Controls on streamwater age in a saturation overland flow-dominated catchment

Dana A Lapides[1], W. Jesse Hahm[2], Daniella Rempe[3], William E Dietrich[4], and David N Dralle[5]

[1]Pacific Southwest Research Station, United States Forest Service — Simon Fraser University
[2]Simon Fraser University
[3]University of Texas at Austinn
[4]Universith of California Berkeley
[5]Pacific Southwest Research Station, United States Forest Service

November 22, 2022

Abstract

Water age and flow pathways should be related; however, it is still generally unclear how integrated catchment runoff generation mechanisms result in streamflow age distributions at the outlet. Here, we combine field observations of runoff generation at the Dry Creek catchment with StorAge Selection (SAS) age models to explore the relationship between streamwater age and runoff pathways. Dry Creek is a 3.5 km2 catchment in the Northern California Coast Ranges with a Mediterranean climate, and, despite an average rainfall of ˜1,800 mm/yr, is an oak savannah due to the limited water storage capacity. Runoff lag to peak—after initial seasonal wet-up—is rapid (˜1-2 hours), and total annual streamflow consists predominantly of saturation overland flow, based on field mapping of saturated extents and an inferred runoff threshold for the expansion of saturation extent beyond the geomorphic channel. SAS modeling based on daily isotope sampling reveals that streamflow is typically older than one day. Because streamflow is mostly overland flow, this means that a significant portion of overland flow must not be event-rain but instead derive from older, non-event groundwater returning to the surface, consistent with field observations of exfiltrating head gradients, return flow through macropores, and extensive saturation days after storm events. We conclude that even in a landscape with widespread overland flow, runoff pathways may be longer and slower than anticipated. Our findings have implications for the assumptions built into widely used hydrograph separation inferences, namely, the assumption that overland flow consists of new (event) water.

Controls on streamwater age in a saturation overland flow-dominated catchment

Dana A. Lapides[1,2], W. Jesse Hahm[1], Daniella M. Rempe[3], William E. Dietrich[4], David N. Dralle[2]

[1]Department of Geography, Simon Fraser University, Burnaby, BC, Canada

[2]Pacific Southwest Research Station, United States Forest Service, Davis, CA, USA

[3]University of Texas, Austin, Austin, TX, USA

[4]University of California, Berkeley, Berkeley, CA, USA

Key Points:

- Field observations of surface flow, groundwater and saturated extents indicate that saturation overland flow dominates streamflow
- Stable isotope tracers show that stream water age decreases as streamflow increases
- Streamflow is nevertheless mainly water greater than one day old, meaning that even overland flow is mostly not event water

Corresponding author: Dana A. Lapides, `dlapides@sfu.ca`

Abstract

Water age and flow pathways should be related; however, it is still generally unclear how integrated catchment runoff generation mechanisms result in streamflow age distributions at the outlet. Here, we combine field observations of runoff generation at the Dry Creek catchment with StorAge Selection (SAS) age models to explore the relationship between streamwater age and runoff pathways. Dry Creek is a 3.5 km^2 catchment in the Northern California Coast Ranges with a Mediterranean climate, and, despite an average rainfall of ≈1,800 mm/yr, is an oak savannah due to the limited water storage capacity. Runoff lag to peak—after initial seasonal wet-up—is rapid (≈ 1-2 hours), and total annual streamflow consists predominantly of saturation overland flow, based on field mapping of saturated extents and an inferred runoff threshold for the expansion of saturation extent beyond the geomorphic channel. SAS modeling based on daily isotope sampling reveals that streamflow is typically older than one day. Because streamflow is mostly overland flow, this means that a significant portion of overland flow must not be event-rain but instead derive from older, non-event groundwater returning to the surface, consistent with field observations of exfiltrating head gradients, return flow through macropores, and extensive saturation days after storm events. We conclude that even in a landscape with widespread overland flow, runoff pathways may be longer and slower than anticipated. Our findings have implications for the assumptions built into widely used hydrograph separation inferences, namely, the assumption that overland flow consists of new (event) water.

Plain Language Summary

Streams that respond most rapidly to rainfall tend to be fed by a process called overland flow. This study uses high-frequency water tracking measurements to show that even in a watershed fed by overland flow, the water entering the stream during storm events tends to be older than the storm event causing the stream response. Hydrologic measurements made during storm events reveal that water travels through the subsurface before re-emerging as surface flow. The interaction between storm event water and subsurface soils and weathered bedrock likely lead to mixing such that the water entering the stream contains a substantial fraction of water from previous storm events.

1 Introduction

Do distinct runoff pathways give rise to particular streamwater age distributions? Younger streamflow should derive from shorter or faster pathways such as overland flow, whereas older streamflow should derive from longer or slower pathways such as subsurface flow. Streamflow volumes can closely match precipitation input volumes over short timescales (hours-days), but there is widespread evidence–based on early isotopic evidence (e.g., Neal & Rosier, 1990; M. Sklash, 1990; Buttle, 1994) and more recent two-component hydrograph separation approaches (e.g., Freyberg et al., 2018), and fractal (e.g., Kirchner et al., 2000; Godsey et al., 2010) and StorAge Selection (SAS) (e.g., Benettin et al., 2017; Visser et al., 2019; Rodriguez & Klaus, 2019) modeling studies–that stormflow can consist of non-event, older water (sometimes years old) displaced by or driven out of subsurface storage by new water (e.g., Hewlett & Hibbert, 1967). This phenomenon indicates that the celerity of a hydraulic perturbation (e.g. a rainfall event) that triggers a runoff response is much faster than the velocity of water in most catchments (e.g., Wilusz et al., 2020; McDonnell & Beven, 2014). Further evidence for the predominance of old water in streamflow comes from the widespread observation that streams are enriched in cations relative to precipitation and commonly exhibit chemostasis (solute concentrations that are relatively invariant compared to flow) across a range of climates, lithologies, and runoff generation types (Godsey et al., 2009), indicative of the release of water that has resided in the catchment sufficiently long to acquire a characteristic solute concentration. (This timescale may be fairly short in some landscapes, however, if chemical evolution of waters in the vadose zone occurs rapidly; H. Kim et al., 2017; Anderson et al., 2002).

One way to produce young (and dilute) streamwater is for rain to reach the stream by flowing over the ground surface as overland flow (Elsenbeer et al., 1994; Elsenbeer & Lack, 1996; Shanley et al., 2002). Relatively abrupt declines of major cation concentrations have been observed at a saturation-overland flow (SOF) prone catchment (Dry Creek) at runoff rates of around 10 mm/day (W. J. Hahm et al., 2017). Nevertheless, even above these high flow rates when SOF dominated streamflow, perfect dilution of streamwater with rainwater was not observed. This observation suggests either rapid cation exchange reactions that increased the solute concentration of incoming rain as it flowed over the surface (e.g., H. Kim et al., 2017), or significant contribution of relatively high-solute concentration older water to streamflow. These alternative mechanisms are closely related to whether the source of the streamflow generated from SOF is event rain water or pre-event stored water.

SOF occurs when the water table rises from below and intersects the ground surface; the overland component of flow derives both from exfiltrating groundwater (return flow) and direct precipitation on saturated areas (DPSA) (Dunne & Black, 1970a, 1970b; Eshleman et al., 1993). Because the water table is dynamic, the area contributing to SOF can vary over time, which has been referred to as the ‘variable source area’ concept (Dunne & Black, 1970b; Wilson & Dietrich, 1987). SOF commonly occurs within convergent zones above channel heads (Dunne & Black, 1970b; Dunne, 1978; Kidron, 2021) and at the riparian-hillslope interface due to a rapid conversion of the tension saturated zone to atmospheric pressure with a small amount of added moisture from infiltration (Abdul & Gillham, 1984). SOF has also been documented to occur where small-scale heterogeneities in bedrock properties result in local exfiltrating head gradients (Wilson

& Dietrich, 1987). In essence, SOF routes flow over the land surface when the subsurface flow capacity is overwhelmed; this interpretation is commonly reflected in hydrological models, where all water in excess of a shallow subsurface flow capacity threshold is routed to surface flow (e.g., Beven & Kirkby, 1979; Litwin et al., 2020). Thus, the age of SOF water should reflect the dominant source of that runoff, either from the subsurface via return flow (consisting of a mixture of relatively old, pre-event water and event water that has infiltrated) or direct precipitation on saturated areas (DPSA, consisting exclusively of newly arriving event water) that never infiltrates.

The relationship between hillslope runoff generation and the integrated age distribution at the catchment outlet is still largely opaque because few studies have evaluated travel time distribution models in places where runoff generation mechanisms have been directly documented (Wilusz et al., 2020; Rodriguez et al., 2018; Benettin et al., 2017; Putnam et al., 2018). Resolving the impact of runoff generation mechanisms on age distributions would help to address the issue of equifinality in transit time distribution modeling and aid in the interpretation of the controls on stream geochemistry (Li et al., 2020; Torres & Baronas, 2021). Recently, Wilusz et al. (2020) used particle tracking to assess the relationship between runoff generation and transit times, while Rodriguez et al. (2018) compared modeled transit times using a conceptual model of catchment hydrology to empirically calculated transit times with good agreement. Benettin et al. (2017) found that little streamflow throughout the year was younger than 10 days at the Bruntland Burns site in Scotland, where saturation overland flow occurs on relatively flat peat-covered areas. Putnam et al. (2018) found that quickflow–which was primarily generated by SOF–was older than event water (i.e., water that derives from the driving rainfall) at the Pond Branch Catchment in Maryland. M. G. Sklash and Farvolden (1979) found that specific conductance and isotopic composition of overland flow water at the Hillman Creek watershed in Ontario, Canada, implied a strong contribution from groundwater. These findings suggest that SOF can be made up primarily of return flow, but controls on the relative fraction of pre-event and event water in SOF remain poorly understood.

Water transit time distributions (TTDs) describe the distribution of water ages in fluxes exiting a catchment control volume (e.g., Haggerty et al., 2002; Rodhe et al., 1996; Małoszewski & Zuber, 1982). Recently, StorAge Selection (SAS) functions have emerged as a tool for estimating TTDs directly from tracer data with minimal prior assumptions (Botter et al., 2011; Van Der Velde et al., 2012; Harman, 2015). SAS functions define what fraction of outflows (e.g., evapotranspiration and streamflow) derive from different water ages in storage. The SAS function framework is grounded in a catchment mass balance; the integrated collection of water ages in storage gives rise to an observed tracer timeseries in effluxes via preferential 'selection' of different storage ages. Studies have found that SAS functions vary through time as a function of catchment state (e.g., Benettin et al., 2017; Harman, 2015; M. Kim et al., 2016), and that streamflow SAS functions tend to show a preference for younger storage water at wetter states, termed the inverse storage effect (ISE) (e.g., Harman, 2015; Benettin et al., 2017).

Here, we combine field observations at the intensively monitored Dry Creek catchment in Northern California with water age modeling using SAS functions to evaluate how SOF mechanisms impact water ages in streamflow. We interpret catchment-integrated isotopic signals in streamflow with intensive field observations of water storage dynamics, runoff generation, saturated extent, groundwater levels, and head gradients. Specifically, we address the following questions:

1. How old is streamflow in a saturation overland flow-dominated catchment?
2. How does the portion of event water in streamflow change as the dominant runoff generation mechanism shifts through storm events?
3. Using transit time models and field observations of runoff generation, what portion of saturation overland flow comes from return flow vs. direct precipitation on saturated areas (DPSA)?

We found that 75% of streamflow is younger than 32 days on average with only 25% of streamflow younger than 3 days, and that at high flow states, ≈10% of flow derives from the 10% youngest water in storage (on average < 16 hours old). Field observations reveal the presence of return flow on the landscape. Comparison between the calculated fraction of SOF in streamflow from field observations and modeled fraction of streamflow younger than 1 day from SAS modeling revealed that the majority of overland flow must be older than 1 day. By estimating direct precipitation on saturated areas (DPSA), we found that even most DPSA must subsequently follow a subsurface pathway. These findings indicate that SOF is predominantly composed of return flow and allows us to set a lower bound on the fraction of pre-event water in SOF.

2 Methods

2.1 Study Site

The study catchment, Dry Creek (3.5 km^2; outlet at 39.5754°, -123.4642°) is in the Eel River watershed, in the Northern California Coast Ranges (Figure 1a)) about 200 km north of San Francisco, in the traditional territory of the Coast Yuki, the California Dene (Athabaskan), and Pomo (Johnson, 1979; Stewart, 1943; Foster, 1944; Baumhoff & Merriam, 1958). Dry Creek is within a ranch named Sagehorn, which has been part of the Eel River Critical Zone Observatory since 2015. The site experiences a Mediterranean climate, with a mean annual temperature of 13.3°C and mean annual precipitation of 1,800 mm (Group, 2013), almost all of which falls as rain between October-May.

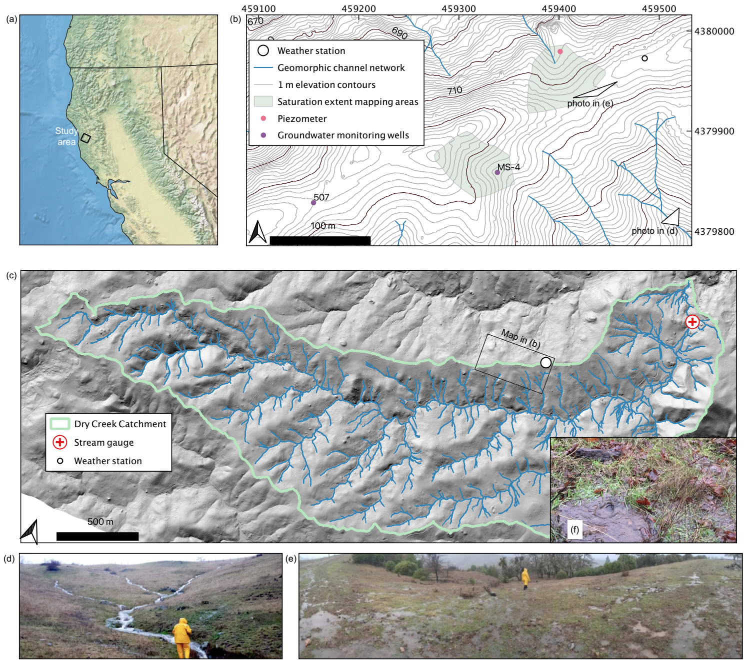

Figure 1. a) Location map of study site in the Northern California Coast Ranges, on Natural Earth hillshade layer. b) Map showing study ridge, with lidar-derived 1-m (thin lines) and 10-m (bold lines) elevation contours. c) Map showing Dry Creek catchment, on lidar-derived hillshade. Blue lines mark the streamflow network calculated from the 1-m DEM. d) Photo of flowing gully network during storm event. e) Panoramic photo of saturated study ridge during storm event. f) Visible return flow through a macropore.

The site is underlain by the Central belt mélange of the Franciscan complex (Jayko et al., 1989). The mélange bedrock is a sheared argillageous matrix with embedded blocks of diverse lithologies, including greywacke (sandstone) and chert. Larger blocks of greywacke cover less than 15% of the site by surface exposure (Lovill et al., 2018). The primary mineralogy of the mélange matrix is quartz, microcline, albite, muscovite, chlorite, illite, titanite, minor gypsum, pumpellyite and lawsonite, and rare kaolinite and carbonate (Cloos, 1983; W. J. Hahm et al., 2019).

Soils developed on the mélange matrix are mollisols (Rittiman Jr & Thorson, 2001; W. J. Hahm et al., 2019). More than 50 pits and augered holes indicate that the soils are typically 50 cm thick (ranging from 30-70 cm), with an upper organic-rich O horizon and a lower clay-rich Bt horizon. Guelph permeameter measurements of saturated hydraulic conductivity document high conductivities in the near surface that are similar to the maximum recorded rainfall intensities (Dralle et al., 2018). Pervasive animal burrowing and plant rooting has resulted in abundant macroporosity in the upper portion of the soil.

Deep drilling across the site (locations denoted with groundwater monitoring wells mapped in Figure 1b, all well locations shown in W. J. Hahm et al., 2019) revealed that the *in situ* mélange beneath the soils is seasonally unsaturated and weathered to depths of 2 - 4 m (W. J. Hahm et al., 2019), with abundant yellow-red oxidation. Below this depth, the parent material is permanently saturated, blue-black in hue, and has extremely low hydraulic conductivity.

Dry Creek drains to the east through a hilly landscape (mean gradient of 28%) typical of the Central belt mélange. A dense gully network is incised into inactive, deep-seated earthflows that have given the site a 'melted ice-cream' appearance (Kelsey, 1978). Grazing by sheep (historically) and cattle (modern) has been relatively light, and no terracettes have formed. The geomorphic channel drainage network (defined by channels with banks and clear elevation contour indentations visible on bare-earth lidar-derived maps) is shown in Figure 1c, and has a relatively high density of 16.9 km/km^2, with an average upslope contributing area of 1,085 m^2 at channel heads (Lovill et al., 2018). Hillslopes are convex-up, with typical divide-to-channel horizontal distances of 10 - 20 m (Figure 1). Dry Creek's catchment-averaged denudation rate, inferred from cosmogenic nuclides in quartz stream sediment, is 0.12 mm/yr (W. J. Hahm et al., 2019). The region has been uplifting and eroding for the past 3 Ma, with the emergence of the Northern California Coast Ranges from sea-level accompanying the northward migration of the Mendocino Triple Junction (Lock et al., 2006; Atwater & Stock, 1998).

The plant community developed on the mélange matrix is an oak savanna (W. Hahm et al., 2017; W. J. Hahm et al., 2018), with primarily European annual herbaceous groundcover that senesces in the summer dry season and a patchy, sparse overstory of winter-deciduous Oregon White Oak (*Quercus garryana*).

2.2 Description of Hydrologic Field Monitoring Infrastructure

The National Center for Airborne Laser Mapping (NCALM) flew lidar at the site in 2015; a 1 m-pixel sized elevation DEM was used to generate the maps in Figure 1. A weather station on the ridgetop records precipitation with a Campbell Scientific TB4 tipping bucket gauge, and is corrected for wind-induced undercatch, as described in (W. J. Hahm et al., 2019). Stream stage is recorded at the outlet with a Solinst Levelogger pressure transducer, with local atmospheric correction. Stream gauging methods are described in (W. J. Hahm et al., 2019).

This study capitalizes on the substantial existing monitoring network at Dry Creek to explore SOF (W. J. Hahm et al., 2019, 2020). Nine groundwater monitoring wells were completed with continuously slotted PVC-wells and outfitted with Solinst Levelogger and Campbell Scientific CS451 pressure transducers to continuously monitor water table fluctuations; two years of groundwater levels for all wells are shown in (W. J. Hahm et al., 2019), and in this study data from two representative wells are used (MS4 and 507). We installed a 2.54 cm solid PVC piezometer (MNP3) via hand auger to a depth of 55 cm in a side-slope about 5 m horizontally above a channel head. The lowest 5 cm was slotted and screened, back-filled with sand and sealed with bentonite. A Solinst pressure transducer was used to monitor head, with 20 cm of casing stick-up above the ground surface to capture possible artesian conditions. Drilling observations revealed that the piezometer opening was below the Bt horizon (which was encountered at 35 cm depth), and within typical smeary, grey-yellow, clay-rich mélange matrix weathered bedrock.

2.3 Precipitation and Streamwater Stable Isotopic Composition

2.3.1 Collection

We measured the stable isotopic composition of hydrogen in both precipitation and stream water as a tracer for interpreting travel times. The isotope sampling program and analysis methods were first described in (W. J. Hahm et al., 2020) in a study of oak water sourcing dynamics. Starting December 10, 2015 through the end of the 2020 water year, precipitation samples were collected daily when sufficient precipitation had fallen, typically between 06:00-08:00, approximately 1.3 km west of the weather station in an open field at an elevation of 645 m.a.s.l, and stored in 30 mL HDPE bottles until analysis. When snow fell (which was rare), it was allowed to melt into the sample collector before sampling. Streamwater samples were collected from near the mouth of Dry Creek when water was present in the channel on a semi-periodic campaign basis that began in Fall 2015, followed by two complete years of daily sampling (typically between 8:00-9:00) during the 2018 and 2019 water years (sampling location = 39°34'22.57"N, 123°27'46.76"W; 3.5 km^2 drainage area). Groundwater samples were collected on a semi-periodic basis via bailer from two monitoring wells (MS4 and 507), from a depth ranging from the water table surface to 1 m below the water table surface.

2.3.2 Analysis

Following the same methodology as described in detail in (W. J. Hahm et al., 2020), all samples were analyzed at the UC Berkeley Center for Stable Isotope Biogeochemistry via Isotope Ratio Mass Spectroscopy on a Thermo Delta PLUS XL instrument. Data are expressed in per mil delta notation (‰) relative to Vienna Standard Mean Ocean Water (VSMOW): δD ‰= $(\frac{R_{sample}}{R_{standard}}-1)1000$, where R is the ratio between the heavy and light isotope (i.e., D to H). The long-term precision is 0.60‰ δD (W. J. Hahm et al., 2020).

To ensure the completeness of the rainfall isotope timeseries, we compared rainfall timeseries from the ridge-top weather station (Figure 1) with the set of timestamps on which precipitation was sampled. We identified all time intervals during the study period for which more than 5 mm of rain fell but no sample collection was recorded in the following 48 hours. These criteria were chosen so that rain events reasonably small enough to evaporate and/or transpire completely would not be detected and so that a rain event sampled the next day would not be recorded as missing. We identified 25 dates with missing data (compared to the existing record of 348 samples). Six of the missing samples were likely misplaced prior to sample analysis, and the remaining nineteen were not sampled. When samples were not collected, any rainfall would mix with samples in the following days until the next sample was collected; thus, the next sample collected would represent the average concentration in rainfall over the intervening rainfall events. We replaced missing dates for which no sample was taken with the next measured isotope value if the next sample was taken within 3 days (1 date).

To fill the remaining missing dates, we performed a linear regression between rainfall isotope concentrations at Sagehorn and the nearby Angelo Coast Range Reserve ('Angelo', 23 km northeast; sampling program is described in Oshun et al., 2016)). For all dates with missing Sagehorn rainfall isotope samples, we identified an Angelo rainfall sample as close in time to the missing sample as possible (no more than 2 days later) and used the linear relationship between Sagehorn and Angelo rainfall isotope data to fill in an appropriate value for the missing Sagehorn data. Only ten dates remained with missing data after this process, representing a negligible fraction of precipitation input during the study period.

2.4 Event Runoff Analysis

2.4.1 Lag to Peak

We quantified the lag from rainfall centroid to peak streamflow response for all storm events with well-defined beginnings and ends for both Dry Creek and for the topographically and geologically similar Hank Creek that neighbors Dry Creek to the north (see Lovill et al. (2018) for a map). Hank Creek has a 56% larger catchment area at the gauging location (see maps in Lovill et al., 2018)). The streamflow sensor sampling frequency is 15 minutes, which represents the precision of the analysis.

2.4.2 Runoff Ratio

Graphical hydrograph separation following the method of Hewlett and Hibbert (1967) was performed for 47 Dry Creek storm events spanning the 2016-2019 water years, to quantify how the amount of 'quickflow' generated (the streamflow generated in excess of pre-event 'baseflow') varies in relation to pre-event catchment storage state (quantified by the streamflow magnitude at the start of the event) and storm event size. Events were chosen in such a way that the hydrograph recession was not interrupted by a new rainfall event. As Latron et al. (2008) note, this hydrograph separation approach is arbitrary, and the water volumes separated are not interpreted in terms of runoff pathway origin or age via this method. Although more sophisticated hydrograph separation methods are available (e.g., Blume et al., 2007), the graphical approach is simple, has seen widespread and sustained use, and is presented here as a diagnostic that informs catchment rainfall response, similar to the lag-to-peak analysis. Here we also report the event runoff ratio (quickflow as a fraction of event precipitation).

2.5 Surface Saturation–Observations and Model

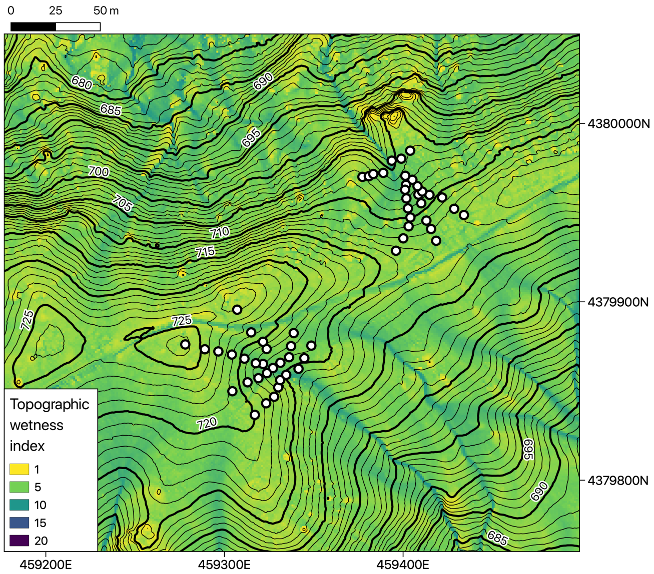

Figure 2. 1-m contour map of the ridge where surface saturation observations were performed. Background colors indicate topographic wetness index, and white circles mark the locations of saturation pits. See Equation 1 for the definition of topographic wetness index.

Over the course of a multi-day storm event in January 2018, surface saturation extents were mapped in two zero-order catchments straddling the northern ridge of Dry Creek (Figure 1). A total of 57 shallow saturation observation pits (see Figure 2) were dug to a depth of approximately two centimeters below the soil surface, and marked with flags to facilitate locating. At seven different times corresponding to a range of different flow values in the stream, the pits were logged as either saturated or not saturated, depending on whether or not a free water surface was observed in the pits, similar to the qualitative wetness classification presented in Rinderer et al. (2012). It was assumed that the presence of a free water surface indicated that the shallow water table at the site had intersected the ground surface at that point, thus potentially contributing to saturation overland flow.

A multi-variate logistic regression was then formulated using the observed saturation data to predict saturation state at all points within the catchment as a function of log-transformed discharge at the catchment outlet, and a topographic wetness index (TWI), calculated as:

$$TWI = \ln\left(\frac{a}{\tan\beta}\right) \tag{1}$$

where a [m] is contributing area per unit length contour (calculated using the r.flow module within GRASS GIS) and β was the topographic slope (Beven & Kirkby, 1979). Calculations were made at a 1 m length scale. Across the observation pits, TWI ranges from 3.1 to 8.8, with a median of 5.0. Across the landscape, the 10th, 50th, and 90th percentile of TWI values are 3.1, 4.6, and 6.4, respectively.

Using the logistic regression model for saturation, the stream discharge record, and a catchment-wide map of TWI, spatially explicit saturation extent maps were generated at all times throughout the period of flow record. At the catchment scale, saturation extent is reported as the percentage of points within the catchment classified as saturated at a given point in time. Note that at the catchment scale, saturated area is effectively a function of discharge in the stream since the spatial distribution of TWI in the catchment is constant. We quantify direct precipitation on saturated areas by multiplying instantaneous rainfall intensities by saturated areas determined from instantaneous streamflow.

2.6 StorAge Selection (SAS) Functions

SAS functions describe quantitatively how waters of different ages are selected from an age-ranked storage distribution to constitute a catchment efflux (ET or streamflow) (Botter et al., 2011; Van Der Velde et al., 2012; Harman, 2015). The basic mass balance is given as:

$$\frac{\delta S_T(T,t)}{\delta t} + \frac{\delta S_T(T,t)}{\delta T} = J(t) - Q(t)\Omega_Q(S_T(T,t),t) - ET(t)\Omega_{ET}(S_T(T,t),t), \tag{2}$$

where t is time [T] and T is age [T]; $S_T(T,t)$ [L] is the system age-rank storage; $J(t)$ [L/T] is precipitation input, $Q(t)$ [L/T] is streamflow output, and $ET(t)$ [L/T] is evapotranspiration output; Ω_Q [$\cdot$] and Ω_{ET} [$\cdot$] are SAS functions for Q and ET respectively that determine the output age cumulative distribution function given the age-rank storage at each time. The corresponding SAS functions ω_Q and ω_{ET} are the derivatives with respect to T of Ω_Q and Ω_{ET}. A boundary condition of $S_T(T = 0, t) = 0$ is assumed, and an initial storage $S_T(T, t = 0)$ must be parameterized. Since initial age-rank storage is never known, a spin-up period is used to identify a reasonable catchment state to use as the initial condition.

A conservative tracer can be used to constrain water age distributions in streamflow and evapotranspiration through the following relation:

$$C_Q(t) = \int_0^\infty C_S(T,t)\overleftarrow{p}_Q(T,t)dt, \tag{3}$$

$$C_{ET}(t) = \int_0^\infty C_S(T,t)\overleftarrow{p}_{ET}(T,t)dt, \tag{4}$$

where C_Q [·] and C_{ET} [·] are the concentrations of tracer in streamflow and ET respectively, C_S [·] is the distribution of tracer concentration in age-ranked storage, $\overleftarrow{p}_Q$ and $\overleftarrow{p}_{ET}$ are the backward transit time distributions in an output given by:

$$\overleftarrow{p}_Q(T,t) = \frac{\delta\Omega_Q(S_T,t)}{\delta S_T}\frac{\delta S_T}{\delta T}, \tag{5}$$

$$\overleftarrow{p}_{ET}(T,t) = \frac{\delta\Omega_{ET}(S_T,t)}{\delta S_T}\frac{\delta S_T}{\delta T}. \tag{6}$$

Streamflow, precipitation, and concentration inputs are derived from the field monitoring campaign. ET is estimated using the Hargreaves equation. The representation of evapotranspiration (ET) used to parameterize the SAS model likely does not fully capture the dynamics of ET in the Dry Creek catchment since storage trends upwards linearly over time. To correct for this, we adjusted ET based on a running mass balance:

$$S = P - ET - Q \tag{7}$$

where S [L/T] is dynamic catchment storage. Over long time periods, catchment storage should remain approximately constant, but due to errors in flux measurements, particularly ET, S grows quickly over time. To resolve this, we fit a linear trend to the storage at the end of each dry season and subtracted this trend from ET to ensure that S remains constant over long timeframes.

We followed the method described by Benettin and Bertuzzo (2018) to calculate the SAS function. Benettin and Bertuzzo (2018) provided a MATLAB implementation of the method, which we translated into the Python programming language (`https://www.python.org/`). An alternate Python implementation was developed by Harman et al. (2019). For a full description of the numerical methods used in this study, see Benettin and Bertuzzo (2018). The only difference is that in our implementation, we use a standard forward Euler numerical scheme, as opposed to the modified Euler method outlined by Benettin and Bertuzzo (2018). Although six options are available in our code, in this study we use a constant power law SAS function for ET:

$$\Omega_{ET} = \left(\frac{S_T(T,t)}{S(t)}\right)^{k_{ET}}, \tag{8}$$

where $S(t)$ is total storage and $k_{ET} \in (0, \infty)$ is a parameter. For the streamflow SAS function, we use a time-varying power law (Benettin et al., 2017):

$$\Omega_Q = \left(\frac{S_T(T,t)}{S(t)}\right)^{k_Q}, \tag{9}$$

where the parameter k_Q [·] varies between a minimum value k_{min_Q} and a maximum value k_{max_Q} with a log-dependence on wetness state wi:

$$k_Q = k_{min_Q} + (k_{max_Q} - k_{min_Q})log[(1 - \text{logfactor}_Q)wi] \tag{10}$$

where wi is the log-transformed instantaneous stream runoff normalized to the maximum log-transformed stream runoff at the outlet, and logfactor_Q [·] is a constant parameter. A time-varying power law has been shown to capture system dynamics well (Benettin et al., 2017), and a log dependence rather than a linear dependence provides more flexibility in how the catchment transitions from a wet to a dry state due to the addition of an extra parameter. We used the time period of October 1, 2017 to October 1, 2018 as a representative spin-up period repeated 10 times to generate an initial condition for age-rank storage. Model calibration was performed using all data through the 2019 water year, with the top 95th percentile of parameter sets retained. Model evaluation was performed on the 2020 water year to evaluate performance of these parameter sets.

Table 1. Parameters tuned in StorAge Selection model using Monte Carlo simulation.

Parameter	Definition
$kmin_Q$	Minimum exponent for Q SAS function as defined in Equation 10 [·]
$kmax_Q$	Maximum exponent for Q SAS function as defined in Equation 10 [·]
logfactor_Q	Scaling between $kmin_Q$ and $kmax_Q$ as defined in Equation 10
k_{ET}	ET SAS function power in Equation 8 [·]
S_0	Initial storage [mm]
C_{S_0}	Initial isotopic concentration in storage [δD‰]

We determined best-fit parameter sets by randomly sampling the parameter space (see Table 1 for a list of tuned parameters) via Monte Carlo simulation on 10,000 parameter sets. Parameter calibration was done using the set of collected data from October 1, 2016 through October 1, 2019. We evaluated model fit using the Nash-Sutcliffe model efficiency coefficient (NSE):

$$NSE = 1 - \frac{\Sigma_{t=1}^{t=t_0}\left(C_m^t - C_0^t\right)^2}{\Sigma_{t=1}^{t=t_0}\left(C_0^t - \bar{C_0}\right)^2}, \tag{11}$$

where time t ranges from the beginning ($t = 1$) to the end ($t = t_0$) of the model simulation, C_m^t is the modeled streamflow concentration at each time, C_0^t is the observed streamflow concentration at each time, and $\bar{C_0}$ is the mean of observed streamflow concentrations (Nash & Sutcliffe, 1970) and Kling-Gupta Efficiency (KGE):

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2}, \tag{12}$$

where r is the linear correlation coefficient between modeled and observed data, $\alpha = \sigma_m/\sigma_o$ is the ratio between modeled and observed standard deviation, and $\beta = (\bar{C_m} - \bar{C_o})/\sigma_o$. NSE> 0 or KGE> -0.41 indicates that the model performs better than a model defined as the mean of the data for all time (Knoben et al., 2019). After parameterization, performance was evaluated on data from October 1, 2019 to October 1, 2020. Previous SAS modeling studies which found model performance to be adequate have found maximum NSE ranging from 0.24 to 0.92 (Rodriguez et al., 2021; Rodriguez & Klaus, 2019; Harman, 2015; Benettin et al., 2017; Smith et al., 2018; Van Der Velde et al., 2012), and Kirchner (2003) suggested that a successful behavioral model has NSE>0.5 and KGE>0.3. We rank model performance by the product of NSE and KGE, with successful behavioral performance above 0.15. Using the top 95th percentile of parameter sets, we calculated ensemble means with 25th-75th percentile and 10th-90th percentile uncertainty ranges for: modeled isotope concentration, median storage and streamflow ages, fraction of streamflow younger than 1 day old, and fraction of streamflow that derives from the youngest 10th percentile of storage.

With a sampling interval of one day, it may be difficult to make robust claims about water ages at or below the daily timescale. Rodriguez and Klaus (2019) found that a composite SAS function was required to represent isotope dynamics on shorter timescales, a finding that suggests that a higher sampling rate could reveal inadequacies in the functional form of the SAS function used in this study that do not appear in our study as designed. Using a synthetic timeseries of stream isotope data with a high fraction of water younger than 1 day, we explored the impact of coarsening sampling frequency (unit, 2x, 4x, 8x, 16x) on model calibration results (Supplemental Information S4). We found that decreasing the sampling frequency from 1 to 2 or 4 days (coarsening by 2x or 4x) had a negligible impact on the estimated fraction of water younger than 1 day (unit frequency), indicating that the fraction of water younger than a unit frequency is fairly robust to coarsening in sampling frequency. Thus, a sampling interval of 1 day should be adequate to have confidence in fraction of water younger than 1 day (or even 12h or 6h).

3 Results

3.1 Catchment Hydrologic Response to Winter Storms

3.1.1 Hydrograph Features and Runoff Sources

At the end of the summer dry season, shallow and deep unsaturated soil moisture stores and weathered rock moisture are depleted at Dry Creek (W. J. Hahm et al., 2020). The first rains increase moisture content in the unsaturated zone without causing a groundwater response (Dralle et al., 2018). Groundwater responds after approximately 100 mm of cumulative rainfall, and about 200 mm is sufficient to raise water tables to or near the ground surface (Dralle et al., 2018). Storage then depletes at the start of the dry season, and, as its name implies, Dry Creek typically ceases to flow by late May or early June (Dralle et al., 2018; Lovill et al., 2018).

During storm events (example in Figure 3a), large volumes of water commonly exfiltrate via macropore flow (see Figure 4a), and artesian conditions and vertical head gradients are ob-

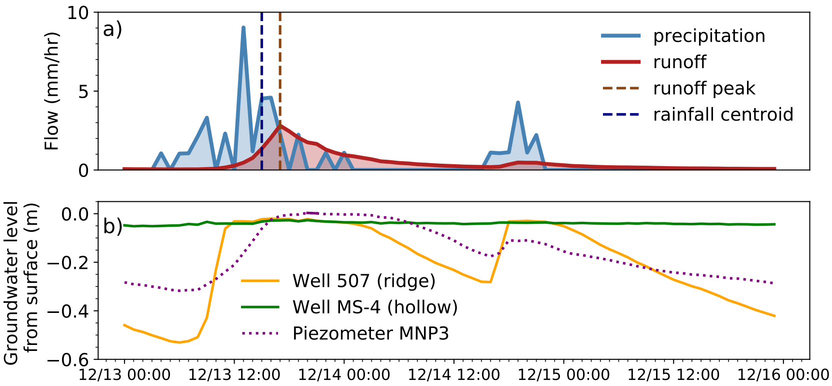

Figure 3. Hydrologic response at Dry Creek in response to a representative wet-season storms on December 12-15, 2018 with a runoff coefficient of 0.54 for the first event. (a) Streamflow is sampled at 15 minute intervals and precipitation is sampled at 5 minute intervals. Both are smoothed to hourly resolution. Lag to peak is 2 hours for the first event. (b) Concurrent groundwater response measured at two wells and one piezometer. Solid line in piezometer data indicates artesian head condition.

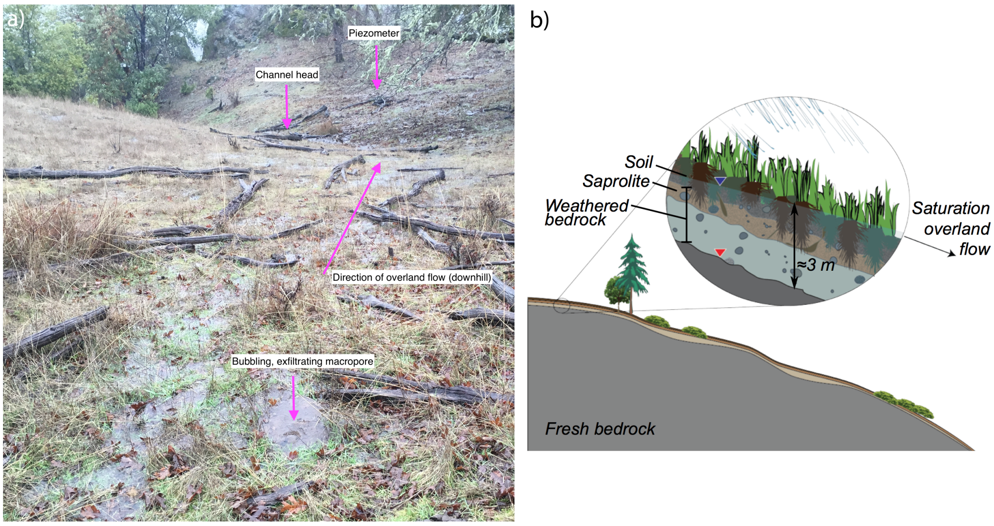

Figure 4. (a) Photo illustrating widespread saturation overland flow, an exfiltrating macropore, and the location of the piezometer on a hillslope above a channel head during a break in the rain on Jan. 17, 2016, 13:20, when the runoff in Dry Creek was 125 mm/day. (b) Conceptual cross-section of the critical zone in the Dry Creek watershed, showing relatively thin weathered zone (≈3 m), location of extreme end-member summer (red) and winter (blue) water table locations via inverted triangles, and runoff generation mechanisms. Modified from W. J. Hahm et al. (2019).

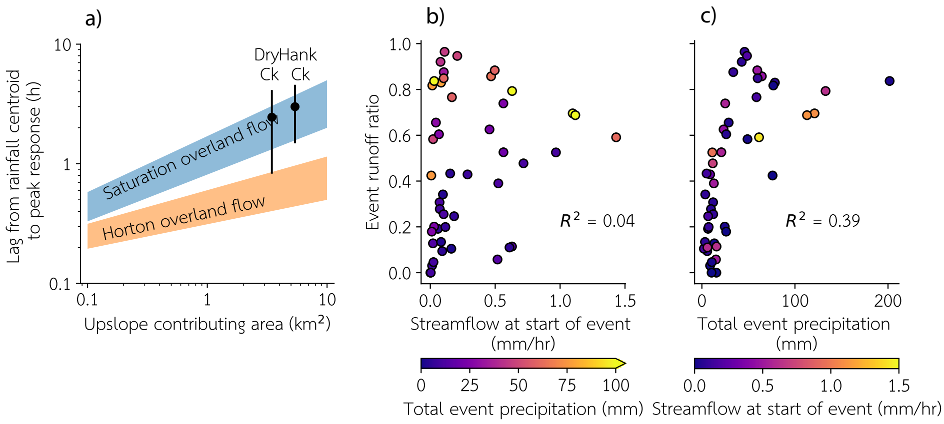

Figure 5. (a) Peak streamflow lag times from rain event centroids (mean±1 s.d.) as a function of drainage area, plotted on regions typical of two overland flow generation mechanisms. Shaded areas and plotting space from Dingman (2015), after Kirkby (1988), based on data from Dunne (1978). (b-c) Event-based runoff ratios at Dry Creek as a function of pre-event streamflow (b) and total event rainfall (c).

served in piezometers (solid line in piezometer data in Figures 3b). Periods of time with artesian head conditions represent a lower bound estimate of the times during which exfiltrating head gradients exist in the catchment. Winter runoff in Dry Creek is dominantly sourced from saturation overland flow (in the sense of Dunne & Black, 1970b; Dunne, 1978) and shallow subsurface flow in the weathered portion (upper few meters) of the subsurface, as illustrated schematically in Figure 4b (Dralle et al., 2018). The subsurface critical zone at Dry Creek consists of a 2 - 4 m thick layer of organic soils and clay-rich weathered bedrock matrix overlying unweathered, perennially saturated mélange, as shown in Figure 4a (W. J. Hahm et al., 2019). The shallow depth to fresh bedrock results in relatively small integrated porosity and water storage capacity, causing widespread saturation overland flow during the winter wet season.

Lag to peak and event runoff coefficients also support widespread SOF. Across analyzed storms, Dry Creek's lag to peak time was on average 2.5±1.6 h (± 1 s.d.), and neighboring Hank Creek's was 3.0±1.5 h, as shown in Figure 5a. These times are typical for catchments of comparable area experiencing saturation overland flow according to the commonly depicted timescales in Dingman (2015)'s *Physical Hydrology* textbook (after Kirkby (1988), based on data from Dunne (1978)). The event-based runoff ratio at Dry Creek is variable and spans the full range from 0 to 1 (Figure 5b-c). The runoff ratio is uncorrelated with the catchment storage state (wetness) at the start of a storm event, quantified via the streamflow just before the intial stream response (Figure 5b). In contrast, the total precipitation in the event explained 39% of the variance in runoff ratio, with events smaller than 25 mm generally producing runoff ratios less than 0.5, and events greater than 25 mm producing runoff ratios greater than 0.5 (Figure 5c).

3.1.2 Surface Saturation in Response to Storms

Saturation extent measured via discrete mapping campaigns correlated with discharge at the catchment outlet (Figure 6); as discharge decreased in both zero-order catchments (a-f and g-l), the number of saturated pits at both catchments decreased as well. These mapping campaigns spanned nearly the full range of discharge throughout the study period (Figure 7a), and the observation locations' TWI range closely matches that of the catchment at large. The logistic regression model shown in Figure 7a used to predict saturation as a function of catchment discharge and topographic wetness index (TWI) has an accuracy of 83% on observed data.

By applying the logistic regression model, we found that the dynamic extent of saturated area grows throughout a storm event and shrinks as the water table recedes from the surface (Figure 7b; Supplementary Video 1), with portions of the surface of the catchment remaining saturated and contributing to overland flow for days following a precipitation event. At runoff rates with the highest relative runoff contribution, the logistic regression model suggests that more than half of the catchment is saturated (Figure 7a). An instantaneous runoff rate of 2 mm/day at the catchment outlet (not shown) was the threshold above which saturation extends beyond the stream channel, according to the logistic regression model. At runoffs of 10 mm/day, saturation is widespread outside of the channel (Figure 7b). Based on these results, as well as field observations of overland flow corresponding to comparable catchment discharge states, we chose 5 mm/day (best estimate; likely range between 2-10 mm/day) as a threshold runoff rate that corresponds with the maximum subsurface flow capacity adjacent to the channel network, such that the streamflow rate above 5 mm/day derives mostly from saturation overland flow.

3.2 Isotope Dynamics

Isotopic composition of 267 precipitation samples, 460 streamflow samples, and 46 groundwater samples is shown for the full range of flow percentiles in Figure 8d. Streamflow isotopic compositions are markedly damped compared to precipitation, as demonstrated by the larger spread of precipitation isotopes (blue) than streamflow isotopes (red) in the timeseries and dual isotope plots of Figure 8. The sensitivity of streamwater isotopes to precipitation inputs over shorter timescale is shown in Figure 8. Individual samples of streamwater isotopic composition tend to follow a highly damped pattern shifting with the long-term mean, with some larger excursions in the direction of individual rainfall inputs. In general, the relationship between precipitation and streamflow isotopic composition can be highly variable on a storm-to-storm basis. In the zoomed-in view in panel b, streamflow isotopic composition can change little with a large rainfall input (first and

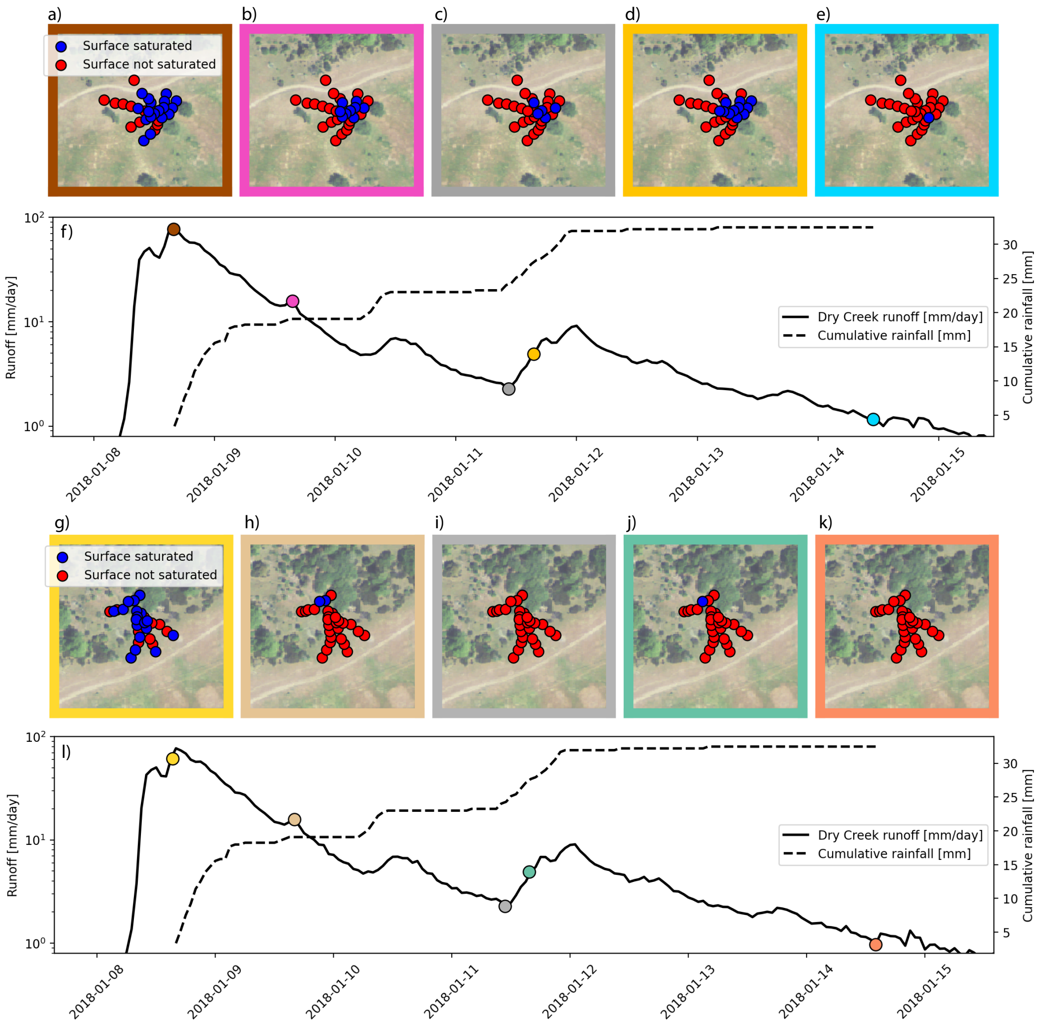

Figure 6. Observations of surface saturation during a streamflow recession in January 2018 at two zero-order catchments (top) in Dry Creek and (bottom) in Hank Creek, bordering Dry Creek. For a map contextualizing the location of the saturation pits, see Figure 2. Panels a-e and g-k show mapped saturation extent during each field visit. Border colors for each panel correspond to the dots with the same color in panel f (a-e) or l (g-k).

last large precipitation events) or be displaced significantly (as in the case of the large negative event) or even the very small negative events in February and March. There is no repeated annual temporal trend in precipitation isotopic composition, unlike the characteristic sinusoidal signature of many continental climates (e.g., DeWalle et al., 1997; Allen et al., 2018, 2019). Instead, we observed a large degree of intra-seasonal scatter in isotopic inputs.

At low discharge at the end of the wet season, streamflow samples show evidence of evaporative enrichment, likely due to evaporation of water in the stream channel during occasional

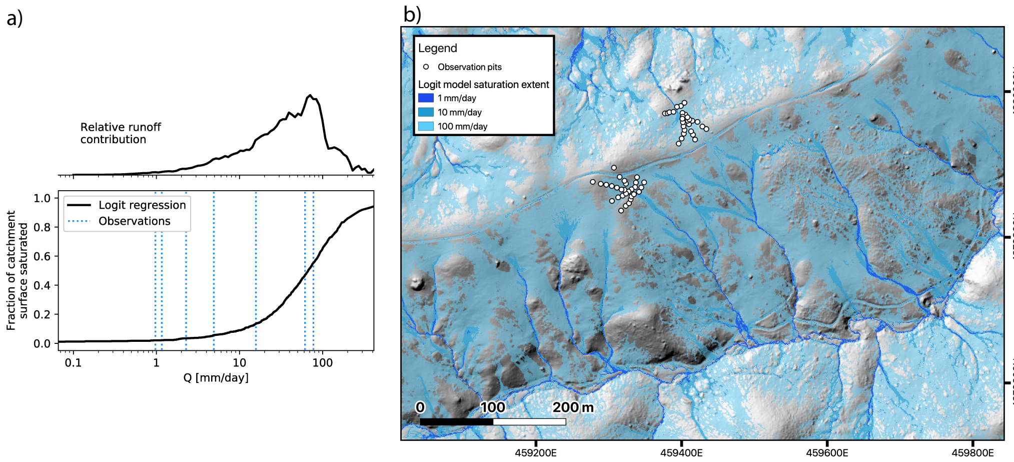

Figure 7. (a) Flow-weighted frequency (top) of instantaneous runoff magnitudes in the Dry Creek catchment. The 5th, 50th, and 95th percentiles flow-weighted frequencies are 20, 90, 320 mm/day, respectively. The median frequency-magnitude flow value coincides with times when a significant (approximately 60% by area) portion of the catchment is saturated, as predicted using the logistic regression model. (b) Saturation extent at different instantaneous streamflow rates. White points show where saturated/not saturated observations were made in field surveys across a range of instantaneous streamflow values. A logistic regression model was fitted using these observations, predicting saturated state at each point in the catchment as a function of log-transformed discharge and topographic wetness index. Blue transparencies over hillshade highlight saturation spatial extent at three discrete streamflow values. Uncolored areas are predicted to not be saturated at an instantaneous streamflow rate of 100 mm/day.

long gaps in rain coupled with high atmospheric temperatures. Since evaporative enrichment is not accounted for in the SAS model, we excluded such samples from the SAS fitting. We identified a flow threshold of 0.05 mm/day, above which all streamflow isotopic data fell on the meteoric water line. At flows below 0.05 mm/day, some streamflow samples were isotopically heavy and fell on a line with a slope shallower than the local meteoric water line (Supplemental Figure S8). While not all flows below 0.05 mm/day show an evaporative enrichment signal, this threshold provides a conservative means of excluding evaporative enrichment from calibration.

3.3 StorAge Selection Modeling

Figure 9 shows SAS modeling results for water year 2019. Results are similar for water year 2020, included in Supplemental Figure S6. Among the top 95th percentile of parameter sets, median NSE and KGE are 0.62 and 0.82, respectively. The range of NSE and KGE values among the top 95th percentile are 0.42 - 0.62 (NSE) and 0.82 - 0.83 (KGE). More details on model pa-

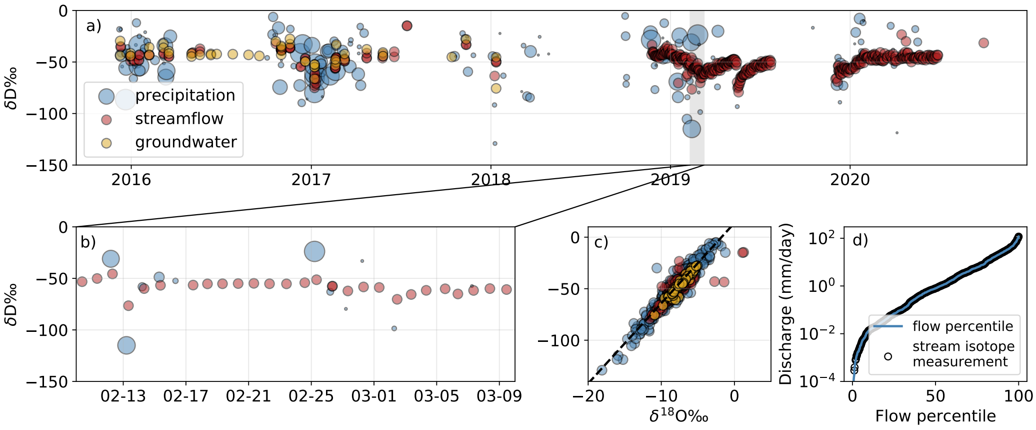

Figure 8. (a) Timeseries of 5 years of daily precipitation sampling, 3 years of episodic and 2 years of daily streamflow sampling, and episodic groundwater sampling with a zoomed-in view for 1 month in (b). In (a) and (b), precipitation isotope markers are scaled by the volume of daily precipitation when the sample was taken. (c) shows dual isotope space for all measurements, and (d) marks the time-weighted flow percentiles at which runoff was sampled.

rameterization can be found in the Supplemental Information S1. As shown in Figure 9b), the SAS model captures the moving average of streamflow isotope data, which shifts in time in response to precipitation inputs (Figure 9a); the model fails to capture the large negative daily excursions January and February and some small positive excursions in December and March. The unexplained large daily excursions suggest that higher temporal resolution in sampling could be beneficial. There is also a period of underestimated streamflow concentration in March-April of 2019, which may be due to a limitation in how the SAS model applies at drier catchment states. While the SAS model has six parameters, results are really only sensitive to two of these parameters (Supplemental Information S1), so additional flexibility in the model structure may be required to capture stream behavior in drier periods. White points, denoting when streamflow is <0.05 mm/day, were excluded from calibration and show an upward trend away from the model, consistent with significant evaporative enrichment (see Supplemental Information S3).

At the end of the dry season, the median ages of water in storage modeled using SAS functions (Figure 9c) are slightly larger than the length of the dry season (5-6 months). At the beginning of the wet season, median streamflow age modeled with SAS functions declines rapidly with high confidence (narrow shaded band) after a short period of rainfall. This timeframe should be be related to the time it takes to fill up approximately half of the catchment's dynamic storage capacity (although not identical since streamflow and ET draw preferentially from younger storage). Indeed, the drop in median storage age within the confidence interval occurs at around 150 mm

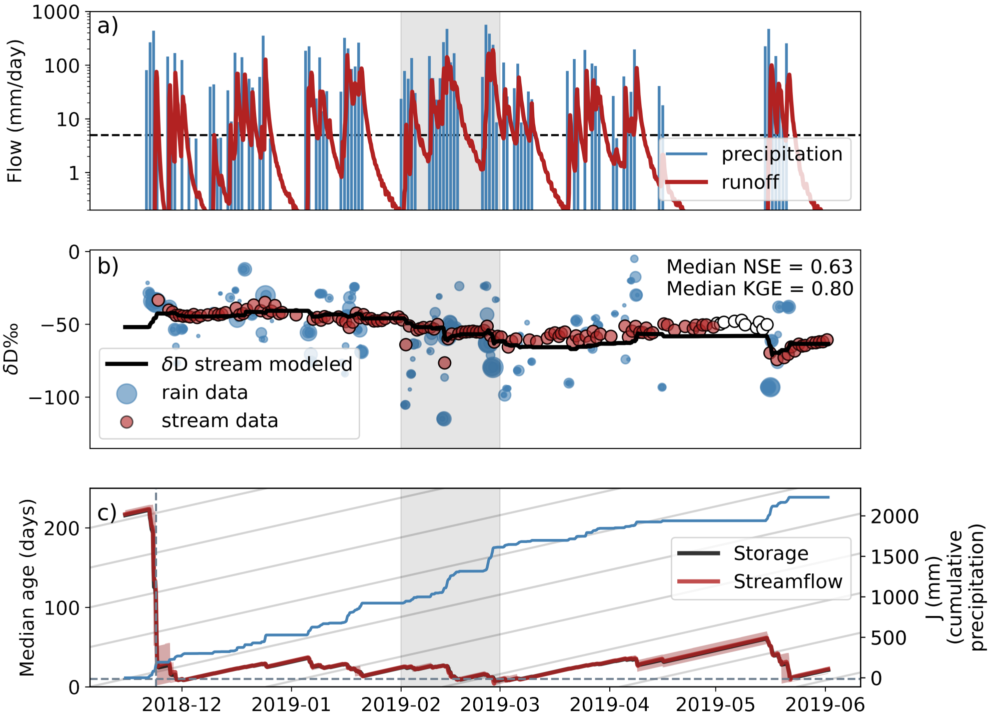

Figure 9. (a) Daily precipitation and instantaneous runoff throughout the wet season 2018-2019. Horizontal dashed black line marks the 5 mm flow threshold above which excess flow is assumed to be SOF. (b) Confidence bars on SAS model predictions (black line) are smaller than the width of the line. The size of plot markers for rainfall data (blue) are scaled by the volume of precipitation. Data shown in white circles are excluded from calibration of the SAS model due to in-channel evaporative enrichment (streamflow <0.05 mm/day). Marked median NSE and KGE are the median values among the top 95th percentile of parameter sets. (c) Shading around median ages indicates 25th-75th percentile of ensemble simulations, and blue line is cumulative precipitation. Storage and streamflow curves lie nearly on top of one another. Vertical dashed line marks cumulative precipitation of 150 mm, and horizontal dashed line marks a median age of 10 days. Shaded vertical bar indicates the timeframe shown in Figure 11.

of cumulative precipitation, just a bit more than half of the estimated approximate dynamic storage capacity of the landscape of 200 mm (Dralle et al., 2018; W. J. Hahm et al., 2019; Dralle et al., 2018). For nearly the whole wet season, median storage age is larger than 10 days (above the horizontal dashed line in Figure 9c).

Median ages of streamflow and storage modeled using SAS functions (Figure 9c) track one another closely throughout the wet season, falling on top of one another with overlapping confidence intervals. Storage age appears young since the storage modeled by the SAS function is

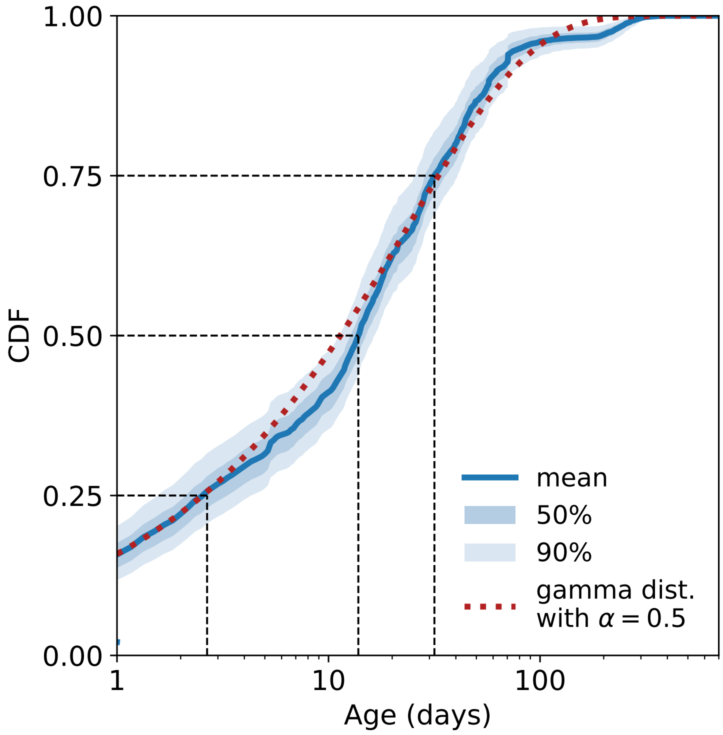

Figure 10. Ensemble mean of flow-weighted average cumulative age distribution function for Dry Creek. Shaded regions show the 25-75th percentiles (50%) and the 5th-95th percentiles (90%) respectively. 25th percentile, median, and 75th percentile of streamflow age are 3 days, 14 days, and 32 days, respectively.

only the dynamic portion of storage during the study period. Older storage may exist, but it accounts for only a small portion of dynamic storage so has essentially no impact on median ages.

Throughout the study period, the mean age distribution that results from the parameterized SAS function indicates that essentially all streamflow is younger than 1 year (Figure 10), the majority of water ($\approx$ 75%) is younger than one month, and about 15% of streamflow is younger than 1 day. More than 90% of streamflow is typically modeled to be younger than 4 months. This finding highlights that the vast majority of streamflow is fairly young, deriving from the current water year (i.e., the current wet season), and little long-term storage is included in catchment discharge.

3.4 Overland Flow is Primarily Pre-event Water.

A summary of streamflow contributions from different runoff sources and water of different ages estimated via SAS modeling is shown in Table 2 and, for a representative month in 2019, in Figure 11. Only one month is shown for legibility, but all winter months in the study period

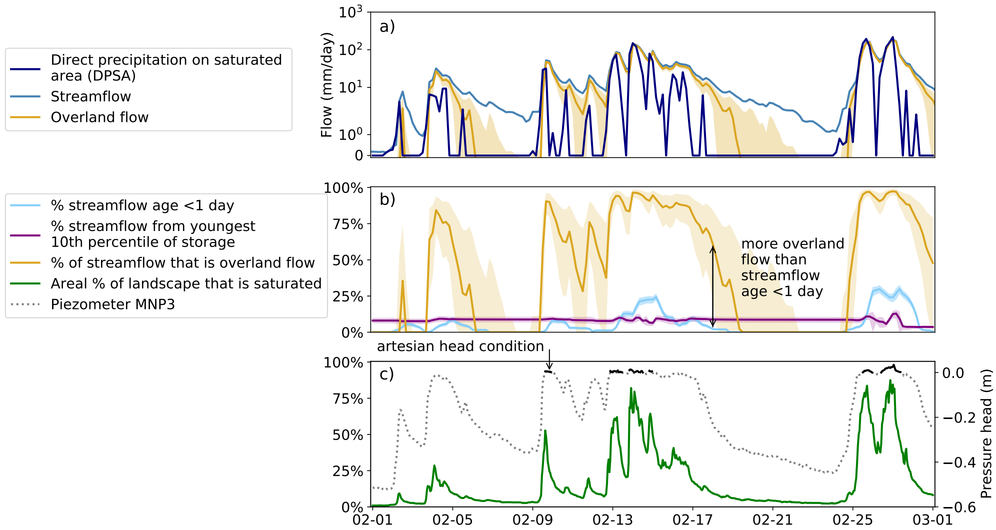

Figure 11. (a) Streamflow, estimated overland flow (streamflow above threshold instantaneous rate), and direct precipitation on saturated area for one representative month in 2019. (b) compares the portion of streamflow derived from overland flow to streamflow water from the youngest 10th percentile of streamflow or water of age <1 day. (c) Piezometer data and areal extents of saturation. Solid portions of piezometer data mark artesian head conditions. Shaded intervals in (a) and (b) denote the 25th-75th percentiles of ensemble simulations except for overland flow. Shaded intervals for overland flow show a range of threshold streamflow values (2-10 mm/day; solid line best estimate of 5 mm/day) for initiation of overland flow throughout the catchment outside of the channel network.

Table 2. Annual streamflow statistics by water year.

Fraction of streamflow that derives from...	WY 2017	WY 2018	WY 2019	WY 2020
overland flow	78%	70%	75%	62%
water age < 1 day	15%	9%	14%	6%
water from youngest 10^{th} percentile of storage	11%	11%	11%	11%
direct precipitation on saturated area	40%	28%	37%	21%

show the same patterns. Using 5 mm/day (likely range of 2-10 mm/day) as the capacity for subsurface flow based on the saturation extent mapping analysis (see Figure 7b), we calculated overland flow as the difference between instantaneous streamflow and a catchment runoff rate of 5 mm/day (Figure 11a). In this analysis, all of this overland flow is considered to be saturation overland flow, as we have not observed any evidence for Horton overland flow at the site. Overland flow constitutes the majority of streamflow, nearly always accounting for more than 50% of streamflow during rainy periods and frequently accounting for more than 90% of flow during large storm events (Figure 11b); in general, overland flow accounts for 62-78% of annual streamflow (Table 2). This result is consistent with sustained high groundwater levels during storms (e.g., Figure 5b) and the prediction that 80% of the landscape is saturated in large storms (Figure 7a).

Figure 11b compares the fraction of streamflow from overland flow to two definitions of new water in streamflow calculated from the 95th percentile of parameter sets for the SAS model: (i) water <1 day old and (ii) water from the youngest 10th percentile of storage. Based on SAS modeling, water from the youngest 10th percentile of storage is consistently about 11% of streamflow, and only about 10% of streamflow is younger than 1 day on an annual basis (Table 2). Since the SAS model parameterizes the relationship between time series of precipitation isotopes and streamflow isotopes, these model results are driven by the highly damped nature of the streamflow timeseries compared to the precipitation time series.

Only 6-15% of annual streamflow is younger than 1 day, but 62-78% of streamflow derives from overland flow (Table 2). Conservative estimates suggest that surface flow paths from the more distal portion of the watershed would reach the outlet within a day. We can approximate the travel paths as consisting of three distinct elements: sheet runoff on the $\approx$40 m long hillslope (e.g. Figure 1e), focused runoff down hollows and tributary channels ($\approx$500 m, e.g., Figure 1d), and travel down the mainstem Dry Creek ($\approx$ 4,000 m). Shallow sheet runoff is likely slow (on the order of 0.1 m/min), while in the hollows and channels velocities can exceed 5 m/min, and in the mainstem channel velocities exceed 10 m/min. These very conservative estimates would lead to the more distal part of the overland region reaching the outlet in about 15 hours. Hence, it is likely that overland flow across this landscape, if it remained on the surface and travelled to the outlet, would do so in less than a day.

Thus, the finding of significantly more overland flow than water younger than one day indicates that a large portion of overland flow must travel through the subsurface to reach the stream. Since all water following a singularly surface flow pathway would reach the outlet in less than 1 day, it is possible to set a limit on pre-event water in overland flow by comparing the fraction of streamflow younger than 1 day (light blue line in Figure 11b) to the fraction of streamflow derived from overland flow (gold line in Figure 11b). The difference between these two curves gives a lower bound on the pre-event water in overland flow, as marked in Figure 12. In Figure 12, we assumed that (at most) all water age <1 day arrived in the stream by overland flow. Then, given the difference in water volumes, at least 82% of overland flow in must be older than 1 day in water years 2019-2020. This finding is not unique to these years; throughout the study period, 81 - 90% of overland flow must be older than 1 day throughout each water year.

Further evidence for the importance of return flow to saturation overland flow comes from estimates of DPSA, calculated as the product of rainfall intensity and the percent saturated area

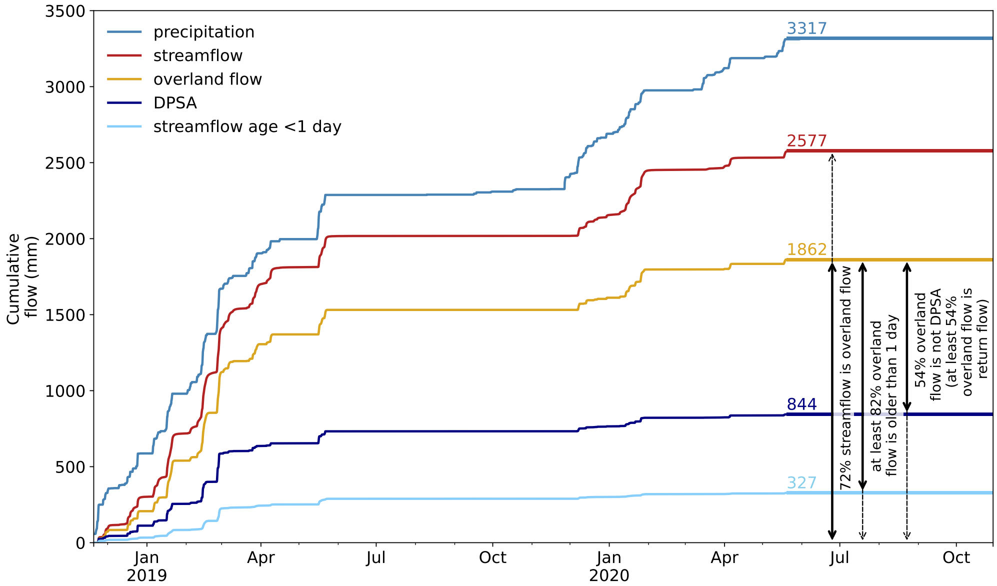

Figure 12. Cumulative amount of precipitation (light blue) compared to streamflow (red), overland flow using a threshold of 5 mm (gold), direct precipitation on saturated area, DPSA, (dark blue), and streamflow age <1 day (light blue). Numbers above each cumulative curve denote cumulative value for WY2019-2020.

(Figure 11a). The difference between this DPSA estimate and the overland flow curve places a different minimum bound on return flow contribution to streamflow since not all rain falling on saturated area necessarily contributes directly to runoff. Again, we see in Table 2 that at most 21 - 40% of streamflow could have been provided by DPSA, whereas overland flow accounts for 62 - 78% of streamflow on an annual basis. Thus, at least 49 - 66% of overland flow must be generated via return flow, providing further evidence that return flow plays an important role in saturation overland flow.

Overland flow accounts for the vast majority of streamflow, but water younger than 1 day and DPSA both account for relatively small fractions of annual runoff. These findings indicate that there must be substantial mixing between surface and subsurface water on the hillslope, which is also apparent in the damped isotopic signal of streamflow compared to rainfall (Figure 8).

4 Discussion

4.1 Pre-event Water in Saturation Overland Flow

In spite of the thin critical zone and dominance of the saturation overland flow mechanism, flow that arrives in the stream at Dry Creek is on average days older than the storm that generated the streamflow. This indicates that: (1) precipitation is stored and overland flow must mix with older, pre-event water; and (2) that water stored between events contributes substantially to saturation overland flow fluxes in events that follow. The storage and mixing have consequences for the conceptualization of runoff generation and water-rock interactions.

During periods of low flow (< approximately 2 mm/day), overland flow is not observed, and groundwater levels are below the ground surface across the borehole network. It is not until sufficient rains arrive to completely saturate the weathered bedrock and soil zone adjacent to the channel network that water tables intersect the ground surface and saturation overland flow is initiated. Further increases in streamflow are sustained by a continued rise of groundwater tables distal to the channel network and accompanying expansion of saturation extent (Figure 7b), leading to an increasing fraction of runoff that can be attributed to saturation overland flow, i.e., variable source area (Dunne & Black, 1970b).

The apparent paradox of fast streamflow response paired with pre-event water has been observed for over 30 years (e.g., Neal & Rosier, 1990; M. Sklash, 1990; Buttle, 1994) and continues to be an active area of hydrologic inquiry (e.g., Kirchner, 2003; Cartwright & Morgenstern, 2018). Overland flow, for instance, results in a quick runoff response, and is often considered to represent new (event) water in hydrograph separation literature (e.g., Uhlenbrook et al., 2002; Kronholm & Capel, 2016; Saraiva Okello et al., 2018; Ogunkoya & Jenkins, 1993). Our findings directly address the "old-water" paradox by demonstrating that, similar to the shallow subsurface stormflow observed by Kienzler and Naef (2008), saturation overland flow delivers pre-event water, and thus is older than the age of water delivered by the storm that generates streamflow. This agrees well with a recent particle tracking study that indicates that overland flow could primarily contain pre-event water while maintaining a streamflow signal that shows a predominance of young water catchment-wide (Wilusz et al., 2020). The behavior we observe at Dry Creek is similar to that of the Sleepers River watershed in Vermont, where saturation overland flow was originally documented. There, large extents (up to 50%) of the landscape can be saturated, saturation overland flow dominates runoff generation, and yet streamflow is nevertheless still largely older water (Shanley et al., 2015). Similarly, Eshleman et al. (1993), working in the Virginia Coastal Plain, found that saturation overland flow must consist primarily of return flow, based on the predominance of old water in streamflow when saturation overland flow was the primary runoff generation mechanism.

Importantly, our results indicate that saturation overland flow and Horton or infiltration excess overland flow should have different signatures in the age distribution of streamflow since in Horton excess overland flow, the interaction with subsurface water pools is likely to be more limited (Horton, 1933, 1945). In the case of Horton overland flow, we would anticipate primarily surface flowpaths and thus delivery of new, event water to streamflow, as has been found in locations where low surface hydraulic conductivity prevents infiltration (e.g., Ribolzi et al., 2007).

4.2 Is There an Inverse Storage Effect in Seasonally Dry Catchments?

The inverse-storage effect (ISE) describes the propensity of catchments to discharge younger water at wetter catchment states (e.g., Harman, 2015; Benettin et al., 2017). This is in contrast to the direct storage effect, where older water makes up the majority of catchment discharge. ISE has been directly observed in laboratory experiments (e.g., M. Kim et al., 2016) and inferred from particle tracking (e.g., Wilusz et al., 2020; Pangle et al., 2017). ISE may be more prevalent in some catchments than others based on particular climates or runoff generation mechanisms. Heidbüchel et al. (2012) found distinct differences in SAS behavior between a semiarid and a humid catchment. Most applications of SAS modeling have been in catchments with limited seasonality, so it is necessary to confirm whether ISE occurs broadly in seasonal catchments; recently, Rodriguez et al. (2018) found that the ISE applies in a catchment with a highly seasonal Mediterranean climate, and in this study we observed a mild ISE effect. Based on parameterization results, the SAS function approximates random sampling behavior as the catchment state becomes drier (less SOF) and exhibits a strong preference for the youngest water in storage at the wettest state (more SOF; see Supplementary Information S1 for details). However, over the course of the study period, the flow-weighted average value of the power exponent k is 0.99 (random sampling is $k = 1$), indicating that most streamflow in Dry Creek is sampled nearly randomly from available storage except during extremely wet periods. Thus, while there is evidence of an ISE at Dry Creek, the overall streamflow signature does not demonstrate a significant inverse storage effect.

While Rodriguez et al. (2018) did find ISE, they found that ISE may not apply at all times in a Mediterranean climate, with a more direct storage effect dominating during transitions between wet and dry seasons in the spring and fall. In our modeling, we allow the SAS function to vary through time according to wetness state, but the relationship between wetness state and SAS function remains constant throughout the study period. As a result, we are unable to determine whether a change in this relationship between wetness state and SAS function behavior occurs at our site. However, Figure 9a (streamflow timeseries in 2019) shows that the runoff goes down to about 0.1 mm/day numerous times over the wet season, indicating significant rapid shifts in catchment wetness throughout the season while the SAS model continues to perform well, missing only a handful of large concentration excursions. Parameterization on 2016-2019 water years also results in similarly good performance on the 2020 water year. There is, however, slightly higher absolute error in modeled concentrations during times of rapid state change versus continuously wet periods (Supplemental Figure S7), and some excursions from the modeled isotopic concentrations correlate with transitions between wet and dry states. This suggests that there may be a direct storage effect during transitions between wet and dry states in the Dry Creek catchment that is not explored in this study. While this effect was not included in our model, these transitions represent a very small portion of the study period so neglecting this effect should have a minimal impact on the results of this study, particularly since our study focuses on SOF, which occurs only once the catchment is wet enough to generate SOF, rather than during transition periods between wet and dry states.

4.3 Assumptions and Limitations

Water age calculations assumed that the entire catchment met a water storage capacity quantified as a streamflow threshold; however, the storage capacity of the landscape is met dynamically through time so that some parts of the landscape may contribute overland flow before the full storage capacity of the subsurface is met. We do not have data to quantify the extent to which this effect may be important at Dry Creek, although results from a particle tracking study conducted by Wilusz et al. (2020) suggest that this effect is minimal. Wilusz et al. (2020) found that maximum groundwater discharge level during different parts of the hydrograph was a function of storage, above which flow derives from overland flow, interflow, or direct runoff (i.e., rain falling directly in the stream channel). Across different portions of the hydrograph, the threshold varied by only about a factor of 2. A constant flow threshold, as used in this study, should provide a reasonable estimate for the fraction of streamflow attributable to overland flow over timescales longer than a few hours. Differences in the time to reach storage capacity across the landscape at this temporal resolution should be negligible, and a difference of a factor of 2 is included in the shaded interval in Figure 11b.

In our analysis, we have assumed that we can scale our hillslope-scale observations (in locations underlain by mélange matrix) to the entire Dry Creek catchment. Lovill et al. (2018), W. J. Hahm et al. (2020), and W. J. Hahm et al. (2019) documented the presence of large sandstone blocks, which cover less than 15% of the catchment by area and behave hydrologically distinctly from the mélange matrix areas. In contrast to the mélange matrix, the sandstone blocks: i) are deeply weathered; ii) have a thick vadose zone (>5 m), below which fluctuates a seasonal groundwater table; and iii) are observed to be the source of springs that persist into the mid-dry season. Because they are a relatively small portion of the landscape and because we are primarily interested in high-flow dynamics, we opted for the sake of simplicity to not separately model these features. The relatively high model performance (NSE = 0.62) provides some justification for this choice, but future work would benefit from extended analysis of the sandstone blocks, which likely have an outsize contribution to streamflow at low flow states (Lovill et al., 2018).

5 Conclusion

In the Dry Creek catchment in the Northern California Coast Ranges, field observations and stream age modeling using StorAge Selection (SAS) functions reveal that saturation overland flow arriving in the channel is pre-event water. Field observations reveal that runoff dynamics are fast (response within a few hours of rainfall), with runoff coefficients as high as 0.9, and that saturation overland flow is the primary storm runoff mechanism. SAS modeling does not indicate much of an inverse storage effect at Dry Creek except at extremely high flows, when younger (as a percentile of storage) streamflow is preferentially discharged. Although streamflow is modeled to be relatively young, the SAS model suggests that streamflow is still almost entirely older than 1 day at all times, meaning that streamflow is modeled to be older than event water. Since streamflow is primarily overland flow, the SAS modeling results imply that overland flow must contain a substantial portion of pre-event water. This finding is supported by field observations of exfiltrating head gradients, return flow through macropores, and extensive saturation days after storm events, which collectively point to a significant subsurface origin (i.e., return flow) for the saturation overland flow. Even in this extreme case of full catchment SOF, our analyses indicate that substantial mixing of overland flow with subsurface storage must occur to explain the observed streamflow ages.

Understanding the relationship between the age of streamflow and runoff generation mechanisms assists in understanding of how water quality may change over time, particularly under climate change. An increase in extreme precipitation with the same mean, as is expected with climate change in some locations, including California where our site is located (Swain et al., 2018), lead to larger overland flow runoff events. This trend of wet season sharpening is likely to make overland flow more important in catchments where overland flow occurs. Increased precipitation volatility is also likely to result in increased relative variability in wetted channel extent (Lapides, Leclerc, et al., 2021), which may apply to saturated area as well. Future studies might consider these interaction and their consequences for kinetic-rate controlled processes like chemical weathering.

Open Research

All data and code associated with the manuscript are available at `https://colab.research.google.com/drive/1fB9BNEY7RzaGpqqnjo7gdeq79Bhqbjvb#scrollTo=znP2tntme3dI` (isotope and groundwater processing code), `https://colab.research.google.com/drive/1EFI1GkU0DlgG56AJ17716UlxXc17W2Yd#scrollTo=1kmoyGsxnuCB` (SAS modeling code), `https://colab.research.google.com/drive/1VDtkjJGjBOr0mXBq1--CLxVHmDLbGifZ?usp=sharing` (logistic regression for saturation extent code), `https://colab.research.google.com/drive/1FzbUSYS6OeKAOI02a35qZfktN72Ypzaz` (event runoff coefficient analysis), `https://colab.research.google.com/drive/1F4H-Mb-DfltsCp8mFvXDOceD7sJVhew5` (lag to peak analysis) and `https://www.hydroshare.org/resource/13244d68f3e74452a8bbcb5d8860768c/` (large data files; Lapides, Hahm, et al., 2021).

Acknowledgments

We gratefully acknowledge field access provided by Marilyn and Jerry Russell, and thank Renee Holleman for helping to collect precipitation and stream samples. Wenbo Yang, Todd Dawson, Collin Bode, Hunter Jamison, Sky Lovill, Colleen Murphy, and Sami Cargill contributed to fieldwork and laboratory analyses. Funding was provided by Simon Fraser University, a Natural Science Engineering Research Council of Canada Discovery Grant, the US National Science Foundation-supported Eel River Critical Zone Observatory (EAR 1331940), Oak Ridge Institute for Science and Education (ORISE), the University of California Natural Reserve System Mildred E. Mathias Graduate Student Research Grant, and the Carol Baird Fund for Graduate Field Science.

References

Abdul, A., & Gillham, R. (1984). Laboratory studies of the effects of the capillary fringe on streamflow generation. *Water Resources Research*, *20*(6), 691–698.

Allen, S. T., Jasechko, S., Berghuijs, W. R., Welker, J. M., Goldsmith, G. R., & Kirchner, J. W. (2019). Global sinusoidal seasonality in precipitation isotopes. *Hydrology and Earth System Sciences*, *23*(8), 3423–3436.

Allen, S. T., Kirchner, J. W., & Goldsmith, G. R. (2018). Predicting spatial patterns in precipitation isotope (δ2h and δ18o) seasonality using sinusoidal isoscapes. *Geophysical Research Letters*, *45*(10), 4859–4868.

Anderson, S., Dietrich, W., & Brimhall, G. (2002). Weathering profiles, mass-balance analysis, and rates of solute loss: Linkages between weathering and erosion in a small, steep catchment. *Geological Society of America Bulletin*, *114*(9), 1143. (00163)

Atwater, T., & Stock, J. (1998). Pacific-north america plate tectonics of the neogene southwestern united states: an update. *International Geology Review*, *40*(5), 375–402.

Baumhoff, M. A., & Merriam, C. H. (1958). *California athabascan groups*. University of California Press.

Benettin, P., & Bertuzzo, E. (2018). tran-sas v1. 0: a numerical model to compute catchment-scale hydrologic transport using storage selection functions. *Geoscientific Model Development*, *11*(ARTICLE), 1627–1639.

Benettin, P., Soulsby, C., Birkel, C., Tetzlaff, D., Botter, G., & Rinaldo, A. (2017). Using sas functions and high-resolution isotope data to unravel travel time distributions in headwater catchments. *Water Resources Research*, *53*(3), 1864–1878.

Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology/un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Journal*, *24*(1), 43–69.

Blume, T., Zehe, E., & Bronstert, A. (2007, October). Rainfall—runoff response, event-based runoff coefficients and hydrograph separation. *Hydrological Sciences Journal*, *52*(5), 843–862. Retrieved 2021-07-16, from https://doi.org/10.1623/hysj.52.5.843 (Publisher: Taylor & Francis _eprint: https://doi.org/10.1623/hysj.52.5.843) doi: 10.1623/hysj.52.5.843

Botter, G., Bertuzzo, E., & Rinaldo, A. (2011). Catchment residence and travel time distributions: The master equation. *Geophysical Research Letters*, *38*(11).

Buttle, J. (1994). Isotope hydrograph separations and rapid delivery of pre-event water from drainage basins. *Progress in physical geography*, *18*(1), 16–41.

Cartwright, I., & Morgenstern, U. (2018). Using tritium and other geochemical tracers to address the "old water paradox" in headwater catchments. *Journal of Hydrology*, *563*, 13–21.

Cloos, M. (1983). Comparative study of melange matrix and metashales from the franciscan subduction complex with the basal great valley sequence, california. *The Journal of Geology*, *91*(3), 291–306.

Craig, H. (1961). Isotopic variations in meteoric waters. *Science*, *133*(3465), 1702–1703.

DeWalle, D., Edwards, P., Swistock, B., Aravena, R., & Drimmie, R. (1997). Seasonal isotope hydrology of three appalachian forest catchments. *Hydrological Processes*, *11*(15), 1895–1906.

Dingman, S. L. (2015). *Physical hydrology*. Waveland press.

Dralle, D. N., Hahm, W. J., Rempe, D. M., Karst, N. J., Thompson, S. E., & Dietrich, W. E. (2018). Quantification of the seasonal hillslope water storage that does not drive streamflow. *Hydrological processes*, *32*(13), 1978–1992.

Dunne, T. (1978). Field studies of hillsope flow processes. *Hillslope hydrology*, 227–293.

Dunne, T., & Black, R. D. (1970a). An experimental investigation of runoff production in permeable soils. *Water Resources Research*, *6*(2), 478–490.

Dunne, T., & Black, R. D. (1970b). Partial area contributions to storm runoff in a small new england watershed. *Water resources research*, *6*(5), 1296–1311.

Elsenbeer, H., & Lack, A. (1996, May). Hydrometric and hydrochemicai evidence for fast flowpaths at La Cuenca, Western Amazonia. *Journal of Hydrology*, *180*(1), 237–250. Retrieved 2019-11-16, from http://www.sciencedirect.com/science/article/pii/0022169495028897 doi: 10.1016/0022-1694(95)02889-7

Elsenbeer, H., West, A., & Bonell, M. (1994, October). Hydrologic pathways and stormflow hydrochemistry at South Creek, northeast Queensland. *Journal of Hydrology*, *162*(1), 1–21. Retrieved 2019-11-16, from http://www.sciencedirect.com/science/article/pii/0022169494900027 doi: 10.1016/0022-1694(94)90002-7

Eshleman, K. N., Pollard, J. S., & O'Brien, A. K. (1993). Determination of contributing areas for saturation overland flow from chemical hydrograph separations.

Water Resources Research, *29*(10), 3577–3587. Retrieved 2021-07-21, from http://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/93WR01811 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/93WR01811) doi: 10.1029/93WR01811

Foster, G. M. (1944). *A summary of yuki culture* (Vol. 50). University of California Press Berkeley, CA.

Freyberg, J. v., Studer, B., Rinderer, M., & Kirchner, J. W. (2018). Studying catchment storm response using event-and pre-event-water volumes as fractions of precipitation rather than discharge. *Hydrology and Earth System Sciences*, *22*(11), 5847–5865.

Gallart, F., Valiente, M., Llorens, P., Cayuela, C., Sprenger, M., & Latron, J. (2020). Investigating young water fractions in a small mediterranean mountain catchment: Both precipitation forcing and sampling frequency matter. *Hydrological Processes*, *34*(17), 3618–3634.

Godsey, S. E., Aas, W., Clair, T. A., De Wit, H. A., Fernandez, I. J., Kahl, J. S., . . . others (2010). Generality of fractal 1/f scaling in catchment tracer time series, and its implications for catchment travel time distributions. *Hydrological Processes*, *24*(12), 1660–1671.

Godsey, S. E., Kirchner, J. W., & Clow, D. W. (2009). Concentration–discharge relationships reflect chemostatic characteristics of us catchments. *Hydrological Processes: An International Journal*, *23*(13), 1844–1864.

Group, P. C. (2013). *PRISM Climate Data* (Tech. Rep.). Oregon State University. Retrieved from http://prism.oregonstate.edu (00000)

Haggerty, R., Wondzell, S. M., & Johnson, M. A. (2002). Power-law residence time distribution in the hyporheic zone of a 2nd-order mountain stream. *Geophysical Research Letters*, *29*(13), 18–1.

Hahm, W., Dralle, D., Lovill, S., Rose, J., Dawson, T., & Dietrich, W. (2017). Exploratory tree survey (2016- eel river critical zone observatory- sagehorn- central belt melange, franciscan complex, northern california coast ranges, usa). hydroshare. *HydroShare*.

Hahm, W. J., Dietrich, W. E., & Dawson, T. E. (2018). Controls on the distribution and resilience of quercus garryana: ecophysiological evidence of oak's water-limitation tolerance. *Ecosphere*, *9*(5), e02218.

Hahm, W. J., Dralle, D. N., Dawson, T. E., & Dietrich, W. E. (2020). Oak transpiration drawn from the weathered bedrock vadose zone in the summer dry season. *Water Resources Research*.

Hahm, W. J., Rempe, D. M., Dralle, D. N., Dawson, T. E., Lovill, S. M., Bryk, A. B., . . . Dietrich, W. E. (2019). Lithologically controlled subsurface critical zone thickness and water storage capacity determine regional plant community composition. *Water*

Resources Research, *55*(4), 3028–3055.

Hahm, W. J., Wang, J., Druhan, J. L., Rempe, D., & Dietrich, W. E. (2017). Relating runoff generation mechanisms to concentration-discharge relationships in catchments with well-characterized critical zone structures and hydrologic dynamics. *AGUFM*, *2017*, B43E–2172.

Harman, C. J. (2015). Time-variable transit time distributions and transport: Theory and application to storage-dependent transport of chloride in a watershed. *Water Resources Research*, *51*(1), 1–30.

Harman, C. J., Evans, O., & Lu, F. (2019). *mesas.* (`https://github.com/charman2/mesas`)

Heidbüchel, I., Troch, P. A., Lyon, S. W., & Weiler, M. (2012). The master transit time distribution of variable flow systems. *Water Resources Research*, *48*(6).

Hewlett, J. D., & Hibbert, A. R. (1967). Factors affecting the response of small watersheds to precipitation in humid areas. *Forest hydrology*, *1*, 275–290.

Horton, R. E. (1933). The role of infiltration in the hydrologic cycle. *Eos, Transactions American Geophysical Union*, *14*(1), 446–460.

Horton, R. E. (1945). Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology. *Geological society of America bulletin*, *56*(3), 275–370.

Jayko, A., Blake, M., McLaughlin, R., Ohlin, H., Ellen, S., & Kelsey, H. (1989). *Reconnaissance geologic map of the covelo 30-by 60-minute quadrangle, northern california* (Tech. Rep. No. MF-2001). US Government Printing Office.

Johnson, S. G. (1979). *The land-use history of the coast range preserve, Mendocino County, California* (Unpublished doctoral dissertation). San Francisco State University, San Francisco, CA.

Kelsey, H. M. (1978). Earthflows in franciscan melange, van duzen river basin, california. *Geology*, *6*(6), 361–364.

Kidron, G. J. (2021). Comparing overland flow processes between semiarid and humid regions: does saturation overland flow take place in semiarid regions? *Journal of Hydrology*, *593*, 125624.

Kienzler, P. M., & Naef, F. (2008). Subsurface storm flow formation at different hillslopes and implications for the 'old water paradox'. *Hydrological Processes: An International Journal*, *22*(1), 104–116.

Kim, H., Dietrich, W. E., Thurnhoffer, B. M., Bishop, J. K., & Fung, I. Y. (2017). Controls on solute concentration-discharge relationships revealed by simultaneous hydrochemistry observations of hillslope runoff and stream flow: The importance of critical zone structure. *Water Resources Research*, *53*(2), 1424–1443.

Kim, M., Pangle, L. A., Cardoso, C., Lora, M., Volkmann, T. H., Wang, Y., . . . Troch, P. A. (2016). Transit time distributions and s tor a ge s election functions in a sloping soil lysimeter with time-varying flow paths: Direct observation of internal and external transport variability. *Water Resources Research*, *52*(9), 7105–7129.

Kirchner, J. W. (2003). A double paradox in catchment hydrology and geochemistry. *Hydrological processes*, *17*(4), 871–874.

Kirchner, J. W., Feng, X., & Neal, C. (2000). Fractal stream chemistry and its implications for contaminant transport in catchments. *Nature*, *403*(6769), 524–527.

Kirkby, M. (1988). Hillslope runoff processes and models. *Journal of Hydrology*, *100*(1-3), 315–339.

Knoben, W. J., Freer, J. E., & Woods, R. A. (2019). Inherent benchmark or not? comparing nash–sutcliffe and kling–gupta efficiency scores. *Hydrology and Earth System Sciences*, *23*(10), 4323–4331.

Kronholm, S. C., & Capel, P. D. (2016). Estimation of time-variable fast flow path chemical concentrations for application in tracer-based hydrograph separation analyses. *Water Resources Research*, *52*(9), 6881–6896.

Lapides, D. A., Hahm, W. J., Rempe, D. M., Dietrich, W. E., & Dralle, D. N. (2021). *Calculating streamwater age using storage selection functions at dry creek, ca.* (HydroShare, https://www.hydroshare.org/resource/13244d68f3e74452a8bbcb5d8860768c/) doi: 10.4211/hs.13244d68f3e74452a8bbcb5d8860768c

Lapides, D. A., Leclerc, C. D., Moidu, H., Dralle, D. N., & Hahm, W. J. (2021). Variability of stream extents controlled by flow regime and network hydraulic scaling. *Hydrological Processes*. doi: 10.1002/hyp.14079

Latron, J., Soler, M., Llorens, P., & Gallart, F. (2008). Spatial and temporal variability of the hydrological response in a small Mediterranean research catchment (Vallcebre, Eastern Pyrenees). *Hydrological Processes*, *22*(6), 775–787. Retrieved 2021-04-14, from https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.6648 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.6648) doi: https://doi.org/10.1002/hyp.6648

Li, L., Sullivan, P. L., Benettin, P., Cirpka, O. A., Bishop, K., Brantley, S. L., . . . others (2020). Toward catchment hydro-biogeochemical theories. *Wiley Interdisciplinary Reviews: Water*, e1495.

Litwin, D. G., Tucker, G. E., Barnhart, K. R., & Harman, C. J. (2020). Groundwaterdupuitpercolator: A landlab component for groundwater flow. *Journal of Open Source Software*, *5*(46), 1935.

Lock, J., Kelsey, H., Furlong, K., & Woolace, A. (2006). Late neogene and quaternary landscape evolution of the northern california coast ranges: Evidence for mendocino triple junction tectonics. *Geological Society of America Bulletin*, *118*(9-10), 1232–1246.

Lovill, S., Hahm, W., & Dietrich, W. (2018). Drainage from the critical zone: Lithologic controls on the persistence and spatial extent of wetted channels during the summer dry season. *Water Resources Research*, *54*(8), 5702–5726.

Małoszewski, P., & Zuber, A. (1982). Determining the turnover time of groundwater systems with the aid of environmental tracers: 1. models and their applicability. *Journal of hydrology*, *57*(3-4), 207–231.

McDonnell, J. J., & Beven, K. (2014). Debates—the future of hydrological sciences: A (common) path forward? a call to action aimed at understanding velocities, celerities and residence time distributions of the headwater hydrograph. *Water Resources Research*, *50*(6), 5342–5350.

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.

Neal, C., & Rosier, P. T. (1990). Chemical studies of chloride and stable oxygen isotopes in two conifer afforested and moorland sites in the british uplands. *Journal of Hydrology*, *115*(1-4), 269–283.

Ogunkoya, O., & Jenkins, A. (1993). Analysis of storm hydrograph and flow pathways using a three-component hydrograph separation model. *Journal of Hydrology*, *142*(1-4), 71–88.

Oshun, J., Dietrich, W. E., Dawson, T. E., & Fung, I. (2016). Dynamic, structured heterogeneity of water isotopes inside hillslopes. *Water Resources Research*, *52*(1), 164–189.

Pangle, L. A., Kim, M., Cardoso, C., Lora, M., Meira Neto, A. A., Volkmann, T. H., . . . Harman, C. J. (2017). The mechanistic basis for storage-dependent age distributions of water discharged from an experimental hillslope. *Water Resources Research*, *53*(4), 2733–2754.

Putnam, S. M., et al. (2018). *The influence of landscape structure on storage and streamflow generation in a piedmont catchment* (Unpublished doctoral dissertation). Johns Hopkins University, Baltimore, MD.

Ribolzi, O., Karambiri, H., Bariac, T., Benedetti, M., Caquineaux, S., Descloitres, M., & Aventurier, A. (2007). Mechanisms affecting stormflow generation and solute behaviour in a sahelian headwater catchment. *Journal of hydrology*, *337*(1-2), 104–116.

Rinderer, M., Kollegger, A., Fischer, B. M., Stähli, M., & Seibert, J. (2012). Sensing with boots and trousers—qualitative field observations of shallow soil moisture patterns. *Hydrological Processes*, *26*(26), 4112–4120.

Rittiman Jr, C., & Thorson, T. (2001). Soil survey of mendocino county. *California, Western*

Part, Mendocino County Resource Conservation District.

Rodhe, A., Nyberg, L., & Bishop, K. (1996). Transit times for water in a small till catchment from a step shift in the oxygen 18 content of the water input. *Water Resources Research*, *32*(12), 3497–3511.

Rodriguez, N. B., & Klaus, J. (2019). Catchment travel times from composite storage selection functions representing the superposition of streamflow generation processes. *Water Resources Research*, *55*(11), 9292–9314.

Rodriguez, N. B., McGuire, K. J., & Klaus, J. (2018). Time-varying storage–water age relationships in a catchment with a mediterranean climate. *Water Resources Research*, *54*(6), 3988–4008.

Rodriguez, N. B., Pfister, L., Zehe, E., & Klaus, J. (2021). A comparison of catchment travel times and storage deduced from deuterium and tritium tracers using storage selection functions. *Hydrology and Earth System Sciences*, *25*(1), 401–428.

Saraiva Okello, A. M. L., Uhlenbrook, S., Jewitt, G. P., Masih, I., Riddell, E. S., & Van der Zaag, P. (2018). Hydrograph separation using tracers and digital filters to quantify runoff components in a semi-arid mesoscale catchment. *Hydrological Processes*, *32*(10), 1334–1350.

Shanley, J. B., Kendall, C., Smith, T. E., Wolock, D. M., & McDonnell, J. J. (2002). Controls on old and new water contributions to stream flow at some nested catchments in Vermont, USA. *Hydrological Processes*, *16*(3), 589–609. Retrieved 2021-07-21, from http://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.312 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.312) doi: 10.1002/hyp.312

Shanley, J. B., Sebestyen, S. D., McDonnell, J. J., McGlynn, B. L., & Dunne, T. (2015). Water's Way at Sleepers River watershed – revisiting flow generation in a post-glacial landscape, Vermont USA. *Hydrological Processes*, *29*(16), 3447–3459. Retrieved 2021-07-21, from http://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.10377 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.10377) doi: 10.1002/hyp.10377

Sklash, M. (1990). Environmental isotope studies of storm and snowmelt runoff generation. *Process studies in hillslope hydrology*, 401–436.

Sklash, M. G., & Farvolden, R. N. (1979). The role of groundwater in storm runoff. *Journal of Hydrology*, *43*(1-4), 45–65.

Smith, A., Tetzlaff, D., & Soulsby, C. (2018). On the use of storage selection functions to assess time-variant travel times in lakes. *Water Resources Research*, *54*(7), 5163–5185.

Stewart, O. C. (1943). *Notes on pomo ethnogeography*. University of California Press.

Swain, D. L., Langenbrunner, B., Neelin, J. D., & Hall, A. (2018, May). Increasing precipitation volatility in twenty-first-century California. *Nature Climate Change*, *8*(5), 427–

433. doi: 10.1038/s41558-018-0140-y

Torres, M. A., & Baronas, J. J. (2021). Modulation of Riverine Concentration-Discharge Relationships by Changes in the Shape of the Water Transit Time Distribution. *Global Biogeochemical Cycles*, *35*(1), e2020GB006694. Retrieved 2021-07-16, from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020GB006694 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020GB006694) doi: 10.1029/2020GB006694

Uhlenbrook, S., Frey, M., Leibundgut, C., & Maloszewski, P. (2002). Hydrograph separations in a mesoscale mountainous basin at event and seasonal timescales. *Water Resources Research*, *38*(6), 31–1.

Van Der Velde, Y., Torfs, P., Van Der Zee, S., & Uijlenhoet, R. (2012). Quantifying catchment-scale mixing and its effect on time-varying travel time distributions. *Water Resources Research*, *48*(6).

Visser, A., Thaw, M., Deinhart, A., Bibby, R., Safeeq, M., Conklin, M., . . . Van der Velde, Y. (2019). Cosmogenic isotopes unravel the hydrochronology and water storage dynamics of the southern sierra critical zone. *Water Resources Research*, *55*(2), 1429–1450.

Wilson, C., & Dietrich, W. (1987). The contribution of bedrock groundwater flow to storm runoff and high pore pressure development in hollows. *IAHS-AISH publication*(165), 49–59.

Wilusz, D., Harman, C., Ball, W., Maxwell, R., & Buda, A. (2020). Using particle tracking to understand flow paths, age distributions, and the paradoxical origins of the inverse storage effect in an experimental catchment. *Water Resources Research*, *56*(4), e2019WR025140.

Supporting Information for "Controls on streamwater age in a saturation overland flow-domianted catchment"

[1, 2] Dana A. Lapides

[1] W. Jesse Hahm

[3] Daniella M. Rempe

[4] William E. Dietrich

[2] David N. Dralle

[1]Department of Geography, Simon Fraser University, Burnaby, BC, Canada

[2]Pacific Southwest Research Station, United States Forest Service, Davis, CA, USA

[3]University of Texas, Austin, Austin, TX, USA

[4]University of California, Berkeley, Berkeley, CA, USA

Contents of this file

Introduction

These supporting information include supplementary analyses and figures for "Controls on streamwater age in a saturation overland flow-dominated catchment." These analyses consist of detailed information about StorAge Selection (SAS) parameterization (S1, Figures S1-S5), additional SAS results not shown in the main text (S2, Figures S6-S7), analysis identifying likely kinetic fractionation at low flows (S3, Figure S8), and a synthetic analysis testing the impact of sampling interval on study results (S4, Figure S9). See the main text Section 2.6 for a description of SAS methodology used in this study.

S1. Parameterization of SAS functions

The range of shapes for the streamflow SAS function is shown in Figure S1a for the median values of $kmin_Q = 0.45$ and $kmax_Q = 1.04$ among the top 95th percentile of parameter sets sampled. As shown in Figure S1a, the SAS function varies as a function of the wetness state of the catchment. Assuming a uniform storage distribution, at the wettest state the 50th percentile of streamflow age derives from the 25th percentile of storage ages, while at the driest state the 50th percentile of streamflow age derives from the 50th percentile of storage ages.

Histograms of the top 95th percentile of parameter sets in Figure S1b-g show that a good model fit (maximum KGE = 0.83, maximum NSE = 0.66) can arise from nearly any parameter value of: i) the minimum power for the streamflow SAS function ($kmin_Q$), ii) the scaling factor for the log-dependence on wetness state in the streamflowSAS function (logfactor_Q), iii) the power law exponent for the ET SAS function (k_{ET}), and iv) the initial isotope concentration in storage (C_{S_0}). In contrast, the model performs best (shown via a peak in the histogram of parameter distributions) for: c) maximal power for the streamflow SAS function ($kmax_Q$) near 1 and f) initial storage volume (S_0) near our estimated maximal storage volume of 200 mm. Relationships between parameters and model performance are shown via a scatterplot matrix in Supplemental Information 1.

We parameterized the StorAge Selection (SAS) function using a Monte Carlo simulation of 10,000 parameter sets. The relationship between all parameter pairs in the calibration time period (2016-2019 water years), is shown in Figure S2 colored by NSE and in Figure S3 colored by KGE. There is no pattern in the relationships among any parameters except $kmax_Q$ and S_0 that indicate a preferential relationship between these parameters. Any value of $kmin_Q$, logfactor_Q, k_{ET}, and C_{S0} can yield a model with a high NSE and KGE, with no relationship among the values that yield high NSE/KGE between these parameters. The value of $kmax_Q$ has the biggest impact on NSE, and the value of $S0$ has the biggest impact on KGE. The same patterns appear in evaluation results of the top 95th percentile of parameters in the 2020 water year, shown in Figure S4 for NSE and Figure

S5 for KGE. The parameter relationships among $kmin_Q$, logfactor$_Q$, $k_E T$, and C_{S0} fill the full space, while values are preferentially excluded from the 95th percentile parameter sets for $kmax_Q$ and $S0$.

S2. SAS results across more conditions

While only WY2019 results are shown in the main text, results of StorAge Selection modeling are similar for WY 2020 (Figure S6).

We also compare the distribution of absolute error in predicted streamflow concentration during periods of rapid wet/dry state evolution versus periods of consistent wetness (Figure S7). A transition from wet to dry state is considered to happen at an instantaneous flow rate of 0.1 mm/day (as marked in Figure S7a), with the two days before and after defined as the transition period or time when the catchment is switching states. The catchment is in a wet state any other time when streamflow is above 0.1 mm/day when not switching states. Figure S7b shows histograms of absolute error in modeled streamflow concentration when the catchment is in a wet state (blue) or switching states (red). The modes of the two histograms are very similar, although the error for switching state is slightly larger. When in a wet state, error is never above 5 ‰; however, when switching states, the tail is longer.

S3. Kinetic fractionation at low flow

To confirm that evaporative enrichment occurs at flows below 0.05 mm/day, we examined one representative flow event in 2019. From April 15-March 15, flow decreases steadily with no precipitation. In Figure S8, points are colored by date, moving from white to dark blue with increasing time and decreasing streamflow, and points outlined in red fall below the 0.05 mm/day threshold used to exclude data points from the SAS model calibration. Figure S8 demonstrates that as flow decreases, isotopic concentration increases following a slope much shallower (≈ 4) than that the local meteoric line (≈ 8), suggesting evaporative enrichment (Craig, 1961). We examined a similar end-of-season period in 2020 but did not note evaporative fractionation, indicating that evaporative enrichment may not always occur below 0.05 mm/day, but this threshold provides a conservative cutoff to prevent calibration from being impacted by evaporative fractionation.

S4. Confidence in information on daily timescale

As described in the discussion (Section 2.6), the stream sampling interval frequency (daily, in our case) may limit the ability of the model to resolve the lower bound (youngest) water transit times. To explore this, we performed a synthetic analysis to determine the potential impact of a daily sampling interval on our results. We generated a timeseries of [dD] in streamflow with about 50% of streamflow younger than 1 day. This timeseries was generated using the form of StorAge Selection (SAS) function used in this study, i.e. a time-varying power law that varies with catchment wetness state (Equation 10 in main text) with $kmin_Q = 0.85$, $kmax_Q = 1.0$, $\mathrm{logfactor}_Q = 45$, $k_{ET} = 1$, $S_0 = 170$ mm, and $C_{S_0} =$ -50 ‰. To enhance new (< 1 day old) water production at short timescales, we set a threshold wetness value (wi=0.3) above which the power in the SAS function exponent is $k = 0.1$. To explore the impact of sampling interval on the parameterization results, we simulated a range of sampling intervals using the observed rainfall timeseries and the generated streamflow timeseries with enhanced new water fraction. We explored the effect of sampling frequency on resolved age distributions by down-sampling from daily frequency to 2, 4, 8, and 16 day sampling intervals for the streamflow by dropping additional data. (Isotope concentrations in streamflow are instantaneous rather than aggregate measurements). In contrast, for precipitation, we averaged concentrations over the previous 2, 4, 8, or 16 days to downsample. We did not examine sampling intervals shorter than daily since we do not have data on how concentrations in rainfall vary at subdaily timescales. For each sampling interval, we ran 5,000 Monte Carlo simulations with the same range of parameters and same SAS functional forms used in the study and evaluated performance using Nash-Sutcliffe Efficiency (NSE) and Kling-Gupta Efficiency (KGE). The top 95th percentile of parameter sets for each sampling interval were retained. Model performance dropped with increasing sampling interval (Figure S9a), with the median NSE dropping below 0.5 by a sampling interval of 8 days. Generally, model performance remains quite good over a range of sampling intervals, with very comparable performance at sampling intervals of 1-4 days.

Using the top 95th percentile of parameter sets for each sampling interval, we calculated the overall fraction of water younger than 1 day throughout the study period (Figure S9b). The fraction of new water calculated for sampling interval 1-4 days have overlapping 99% confidence intervals, while much more new water is predicted for longer sampling intervals. This finding of more new water with a longer sampling interval is contrary to recent findings that indicate a shorter sampling interval may reveal more young water Gallart et al. (2020). Both sites have a Mediterranean climate and high runoff coefficients with up to 90% of streamflow made up of event water. The difference in results is likely due to the very different definition of young water. In the present study, we consider water younger than 1 day, comparable to the dominant runoff generation processes at Dry Creek, whereas Gallart et al. (2020) consider water younger than 73 days, much longer than dominant runoff generation mechanisms. Given that the new water fraction can be captured consistently with 1-4 day long sampling intervals, the new water fraction inferred from the relatively short sampling interval of 1 day used in this study should provide confidence on the fraction of water younger than 1 day at Dry Creek.

References

Craig, H. (1961). Isotopic variations in meteoric waters. *Science*, *133*(3465), 1702–1703.

Gallart, F., Valiente, M., Llorens, P., Cayuela, C., Sprenger, M., & Latron, J. (2020). Investigating young water fractions in a small mediterranean mountain catchment: Both precipitation forcing and sampling frequency matter. *Hydrological Processes*, *34*(17), 3618–3634.

Kirchner, J. W. (2003). A double paradox in catchment hydrology and geochemistry. *Hydrological processes*, *17*(4), 871–874.

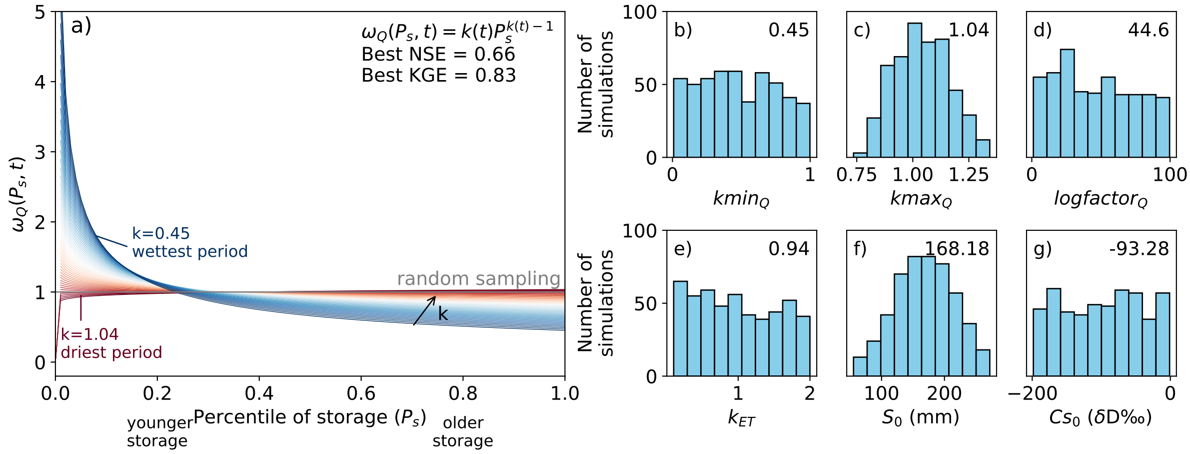

Figure S1. (a) ω_Q is the StorAge Selection (SAS) function described by the general functional form in the upper right corner of the plot. It describes the relative tendency for the stream to draw water from each age percentile of storage P_s. The displayed range of SAS functions for streamflow ω_Q is represented by the median values of k_{min} and k_{max} in the 95th percentile of parameter sets. From wettest to driest conditions assuming a uniform storage distribution, the 50th percentile of streamflow ages range from the youngest 24th-47th percentiles of storage. Histograms of parameters for the top 95th percentile of Monte Carlo simulations of 10,000 parameter sets (b-g) results in a median NSE=0.62 and maximum NSE=0.66, with the median value for each parameter shown in the upper right corner of each panel. k_{min_Q}, k_{max_Q}, and logfactor$_Q$ are parameters for the streamflow SAS function ω_Q. k_{ET} is a parameter for the evapotranspiration SAS function ω_{ET}. S_0 is the initial catchment storage, and C_{S_0} is the initial isotopic concentration. See Section 2.6 for full details.

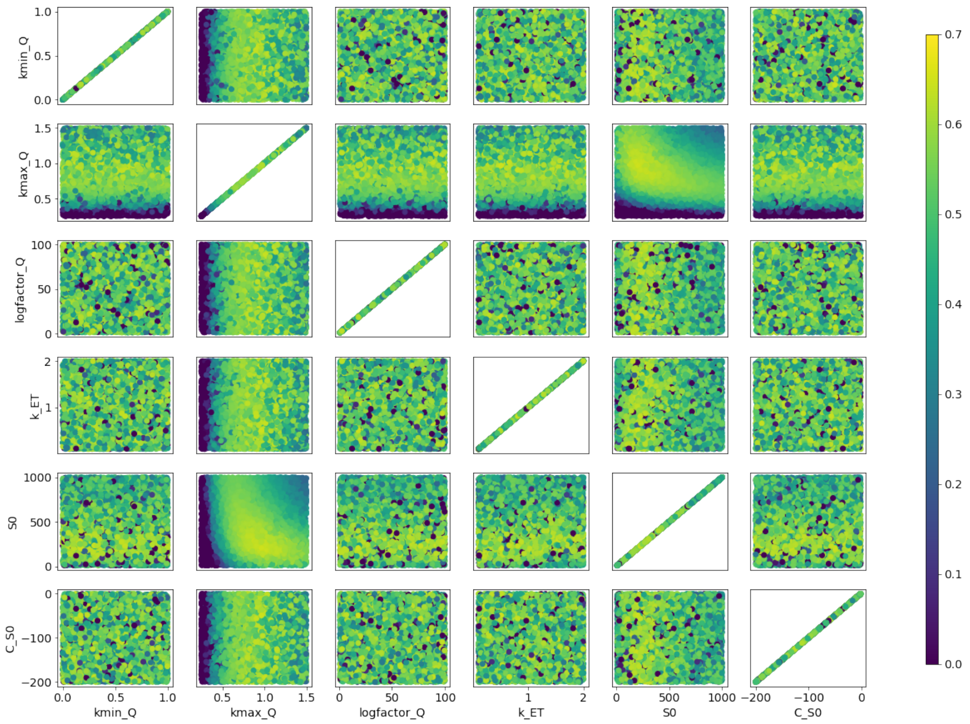

Figure S2. Covariance between parameter performance for all 10,000 parameter sets tested in Monte Carlo simulations. Points are colored by NSE.

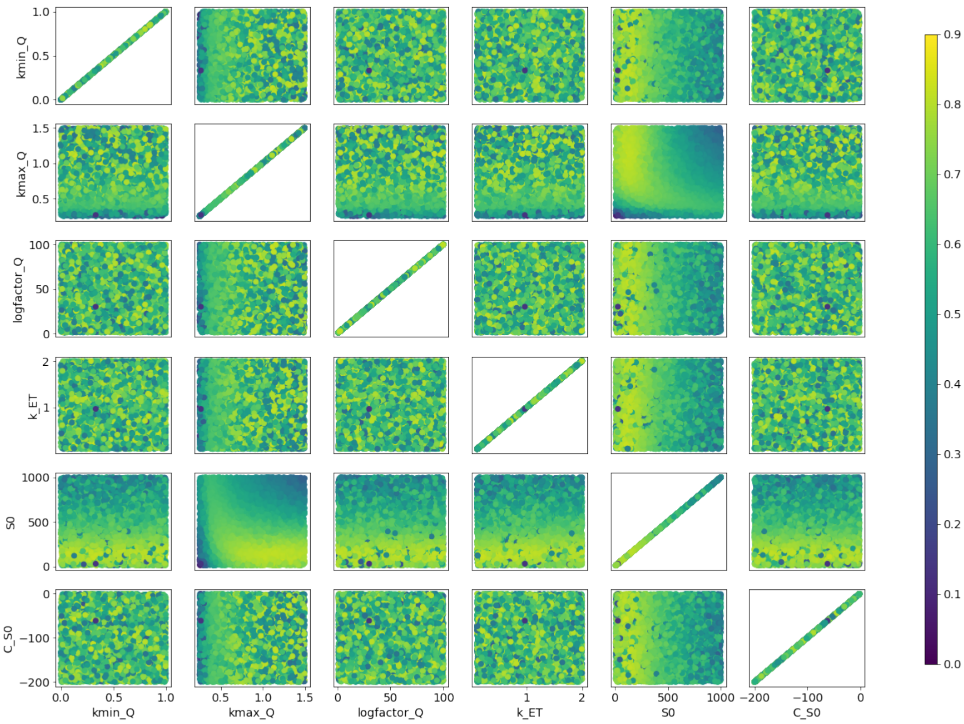

Figure S3. Covariance between parameter performance for all 10,000 parameter sets tested in Monte Carlo simulations. Points are colored by KGE.

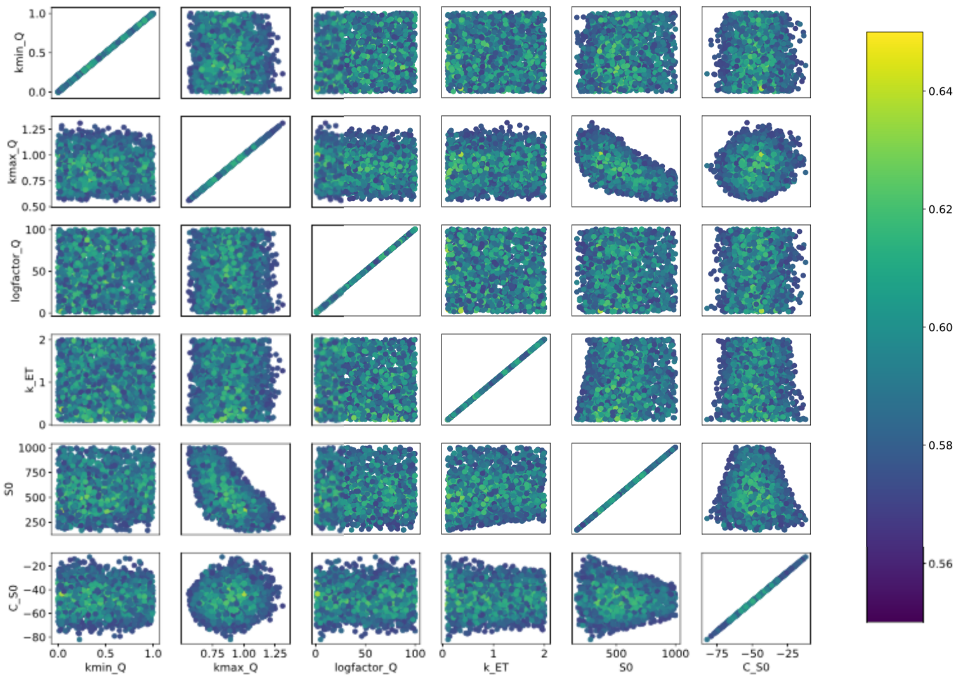

Figure S4. Covariance between parameter performance for all the 95th percentile of parameter sets tested in Monte Carlo simulations. Points are colored by NSE.

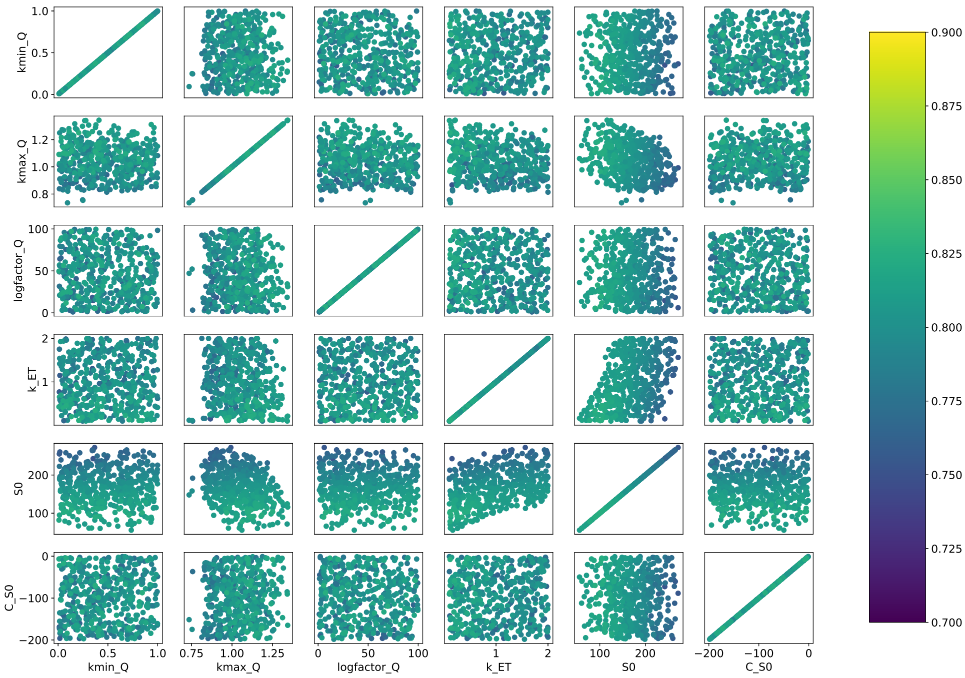

Figure S5. Covariance between parameter performance for all the 95th percentile of parameter sets tested in Monte Carlo simulations. Points are colored by KGE.

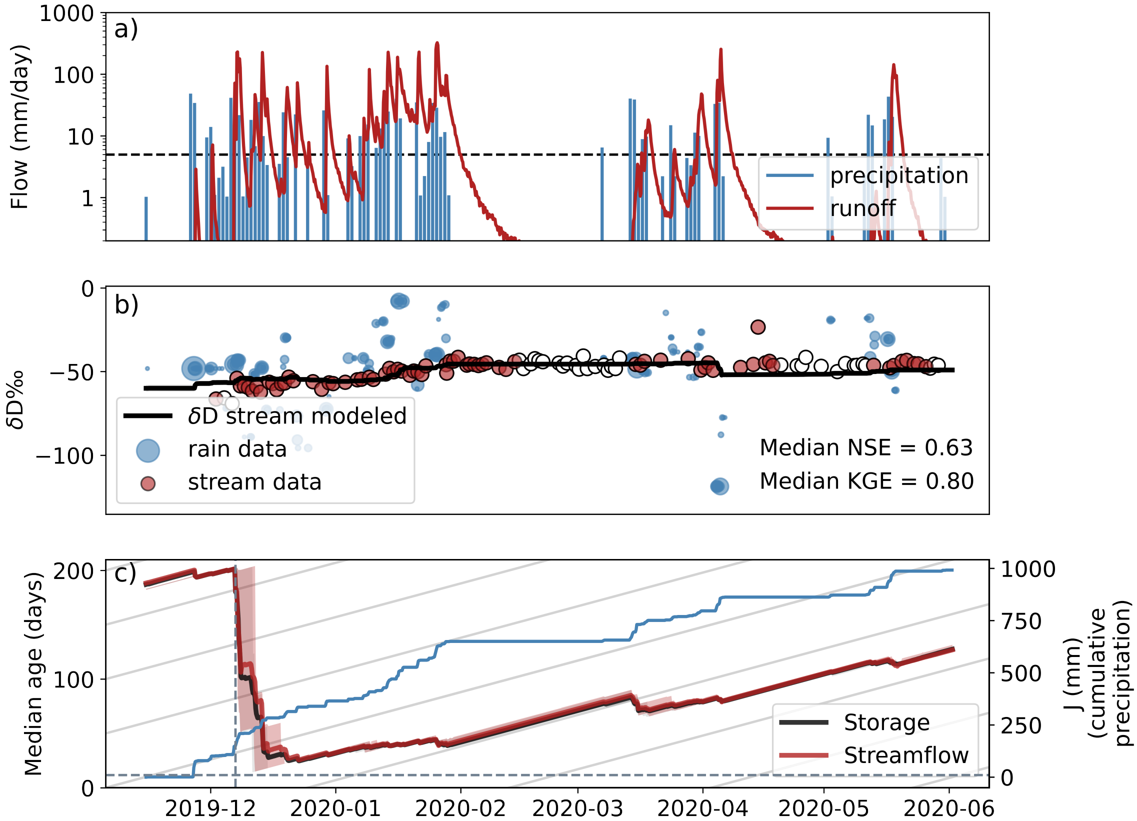

Figure S6. (a) Daily precipitation and instantaneous runoff throughout the wet season 2019-2020. Vertical dashed grey line marks the 5 mm flow threshold above which excess flow is assumed to be SOF. Horizontal dashed grey line marks a median age of 10 days. (b) Confidence bars on SAS model predictions (black line) are smaller than the width of the line. The size of plot markers for rainfall data (blue) are scaled by the volume of precipitation. Data shown in white circles are excluded from calibration of the SAS model due to in-channel evaporative enrichment (streamflow ¡0.05 mm/day). Marked median NSE is the median value among the top 95th percentile of parameter sets. (c) Shading around median ages indicates 25th-75th percentile of ensemble simulations, and blue line is cumulative precipitation. Vertical dashed line marks cumulative precipitation of 150 mm, and horizontal dashed line marks a median age of 10 days.

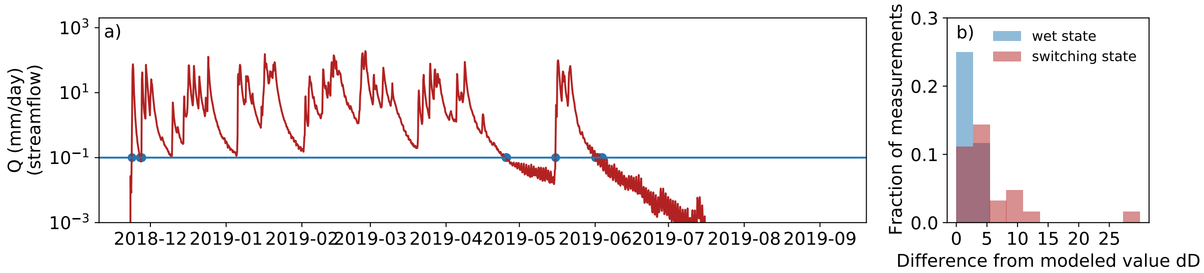

Figure S7. (a) Streamflow at Dry Creek throughout the study period with switches between wet and dry states marked as crossing the blue 0.1 mm/day threshold. Points along the hydrograph where the threshold is crossed are marked with blue dots. (b) Density histograms of absolute error in streamflow concentration between the SAS model and measured streamflow concentration. Times when the catchment is 'switching states' are defined by the 2 days before and after the threshold in panel a is crossed. The 'wet state' is all times not during a switching state when streamflow is above 0.1 mm/day.

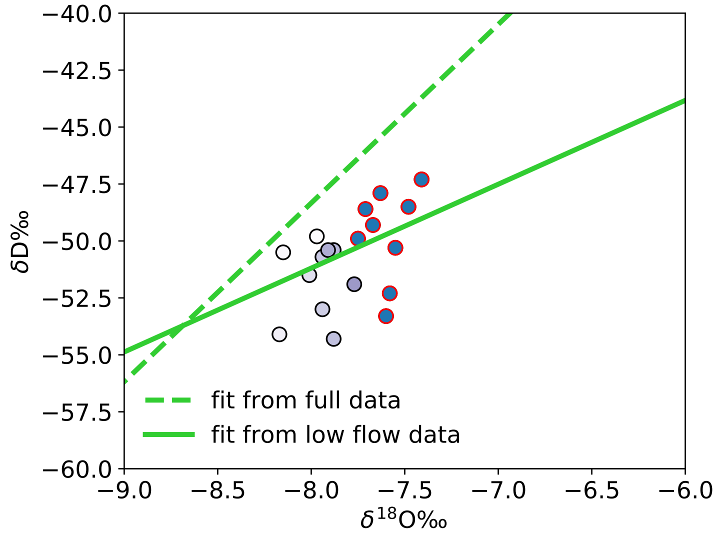

Figure S8. In dual isotope space, the isotope concentration increased from March 15-April 15, 2019 (white to dark blue) as streamflow decreased with no precipitation following a much shallower slope than that for the full set of isotope data. Red outlines indicate that streamflow falls below the 0.05 mm/day threshold for exclusion from calibration.

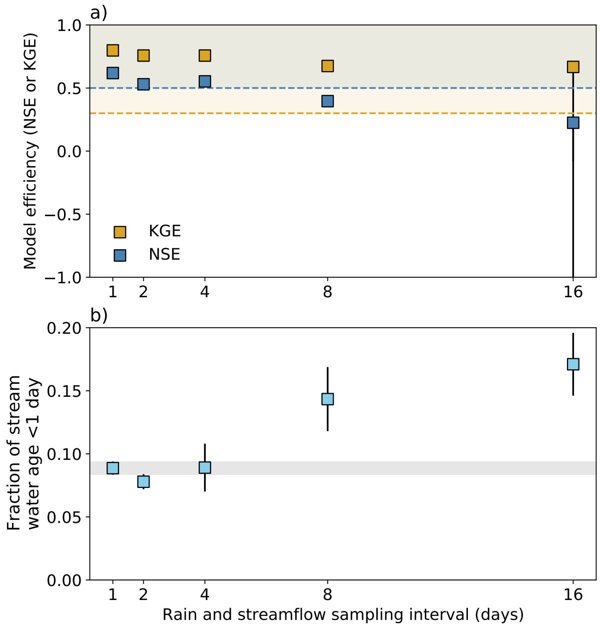

Figure S9. (a) Median model performance of parameterized SAS models as measured by NSE and KGE over a range of sampling intervals for input data. Vertical black lines show 25-75th percentile of model efficiency values. Where not visible, 25-75th percentile range is smaller than the size of the marker. The shaded gold and blue regions indicate the range of KGE and NSE values for a behavioral model (Kirchner, 2003). (b) Mean new water fraction over the full study period with vertical 99% confidence intervals. The horizontal shaded bar is the confidence interval for the 1-day sampling interval.