

## RESEARCH LETTER

10.1002/2017GL074139

## Key Points:

- New method predicts standard deviation of annual/seasonal hydrologic variables
- Stochastic daily flow process successfully scales to describe seasonal and annual variables
- Two case studies are presented: Seasonal flow totals and net seasonal suspended sediment export

## Supporting Information:

- Supporting Information S1

## Correspondence to:

D. Dralle,  
dralle@berkeley.edu

## Citation:

Dralle, D., N. Karst, M. Müller, G. Vico, and S. E. Thompson (2017), Stochastic modeling of interannual variation of hydrologic variables, *Geophys. Res. Lett.*, 44, doi:10.1002/2017GL074139.

Received 12 MAY 2017

Accepted 7 JUL 2017

Accepted article online 12 JUL 2017

## Stochastic modeling of interannual variation of hydrologic variables

David Dralle<sup>1</sup> , Nathaniel Karst<sup>2</sup>, Marc Müller<sup>3</sup>, Giulia Vico<sup>4</sup> , and Sally E. Thompson<sup>5</sup> 

<sup>1</sup>Department of Earth and Planetary Science, University of California, Berkeley, California, USA, <sup>2</sup>Mathematics and Science Division, Babson College, Wellesley, Massachusetts, USA, <sup>3</sup>Department of Civil and Environmental Engineering and Earth Sciences, University of Notre Dame, Notre Dame, Indiana, USA, <sup>4</sup>Department of Crop Production Ecology, Swedish University of Agricultural Sciences, Uppsala, Sweden, <sup>5</sup>Department of Civil and Environmental Engineering, University of California, Berkeley, California, USA

**Abstract** Quantifying the interannual variability of hydrologic variables (such as annual flow volumes, and solute or sediment loads) is a central challenge in hydrologic modeling. Annual or seasonal hydrologic variables are themselves the integral of instantaneous variations and can be well approximated as an aggregate sum of the daily variable. Process-based, probabilistic techniques are available to describe the stochastic structure of daily flow, yet estimating interannual variations in the corresponding aggregated variable requires consideration of the autocorrelation structure of the flow time series. Here we present a method based on a probabilistic streamflow description to obtain the interannual variability of flow-derived variables. The results provide insight into the mechanistic genesis of interannual variability of hydrologic processes. Such clarification can assist in the characterization of ecosystem risk and uncertainty in water resources management. We demonstrate two applications, one quantifying seasonal flow variability and the other quantifying net suspended sediment export.

**Plain Language Summary** We present a method to predict interannual/interseasonal streamflow variation. Predicting such variation is a long-standing target of hydrological modeling, yet remains challenging to achieve using existing methods, which are generally derived from purely meteorological data. Here we develop an alternative approach that accounts for the random occurrence of rainfall and the process by which rainfall is transformed (via infiltration into the soil and eventual drainage through a shallow groundwater system) into stream discharge. By summing over these variations, the method addresses a major source of interannual variability, arising from the stochasticity of precipitation. The proposed technique can also be generalized to predict interannual variability of physical processes that depend on flow (e.g., suspended sediment export), meaning that the method has broad applicability across the geosciences.

## 1. Introduction

Skillful prediction of the interannual variability of hydrologic processes is valuable to inform flow forecasts [Farmer *et al.*, 2003; Bárdossy, 2007], to manage the design of infrastructure (e.g., reservoir sizing), and to characterize and protect stream ecosystem function [Löff and Hardison, 1966; Poff *et al.*, 1997; Vogel and Wilson, 1996; Dettinger and Diaz, 2000; Biggs *et al.*, 2005; Monk *et al.*, 2008; Blöschl, 2013]. Variability in annual runoff determines the uncertainty associated with water availability in given river systems [McMahon *et al.*, 2013], with implications for hydropower production [Castellarin *et al.*, 2004], drinking water systems [Löff and Hardison, 1966], ecosystem health [Poff *et al.*, 1997], landscape evolution [Rossi *et al.*, 2016], and environmental quality in receiving waters [Anderson *et al.*, 2005; Ahearn *et al.*, 2005; Morehead *et al.*, 2003]. Understanding the connection between annual (or seasonal) flow variations and the statistical properties of climate as mediated by the physical properties of a catchment remains a fundamental and incompletely realized goal of hydrological prediction [McMahon *et al.*, 2007].

Although stream discharge varies continuously in time, most discharge records report daily mean flows, and a wide variety of hydrologic models focus on the prediction of daily flow values. For example, simple modeling approaches from stochastic hydrology [Botter *et al.*, 2007; Müller *et al.*, 2014; Basso *et al.*, 2015, 2016]

can often skillfully predict daily discharge variability in natural catchments using a constrained set of predictors derived from rainfall climatology, storage capacity of the unsaturated zone, vegetation water use, and the drainage characteristics of the catchment groundwater system. Subsequent extensions of this framework have been used to study run-of-river hydropower generation [Müller *et al.*, 2016; Lazzaro and Botter, 2015], flood risk assessment [Basso *et al.*, 2016], river macroinvertebrate habitat extent [Ceola *et al.*, 2014], and riparian vegetation extent [Doulatyari *et al.*, 2014], among other applications across the geological and ecological sciences. With such a model for daily flows, climatic and catchment-scale properties could be linked to annual measures of flow variability by computing seasonal or annual sums of the (predicted) daily flow variable. For instance, annual flow volumes ( $Q_{AF}$  (L)) can be computed as

$$Q_{AF} = \sum_{i=1}^T Q_i \Delta t, \quad (1)$$

where  $Q_i$  ( $L T^{-1}$ ) is the value of daily mean discharge on day  $i$ ,  $T$  is the number of days over which the sum is taken, and  $\Delta t = 1$  day. When annual or seasonal properties of flow-derived variables ( $X[*]$ , unspecified units) are of interest, equation (1) can be generalized as

$$X = \sum_{i=1}^T f[Q_i] \Delta t. \quad (2)$$

The function  $f[* \cdot T^{-1}]$  describes the transformation from flow to the variable of interest, at the daily time scale. For example, suspended river sediment discharge mass flux,  $L [MT^{-1}]$  (or concentration, which can be converted to a mass flux by multiplication with  $Q$ ), is commonly described through the sediment rating curve as a power law function of daily flow,  $L = f(Q) = \beta Q^\delta$  [Runkel *et al.*, 2004]. In this case, seasonal or annual suspended sediment export ( $L_{net}$ ) can be derived from the sum of the transformed daily flow variable as

$$L_{net} = \sum_{i=1}^T \beta Q_i^\delta \Delta t. \quad (3)$$

Variability of such seasonal or annual measures of flow (or a flow-derived process) can be parsimoniously described by the measure's standard deviation ( $\sigma_x$ ):

$$\sigma_x = \sqrt{\text{Var}[X]} = \sqrt{\text{Var} \left[ \sum_{i=1}^T f[Q_i] \right]}. \quad (4)$$

Reliable estimates of  $\sigma_x$  are often difficult to make because extensive empirical records are required to ensure representative sampling of interannual variability in time. Such lengthy records of flow are available in relatively few basins worldwide and fewer still in the case of water quality data such as suspended sediment [Grabs, 2009; Sene and Farquharson, 1998]. If, however, interannual variation arises primarily from the year-to-year realization of stationary daily climate statistics (e.g., mean rainfall depth), it would be appealing to estimate seasonal or interannual variations from the known probability distribution of daily streamflows. Upscaling daily variability to seasonal or greater time scales in natural basins, however, requires accounting for the autocorrelated structure of the underlying daily flow generation process. For a stationary time series that exhibits autocorrelation, calculating the standard deviation of the aggregated variable requires estimating the full autocorrelation function for lags from 1 to  $T - 1$ , where  $T$  is the length of the summation period:

$$\sigma_x = \sqrt{\text{Var} \left[ \sum_{i=1}^T f[Q_i] \right]} = \sqrt{\sum_{i=1}^T \sum_{j=1}^T \text{Cov} (f[Q_i], f[Q_j])} = \sqrt{\sum_{i=1}^T \sum_{j=1}^T \sigma_Q^2 \rho (f[Q_i], f[Q_j])}. \quad (5)$$

Here  $\sigma_Q^2$  is the variance of the (assumed stationary) daily flow random variable, and  $\text{Cov}$  and  $\rho$  are its covariance and autocorrelation functions, respectively. Failing to properly account for the serially correlated nature of the hydrograph would underestimate the spread of the aggregated distribution ( $\sigma_x$ ) because all the  $j \neq i$  terms of the summation in equation (5) would be omitted.

The aim of this study is to use a predicted probability distribution of daily flow [Botter *et al.*, 2007; Basso *et al.*, 2015, 2016; Müller *et al.*, 2014] to derive the covariance function of the daily discharge or discharge-derived variable, and thus the standard deviation of the associated integrated annual or seasonal variable. A similar framework was proposed by Zanardo *et al.* [2012] to examine interannual variations of the Budyko evaporation index as a function of the interannual rainfall variability. However, Zanardo *et al.* [2012] did not address serial

correlation of the hydrograph, thus limiting generality of the analysis. Additionally, because many families of probability distributions are uniquely defined by their mean and standard deviation, we also demonstrate how “moment matching” can be used to obtain approximate probability distributions of seasonal and annual variables [e.g., *Jiang et al.*, 2017]. These distributions could provide relevant management-oriented information, such as estimates of annual flow return intervals. In particular, we focus on the application of the two-parameter gamma distribution, which has been used to successfully represent distributions of sums of environmental variables [*Kotz and Neumann*, 1963], including annual and seasonal total flow volumes [e.g., *McMahon et al.*, 2007; *Yu et al.*, 2015]. Two case studies demonstrate the utility of the method for diverse applications: Case Study 1 examines interannual variation of seasonal flow totals for seven watersheds (and multiple seasons) across the continental United States and Puerto Rico, and Case Study 2 uses the model to predict net seasonal export of suspended sediment in four U.S. rivers.

## 2. Methods

### 2.1. Daily Streamflow Model

Catchment discharge behavior is specified according to the stochastic streamflow model developed by *Botter et al.* [2007]. This approach assumes that a watershed can be modeled as a linear reservoir with area-normalized discharge  $Q$  [ $L/T$ ] and response time scale  $1/k$  [ $T$ ], resulting in an exponential catchment recession governed by  $dQ/dt = -kQ$ . *Botter et al.* [2007] model the sequence of recharge events that generate streamflow as a marked Poisson process with mean interarrival time  $1/\lambda$  [ $T$ ], resulting in exponentially distributed discharge increments with mean  $1/\gamma_Q$  [ $L/T$ ]. The timing and volume of recharge/discharge events depend upon rainfall climatology, the storage capacity of the unsaturated zone, and vegetation water use. Accordingly, the parameters  $\lambda$  and  $\gamma_Q$  are determined from catchment rainfall features, catchment soil textural properties, rooting depth, and vegetation water demand [*Milly*, 1993; *Porporato et al.*, 2004; *Botter et al.*, 2007]. It is assumed that these parameters can be taken as constant for the seasonal or annual period of interest. With this set of assumptions, *Botter et al.* [2007] show that the steady state daily catchment discharge follows a two-parameter gamma distribution with a probability density function given by

$$p_Q(q) = \frac{\gamma_Q^m}{\Gamma(m)} q^{m-1} \exp(-\gamma_Q q), \quad (6)$$

with  $m = \lambda/k$ .

### 2.2. Standard Deviation of a Sum

In general, the covariance function from equation (5) can be computed as

$$\text{Cov}(f[Q_i], f[Q_j]) = \int_0^\infty \int_0^\infty (f(q_i) - \mu_f)(f(q_j) - \mu_f) p_{Q_i, Q_j}(q_i, q_j) dq_i dq_j, \quad (7)$$

where  $p_{Q_i, Q_j}(q_i, q_j)$  is the joint probability density function for discharge on days  $i$  and  $j$  and where  $\mu_f$  is the daily mean of  $f(Q)$ :

$$\mu_f = \int_0^\infty f(q) p_Q(q) dq, \quad (8)$$

with  $p_Q$  given by equation (6).

The joint distribution  $p_{Q_i, Q_j}(q_i, q_j)$  can be decomposed into the product of a conditional distribution and the marginal distribution:  $p_{Q_i, Q_j}(q_i, q_j) = p_{Q_i|Q_j}(q_i|q_j) p_Q(q_j)$ , where again  $p_Q$  comes from equation (6). An analytical expression for  $p_{Q_i|Q_j}(q_i, q_j)$  was obtained by *Viola et al.* [2008] (for the case of soil moisture) using the same assumptions about reservoir linearity and rainfall statistics as *Botter et al.* [2007]:

$$p_{Q_i|Q_j}(q_i|q_j) = e^{-\lambda\tau} \delta(q_i - q_j e^{-k\tau}) + e^{-\lambda\tau - \gamma(q_i - q_j e^{-k\tau})} \frac{\lambda\gamma}{k} \cdot (e^{k\tau} - 1) {}_1F_1[1 - \lambda/k, 2, \gamma(q_i - q_j e^{-k\tau})(1 - e^{k\tau})], \quad (9)$$

where  $i \geq j$ ,  $\tau = i - j$ ,  $\delta(*)$  is the Dirac delta function, and  ${}_1F_1[*, *, *]$  is the Kummer confluent hypergeometric function [*Abramowitz and Stegun*, 1965]. Equation (9) is a mixed type distribution, including both a discrete atom of probability at  $q_i = q_j e^{-k\tau}$  and a continuous distribution on the interval  $q_i \in [q_j e^{-k\tau}, \infty)$ . The atom represents the probability that a recharge event has not occurred between day  $i$  and day  $j$  (with  $j - i = \tau$ ), which happens with probability  $e^{-\lambda\tau}$ , according to the properties of the Poisson recharge process. Conversely, the continuous portion of the distribution represents the possible values of  $Q_i$  if one or more recharge events

occur between  $i$  and  $j$ , which happens with probability  $1 - e^{-\lambda\tau}$ . It is known (and we ourselves have found) that the hypergeometric function in equation (9) is numerically unstable for large values of  $\tau$  [Muller, 2001]. The instability often leads to divergence of the covariance function between daily flows, making it difficult to confidently predict the standard deviation of the aggregated variable. We therefore approximate the continuous portion of  $p_{Q_i|Q_j}(q_i, q_j)$  with a more stable, shifted gamma distribution that is parameterized to match the mean ( $M$ ) and variance ( $V$ ) of equation (9). This shifted gamma distribution takes the form

$$\tilde{p}_{Q_i|Q_j}(q_i|q_j) = e^{-\lambda\tau} \delta(q_i - q_j e^{-k\tau}) + \left( \frac{1 - e^{-\lambda\tau}}{(q_i - q_j e^{-k\tau})^{1-a}} \right) \frac{b^{-a}}{\Gamma(a)} \exp\left(-\frac{q_i - q_j e^{-k\tau}}{b}\right). \quad (10)$$

The parameters of this gamma distribution ( $a$  and  $b$ ) will be determined from  $M$  and  $V$ , which can be obtained from the moment generating function (where  $u$  is the Laplace variable) of equation (9), also presented in Viola *et al.* [2008]:

$$p_{Q_i|Q_j}^*(u, \tau, q_j) = e^{q_j u e^{-k\tau}} \left( \frac{u e^{-k\tau} + \gamma}{\gamma + u} \right)^{\frac{\lambda}{k}}. \quad (11)$$

The properties of the moment generating function give  $M = \left. \frac{dp_{Q_i|Q_j}^*}{du} \right|_{u=0}$ , and  $V = \left. \frac{d^2 p_{Q_i|Q_j}^*}{du^2} \right|_{u=0}$ . We compute expressions for the mean and variance of the approximate distribution, set them equal to the mean ( $M$ ) and variance ( $V$ ) of the exact distribution, and solve for the parameters of the shifted gamma distribution ( $a$  and  $b$ ); the solution is

$$a = -\frac{(M - q_j e^{-k\tau})^2}{e^{-\lambda\tau} (M - q_j e^{-k\tau})^2 + V(e^{-\lambda\tau} - 1)} \quad (12)$$

$$b = \frac{e^{\lambda\tau} (M - q_j e^{-k\tau})}{e^{-\lambda\tau} - 1} - \frac{V}{q_j e^{-k\tau} - M}. \quad (13)$$

In the supporting information (Figure S4), we present an analysis that demonstrates that the approximation is extremely accurate with error that is nearly immeasurable across a wide range of lags ( $\tau$ ) and model parameterizations. With this approximation, the covariance function is computed by substituting  $\tilde{p}_{Q_i|Q_j}$  into equation (7):

$$\text{Cov}(f[Q_i], f[Q_j]) = \int_0^\infty \int_0^\infty (f(q_i) - \mu_f)(f(q_j) - \mu_f) \tilde{p}_{Q_i|Q_j}(q_i|q_j) p_Q(q_j) dq_i dq_j. \quad (14)$$

We note that while we use equation (14) to compute the covariance, a mathematically equivalent analytical expression for the covariance is already available for the special case of  $f(Q) = Q$  used in Case Study 1 [Muneepeerakul *et al.*, 2010]:

$$\text{Cov}(Q_i, Q_j) = \sigma_Q^2 e^{-|i-j|}, \quad (15)$$

where  $\sigma_Q$  is the standard deviation of the daily flow variable. Nevertheless, such an expression has not been derived generally for nonlinear  $f$ , as required here.

Finally, we determine  $\sigma_x$  by substituting equation (14) into equation (5), where the double sum in equation (5) is obtained by numerical integration of equation (14) for all lags  $|i-j| \in [1, T-1]$  with  $i, j \in [1, T]$  and use  $\sigma_x$  and  $\mu_x$  to moment match and approximate  $p_x(x)$ .

To elucidate the details of the above methods, we created a well-commented Python (<https://www.python.org/>) code in a Jupyter Notebook (<http://jupyter.org/>) that computes the standard deviation for any power law function of streamflow,  $f(Q)$  (defined by a user-specified  $\beta$  and  $\delta$ ), given the model parameters  $\gamma_Q$ ,  $\lambda$ ,  $k$ , and  $T$ . This notebook can be found at <https://github.com/daviddralle/sdSum>.

### 2.3. Case Studies

Two case studies are considered: predicting the standard deviation and PDF of seasonal flow totals and predicting the standard deviation and PDF of seasonal sediment export. Both case studies require specification of a summation duration ( $T$ ) and the three parameters of the underlying model for stream discharge [Botter *et al.*, 2007]: the streamflow recession constant ( $k$ ), the inverse of the mean increment in discharge during runoff generation events ( $\gamma_Q$ ), and the frequency of runoff generation events ( $\lambda$ ). While  $\lambda$  and  $\gamma_Q$  can be found using information on catchment soil, vegetation, and climatic features [Botter *et al.*, 2007], we examine

translation of the daily model to seasonal and annual time scales and therefore estimate parameters of the daily model by extracting values of  $\lambda$  and  $\gamma_Q$  from the hydrograph. The mean interarrival time of runoff generation events ( $\lambda$ ) is calculated by identifying the number of hydrograph peaks ( $N$ ) over all summation periods and dividing this by the total number of days over which all sums are computed ( $P$  days =  $n$  years of data  $\times T$  days per year). The recession constant ( $k$ ) is determined by first restricting our attention to hydrograph peaks during the summation period and selecting only those for which runoff both decreases (negative first derivative) and is convex (positive second derivative or concave up) for at least 4 days following the recession peak [Dralle *et al.*, 2017]. Recession constants are computed using nonlinear least squares regression to fit the model  $Q_i = Q_0 e^{-kt}$  to each recession, and the median  $k$  of all extracted recessions is used in the model. Finally, the mean discharge increment is calculated by first computing the long-term mean of daily discharge during the summation periods ( $\mu_Q$ ) and noting that  $\mu_Q = \lambda/(k\gamma_Q) \implies \gamma_Q = \lambda/(\mu_Q k)$ . We use this relationship, along with the previously computed values of  $k$  and  $\lambda$ , to estimate  $\gamma_Q$ .

### 2.3.1. Case Study 1: Seasonal Flow Totals

For a given season length in days ( $T$ ), the seasonal flow can be straightforwardly expressed as

$$Q_{\text{season}} = \sum_{i=1}^T Q_i \Delta t, \quad (16)$$

where  $Q_i$  is the flow value (in units of cm/d) on day  $i$ . The mean of the distribution of  $Q_{\text{season}}$  is obtained as  $\mu_{Q_{\text{season}}} = T \cdot \mu_Q$ , where  $\mu_Q$  is the daily mean flow over all summation periods, and the standard deviation  $\sigma_{Q_{\text{season}}}$  can be found using equation (5), with  $f(Q) = Q$ . Finally, we generate an approximate gamma probability distribution  $p_{Q_{\text{season}}}$  for the random variable  $Q_{\text{season}}$  with shape ( $c$ ) and rate ( $d$ ) parameters that are consistent with  $\mu_{Q_{\text{season}}}$  and  $\sigma_{Q_{\text{season}}}^2$ , or

$$c = \mu_{Q_{\text{season}}}^2 / \sigma_{Q_{\text{season}}}^2 \quad (17)$$

$$d = \mu_{Q_{\text{season}}} / \sigma_{Q_{\text{season}}}^2 \quad (18)$$

$$p_{Q_{\text{season}}}(q) = \frac{d^c}{\Gamma(c)} q^{c-1} e^{-dq}. \quad (19)$$

In Figure S1 in the supporting information, we demonstrate that the choice of a gamma distribution well represents actual distributions of flow sums.

To assess goodness of fit of the distributions, we use the Nash-Sutcliffe efficiency applied to the percentiles of the predicted and empirical probability distributions of  $Q_{\text{season}}$  [e.g., Castellarin *et al.*, 2004; Müller *et al.*, 2014]:

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{99} (Q_m^i - Q_o^i)^2}{\sum_{i=1}^{99} (Q_o^i - \overline{Q_o})^2}, \quad (20)$$

where  $Q_m^i$  is the model predicted  $i$ th percentile of  $Q_{\text{season}}$  and  $Q_o^i$  the corresponding empirical percentile. The NSE achieves a value of 1 for perfect agreement between the model and observation.

### 2.3.2. Case Study 2: Sediment Export

The total sediment export  $L_{\text{season}}$  (kg) over a season of length  $T$  (days) can be approximated as the sum of daily observations  $L_i$  (kg/d):

$$L_{\text{season}} = \sum_{i=1}^T L_i \Delta t. \quad (21)$$

To calculate the standard deviation of  $L_{\text{season}}$ , one need not consider the steady state distribution or covariance of  $L_i$  directly. Instead,  $L_i$  can be formulated as a derived random variable of  $Q_i$  whose covariance is governed by equation (7), once a suitable parameterization of the sediment rating curve  $L_i = \beta Q_i^\delta$  has been established. We fit the rating curve for each site using nonlinear least squares regression on the  $(Q_i, L_i)$  pairs observed over the period of record for which both  $Q_i$  and  $L_i$  are available. The mean of the distribution of  $L_{\text{season}}$  is then defined as  $T \cdot \mu_L$ , where  $\mu_L$  is the mean of the daily sediment export distribution. This mean  $\mu_L$  and the associated variance of the seasonal variable  $\sigma_{L_{\text{season}}}^2$  can be computed using  $f(Q) = \beta Q^\delta$  in equation (8) and equation (14), respectively. We again use  $\mu_{L_{\text{season}}}$  and  $\sigma_{L_{\text{season}}}^2$  to determine a moment consistent gamma distribution for the probability density function of  $L_{\text{season}}$  ( $p_{L_{\text{season}}}$ ) and measure model performance using the Nash-Sutcliffe efficiency.

### 3. Results and Discussion

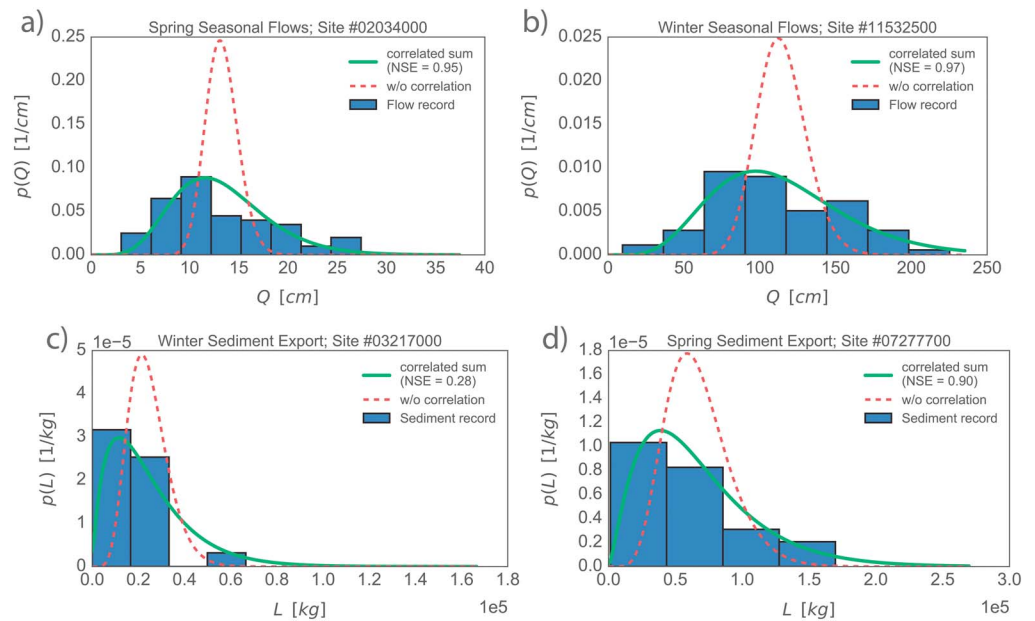
#### 3.1. Case Study 1: Seasonal Flow Totals

In Figures 1a and 1b, we plot the empirical frequency histograms for total winter season flows on the Smith River (December through February;  $T = 92$  days), and total spring flows on the Rivanna River (March–May;  $T = 93$  days). The Smith River is highly seasonal, with most annual precipitation (annual rainfall average approximately equal to 2 m) falling during a 6 month wet season spanning from November to April. The Rivanna River has a fairly persistent rainfall regime, with little seasonal variation in monthly rainfall totals, and an annual mean total rainfall of approximately 1.2 m. The number of bins for each histogram equals the rounded square root of the number of years in the flow record [Montgomery and Runger, 2010]. On top of these histograms, we plot the gamma probability distribution of seasonal flows for two cases: (1) where daily flows are independent and identically (gamma) distributed according to equation (6) (red hashed line); and (2) where daily flows are identically distributed according to equation (6) and autocorrelated in time (green solid line). In the latter case, the flow correlation between days is accounted for by the methods from section 2.2. Differences between the independent and autocorrelated cases are expected (and included) for reference, but the key result is the good match between empirical histogram and the analytical prediction of the solid curve. Both flow plots in Figure 1 demonstrate that accounting for serial correlation is necessary to correctly estimate the variability of seasonal sums. The importance of the contribution of hydrograph autocorrelation is a function of both the autocorrelation time scale of the hydrograph and the duration of the period over which the sum is computed. In particular, we find that the error increases sharply as the summation period becomes longer, and, for a given summation period, the error increases with the autocorrelation length. This sensitivity is explored in Figure S2 in the supporting information. Results for other seasons are similar to those plotted in Figure 1. Table 1 catalogues NSE of model performance for six other USGS study watersheds.

Error patterns can be attributed to the simplified model formulation. For example, standard deviations are systematically underestimated, likely due to interannual variations in climatic model parameters ( $\gamma_Q$ ,  $\lambda$ ), which are here assumed constant. Appropriate representations of year-to-year stochasticity of  $\lambda$  and  $\gamma_Q$  [e.g., Porporato *et al.*, 2006] could be integrated into the model to account for this variability, analogously to the stochastic soil moisture model of Feng *et al.* [2017]. Still, generally good model performance suggests that interannual variations in seasonal flow totals are largely explained by the statistical properties of daily flows. That is, most of the observed variability does not stem from interannual variations in climate statistics but instead from the manner in which a particular set of weather statistics are realized and propagated through a catchment.

Model performance is lowest during the summer period in eastern U.S. watersheds. We find that almost universally, these watersheds exhibit strong downward flow trends each summer (followed by a strongly non-stationary upward trend in the fall), despite relatively constant rainfall patterns. We suspect that this is due to a combination of increasing transpiration over the course of the summer, or increasing interception in deciduous forests, both of which would limit volumes of groundwater recharge and thus streamflows [Zimmer and McGlynn, 2017]. Such forms of nonstationarity violate the assumptions of the underlying model [Botter *et al.*, 2007]. This does not represent a failure of the method presented here for accounting for autocorrelation but rather an important limitation to the applicability of existing probabilistic models for the daily flow PDF. For illustration, we present summer flows for two catchments (Watauga River and Rivanna River) but omit the fall and summer seasons in other study watersheds. We examine sensitivity of seasonal flow variability to intraseasonal transpiration nonstationarity in Figure S3 in the supporting information.

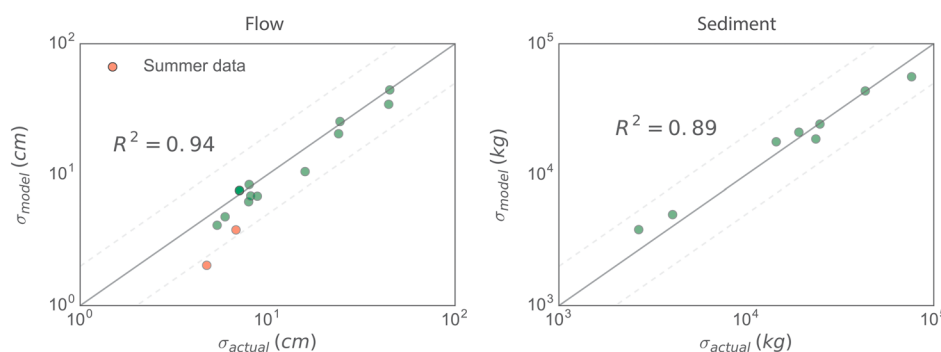
Overall, model performance is comparable to other methods for examining interannual flow variations. Peel *et al.* [2000] calibrated a seven-parameter hydrologic model (SIMHYD) to predict monthly flow variations in 331 Australian catchments. Their model is not directly comparable to the present study (we predict by season, instead of month; Peel *et al.* [2000] examine the coefficient of variation, not the standard deviation of flows). Nevertheless, as a coarse comparison, 80% of the validation data sets from Peel *et al.* [2000] exhibit percent error of less than 20%, compared to 22% in this study (ignoring the summer season, which fails because model stationarity assumptions are violated). Simpler statistical methods are also available to predict annual flow variability. For example, McMahon *et al.* [2011] used 10 variants of a meteorologically based statistical technique to predict the standard deviation of annual flows, with model  $R^2$  values ( $\sigma_{\text{predicted}}$  versus  $\sigma_{\text{actual}}$ ) ranging from 0.52 to 0.58. In comparison, our model achieves an  $R^2$  equal to 0.94 (see Figure 2), although we note that these bulk measures of performance should be interpreted with caution, given the relatively small data set used in the present study.



**Figure 1.** Distributions of seasonal flows for (a) Rivanna River (spring: March–May) and (b) Smith River (winter season: December–February), and seasonal net suspended sediment export for (c) Tygart’s Creek (winter: December–February) and (d) Hickahala Creek (spring: March–May). The legend indicates a green (red) line for the “moment consistent” gamma distribution approximation with (without) accounting for correlation in summing.

**Table 1.** Flow and Sediment Performance

USGS Gage ID	Name	Season	Theoretical (Empirical) $\sigma$	NSE	CV <sub>Q</sub>
<i>Flow Performance (<math>\sigma</math> (cm))</i>					
03491000	Big Creek, Tennessee	Winter	6.95 (8.13)	0.95	1.55
03491000	Big Creek, Tennessee	Spring	6.95 (8.13)	0.95	1.06
11475560	Elder Creek, California	Spring	20.52 (23.92)	0.97	1.29
11475560	Elder Creek, California	Winter	34.45 (44.26)	0.92	1.5
50138000	Rio Guanajibo, Puerto Rico	Fall	10.53 (15.85)	0.90	1.31
07277700	Hickahala Creek, Mississippi	Spring	6.23 (7.93)	0.96	2.51
07277700	Hickahala Creek, Mississippi	Winter	6.87 (8.81)	0.94	2.27
02034000	Rivanna River, Virginia	Spring	4.78 (5.93)	0.95	1.18
02034000	Rivanna River, Virginia	Summer	2.04 (4.73)	0.67	1.53
02034000	Rivanna River, Virginia	Winter	4.12 (5.38)	0.94	1.26
11532500	Smith River, California	Winter	44.38 (44.85)	0.97	1.35
11532500	Smith River, California	Spring	25.45 (24.34)	0.96	1.02
03479000	Watauga River, North Carolina	Summer	3.82 (6.86)	0.85	1.09
03479000	Watauga River, North Carolina	Spring	8.40 (7.97)	0.98	1.06
03479000	Watauga River, North Carolina	Winter	7.57 (7.08)	0.98	1.20
<i>Sediment Performance (<math>\sigma</math> (kg))</i>					
07277700	Hickahala Creek, Mississippi	Spring	43,594.97 (43,098.30)	0.90	2.51
07277700	Hickahala Creek, Mississippi	Winter	55,944.55 (76,173.64)	0.87	2.27
01666400	Rappahannock River, Virginia	Spring	24,322.43 (24,697.13)	0.51	1.06
01666400	Rappahannock River, Virginia	Winter	18,710.69 (23,482.05)	-0.07	1.20
03383000	Tradewater River, Kentucky	Spring	4,964.76 (4,041.28)	0.85	1.37
03383000	Tradewater River, Kentucky	Winter	3,812.57 (2,664.57)	0.44	1.70
03217000	Tygart’s Creek, Kentucky	Spring	21,083.86 (19,066.67)	0.72	1.70
03217000	Tygart’s Creek, Kentucky	Winter	17,809.00 (14,332.30)	0.28	1.85



**Figure 2.** Coefficients of determination for model-predicted standard deviations versus actual standard deviations, for both seasonal flow totals and seasonal net suspended sediment export. Dashed lines indicate a factor of 2 error envelope.

### 3.2. Case Study 2: Sediment Export

We plot suspended sediment results for Hickahala Creek and Tygart’s Creek (both of which exhibit fairly persistent rainfall regimes) in Figures 1c and 1d. The plots are generated analogously to Figures 1a and 1b and demonstrate that complex sums of functions of flow are also amenable to the presented theory. Relatively lower NSEs (Figure 1) for sediment PDFs may result from shorter data records or inadequacy of gamma probability distributions for this application; predicted versus actual standard deviations (Figure 2) are fairly accurate, with an  $R^2$  value equal to 0.89.

Other error results from the limited usefulness of the sediment rating curve approach in some basins, where hysteretic or nonfunction sediment-flow relations are common [Seeger *et al.*, 2004]. Using the sediment rating curve to “project” sediment records with the flow variable (that is, transforming daily flows with the rating curve function as an estimate for daily sediment fluxes) significantly improves seasonal sediment model performance, with, for example, the winter sediment Rappahannock Creek NSE rising from  $-0.07$  to  $0.91$ . This demonstrates that the sediment rating curve is likely the dominant source of error in Table 1. Alternative rating curve models, leading to more complex functions of flow,  $f(Q)$  [see Runkel *et al.*, 2004], could be applied without altering our approach.

More generally, the approach could be straightforwardly extended to other applications. Simple functions,  $f$ , could be used to specify water abstraction or irrigation strategies as a function of daily flow; stochastic sums, in this instance, would be necessary to examine the variability of total water withdrawals for a given strategy. This suggests useful applications of the model for alternative sustainable management strategies in regions where poorly planned surface water withdrawals are detrimental to stream ecosystems [Bauer *et al.*, 2015]. Similar functions of daily flow have also been used to model power output of run-of-river hydropower facilities [Basso and Botter, 2012]. Whereas hydrograph autocorrelation limited Basso and Botter [2012] to a lumped, multiyear analysis of energy output, the methods presented here would make it possible to examine annual or seasonal variability of total energy output potential for a given season.

### 3.3. Model Limitations

As previously mentioned, interannual variations of the climatic parameters  $\gamma_Q$  and  $\lambda$ , or within-season nonstationarity (due to increasing ET, for example) are not presently accounted for by the model, which nevertheless performs well given that model parameters are assumed constant over 3 month periods. Other more complex forms of interannual variability or covariation between forcing variables, such as observed covariation between ET and annual rainfall totals [McMahon *et al.*, 2011], may play an important role in determining interannual flow variations, which has not been examined here. Nevertheless, the relative success of the model suggests that the majority of interannual variation in the presented case studies results from the year-to-year realization of the assumed stationary sequence of recharge pulses, as determined by the (constant) mean interarrival time between recharge events ( $1/\lambda$ ) and the mean flow increment during recharge events ( $1/\gamma_Q$ ).

Other extensions or improvements of the model might help to alleviate error. For example, some error may also derive from the simplified linear reservoir used to model stream discharge. Botter *et al.* [2009] extend the framework to include nonlinear reservoirs, and Botter [2010] examine stochasticity in the rate of recession ( $k$ ) itself, but we are unaware of any simple procedure for computing the covariance function in these cases.



Although we present a relatively straightforward method for computing the standard deviation of aggregated flow variables, the technique must be implemented numerically for any nonlinear function  $f$ . Analytical expressions for the covariance for nonlinear  $f$  would greatly increase speed of computation and facilitate interpretation of model output.

#### 4. Conclusion

We developed a process-oriented method to predict variability of catchment discharge and discharge-derived variables at seasonal and annual time scales. Results demonstrate that the statistics and covariance structure of daily flows are the primary drivers of hydrologic variability on longer time scales. By accounting for the serially correlated nature of the flow time series, the underlying flow model is easily extended to variables that functionally depend on daily flow values. We demonstrated this linkage using the example of a power law model for suspended sediment export and suggest other applications relating to water resources management. Predictive fidelity of the model is on par with or better than other more highly parameterized process models and statistical methods, providing a new, powerful tool for predicting the interannual and interseasonal variability of aggregated hydrologic variables.

#### Acknowledgments

D. Dralle and S.E. Thompson acknowledge support from the National Science Foundation CZP EAR-1331940 for the Eel River Critical Zone Observatory. Streamflow and suspended sediment data for this study are generously provided by the United States Geological Survey. Code to calculate the standard deviation of a stochastic sum of a power law function of  $Q$  can be found at <https://github.com/daviddralle/sdSum>.

#### References

- Abramowitz, M., and I. Stegun (1965), *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables*, National Bureau of Standards Applied Mathematics Series, vol. 55, 8 pp., U.S. Govt. Printing Office, Washington, D. C.
- Ahearn, D. S., R. W. Sheibley, R. A. Dahlgren, M. Anderson, J. Johnson, and K. W. Tate (2005), Land use and land cover influence on water quality in the last free-flowing river draining the western Sierra Nevada, California, *J. Hydrol.*, *313*(3), 234–247.
- Anderson, J., E. Jeppesen, and M. Søndergaard (2005), Ecological effects of reduced nutrient loading (oligotrophication) on lakes: An introduction, *Freshwater Biol.*, *50*(10), 1589–1593.
- Bárdossy, A. (2007), Calibration of hydrological model parameters for ungauged catchments, *Hydrol. Earth Syst. Sci. Discuss.*, *11*(2), 703–710.
- Basso, S., and G. Botter (2012), Streamflow variability and optimal capacity of run-of-river hydropower plants, *Water Resour. Res.*, *48*(10), doi:10.1029/2012WR012017.
- Basso, S., A. Frascati, M. Marani, M. Schirmer, and G. Botter (2015), Climatic and landscape controls on effective discharge, *Geophys. Res. Lett.*, *42*, 8441–8447, doi:10.1002/2015GL066014.
- Basso, S., M. Schirmer, and G. Botter (2016), A physically based analytical model of flood frequency curves, *Geophys. Res. Lett.*, *43*, 9070–9076, doi:10.1002/2016GL069915.
- Bauer, S., J. Olson, A. Cockrill, M. van Hattem, L. Miller, M. Tauzer, and G. Leppig (2015), Impacts of surface water diversions for marijuana cultivation on aquatic habitat in four northwestern California watersheds, *PLoS one*, *10*(3), e0120016.
- Biggs, B. J., V. I. Nikora, and T. H. Snelder (2005), Linking scales of flow variability to lotic ecosystem structure and function, *River Res. Appl.*, *21*(2-3), 283–298, doi:10.1002/rra.847.
- Blöschl, G. (2013), *Runoff Prediction in Ungauged Basins: Synthesis Across Processes, Places and Scales*, Cambridge Univ. Press, Cambridge, U. K.
- Botter, G. (2010), Stochastic recession rates and the probabilistic structure of stream flows, *Water Resour. Res.*, *46*, W12527, doi:10.1029/2010WR009217.
- Botter, G., A. Porporato, I. Rodriguez-Iturbe, and A. Rinaldo (2007), Basin-scale soil moisture dynamics and the probabilistic characterization of carrier hydrologic flows: Slow, leaching-prone components of the hydrologic response, *Water Resour. Res.*, *43*, W02417, doi:10.1029/2006WR005043.
- Botter, G., A. Porporato, I. Rodriguez-Iturbe, and A. Rinaldo (2009), Nonlinear storage-discharge relations and catchment streamflow regimes, *Water Resour. Res.*, *45*, W10427, doi:10.1029/2008WR007658.
- Castellarin, A., G. Galeati, L. Brandimarte, A. Montanari, and A. Brath (2004), Regional flow-duration curves: Reliability for ungauged basins, *Adv. Water Resour.*, *27*(10), 953–965.
- Ceola, S., E. Bertuzzo, G. Singer, T. J. Battin, A. Montanari, and A. Rinaldo (2014), Hydrologic controls on basin-scale distribution of benthic invertebrates, *Water Resour. Res.*, *50*, 2903–2920, doi:10.1002/2013WR015112.
- Dettinger, M. D., and H. F. Diaz (2000), Global characteristics of stream flow seasonality and variability, *J. Hydrometeorol.*, *1*(4), 289–310.
- Doulatyari, B., S. Basso, M. Schirmer, and G. Botter (2014), River flow regimes and vegetation dynamics along a river transect, *Adv. Water Resour.*, *73*, 30–43.
- Dralle, D. N., N. J. Karst, K. Charalampous, A. Veenstra, and S. E. Thompson (2017), Event-scale power law recession analysis: Quantifying methodological uncertainty, *Hydrol. Earth Syst. Sci.*, *21*(1), 65–81.
- Farmer, D., M. Sivapalan, and C. Jothityangkoon (2003), Climate, soil, and vegetation controls upon the variability of water balance in temperate and semiarid landscapes: Downward approach to water balance analysis, *Water Resour. Res.*, *39*(2), 1035, doi:10.1029/2001WR000328.
- Feng, X., T. E. Dawson, D. D. Ackerly, L. S. Santiago, and S. E. Thompson (2017), Reconciling seasonal hydraulic risk and plant water use through probabilistic soil-plant dynamics, *Global Change Biol.*, doi:10.1111/gcb.13640.
- Grabs, W. (2009), *The United Nations World Water Development Report 3: Water in a Changing World*, chap. Bridging the observational GAP, UNESCO World Water Assessment Programme/London: Earthscan, Paris.
- Jiang, C., L. Xiong, S. Guo, J. Xia, and C. Xu (2017), A process-based insight into nonstationarity of the probability distribution of annual runoff, *Water Resour. Res.*, *53*, 4214–4235, doi:10.1002/2016WR019863.
- Kotz, S., and J. Neumann (1963), On the distribution of precipitation amounts for periods of increasing length, *J. Geophys. Res.*, *68*(12), 3635–3640, doi:10.1029/JZ068i012p03635.
- Lazzaro, G., and G. Botter (2015), Run-of-river power plants in alpine regions: Whither optimal capacity?, *Water Resour. Res.*, *51*(7), 5658–5676, doi:10.1002/2014WR016642.
- Löf, G. O., and C. H. Hardison (1966), Storage requirements for water in the United States, *Water Resour. Res.*, *2*(3), 323–354, doi:10.1029/WR002i003p00323.

- McMahon, T. A., R. M. Vogel, M. C. Peel, and G. G. Pegram (2007), Global streamflows—Part 1: Characteristics of annual streamflows, *J. Hydrol.*, *347*(3), 243–259.
- McMahon, T. A., M. C. Peel, G. G. Pegram, and I. N. Smith (2011), A simple methodology for estimating mean and variability of annual runoff and reservoir yield under present and future climates, *J. Hydrometeorol.*, *12*(1), 135–146.
- McMahon, T. A., et al. (2013), Prediction of annual runoff in ungauged basins, in *Runoff Prediction in Ungauged Basins: Synthesis Across Processes, Places and Scales*, pp. 70–100, Cambridge, U. K.
- Milly, P. (1993), An analytic solution of the stochastic storage problem applicable to soil water, *Water Resour. Res.*, *29*(11), 3755–3758, doi:10.1029/93WR01934.
- Monk, W. A., P. J. Wood, D. M. Hannah, and D. A. Wilson (2008), Macroinvertebrate community response to annual and regional river flow regime dynamics, *River Res. Appl.*, *24*(7), 988–1001, doi:10.1002/rra.1120.
- Montgomery, D. C., and G. C. Runger (2010), *Applied Statistics and Probability for Engineers*, John Wiley, Hoboken, N. J.
- Morehead, M. D., J. P. Syvitski, E. W. Hutton, and S. D. Peckham (2003), Modeling the temporal variability in the flux of sediment from ungauged river basins, *Global Planet. Change*, *39*(1), 95–110.
- Muller, K. E. (2001), Computing the confluent hypergeometric function,  $M(a, b, x)$ , *Numer. Math.*, *90*(1), 179–196.
- Müller, M. F., D. N. Dralle, and S. E. Thompson (2014), Analytical model for flow duration curves in seasonally dry climates, *Water Resour. Res.*, *50*, 5510–5531, doi:10.1002/2014WR015301.
- Müller, M. F., S. E. Thompson, and M. N. Kelly (2016), Bridging the information GAP: A webGIS tool for rural electrification in data-scarce regions, *Appl. Energy*, *171*, 277–286.
- Muneepeerakul, R., S. Azale, G. Botter, A. Rinaldo, and I. Rodriguez-Iturbe (2010), Daily streamflow analysis based on a two-scaled gamma pulse model, *Water Resour. Res.*, *46*, W11546, doi:10.1029/2010WR009286.
- Peel, M. C., F. H. Chiew, A. W. Western, and T. A. McMahon (2000), Extension of unimpaired monthly streamflow data and regionalisation of parameter values to estimate streamflow in ungauged catchments, Melbourne, Australia.
- Poff, N. L., J. D. Allan, M. B. Bain, J. R. Karr, K. L. Prestegard, B. D. Richter, R. E. Sparks, and J. C. Stromberg (1997), The natural flow regime, *BioScience*, *47*(11), 769–784.
- Porporato, A., E. Daly, and I. Rodriguez-Iturbe (2004), Soil water balance and ecosystem response to climate change, *Am. Natural.*, *164*(5), 625–632.
- Porporato, A., G. Vico, and P. A. Fay (2006), Superstatistics of hydro-climatic fluctuations and interannual ecosystem productivity, *Geophys. Res. Lett.*, *33*, L15402, doi:10.1029/2006GL026412.
- Rossi, M. W., K. X. Whipple, and E. R. Vivoni (2016), Precipitation and evapotranspiration controls on daily runoff variability in the contiguous United States and Puerto Rico, *J. Geophys. Res. Earth Surf.*, *121*, 128–145, doi:10.1002/2015JF003446.
- Runkel, R. L., C. G. Crawford, and T. A. Cohn (2004), Load Estimator (LOADEST): A FORTRAN program for estimating constituent loads in streams and rivers, Tech. Rep.
- Seeger, M., M.-P. Errea, S. Begueria, J. Arnáez, C. Martí, and J. Garcia-Ruiz (2004), Catchment soil moisture and rainfall characteristics as determinant factors for discharge/suspended sediment hysteretic loops in a small headwater catchment in the Spanish Pyrenees, *J. Hydrol.*, *288*(3), 299–311.
- Sene, K., and F. Farquharson (1998), Sampling errors for water resources design: The need for improved hydrometry in developing countries, *Water Resour. Manage.*, *12*(2), 121–138.
- Viola, F., E. Daly, G. Vico, M. Cannarozzo, and A. Porporato (2008), Transient soil-moisture dynamics and climate change in Mediterranean ecosystems, *Water Resour. Res.*, *44*, W11412, doi:10.1029/2007WR006371.
- Vogel, R. M., and I. Wilson (1996), Probability distribution of annual maximum, mean, and minimum streamflows in the United States, *J. Hydrol. Eng.*, *1*(2), 69–76.
- Yu, K.-x., L. Gottschalk, L. Xiong, Z. Li, and P. Li (2015), Estimation of the annual runoff distribution from moments of climatic variables, *J. Hydrol.*, *531*, 1081–1094.
- Zanardo, S., C. Harman, P. Troch, P. Rao, and M. Sivapalan (2012), Intra-annual rainfall variability control on interannual variability of catchment water balance: A stochastic analysis, *Water Resour. Res.*, *48*, W00J16, doi:10.1029/2010WR009869.
- Zimmer, M. A., and B. L. McGlynn (2017), Ephemeral and intermittent runoff generation processes in a low relief, highly weathered catchment, *Water Resour. Res.*, doi:10.1002/2016WR019742.