely lower at larger depths for instance due to a higher skeleton fraction. This is no explanation for the Colpach catchment as porosity was reduced in deeper layers with respect to the skeleton fraction. In this context it is interesting to note that quite a few of the soil moisture observations are suspiciously low with average values around 0.2. The resulting pF values of around 3.8 and 4.1 in the Colpach and Wollefsbach, respectively, indicate dry soils even in the wet winter period. This fact has two implications: The first is that the chosen model is almost not capable to simulate such small values, because root water uptake stops at the permanent wilting point and is small at these pF values. The second is that these sensors may have systematic measurement errors, possibly due to entrapped air between the probe and the soil. This entrapped air decreases the dielectric permittivity close to the sensor (Graeff et al., 2010), which implies that measured values will be systematically too low. From this we may conclude that average soil moisture dynamics in both catchments might be higher and the spatial variability of soil moisture time series in turn lower as it appears from the measurements.

Additional to the mismatch of the soil moisture simulations, the model fails in reproducing the strong decline in observed soil moisture between May and July 2014. A likely reason for this is that plant roots in the model extract water uniformly within the root zone, while this process is in fact much more variable (Hildebrandt et al., 2015).

5.2.2 Simulated transpiration and sap velocities

It is no surprise that evapotranspiration in our two research catchments is - with a share of around 50 % of the annual water balance - equally important as stream flow. It is also no surprise that evapotranspiration is dominated by transpiration as both catchments are almost entirely covered by vegetation. However, measuring transpiration remains a difficult task, and a lack of reliable transpiration data often hinders the evaluation of hydrological models with respect to this important flux. While it is possible to calculate annual or monthly evapotranspiration sums based on the water balance, more precise information about the temporal dynamics of transpiration is difficult to obtain. Therefore we decided to evaluate our transpiration routine with available sap flow velocity data, because although the absolute values are somewhat error-prone, the dynamics are quite reliable. We tried to account for the uncertainties of the measurements by deriving a three-day-rolling median of 28 observations instead of using single sap flow velocity measurements. As we are comparing sap flow velocity to the simulated transpiration as a normalized flow, we only compare the dynamics of both variables. It is remarkable that despite the uncertainties in the sap flow velocity measurements and our ad-hoc parametrization of the vegetation properties, the comparison of sap flow velocity and simulated transpiration provides additional information, which cannot be extracted from the double mass curve or discharge data. For example, based on the comparison with sap flow velocities we were able to evaluate if the bud break of the dormant trees was specified correctly by the temperature index model of Menzel et al. (2003). Additionally, we could identify that the spring and autumn dynamics of transpiration, in April as well as in August and September, are matched poorly by the model while the pattern corresponds well in May, June and July. We attribute this discrepancy to the lack of measured LAI values in spring and autumn and to our possibly overly simple vegetation parametrization including several parameters like root depth or plant albedo which are held constant throughout the entire vegetation period. We are aware that this comparison of modeled transpiration with sap flow velocity is only a first, rather simple test; however it encourages the use of sap flow measurements for hydrological modeling. It shows furthermore that the concept of a representative hillslope offers various opportunities for integrating diverse field observations and testing the model's hydrological consistency, for example evaluating it against soil water retention data and sap flow velocities.

5.3 The concept of representative hillslope models

The attempt to model catchment behavior using a two-dimensional representative hillslope implies a symmetry assumption in the sense that the water balance is dominated by the interplay of hillslope parallel and vertical fluxes and the related driving gradients (Zehe et al., 2014). This assumption is corroborated by the acceptable but yet seasonally dependent performance of both hillslope models with respect to matching the water balance and the hydrographs. We particularly learn that the timing of runoff events in these two catchments is dominantly controlled by the structural properties of the hillslopes. This is remarkable for the Colpach catchment which has a size of 19.4 km^2, but in line with Robinson et al., (1995) who showed that catchments of up to 20 km^2 can still be hillslope dominated.

An example of the limitations of our single hillslope approach is the deficiency of both models in capturing flashy rainfall-runoff events in the vegetation period. Besides the existence of emergent structures, these events might likely be caused by localized convective storms, probably with a strong contribution of the riparian zones (Martínez-Carreras et al., 2015) and forest roads in the Colpach catchment, and by localized overland flow in the Wollefsbach catchment (Martínez-Carreras et al., 2012). Such fingerprints of a non-uniform rainfall forcing are difficult to be captured by a simulation with a spatially aggregated model; and might require an increase in model complexity. Nevertheless, we suggest that a representative hillslope model provides the right start-up for parameterization of a functional unit when setting up a fully distributed catchment model consisting of several hillslopes and an interconnecting river network. Simulations with distributed rainfall and using the same functional unit parameterization for all hillslopes would tell how the variability in response and storage behavior can be explained compared to the single hillslope. If different functional units are necessary to reproduce the variability of distributed fluxes and storage dynamics, these can for example be generated by stochastic perturbation. We further conclude that the idea of hillslope-scale functional units, which act similarly with respect to runoff generation and might hence serve as building blocks for catchment models, has been corroborated. This is particularly underpinned by the fact that the parameterization of both models was – without tuning – successfully transferred to headwaters in the same geological setting and worked also well for other hydrological years.

6. Conclusions

The exercise to picture and model the functioning of an entire catchment by using a single representative hillslope proved to be successful and instructive. The picturing approach allowed us to consider both quantitative and qualitative information in the physically-based modeling process. This concept made an automated parameter calibration unnecessary and lead to overall acceptable stream flow simulations in two lower-mesoscale catchments. A closer look, however, revealed limitations arising from the drawn perceptual models, the chosen hydrological model or the applicability of the concept itself.

Distilling a catchment into a representative hillslope model obviously cannot reflect the entire range of the spatially distributed catchment characteristics. But as the stream flow dynamics of the catchments were simulated reasonably well and the models were even transferable to different catchments it seems that, the use of physically-based models and the large heterogeneities in subsurface characteristics must not prevent meaningful simulations. Additionally, our results highlight the importance of considering non-stationarity of catchment properties in hydrological models on seasonal time scales and emphasize once more the value of multi-response model evaluation. A representative hillslope model for a catchment is, hence, perhaps less accurate than a fully distributed model, but in turn also requires considerably less data and reduced efforts for setup and computation. Therefore, this approach provides a convenient means to test different perceptual models and it can serve as a starting point for increasing model complexity through combination of different hillslopes and a river network to model a catchment in a more distributed manner.

Acknowledgements

This research contributes to the "Catchments As Organized Systems (CAOS)" research group (FOR 1598) funded by the German Science Foundation (DFG). Laurent Pfister and Jean-Francois Iffly from the Luxembourg Institute of Science and Technology (LIST) are acknowledged for organizing the permissions for the experiments and providing discharge data for Wollefsbach and Colpach. We also thank the whole CAOS team of phase I & II. Particular we thank Malte Neuper (KIT) for support and discussions on the rainfall data and Markus Weiler, Theresa Blume and Britta Kattenstroth for providing and collecting the soil moisture data. Finally we would like to thank the two anonymous reviewers as they significantly helped to improve and restructure this manuscript.

References

Angermann, L., Jackisch, C., Allroggen, N., Sprenger, M., Zehe, E., Tronicke, J., Weiler, M. and Blume, T.: In situ investigation of rapid subsurface flow: Temporal dynamics and catchment-scale implication, Hydrol. Earth Syst. Sci. Discuss., 2016(May), 1–34, doi:10.5194/hess-2016-189, 2016.

Bachmair, S. and Weiler, M.: Forest Hydrology and Biogeochemistry, edited by D. F. Levia, D. Carlyle-Moses, and T. Tanaka, Springer Netherlands, Dordrecht., 2011.

Bear, J.: Dynamics of Fluids in Porous Media, American Elsevier, New York., 1972.

Bergström, S. and Forsman, A.: Development of a conceptual deterministic rainfall-runoff-model, Hydrol. Res., 4(3), 1973.

Berne, A., Uijlenhoet, R. and Troch, P. A.: Similarity analysis of subsurface flow response of hillslopes with complex geometry, Water Resour. Res., 41(9), n/a–n/a, doi:10.1029/2004WR003629, 2005.

Beven, K.: Changing ideas in hydrology—the case of physically-based models, J. Hydrol., 105, 157–172, 1989.

Beven, K.: Searching for the Holy Grail of scientific hydrology, , 609–618, 2006a.

Beven, K.: Streamflow Generation Processes: Introduction, 1st edtion., edited by K. J. Beven, University of Lancaster, Lancaster, UK., 2006b.

Beven, K. and Germann, P.: Macropores and water flow in soils revisited, Water Resour. Res., 49(6), 3071–3092, doi:10.1002/wrcr.20156, 2013.

Bishop, J. M., Callaghan, M. V., Cey, E. E. and Bentley, L. R.: Measurement and simulation of subsurface tracer migration to tile drains in low permeability, macroporous soil, Water Resour. Res., 51(6), 3956–3981, doi:10.1002/2014WR016310, 2015.

Bos, R. van den, Hoffmann, L., Juilleret, J., Matgen, P. and Pfister, L.: Conceptual modelling of individual HRU ' s as a trade-off between bottom-up and top-down modelling , a case study ., in Conf. Environmental Modelling and Software. Proc. 3rd Biennal meeting of the international Environmental Modelling and Software Society. Vermont, USA., 1996.

Breuer, L., Eckhardt, K. and Frede, H.-G.: Plant parameter values for models in temperate climates, Ecol. Modell., 169(2-3), 237–293, doi:10.1016/S0304-3800(03)00274-6, 2003.

Bronstert, A. and Plate, E. J.: Modelling of runoff generation and soil moisture dynamics for hillslopes and micro-catchments, J. Hydrol., 198(1-4), 177–195, doi:10.1016/S0022-1694(96)03306-9, 1997.

Brunner, P. and Simmons, C. T.: HydroGeoSphere: A Fully Integrated, Physically Based Hydrological Model, Ground Water, 50(2), 170–176, doi:10.1111/j.1745-6584.2011.00882.x, 2012.

Burgess, S. S., Adams, M. a, Turner, N. C., Beverly, C. R., Ong, C. K., Khan, a a and Bleby, T. M.: An improved heat pulse method to measure low and reverse rates of sap flow in woody plants., Tree Physiol., 21(9), 589–598, doi:10.1093/treephys/21.9.589, 2001.

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

Celia, M. A., Bouloutas, E. T. and Zarba, R. L.: A general mass-conservative numerical solution for the unsaturated flow equation, Water Resour. Res., 26(7), 1483–1496, doi:10.1029/WR026i007p01483, 1990.

Coenders-Gerrits, A. M. J., Hopp, L., Savenije, H. H. G. and Pfister, L.: The effect of spatial throughfall patterns on soil moisture patterns at the hillslope scale, Hydrol. Earth Syst. Sci., 17(5), 1749–1763, doi:10.5194/hess-17-1749-2013, 2013.

Dooge, J. C. I.: Looking for hydrologic laws, Water Resour. Res., 22(9), 46S, doi:10.1029/WR022i09Sp0046S, 1986.

Dunne, T. and Black, R. D.: Partial Area Contributions to Storm Runoff in a Small New England Watershed, Water Resour. Res., 6(5), 1296–1311, doi:10.1029/WR006i005p01296, 1970.

Ebel, B. a. and Loague, K.: Physics-based hydrologic-response simulation: Seeing through the fog of equifinality, Hydrol. Process., 20(13), 2887–2900, doi:10.1002/hyp.6388, 2006.

Ebel, B. a., Loague, K., Montgomery, D. R. and Dietrich, W. E.: Physics-based continuous simulation of long-term near-surface hydrologic response for the Coos Bay experimental catchment, Water Resour. Res., 44(7), 1–23, doi:10.1029/2007WR006442, 2008.

Fenicia, F., Kavetski, D., Savenije, H. H. G., Clark, M. P., Schoups, G., Pfister, L. and Freer, J.: Catchment properties, function, and conceptual model representation: is there a correspondence?, Hydrol. Process., 28(4), 2451–2467, doi:10.1002/hyp.9726, 2013.

Francke, T., Güntner, a., Mamede, G., Müller, E. N. and Bronstert, a.: Automated catena-based discretization of landscapes for the derivation of hydrological modelling units, Int. J. Geogr. Inf. Sci., 22(2), 111–132, doi:10.1080/13658810701300873, 2008.

Freeze, R. A. and Harlan, R. L.: Blueprint for a physically-based, digitally-simulated hydrologic response model, J. Hydrol., 9(3), 237–258, doi:10.1016/0022-1694(69)90020-1, 1969.

Graeff, T., Zehe, E., Reusser, D., Lück, E., Schröder, B., Wenk, G., John, H. and Bronstert, A.: Process identification through rejection of model structures in a mid-mountainous rural catchment: observations of rainfall-runoff response, geophysical conditions and model inter-comparison, Hydrol. Process., 23(5), 702–718, doi:10.1002/hyp.7171, 2009.

Graeff, T., Zehe, E., Schlaeger, S., Morgner, M., Bauer, A., Becker, R., Creutzfeldt, B. and Bronstert, A.: A quality assessment of Spatial TDR soil moisture measurements in homogenous and heterogeneous media with laboratory experiments, Hydrol. Earth Syst. Sci., 14(6), 1007–1020, doi:10.5194/hess-14-1007-2010, 2010.

Grayson, R. B., Moore, I. D. and McMahon, T. A.: Physically based hydrologic modeling: 2. Is the concept realistic?, Water Resour. Res., 28(10), 2659–2666, doi:10.1029/92WR01259, 1992.

Gupta, H. V., Kling, H., Yilmaz, K. K. and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377(1-2), 80–91, doi:10.1016/j.jhydrol.2009.08.003, 2009.

Gupta, H. V., Clark, M. P., Vrugt, J. a., Abramowitz, G. and Ye, M.: Towards a comprehensive assessment of model structural adequacy, Water Resour. Res., 48(8), 1–16, doi:10.1029/2011WR011044, 2012.

Hazenberg, P., Broxton, P., Gochis, D., Niu, G.-Y., Pangle, L. A., Pelletier, J. D., Troch, P. A. and Zeng, X.: Testing the hybrid-3-D hillslope hydrological model in a controlled environment, Water Resour. Res., 52(2), 1089–1107, doi:10.1002/2015WR018106, 2016.

Hildebrandt, A., Kleidon, A. and Bechmann, M.: A thermodynamic formulation of root water uptake, Hydrol. Earth Syst. Sci. Discuss., 12(12), 13383–13413, doi:10.5194/hessd-12-13383-2015, 2015.

Hopp, L. and McDonnell, J. J.: Connectivity at the hillslope scale: Identifying interactions between storm size, bedrock permeability, slope angle and soil depth, J. Hydrol., 376(3-4), 378–391, doi:10.1016/j.jhydrol.2009.07.047, 2009.

Hopp, L., Harman, C., Desilets, S., Graham, C., McDonnell, J. and Troch, P.: Hillslope hydrology under glass: confronting fundamental questions of soil-water-biota co-evolution at Biosphere 2, Hydrol. Earth Syst. Sci. Discuss., 6(3), 4411–4448, doi:10.5194/hessd-6-4411-2009, 2009.

Jackisch, C.: Linking structure and functioning of hydrological systems., KIT - Karlsruher Institut of Technology., 2015.

Jackisch, C., Angermann, L., Allroggen, N., Sprenger, M., Blume, T., Weiler, M., Tronicke, J. and Zehe, E.: In situ investigation of rapid subsurface flow: Identification of relevant spatial structures beyond heterogeneity, Hydrol. Earth Syst. Sci. Discuss., 2016, 1–32, doi:10.5194/hess-2016-190, 2016.

Jakeman, A. J. and Hornberger, G. M.: How much complexity is warranted in a rainfall-runoff model?, Water Resour. Res., 29(8), 2637–2649, doi:10.1029/93WR00877, 1993.

Jarvis, P. G.: The Interpretation of the Variations in Leaf Water Potential and Stomatal Conductance Found in Canopies in the Field, Philos. Trans. R. Soc. B Biol. Sci., 273(927), 593–610, doi:10.1098/rstb.1976.0035, 1976.

Juilleret, J., Iffly, J. F., Pfister, L. and Hissler, C.: Remarkable Pleistocene periglacial slope deposits in Luxembourg (Oesling): pedological implication and geosite potential, Bull. la Société des Nat. Luxemb., 112(1), 125–130, 2011.

Kirkby, M.: Tests of the random network model, and its application to basin hydrology, Earth Surf. Process., 1(August 1975), 197–212, doi:10.1002/esp.3290010302, 1976.

Klaus, J. and Zehe, E.: Modelling rapid flow response of a tile-drained field site using a 2D physically based model: assessment of “equifinal” model setups, Hydrol. Process., 24(12), 1595–1609, doi:10.1002/hyp.7687, 2010.

Klaus, J. and Zehe, E.: A novel explicit approach to model bromide and pesticide transport in connected soil structures, Hydrol. Earth Syst. Sci., 15(7), 2127–2144, doi:10.5194/hess-15-2127-2011, 2011.

Klemeš, V.: Operational testing of hydrological simulation models, Hydrol. Sci. J., 31(1), 13–24, doi:10.1080/02626668609491024, 1986.

Lee, H., Zehe, E. and Sivapalan, M.: Predictions of rainfall-runoff response and soil moisture dynamics in a microscale catchment using the CREW model, Hydrol. Earth Syst. Sci. Discuss., 3(4), 1667–1743, doi:10.5194/hessd-3-1667-2006, 2006.

Lehmann, P., Hinz, C., McGrath, G., Tromp-van Meerveld, H.-J. and McDonnell, J. J.: Rainfall threshold for hillslope outflow: an emergent property of flow pathway connectivity, Hydrol. Earth Syst. Sci. Discuss., 3(5), 2923–2961, doi:10.5194/hessd-3-2923-2006, 2006.

Lindsay J.B.: The Whitebox Geospatial Analysis Tools project and open-access GIS, Proc. GIS Res. UK 22nd Annu. Conf., (001), 8, doi:10.13140/RG.2.1.1010.8962, 2014.

Loague, K. and VanderKwaak, J. E.: Physics-based hydrologic response simulation: Platinum bridge, 1958 Edsel, or useful tool, Hydrol. Process., 18(15), 2949–2956, doi:10.1002/hyp.5737, 2004.

Loke, M.: Rapid 2D Resistivity & IP Inversion using the least-squares method, Geotomo Software, Man., 2003.

Martínez-Carreras, N., Krein, A., Gallart, F., Iffly, J.-F., Hissler, C., Pfister, L., Hoffmann, L. and Owens, P. N.: The Influence of Sediment Sources and Hydrologic Events on the Nutrient and Metal Content of Fine-Grained Sediments (Attert River Basin, Luxembourg), Water, Air, Soil Pollut., 223(9), 5685–5705, doi:10.1007/s11270-012-1307-1, 2012.

Martínez-Carreras, N., Wetzel, C. E., Frentress, J., Ector, L., McDonnell, J. J., Hoffmann, L. and Pfister, L.: Hydrological connectivity inferred from diatom transport through the riparian-stream system, Hydrol. Earth Syst. Sci., 19(7), 3133–3151, doi:10.5194/hess-19-3133-2015, 2015.

Maurer, T.: Physikalisch begründete zeitkontinuierliche Modellierung des Wassertransports in kleinen ländlichen Einzugsgebieten., Karlsruher Institut für Technologie., 1997.

Mendoza, P. A., Clark, M. P., Barlage, M., Rajagopalan, B., Samaniego, L., Abramowitz, G. and Gupta, H.: Are we unnecessarily constraining the agility of complex process-based models?, Water Resour. Res., 51(1), 716–728, doi:10.1002/2014WR015820, 2015.

Menzel, A., Jakobi, G., Ahas, R., Scheifinger, H. and Estrella, N.: Variations of the climatological growing season (1951-2000) in Germany compared with other countries, Int. J. Climatol., 23(7), 793–812, doi:10.1002/joc.915, 2003.

Mueller, E. N., Güntner, A., Francke, T. and Mamede, G.: Modelling sediment export, retention and reservoir sedimentation in drylands with the WASA-SED model, Geosci. Model Dev., 3(1), 275–291, doi:10.5194/gmd-3-275-2010, 2010.

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M. and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, J. Geophys. Res., 116(D12), D12109, doi:10.1029/2010JD015139, 2011.

Pérez, A. J., Abrahão, R., Causapé, J., Cirpka, O. A. and Bürger, C. M.: Simulating the transition of a semi-arid rainfed catchment towards irrigation agriculture, J. Hydrol., 409(3-4), 663–681, doi:10.1016/j.jhydrol.2011.08.061, 2011.

Peters, A. and Durner, W.: Simplified evaporation method for determining soil hydraulic properties, J. Hydrol., 356(1-2), 147–162, doi:10.1016/j.jhydrol.2008.04.016, 2008.

Pfister, L., Humbert, J. and Hoffmann, L.: Recent trends in rainfall-runoff characteristics in the Alzette River basin, Luxembourg, Clim. Change, 45(1996), 323–337, doi:10.1023/A:1005567808533, 2000.

Pfister, L., Iffly, J.-F., Hoffmann, L. and Humbert, J.: Use of regionalized stormflow coefficients with a view to hydroclimatological hazard mapping, Hydrol. Sci. J., 47(3), 479–491, doi:10.1080/02626660209492948, 2002.

Refsgaard, J. and Storm, B.: MIKE SHE in Computer models of watershed hydrology, edited by V. P. Singh, Water Resources Publications., 1995.

Refsgaard, J. C., Thorsen, M., Jensen, J. B., Kleeschulte, S. and Hansen, S.: Large scale modelling of groundwater contamination from nitrate leaching, J. Hydrol., 221(3-4), 117–140, doi:10.1016/S0022-1694(99)00081-5, 1999.

Reggiani, P. and Rientjes, T. H. M.: Flux parameterization in the representative elementary watershed approach: Application to a natural basin, Water Resour. Res., 41(4), n/a–n/a, doi:10.1029/2004WR003693, 2005.

Robinson, J. S., Sivapalan, M. and Snell, J. D.: On the relative roles of hillslope processes, channel routing, and network geomorphology in the hydrologic response of natural catchments, Water Resour. Res., 31(12), 3089–3101, doi:10.1029/95WR01948, 1995.

Savenije, H. H. G. and Hrachowitz, M.: Opinion paper: How to make our models more physically-based, Hydrol. Earth Syst. Sci. Discuss., 0(August), 1–23, doi:10.5194/hess-2016-433, 2016.

Schaefli, B. and Gupta, H. V: Do Nash values have value?, Hydrol. Process., 21(15), 2075–2080, doi:10.1002/hyp.6825, 2007.

Schierholz, I., Schäfer, D. and Kolle, O.: The Weiherbach data set: An experimental data set for pesticide model testing on the field scale, Agric. Water Manag., 44(1-3), 43–61, doi:10.1016/S0378-3774(99)00083-9, 2000.

Schoups, G., Van De Giesen, N. C. and Savenije, H. H. G.: Model complexity control for hydrologic prediction, in Water Resources Research., 2008.

Scudeler, C., Pangle, L., Pasetto, D., Niu, G.-Y., Volkmann, T., Paniconi, C., Putti, M. and Troch, P.: Multiresponse modeling of an unsaturated zone isotope tracer experiment at the Landscape Evolution Observatory, Hydrol. Earth Syst. Sci. Discuss., (May), 1–29, doi:10.5194/hess-2016-228, 2016.

Seibert, J. and McDonnell, J. J.: On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, Water Resour. Res., 38(11), 23–1–23–14, doi:10.1029/2001WR000978, 2002.

Seibert, S. P., Ehret, U. and Zehe, E.: Disentangling timing and amplitude errors in streamflow simulations, Hydrol. Earth Syst. Sci. Discuss., (March), 1–37, doi:10.5194/hess-2016-145, 2016.

Simunek, J., Genuchten, M. Van and Sejna, M.: The HYDRUS software package for simulating the two-and three-dimensional movement of water, heat, and multiple solutes in variably-saturated media, Tech. Man., 2006.

Troch, P., Paniconi, C. and Loon, E. Van: Hillslope-storage Boussinesq model for subsurface flow and variable source areas along complex hillslopes: 1. Formulation and characteristic response, Water Resour. Res., 39(11), 1316, 2003.

Tromp-Van Meerveld, H. J. and McDonnell, J. J.: Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis, Water Resour. Res., 42(August 2005), 1–11, doi:10.1029/2004WR003800, 2006.

Wienhöfer, J. and Zehe, E.: Predicting subsurface stormflow response of a forested hillslope – the role of connected flow paths, Hydrol. Earth Syst. Sci., 18(1), 121–138, doi:10.5194/hess-18-121-2014, 2014.

Wrede, S., Fenicia, F., Martínez-Carreras, N., Juilleret, J., Hissler, C., Krein, A., Savenije, H. H. G., Uhlenbrook, S., Kavetski, D. and Pfister, L.: Towards more systematic perceptual model development: a case study using 3 Luxembourgish catchments, Hydrol. Process., 29(12), 2731–2750, doi:10.1002/hyp.10393, 2015.

Zehe, E. and Sivapalan, M.: Editorial Towards a new generation of hydrological process models for the meso-scale: an introduction, 2007.

Zehe, E., Maurer, T., Ihringer, J. and Plate, E.: Modeling water flow and mass transport in a loess catchment, Phys. Chem. Earth, Part B Hydrol. Ocean. Atmos., 26(7-8), 487–507, doi:10.1016/S1464-1909(01)00041-7, 2001.

Zehe, E., Becker, R., Bárdossy, A. and Plate, E.: Uncertainty of simulated catchment runoff response in the presence of threshold processes: Role of initial soil moisture and precipitation, J. Hydrol., 315(1-4), 183–202, doi:10.1016/j.jhydrol.2005.03.038, 2005.

Zehe, E., Lee, H. and Sivapalan, M.: Dynamical process upscaling for deriving catchment scale state variables and constitutive relations for meso-scale process models, Hydrol. Earth Syst. Sci., 10(6), 981–996, doi:10.5194/hess-10-981-2006, 2006.

Zehe, E., Graeff, T., Morgner, M., Bauer, A. and Bronstert, A.: Plot and field scale soil moisture dynamics and subsurface wetness control on runoff generation in a headwater in the Ore Mountains, Hydrol. Earth Syst. Sci., 14(6), 873–889, doi:10.5194/hess-14-873-2010, 2010.

Zehe, E., Ehret, U., Blume, T., Kleidon, a., Scherer, U. and Westhoff, M.: A thermodynamic approach to link self-organization, preferential flow and rainfall–runoff behaviour, Hydrol. Earth Syst. Sci., 17(11), 4297–4322, doi:10.5194/hess-17-4297-2013, 2013.

Zehe, E., Ehret, U., Pfister, L., Blume, T., Schröder, B., Westhoff, M., Jackisch, C., Schymanski, S. J., Weiler, M., Schulz, K., Allroggen, N., Tronicke, J., Dietrich, P., Scherer, U., Eccard, J., Wulfmeyer, V. and Kleidon, A.: HESS Opinions: Functional units: a novel framework to explore the link between spatial organization and hydrological functioning of intermediate scale catchments, Hydrol. Earth Syst. Sci. Discuss., 11(3), 3249–3313, doi:10.5194/hessd-11-3249-2014, 2014.

Zhang, G. P. and Savenije, H. H. G.: Rainfall-runoff modelling in a catchment with a complex groundwater flow system: application of the Representative Elementary Watershed (REW) approach, Hydrol. Earth Syst. Sci. Discuss., 2(3), 639–690, doi:10.5194/hessd-2-639-2005, 2005.

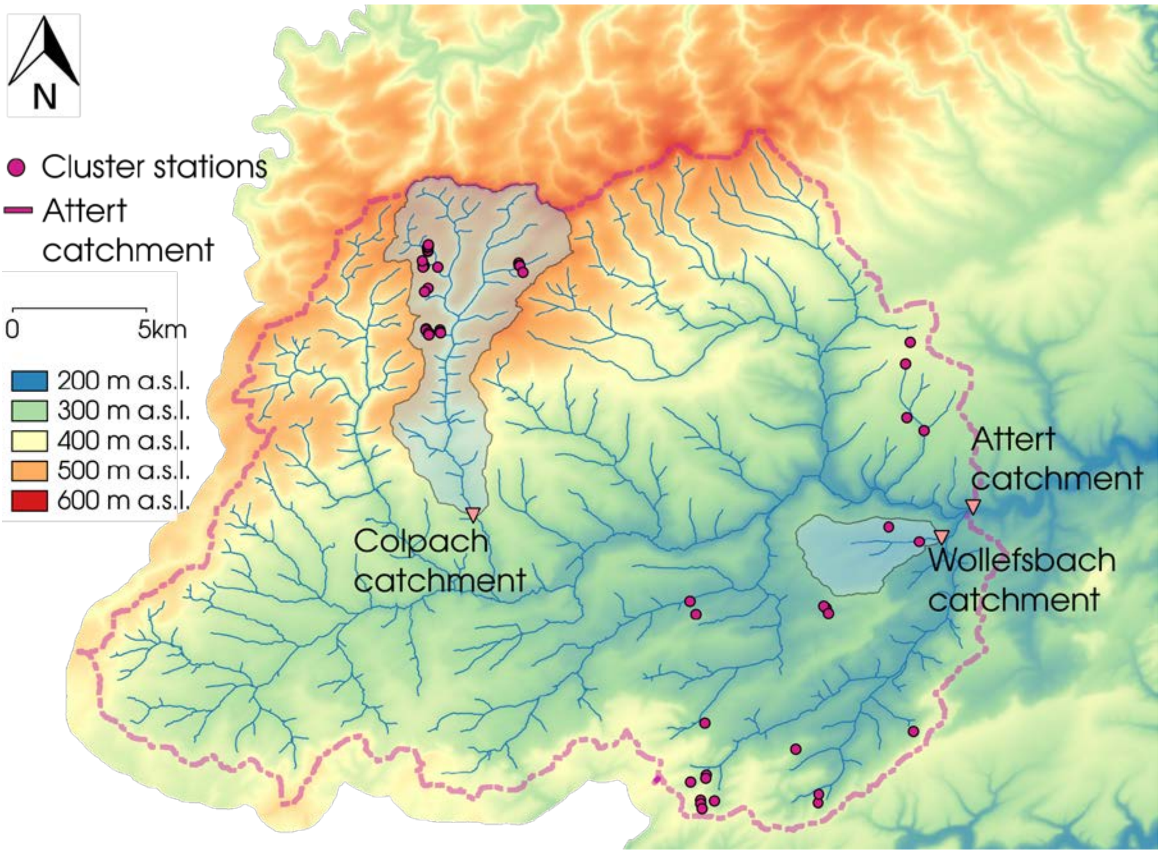


Figure 1 Map of the Attert basin with the two selected headwater catchments of this study (Colpach and Wollefsbach). In addition, the cluster sites of the CAOS research unit are displayed.

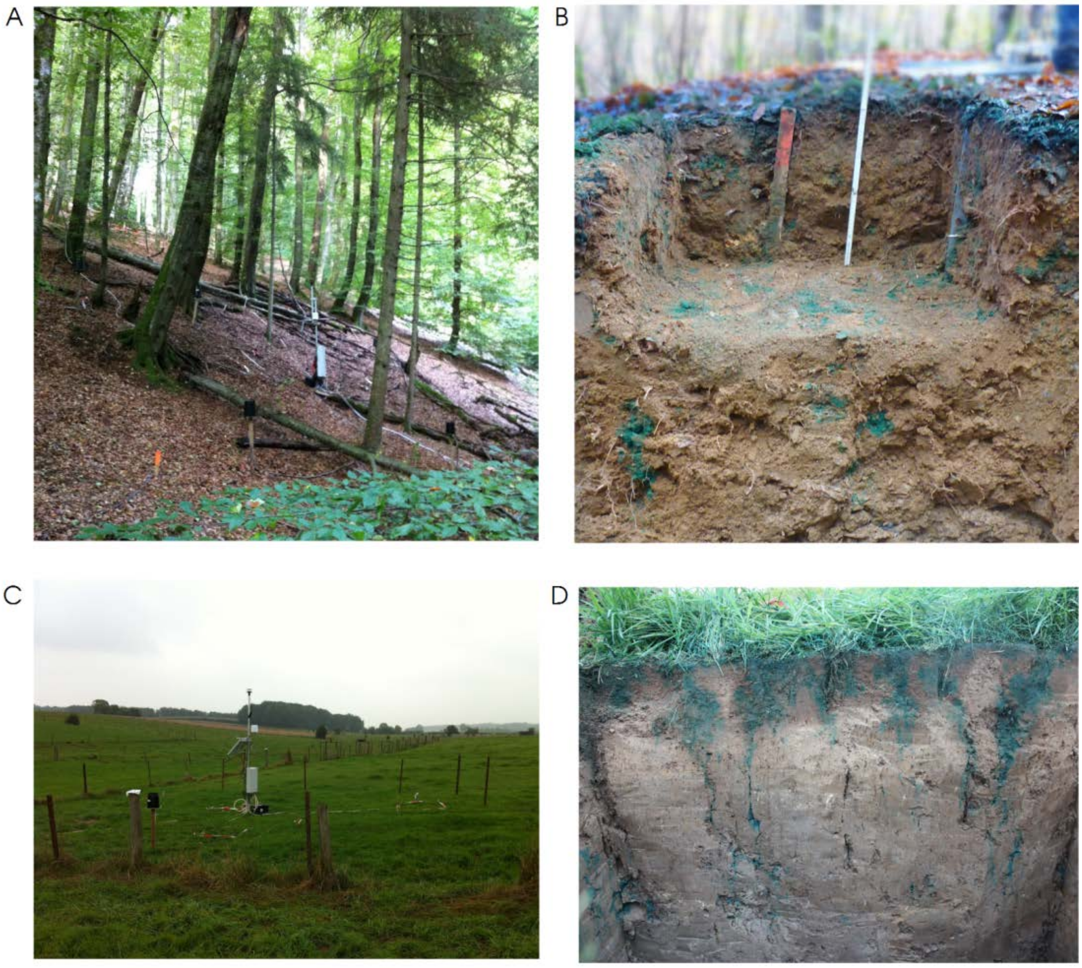


Figure 2 (A) Typical steep forested hillslope in the Colpach catchment; (B) Soil profile in the Colpach catchment after a brilliant blue sprinkling experiment was conducted. The punctual appearance of blue color illustrates the influence of vertical structures on soil water movement in this schist area. (C) Plain pasture site of the Wollefsbach catchment; (D) Soil profile in the Wollefsbach catchment after a brilliant blue experiment showing the influence of soil cracks and vertical structures on the soil water movement.

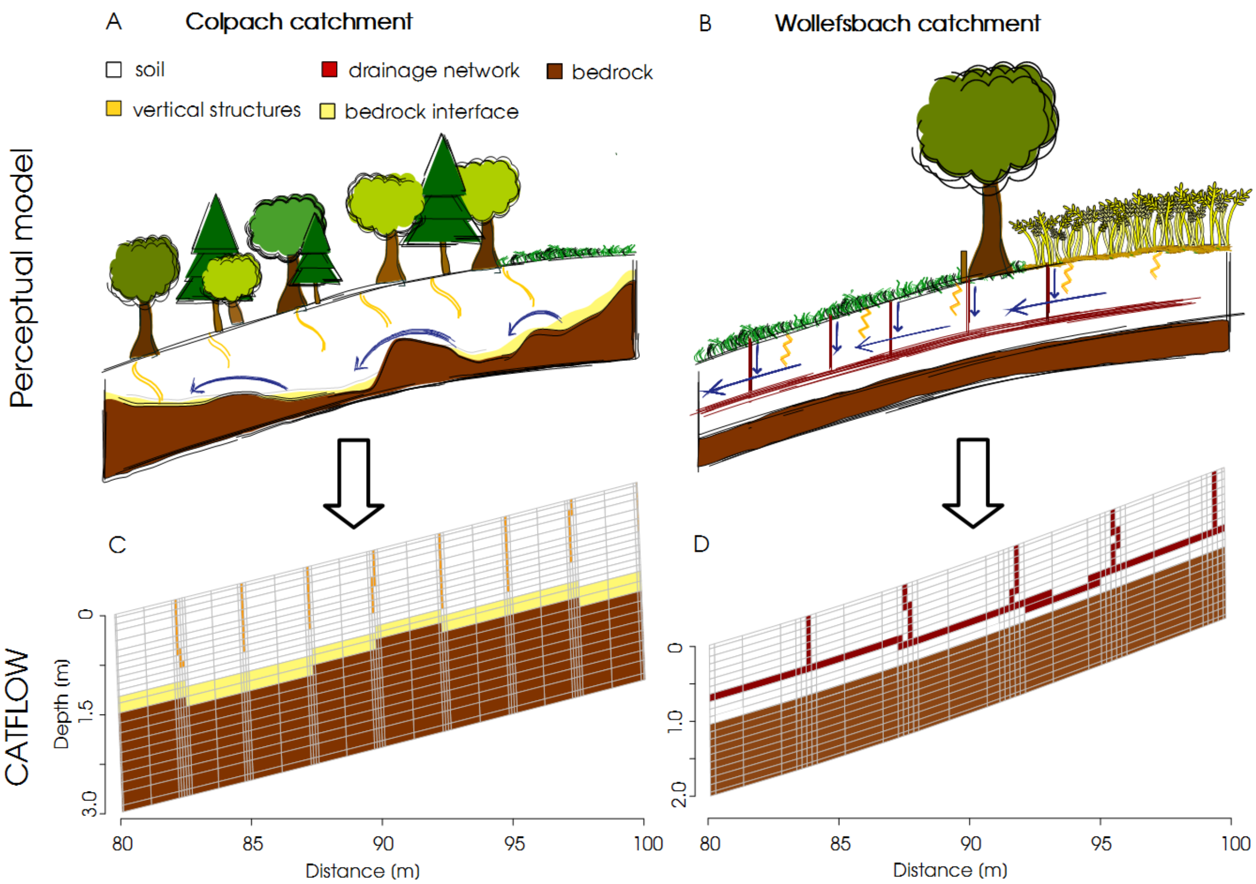


Figure 3 Perceptual models of the (A) Colpach and (B) Wollefsbach and their translation into a representative hillslope model for CATFLOW. It is important to note that only small sections of the model hillslope are displayed (C Colpach; D Wollefsbach) and not the entire hillslope.

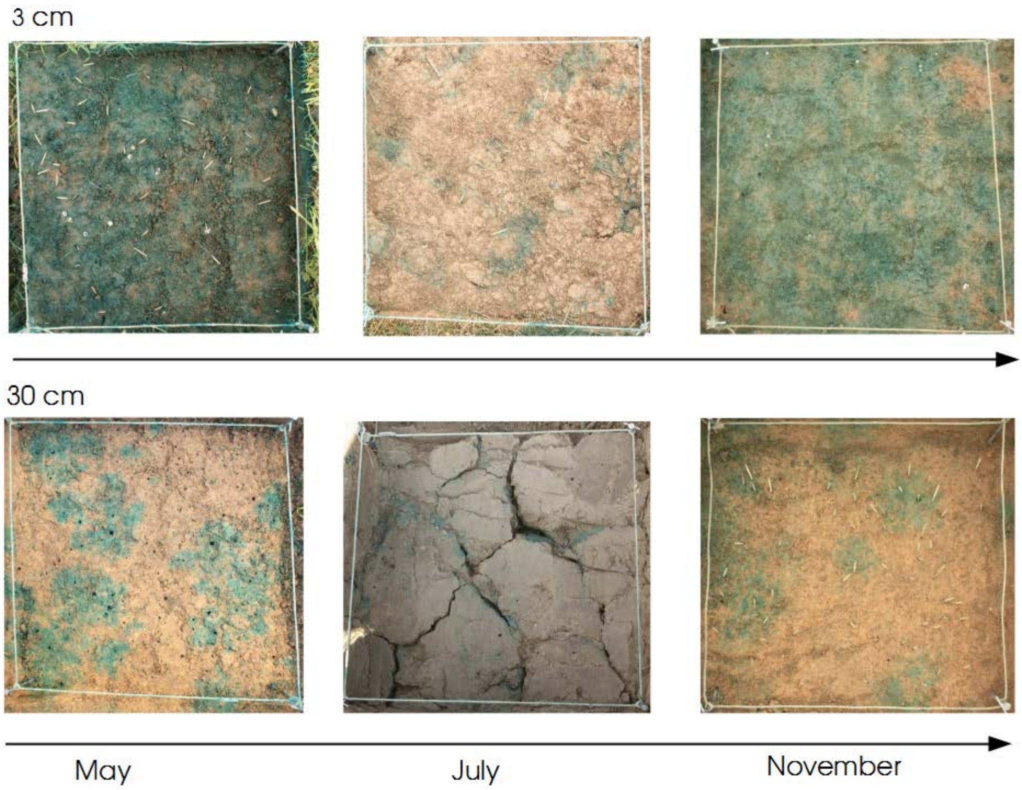


Figure 4 Emergent structures in the Wollefsbach catchment for the sampling dates. In May macropore flow through earth worm burrows dominates infiltration, while in July clearly visible soil cracks occur. In contrast, a more homogenous infiltration pattern is visible in November, especially at 3 cm depth.

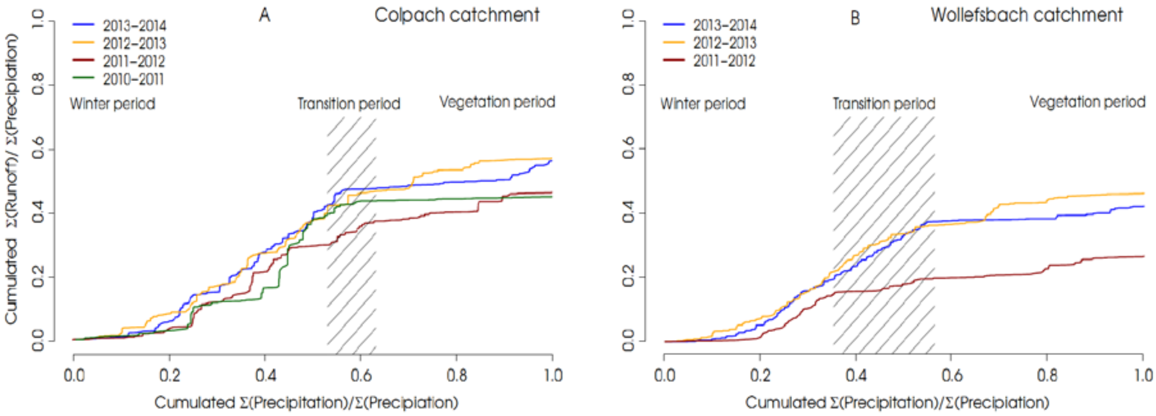


Figure 5 Normalized double mass curves for each hydrological year from 2010 to 2014 in the Colpach catchment (A) and from 2011 to 2014 in the Wollefsbach catchment (B). The transition period marks the time of the years when the catchment shifts from the winter period to the vegetation period. The separation of the seasons is based on a temperature index model from Menzel et al., (2003). Since the season shift varies between the hydrological years the transition period is displayed as an area.

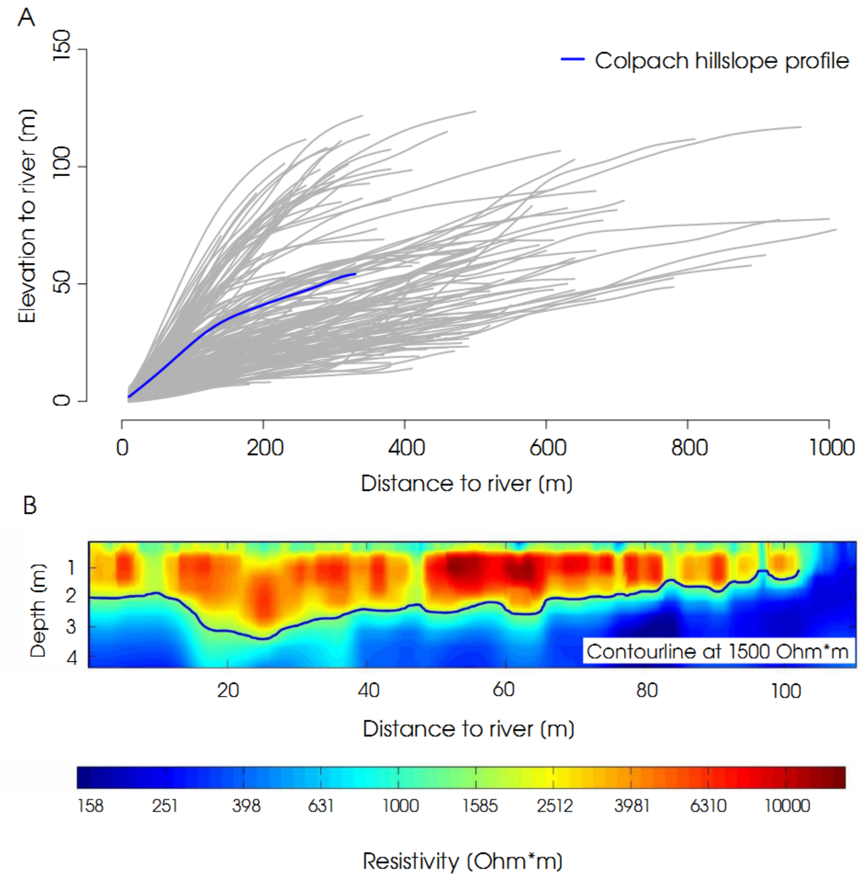


Figure 6 (A) Profile of all hillslope extracted from a DEM in the Colpach catchment. Hillslope profile we used in this study highlighted in blue. (B) Bedrock topography of a hillslope in the Schist area measured using ERT. The contour line displays the 1500 Ωm isoline which is interpreted as soil bedrock interface.

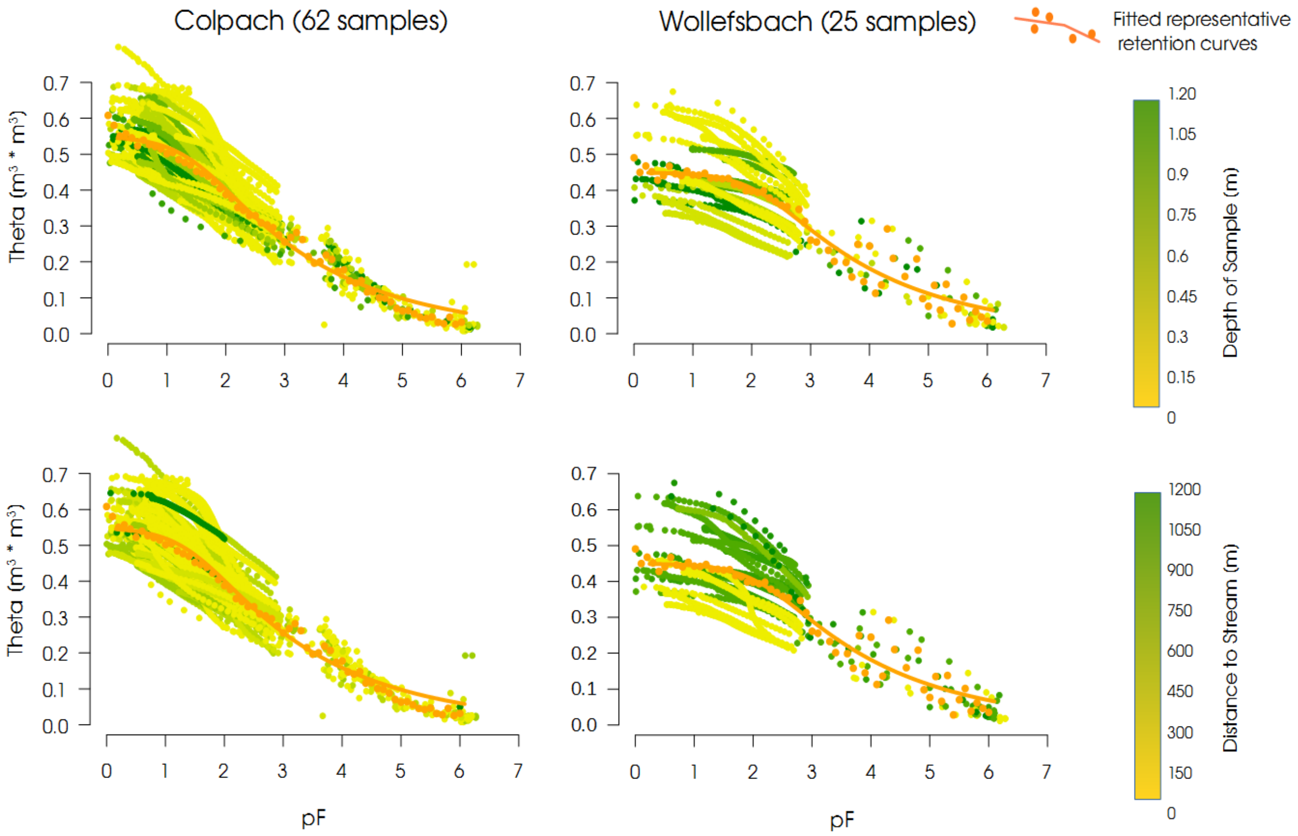


Figure 7 Fitted soil water retention curves and measured soil water retention relationships for the Colpach (A) and Wollefsbach (B) catchment.

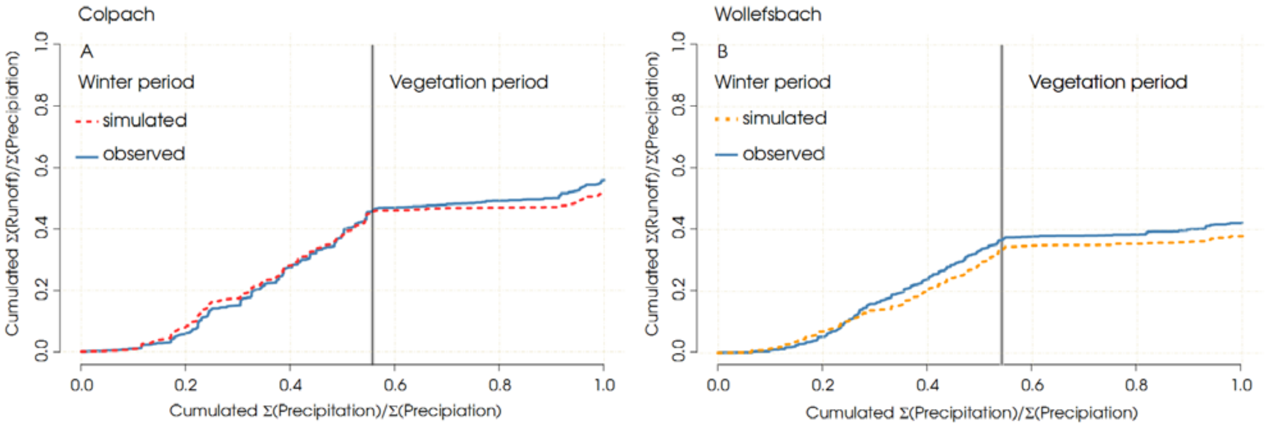
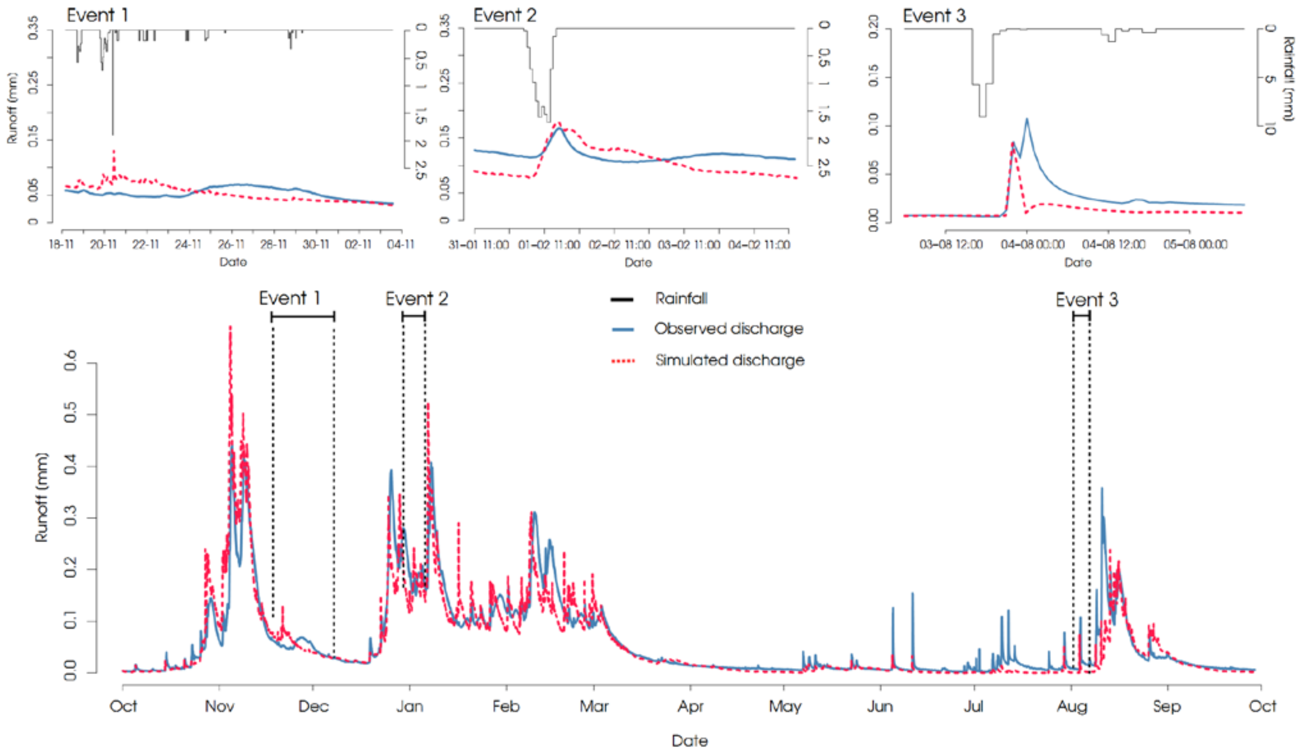


Figure 8 Simulated and observed normalized double mass curves of (A) the Colpach and (B) the Wollefsbach catchment. The double mass curves are separated into a winter and a vegetation period after Menzel et al. (2003).

1182



1183

1184 **Figure 9 Observed and simulated runoff of the Colpach catchment. Moreover, three rainfall runoff events are**
1185 **highlighted and displayed separately.**

1186

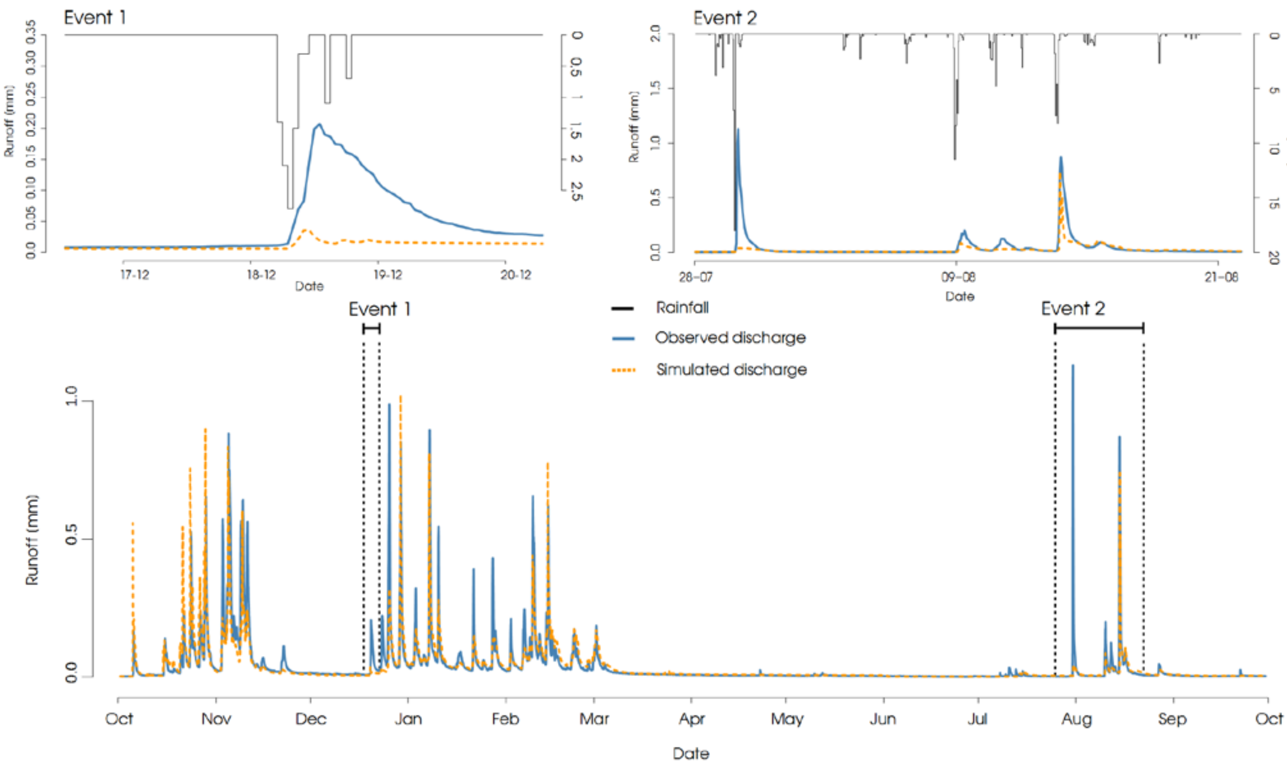


Figure 10 Observed and simulated runoff of the Wollefsbach catchment. Two rainfall runoff events are highlighted and displayed separately.

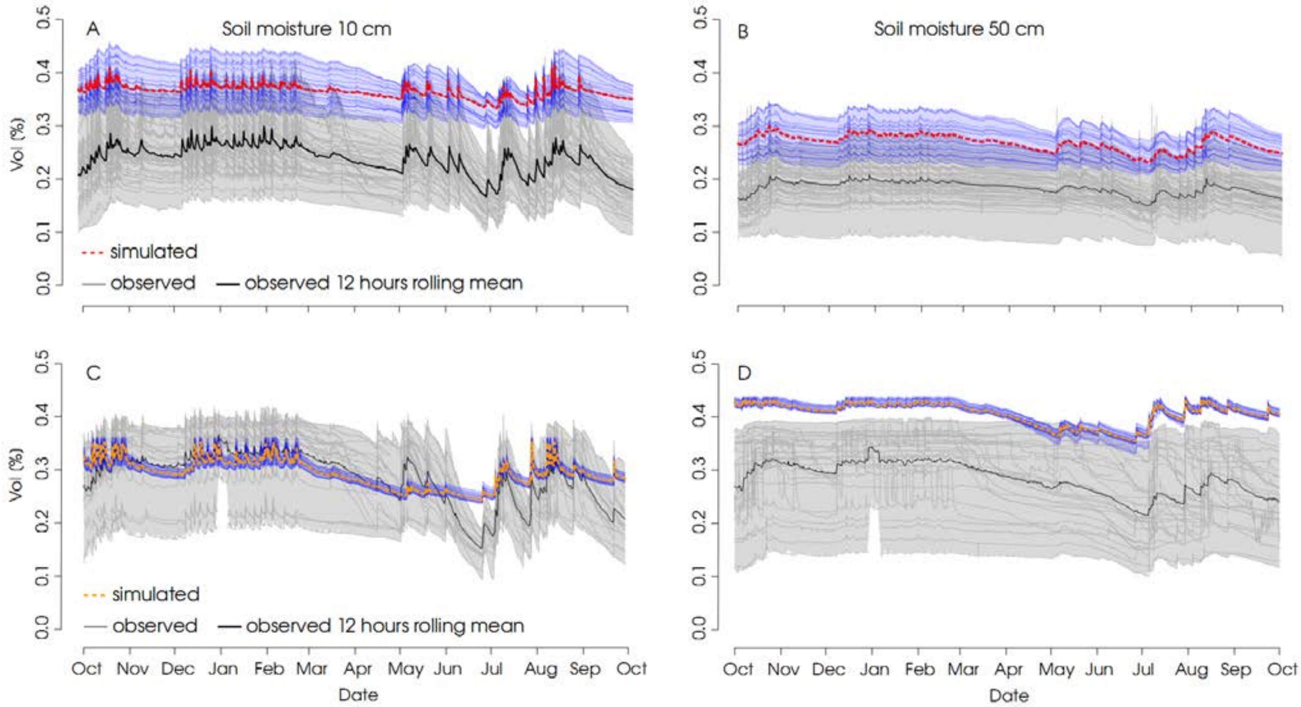


Figure 11 Observed soil moisture at 10 and 50 cm depths in the schist (A and B) and marl (C and D) area of the Attert catchment. Additionally the 12 hours rolling median (black) derived from the soil moisture observations and the simulated soil moisture dynamics at the respective depths (red Colpach; orange Wollefsbach) are displayed.

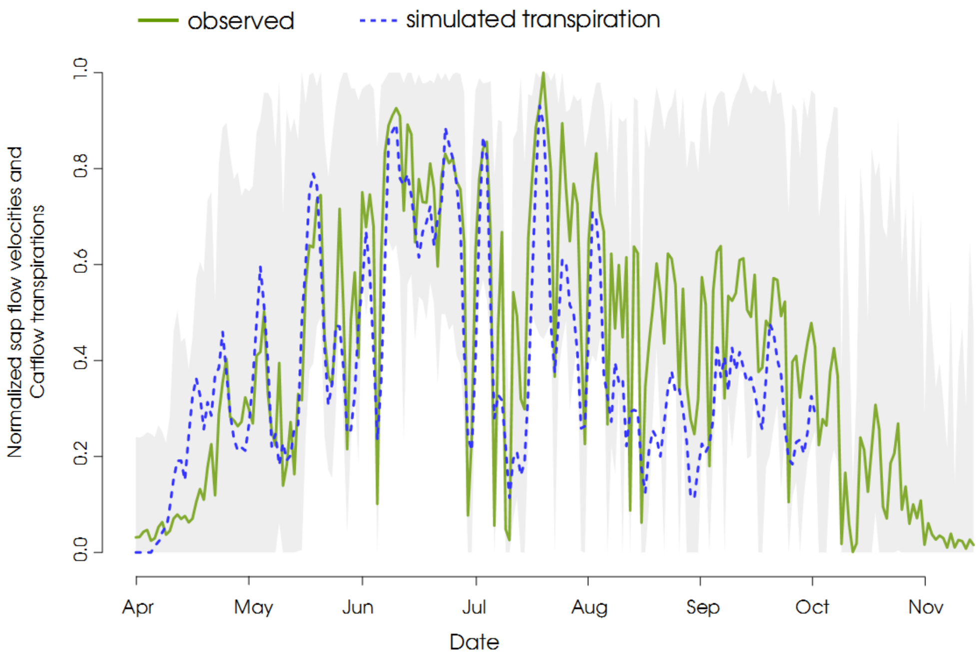


Figure 12 Normalized observed average sap velocities of 28 trees in the Colpach catchment (green) and normalized simulated transpiration from the Colpach model smoothed with a three-day rolling mean (dashed blue). Additionally the ensemble of all 28 sap flow measurements is displayed in grey.

Table 1 Hydraulic and transport parameter values used for different materials in the model setups.

Type of structure	Saturated hydraulic conductivity K_s (m s^{-1})	Total porosity Θ_s (–)	Residual water content Θ_r (–)	Alpha value α (m^{-1})	Shape parameter n (–)
Colpach					
Soil layer	5×10^{-4}	0.57	0.05	4.93	1.05
Macropores & soil bedrock interface	1×10^{-3}	0.25	0.1	7.5	1.5
Bedrock	1×10^{-9}	0.2	0.05	0.5	2
Wollefsbach					
Soil layer	2.92×10^{-4}	0.46	0.05	0.66	1.05
Drainage system	1×10^{-3}	0.25	0.1	7.5	1.5
Bedrock	1×10^{-9}	0.2	0.05	0.5	2

Table 2 Vegetation parameter values for the different land use forms in the model setup.

	Start / End of the Vegetation period [doy]	LAI [-]	Root depth [m]	Through fall rate [%]	Plant height [m]	Interception [mm]	Maximum stomata conductance [mm s−1]	Albedo [-]
Colpach:								
Forest (*Fagus sylvatica*)	97 / 307	6.3[4]	1.8	95	24 [4]	2	5	0.2
Wollefsbach								
Corn (*Zea mays*)	97 / 307	4[2]	1.2[1]	100	2	3	2.5	0.2
Pasture	97 / 274	6[2]	1.3[3]	100	0.4	3.1[3]	2.5	0.2

[1] value for gley brown soils; [2] mean value (Breuer et al., 2003); [3] Trifolium spec., [4] observed

Table 3 Benchmarks for simulated double mass curves and simulated discharge for all model setups used in this study.

Model setup	*Double mass curve:*	*Discharge*:		
	KGE	KGE	NSE	logNSE
Colpach models				
Reference Colpach model:	0.92	0.88	0.79	0.25
Performance winter :	0.95	0.88	0.75	0.93
Performance summer:	0.49	0.52	0.51	0.62
Wollefsbach models				0.87
Reference Wollefsbach model:	0.9	0.71	0.68	
				0.84
Peformance winter:	0.85	0.74	0.7	0.57
Performance summer:	0.74	0.28	0.33	

Appendix

A1 Subsurface structure and bedrock topography

Spatial subsurface information of representative hillslopes were obtained from 2-D ERT sections collected using a GeoTom (GeoLog) device at seven profiles on two hillslopes in the Colpach catchment. We used a Wenner configuration with electrode spacing of 0.5 m and 25 depth levels: electrode positions were recorded at a sub-centimeter accuracy using a total station providing 3D position information. Application of a robust inversion scheme as implemented in Res2Dinv (Loke, 2003) resulted in the two-layered subsurface resistivity model shown in Figure 6 B. The upper 1-3 m are characterized by high resistivity values larger than 1500 Ω*m. This is underlain by a layer of generally lower resistivity values smaller than 1500 Ω*m. In line with the study of Wrede et al. (2015) and in correspondence with the maximum depth of the local auger profiles, we interpreted the transition from high to low resistivity values to reflect the transition zone between bedrock and unconsolidated soil. In consequence, we regard the 1500 Ωm isoline as being representative for the soil-bedrock interface. For our modeling study we have access to seven ERT profiles within the Colpach area (example see Figure 6 B).

A2 Soil hydraulic properties, infiltrability and dye staining experiments

Saturated hydraulic conductivity was measured with undisturbed 250 ml ring samples with the KSAT apparatus (UMS GmbH). The apparatus records the falling head of the water supply though a highly sensitive pressure transducer which is used to calculate the flux. The soil water retention curve of the drying branch was measured with the same samples in the HYPROP apparatus (UMS GmbH) and subsequently in the WP4C dew point hygrometer (Decagon Devices Inc.). The HYPROP records total mass and matric head in two depths in the sample over some days when it was exposed to free evaporation (Peters and Durner, 2008, Jackisch 2015 for further details). For both geological settings we estimated a mean soil retention curve by grouping the observation points of all soil samples (62 and 25 for schist and marl, respectively), and averaging them in steps of 0.05 pF. We then fitted a van Genuchten-Mualem model using a maximum likelihood method to these averaged values (Table 1 and Figure 7). We used a representative soil water retention curve because the young soils on periglacial slope deposits prevail in the both headwaters exhibit large heterogeneity which cannot be grouped in a simple manner. This is due to a) the general mismatch of the scale of 250 mL undisturbed core samples with the relevant flow paths and b) the high content of gravel and voids, which affect the retention curve especially above field capacity and concerning its scaling with available pore space (Jackisch 2015, Jackisch et al. 2016). The dye tracer images, Figure 2 B and D, were obtained with high rainfall intensities of 50 mm in 1 h on 1 m^2 and the sprinkling water was enriched with 4.0 g l^{-1} Brilliant Blue dye tracer (Jackisch et al. 2016). The aim of these rainfall simulations was to visualize the macropore networks in the topsoil, to gather information on the potential preferential flow paths relevant for infiltration.

A3 Physically-based model CATFLOW

The model CATFLOW has been successfully used and specified in numerous studies (e.g. Zehe et al., 2005; Zehe et al. 2010; Wienhöfer and Zehe, 2014; Zehe et al., 2014). The basic modeling unit is a two-dimensional hillslope. The hillslope profile is discretized by curvilinear orthogonal coordinates in vertical and downslope directions; the third dimension is represented via a variable width of the slope perpendicular to the slope line at each node. Soil water dynamics are simulated based on the Richards equation in the pressure based form and numerically solved using an implicit mass conservative "Picard iteration" (Celia et al., 1990). The model can simulate unsaturated and saturated subsurface flow and hence has no separate groundwater routine. Soil hydraulic functions after van Genuchten-Mualem are commonly used, though several other parameterizations are possible. Overland flow is simulated using the diffusion wave approximation of the Saint-Venant equation and explicit upstreaming. The hillslope module can simulate infiltration excess runoff, saturation excess runoff, re-infiltration of surface runoff, lateral water flow in the subsurface as well as return flow. For catchment modeling several hillslopes can be interconnected by a river network for collecting and routing their runoff contributions, i.e. surface runoff or subsurface flow leaving the hillslope, to the catchment outlet. CATFLOW has no routine to simulate snow or frozen soil.

A3.1 Evaporation controls, root water uptake and vegetation phenology

Soil evaporation, plant transpiration and evaporation from the interception store is simulated based on the Penman–Monteith equation. Soil moisture dependence of the soil albedo is also accounted for as specified in Zehe et al. (2001). Annual cycles of plant phenological parameters, plant albedo and plant roughness are accounted for in the form of tabulated data (Zehe et al., 2001). Optionally, the impact of local topography on wind speed and on radiation may be considered, if respective data are available. The atmospheric resistance is equal to wind speed in the boundary layer over the squared friction velocity [mm d^{-1}]. The former depends on observed wind speed, plant roughness and thus plant height. The friction velocity depends on observed wind speed as well as atmospheric stability, which is represented through six stability classes depending on prevailing global radiation, air temperature and humidity. The canopy resistance is the product of leaf area index and leaf resistance, which in turn depends on stomata and cuticular resistance. The stomata resistance varies around a minimum value, which depends on the Julian day as well as on air temperature, water availability in the root zone, the water vapor saturation deficit and photosynthetic active radiation (Jarvis, (1976). The resulting root water uptake is accounted for as a sink in the Richards equations term, and is specified as a flux per volume, which is extracted uniformly along the entire root depth.

A4 Soil moisture observations

Figure A1 shows the soil moisture observations of the Colpach catchment group by their position at the hillslope. This figure highlight, similar to Figure 7 for the soil water retention properties, that the small-scale variability of the prevailing soils make a simple grouping by the landscape position difficult.

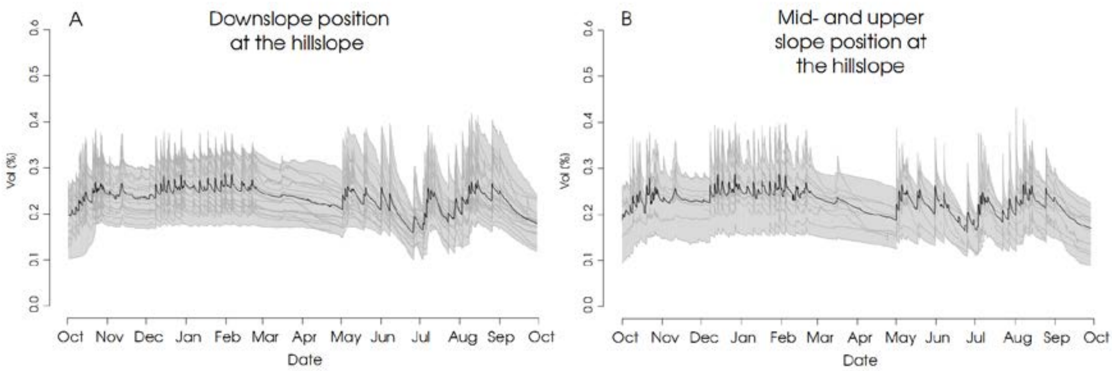


Figure A1 Soil moisture observations grouped by their landscape position. (A) Soil moisture observations at the hillslope foot and hence close to the river. (B) Soil moisture observations at the upper part of the hillslope.