

Water Resources Research

RESEARCH ARTICLE

10.1029/2018WR023736

Key Points:

- Soil and macropore properties predominantly control storage-discharge relationships in shale and sandstone catchments
- Dynamic water storage increases with catchment size because hillslope-stream connectivity increases in larger catchments
- Swap experiments can assess the impacts of topography and catchment size that a sensitivity analysis cannot

Correspondence to:

L. Li,
lili@engr.psu.edu

Citation:

Xiao, D., Shi, Y., Brantley, S. L., Forsythe, B., DiBiase, R., Davis, K., & Li, L. (2019). Streamflow generation from catchments of contrasting lithologies: The role of soil properties, topography, and catchment size. *Water Resources Research*, 55, 9234–9257. <https://doi.org/10.1029/2018WR023736>

Received 19 JUL 2018

Accepted 10 SEP 2019

Accepted article online 16 SEP 2019

Published online 19 NOV 2019

Streamflow Generation From Catchments of Contrasting Lithologies: The Role of Soil Properties, Topography, and Catchment Size

Dacheng Xiao¹ , Yuning Shi² , Susan L. Brantley^{3,4} , Brandon Forsythe⁴ , Roman DiBiase^{3,4} , Kenneth Davis^{4,5} , and Li Li¹ 

¹Department of Civil and Environmental Engineering, The Pennsylvania State University, University Park, PA, USA,

²Department of Ecosystem Science and Management, The Pennsylvania State University, University Park, PA, USA,

³Department of Geosciences, The Pennsylvania State University, University Park, PA, USA, ⁴Earth and Environmental Systems Institute, The Pennsylvania State University, University Park, PA, USA, ⁵Department of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, PA, USA

Abstract Understanding streamflow generation and its dependence on catchment characteristics requires large spatial data sets and is often limited by convoluted effects of multiple variables. Here we address this knowledge gap using data-informed, physics-based hydrologic modeling in two catchments with similar vegetation and climate but different lithology (Shale Hills [SH], shale, 0.08 km², and Garner Run [GR], sandstone, 1.34 km²), which influences catchment topography and soil properties. The sandstone catchment, GR, is characterized by lower drainage density, extensive valley fill, and bouldery soils. We tested the hypothesis that *the influence of topographic characteristics is more significant than that of soil properties and catchment size*. Transferring calibration coefficients from the previously calibrated SH model to GR cannot reproduce monthly discharge until after incorporating measured boulder distribution at GR. Model calibration underscored the importance of soil properties (porosity, van Genuchten parameters, and boulder characteristics) in reproducing daily discharge. Virtual experiments were used to swap topography, soil properties, and catchment size one at a time to disentangle their influence. They showed that clayey SH soils led to high nonlinearity and threshold behavior. With the same soil and topography, changing from SH to GR size consistently increased dynamic water storage (S_d) from ~0.12 to ~0.17 m. All analyses accentuated the predominant control of soil properties, therefore rejecting the hypothesis. The results illustrate the use of physics-based modeling for illuminating mechanisms and underscore the importance of subsurface characterization as we move toward hydrological prediction in ungauged basins.

1. Introduction

Forecasting streamflow and extreme hydrological events (e.g., flooding and droughts) is essential for our society as the pace of climate change accelerates (Hrachowitz et al., 2013; Montanari et al., 2013; Sivapalan, 2003; Vorosmarty et al., 2010). Although forecasting capability has progressed significantly in the recent decades with rapid data accumulation and model development, hydrological prediction in ungauged basins (PUB) remains a grand challenge (Hrachowitz et al., 2013). Model Transferability tests directly from gauged to ungauged watersheds have yielded mixed results (Fenicia et al., 2016; Heuvelmans et al., 2004; Li et al., 2012; Smith, Hayes, et al., 2016; van der Linden & Woo, 2003), underscoring challenges in understanding how and how much catchment characteristics (e.g., lithology, land cover, topography, and size) influence streamflow generation.

Studies on streamflow generation have revolved around the dynamics of connectivity, storage (S) and discharge (Q) relationships, and threshold behaviors that emerge at the catchment scale. Storage-discharge relationships have been explored since the 1930s (Horton, 1936) and are generally recognized as highly nonlinear, often taking a power law form (Wittenberg, 1999). Water storage is often conceptualized as partitioning between distinct reservoirs: a small, active reservoir that rapidly responds to hydroclimatic forcing (dynamic storage S_d) and a passive reservoir that is characterized by longer residence times before water reemerges in streams and rivers (Dunn et al., 2010; Kobierska et al., 2015; McNamara et al., 2011; van der Velde et al., 2015). Transit time estimates using water isotopes or tracers have shown that S_d can be more than an order of magnitude smaller than the total storage (S_t) inferred from soil

©2019. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

porosity and depth (Birkel et al., 2011; Bishop et al., 2011; McNamara et al., 2011). Such dramatic reduction in S_d has been attributed to limited hydrologic connectivity between hillslopes and streams (Bracken et al., 2013; Jencso et al., 2010; McGlynn & McDonnell, 2003; Wlostowski et al., 2016), large water storage, and the existence of thresholds that must be surpassed to initiate hydrological response (Ali et al., 2015; James & Roulet, 2007; Lehmann et al., 2007; Seibert et al., 2011; Tromp-van Meerveld & McDonnell, 2006).

Streamflow generation is often influenced by a multitude of competing factors including external forcings (e.g., climate) and internal structure characteristics (vegetation, land use, topography, and lithology). Among these, the role of lithology and topographic characteristics has been extensively studied. Brantley et al. (2017) argued that weathering affects subsurface permeability, water infiltration, and flow partitioning between shallow and deep subsurface. Kuentz et al. (2017) concluded that the base flow index correlates more strongly with lithology whereas topography primarily controls streamflow flashiness. Hale and McDonnell (2016) observed that catchments derived from permeable sandstone bedrock have longer mean transit times than those with tight volcanic bedrock (Hale & McDonnell, 2016). Pfister et al. (2017) showed that catchment storage decreases whereas streamflow flashiness increases as the percentage of impermeable bedrock increases (Pfister et al., 2017). On the other hand, topographic characteristics, including relief and riparian versus hillslope areas, have long been demonstrated as governing streamflow. Catchment topography and topology have been observed to correlate to hillslope-stream connectivity and runoff source area, therefore exerting a first-order control on streamflow (Jencso et al., 2009). Mean residence time has been shown to correlate strongly with topographic characteristics that determine flow path distance and gradients (McGuire et al., 2005).

In addition to lithologic and topographic properties, soil properties generally regulate water holding and storage capacity (Van Genuchten, 1980), base flow indexes (Zimmer & Gannon, 2018), and streamflow dynamics (Zimmer & McGlynn, 2017). Shi et al. (2015) showed that soil moisture is controlled primarily by soil properties and secondarily by topography and depth to bedrock. Fenicia et al. (2016) found that models with geology-based hydrological response units (HRUs) are more robust in reproducing spatial variations in streamflow compared to those using topography-based HRUs. This observation supports the idea that geology and soil properties exert a stronger control than topography. Topography and soil properties can also drive runoff generation differently under different conditions. In the Tenderfoot watershed in Montana, runoff is topographically driven with lateral redistribution of water and hydrologic connectivity under wet conditions (Jencso et al., 2009); under dry conditions, the influence of geologic controls is more pronounced (Payn et al., 2009).

Streamflow generation also varies with catchment size (Pilgrim et al., 1982). For example, in wet catchments at Maimai, New Zealand, with relatively similar geology, topography, and soil depths, riparian zone groundwater levels and runoff correlate strongly in small headwater catchments but not in large catchments (McGlynn et al., 2004). In nested headwater catchments in New York, USA, estimated event-water contribution during intensive storms inversely correlate to catchment size (Brown et al., 1999). Those authors argued that shallow subsurface flow contributes to summer stormflow substantially in small catchments but not in large catchments. A geomorphology-based model of runoff routing was used to argue that small catchment response is governed primarily by hillslope processes whereas large catchment response is governed by the structure of the stream network (Robinson et al., 1995).

The above-mentioned examples highlight the challenges of disentangling the effects of interdependent controlling variables. While general relationships may exist, their relative influence on streamflow generation remains equivocal. Often, catchments are carefully chosen to highlight the role of one variable (e.g., topography, soil properties, and size) while keeping others relatively similar. Nonetheless, large differences in natural catchments inevitably exist. An additional challenge is that streamflow generation is a catchment-scale emergent behavior and investigating its mechanisms often requires intensive spatial measurements of soil moisture and groundwater levels. Previous field studies on connectivity involve >147 storm analyses (Tromp-van Meerveld & McDonnell, 2006), 13 spatial distributions of soil moisture each with >500 measurements (Western et al., 1999), water table analyses in 84 recording wells distributed across 24 transects (Jencso et al., 2009), and ~30 soil moisture monitoring sites (Lin et al., 2006). One way to reduce and circumvent these challenges is to compare catchments of distinctive characteristics

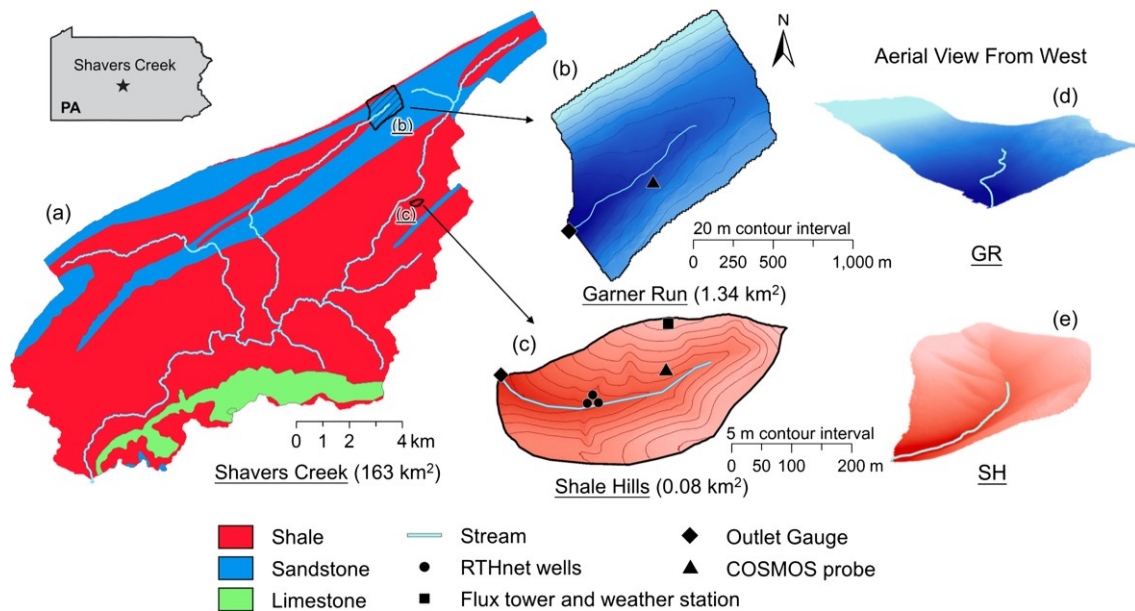


Figure 1. A simplified geological map of Shavers Creek watershed (the Susquehanna Shale Hills Critical Zone Observatory), showing the location, dominant lithology, and topography (elevation contours) for Shale Hills (SH) and Garner Run (GR) catchments. The catchments experience similar land use, climate, and tectonic histories but differ in lithology. These lithological differences have resulted in large differences in terms of relief, catchment size, drainage density, riparian/hillslope ratios, and so on. Both today's climate and past periglacial conditions have influenced the underlying aquifer and surficial soil properties.

in controlled virtual experiments while changing only one variable at a time. In that way, the role of structural variables can be distinguished at the same time that general, emergent patterns can be highlighted. Such an approach could be more effective than focusing on the “idiosyncrasies of yet another experimental catchment” (McDonnell, 2003).

Here we combined data and model approaches to understand the relative influence of soil properties, topography, and catchment size on streamflow generation in two first-order, monolithological catchments in central Pennsylvania experiencing the same climate (temperate) and land use (forestland). One is the shale-underlain Shale Hills (SH, 0.08 km²); the other is the sandstone-based Garner Run (GR, 1.34 km²). The two catchments derive from different parent materials and differ in topography (relief, size of riparian zones, and slope), soil and macropore properties, and catchment size, all of which emerge from landscape evolution and ultimately depend on lithology (Reinhardt & Ellis, 2015). Streamflow monitoring data show that GR discharge is less flashy than SH (Hoagland et al., 2017; Shi et al., 2013). Given that GR has more permeable, bouldery sandy soils that often lead to flashy discharge, we hypothesize that streamflow response and storage-discharge relationships are more affected by topography than by soil properties and catchment size. In other words, we hypothesize that the influence of topographic characteristics (a flatter slope, longer slope length, larger riparian zone) is more significant than that of soil properties and catchment size, leading to a dampened streamflow response and a linear S - Q relationship at GR compared to SH.

This hypothesis was tested using three analyses. The first was a model transferability test, treating GR as if it was a new, ungauged catchment to which we transferred model calibration information from the previously modeled SH using the physically based, spatially distributed land surface hydrologic model Flux-Penn State Integrated Hydrologic Model (PIHM; Shi et al., 2013). The data used in GR were restricted to those that would be typically available for ungauged catchments. In this way, we explored the idea of transferring calibrated model parameters from one measured catchment to a nearby, ungauged one. If the hypothesis is true, the model calibration from SH should be directly transferable and should produce satisfactory discharge because topography is explicitly represented in Flux-PIHM. The second analysis included model calibration and sensitivity analysis that identified key process parameters. If the hypothesis is true, the soil property parameters should have relatively small influence on reproducing discharge and soil moisture data. Third,

Table 1
Physiographic and Hydrologic Characteristics of the Catchments

Characteristics		Shale Hills	Garner Run	Data sources
Area	km ²	0.08	1.34	NHD
Elevation	m	250–300	450–650	NED 3 m
Average slope	N-facing (°)	21	12	
	S-facing (°)	14	17	
Max slope length	N-facing (m)	115	357	
	S-facing (m)	126	670	
Riparian zone ^a	% of catchment area	10.6	34.7	NLDAS-2
Precipitation	Mean annual (mm)	1,000	1,000	
Lithology		Rose Hill gray shale	Tuscarora sandstone	Brantley et al. (2016)
Bedrock depth	m	0.25–1.87 ^b	0.58–2.42 ^c	Lin et al. (2006) and Brantley et al. (2016)
Soil	Sand (%)	29	48	SSURGO and Jin et al. (2010)
	Silt (%)	32	34	
	Clay (%)	39	18	
Land cover	Deciduous forest (%)	87	87	NLCD
	Others (%)	13	13	MODIS
	Max LAI in 2015	6.4	5.4	

Note. LAI = leaf area index; NED = National Elevation Dataset; NHD = National Hydrography Dataset; NLCD = National Land Cover Database; NLDAS-2 = North American Land Data Assimilation Systems phase 2.

^aRiparian zone was defined as those with simulated saturated water storage higher than the threshold indicator value (TIV) in August 2015, the driest month of the year. ^bFor SH, the hand-augerable soil depth is smaller than the water table depth in the valley. As a result, an additional 1.2 m were added in the model to represent the top layer of weathered bedrock that interacts with the stream. (Shi et al., 2013). This results in soil depths from 1.45 to 3.07 m in the SH simulation with an area-weighted average of 1.7 m. ^cFor GR, soil depth is based on ground-penetrating radar (GPR) measurements.

swap experiments were carried out where catchment characteristics were swapped one at a time, to tease apart the relative influence of topography, soil properties, and size. These virtual experiments assessed the influence of topography compared to soil properties and catchment size. The hypothesis suggests that the largest differences are expected for virtual catchments with different topography. The mechanistic understanding gleaned from these virtual experiments can facilitate the development of parsimonious models, reproduce streamflow dynamics, and advance our forecasting capabilities in PUB (Sivapalan et al., 2003).

2. Study Sites and Data

2.1. Study Sites

SH and GR are first-order catchments that are nested within the Susquehanna Shale Hills Critical Zone Observatory (SSHCZO). The CZO consists of the entire Shavers Creek watershed, a hydrologic unit code 10 watershed (Figure 1 and Table 1). The streams in SH and GR are intermittent and perennial, respectively. SH is steeper (mean hillslope angle = 14–21°) than GR (mean hillslope angle = 12–17°), and GR has a much larger riparian zone with an extensive valley fill up to 300 m wide (~30% of the catchment width), compared to about 25 m (10% of catchment width) at SH. Both catchments are temperate forests dominated by deciduous broadleaf species with a small component of evergreen conifers and understory shrubs (Brubaker et al., 2018; Smith, Eissenstat, et al., 2016). SH generally has denser vegetation cover (60% more biomass and 19% larger maximum leaf area index [LAI]) than GR. Aspect plays a very different role between the catchments. While there is an overall similar pattern across slope position, both sites have higher biomass values at the valley floor compared to the ridge top position (Brubaker et al., 2018). Colocated measurements have shown that soil moisture covaries with leaf production and LAI, suggesting that trees influence soil moisture across space and time (Naithani et al., 2013). Detailed, high-resolution measurements of temporal LAI data however are not available such that we assume uniform LAI in the simulations.

SH is nearly 100% situated on the Rose Hill Shale Formation (Brantley et al., 2016). The Weikert soil dominates hillslopes while the Berks, Blairton, Rushtown, and Ernest soils prevail in swales and valley floor (Lin et al., 2006). Given that these soils are formed on the same rock in the same climate and with the same vegetation, the difference in soil types is largely a function of landscape position (soils in valley are wetter because

water drains downhill and organic matter accumulates in the valley) and aspect (one side is sunnier than the other). GR overlies the Tuscarora Formation (Flueckinger, 1969), which consists of almost pure quartzitic sandstone with minor interbedded shales (Hettinger, 2001). The Natural Resources Conservation Service Soil Survey Geographic Database (SSURGO; <http://www.nrcs.usda.gov/wps/portal/nrcs/main/soils/survey/>) and CZO observations (Brantley et al., 2016; Del Vecchio et al., 2018; Hoagland et al., 2017) show that GR bouldery soil has higher sand content and less silt and clay content compared to SH, which is consistent with higher hydraulic conductivity, larger pore size and porosity in GR. The residence times of the soils at the two sites differ because of differing erodibility and reactivity of the parent rock (Brantley et al., 2018; Del Vecchio et al., 2018; Li et al., 2018). Jin et al. (2010) reported that the SH soil has 38 wt.% (ridge top) to 64 wt.% (valley floor) clay (summation of illite, chlorite, and kaolinite; Table 2 in Jin et al., 2010). In contrast, the clay content in GR soils varies from 8% to 32%.

2.2. Data

The SH catchment was established in the 1970s (Lynch, 1976) and expanded as the SSHCZO (Anderson et al., 2008; Brantley et al., 2007; Qu & Duffy, 2007). It has been extensively measured for soil properties, discharge, soil moisture, water table depth, tree population, and stream chemistry, among other attributes (Jin et al., 2010; Kuntz et al., 2011; Lin & Zhou, 2008; Ma et al., 2015; Zhu et al., 2010). The GR site has been established relatively recent as part of the SSHCZO (Brantley et al., 2016).

Given that we wanted to treat GR as if it was a new, ungauged catchment to which we transfer model parameters and knowledge from the well-studied SH, the model setup was mostly based on satellite and national data that were available even for ungauged river basins, allowing a test of transferability. These included data from the U.S. Geological Survey (USGS), National Elevation Dataset (NED, elevation map, <https://lta.cr.usgs.gov/NED>), the National Land Cover Database (NLCD, vegetation map, <https://www.mrlc.gov/nlcd2011.php>), the National Hydrography Dataset (hydrographic data, <https://www.usgs.gov/core-science-systems/ngp/national-hydrography/national-hydrography-dataset>), the North American Land Data Assimilation Systems Phase 2 (hourly meteorological forcing, <https://ldas.gsfc.nasa.gov/nldas/NLDAS2forcing.php>), and the Moderate Resolution Imaging Spectroradiometer (MODIS, LAI every 8 days, <https://modis.gsfc.nasa.gov/data/dataproduct/mod15.php>). For precipitation, although we had local measurements at SH, we did not have those measurements at GR. To compare the response of SH and GR to the same precipitation forcing, we used the same North American Land Data Assimilation Systems Phase 2 precipitation forcing data for the two catchments. The discharge was measured using stream stage (HOBO Pressure Transducer, Onset Computing) and a Parshall. The averaged surface soil moisture was measured using the cosmic-ray soil moisture observing system (COSMOS) that counts neutron intensities at the vicinity of the ground surface. Discharge and COSMOS data in 2015 were used as model constraints. The averaged model output of the topsoil moisture (<10 cm) within the effective footprint radius of ~300 m were compared to the COSMOS data. The COSMOS measurements are affected by water content in soil organic matter (e.g., O horizons) and vegetation. The top 10 cm was chosen to be consistent with the calculation that ~86% of the fast neutron counts was found to within the top 10 cm of the soil (Zreda et al., 2008).

At GR, boulders cover a fraction of the hillslopes; in contrast, at SH, rock fragments are observed to emerge only near trees or tree throws. The spatial distributions of boulders at GR were mapped at a resolution of 5 m and were grouped into four categories of volume fractions (<0.10, 0.10–0.67, >0.67 with trees, and >0.67 without trees; Del Vecchio et al., 2018). At GR, depth to refusal with respect to hand augering (midslope and ridge) or depth to inferred bedrock (excavated with a jack hammer in the valley floor pit) ranges from 0.58 to 2.42 m with an average of 1.8 m

3. The Hydrology Model Flux-PIHM

3.1. Model Processes

Flux-PIHM is a physically based, spatially distributed model (Shi et al., 2013) that couples a land surface scheme adapted from the Noah Land Surface Model with the PIHM (Qu & Duffy, 2007). In addition, there is a family of PIHM-related codes with different simulation capabilities (Duffy et al., 2014), including landscape evolution (Zhang et al., 2016), ecosystem biogeochemistry (Shi et al., 2018), and catchment-scale reactive transport (Bao et al., 2017; Li, 2019; Zhi et al., 2019). The code discretizes the land surface into triangular elements and rivers into rectangular segments that are projected vertically down to the bedrock to generate

prismatic volumes. Simulated hydrological processes include precipitation, canopy interception, evapotranspiration, channel flow, overland flow, infiltration, recharge from the unsaturated to saturated zone, lateral flow connecting hillslope to the stream, and snow melt. The model outputs include spatial and temporal distribution of water content, from which water fluxes and budgets can be calculated.

Flux-PIHM assumes a no-flow boundary corresponding to the soil-weathered rock interface (Qu & Duffy, 2007; Shi et al., 2013). It therefore does not take into account recharge vertically from soil into regional groundwater in deep aquifers (Brantley et al., 2013). Such vertical flow is typically small given the sharp, often orders of magnitude, permeability contrast at the soil-weathered rock interface (Kuntz et al., 2011; Welch & Allen, 2014) and is typically ignored when solving water balance for a catchment (Kirchner, 2009). At GR, the valley is filled with high-permeability colluvial materials up to 9 m thick such that water can be lost to the deeper aquifer (Schaller & Fan, 2009). Water is also lost in the subsurface of SH, but this may be smaller in SH, as it does not have such high-permeability fluvial materials. Given that Flux-PIHM does not include this potential water loss into the deep aquifer, the code essentially lumps this water loss into evapotranspiration (ET) to conserve water balance when stream discharge and soil moisture are used for calibration. Here we note the calculated ET as ET^* , which includes potential recharge into the deep aquifer.

3.2. Model Setup

Catchment characteristics include topography (e.g., surface elevation), vegetation properties (e.g., land cover type, rooting depth, LAI, and stomatal conductance), and soil properties. All vegetation parameters were from previous work (Shi et al., 2013). The LAI values from MODIS at a frequency of 8 days were used as external forcing. A uniform LAI value was used for the whole watershed at the spatial resolution of MODIS (1 km²). Soil properties include matrix properties such as depth, hydraulic conductivity, porosity, and van Genuchten parameters (α , n , θ_s , and θ_r) and macropore properties such as depth and conductivities (horizontal and vertical) that reflect boulder characteristics. The code read elevation maps from the USGS NED and land cover maps from the NLCD. These data set up the domain of virtual catchments such that they represented the topography and land cover of the real catchments. Initial and boundary conditions included subsurface characteristics (e.g., soil depth), initial water distribution, water table, snow cover, canopy storage, and watershed boundary fluxes. Soil matrix properties were from the SSURGO data initially, and macropore properties were based on soil survey data, as will be discussed later.

3.3. Calibration

Like any spatially explicit model, Flux-PIHM is parameter intensive. There are 8 soil types in GR from the SSURGO database and 13 distributed parameters for each soil type, including 3 macropore parameters. If each parameter was calibrated independently in Flux-PIHM, a total of 104 soil parameters is needed. This overparameterization was circumvented by using parameter regularization that utilized spatially distributed a priori soil parameters and a single global calibration coefficient (GCC) for each parameter, a common practice in physically based hydrologic modeling (Smith et al., 2004).

4. Measures of Emergent Dynamics at the Catchment Scale

4.1. Storage-Discharge (S - Q) Relationship

Streamflow generation is known to depend largely on S - Q relationships. Based on the rainfall-runoff equation, Kirchner (2009) developed the following linking streamflow to storage:

$$Q \propto Q_{\text{ref}} \left(\frac{S - S_{\min}}{S} \right)^{1/b-2} \quad (1)$$

Under the condition $b < 2$, Q_{ref} is a reference discharge based on the best fit to the S - Q relationship; S is the water storage calculated from model output averaged across the whole catchment; S_d (or originally k_1 in Kirchner, 2009) represents a scaling constant that has the dimension of storage; $S_{\min}$ (or originally S_0 in Kirchner, 2009) is the storage when discharge drops to zero; and b is the exponential term in $\ln \frac{dQ}{dt} \propto -aQ^b$ (Brutsaert & Nieber, 1977), representing the S - Q nonlinearity. Water storage is often considered as

composed of “dynamic” storage that responds rapidly to hydrological events and “passive” storage that is often deeper and takes longer time for water parcels to route through (Hrachowitz et al., 2016). Summation of the two is the total storage. In this work, we consider S_d in equation (1) as the dynamic storage. It was calculated as the difference in the modeled water storage at the maximum and minimum discharges, as will be shown later in the S - Q figures in section 6. The total storage S_t was calculated as the total pore volume per land surface area (soil porosity $\times$ depth/land surface area). By fixing $S_{\min}$ and S_d , values of Q_{ref} and b were obtained by fitting the calculated Q and S relationship from the calibrated Flux-PIHM with equation (1) using the trust-region algorithm and “Curve Fitting” tool in MATLAB 2016a.

4.2. Connectivity

Hydrologic connectivity at the catchment scale is transient and depends on the spatial distribution of water content. Here we calculated hydrologic connectivity based on the model output of saturated water storage and flow connecting the uphill to the stream. A MATLAB code was developed based on approaches in literature (Allard, 1994; Western et al., 2001). The connectivity function $\tau(h)$ is defined as the probability (P) of two grid blocks being connected at a separation distance of h (in Euclidean distance):

$$\tau(h) = P(\delta x \leftrightarrow x + h | x \in A; x + h \in G); \quad (2)$$

where x and $x + h$ represent the locations of two grid blocks; G is the set of grid blocks in the domain; A is the subset of G with saturated water storage higher than a threshold indicator value (TIV), defined as the 75th percentile of saturated storage over the whole year and the whole catchment (James & Roulet, 2007; Western et al., 2001); and the “ $\leftrightarrow$ ” sign indicates two grid blocks (at x in set A and at $x + h$ in set G) are connected if there is a continuous path of neighboring grid blocks between them with indicator values larger than the TIV.

The approach involves three main steps: (1) Identify grid blocks with saturated water storage higher than TIV; (2) use a recursive algorithm to identify and label continuous paths that consist of grid blocks with saturated water storage higher than TIV and are connected to the stream; and (3) calculate the integral connectivity scale (I_{cs}) by looping through all grid blocks as follows:

$$I_{cs} = \int_0^\infty \tau(h) h dh; \quad (3)$$

The I_{cs} has the units of length and can be interpreted as the average length of flow paths from the uphill saturated zone to the stream. The relative connectivity I_{cs}/W , where W is the average width of the catchment in the direction perpendicular to the stream, quantifies the proportional flow path length connected to the stream.

5. Data Model Analysis

Three lines of analysis were carried out and are described in subsequent sections. The first was the direct parameter transferability test from the calibrated modeled SH to GR. This assessed the degree to which discharge and soil moisture data can be reproduced in GR directly using the SH calibration information. The second was the calibration and sensitivity analysis that identified key parameters to reproduce data at GR. The third was the swap experiments that swapped characteristics (soil properties, topography, and size) of SH and GR one at a time, in an effort to quantify the relative significance of catchment characteristics on water storage, connectivity, and S - Q relationships.

5.1. Parameter Transferability Test

Flux-PIHM has a large number of parameters and therefore requires time- and labor-intensive model calibration. Here we transferred model calibration coefficients from the calibrated SH model to GR without any calibration. We considered two soil data sets for SH: the data from the SSURGO database and the data from the soil survey in the field (Lin et al., 2006). The requisite parameters for the GR soil matrix (e.g., van Genuchten parameters) were only available from the SSURGO soil database. To take into account the differences in the soil data, three transferability tests were carried out to assess different approaches of parameter transfer. In the “SSURGO” case, both catchments were parameterized using the SSURGO soil database. The GCC values for SH, calibrated using SSURGO soil parameters and discharge data (Shi et al., 2015), were directly

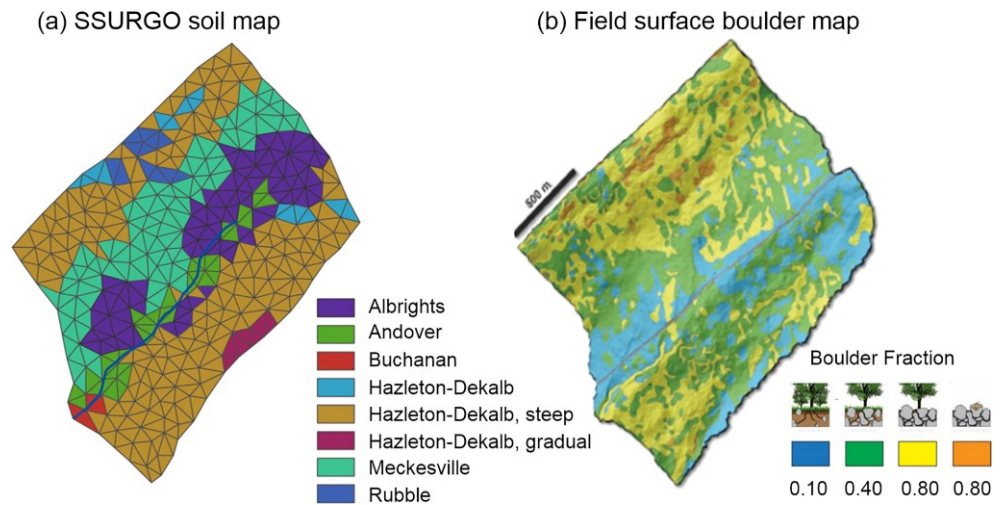


Figure 2. (a) Soil map from the Soil Survey Geographic Database (SSURGO) for Garner Run and (b) field-surveyed boulder map showing the fraction of the land surface mantled by boulders (Del Vecchio et al., 2018). The macropore fraction in Flux-PIHM was set based on the boulder fraction: 0.10, 0.40, 0.80 (with trees), and 0.80 (without trees) from blue to orange in the figure.

transferred to the GR model that used SSURGO soil parameters. In the “Scaled” case, the SH model parameters were calibrated by discharge and COSMOS data using the a priori field soil survey. For GR, we rescaled the SH GCC values to reconcile the average parameter differences between the soil survey and

SSURGO soil parameters by using the scaling equation $\chi_{GR, \text{scaled}} = \frac{\chi_{SH, \text{survey}}}{\chi_{SH, \text{SSURGO}}} \times \chi_{GR, \text{SSURGO}}$, where

$\frac{\chi_{SH, \text{survey}}}{\chi_{SH, \text{SSURGO}}}$ is the factor that scales a priori SSURGO soil parameters to corresponding soil survey parameters. Here $\chi_{SH, \text{survey}}$ and $\chi_{SH, \text{SSURGO}}$ are the area-weighted average parameters from the soil survey and from SSURGO, respectively. The assumption was that the scaling factor between SSURGO and field survey data was the same for the two catchments. However, we did observe a high land surface area covered by boulders at GR. Therefore, a third “Scaled + Boulder Map” case was set up utilizing the GCC values in the Scaled case but additionally included the surveyed boulder map for GR (Del Vecchio et al., 2018). The SSURGO and boulder maps (Figure 2) shared similar characteristics and distribution patterns. The boulder map however was more detailed with higher resolution. The boulders included clasts with large grain size that were characterized by cracks between rock and soil and macropores. In Flux-PIHM, the boulder content was represented by setting a distribution of macropore volume fraction and depth based on the measured boulder map. The boulder survey was attained from intensive field work and was thus an unusual data set that would generally not be available without intensive mapping.

5.2. Uncertainty-Based Calibration

Unsaturated water dynamics were described by the van Genuchten equation: $\theta(h) = \theta_r + \frac{\theta_s - \theta_r}{1 + |\alpha|h^n}^{\frac{1}{1-n}}$, where $\theta(h)$ is the water content ($[L^3 L^{-3}]$); $|h|$ is the water pressure (L or water head); θ_s and θ_r are saturated and residual water content ($[L^3 L^{-3}]$), respectively, with θ_s the same as porosity; and α and n are van Genuchten parameters describing the shape of the water retention curve. Previous studies in SH (Shi et al., 2014; Yu et al., 2013; Yu et al., 2014) identified six parameters that were most sensitive: porosity, van Genuchten parameters (α and n), macropore depth (D_{mac} , [L]), and conductivities (horizontal K_{macH} and vertical K_{macV} , $[L T^{-1}]$). Here we calibrated these parameters using stream discharge and COSMOS data. The model was calibrated under two scenarios: one based directly on the SSURGO soil map (GR-without boulder) and the other based on SSURGO combined with the measured boulder map (GR-with boulder). The comparison of the two cases assesses the importance of including the boulder map in calibration.

Table 2
Averaged Soil Parameters at SH and GR After Calibration

	Porosity (m^3/m^3)	α (m^{-1})	n (–)	D_{mac} (m)	K_{macH} (m/s)	K_{macV} ($\times 10^{-4}$ m/s)
SH	0.29	11.43	1.29	1.12	0.07	1.79
GR-without boulder	0.34 (± 0.05)	5.40 (± 2.16)	1.59 (± 0.13)	0.84 (± 0.54)	0.04 (± 0.08)	2.29 (± 2.66)
GR-with boulder	0.35 (± 0.04)	4.98 (± 2.36)	1.63 (± 0.15)	0.47 (± 0.32)	0.06 (± 0.09)	1.75 (± 2.54)
GR-average	0.35	5.19	1.61	0.65	0.05	2.02

Note. Area-averaged soil depths are 1.7 and 1.8 m for SH and GR, respectively. The soil parameters are porosity, van Genuchten parameters (α and n), which describe the shape of the water retention curve, Macropore parameters: include macropore depth (D_{mac}), and conductivities (horizontal K_{macH} and vertical K_{macV}).

The Hornberger-Spear-Young approach (Hornberger & Spear, 1981) was used to calibrate the model. GCC values were sampled using the Latin hypercube sampling method (McKay et al., 1979) for 500 simulations for the two scenarios (GR-with boulder and GR-without boulder) for a total of 1,000 simulations. The daily Nash-Sutcliffe efficiency (NSE; Nash & Sutcliffe, 1970), percent bias (PBIAS; Gupta et al., 1999), and RMSE-observations standard deviation ratio (RSR; Singh et al., 2005) were used for model performance evaluation using the satisfactory range of $\text{NSE} > 0.5$, $\text{RSR} \leq 0.7$, and $|\text{PBIAS}| < 25\%$ (Moriassi et al., 2007). The mean and uncertainty of each parameter were calculated using cases within “acceptable” statistical criteria and were normalized for comparison using $\frac{\chi - \chi_{\min}}{\chi_{\max} - \chi_{\min}}$, where χ represents the value of each parameter and $\chi_{\max}$ and $\chi_{\min}$ are the maximum and minimum values of model parameters.

5.3. Swap Experiments

The model used the digital elevation data from the USGS NED to define the size and topography. The same homogeneous vegetation type (deciduous) and LAI values (time dependent from MODIS) were used for SH and GR to eliminate the influence of vegetation. The soil property and land cover parameters were then distributed for each grid. In this way, the code built digital catchments that represented the spatial structure of real catchments.

SH and GR differ in three variables: soil properties, topography, and size. Although vegetation types, total biomass, and LAI also differ in these two catchments (Brubaker et al., 2018), the magnitude of these differences are generally smaller compared to the three variables (Table 1). As discussed earlier, their interdependent effects present a major barrier for quantifying the relative importance of individual variables. Although parameter sensitivity analysis is often used to assess soil property parameters, it cannot evaluate effects of topography and catchment size as they are typically held constant. The swap experiments here aim to circumvent such limitations. For each swap experiment (Table 2), we first picked the topography of one catchments to set up the simulation domain. The size can be reduced or expanded by scaling the elevation and area of each grid proportionally so that topography features such as slope gradients, proportional slope length, and area of riparian zone were preserved. The averaged soil property parameters, including porosity, van Genuchten parameters, and macropore parameters, were then assigned uniformly to each grid block based on calibrated parameter sets from the base case.

As an example, in setting up the simulation labeled $\text{Soil}_{\text{SH}}\text{Tp}_{\text{GR}}\text{Size}_{\text{SH}}$, we took the GR digital elevation map (Tp_{GR}) but reduced the elevation and length of each grid proportionally to the size of SH to maintain the GR topography. The averaged soil properties of SH were then assigned uniformly across the domain (Soil_{SH}). Similarly, $\text{Soil}_{\text{GR}}\text{Tp}_{\text{GR}}\text{Size}_{\text{SH}}$ had the topography of GR and the size of SH but averaged GR soil properties across the domain. The comparison between these two cases quantified the effects of soil properties because this was the only catchment variable that differed between the two cases. Similarly, the difference between $\text{Soil}_{\text{SH}}\text{Tp}_{\text{GR}}\text{Size}_{\text{SH}}$ and $\text{Soil}_{\text{SH}}\text{Tp}_{\text{SH}}\text{Size}_{\text{SH}}$ quantified the topography influence. Note that the $\text{Soil}_{\text{SH}}\text{Tp}_{\text{SH}}\text{Size}_{\text{SH}}$ case used averaged, uniformly assigned soil properties and therefore differed from the calibrated SH real case (SH) where soil properties (and therefore parameters) were spatially heterogeneous across the catchment. It also differed from the hypothetical GR- Size_{SH} that preserved the heterogeneous soil distribution and topography of the real GR case (the calibrated model) but reduced to the size of SH.

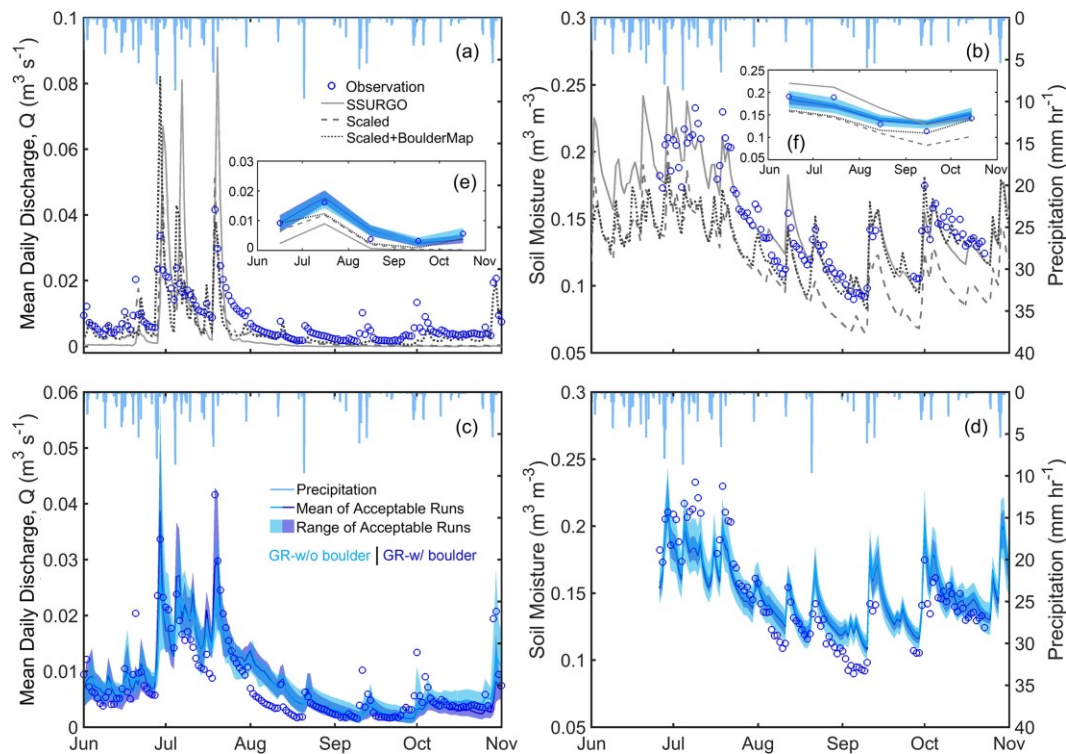


Figure 3. Comparison of Garner Run observations in 2015 (blue circles) and model output from direct transferability tests (gray lines) and after calibration (blue lines and shades). Comparison is for daily discharge (a and c) and areal averaged topsoil moisture from COSMOS within ~10 cm depth (b and d). The inserts compare corresponding monthly discharge (e) and areal averaged topsoil moisture from COSMOS (f). Without any calibration, the “Scaled + Boulder Map” case can reproduce the monthly dynamics (i.e., within the standard of $NSE > 0.5$ and $R^2 > 0.5$) but not daily dynamics. The other two cases (SSURGO and Scaled) without boulder information cannot reproduce monthly and daily water dynamics adequately, indicating the importance of measured boulder distribution. After calibration, acceptable runs without the boulder data (GR-without boulder) and with boulder data (GR-with boulder) similarly reproduce the daily and monthly dynamics.

6. Results

6.1. Transferability Test and the Calibrated Model

Figures 3a and 3b compare the modeled and measured discharge and COSMOS soil moisture in the transferability tests (gray) and from the model after calibration (blue) at GR. GCC values calibrated using discharge data (Shi et al., 2015) were directly transferred from SH to GR for the SSURGO case (where SSURGO soil parameters were used). The model reproduced the temporal trend of monthly average topsoil moisture but only marginally reflected discharge peaks responding to large rainfall events. The model generally overestimated discharge peaks and topsoil moisture and underestimated low flow. The “Scaled” case underestimated discharge and shallow soil moisture to an even larger extent, especially under dry conditions. In the “Scaled + Boulder Map” case where the macropore fraction was spatially distributed based on the boulder map, the model output came closer to measured soil moisture and daily discharge, especially under dry conditions. Model prediction of monthly discharge an NSE value of 0.78 (>0.5); the topsoil moisture prediction also came closer to COSMOS measurements compared to the cases without boulders. That is, when we incorporated the measured boulder map, the model can capture monthly dynamics although not daily dynamics. Comparison of model output between the “Scaled” and “Scaled + Boulder Map” cases showed that the addition of boulder map increased infiltration rates and lateral flow by about 6% of precipitation (data not shown). However did not change much of the already-negligible surface runoff ($<1\%$) at GR. The model with the boulder map produced more flashy peaks under wet conditions.

Figure 3 also compares observations and calibrated output for acceptable runs that satisfied all three performance criteria with and without boulder information. At GR, frequent large rainfalls in July and August led to large discharge peaks and high soil moisture conditions. After August, the catchment became increasingly

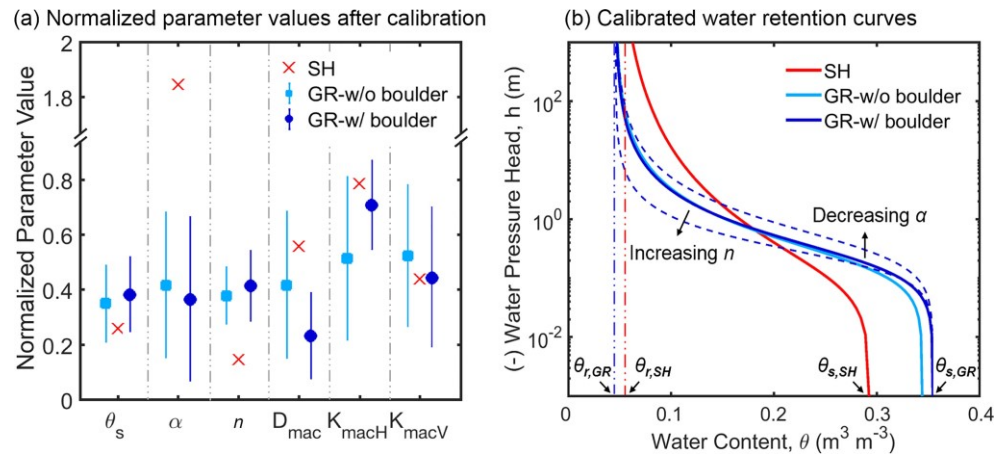


Figure 4. (a) Calibrated (average and normalized) soil parameter values (not global calibration coefficient values): porosity (θ_s), van Genuchten α and n , macropore depth (D_{mac}), and horizontal and vertical macropore conductivities (K_{macH} and K_{macV}) for both sites. Error bars for Garner Run (GR) parameters indicate one standard deviation in all acceptable runs. Shale Hills (SH) parameters are from the one calibration from Shi et al. (2013) and were normalized by the range of corresponding GR parameters so that the relative magnitude of parameters in the two sites could be compared. The θ_s and n values are the most sensitive parameters with the smallest range. (b) Water retention curves with averaged soil parameters (Table 2). The curved dashed lines demonstrate the effects at GR of increasing n from 1.6 to 2.0 and decreasing α from 5.0 to 3.0 m^{-1} . The range of water content ($\theta_s - \theta_r$) is a measure of the potentially mobile water storage capacity per unit soil depth.

drier with infrequent rainfall. The model output reproduced monthly and daily response to rainfall events, overcoming the systematic underestimation of discharge and soil moisture observed in the transferability test. A total of 23 (without boulder map) and 10 (with boulder map) acceptable runs out of 500 runs reproduced the data, manifesting model equifinality (Beven & Freer, 2001). The variation in parameters across different runs was used to estimate the uncertainty range.

6.2. Comparing SH and GR

6.2.1. Differences in Soil Parameters

Table 2 and Figure 4a compared the calibrated soil parameters between SH (red) and GR (blue). Each parameter for GR was an area-weighted average of different soil types across acceptable runs. The most notable difference was the higher porosity, smaller α , larger n , and lower D_{mac} at GR. Small α values arose from large pore sizes and large n values described more drainable soils with lower water retention.

The parameters (Figure 4a) also showed different standard deviations (error bars), where smaller deviation indicates higher sensitivity. Porosity (θ_s) and n were the most sensitive with the smallest acceptable range, indicating these two parameters had to be in a narrow range in order to reproduce daily discharge and soil moisture. Water retention curves depended on α , n , θ_r (residual water content), and θ_s (saturated water content or porosity). The value of α controlled the position whereas n determined the steepness of the water retention curve (Figure 4b). Flatter curves (larger n) represented highly drainable and sandy soils, and steeper curves represented clayey soils where water does not drain as easily. The macropore depth in SH almost doubled the average of GR, and macropore hydraulic conductivities were surprisingly similar, indicating potentially different causes of macropores. We hypothesize that at SH, vegetation roots, which could go deep, might generate macropores whereas at GR, boulders distributed on the ground surface and at shallow soils might be the predominant contributor to macropore generation. Including the boulder map narrowed the range of macropore parameters and therefore reduced uncertainty but did not change the average soil parameters. Comparing the cases with and without boulders, the explicit expression of the greater macropore fraction in the with boulder case led to a lower D_{mac} . In any case, the no boulder cases reproduced very similar discharge and soil moisture data as the with boulder case, indicating its model parameters effectively represented both soil matrix and macropore characteristics without explicitly including boulder information.

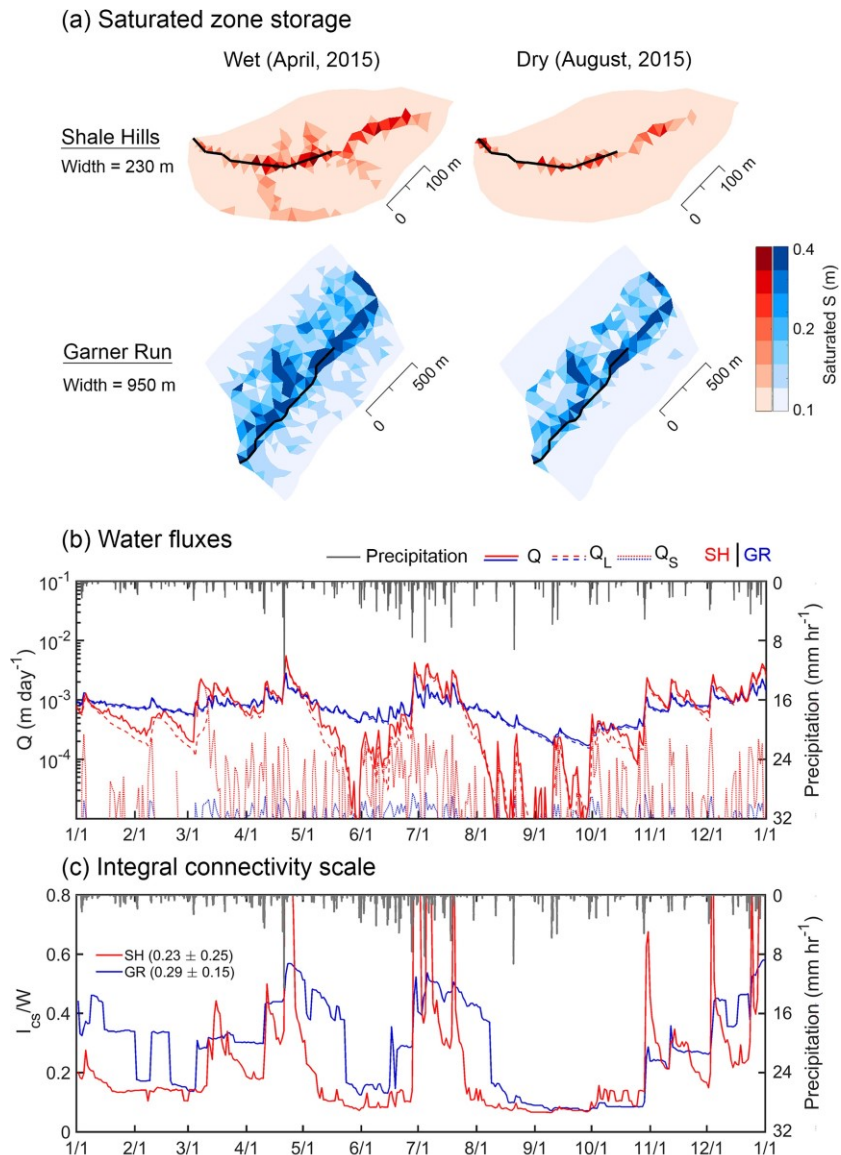


Figure 5. (a) Spatial distribution of saturated zone storage under wet (April) and dry (August) conditions; (b) modeled temporal patterns of different flows in year 2015: stream discharge Q , surface runoff Q_s , and subsurface lateral flow Q_l ; (c) temporal pattern of I_{cs}/W (width-normalized integral connectivity scale). The GR has higher I_{cs}/W and fluctuate less.

6.2.2. Spatial and Temporal Patterns of Water Storage and Fluxes

Figure 5 shows spatial patterns of saturated water storage at different times. GR has a relatively large riparian zone and shows a persistent water presence under both dry and wet conditions (34.7% of total area). SH has a narrow riparian zone (10.6% of total area), mostly consisting of swales with convergent flow. Under dry conditions, only the riparian area was connected to the stream, whereas under wet conditions some uphill area was also connected to the stream. The riparian zone therefore set the baseline connectivity such that GR rarely fell below 0.2 unless under very dry conditions. Under dry conditions, although the riparian zone at GR had relatively higher water content, these waters were not sufficient to form flow connecting to the stream. This may explain similar minimum connectivities at GR and SH. The riparian zone retained 48% of total water storage at GR annually, compared to 20% at SH. During rainfall events, surface runoff occurred as short-lived pulses followed by subsurface lateral flows that dominated and sustained for longer periods (Figure 5b). The discharge at SH fluctuated more than GR, consistent with lower porosity as indicated by

the water retention curve (Figure 5b). In both catchments, surface runoff Q_s was small but higher in SH than in GR (8.9% vs. 1.2%). This was expected as SH soils are clay rich, and water cannot infiltrate as easily as in the sandstone-derived GR with macropores. Figure 5c showed significantly different connectivities, that is, flashy connectivity in SH compared to GR. GR had a larger normalized I_{cs}/W because of the larger soil water storage capacity and larger riparian zone.

In the models, the precipitation (P) is partitioned into evapotranspiration (ET^*) and discharge (Q), and Q is further split into surface runoff (Q_s) and subsurface lateral flow (Q_i) that lumps shallow soil water and some groundwater that contributes to the stream. Further partitioning of ET^* (from model output) into E (evaporation) and T (transpiration) indicated that T dominated the ET^* term in both catchments at ~75% of the ET^* . Water balance calculations from model output after 3-year spin-up indicated relatively similar ET^*/P (~65%) in GR and SH. This is somewhat surprising as GR is characterized by a lower LAI and vegetation density. The LAI values from MODIS (~1-km spatial resolution and 8-day temporal resolution) in 2015 indicated that the SH LAI on average is ~5.4% larger than the GR LAI, although the MODIS resolution was low and may not accurately represent the LAI differences at the two sites. As noted earlier, ET^* represents the total water outflow other than discharge ($P = Q + ET^*$), which can include water loss into the regional aquifer that was not counted in the model.

6.3. Swap Experiments: Comparing the Effects of Soil Property, Topography, and Size

6.3.1. Soil Saturation and Runoff Ratio

With the same topography and size (paired comparison of Figures 6a and 6b, and 6c and 6d), SH soil held more water in the hillslope, and the water content was higher across a wider area compared to the GR soil. Changing from SH soil to GR soil, the runoff ratio (Q/P) increased, again because of the lower water holding capacity of the GR soil. With the same soil properties, changing from SH to GR topography and size decreased Q/P and increased ET^*/P . These trends were generally true for all cases in Table 3. This highlighted that the SH clay-rich soil held more water, reducing Q/P , whereas the steep SH topography and small size increased discharge and Q/P . These two compensating effects (soil vs. topography and size) led to very similar Q/P values between the two real cases SH and GR and between the two hypothetical cases $Soil_{SH}TP_{SH}Size_{SH}$ and $Soil_{GR}TP_{GR}Size_{GR}$, therefore masking the effects of individual variables.

6.3.2. Storage-Discharge Relationship

The S - Q relationship (equation (1)) and parameters (Table 3) quantified the response of discharge to changing water storage (Kirby et al., 1991; Kirchner, 2009). Figure 7 shows the model output of discharge versus averaged storage for hypothetical catchments. The catchments generally did not produce discharge until a minimum water storage (S_{min}) was reached. Beyond that, the discharge increased exponentially with storage, until reaching a maximum water storage (S_{max}). The difference between these storage values quantified the dynamic storage S_d . The S_d in SH varied from about 0.20 to 0.31 m, corresponding to discharge varying by more than 2 orders of magnitude ($\sim 10^{-5}$ – 8.2×10^{-3} m/day). At GR, water storage changed from 0.23 to 0.39 m corresponding to Q from 1.7×10^{-4} to 4.4×10^{-3} m/day. These corresponded to 0.11 and 0.16 m of streamflow-generating S_d at SH and GR, respectively. These dynamic storages were much smaller than the S_t value of 0.49 m in SH and 0.63 m at GR. Note that S_{max} values in all cases never reached S_t of 0.49 m at SH and 0.63 m at GR, indicating the catchments never reached the total storage capacity and never were fully saturated. Comparing across different cases, catchment size had significant effects on S_d . The S_d in the real SH case had the lowest value of about 0.11 m. Despite the differences in topography and soil properties, all cases with SH size had about 0.11 to 0.13 m, whereas all cases with GR size had S_d at about 0.16 to 0.17 m. The results suggested that as catchment size increased, S_d also increased due to the expanded area that was connected to the stream (larger I_{cs}), even though the relative I_{cs}/W decreased with catchment size (Figure 8).

The sensitivity of discharge to storage was prescribed by the b values, or the nonlinearity of the S - Q curve. The red cases with SH soil in Figure 7a had b values at 1.73–1.86 (Table 3), compared to values of 1.34–1.59 in blue cases with GR soil. Catchment size had some impacts on b values but was not as influential as soil properties. Increasing catchment size from SH to GR without changing soil properties and topography decreased b values slightly from 1.79 to 1.73 for SH (from SH to SH-Size_{GR}), and from 1.50 to 1.34 for GR (from GR to GR-Size_{SH}). Combining all cases, the soil properties had a predominant control on the nonlinearity of S - Q relationships.

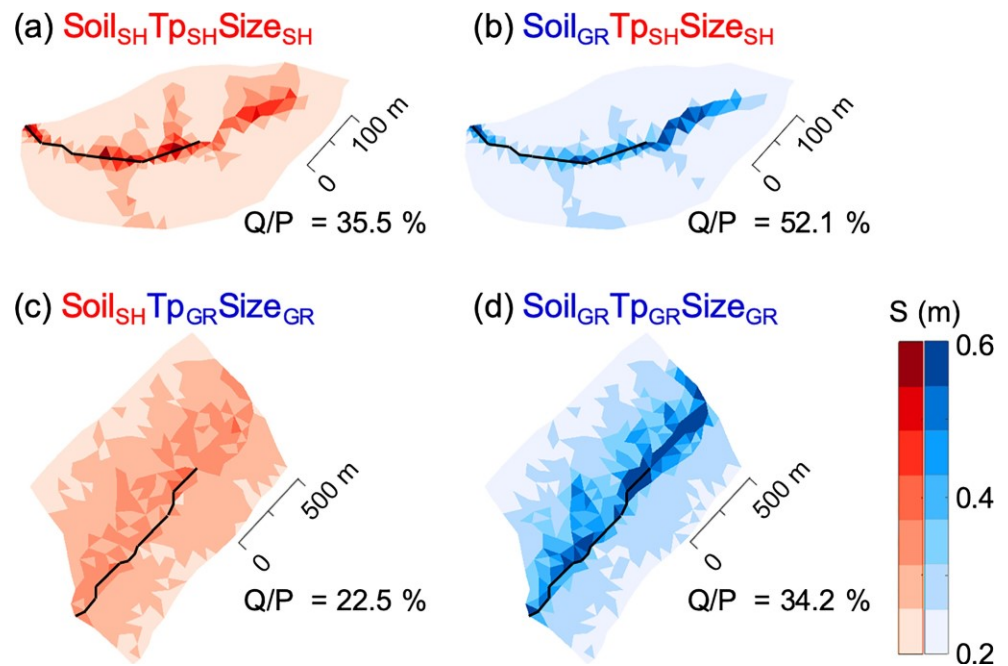


Figure 6. Water balance (Q/P) and saturated storage (S) distribution in four hypothetical cases where soil properties, topography, and vegetation are uniformly assigned with area-weighted averages. The “Soil_{SH}” and “Soil_{GR}” refer to cases using SH and GR soil properties, respectively; “Tp_{SH}” and “Tp_{GR}” refer to cases using SH and GR topography, respectively; and “Size_{SH}” and “Size_{GR}” refer to cases using their respective sizes. With everything else being the same, the GR soil generates more runoff (Q/P) than SH. The GR topography and size lead to lower runoff due to its large riparian zone and water storage capacity.

Table 3
Water Balance, Parameters for Storage-Discharge Relationships, and Connectivity

Cases ^a	Water balance		$Q \propto \rho_{ref} \frac{S - S_{min}}{S_d} \quad 1 = 0.2 - b \rho$						Connectivity		
	Q/P (%)	Q (10^{-3} m/day)		Q_{ref}	Water storage ^b (m)		b	R^2	Mean I_{cs} (m)	Mean I_{cs}/W ($\pm \frac{std}{I_{cs}}$)	S_c^d (I_{cs}/W_c)
		Q_{min}	Q_{max}		S_{min}	S_d^c					
SH	33.8	0.01	8.19	4.35	0.20	0.11	1.79	0.89	53	0.23 ($\pm 109\%$)	0.28 (0.27)
SH-Size _{GR}	33.3	0.02	10.50	7.80	0.22	0.16	1.73	0.77	92	0.10 ($\pm 172\%$)	0.32 (0.19)
Soil _{SH} Tp _{SH} Size _{SH}	35.5	0.01	8.11	5.43	0.19	0.13	1.77	0.91	69	0.30 ($\pm 117\%$)	0.27 (0.19)
Soil _{SH} Tp _{SH} Size _{GR}	24.5	0.01	6.68	5.73	0.19	0.17	1.77	0.84	165	0.18 ($\pm 121\%$)	0.29 (0.24)
Soil _{SH} Tp _{GR} Size _{GR}	22.5	0.01	8.67	2.59	0.20	0.17	1.86	0.80	209	0.22 ($\pm 100\%$)	0.28 (0.24)
GR	33.5	0.17	4.35	1.87	0.23	0.16	1.34	0.70	276	0.29 ($\pm 52\%$)	0.31 (0.28)
GR-Size _{SH}	37.0	0.10	5.18	3.86	0.23	0.13	1.50	0.84	98	0.42 ($\pm 29\%$)	—
Soil _{GR} Tp _{SH} Size _{SH}	52.1	0.01	4.12	2.73	0.17	0.13	1.59	0.85	70	0.30 ($\pm 83\%$)	0.28 (0.50)
Soil _{GR} Tp _{GR} Size _{SH}	38.8	0.10	3.12	2.70	0.22	0.12	1.49	0.89	102	0.44 ($\pm 25\%$)	—
Soil _{GR} Tp _{GR} Size _{GR}	34.2	0.24	1.75	1.56	0.21	0.17	1.34	0.93	380	0.40 ($\pm 35\%$)	0.29 (0.37)

Note. GR = Garner Run; SH, Shale Hills. The bold rows are for the real SH and GR cases

^aSH and GR are the calibrated cases using measurements. All other cases are hypothetical. SH-Size_{GR} and GR-Size_{SH} are based on the real SH and GR with heterogeneous distribution of soil but shrunk or expanded to the size of the other catchment. All other cases have uniformly assigned vegetation cover and soil properties based on area-averaged values. ^bThe total water storage S_t was estimated as soil depth $\times$ porosity θ_s . With the area-averaged soil depths of 1.7 and 1.8 m for SH and GR, respectively, the S_t for GR = $1.8 \text{ m} \times 0.35 = 0.63 \text{ m}$, and $1.7 \text{ m} \times 0.29 = 0.49 \text{ m}$ for SH. Note that this differs from the potentially mobile soil water storage directly from the soil depth and mobile porosity ($\theta_s - \theta_r$), which should be $1.8 \text{ m} \times (0.35 - 0.04) = 0.56 \text{ m}$ for GR, and $1.7 \text{ m} \times (0.29 - 0.05) = 0.39 \text{ m}$ for SH. ^cThe dynamic water storage $S_d = S_{max} - S_{min}$ is the same as k_1 in (Kirchner, 2009), corresponding to differences in storage values that produce Q_{max} and Q_{min} , maximum and minimum discharge. The Q_{ref} represents a reference Q between minimum and maximum Q values. ^d S_c is the critical storage value at their corresponding critical connectivity values (I_{cs}/W_c in the parenthesis) beyond which discharge increases significantly. For some hypothetical catchments, the critical connectivity cannot be easily observed so there are no values. (Figure 9)

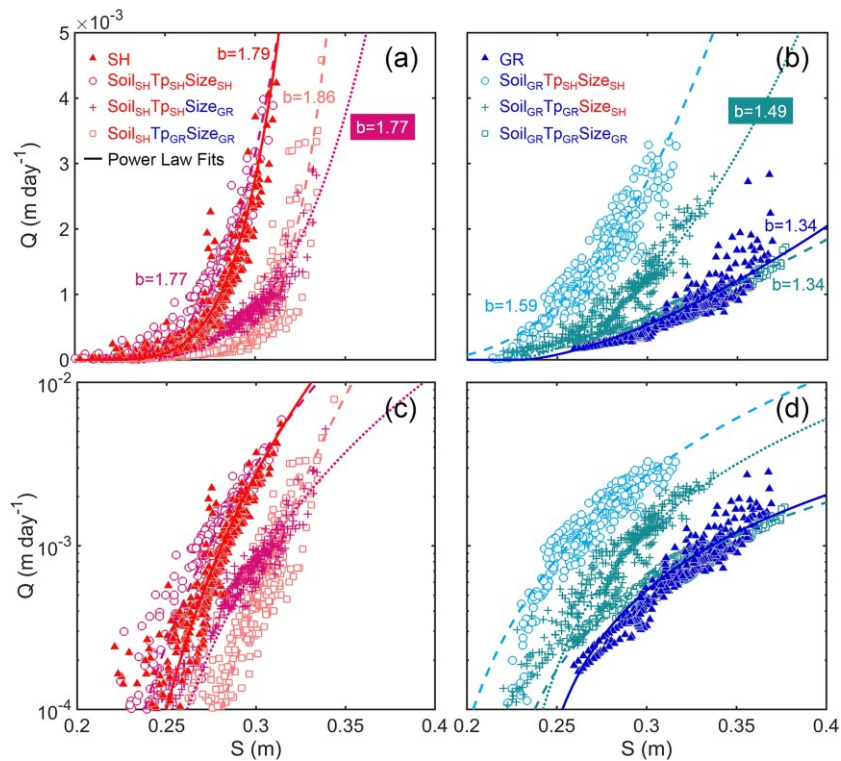


Figure 7. Storage-discharge (S - Q) relationship with Kirchner (2009) fitting curves (equation (1)) on (a and c) linear and (b and d) logarithmic scales. Soil properties have the first order control on the nonlinearity of the S - Q relationship.

6.3.3. Connectivity

Figures 9a and 9b showed significantly contrasting connectivity-storage relationships. SH soil, the connectivity exhibited a pronounced threshold behavior: The connectivity remained low at low water storage until reaching a “critical” storage (S_c) beyond which it increased suddenly and dramatically. All cases with SH soil had S_c values between 0.27 and 0.32 m and at corresponding connectivity values between 0.19 and 0.27 (last column in Table 3). Cases with GR soil also showed critical storage values; however, the connectivity tended

to “jump” from values at low storage ($S < S_c$) to values at $S > S_c$, but the connectivity values increased linearly with small variations instead of exponentially (lower $\sigma_{I_{cs}=W}$ in Table 3). In some cases, no critical storage and connectivity values could be derived. For cases that did exhibit threshold behavior, the S_c values were surprisingly similar to those with SH soil cases. However, the corresponding critical connectivity was much higher. Table 3 also indicated that connectivity remained similar or increased when comparing heterogeneously distributed soils to uniform soils with area-averaged parameters. Figures 9c and 9d showed that discharge had a strong dependence on connectivity at low connectivity but tended to scatter at high connectivity beyond critical connectivity values.

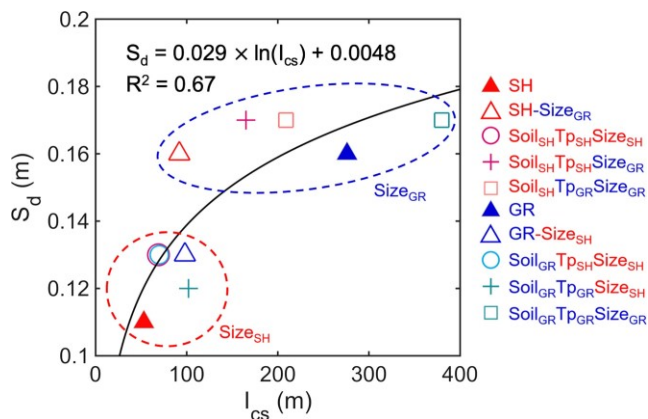


Figure 8. The dynamic water storage (S_d) versus the mean of integrated connectivity (I_{cs}). The logarithmic fitting result has the highest R^2 compared to linear and exponential relationships. Increasing the catchment size without changing other factors provides the larger S_d values, which indicate that S_d increases with catchment size.

7. Discussion

Topography and soil properties emerge as a result of the lithological starting material and the climate and tectonic forcings (Jenny, 1941). Results from three analyses underscore the predominant control of soil properties, instead of topography, in regulating the nonlinearity of the S - Q relationship. This result contrasts with the original hypothesis that topography plays a more important role. The transferability test cannot reproduce

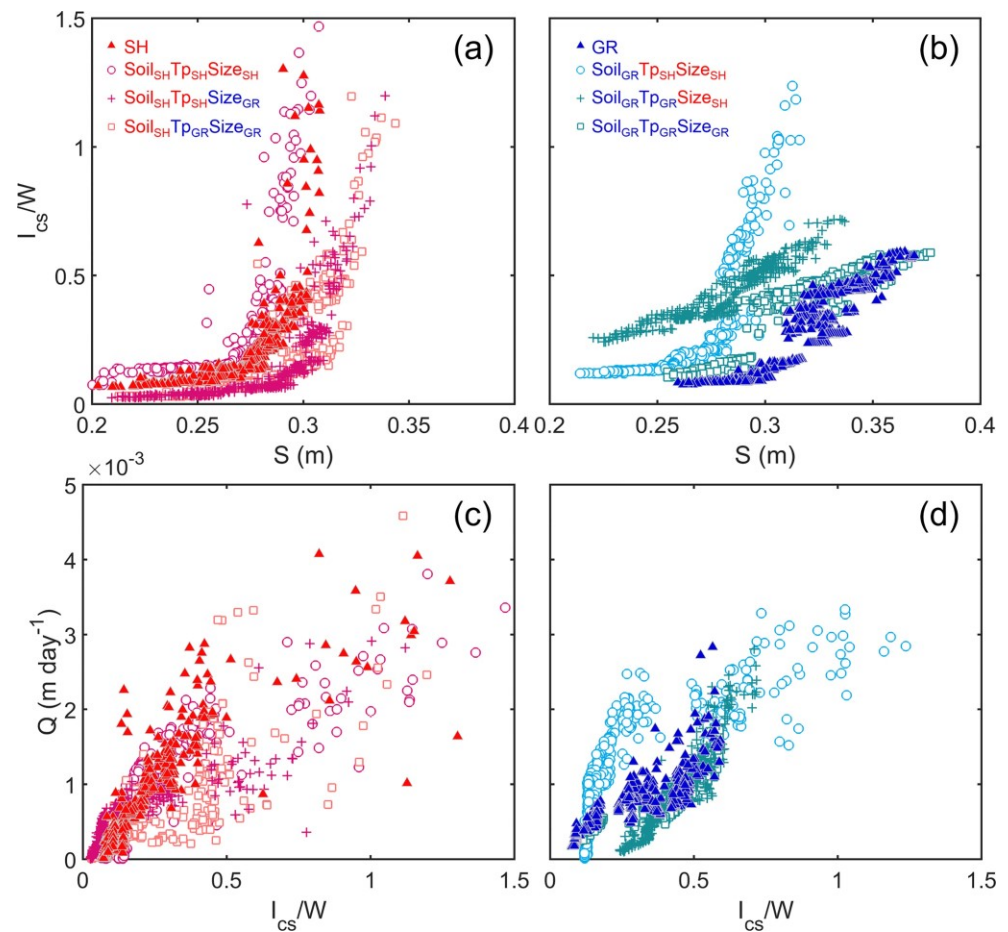


Figure 9. Relationship of normalized integral connectivity scale (I_{cs}/W) with water storage (S) (a and b) and discharge-connectivity relationship (c and d) for swap experiment cases. All cases with Shale Hills soil have a pronounced threshold behavior where connectivity does not change much until water storage reaches a critical value, beyond which the connectivity increase significantly. The cases with Garner Run soil generally have much more gradual increase in connectivity and do not exhibit threshold behavior, except the case with Shale Hills topography and size.

daily or monthly discharge until the boulder distribution is incorporated in the model. The calibration and sensitivity analysis underscore the importance of the parameters describing soil properties. These parameters reflect both soil matrix and macropore/boulder characteristics (Figure 4). The swap experiments further revealed two pronounced observations (Figure 10 and Table 3). First, differences in soil matrix and macropore properties arising from lithology (shale vs. sandstone) exert a predominant control: GR bouldery soils and large macropores consistently have lower b values, higher Q/P , and higher connectivity, as illustrated in the larger differences in symbol size in cases with different soil properties (vertical direction) (Figures 10a to 10c). Second, catchment size strongly influences S_d , changing catchment size from SH to GR size increases S_d from ~ 0.12 to ~ 0.17 m, illustrated with solid-line and dashed-line circles in Figure 10d. The Q/P ratios are higher in all GR soil cases compared to their corresponding SH soil cases, indicating that although SH soils generate high flow in large storms (higher flooding tendency), they generate less discharge in the long term (at the annual time scale) because of a larger water holding capacity.

7.1. Soil Properties and S - Q Relationship

Clayey and sandy soils release water very differently: Sandy soils with large pores and low clay content release water rapidly, whereas clayey soils with small pores tend to hold water. Soils derived from different lithologies differ not only in clay content, soil aggregation, and spatial arrangement (Du et al., 2016) but also

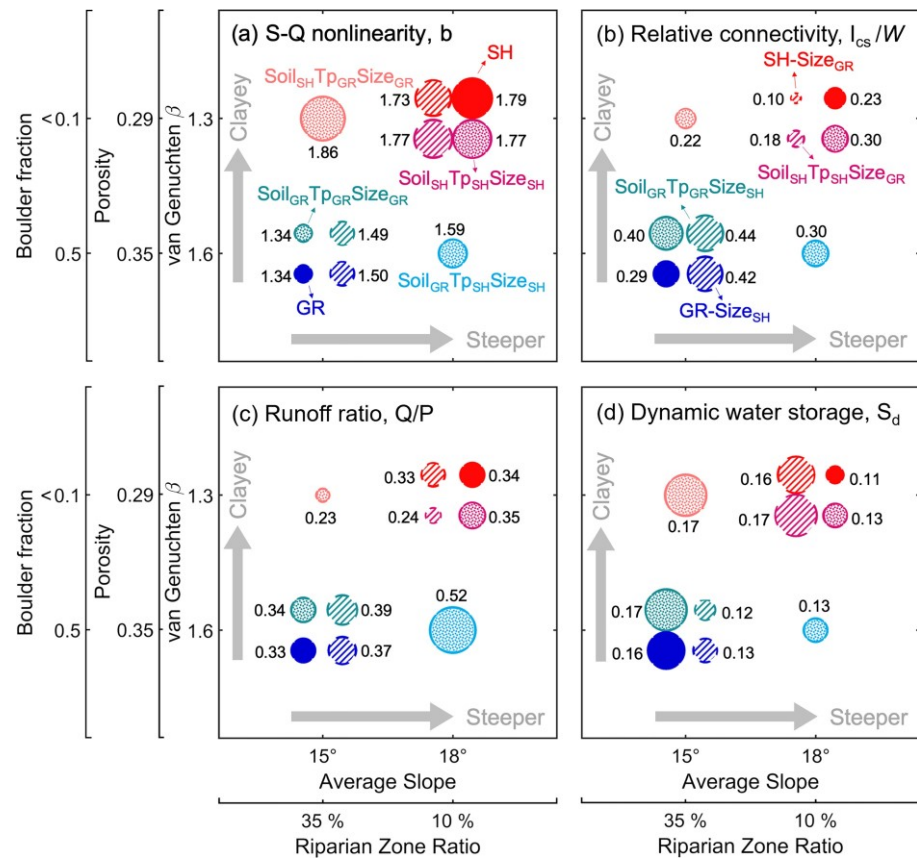


Figure 10. Comparing (a) S - Q nonlinearity b , (b) relative connectivity I_{cs}/W , calculated as averaged normalized connectivity, (c) runoff ratio Q/P , and (d) dynamic water storage S_d (or k_1) for the 10 simulated cases. The symbols size indicates the magnitude of each quantity. The warm and cold colors represent cases with SH and GR soils, respectively, the same as in previous figures. The dashed line circles filled with lined texture represent cases where only catchment size is changed whereas soil and topography characteristics are the same as the solid line circles with same colors. Symbols size comparison between horizontal cases shows the effects of topography; comparison between vertical cases indicates effects of soil properties (boulder fraction and van Genuchten parameter n). Larger symbol size differences mean more pronounced effects. The figure indicates that steep hills with small riparian zones and clayey soils of high water retention capability (i.e., SH) tend to have small dynamic water storage (S_d and smaller and more variable connectivity (b) with higher nonlinearity of S - Q relationships (a). The catchment size influences S_d but not as much on b values and relative connectivity.

in the amount of colluvial materials and boulders that generate macropores (Del Vecchio et al., 2018). In both catchments, surface runoff is negligible. The SH soil has been shown to form preferential flow paths at the interfaces of soil horizons and soil-rock contrasts (Jin et al., 2011). These distinctive characteristics possibly lead to different flow generation mechanisms. In GR, abundant colluvial materials and bouldery soil lead to more infiltration and fast flow activation via macropores. In SH, the strong water holding capacity may prevent the formation of preferential flow until after reaching threshold values, beyond which fast flow triggers highly nonlinear, threshold behavior. This finding agrees with threshold behaviors observed when dry-soil barriers are breached (McNamara et al., 2005), when subsurface saturated areas are connected to the trench (Tromp-van Meerveld & McDonnell, 2006), and when threshold soil moisture values are reached (Seeger & Weiler, 2014). Rapid flow activation beyond thresholds has been observed with intensive soil moisture measurements at SH (Lin et al., 2006). These observations related to soil properties have strong implications on how to represent the first-order control of stream flow generation in parsimonious models (Fenicia et al., 2014).

The importance of soil properties emphasized here corroborates findings of related work. Spatial analysis of nested watersheds comparing topography and soil property indicates that transit time characteristics

correlate better with soil maps than with topographic indices in northern, wet catchments with wetland riparian areas (Geris, Tetzlaff, McDonnell, et al., 2015; Soulsby et al., 2006; Tetzlaff et al., 2009). In another study of 24 mesoscale catchments in Switzerland, only weak correlations were observed between transit time measures and topographic indices (Seeger & Weiler, 2014). The catchment storage derived from mean transit times and mean discharge did not show a clear relation to any catchment properties. Zimmer and Gannon (2018) examined 20 years of daily runoff from 73 regional watershed scale USGS stream gaging sites across North Carolina, United States. Their work suggests that differences in soil and bedrock properties outweigh topographic and climatic differences, because soil-bedrock interfaces and low-hydraulic conductivity layers strongly influence the partitioning of infiltrated water.

Based on an analysis of streamflow generation, Kirchner (2009) proposed that the steepness of the S - Q slope (dQ/dS) or the sensitivity of the S - Q relationship ($g(Q)$) positively relate to catchment slope (m , topography descriptor) and soil hydraulic conductivity (κ , soil property indicator) and negatively relate to soil porosity (θ , soil property indicator) in the form of $g(Q) \propto \frac{m\kappa}{\theta}$ (Kirchner, 2009). From this equation one might infer that soil permeability is the predominant control on streamflow generation because the permeability of sandy soil is typically orders of magnitude higher than for clayey soil, whereas m and θ have a narrow range and do not change as much. In contrast, our results indicate that the van Genuchten parameters (α , n , θ_r , and θ_s) and macropore properties may be better measures of water release and flow generation. In addition, θ here may be replaced by the hydrologically responsive S_d , instead of total porosity, which includes both dynamic and passive storage. In other words, the correlation equation may need to be modified to take into account of soil and macropore properties topographic features such as the size of the riparian zones.

7.2. Dynamic Water Storage, Connectivity, and Catchment Size

Dynamic storage is defined as the portion of storage “that is hydrologically active and directly contributes to streamflow” (Hrachowitz et al., 2016; McNamara et al., 2011) or controls streamflow dynamics (Buttle, 2016). In this work, values of S_d (0.11–0.17 m) are much smaller than S_t calculated from soil depth and porosity (0.49 m for SH and 0.63 m for GR). This indicates that small changes in soil structure and porosity can have dramatic impacts on S_d . The difference of 0.05 m (5 cm) in S_d may appear small. It is however 42% and 30% of S_d in SH and GR, respectively. The largest storms in Pennsylvania occur at the rate of ~5–10 mm/hr. A 5 cm difference in S_d can well make the difference between flooding or not.

Water storage estimation remains one of the largest uncertainties in predicting streamflow generation. Dynamic storage calculated from different approaches often yields estimates that differ by up to an order of magnitude (Staudinger et al., 2017). Buttle (2016) used the Kirchner (2009) framework to calculate S_d for five drainage basins (~11 to 270 km²). The thick sand and gravel deposits of the Oak Ridges Moraine in southern Ontario, Canada. They did not observe a clear correlation between S_d and catchment size but found that S_d inversely correlated to the ratio of stream water deuterium variability relative to that of precipitation, suggesting larger S_d associated with shorter water mean transit time (Buttle, 2016). Birkel et al. (2011) estimated S_d based on water isotope data and modeling for two nested Scottish catchments (3.6 and 30.4 km²). They estimated smaller S_d (40 mm out of 500-mm total storage) for the larger catchment and larger S_d (55 mm out of 900-mm total storage) for the smaller catchment, respectively. Although carefully chosen for their similarities, the catchments in these studies still vary in soil properties, topography, and land cover, any of which could mask effects of catchment size.

Our results indicate that S_d correlates with hydrological connectivity and increases with catchment size (Figure 8). This correlation is hardly surprising: Streamflow is generated when stored water in different parts of catchments connects and forms flow. In fact, hydrologic connectivity between upland and stream has long been considered essential for transmitting water to streams (Bracken et al., 2013; Spence, 2010). The internal catchment structure, that is, the spatial arrangement and size of hillslope and riparian zones, and their connectivity have been shown to exert a first-order control on the activation of fast flow paths and release of old water (Stockinger et al., 2014) and the magnitude and timing of water and solute release (Jencso et al., 2009; Jencso et al., 2010; Nippgen et al., 2015). In the swap experiments where catchment size changes whereas the topography and soil properties remain constant, the runoff generating area that connects to the stream inevitably changes. Larger catchments offer a larger zone of source water that dynamically connects to the

stream, as indicated by the increasing I_{cs} in Figure 8 and Table 3 when the modeled catchments increase in size from that of SH to GR.

7.3. Lack of Model Transferability Suggests Challenges in a PUB Framework

Model transferability from SH to GR did not work well even though the two catchments are both first order and lie within a few kilometers of one another. In fact, the models required detailed information about the boulder distribution to reproduce monthly discharge because the boulder significantly affects macropore distribution. The importance of soil properties has also been emphasized in other parameter sensitivity studies using the Observing System Simulation Experiments (Shi et al., 2014). Other published transferability studies have shown that availability of a soil map can guide a priori parameterization and improve model performance (Smith et al., 2004). They also showed that parameter transfer between neighboring catchments is the most unsatisfactory for catchments with differing lithology; transfers between nonneighboring catchments with analogous soils and with different vegetation were observed to be more successful (Heuvelmans et al., 2004). Hydrologic models using geology-based HRUs have been found to capture the spatial variability of streamflow time series better than those using topography-based HRUs (Fenicia et al., 2016). Together, these studies suggest potentially higher parameter transferability between catchments with relatively similar soil properties.

Insights gained from these studies suggest that measurements of soil properties are crucial for successful model transfer. Among the many variables that govern streamflow generation, information on climate, topography (relief, riparian zone, and hillslopes), and land cover (organisms) has become increasingly available from satellites and other observation instruments even in remote areas (McCabe et al., 2017). Below ground characteristics such as soil depth, boulder spatial distribution, and water retention properties, however, cannot be observed directly using similar large instruments. In situ conductivities and their spatial variation are challenging to obtain and often need to be inferred from tracer tests (Jackson et al., 2016; Kuntz et al., 2011; Li et al., 2010). Their measurements are often local, at scales much smaller than the relevant scale for streamflow prediction. The boulder information in this work is an idiosyncrasy of the site and is available for this work only because of another study. Such information however is not available for most watersheds. Current practices in assigning soil property parameters typically use pedotransfer functions in combination with soil properties obtained from databases such as SSURGO (Shi et al., 2013). In fact, generating the boulder map involves intensive field mapping and thus is about the furthest from the ungauged basin that one could possibly envision. It suggests that unless we have a good grasp of the subsurface properties, we may have little chance in making reasonable predictions in ungauged basins.

Our work therefore highlights a general need to understand lithology and parent materials in determining soil properties. It is impossible to measure soil properties everywhere. We need methods to predict subsurface properties based on a small number of measurements that are easy to assess. In fact, it may be possible that soil formation models developed from simulating weathering processes can eventually allow meaningful prediction of soil properties that can be used for predicting streamflow dynamics (Heidari et al., 2017; Lebedeva et al., 2010). With such efforts, the need to measure soil properties everywhere might be avoided.

7.4. Model Limitations and Strengths

McDonnell (2003) argued that it may be more fruitful to examine the common characteristic forms of nonlinearity and feedbacks than characterizing the peculiarity of specific catchments. Generally speaking, however, field studies of catchment-scale emergent dynamics and catchment classification are challenging because of the confounding nature of multiple factors, and the large data sets required for parameterization. This is one reason for the popularity of experiments that incorporate paired catchments where some variables are held the same. Physically based models and carefully thought virtual experiments may help circumvent such limitations. Spatially explicit hydrology models typically require a large parameter set with large uncertainty, leading to issues with respect to equifinality (Beven & Freer, 2001). A good example of this is the many cases that can reproduce data shown in this work. Despite these limitations, spatially explicit, process-based models have the advantage of generating “digital” catchments in an era with exponentially growing Earth surface data and investigating the convoluted mechanisms of multiple processes using numerical experiments. The thought (swap) experiments here, by systematically controlling variables via a hypothesis testing approach (Clark et al., 2011), offer mechanistic insights and quantify the relative effects

of interdependent variables. In particular, the swap experiments enable the assessment of topography and catchment size—two variables that are often held constant in typical sensitivity analyses.

Given spatially and temporally sparse field measurements, these models can also guide when and where critical measurements are needed. Data-informed numerical experiments could provide an alternate, cost-effective approach to test hypotheses against field data (Fatichi et al., 2016; Li et al., 2017; Weiler & McDonnell, 2004), especially where field data are collected systematically for catchments chosen on the basis of important variables such as climate, biota, relief, and parent material. They can also be used to “discover” general principles across space and time, therefore complementing a growing body of literature using statistical approaches and cross-site synthesis (Jasechko et al., 2016; Jasechko et al., 2017; Kuentz et al., 2017; Wagener et al., 2010). This will ultimately facilitate catchment classification that will lead to better hydrologic theory for PUB (Hrachowitz et al., 2013; McDonnell & Woods, 2004; Wagener et al., 2007).

8. Conclusions

This study examined the emergent dynamics of streamflow generation including the nonlinearity of S - Q relationships, connectivity, and dynamic water storage in two first-order monolithological catchments: the shale-underlain SH (0.08 km²) and the sandstone-underlain GR (1.34 km²). Both catchments are located in the temperate climate in central Pennsylvania, USA. SH has steeper slopes and a narrow riparian zone (~10% of catchment area) whereas GR is not as steep and has a relatively large riparian zone (~34%). We tested the hypothesis that the influence of topographic characteristics (a flatter slope, longer slope length, larger riparian zone) is more significant than that of soil properties and catchment size, leading to a dampened streamflow response and a more linear S - Q relationship at GR compared to SH. The hypothesis was tested combining streamflow, soil moisture data, and a spatially explicit, process-based model Flux-PIHM. Three lines of analyses, including model transferability tests, calibration and parameter sensitivity analyses, and swap experiments, all led to the rejection of the original hypothesis. These results underscore the predominant role of soil properties.

The transferability test showed that direct transfer of calibration information from SH to GR in models with catchment topography (digital elevation) data reproduced the monthly trend but overestimated daily discharge peaks and underestimated monthly discharge and topsoil moisture. Only when the boulder map from the field survey was incorporated did the SH calibration information used for GR predict monthly discharge at GR because the boulders enhanced infiltration and accelerated water release to the stream (Figure 3). Thus, the role of topography was small compared to that of soil properties. If the topography was the primary control, the model calibration information would have been directly transferable from SH to GR because topography was explicitly represented in each of the model domains. Model calibration and sensitivity analysis for GR (500 simulation runs) identified important soil parameters including soil matrix and macropore properties that reflected boulder characteristics (Figure 4). The analysis indicated that the GR soil has higher porosity, lower water retention (van Genuchten n), higher water storage capacity (θ_s), and shallower macropore depths compared to the SH soil. Comparison between the two catchments indicated more flashy streamflow response and hydrological connectivity variation in SH than GR (Figure 5).

Swap experiments comparing cases disentangling the effects of interdependent variables (topography, soil properties, and size) accentuated the significance of clayey soils in holding water (Figures 6 and 7). In addition, they revealed two overarching observations (Figure 10 and Table 3). First, soil properties exerted a predominant control: cases with GR bouldery soil consistently had lower b values, higher Q/P , and higher connectivity. Second, dynamic water storage S_d increased with catchment size. For example, an increase in size from that of SH to GR resulted in an increase in S_d from about ~0.12 to ~0.17 m, largely because of the higher connectivity related to a larger area connected to the stream in the larger catchments (Figure 8). Cases with GR soil had more linear S - Q relationships without pronounced threshold behavior: in some cases, critical connectivity could not be identified (Figure 9). The Q/P ratios were higher in all GR soil cases compared to their corresponding SH soil cases, indicating that although SH soil tended to have high flow in large storms (more flooding tendency), it generated less discharge in the longer term (i.e., years) due to its large water holding capacity.

These results underscore the importance of soil properties as compared to topography. This predominant role of soil properties presents a grand challenge for PUB, as soil properties are more challenging to measure and are not as readily available as earth surface characteristics such as topography and land cover. This work also illustrates that controlled virtual experiments (comparisons between model outputs for different catchment characteristics) can systematically disentangle convoluted effects of interdependent variables and can offer mechanistic understanding of catchment-scale emergent dynamics.

Acknowledgments

Simulation data from this work will be archived in the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) data website (<http://www.hydroshare.org/resource/afecce37672e42c9a248e018f3e6fb8c>). We appreciate the constructive and comprehensive comments from Editor Ilja van Meerveld, the Associate Editor, and three anonymous reviewers who have helped sharpen the message. In particular, the anonymous Associate Editor has read the manuscript and related publications carefully, paid particular attention to details, and have offered thoughtful comments that have significantly improved the manuscript. We appreciate discussion with Joannmarie Del Vecchio on geomorphological processes, with David M. Eissenstat on vegetation and evapotranspiration, with Tess Russo and Beth Hoagland on discharge data. Financial support was provided by National Science Foundation Grant EAR – 0725019 (C. Duffy), EAR – 1239285 (S. Brantley), and EAR – 1331726 (S. Brantley) for the Susquehanna Shale Hills Critical Zone Observatory (SSHCZO). Logistical support and/or data were provided by the NSF-supported SSHCZO. This research was conducted in Penn State's Stone Valley Forest, which is funded by the Penn State College of Agriculture Sciences, Department of Ecosystem Science and Management, and managed by the staff of the Forestlands Management Office. This research was conducted in Rothrock State Forest that is funded and managed by the Pennsylvania Department of Conservation and Natural Resources, Bureau of Forestry. The field data have been digitized and are accessible through national CZO data portal (<http://criticalzone.org/shale-hills/data/datasets/>). The source code of Flux-PIHM and the input files necessary to reproduce the results are available from the authors upon request (lili@engr.psu.edu).

References

- Ali, G., Tetzlaff, D., McDonnell, J. J., Soulsby, C., Carey, S., Laudon, H., et al. (2015). Comparison of threshold hydrologic response across northern catchments. *Hydrological Processes*, 29(16), 3575–3591. <https://doi.org/10.1002/hyp.10527>
- Ali, G. A., & Roy, A. G. (2009). Revisiting hydrologic sampling strategies for an accurate assessment of hydrologic connectivity in humid temperate systems. *Geography Compass*, 3(1), 350–374. <https://doi.org/10.1111/j.1749-8198.2008.00180.x>
- Allard, D. (1994). Simulating a geological lithofacies with respect to connectivity information using the truncated Gaussian model. In *Geostatistical simulations*, (pp. 197–211). Springer.
- Anderson, S. P., Bales, R. C., & Duffy, C. J. (2008). Critical Zone Observatories: Building a network to advance interdisciplinary study of Earth surface processes. *Mineralogical Magazine*, 72(1), 7–10. <https://doi.org/10.1180/minmag.2008.072.1.7>
- Bao, C., Li, L., Shi, Y., & Duffy, C. (2017). Understanding watershed hydrogeochemistry: 1. Development of RT-Flux-PIHM. *Water Resources Research*, 53, 2328–2345. <https://doi.org/10.1002/2016WR018934>
- Beven, K., & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology*, 249(1–4), 11–29. [https://doi.org/10.1016/S0022-1694\(01\)00421-8](https://doi.org/10.1016/S0022-1694(01)00421-8)
- Birkel, C., Soulsby, C., & Tetzlaff, D. (2011). Modelling catchment-scale water storage dynamics: Reconciling dynamic storage with tracer-inferred passive storage. *Hydrological Processes*, 25(25), 3924–3936. <https://doi.org/10.1002/hyp.8201>
- 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.11488>
- Bracken, L. J., Wainwright, J., Ali, G. A., Tetzlaff, D., Smith, M. W., Reaney, S. M., & Roy, A. G. (2013). Concepts of hydrological connectivity: Research approaches, pathways and future agendas. *Earth-Science Reviews*, 119, 17–34. <https://doi.org/10.1016/j.earscirev.2013.02.001>
- Brantley, S. L., DiBiase, R. A., Russo, T. A., Shi, Y., Lin, H., Davis, K. J., et al. (2016). Designing a suite of measurements to understand the critical zone. *Earth Surface Dynamics*, 4(1), 211–235. <https://doi.org/10.5194/esurf-4-211-2016>
- Brantley, S. L., Goldhaber, M. B., & Ragnarsdottir, K. V. (2007). Crossing disciplines and scales to understand the critical zone. *Elements*, 3(5), 307–314. <https://doi.org/10.2113/gselements.3.5.307>
- Brantley, S. L., Holleran, M. E., Jin, L., & Bazilevskaya, E. (2013). Probing deep weathering in the Shale Hills Critical Zone Observatory, Pennsylvania (USA): The hypothesis of nested chemical reaction fronts in the subsurface. *Earth Surface Processes and Landforms*, 38(11), 1280–1298. <https://doi.org/10.1002/esp.3415>
- Brantley, S. L., Lebedeva, M. I., Balashov, V. N., Singha, K., Sullivan, P. L., & Stinchcomb, G. (2017). Toward a conceptual model relating chemical reaction fronts to water flow paths in hills. *Geomorphology*, 277, 100–117. <https://doi.org/10.1016/j.geomorph.2016.09.027>
- Brantley, S. L., White, T., West, N., Williams, J. Z., Forsythe, B., Shapich, D., et al. (2018). Susquehanna Shale Hills Critical Zone Observatory: Shale Hills in the context of Shaver's Creek Watershed. *Vadose Zone Journal*, 17(1). <https://doi.org/10.2136/vzj2018.04.0092>
- Brown, V. A., McDonnell, J. J., Burns, D. A., & Kendall, C. (1999). The role of event water, a rapid shallow flow component, and catchment size in summer stormflow. *Journal of Hydrology*, 217(3), 171–190. [https://doi.org/10.1016/S0022-1694\(98\)00247-9](https://doi.org/10.1016/S0022-1694(98)00247-9)
- Brubaker, K. M., Johnson, Q. K., & Kaye, M. W. (2018). Spatial patterns of tree and shrub biomass in a deciduous forest using leaf-off and leaf-on lidar. *Canadian Journal of Forest Research*, 48(9), 1020–1033. <https://doi.org/10.1139/cjfr-2018-0033>
- Brutsaert, W., & Nieber, J. L. (1977). Regionalized drought flow hydrographs from a mature glaciated plateau. *Water Resources Research*, 13(3), 637–643. <https://doi.org/10.1029/WR013i003p00637>
- Buttle, J. M. (2016). Dynamic storage: A potential metric of inter-basin differences in storage properties. *Hydrological Processes*, 30(24), 4644–4653. <https://doi.org/10.1002/hyp.10931>
- Clark, M. P., Kavetski, D., & Fenicia, F. (2011). Pursuing the method of multiple working hypotheses for hydrological modeling. *Water Resources Research*, 47, W09301. <https://doi.org/10.1029/2010WR009827>
- Del Vecchio, J., DiBiase, R. A., Denn, A. R., Bierman, P. R., Caffee, M., & Zimmerman, S. R. (2018). Record of coupled hillslope and channel response to Pleistocene erosion and deposition in a sandstone headwater valley, central Pennsylvania. *Geological Society of America Bulletin*. <https://doi.org/10.1130/B31912.1>
- Du, E., Rhett Jackson, C., Klaus, J., McDonnell, J. J., Griffiths, N. A., Williamson, M. F., et al. (2016). Interflow dynamics on a low relief forested hillslope: Lots of fill, little spill. *Journal of Hydrology*, 534, 648–658. <https://doi.org/10.1016/j.jhydrol.2016.01.039>
- Duffy, C., Shi, Y., Davis, K., Slingerland, R., Li, L., Sullivan, P. L., et al. (2014). Designing a suite of models to explore critical zone function. *Procedia Earth and Planetary Science*, 10, 7–15. <https://doi.org/10.1016/j.proeps.2014.08.003>
- Dunn, S., Birkel, C., Tetzlaff, D., & Soulsby, C. (2010). Transit time distributions of a conceptual model: Their characteristics and sensitivities. *Hydrological Processes*, 24(12), 1719–1729. <https://doi.org/10.1002/hyp.7560>
- Faticchi, S., Vivoni, E. R., Ogden, F. L., Ivanov, V. Y., Mirus, B., Gochis, D., et al. (2016). An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. *Journal of Hydrology*, 537, 45–60. <https://doi.org/10.1016/j.jhydrol.2016.03.026>
- Fenicia, F., Kavetski, D., Savenije, H. H., & Pfister, L. (2016). From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions. *Water Resources Research*, 52, 954–989. <https://doi.org/10.1002/2015WR017398>
- Fenicia, F., Kavetski, D., Savenije, H. H. G., Clark, M. P., Schoups, G., Pfister, L., & Freer, J. (2014). Catchment properties, function, and conceptual model representation: Is there a correspondence? *Hydrological Processes*, 28(4), 2451–2467. <https://doi.org/10.1002/hyp.9726>
- Flueckinger, L. A. (1969). *Geology of a portion of the Allensville quadrangle, Centre and Huntingdon counties*. Pennsylvania: Commonwealth of Pennsylvania, State Planning Board, Bureau of Topographic and Geologic Survey.

- Geris, J., Tetzlaff, D., McDonnell, J., & Soulsby, C. (2015). The relative role of soil type and tree cover on water storage and transmission in northern headwater catchments. *Hydrological Processes*, 29(7), 1844–1860. <https://doi.org/10.1002/hyp.10289>
- Geris, J., Tetzlaff, D., & Soulsby, C. (2015). Resistance and resilience to droughts: hydrogeological controls on catchment storage and runoff response. *Hydrological Processes*, 29(21), 4579–4593. <https://doi.org/10.1002/hyp.10480>
- Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1999). Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *Journal of Hydrologic Engineering*, 4(2), 135–143. [https://doi.org/10.1061/\(ASCE\)1084-0699\(1999\)4:2\(135\)](https://doi.org/10.1061/(ASCE)1084-0699(1999)4:2(135))
- Hale, V. C., & McDonnell, J. J. (2016). Effect of bedrock permeability on stream base flow mean transit time scaling relations: 1. A multiscale catchment intercomparison. *Water Resources Research*, 52, 1358–1374. <https://doi.org/10.1002/2014wr016124>
- Heidari, P., Li, L., Jin, L., Williams, J. Z., & Brantley, S. L. (2017). A reactive transport model for Marcellus shale weathering. *Geochimica et Cosmochimica Acta*, 217(Supplement C), 421–440. <https://doi.org/10.1016/j.gca.2017.08.011>
- Hettinger, R. D. (2001). *Subsurface correlations and sequence stratigraphic interpretations of Lower Silurian Strata in the Appalachian Basin of northeast Ohio, southwest New York, and northwest Pennsylvania: The Survey*.
- Heuvelmans, G., Muys, B., & Feyen, J. (2004). Evaluation of hydrological model parameter transferability for simulating the impact of land use on catchment hydrology. *Physics and Chemistry of the Earth, Parts A/B/C*, 29(11), 739–747. <https://doi.org/10.1016/j.pce.2004.05.002>
- Hoagland, B., Russo, T. A., Gu, X., Hill, L., Kaye, J., Forsythe, B., & Brantley, S. L. (2017). Hyporheic zone influences on concentration-discharge relationships in a headwater sandstone stream. *Water Resources Research*, 53, 4643–4667. <https://doi.org/10.1002/2016WR019717>
- Hornberger, G. M., & Spear, R. C. (1981). Approach to the preliminary analysis of environmental systems. 12.
- Horton, R. E. (1936). Natural stream channel-storage. *Eos, Transactions American Geophysical Union*, 17(2), 406–415. <https://doi.org/10.1029/TR017i002p00406>
- Hrachowitz, M., Benettin, P., van Breukelen, B. M., Fovet, O., Howden, N. J. K., Ruiz, L., et al. (2016). Transit times—The link between hydrology and water quality at the catchment scale. *Wiley Interdisciplinary Reviews Water*, 3(5), 629–657. <https://doi.org/10.1002/wat2.1155>
- Hrachowitz, M., Savenije, H., Blöschl, G., McDonnell, J., Sivapalan, M., Pomeroy, J., et al. (2013). A decade of Predictions in Ungauged Basins (PUB)—A review. *Hydrological Sciences Journal*, 58(6), 1198–1255. <https://doi.org/10.1080/02626667.2013.803183>
- Jackson, C. R., Du, E., Klaus, J., Griffiths, N. A., Bitew, M., & McDonnell, J. J. (2016). Interactions among hydraulic conductivity distributions, subsurface topography, and transport thresholds revealed by a multitracer hillslope irrigation experiment. *Water Resources Research*, 52, 6186–6206. <https://doi.org/10.1002/2015WR018364>
- James, A. L., & Roulet, N. T. (2007). Investigating hydrologic connectivity and its association with threshold change in runoff response in a temperate forested watershed. *Hydrological Processes*, 21(25), 3391–3408. <https://doi.org/10.1002/hyp.6554>
- Jasechko, S., Kirchner, J. W., Welker, J. M., & McDonnell, J. J. (2016). Substantial proportion of global streamflow less than three months old. *Nature Geoscience*, 9(2), 126–129. <https://doi.org/10.1038/NGEO2636>
- Jasechko, S., Perrone, D., Befus, K. M., Cardenas, M. B., Ferguson, G., Gleeson, T., et al. (2017). Global aquifers dominated by fossil groundwaters but wells vulnerable to modern contamination. *Nature Geoscience*, 10(6), 425–429. <https://doi.org/10.1038/NGEO2943>
- 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>
- Jenny, H. (1941). *Factors of soil formation: A system of quantitative pedology*. New York: McGraw-Hill Book Company Inc.
- Jin, L., Andrews, D. M., Holmes, G. H., Lin, H., & Brantley, S. L. (2011). Opening the “black box”: Water chemistry reveals hydrological controls on weathering in the Susquehanna Shale Hills Critical Zone Observatory. *Vadose Zone Journal*, 10(3), 928–942. <https://doi.org/10.2136/vzj2010.0133>
- Jin, L., Ravella, R., Ketchum, B., Bierman, P. R., Heaney, P., White, T., & Brantley, S. L. (2010). Mineral weathering and elemental transport during hillslope evolution at the Susquehanna/Shale Hills Critical Zone Observatory. *Geochimica et Cosmochimica Acta*, 74(13), 3669–3691. <https://doi.org/10.1016/j.gca.2010.03.036>
- Kirby, C., Newson, M., & Gilman, K. (1991). Plynlimon research: The first two decades: Institute of Hydrology.
- Kirchner, J. W. (2009). Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward. *Water Resources Research*, 45, W02409. <https://doi.org/10.1029/2008WR006912>
- Kobierska, F., Jonas, T., Kirchner, J. W., & Bernasconi, S. M. (2015). Linking baseflow separation and groundwater storage dynamics in an alpine basin (Dammagletscher, Switzerland). *Hydrology and Earth System Sciences*, 19(8), 3681–3693. <https://doi.org/10.5194/hess-19-3681-2015>
- Kuentz, A., Arheimer, B., Hündecha, Y., & Wagener, T. (2017). Understanding hydrologic variability across Europe through catchment classification. *Hydrology and Earth System Sciences*, 21(6), 2863–2879. <https://doi.org/10.5194/hess-21-2863-2017>
- Kuntz, B. W., Rubin, S., Berkowitz, B., & Singha, K. (2011). Quantifying solute transport at the Shale Hills Critical Zone Observatory. *Vadose Zone Journal*, 10(3), 843–857. <https://doi.org/10.2136/vzj2010.0130>
- Lebedeva, M. I., Fletcher, R. C., & Brantley, S. L. (2010). A mathematical model for steady-state regolith production at constant erosion rate. *Earth Surface Processes and Landforms*, 35(5), 508–524. <https://doi.org/10.1002/esp.1954>
- Lehmann, P., Hinz, C., McGrath, G., Meerveld, T.-V., & McDonnell, J. (2007). Rainfall threshold for hillslope outflow: An emergent property of flow pathway connectivity. *Hydrology and Earth System Sciences*, 11(2), 1047–1063. <https://doi.org/10.5194/hess-11-1047-2007>
- Li, C. Z., Zhang, L., Wang, H., Zhang, Y. Q., Yu, F. L., & Yan, D. H. (2012). The transferability of hydrological models under nonstationary climatic conditions. *Hydrology and Earth System Sciences*, 16(4), 1239–1254. <https://doi.org/10.5194/hess-16-1239-2012>
- Li, L. (2019). Watershed reactive transport. In J. Druhan, & C. Tournassat (Eds.), *Reviews in Mineralogy & Geochemistry: Reactive Transport in Natural and Engineered Systems*, Mineralogical Society of America, (Vol. 85). Chapter 13, pp. 381–418
- Li, L., DiBiase, R. A., Del Vecchio, J., Marcon, V., Hoagland, B., Xiao, D., et al. (2018). The effect of lithology and agriculture at the Susquehanna Shale Hills Critical Zone Observatory. *Vadose Zone Journal*, 17(1). <https://doi.org/10.2136/vzj2018.03.0063>
- Li, L., Maher, K., Navarre-Sitchler, A., Druhan, J., Meile, C., Lawrence, C., et al. (2017). Expanding the role of reactive transport models in critical zone processes. *Earth-Science Reviews*, 165, 280–301. <https://doi.org/10.1016/j.earscirev.2016.09.001>

- Li, L., Steefel, C., Kowalsky, M., Englert, A., & Hubbard, S. (2010). Effects of physical and geochemical heterogeneities on mineral transformation and biomass accumulation during a biostimulation experiment at Rifle, Colorado. *Journal of Contaminant Hydrology*, 112, 45–63.
- Lin, H., Kogelmann, W., Walker, C., & Bruns, M. (2006). Soil moisture patterns in a forested catchment: A hydrogeological perspective. *Geoderma*, 131(3), 345–368. <https://doi.org/10.1016/j.geoderma.2005.03.013>
- Lin, H., & Zhou, X. (2008). Evidence of subsurface preferential flow using soil hydrologic monitoring in the Shale Hills catchment. *European Journal of Soil Science*, 59(1), 34–49. <https://doi.org/10.1111/j.1365-2389.2007.00988.x>
- Lynch, J. A. (1976). Effects of antecedent soil moisture on storm hydrographs. Retrieved from
- Ma, L., Teng, F.-Z., Jin, L., Ke, S., Yang, W., Gu, H.-O., & Brantley, S. L. (2015). Magnesium isotope fractionation during shale weathering in the Shale Hills Critical Zone Observatory: Accumulation of light Mg isotopes in soils by clay mineral transformation. *Chemical Geology*, 397, 37–50. <https://doi.org/10.1016/j.chemgeo.2015.01.010>
- McCabe, M. F., Rodell, M., Alsdorf, D. E., Miralles, D. G., Uijlenhoet, R., Wagner, W., et al. (2017). The future of Earth observation in hydrology. *Hydrology and Earth System Sciences*, 21(7), 3879–3914. <https://doi.org/10.5194/hess-21-3879-2017>
- McDonnell, J. J. (2003). Where does water go when it rains? Moving beyond the variable source area concept of rainfall-runoff response. *Hydrological Processes*, 17(9), 1869–1875. <https://doi.org/10.1002/hyp.5132>
- McDonnell, J. J., & Woods, R. (2004). On the need for catchment classification. *Journal of Hydrology*, 299, 2–3. <https://doi.org/10.1016/j.jhydrol.2004.09.003>
- McGlynn, B. L., & McDonnell, J. J. (2003). Quantifying the relative contributions of riparian and hillslope zones to catchment runoff. *Water Resources Research*, 39(11), 1310. <https://doi.org/10.1029/2003WR002091>
- McGlynn, B. L., McDonnell, J. J., Seibert, J., & Kendall, C. (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>
- McGuire, K. J., McDonnell, J. J., Weiler, M., Kendall, C., McGlynn, B. L., Welker, J. M., & Seibert, J. (2005). The role of topography on catchment-scale water residence time. *Water Resources Research*, 41, W05002. <https://doi.org/10.1029/2004WR003657>
- McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2), 239–245. <https://doi.org/10.1080/00401706.1979.10489755>
- 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>
- McNamara, J. P., Tetzlaff, D., Bishop, K., Soulsby, C., Seyfried, M., Peters, N. E., et al. (2011). Storage as a metric of catchment comparison. *Hydrological Processes*, 25(21), 3364–3371. <https://doi.org/10.1002/hyp.8113>
- Montanari, A., Young, G., Savenije, H. H. G., Hughes, D., Wagener, T., Ren, L. L., et al. (2013). “Panta Rhei—Everything Flows”: Change in hydrology and society—The IAHS Scientific Decade 2013–2022. *Hydrological Sciences Journal*, 58(6), 1256–1275. <https://doi.org/10.1080/02626667.2013.809088>
- Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3), 885–900. <https://doi.org/10.13031/2013.23153>
- Naithani, K. J., Baldwin, D. C., Gaines, K. P., Lin, H., & Eissenstat, D. M. (2013). Spatial distribution of tree species governs the spatio-temporal interaction of leaf area index and soil moisture across a forested landscape. *PLoS ONE*, 8(3), e58704. <https://doi.org/10.1371/journal.pone.0058704>
- 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](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>
- Payn, R., Gooseff, M., McGlynn, B., Bencala, K., & Wondzell, S. (2009). Channel water balance and exchange with subsurface flow along a mountain headwater stream in Montana, United States. *Water Resources Research*, 45, W11427. <https://doi.org/10.1029/2008WR007644>
- Plister, L., Martínez-Carreras, N., Hissler, C., Klaus, J., Carrer, G. E., Stewart, M. K., & McDonnell, J. J. (2017). Bedrock geology controls on catchment storage, mixing, and release: A comparative analysis of 16 nested catchments. *Hydrological Processes*, 31(10), 1828–1845.
- Pilgrim, D. H., Cordery, I., & Baron, B. C. (1982). Effects of catchment size on runoff relationships. *Journal of Hydrology*, 58(3), 205–221. [https://doi.org/10.1016/0022-1694\(82\)90035-X](https://doi.org/10.1016/0022-1694(82)90035-X)
- Qu, Y., & Duffy, C. J. (2007). A semidiscrete finite volume formulation for multiprocess watershed simulation. *Water Resources Research*, 43, W08419. <https://doi.org/10.1029/2006WR005752>
- Reinhardt, L., & Ellis, M. A. (2015). The emergence of topographic steady state in a perpetually dynamic self-organized critical landscape. *Water Resources Research*, 51, 4986–5003. <https://doi.org/10.1002/2014WR016223>
- Robinson, J. S., Sivapalan, M., & Snell, J. D. (1995). On the relative roles of hillslope processes, channel routing, and network geomorphology in the hydrologic response of natural catchments. *Water Resources Research*, 31(12), 3089–3101. <https://doi.org/10.1029/95WR01948>
- Schaller, M. F., & Fan, Y. (2009). River basins as groundwater exporters and importers: Implications for water cycle and climate modeling. *Journal of Geophysical Research*, 114, D04103. <https://doi.org/10.1029/2008JD010636>
- Seeger, S., & Weiler, M. (2014). Reevaluation of transit time distributions, mean transit times and their relation to catchment topography. *Hydrology and Earth System Sciences*, 18(12), 4751–4771. <https://doi.org/10.5194/hess-18-4751-2014>
- Seibert, J., Bishop, K., Nyberg, L., & Rodhe, A. (2011). Water storage in a till catchment. I: Distributed modelling and relationship to runoff. *Hydrological Processes*, 25(25), 3937–3949. <https://doi.org/10.1002/hyp.8309>
- Shi, Y., Baldwin, D. C., Davis, K. J., Yu, X., Duffy, C. J., & Lin, H. (2015). Simulating high-resolution soil moisture patterns in the Shale Hills watershed using a land surface hydrologic model. *Hydrological Processes*, 29(21), 4624–4637. <https://doi.org/10.1002/hyp.10593>
- Shi, Y., Davis, K. J., Duffy, C. J., & Yu, X. (2013). Development of a coupled land surface hydrologic model and evaluation at a critical zone observatory. *Journal of Hydrometeorology*, 14(5), 1401–1420. <https://doi.org/10.1175/JHM-D-12-0145.1>
- Shi, Y., Davis, K. J., Zhang, F., & Duffy, C. J. (2014). Evaluation of the parameter sensitivities of a coupled land surface hydrologic model at a critical zone observatory. *Journal of Hydrometeorology*, 15(1), 279–299. <https://doi.org/10.1175/JHM-D-12-0177.1>
- Shi, Y., Eissenstat, D. M., He, Y., & Davis, K. J. (2018). Using a spatially-distributed hydrologic biogeochemistry model with a nitrogen transport module to study the spatial variation of carbon processes in a Critical Zone Observatory. *Ecological Modelling*, 380, 8–21. <https://doi.org/10.1016/j.ecolmodel.2018.04.007>
- Singh, J., Knapp, H. V., Arnold, J., & Demissie, M. (2005). Hydrological modeling of the iroquois river watershed using HSPF and SWAT. In: Wiley Online Library.

- Sivapalan, M. (2003). Prediction in ungauged basins: A grand challenge for theoretical hydrology. *Hydrological Processes*, 17(15), 3163–3170. <https://doi.org/10.1002/hyp.5155>
- Sivapalan, M., Takeuchi, K., Franks, S., Gupta, V., Karambiri, H., Lakshmi, V., et al. (2003). IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. *Hydrological Sciences Journal*, 48(6), 857–880. <https://doi.org/10.1623/hysj.48.6.857.51421>
- Smith, L. A., Eissenstat, D. M., & Kaye, M. W. (2016). Variability in aboveground carbon driven by slope aspect and curvature in an eastern deciduous forest, USA. *Canadian Journal of Forest Research*, 47(2), 149–158. <https://doi.org/10.1139/cjfr-2016-0147>
- Smith, M. B., Seo, D.-J., Koren, V. I., Reed, S. M., Zhang, Z., Duan, Q., et al. (2004). The distributed model intercomparison project (DMIP): Motivation and experiment design. *Journal of Hydrology*, 298(1-4), 4–26. <https://doi.org/10.1016/j.jhydrol.2004.03.040>
- Smith, T., Hayes, K., Marshall, L., McGlynn, B., & Jencso, K. (2016). Diagnostic calibration and cross-catchment transferability of a simple process-consistent hydrologic model. *Hydrological Processes*, 30(26), 5027–5038. <https://doi.org/10.1002/hyp.10955>
- Soulsby, C., Tetzlaff, D., Rodgers, P., Dunn, S., & Waldron, S. (2006). Runoff processes, stream water residence times and controlling landscape characteristics in a mesoscale catchment: an initial evaluation. *Journal of Hydrology*, 325(1-4), 197–221.
- Spence, C. (2010). A paradigm shift in hydrology: Storage thresholds across scales influence catchment runoff generation. *Geography Compass*, 4(7), 819–833. <https://doi.org/10.1111/j.1749-8198.2010.00341.x>
- Staudinger, M., Stoelzle, M., Seeger, S., Seibert, J., Weiler, M., & Stahl, K. (2017). Catchment water storage variation with elevation. *Hydrological Processes*, 31(11), 2000–2015. <https://doi.org/10.1002/hyp.11158>
- Stockinger, M. P., Bogen, H. R., Lücke, A., Diekkrüger, B., Weiler, M., & Vereecken, H. (2014). Seasonal soil moisture patterns: Controlling transit time distributions in a forested headwater catchment. *Water Resources Research*, 50, 5270–5289. <https://doi.org/10.1002/2013WR014815>
- Tetzlaff, D., Seibert, J., & Soulsby, C. (2009). Inter-catchment comparison to assess the influence of topography and soils on catchment transit times in a geomorphic province; the Cairngorm mountains, Scotland. *Hydrological Processes: An International Journal*, 23(13), 1874–1886. <https://doi.org/10.1002/hyp.7318>
- 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>
- van der Linden, S., & Woo, M.-k. (2003). Transferability of hydrological model parameters between basins in data-sparse areas, subarctic Canada. *Journal of Hydrology*, 270(3), 182–194. [https://doi.org/10.1016/S0022-1694\(02\)00295-0](https://doi.org/10.1016/S0022-1694(02)00295-0)
- van der Velde, Y., Heidbüchel, I., Lyon, S. W., Nyberg, L., Rodhe, A., Bishop, K., & Troch, P. A. (2015). Consequences of mixing assumptions for time-variable travel time distributions. *Hydrological Processes*, 29(16), 3460–3474. <https://doi.org/10.1002/hyp.10372>
- Van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal*, 44(5), 892–898. <https://doi.org/10.2136/sssaj1980.03615995004400050002x>
- Vorosmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., et al. (2010). Global threats to human water security and river biodiversity. *Nature*, 467(7315), 555–561. <https://doi.org/10.1038/nature09440>
- Wagner, T., Sivapalan, M., Troch, P., & Woods, R. (2007). Catchment classification and hydrologic similarity. *Geography Compass*, 1(4), 901–931. <https://doi.org/10.1111/j.1749-8198.2007.00039.x>
- Wagner, T., Sivapalan, M., Troch, P. A., McGlynn, B. L., Harman, C. J., Gupta, H. V., et al. (2010). The future of hydrology: An evolving science for a changing world. *Water Resources Research*, 46, W05301. <https://doi.org/10.1029/2009WR008906>
- Weiler, M., & McDonnell, J. (2004). Virtual experiments: A new approach for improving process conceptualization in hillslope hydrology. *Journal of Hydrology*, 285(1), 3–18. [https://doi.org/10.1016/S0022-1694\(03\)00271-3](https://doi.org/10.1016/S0022-1694(03)00271-3)
- Welch, L. A., & Allen, D. M. (2014). Hydraulic conductivity characteristics in mountains and implications for conceptualizing bedrock groundwater flow. *Hydrogeology Journal*, 22(5), 1003–1026. <https://doi.org/10.1007/s10040-014-1121-5>
- Western, A. W., Blöschl, G., & Grayson, R. B. (2001). Toward capturing hydrologically significant connectivity in spatial patterns. *Water Resources Research*, 37(1), 83–97. <https://doi.org/10.1029/2000WR900241>
- Western, A. W., Grayson, R. B., Blöschl, G., Willgoose, G. R., & McMahon, T. A. (1999). Observed spatial organization of soil moisture and its relation to terrain indices. *Water Resources Research*, 35(3), 797–810. <https://doi.org/10.1029/1998WR900065>
- Wittenberg, H. (1999). Baseflow recession and recharge as nonlinear storage processes. *Hydrological Processes*, 13(5), 715–726. [https://doi.org/10.1002/\(SICI\)1099-1085\(19990415\)13:5<715::AID-HYP775>3.0.CO;2-N](https://doi.org/10.1002/(SICI)1099-1085(19990415)13:5<715::AID-HYP775>3.0.CO;2-N)
- Wlostowski, A. N., Gooseff, M. N., McKnight, D. M., Jaros, C., & Lyons, W. B. (2016). Patterns of hydrologic connectivity in the McMurdo Dry Valleys, Antarctica: A synthesis of 20 years of hydrologic data. *Hydrological Processes*, 30(17), 2958–2975.
- Yu, X., Bhatt, G., Duffy, C., & Shi, Y. (2013). Parameterization for distributed watershed modeling using national data and evolutionary algorithm. *Computers & Geosciences*, 58, 80–90. <https://doi.org/10.1016/j.cageo.2013.04.025>
- Yu, X., Duffy, C., Baldwin, D. C., & Lin, H. (2014). The role of macropores and multi-resolution soil survey datasets for distributed surface-subsurface flow modeling. *Journal of Hydrology*, 516, 97–106. <https://doi.org/10.1016/j.jhydrol.2014.02.055>
- Zhang, Y., Slingerland, R., & Duffy, C. (2016). Fully-coupled hydrologic processes for modeling landscape evolution. *Environmental Modelling & Software*, 82, 89–107. <https://doi.org/10.1016/j.envsoft.2016.04.014>
- Zhi, W., Li, L., Dong, W., Brown, W., Kaye, J., Steefel, C., & Williams, K. H. (2019). Distinct source water chemistry shapes contrasting concentration-discharge patterns. *Water Resources Research*, 55, 4233–4251. <https://doi.org/10.1029/2018WR024257>
- Zhu, Q., Lin, H., & Doolittle, J. (2010). Repeated electromagnetic induction surveys for determining subsurface hydrologic dynamics in an agricultural landscape. *Soil Science Society of America Journal*, 74(5), 1750–1762. <https://doi.org/10.2136/sssaj2010.0055>
- Zimmer, M. A., & Gannon, J. P. (2018). Run-off processes from mountains to foothills: The role of soil stratigraphy and structure in influencing run-off characteristics across high to low relief landscapes. *Hydrological Processes*, 32(11), 1546–1560. <https://doi.org/10.1002/hyp.11488>
- Zimmer, M. A., & McGlynn, B. L. (2017). Ephemeral and intermittent runoff generation processes in a low relief, highly weathered catchment. *Water Resources Research*, 53, 7055–7077. <https://doi.org/10.1002/2016WR019742>
- Zreda, M., Desilets, D., Ferré, T., & Scott, R. L. (2008). Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons. *Geophysical Research Letters*, 35, L21402. <https://doi.org/10.1029/2008GL035655>