Water Resources Research

RESEARCH ARTICLE
10.1029/2018WR023886

From Points to Patterns: Using Groundwater Time Series Clustering to Investigate Subsurface Hydrological Connectivity and Runoff Source Area Dynamics

M. Rinderer[1,2], **H. J. van Meerveld**[3], **and B. L. McGlynn**[1]

[1]Nicholas School of the Environment, Duke University, Durham, NC, USA, [2]Chair of Hydrology, University of Freiburg, Freiburg, Germany, [3]Department of Geography, University of Zurich, Zurich, Switzerland

Key Points:
- We present a data-driven approach composed of time series clustering and topography-based upscaling of groundwater data to the catchment
- The groundwater activation maps highlight the dynamic expansion and contraction of subsurface runoff source areas along the channel network
- Results agree with the variable source area concept, but isolated groundwater response areas were not hydrologically connected to the stream

Correspondence to:
M. Rinderer,
michael.rinderer@hydrology.uni-freiburg.de

Citation:
Rinderer, M., van Meerveld, H. J., & McGlynn, B. L. (2019). From points to patterns: Using groundwater time series clustering to investigate subsurface hydrological connectivity and runoff source area dynamics. *Water Resources Research*, *55*, 5784–5806. https://doi.org/10.1029/2018WR023886

Received 11 AUG 2018
Accepted 17 JUN 2019
Accepted article online 26 JUN 2019
Published online 18 JUL 2019

Abstract Groundwater levels are typically measured at only a limited number of points in a catchment. Thus, upscaling these point measurements to the catchment scale is necessary to determine subsurface flow paths and runoff source areas. Here we present a data-driven approach composed of time series clustering and topography-based upscaling of shallow, perched groundwater dynamics using groundwater data from 51 monitoring sites in a 20-ha prealpine headwater catchment in Switzerland. The agreement between the upscaled (modeled) and measured groundwater dynamics was strong for most of the 19-month study period for the upslope and footslope locations but weaker at the beginning of events and for the midslope locations. However, these differences between measured and modeled groundwater levels did not significantly affect modeled groundwater activation, that is, the time when groundwater levels were within the more transmissive soil layers near the soil surface. The resulting groundwater activation maps represent the groundwater response across the catchment and highlight the dynamic expansion and contraction of the subsurface runoff source areas, particularly along the channel network. This is in agreement with the variable source area concept. However, there were also isolated active zones that did not get connected to the stream during rainfall events, highlighting the need to distinguish between variable active and variable stream-connected runoff source areas. Our data-driven approach to upscale point measurements of shallow groundwater levels appears useful for studying catchment-scale variations in groundwater storage and connectivity and thus may help to better understand runoff generation in mountain catchments.

Plain Language Summary For a better understanding of how runoff in streams is generated, we need to know how groundwater levels respond across a catchment. However, groundwater can usually only be measured at a few selected points, and interpolation between these points does often not result in realistic groundwater response patterns. Here we present a data-driven approach based on groundwater level data from 51 sites in a catchment in Switzerland for a 19-month study period. We grouped the monitoring sites into six clusters with similar groundwater level dynamics. We then determined the topographic characteristics of the sites in each cluster and assigned the average relative groundwater level for the monitoring sites in a cluster to all other sites in the catchment with similar topographic characteristics. By doing so, we created sequences of maps of groundwater levels across the entire study catchments. These maps show an expansion and contraction of the areas where the groundwater level is close to the surface and which of these areas are connected to the stream channels. These maps are useful to identify from which parts of the catchment streamwater may come during a rain event, which helps to improve our understanding of runoff generation processes.

1. Introduction

Mapping runoff source areas and their connectivity to the stream network and understanding the processes and factors that affect their dynamics improve our understanding of catchment-scale streamflow generation processes (Jencso et al., 2009; McGlynn & Seibert, 2003; Nippgen et al., 2015; Tetzlaff, Soulsby, Waldron, et al., 2007). The spatial and temporal change in the extent of runoff source areas and their hydrologic connectivity to the stream are also of key importance to understand the temporal and spatial variability in nutrient transport (Creed et al., 1996; Dick et al., 2015; Thompson et al., 2012; Vidon & Hill, 2004), stream solute concentrations (Ali et al., 2010; Fischer et al., 2015; McGlynn & McDonnell, 2003), mean transit

times (McGlynn, 2004; McGuire, 2005; Tetzlaff et al., 2009; Yang et al., 2018), and aquatic habitats (Soulsby et al., 2016; Tetzlaff, Soulsby, Bacon, et al., 2007).

In most undisturbed humid catchments, saturated overland flow occurs only in a small part of the catchment, mainly near the stream channel. This area expands and contracts during events and seasonally (variable source area concept; Cappus, 1960; Tsukamoto, 1961; Hewlett & Hibbert, 1967; Kulasova et al., 2014; Rinderer et al., 2012). Subsurface stormflow is generally a more important contributor to streamflow than overland flow (Hewlett & Hibbert, 1963; Tsukamoto, 1963) and usually requires the development of a perched groundwater table above a less permeable layer (e.g., at the soil-bedrock interface) or the rise of the water level into higher permeability soil layers. Subsurface contributing areas can expand upslope (saturated wedge theory; Weyman, 1973) during rainfall events, but field studies have shown that subsurface flow may also start in shallow soil regions near the ridge and expand downward (Penna et al., 2015; Tromp-van Meerveld & McDonnell, 2006; van Meerveld et al., 2015) or that different parts of a catchment connect throughout an event (Jencso et al., 2009; Nippgen et al., 2015). Because not all areas with shallow groundwater have a continuous hydrologic connection to the stream and can directly contribute to streamflow, it is important to distinguish between hydrologically active areas and hydrologically connected areas. Active areas are defined here as areas in a catchment for which the groundwater level is within the transmissive soil layers, whereas connected areas are defined here as the spatial extent of all active areas for which there is a continuous connection via other active areas to the stream (Ambroise, 2004; Engman, 1974; Nippgen et al., 2015). It is generally assumed that there is lateral flow of water toward the stream from all the connected areas. However, it is important to note that this water not necessarily reaches the stream on the even time scale, if the distance to the stream is too long or the hydraulic conductivity of the transmissive soil layer is too low. The water can also be partly subject to infiltration into the underlying soil layer before reaching the stream (Klaus & Jackson, 2018). The slow responding regional groundwater flow system can also contribute to streamflow (Toth, 1963; Winter, 2001), and intercatchment groundwater flow may be important for some catchments (Iwagami et al., 2010; Uchida & Asano, 2010; Welch et al., 2012). However, for rainfall events in small headwater catchments, hydrological connectivity of fast subsurface stormflow is often a dominant factor influencing the streamflow response (Blumstock et al., 2016; McNamara et al., 2005; Penna et al., 2015; Seyfried et al., 2009; Stieglitz et al., 2003; Tromp-van Meerveld & McDonnell, 2006).

Surface runoff areas can be determined by mapping the extent of saturated areas seasonally or during events (Dunne & Black, 1970; Rinderer et al., 2012). While this mapping is restricted to relatively small areas, the resulting temporally discontinuous maps of contributing areas are a simple (but time consuming) approach to calibrate models or to verify remote sensing data (Ali et al., 2014; Birkel et al., 2010; Blazkova et al., 2002). Thermal infrared cameras can be used to obtain more continuous information on the changes in saturated areas over time for small areas in the riparian zone or at the foot of hillslopes (e.g., Glaser et al., 2016; Glaser et al., 2018; Pfister et al., 2010).

Subsurface contributing areas are more difficult to determine and are generally inferred from the presence or absence of a perched water level above a low conductivity layer (Ali et al., 2011; Jencso et al., 2009; Jencso & McGlynn, 2011; van Meerveld et al., 2015) or the rise of the water level into a higher conductivity layer (Bishop et al., 2011; Rinderer et al., 2014). However, groundwater level information is typically only available for a few monitoring wells in a catchment, and the spatial variability in groundwater levels is high (Bachmair & Weiler, 2012; Lyon et al., 2006). This makes interpolation of groundwater levels difficult. Basic interpolation methods such as Inverse Distance Weighting (Shepard, 1968) or Kriging (Delhomme, 1978) assume that spatial distance is a good criterion for estimating similarity in groundwater levels (Tobler, 1970). While this may hold for flatter areas or gentle slopes, several studies have shown that these methods are less appropriate for interpolation of the depth to groundwater in mountainous environments where groundwater levels can vary significantly over a distance of just a few meters (Bachmair & Weiler, 2012; Tromp-van Meerveld & McDonnell, 2006) and landscape position and local morphology are more important predictors of groundwater dynamics (McDonnell, 1990; Mosley, 1979; Rinderer et al., 2014; Sklash et al., 1986). In such terrain, interpolation methods that incorporate auxiliary information, such as topography, landscape-category, or vegetation patterns (e.g., indicator kriging or kriging with external drift), may work better (Desbarats et al., 2002; Lyon et al., 2006; Peeters et al., 2010). Geostatistical methods have also been used to study soil moisture patterns (Grayson, 2001; Rosenbaum et al., 2012; Western et al., 1998);

the semivariogram has been shown useful to determine characteristic length scales of groundwater and soil moisture data (Western et al., 2001).

Topographic indices derived from Digital Elevation Models (DEMs) have been used to identify areas in the catchment that are likely to generate runoff (Ali et al., 2014; Ali & Roy, 2010; Jencso et al., 2009; Lessels et al., 2016; Tetzlaff, Soulsby, Waldron, et al., 2007). The Topographic Wetness Index (TWI; Beven & Kirkby, 1979), for instance, is often used in conceptual hydrological models (e.g., TOPMODEL) to simulate the distribution of the groundwater table across a catchment. A number of empirical studies (Anderson & Burt, 1978; Burt & Butcher, 1985; Detty & McGuire, 2010; Rinderer et al., 2014; Rinderer et al., 2016) have shown that it is indeed a good predictor of the median groundwater level and the groundwater response timing.

Approaches that combine groundwater level data from wells and topographic analyses can be very useful to understand catchment-scale groundwater dynamics. Empirical data from 84 recording groundwater wells in the Tenderfood Creek Experimental Forest in Montana revealed that the fraction of time that hillslopes were hydrologically connected to the stream was highly correlated with the upslope accumulated area (Jencso et al., 2009). This functional relation was then applied to the entire catchment to derive an annual landscape connectivity duration curve, which was highly correlated to the annual flow duration curve at the catchment outlet (Jencso et al., 2009). Smith et al. (2013) divided the catchment into stream reaches and determined the change in storage for each adjacent hillslope based on a simple mass balance approach. By applying a hillslope-specific storage threshold, which was a function of the upslope contributing area so that large hillslopes exceed this threshold more often, the total outflow of a hillslope became a function of the duration of hydrological connectivity. The empirical network connectivity curve (Jencso et al., 2009) was also used to calibrate a watershed-scale storage threshold that was applied to individual cells of a spatially distributed hydrological model (Nippgen et al., 2015). Cells that exceeded the threshold were considered active, and continuous chains of active cells along the surface gradient were connected to allow identification of the temporal changes in runoff source areas and connectivity patterns at the catchment scale. While these studies for the Tenderfoot Creek Experimental Forest highlight the use of point measurements of groundwater together with topographic analyses to understand hydrologic connectivity and streamflow dynamics, the groundwater measurements were only taken at the base of hillslopes (Jencso et al., 2009; Jencso et al., 2010; Jencso & McGlynn, 2011), and the derived indices of hillslope-riparian zone-stream connectivity ignored the spatial variability in connectivity within hillslopes. Therefore, it is useful to also explore other ways to upscale point measurements of groundwater levels to study hydrological connectivity across catchments.

Hydrological models can also be used to obtain information on the likely subsurface source areas. Even though there is a growing number of conceptual or physically based modeling studies that investigate the spatial and temporal source area dynamics (Hopp & McDonnell, 2009; Weiler & McDonnell, 2004; Yang et al., 2018), validation of the simulated spatial patterns in groundwater levels remains limited (Loague & Ebel, 2016). Field studies on catchment-scale patterns of runoff sources are still rare (Burt & McDonnell, 2015), and the data to test (and/or parameterize) the models are thus often not available.

Here we use shallow perched groundwater level data from 51 groundwater monitoring wells, time series clustering, and topography-based upscaling to extrapolate point-scale groundwater level measurements to the catchment scale and then use these modeled catchment-scale groundwater patterns to obtain information about the spatial and temporal dynamics of the subsurface runoff source areas. We focus on hydrological connectivity, that is, conditions by which disparate regions of a catchment (i.e., hillslopes) can provide water to the stream network via a continuous subsurface flowpath (e.g., a conductive soil layer or preferential flow path); (after Hornberger et al., 1994; Creed & Band, 1998). We acknowledge that hydraulic connectivity is another important mechanism that contributes water from near-stream areas (often the riparian zone) to the stream via a pressure response of hydraulically connected areas of the catchment, but this is not subject of this study. More specifically, we address the following questions:

1. How robust are predictions of groundwater levels for an entire catchment based on time series clustering and topographic characteristics, and is the robustness of the predictions affected by the chosen groundwater activation threshold, antecedent wetness conditions or rainfall event characteristics?
2. Which portions of the landscape are hydrologically connected to the stream network during storm events and different seasons? Where are the areas that are always/never hydrologically connected to the stream,

and to what extent are the dynamics of the active and the connected areas during rainfall events consistent with the variable source area concept?

2. Methods

2.1. Study Catchment

For the analyses, we used shallow groundwater and discharge data from the 20-ha prealpine "Studibach" (C7) headwater catchment. The catchment is located in the Alptal, 40 km southeast of Zurich, Switzerland (Figure 1), and ranges in elevation from 1,270 to 1,650 m above sea level (a.s.l.). The main gauging station is located ~1 km (380-m elevation) downstream of the water divide and ~1 km (170-m elevation) upstream of the confluence with the Zwackentobel in the main valley bottom. The catchment has steep slopes that alternate with more gentle parts and depressions that originated from landslides and soil creep. The average slope is 35%. There are almost no flat valley bottom areas with a riparian zone (<1% of the catchment area). Due to the high annual precipitation (2,300 mm/year, of which one third falls as snow) and low permeability gleysols, the water levels are close to the surface in many areas of the catchment. Soil depth varies between 0.5 and 1 m on ridge sites and between 1 and 2.5 m in depressions. The bedrock consists of Flysch, a clay-rich and poorly draining bedrock comprising of calcareous sandstone and argillite and bentonite schist layers (Mohn et al., 2000).

The mineral soil in the moor areas is classified as a mollic Gleysol, with typically a 10- to 30-cm-thick Ah horizon (43% sand, 43% silt, and 14% clay), a permanently reduced Bg horizon (27% sand, 45% silt, and 28% clay), and an underlaying Cv horizon (32% sand, 38% silt, and 30% clay); (Weiler et al., 1998). The steeper, better drained areas are covered by forest (typically: *Picea abies* L. with an understory of *Vaccinium* sp.; Hagedorn et al., 2000) and have an umbric Gleysol with an Ah horizon (21% sand, 41% silt, and 38% clay), an oxidized Bw horizon (15% sand, 45% silt, and 40% clay), and a Cv horizon (23% sand, 39% silt, and 38% clay); (Weiler et al., 1998). The vertical and horizontal saturated hydraulic conductivity of a neighboring grassland site is 2×10^{-5}, 3×10^{-6}, and 3×10^{-6} m/s and 1.7×10^{-3}, 5×10^{-5}, and 2×10^{-5} m/s for the Ah, Bg, and Cv soil horizon, respectively. The corresponding values of a forested site are 1×10^{-5}, 1×10^{-5}, and 4×10^{-7} m/s and 8×10^{-5}, 5×10^{-5}, and 4×10^{-7} m/s for the Ah, Bw, and Cv soil horizon, respectively (Weiler et al., 1998). The lateral conductivity is up to 2 orders of magnitude larger than the vertical hydraulic conductivity, promoting lateral flow on the steep hillslopes (Feyen et al., 1999; Weiler et al., 1998). Given that the saturated hydraulic conductivity decreases sharply with soil depth, fast subsurface flow is expected only to occur in the uppermost 0.3 m of the soil profile. For a detailed description of the catchment, see Rinderer et al. (2014) and van Meerveld et al. (2018).

2.2. Field Measurements

Shallow (perched) groundwater levels were measured in 51 shallow groundwater wells located across the study catchment. The groundwater monitoring sites were selected based on the frequency distribution of the TWI (Beven & Kirkby, 1979), which is a good predictor of median groundwater levels in this catchment (Rinderer et al., 2014). The frequency distribution of the TWI was divided into eight classes, and the coordinate of the pixel with the median TWI value for each class was chosen for installation of the well. This was done for the entire catchment and for six nested subcatchments (for more details, see Rinderer et al., 2014). The monitoring sites included 8 ridge sites, 22 midslope locations, and 21 footslope or depression locations. Twenty sites were forested, and 31 were located in grassland; 25 were located in a mollic Gleysol and 26 in an umbric Gleysol. The wells were installed by hand to the depth of refusal, which on average was 1.06 m below the surface (range: 0.45 to 2.16 m). The wells were screened over their entire length, except the top 10 cm, and sealed with bentonite at the soil surface to avoid surface runoff entering the bores. Groundwater levels were measured at a 5-min interval during the 2010 to 2014 summer months using Odyssey capacitance water level loggers (Dataflow Systems Pty Limited).

Stream stage was measured at a natural cross section at the catchment outlet at a 5-min time interval using a pressure transducer (DL/N 70 by STS, Sensor Technik Sirnach AG). Stage was converted into streamflow using a site-specific rating curve based on salt dilution measurements that covered 58% of the range of observed stages. Stream stage was higher than the measurements used to create the stage-discharge relationship for only 1% of the 4-year study period.

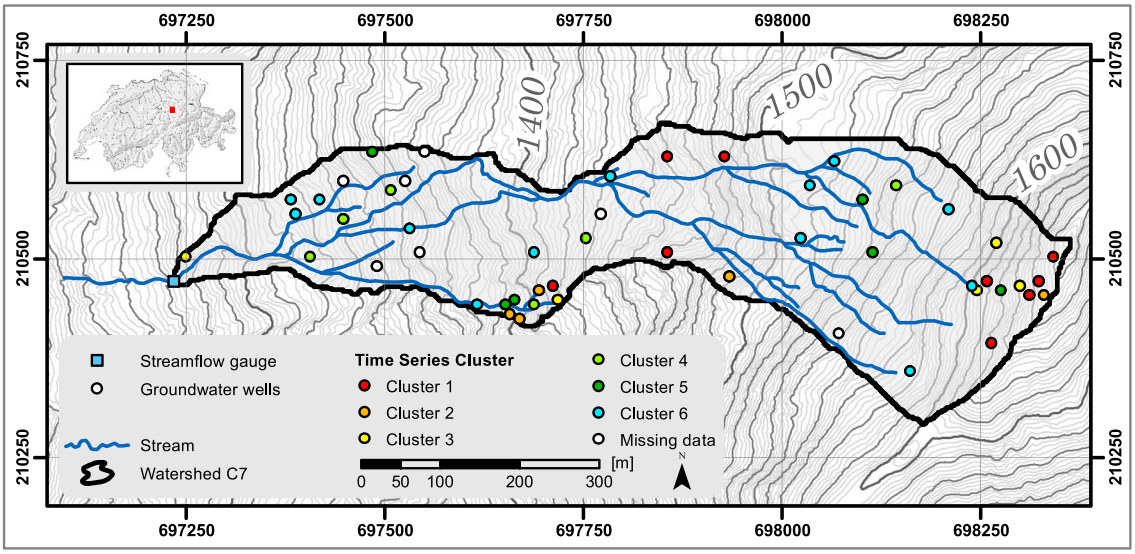

Figure 1. Map of the study catchment with the 51 groundwater monitoring sites. The colors represent the groundwater dynamics clusters to which the monitoring sites assigned (for a description of the characteristics of the clusters, see Table 1).

Air temperature, precipitation, and barometric pressure were measured at a weather station located 1 km northwest of the catchment at 1219 m a.s.l. We do not have reliable information on the spatial variability of the meteorological variables within the 20-ha catchment for the study period, but based on the relatively small size of the catchment, we assume that this variability is small.

Topographic characteristics (i.e., TWI, slope, upslope contributing area, and curvature) of the monitoring sites were derived from a 6 × 6 m DEM using a multiple flow direction algorithm (Seibert & McGlynn, 2007). Similar to Rinderer et al. (2014, 2016, 2017), sites with a TWI >6, a local slope <30%, and an upslope contributing area >600 m^2 are defined as footslope sites. Upslope sites are defined as sites with a TWI <4, a local slope >50%, and an upslope contributing area <200 m^2, whereas midslope sites have a TWI between 4 and 6, a local slope between 30% and 50%, and an upslope contributing area between 200 and 600 m^2.

2.3. Data Analysis Methods

2.3.1. Upscaling of the Measured Groundwater Responses to the Catchment Scale

2.3.1.1. Step 1: Clustering of Monitoring Sites Based on the Groundwater Time Series Similarity

For the analyses, we selected 19 months between 2010 and 2014 for which data were available for at least two thirds of all groundwater monitoring sites. The groundwater level and streamflow time series were subsampled at a 15-min time interval to reduce noise in the data. The groundwater level time series were normalized by the soil depth (1 at the soil surface and 0 at the bottom of the well) to allow extrapolation across the study site because soil depth data are not available for the entire catchment (cf. Rinderer et al., 2017).

To quantify the similarity between the streamflow and groundwater level time series, we determined the integral (sum of distances) between the two time series and used this similarity measure in the subsequent time series clustering procedure. Because small differences in the timing of the water level loggers, for instance, due to an offset of the internal clocks, would cause an artificial shift between the time series and thus strongly influences the similarity measure, we applied Dynamic Time Warping as a preprocessing step (Sakoe & Chiba, 1978). This method involves a limited stretching or compression of the time axis to compensate for small differences in the timing. We allowed for a maximum warping of 30 min in order to prevent distortion of the original time series (for more information and a graphical illustration of the method, see Rinderer et al., 2017). The Distance-based Time-Warped similarity (DTW) was calculated using the MATLAB function of Skov et al. (2006).

The groundwater level dynamics at the different wells were grouped using the DTW-based similarity measure and hierarchical clustering for each of the 19 months. The optimal number of clusters was based on the Calinski criterion (Calinski & Harabasz, 1974) and varied for the different months between four and

seven clusters. For consistency, a classification with six clusters was chosen because it was able to split the groundwater time series according to the mean water level (deep, intermediate, and shallow) and amplitude of the variations (small and large). Classifications with four clusters had higher within-cluster variability of topographic site characteristics, whereas classifications with seven clusters resulted in two clusters with only one member.

The persistence of the assignment of individual groundwater monitoring sites to one of the six clusters was quantified by the Krippendorff's Alpha (K), a measure to assess the degree of agreement between multiple-classifications (Krippendorff, 2004). Krippendorff's Alpha has a value of 1 if the monitoring sites are assigned to the same cluster for each of the 19 months and a value of 0 if the classifications are assigned randomly (Krippendorff, 2011). The final assignment of each groundwater monitoring site to one of the six clusters was based on the most frequent cluster assignment for the 19 months. All following data analyses, including the analysis of the seasonal connectivity patterns, were based on this final (i.e., most frequent) cluster assignment (named *time series-based classification* hereafter).

2.3.1.2. Step 2: Assignment of All Sites in the Catchment to a Groundwater Cluster Based on Topography

The topographic site characteristics that were distinctly different for the six clusters (TWI and curvature; see section 3.2) were used to assign each pixel in the catchment to one of the six groundwater clusters. In the following, this catchment-scale classification is named *topography-based classification*.

The agreement between the time series-based classification of a groundwater monitoring site and the topography-based classification for the corresponding pixel was assessed using several measures. A true positive class assignment means that the time series-based classification (true condition) and the topography-based classification (predicted condition) assigned a site to the same cluster. A true negative classification means that a site is not assigned to a certain cluster by neither the time series-based clustering nor the topography-based clustering. A false positive classification means that a site is assigned to a certain cluster (e.g., cluster 1) by the topography-based classification (predicted condition) when, in fact, it is not a member of this cluster according to the time series-based classification (true condition; type-I error). A false-negative class assignment means that a site is assigned to a certain groundwater cluster by the time series-based classification (true condition), but the topographically based classification (predicted condition) results in a different cluster assignment (type-II error).

1. The confusion matrix (Table 2) shows the classification offset between the two classifications, that is, the number of sites that belong to a certain cluster according to the time series-based classification and according to the topography-based classification.
2. The Positive Predictive Value (PPV; also called Producer's Accuracy) quantifies the proportion of classifications that are the same for the time series-based classifications and the topography-based classifications (PPV close to 1 is good). More specific, PPV is defined as the ratio between the sum of the true-positive class assignments divided by the sum of true-positive and false-positive class assignments (Stehman, 1997).
3. The True Positive Rate (TPR; also called User's Accuracy) describes the proportion of sites that are assigned to a certain class by both classifications relative to the sum of sites that either consistently have or have not been assigned to that class by both classifications (TPR close to 1 is good). More specific, TPR is defined as the ratio between the sum of true-positive class assignments divided by the sum of true-positive and false-negative cluster assignments (Stehman, 1997).
4. The Accuracy is defined as the ratio between the sum of true-positive and true-negative cases divided by the sum of all cases in the total population (Stehman, 1997). The Accuracy is the proportion of the sites that either have or have not been assigned to a certain class by both classification schemes relative to the total number of classifications (Accuracy close to 1 is good).

In addition, we calculated Cohen's Kappa (CK) (Cohen 1960), a statistical measure to assess the agreement of categorical data set, such as classifications. CK is 0 if two classifications do not agree at all and is theoretically 1 if they fully agree. However, when classes are not assigned with equal frequency, the maximum attainable CK values (CK_{max}) is smaller than 1. Therefore, we use the ratio of CK/CK_{max} as a measure of agreement (see Sim & Wright, 2005, and Rinderer et al., 2012, for equations and a more detailed explanation of CK).

2.3.1.3. Step 3: Assigning a Water Level to Each Pixel

For each time step, we calculated the mean of the relative groundwater levels for all wells in a cluster and assigned this mean value to each pixels of that cluster. This yielded the time series of estimated (perched) groundwater levels relative to the soil depth for each pixel (called *modeled relative groundwater levels* in the following). These modeled relative groundwater level values were converted to absolute groundwater levels below the surface using the estimated soil depth for each pixel. Detailed maps of soil depth are not available, and therefore, the soil depth was estimated for each pixel based on the linear relation between the measured soil depth (D, cm) and local slope (S, %) at the 51 monitoring sites ($D = 152.5 - 1.31\ S$; adjusted R^2: 0.19, $p < 0.001$, standard error: 0.34 m). The low correlation and relatively large standard error causes a mismatch between the modeled and observed soil depth for each pixel. It results in a bias in the modeled total groundwater storage but has a much smaller effect on the dynamics of the total groundwater storage. Furthermore, this effect is expected to largely balance out over the large number of pixels for the entire catchment ($n = 5685$). We tested other topographic site characteristics and combinations of site characteristics to predict the soil depth, but none of them led to a better prediction (i.e., higher correlation or lower standard error). To test the influence of the soil depth estimation on the active and connected area (see section 2.3.2), we also generated an alternative soil depth map based on kriging with external drift. This is a more sophisticated interpolation methods that incorporates the spatial autocorrelation of soil depth and additional auxiliary variables (in our case slope) to interpolate soil depth for the catchment. For this analysis, we fitted an exponential function to the variogram with a cutoff equal to 150 m.

2.3.2. Determination of Active and Connected Areas Across the Catchment

Because the hydraulic conductivity decreases sharply with depth below the soil surface in the Gleysols of this catchment, significant lateral subsurface flow is expected to occur only in the shallow soil layers close to the soil surface (Feyen et al., 1999). Because the soil profile and soil hydraulic properties are not known for each pixel, several response thresholds were tested: A pixel was considered to be hydrologically active when the water level was within the top 10, 20, or 30 cm of the soil profile. The 10-cm response threshold reflects near-saturated soil conditions, while the 20- and 30-cm thresholds represent water levels in the higher conductivity topsoil. The results shown in the graphs and tables in this manuscript are for the 20-cm threshold scenario, unless explicitly stated otherwise.

The subsurface flow network was represented as a graph network composed of 5685 nodes (i.e., the 6 by 6 m pixels for the study catchment) with the steepest downslope connection to one of the eight neighboring pixels (D8 flow direction algorithm; O'Callaghan & Mark, 1984) as the direction to which most subsurface flow was expected to occur. Based on graph theory, all active pixels with a continuous connection to the nearest stream via other active pixels (called *connected pixels* in the following) could be separated from *active pixels* where groundwater levels exceeded the threshold but the pixel had no continuous connection via other active pixels to a stream channel. We used the R-package *igraph* for building the graph network, which consists of a network of nodes (pixels) connected by their DEM-derived flowpaths. This graph-based spatial representation of the catchment allows very efficient queries on the active or connected status of neighboring pixels and has computational advantages compared to a raster-based approach. For each time step, we determined for each pixel if it was active and hydrologically connected, resulting in maps of modeled active and connected groundwater areas across the catchment.

The accuracy of the catchment-scale maps of active areas was assessed by calculating the fraction of wells for which the active state agreed (i.e., the measured water level was above the threshold and the modeled state for the corresponding pixel was active as well, and vice versa) This analysis was done for all time steps of the 19-month study period, resulting in time series of the fraction of agreement between measured and modeled groundwater activation. Since the descriptive statistics for this time series are biased toward baseflow conditions (which occur more often than the high flows), we also calculated the fraction of agreement between measured and modeled groundwater activation for 49 rainfall events. To select the rainfall events, first all rainfall events with total rainfall ≥5 mm or a maximum rainfall intensity of ≥3 mm/15 min were identified and events with multiple peaks, without a clear recession (i.e., specific discharge did not decline to a value that was similar to that at the beginning of the event), and events with a groundwater response but no streamflow response or vice versa were excluded from the set of events (cf. Rinderer et al., 2016). The remaining 49 rainfall events were grouped into four rainfall event types (1a: low-intensity/dry conditions [$n = 5$], 1b: low-intensity/wet conditions [$n = 8$], 2a: moderate-intensity/dry conditions [$n = 18$], and 2b:

moderate-intensity/wet conditions [$n = 18$]) based on the mean event rainfall intensity of 1.8 mm/hr that caused a water level response for at least 10% of all sites and the mean 3-day sum of antecedent rainfall, which was 10 mm. These thresholds were chosen to be consistent with Rinderer et al. (2016), even though the number of events differed for the different rainfall event types and the sample size for some of the event types was small. The significance of the difference in the median of the average fraction of agreement (i.e., agreement between measured and modeled groundwater activation during the event) for the four rainfall event types was tested using a Kruskal-Wallis significance test (Kruskal & Wallis, 1952) with a Bonferroni correction (Dunn, 1961). The difference in the median of the average fraction of agreement between measured and modeled groundwater activation for the different activation thresholds (i.e., 10, 20, and 30 cm) was assessed similarly.

All data preprocessing and data analyses were done in R (R Developer Community) using custom-made functions and functions from the *stats*, *hclust*, *irr*, *igraph*, and *hydroGOF* packages. For all statistical analyses, we used a 0.05 significance level.

2.3.3. Uncertainty Assessment

We tested the sensitivity of the active and connected areas to potential misclassifications of sites to the groundwater response clusters. To do this, we randomly assigned pixels to a new cluster according to the frequency of the disagreement between the topography-based and time series-based groundwater clusters. We did this 100 times and calculated the maps and time series of active and connected area for each new classification for one of the study months (August 2013, which was characterized by frequent groundwater responses of different magnitudes). Computational demand prevented us from repeating this procedure more often. We then compared the time series of the fraction of the catchment area that was active and connected based on the classification that we used in the study (i.e., based on the most frequent cluster assignment; called *baseline classification* in the following) with the time series of the median fraction of the catchment area that was active and connected for the 100 alternative classifications, using the root mean squared error (RMSE), the correlation coefficient (COR), and the Nash & Sutcliff Efficiency (NSE) (Nash & Sutcliffe 1970). It should be noted that the results of this sensitivity study are conservative as the random reassignment of pixels to another cluster does not account for spatial autocorrelation.

In addition, we tested the sensitivity of the time series of the active and connected areas to the soil depth estimates. We compared the time series of the fraction of the catchment that was active and connected based on the soil depth map from the linear regression with slope and soil depth map determined by kriging with external drift (see details above). As the RMSE was very low (0.0003, 0.0004, and 0.0000 [relative soil depth in %] for the active area and 0.0005, 0.0009, and 0.0000 [relative soil depth in %] for the connected area and the 10-, 20-, and 30-cm threshold scenario, respectively) and the COR and the NSE (Nash & Sutcliffe, 1970) were close to 1 (0.9960, 0.9995, and 1.0000 for the active area and 0.9955, 0.9996, and 1.000 for the connected area and the 10-, 20-, 30-cm threshold scenario, respectively), we only show the time series and maps of the active and connected areas based on the soil depth map derived from the linear regression of soil depth on slope.

3. Results

3.1. Groundwater Time Series Clustering

The DTW-based similarity measure of Rinderer et al. (2017) was used as the distance measure in the hierarchical clustering to group the 51 groundwater monitoring sites into six groundwater response clusters. Five sites were assigned to the same cluster for all 19 study months, 19 sites were assigned to two different clusters, 17 sites to three different clusters, 9 sites to four different clusters, and 1 site was assigned to five different clusters. The overall agreement of the cluster assignment for the 19 months, quantified by the Krippendorff's Alpha (K), was 0.49. The final assignment of a monitoring site to a cluster based on the most frequent cluster assignment for each site resulted in a relatively similar number of sites in each cluster (cluster 1: 9 sites, cluster 2: 5 sites, cluster 3: 5 sites, cluster 4: 6 sites, cluster 5: 6 sites, and cluster 6: 13 sites). Seven of the 51 groundwater monitoring sites could not be assigned to a cluster because the data were available for less than two thirds of the study period. The relative groundwater levels for the wells in each cluster were significantly different in terms of the mean relative groundwater level, the amplitude, and the rate of recession (Figure 3b).

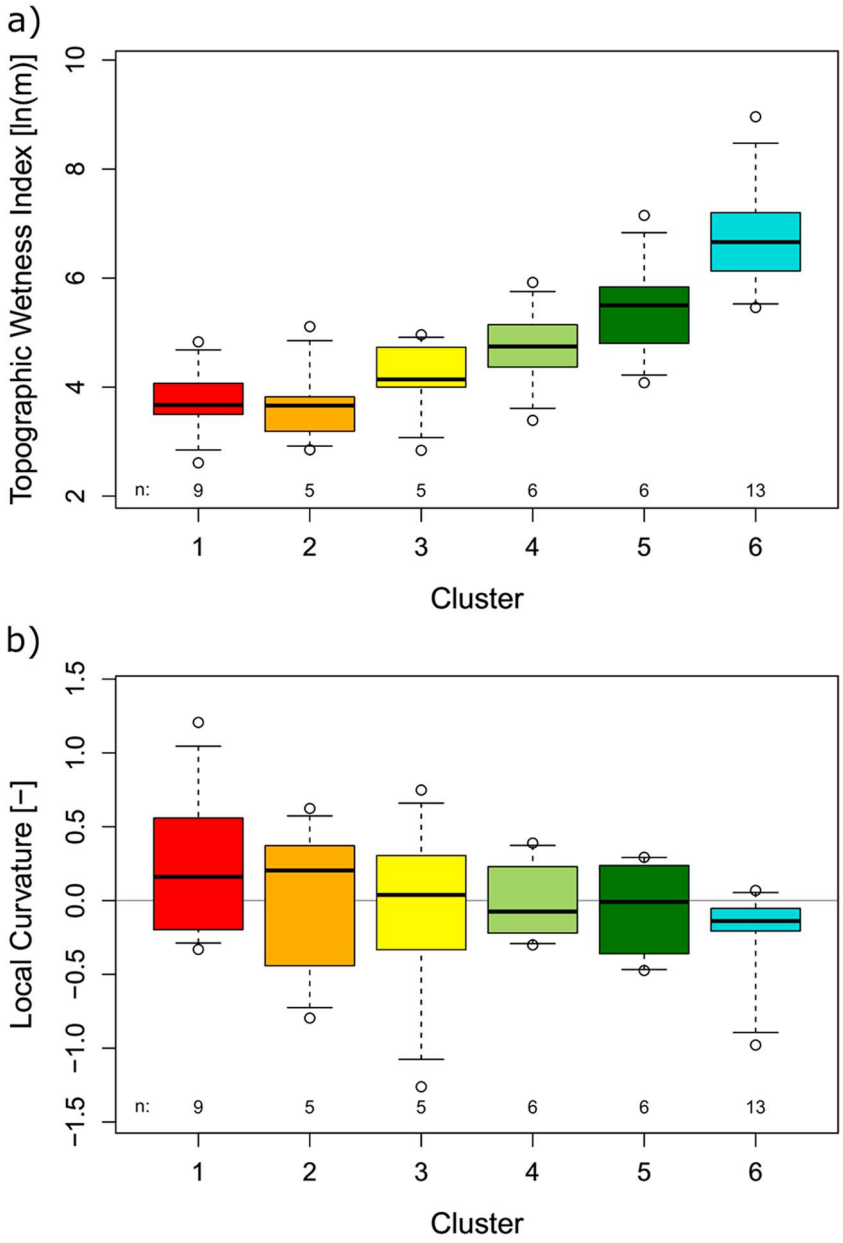

Figure 2. Box plots representing the within-cluster variability of Topographic Wetness Index (a) and curvature (b) for the groundwater monitoring sites within each cluster.

3.2. Distribution of Clusters Across the Catchment

In order to be able to assign all pixels in the catchment to a cluster, the within- and between-cluster variability of various topographic site characteristics for the monitoring sites in a cluster were examined. TWI was the most suitable characteristic to describe the six clusters. The median TWI, as well as the lower and upper quartiles for the six clusters, had limited overlap, except for clusters 1 and 2 (Figure 2a). However, the Kruskal-Wallis test with Bonferroni-adjusted p values indicated that only the median TWI of sites in clusters 1–4 were statistically significantly different from the median TWI of sites in cluster 6 (Figure 2a), likely in part due to the small number of sites in some of the clusters. Still the lower and upper quartiles of the TWI for each cluster were used to assign each pixel to a cluster (Table 1). The median TWI was similar for clusters 1 (TWI: 3.67) and 2 (TWI: 3.66), but the distribution of curvature of the monitoring sites was positively skewed for cluster 1 and negatively skewed for cluster 2 (median for both: 0.18; Figure 2b). Therefore,

Table 1
Description of the Six Time Series Clusters and the Ranges of the TWI and Curvature Used to Assign Each Pixel to a Cluster (see also Figure 2).

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Typical groundwater dynamics	Low median level and small amplitude	Low median level and large amplitude	Low median level, large amplitude, and longer tails	Intermediate median level, large amplitude, and flashy (quick rise and recession)	Intermediate median level, large amplitude, and less flashy	High median level and small amplitude
Typical site morphology	Ridge sites in steep upslope locations	Hollow sites in steep upslope locations	Steep midslope locations with small upslope contributing area	Midslope locations with intermediate upslope contributing area	Midslope locations with large upslope contributing area	Hollow and footslope locations (often near a stream or a depression)
Mean relative groundwater level	0.09	0.17	0.36	0.62	0.77	0.94
IQR relative groundwater levels	0.02	0.10	0.17	0.17	0.19	0.08
TWI (ln(m))	<4.0	<4.0	4.0–4.5	4.5–5.0	5.0–6.0	>6.0
Curvature (−)	≥0.18	<0.18				

Note. The relative groundwater level is zero at the soil bedrock interface and one at the soil surface. IQR = interquartile range; TWI = Topographic Wetness Index.

monitoring sites with a TWI < 4 and a curvature ≥ 0.18 were assigned to cluster 1 and sites with a TWI < 4 and curvature < 0.18 were assigned to cluster 2.

The assignment of all pixels in the catchment based on these TWI and curvature boundaries (Table 1) led to the following distribution of groundwater clusters across the catchment: 14% of the catchment area (3 ha) was assigned to cluster 1 (typically ridge sites in steep upslope locations), 5% of the area (1 ha) to cluster 2 (typically hollows in steep upslope locations), 12% (2.5 ha) to cluster 3 (mainly steep midslope locations with small upslope contributing areas), 13% (2.7 ha) to cluster 4 (mainly midslope locations with intermediate upslope contributing area), 23% (4.8 ha) to cluster 5 (midslope locations with large upslope contributing areas), and 23% (6.5 ha) to cluster 6 (typically hollow and footslope locations near the streams; Figure 3a).

The CK for the agreement between the time series-based cluster assignment of the groundwater monitoring sites and the topography-based cluster assignment for the corresponding pixels was 0.35 (CK_{max} = 0.89; CK/CK_{max} = 0.40). The (total) Accuracy was 0.84. The PPV (Producer's Accuracy) for the six clusters ranged between 0.20 to 0.91, with the lowest value for cluster 3 and the highest value for cluster 6 (Table 2). The TPR (User's Accuracy) ranged between 0.20 and 0.77, with the lowest values for clusters 2 and 3 and the highest value for cluster 6 (Table 2). The kernel distributions of the time series-based cluster assignment and the topography-based cluster assignments were not significantly different (p value of Kolmogorov-Smirnov test = 0.99).

3.3. Agreement Between the Measured and Modeled Groundwater Activation Time Series

The time series of the mean relative groundwater level for each cluster was used together with the topography-based cluster assignments to model the relative groundwater level for each pixel in the catchment. These relative groundwater levels were converted to actual groundwater levels based on the estimated soil depth derived from the regression with the local slope. Finally, for every time step, it was determined whether a pixel was active or not based on the value of the activation threshold (i.e., a rise of groundwater level to within, e.g., 10, 20, or 30 cm of the soil surface). While there were differences between the measured and predicted groundwater level time series for the monitoring sites, these differences had a very limited effect on the time that groundwater levels were above the activation threshold (Figure 4). The average fraction of monitoring sites for which the measured groundwater activation agreed with the modeled activation for the corresponding pixel for the study period was 0.91 (25th, 50th, and 75th percentile: 0.88, 0.92 and 0.95, respectively). The fraction of agreement dropped by 10% to 15% at the beginning of most rainfall events (example in Figure 5), but the average fraction of agreement during an event was high for all 49 events (Figure 6a). The average fraction of agreement differed for the 10-, 20- and 30-cm activation thresholds (Figure 6a) and was highest for the 10-cm activation threshold (mean of the average fraction of agreement during an event of 0.91, 0.88, and 0.87 for the 10-, 20- and 30-cm threshold, respectively; n = 49 events).

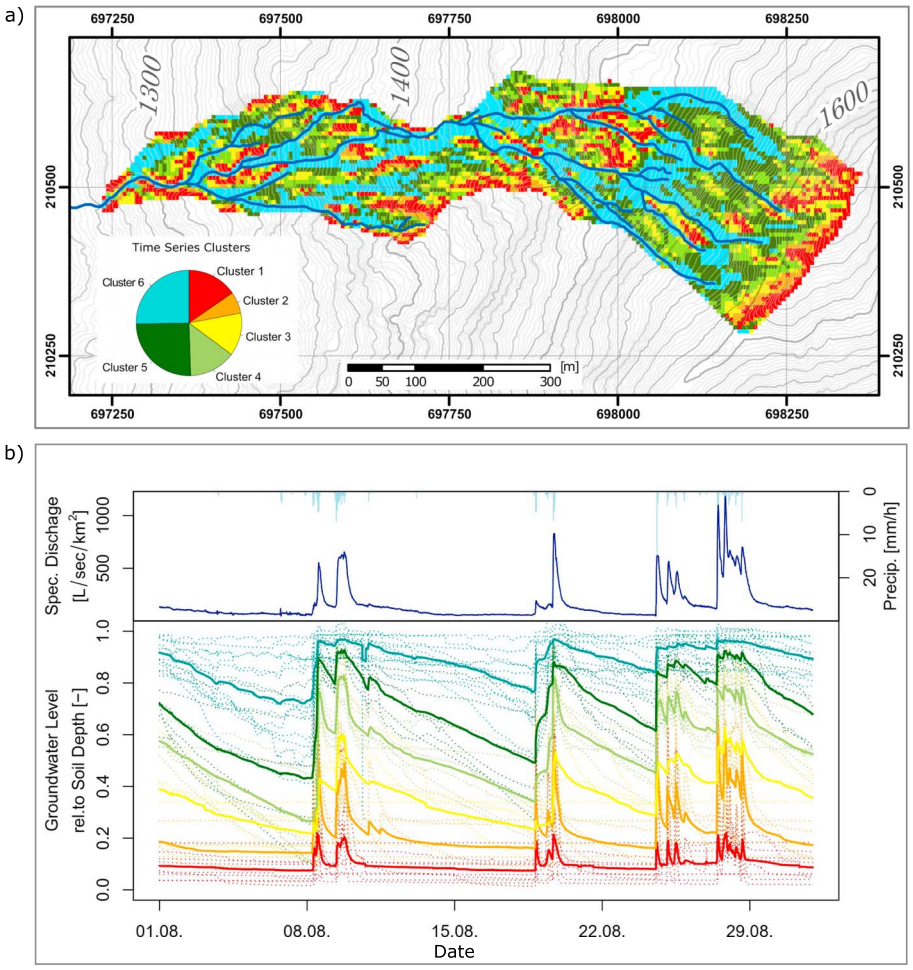

Figure 3. (a) Map of the topography-based (i.e., Topographic Wetness Index and curvature based) groundwater clusters. The insert shows the fraction of the total catchment area for each cluster (20 ha = 100%). (b) Time series of specific discharge, daily precipitation, the mean relative groundwater level for individual sites in each cluster (dotted lines), and the mean for all monitoring sites in a cluster (solid line) for August 2013.

The mean fraction of agreement was not significantly different for the four rainfall event types (Figure 6b). When the measured soil depth, instead of the regression-based soil depth, was used for each monitoring site, the mean fraction of agreement was higher and not significantly different for the three thresholds (mean of the average fraction of agreement during the events: 0.91, 0.91, and 0.90 for the 10-, 20-, and 30-cm threshold

Table 2
Confusion Matrix, Positive Prediction Value, and True Positive Rate for the Time Series-Based Cluster Assignment and the Topography-Based Cluster Assignment of the Groundwater Monitoring Sites (n = 44, Seven Sites Were Not Classified Because They Had Data for Less Than Two Thirds of the Time Analyzed)

		Time series-based cluster						Marginal sum	Positive predictive value
		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6		
Topography-based cluster	Cluster 1	4	3	2	1	0	0	10	0.40
	Cluster 2	2	1	0	0	0	0	3	0.33
	Cluster 3	2	0	1	1	1	0	5	0.20
	Cluster 4	1	0	2	2	1	0	6	0.33
	Cluster 5	0	1	0	2	3	3	9	0.33
	Cluster 6	0	0	0	0	1	10	11	0.91
Marginal sum		9	5	5	6	6	13		
True positive rate		0.44	0.20	0.20	0.33	0.50	0.77		

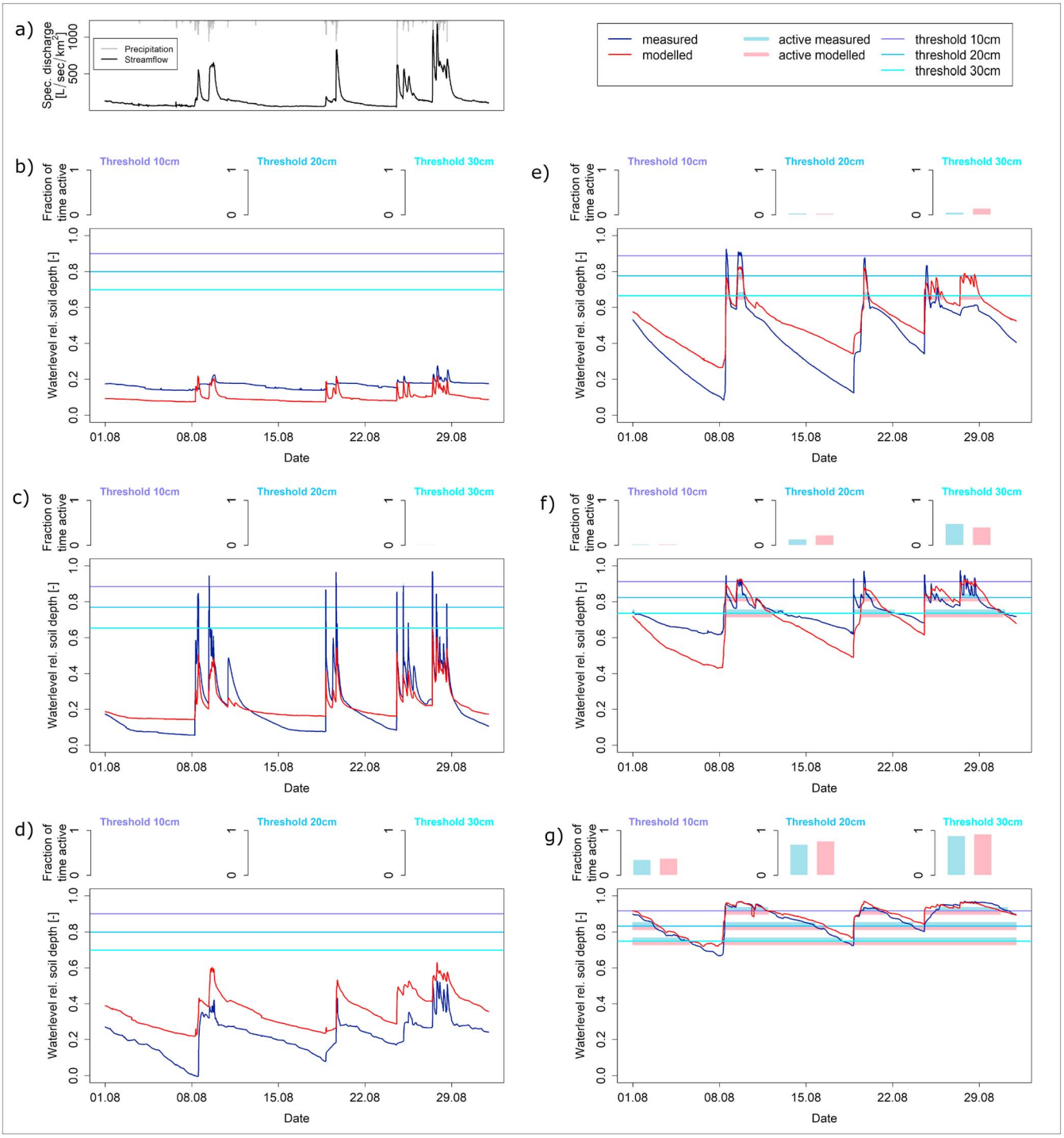

Figure 4. Time series of the (a) specific discharge (black) and precipitation (gray), (b-g) measured (blue) and modeled (red) relative groundwater levels (1 = at the soil surface and 0 = at the soil bedrock interface) for a groundwater monitoring site that belongs to groundwater cluster 1 (b), cluster 2 (c), cluster 3 (d), cluster 4 (e), cluster 5 (f), and cluster 6 (g) for August 2013. The activation thresholds (10, 20 and 30 cm) are indicated with thin horizontal lines. The time that the measured and modeled groundwater levels were above the activation thresholds is indicated by thick horizontal lines. The bar plots in the top row of (b) through (g) indicate the total fraction of time that the site was considered to be active (observed = blue and modeled = red) for the three activation thresholds. We conclude that differences between the measured and modeled groundwater time series had a limited effect on the fraction of time the sites were active.

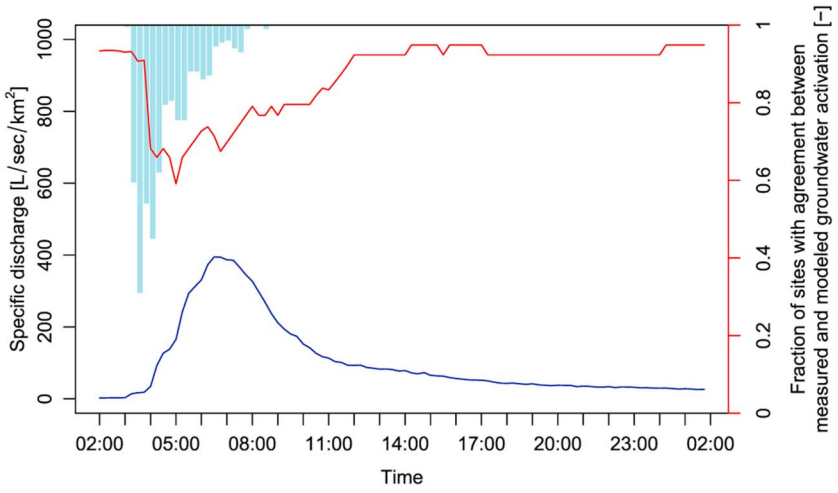

Figure 5. Fraction of sites for which there was agreement between measured and modeled groundwater activation (or inactivation); (red), as well as the specific discharge at the catchment outlet (dark blue) and precipitation (light blue) during a rainfall event on 16 August 2012.

scenarios, respectively). This suggests that at least part of the disagreement in the observed and predicted activation is due to the lack of soil depth data across the catchment. The median values for the agreement between measured and modeled activation were comparable to the mean values.

3.4. Catchment-Scale Modeled Active and Connected Areas

The modeled time series of activation for each pixel allowed us to derive maps of the modeled active and connected areas for each time step (Figure 7). During baseflow conditions, the groundwater levels were below the activation threshold in most locations, and active and connected areas were limited to the near-stream areas. Soon after the start of a rainfall event, the footslope and midslope locations became active. The connected areas were largely aligned with the stream channels and expanded laterally and in an upstream direction with increasing rainfall (Figure 7). However, some parts of the landscape, particularly in midhill and

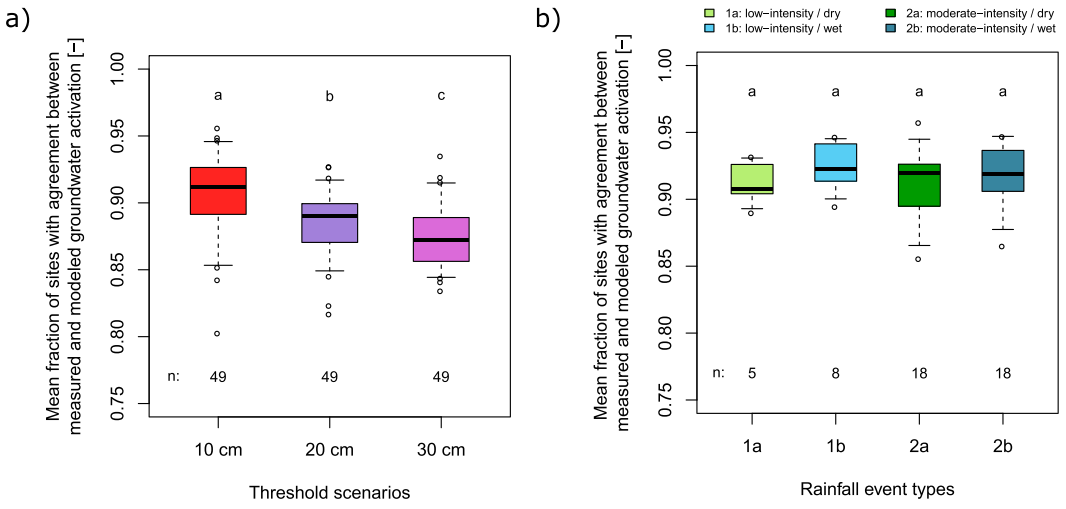

Figure 6. Box plots of the average fraction of monitoring sites for which measured and modeled groundwater activation agreed during an event for the three threshold scenarios (10, 20 and 30 cm; (a)), and the four different rainfall event types for the 20-cm threshold (b). The box indicates the 25% and 75% percentiles, the whiskers extends to the data points within 1.5 times the interquartile range, the dots indicate outliers, and the solid line indicates the median. Kruskal-Wallis test with the Bonferroni-adjusted p values showed that the median of the mean agreement during an event was significantly different for the different groundwater activation thresholds but not for the four rainfall event types.

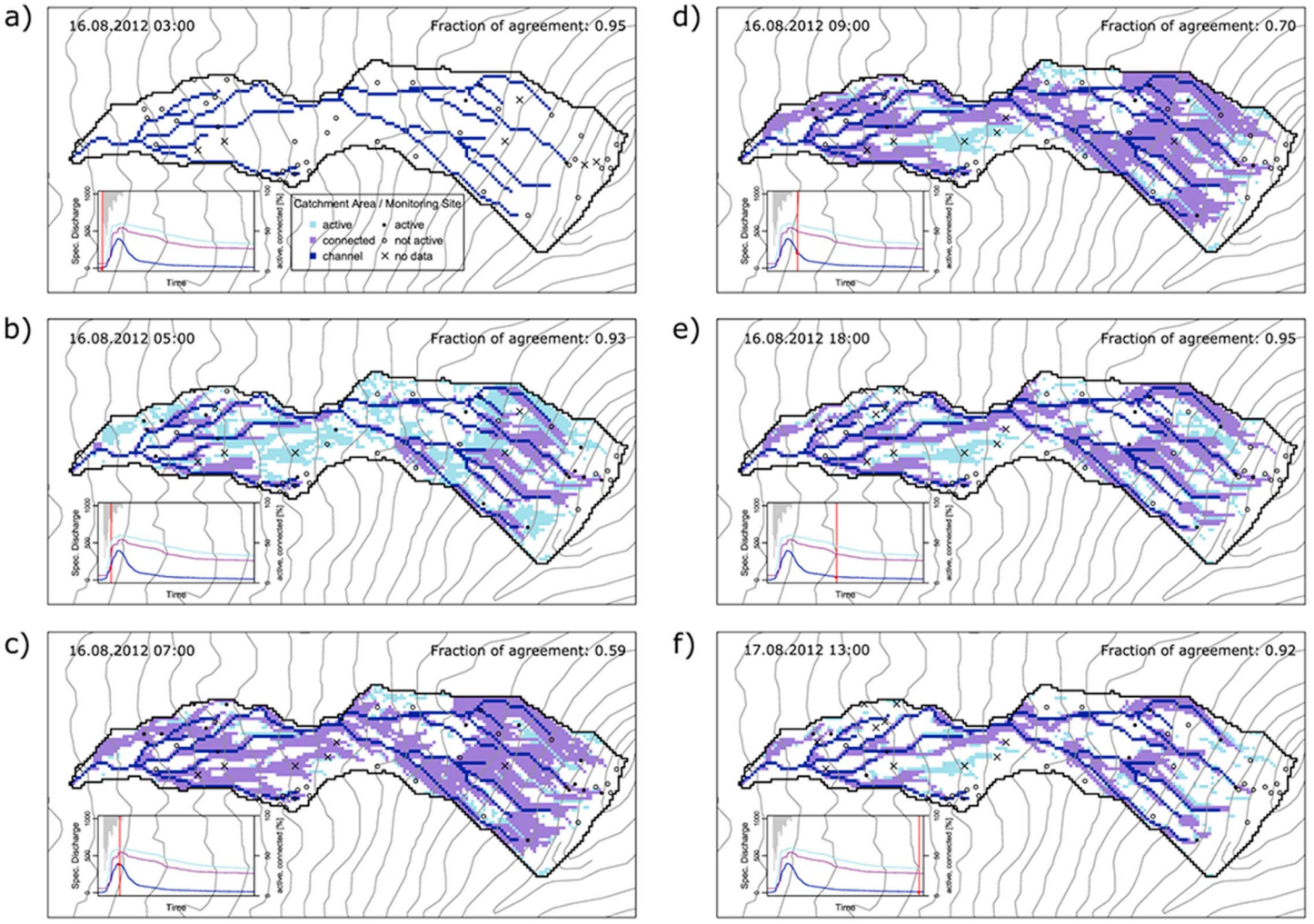

Figure 7. Modeled spatial patterns of active and connected areas during the 42-mm rainfall event on 16 August 2012. The inset shows the time series of the specific discharge and the percentage of the catchment area that is active and connected. The fraction of monitoring sites for which measured and modeled groundwater activation agreed is shown for each time step in the upper right corner of the subplot.

uphill locations, were only connected to the streams for a short time during the rainfall events. For the 42-mm rainfall event and the 20-cm activation threshold shown in Figure 7, 40% (8.2 ha) of the catchment area did not become active, and 45% (9.2 ha) of the catchment area did not become connected to the stream (see white areas in Figure 7; e.g., steep ridge locations close to the water divide). The difference between the active area and connected area varied between events but never exceeded 19% (3.8 ha). The part of the catchment that did not become active during an event ranged between 21% (4.3 ha) and 48% (9.8 ha) for the 49 rainfall events, whereas the part that did not become connected ranged between 23% (4.7 ha) and 67% (13.7 ha). For the 19-month study period, 18% of the area (3.6 ha) never became connected to the stream, and 40% of the area (8.2 ha) was connected to the stream for less than 5% of the time. In contrast, 10% of the area (2.0 ha) was connected to the stream for more than 95% of the time, and 9% of the area (1.9 ha) was always connected (Figure 8).

The extent and frequency of connectivity varied seasonally. For example, the modeled connected area (for the 20-cm activation threshold) in March 2013 was relatively small and limited to the near-stream areas (5.5 ha or 27%; Figure 9a, Table 4), but these areas were connected to the stream during the entire month. A large part of the catchment (13.3 ha or 64%) did not become connected during the entire month. In April 2013, connectivity was more widespread, so that only 5.5 ha (27%) of the catchment did not connect

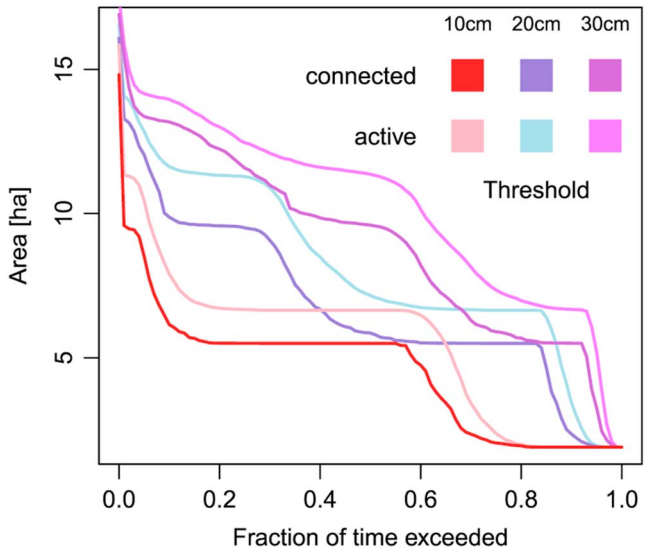

Figure 8. Duration curves of active and connected areas for the 19-month study period for the three different activation thresholds (10, 20, and 30 cm).

to the stream during the entire month (Figure 9b). In May 2013, 15.9 ha (78%) of the catchment was connected to the stream at least once during the month, and 8.9 ha (44%) was connected for more than 3 weeks (Figure 9c). In June 2013, the area that was connected to the stream for the entire month decreased to 5.5 ha (27%), but only 3.8 ha (19%) of the catchment was never connected (Figure 9d). This changed in July, when only 1.9 ha (9%) of the catchment was always connected to the stream and the area that did not connect to the stream at all increased to 6.8 ha (33%; Figure 9e). The extent and frequency of connectivity in July 2013 was characteristic of the overall connectivity pattern for the entire 19-month study period (Figure 9f).

3.5. Effect of Misclassification on Active and Connected Areas

Although errors in the cluster assignment had a larger effect on the total connected area than the total active area, the dynamics of the active and connected areas (i.e., expansion and contraction over time) were not significantly influenced by the misclassification of sites to a cluster. For the 20-cm activation threshold scenario, the RMSE for the relation between the active area based on the *baseline* classification and the median of the active area for all the Monte Carlo runs for the adjusted cluster assignments was 0.81 ha (4%), the COR 1.0, and the NSE 0.94. For the connected area, the RMSE was 1.55 ha (8%), the COR was 0.99, and the NSE was 0.67. The agreement in the modeled active and connected areas for the baseline classification and the median for the Monte Carlo simulations was better for the 10-cm threshold but slightly worse for the 30-cm activation threshold (Table 3).

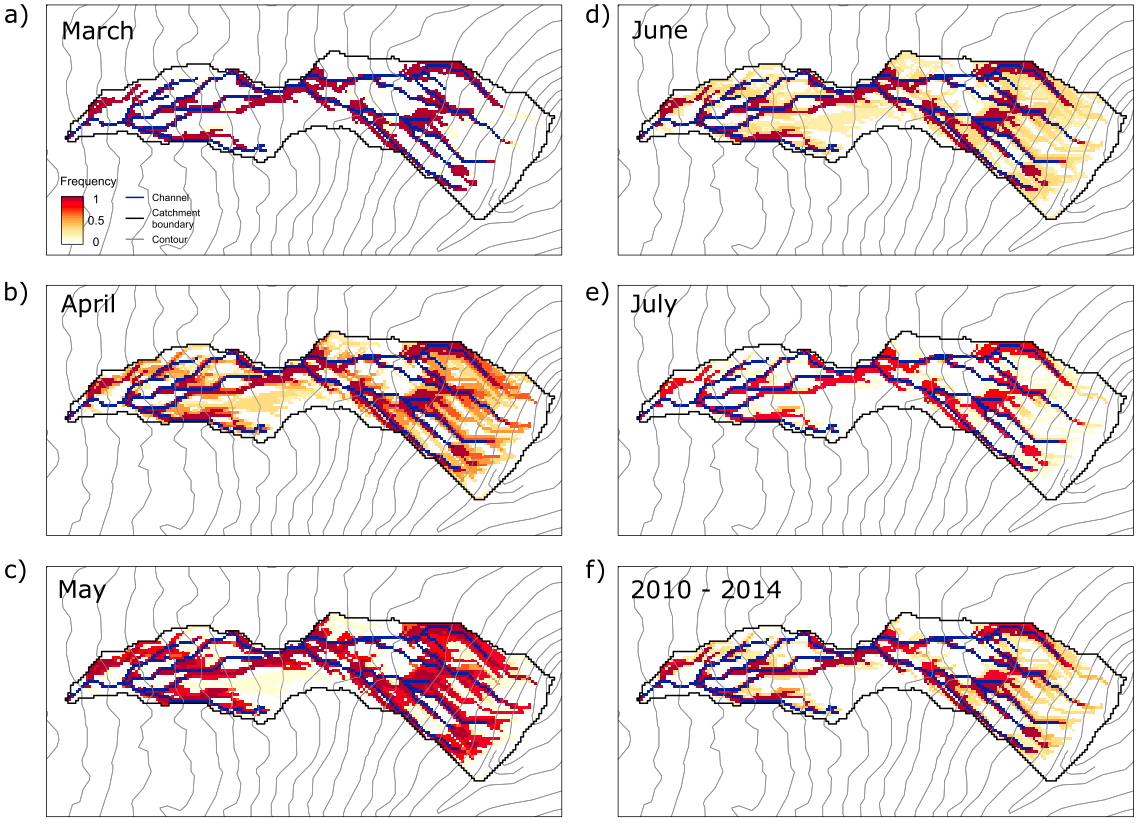

Figure 9. Maps showing the fraction of time that a pixel was connected to the stream during March (a), April (b), June (c), July 2013 (d), and the entire study period (2010 to 2014) (f). Monthly rainfall was 91, 107, 310, 292, and 192 mm for March, April, May, June, and July 2013, respectively.

Table 3
Agreement Between the Time Series of the Modeled Active Area and Connected Area for August 2013 for the Baseline Scenario and the Time Series of the Median of the Modeled Active and Connected Areas for the 100 Monte Carlo Simulations for Which Pixels Were Assigned to a Different Cluster

Threshold scenario	10 cm		20 cm		30 cm	
Area	Active	Connected	Active	Connected	Active	Connected
RMSE	0.69	0.74	0.81	1.55	0.83	2.18
COR	0.99	1.00	1.00	0.99	1.00	1.00
NSE	0.92	0.85	0.94	0.67	0.92	0.46

Note. RMSE = root mean square error (relative soil depth in percent); COR = correlation; NSE = Nash-Sutcliff Efficiency.

4. Discussion

4.1. Upscaling of the Groundwater Level Dynamics Based on Time Series Clustering and Topographic Characteristics

4.1.1. Uncertainties in Time Series Clustering

The time series-based cluster assignment and topography-based cluster assignment agreed moderately well (Table 2 and CK values). The agreement in cluster assignment was best for upslope and footslope sites but more uncertain for midslope sites (Table 2). These midslope sites are characterized by low to intermediate median groundwater levels but large response amplitudes and comprise 33% of the total catchment area (clusters 2–4; Figure 3). Midslope locations have previously been identified as zones with a high variability in median groundwater levels and groundwater response timing (Rinderer et al., 2014; Rinderer et al., 2016). Midslope locations have also been highlighted as important locations for connectivity between upslope runoff source areas and the riparian area in catchments where saturation started in upslope locations with shallow soils (Jencso et al., 2009; McNamara et al., 2005; Ocampo et al., 2006; Stieglitz et al., 2003; Tetzlaff et al., 2014; van Meerveld et al., 2015). The poorer correlation between groundwater levels and topography that has been reported in previous studies (Rinderer et al., 2014; Rinderer et al., 2016) in these midslope areas may reflect differences in soil depth, flow pathways from upslope areas, and storage deficits, which cause a larger temporal and spatial variability in the groundwater responses, particularly at the beginning of the events. This is when the steady-state assumption of the topography-based groundwater level estimates (i.e., that groundwater tables rise and fall in unison) is most likely not fulfilled (Rinderer et al., 2014). This likely explains the drop in the agreement between the measured activation and the modeled activation at the start of the events (Figure 5).

Another source of uncertainty in the time series clustering method is the within- and between-cluster variability of the site characteristics (TWI and curvature) that is used to upscale the mean cluster groundwater response to all nonmonitored pixels. However, the sensitivity analysis showed that misclassifications had a minor influence on the active area. It resulted in an offset in the total connected area (see RMSE and NSE in Table 3) but did not affect the dynamics of the connected area, that is, the way the connected area expands and contracts (see COR in Table 3). The grouping of the groundwater dynamics into six clusters and the assumption that all sites in this cluster respond similar to the mean relative groundwater level of all monitoring sites in a cluster are of course gross simplifications as there are obviously more than six types of groundwater responses. On the other hand, this clustering is a necessary step to synthesize site-specific groundwater dynamics to a general groundwater behavior that is more appropriate for transferring or upscaling to other sites. Using more clusters would allow for a better discretization of the naturel variability in groundwater dynamics but would also make it harder to identify site characteristics that separate the clusters, so that overall, this will likely result in more misclassifications of sites.

4.1.2. Groundwater Activation Thresholds, Antecedent Wetness, and Rainfall Event Characteristics

The differences between the measured and modeled groundwater time series had a limited effect on the time that the modeled groundwater levels exceeded the activation threshold (Figure 4). The agreement between measured and modeled groundwater activation did not significantly depend on antecedent wetness conditions or rainfall intensity (Figure 6) but was lowest at the beginning of rainfall events (Figure 5). This

finding is in agreement with results from a previous study that indicated that the relation between groundwater level and topographic indices (i.e., TWI) was less strong at the beginning of rainfall events (Rinderer et al., 2014). This is likely because spatial differences in antecedent wetness (and interception losses) have a large effect on the onset of the groundwater level rise but only a small effect on the groundwater dynamics later in the event.

While the agreement between measured and modeled groundwater activation was high for all activation thresholds, the choice of the activation threshold had a significant effect on the agreement (Figure 6a). The agreement between measured and modeled groundwater activation was the highest for the 10-cm threshold scenario, but this threshold results in a very conservative estimate of active areas because it represents only lateral flow in the near surface layer, while fast lateral flow can occur in lower soil layers as well (e.g., through macropores that extend deeper). The activation threshold should thus be based on knowledge of the soil properties, particularly the variation in hydraulic conductivity with depth. The choice of our threshold scenarios (i.e., 10, 20 and 30 cm) was informed by tracer tests for grassland areas in the Alptal region that indicated that most lateral flow occurred through the A horizon and highlighted the importance of preferential flow in the underlaying B horizon for forested areas (Feyen et al., 1999; Weiler et al., 1998). In reality, there will not be a clear activation threshold (i.e., there is no sharp boundary) but a zone over which lateral flow increases, and this may differ for different land uses. Activation thresholds have been used in several model-based runoff source area investigations (Birkel et al., 2010;Nippgen et al., 2015 ; Smith et al., 2013), but thresholds needed either to be calibrated to fit the extent of modeled and observed saturated areas (Blazkova et al., 2002), to fit the cumulative frequency distribution of the connectivity duration curve to the flow duration curve (Nippgen et al., 2015; Smith et al., 2013), or to match modeled and observed hydrochemical or stable water isotope signatures (Birkel et al., 2010).

The estimated soil depth affected the agreement between measured and modeled groundwater activation. When the measured soil depth, instead of the regression-based soil depth, was used the fraction of sites for which measured and modeled groundwater activation agreed was ≥0.90 and not statistically significantly different for the three threshold scenarios, a more sophisticated method to estimate the soil depth (i.e., kriging with external drift) was not successful in providing a more realistic estimate of soil depth because the spatial autocorrelation in the available soil depth data and topographic predictor variables was low. For instance, the spatial autocorrelation of TWI (calculated sensus Western et al., 1998) was 22 m.

4.1.3. Upscaling Based on Surface Topography

The approaches used in this study are based on assumptions and simplifications that might only be justified in catchments with steep topography and groundwater tables close to the soil surface. In the graph-theory approach used here to assess connectivity, the flow network is based on the surface topography, is static and does not account for multiple flow direction. We argue that in soils with a sharp decline in the saturated hydraulic conductivity with depth and therefore efficient lateral flow only within the uppermost 10 to 30 cm (see scenario simulations), the flow direction is likely similar to the slope of the soil surface and can be derived from a DEM. Because the catchment is steep (average slope is 35%), small differences in hydraulic head within the more conductive upper 10- to 30-cm soil layer are likely smaller than the differences in gravitational potential, so that the flow direction is mainly determined by the surface topography. To prove that a static flow direction based on surface topography is reasonable, we calculated the change in flow directions (as described by Rodhe & Seibert, 2011; van Meerveld et al., 2015) based on the hydraulic head of 27 groundwater triplets (i.e., three neighboring groundwater wells) for all time steps ($n = 2624$) of the 49 rainfall events analyzed in this study. The mean difference between the groundwater level-based and topography-based flow directions ranged between <0.01° and 0.3°. The standard deviation of the groundwater level-based flow direction ranged between 0.1° and 3.7°. We thus concluded that for a steep catchment with shallow (perched) groundwater levels as in our study site, the surface topography can be used to determine the main flow directions to describe connectivity. In less steep catchments with more permeable soils (e.g., as shown in Rodhe & Seibert, 2011; van Meerveld et al., 2015), this is likely not the case. The effect of the choice of the flow direction algorithm (e.g., D8 instead of D-Infinity) is likely also smaller than for flatter catchments. Because flat areas and thus divergent flow is rather seldom in our catchment, the results are likely not significantly affected by the choice of the flow algorithm and the assumption of a single flow direction. Furthermore, riparian areas, where the flow direction can vary significantly (Rodhe & Seibert, 2011), cover less than 1% of the area in our study catchment.

Despite these uncertainties and assumptions, the upscaled groundwater dynamics corresponded well with the measured groundwater time series at the monitoring locations, suggesting that topography based upscaling of clustered groundwater dynamics is a promising method to derive subsurface runoff source area in catchments where the groundwater tables are close to the soil surface. A question remains whether this approach is applicable, if an extended groundwater monitoring network, like ours, is not in place? This study and previous work by Rinderer et al. (2017) showed that if groundwater is monitored in a few representative landscape positions, it can be transferred to other sites with similar environmental conditions — at least in catchments with steep slopes and shallow groundwater tables. As a consequence, the effort and costs for groundwater monitoring can be considerably reduced by an informed choice of representative monitoring locations based on landscape position. The results of this study show that the data-driven approach used here can be used to infer catchment-scale groundwater responses and connectivity patterns from point-scale monitoring data. Data-driven approaches, like the one presented here, depend on the amount of data available and the assumption that sites with similar landscape characteristics also have similar groundwater dynamics. At least for our study site with steep terrain and shallow groundwater tables, this could be proven (Rinderer et al., 2017). We, therefore, expect this method to be also useful for other catchments where the differences in groundwater level response dynamics are related to topographic indices or other spatial information.

4.2. Runoff Source Area Dynamics

4.2.1. The Variable Source Area Concept: Is It also Representative for Subsurface Flow?

The upscaled groundwater level responses suggest that groundwater activation during rainfall events was widespread but that the connected areas were primarily located along the stream network and expanded laterally and longitudinally during rain events (Figure 7). This observation is in agreement with the variable source area concept for overland flow (Hewlett & Hibbert, 1963; Tsukamoto, 1963). Most likely the very rapid runoff response at the beginning of events (see Figure 3b) is caused by discharge of water from near-stream areas. The runoff during peakflow and the recession limb is more likely from water that has been transported via lateral subsurface flow from areas of the catchment with a hydrological connection (i.e., a continuous flow path) to the stream. This fast lateral subsurface flow near the surface has also been reported in other experimental studies in neighboring catchments (Feyen et al., 1999; Weiler et al., 1998). Tracer tests would be necessary to corroborate this concept of runoff generation at the catchment scale and to test in how far the connected areas actually contribute to runoff at the event time scale.

Typically, the storage deficits are satisfied first in footslope locations near the channel where there is less available storage capacity, and therefore, these areas are the first to respond. However, in some circumstances, storage capacity can be limited in upslope locations (e.g., due to shallow soils at upslope locations), and therefore, these areas respond first. For instance, in the Panola Mountain Research Watershed, near Atlanta, Georgia, groundwater responded first in upslope locations with thin soils near bedrock outcrops, and hydrological connectivity to the bottom of the hillslope was not established until subsurface depressions in the middle of the hillslope were filled and facilitated further flow downslope (Tromp-van Meerveld & McDonnell, 2006; van Meerveld et al., 2015). Similar differences in response times were observed in the Bridge Creek catchment in the Italian Dolomites, where thinner soils in upslope locations caused groundwater levels to respond earlier than in footslope locations, and anticlockwise hysteresis between streamflow and connectivity suggested that hillslopes became connected to the stream only during the falling limb of the hydrograph (Penna et al., 2015; Zuecco et al., 2016; Zuecco et al., 2018). This downslope expansion of active areas was not seen for the Alptal study catchment.

The modeled pattern of expansion and contraction of the connected area along the channel network is similar to the expansion of saturated areas that produced overland flow in the steep, well-drained W4 headwater catchment at Sleepers River in Vermont. The saturated area expanded mainly longitudinally from the narrow valley bottoms (3% of the catchment) into the steeper headwater areas (11% of the catchment; Dunne et al., 1975). In contrast, the saturated areas in the gentler and poorly to moderately drained W2 catchment at Sleepers River mainly expanded laterally into the hillslopes (from 11% to 51% of the catchment) and less into the headwaters (Dunne et al., 1975).

Table 4
Catchment Area in Hectare and Percent That Was Connected to the Stream for Different Fractions of Time for March, April, May, June, and July 2013 and the Entire Study Period 2010 to 2014 Based on an Activation Threshold of 20 cm (See Figure 9 for Corresponding Maps)

Duration connected (fraction of time)	March		April		May		June		July		2010–2014	
	(ha)	(%)	(ha)	(%)	(ha)	(%)	(ha)	(%)	(ha)	(%)	(ha)	(%)
0	13.1	64	5.5	27	4.5	22	3.8	19	6.8	33	3.6	18
0–0.25	1.8	9	2.1	10	6.3	31	6.8	33	7.6	37	7.3	36
0.25–0.50	0.0	0	3.9	19	0.1	0	4.3	21	0.6	3	3.8	19
0.50–0.75	0.0	0	3.5	17	0.7	3	0.0	0	0.0	0	0.2	1
0.75–1	0.0	0	0.0	0	3.4	17	0.0	0	3.6	18	3.6	18
1	5.5	27	5.5	27	5.5	27	5.5	27	1.9	9	1.9	9

4.2.2. Variable Active Versus Variable Connected Areas

The sequence of maps with the active and connected areas during rainfall events also allowed identification of areas where the groundwater table rose into the more transmissive soil layer but did not have a continuous connection to the stream channel (Figure 7). Small, spatially restricted areas in uphill positions were active but not connected unless so called *gatekeeper* sites facilitated a continuous flowpath to the stream. These gatekeeper sites were predominantly located at midslope positions and members of clusters 4 and 5. There were differences between the active and connected areas for all individual events and for different seasons. Contrary, Nippgen et al. (2015) showed that the difference between the active and connected area was the largest during snowmelt and disappeared during the remainder of the year. The results of both studies highlight the need to distinguish between the active and connected area in order to avoid that models overestimate the extent of the runoff source areas. In our study site, where the source areas mainly expanded along the dense stream network, the difference between the active and connected area were relatively small (max difference: 19%), and the dynamics of the active and connected areas were highly correlated. In other catchments with more permeable soils, however, the difference between active and connected area is probably much larger, and a distinction between active and connected area is even more important.

4.2.3. Seasonal Patterns of Hydrological Connectivity

Changes in the area that was connected to the stream were the largest during the transition from snowmelt to rainfall-dominated conditions (Figure 9 and Table 4). While the connected area in our study catchment was restricted to a narrow area next to the stream channels (27% of the catchment area) when the catchment was snow covered and no or very little water infiltrated from the snowpack in to the soil, connectivity was widespread as snowmelt reached its maximum in late April/early May (Figure 9). Toward the end of the snowmelt in May 2013, 78% of the catchment was at least once connected to the stream. In the Tenderfood Creek Experimental Forest, a snowmelt-dominated catchment in Montana, United States, the active area during peak flow in spring reached 78% of the total catchment area but was only 2% during the rest of the year (Nippgen et al., 2015). However, a shift in the runoff source area from near-stream areas to footslopes and finally upper hillslopes, as reported from the Tenderfood Creek Experimental Forest (Nippgen et al., 2015), could not be observed in our study catchment. This is surprising as the elevation difference of the study catchment is 380 m and differences in the depth of the snowpack at the lower and upper parts of the catchment at the end of the winter are remarkable. The average snow water equivalent in open grassland sites for the winter periods 1997 to 2014 was 165 mm at 1260 m a.s.l. and 280 mm at 1430 m a.s.l. (monthly snow surveys by the Federal Institute for Forest, Snow and Landscape Research WSL).

The area with a persistent connection to the stream shrank from 5.5 ha (27%) in spring to 1.9 ha (9%) during the summer months (Figure 9). While in June, the groundwater levels were still high, and therefore, 16.7 ha (81%) of the catchment was at least once connected to the stream; 6.8 ha (33%) of the catchment did not become connected during July (Figure 9 and Table 4). The connected areas were located mainly close to the stream. This is partly in agreement with Fischer et al. (2015) who suggested that springs in the upper part of the catchment and near-stream areas are the main source areas for baseflow based on dissolved organic carbon (DOC) DOC and calcium concentrations and stable water isotope data for three snapshot sampling campaigns in the study catchment and neighboring catchments. The frequent rainstorms in summer caused connected areas to temporarily expand (Figure 7), but this increased connectivity was transient so that

hillslopes were connected to the streams for only relatively short times (Figure 8). This connection of more distant areas to the stream during events could deliver water with potentially a different hydrochemical signature and age distribution to the stream. This was the case for an agricultural catchment in Germany, where stream nitrate concentrations were high during wet periods when the source areas extended into agricultural areas (i.e., fertilization) and the flow paths were shallow and fast so that denitrification could not attenuate nitrate fluxes to the stream (Yang et al., 2018). However, when analyzing hydrochemical concentrations or fluxes, we need to consider that solute transport connectivity (Knudby & Carrera, 2005) is not equivalent to hydrological connectivity because of dispersion. In other words, if the transport connectivity is low, the solute is spread out due to a lack of fast flow paths. Information on transport connectivity is contained in the breakthrough curve of the transported hydrochemical (Knudby & Carrera, 2006).

5. Conclusions

We used a data-driven approach to obtain catchment-scale groundwater level dynamics and maps of active and connected areas. This method consists of time series clustering of shallow groundwater measurements and subsequent topography-based upscaling of the point-scale groundwater data to the catchment scale. The resulting maps of active and connected areas for individual events, different study months, and seasons showed that groundwater responded throughout the catchment but that shallow subsurface runoff source areas (i.e., those areas with a hydrological connection to the stream) expanded and contracted mainly along the channel network. This observation is in agreement with the variable source area concept for overland flow (Hewlett & Hibbert, 1963; Tsukamoto, 1963).

The upscaled (or modeled) groundwater dynamics agreed well with the measurements, particularly for upslope and footslope locations. However, the agreement was lower at the beginning of events and for midslope locations. The mean fraction of sites for which measured and modeled groundwater activation agreed depended on the assumed depth of the more transmissive soil layer (i.e., the activation threshold) but did not differ for the different rainfall event types or antecedent conditions. A sensitivity study showed that the modeled active area was not sensitive to potential misclassification of sites to a certain cluster. Therefore, the method appears to be useful and robust and likely applicable in other catchments where distinct differences in groundwater level responses can be related to site characteristics, such as topography (as in this study) or other characteristics (e.g., upslope distance or soil depth).

Acknowledgments
Special thanks to Colin Rundel, Department of Statistical Sciences, Duke University, for his valuable input on time series regression. Thanks also to Ellen Cerwinka, Jana Dusik, Benjamin Fischer, Peter Herrmann, Seraina Kauer, Nadja Lavanga, Claudia Müller, Stephan Müller, Stefan Plötner, Sandra Pool, Paribesh Pradhan, Sandra Schärer, Karl Steiner, Ivan Woodrich, and Mirjam Zehnder for field and lab assistance to obtain the groundwater level data. We also thank the Swiss Federal Institute for Forest, Snow and Landscape Research WSL for meteorological data, as well as the OAK—Gemeinde Alpthal and the Canton of Schwyz—Amt für Natur, Jagd und Fischerei, for their cooperation. The groundwater data are available on the data repository of WSL (https://www.envidat.ch/dataset/groundwater-time-series-studibach-rinderer-et-al-2019-wrr). M. R. gratefully acknowledges the financial support provided by the Swiss National Science Foundation (SNF) for the project: ICaRuS — Investigating Catchment Runoff Response by Assessing Spatial Patterns of Groundwater Dynamics.

References

Ali, G., Birkel, C., Tetzlaff, D., Soulsby, C., McDonnell, J. J., & Tarolli, P. (2014). A comparison of wetness indices for the prediction of observed connected saturated areas under contrasting conditions. *Earth Surface Processes and Landforms*, *39*(3), 399–413. https://doi.org/10.1002/esp.3506

Ali, G. A., L'Heureux, C., Roy, A. G., Turmel, M.-C., & Courchesne, F. (2011). Linking spatial patterns of perched groundwater storage and stormflow generation processes in a headwater forested catchment. *Hydrological Processes*, *25*(25), 3843–3857. https://doi.org/10.1002/hyp.8238

Ali, G. A., & Roy, A. G. (2010). Shopping for hydrologically representative connectivity metrics in a humid temperate forested catchment. *Water Resources Research*, *46*, W12544. https://doi.org/10.1029/2010WR009442

Ali, G. a., Roy, A. G., Turmel, M. C., & Courchesne, F. (2010). Source-to-stream connectivity assessment through end-member mixing analysis. *Journal of Hydrology*, *392*(3-4), 119–135. https://doi.org/10.1016/j.jhydrol.2010.07.049

Ambroise, B. (2004). Variable 'active' versus 'contributing' areas or periods: A necessary distinction. *Hydrological Processes*, *18*(6), 1149–1155. https://doi.org/10.1002/hyp.5536

Anderson, M. G., & Burt, T. P. (1978). The role of topography in controlling throughflow generation. *Earth Surface Processes and Landforms*, *3*(4), 331–344. https://doi.org/10.1002/esp.3290030402

Bachmair, S., & Weiler, M. (2012). Hillslope characteristics as controls of subsurface flow variability. *Hydrology and Earth System Sciences*, *16*(10), 3699–3715. https://doi.org/10.5194/hess-16-3699-2012

Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, *24*(1), 43–69. https://doi.org/10.1080/02626667909491834

Birkel, C., Tetzlaff, D., Dunn, S. M., & Soulsby, C. (2010). Towards a simple dynamic process conceptualization in rainfall–runoff models using multi-criteria calibration and tracers in temperate, upland catchments. *Hydrological Processes*, *275*, 260–275.

Bishop, K., Seibert, J., Nyberg, L., & Rodhe, A. (2011). Water storage in a till catchment. II: Implications of transmissivity feedback for flow paths and turnover times. *Hydrological Processes*, *25*(25), 3950–3959. https://doi.org/10.1002/hyp.8355

Blazkova, S., Beven, K., Tacheci, P., & Kulasova, A. (2002). Testing the distributed water table predictions of TOPMODEL (allowing for uncertainty in model calibration): The death of TOPMODEL? *Water Resources Research*, *38*(11), 1257. https://doi.org/10.1029/2001WR000912

Blumstock, M., Tetzlaff, D., Dick, J. J., Nuetzmann, G., & Soulsby, C. (2016). Spatial organisation of groundwater dynamics and streamflow response from different hydropedological units in a montane catchment. *Hydrological Processes*, 1–18.

Burt, T. P., & Butcher, D. P. (1985). Topographic controls of soil moisture distributions. *Journal of Soil Science*, *36*(3), 469–486. https://doi.org/10.1111/j.1365-2389.1985.tb00351.x

Burt, T. P., & McDonnell, J. J. (2015). Whither field hydrology? The need for discovery science and outrageous hydrological hypotheses. *Water Resources Research*, *51*, 5919–5928. https://doi.org/10.1002/2014WR016839

Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, *3*(1), 1–27.

Cappus, P. (1960). Bassin experimental d'Alrance—´ Etude des lois ´ de l'ecoulement—Application au calcul et ´ a la pr ` evision des ´ debits. *La Houille Blanche, A*, *60*, 493–520. https://doi.org/10.1051/lhb/1960007

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Creed, I. F., & Band, L. E. (1998). Exploring functional similarity in the export of nitrate-N from forested catchments: A mechanistic modeling approach. *Water Resources Research*, *34*(11), 3079–3093. https://doi.org/10.1029/98WR02102

Creed, I. F., Band, L. E., Foster, N. W., Morrison, I. K., Nicolson, J. A., Semkin, R. S., & Jeffries, D. S. (1996). Regulation of nitrate-N release from temperate forests: A test of the N flushing hypothesis. *Water Resources Research*, *32*(11), 3337–3354. https://doi.org/10.1029/96WR02399

Delhomme, J. P. (1978). Kriging in the hydrosciences. *Advances in Water Resources*, *1*(5), 251–266. https://doi.org/10.1016/0309-1708(78)90039-8

Desbarats, A. J., Logan, C. E., Hinton, M. J., & Sharpe, D. R. (2002). On the kriging of water table elevations using collateral information from a digital elevation model. *Journal of Hydrology*, *255*(1–4), 25–38.

Detty, J. M., & McGuire, K. J. (2010). Topographic controls on shallow groundwater dynamics: Implications of hydrologic connectivity between hillslopes and riparian zones in a till mantled catchment. *Hydrological Processes*, *24*(16), 2222–2236. https://doi.org/10.1002/hyp.7656

Dick, J. J., Tetzlaff, D., Birkel, C., & Soulsby, C. (2015). Modelling landscape controls on dissolved organic carbon sources and fluxes to streams. *Biogeochemistry*, *122*(2–3), 361–374. https://doi.org/10.1007/s10533-014-0046-3

Dunn, O. (1961). Multiple comparisons among means. *Journal of the American Statistical*, *56*(293), 52–64. https://doi.org/10.1080/01621459.1961.10482090

Dunne, T., & Black, R. (1970). Partial area contribution to storm runoff in a small New England watershed. *Water Resources Research*, *6*(5), 1296–1311. https://doi.org/10.1029/WR006i005p01296

Dunne, T., Moore, T. R., & Tayloer, C. H. (1975). Recognition and prediction of runoff producing zones in humid regions. *Hydrological Sciences Bulletin*, *20*(3), 305–327.

Engman, E. (1974). Partial area hydrology and its application to water resources. *Journal of the American Water Resources Association*, *10*(3), 512–521. https://doi.org/10.1111/j.1752-1688.1974.tb00592.x

Feyen, H., Wunderli, H., Wydler, H., & Papritz, A. (1999). A tracer experiment to study flow paths of water in a forest soil. *Journal of Hydrology*, *225*(3–4), 155–167. https://doi.org/10.1016/S0022-1694(99)00159-6

Fischer, B. M. C., Rinderer, M., Schneider, P., Ewen, T., & Seibert, J. (2015). Contributing sources to baseflow in pre-alpine headwaters using spatial snapshot sampling. *Hydrological Processes*, *29*(26), 5321–5336. https://doi.org/10.1002/hyp.10529

Glaser, B., Antonelli, M., Chini, M., Pfister, L., & Klaus, J. (2018). Technical note: Mapping surface saturation dynamics with thermal infrared imagery. *Hydrology and Earth System Sciences*, *22*(11), 5987–6003. https://doi.org/10.5194/hess-22-5987-2018

Glaser, B., Klaus, J., Frei, S., Frentress, J., Pfister, L., & Hopp, L. (2016). On the value of surface saturated area dynamics mapped with thermal infrared imagery for modeling the hillslope-riparian-stream continuum. *Water Resources Research*, *52*, 8317–8342. https://doi.org/10.1002/2015WR018414

Grayson, R. B. (2001). Toward capturing hydrologically significant connectivity in spatial patterns. *Water Resources*, *37*(1), 83–97.

Hagedorn, F., Schleppi, P., Waldner, P., & Flühler, H. (2000). Export of dissolved organic carbon and nitrogen from Gleysol dominated catchments—The significance of water flow paths. *Biogeochemistry*, *50*, 137–161.

Hewlett, J., & Hibbert, A. (1967). Factors affecting the response of small watersheds to precipitation in humid areas. In W. E. Sopper, & H. W. Lull (Eds.), *Proceedings of first International Symposium on Forest Hydrology*, (pp. 275–290). New York: Pergamon Press.

Hewlett, J. D., & Hibbert, A. R. (1963). Moisture and energy conditions within a sloping soil mass during drainage. *Journal of Geophysical Research*, *68*(4), 1081.

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

Hornberger, G. M., Bencala, K. E., & McKnight, D. M. (1994). Hydrological controls on dissolved organic carbon during snowmelt in the Snake River near Montezuma, Colorado. *Biogeochemistry*, *25*(3), 147–165. https://doi.org/10.1007/BF00024390

Iwagami, S., Tsujimura, M., Onda, Y., Shimada, J., & Tanaka, T. (2010). Role of bedrock groundwater in the rainfall-runoff process in a small headwater catchment underlain by volcanic rock. *Hydrological Processes*, *24*(19), 2771–2783. https://doi.org/10.1002/hyp.7690

Jencso, K. G., & McGlynn, B. L. (2011). Hierarchical controls on runoff generation: Topographically driven hydrologic connectivity, geology, and vegetation. *Water Resources Research*, *47*, W11527. https://doi.org/10.1029/2011WR010666

Jencso, K. G., McGlynn, B. L., Gooseff, M. N., Bencala, K. E., & Wondzell, S. M. (2010). Hillslope hydrologic connectivity controls riparian groundwater turnover: Implications of catchment structure for riparian buffering and stream water sources. *Water Resources Research*, *46*, W10524. https://doi.org/10.1029/2009WR008818

Jencso, K. G., McGlynn, B. L., Gooseff, M. N., Wondzell, S. M., Bencala, K. E., & Marshall, L. A. (2009). Hydrologic connectivity between landscapes and streams: Transferring reach- and plot-scale understanding to the catchment scale. *Water Resources Research*, *45*, W04428. https://doi.org/10.1029/2008WR007225

Klaus, J., & Jackson, C. R. (2018). Interflow is not binary: A continuous shallow perched layer does not imply continuous connectivity. *Water Resources Research*, *54*, 5921–5932. https://doi.org/10.1029/2018WR022920

Knudby, C., & Carrera, J. (2005). On the relationship between indicators of geostatistical, flow and transport connectivity. *Advances in Water Resources*, *28*(4), 405–421. https://doi.org/10.1016/j.advwatres.2004.09.001

Knudby, C., & Carrera, J. (2006). On the use of apparent hydraulic diffusivity as an indicator of connectivity. *Journal of Hydrology*, *329*(3–4), 377–389. https://doi.org/10.1016/j.jhydrol.2006.02.026

Krippendorff, K. (2004). *Content analysis, an introduction to its methodology*, (2nd ed.). Thousand Oaks: Sage Publications.

Krippendorff, K. (2011). *Computing Krippendorff's alpha-reliability*. Philadelphia: University of Pennsylvania.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, *47*(260), 583–621. https://doi.org/10.1080/01621459.1952.10483441

Kulasova, A., Beven, K. J., Blazkova, S. D., Rezacova, D., & Cajthaml, J. (2014). Comparison of saturated areas mapping methods in the Jizera Mountains, Czech Republic. *Journal of Hydrology and Hydromechanics*, *62*(2), 160–168. https://doi.org/10.2478/johh-2014-0002

Lessels, J. S., Tetzlaff, D., Birkel, C., Dick, J., & Soulsby, C. (2016). Water sources and mixing in riparian wetlands revealed by tracers and geospatial analysis. *Water Resources Research*, *52*, 456–470. https://doi.org/10.1002/2015WR017519

Loague, K., & Ebel, B. A. (2016). Finite-element modelling of physics-based hillslope hydrology, Keith Beven, and beyond. *Hydrological Processes*, *30*(14), 2432–2437. https://doi.org/10.1002/hyp.10762

Lyon, S. W., Lembo, A. J., Walter, M. T., & Steenhuis, T. S. (2006). Defining probability of saturation with indicator kriging on hard and soft data. *Advances in Water Resources*, *29*(2), 181–193. https://doi.org/10.1016/j.advwatres.2005.02.012

Lyon, S. W., Seibert, J., Lembo, A. J., Walter, M. T., & Steenhuis, T. S. (2006). Geostatistical investigation into the temporal evolution of spatial structure in a shallow water table. *Hydrology and Earth System Sciences*, *10*(1), 113–125. https://doi.org/10.5194/hess-10-113-2006

McDonnell, J. J. (1990). A rationale for old water discharge through macropores in a steep, humid catchment. *Water Resources Research*, *26*(11), 2821–2832. https://doi.org/10.1029/WR026i011p02821

McGlynn, B. L. (2004). Scale effects on headwater catchment runoff timing, flow sources, and groundwater-streamflow relations. *Water Resources Research*, *40*, W07504. https://doi.org/10.1029/2003WR002494

McGlynn, B. L., & McDonnell, J. J. (2003). Role of discrete landscape units in controlling catchment dissolved organic carbon dynamics. *Water Resources Research*, *39*(4), 1090. https://doi.org/10.1029/2002WR001525

McGlynn, B. L., & Seibert, J. (2003). Distributed assessment of contributing area and riparian buffering along stream networks. *Water Resources Research*, *39*(4), 1082. https://doi.org/10.1029/2002WR001521

McGuire, K. J. (2005). The role of topography on catchment-scale water residence time. *Water Resources Research*, *41*, W05002. https://doi.org/10.1029/2004WR003657

McNamara, J. P., Chandler, D., Seyfried, M., & Achet, S. (2005). Soil moisture states, lateral flow, and streamflow generation in a semi-arid, snowmelt-driven catchment. *Hydrological Processes*, *19*(20), 4023–4038. https://doi.org/10.1002/hyp.5869

Mohn, J., Schürmann, A., & Hagedorn, F. (2000). Increased rates of denitrification in nitrogen-treated forest soils. *Forest Ecology and Management*, *137*(1-3), 113–119. https://doi.org/10.1016/S0378-1127(99)00320-5

Mosley, M. R. (1979). Streamflow generation in a forested watershed, New Zealand. *Water Resources Research*, *15*(4), 795–806. https://doi.org/10.1029/WR015i004p00795

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Nippgen, F., McGlynn, B. L., & Emanuel, R. E. (2015). The spatial and temporal evolution of contributing areas. *Water Resources Research*, *51*, 4550–4573. https://doi.org/10.1002/2014WR016719

O'Callaghan, J. F., & Mark, D. M. (1984). The extraction of drainage networks from digital elevation data. *Computer Vision, Graphics, and Image Processing*, *28*(3), 323–344. https://doi.org/10.1016/S0734-189X(84)80011-0

Ocampo, C. J., Sivapalan, M., & Oldham, C. (2006). Hydrological connectivity of upland-riparian zones in agricultural catchments: Implications for runoff generation and nitrate transport. *Journal of Hydrology*, *331*(3–4), 643–658. https://doi.org/10.1016/j.jhydrol.2006.06.010

Peeters, L., Fasbender, D., Batelaan, O., & Dassargues, A. (2010). Bayesian data fusion for water table interpolation: Incorporating a hydrogeological conceptual model in kriging. *Water Resources Research*, *46*, W08532. https://doi.org/10.1029/2009WR008353

Penna, D., Mantese, N., Hopp, L., Borga, M., & Dalla Fontana, G. (2015). Spatio-temporal variability of piezometric response on two steep alpine hillslopes. *Hydrological Processes*, *29*(2), 198–211. https://doi.org/10.1002/hyp.10140

Pfister, L., McDonnell, J. J., Hissler, C., & Hoffmann, L. (2010). Ground-based thermal imagery as a simple, practical tool for mapping saturated area connectivity and dynamics. *Hydrological Processes*, *24*(21), 3123–3132. https://doi.org/10.1002/hyp.7840

Rinderer, M., Kollegger, A., Fischer, B. M. C., Stähli, M., & Seibert, J. (2012). Sensing with boots and trousers—Qualitative field observations of shallow soil moisture patterns. *Hydrological Processes*, *26*(26), 4112–4120. https://doi.org/10.1002/hyp.9531

Rinderer, M., McGlynn, B. L., & van Meerveld, H. J. (2017). Groundwater similarity across a watershed derived from time-warped and flow-corrected time series. *Water Resources Research*, *53*, 3921–3940. https://doi.org/10.1002/2016WR019856

Rinderer, M., van Meerveld, I., & Seibert, J. (2014). Topographic controls on shallow groundwater levels in a steep, prealpine catchment: When are the TWI assumptions valid? *Water Resources Research*, *50*, 6067–6080. https://doi.org/10.1002/2013WR015009

Rinderer, M., van Meerveld, I., Stähli, M., & Seibert, J. (2016). Is groundwater response timing in a pre-alpine catchment controlled more by topography or by rainfall? *Hydrological Processes*, *30*(7), 1036–1051. https://doi.org/10.1002/hyp.10634

Rodhe, A., & Seibert, J. (2011). Groundwater dynamics in a till hillslope: Flow directions, gradients and delay. *Hydrological Processes*, *25*(12), 1899–1909. https://doi.org/10.1002/hyp.7946

Rosenbaum, U., Bogena, H. R., Herbst, M., Huisman, J. a., Peterson, T. J., Weuthen, A., et al. (2012). Seasonal and event dynamics of spatial soil moisture patterns at the small catchment scale. *Water Resources Research*, *48*, W10544. https://doi.org/10.1029/2011WR011518

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *26*(1), 43–49. https://doi.org/10.1109/TASSP.1978.1163055

Seibert, J., & McGlynn, B. L. (2007). A new triangular multiple flow direction algorithm for computing upslope areas from gridded digital elevation models. *Water Resources Research*, *43*, W04501. https://doi.org/10.1029/2006WR005128

Seyfried, M. S., Grant, L. E., Marks, D., Winstral, A., & McNamara, J. (2009). Simulated soil water storage effects on streamflow generation in a mountainous snowmelt environment, Idaho, USA. *Hydrological Processes*, *23*(6), 858–873. https://doi.org/10.1002/hyp.7211

Shepard, D., (1968). A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 ACM National Conference*. pp. 517–524.

Sim, J., & Wright, C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy*, *85*(3), 257–268.

Sklash, M., Stewart, M., & Pearce, A. (1986). Strom runoff generation in humid headwater catchments 2. A case study of hillslope and low-order stream response. *Water Resources Research*, *22*(8), 1273–1282. https://doi.org/10.1029/WR022i008p01273

Skov, T., van den Berg, F., Tomasi, G., & Bro, R. (2006). Automated alignment of chromatographic data. *Journal of Chemometrics*, *20*(11–12), 484–497. https://doi.org/10.1002/cem.1031

Smith, T., Marshall, L., McGlynn, B., & Jencso, K. (2013). Using field data to inform and evaluate a new model of catchment hydrologic connectivity. *Water Resources Research*, *49*, 6834–6846. https://doi.org/10.1002/wrcr.20546

Soulsby, C., Birkel, C., & Tetzlaff, D. (2016). Modelling storage-driven connectivity between landscapes and riverscapes: Towards a simple framework for long-term ecohydrological assessment. *Hydrological Processes*, *30*(14), 2482–2497. https://doi.org/10.1002/hyp.10862

Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, *62*(1), 77–89. https://doi.org/10.1016/S0034-4257(97)00083-7

Stieglitz, M., Shaman, J., McNamara, J., Engel, V., Shanley, J., & Kling, G. W. (2003). An approach to understanding hydrologic connectivity on the hillslope and the implications for nutrient transport. *Global Biogeochemical Cycles*, *17*(4), 1105. https://doi.org/10.1029/2003GB002041

Tetzlaff, D., Birkel, C., Dick, J., Geris, J., & Soulsby, C. (2014). Storage dynamics in hydropedological units control hillslope connectivity, runoff generation, and the evolution of catchment transit time distributions. *Water Resources Research*, *50*, 969–985. https://doi.org/10.1002/2013WR014147

Tetzlaff, D., Seibert, J., McGuire, K. J., Laudon, H., Burns, D. A., Dunn, S. M., & Soulsby, C. (2009). How does landscape structure influence catchment transit time across different geomorphic provinces? *Hydrological Processes*, *23*(6), 945–953. https://doi.org/10.1002/hyp.7240

Tetzlaff, D., Soulsby, C., Bacon, P. J., Youngson, A. F., Gibbins, C., Malcolm, I. A., & Cooper, M. (2007). Connectivity between landscapes and riverscapes—A unifying theme in integrating hydrology and ecology in catchment science? *Hydrological Processes*, *21*(10), 1385–1389. https://doi.org/10.1002/hyp.6701

Tetzlaff, D., Soulsby, C., Waldron, S., Malcolm, I. A., Bacon, P. J., Dunn, S. M., et al. (2007). Conceptualization of runoff processes using a geographical information system and tracers in a nested mesoscale catchment. *Hydrological Processes*, *1307*, 1289–1307.

Thompson, J. J. D., Doody, D. G., Flynn, R., & Watson, C. J. (2012). Dynamics of critical source areas: Does connectivity explain chemistry? *Science of the Total Environment*, *435–436*, 499–508.

Tobler, W. R. (1970). A computer movie simulation urban growth in Detroit Region. *Economic Geography*, *46*(332), 234–240. https://doi.org/10.2307/143141

Toth, J. (1963). A theoretical analysis of ground water flow in small drainage basins. *Journal of Geophysical Research*, *68*(16), 4795–4812. https://doi.org/10.1029/JZ068i016p04795

Tromp-van Meerveld, H. J., & McDonnell, J. J. (2006). Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis. *Water Resources Research*, *42*, W02411. https://doi.org/10.1029/2004WR003800

Tsukamoto, Y. (1961). An experiment on sub-surface flow. *Journal of the Japanese Forestry Society*, *43*, 62–67.

van Meerveld, H. J., Seibert, J., & Peters, N. E. (2015). Hillslope-riparian-stream connectivity and flow directions at the Panola Mountain Research Watershed. *Hydrological Processes*, *29*(16), 3556–3574. https://doi.org/10.1002/hyp.10508

van Meerveld, H. J. I., Fischer, B. M. C., Rinderer, M., Stähli, M., & Seibert, J. (2018). Runoff generation in a pre-alpine catchment: A discussion between a tracer and a shallow groundwater hydrologist. *Cuadernos de Investigación Geográfica*, *44*(2), 452. https://doi.org/10.18172/cig.3349

Tsukamoto, Y. (1963). Storm discharge from an experimental watershed. *Journal of the Japanese Forestry Society*, *45*(6).

Uchida, T., & Asano, Y. (2010). Spatial variability in the flowpath of hillslope runoff and streamflow in a meso-scale catchment. *Hydrological Processes*, *24*(16), 2277–2286. https://doi.org/10.1002/hyp.7767

Vidon, P. G. F., & Hill, A. R. (2004). Landscape controls on nitrate removal in stream riparian zones. *Water Resources Research*, *40*, W03201. https://doi.org/10.1029/2003WR002473

Weiler, M., & McDonnell, J. (2004). Virtual experiments: A new approach for improving process conceptualization in hillslope hydrology. *Journal of Hydrology*, *285*(1–4), 3–18. https://doi.org/10.1016/S0022-1694(03)00271-3

Weiler, M., Naef, F., & Leibundgut, C. (1998). Study of runoff generation on hillslopes using tracer experiments and a physically-based numerical hillslope model. *IAHS Publication*, *248*, 353–362.

Welch, L. A. A., Allen, D. M. M., & van Meerveld, H. J. (2012). Topographic controls on deep groundwater contributions to mountain headwater streams and sensitivity to available recharge. *Canadian Water Resources Journal/Revue canadienne des ressources hydriques*, *37*(4), 349–371. https://doi.org/10.4296/cwrj2011-907

Western, A. W., Bloeschl, G., & Grayson, R. B. (1998). Geostatistical characterisation of soil moisture patterns in the Tarrawarra catchment. *Journal of Hydrology*, *205*(1-2), 20–37. https://doi.org/10.1016/S0022-1694(97)00142-X

Western, A. W., Blöschl, G., & Grayson, R. B. (2001). Toward capturing hydrologically significant connectivity in spatial patterns. *Water Resources*, *37*(1), 83–97. https://doi.org/10.1029/2000WR900241

Weyman, D. R. (1973). Measurements of downslope flow of water in a soil. *Journal of Hydrology*, *20*(3), 267–288. https://doi.org/10.1016/0022-1694(73)90065-6

Winter, T. C. (2001). The concept of hydrologic landscapes. *Journal of the American Water Resources Association*, *37*(2), 335–349. https://doi.org/10.1111/j.1752-1688.2001.tb00973.x

Yang, J., Heidbüchel, I., Musolff, A., Reinstorf, F., & Fleckenstein, J. H. (2018). Exploring the dynamics of transit times and subsurface mixing in a small agricultural catchment. *Water Resources Research*, *54*, 2317–2335. https://doi.org/10.1002/2017WR021896

Zuecco, G., Penna, D., Borga, M., & van Meerveld, H. J. (2016). A versatile index to characterize hysteresis between hydrological variables at the runoff event timescale. *Hydrological Processes*, *30*(9), 1449–1466. https://doi.org/10.1002/hyp.10681

Zuecco, G., Rinderer, M., Penna, D., Borga, M., & van Meerveld, H. J. (2018). Quantification of subsurface hydrologic connectivity in four headwater catchments using graph theory. *The Science of the Total Environment*, *646*, 1265–1280.