ESTIMATING THE EFFECTIVE HYDRAULIC PROPERTIES OF THE SUBSURFACE AND THEIR
SPATIOTEMPORAL RESPONSE TO CLIMATE USING A MODIFIED STREAMFLOW RECESSION ANALYSIS

Arik Tashie

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial
fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Geological
Sciences in the School of Arts and Sciences

Chapel Hill
2020

Approved by:

Lawrence E. Band

Ryan E. Emanuel

Jonathan M. Lees

Tamlin M. Pavelsky

Kevin G. Stewart

ABSTRACT

Arik Tashie: Estimating the Effective Hydraulic Properties of the Subsurface and their Spatiotemporal Response to Climate Using a Modified Streamflow Recession Analysis
(Under the direction of Tamlin Pavelsky)

Between periods of precipitation or snowmelt, the volume and timing of streamflow is largely determined by the properties of the subsurface and the time-varying distribution of groundwater storage. While streamflow during these periods (i.e., baseflow) is commonly treated according to a unique storage-discharge relationship, recent innovations in streamflow recession analysis have allowed novel findings regarding the variability of both the stability of baseflow and its nonlinearity (i.e., the concavity of the hydrograph), as well as the regional clustering of these characteristics. Here, I assess traditional and novel models of streamflow recession behavior using historical streamflow data from over 1,000 watersheds in the continental United States (US). Observed streamflow behavior from only nine watersheds often conforms to traditional models, and streamflow behavior from the vast majority (>99%) of watersheds typically conforms to a parsimonious parallel aquifer model which accounts for subsurface heterogeneity. I then apply this conceptual model alongside remotely-sensed estimates of watershed-scale groundwater storage and climate reanalysis estimates of watershed-scale soil moisture, rates of evapotranspiration, and cumulative precipitation to investigate seasonal patterns in both the stability and nonlinearity of streamflow that vary systematically across large regions. I find that coincident watershed storage is the best predictor of baseflow stability in many regions (particularly the Appalachian Mountains) while evapotranspiration from two to three months previous is the best predictor of baseflow stability in other regions (particularly the Pacific Northwest), and discuss the novel finding that streamflow nonlinearity has increased significantly in most watersheds across the US since

1980. Then, I estimate the effective hydraulic properties of all gaged watersheds in the continental US that are largely dam-free by adapting traditional methods of streamflow recession analysis to account for subsurface heterogeneity. Using these results, I develop models of effective hydraulic properties based on estimates of watershed topography, soils, bedrock, and climate, and apply these models to predict the effective hydraulic properties of all watersheds in the continental US. Key practical results of this analysis include: 1) the finding that streamflow is more stable during periods of extended drought than generally predicted; 2) the identification of regional patterns in the response of streamflow to climate change; and 3) a novel dataset representing the effective hydraulic properties of the subsurface for the entire continental US for use in regional-scale hydrological models.

ACKNOWLEDGEMENTS

To my colleagues and collaborators, many thanks for your time and support these last few years. To my advisor Tamlin Pavelsky, your kindness and patience were as essential to completing this work as was your scientific insight. To my lovely and ever supportive wife Carolyn, thanks for being there every step of the way. And to my little ones Evander and Darian, you really didn't help much at all, but I love you anyways.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

a	Scalar in power law relating streamflow to its first time derivative
A	Area
b	Exponent in power law relating streamflow to its first derivative
B	Breadth of aquifer
BFI	Baseflow index
$C_{circles}$	Curvature of recession based on the fitting of a circular model of regression
$C_{exponentials}$	Curvature of recession based on logarithmic models
CLM4.5	Community Land Model Version 4.5
cm	Centimeter
C_{multi}	Curvature of recession based on fitting of all curvature models
$C_{residuals}$	Curvature of recession based on residuals of linear model
C_{slopes}	Curvature of recession based on slope of early and late periods
$C_{squares}$	Curvature of recession based on exponential models
d	Day
D	Depth of aquifer
dQ	Change in streamflow
dt	Change in time
Dwt	Depth of water table
ET	Evapotranspiration
f	Drainable porosity
GAGES-II	Geospatial Attributes of Gages for Evaluating Streamflow
GRACE	Gravity Recovery and Climatology Experiment
h	Water table elevation
HLR	Hydrologic Landscape Regions
HUC	Hydrologic Unit Code
i	Slope of aquifer
K	Hydraulic conductivity

L	Length of stream network
LLO-CV	Leave-location out block cross-validation
LSM	Land surface model
m	Meter
MAE	Mean absolute error
mm	Millimeter
MERRA-2	Modern-Era Retrospective analysis for Research and Applications, Version 2
n	Number
p	Fitting parameter
PET	Potential evapotranspiration
PPT	Precipitation
Q	Streamflow
Q_{recess}	Streamflow during periods of recession
Rday	Recession based on day of year
Rk	Recession based on binned data
Rl	Recession based on the lower envelope of the data point cloud
Rmed	Median value of recession parameters derived from individual events
RMSE	Root mean square error
Rpc	Recession based on the data point cloud
Ru	Recession based on the upper envelope of the data point cloud
S	Drainable storage
SD	Standard deviation
t	time
TOPMODEL	TOPography-based hydrologic MODEL
US	United States
USGS	United States Geological Survey
WT	Water table

CHAPTER 1: AN EMPIRICAL REEVALUATION OF STREAMFLOW RECESSION ANALYSIS AT THE CONTINENTAL SCALE[1]

Section 1: Introduction

Section 1.1: Summary

Streamflow recession analysis is a widely used hydrologic tool that relies on readily available discharge measurements to estimate otherwise unmeasurable watershed-scale properties, to predict low flows, and to parameterize many lumped hydrologic models. Traditional methods apply the simplifying assumptions of outflow from a Boussinesq aquifer, which predicts the slope of the recession curve relating streamflow to its derivative in log-log space to decrease from early-stage to late-stage recession. However, this prediction has not been validated in actual watersheds. Also, recent studies have shown that slopes of observed recession events are often much greater than traditional methods which rely on data point clouds. We analyze recession behavior of 1,027 streams from across the continental US for periods of 10 to 118 years, identifying over 155,000 individual recession events. We find that the average slope of observed recession events is greater than that of the point cloud for all streams. Further, the recession slopes of observed events decrease with time in only 10% of cases, and instead increase with time in 74% of cases. We identify only nine watersheds where observed streamflow behavior often conforms to the predictions of traditional recession analysis, each of which is arid and flat with low permeability. Analysis of our extensive empirical results with a regionalization of catchment hydrologic characteristics indicates that heterogeneity of subsurface flows paths increases the nonlinearity and convexity of observed recession, likely as a function of watershed memory. The

---

practical implications of our analysis are that streamflow is more stable during periods of extended drought than generally predicted.

Section 1.2: Background

Understanding the low flow characteristics of streams is essential for predicting water supply and describing aquatic habitat in headwater systems. Hydrograph recession analysis, a mathematical tool introduced in its modern form by Brutsaert and Nieber [1977], provides a standard approach for investigations of low flow characteristics [Troch et al. 2013]. It is also routinely used to estimate watershed-scale hydrologic features [Pauritsch et al. 2015], perform baseflow separation [Wittenberg and Sivapalan 1999], calculate dynamic storage of a watershed [Krakauer and Temimi 2011], and help calibrate hydrologic models [Jepsen et al. 2016].

Invoking the Dupuit assumptions [Dupuit 1863] that groundwater flow is generally horizontal, and neglecting capillarity, Brutsaert and Nieber applied the Boussinesq equation to describe Darcian flow from an unconfined rectangular aquifer atop a flat impermeable layer into a fully penetrating stream (Figure 1a). A fully recharged aquifer of this type will initially exhibit a flat water table elevation (h) profile that remains fixed at some thickness (Dwt) over much of the width of the aquifer. "Early" recession transitions to "late" recession as the aquifer, draining more quickly nearer the stream, generates a curvilinear h profile approximating an incomplete beta function with h being equal to Dwt only at the boundary furthest from the stream [Brutsaert and Lopez 1998]. These dynamics in the water table profile have a predictable effect on the hydrograph, generating a highly concave hydrograph (i.e., high "nonlinearity") in the early period and a less concave hydrograph (i.e., low "nonlinearity") in the late period (Figure 1b-c).

Application of the Boussinesq equation to hillslope recession allows unit discharge (q) to the stream network can be solved simply as a function of hydrological conductivity (K), drainable porosity (f), aquifer depth (D), aquifer breadth (B), impermeable layer slope (i), and h profile, while watershed

discharge (Q) may be estimated as the product of q and 2x the length of the stream (L), assuming similar

aquifer characteristics throughout the watershed. In parallel, Brutsaert and Nieber [1977] describe how

if all other inflows and outflows to a watershed are negligible, Q may be modeled as a function of

aquifer storage (S), and the change in streamflow (dQ/dt) assessed as a function of time-independent Q:

$$\frac{-dQ}{dt} = aQ^b \qquad \text{Eq. 1}$$

where a is a function of static watershed characteristics (K, f, D, B, L, and i), and b is a function of the

geometry of the contributing aquifer as well as the h profile that defines early and late periods of

recession. Since a and b can be estimated by plotting Q against dQ/dt in log-log space, it then becomes

possible to estimate basin-scale hydrological properties (e.g. K, D, or L) that are otherwise very difficult

to measure [Szilagyi et al. 1998].
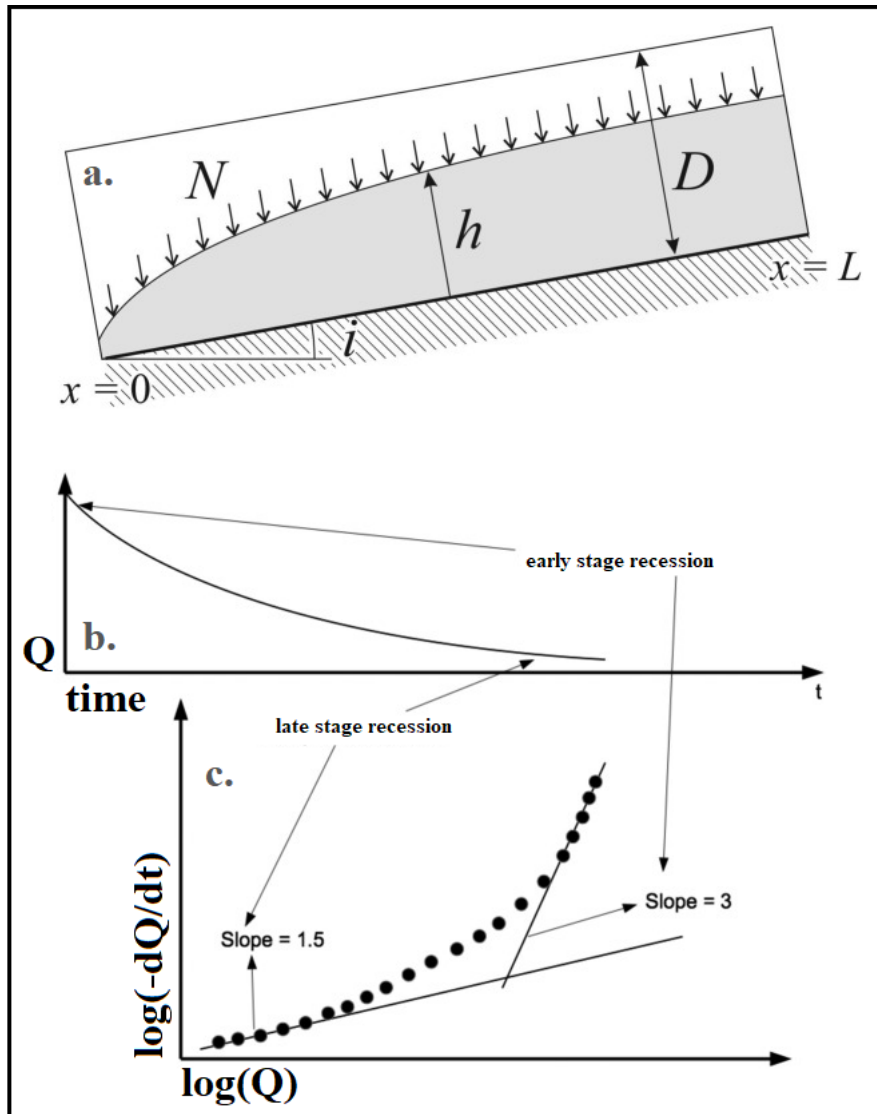
Figure 1: (a) definition sketch of a Dupuit-Boussinesq aquifer under a constant rate of recharge (N) atop an impermeable layer with a slope of i, depth of D, length of L, and water table elevation of h; (b) a hydrograph and (c) a recession plot illustrating early and late stage recession. Figure adapted from Troch et al., 2013.

Fundamental predictions of the Boussinesq aquifer relating to the slope of the recession curve have not been validated with extensive catchment-scale field observations, but have been repeatedly validated with computer models [e.g, Rupp and Selker 2006; Szilagyi et al. 1998], with table-top models [Guerin et al. 2014; Luo et al. 2018], and in highly monitored hillslopes [Clark et al. 2009], as well as analytically. Theoretical outflow from a Boussinesq aquifer yields two flow regimes, with higher initial b values (around 3) when Dwt approximates the h profile for much of the aquifer and which drop to around 1 or 1.5 when the h profile only approximates Dwt at the boundary furthest from the stream (Figure 1c). As a result, Brutsaert and Nieber [1977] graphically defined an upper and lower envelope for

the data point cloud of all recession data to describe early- and late-stage recession, respectively (Figure 2). Many studies have followed suit, though because of the difficulty in developing a rigorous methodology for delineating the upper from the lower envelope, most subsequent studies have used a single line of regression to the data point cloud as a whole [Tallaksen 1995], to the central tendency of the point cloud [Kirchner 2009], or have used a combined approach incorporating multiple methods [e.g., Sanchez-Murillo 2015]. Values of b between 1 and 2 are generally deemed to be appropriate, while significantly higher values may be indicative of processes in violation of the Dupuit-Boussinesq assumptions [Tallaksen 1995] or the result of errors in methodologies or data uncertainty, like discharge estimates from a rating curve [Westerberg and McMillan 2015].

While the original assumptions underlying recession analysis pioneered by Brutsaert and Nieber [1977] are rigid, analytical solutions have been extended to accommodate sloping aquifers [Rupp and Selker 2006] (Figure 1a), nonnegligible groundwater outflows or groundwater withdrawals [Thomas et al. 2013; Vogel and Famiglietti 2015], steep hillslopes [Hunter et al. 2005], substantial unsaturated zone flow [Luo et al. 2018], substantial sources and sinks of water at the bedrock interface [Bartlett and Porporato 2018], and flow paths that are not completely horizontal [Harman and Kim 2019]. In fact, many researchers have isolated and estimated watershed features or processes, such as the consumption of groundwater by evapotranspiration, in violation of the Dupuit-Boussinesq assumptions through investigation of anomalies in streamflow records [Palmorth et al. 2010; Szilagyi et al. 2007]. Similarly, violations of Dupuit-Boussinesq assumptions have been invoked as potential explanations for the spread of observed recession values in recession point clouds [Brutsaert and Nieber 1977, Shaw and Riha 2012].

Indeed, a holistic accounting of recession behavior must include a description of the shape of the point cloud itself, as well as the shape of the individual events that define it. Apart from measurement error, uncertainty in Q calculated with a rating curve, or effects of evapotranspiration

[Szilyagi et al. 2007], an ideal Boussinesq aquifer predicts vertical scatter where early-stage recession is captured. According to this interpretation of the recession point cloud, the lower envelope defines actual late-stage recession for all values of Q, and the scatter in the vertical direction represents a series of early-stage recession flows stacked horizontally beside one another, while the upper envelope represents early-stage recession during the period of highest Q captured in the point cloud.

However, many recent studies relying on the properties of individual recession events have found that observed recession often violates many of these predictions. For instance, the median b value of individual events has been shown to be generally greater than b values generated using the data point cloud as a whole in several watersheds [e.g., Shaw and Riha 2012]. One possible explanation is that some individual recession events represent an early stage of recession, when the h profile is still steep, and that given time the b values of these events would decay to a value that agrees with the point cloud. In such cases, the slope of these individual recession events should be concave, as the recession slope transitions from early- to late-stage recession. This concavity has been well documented in analytical solutions [Rupp and Selker 2006] as well as in table-top models [Guerin et al. 2015]. For observed recession events in natural watersheds, however, evidence of concavity is absent in the literature. On the contrary, the convexity of individual recession events has been noted in several studies, in direct contradiction of hydraulic aquifer theory illustrated in Figure 1, though this discrepancy has not been thoroughly investigated [Ghosh et al. 2016; Shaw and Riha 2012; Tashie et al. 2019; McMillan et al. 2011].

Further, several studies have identified b values that consistently exceed the theoretical early stage value of 3 [Mutzner et al. 2013; Gosh et al. 2016; Tashie et al. 2019]. Anomalous b values have been described as resulting from potential data issues like observational error or the limited data points available for individual recession events [Shaw and Riha 2012], spatial variation in rainfall [Biswal and Kumar 2014], the effects of evapotranspiration [Shaw and Riha 2012], seepage or leakage to a deep

aquifer [Wang and Cai 2010], persistent groundwater recharge [Li et al. 2017], or the persistence of overland flow or stormflow [Ye et al. 2014].

We hypothesize that the recession of streamflow in actual watersheds consistently violates traditional assumptions of recession behavior. Specifically, we test the following hypotheses:

1)   the observed recession of individual events tends to be significantly more nonlinear than recession determined using any data point cloud method;

2)   the nonlinearity of observed recession events does not tend to decrease with time (i.e., concave recession), but instead tends to increase with time (i.e., convex recession);

3)   and the typical behavior of observed recession events is not primarily a function of hillslope hydraulics as classically proposed, but is instead a function of the heterogeneity of catchment-scale hydrologic properties and topoclimate-induced variability in the distribution of water storage.

To test these hypotheses, we analyze the streamflow records of watersheds across the continental United States (US) using three methods of traditional streamflow recession analysis as well as several newly developed methods for analysis of individual events. We explore the recession characteristics of over 1,000 watersheds in relation to their physio-climatic features, and discuss the implications of our results in relation to the application of recession analysis more generally.

Section 2: Methods

Sectin 2.1: Recession Analysis

We analyzed streamflow records of all 1946 watersheds across the continental US that are gaged by the US Geological Survey (USGS) and had been relatively unimpacted by human development as of 2009, as described by the Geospatial Attributes of Gages for Evaluating Streamflow (GAGES-II) data set [Falcone et al. 2010]. All flow data is reported daily, and we removed from analysis all time periods when flow was identified as obstructed or affected by measurement error.

Because of the relative novelty of analyzing the curves of individual recession events, there

exists no universally agreed-upon methodology for selecting periods of recession. Dralle et al. [2017]

recently quantified the methodological uncertainty of selection criteria, which affected the actual value

of parameters and model goodness-of-fit, though the ranking of typical event-scale parameters among

different watersheds was robust. In our primary analysis, we followed their recommendations that a

recession period is identified when both Q and -dQ/dt decrease for consecutive days, no days are

removed from the beginning or end of an event, and recession flow is not delimited by a minimum or

maximum flow threshold. Though Dralle et al. [2017] recommend a minimum of 4 consecutive days to

identify a recession event, because one of our methods for assessing recession curvature relies on linear

regression of early and late segments of individual curves, we required at least 5 consecutive days per

event. It is also important to note that the relatively unrestricted event selection methods

recommended by Dralle et al. [2017] do not eliminate from analysis days during which precipitation falls

in quantities sufficient to affect the shape of the recession curve due to decreases in -dQ/dt but

insufficient to actually increase Q. However, because of the requirement of monotonic decrease in -

dQ/dt, this potential effect should generally be restricted to the final single day of selected events, as

the cessation of rainfall generally results in an increase in -dQ/dt. To ensure that our analysis for each

watershed was robust, we only report results watersheds for which we could identify at least 50 unique

recession events.

To ensure our results were robust and not sensitive to selection criteria, we also reanalyzed all

data according to three sets of alternative selection criteria. In the first alternative scenario (A1), we

eliminated the first 2 days following the initiation of a recession event to ensure that recession events

included no periods of time when streamflow was affected by potential stormflow. To preserve our

methods of assessing nonlinearity, we also increased the minimum number of days required for

identifying a recession event from 5 to 7. A1 generated 32,008 individual recession events across 355

8

watersheds. In the second alternative scenario (A2), we assessed recession only during extended dry periods by doubling the minimum number of days required for identifying recession to ten days, and we eliminated the first two days and final single day of each recession event. We also decreased from 50 to 35 the minimum number of events that needed to be captured in order for us to include a watershed in our analysis. A2 generated 1,561 individual recession events across 31 watersheds. In the third alternative scenario (A3), we eliminated the final day of recession, increased the number of days required for identifying a recession event from 5 to 6, and decreased from 50 to 40 the minimum number of events that needed to be captured in order for us to include a watershed. A3 generated 82,160 individual recession events across 685 watersheds. The results from these analyses are summarized below and presented in detail in the supplementary material (S1).

For each watershed, we performed recession analysis on the entire data point cloud of daily average $-dQ/dt$ plotted against Q in log-log space. We relied on three of the most common methods of traditional analysis: 1) point cloud linear regression (Rpc) [Vogel and Kroll 1992]; 2) the central tendency of the point cloud according to the methods of Kirchner [2009] (Rk); 3) and the upper (Ru) and lower (Rl) envelopes as suggested by Brutsaert and Nieber [1977]. While Ru and Rl represent the earliest and most theoretically sound implementation of traditional recession analysis, they are rarely used because of the subjectivity of their implementation. When comparing Rl with Rpc and Rk, Stoelzle et al. [2013] defined Rl using quantile regression, noting that previous studies had parameterized a model of recession with 5% of the data below it. However, these methods convolve Rl and Ru, and analyzing the transition from early-stage to late-stage recession requires they be differentiated. To deconvolve Rl from Ru, we defined a "knickpoint" separating these two periods of recession as the point that maximally differentiates estimates of their slopes. First, we apportioned the data into 15 bins according to ascending values of Q, then took the median value of Q and the median value of the lowest 20% of dQ/dt values. We estimated values of Ru and Rl according to potential knickpoints located between the 4th and the 11th bins using

9

linear models. We estimated the knickpoint as the bin where the difference between these slopes was greatest, and we estimated Rl and Ru using bins below and above the knickpoint (Figure 2). The b value of each individual recession event (Figure 3) was estimated using the slope of the linear model (Figure 4).

Because multiple previous studies have identified seasonal variability in recession characteristics using both traditional point cloud-based methods and event-based methods [e.g., Wittenberg 2003; Shaw and Riha 2012], we attempted to isolate seasonal variability in the $Q - dQ/dt$ relationship for each watershed. To do so, we identified a median value of both Q and -dQ/dt for each day of the year within a two-sided 30-day moving window. A value for b relating the seasonal dynamics of each watershed (Rday) was then computed using a linear model of these average values (Figure 2).
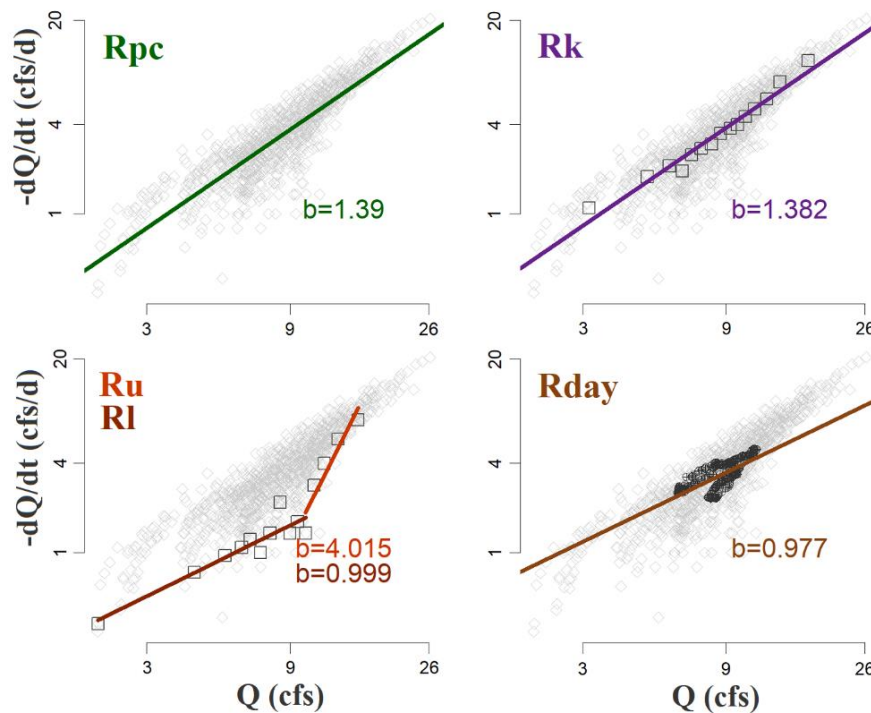
Figure 2: example of each type of cloud-based regression applied to the Piscataquis River near Blanchard, Maine (USGS gage 01031300). Each grey dot represents the daily mean value of Q and -dQ/dt during a recession event. Clockwise from top left: linear regression of the point cloud (Rpc, in green), the Kirchner method (Rk, in purple), median values of days of the year (Rday, in brown), and the upper envelope (Ru, in light red) along with the lower envelope (Rl, in dark red).
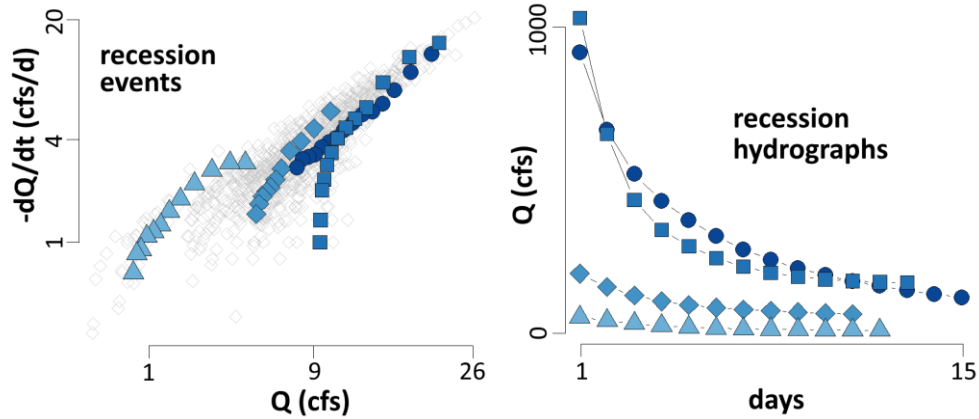
Figure 3: four individual recession events identified at the Piscataquis River near Blanchard, Maine (USGS gage 01031300) represented on a recession plot (left) and a hydrograph (right).

Section 2.2: Recession Curvature

A complicating factor in our analysis was the lack of established methods in hydrology for assessing the curvature (i.e., concavity or convexity) of a small sequence of data points. The curvature of each recession event was initially assessed according to the residual pattern of each linear model. While visual assessment of a small subsample of events often led to identification of curvature, most recession events had far too few data points for quantitative assessment, e.g. the Shapiro-Wilk test [Royston 1995]. Therefore, we developed a suite of 5 methods for ascertaining the relative likelihood of convexity versus concavity for each recession event (Figure 4):

- $C_{residuals}$: we summed the value of the first and final residual of each linear model, with a positive sum indicating a population more likely to be concave and a negative sum indicating a population more likely to be convex.

- C_{slopes}: we applied a linear model to the first half of each recession event and compared the computed b value with that of the final half of the same event, with concavity being indicated by higher b values in the early period and convexity by lower b values in the early period.

11

- $C_{exponential}$: we assessed the r2 value of models predicting log(-dQ/dt) using transformations of log(Q). Recession events were identified as concave where log(-dQ/dt) was better predicted using log transformations of the data than exponential transformations.

- $C_{squares}$: similarly, concavity was identified where log(-dQ/dt) was better predicted using the square-root of log(Q) than the square of log(Q).

- $C_{circles}$: finally, we fitted a circle to each recession event using circular regression [http://people.cas.uab.edu/~mosya/cl/]. Concavity was identified when the center of the circle was above the line of best fit from a linear model, and convexity identified when the center of the circle was below that line.

Each method has limitations in terms of small sample size and random variation effects, so we used a balance of evidence approach. Specifically, we combined these 5 individual metrics into a single indicator (C_{multi}) according to the following method: where the same recession event was identified as concave (or convex) by at least 4 of the 5 individual metrics, we defined that event as likely concave (or convex), and we defined recession events where concavity (or convexity) was identified by fewer than 4 methods as indeterminate.

We assessed the observed slope and curvature of each individual event in relation to theoretical values of early-stage recession, late-stage recession, and the transition from early- to late-stage recession. Because theoretical late-stage b values can range from 1 to 1.5, we used empirical values derived from point cloud recession to establish a range of likely late-stage b values. We used Rpc as the standard measure of comparison because it is both robust (compared with Ru, Rl, and Rday, as described below in Results) and the most commonly used method in the field of recession analysis. First, we defined each event as shallow, equivalent, or steep where the b value of that event was less than 90%, between 90% and 110%, or over 110% of the Rpc b value for that watershed. Then, we characterized each event as showing concavity, convexity, or indeterminacy according to C_{multi}, thereby
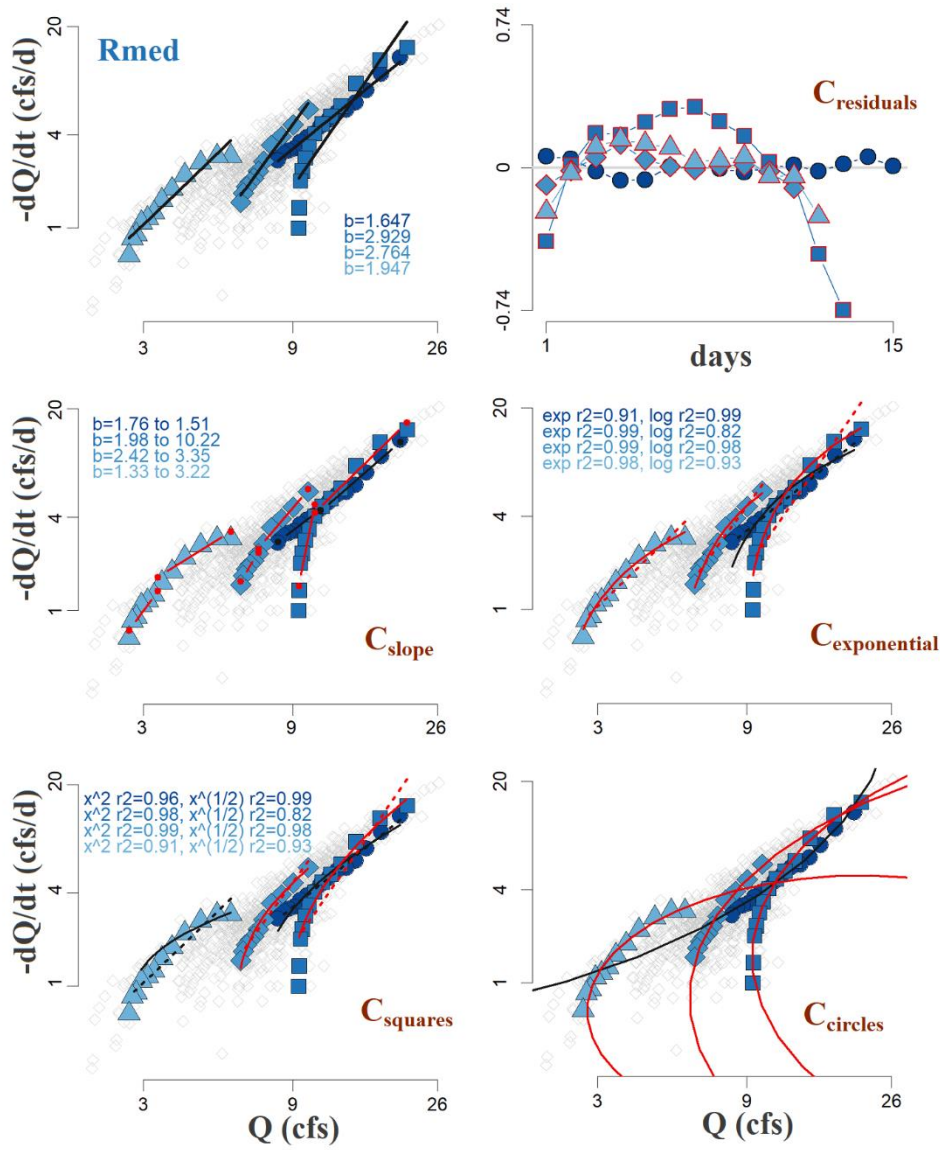
Figure 4: examples of event-based b values (top left), and each method of assessment of recession slope curvature applied to four recession events at the Piscataquis River near Blanchard, Maine (USGS gage 01031300): (Cresiduals) residual analysis; (Cslopes) comparison of b values of linear models of the first and second half of each event; (Cexponential) comparison of the fit of exponential (solid line) versus logarithmic (dashed line) models of each event; (Csquares) comparison of square (solid line) versus square root (dashed line) models of each event; and (Ccircles) the circle of best fit. For each metric of concavity, events identified as concave are in black, while events identified as convex are in red.

generating 9 categories, and calculated the fraction of individual events per watershed that fell into

each category.

Finally, we related b values and curvature to summaries of catchment geomorphology, soil

properties, vegetation, and climate. We relied on summary attributes in the GAGES-II database to

13

describe the relationship between recession characteristics and individual physical drivers (e.g., annual

precipitation or stream network density). To identify multivariate drivers of recession we used

Hydrologic Landscape Regions (HLR) [Wolock 2003], a regionalization of hydrologic characteristics based

on catchment similarity according to 8-digit hydrologic unit codes.

Section 3: Results

Section 3.1: Nonlinearity, or Values of b

We identified 169,960 individual streamflow recession events across the 1946 watersheds

included in our study. After removing all watersheds that failed to meet our threshold of a minimum of

50 events, we retained for further analysis a total of 155,309 individual recession events across 1,027

watersheds. The b values generated by Rpc and Rk were well within the expected range of values and

consistent with values previously reported in the literature. The median b value of all watersheds

according to Rpc was 1.319, with first and third quartile values of 1.162 and 1.541 (Figure 5) and the

values of b derived using Rk were nearly identical (first, second, and third quartiles being 1.159, 1.320,
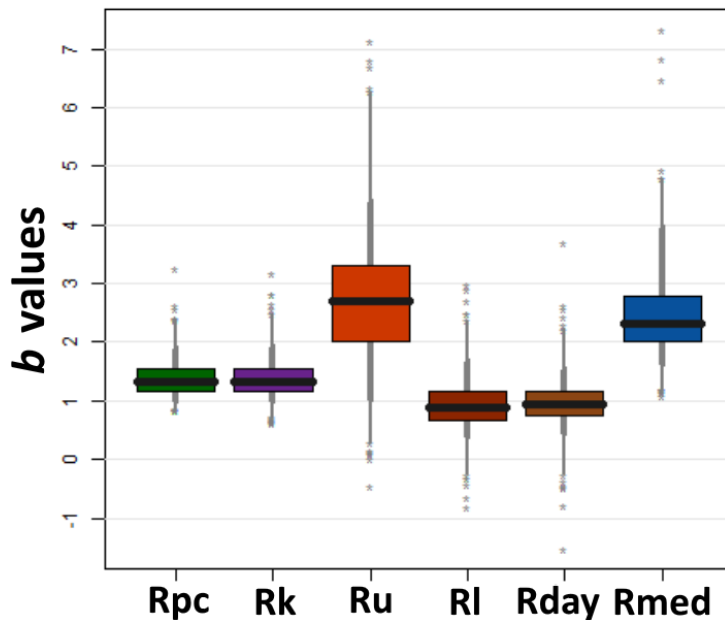
Figure 5: values of b calculated for each watershed according to point cloud regression (Rpc), Kirchner-style regression (Rk), the upper envelope (Ru), the lower envelope (Rl), median values of days of the year (Rday), and the median value of all events (Rmed). Thick black lines represent median values, boxes represent the 25th to 75th percentiles, thick grey lines represent the 5th to 95th percentiles, thin grey lines represent 99% of all data, and stars represent outliers.

and 1.539). Also, b values according to Rpc and Rk were strongly correlated, with minimal bias (Figure 6), supporting similar findings from a smaller subset of watersheds by Stoelzle et al. [2013].

The median values of b according to Ru and Rl were close to the values predicted by recession from a Boussinesq aquifer during early- and late-period recession, in agreement with results from traditional models. Median Ru b values were 2.701, though for 45 watersheds they were below 1. Similarly, typical values of b for Rl were close to predicted values for a linear aquifer (median of 0.874), with several outliers including 11 watersheds where b values were negative and 20 watersheds where b values were above 2. The outliers and physically unrealistic values for both Ru and Rl are likely explained by data limitations and the mechanistic approach we applied. Lines of regression used to calculate Ru and Rl were defined by as few as 15 data points reflective of values of daily Q and -dQ/dt measurements. Therefore, our estimates of Ru and Rl are likely only robust in assessing typical values across multiple watersheds, while estimates of Ru and Rl for individual watersheds are subject to uncertainty.

The median b value of all individual recession events at each watershed (Rmed) were much higher than those of all other methods except Ru, implying that Rpc and Rk fundamentally underestimate the nonlinearity of observed streamflow recession. Rmed b values ranged from a minimum of 1.084 to a maximum of 7.337, though most watersheds had b values between 2 and 3 with a median of 2.311. Rmed b values were greater than those generated by either Rpc or Rk in all 1,027 watersheds, as well as those generated by Rl in all but one watershed. Rmed b values were poor predictors of b values for Ru, Rl, and Rday, though they strongly correlated with Rpc and Rk, with r2 values of 0.496 and 0.453, respectively. Model performance improved when Rmed and Rday were simultaneously considered for predicting Rpc and Rk, with r2 values of 0.559 and 0.516.

Relative b values according to Rpc, Rk, Rl, and Rmed exhibited a consistent geographic pattern, with relatively elevated values in the southern Appalachian Mountains, the Pacific Northwest, and the Gulf Coast (Figure 7). While availability of data was too limited to draw strong conclusions in much of
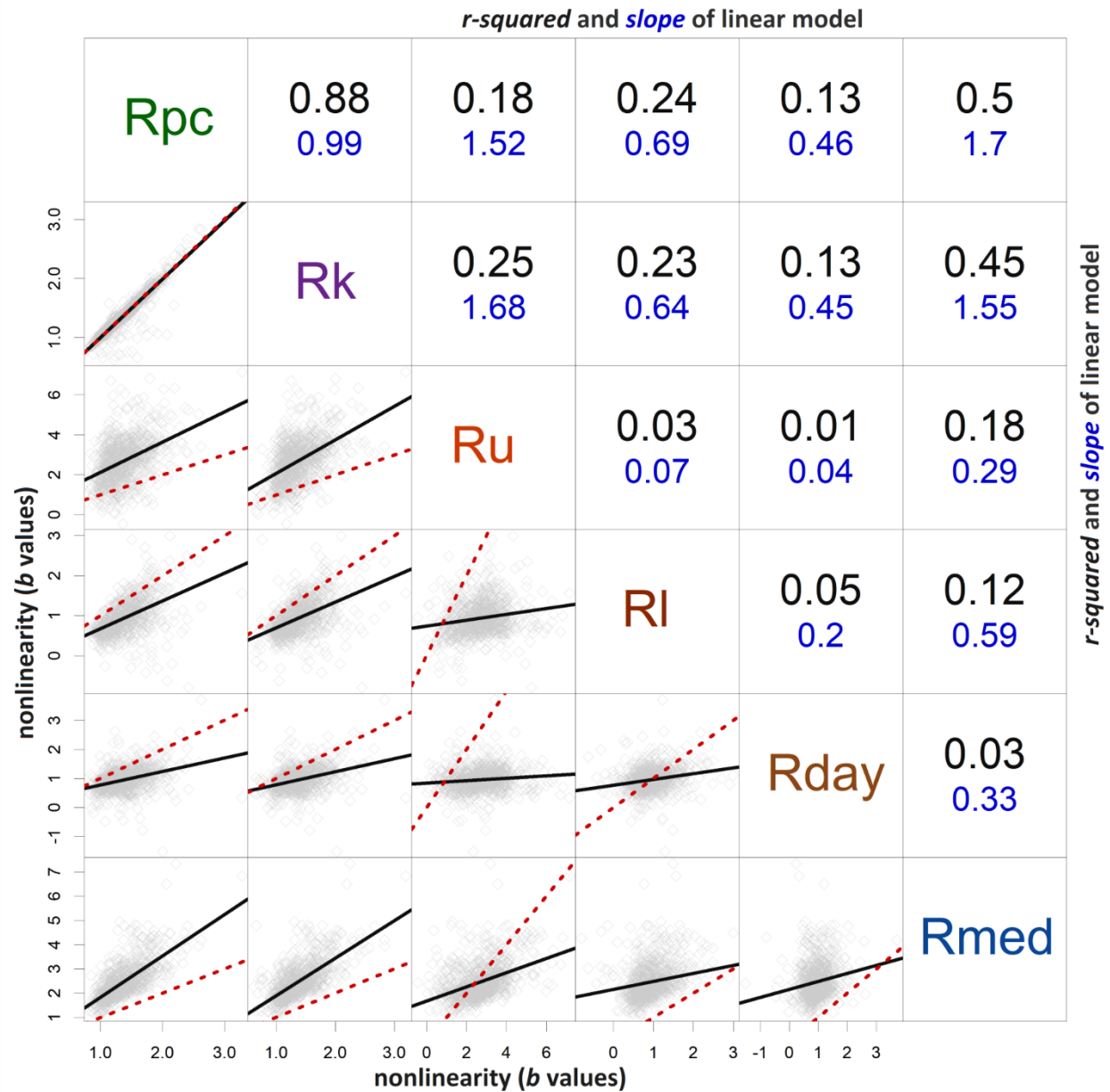
Figure 6: regression analysis comparing all methods calculating b values. Plots on the lower left represent typical b values calculated for each watershed using each method identified along the diagonal, with a black line representing a linear regression of those values, and a dashed red line a one-to-one line. Values in the top right represent the r-squared value (black) of each linear model as well as the slope of that model (blue).

the US between the Appalachian Mountains and the Rocky Mountains, potential hotspots of event-based nonlinearity are also apparent along the east coast of Lake Michigan, around the New Madrid Seismic Zone centered on the state of Missouri, and in the northern portion of the American Rocky Mountains. These patterns were more strongly evident in Rmed than Rpc, Rk, or Rl, and no strong geographic patterns were evident in Ru and Rday.
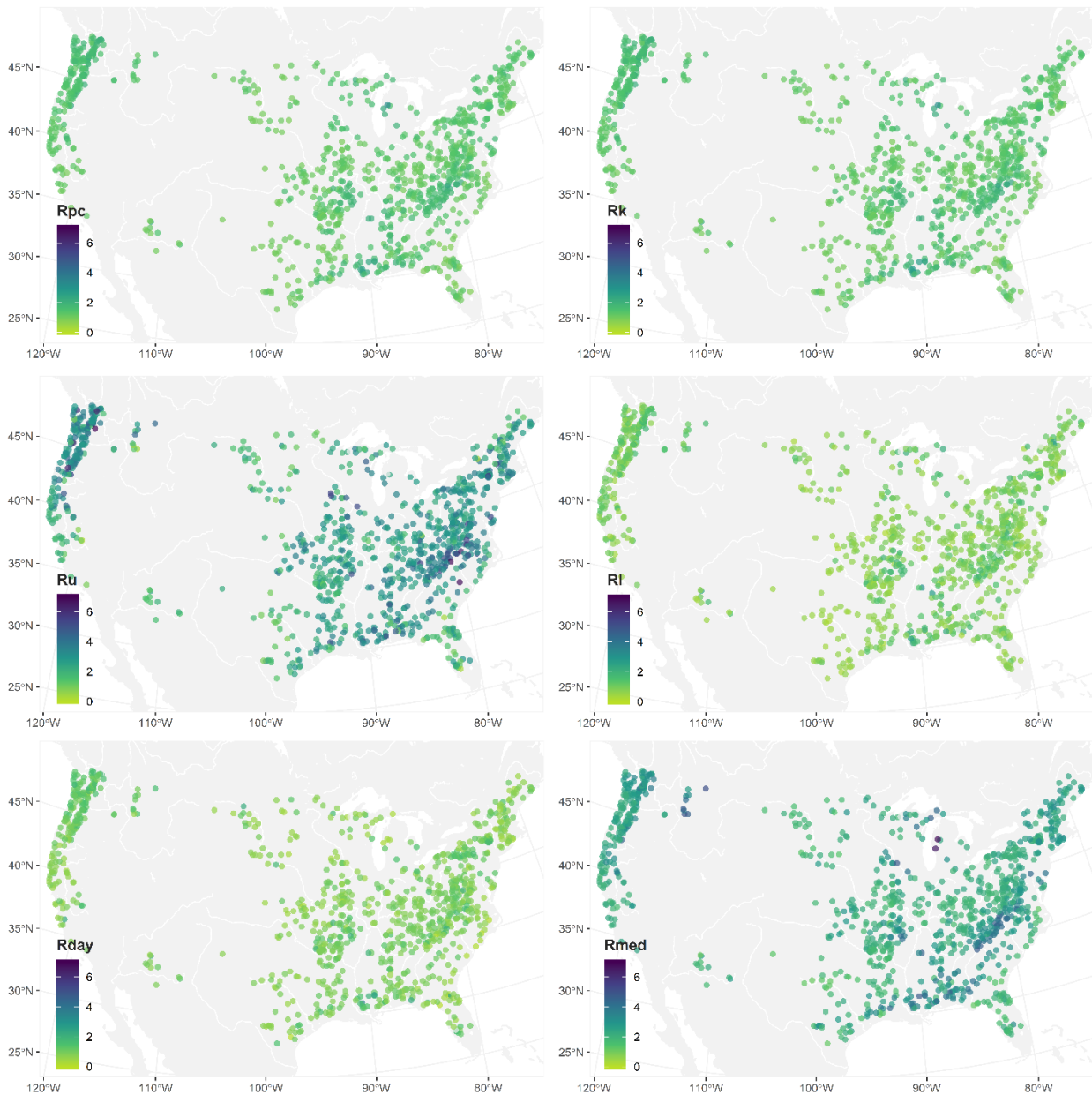
Figure 7: catchment nonlinearity estimated according to Rpc (top left), Rk (top right), Ru (middle left), Rl (middle right), Rday (bottom left), and Rmed (bottom right).

Section 3.2: Recession Curvature

The majority of recession events exhibited convexity, not concavity, according to all 5 methods of assessing nonlinearity. Of the 154,662 individual recession events captured in this study, 18.0%, 21.0%, 38.7%, 40.1%, and 17.9% were identified as concave according to $C_{residuals}$, C_{slopes}, $C_{exponential}$, $C_{squares}$, and $C_{circles}$, respectively (Figure 8). The remainder were identified as convex. According to the more robust indicator of C_{multi}, in only one watershed were more than half of recession events concave (at 52.1%). In only nine watersheds were recession events more likely to be concave than convex (Figure 9). This represents less than 1% of all watersheds in this study. A typical watershed exhibited recessions of concave, convex, or indeterminate shape in 9.25%, 76.7%, and 13.7% of events, respectively. Similarly, across all watersheds, 10.1%, 74.5%, and 15.4% of events were identified as concave, convex, or indeterminate.
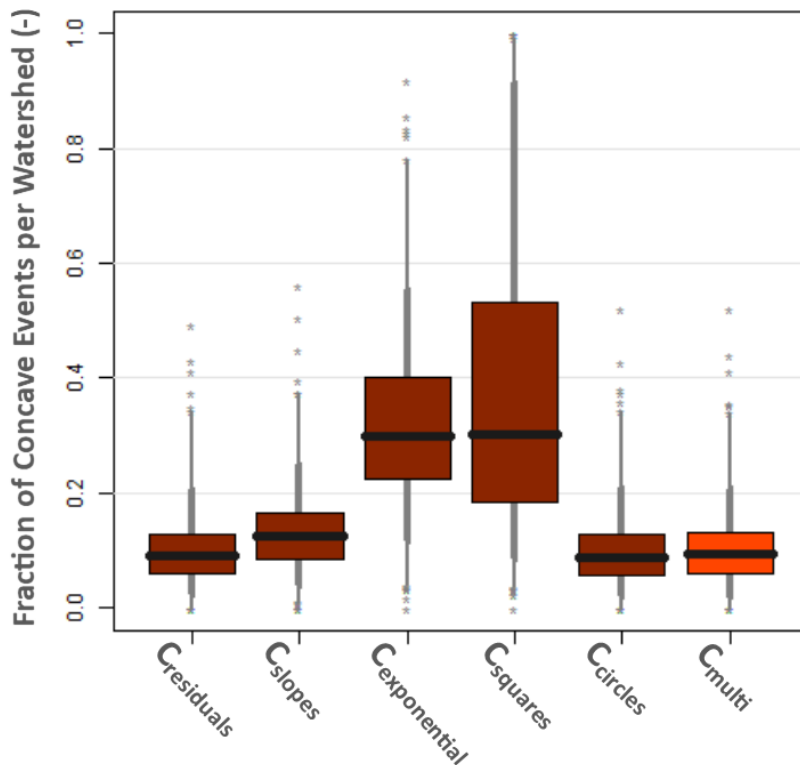
Figure 8: assessment of concavity according to each method: residual analysis ($C_{residuals}$); early versus late slope comparison (C_{slopes}); exponential versus logarithmic model ($C_{exponential}$); square versus square root model ($C_{squares}$); circular regression ($C_{circles}$); and multi-metric indicator (C_{multi}). Thick black lines represent median values, boxes represent the 25th to 75th percentiles, thick grey lines represent the 5th to 95th percentiles, thin grey lines represent 99% of all data, and stars represent outliers.
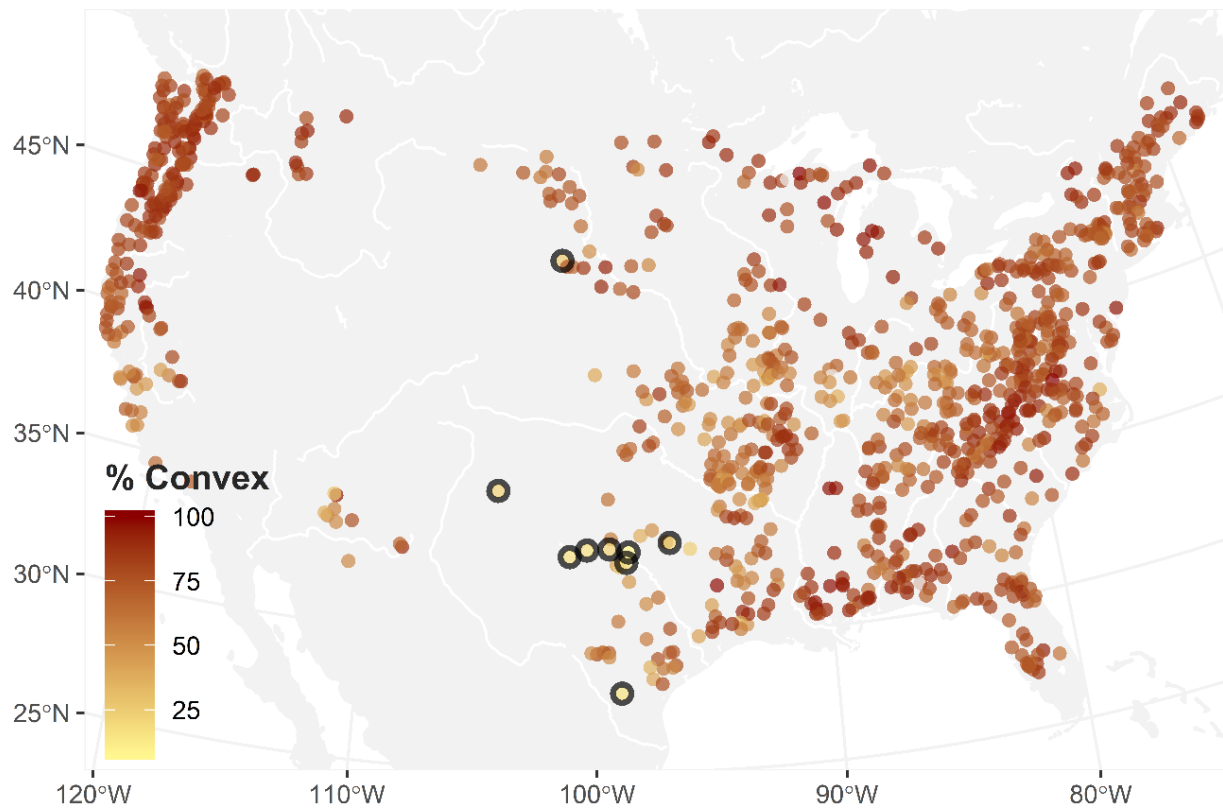
Figure 9: frequency of convexity of recession. Watersheds where recession concavity is more likely than recession convexity are outlined with a black circle.

We assessed whether curvature exacerbated or moderated the variance of typical b values among watersheds by analyzing the likelihood of convexity (or concavity) in relation to values of b. b values from each recession method were positively correlated with convexity and negatively correlated with both concavity and indeterminate shape (Table 1). While Rpc was the strongest predictor of concavity, with an r2 of 0.270, Rmed was the best predictor of both indeterminate shape and convexity, with an r2 of 0.359 and 0.377 respectively.

Table 1: correlation of recession slopes with recession curvature. Analysis performed using a linear model predicting watershed-average shape of recession curves (concave, indeterminate, or convex)

		Rpc	Rk	Ru	Rl	Rday	Rmed
Concave	r^2	0.270	0.247	0.089	0.082	0.036	0.223
	slope	*-0.11*	*-0.10*	*-0.02*	*-0.04*	*-0.03*	*-0.04*
Indeterminate	r^2	0.350	0.309	0.120	0.137	0.043	0.359
	slope	*-0.38*	*-0.34*	*-0.06*	*-0.17*	*-0.11*	*-0.16*
Convex	r^2	0.340	0.296	0.119	0.145	0.040	0.377
	slope	*0.28*	*0.25*	*0.05*	*0.13*	*0.08*	*0.12*

Categorization of each observed recession event according to its slope and curvature in relation to theoretical values of early-stage recession, late-stage recession, or the transition from early- to late-stage recession, is related in Figure 10. For the vast majority of watersheds, most observed recession events were both steeper than Rpc and convex. Observed recession events which were steeper than Rpc and of indeterminate curvature were also common in most watersheds, while observed recession events which were steeper than Rpc and concave comprised fewer than 10% of events in most watersheds. All other types of recession curves were very rare.

Section 3.3 Catchment-Scale Drivers

The GAGES-II data set provides an extensive list of summary attributes for each catchment. We assessed each for its predictive power in describing both Rmed b values and the likelihood of convexity using a linear model, a logarithmic model, and an exponential model. Confirming Patnaik et al.'s [2018] results from a smaller data set, b values were poorly predicted by each individual catchment attribute. The single best predictor of b was the log of the fraction of the watershed comprised of soil group HGB (a deep permeable soil; r2 = 0.147), with other properties of soil, precipitation (PPT), and topography being the next best predictors with r^2 values generally below 0.06. The likelihood of convexity was

poorly predicted by all catchment attributes. The best predictor was by the log of catchment potential evapotranspiration (PET; $r^2 = 0.086$) followed by annual average PPT. Both values of b and likelihood of convexity were negatively correlated with drier conditions (i.e., higher PET or lower PPT), low relief, and impermeability.

We also analyzed multivariate drivers of Rmed b values and convexity by partitioning watersheds according to Hydrologic Landscape Regions (HLR) [Wolock 2003], a regionalization of hydrologic characteristics based on catchment similarity in geology, topography, soils, vegetation, and
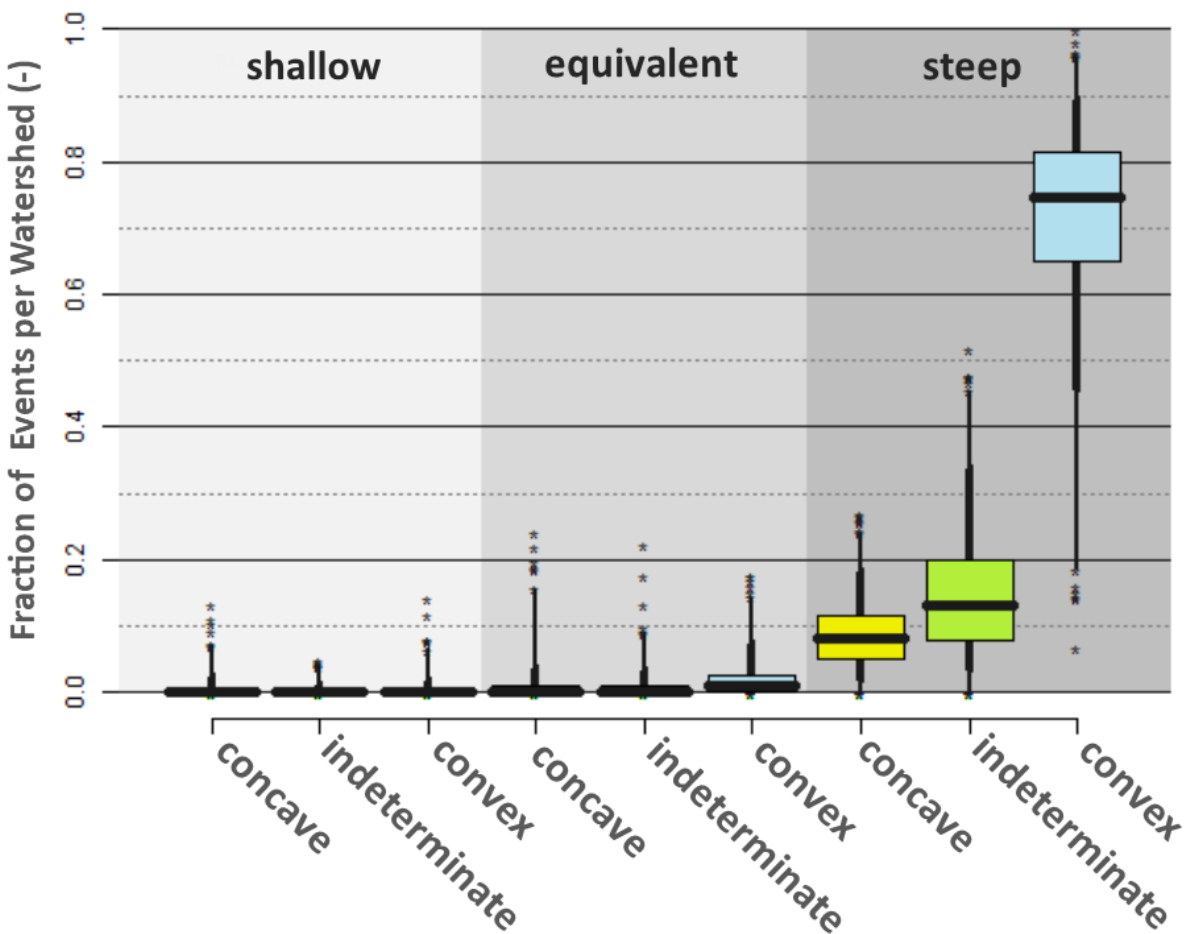
Figure 10: categories of recession form. For each watershed, the fraction of individual events whose b values are both shallower than (light grey on left), equivalent to (medium grey in middle), or steeper than (dark grey on right) that of Rpc, as well as identified as likely concave (yellow), indeterminate (green), or convex (blue). Thick, horizontal black lines represent median values, boxes represent the 25th to 75th percentiles, thick grey lines represent the 5th to 95th percentiles, thin grey lines represent 99% of all data, and stars represent outliers.

climate based on 8-digit hydrologic unit codes (Figure 11; see caption for a summary of the salient

features of each HLR). HLRs with the highest typical values of both b and convexity were inconsistent,

being variously semiarid to humid and mountainous to flat with all combinations of (im)permeable

bedrock and soils. HLRs with the lowest typical values of b and convexity were, however, strikingly
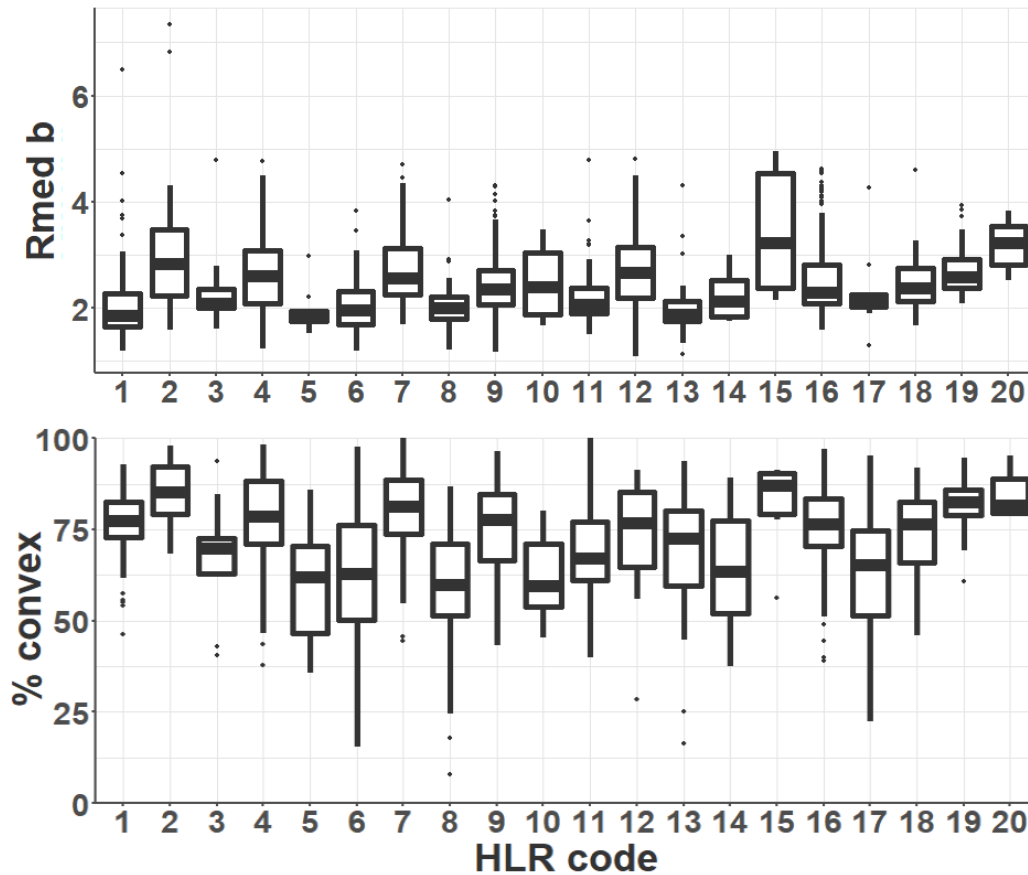
Figure 11: the percent of convex recession events of each watershed organized according to Hydrologic Landscape Regions (HLR): (1) Subhumid plains with permeable soils and bedrock; (2) Humid plains with permeable soils and bedrock; (3) Subhumid plains with impermeable soils and permeable bedrock; (4) Humid plains with permeable soils and bedrock; (5) Arid plains with permeable soils and bedrock; (6) Subhumid plains with impermeable soils and bedrock; (7) Humid plains with permeable soils and impermeable bedrock; (8) Semiarid plains with impermeable soils and bedrock; (9) Humid plateaus with impermeable soils and permeable bedrock; (10)Arid plateaus with impermeable soils and permeable bedrock; (11) Humid plateaus with impermeable soils and bedrock; (12) Semiarid plateaus with permeable soils and impermeable bedrock; (13) Semiarid plateaus with impermeable soils and bedrock; (14) Arid playas with permeable soils and bedrock;(15) Semiarid mountains with impermeable soils and permeable bedrock;(16) Humid mountains with permeable soils and impermeable bedrock;(17) Semiarid mountains with impermeable soils and bedrock;(18) Semiarid mountains with permeable soils and impermeable bedrock;(19) Very humid mountains with permeable soils and impermeable bedrock;(20) Humid mountains with permeable soils and impermeable bedrock

consistent. The five HLRs with the lowest typical b values (1, 5, 6, 8, and 13) as well as lowest likelihood

of convexity (5, 8, 9, 10, and 14) were arid to subhumid plains, plateaus, or playas with generally

impermeable soils.

Section 3.4: Alternative Scenarios

Because of the possibility that the first few days following a substantial rain event obscure

actual streamflow recession, or that five days of Q and –dQ/dt values represents too few data points for

characterization of an individual event, we reanalyzed all data according to A1 (minimum recession

length is 7 days, with the first 2 days removed) and A2 (minimum recession length is 10 days, with the

first 2 days and final single day removed). These data are presented in detail in the supplementary

material (S1). However, the results from each of these two alternative scenarios are not functionally

different from our primary results. According to A1, at none of the 355 watersheds retained for analysis

were concave recession curves more likely than convex recession curves, nor were at least half of

recession events likely to represent Boussinesq outflow. According to A2, only 1 of the 31 watersheds

generated recession events that could be representative of Boussinesq outflow in half or more cases,

and only 3 watersheds generated more concave recession curves than convex recession curves.

Section 4: Discussion

Section 4.1: Analysis of Results

For every reference USGS watershed in the continental US, typical b values of observed

recession events were higher than the b values derived from the entire recession point cloud. In fact,

despite the high variability and range of b values among events, fewer than 5% of observed events

generated b values near or below those generated by Rpc. Even fewer observed events generated b

values near or below those generated using Rk. Therefore, streamflow recession analysis performed

using point cloud data universally underestimates the slope of observed streamflow recession events.

These findings confirm results from smaller-scale studies [e.g., Biswal and Marani 2010; Shaw and Riha

2012; Dralle, Karst, and Thompson 2015]. On a practical level, these results imply that estimates of

streamflow recession using traditional point cloud methods systematically under-predict streamflow

decay in the early period soon after the cessation of rainfall and systematically over-predict streamflow

decay during extended dry periods.

These results, however, are not necessarily in violation of the predictions of outflow from a

Boussinesq aquifer that underlie traditional recession analysis. A median b value for all watersheds of

2.311 is well within the range of expected values if our method of event selection captures the period

primarily during the transition from early- to late-stage recession. However, while outflow from a

Boussinesq aquifer generates a concave recession curve during this transition, our study shows that

observed recession is overwhelmingly convex. Previous studies have noted that convexity of the

recession curve indicates recession behavior in violation of the Boussinesq assumptions, due to the

existence of multiple reservoirs [McMillan et al. 2011], ongoing groundwater recharge [Li et al. 2017],

transition from active drainage network contraction to q decay as the primary mechanism of decreasing

-dQ/dt [Ghoshet al. 2016], deep subsurface groundwater contributions [Wang and Cai 2010], or

substantial unsaturated zone contribution to streamflow [Tashie et al. 2019]. Our results indicate that

for the overwhelming majority of watersheds in the continental US, observed recession curves are not

only more nonlinear than previously indicated, but that their nonlinearity also tends to increase with

length of recession. Analysis of results from alternative recession selection criteria confirms that this

holds true even when assessing only the longest periods of recession in the historical record. Therefore,

predictions of low flows using traditional methods are not only biased towards underpredictions of low

flows, but this bias also tends to increase during extended rainless periods, when accurate estimations

are most necessary.

While most recession events for most watersheds violated the assumptions of traditional

recession analysis, we attempted to identify watersheds where streamflow usually receded according to

the predictions of either a single (non)linear reservoir or a Boussinesq aquifer. In no watershed were observed b values likely to be constant in time, therefore in no watershed was observed recession likely dependent on a single (non)linear reservoir. We classified streamflow recession events as non-Boussinesq where the curvature was convex, or where event-based b values were significantly greater than Rpc b values but not concave. We classified all other types of recession events as possible Boussinesq outflow. In only two watersheds were more than 50% of recession events likely to reflect possible Boussinesq outflow: Croton Creek near Jayton, Texas, (USGS station 08081200) and Briar Creek near Graham, Texas, (USGS station 08088300) at 57.5% and 52.0%, respectively. We also identified every watershed for which more recession curves were concave than convex according to Cmulti, of which there were an additional seven watersheds (illustrated on Figure 9). The scale of these watersheds extended from 10s to 1,000s of square kilometers. We visually assessed hydrographs from each (Figure 12) and investigated their physical properties as summarized by the GAGES-II data set. All nine watersheds were in the western US (seven in Texas, and one each in New Mexico and South Dakota), dry (median precipitation 588 mm per year), flat (median slope 1.2 degrees), and with estimated percolation rates about 70% lower than the national average. Because of low rates of precipitation and high stormflow, each watershed runs dry for at least several months of the year (Figure 12).
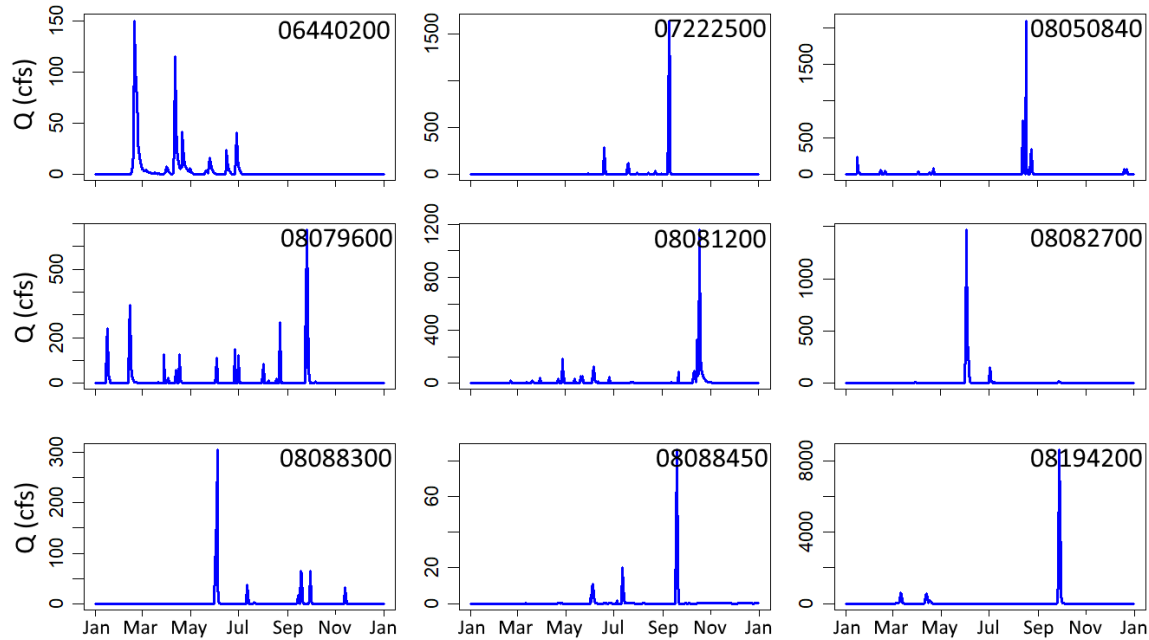
Figure 12: hydrograph of the final full year of each watershed for which more recession curves were concave than convex. In order from top left to bottom right are USGS gages 06440200, 07222500, 08050840, 08079600, 08081200, 08082700, 08088300, 08088450, and 08194200.

The result that observed streamflow recession compatible with Boussineq-type behavior is most likely in dry, flat, impermeable catchments may be seem counterintuitive. However, this is consistent with the known effects of watershed memory [Jachens et al. 2019] and catchment heterogeneity [Harman et al. 2009] in increasing the b values of observed recession. The implications of these results for mechanistic alternatives to hillslope hydraulics as the foundation of Boussinesq-type behavior are discussed in detail below. Functionally, however, low volumes of PPT, long periods between PPT, high PET, and high impermeability all serve to limit the effects of past water inputs on future periods of recession, while low relief limits the potential heterogeneity of subsurface flow paths and topoclimate effects on hillslope aquifers. These 9 watersheds also highlight the results addressed more tentatively above, that while aridity, low relief, and impermeability each acts to independently constrain b values and convexity only weakly, these factors acting in concert are a powerful constraint on both b and convexity. Similarly, high catchment-scale humidity, relief, or permeability may each independently

generate both high b values high likelihood of convexity. That is, baseflow recession as classically understood exists according to the Anna Karenina principle: all low nonlinearity (or convexity) recession is alike, but all high nonlinearity (or convexity) recession is nonlinear (or convex) in its own way.

Section 4.2: Alternative Models

Our analysis is hardly the first to identify uncertainties or inconsistencies in the application of recession analysis, which has been described as "a technique in disarray" [Stewart 2016]. Indeed, alternative hypotheses to the h profile of a dominant hillslope aquifer being the physical basis for the value of b have become increasingly common. Heterogeneity of hydrologic characteristics within or among subsurface features is often invoked to describe recession curvature. Harman et al. [2009] noted that heterogeneity among hillslopes may be sufficient to explain not only the value of b, but the decay in b values from early to late recession, thus accounting for the concave shape of classical recession curves. Further, additional analysis of their results indicates that between-hillslope heterogeneity is also capable of producing convex recession curves. Gao et al. [2017] described a similar effect via vertical heterogeneity within a hillslope (e.g., a perched aquifer), and Wang and Cai [2010] described how macro-scale leakage to (or recharge from) a deep aquifer may force a recession curve to approach some value of -dQ/dt (or Q) asymptotically, yielding a concave (or convex) curve. Clark et al. [2011] noted that neither a single nonlinear reservoir nor two linear reservoirs were capable of producing both the non-unique discharge relationship and the extremely high b values of individual recession events observed at the small, humid catchment at Mahurangi in New Zealand, and therefore recommended a model composed of parallel nonlinear reservoirs. Alternatively, vigorous debate has ensued from the suggestion of Biswal and Marani [2010] that the contraction and expansion of the active drainage network (distinct from the presumed static L) itself may be sufficient to drive baseflow recession. While their analysis generated concave recession curves, Ghosh et al. [2016] observed that the transition from early recession driven by the contraction of an ephemeral drainage network to late recession governed

27

by groundwater hydraulics was capable of producing extremely convex recession in a small headwater catchment at the Panola research station in Georgia, US.

Apart from the shape of individual recession events, a comprehensive description of recession behavior must account for the spread of individual events across the data point cloud (Figure 3). As described above, the Boussinesq model has repeatedly been shown to generate data point clouds with a lower envelope comprised of all late-stage recession and vertical scatter above this line comprised of concave limbs of early-stage recession. However, our analysis indicates that recession is predominantly convex (not concave), and the apparent lower envelope is not comprised of late-stage recession.

The data point cloud may be reinterpreted according to outflow from parallel aquifers that vary in hydrologic conductivity (or drainable porosity, aquifer depth, or storage) and have been shown to generate recession curves of highly variable shape [Clark et al. 2009; Gao et al. 2017; McMillan et al. 2011]. Discharge from the lower K aquifer would define the low, flat slope of the lower envelope, while discharge from the higher K aquifer would define the high, flat slope of an "upper lid" of sorts, and individual events would exhibit high b when streamflow was transitioning from dominance by one end-member to the other. Shaw and Riha [2012] proposed that the point clouds of several glacially impacted watersheds in New York were composed of a series of relatively steep individual recession events that were shifted up / left (higher -dQ/dt for value of Q) as rates of evapotranspiration increased. In a humid headwater catchment with high transmissivity soils at Coweeta, North Carolina, Tashie et al. [2019] further developed this model, noting seasonal hysteretic loops in the relationship between the recession characteristics of individual events and both rates of evapotranspiration and water table elevation, indicating a seasonal switching between dominant controls on recession behavior. They proposed that b is variable and governed by the relative contribution of shallow unsaturated zone storage and deep saturated zone storage, and that a values are governed by rates of evapotranspiration for part of the year and storage conditions for the remainder of the year.

These alternative models predict seasonal variability in typical Q-to-dQ/dt relationships in temperate climates due to the seasonal variability of rates of evapotranspiration and aquifer storage conditions. In our case, all such seasonal processes are convolved in the metric Rday. A possibility of much potential utility is that the upper envelope is defined by observed event recession and the lower envelope by seasonal shifts in typical behavior. The point cloud, therefore, would be comprised of a series of seasonally stacked individual recessions, similar to the proposition by Shaw and Riha [2012], but with moderate and systematic variability in b. In fact, the typical value and distribution of b for individual events (Rmed) are extremely similar to those of the upper envelope (Ru). The same holds true for b values of the lower envelope (Rl) and typical seasonal values (Rday) (Figure 6). Nonetheless, our results indicate this potential relationship may be illusory. For each watershed, the b value according to Ru (or Rl) is an extremely poor predictor of Rmed (or Rday) (r2 of 0.18 and 0.05, respectively). Future investigations into the relationship between the upper (or lower) envelope and individual events (or seasonal dynamics) may be more fruitful if focused on the physical processes driving recession in individual watersheds.

Section 4.3: Alternative Methods and Study Limitations

To assess if our results were robust or biased during periods of extended drought, we reassessed each watershed according to our primary methods while only retaining the 10 longest individual recession events from each watershed for analysis (A4). Using this method, out of 1,027 watersheds we identified 11 for which a majority of recession events might be illustrative of outflow from a Boussinesq aquifer, and 73 watersheds for which more individual recession curves were concave than convex (S1). That is, the longest streamflow recession events, while still overwhelmingly convex and steep compared with Rpc, have slightly lower b values and lesser convexity than shorter recession events. This difference may indicate that in some watersheds, Boussinesq conditions may be met at the watershed scale only after very long periods of recession. Alternatively, because the uncertainty in

calculating –dQ/dt from a rating curve, it is difficult to estimate –dQ/dt precisely at low values, and therefore steep recession curves may be overwhelmed by error after a relatively short period of recession. Because of the strong correlation between b values and convexity (Table 1), this low signal-to-noise ratio would also account for the decreased convexity of long recession events. In any case, Boussinesq outflow is a possible source for a majority of the longest streamflow recession events in only 1.1% of watersheds.

Visual analysis of individual recession curves (Figure 4) suggests that the convexity of a recession curve is often especially pronounced in the latter stages of recession. Results from alternate scenarios A2 and A3, which eliminated the final day of recession from analysis, generate a higher ratio of concave-to-convex curves than the other methodologies tested here (though the vast majority of curves are still identified as convex in A2 and A3). Similarly, metrics for assessing concavity that deprioritize end-member values ($C_{exponential}$ and $C_{squares}$) are more likely to identify concave recession than the method which exclusively relies on them ($C_{residuals}$). These effects may result from a general increase in the convexity of recession curves with time, or from the occasional input of PPT which is insufficient to increase Q yet substantial enough to limit the magnitude of decrease of –dQ/dt. In any case, the variability among our metrics and methods does indicate substantial uncertainty, especially when evaluating individual events.

Finally, it is important to note that our identification of watershed-scale drivers of b and likelihood of convexity rely on coarse resolution catchment-average summary variables. While we are nonetheless able to identify general patterns in recession behavior among watersheds in similar HLUs, there remains substantial uncertainty in this analysis. For instance, note the outliers in HLU 13 (Figure 11). More focused analysis of individual watersheds may help identify within- or between-hillslope heterogeneities in hydraulic properties (e.g., bedrock fracture networks), flow paths (e.g., soil structure),

or storage (e.g., differential ET due to aspect or vegetation) which are key to deriving a fully mechanistic explanation of streamflow recession.

Section 5: Conclusion

Through evaluation of over 150,000 recession events from over 1,000 watersheds across the continental US, we have illustrated that observed streamflow recession consistently violates predictions underlying traditional recession analysis. Specifically:

1)   for all watersheds, observed recession during individual events is substantially more nonlinear than recession determined using the data point cloud as a whole;

2)   for over 99% of watersheds, the nonlinearity of observed recession events tends to increase with time, generating a "convex" recession curve;

3)   and while the average nonlinearity (or convexity) of observed recession events is poorly predicted by individual physio-climatic metrics, low nonlinearity (or convexity) values are common only in dry, flat watersheds with low permeability substrate.

The direct practical implications of these results are that predictions of low flows using the data point cloud are biased towards underpredictions, and that this bias increases during extended rainless periods. Baseflow separation methods and modeled baseflow behavior based on recession analysis or the application of a single (non)linear reservoir are likely to be similarly biased. These predictions of enhanced low flows may serve as an essential buffer to streamflow during uncertain future climate conditions.

Traditional models of recession behavior are evidently most likely to be accurate in "simple" catchments which are flat and dry with low permeability. Observed recession in watersheds that are either high-relief, humid, or highly permeable tend to be increasingly nonlinear and convex, indicating inherently non-unique watershed-scale storage-discharge relationships. An accurate description of recession behavior in these watersheds must account for the time-varying distribution of water. Our

analysis also indicates the potential effects of watershed memory on recession due to variable storage among hillslopes, confirming results from recent studies highlighting seasonal variability in recession behavior.

CHAPTER 2: SPATIAL AND TEMPORAL PATTERNS IN BASEFLOW RECESSION IN THE CONTINENTAL US[2]

Section 1: Introduction

Section 1.1: Summary

Baseflow is often treated according to a unique storage-discharge relationship. However, recent innovations in baseflow recession analysis have allowed novel findings regarding the variability of both the stability of baseflow and its nonlinearity (i.e., the concavity of the hydrograph), as well as the regional clustering of these characteristics. We investigate spatial and temporal patterns in the character of baseflow recession for over 1,000 watersheds in the continental United States (US). We discover seasonal patterns in both the stability and nonlinearity of baseflow which vary systematically across large regions. Further, we relate these baseflow characteristics to their potential physical drivers, including estimates of evapotranspiration, watershed storage, the distribution of watershed storage, and precipitation. While coincident watershed storage is the best predictor of baseflow stability in many regions (particularly the Appalachian Mountains), evapotranspiration from two to three months previous is the best predictor of baseflow stability in other regions (particularly the Pacific Northwest). We also discuss the novel finding that baseflow nonlinearity has increased significantly in most watersheds across the US since 1980.

Section 1.2: Background

Following periods of precipitation, streamflow in headwater catchments is sustained by water exfiltrating from the subsurface. It tends to recede monotonically as storage is depleted [Hewlett and

---

[2] This chapter previously appeared as an article in *Water Resources Research*. The original citation is as follows: Tashie, A., Pavelsky, T., & Emanuel, R. E. (2020). Spatial and Temporal Patterns in Baseflow Recession in the Continental United States. Water Resources Research, 56(3), e2019WR026425.

Hibbert 1963]. This exfiltrating water, or "baseflow", supplies a majority of streamflow in many

watersheds [Santhi et al 2008] and is essential for sustaining habitat [Boulton 2003] and downstream

water resources [Hurd et al. 1999]. It is often described by hydrologists and engineers according to a

unique, nonvarying formulation derived from historical streamflow data [Hall 1968, Nippgen et al.

2016].

Traditionally, recession analysis has been performed by calculating a unique value of the

recession characteristics *a* and *b* by regression on a data point cloud of Q against -dQ/dt in log-log space.

Recession characteristics are then used to estimate basin-scale hydraulic properties [Bartlett and

Proporato 2018, Mendoza et al. 2003, Parlange et al. 2002, Pauritsch et al. 2015, Szilagyi et al. 1998,

Troch et al. 2013] to calibrate hydrologic models [Jepsen et al. 2016, Vaud et al. 2014], to estimate

active stream network length [Wei et al. 2017], to calculate basin-scale evapotranspiration (ET)

[Palmroth et al. 2010, Szilagyi et al. 2007], to predict low flows and their recurrence rate [Charron and

Ouarda 2015], to characterize streamflow sensitivity to catchment storage [Berghuijs et al. 2016], or to

explain geomorphological properties [Zecharias and Brutsaert 1988].
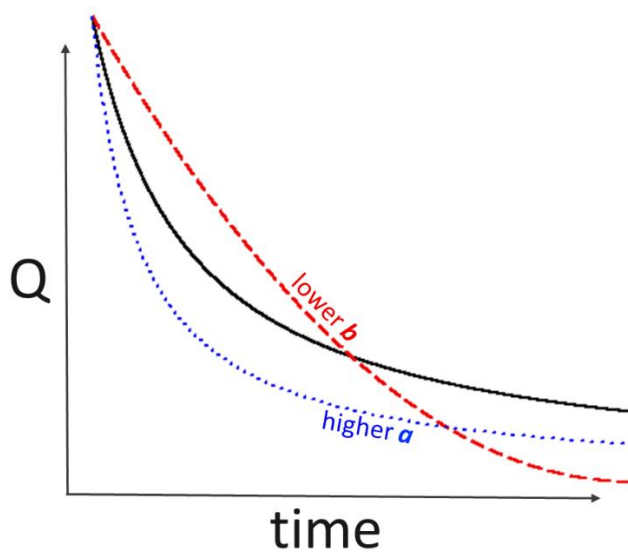
Figure 1: a representative hydrograph (solid black line) shown in relation to a hydrograph with lower nonlinearity (i.e., lower b value shown with a red dashed red line) and decreased streamflow stability (i.e., higher a value shown with a dotted blue line). Adapted from Tashie et al. [2019].

However, it is becoming increasingly common to interpret the data point-cloud as a mathematical artefact [Jachens et al. 2019, Sanchez-Murillo et al. 2015] and to acknowledge that point-cloud based regression methods systematically underestimate the nonlinearity of observed recession events [Santos et al. 2019, Tashie et al. 2020]. Instead, many researchers have begun to assess watersheds according to the typical values of recession parameters calculated using individual recession events [e.g., Dralle et al. 2017, Shaw and Riha 2012]. Promisingly, the typical nonlinearity of individual events strongly predicts the nonlinearity of data point clouds (as well as multiple baseflow indices), indicating that regional-scale physical mechanisms likely underlie recession nonlinearity [Tashie et al. 2020]. However, these physical mechanisms have proven difficult to identify [Patnaik et al. 2018].

Attempts to estimate representative basin-scale recession parameters have also been complicated by results showing that a as estimated from individual events is highly variable [Shaw and Riha 2012]. To accommodate this variability, researchers have developed a new paradigm in which a is not a static metric that integrates the hydraulic properties of a representative Boussinesq hillslope aquifer [i.e., Brutsaert and Nieber 1977], but instead is a variable indicator of watershed state and flux conditions. For example, when a and b are effectively decorrelated, an increase in the relative value of a between events indicates a relative decay in the stability of streamflow, though it is still debated whether this decay is due to direct competition between transpiration and streamflow for water resources, a decline in the water table, a contraction of the stream network, or some seasonally transient combination of factors [Biswal and Kumar 2014, Biswal and Marani 2014, Li et al. 2017, Patnaik et al. 2015, Sanchez-Murillo et al. 2015, Tashie et al. 2019, Wei et al. 2017]. The nonlinearity of recession also varies among individual events [e.g., Jachens et al. 2019, Mutzner et al. 2013, Sanchez-Murillo et al. 2015], often according to a distinct seasonal pattern [Karlsen et al. 2019, Tashie et al. 2019] governed by mechanisms that are not yet agreed upon.

Previous studies have investigated the effects of geomorphologic, soil hydraulic, and climatologic properties on basin-average recession characteristic values at the regional scale [Mutzner et al. 2013, Patnaik et al. 2015, Patnaik et al. 2018, Sanchez-Murillo 2015, Tashie et al. 2020, Ye et al. 2014] and of temporally variable drivers like ET, precipitation (PPT), soil moisture, or groundwater at the local scale [Karlsen et al. 2019, Shaw and Riha 2012, Tashie et al. 2019a]. However, potential regional patterns and trends in the seasonal dynamics of recession characteristics (and their physical drivers) remain poorly understood.

We investigate the physical processes driving streamflow recession at the continental scale according to the following paradigm: the stability of streamflow ($\log(a)$) depends on the volume of deep mobile groundwater as well as the effective hydraulic properties of the reservoirs transmitting this flow, while recession nonlinearity (b) is a function of the relative contribution to streamflow by subsurface features of various hydraulic properties, as proposed by Harman et al. [2009]. Therefore, changes in streamflow stability should integrate basin-scale dynamics in groundwater, while the variability of nonlinearity among events should reflect the relative distribution of mobile water, and static or strongly characteristic values of nonlinearity may be interpreted as general measures of catchment complexity [Dralle et al. 2017]. Specifically, we propose the following hypotheses:

1)   variance in the stability of streamflow ($\log(a)$) during recession is a function of regional storage dynamics, especially seasonal storage dynamics in temperate climates;

2)   ET primarily affects streamflow stability ($\log(a)$) indirectly, via the consumption of potential recharge;

3)   dynamics in the values of nonlinearity (b) are driven by the distribution of water within a watershed between fast and slow responding features;

4)   and long-term climate trends should express themselves in interdecadal shifts in the nonlinearity (b) and stability ($\log(a)$) of streamflow.

We assess these hypotheses through statistical analysis of 175,034 recession events identified in long-term streamflow records from 1,093 gauges across the continental United States, alongside estimates of climate forcing variables and antecedent catchment conditions.

Section 2: Methods

Section 2.1: Recession Analysis

We performed recession analysis on all gaged streams in the continental United States (US) for which daily data is available for at least 20 years and that are identified as "reference gages" according to the Geospatial Attributes of Gages for Evaluating Streamflow version II (GAGES-II) dataset [Falcone et al. 2011]. Reference gages are identified as the least-disturbed watersheds within each of the 12 major ecoregions of the US and are generally free of obstructions and major dams [Falcone et al. 2011]. Though there is no general consensus on methods for defining and selecting recession events, robust analyses of sensitivity and uncertainty for event-based recession selection techniques for 16 watersheds in the Pacific Northwest, US, [Dralle et al. 2017] and 40 watersheds in Switzerland [Santos et al. 2019] have argued for a generally unrestrictive selection procedure, with the exception that both streamflow and the absolute value of its derivative must decrease monotonically (i.e., a concave hydrograph). We have broadly followed these recommendations.

Specifically, we define a recession event as a period of at least five consecutive days when both daily average streamflow (Q) and the absolute value of its time derivative (dQ/dt) decrease monotonically. We removed only the first day of each recession event to account for potential overland flow and placed no minimal or maximal flow requirement on the antecedent flow peak. We also rescaled all discharge measurements according to watershed area. All subsequent references to Q (or −dQ/dt) in this text are made with respect to specific discharge, not volumetric discharge. It is important to note that there remain several sources of uncertainty in the estimation of the absolute value of recession parameters. These include uncertainty in measurements (e.g., stage-discharge relationships)

or undetected sources of error (e.g., small amounts of rainfall) [Troch et al. 2013]. While these

uncertainties can be reduced through use of (sub)hourly discharge data and a variable time step in the

calculation of Q and –dQ/dt [Roques et al. 2017, Rupp and Selker 2006], the scale of our analysis

demands that we rely on readily available daily data from the USGS. Thus, the *absolute* values of the

recession parameters calculated here are uncertain. However, uncertainty analysis of event-based

recession has shown that the *relative* values of recession parameters are robust between events and

among watersheds [Dralle et al. 2017, Tashie et al. 2020]. Therefore, all subsequent analysis relies on

the *relative* values of recession parameters.

The nonlinearity (*b*) of each recession event was extracted according to the slope of a linear

model fitted to Q and -dQ/dt in log-log space. Though we rescaled streamflow by basin area, the

covariance of log(*a*) with *b* makes it impossible to directly compare values of log(*a*) among events or

across watersheds if *b* is allowed to simultaneously vary. While Dralle et al. [2015] did develop a method

for maximally decorrelating log(*a*) from *b* while allowing each to vary simultaneously, seasonal

covariance in the typical values of log(*a*) and *b* may introduce additional bias [Tashie et al. 2019].

Therefore, after defining a unique value of *b* for each event, we reran our analysis using the median

value of *b* for each watershed to estimate the value of log(*a*) for each event, according to the methods

of Shaw and Riha [2012]. To ensure our results were robust, we also needed to capture individual events

across a wide variety of catchment conditions. After qualitative assessment of various methods, we

settled on a minimum of 25 individual events per watershed; this threshold was high enough to be

sufficiently representative but low enough to include a majority of reference watersheds. We retained

for analysis 175,034 total events across 1,093 watersheds that range in size from 1.5 to 25,791.0 km^2.

Unfortunately, due to a less dense gage network and less frequent rainfall in the arid West (i.e., the

Rocky Mountains, the Great Plains, and the Southwest), we were unable to capture sufficient recession

events per watershed across much of this region (Figure 2). To allow comparison of our results with

historical analyses, we also calculated the value of b according to regression to the data point cloud

comprised of all days of recession, a well-established recession analysis method [Vogel and Kroll 1992].
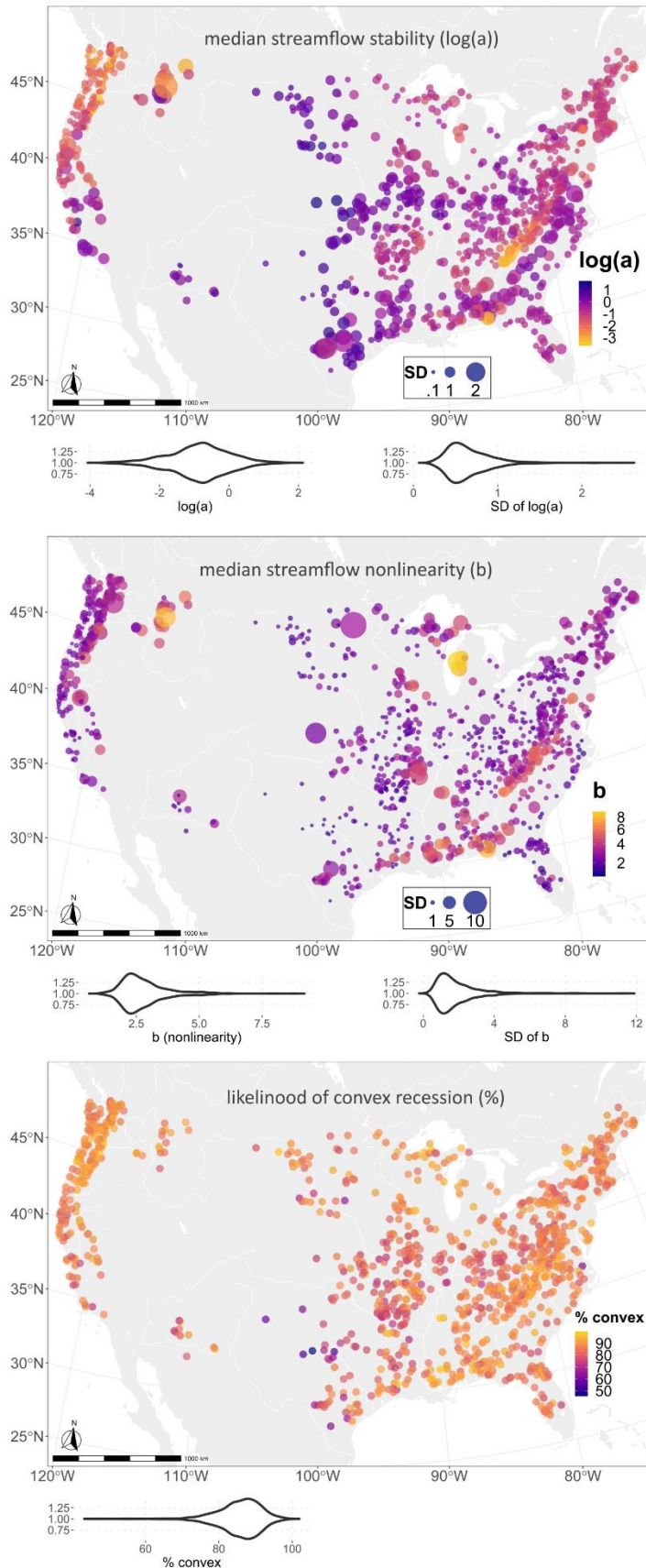
Figure 2: distribution of static recession variables and their variance. Top: typical streamflow instability as approximated with the median value of log(a) of all events, with brighter colors indicating greater stability (i.e., lower values of log(a)) and larger points indicating greater variance. Middle: typical recession nonlinearity as approximated with the median value of b of all events, with brighter colors indicating greater nonlinearity (i.e., higher values of b) and larger points indicating greater variance. Bottom: percent of individual recession events which were convex in log-log recession plots, with brighter colors indicating a greater likelihood of convexity. Violin plots describe the distribution of values of all watersheds (bottom left of each subplot) and the standard deviation of those values across all watersheds (bottom right of each subplot). The location of the New Madrid Seismic Zone is indicated by a star in the top plot.

Section 2.2: Concavity (or Convexity) of Recession

Though the nonlinearity of recession is generally treated as constant over the course of a single event, the nonlinearity of hillslope-scale recession is actually expected to decrease with transition from early-period recession to late-period recession [Brutsaert and Nieber 1977], and the nonlinearity of recession in actual watersheds has often been identified as tending to increase [e.g., Clark et al. 2009, Rupp and Selker 2006, Tashie et al. 2020, Wang 2011]. Adapting methods form Tashie et al. [2020], we estimated the concavity (decreasing nonlinearity) or convexity (increasing nonlinearity) of each recession event according to the consensus of four metrics: 1) the slope of the first half of each recession versus the second half; 2) the center of a fitted circular model relative to the line of best fit of a linear model; 3) the sum of the first and final residuals of a linear model; and 4) the fit of an exponential model relative to the fit of a logarithmic model. We identified each recession event as either concave or convex when at least three of the four metrics agreed, and as indeterminate when there was no consensus among the metrics. To assess potential static controls on average recession characteristics, we relied on summary statistics of catchment-scale geomorphology, soil hydraulics, river morphology, and climate included in the GAGES-II dataset.
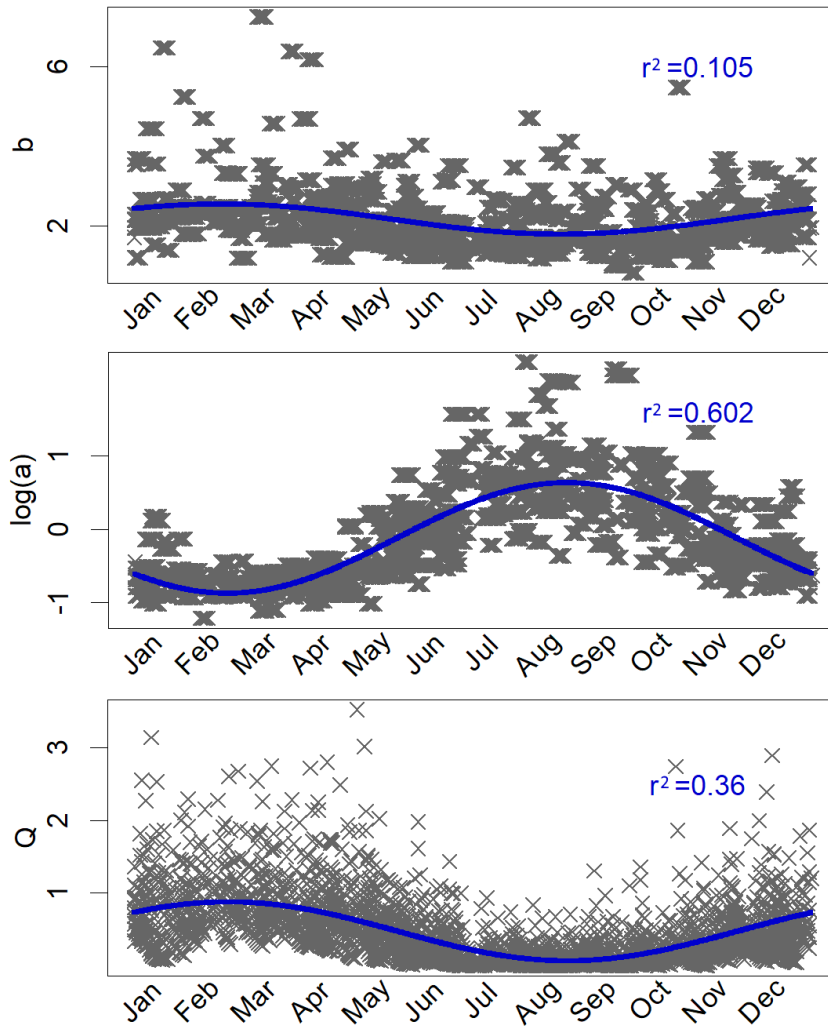
Figure 3: example of sinusoidal models fitted to b, log(a) and Q_recess for USGS gage 14141500 at Little Sandy River near Bull Run, Oregon.

Section 2.3: Seasonality

We investigated the seasonality of each recession characteristic for each watershed using the phase shift and amplitude of a fitted sinusoidal wave (Figure 3). We assessed the significance of each relationship at the three-sigma level and removed from further analysis all insignificant relationships. To validate investigation of the seasonality of log(a) as an indicator of streamflow stability independent of simple flow volumes, we also investigated the phase shift and amplitude of a fitted sinusoidal wave to Q during periods of recession (Q_{recess}) for comparison.

Section 2.4: Regional Stores and Fluxes

To assess regional changes in mobile water, we relied on Gravity Recovery and Climatology Experiment (GRACE) estimates of monthly anomalies downscaled from 3 to 0.5 degree resolution [Wiese et al. 2018]. For climate factors and the vertical distribution of regional water, we relied on output from Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) which is resolved at 0.5 degrees latitude [Gelaro et al. 2017]. Specifically, we extracted estimates of total monthly PPT, totally monthly ET, and monthly average volumetric water content in the soil layers at depths of 0-10cm, 10-40cm, 40-100cm, and 100-200cm. We identified the most relevant mascon or grid-tile using the least distance from the centroid of each watershed and linearly interpolated the monthly data to generate an average value coincident with recession data. It is important to note that downscaled GRACE mascons and MERRA-2 grid tiles are an order of magnitude larger than the typical watershed used in this study (over 3,000 km^2 compared with a median watershed size of 218 km^2). Therefore, we relate watershed recession characteristics to regional monthly estimates of storage, ET, and PPT.

We also assessed the predictive power of each variable on each recession characteristic at time lags of between 0 and 12 months, retaining for analysis the relationship at the optimal lag (Figure 4). We limited our analysis to the time period when the relevant data was available. Further, we ensured the robustness of our analysis for each forcing variable by eliminating watersheds that failed a three-sigma test for significance governing the relationship between each recession characteristic and the forcing variable in question. GRACE data are available from March 2002 to October 2017, and MERRA-2 data are available from January 1980 to present.
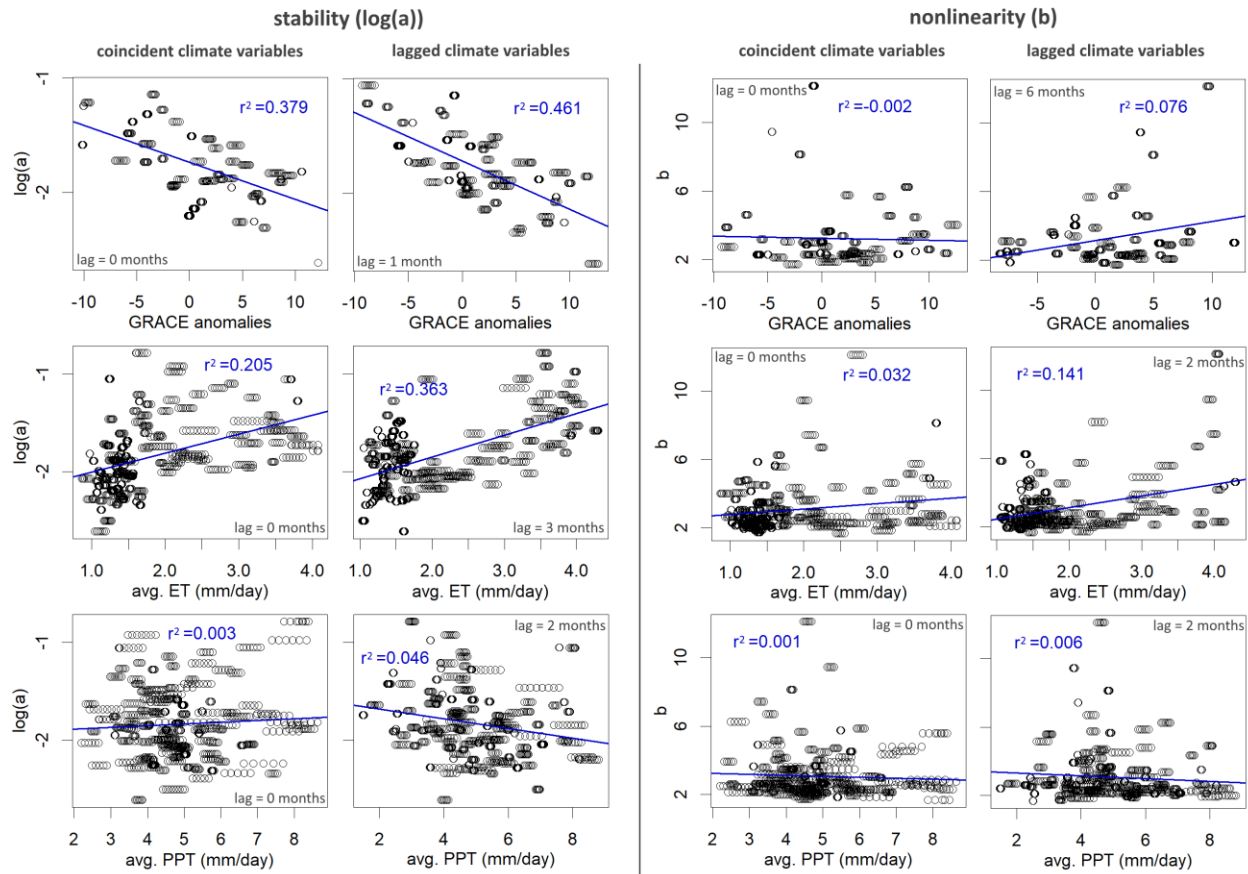
Figure 4: example linear models of log(a) (left columns) and b (right columns) fitted to GRACE anomalies (top row), MERRA-2 ET estimates (middle row), and MERRA-2 PPT estimates (bottom row). Paired plots illustrate the improvement of model performance when comparing recession variables to coincident climate variables (left) and time-lagged climate variables (right). Data is from USGS gage 14141500 at Little Sandy River near Bull Run, Oregon.

Section 2.5: Long Term Trends

We also attempted to identify potential interdecadal trends in the form of streamflow recession by comparing typical values before 1980 with typical values during or after 1980. To control for potential bias resulting from incidental capture of seasonally high (or low) values, we compared the mean values of sinusoidal models fitted to the early and late periods instead of the median value of captured events. To ensure the robustness of our results, we applied a three-sigma test to each of the sinusoidal models and a two-tailed Kolmogorov-Smirnov test to the samples of early- and late-period recession events. We

removed from analysis all watersheds for which any of these relationships were insignificant. Results

were then normalized according to the available period of record and reported as a change per decade.

Section 3: Results

Section 3.1: Typical Values of the Recession Coefficient

Nonlinearity (*b*) calculated according to the median value of individual events was greater than

nonlinearity calculated according to the point cloud for all watersheds, and recession was more likely to

be convex than concave in over 99% of watersheds (Table 1), as reported previously [see Tashie et al.

2020]. Spatial patterns of *b* (and likelihood of convexity) were elevated in the southern Appalachians,

along the Gulf Coast, to the west of the New Madrid Seismic Zone, along the east coast of Lake

Michigan, in the northern Rocky Mountains, and in the Pacific Northwest (Figure 2; see Tashie et al.

2020 for a full discussion). Interestingly, the within-watershed standard deviation of event *b* values

(median of 1.49) was greater than standard deviation of median *b* values among watersheds (0.94)

(Table 1). Also, the regions with the highest variance of *b* values tended to have the highest typical *b*

values (r^2 of linear model = 0.50; Figure 2).

Table 1: summary statistics of variables across all watersheds.

	1st quartile	Median	Mean	3rd quartile	SD
b values (point cloud)	1.09	1.25	1.30	1.47	0.29
b values (median of events)	2.18	2.55	2.78	3.13	0.94
b SD (median of events)	1.06	1.49	1.82	2.27	1.15
b amplitude (seasonal)	0.29	0.52	0.75	0.92	0.78
b convexity (fractional likelihood)	0.83	0.87	0.86	0.90	0.06
log(a) values (median of events)	-1.33	-0.79	-0.85	-0.30	0.86
log(a) SD (median of events)	0.47	0.59	0.65	0.78	0.26
log(a) amplitude (seasonal)	0.43	0.57	0.61	0.73	0.28

While the scale dependence of the scalar *a* (streamflow stability) precludes a quantitative

analysis of differences in *a* among watersheds, the spatial distribution of median event log(*a*) values

yields interesting patterns. Somewhat axiomatically, log(*a*) tends to covary with *b*, such that high

nonlinearity yields high stability, with streamflow tending to be most stable in the southern

Appalachians and in the northern Cascades, and streamflow tending to be the least stable in the Great

Plains. However, the variance of within-watershed event log(a) values was not well associated with

anomalous log(a) values, b values, or b variance (Figure 2). For example, southern Texas exhibited high

variance in log(a) but low variance in b, low typical values of b, and high typical values of log(a), while

the Pacific Northwest exhibited low variance in log(a) but high variance in b, high typical values of b, and

low typical values of log(a).

Section 3.2: Seasonality

Streamflow stability (log(a)) showed a significant seasonal cycle for over 99% of basins

examined. Only eight watersheds did not show significant seasonality of log(a), and all were in the

western plains, primarily in Texas. The stability of streamflow reached a minimum (i.e., highest log(a)) by

August across most of the continent (Figure 5). Beyond this general observation, four specific patterns

regarding the timing of streamflow minima were evident at this scale. First, along the west coast, the

date of minimum streamflow stability tended to be *earlier* in the year the further south one goes. The

median date of the minimum stability was 13 days earlier in the Pacific Northwest than in northern

California, representing a change of about 3 days per 100 km north to south. Second, along the east

coast, the date of minimum stability tended to be *later* in the year the farther south one goes. The

median date of minimum stability was over 88 days later in Florida than it was in New England,

representing a change of about 5.5 days per 100 km north to south. Third, stability tended to minimize

in the southern Appalachians and the New Madrid Seismic Zone about a month or two earlier than in

the surrounding regions. Finally, while broad regional trends tended to be evident along the Gulf Coast,

east coast, and west coast, the seasonality of streamflow stability was highly inconsistent in the Great

Plains.
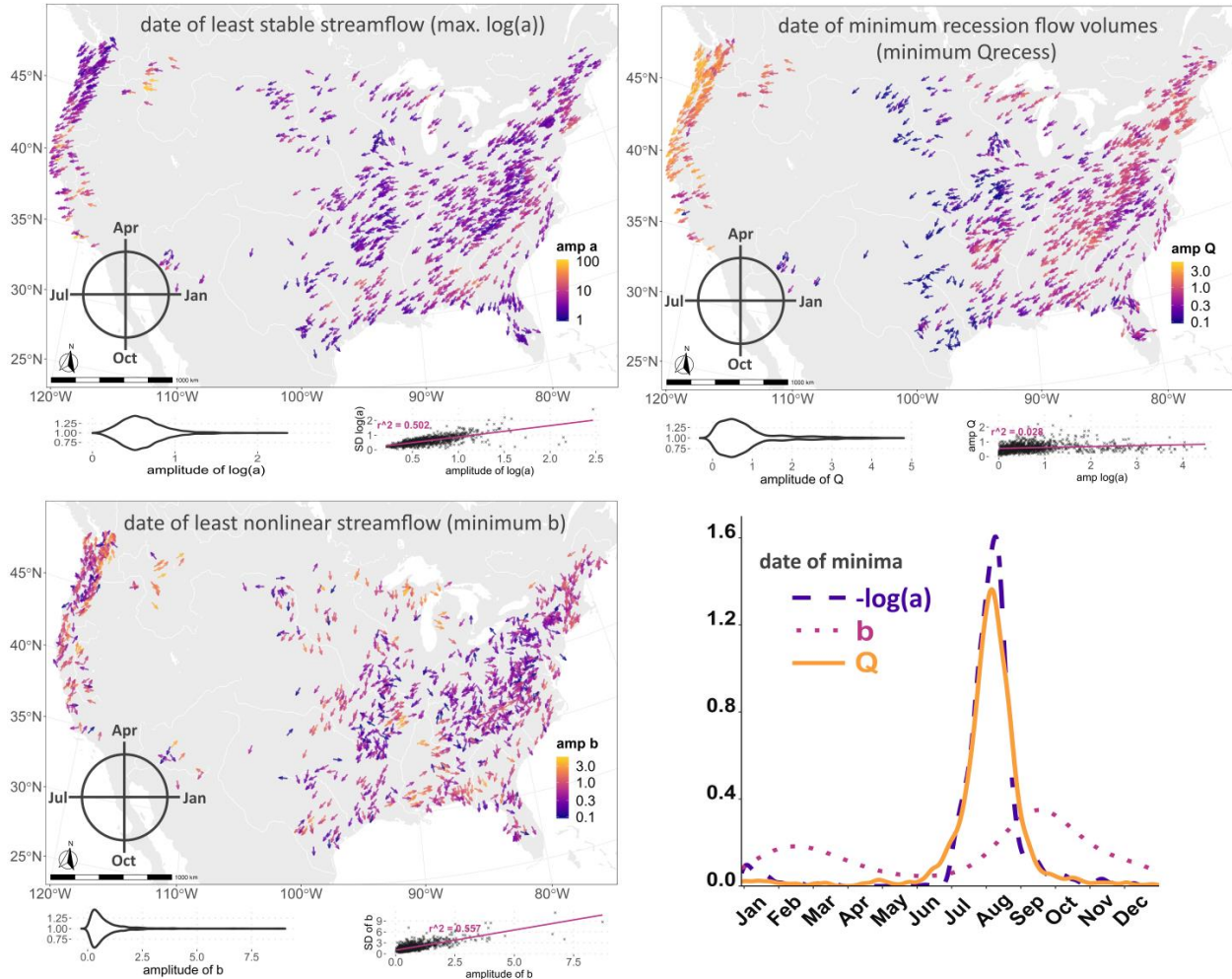
Figure 5: seasonality of streamflow recession. Arrows indicate the seasonal minima of streamflow stability (top left), recession flow volumes in mm/day (top right), and nonlinearity (bottom left). Colors represent the amplitude of the seasonal signal in log space with brighter colors indicating greater amplitudes. Violin plots describe the distribution of amplitudes of all watersheds. Scatterplots illustrate the relationship between the standard deviation of a and the amplitude of the seasonal signal of log(a) (top left), the amplitude of the seasonal signal of log(a) and the amplitude of the seasonal signal of Q_{recess}, (top right), and the standard deviation of b and the amplitude of the seasonal signal of b (bottom left). The density plot on the bottom right describes the date of seasonal minima of streamflow stability (-log(a); blue dashed line), nonlinearity (b; purple dotted line), and recession flow volumes in mm/day (Q_{recess}; orange solid line).

The seasonal amplitude of streamflow stability predicted the standard deviation of streamflow stability quite well ($r^2 = 0.50$). However, though the statistical distribution of amplitude and standard deviation values is nearly identical (Table 1), their spatial patterns are often inconsistent (Figures 2 [top] and 5 [top left]).

Recession flow volume (Q_{recess}) minima tended to coincide with streamflow stability minima, though their distribution exhibited less kurtosis, and the amplitudes of the seasonal signals are strikingly different (Figure 3). Most notably, the west coast exhibits by far the highest amplitude in seasonal recession flows volumes, which diminish as one moves south, while the seasonal amplitude of the streamflow stability signal increases as one moves south.

Regional patterns in the amplitude of recession nonlinearity (b) were more difficult to discern (Figure 5, bottom left) and did not conform to regional patterns in the variance of within-watershed event nonlinearity (Figure 2, bottom left). Nonetheless, the seasonal amplitude of b and the within-watershed standard deviation of b were strongly correlated (r^2 = 0.57).

The timing of minima in seasonal nonlinearity was similarly more erratic within regions than the timing of streamflow stability or recession flow volumes. However, a continental pattern in timing was apparent (Figure 5, bottom plots). For most of the eastern US, recession nonlinearity reached a minimum around mid-September (i.e., about 1-2 months following the average streamflow stability minimum), while along the west coast the timing of recession nonlinearity and streamflow stability were inverted, with minimum nonlinearity (around February, in Figure 5) preceding the streamflow stability minimum (around August, in Figure 5) by as much as 6 months.

Section 3.3: Interdecadal Trends

The nonlinearity of recession has increased substantially in most watersheds across the US since 1980, with an average increase in mean event nonlinearity of about 0.152 per decade (Table 2). While a decrease in nonlinearity was evident in 22% of watersheds, the magnitude of these decreases (mean decrease of 0.140 per decade) was generally much smaller than the magnitude of change in the 78% of watersheds with increasing nonlinearity (mean increase of 0.230 per decade). Watersheds showing a decrease in nonlinearity were seemingly scattered at random, but a general increase in nonlinearity was evident in some regions – especially in the southern Appalachians and along the Pacific coast in northern

California, Oregon, and Washington (Figure 6). The number of watersheds seeing an increase or decrease in stability was relatively evenly distributed (Table 2) though regional patterns were evident. The southeastern US has experienced a strong and nearly universal decrease in both stability and recession flow volumes, while the Great Plains and the Midwest have experienced a general increase in streamflow stability. The Sierra Nevada has seen a general increase in streamflow stability, northern California has seen a general decrease, and the signal is mixed in the Pacific Northwest.

Nearly 75% of watersheds saw a decline in recession flow volumes, though these declines are generally small (Table 2). Only a quarter of watersheds saw an annual decrease of recession flow volumes of greater than 40 mm per decade. Declines were nearly universal in the southeast and the Pacific Northwest. The signal was mixed for much of the rest of the US, though areas of increasing recession flow volumes were evident in parts of the Midwest, the northern Great Plains, and the Sierra Nevada. However, regional patterns were potentially masked because the seasonal signal of recession flow volumes was less likely to be statistically significant (n=369) than that of nonlinearity (n=421) or stability (n=698).
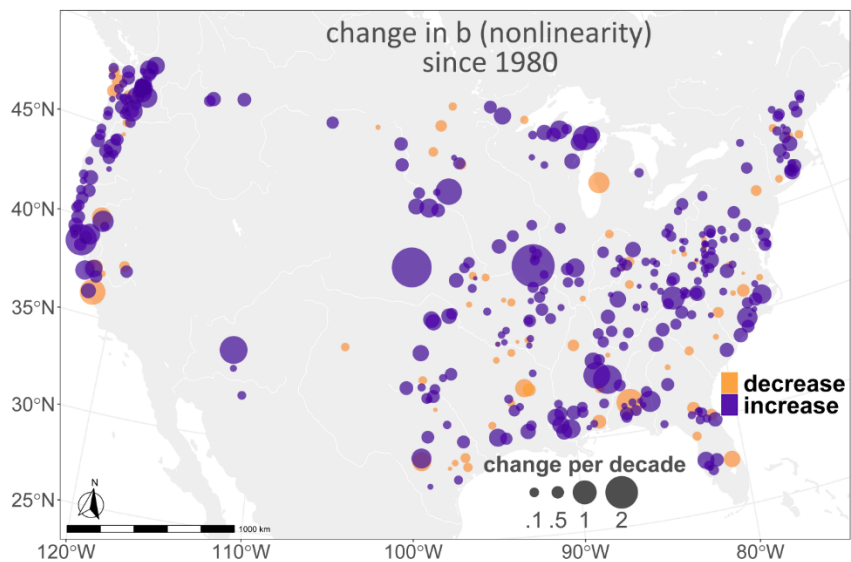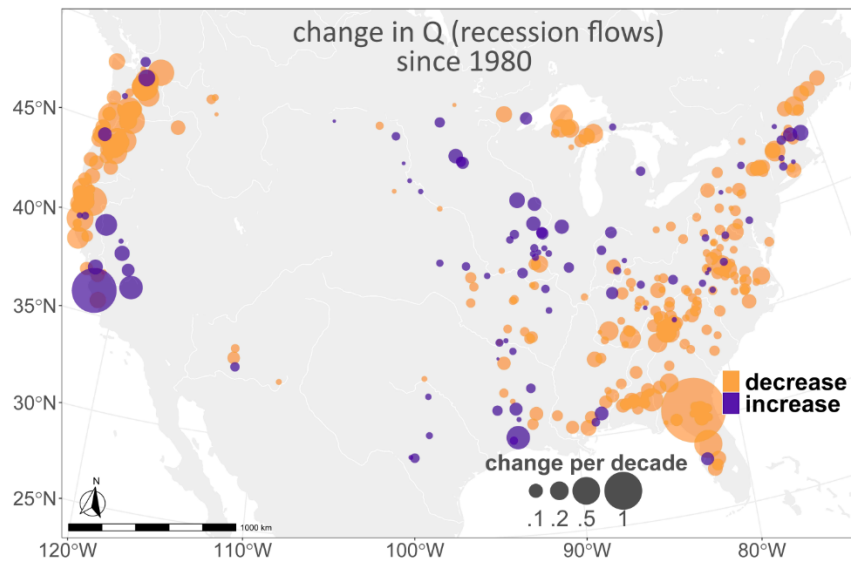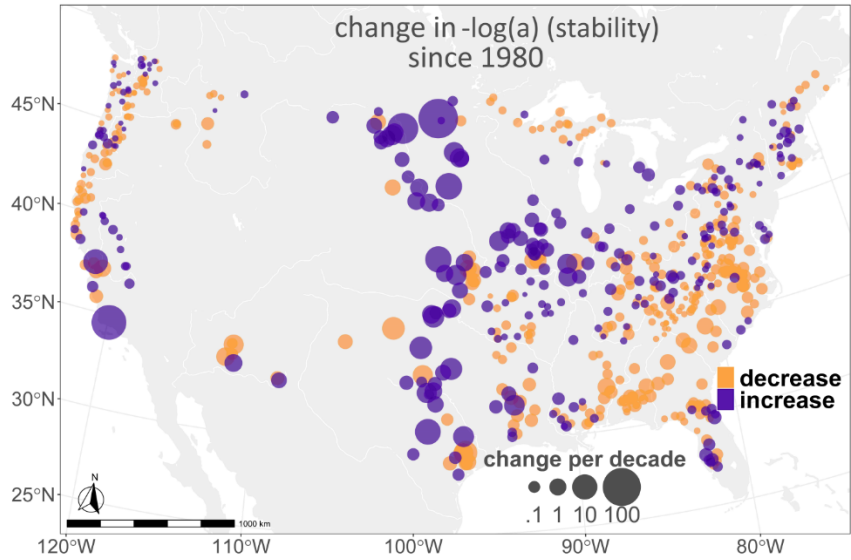
Figure 6: statistically significant interdecadal (pre vs. post 1980) trends in streamflow stability (top left), recession flow volumes (top right), and nonlinearity (bottom left). Purple indicates decreased log(a) (i.e., increased streamflow stability), increased Q_{recess} (i.e., increased recession flow volumes), and decreased b (i.e., decreased nonlinearity), while yellow indicates the reverse trend. The size of each dot represents the strength of the trend.

Metric (change per decade)	1st quartile	Median	Mean	3rd quartile	SD
log(a) (values)	-0.007	0.001	-0.253	0.014	2.929
log(a) (amplitudes)	-0.227	0.030	-0.126	0.298	11.392
Q_{recess} (values)	-0.107	-0.035	-0.066	0.001	0.217
Q_{recess} (amplitudes)	-0.070	-0.012	-0.044	0.021	0.241
b (values)	0.022	0.094	0.152	0.220	0.356
b (amplitudes)	-0.015	0.040	0.109	0.148	0.416

Table 2: interdecadal trends in streamflow stability (-log(a)), recession flow volumes (Q_{recess}), and nonlinearity (b).

Section 3.4: Climate Predictors

GRACE water anomalies predicted streamflow stability reasonably well (i.e., a negative

relationship to log(*a*); Table 3), especially in the Appalachian Mountains, the New Madrid Seismic Zone,

and the Pacific Northwest, where r^2 values often exceeded 0.50 (Figure 7). For most watersheds,

coincident (lag = 0 months) or recent (lag = 1 month) regional estimates of water anomalies were the

best predictors of streamflow stability (Figure 7, bottom right). This was nearly universally true in the

Appalachian Mountains. However, in many watersheds in the Pacific Northwest and to the west of the

Appalachian Mountains, GRACE anomalies lagged by 10-12 months provided the best predictions. The

predictive power of GRACE water anomalies was relatively weak along the Coastal Plains, the Great

Plains, around the Great Lakes, and in the Southwest.

Table 3: r^2 of linear models relating log(a) and b to climate predictors at the optimal time lag.

Climate predictors of stability and nonlinearity	r^2 values				
	1st quartile	Median	Mean	3rd quartile	SD
log(a) to GRACE	0.191	0.347	0.356	0.499	0.208
log(a) to ET	0.200	0.356	0.356	0.491	0.193
log(a) to PPT	0.065	0.130	0.169	0.238	0.146
b to GRACE	0.028	0.066	0.113	0.134	0.151
b to ET	0.032	0.079	0.120	0.152	0.140
b to PPT	0.021	0.046	0.088	0.099	0.131

Problematically, the signal of individual components of storage was not sufficiently disaggregated from that of storage as a whole. MERRA-2 soil moisture data exhibited relationships in the same direction as but of a smaller magnitude than those of GRACE water storage anomalies. Because the distribution of storage in a watershed has been shown to be a major driver of recession characteristics both theoretically [e.g., Harman et al. 2009] and in actual watersheds [e.g., Ghosh et al.
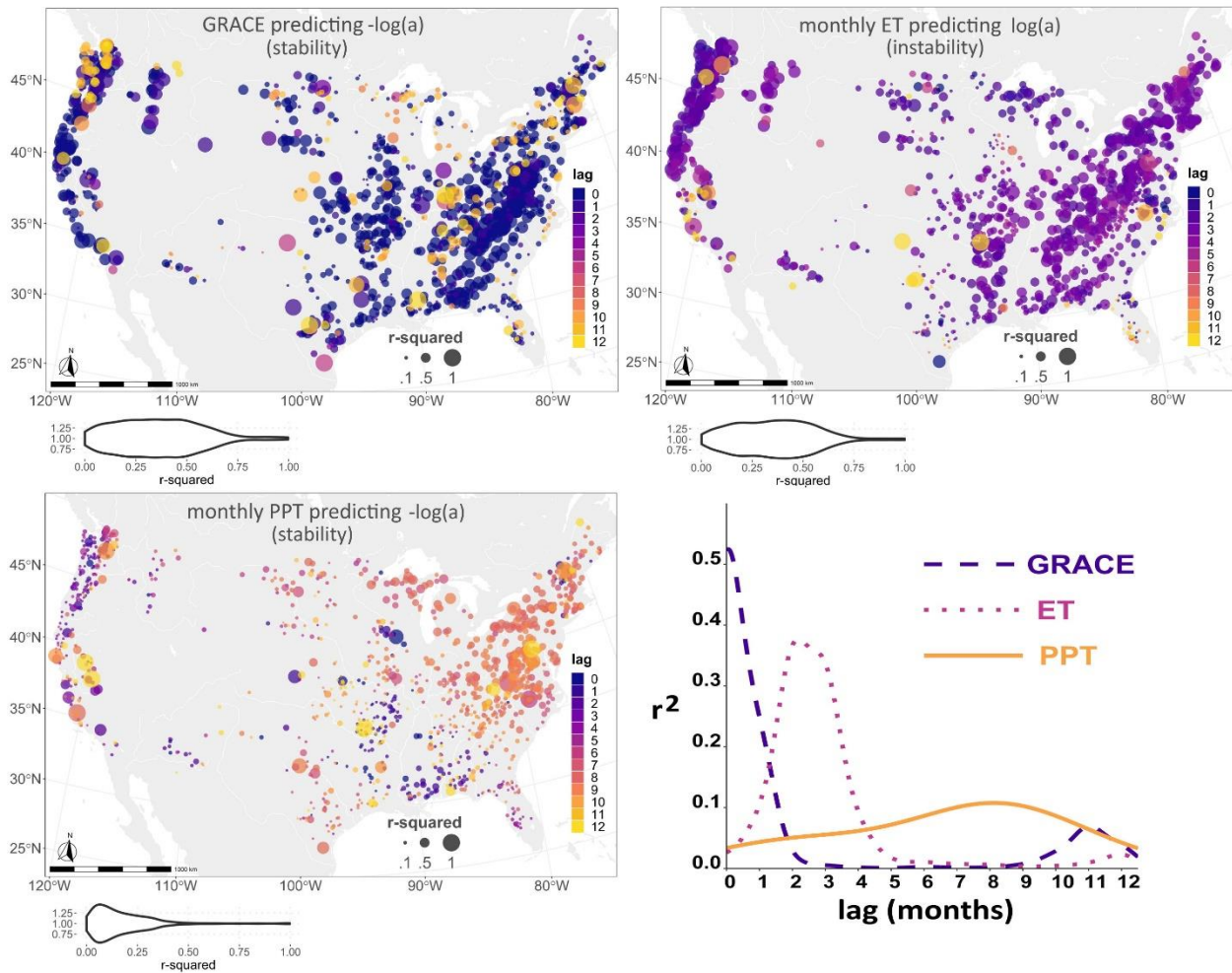
Figure 7: predictive power of streamflow stability (-log(a)) by GRACE storage anomalies (top left), streamflow instability (log(a)) by monthly ET (top right), and streamflow stability (-log(a)) by monthly PPT (bottom left). The strength of that relationship (i.e., the r^2 value of a linear model) is indicated by the size of each dot. Violin plots give the distribution of r^2 values from all watersheds. Colors indicate the time lag (in months) which generated the strongest relationship (highest r^2 value) for each watershed, with brighter colors indicating a longer time lag. The density plot (bottom right) gives the time lag (in months) which best predicts streamflow stability by GRACE storage anomalies (blue dashed line), streamflow instability by monthly ET (pink dotted line), and streamflow stability by monthly PPT (yellow solid line).

2016, Tashie et al. 2019], these poor relationships may be due to uncertainty, poor reliability of MERRA-2 soil moisture data, or low spatio-temporal resolution. Therefore, our primary analysis proceeded using only seasonality, regional storage, ET, and PPT.

Rates of ET were a surprisingly strong predictor of streamflow *in*stability (i.e., a positive relationship to log(a); median r^2 = 0.36) (Table 3). Indeed, in the Northeast and the Pacific Northwest, ET was a better predictor of streamflow instability than GRACE water anomalies. Generally, ET from about 2-3 months previous was the best predictor of streamflow instability (Figure 7, bottom right). However, in the Coastal Plains, the Great Lakes region, and in the Pacific Northwest, ET from 0-2 months previous was the best predictor. Also, ET of any time lag was a poor predictor of streamflow instability in most of the Coastal Plains, the Great Plains, and in southern California.

Monthly PPT was generally the poorest predictor of streamflow stability (i.e., a negative relationship to log(a); median r^2 = 0.13) (Table 3), except in the Mid-Atlantic where r^2 values often exceeded 0.40. PPT of any time lag was an especially poor predictor of streamflow stability in the southeast, the Great Plains, and the Pacific Northwest. Unlike with GRACE water anomalies and rates of ET, the time lag generating the best prediction of streamflow stability was inconsistent among various regions. For instance, recent or coincident monthly PPT was the best predictor of streamflow stability along the central Gulf Coast, while 2-4 months previous PPT was the best predictor in the Pacific Northwest, and much older (7-11 months previous) PPT was the best predictor in the Mid-Atlantic.

GRACE water anomalies, rates of ET, and monthly PPT were generally poor predictors of recession nonlinearity (b) across broad regions at any time lag (median r^2 of 0.07, 0.079, and 0.046) (Figure 8 and Table 3). A few regional patterns are worth noting, though. GRACE water anomalies from 10-12 months previous were the best predictors of nonlinearity in New England, while more recent anomalies were the best predictors for most other watersheds. For ET, coincident rates were the best predictors on the west coast, recent (2-5 months previous) rates were the best predictors in the
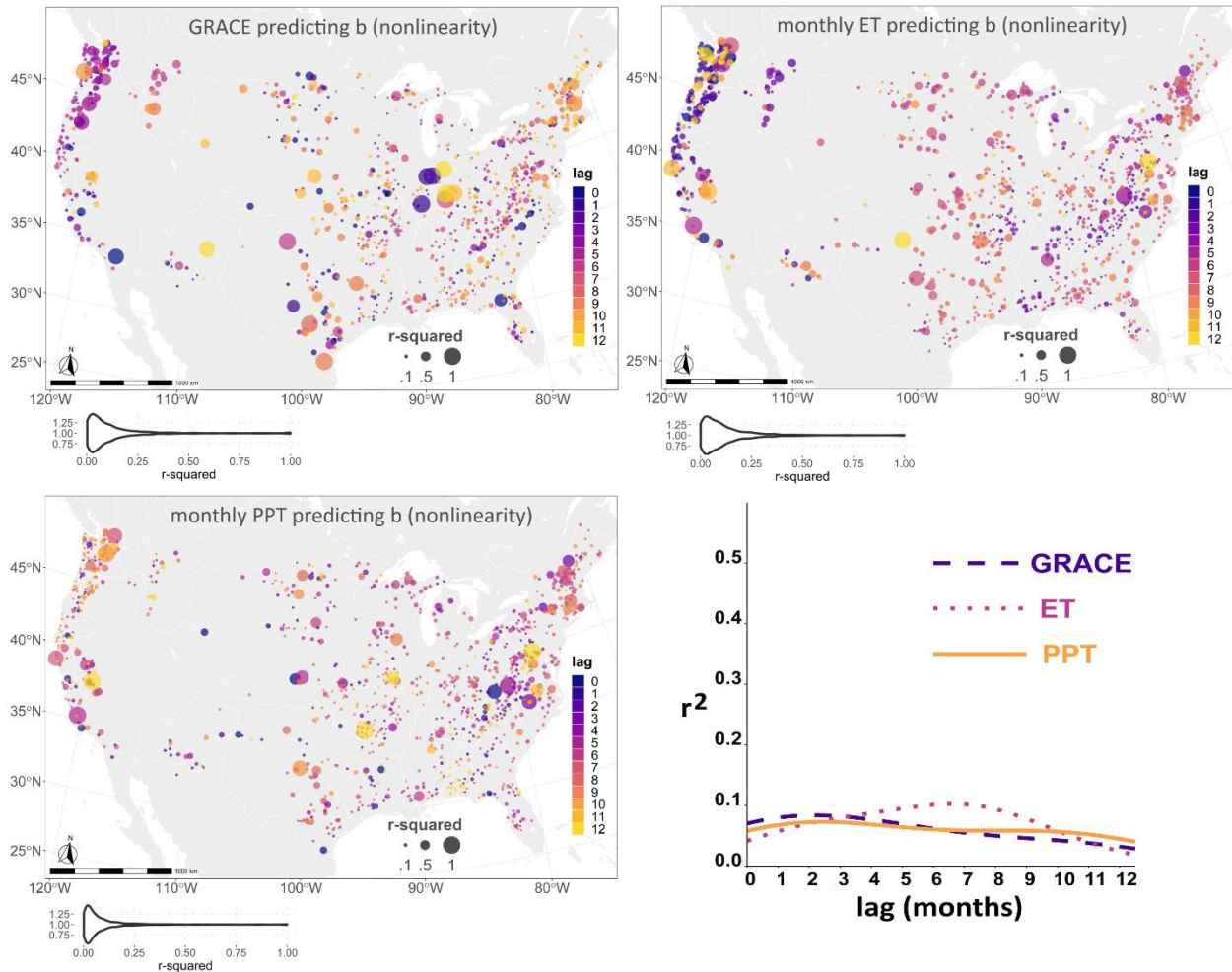
Figure 8: predictive power of increased nonlinearity (b) by GRACE storage anomalies (top left), decreased nonlinearity (-b) by monthly ET (top right), and increased nonlinearity (b) by monthly PPT (bottom left). The strength of that relationship (i.e., the r^2 value of linear model) is indicated by the size of each dot. Violin plots give the distribution of r^2 values from all watersheds. Colors indicate the time lag (in months) which generated the strongest relationship (highest r^2 value) for each watershed, with brighter colors indicating a longer time lag. The density plot (bottom right) gives the time lag (in months) which best predicts nonlinearity by GRACE storage anomalies (blue dashed line), monthly ET (pink dotted line), and monthly PPT (yellow solid line).

Appalachian Mountains and in the New Madrid Seismic Zone, while much older rates (6-10 months previous) were the best predictors for most other watersheds.

Section 4: Discussion

Section 4.1: Temporal Drivers of Recession Characteristics

Streamflow stability (log(a)) is evidently driven by watershed-scale seasonal dynamics, while nonlinearity (b) is not. Specifically, though nonlinearity is highly variable among events (Table 1), this variability is poorly predicted by seasonality, with the median r^2 of the seasonal signal computed according to a sine wave being only 0.06 (Table 2). Streamflow stability, however, tends to be strongly governed by seasonal dynamics, with the median r^2 of a simple annual-scale sinusoidal model being 0.41. Further, the amplitude of the seasonal signal was a strong, unbiased predictor of the standard deviation of streamflow stability (median r^2 = 0.56).

GRACE-based estimates of regional storage anomalies were significantly correlated with streamflow stability for most watersheds, and they explained half (or more) of the total variance of streamflow stability in a quarter of watersheds (Table 3). Because storage dynamics are often strongly seasonal [Alley et al. 2002, Strassberg et al. 2007, Tashie et al. 2016, Wang and Alimohammadi 2012], we assessed the covariance of the predictive power of regional storage with the predictive power of the sinusoidal seasonal model. Surprisingly, there was essentially no correlation (r^2 = 0.01) between the predictive power of a sinusoidal seasonal model and that of GRACE-based regional storage anomalies (Figure 9). That is, regional storage governs streamflow stability independently of seasonal cycling.
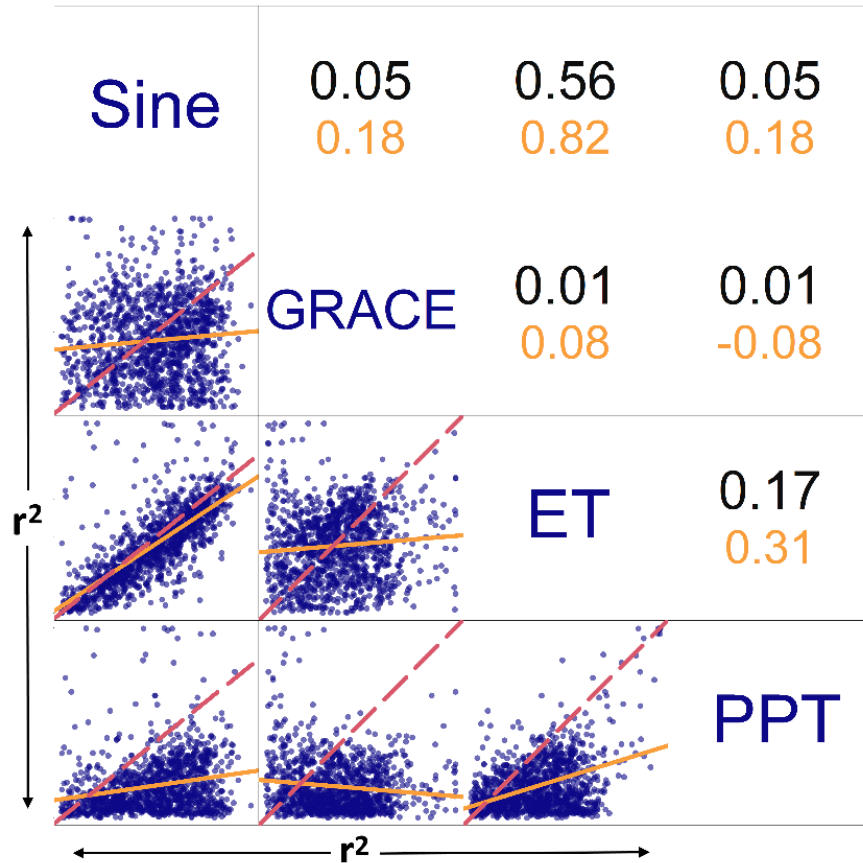
Figure 9: correlation of models predicting the (in)stability of streamflow). Scatterplots give r^2 values of models predicting: streamflow stability by an annual sine wave, stability (-log(a)) by GRACE water anomalies, and instability (log(a)) by monthly ET (on the x-axes); and stability (-log(a)) by GRACE water anomalies, instability (log(a)) by monthly ET, and stability (-log(a)) by monthly PPT (on the y-axes). Fitted linear models are represented by a solid gold line, and one-to-one lines by a dashed pink line. The r^2 values of these models are given in black and the slope of these models are given in gold.

These results expand on smaller-scale studies that have relied on indirect measures of storage. For example, Bart and Tague [2017] showed that cumulative antecedent PPT (as proxy for watershed wetness conditions) was strongly correlated (overall $r^2 > 0.5$) with streamflow stability (*a*) for eight California watersheds, while Patnaik et al. [2015] showed that streamflow stability (*a*) linearly correlated with lagged *Q* (also a proxy for watershed wetness conditions) in log-log space for 358 watersheds across the continental US. By relying on "direct" measures of regional storage anomalies (i.e., GRACE), our study confirms that storage is a primary driver of streamflow stability in the continental US, and it

offers the first evidence that storage governs streamflow stability independently of any confounding seasonal signal [Patnaik et al. 2015].

While the relationships between storage and streamflow stability reported here are strong, it is important to recall the incongruity of scales used in this study: downscaled GRACE mascons are at least an order of magnitude larger than most watersheds. This source of uncertainly provides a likely explanation for the small number of watersheds where the optimal time lag is between 10 to 12 months (Figure 7, bottom right). In these watersheds, long-term trends and seasonality may be the only signals that significantly link regional GRACE anomalies (integrating groundwater, soil moisture, and surface water) to watershed-scale mobile groundwater. We suggest that future studies with higher resolution, disaggregated groundwater data may generate more robust results.

While it has long been known that transpiration is linked to diminished rates of streamflow during the growing season [Dunford and Fletcher 1947, Federer 1973, Wondzell et al. 2007], researchers have generally been unable to show that ET directly affects streamflow recession at the catchment scale [Biswal and Kumar 2014, Shaw and Riha 2012]. Bart and Tague [2017] did identify weak (overall $r^2 < 0.2$) but statistically significant relationships between daily PET and streamflow stability (a) in four out of eight watersheds in California. Meanwhile, Tashie et al. [2019] recently noted a strong but transient ($r^2 >$ 0.5 for two consecutive months) signal of transpiration in streamflow recession during the early growing season in a small, humid, headwater catchment in North Carolina. They inferred that recession may be directly affected by transpiration only when an elevated water table intersects with the root zone of an extensive portion of a watershed thus allowing transpiration to directly compete with streamflow for groundwater resources. Therefore, the effect of ET on recession may be realized primarily in the consumption of infiltrated water which would otherwise recharge groundwater [Gabrielli and McDonnel 2018]. Our results confirm that coincident regional-scale ET is not the first-order driver of streamflow recession for most watersheds in the continental US. Instead, the effect of ET on streamflow recession is

generally delayed by 2-3 months. This lag highlights the indirect influence of vegetation on streamflow

through altering moisture in the upper (i.e., active) soil layers, setting the stage for groundwater

recharge and (therefore) baseflow later in the season [Emanuel et al., 2010].

Still, the delayed effect of ET is a strong predictor of streamflow stability (median r^2 = 0.36) and

is, in fact, a better predictor of streamflow stability than GRACE-based estimates of regional storage for

49% of watersheds. Further, the predictive power of ET on streamflow stability is almost completely

uncorrelated with the predictive power of regional storage on streamflow stability (r^2 of a linear model

is 0.01), though the predictive power of ET on streamflow stability is very strongly correlated with the

predictive power of seasonality on streamflow stability (r^2 of a linear model is 0.56), with little bias

(Figure 9). Taken together, this close correspondence implies that the seasonality of streamflow stability

is due to the delayed effects of the consumption of infiltrated water by ET, while the non-seasonal

variance of streamflow stability is due to variations in coincident storage.

Further, in many watersheds, the coincident effects of ET on streamflow stability are not

negligible (r^2 > 0.14 in 25% of watersheds, concentrated in the Pacific Northwest) and are likely

underestimated in this study due to the coarse spatio-temporal resolution of our data. Contrary to our

results, He et al. [2016] recently identified an extremely strong correlation between streamflow stability

(*a*) and coincident daily evaporation from an eddy flux tower (r^2 = 0.83), temperature (r^2 = 0.82), and

atmospheric vapor pressure (r^2 = 0.71) in a subtropical monsoonal watershed (864 km^2) in China.

Therefore, especially in humid watersheds with high water tables, ET may directly compete with

streamflow for water resources over great spatial extents for significant portions of the year.

Of the four potential physical drivers of streamflow recession investigated in this study, PPT

showed the weakest correlation with streamflow stability and nonlinearity for most watersheds, both

when assessed coincidentally and given an optimal time lag. Recent work has highlighted the strong

effects that time-lagged and cumulative PPT can have on baseflow [Anderson et al. 2008] and

streamflow signatures like runoff ratios [Nippgen et al. 2016], extreme flows [Carlier et al. 2018], and

streamflow stability (*a*) [Bart and Tague 2017]. Notably, these strong correlations were identified using

daily PPT data, whereas the scope of our study demanded we rely on spatially coarse monthly data

which may have masked potential relationships.

Nonlinearity dynamics were poorly predicted by each variable assessed in this study. That is,

nonlinearity was not determined seasonally or by regional monthly-average storage anomalies, PPT, or

rates of ET. Given that nonlinearity is generally nonstationary even during individual events (i.e.,

recession curves are usually convex), this result is not unexpected. Indeed, it gives further evidence that

recession nonlinearity is a function of highly transient watershed features.

Though the variance of nonlinearity among events is greater than the variance of catchment-

average nonlinearity among watersheds, catchment-average values of nonlinearity do show distinct

regional patterns (Figure 2). These regional patterns potentially indicate some physical mechanism(s) for

driving up (or down) the average nonlinearity of individual events. Indeed, for 26 watersheds in the

Pacific Northwest, Sanchez-Murillo et al. [2015] indicated a relationship between nonlinearity and both

climate factors (i.e., aridity index) and geology. Patnaik et al. [2018] also discovered multiple significant

relationships between catchment-average nonlinearity and characteristics of soil, climate, and

geomorphology for 358 basins across the US. Importantly, they also found that the significance of these

relationships varied among regions, such that (for example) soil class was strongly correlated with

nonlinearity in the eastern US but totally uncorrelated in the western US. Similarly, Tashie et al. [2020]

found that various forms of catchment complexity (e.g., climate or the hydraulic conductivity of soils or

bedrock) were each independently sufficient to enhance catchment-average nonlinearity across a range

of different hydrologic landscape regions. Therefore, while there does not yet appear to be a generally

applicable model for predicting catchment-average values of nonlinearity, future studies focused on

physio-climatically similar regions or landscapes may allow a more nuanced analysis of the physical

processes underlying recession nonlinearity.

Section 4.2: Seasonality and Trends

For most of the eastern US, the dates of streamflow stability minima are later in the year the

farther south or closer to the coast one goes. Given the lagged effects of ET in depressing streamflow

stability, this pattern is apparently due to growing seasons being longer in the southern and coastal

portions of the US than elsewhere. Curiously, the reverse is true on the west coast: streamflow stability

minima occur earlier in the year the farther south one goes (Figure 5). A second curious trend is also

apparent in the western US: the farther south one goes, the longer the optimal time lag for maximizing

the relationship between ET and streamflow instability (Figure 7), indicating a greater delay between

infiltration and recharge. Therefore, in the west, the north-south shift in the timing of streamflow

stability minima is not a function of the length of the growing season (as is likely in the east) but likely a

function of water table depth, as well as monsoonal patterns of precipitation and snowmelt.

Most watersheds showed statistically significant interdecadal trends in recession characteristics,

with a slight tendency towards decreased stability and average recession flow volumes (Figure 6).

Streamflow stability was a more sensitive indicator than recession flows, generating statistically

significant relationships in about 25% more watersheds. Our results confirm and extend results from the

west coast of the US from 1988 to 2009 [Sawaske and Freyburg 2014] that showed increased minimum

recession flows in southern California and decreased minimum recession flows in the Pacific Northwest,

confirm at a finer scale a trend in increased low flows in the Great Plains [Gudmundsson et al. 2019],

and discover a strong trend in decreased streamflow stability in the southeastern US.

Intriguingly, long term trends in the nonlinearity of streamflow are consistently positive across

the continent, with streamflow nonlinearity being four times more likely to increase after 1980 than to

decrease (Figure 6). Therefore, watersheds may be growing increasingly dependent on a combination of

shallow stormflow and deep regional groundwater flow, with decreasing contributions to streamflow from groundwater near the root zone. While the evidence here is too broad and coarse to draw such a fine-detailed, process-based conclusion, these results do align with the scientific consensus that climate change has generally resulted in increased demands on soil moisture in the root zone due to increased PET [Rind et al. 1990] and decreased episodic recharge through an increase in the intensity of rainfall during storms [Tashie et al. 2016]. Regardless of the underlying processes, these trends may have implications for water management. Shifts in the sources supplying streamflow may affect water quality, and flow extremes may become more sensitive to climate variability, groundwater pumping, landcover alterations, and other environmental impacts.

The strong, regionally consistent interdecadal trends noted here illustrate the sensitivity, and potential utility, of event-based recession analysis in investigating the effects of climate change on streamflow, a novel application. We suggest that by integrating all recession flows according to known seasonal patterns, it is possible to identify trends that are more robust and less sensitive to stochastic interannual variability than when using traditional indices like annual flow minima [e.g., Gudmundsson et al. 2019]. However, while our analysis does extend across multiple ENSO cycles, the potential effects of ENSO and PDO anomalies on streamflow forcing variables (e.g., PPT and runoff) are well documented [e.g., Rice and Emanuel 2017]. Therefore, the trends in recession characteristics documented here should be treated cautiously and deserve further analysis.

Section 5: Conclusions

Through systematic analysis of streamflow records alongside estimates of climate forcing variables and antecedent catchment conditions for over 1,000 watersheds across the continental US, we describe previously unidentified patterns in the character of streamflow recession. Specifically, we find that:

1) the stability of streamflow is highly seasonal, though, contrary to our hypothesis, this seasonality is largely a function of time-lagged ET, while the control of regional storage on streamflow stability is generally independent of seasonality;

2) ET primarily affects streamflow stability indirectly via consumption of potential recharge, though ET does also directly compete with streamflow in many watersheds;

3) dynamics in the values of nonlinearity are not driven by changes in monthly-average forcing variables;

4) since 1980, the nonlinearity of streamflow has increased across the US, while the stability of streamflow has increased in some regions (e.g., the Great Plains and southern California) and decreased in others (e.g., the southeast and northern California).

For most watersheds, we are able to ascribe the majority of the variance in the stability of streamflow recession to seasonal dynamics and regional-scale monthly-average forcing variables. The variance of nonlinearity, though, cannot be explained by variables at such coarse resolution. Finally, while broad regional and temporal recession trends are consistent across the US, we also identify several regions of anomalous behavior that may serve as useful case studies for future investigations into the physical origins of streamflow recession (e.g., the southern Appalachian Mountains, the New Madrid Seismic Zone, and the Cascade Range).

CHAPTER 3: EFFECTIVE HYDRAULIC CONDUCTIVITY AND DRAINABLE STORAGE FOR THE CONTINENTAL UNITED STATES[3]

Section 1: Introduction

Section 1.1: Summary

In regional and global hydrologic models, the hydraulic properties of the deep subsurface are commonly estimated according to the texture of soils at the earth's surface. This approach ignores macropores and fracture flow, as well as heterogeneity in soils and bedrock. Using hydrograph recession analysis, we constrain estimates of the effective saturated hydraulic conductivity (K) and effective drainable aquifer storage (S) of all reference watersheds in the continental US for which sufficient streamflow data are available (n=1561). Then, we use machine learning methods to model these properties across the continental US using catchment geomorphology, soils, climate, and geology as model inputs. Model validation results in high confidence for estimates of log(K) (r^2 > 0.89; 1% < bias < 9%) and reasonable confidence for S (r^2 > 0.83; -70% < bias < -18%). Our estimates of effective K are, on average, two orders of magnitude higher than comparable estimates from soil texture alone, confirming the importance of soil structure and preferential flow pathways at the watershed scale. Our estimates of effective S compare favorably with recent global estimates of mobile groundwater and are spatially heterogeneous (5-3355mm). Because estimates of S are much lower than the global maximums generally used in land surface models (e.g., 5000mm), they may serve both to limit model spin-up time and to constrain model parameters to more realistic values. These results represent the first attempt to constrain estimates of watershed-scale effective hydraulic properties for the entire continental US.

Section 1.2: Background

---

[3] Note to self: add HESS early access citation info

Regional hydrologic models and land surface models (LSMs) are powerful tools for investigating the distribution of water at the Earth's surface and its response to changing land use and climate. However, while models of this type often incorporate multiple layers of soil structure and the relatively computationally expensive Richard's equation to represent the flow of water in the shallow subsurface, groundwater and baseflow processes remain poorly constrained [Clark et al. 2008, Clark et al. 2015, Fan et al. 2019, Fatichi et al. 2020]. Because groundwater sustains half of global streamflow at annual timescales [Beck et al. 2013], buffers streamflow against changes in temperature, nutrients, and precipitation (PPT) [Ficklin et al. 2016], and sustains baseflow and evapotranspiration (ET) during extended dry periods [Yang et al. 2011], modeling it correctly is important for accurately representing the water cycle and water resources.

In early LSMs, baseflow depended on one-dimensional free drainage below a thin (2-3 meters) soil layer, though many aquifers are known to be tens of meters deep [e.g., Winter et al. 1998, Goodfellow et al. 2014]. This simplification results in underestimates of seasonal storage, enhanced drainage during wet periods, and inhibited dry season ET [Baker et al. 2008, Brunke et al. 2016, Fan et al. 2017, Kuppel et al. 2017, Milly and Shmakin 2002, Miguez-Macho and Fan 2012a, Miguez-Macho and Fan 2012b, Pokhrel et al. 2013]. In an attempt to resolve these issues, many LSMs have incorporated a TOPMODEL approach to runoff [e.g., Niu et al. 2005 or Oleson et al 2010] or added a (linear) groundwater reservoir below the soil profile [e.g., Liang et al. 2003 or Niu et al. 2007]. However, this additional model complexity [NOAA 2016] has failed to generate clear improvements in model performance [Yang et al. 2011, Gan et al. 2019].

A potential confounding issue is that the hydraulic properties of the subsurface components that are contributing to baseflow are poorly constrained [Dai et al. 2019b]. Generally, they are calculated as a function of properties of the overlying soil as estimated according to soil textural class, following Gedney and Cox [2003]. However, soil textural classes may be expected to constrain values of

hydraulic conductivity (K) only to within a range of several orders of magnitude [Freeze and Cherry 1979, Gleeson et al. 2011, Huscroft et al. 2018, Zhang and Schaap 2019]. Uncertainties are known to be even greater in non-temperate climates where data are sparser [Hengl et al. 2017, Huscroft et al. 2018] and there is some disagreement even among common pedotransfer functions [Dai et al. 2019a, Zhang and Schaap 2019]. Further, soil texture is known to be a poorer predictor of hydraulic properties than other soil characteristics like hydraulic radius, structure, and sorting [Zhang and Schaap 2019]. Soil-texture based predictions may also under use additional information available in soil surveys, and do not address deeper systems like saprolite and bedrock.

Richard's equation and estimates of K from soil texture also explicitly ignore macropores [Beven and Germann 2013]. Even direct lab measurements of the K of a soil sample (on which pedotransfer functions are based) may be order(s) of magnitude lower than in-situ measurements that do incorporate contributions from macropores [Zecharias and Brutsaert 1988, Mendoza et al. 2004]. Macropores have been shown to sustain discharge not only in the shallow subsurface (e.g., bioturbation or root rot), but also at the interface between weathering bedrock and soils (e.g., saprolite) [Tromp-van Meerveld and McDonnell 2006, Beven and Germann 2013].

The geologic units used in global maps of hydraulic properties bring additional uncertainty, as they often explicitly ignore fault and fracture networks [e.g., Gleeson et al. 2014] despite these networks' enhancement of the effective porosities and K of their parent rock by up to several orders of magnitude [Freeze and Cherry 1979]. Further, regional (and global) maps of soil and geologic units are often stitched together from multiple sources, resulting in sharp boundaries in unit classes at political boundaries [Gleeson et al. 2011, Hartmann and Moosdorf 2012, Gleeson et al. 2014, Huscroft et al. 2018, Dai et al. 2019b] and are based on surveys that rarely extend more than 2m below the surface [Dai et al. 2019a].

Even where the three-dimensional structure of the subsurface is well mapped, assessing the effective hydraulic properties of the components of the subsurface that actively contribute to streamflow remains problematic [Binley et al. 1989]. These complexities are compounded in LSMs that aggregate the hydraulic properties of the soils and/or geologic material underlying a watershed whether or not those materials are actually likely to be saturated. Nevertheless, because the time-varying sourcing of water within the three-dimensional structure of a watershed has a major impact on the stability of streamflow [Barnes 1939], as well as its chemistry and biology [Zhi et al. 2019], researchers are making strides towards incorporating these processes into modern LSMs [Fan et al. 2019]. For instance, including estimates of depth to water table in LSMs helps modulate soil moisture distribution, rates of ET, and streamflow response [Miguez-Macho et al. 2007, Koirala et al. 2018]. Similarly, Brunke et al. [2016] showed that incorporating spatial differences in depth of unconsolidated material has major impacts on patterns of baseflow, soil moisture, and storage, as well as heat fluxes. However, no regional databases of the effective hydraulic properties of the subsurface units that actually sustain baseflow currently exist.

Section 1.3: Recession Analysis

The effective hydraulic properties of the aquifer(s) that actually sustain baseflow have long been estimated by watershed hydrologists using hydrograph recession analysis. Here, we briefly summarize the assumptions and applications of hydrograph recession analysis, though we refer readers to recession literature for a detailed accounting [Brutsaert and Nieber 1977, Harman and Kim 2019, Kirchner 2009, Tallaksen 1995, Tashie et al. 2020b, Troch et al. 2013] and give a more thorough description of the specific methods we use and their underlying uncertainties below.

Applying the Dupuit assumptions [Dupuit 1863] that groundwater flow is horizontal and neglecting capillarity, the Boussinesq equation [Boussinesq 1877] is used to describe unit discharge (q) as a function of effective saturated hydraulic conductivity (K), drainable porosity (f), drainable aquifer

thickness (D), aquifer breadth (B), slope (i), stream network length (L), and the water table elevation (h) profile. When all other inflows and outflows to a watershed are negligible, total discharge (Q) is a function of drainable aquifer storage (S = Df) and the distribution of water within the aquifer. Brutasert and Nieber [1977] showed that if change in streamflow (dQ/dt) is expressed as a function of time-independent Q according to the power law in Eq. 1 ($dQ/dt = aQ^b$), then a is a function of static watershed characteristics (K, f, D, B, L, and i) while b is a function of aquifer geometry and the profile of h. Because a and b are readily calculated by logarithmic regression of Q on -dQ/dt, it is possible to use historical streamflow data to estimate watershed-average hydraulic properties (e.g. K, S, D, or f) which are otherwise difficult to measure [Szilagyi et al. 1998].

Though the assumptions underlying recession analysis are restrictive, this technique has repeatedly been validated analytically, using computer models [e.g., Rupp and Selker 2006, Pauritsch et al. 2015], table-top models [Guerin et al. 2014, Luo et al. 2018], tracer experiments [Winkler et al. 2016], in highly monitored hillslopes [Clark et al. 2009], and in many actual catchments [e.g., Mendoza et al. 2003, Zecharias and Brutsaert 1998, Troch et al. 1993, Parlange et al. 2001, Pauritsch et al. 2015]. However, recent work has illustrated the need to incorporate the heterogeneity of subsurface hydraulic properties to minimize biases [Winkler et al. 2018, Tashie et al. 2020a]. Because the character of baseflow recession (and therefore the properties of the subsurface which are sustaining it) has been shown to vary systematically with antecedent conditions [Shaw and Riha 2012, Tashie et al. 2020b], recession analysis is better seen as providing a snapshot of the effective hydraulic properties that are instantaneously sustaining baseflow [Tashie et al. 2020a]. "Parallel" reservoirs representing the end-member properties of subsurface heterogeneity have been successfully invoked to describe dynamic recession patterns [Clark et al. 2009, McMillan et al. 2011, Bart and Hope 2014, Winkler et al. 2016, Gao et al. 2017, Mateo-Lázaro et al. 2018, Tashie et al. 2019]. By applying recession analysis only during periods when baseflow is known to be dependent on specific components of subsurface storage (e.g.,

during winter in Swiss catchments [Pauritsch et al. 2015]), uncertainty in estimates of the effective

hydraulic properties and drainable storage volumes of those subsurface features may be minimized.

Section 1.4: Major Aims

Here, we estimate the effective hydraulic properties (i.e., K and S) of the subsurface units that

actually sustain baseflow for all watersheds in the continental US. First, we apply recession analysis

according to best practices to estimate effective K and S while accounting for the dynamic partitioning of

water among heterogeneous features of the subsurface. We do so for all USGS reference watersheds in

the US for which sufficient streamflow data exists (n = 1561), selected from all watersheds gauged by

the US Geological Survey (USGS) that are considered to be minimally impacted by human interference

[Falcone 2010] (n=1946). We calculate the methodological uncertainty in our methods, then develop a

model for predicting effective K and S using catchment average variables describing soils, geology,

climate, topography, and geomorphology. After validating this model, we apply it to all watersheds in

the US at the Hydrologic Unit Code-12 (HUC-12) scale. Finally, we contextualize our results and provide a

brief overview of their limitations and implications for large-scale hydrologic modeling.

Section 2: Methods

Section 2.1: Recession Analysis

To develop a nationwide dataset of effective hydraulic properties derived from observations, we

analyzed the historical streamflow from all watersheds gauged by the US Geological Survey (USGS) that

are considered to be minimally impacted by human interference [Falcone 2011] (n=1946).

Recession analysis is performed on historical streamflow data during time periods when

additions to watershed storage (e.g., PPT or snowmelt) as well as abstractions (e.g., ET or groundwater

withdrawal) are minimal, i.e. "recession periods." Selection criteria for identifying recession periods are

chosen to minimize interference by these processes, and have been shown to affect parameter values

[e.g., Stoelzle et al. 2013]. We rely on recommendations from Dralle et al.'s [2017] uncertainty analysis

for selection criteria: both Q and -dQ/dt must decrease for at least four consecutive days, no antecedent

flow conditions are imposed, and no days are removed from the beginning or end of a period of

recession. We estimated –dQ/dt by (Qi – Qi-1) / (ti – ti-1) and adjusted Q to reflect this time delay, such

that Q = (Qi + Qi-1) / 2, following Shaw and Riha [2012]. While recent research has shown that

incorporating a variable time step may reduce noise and improve parameter estimation [Roques et al.

2017], these methods require subdaily (hourly or less) Q records that are either not available or not

efficiently accessible for most USGS gages. Instead, we relied on daily Q time series and minimized noise

in the recession cloud through recession selection criteria (i.e., requiring a monotonic decrease in –

dQ/dt) and minimized bias in recession parameter estimation by applying quantile regression, as

described below.

There are multiple methods for implementing recession analysis, each designed according to a

suite of different simplifying assumptions [Tallaksen 1995, Troch et al. 2013]. Pauritsch et al. [2015]

recently provided uncertainty analysis for several of these methods and suggested the solutions of

Brutsaert [1994] as performing the best across a range of aquifer conditions. Two solutions are

presented, such that b (from Eq.1) is set to 3 during the early time period of recession and a is solved as

$$a_{early} = \frac{1.33}{KfD^3L^2\cos(i)} \qquad\qquad Eq.2$$

and *b* is set to 1 during the late time period of recession and *a* is solved as

$$a_{late} = \frac{\pi^2 pKDL^2}{fA^2}\cos(i)\left[1 + \left(\frac{\frac{B}{D}\tan(i)}{\pi p}\right)^2\right] \qquad\qquad Eq.3$$

To ensure methodological consistency, we applied these solutions to all watersheds. We refer readers to

Brutsaert [1994] for the derivation of these solutions and their detailed description.

In addition to traditional assessments of effective hydraulic properties using "lumped" recession

analysis [e.g., Berghuijs et al. 2016, Hinzmann et al. 2020, Ploum et al. 2019], it has become increasingly

common to perform recession analysis on individual events with a focus on differences in recession form

between events [e.g., Dralle et al. 2017, Shaw et al. 2012, Tashie et al. 2020a]. These event-based

studies have highlighted important temporal dynamics in recession form that have been attributed to

the extension and contraction of the stream network (L) [Biswal and Marani 2010] as well as

precipitation patterns and ET [Bart and Tague 2017]. However, these dynamics are most commonly

attributed to the partitioning of S among heterogeneous subsurface features [Bart and Hope 2014, Clark

et al. 2009, Gao et al. 2017, Harman and Kim 2019, Mcmillan et al. 2011, Shaw et al. 2016, Tashie et al.

2019, Wang and Cai 2010], as was recently confirmed empirically for over 1,000 watersheds across the

continental US [Tashie et al. 2020b].

To account for the effects of the dynamic distribution of S among heterogeneous subsurface

features on estimates of effective hydraulic properties, we took the following steps. Broadly, we

partitioned the streamflow time series into periods of relatively low S and periods of relatively high S

through evaluation of antecedent streamflow records, a well-established proxy [Mishra et al. 2003,

Biswal and Kumar 2014, Bart and Hope 2014, Patnaik et al. 2015]. A confounding issue is that watershed

memory is strongly dependent on the physio-climatic properties of the watershed in question [e.g.

Tromp-van Meerveld and McDonnell 2006, Tetzlaff et al. 2009, Nieber and Sidle 2010, Spence et al.

2010, and Patnaik et al. 2015]. Indeed, Bart and Hope [2014] showed that streamflow from six months

previous was a better predictor of temporally variable recession parameters than more recent

streamflow for several watersheds in the California Sierra Nevada. Therefore, we adapted the methods

of Patnaik et al. [2015] to partition each day of streamflow recession for each watershed into periods of

relatively low S and high S according to cumulative antecedent Q. For each day of recession, we

calculated the cumulative antecedent Q at multiple time lags (7, 15, 30, 45, 90, and 180 days). Then, for

each watershed at each time lag, we classified each day of recession into "low S" (lowest 20th-percentile

cumulative antecedent Q) and "high S" (highest 20th-percentile cumulative antecedent Q). For each

watershed, we assessed the effectiveness of each time lag in partitioning recession flows into these two

categories by calculating the Euclidean distance in a log-log recession plot between the median value of
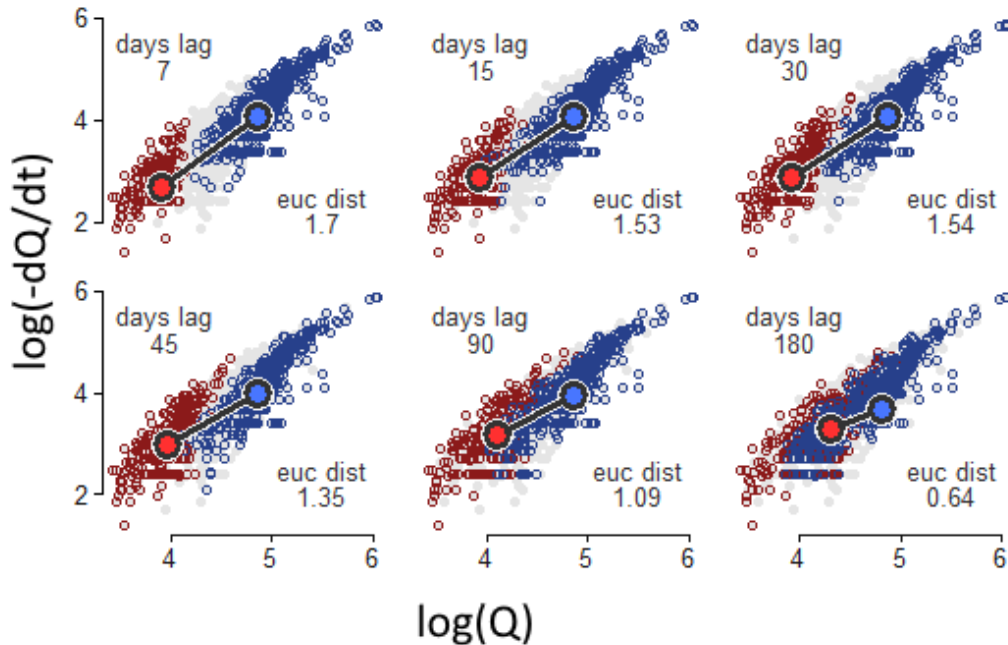
Figure 1: example of identifying the time step which maximally disaggregates periods of "low storage" from periods of "high storage." Recession plots illustrate the log base 10 of Q (in m^3/s) (x-axes) plotted against the log base 10 of the time derivative of Q (y-axes). For each time lag, the lowest 20th percentile storage states are indicated in red, and the highest 20th percentile storage states in blue. Larger red (and blue) dots indicate median value of Q and -dQ/dt during low S (and high S) periods, with the Euclidean distance between them given in the lower right.

recession points during days which occurred during the lowest (and highest) 20th-percentile cumulative antecedent Q (Figure 1). We used the time lag that maximized this distance for each watershed. We retained for analysis only watersheds for which we identified at least 500 total days of recession, with at least 25 days of recession occurring during periods of both low S and high S (n = 1561).

Recession analysis exploits the fact that the shape of hydrographs during early periods of recession are distinct from the shape of hydrographs during late periods of recession to develop two empirical solutions to the recession equations (Eq.2 and Eq. 3). These distinct hydrograph shapes express themselves in "recession plots" (i.e., a plot of Q and –dQ/dt in log-log space) as the "early envelope" during the early period and the "late envelope" during late period recession. The exact delineation of the early and late envelopes is somewhat arbitrary [Troch et al. 2013] and should not be seen as an expression of actual hillslope hydraulic behavior at a particular moment in time [Tashie et al.
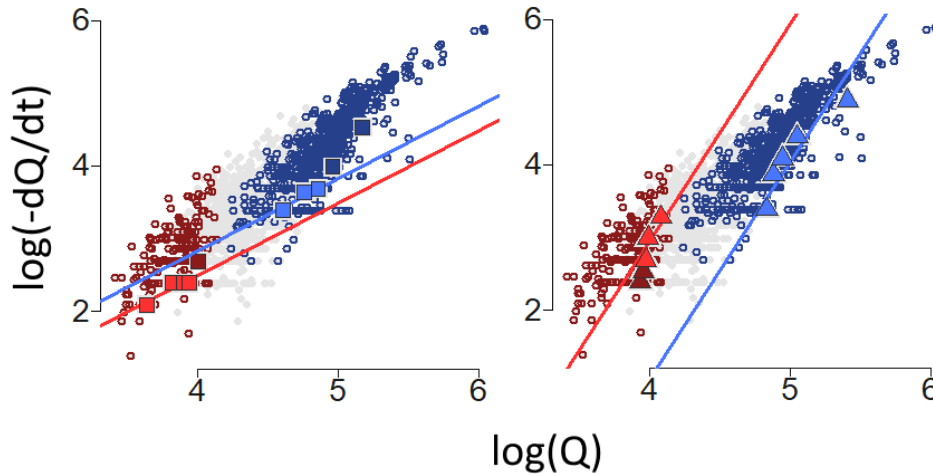
Figure 2: identifying the line of best fit for the late envelope (left) and early envelope (right) during states of low storage (red) and high storage (blue). Recession plots illustrate the log base 10 of Q (in m^3/s) (x-axes) plotted against the log base 10 of the time derivative of Q (y-axes). Grey dots indicate days of recession during mesic conditions. Red (and blue) dots illustrate days of recession during low (and high) storage. Triangles indicate the median values of the lowest 20th percentile of Q according to binned quintiles of -dQ/dt. Squares indicated the median values of the lowest 20th percentile of -dQ/dt according to binned quintiles of Q. Lighter colored triangles (and squares) indicate quintiles which were included in the modeled best fit for defining the early (and late) envelopes.

2020a], but instead as a lower boundary on the effective hydraulic properties of the aquifer(s) that actually sustain baseflow during a particular state of S [e.g., Gao et al. 2017]. Automated methods depending on fitting a line of a defined slope such that some specific fraction of data points are below the line (either 2% [Vannier et al. 2014], 5% [Troch et al. 1993, Stoelzle et al. 2013], or 10% [Zecharias and Brutsaert 1988, Mendoza et al. 2003, Tashie et al. 2020a]. Commonly acknowledged problems include sensitivity to outliers and identification of the "knickpoint" where early-period recession (i.e., the early envelope) transitions to late-period recession (i.e., the late envelope), which has a significant impact on parameter estimation [Pauritsch et al. 2015]. Here, we adapt the methods of Tashie et al. [2020a] to delineate the late envelope according to the following steps: 1) data are binned by quintile of Q; 2) for each bin, the lower 20% of data according to -dQ/dt are selected; 3) a line with a slope of 1 (as determined by the late-period solution to recession flow provided by Brutsaert [1994]) is fitted to the data of the first three bins, the first four bins, and all five bins in order of increasing Q; 4) the line with the maximum r2 is selected as the best solution. This process is repeated for the early envelope, except

data are binned by quintile of -dQ/dt in step 1, selected according to the lower 20% of Q in step 2, the

fitted line has a slope of 3 in step 3 (as determined by the early-period solution to recession flow

provided by Brutsaert [1994]), and lines are fitted in order of decreasing Q. An example of this process is

illustrated in Figure 2.

Section 2.2: Estimating D and K

There remains the issue that Brutsaert's [1994] equations have seven unknown parameters (K, f,

D, B, L, i, and the fitting parameter p) and empirical solutions from only two equations. Following

common practice [e.g., Mendoza et al. 2003, Vannier et al. 2014], we set p to 1 and estimated L, B, and i

using readily available geomorphological descriptions of perennial stream network length, catchment
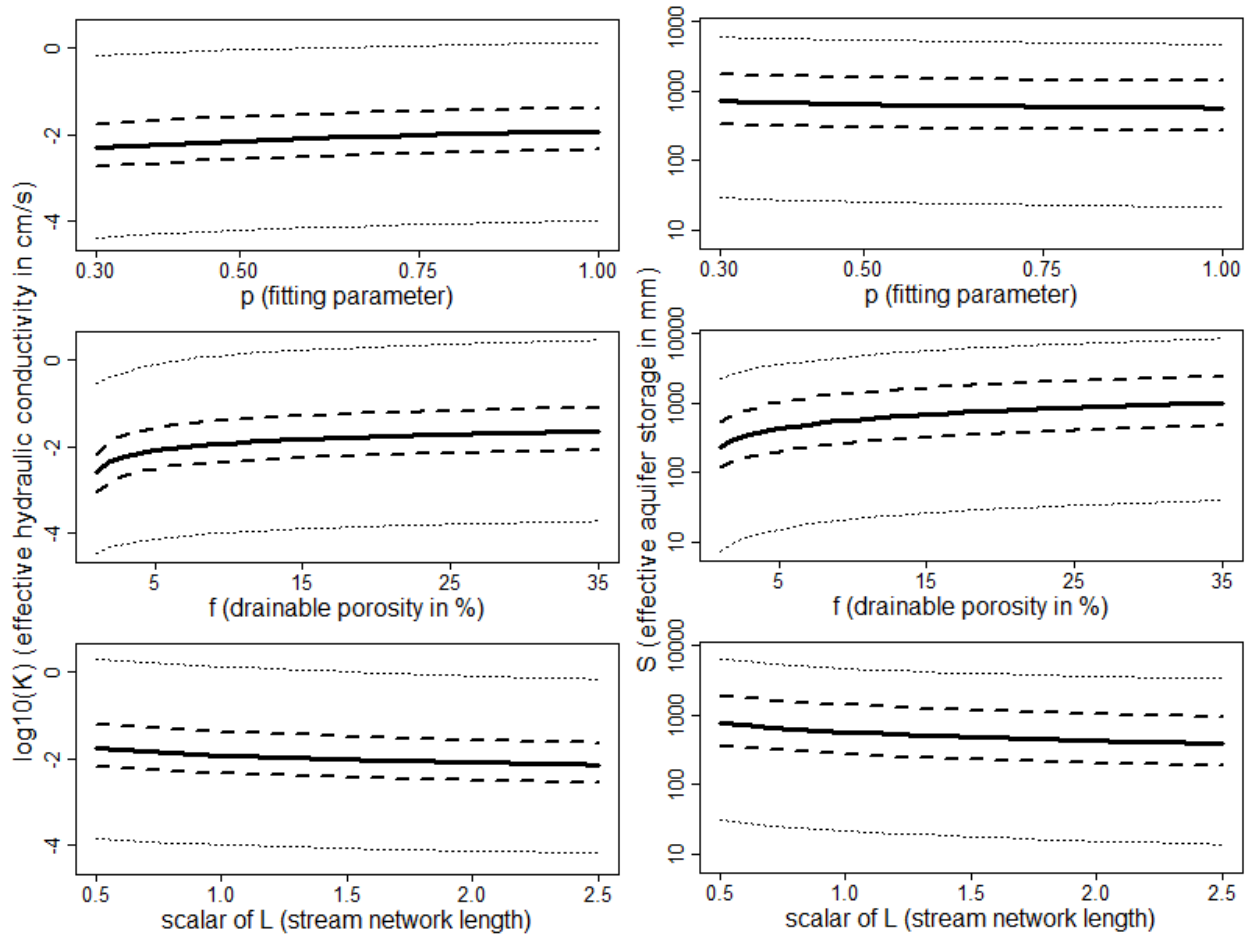

Figure 3: sensitivity of K (left) and S (right) to changes in p (top), f (middle), and scalar of L (bottom). Bold solid lines are median values, dashed lines are 1st and 3rd quartile values, and thin dotted lines are minimum and maximum values.

area (A) divided by L, and median topographic slope provided by Falcone et al. [2010]. Because f is generally the most narrowly constrained of the three remaining parameters in the published literature, we made a reasonable estimate of its global value and solved for K and D, an approach common in recession analysis [Szilagyi et al. 1998, Brutsaert and Lopez 1998, Mendoza et al. 2003, Vannier et al. 2014]. We chose a value for f of 0.1, which is near the estimated global average near-surface value of 0.14 [Gleeson et al. 2014] and is in the range of previous recession analysis studies [e.g., Vannier et al. 2014]. Because we used a universal value of f, we chose to convert D (effective drainable aquifer thickness) to the more intuitive and useful variable of S (effective drainable storage) for subsequent analysis.

To ensure our results were robust, we assessed the sensitivity of the recession equation to uncertainty in parameter estimation. Because the length of active drainage networks varies by a factor of 2 or 3 under typical flow conditions relative to a presumed static L [Godsey and Kirchner 2014], we reran all analysis with L scaled by a factors of 0.5 to 2.5. Similarly, since f is known to range from 0.02 to 0.22 in common soils [Heath 1983], we reran all analysis with values of f ranging from 0.01 to 0.35. Finally, while it is common to set p to 1 and it is rarely set below 0.5, we reran all analysis with values of p ranging from 0.3 to 1. Both L and p had only a moderate impact on estimates of K and S; median values of log(K) and S for every value of L and p were greater than the 1st quartile values of log(K) and S for every value of L and p, and (similarly) median values of log(K) and S for every value of L and p were less than the 3rd quartile values of log(K) and S for every value of L and p (Figure 3). However, estimates of K and S were somewhat more sensitive to extremely low values of f, with an increase in values of f from 0.01 to 0.06 resulting in an increase of S by a factor of 2 and an increase in log(K) by a factor of 3.5. K and S are less sensitive to changes in f where f is greater than 0.05.

Section 2.3: Predicting S and K from Catchment Average Data

Because baseflow is strongly influenced by climate, soils, geology, and geomorphology in a generally nonlinear fashion [e.g., Winter 2007, van Dijk 2010, Beck et al. 2013], we assessed the strength of the relationships of calculated values of K and S with catchment-average physio-climatic properties using random forest models. We primarily relied on the USGS NHD Plus Version 2.1 database [Wieczorek et al. 2018] for watershed summary data. We narrowed the potential number of watershed summary metrics from roughly a thousand to the 99 most hydrologically relevant metrics describing climate, topography, geology, and soils. These are listed in the supplementary material. We also summarized and assessed distributed soil and geology attributes from a global database [Beck et al. 2013, Gleeson et al. 2011, Gleeson et al. 2014] though none of these metrics were ultimately included in our model because they did not improve performance. Additionally, we calculated the ratio of annual potential evapotranspiration (PET) to PPT (i.e., the aridity index) and derived two metrics of seasonality (one for baseflow and one for PET), where seasonality is defined as $Seasonality =$

$$\frac{\sum \left| monthly\ value - \frac{annual\ value}{12} \right|}{annual\ value}.$$

Overfitting of machine learning models is a well-documented phenomenon [Domingos 2012] that can be exacerbated in datasets with spatially autocorrelated inputs [Meyer et al. 2019]. We took extensive measures to minimize overfitting within our random forest models following the methods of Meyer et al. [2018]. Key steps included: 1) rounding all input features to 3 significant digits to avoid input features acting as unique identifiers; 2) forward-feature selection based on target-oriented performance to reduce the overall parameter space; and 3) model development using leave-location out block cross-validation (LLO-CV). In combination, these steps significantly reduce overfitting within spatially explicit random forest algorithms [Meyer et al. 2018]. Resulting inputs from the forward feature selection for each dependent variable are in Table 1. LLO-CV was done using HUC-2 watersheds as blocks; we partitioned them into 5 equally proportioned folds. Final models were trained through

Table 1: selected independent variables in order of importance for each dependent variable. Performance metric (r^2) is cumulative. Note that the random forest model implemented here does not allow the calculation of performance metrics using a single independent variable. Variable names are adapted from Wieczorek et al. [2018].

Number of variables	K (low S periods)		K (high S periods)		S (low S periods)		S (high S periods)	
	variable	r^2	variable	r^2	variable	r^2	variable	r^2
1	stream elevation change	NA	max elevation	NA	BFI	NA	max elevation	NA
2	max elevation	0.705	depth to WT	0.83	annual PPT	0.732	annual PPT	0.841
3	% sand	0.768	stream elevation change	0.881	% soil group HGA	0.796	% organic matter	0.859
4	depth to WT	0.803	% organic matter	0.895	% sediment <5mm	0.81	no20	0.888
5	BFI	0.829	BFI	0.903	annual mean temperature	0.823	stream elevation change	0.902
6	% colluvium	0.843	% sand	0.908	annual baseflow	0.83	annual max temperature	0.904
7	watershed area	0.858	depth to bedrock	0.913				
8	% silt	0.875	% sediment <5mm	0.916				
9	mean temperature	0.888						
10	% clay	0.89						

LLO-CV using 80% of the included watersheds and RMSE as the evaluation term. The remaining 20% of watersheds were set aside as hold-out data for model testing. The number of candidate variables tested at each split ('mtry' in random forest) was identified using a grid search approach from 1 to the total number of feature inputs for a given model which resulted in final mtry values of 5-7 features per tree, depending on the model. All analysis was done using R version 3.6.0 statistical software ["R Core Team" 2017]. We then assessed final models using the 20% hold-out data as described below in Results.

Roughly 3% of watersheds in the continental US lacked sufficient watershed-scale summary data to apply our model. We estimated values of K and S for these watersheds according to the USGS product
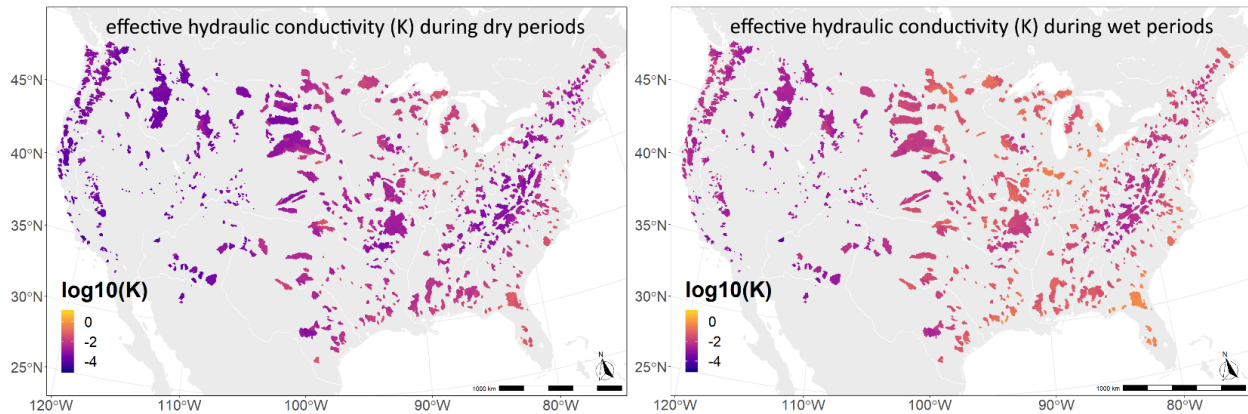
Figure 4: the effective hydraulic conductivity (K) for all watersheds as determined by recession analysis. Effective K during periods of low storage is given in the figure to the left, and during periods of high storage in the figure to the right. Units are the log10 of cm/s.

Hydrologic Landscape Regions (HLR) [Wolock 2003]. First, we organized all watersheds according to their classification within the HLR framework. We calculated average values of K and S for each HLR according to the median value of all available data. Then, we assigned these values of K and S to the 3% of watersheds with missing data according to their HLR classification (Figures S1-4).

Section 3: Results

Calculated values of effective K range from below 3.2 x 10-5 cm/s to as high as 1.3 x 10-1 cm/s (Table 2, Figure 4). This range is typical of semi-permeable to permeable soils (layered clays, silts, sands, and gravels) and bedrock (sandstones, fractured crystalline rocks, and karst limestones) [Freeze and Cherry 1979, Domenico and Schwartz 1990] and is similar to results from previous applications of recession analysis in comparable catchments [e.g., Mateo-Lázaro et al. 2018, Mendoza et al 2004, Pauritsch et al. 2015, Winkler et al. 2016]. Effective K was moderately elevated during high S periods at most watersheds, but ranges generally shifted by less than a full order of magnitude, similar to the results of Pauritsch et al. [2015].

Table 2: summary of calculated values of K and S for all 1561 reference USGS watersheds for which sufficient data exists.

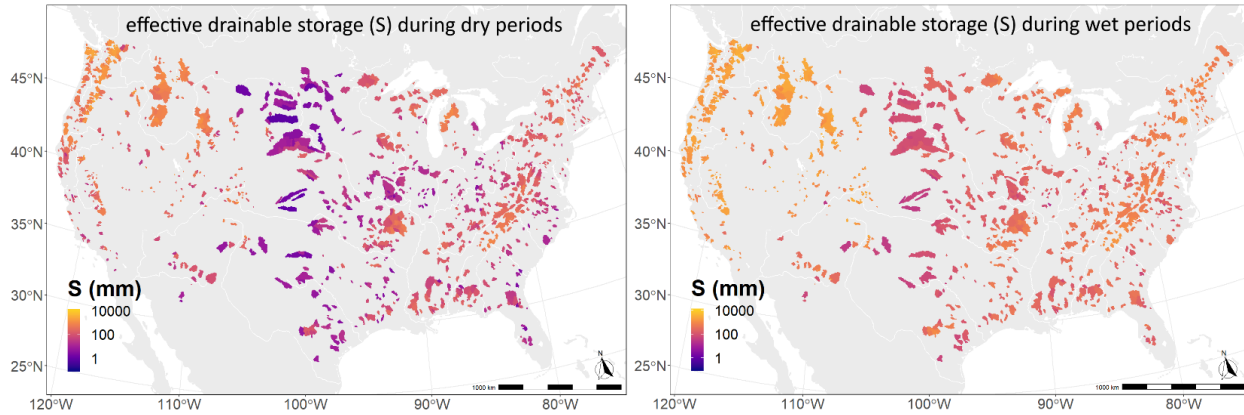|		Min	Q1	Med	Mean	Q3	Max
log10(K)	*high S periods*	-4.0	-2.4	-1.9	-1.9	-1.4	0.1
(cm/s)	*low S periods*	-4.5	-3.3	-2.8	-2.7	-2.2	-0.7
S	*high S periods*	21	268	575	912	1397	4680
(mm)	*low S periods*	1	49	155	318	366	3309

Figure 5: the effective drainable storage (S) for all watersheds as determined by recession analysis. Effective S during periods of low storage is given in the figure to the left, and during periods of high storage in the figure to the right. Units are in mm.

Calculated values of *S* ranged from 1 to 3300 mm during low S periods and 20 to 4700 mm during high S periods (Table 2, Figure 5). Given our assumption of a static areal aquifer extent and an *f* of 0.1, 1st to 3rd quartiles of our calculated effective aquifer thickness (*D*) range from 0.5 to 3.7 m during low S periods and 2.7 to 14 m during high S periods, similar to previous results from regional-scale recession analysis [e.g., Mateo-Lázaro et al. 2018, Mendoza et al 2004, Pauritsch et al. 2015]. Recall that *D* as estimated by recession analysis represents the maximum thickness of the aquifer(s) that is actively contributing to streamflow during periods of recession. Therefore, 1-3300 mm (or 20-4700mm) represents the range of contributing subsurface storage volumes of watersheds during low S periods (or high S periods) during the actual historical record. Immobile soil water and effectively immobile groundwater in impermeable substrate that is not actively supplying baseflow is masked. Similarly, potential storage volumes that have not been activated during the historical record are also masked.

78

Validation results of the random forest models used to extend these calculations to all watersheds in the continental US are reported in Table 3. While all random forest models performed well ($r2 > 0.8$), they were better able to predict both $\log(K)$ and S during high S periods ($r2$ of 0.92 and 0.90) than low S periods ($r2$ of 0.89 and 0.83). The mean absolute error (MAE) of K for both high S and low S periods was small ($< 20\%$) relative to the absolute range of effective K of watersheds in this study (5 orders of magnitude) and the range of K of common materials (15 orders of magnitude). The MAE of S was similarly well constrained (< 170 mm) relative to the range of calculated values which were between 0 and 4000 mm. Further, the models made no predictions of hydrologically unrealistic values of either K (e.g., > 10 cm/s) or S (e.g., < 0 mm). Nonetheless, all models were biased, tending to underpredict extreme values, especially in the case of S during low S periods (Figure 6). For this reason, the range of values of K and S predicted by our models (Table 4) are narrower than the range of values calculated using recession analysis (Table 2).

Table 3: validation of model performance

		MAE	RMSE	r^2	% bias	SD of residuals
log10(K)	*high S periods*	0.14	0.19	0.92	8%	0.17
(cm/s)	*low S periods*	0.17	0.23	0.89	1%	0.21
S	*high S periods*	166	276	0.90	-18%	247
(mm)	*low S periods*	108	200	0.83	-70%	157
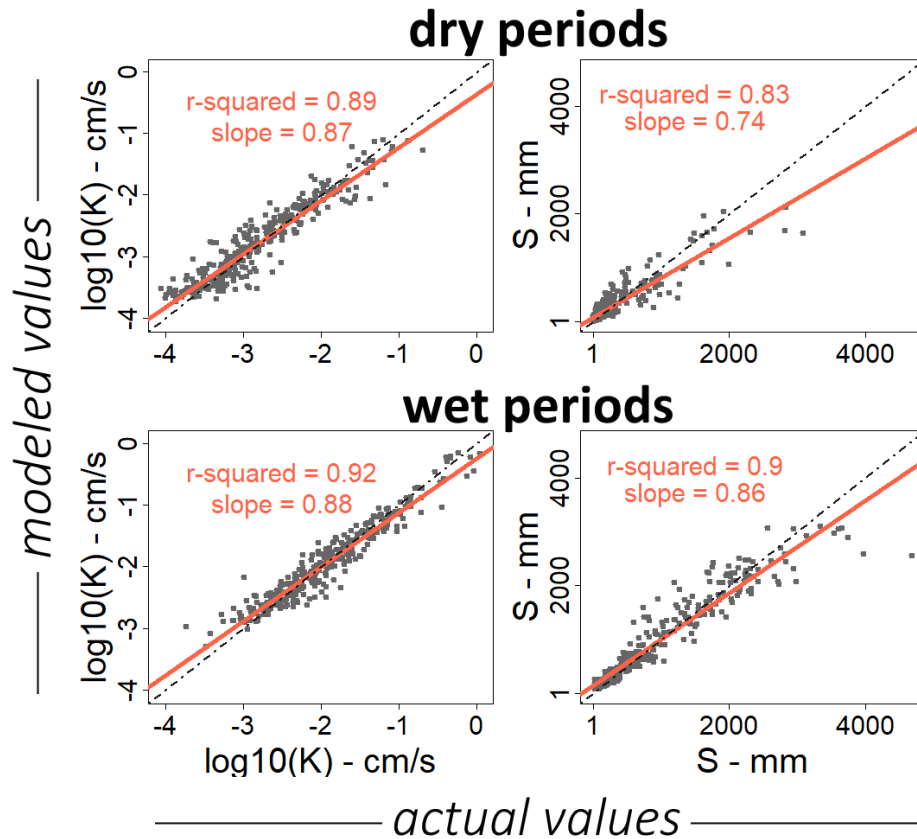
Figure 6: validation of model performance. All data points are from hold-out data (n=300). X-axes are of actual values calculated using recession analysis and y-axes are of estimated values using the random forest model. Left plots are of K and right plots are of S. Top plots are estimates during low S periods and bottom plots are estimates during high S periods.

Section 4: Discussion

Section 4.1: Effective K

Model estimates of effective K spanned fewer than 5 orders of magnitude, within the range of values expected of permeable and semi-permeable materials (Table 4). Indeed, the effective K of most watersheds was greater than the global geometric mean of K for unconsolidated materials [Huscroft et al. 2018]. This relatively tight range of relatively high values of K across the US may seem counterintuitive, given that the hydraulic properties of (un)consolidated materials commonly found in

the landscape may span more than ten orders of magnitude [Heath 1983]. Instead, these results should

be seen to highlight the axiomatic principle that (im)permeable materials are (im)permeable. That is, a

majority of baseflow is everywhere dependent on materials with a relatively narrow range of relatively

high K values, because water stored in low K materials does not flow to streams efficiently.

Table 4: summary of modeled values of K and S for the continental US.

		Min	**Q1**	**Med**	**Mean**	**Q3**	**Max**
log10(K)	*high S periods*	-3.6	-2.0	-1.4	-1.5	-0.9	-0.1
(cm/s)	*low S periods*	-3.7	-2.5	-1.9	-2.0	-1.5	-1.0
S	*high S periods*	71	216	316	430	510	3355
(mm)	*low S periods*	5	35	80	138	176	2176

Our model suggests climate plays a strong role in governing effective K in watersheds in the

continental US. Indeed, the "100th meridian" that delineates the arid West (where PET > PPT) from the

humid East (where PET < PPT) [Powell et al. 1879] is readily apparent in the estimates of effective K

during low S periods (Figure 7, top). Climate controls on effective K are driven by two independent

processes. First, water facilitates the weathering of bedrock and enhances soil structure through

increased biomass and bioturbation; each of these processes is known to increase the hydraulic

conductivity of substrate by order(s) of magnitude [Fatichi et al 2020]. More fundamentally, because

most watersheds are heterogeneous and because highly conductive hillslopes (or aquifers) drain more

quickly than less conductive hillslopes (or aquifers), streamflow becomes increasingly dependent on less

conductive hillslopes (or aquifers) during extended low S periods [e.g., Clark et al. 2009]. This process

drives down the effective K of a watershed as it dries out [e.g., Binley et al. 1989], as is evident for nearly

all watersheds in this study (Figure 7). Our results contrast with the analysis of Hirmas et al. [2018] who

found that the effective porosity (and therefore K) of soils was positively correlated with hotter, drier

conditions in the continental US. A likely explanation for our contrasting results lies in our underlying

assumptions: while their approach explicitly ignored organic material and the bedrock interface, our

approach explicitly ignores all materials which do not substantially contribute to streamflow.

Nevertheless, both their study and ours reach the important conclusion that climate has a strong

influence on the hydraulic properties of the subsurface, with implications for water-cycle feedbacks
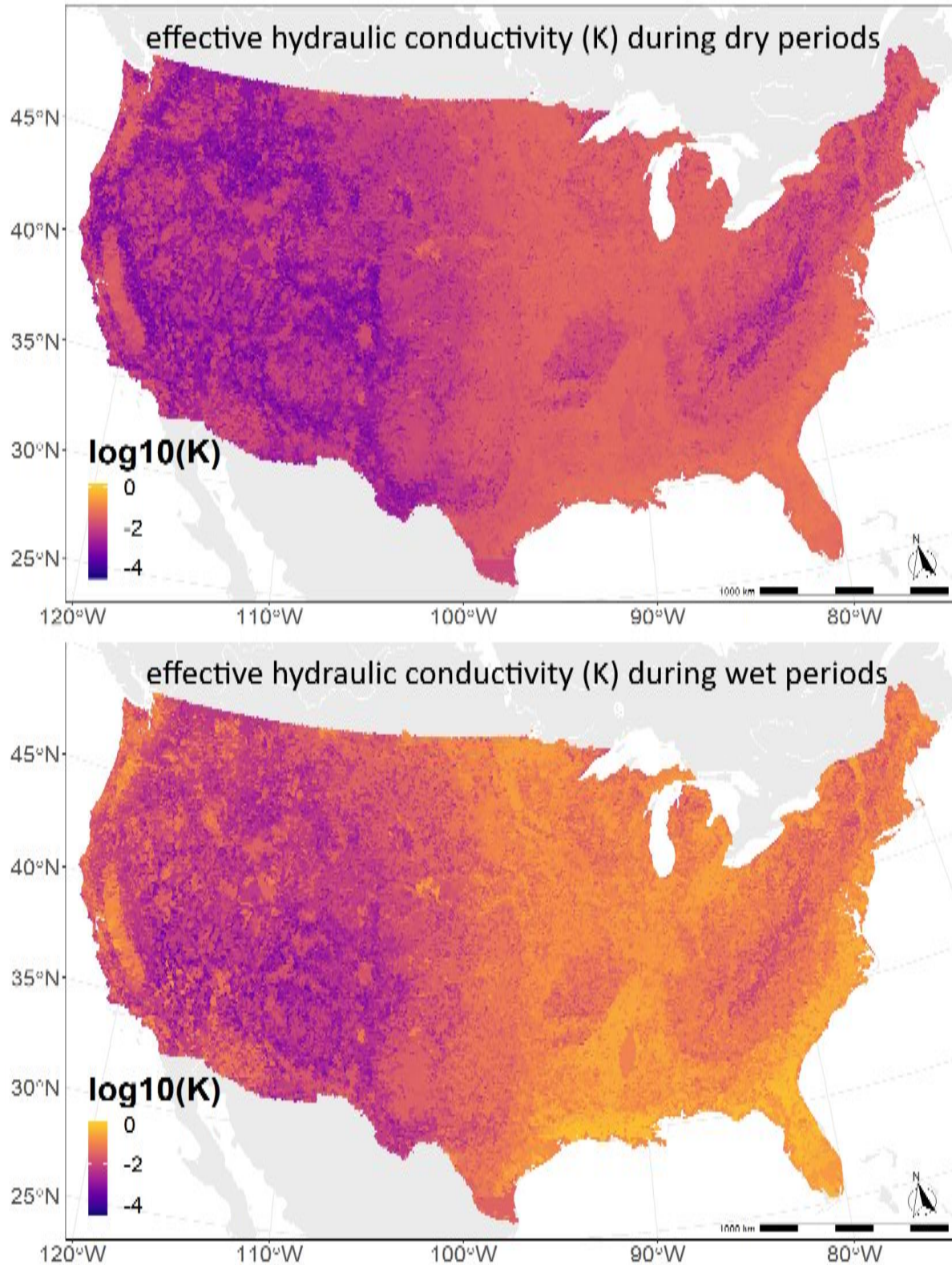
under a changing climate.

Figure 7: modeled estimates of the effective saturated hydraulic conductivity (K) of HUC12 watersheds. Effective K during periods of low storage is given in the top figure, and during periods of high storage in the bottom figure. Units are in cm/s.

Section 4.2: Effective S

S is primarily a function of climate and topography in our model. The lowest values of S are in the low topography regions of arid Great Plains west of the 100th meridian, while the highest values of S are in humid and mountainous regions like the Pacific Northwest and southern Appalachian mountains. Similarly, the greatest changes in effective S are in high topography regions with strong seasonal patterns in PPT, especially where annual recharge is dependent on spring snowmelt, as in the California Sierra Nevada and the Rocky Mountains of Idaho and Montana. While the effects of climate on S are readily apparent, it is important to recall the specific definition of S used here and its relationship to topography: S is the drainable groundwater that may sustain streamflow within a particular watershed. Soil water is not included in S, nor is any groundwater that is either deeper than the watershed outlet or is of insufficient hydraulic head gradient to sustain measurable surface flow. This definition has two important implications for the use of our data in models.
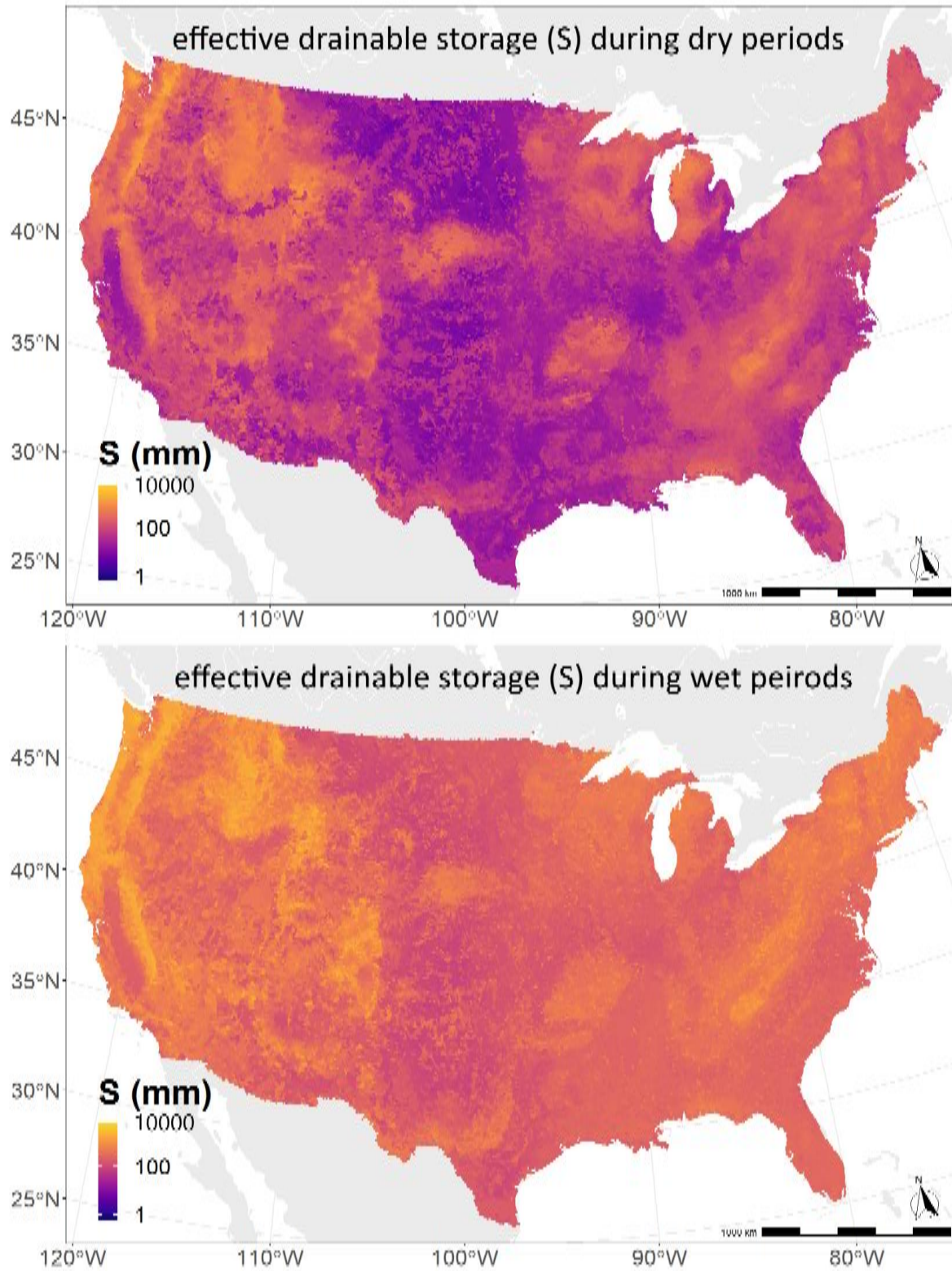
Figure 8: modeled estimates of the effective drainable storage (S) of HUC12 watersheds. Effective S during periods of low storage is given in the top figure, and during periods of high storage in the bottom figure. Units are in mm.

First, a difficulty in the implementation of groundwater modules in large-scale hydrologic

models has been the long spin-up times required for annual-average groundwater storage volumes to

equilibrate, especially in arid regions [Yang et al. 1995, Niu et al. 2007, Ma et al. 2017, Gan et al. 2019].

By masking immobile water and potential storage volumes which have not actually contributed to

baseflow during the historical record, our results may be used to constrain potential active storage

volumes to ranges that are much narrower and more realistic. Our ranges of mobile S of 70-3400 mm

and 5-2200 mm during periods of relatively high and low storage contrast sharply with the much higher,

universal global maximum parameterization of 5000 mm that is common in the Community Land Model,

Version 4.5 (CLM4.5) [Oleson et al. 2010] and Noah-Multiparameterization Land Surface Model (Noah-

MP) [Niu et al. 2011, Brunke et al. 2016], for example. Further, our results of effective S are comparable

to recent global estimates of modern groundwater storage [Gleeson et al. 2016] as well as recent

estimates of the total drainable storage of the Mississippi River Basin [Ehalt Macedo et al. 2019]. By

providing a realistic upper limit to mobile groundwater storage for all watersheds in the continental US,

these data may be used to reduce model spin-up time considerably.

Also, considering mobile groundwater as a unit of storage distinct from deep, immobile

groundwater has important implications for watershed response to climate change. Models that fail to

distinguish mobile storage from deep, immobile storage [e.g., Niu et al 2007] may underestimate the

effects of relatively small losses in total groundwater storage on streams that depend on mobile storage

in highly conductive, relatively shallow aquifers. Even small losses of baseflow to streams may have

major implications for temperature, quality, connectivity, and extent that are essential to their biological

integrity and to managers of water resources [Allan and Castillo 2007, Botter et al. 2013]. Conversely,

calculating baseflow as free drainage from all deep groundwater storage within a watershed may

overestimate the loss of groundwater during extended low S periods by allowing streams to continue to

drain groundwater long after the hydraulic head gradient in real watersheds would have reversed.

Section 4.3: Limitations and Future Efforts

Our sensitivity analysis shows that estimates of K and S varied by only a factor of less than 3 in relation to changes in scalars of L (i.e., the length of the stream network) from 0.5 to 2.5. While this magnitude of change in K and S is relatively small compared with the range of values of K among watersheds (4 and 2 orders of magnitude, respectively), dynamic values of L may account for much of the change in values of K and S between periods of high and low storage. In some watersheds, streamflow recession may be controlled in part by an expansion and contraction of the stream network [Biswal and Marani 2010] rather than by hillslope hydraulics. Similarly, the fraction of a watershed which actively contributes to streamflow has been shown to expand and contract in response to climate factors [Nippgen et al. 2015]. Therefore, some of the change in the values of K and S between periods of high and low storage may be an artefact of the contraction of both the stream network and the active contributing area of the watershed. Interestingly, while catchment area (A) was included as a potential independent variable in all random forest models, it was a significant contributor to estimates only of K during low S periods, suggesting that these results are nonetheless robust across scales.

Our methods also do not account for the potential import or export of water by deep regional flow paths, i.e. "leaky" groundwater systems [Fan et al. 2019]. For instance, many of the watersheds used to develop this model are generally in headwaters and unlikely to receive substantial regional groundwater influx. Our methods do not account for deep regional flow paths which may generate a net export of groundwater in hilly headwaters and a net import of groundwater in downstream flatter regions [e.g., Toth 1967]. This in part may explain the low values of S and high values of K estimated in the High Plains and the Coastal Plains, where low order streams are dependent on baseflow from shallow sediment and riparian aquifers, though some higher order streams may be sustained by regional groundwater flow.

The primary motivation for limiting this study to the continental US is data quality and abundance. While global datasets describing climate and geomorphology comparable to those used in this study are available, their resolution and accuracy is generally much poorer. For example, the summary variables provided Wieczorek et al. [2018] use SSURGO [SSURGO 2019] soils data that estimate soil texture, hydraulic properties, and depth at a resolution of 10 m for the US, while global datasets offer much more limited descriptions of soils [Hengl et al. 2017], permeability [Huscroft et al. 2018], and depth to bedrock [Shangguan et al. 2017] at resolutions of 250m to over 1km. For instance, we did attempt to include the unconsolidated material permeability parameters available in Huscroft et al. [2018], but while they were significantly correlated with K and S, they were not nearly as strongly correlated as similar variables available in SSURGO and ultimately did not meet our selection criteria for building our models. Extension of our methods to the global scale is therefore feasible, although much work would need to be done to constrain uncertainty.

Section 5: Conclusions

We estimated the effective hydraulic conductivity (K) and effective drainable storage (S) of all minimally impacted USGS-gauged watersheds in the continental US during periods of high S and low S. Using machine learning methods, we predict K and S for all watersheds in the continental US at the HUC-12 scale. We find that typical values of effective K are roughly 2 orders of magnitude higher than is predicted using soil texture data alone, highlighting the importance of soil structure and preferential flow pathways. Typical values of effective S are more than an order of magnitude lower than commonly used global maximum storage volumes. These data may be used to constrain regional hydrologic models and Land Surface Models to more hydrologically realistic parameterizations and to reduce model spin-up time.

REFERENCES

Allan, J. D., & Castillo, M. M. (2007). Stream ecology: structure and function of running waters. Springer Science & Business Media.

Alley, W. M., Healy, R. W., LaBaugh, J. W., & Reilly, T. E. (2002). Flow and storage in groundwater systems. science, 296(5575), 1985-1990.

Anderson Jr, WP, and RE Emanuel. (2008). Effect of interannual and interdecadal climate oscillations on groundwater in North Carolina, Geophysical Research Letters, 35(23).

Appleby, V. C. (1970). Recession and the baseflow problem, Water Resour. Res., 6(5), 1398-1403

Baker, I. T., Prihodko, L., Denning, A. S., Goulden, M., Miller, S., & da Rocha, H. R. (2008). Seasonal drought stress in the Amazon: Reconciling models and observations. Journal of Geophysical Research, 113, G00B01.

Barnes BS. 1939. The structure of discharge recession curves. Transactions of the American Geophysical Union 20: 721–725

Bart, R., & Hope, A. (2014). Inter-seasonal variability in baseflow recession rates: The role of aquifer antecedent storage in central California watersheds. Journal of Hydrology, 519, 205-213.

Bart, R. R., & Tague, C. L. (2017). The impact of wildfire on baseflow recession rates in California. Hydrological Processes, 31(8), 1662– 1673. https://doi-org.libproxy.lib.unc.edu/10.1002/hyp.11141

Bartlett, MS, and A Porporato. (2018). A Class of Exact Solutions of the Boussinesq Equation for Horizontal and Sloping Aquifers, Water Resources Research, 54:767-778.

Beck, H. E., van Dijk, A. I., Miralles, D. G., de Jeu, R. A., Bruijnzeel, L. S., McVicar, T. R., & Schellekens, J. (2013). Global patterns in base flow index and recession based on streamflow observations from 3394 catchments. Water Resources Research, 49(12), 7843-7863.

Berghuijs, W. R., Hartmann, A., & Woods, R. A. (2016). Streamflow sensitivity to water storage changes across Europe. Geophysical Research Letters, 43, 1980– 1987. https://doi-org.libproxy.lib.unc.edu/10.1002/2016GL067927

Beven, K., & Germann, P. (2013). Macropores and water flow in soils revisited. Water Resources Research, 49(6), 3071-3092.

Binley, A., Beven, K., & Elgy, J. (1989). A physically based model of heterogeneous hillslopes: 2. Effective hydraulic conductivities. Water Resources Research, 25(6), 1227-1233.

Biswal, B., & Marani, M. (2010). Geomorphological origin of recession curves. Geophysical Research Letters, 37(24).

Biswal, B., & Nagesh Kumar, D. (2014). Study of dynamic behaviour of recession curves. Hydrological Processes, 28(3), 784-792.

Botter, G., Basso, S., Rodriguez-Iturbe, I., & Rinaldo, A. (2013). Resilience of river flow regimes. Proceedings of the National Academy of Sciences, 110(32), 12925-12930.

Boulton, AJ. (2003). Parallels and contrasts in the effects of drought on stream macroinvertebrate assemblages. Freshwater Biology48, 1173–1185.

Boussinesq, J. (1877). Essai sur la théorie des eaux courantes. Impr. nationale.

Brodie, RS, and S Hostetler. (2005). A review of techniques for analyzing baseflow separation from stream hydrographs. Proceedings of the NZHS-IAH-NZSSS 2005 Conference, 28 November-2 December, 2005, Aukland, New Zeeland.

Brunke, M. A., Broxton, P., Pelletier, J., Gochis, D., Hazenberg, P., Lawrence, D. M., Leung, L. R., Niu, G., Troch, P. A., & Zeng, X. (2016). Implementing and evaluating variable soil thickness in the Community Land Model, version 4.5 (CLM4.5). Journal of Climate, 29, 3441–3461

Brutsaert, W., & Nieber, J. L. (1977). Regionalized drought flow hydrographs from a mature glaciated plateau. Water Resources Research, 13(3), 637-643. DOI: 10.1029/WR013i003p00637

Brutsaert, W. (1994). The unit response of groundwater outflow from a hillslope. Water Resources Research, 30(10), 2759-2763.

Brutsaert, W, and JP Lopez. (1998). Basin-scale geohydrologic drought flow features of riparian aquifers in the southern Great Plains, Water Resources Research, 34(2):233-240.

Carlier, C, SB Wirth, F Cochand, D Hunkeler, and P Brunner. (2018). Geology controls streamflow dynamics, Journal of Hydrology, 566, 756-769.

Clark, MP, HK McMillan, DBG Collins, D Kavetski, and RA Woods. (2011). Hydrological field data from a modeller's perspective: Part 2: process-based evaluation of model hypotheses, Hydrological Processes, 25:523-543.

Clark MP, Slater AG, Rupp DE, Woods RA, Vrugt JA, Gupta HV, Wagener T, Hay LE. (2008). Framework for Understanding Structural Errors (FUSE): a modular framework to diagnose differences between hydrological models. Water Resources Research: 44: W00B02, DOI: 10.1029/2007WR006735.

Clark MP, Rupp DE, Woods RA, Tromp-van Meerveld HJ, Peters NE, Freer JE. (2009). Consistency between hydrological models and field observations: linking processes at the hillslope scale to hydrological responses at the watershed scale. Hydrological Processes 23: 311–319. DOI: 10.1002/hyp.7154

Clark, M.P., Y. Fan, D.M. Lawrence, J.C. Adam, D. Bolster, D.J. Gochis, R.P. Hooper, M. Kumar, L.R. Leung, D.S. Mackay, R.M. Maxwell, C.Shen, S.C. Swenson, X. Zeng. (2015). Improving the representation of hydrologic processes in Earth System Models, Water Resources Research, 51(8), 5929-5956.

Charron, C, and UBMJ Ouarda. (2015). Regional low-flow frequency analysis with a recession parameter from a non-linear reservoir model, Journal of Hydrology, pp 468-475

Dai, Y., Shangguan, W., Wei, N., Xin, Q., Yuan, H., Zhang, S., ... & Yan, F. (2019). A review of the global soil property maps for Earth system models. Soil, 5(2), 137-158.

Dai, Y., Xin, Q., Wei, N., Zhang, Y., Shangguan, W., Yuan, H., ... & Lu, X. (2019). A Global High-Resolution Data Set of Soil Hydraulic and Thermal Properties for Land Surface Modeling. Journal of Advances in Modeling Earth Systems, 11(9), 2996-3023.

Domingos, P. (2012). A few useful things to know about machine learning. Communications of the ACM, 55(10), 78. https://doi.org/10.1145/2347736.2347755

Dralle, D, N Karst, and SE Thompson. (2015). a, b careful: The challenge of scale invariance for comparative analyses in power law models of the streamflow recession. Geophysical Research Letters, 42(21):9285-9293.

Dralle, D. N., Karst, N. J., Charalampous, K., Veenstra, A., & Thompson, S. E. (2017). Event-scale power law recession analysis: quantifying methodological uncertainty. Hydrology and Earth System Sciences, 21(1), 65-81.Dunford, E.G., and P.W. Fletcher. (1947). Effect of removal of stream-bank vegetation upon water yield, Transactions, American Geophysical Union, 28

Dupuit, J. (1863). Etudes Théoriques et Pratiques sur le mouvement des Eaux dans les canaux découverts et à travers les terrains perméables (Second ed.). Paris: Dunod.ld

Eckhardt, K. (2005). How to construct recursive digital filters for baseflow separation, Hydrol. Process.,19,507–515. doi:10.1002/hyp.5675

Ehalt Macedo, H., Beighley, R. E., David, C. H., & Reager, J. T. (2019). Using GRACE in a streamflow recession to determine drainable water storage in the Mississippi River basin. Hydrology & Earth System Sciences, 23(8).

Emanuel, R.E., H.E. Epstein, B.L. McGlynn, D.L. Welsch, D.J. Muth, and P. D'Odorico. (2010). Spatial and temporal controls on watershed ecohydrology in the northern Rocky Mountains, Water Resources Research, 46(11).

Falcone, J. A., Carlisle, D. M., Wolock, D. M., & Meador, M. R. (2010). GAGES: A stream gage database for evaluating natural and altered flow conditions in the conterminous United States: Ecological archives E091-045. Ecology, 91(2), 621-621.

Fan, Y. (2019). Are catchments leaky? Wiley Interdisciplinary Reviews: Water, 6(6), e1386.

Fan, Y, M Clark, DM Lawrence, S Swenson, LE Band, SL Brantley, PD Brooks, WE Dietrich, A Flores, G Grant, JW Kirchner, DS Mackay, JJ McDonnell, PCD Milly, PL Sullivan, C Tague, H Ajami, N Chaney, A Hartmann, P Hazenberg, J McNamara, J Pelletier, J Perket, E Rouholahnejad-Freund, T Wagener, X Zeng, E Beighley, J Buzan, M Huang, B Lizneh, BP Mohanty, B Nijssen, M Safeeq, C Shen, W van Verseveld, J Volk, and D Yamazaki. (2019). Hillslope hydrology in global change research and Earth system modeling. Water Resources Research 55(2): 1737-1772. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018WR023903

Fan, Y., Miguez-Macho, G., Jobbágy, E. G., Jackson, R. B., & Otero-Casal, C. (2017). Hydrologic regulation of plant rooting depth. Proceedings of the National Academy of Sciences of the United States of America, 114(40), 10,572–10,577

Fatichi, S., Or, D., Walko, R., Vereecken, H., Young, M. H., Ghezzehei, T. A., ... & Avissar, R. (2020). Soil structure is an important omission in Earth System Models. Nature Communications, 11(1), 1-11. https://www.nature.com/articles/s41467-020-14411-z

Ficklin, D. L., Robeson, S. M., & Knouft, J. H. (2016). Impacts of recent climate change on trends in baseflow and stormflow in United States watersheds. Geophysical Research Letters, 43(10), 5079-5088.

Fiori, A, V Cvetkovic, G Dagan, S Attinger, A Bellin, P Dietrich, A Zech, and G Teutsch. (2015). Debates -- Stochastic subsurface hydrology from theory to practice: The relevance of stochastic subsurface hydrology to practical problems of contaminant transport and remediation. What is characterization and stochastic theory good for? Water Resources Research, 52(12): 9228-9234

Freeze, R. A., and J. A. Cherry (1979), Groundwater, 604 pp., Prentice-Hall, Englewood Cliffs, N. J.

Gabrielli, CP, and JJ McDonnel. (2018). No Direct Linkage between Event-Based Runoff Generation and Groundwater Recharge on the MaiMai Hillslope, Water Resources Research, 54(11), 8718-8733

Gan, Y., Liang, X. Z., Duan, Q., Chen, F., Li, J., & Zhang, Y. (2019). Assessment and Reduction of the Physical Parameterization Uncertainty for Noah-MP Land Surface Model. Water Resources Research.

Gao, H, Q Tang, X Shi, C Zhu, T Bohn, F Su, M Pan, J Sheffield, D Lettenmaier, and E Wood. (2010). Water budget record from Variable Infiltration Capacity (VIC) model, 120-173.

Gao, M., X. Chen, J. Liu, Z. Zhang, and Q. Cheng. (2017). Using two parallel linear reservoirs to express multiple relations of power-law recession curves, J. Hdrol. Eng., 22(7).

Gedney, N., and P. M. Cox (2003), The sensitivity of global climate model simulations to the representation of soil moisture heterogeneity, J. Hydrometeorol., 4, 1265–1675.

Gelaro, R, W McCarty, MJ Suárez, R Todling, A Molod, L Takacs, CA Randles, A Darmenov, MG Bosilovich, R Reichle, and K Wargan. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). Journal of Climate, 30(14), 5419-5454.

Ghosh, DK, D Wang, and T Zhu. (2016). On the transition of base flow recession from early stage to late stage, Advances in Water Resources, 88:8-13.

Gleeson, T., Smith, L., Moosdorf, N., Hartmann, J., Dürr, H. H., Manning, A. H., ... & Jellinek, A. M. (2011). Mapping permeability over the surface of the Earth. Geophysical Research Letters, 38(2).

Gleeson, T., Moosdorf, N., Hartmann, J., & Van Beek, L. P. H. (2014). A glimpse beneath earth's surface: GLobal HYdrogeology MaPS (GLHYMPS) of permeability and porosity. Geophysical Research Letters, 41(11), 3891-3898. https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2014GL059856

Gleeson, T., Befus, K. M., Jasechko, S., Luijendijk, E., & Cardenas, M. B. (2016). The global volume and distribution of modern groundwater. Nature Geoscience, 9(2), 161.

Global Modeling and Assimilation Office (GMAO). (2008). tavgM_2d_mld_Nx: MERRA Simulated 2D Incremental Analysis Update (IAU) MERRA-Land reanalysis, GEOSldas-MERRALand, Time Average Monthly Mean V5.2.0, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: June 2019, 10.5067/K9PCGOMQ1XP1

Gnann, S. J., Woods, R. A., & Howden, N. J. (2019). Is there a baseflow Budyko curve?. Water Resources Research, 55(4), 2838-2855.

Goodfellow, B. W., Chadwick, O. A., Hilley, G. E., & Landforms (2014). Depth and character of rock weathering across a basaltic-hosted climosequence on Hawai ʻi. Earth Surface Processes and Landforms, 39(3), 381–398.

Grigg, AH, and JD Hughes. (2018). Nonstationarity driven by multidecadal change in catchment groundwater storage: A test of modifications to a common rainfall-runoff model, Water Resources Research, 32(24):3675-3688. https://onlinelibrary.wiley.com/doi/10.1002/hyp.13282

Guérin, A., Devauchelle, O., & Lajeunesse, E. (2014). Response of a laboratory aquifer to rainfall. Journal of Fluid Mechanics, 759.

Guerin, A, O Devauchelle, and E Lajeunesse. (2014). Response of a laboratory aquiver to rainfall, Journal of Fluid Mechanics, 795, R1.

Gudmundsson, L., Leonard, M., Do, H. X., Westra, S., & Seneviratne, S. I. (2019). Observed trends in global indicators of mean and extreme streamflow. Geophysical Research Letters, 46(2), 756-766.

Hall, F. R. (1968). Base-flow recessions—A review. Water resources research, 4(5), 973-983.

Harman, C. J., & Kim, M. (2019). A low-dimensional model of bedrock weathering and lateral flow coevolution in hillslopes: 1. Hydraulic theory of reactive transport. Hydrological processes, 33(4), 466-475.

Harman CJ, M Sivapalan, and P Kumar. (2009). Power law catchment-scale recessions arising from heterogeneous linear small-scale dynamics, Water Resources Research, 45(9), DOI:10.1029/2008WR007392.

Hartmann, J., & Moosdorf, N. (2012). The new global lithological map database GLiM: A representation of rock properties at the Earth surface. Geochemistry, Geophysics, Geosystems, 13(12).

He, S, S Li, R Xie, and J Lu. (2016). Baseflow separation based on a meteorology corrected nonlinear algorithm in a typical rainy agricultural watershed. J Hydrol., 535, 418-428

Heath, R.C., 1983. Basic ground-water hydrology, U.S. Geological Survey Water-Supply Paper 2220, 86p.

Hengl, T., de Jesus, J. M., Heuvelink, G. B., Gonzalez, M. R., Kilibarda, M., Blagotić, A., ... & Guevara, M. A. (2017). SoilGrids250m: Global gridded soil information based on machine learning. PLoS one, 12(2), e0169748.

Hewlett, JD, and AR Hibbert. (1963). Moisture and energy conditions within a sloping soil mass during drainage. Journal of geophysical research, 68(4), 1081-1087.

Hinzman, A. M., Sjöberg, Y., Lyon, S. W., Ploum, S. W., & van der Velde, Y. (2020). Increasing non-linearity of the storage-discharge relationship in sub-Arctic catchments. Hydrological Processes.

Hirmas, D. R., Giménez, D., Nemes, A., Kerry, R., Brunsell, N. A., & Wilson, C. J. (2018). Climate-induced changes in continental-scale soil macroporosity may intensify water cycle. Nature, 561(7721), 100.

Hunter, MA, T Quinn, and MP Hayes. (2005). Low flow spatial characteristics in forested headwater channels of southwest Washington, Journal of the American Water Resources Association, 41(3):503–516.

Hurd, B., Leary, N., Jones, R., & Smith, J. (1999). RELATIVE REGIONAL VULNERABILITY OF WATER RESOURCES TO CLIMATE CHANGE 1. JAWRA Journal of the American Water Resources Association, 35(6), 1399-1409.

Hurst, H. E. (1956). Methods of using long-term storage in reservoirs. Proceedings of the Institution of Civil Engineers, 5(5), 519-543.

Huscroft, J., Gleeson, T., Hartmann, J., & Börker, J. (2018). Compiling and mapping global permeability of the unconsolidated and consolidated Earth: GLobal HYdrogeology MaPS 2.0 (GLHYMPS 2.0). Geophysical Research Letters, 45(4), 1897-1904.

Jachens, E. R., Rupp, D. E., Roques, C., & Selker, J. S. (2019). Recession analysis 42 years later–work yet to be done. Hydrology and Earth System Sciences Discussions, 1-16.

Jepsen, S.M., T.C. Harmon, and Y. Shi. (2016). Watershed model calibration to the baseflow recession curve with and without evapotranspiration effects, Water Resources Research, 52, 2919-2933.

Kavetski, D, G Kuczera, and SW Franks. (2003). Semidistributed hydrological modeling: A "saturation path" perspective on TOPMODEL and VIC, Water Resources Research, 39(9).

Kirchner, JW. (2009). Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backwards. Water Resources Research, 45(2).

Koirala, S., Kim, H., Hirabayashi, Y., Kanae, S., & Oki, T. (2019). Sensitivity of global hydrological simulations to groundwater capillary flux parameterizations. Water Resources Research, 55(1), 402-425.

Krakauer, Nir Y., and Marouane Temimi. (2011). Stream recession curves and storage variability in small watersheds, Hydrology and Earth System Sciences 15(7):2377-2389.

Kuppel, S., Fan, Y., & Jobbagy, E. G. (2017). Seasonal hydrologic buffer on continents: Patterns, drivers and ecological benefits. Advances in Water Resources, 102, 178–187.

Li, W, K Zhang, Y Long, and L Feng. (2017). Estimation of Active Stream Network Length in a Hilly Headwater Catchment Using Recession Flow Analysis, Water, 9(5):348.

Liang, X., Xie, Z., & Huang, M. (2003). A new parameterization for surface and groundwater interactions and its impact on water budgets with the variable infiltration capacity (VIC) land surface model. Journal of Geophysical Research: Atmospheres, 108(D16).

Luo, Z, C Shen, J Kong, G Hua, X Gao, Z Zhao, H Zhao, and L Li. (2018). Effects of Unsaturated Flow on Hillslope Recession Characteristics, Water Resources Research, 54(3):2037-2056.

Ma, N., Niu, G. Y., Xia, Y., Cai, X., Zhang, Y., Ma, Y., & Fang, Y. (2017). A systematic evaluation of Noah-MP in simulating land-atmosphere energy, water, and carbon exchanges over the continental United States. Journal of Geophysical Research: Atmospheres, 122(22), 12-245.

Machlica, A, O Horvát, S Horáček, J Oosterwijk, AF Van Loon, M Fendeková, and HAJ Van Lanen. (2012). Influence of model structure on base flow estimation using Bilan, frier and HBV-light models, Journal of Hydrology and Hydromechanics, 60(4):242-251.

McMillan, HK, MP Clark, WB Bowden, M Duncan, and R Woods, 2011. Hydrological field data from a modeller's perspective: Part 1. Diagnostic tests for model structure, Hydrological Processes, 25, 511-522

Mendoza, GF, TS Steenhuis, MT Walter, and JY Parlange. (2003). Estimating basin-wide hydraulic parameters of a semi-arid mountainous watershed by recession-flow analysis, Journal of Hydrology, 279(1-4):57–69.

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., & Nauss, T. (2018). Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. Environmental modelling & software, 101, 1-9.

Meyer, H., Reudenbach, C., Wöllauer, S., & Nauss, T. (2019). Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. http://arxiv.org/abs/1908.07805

Miguez-Macho, G., Fan, Y., Weaver, C. P., Walko, R., & Robock, A. (2007). Incorporating water table dynamics in climate modeling: 2. Formulation, validation, and soil moisture simulation. Journal of Geophysical Research: Atmospheres, 112(D13).

Miguez-Macho, G., & Fan, Y. (2012a). The role of groundwater in the Amazon water cycle: 1. Influence on seasonal streamflow, flooding and wetlands. Journal of Geophysical Research, 117, D15113. https://doi.org/10.1029/2012JD017539

Miguez-Macho, G., & Fan, Y. (2012b). The role of groundwater in the Amazon water cycle: 2. Influence on seasonal soil moisture and evapotranspiration. Journal of Geophysical Research, 117, D15114. https://doi.org/10.1029/2012JD017540

Milly, P. C., & Shmakin, A. B. (2002). Global modeling of land water and energy balances. Part I: The land dynamics (lad) model. Journal of Hydrometeorology, 3, 283–299. https://doi.org/10.1175/1525-541(2002)003<0283:GMOLWA>2.0.CO;2

Mishra, A., Hata, T., Abdelhadi, A. W., Tada, A., & Tanakamaru, H. (2003). Recession flow analysis of the Blue Nile River. Hydrological Processes, 17(14), 2825-2835.

Mutzner, R., Bertuzzo, E., Tarolli, P., Weijs, S. V., Nicotina, L., Ceola, S., Tomasic, N., Rodríguez-Iturbe, I., Parlange, M. B., and Rinaldo, A. (2013). Geomorphic signatures on Brutsaert base flow recession analysis. Water Resources Research, 49(9), 5462-5472.

Nash, J. E., & HRS. (1960). A unit hydrograph study, with particular reference to British catchments. Proceedings of the Institution of Civil Engineers, 17(3), 249-282.

Neitsch, S.L., J.G. Arnold, J.R. Kiniry, J.R. Williams. (2005). Soil and Water Assessment Tool Theoretical Documentation, Version 2000, Grassland, soil and water research service, Temple, TX.

Nieber JL, Sidle RC. 2010. How do disconnected macropores in sloping soils facilitate preferential ow? Hydrological Processes 24: 1582–1594.

Nippgen, F., B.L. McGlynn, R.E. Emanuel, and J.M. Vose. (2016) Watershed memory at the Coweeta Hydrologic Laboratory: The effect of past precipitation and storage on hydrologic response, Water Resources Research, 52, 1673-1695

Niu, G. Y., Yang, Z. L., Dickinson, R. E., & Gulden, L. E. (2005). A simple TOPMODEL-based runoff parameterization (SIMTOP) for use in global climate models. Journal of Geophysical Research: Atmospheres, 110(D21).

Niu, G. Y., Yang, Z. L., Dickinson, R. E., Gulden, L. E., & Su, H. (2007). Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data. Journal of Geophysical Research: Atmospheres, 112(D7).

Niu, GY, ZL Yang, KE Mitchell, F Chen, MB Ek, M Barlage, A Kumar, K Manning, D Niyogi, E Rosero, and M Tewari. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, Journal of Geophysical Research: Atmospheres, 116(D12).

NOAA (2016), National Water Model: Improving NOAA's Water Prediction Services, http://water.noaa.gov/documents/wrn-national-water-model.pdf. 2 pp.

Oleson, K. W., Lawrence, D. M., Bonan, G. B., Flanner, M. G., Kluzek, E., Lawrence, P. J., ... & Decker, M. (2010). Technical description of version 4.5 of the Community Land Model (CLM), NCAR Tech. Notes (NCAR/TN-478+ STR).

Palmroth S, GG Katul, D Hui, HR McCarthy, RB Jackson, and R Oren. (2010). Estimation of long-term basin scale evapotranspiration from streamflow time series, Water Resour Res, 46(10), W10512, 10.1029/2009WR008838.

Parlange, J.Y., Stagnitti, F., Heilig, A., Szilagyi, J., Parlange, M.B., Steenhuis, T.S., Hogarth, W.L., Barry, D.A. and Li, L. (2001). Sudden drawdown and drainage of a horizontal aquifer. Water resources research, 37(8), pp.2097-2101.

Patnaik, S., Biswal, B., Kumar, D. N., and Sivakumar, B. (2015). Effect of catchment characteristics on the relationship between past discharge and the power law recession coefficient, J. Hydrol., 528, 321-328.

Patnaik, S., B. Biswal, D.N. Kumar, and B. Sivakumar. (2018). Regional Variation of recession flow power-law exponent, Hydrological Processes

Pauritsch, M., S Birk, T Wagner, S Hergarten, and G Winkler. (2015) Analytical approximations of discharge recessions for steeply sloping aquifers in alpine catchments, Water Resour. Res., 51, 8729–8740.

Pauwels VRN, and PA Troch. (2010). Estimation of aquifer lower layer hydraulic conductivity values through base flow hydrograph rising limb analysis, Water Resources Research 46(3), DOI: 10.1029/2009WR008255.

Pauwels, VRN, and R Uijlenhoet. (2018). Confirmation of a Short-Time Expression for the Hydrograph Rising Limb of an Initially Dry Aquifer Using Laboratory Hillslope Outflow experiments, Water Resources Research, 54(12).

Payn, R.A., Gooseff, M.N., McGlynn, B.L., Bencala, K.E., Wondzell, S.M. (2012(. Exploring changes in the spatial distribution of stream baseflow generation during a seasonal recession. Water Resour. Res. 48, W04519. http://dx.doi.org/10.1029/2011WR011552.

Ploum, S. W., Lyon, S. W., Teuling, A. J., Laudon, H., & van der Velde, Y. (2019). Soil frost effects on streamflow recessions in a subarctic catchment. Hydrological Processes, 33(9), 1304–1316. https://doi.org/10.1002/hyp.13401

Powell, J. W., Gilbert, G. K., Dutton, C. E., Thompson, A. H., & Drummond, W. (1879). Report on the Lands of the Arid Region of the United States: With a More Detailed Account of the Lands of Utah. With Maps. US Government Printing Office.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rice, J.S., and R.E. Emanuel. (2017). How are streamflow responses to the El Nino Southern Oscillation affected by watershed characteristics? Water Resources Research, 53, 4393-4406

Rind, D, R Goldberg, J Hansen, C Rosenzweig, and R Ruedy. (1990). Potential evapotranspiration and the likelihood of future drought, Journal of Geophysical Research: Atmospheres, 95(D7): 9983-10004.

Roques, C., Rupp, D. E., & Selker, J. S. (2017). Improved streamflow recession parameter estimation with attention to calculation of– dQ/dt. Advances in water resources, 108, 29-43.

Royston, P. (1995). A Remark on Algorithm AS 181: The W-test for Normality, Journal of the Royal Statistical Society. Series C (Applied Statistics), 44(4): 547-551.

Rupp, D.E.; Selker, J.S. (2006). On the use of the Boussinesq equation for interpreting recession hydrographs from sloping aquifers. Water Resour. Res. 2006, 42

Sanchez-Murillo, R, ES Brooks, ES, WJ Elliot, W Gazel, and J Boll. (2015). Baseflow recession analysis in the inland Pacific Northwest of the United States, Hydrogeology Journal, 23(2):287-303.

Santhi, C., Allen, P. M., Muttiah, R. S., Arnold, J. G., & Tuppad, P. (2008). Regional estimation of base flow for the conterminous United States by hydrologic landscape regions. Journal of Hydrology, 351(1-2), 139-153.

Santos, AC, MM Portela, A Rinaldo, and B Schaefli. (2019). Estimation of streamflow recession parameters: New insights from an analytic streamflow distribution model, HP https://onlinelibrary.wiley.com/doi/10.1002/hyp.13425

Sawaske, SR, and DL Freyberg. (2014). An analysis of trends in baseflow recession and low-flows in rain-dominated coastal streams of the pacific coast, Journal of Hydrology, 519, 599-610

Shangguan, W., Hengl, T., Mendes de Jesus, J., Yuan, H. & Dai, Y. Mapping the global depth to bedrock for land surface modeling. J. Adv. Model. Earth Syst. 9, 65–88 (2017).

Shaw, S. B. (2016). Investigating the linkage between streamflow recession rates and channel network contraction in a mesoscale catchment in New York state. Hydrological Processes, 30, 479– 492. https://doi-org.libproxy.lib.unc.edu/10.1002/hyp.10626

Shaw, SB, and SJ Riha. (2012). Examining individual recession events instead of a data cloud: Using a modified interpretation of dQ/dt-Q streamflow recession in glaciated watersheds to better inform models of low flow, Journal of Hydrology, 434:46-54.

Singh, N.K., R.E. Emanuel, and B.L. McGlynn. (2016). Variability in isotopic composition of base flow in two headwater streams of the southern Appalachians, Water Resources Research, 52(6): 4264-4279

Spence C, Guan XJ, Phillips R, Hedstrom N, Granger R, Reid B. 2010. Storage dynamics and streamï¬‚ow in a catchment with a variable contributing area. Hydrological Processes 24: 2209–2221.

Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Soil Survey Geographic (SSURGO) Database for continental United States. Available online. Accessed 12/05/2019.

Stewart, MK. (2016. Promising new baseflow separation and recession analysis methods applied to streamflow at Glendhu Catchment, New Zealand, Hydrology and Earth System Sciences, 19:2587-2603.

Stoelzle, M, K Stahl, and M Weiler. (2013). Are streamflow recession characteristics really characteristic? Hydrology and Earth Systems Sciences, 17:817-828.

Szilagyi, J, Z Gribovszki, and P Kalicz. (2007). Estimation of catchment-scale evapotranspiration from base flow recession data: Numerical model and practical application results, Journal of Hydrology, 336:206-217, DOI:10.1016/j.jhydrol.2007.01.004.

Szilagyi, J., M.B. Parlange, and J.D. Albertson. (1998). Recession flow analysis for aquifer parameter determination, Water Resources Research, 34(7), 1851-1857.

Tallaksen, L. M. (1995). A review of baseflow recession analysis. Journal of hydrology, 165(1-4), 349-370.

Team, R. C. (2013). R: A language and environment for statistical computing.

Tashie, A., Scaife, C. I., & Band, L. E. (2019). Transpiration and subsurface controls of streamflow recession characteristics. Hydrological Processes, 33(19), 2561-2575.

Tashie, A., Pavelsky, T., & Band, L. E. (2020a). An Empirical Reevaluation of Streamflow Recession Analysis at the Continental Scale. Water Resources Research, e2019WR025448.

Tashie, A., Pavelsky, T., & Emanuel, R. E. (2020b). Spatial and temporal patterns in baseflow recession in the continental United States. Water Resources Research, 55, 2.

Tetzlaff D, Seibert J, McGuire KJ, Laudon H, Burns DA, Dunn SM, Soulsby C. 2009. How does landscape structure inﬂuence catchment transit times across different geomorphic provinces? Hydrological Processes 23: 945–953.

Thomas, BF, RM Vogel, CN Kroll, and JS Famiglietti, (2013). Estimation of the base flow recession constant under human interference, Water Resources Research, 49(11):7366-7379.

Thomas, BF, Vogel, RM, and Famiglietti, JS (2015). Objective hydrograph baseflow recession analysis, Journal of Hydrology, 525:102-112.

Toth, J. (1963). A theoretical analysis of groundwater flow in small drainage basins. Journal of geophysical research, 68(16), 4795-4812.

Troch, PA, A Berne, P Bogaart, C Harman, AGJ Hilberts, SW Lyon, C Paniconi, VRN. Pauwels, DE Rupp, JS Selker, AJ Teuling, R Uijlenhoet, and NEC Verhoest. (2013). The importance of hydraulic groundwater theory in catchment hydrology: The legacy of Wilfried Brutsaert and Jean-Yves Parlange, Water resources Research, 49(9):5099-5116.

Troch, PA, FP De Troch, and W Brutsaert. (1993). Effective water table depth to describe initial conditions prior to storm rainfall in humid regions, Water Resources Research, 29(2):427–434.

Tromp-van Meerveld HJ, McDonnell JJ. (2006). Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis. Water Resources Research 42: W02411. DOI: 10.1029/2004WR003800.

Tromp-van Meerveld, H. J., & McDonnell, J. J. (2006). Threshold relations in subsurface stormflow: 1. A 147-storm analysis of the Panola hillslope. Water Resources Research, 42, W02410.

Van Dijk, A. I. J. M. (2010), Climate and terrain factors explaining streamflow response and recession in Australian catchments, Hydrol. Earth Syst. Sci., 14, 159–169.

Vannier, O., Braud, I., Anquetin, S. (2014). Regional estimation of catchment-scale soil properties by means of streamflow recession analysis for use in distributed hydrological models. Hydrol. Process. 2014, 28, 6276–6291.

Vogel, RM, and CN Kroll. (1992). Regional geohydrologic-geomorphic relationships for the estimation of low-flow statistics, Water Resources Research, 38(9):2451–2458.

Wang D. (2011). On the base flow recession at the Panola Mountain Research Watershed, Georgia, United States, Water Resources Research 47 (3)

Wang, D., & Alimohammadi, N. (2012). Responses of annual runoff, evaporation, and storage change to climate variability at the watershed scale, Water Resources Research, 48(5).

Wang, D., & Cai, X. (2010). Recession slope curve analysis under human interferences. Advances in Water Resources, 33, 1053–1061. https://doi-org.libproxy.lib.unc.edu/10.1016/j.advwa

Western, W., and T. A. McMahon (2000), Extension of unimpaired monthly streamflow data and regionalisation of parameter values to estimate streamflow in ungauged catchments, Rep. prepared for the Australian National Land and Water Resources Audit, Cent. for Environ. Appl. Hydrol., Univ. of Melbourne, Parkville, Victoria.

Westerberg, IK, and HK McMillan. (2015). Uncertainty in hydrological signatures, Hydrology and Earth System Sciences, 19:3951-3968.

Wieczorek, M. E., Jackson, S. E., & Schwarz, G. 2. (2018). Select attributes for NHDPlus Version 2.1 reach catchments and modified network routed upstream watersheds for the conterminous United States. Reston, VA: US Geological Survey.

Wiese, D. N., D.-N. Yuan, C. Boening, F. W. Landerer, M. M. Watkins. (2018). JPL GRACE Mascon Ocean, Ice, and Hydrology Equivalent Water Height Release 06 Coastal Resolution Improvement (CRI) Filtered Version 1.0. Ver. 1.0. PO.DAAC, CA, USA. Dataset accessed 2019-05-01 at http://dx.doi.org/10.5067/TEMSC-3MJC6.

Winkler, G., Wagner, T., Pauritsch, M., Birk, S., Kellerer-Pirklbauer, A., Benischke, R., ... & Hergarten, S. (2016). Identification and assessment of groundwater flow and storage components of the relict Schöneben Rock Glacier, Niedere Tauern Range, Eastern Alps (Austria). Hydrogeology journal, 24(4), 937-953.

Winter, T. C., Harvey, J. W., Franke, O. L., & Alley, W. A. (1998). Ground water and surface water: A single resource. U.S. Geological Survey Circular, 1139, 79.

Winter, T. C. (2001). The concept of hydrologic landscapes 1. JAWRA Journal of the American Water Resources Association, 37(2), 335-349.Wittenberg, H. (1994) Nonlinear analysis of low flow recession curves. FRIENDS: Flow Regimes from International and Experimental Network Data. IAHS Publ. No 221:61-67.

Wittenberg, H. (1999). Baseflow recession and recharge as nonlinear storage processes, Hydrological Processes, 13(5):715-726.

Wittenberg, H. (2003). Effects of season and man-made changes on baseflow and recession: case studies, Hydrological Processes, 17, 2113-2123.

Wittenberg, H, and M Sivapalan. (1999). Watershed groundwater balance estimation using streamflow recession analysis and baseflow separation, Journal of Hydrology 219:20-33.

Wolock, David M. Hydrologic landscape regions of the United States. No. 2003-145. US Geological Service, 2003.

Wondzell, SM, MN Gooseff, and BL McGlynn. (2007). Flow velocity and the hydrologic behavior of streams during baseflow, Geophysical Research Letters, 34(24).

Yang, Z.-L., R. E. Dickinson, A. Henderson-Sellers, and A. J. Pitman (1995), Preliminary study of spin-up processes in landsurface models with the first stage data of Project for Intercomparison of Land Surface Parameterization Schemes Phase 1(a), J. Geophys. Res., 100(D8), 16,553–16,578.

Yang, Z.-L., Cai, X., Zhang, G., Tavakoly, A. A., Jin, Q., Meyer, L. H., & Guan, X. (2011). The community Noah land surface model with multi-parameterization options (Noah-MP). Austin, TX: The University of Texas at Austin.

Ye, S, H Li, M Huang, M Ali, G Leng, LR Leung, S Wang, and M Sivapalan. (2014). Regionalization of subsurface stormflow parameters of hydrologic models: Derivation from regional analysis of streamflow recession curves, Journal of Hydrology, 519:670–682.

Zecharias, Y. B., & Brutsaert, W. (1988). Recession characteristics of groundwater outflow and base flow from mountainous watersheds. Water Resources Research, 24(10), 1651-1658.

Zhang, Y., & Schaap, M. G. (2019). Estimation of saturated hydraulic conductivity with pedotransfer functions: A review. Journal of Hydrology, 575, 1011–1030.

Zhi, Wei, Li Li, Wenming Dong, Wendy Brown, Jason Kaye, Carl Steefel, and Kenneth H. Williams. (2019). Distinct Source Water Chemistry Shapes Contrasting Concentration-Discharge Patterns. Water Resources Research. https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2018WR024257