Journal of Hydrology 505 (2013) 35–46

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

The use of broken power-laws to describe the distributions of daily flow above the mean annual flow across the conterminous U.S.

Catalina Segura [a,b,*], Davide Lazzati [c], Arumugam Sankarasubramanian [d]

[a] Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC, USA
[b] Department of Marine, Earth, and Atmospheric Sciences, North Carolina State University, Raleigh, NC 27695, USA
[c] Department of Physics, North Carolina State University, Raleigh, NC 27695, USA
[d] Department of Civil, Construction, and Environmental Engineering, North Carolina State University, Raleigh, NC 27695, USA

ARTICLE INFO

Article history:
Received 20 February 2013
Received in revised form 5 August 2013
Accepted 16 September 2013
Available online 21 September 2013
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Purna Chandra Nayak, Associate Editor

Keywords:
Stream flow distribution
Ungauged basins
Broken power law
Daily flows
Statistical description of flow distribution

ABSTRACT

A recent study employed a broken power-law (BPL) distribution for understanding the scaling frequency of bankfull discharge in snowmelt-dominated basins. This study, grounded from those findings, investigated the ability of a BPL function to describe the distribution of daily flows above the mean annual flow in 1217 sites across the conterminous U.S. (CONUS). The hydrologic regime in all the sites is unregulated and spans a wide range in drainage areas (2–120,000 km^2) and elevation (0–3000 m). Available daily flow records in all sites varied between 15 and 108 years. Comparing the performance of BPL distribution and the traditionally used lognormal distribution, we found that BPL provides stronger fit in ~80% of the sites. Thus the BPL function provides a suitable tool to model daily flows in most areas of the CONUS. The potential for developing a model for predicting the frequency distribution of daily flows in ungauged sites was analyzed. We found that such model is possible using drainage area, mean basin elevation, and mean annual precipitation as predicting variables for any site located above 600 m across the CONUS. We also found strong continental-wide correlations between 3 of the 4 parameters that describe the BPL and basin characteristics. Our results indicate that the BPL function provides a robust alternative to traditional functions such as the lognormal to model the statistical variation of daily flows above the mean annual in most basins of the CONUS.

1. Introduction

Predicting the distribution of daily flows for ungauged sites or for river systems that lack hydrologic records is critical for a number of diverse problems in hydrology and water management that includes the estimation of the daily distributions of nutrients, pollutants, water temperature, watershed sediment yields, and for the estimation of design yield for runoff-the-river reservoir systems (Vogel and Fennessey, 1995). From the geomorphological perspective, predictions of the frequency of intermediate to high flows are of special interest since these flows transport the majority of the sediment load and determine the morphology of the river channel (Andrews, 1994; Emmett and Wolman, 2001; Whiting et al., 1999; Wolman and Miller, 1960).

In most drainage basins only a small fraction of the stream network is gauged. Hydrologic records from these gauges are vital, however, for developing fluvial-hydraulic models, which can then be applied to ungauged sites. Predicting flows in ungauged basins (PUB) have rich literature in hydrology (Sivapalan, 2003; and references therein). Approaches available to PUB include the application of hydrologic models (conceptual or physically based) and the extrapolation of response information from gauge to ungauged basins. Statistically the approach commonly used for formulating flow frequency distributions at ungauged locations involves two steps: First, a function that describes the frequency of flows at gauged locations needs to be found. Then, a proper scaling relationship has to be identified to extend the model from gauged to ungauged locations, based on one or more properties of the basin that correlate with the parameters of the function that describes the flow frequency distribution. These assumptions form the basis of the *index flood* procedure, which hydrologists have used for years in developing regional relationship for estimating peak flows (Pitlick, 1994; Potter and Faulkner, 1987; Vogel and Wilson, 1996). In principle, any other streamflow characteristic, including extreme flows and daily flows, can be regionalized with this same approach. However, the frequency distribution of daily flows is often complex (e.g., biomodal and non-symmetric, Vogel and Fennessey, 1995), thereby the identification of a function that describes it and the prediction of the frequency distribution to ungauged sites is difficult. To illustrate this complexity, Fig. 1 presents

* Corresponding author at: Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC, USA. Tel.: +1 919 985 6353.
E-mail address: csegura@ncsu.edu (C. Segura).

0022-1694/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jhydrol.2013.09.016

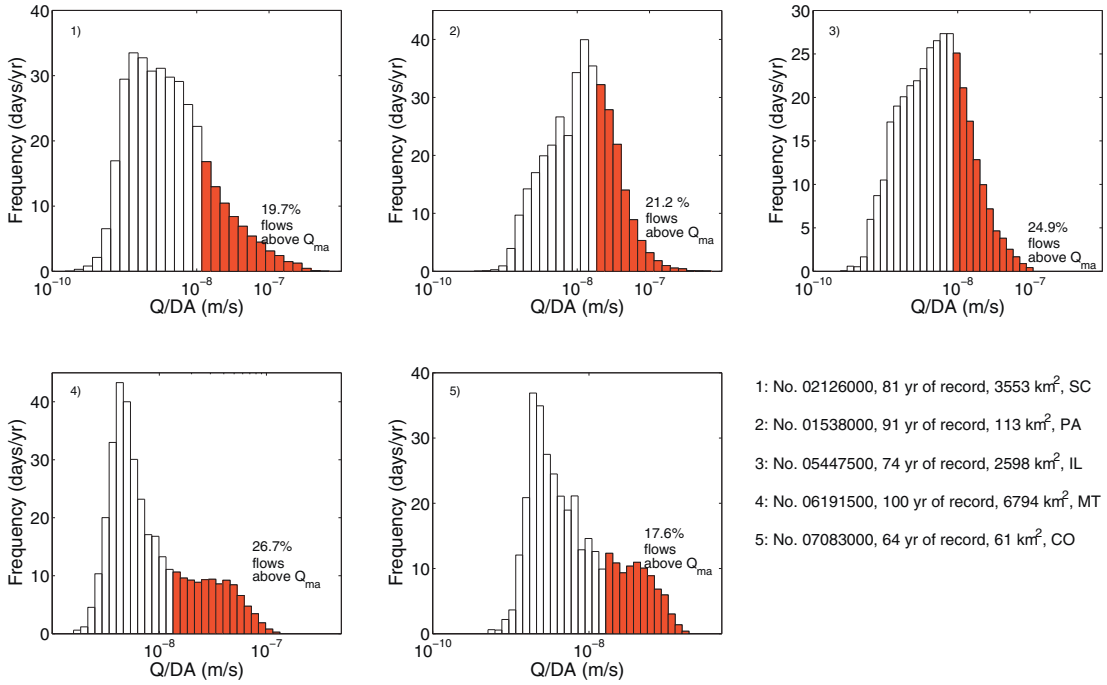

Fig. 1. Frequency distributions of daily flows normalized by drainage area, *DA*, of 5 basins across the CONUS. The red area shows flows above the mean annual, Q_{ma} (i.e. range considered in this paper). The percentage of the flows above the mean annual flow varies between 17% and 25% and it is indicated in each pannel. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

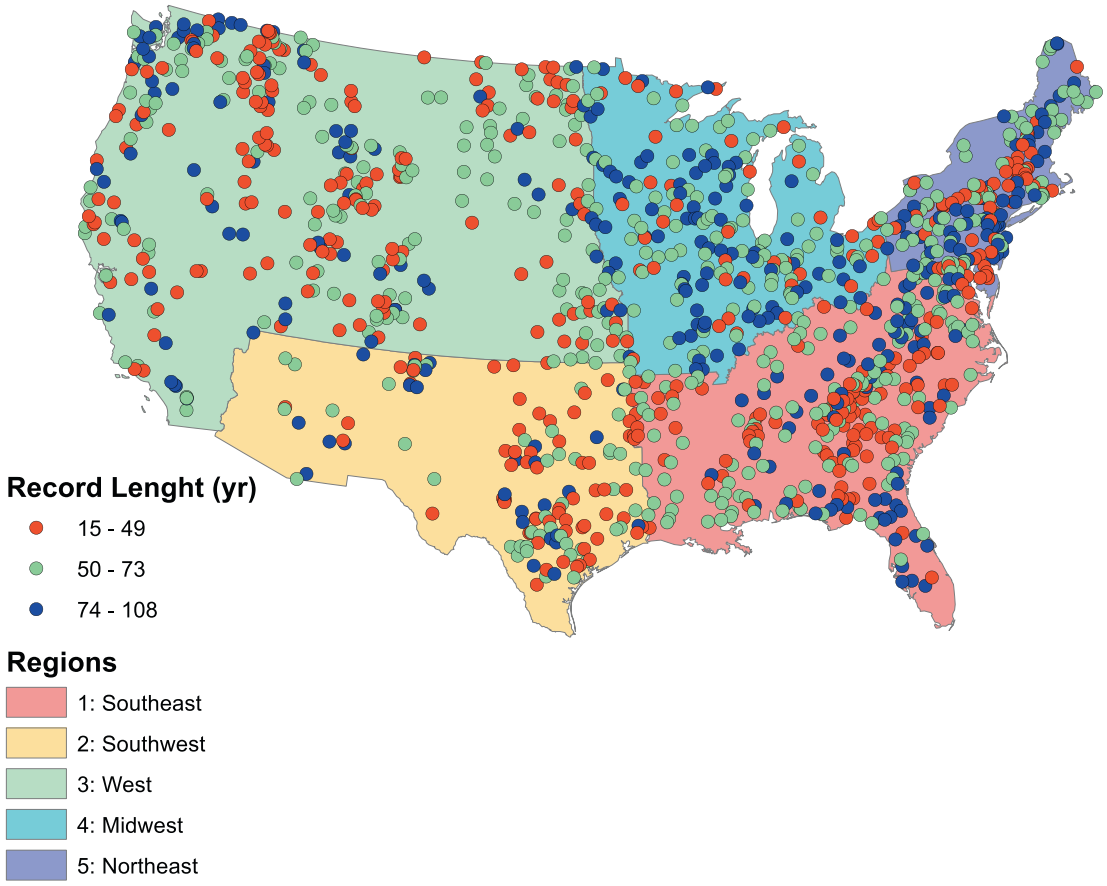

Fig. 2. Location and length of record of each studied gauge (n = 1217). Five regions considered in the analysis are also indicated. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the histograms for the daily flow frequency distribution of 5 basins in different locations across the conterminous U.S. (CONUS). The primary interest to this paper is on the distribution of flows above the mean annual flows (shaded area in Fig. 1). From Fig. 1, basins with significant snowmelt (e.g. Site 4 and 5) have a lower number of flows in the right tail indicating the low frequency of extreme events. In these systems over 99% of all annual peak flows occurs during May–July and no peak flows have been even reported between September and January. In contrast, rainfall dominated regimes (e.g. sites 1, 2, and 3) produced distributions dominated by less sharp changes in the frequency of high flow events (i.e. right tail) indicating a more unpredictable hydrologic regime and a higher frequency of high flow events. In these sites the available record indicates that the annual peak flow has occurred in all months but mainly between January and May. Many statistical distributions have been used to model the frequency distribution of daily flows that include lognormal, exponential, gamma, pareto, and kappa (Castellarin et al., 2004; Fennessey and Vogel, 1990; Goodwin, 2004; Mueller and Pitlick, 2005; Nash, 1994; Potter, 2001). However, to our knowledge none of these functions has been proven to be an appropriate choice to model daily flows across the whole CONUS. An alternative function called the broken power law, BPL, was found to be a strong candidate to describe the distributions of daily flows above the mean annual in snowmelt dominated systems of Colorado and Idaho (Segura and Pitlick, 2010). The objectives of this paper are to (1) determine if the BPL function is adequate to describe the distribution of daily flows above the mean annual flow across unregulated sites of the CONUS and (2) explore correlations between basin characteristics and the parameters of the BPL function that could enable a model to predict the distribution of daily flows at ungauged sites.

This paper is organized as follows: The methodology section presents the data sources and candidate distributions for fitting the distribution of daily flows over the CONUS. Following that, in the results and discussion section, we analyzed the performance of 4 functions in over 1200 watersheds across the CONUS, and compared their performance. Then we seek correlations between the parameters of the best function (i.e. BPL), basin characteristics, and climate (i.e. mean annual precipitation). Finally, the conclusions summarize the salient findings from the study along with implications for extending the findings for predicting the distribution of daily flows at ungauged sites.

2. Methodology

2.1. Data sources

We selected 1217 watersheds across the CONUS (Fig. 2) from the Hydro-Climatic Data Network, HCDN (Slack et al., 1993) to quantify the spatial variability in the distribution of daily flows above mean annual flow. All sites have unregulated flow regimes with at least 15 years of daily streamflow data on record (Fig. 2). HCDN watersheds, identified by Slack et al. (1993), are basins whose streamflow records are minimally affected by anthropogenic influences such as upstream storages or groundwater pumping. The lowest accuracy rating of any of the data records was categorized as "good". The original HCDN data set includes 1474 sites with acceptable flow time series for daily flow analysis (Slack et al., 1993). However, we exclude 257 sites that, according to information in the USGS website, have minor diversions. The drainage area (DA) of the selected basins spans 5 orders of magnitude between 3 and 120,000 km^2. The daily flow records from available complete water years between 1902 and 2010 were downloaded from the USGS website on each site. The average number of years of data available in all sites is 61.8 ranging from 15 and 108 years. Several studies have employed HCDN sites for understanding the scaling properties of annual flows over the continental US (Vogel and Sankarasubramanian, 2000) as well as for relating the long-term water balance to various basin characteristics (Sankarasubramanian and Vogel, 2002). Basin characteristics (DA and mean basin elevation, BE) were also accessed from the USGS website. Precipitation normals between 1981 and 2010 were obtained from the 4 km gridded data developed by the PRISM Climate Group, Oregon State University (http://prism.oregonstate.edu, created 20 Jan 2012).

2.2. Distribution functions considered

We consider four functions for characterizing the daily flow distributions above the mean annual flow of the selected sites. Some of these functions are not integrable in the positive domain, but they can be used as probability density functions when they are truncated at the mean-annual flow. Three of the functions are power laws. The first is a simple power law:

$$\frac{dN}{dQ} = b_0 Q^{b_1} \quad Q_{ma} \leqslant Q \leqslant \infty; \quad b_1 < -1 \tag{1}$$

where N is the number of days per year, Q is the discharge, Q_{ma} is the mean annual discharge, and b_0 and b_1 are the normalization parameter and the slope of the discharge-frequency relation, respectively. The second function consider corresponds to a broken power law (BPL). This function has been successfully used to describe the distribution of daily flows above the mean annual in snowmelt dominated systems (Segura and Pitlick, 2010):

$$\frac{dN}{dQ} = \frac{a_0/a_1}{(Q/a_1)^{\alpha} + (Q/a_1)^{\beta}} \quad Q_{ma} \leqslant Q \leqslant \infty; \quad \alpha < \beta; \quad \beta > 1 \tag{2}$$

where a_0 is a normalization parameter, a_1 is a value of Q corresponding to the inflection point in the distribution, and α and β are the slopes of the two separated power law segments of the function for low and high flows, respectively. The last power function consider is also a broken power law but inverted, having the slope for the second segment (θ) smaller than the slope of the first segment (ϕ). We refer to this function as an inverted broken power law (IBPL):

$$\frac{dN}{dQ} = c_0 \left(\left[\frac{Q}{c_1}\right]^{-\phi} + \left[\frac{Q}{c_1}\right]^{-\theta} \right) \quad Q_{ma} \leqslant Q \leqslant \infty; \quad \phi < \theta; \quad \theta > 1 \tag{3}$$

where c_0 is a normalization parameter and c_1 is a value of Q corresponding to the inflection point.

The fourth and final distribution function consider is the lognormal (LN). This function has been widely used to describe daily flows (Doyle and Shields, 2008; Fennessey and Vogel, 1990; Vogel et al., 2003). We use a truncated version of this function to model daily flows above the mean annual:

$$\frac{dN}{dQ} = \frac{d_0}{Q\sigma\sqrt{2\pi}} \exp\left[\frac{(\ln(Q) - \mu)^2}{2\sigma^2}\right] \quad Q_{ma} \leqslant Q \leqslant \infty \tag{4}$$

where μ and σ are the mean and standard deviation of the natural logarithm of Q when the whole range of flows is consider and d_0 is a parameter introduce to correct the normalization for the effect of truncating the distribution at Q_{ma}. Fig. 3 shows the fit for data above the mean annual flows for a site, USGS Gauge No. 05447500 – Green River near Geneseo, IL, for three of the candidate distributions considered. The inverted BPL is not shown because the best fit for this site yields $\alpha = \beta$, $\alpha = \beta$, which is the same as the PL.

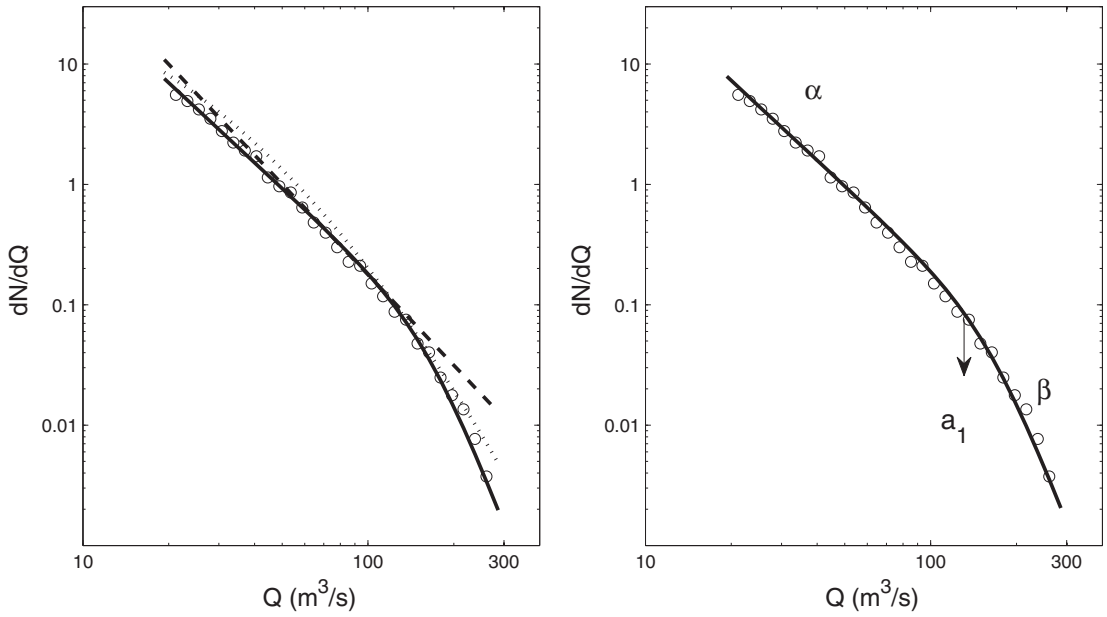

Fig. 3. Three of the distribution functions considered to model the daily flow distributions of 1217 sites across the CONUS. The left panel presents the data from one of the sites (USGS Gauge No. 05447500 – Green River near Geneseo, IL) with the best-fit broken power law, BPL, simple power law, PL, and log-normal, LN, functions. The right panel presents the same data with the BPL fit (the overall best fitting function). The parameters of the function are also shown.

2.3. Fitting methodology

Each probability density function described above is characterized by a set of n free parameters, $n-1$ which correspond to shape parameters plus a normalization parameter (a_0, b_0, c_0, *or* d_0). The most probable set of the $n-1$ shape parameters of each candidate function is obtained by maximizing the likelihood of the specified probability density function within a grid of values for the parameters (Bevington and Robinson, 2003). The range of parameter values tested for each function was wide enough to ensure capturing all possible variations (Table 1). The logarithm of the Likelihood ($\log L$) is computed as:

$$\log \mathcal{L} = \sum_i \log[f(Q_i)] \tag{5}$$

where $\mathcal{L}$ is the likelihood, which is a function of the $n-1$ shape parameters, f is each function (PL, BPL, IBPL, or LN), and Q_i are the daily flow data points. A value of $\log L$ is computed for each choice of the parameter values and the parameters corresponding to the highest $\log L$ are retained as the best set. The same procedure was repeated for each function f, yielding four sets of best parameters per site.

The integral of any of the distributions must give the average number of days ($\#_{ma}$) in which the flow is above the Q_{ma}:

$$\int_{Q_{ma}}^{\infty} \frac{dN}{dQ} dQ = \#_{ma} \tag{6}$$

The normalization parameter is computed from the data (i.e., $\#_{ma}$) using Eq. (6) once the shape parameters have been determined.

Once the best parameter set is determined for every function, the strength of the fit is evaluated using χ^2 minimization, since maximum likelihood can be used to compare fits but not to establish an absolute measure of the goodness of fit (Bevington and Robinson, 2003; Press et al., 2007; Segura and Pitlick, 2010). The χ^2 is defined as

$$\chi^2 = \sum_{i=1} \frac{[f_i - f(Q_i)]^2}{\sigma_i^2} \tag{7}$$

where f_i are the number of flows falling in a bin centered on flow Q_i and of width ΔQ and σ_i is the uncertainty associated with f_i. The uncertainty is assumed to be $2\sqrt{f_i}$ (Segura and Pitlick, 2010). The frequency distributions of all sites are constructed following the methodology described by Segura and Pitlick (2010). Histograms of daily flows above mean annual are made over 100 intervals of discharge equally sized in logarithmic space. The number of counts per bin is normalized by the bin width, ΔQ, to produce a frequency distribution where the frequency f_i is independent of the chosen interval width (Newman, 2005; Segura and Pitlick, 2010). If the number of observations in a given bin is less than ten, two consecutive bins are joined to improve statistics.

Our parameter fitting procedure assumes that the data are independent. However the distributions of daily flows are serially correlated and therefore the assumption is not met. In this case, the tested distributions are used only as mathematical functions for fitting the observations, and the Maximum likelihood scores are not used as the basis for hypothesis testing. To our knowledge

Table 1
Range of values considered for every free parameter (x_{1-n}) and range of values obtained for the normalization parameters (x_0) of each distribution considered. PL = Power law, BPL: Broken power law, IBPL: Inverted broken power law, LN: Log normal.

Parameter	Function	Range of values tested	Units*
b_o	PL	16–4.8 × 10^{16}	d/y
b_1	PL	1.1–10	–
a_0	BPL	0.008–1320	d/y
a_1	BPL	1.1–100	Q/Q_{ma}
α	BPL	−5–7	–
β	BPL	1.1–20	–
c_0	IBPL	8.3 × 10^{-13}–30.2	d/y
d_1	IBPL	1.1–100	Q/Q_{ma}
ϕ	IBPL	1.1–20	–
θ	IBPL	1.1–20	–
d_0	LN	0.2–10.6	
μ	LN	10^{-10}–10^{10}	d/y
σ	LN	0.1–10	–

d/y: Days per year; Q: discharge, Q_{ma}: mean annual flow.

there is no goodness-of-fit test that overcomes the problem of serial correlation in daily flows (Segura and Pitlick, 2010).

3. Results and discussion

3.1. Best function to describe the distributions of daily flows

We found that power law functions (PL, BPL, and IBPL) are superior to describe the daily flow distribution across the CONUS compare to the lognormal (LN) distribution. The maximum_$\mathcal{L}$ of the fit for the power functions is higher for 1008 gauges (83%) whereas it is higher for the LN in 209 sites (17%, Fig. 4a). Even though the LN function fit yields a higher maximum_$\mathcal{L}$ for 17% of the sites, the difference between the maximum_$\mathcal{L}$ for the power fits versus the LN fit is very small (Fig. 4b) and varied between 0.008 and 53. In fact the ratio between the maximum_$\mathcal{L}$ of the LN function to the best maximum_$\mathcal{L}$ of the 3 power functions varies between 0.98 and 1.03 in these 209 sites (Fig. 4b). On the other hand, the difference between the maximum_$\mathcal{L}$ of the power functions versus themaximum_$\mathcal{L}$ of the LN distribution for sites where the former is best varies between 0.0008 and 2241 with a ratio of maximum_$\mathcal{L}$_LN to maximum_$\mathcal{L}$_power functions between −0.15 and 1.53. In other words, while the LN is only

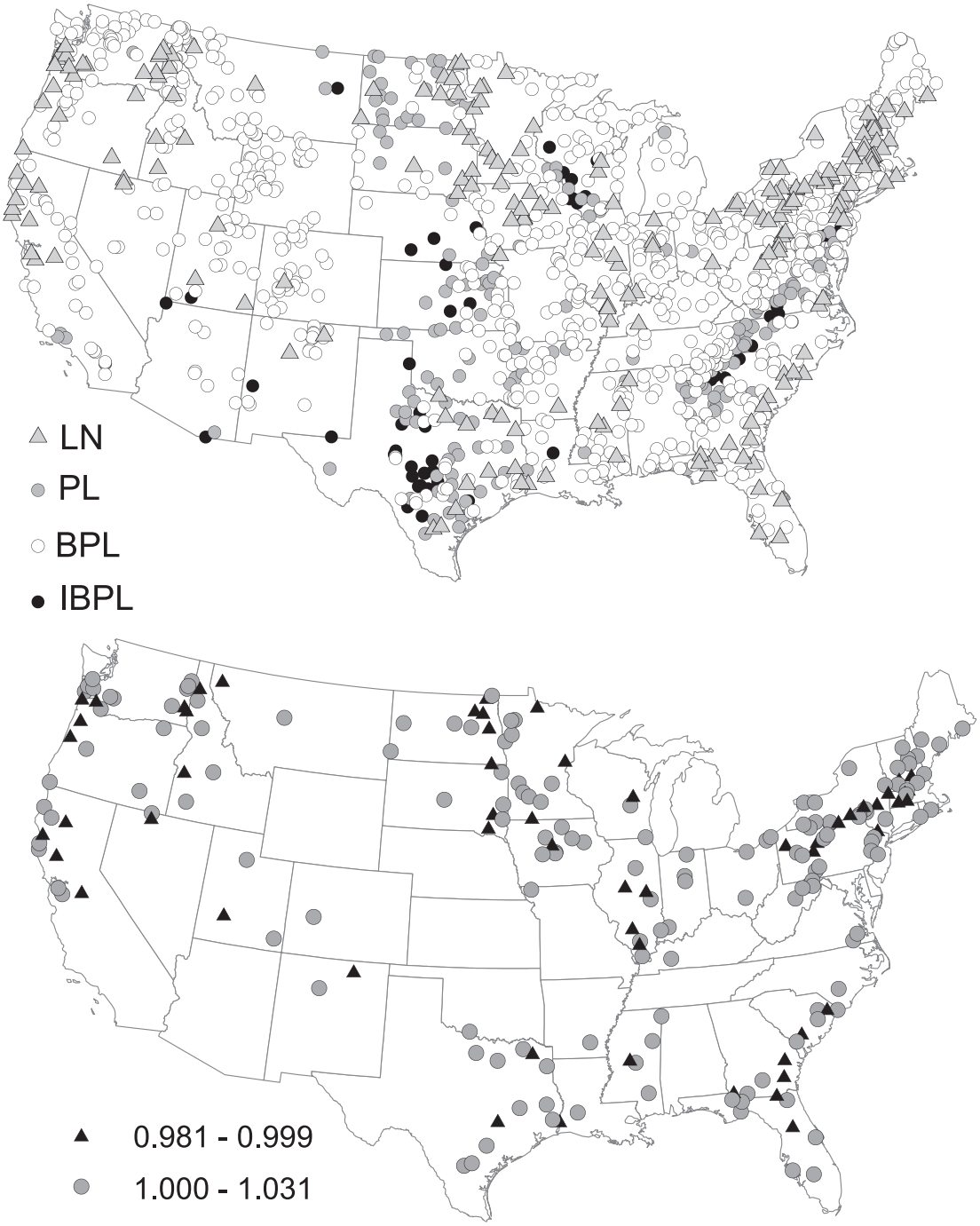

Fig. 4. (a) Performance of functions considered as determined by the maximum_$\mathcal{L}$. Circle markers show 1008 gauges better fitted by power-law functions (inverted Power Law, IBPL, power law, PL, and broken power law, BPL) and triangles represent 209 gauges best fitted by the log-normal function. Panel (b) ratio of the Maximum likelihood of the LN fit to the best power functions (PL, BPL, or IBPL) in the 209 sites where the LN is best.

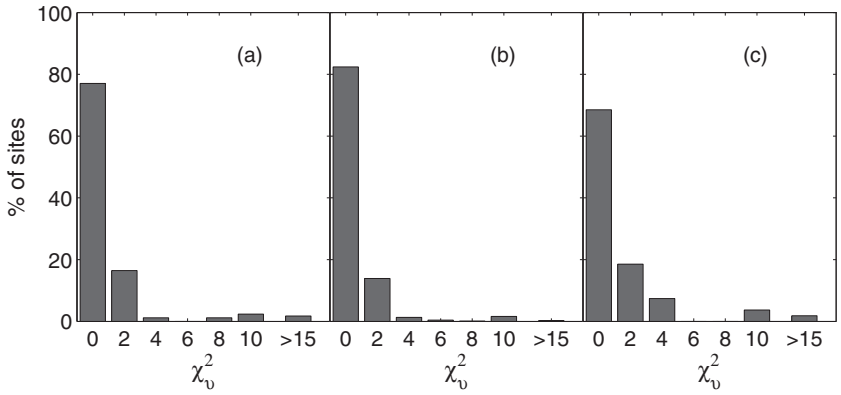

Fig. 5. Histograms of the reduced χ^2 of the three power-law function to 1217 gauges in the CONUS. Panel a shows the 170 sites best described by the simple power law function, panel b shows the 993 sites best fitted by the broken power law function, and panel c shows the 54 sites best fitted by the inverted broken power law function.

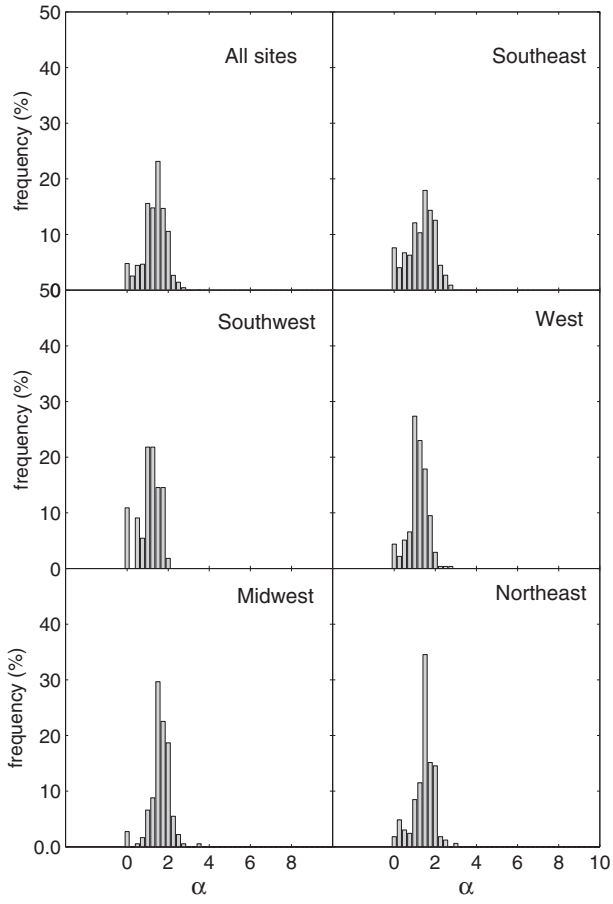

Fig. 6. Distributions of the α parameter of the BPL by geographic region (Fig. 2).

marginally superior in some cases, there are many others in which the power-law functions are strongly superior. Therefore the LN distribution was discarded from the rest of the analysis. In order to compare models with different number of free parameters we use the likelihood ratio technique (Kendall and Stuart, 1977; Mood and Graybill, 1963). This technique is based on the fact that the ratio of the likelihoods of two models with different numbers of free parameters is distributed as a χ^2 with a number of degrees of freedom equal to the difference between free parameters in the two models. For example, when comparing a simple power-law (2 parameters) with a broken power-law (4 parameters), the latter is consider superior only if the likelihood of BPL is higher than the likelihood of the PL by 13.82 corresponding to a probability of 99.9% (for 2 degrees of freedom) that the BPL is superior than the PL ($L_{BPL} > L_{PL} + 13.82$). When instead comparing LN and BPL models (with 3 and 4 parameters, respectively) the latter is considered superior only if its likelihood exceeded by 10.8 the former.

Among power functions alone (i.e. not considering the LN distribution) the broken power law is the best at describing the daily flow frequency distributions, yielding the best maximum_$\mathcal{L}$ for 993 sites (82%) compared to 54 sites for the IBPL (4%), and 170 sites for the PL (14%). The comparison of maximum_$\mathcal{L}$ scores among functions allows finding the best set of parameters of each function to fit the daily flow data of every site. It is also useful to compare models and establish which function is the best at describing the data. However, it is insufficient to evaluate how strong the best fit is in absolute terms. To that end we used the χ^2 statistic as a goodness of fit measure and the reduced χ^2 (χ^2_ν) to have a direct comparison between the χ^2 and the number of degrees of freedom. When the χ^2_ν is below or equal to one the fit is consider strong. The computed χ^2_ν for BPL fits in the 993 sites ranged between 0.17 and 1700; it ranged between 0.17 and 48 for the 170 sites best fitted by the simple power law function, and between 0.26 and 57 for the 54 sites best described by the IBPL (Fig. 5).

Table 2
Variability of the α and β parameter of the BPL per region. $\bar{x}$: Mean; σ: standard deviation, α_5, β_5: fifth percentile; α_{50}, β_{50}: median value; α_{95}, β_{95}: nine-fifth percentile.

Region	N	α					β				
		$\bar{x}$	σ	α_5	α_{50}	α_{95}	$\bar{x}$	σ	β_5	β_{50}	β_{95}
All sites	899	1.31	0.64	0.20	1.40	2.10	4.8	1.9	2.70	4.30	8.40
Southeast	223	1.26	0.77	0.00	1.40	2.30	4.2	1.5	2.80	3.90	6.61
Southwest	56	1.00	0.78	−1.00	1.20	1.80	3.7	1.4	1.93	3.20	6.78
West	273	1.14	0.53	0.22	1.20	1.80	5.7	2.2	2.52	5.40	9.30
Midwest	182	1.61	0.50	0.86	1.65	2.20	4.8	1.7	2.96	4.50	7.90
Northeast	165	1.43	0.54	0.30	1.50	2.10	4.4	1.5	3.10	4.20	6.63

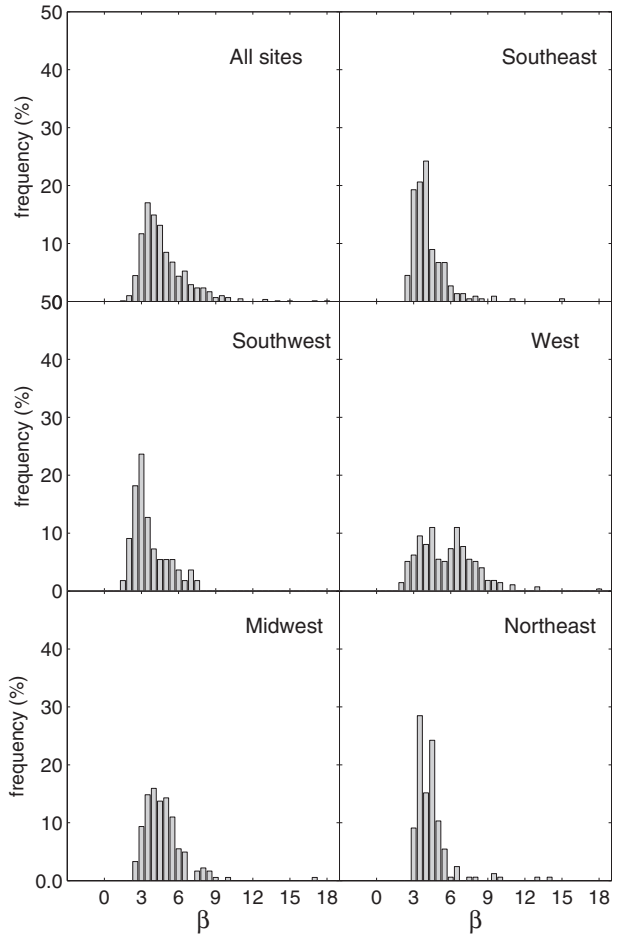

Fig. 7. Distributions of the β parameter of the BPL by geographic region (Fig. 2).

Since a strong fit yields a χ^2_ν near one we set up a threshold of 1.5 to discriminate between weak and strong fits and found that 150 sites were strongly fitted by the PL, 43 sites by the IBPL, and 899 by the BPL. The following sections discuss the parameter variability of the BPL in the 899 sites with reduced χ^2 below 1.5. These sites cover the entire CONUS and span over all the five regions considered.

3.2. Variability of the broken power law function parameters

In this section we describe the statistical summary and the spatial variability of the four parameters of the BPL function in 899 sites across the CONUS. We also explore linear relationships between these parameters and basin characteristics including *DA*, *BE*, climate (mean annual precipitation, *MAP*), and the predictability of the hydrologic regime, measured as the lag-1 annual autocorrelation, *R-1* (see explanation below), considering 5 geographic regions (Fig. 2). The motivation behind relating the parameters to basin characteristics is to develop a model that can predict the distribution of daily flows above the mean annual flows at ungauged sites.

3.2.1. Parameters α and β

The parameters that define the slopes of the two liner segments of the distribution show contrasting results. The parameter α (slope of the first power segment) varies over a smaller range than the parameter β (slope of the second power segment). In addition the shape of the distribution of α over the CONUS is more symmetric than the distribution of β (Fig. 6). The mean value and standard deviation of the α across all sites was 1.3 and 0.64 respectively compared to a mean value of 4.8 and standard deviation of 1.9 for β (Table 2, Figs. 6 and 7). The regional average of α ranged between 1.00 and 1.61 (Table 2). α did not reveal any trend or correlation with respect to *BE*, *DA*, *MAP*, or the regime predictability (i.e. *R-1*, Table 3). Since the flows covered by the parameter α are primarily over the normal range of flow, the parameter exhibits low variability. Previous studies have suggested setting up a regional mean value for the parameter α to develop a regional scale model (Segura and Pitlick, 2010). Considering the low regional variability observed here for this parameter (Table 2, Fig. 6) it could be set to a mean regional constant. By doing this, the number of free parameters of the BPL function are reduced from 4 to 3. As expected, the variability of β is larger since the parameter captures the extreme daily flows (Fig. 7). Therefore, assigning a constant regional value to this parameter was not an appropriate choice. Spatial inspection of β values across the CONUS revealed higher values in western mountainous basins than in the rest of the country (Fig. 8). The mean β for basins in Region 3 (i.e. Western) was the highest of all (Table 2). We found for snowmelt dominated systems in Colorado and Idaho that β was nearly constant and equal to 7 (Segura and Pitlick, 2010). The steeper the beta parameter the less likely is the occurrence of high extreme events in any given site and the extreme flows under these sites are predominantly controlled by snow storage. The correlation between *BE* and β was significant ($r = 0.47$, $p < 1E\text{-}15$, Table 3). The mean β value for basins located above 500 m was significantly higher ($p < 0.001$) than the mean value for sites at lower elevation. The variability of β with elevation (Fig. 9) indicated a strong correlation for sites above 500 m ($r = 0.98$, $p = 0.016$, Fig. 9). Conversely β could be set to a constant value of 4.4 for sites below 500 m (Fig. 9).

The annual hydrograph of western sites is often characterized by snowmelt regimes with peak annual flows between May and July. These regimes are therefore highly predictable compared to those in catchments in which high flows are dominated by the much less predictable occurrence of convective and frontal precipitation events. The degree of predictability of the hydrologic

Table 3
Pearson correlation matrix between the parameters of the BPL, number of days the daily flow is above the mean annual flow, #days, mean annual flow, Q_{ma}, and drainage area, *DA*, mean basin elevation, *BE*, one-year autocorrelation, *R-1*, and mean annual precipitation, *MAP*. Values in parenthesis correspond to *p*-values. Significant relations are in bold.

Depended variable	DA	BE	R-1	MAP
α	0.068 (0.04)*	−0.14 (0.000001)*	−0.095 (0.004)*	0.0013(0.6)
β	0.011 (0.73)	**0.47 (<1E-15)**	**0.66 (<1E-15)**	−0.29 (<1E-15)*
a_1	**0.74 (<10E-15)**	−0.17 (0.000001)*	0.05 (0.17)	**0.17 (0.0000002)**
a_0	−0.14 (0.00002)*	−0.05 (0.098)	−0.09 (0.004)*	**0.30 (<1E-15)**
#days	−0.08 (0.016)*	−0.13 (0.0004)*	0.14 (0.0002)*	**0.41 (<1E-15)**
Q_{ma}	**0.77 (<10E-15)**	−0.24 (<10E-10)*	0.02 (0.43)	**0.32 (<10E-15)**

* Indicates that while the linear relationship between the variables is significant ($p < 0.05$) the independent variable is a poor predictor of the dependent variable.

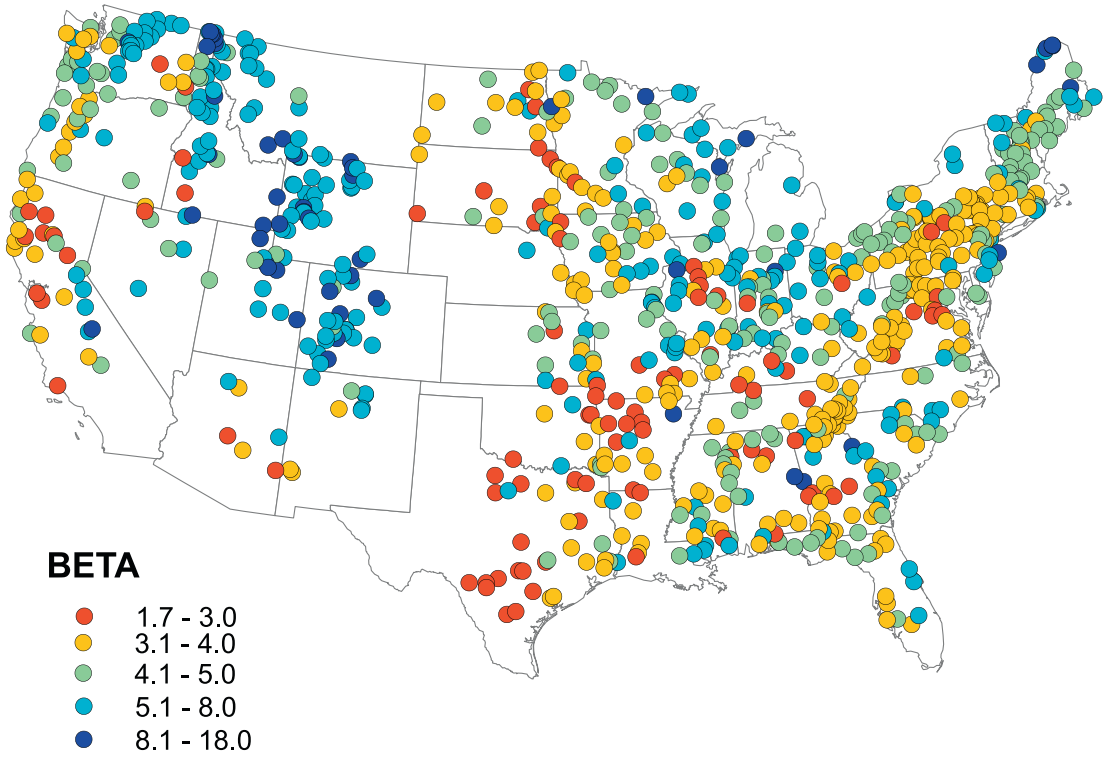

Fig. 8. Spatial distribution of the β parameter values across the CONUS (n = 899). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

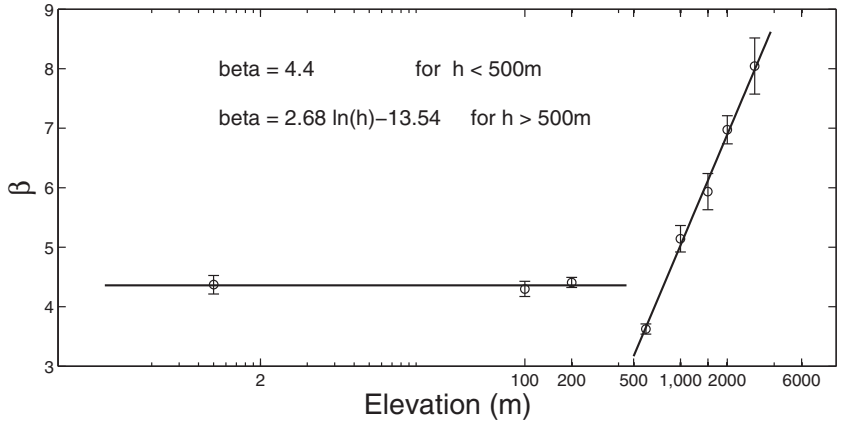

Fig. 9. Relationship between mean basin elevation and the parameter β. Error bars show the standard error.

regime of a given site is thus likely to be positively correlated to β. We characterize the degree of predictability of the hydrologic regimen with the lag-1 annual autocorrelation, *R-1:*

$$R-1=\frac{\int Q(t)Q(t+\delta t)dt}{\int Q^2(t)dt} \quad (8)$$

where δt is the time shift (in this case 1 year), and $Q(t)$ is discharge. Values of *R-1* close to one represent locations in which the flow experienced in any given day is significantly correlated to the flow experienced one year after that day and one year before that day. The *R-1* values for the 899 sites varied between −0.01 and 0.82 (Fig. 10). Sites with predictable regimes (i.e. *R-1* > 0.6) dominate in snowmelt-dominated systems in mountainous regions of Colorado, Idaho, Montana, and Wyoming, all located in Region 3 (Fig. 11). These sites exhibit a β value between 4.8 and 18 and a mean of 7.6. Low values of *R-1* (i.e. *R-1* < 0.35) characterize the Southeast (Region 2, Fig. 11) with β values varying widely between 2.3 and 15 and a mean of 4.2. Region 2 is predominantly a rainfall-runoff regime indicating a less role of storage in influencing the extreme flow values, which results in relatively lower values of β. The correlations between *R-1* and β are strong and significant in all regions (r = 0.24–0.77, p = 0.001-1E-15, Fig. 11, Table 3) and indicate that the frequency of extreme events decreases with increasing level of predictability (i.e. there is a positive relation between *R-1* and β). This relation was also evaluated considering *BE* (not shown here for brevity). The relationship for sites located above 500 m is stronger (r = 0.81, p < 1E-15) than for sites located below 500 m (r = 0.39, p < 1E-15). This again indicates the ability to determine β is much easier for basins at higher elevations.

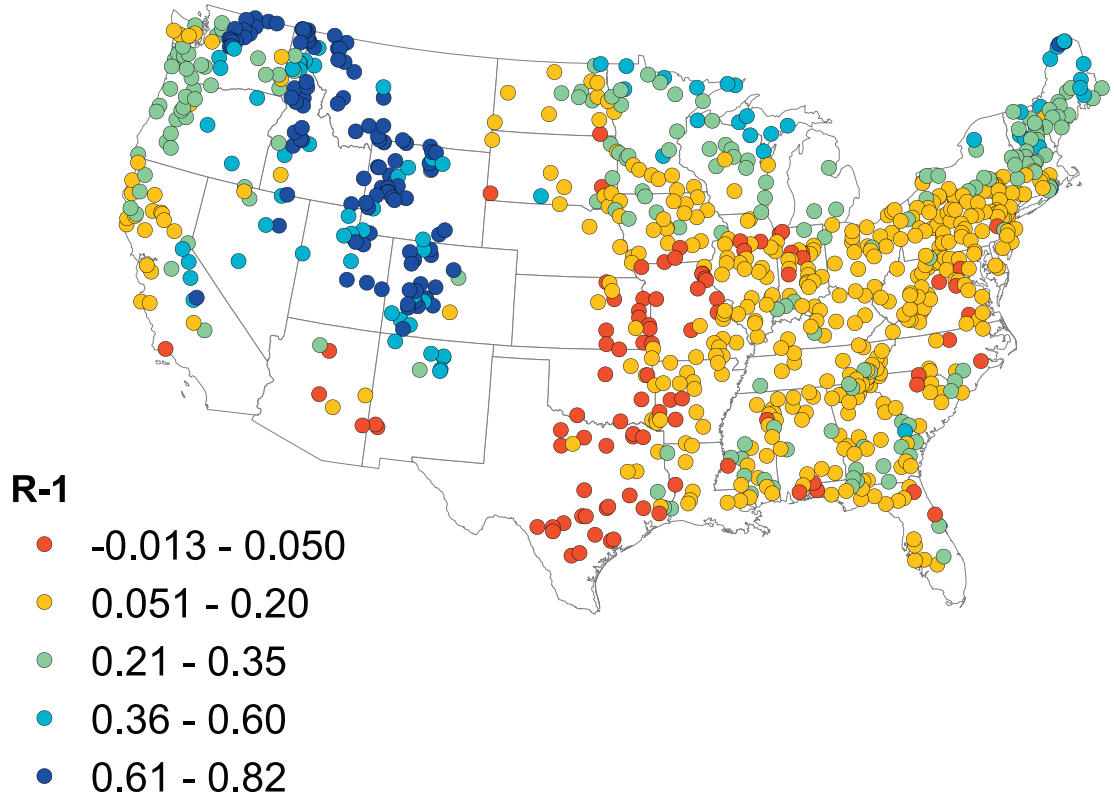

Fig. 10. Spatial distribution of the 1 year autocorrelation (*R-1*) across the CONUS. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

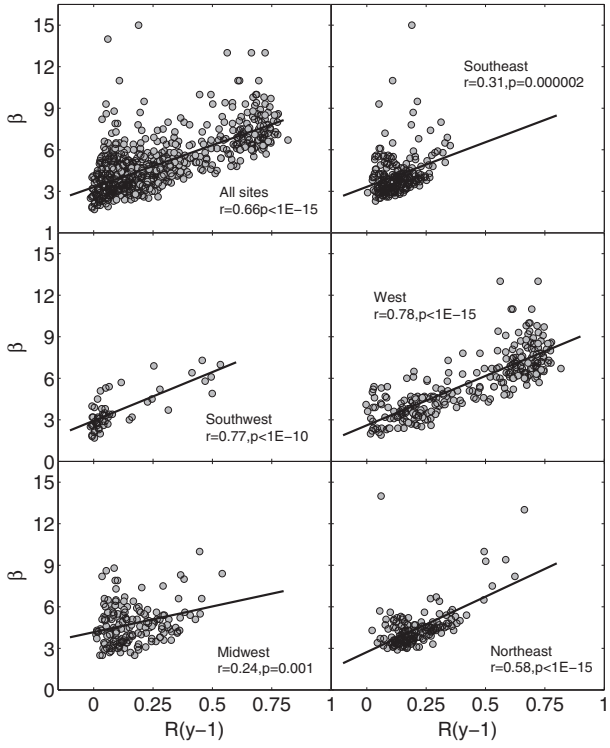

Fig. 11. Correlation between the one-year autocorrelation (*R-1*) and the β parameter.

3.2.2. Parameters a_0 and a_1

The flow at which the break in the broken power distribution occurs corresponds to the a_1 parameter. This discharge level varies between 0.57 and 3492 m^3/s over the 899 sites considered. This discharge divides intermediate and large flow levels and is strongly correlated to *DA*, ($r = 0.74$, $p < 0.000001$, Fig. 12a). The slope of this relation is less than one, which indicates that the frequency of extreme events decreases with drainage area. This is further emphasized by normalizing the a_1 values by *DA* (Fig. 12b). If the inflexion in the distribution would scale linearly with drainage area, the slope of this relation would be zero; instead it is −0.19. The physical explanation for this finding is beyond the scope of this paper however we found for 64 snowmelt dominated systems that it may be related to scale-dependent variations in runoff and sediment supply, which influence downstream trends in the bankfull channel geometry and intensity of sediment transport (Segura and Pitlick, 2010). Some of the variability of a_1 was also explained by *MAP* ($r = 0.17$, $p = 0.0000002$, Table 3). This parameter was not correlated to either *BE* or *R-1* (Table 3). The values of a_0 are directly computed from the data (Eq. (6)) based on the number of days the daily discharge is above the mean annual flow (#days) and the mean annual flow (Q_{ma}). a_0 varied between 0.48 and 548.6 with a mean value of 68.6 and was correlated to *MAP* ($r = 0.30$ $p <$ 1E-15, Table 3). However this correlation had significant scatter (not shown here for brevity). Considering that a_0 is computed based on the #days and Q_{ma} (i.e. Eq. (6)) we also explored the correlations between these variables and *DA*, *BE*, *R-1*, and *MAP* (Table 3). We found that #days is significantly correlated to *MAP* ($r = 0.41$ $p <$ 1E-15, Table 3, and Fig. 13). This relation is stronger for sites located above 600 m (Fig. 13, panel a) than for sites located at lower elevations (Fig. 13, panel b). The Q_{ma} is found to be strongly correlated to *DA* and *MAP* ($r = 0.32$–0.77, $p <$ 10E-15, Table 3, and Fig. 13) and to the product of *MAP* and *DA* (i.e. total volume of precipitation, Fig. 13, panel d). Based on the mentioned relations (Fig. 13a and d) a_0 can be computed at ungauged sites located above 600 m (i.e. using Eq. (6)).

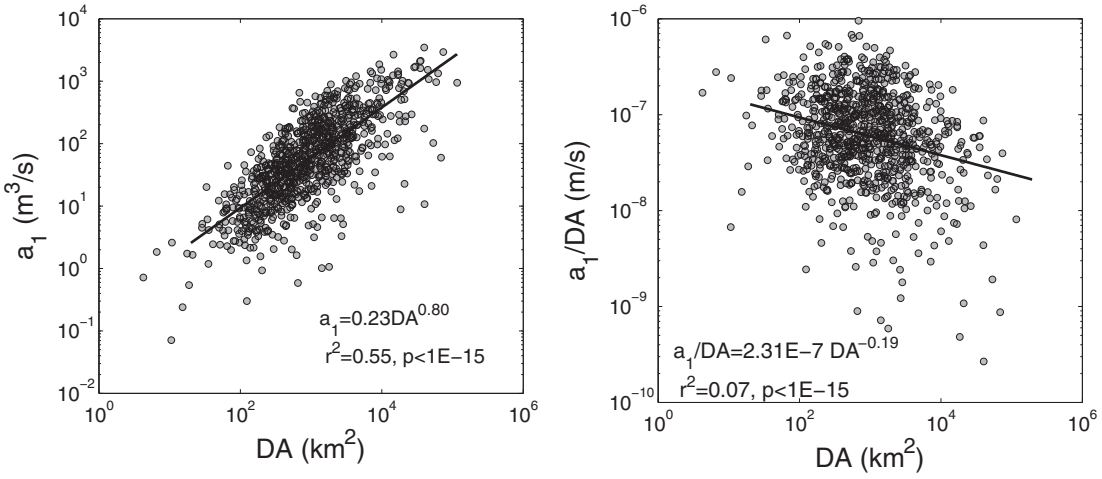

Fig. 12. Relationship between drainage, *DA*, area and the parameter a_1 of the broken power law (BPL) function. Panel a shows the relation between *DA* and a_1 and panel b the relation between a_1 normalized by *DA* and *DA*.

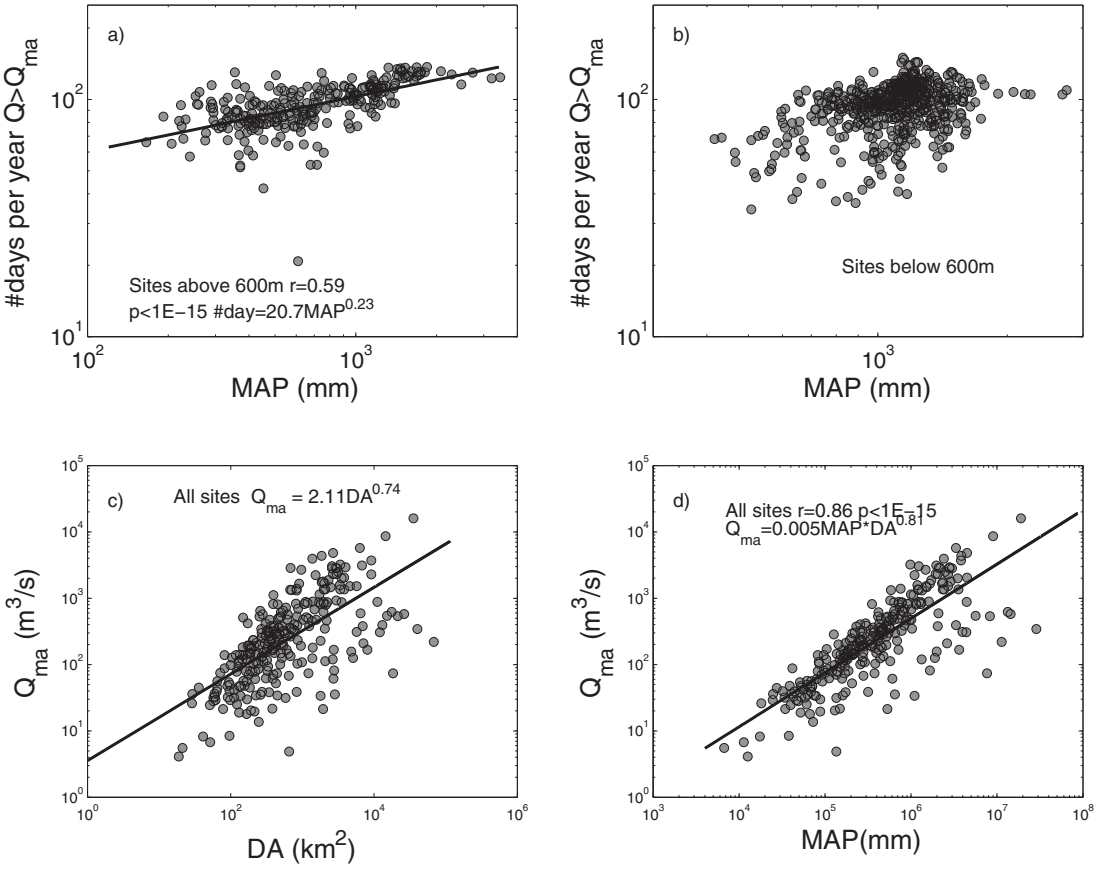

Fig. 13. Relationship between mean annual precipitation, *MAP*, and the number days the daily discharge is above the mean annual flow (#days) and mean annual flow (Q_{ma}) and between drainage area (*DA*) and Q_{ma}. Panel a presents the relation between *MAP* and #days for sites located above 600 m and panel b presents this relation for sites located below 600 m. Panel c presents the relation between *DA* and Q_{ma} for all sites and panel d the relation between *MAP* and Q_{ma} for all sites.

3.3. Approximation to a regional scale model

According to the results presented here, a rigorous model to predict the frequency distribution of daily flow above the mean annual flow at the CONUS scale is not possible. However, we found that it is possible to (1) set the α parameter to a constant in each of the 5 geographic regions (see Table 2); (2) predict the slope of the second segment of the distribution, β based on *BE*, (3) predict the inflection discharge of the distribution, a_1, based on *DA*, and (4) predict the normalization parameter a_0 based on *MAP* and *DA* for any site located above 600 m. The relations to predict β, a_1, and a_0 based on *MAP*, *DA*, and *BE* are shown for all sites in Figs. 9,12,13 and Table 3. These relations were adjusted including only sites above 600 m:

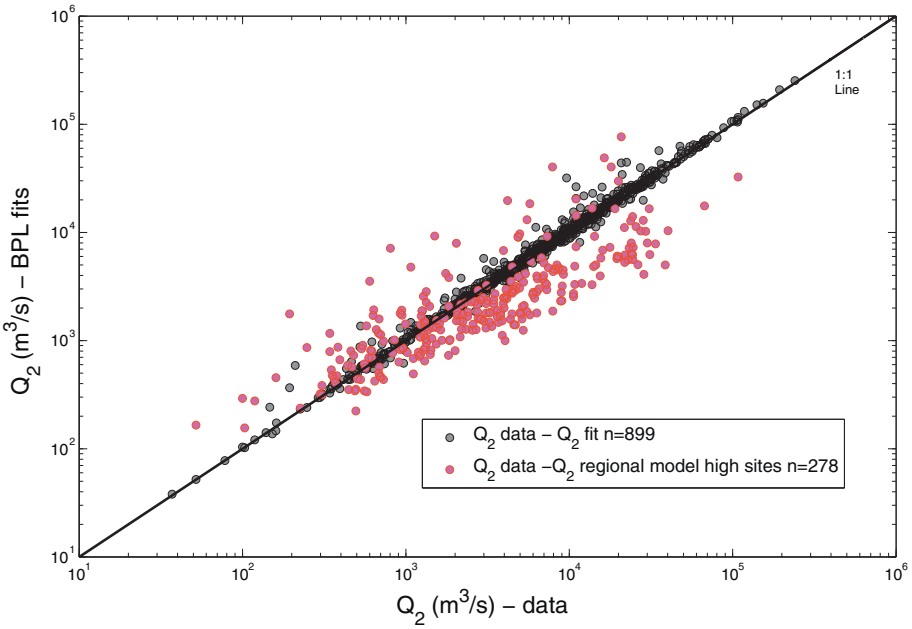

Fig. 14. (a) Comparison of the two-year flow (Q_2) computed based on the observations at each site and the predicted (Q_2) based on the fitted parameters of the BPL in all sites (gray markers) and (b) comparison between and Q_2 computed based on the observations and Q_2 computed based on BPL parameters calculated based drainage area, precipitation, and mean basin elevation (Eqs. (9)–(13)) for 278 sites located above 600 m (red markers). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$\beta = 2.68ln(BE) - 13.54 \tag{9}$$

$$a_1 = 15.3DA^{0.67} \tag{10}$$

$$\#_{ma} = 20.7MAP^{0.23} \tag{11}$$

$$Q_{ma} = 0.0054(MAP \times DA)^{0.81} \tag{12}$$

$$a_0 = \frac{a_1 \times \#_{ma}}{\int_{Q_{ma}}^{\infty} \frac{dQ}{(Q/a_1)^{\alpha}+(Q/a_1)^{\beta}}} \tag{13}$$

In order to test the predicting power of a model based on a constant alpha (Table 2) and Eqs. (9)–(13) for predicting the daily flow distribution in ungauged sites, we computed the 2 year flow (Q_2) at 278 sites above 600 m elevation by using both the best fit BPL function as well as a BPL function with the parameters derived with Eqs. (9)–(13). Fig. 14 presents a comparison between the Q_2 computed directly from the data on every site versus Q_2 computed by the fitted BPL. The correlation between these two estimates of Q_2 is strong ($r^2 = 0.99$, $p < 1E\text{-}15$) and highlights the strength of the fit to the daily flow frequency distributions. The relationship between Q_2 computed from the observations and Q_2 computed from the parameters estimated with the mentioned relations based on *DA*, *MAP*, and *BE* for sites located above 600 m (Eqs. (9)–(13)) has more scatter around the 1:1 but it is strong ($r^2 = 0.83$ $p < 1E\text{-}15$, Fig. 14). This means that knowing only basic watershed data (i.e. *DA*, *BE*, and *MAP*) one can use the model to estimate, at the order of magnitude level, the flow quartiles of an ungauged basin with $BE > 600$ mn.

4. Conclusions

The goal of this paper was to determine if the broken power law function was adequate to describe the flow frequency distribution of daily flows above the mean annual in 1217 sites across the CONUS and to study the spatial variability of the function parameters. The results indicate that the frequency distribution of daily flows is better described by power law functions than by the log-normal distribution. Among power functions, the 4-parameter broken power law (BPL) was superior over the other two alternatives power functions in 82% of the considered sites. The distribution of the BPL parameter that describes the slope of the first log segment, α, revealed low variability and highlighted the possibility of setting this parameter to a regional average reducing the number of free parameter from 4 to 3. The β parameter of the BPL that describes the slope of the second segment (i.e. frequency of extreme events) increases as a function of both elevation and regime predictability measured with the one-year autocorrelation indicating that predictable regimes (e.g. snowmelt) are characterize by high values of beta and therefore lower probability of high flows. We found that the a_1 parameter that defines the inflection point in the BPL distribution increases with drainage area, *DA*; however, we find that the relation between a_1 and *DA* is nonlinear. This implies a reduction in the frequency of large flows in the downstream direction. Finally the parameter a_0, can be computed for any location above 600 m based on *DA* and *MAP*. We believe that the broken power law function offers advantages over other functions to model the distributions of daily flows and should be considered in regional models to predict daily flows at ungauged locations.

References

Andrews, E.D., 1994. Marginal bed-load transport in a gravel-bed stream, Sagehen Creek, California. Water Resour. Res. 30.
Bevington, P., Robinson, D., 2003. Data Reduction and Error Analysis for the Physical Sciences. McGraw-Hill, Boston.
Castellarin, A., Galeati, G., Brandimarte, L., Montanari, A., Brath, A., 2004. Regional flow-duration curves reliability for ungauged basins. Adv. Water Resour. 27.
Doyle, M.W., Shields, C.A., 2008. An alternative measure of discharge effectiveness. Earth Surf. Process. Landforms 33, 308–316.
Emmett, W.W., Wolman, M.G., 2001. Effective discharge and gravel-bed rivers. Earth Surf. Process. Landforms 26.
Fennessey, N., Vogel, R., 1990. Regional flow-duration curves for ungauged sites in Massachusetts. J. Water Resour. Plan. Manage. ASCE 116, 530–549.
Goodwin, P., 2004. Analytical solutions for estimating effective discharge. J. Hydraulic Eng. Asce 130.
Kendall, M., Stuart, A., 1977. The Advanced Theory of Statistics. Macmillan, New York.

Mood, A., Graybill, F., 1963. Introduction to the Theory of Statistics. McGraw-Hill, New York.
Mueller, E., Pitlick, J., 2005. Morphologically based model of bed load transport capacity in a headwater stream. J. Geophys. Res.-Earth Surf. 110, F02016.
Nash, D.B., 1994. Effective sediment-transporting discharge from magnitude-frequency analysis. J. Geol. 102.
Newman, M., 2005. Power laws, Pareto distributions and Zipf's law. Contemporary Phys. 46, 323–351.
Pitlick, J., 1994. Relation between peak flows, precipitation, and physiography for 5 mountainous regions in the Western USA. J. Hydrol. 158.
Potter, K.W., 2001. A simple method for estimating baseflow at ungaged locations. J. Am. Water Resour. Assoc. 37.
Potter, K.W., Faulkner, E.B., 1987. Catchment response–time as a predictor of flood quantiles. Water Resour. Bull. 23.
Press, W., Teukolsky, S., Vetterling, W., Flannery, B., 2007. Numerical Recipes: The Art of Scientific Computing. Cambridge University Press, New York.
Sankarasubramanian, A., Vogel, R.M., 2002. Annual hydroclimatology of the United States. Water Resour. Res. 38, 1083.
Segura, C., Pitlick, J., 2010. Scaling frequency of channel-forming flows in snowmelt-dominated streams. Water Resour. Res. 46, W06524.
Sivapalan, M., 2003. Prediction in ungauged basins: a grand challenge for theoretical hydrology. Hydrol. Process. 17, 3163–3170.
Slack, J., Lumb, R., Landwehr, J., 1993. Hydro-climatic data network (HCDN) streamflow data set, 1874–1988. US Geol. Survey Water Resour. Invest. Rep., 93–4076.
Vogel, R.M., Fennessey, N.M., 1995. Flow duration curves. 2. A review of applications in water-resources planning. Water Resour. Bull. 31.
Vogel, R.M., Sankarasubramanian, A., 2000. Spatial scaling properties of annual streamflow in the United States. Hydrolo. Sci. J. Des. Sci. Hydrologiques. 45.
Vogel, R.M., Wilson, I., 1996. Probability distribution of annual maximum mean and minimum streamflows in the United States. J. Hydrol. Eng. 1.
Vogel, R.M., Stedinger, J.R., Hooper, R.P., 2003. Discharge indices for water quality loads. Water Resour. Res. 39, 1273.
Whiting, P.J., Samm, J.F., Moog, D.B., Orndorff, R.L., 1999. Sediment-transporting flows in headwater streams. Geol. Soc. Am. Bull. 111.
Wolman, M.G., Miller, J.P., 1960. Magnitude and frequency of forces in geomorphic processes. J. Geol. 68.