Water Resources Research

RESEARCH ARTICLE

10.1029/2018WR022732

Key Points:

- Climatology of tails of hourly precipitation extremes over the United States shows heavier tails than those commonly used in practice
- A striking coherent spatial pattern of tail's heaviness is revealed with mountainous regions showing heavier tails
- Simple parametric *law* found to describe the nonlinear increase in the heaviness of the tail with elevation

Correspondence to:
S. M. Papalexiou,
sm.papalexiou@usask.ca

Citation:
Papalexiou, S. M., AghaKouchak, A., & Foufoula-Georgiou, E. (2018). A diagnostic framework for understanding climatology of tails of hourly precipitation extremes in the United States. *Water Resources Research*, *54*, 6725–6738. https://doi.org/10.1029/2018WR022732

Received 7 FEB 2018
Accepted 13 AUG 2018
Accepted article online 24 AUG 2018
Published online 22 SEP 2018

A Diagnostic Framework for Understanding Climatology of Tails of Hourly Precipitation Extremes in the United States

Simon Michael Papalexiou[1,2,3], Amir AghaKouchak[3], and Efi Foufoula-Georgiou[3]

[1]Department of Civil, Geological and Environmental Engineering, University of Saskatchewan, Canada, [2]Global Institute for Water Security, Canada, [3]Department of Civil and Environmental Engineering, University of California, Irvine, CA, USA

Abstract Hourly precipitation extremes are crucial in hydrological design. Their frequency and magnitude is encapsulated in the probability distribution tail. Traditional extreme-analysis methods rely on theorems, like the Pickands-Balkema-de Haan, indicating specific type of tails assuming asymptotic convergence—a questionable assumption for real-world samples. Moreover, popular stochastic models for hourly precipitation presume light-tailed distributions to facilitate their mathematical formulation. In practice, limited information on hourly precipitation extremes makes identifying and quantifying their tail highly uncertain, especially on a station-by-station basis. Yet no comprehensive regional analysis of tails has been undertaken to quantify a *climatology of tails* for diagnostic and prognostic purposes. Here we undertake such an analysis for the conterminous United States. We introduce a novel Bayesian-adjustment approach to assess the best model between power-type and stretched-exponential tails showing that the latter performs better. We present climatology of the tail and quantify its heaviness in over 4,000 hourly precipitation records across the United States and present three main conclusions. First, we show that hourly precipitation tails are heavier than those commonly used with important implications including underestimation of extremes. Second, we provide spatial maps of the tail behavior which reveal some strikingly coherent spatial patterns that can be used for inference in the absence of local observations. Third, we find a nonlinear increase in the tail heaviness with elevation and we formulate parametric functions to describe this *law*. These results can improve the accuracy of frequency analysis, probabilistic prediction, rainfall-runoff modeling, and downscaling of historical observations and climate model projections.

1. Introduction and Motivation

The National Weather Service of the United States reports a recent 30-year average flood loss of $7.96 billion per year and 82 fatalities per year (Downton et al., 2005; see also Pielke et al., 2002), naming flash floods as the No. 1 weather-related hazard. Flash floods result from heavy storm events lasting several minutes to a few hours (e.g., Brooks & Stensrud, 2000; Georgakakos, 1986) highlighting the importance of studying extreme precipitation intensities and their related probabilities at subdaily timescales, such as hourly. Moreover, the output of rainfall-runoff models typically forced with hourly or subhourly precipitation series for decision support and operation depends on the quality of the precipitation input (e.g., Beven, 2012; Moradkhani & Sorooshian, 2009).

Over the recent years, predictive studies on future trends of hourly precipitation have mainly focused on the Clausius-Clapeyron relation and how hourly precipitation scales with air temperature, suggesting in some cases higher increases than the theoretically predicted Clausius-Clapeyron value of 7% per degree Celsius (Berg et al., 2009; Formayer & Fritz, 2017; Lenderink et al., 2011; Lenderink & van Meijgaard, 2008, 2010; Mishra et al., 2012; Shaw et al., 2011). Trends in hourly precipitation extremes have been investigated in only a few studies over the United States, for example, Muschinski and Katz (2013) analyzed a limited number of United states stations, and Yu et al. (2016) compared hourly and daily precipitation extremes and reported larger increases at the hourly scale; however, another study reported similar increases for daily and hourly scales in North America (Barbero et al., 2017).

Typically, stochastic modeling of hourly precipitation is mainly based on point process models like the Bartlett-Lewis and Neyman-Scott (Foufoula-Georgiou & Guttorp, 1987; Onof & Wheater, 1993; Rodriguez-Iturbe et al., 1987; Verhoest et al., 1997; Wheater et al., 2005). These models, however, are based on Exponential and Gamma distributions for nonzero hourly precipitation, that may underestimate return levels, an acknowledged fact that led to model alternatives trying to improve extremes' behavior (see e.g., Onof et al., 2005). Additionally, simulations from climate models, used for assessing future climate under different emission scenarios, are typically too coarse to be used directly as inputs into hydrologic/land-surface models. Thus, despite recent significant advances in regional climate modeling that can provide more reliable climate information, for example, convection-permitting models (Kendon et al., 2012; Prein et al., 2015), a realistic representation of precipitation still remains a challenge. For this reason, over the recent decades, many different downscaling schemes have been developed (e.g., regression and weather pattern methods, stochastic weather generators), several of which, however, acknowledge the underestimation of extreme precipitation (see e.g., Bárdossy & Pegram, 2011; Benestad, 2010; Fatichi et al., 2013; Foufoula-Georgiou et al., 2014; Hanel & Buishand, 2010; Maraun et al., 2010; Mezghani & Hingray, 2009; Perica & Foufoula-Georgiou, 1996; Venugopal et al., 1999; Vrac & Naveau, 2007; Wilby et al., 1998; Wilby & Wigley, 1997). Recent advances on stochastic modeling and disaggregation, however, make feasible to easily generate time series preserving any desired marginal distribution and correlation structure (including features like intermittency), allowing thus, realistic modeling of precipitation at any timescale including the hourly (Papalexiou, 2018; Papalexiou et al., 2018).

A systematic and extensive probabilistic investigation of hourly extremes is still an issue of theoretical and practical significance for improving probabilistic prediction, rainfall-runoff simulation of extremes, and downscaling of satellite precipitation products and outputs of climate model projections with emphasis on hazard-causing extreme events. While there are a number of studies on tail behavior of daily precipitation, a comprehensive study on hourly extremes is lacking. Here we analyze 4,137 hourly precipitation records across the United States exploring the behavior of hourly extremes. Our aim is threefold: (a) to outline a Bayesian-adjustment approach understanding and diagnosing the climatology of tail including the type of distribution that better describes hourly precipitation extreme; (b) to assess the tails' index values (or else shape parameters) across different climate/geographical regions quantifying the degree of heaviness, and thus the likelihood to generate extreme events; and (c) to investigate the spatial pattern of tail behavior across the United States trying to reveal their *climatology*.

2. Hourly Precipitation Data

We use the largest database of hourly precipitation records, available from NOAA, comprising 7,127 United states stations (for detailed information see Hammer & Steurer, 1997). Since we study the tail behavior, we need records with sufficiently large number of nonzero hourly precipitation values. Records vary in length ranging from 1 to 112 years but the total number of nonzero values depends also on the probability of dry hours and the percentage of missing values. Thus, we select stations having at least 3,000 nonzero values, a number we deem is large enough to study the tail behavior. Screening with this criterion results in a set of 4,316 stations; we show the selected station locations in Figure 1 (the set includes 16 and 105 stations not shown in the map, which are located, respectively, in Alaska and Hawaii). The average record length of the selected stations is approximately 43 years, while the average number of nonzero values is as high as 13,160 values. Note that several hourly records show inhomogeneities and changes in measurement resolution that have been acknowledged also in other studies (e.g., Barbero et al., 2017). These quality data issues may lead to inconsistencies when the whole sample is analyzed, yet they do not affect the results when the focus is on the extremes. Particularly, we used a subset of 1,684 records which do not show the aforementioned quality issues and the results were almost identical. Therefore, we used the whole data set to obtain a more robust representation of the spatial pattern of the tails.

3. Methods

3.1. Types and Definition of Tail

The tail function is typically defined as $\overline{F}_X(x) = 1 - F_X(x)$, which is the complementary cumulative distribution function, also known as the survival function (e.g., Klüppelberg, 1988). The term *tail*, however, is linked with extremes or large values and refers to complementary cumulative distribution function's upper part, that is, it quantifies the likelihood for extremes to occur. Regions with heavy tails should expect more frequent and larger extremes in relation to their *mean* or anticipated precipitation values. As a function, $\overline{F}(x)$ expresses how fast the exceedance probability tends to zero and depends on its mathematical form. Several attempts have been made to group tails according to their general properties and limiting

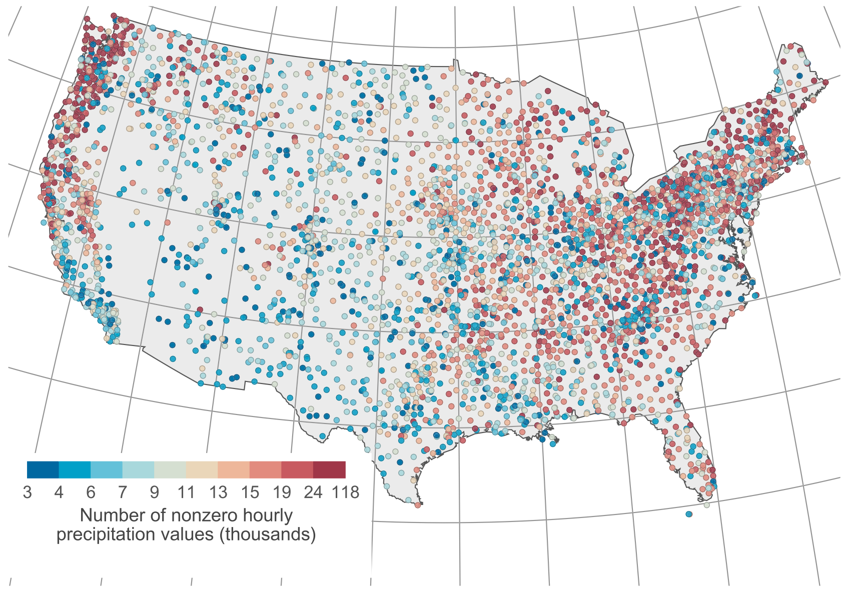

Figure 1. Location of the hourly precipitation stations; a total of 4,137 records having more than 3,000 nonzero values.

behavior (El Adlouni et al., 2008; Goldie & Klüppelberg, 1998; Werner & Upper, 2004). A major classification results by comparing a tail with the exponential one, characterizing slower or faster decreasing tails as subexponential or hyper-exponential, respectively. Also, many tails can be identified as equivalent, that is, the tails $\overline{F}(x)$ and $\overline{G}(x)$ are equivalent if $\lim_{x\to\infty}\overline{F}(x)/\overline{G}(x) = c$, where c is a constant. Here we use and compare two major types of tails based on their vast popularity across many scientific fields, that is, power-type, or else Pareto tails (e.g., Newman, 2005; Schroeder, 1991) and Weibull tails (e.g., Laherrère & Sornette, 1998) which include stretched-exponential, exponential, and hyper-exponential tails. Lognormal (LN) and Gamma (G) distributions are also popular models in hydrology, yet the LN distribution tail is similar to a power-type tail (Mitzenmacher, 2004) and the G distribution has essentially an exponential tail, and thus, can be obviously approximated very well by the Weibull tail which includes as special case (shape parameter equal to 1) the exponential.

Power-type tails (also known as algebraic tails) are mathematically defined by the tail of the Pareto II (PII) distribution (also known as Lomax)—the simplest power-type distribution defined in $(0, \infty)$. Its tail function is

$$\overline{F}_{\mathrm{PII}}(x) = \left(1 + \gamma \frac{x}{\beta}\right)^{-\frac{1}{\gamma}}, \tag{1}$$

where $\beta > 0$ and $\gamma > 0$ are scale and shape parameters, respectively, with the latter also known as the tail index (for $\gamma \to 0$ the tail becomes exponential). Note that the classical Pareto distribution (a straight line in log-log plot) is inconsistent with hydroclimatic positive variables as it cannot be defined in $(0, \infty)$, yet as x gets very large the PII tail gets equivalent to the classical Pareto tail, that is, $x^{-1/\gamma}$. Many well-known distributions are tail equivalent to the PII distribution, for example, the Generalized Beta of the second kind (Mielke Jr & Johnson, 1974); the Burr type III (known also as Dagum) and type XII (known also as Singh-Maddala) distributions (Burr, 1942; Papalexiou & Koutsoyiannis, 2012; Tadikamalla, 1980); the Beta of the second kind (McDonald & Xu, 1995); the log-logistic; the inverse generalized Gamma; the inverse Gamma; and others.

The Weibull (W) tail is of exponential form and defined by

$$\overline{F}_{\mathrm{W}}(x) = \exp\left(-\left(\frac{x}{\beta}\right)^{\gamma}\right), \tag{2}$$

with scale parameter $\beta > 0$ and tail index $\gamma > 0$. The W tail is subexponential (or else stretched-exponential), exponential, and hyper-exponential for tail index $\gamma > 1$, $\gamma = 1$, and $\gamma < 1$, respectively. Weibull tails are popular

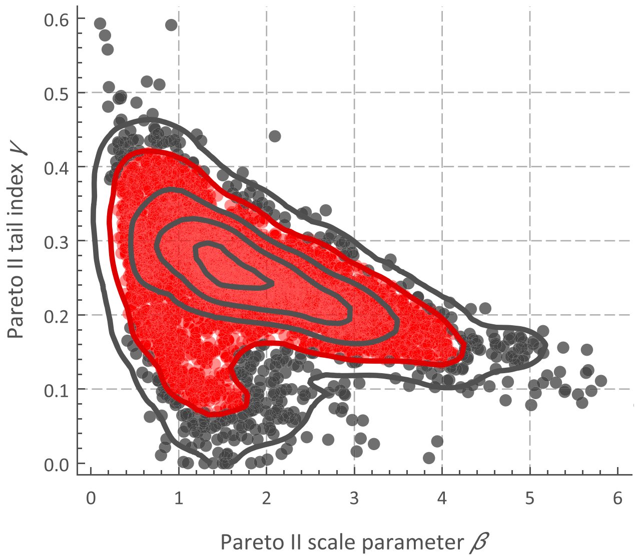

Figure 2. Estimated $(\hat{\beta}, \hat{\gamma})$ parameters for the Pareto II tail and for tail sample $S_{10\%}$. The contour lines indicate frequency; the red shaded region contains 90% of the points and is used to create the sampling distribution for the Monte Carlo simulations.

as the W distribution is extensively used in statistical and engineering literature as well as in other fields, for example, life sciences (Laherrère & Sornette, 1998). Popular distributions like Gamma (G) and Generalized Gamma (GG; Papalexiou & Koutsoyiannis, 2012; Stacy, 1962) are not exactly tail equivalent with the Weibull, yet inspection of their probability density functions, that is, $f_G(x) \sim x^{\gamma-1} \exp(-x)$ and $f_{GG}(x) \sim x^{\gamma_1 - 1} \exp(-x^{\gamma_2})$ reveals their structural similarity with the Weibull probability density function $f_W(x) \sim x^{\gamma-1} \exp(-x^{\gamma})$; note also that the exponential term quickly dominates the polynomial part.

The empirical tail cannot be uniquely defined and different approaches exist, for example, peak above threshold definitions using a percentage of the largest sample values or using a fixed number of peaks, for example, m largest values in an m-year sample (Buishand, 1989). Here we define the empirical tail, or the tail sample $S_{p\%}$, for tail level p% as

$$S_{p\%} = \{x_i | x_i \geq Q_N(1-p)\}, \tag{3}$$

where $Q_N(u) = F_N^{-1}(u)$ is the empirical quantile function. For example, the $S_{10\%}$ tail sample comprises the largest 10% nonzero values of an hourly precipitation record.

To investigate how the empirical tail affects the underlying tail assessment, we form for each precipitation record tail samples $S_{p\%}$ for $p = \{10, 5, 2, 1, 0.5\}$ and fit PII and W tails by minimizing the probability ratio mean square error (PRMSE; see Papalexiou et al., 2013, for a detailed assessment of the PRMSE norm) defined by

$$\text{PRMSE} = \frac{1}{n} \sum_{x_i \in S_{p\%}} \left(\frac{\overline{F}(x_i)}{\overline{F}_N(x_i)} - 1 \right)^2, \tag{4}$$

where the sum is over x_i values forming the tail sample $S_{p\%}$; n is the sample size of $S_{p\%}$; $\overline{F}(x_i)$ is the exceedance probability of x_i according to the theoretical tail (PII or W); $\overline{F}_N(x_i) = 1 - R(x_i)/(N+1)$ is the empirical exceedance probability (according to the Weibull plotting position) with $R(x_i)$ indicating the rank of x_i in the ascending ordered sample of all nonzero values; and N is the sample size of all nonzero values. Note that the PRMSE norm, which is a function of the parameters β and γ of the theoretical tail, uses relative errors between theoretical and empirical values, and thus, each point contributing in the sum is equally *weighted*.

Finally, we note that the analysis performed here can be also implemented seasonally, for example, investigating and comparing the tails in a monthly basis. However, here we studied the whole tail, as for many practical applications, we are not interested on the season that an event might occur. For example, Intensity-Duration-Frequency-Curves, probably the most commonly applied tool in hydrological design, are rarely developed in a seasonal basis.

3.2. Bayesian Adjustment Method

We assess the precision of the tail-fitting method to discriminate between PII and W tails through Monte Carlo (MC) simulations as described in the following steps: (a) we generate 1,000 random samples from each distribution with sample size, scale, and shape parameters varying randomly according to the observed empirical distributions, for example, using the distributions of tail indices given as box plots in Figure 5 (section 4.1 describes the details of results shown in Figure 5); and (b) we fit and compare the two tails according to the tail-fitting method applied to the observed hourly precipitation tails.

To clarify further the step (a) we note that in order for the MC simulation to be as realistic as possible, we studied the bivariate distribution of the estimated scale parameter and tail index, for example, Figure 2 depicts the estimated $(\hat{\beta}, \hat{\gamma})$ points of the PII tail for the $S_{10\%}$ precipitation samples, as well as various contour lines indicating the frequency of the points; red colored points mark a region comprising 90% of the points. Clearly, the two parameters are correlated with a correlation coefficient equal to -0.46 in this case. This

10.1029/2018WR022732

Table 1
Success and Failure Probabilities of the True Tail to be Identified Correctly and Erroneously, Respectively, Based on Monte Carlo Simulations

	Success and failure rates of the outcome									
	10		5		2		1		0.5	
p%	$\widehat{\text{PII}}$	$\widehat{\text{W}}$	$\widehat{\text{PII}}$	$\widehat{\text{W}}$	$\widehat{\text{PII}}$	$\widehat{\text{W}}$	$\widehat{\text{PII}}$	$\widehat{\text{W}}$	$\widehat{\text{PII}}$	$\widehat{\text{W}}$
PII	88.9	11.1	80.6	19.4	75.5	24.5	70.1	29.9	66.5	33.5
W	12.4	87.6	25.4	74.6	34.4	65.6	45.5	54.5	52.9	47.1

Note. True tails are in the first column. PII = Pareto II; W = Weibull.

implies that the (β, γ) points should be randomly sampled from a bivariate distribution. Here we use the 90% of *central* points (red-colored in Figure 2) to construct a smoothed nonparametric bivariate distribution that is used for random sampling. The same process, that is, of forming the nonparametric bivariate distribution, was applied for the parameters of both tails and for each of the five tail levels as parameter estimates differ among the different tail levels; in total, 10 sampling bivariate distributions were formed. We chose the 90% of central points to form the sampling distribution, instead of all the points, as the a posteriori $\left(\widehat{\beta}, \widehat{\gamma}\right)$ point estimates will generate a more *dispersed* distribution than the sampling one.

We present the MC results in Table 1 and depict success rates of each distribution tail in Figure 3. For example, for the $S_{5\%}$ case there is 80.6% probability to identify correctly a PII tail given that the true tail is PII, and 74.6% to identify correctly a W tail given that the true tail is W. There are two noteworthy findings in these results: first, the precision to identify a PII and a W tail differs with the PII tail having in general a higher probability to be identified correctly (exception is only the $S_{10\%}$ sample), and second, as we move from large tail samples to smaller ones, that is, from $S_{10\%}$ to $S_{0.5\%}$, the precision decreases for both tails.

Once the success/failure probabilities are known, we can correct the estimated percentage values given in Figure 4a based on Bayes formula which for two events A and B is given by

$$\Pr(A|B) = \Pr(A)\ \Pr(B|A)/\Pr(B). \tag{5}$$

The results presented in Figure 4a are percentage estimates, or else, probability estimates of an empirical tail to be PII or W; we denote these estimates (using hat notation) by $\widehat{p}_{\text{PII}} = \Pr\left(\widehat{\text{PII}}\right)$ and $\widehat{p}_{\text{W}} = \Pr\left(\widehat{W}\right)$, for example, for the $S_{10\%}$ sample $\widehat{p}_{\text{PII}} = 0.326$ and $\widehat{p}_{\text{W}} = 0.263$. Let us denote the true probabilities that we are trying to estimate with $p_{\text{PII}} = \Pr(\text{PII})$ and $p_{\text{W}} = \Pr(W)$. Also, we denote the results of the MC simulation given in Table 1 with the standard conditional notation, that is,

$$\begin{matrix} p_{11} = \Pr\left(\widehat{\text{PII}}|\text{PII}\right) & p_{12} = \Pr\left(\widehat{W}|\text{PII}\right) \\ p_{21} = \Pr\left(\widehat{\text{PII}}|W\right) & p_{22} = \Pr\left(\widehat{W}|W\right) \end{matrix}, \tag{6}$$

so that, for example, $p_{11} = \Pr\left(\widehat{\text{PII}}|\text{PII}\right)$ denotes the probability of an empirical tail to be identified as a PII given that the true tail is PII. If we assume that the real tail is either PII or W then it is obvious that

$$\begin{matrix} \Pr\left(\text{PII}|\widehat{\text{PII}}\right) + \Pr\left(W|\widehat{\text{PII}}\right) = 1 \\ \Pr\left(\text{PII}|\widehat{W}\right) + \Pr\left(W|\widehat{W}\right) = 1 \end{matrix}, \tag{7}$$

as for any given outcome (either $\widehat{\text{PII}}$ or $\widehat{W}$) the true tail is either PII or W since these are the only options, and thus these conditional events are clearly complementary. Now, if we replace the terms in equation (7) using Bayes formula and the definitions given in equation (6) we get

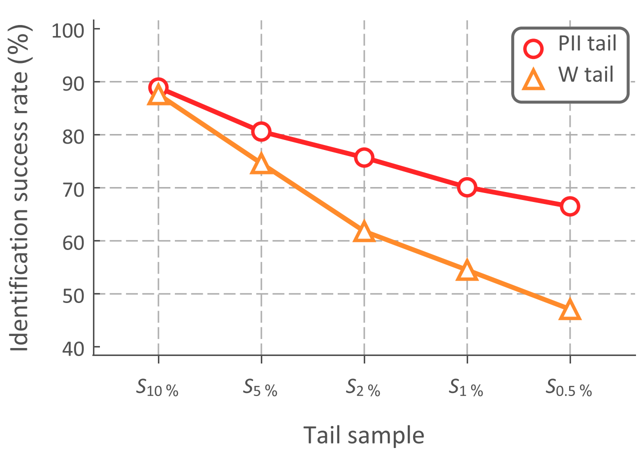

Figure 3. Success and failure probability for a tail to be identified correctly. PII = Pareto II; W = Weibull.

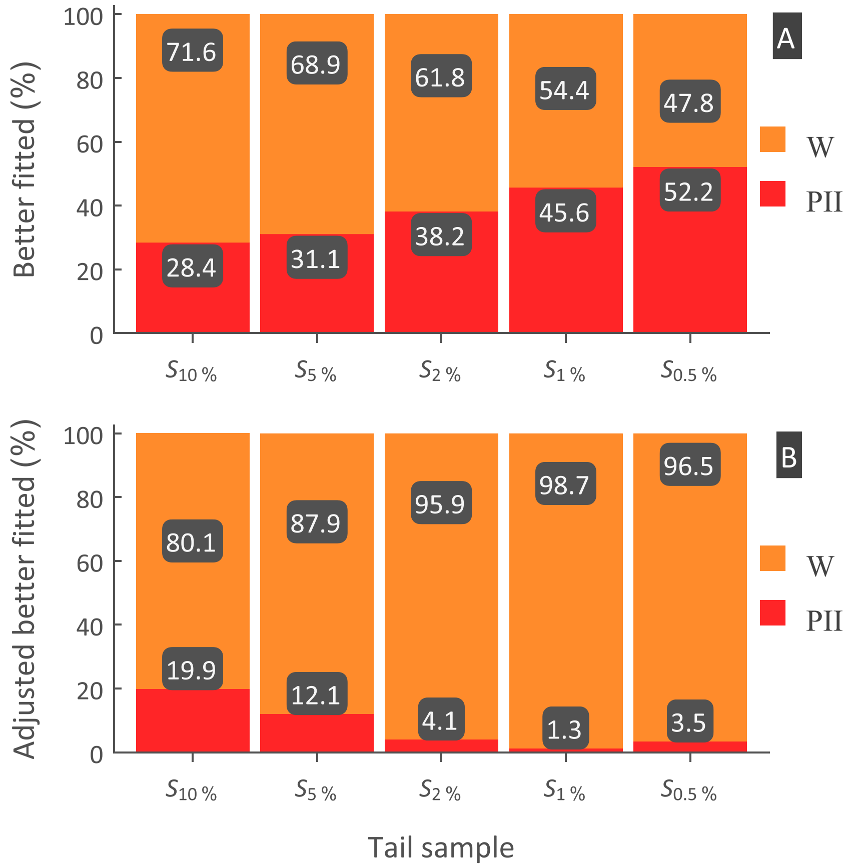

Figure 4. (a) Percentage of better-fitted tails to hourly precipitation extremes for various empirical tail samples; (b) Bayesian adjusted values. PII = Pareto II; W = Weibull.

$$p_{\mathrm{PII}}\frac{p_{11}}{\widehat{p}_{\mathrm{PII}}}+p_{\mathrm{W}}\frac{p_{21}}{\widehat{p}_{\mathrm{PII}}}=1 \tag{8}$$

$$p_{\mathrm{PII}}\frac{p_{12}}{\widehat{p}_{\mathrm{W}}}+p_{\mathrm{W}}\frac{p_{22}}{\widehat{p}_{\mathrm{W}}}=1.$$

Solving this linear system, we obtain the equations

$$p_{\mathrm{PII}}=\frac{p_{22}\widehat{p}_{\mathrm{PII}}-p_{21}\widehat{p}_{\mathrm{W}}}{p_{11}p_{22}-p_{12}p_{21}}, \tag{9}$$

$$p_{\mathrm{W}}=\frac{p_{11}\widehat{p}_{\mathrm{W}}-p_{12}\widehat{p}_{\mathrm{PII}}}{p_{11}p_{22}-p_{12}p_{21}}$$

which are used to adjust the initial estimates given in Figure 4a.

4. Results

4.1. Fitting Results

We compare the fitted tails in terms of the resulting PRMSE, that is, for a given tail level (e.g., 10%) we fit in every tail sample the PII and W tails identifying as better fit the tail with the smaller PRMSE. Eventually, we calculate how many times (percentage of stations) each tail performed better than the other, repeating this process for the five tail levels p we study; see Figure 4a for the estimated percentages. We observe a change in the tails' fitting performance among tail levels. There is a systematic *trend* indicating that the PII tail increases its performance from 28.4% to 52.2% as we move from $S_{10\%}$ to $S_{0.5\%}$. One could rationally

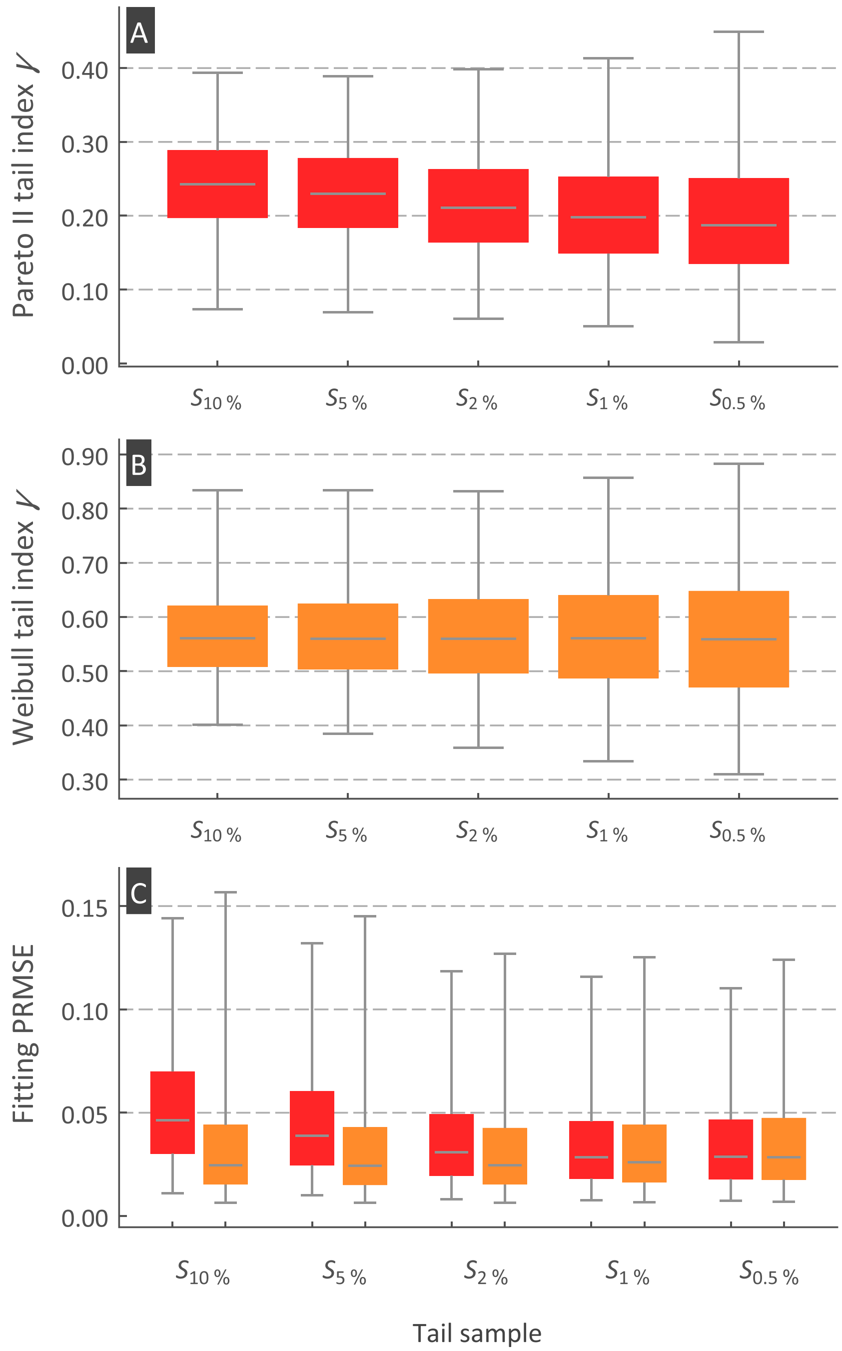

Figure 5. Box plots of the estimated (a) Pareto II and (b) Weibull tail indices as well as the resulting (c) fitting error. The whiskers define the 95% empirical confidence interval.

assume that the PII tail would increase further its fitting performance if additional data were available and the study of empirical tails at levels smaller than 0.5% was meaningful. Given this, and combined with the fact that lower level tail samples are, at least theoretically, more representative of the true tail as they can provide a clearer picture of tail's behavior, one may draw the conclusion that the PII tail is a better model (compared to the W tail) for the description of extreme hourly precipitation.

We stress that identifying correctly the type of tail is not trivial and any method used to compare and assess which model is better cannot be absolutely precise. Knowing the success/failure identification

10.1029/2018WR022732

rates of a method may significantly alter the results. In MC simulations we know a priori the true tail, that is, the model used to generate a random sample, yet if we fit and compare two models, the true model is not always identified as the best in terms of fitting due to sample variations. In section 3.2 we estimate the success/failure identification rates based on MC simulations and we adjust the empirical results based on the Bayesian framework described. We find that the adjusted results, given in Figure 4b, provide a different picture than the results in Figure 4a. Particularly, the adjusted results show that the performance of the PII tail actually decreases as lower tail levels are considered, leading eventually to complete domination of the W tail at the 0.5% level. Note that to estimate the success/failure rate of the method we performed simulations where we assumed that the underlying true tail of the empirical samples is either PII or W and not of any other type (see section 3.2). This assumption could be modified, for example, by including additional models like the LN or G tails, yet here we restricted our choice between the PII and W tails which performed well in our preliminary fitting. We note that both models fit equally well (especially for $S_{1\%}$ and $S_{0.5\%}$; see Figure 5c), supporting both PII and W tails as good tail choices.

These results are enhanced by studying the estimated tail indices presented in the box plots of Figure 5. As we move from $S_{10\%}$ to $S_{0.5\%}$ the PII shape parameters constantly decrease, for example, the median values decrease from 0.24 to 0.19 (see Figure 5a). This implies that the Pareto tail is not a robust choice for hourly precipitation (at least for this data set), as up to the 0.5% tail level we cannot assess the tail index, or the value it converges to, or if it converges at all. The conditional Pareto distribution for $x > x_p$ (where x_p is the quantile value for probability p) is also Pareto with the same tail index, and thus, the tail index should be the same at all tail levels (expecting of course a larger variance in lower tail levels due to smaller samples). In contrast, the W tail index has a constant median equal to 0.56 at all tail levels (see Figure 5b) and a very narrow interquartile range. The upper limit of the 95% empirical confidence intervals, expressed by the outer box-plot fences, is less than 0.9, implying that almost all samples analyzed have a stretched-exponential tail, that is, heavier than the exponential.

4.2. Spatial Variation of Tails and Relation With Elevation

We present the spatial variability of the estimated PII and W tail indices at the 0.5% tail level (Figure 6)—we chose the 0.5% level as it is closer to the true tail and offers better assessment of the PII tail index (see the convergence issue of PII in Figure 5a). Both maps (created by applying ordinary kriging with spherical semi-variogram and search radius five points, and Gaussian smoothing taking into account the eight neighboring pixels) show the same spatial pattern revealing strong spatial coherence but also large variability. Three major regions emerge: the first, comprising the *heaviest* tails ($0.26 \le \gamma_{\mathrm{PII}} < 0.33$ and $0.38 \le \gamma_{\mathrm{W}} < 0.50$), is located at the mountainous regions of the United States, that is, it includes Sierra Nevada and Cascade Mountains, the Great Basin, the Rocky Mountains, and the northernmost part of the Great Plains; relatively heavy tails are also observed in the Appalachian Mountains and the state of Maine. The second region, with relatively *thin* tails ($0.10 \le \gamma_{\mathrm{PII}} < 0.18$ and $0.60 \le \gamma_{\mathrm{W}} < 0.74$), is located in the Southern United States, spanning the Gulf Coastal Plains and including parts of the Atlantic Coastal Plains from Florida to North Carolina; other regions with relatively thin tails are located in most of the Pacific Coast and all over California. The third region, with relatively *moderate* tails ($0.18 \le \gamma_{\mathrm{PII}} < 0.26$ and $0.50 \le \gamma_{\mathrm{W}} < 0.60$), starts from south Arizona and New Mexico extends across the Great Plains and continues in most parts of the Midwest and Northeast regions around the Appalachian Mountains. This first analysis of the spatial pattern of the tail's heaviness seems to reveal a relation with elevation. Heavy tails are noted in high elevations, however, this does not indicate the extent to which elevation is a determinant of the tails of the distribution as many other interacting factors (e.g., synoptic conditions, vertical stability, and moisture convergence) contribute to the intensity of rainfall. A relatively thin tail is observed in southeastern United States where rainfall is mainly convective (short-term intense rain over a smaller spatial extent compared to a typical stratiform event). However, different regions receive various types of storms throughout the year. For example, the southeastern United States receives heavy rainfall through short-term convective storms, tropical cyclones, and large scale frontal systems. For this reason, the tail parameter cannot be simply attributed to a particular storm type or a single regional physical process. The tail parameter represents the overall behavior of extremes considering samples from different storms over a long period of time. A discussion on extreme rainfall and flood peaks mechanisms over the United States can be found, respectively, in Barbero et al. (2018) and Villarini and Smith (2010).

10.1029/2018WR022732

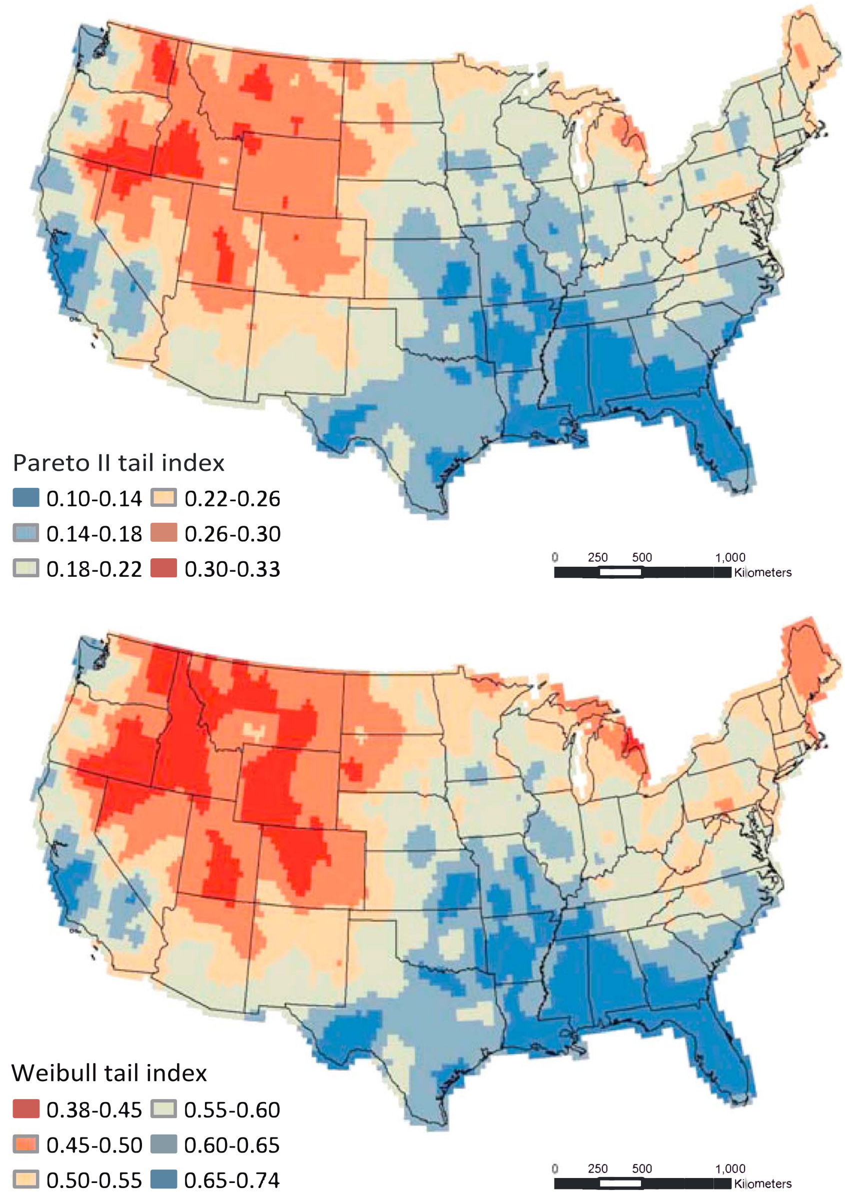

Figure 6. Spatial variation of Pareto II and Weibull tail indices indicating the heaviness of the tail for hourly precipitation.

It is noteworthy that a comparison between the spatial variation of the tail indices (Figure 6) and the main climate classes of the Köppen-Geiger classification system (e.g., Kottek et al., 2006) reveals some similarities. For example, the south-eastern part of the United States which is classified as Warm Temperate shows thinner tails, while heavier tails are observed in regions classified as Arid and Snow. However, the pattern is not clear, for example, regions classified in the Snow class over the United States show different tail heaviness. Of course, some similarities between the spatial pattern of the tail's heaviness and the Köppen-Geiger classes are anticipated. Particularly, the relation of tail's heaviness with elevation is expected to create a link with temperature, as higher elevation typically implies lower temperatures. This in turn could create relations with the Köppen-Geiger classes as this classification system uses temperature values, among other variables like precipitation, to define the climatic class of a region.

The previous comparison of the tail indices indicates a strong relationship with elevation, and yet Pearson's ρ and Kendall's τ correlation coefficients between elevation H and γ_{PII} or γ_{W} (calculated using all 4703 $(H, \gamma_{\mathrm{PII}})$ and (H, γ_{W}) points) show low correlation, that is, $\rho(H, \gamma_{\mathrm{PII}}) = 0.30$, $\tau(H, \gamma_{\mathrm{PII}}) = 0.22$, and

10.1029/2018WR022732

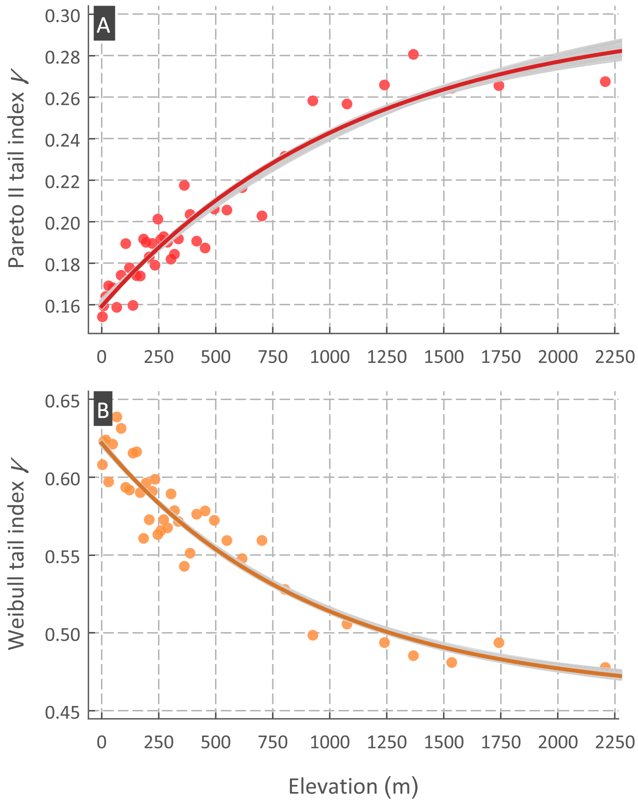

Figure 7. (a) Pareto II and (b) Weibull tail indices versus elevation. The gray shaped areas result from ensemble curves; the colored curves within the gray areas are the *median* curves given in equation (10); the cloud of points is for block size $n = 100$.

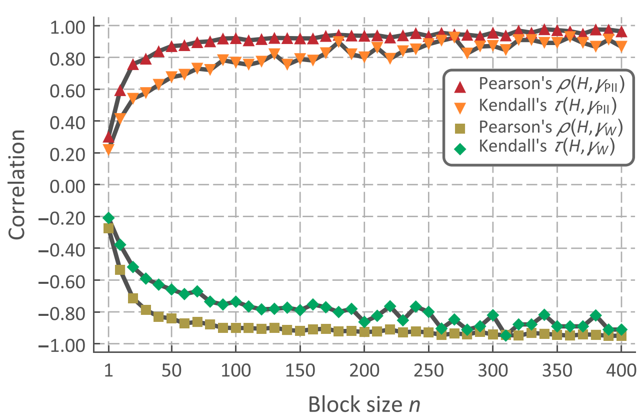

Figure 8. Pearson's and Kendall's correlation coefficients of Pareto II and Weibull tail indices versus elevation for various block sizes.

$\rho(H, \gamma_W) = -0.27$, $\tau(H, \gamma_W) = -0.21$. This fact is not surprising as the tail indices are strongly affected by random sample variations that blur the spatial or elevation-dependent patterns. To reveal the relationship of the tail index with elevation, we sort the 4073 (H, γ_{PII}) and (H, γ_W) points in ascending elevation order and create consecutive blocks of n-points (with n varying from 10 to 400 increasing in steps of 10) and estimate each block's mean elevation $\overline{H}$ and mean $\overline{\gamma}_{PII}$ or $\overline{\gamma}_W$. For example, for $n = 100$, 40 blocks are formed resulting in 40 $(\overline{H}, \overline{\gamma}_{PII})$ or $(\overline{H}, \overline{\gamma}_W)$ points (see Figure 7). This averaging process increases vastly the correlation coefficient values, as the block size increases (see Figure 8), from low to very high values above 0.9.

This increase in correlation (especially in Pearson's ρ) does not necessarily imply a linear relationship between tail indices and elevation. Actually, the cloud of points (see Figures 7a and 7b depicting the points for block size $n = 100$) indicate concave increase and convex decrease, respectively, for γ_{PII} and γ_W over elevation. We deem a function $g(x)$ interpolating these points should be real at $H = 0$ and tend to a constant as $H \to \infty$ (very large values) given that an infinite increase of the tail index with elevation is physically inconsistent. One function with these properties is $g(H) = a + b\exp(-cH)$, with $a \in \mathbb{R}$ and $c > 0$; $g(x)$ is concave for $b > 0$ and convex for $b < 0$, while $g(0) = a + b$ and $\lim_{H\to\infty} g(H) = a$.

We estimate the parameters a, b, and c, by least square fitting, yet for different block sizes different clouds of $(\overline{H}, \overline{\gamma}_{PII})$ and $(\overline{H}, \overline{\gamma}_W)$ points are formed potentially affecting the estimated parameters. Selecting a unique block size cannot be theoretically justified, thus, we fit $g(x)$ to all clouds of points for $n = \{10, 20, \ldots, 400\}$, that is, a total of 40 different clouds. We show the ensemble of fitted curves in Figures 7a and 7b as gray shaded areas (remarkably narrow) revealing that block size does not affect significantly the estimated parameters. We provide characteristic curves for the tail indices γ_{PII} and γ_W versus elevation by calculating the median values of the 40 estimated a, b, and c, parameters, resulting in

$$\begin{aligned}\gamma_{PII}(H) &= 0.30 + 0.14\ \exp(-0.0009H)\\ \gamma_W(H) &= 0.46 - 0.16\ \exp(-0.0011H)\end{aligned}, \tag{10}$$

where elevation H is in meters. The curves are shown in Figures 7a and 7b and for $H = 0$ m the tail-index values are $\gamma_{PII} = 0.16$ and $\gamma_W = 0.62$, while their maxima for $H \to \infty$ are $\gamma_{PII} = 0.30$ and $\gamma_W = 0.46$.

As an example, we use a 50-year hourly record from Lake Charles regional Airport in Louisiana to fit the Gamma, the Burr type XII (BrXII; three-parameter generalization of the PII), and the Generalized Gamma (GG; three-parameter generalization of Weibull or Gamma) distributions using fixed PII and W tail indices for the latter two, selected from the maps of Figure 6. In this way we avoid the uncertainty to assess the tail index of a three-parameter distribution while the remaining two parameters are estimated to preserve the mean and standard deviation of the empirical sample (of course maximum likelihood, L-moments, or other methods can be used to estimate the remaining two parameters). The mean and standard deviation of nonzero values are, respectively, 3.29 and 5.83 mm, with probability of dry hours 95.1%. According to the maps the PII and W tail indices in this area are, respectively, 0.12 and 0.7 (we use the mean of the class). The fitted BrXII and GG distributions (see

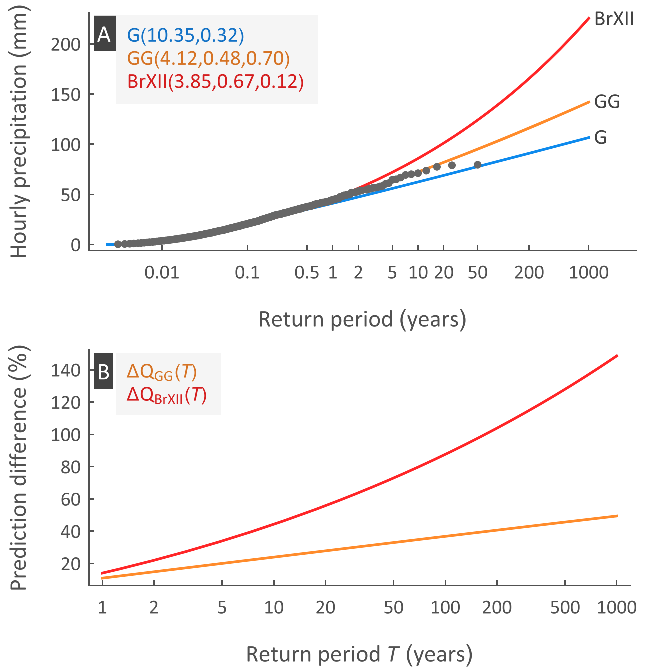

Figure 9. (a) Fitted G, BrXII, and GG distributions for an hourly precipitation record of Lake Charles regional Airport in Louisiana. Tail indices of BrXII and GG distributions were fixed from the maps of Figure 6. (b) Prediction difference of the GG and the BrXII distributions compared to a G distribution for a typical hourly precipitation record in the United States. G = Gamma; BrXII = Burr type XII; GG = Generalized Gamma.

Figure 9a) predict for T = 100 years, respectively, 75% and 25% larger precipitation than the fitted G distribution, while for T = 1,000 years the corresponding values are 111% and 33% larger.

A more general demonstration of the implication that a misjudged tail might have on the estimation of design rainfall amounts is shown by the relative difference of predicted amounts by three-parameter distributions with a fixed tail compared to predicted amounts by a light-tailed Gamma distribution (Figure 9b). Particularly, we consider a characteristic hourly precipitation sample with coefficient of variation $C_V = \sigma/\mu \cong 1.8$ and probability of dry hours $p_0 = 0.90$. Assuming an arbitrary mean value μ (results do not change with different μ as long as C_V remains the same), we fit the G, GG, and BrXII distributions by preserving the mean and standard deviation (obviously $\sigma = \mu \times C_V$) and using as tail index for the Generalized Gamma and BrXII the corresponding median values, that is, 0.56 and 0.19, respectively. Specifically, we show (Figure 9b) the relative differences of GG and BrXII predictions to G predictions defined, respectively, as $\Delta Q_{GG}(T) = Q_{GG}(T)/Q_G(T) - 1$ and $\Delta Q_{BrXII}(T) = Q_{BrXII}(T)/Q_G(T) - 1$. We show that for a typical hourly precipitation record and for T = 1000 years the GG and the BrXII predict approximately 50% and 150% more precipitation than the G estimates.

From the previous analysis, it is clear that models like the G distribution can underestimate return levels since they cannot adapt their tail to the behavior of observed extremes, as quantified and shown in the maps. On the other hand, two parameter models like the PII or the LN, although having a shape parameter to adjust the tail's heaviness, are not flexible enough. More specifically, a two parameter model fitted using the method of moments (either product or L-moments) preserves two moments which usually are the first and second as they represent important location and dispersion measures. Thus, the heaviness of the tail emerges as a *byproduct* of the estimated parameters without explicitly taking into account the behavior of extremes. Certainly, the information summarized in statistics like the mean and variance is not sufficient to characterize either the general shape or the distribution tail. Other generic fitting methods like maximum likelihood or least squares do not preserve any moment, but rather try to determine the *best* distribution describing the data by maximizing or minimizing, respectively, a specific norm. Again, these approaches (unless special modifications are considered, e.g., adding weights in data values) tend to express the majority of the data which clearly are not the extreme values. Thus, two-parameter models do not offer the desired flexibility to describe in general the plethora of observed distribution *shapes*. Actually, regarding precipitation at daily scale, a global investigation showed clearly that two-parameter models cannot adequately reproduce its statistical properties (Papalexiou & Koutsoyiannis, 2016).

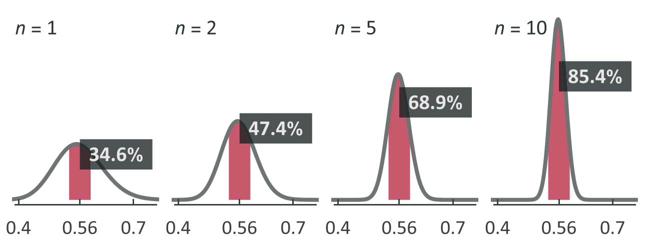

Figure 10. Distribution of the tail index estimated from random samples from a Weibull distribution with theoretical tail index $\gamma = 0.56$ (equal to the observed one for hourly precipitation). Highlighted areas and labels show the probability to estimate γ with error less than 5% based on single *station* estimates or on regional values by averaging estimates in 2, 5, and 10 stations.

Of course three-parameter models are less parsimonious, increasing the parametric uncertainty in estimating the tail index. We can improve the accuracy of this estimation, however, by using regionalized values of tail indices instead of the station ones. Regionalized tail indices are estimated focusing only on extreme values while the regional average is more robust than single-station estimates when clear spatial patterns emerge (as we show in this case) that imply the existence of spatial homogeneity in the behavior of extreme rain. This increased robustness in estimation is expected, as regionalization increases the sample size by substituting space for time. We demonstrate this by a MC experiment, that is, we generate samples from a W distribution with tail index $\gamma = 0.56$ and we calculate the probability of a posteriori estimating γ with error less than 5% assuming the estimates are from a single station (n = 1) or from the average value of $n = \{2, 5, 10\}$ stations. We form the distribution of the estimator for each n and calculate the corresponding probabilities (Figure 10). For example, the probability when $n = 1$ is just 34.6% and doubles for $n = 5$. In this way we can simplify and make more robust the estimation process of fitting a three-parameter distribution by aligning the parameter controlling the tail with the regional tail index and estimate the rest two parameters using any method, for example, we can use the method of moments to preserve the mean and standard deviation of the observed data. Of course, it should be clear that this MC experiment demonstrates the gain in estimation accuracy assuming that stations are independent. In the presence of spatiotemporal correlations this increase in the estimation accuracy is expected to be smaller depending of course on the strength of the correlations (e.g., Douglas et al., 2000).

5. Summary and Conclusions

We analyzed the largest available database of hourly precipitation records in the United States investigating the behavior and climatology of tails of extreme hourly precipitation. From 7,127 stations we selected 4,316 stations with more than 3,000 nonzero values. In this paper, we outlined a novel Bayesian-adjustment approach for understanding and diagnosing the tails of hourly precipitation extremes. We explored three questions: (a) which distribution tail between power-type and Weibull tails better describes hourly precipitation extremes and how we can create accurate diagnostics, (b) what is the range of the tail-index values, and (c) what is their spatial variation, or else their climatology.

Both tails performed well especially for empirical tail samples comprising the most extreme values, that is, for $S_{1\%}$ and $S_{0.5\%}$ (see Figures 4a and 5c), yet a comparison after adjusting the results based on a Bayesian framework indicated the Weibull tail as a more robust model for hourly precipitation extremes over the United States. The Weibull tail index γ_W is remarkably stable across all tail samples studied with a median value equal to 0.56 (Figure 5b). The PII tail index γ_{PII} is affected by the empirical tail sample, with a median value of 0.24 at $S_{10\%}$ constantly decreasing to 0.19 at $S_{0.5\%}$ (Figure 5a), indicating that the Pareto tail may not be a robust choice since the value at which it converges (if it does) is unknown. Both the PII and W tail-index values showed that hourly precipitation has a heavy (subexponential) tail, much heavier than exponential or Gamma tails that would significantly underestimate precipitation over large return periods. The spatial variation of both tail indices shows a coherent pattern (see Figure 6) with mountainous areas exhibiting heavier tails. We revealed a nonlinear increase of the tail's *heaviness* with elevation and we provided corresponding functions for the PII and W tail indices (see equation (10) and Figure 7). However, we stress that elevation cannot be the sole determinant of the tail's heaviness as regional precipitation characteristics and physical mechanisms also affect extremes' behavior. These functions as well as the spatial variation maps in Figure 6 can be used as guidelines, or as a first choice, in selecting tail index values for probabilistic and stochastic modeling of hourly precipitation, which in turn can lead to improved rainfall-runoff modeling. Finally, the spatial patterns of tail's heaviness shown here can also be used to improve downscaling of historical observations and climate model projections.

Acknowledgments
We thank the three reviewers and the Associate Editor for their thoughtful reviews towards improving the manuscript as well as S. Batelis for his help in constructing the maps of Figure 6. This study was partially supported by National Science Foundation (NSF) Grants (NSF WSC: EAR-1209402, NSF LIFE:EAR-1242458, NSF CMMI-1635797), National Aeronautics and Space Administration (NASA) Grant (NNX16AO56G), National Oceanic and Atmospheric Administration (NOAA) Grant (NA14OAR4310222), and California Energy Commission Grant (500-15-005). The data used in this study are freely available at https://www.ncdc.noaa.gov/.

References

Barbero, R., Abatzoglou, J. T., & Fowler, H. J. (2018). Contribution of large-scale midlatitude disturbances to hourly precipitation extremes in the United States. *Climate Dynamics*. https://doi.org/10.1007/s00382-018-4123-5

Barbero, R., Fowler, H. J., Lenderink, G., & Blenkinsop, S. (2017). Is the intensification of precipitation extremes with global warming better detected at hourly than daily resolutions? *Geophysical Research Letters*, *44*, 974–983. https://doi.org/10.1002/2016GL071917

Bárdossy, A., & Pegram, G. (2011). Downscaling precipitation using regional climate models and circulation patterns toward hydrology. *Water Resources Research*, *47*, W04505. https://doi.org/10.1029/2010WR009689

Benestad, R. E. (2010). Downscaling precipitation extremes. *Theoretical and Applied Climatology*, *100*(1–2), 1–21. https://doi.org/10.1007/s00704-009-0158-1

Berg, P., Haerter, J. O., Thejll, P., Piani, C., Hagemann, S., & Christensen, J. H. (2009). Seasonal characteristics of the relationship between daily precipitation intensity and surface temperature. *Journal of Geophysical Research*, *114*, D18102. https://doi.org/10.1029/2009JD012008

Beven, K. J. (2012). *Rainfall-runoff modelling: The primer*. Chichester, UK: John Wiley & Sons.

Brooks, H. E., & Stensrud, D. J. (2000). Climatology of heavy rain events in the United States from hourly precipitation observations. *Monthly Weather Review*, *128*(4), 1194–1201. https://doi.org/10.1175/1520-0493(2000)128<1194:COHREI>2.0.CO;2

Buishand, T. A. (1989). The partial duration series method with a fixed number of peaks. *Journal of Hydrology*, *109*(1–2), 1–9. https://doi.org/10.1016/0022-1694(89)90002-4

Burr, I. W. (1942). Cumulative frequency functions. *The Annals of Mathematical Statistics*, *13*(2), 215–232. https://doi.org/10.1214/aoms/1177731607

Douglas, E. M., Vogel, R. M., & Kroll, C. N. (2000). Trends in floods and low flows in the United States: Impact of spatial correlation. *Journal of Hydrology*, *240*(1-2), 90–105. https://doi.org/10.1016/S0022-1694(00)00336-X

Downton, M. W., Miller, J. Z. B., & Pielke, R. A. Jr. (2005). Reanalysis of US National Weather Service flood loss database. *Natural Hazards Review*, *6*(1), 13–22. https://doi.org/10.1061/(ASCE)1527-6988(2005)6:1(13)

El Adlouni, S., Bobée, B., & Ouarda, T. B. M. J. (2008). On the tails of extreme event distributions in hydrology. *Journal of Hydrology*, *355*(1–4), 16–33. https://doi.org/10.1016/j.jhydrol.2008.02.011

Fatichi, S., Ivanov, V. Y., & Caporali, E. (2013). Assessment of a stochastic downscaling methodology in generating an ensemble of hourly future climate time series. *Climate Dynamics*, *40*(7–8), 1841–1861. https://doi.org/10.1007/s00382-012-1627-2

Formayer, H., & Fritz, A. (2017). Temperature dependency of hourly precipitation intensities – Surface versus cloud layer temperature. *International Journal of Climatology*, *37*(1), 1–10. https://doi.org/10.1002/joc.4678

Foufoula-Georgiou, E., Ebtehaj, A. M., Zhang, S. Q., & Hou, A. Y. (2014). Downscaling satellite precipitation with emphasis on extremes: A variational ℓ1-norm regularization in the derivative domain. *Surveys in Geophysics*, *35*(3), 765–783. https://doi.org/10.1007/s10712-013-9264-9

Foufoula-Georgiou, E., & Guttorp, P. (1987). Assessment of a class of Neyman-Scott models for temporal rainfall. *Journal of Geophysical Research*, *92*(D8), 9679–9682. https://doi.org/10.1029/JD092iD08p09679

Georgakakos, K. P. (1986). On the design of national, real-time warning systems with capability for site-specific, flash-flood forecasts. *Bulletin of the American Meteorological Society*, *67*(10), 1233–1239. https://doi.org/10.1175/1520-0477(1986)067<1233:OTDONR>2.0.CO;2

Goldie, C. M., & Klüppelberg, C. (1998). Subexponential distributions. In *A practical guide to heavy tails: Statistical techniques and application* (pp. 435–459). Cambridge, MA: Birkhauser Boston Inc.

Hammer, G. R., & Steurer, P. M. (1997). Data set documentation for hourly precipitation data. *NOAA/NCDC TD3240 Documentation Series, Asheville, NC*, *18*.

Hanel, M., & Buishand, T. A. (2010). On the value of hourly precipitation extremes in regional climate model simulations. *Journal of Hydrology*, *393*(3–4), 265–273. https://doi.org/10.1016/j.jhydrol.2010.08.024

Kendon, E. J., Roberts, N. M., Senior, C. A., & Roberts, M. J. (2012). Realism of rainfall in a very high-resolution regional climate model. *Journal of Climate*, *25*(17), 5791–5806. https://doi.org/10.1175/JCLI-D-11-00562.1

Klüppelberg, C. (1988). Subexponential distributions and integrated tails. *Journal of Applied Probability*, *25*(01), 132–141. https://doi.org/10.2307/3214240

Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World map of the Koppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, *15*(3), 259–263. https://doi.org/10.1127/0941-2948/2006/0130

Laherrère, J., & Sornette, D. (1998). Stretched exponential distributions in nature and economy: "Fat tails" with characteristic scales. *The European Physical Journal B - Condensed Matter and Complex Systems*, *2*(4), 525–539. https://doi.org/10.1007/s100510050276

Lenderink, G., Mok, H. Y., Lee, T. C., & Van Oldenborgh, G. J. (2011). Scaling and trends of hourly precipitation extremes in two different climate zones–Hong Kong and the Netherlands. *Hydrology and Earth System Sciences*, *15*(9), 3033–3041. https://doi.org/10.5194/hess-15-3033-2011

Lenderink, G., & van Meijgaard, E. (2008). Increase in hourly precipitation extremes beyond expectations from temperature changes. *Nature Geoscience*, *1*(8), 511–514. https://doi.org/10.1038/ngeo262

Lenderink, G., & van Meijgaard, E. (2010). Linking increases in hourly precipitation extremes to atmospheric temperature and moisture changes. *Environmental Research Letters*, *5*(2), 025208. https://doi.org/10.1088/1748-9326/5/2/025208

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., et al. (2010). Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*, *48*, RG3003. https://doi.org/10.1029/2009RG000314

McDonald, J. B., & Xu, Y. J. (1995). A generalization of the beta distribution with applications. *Journal of Econometrics*, *66*(1-2), 133–152. https://doi.org/10.1016/0304-4076(94)01612-4

Mezghani, A., & Hingray, B. (2009). A combined downscaling-disaggregation weather generator for stochastic generation of multisite hourly weather variables over complex terrain: Development and multi-scale validation for the Upper Rhone River basin. *Journal of Hydrology*, *377*(3–4), 245–260. https://doi.org/10.1016/j.jhydrol.2009.08.033

Mielke, P. W. Jr., & Johnson, E. S. (1974). Some generalized beta distributions of the second kind having desirable application features in hydrology and meteorology. *Water Resources Research*, *10*(2), 223–226. https://doi.org/10.1029/WR010i002p00223

Mishra, V., Wallace, J. M., & Lettenmaier, D. P. (2012). Relationship between hourly extreme precipitation and local air temperature in the United States. *Geophysical Research Letters*, *39*(16), L16403. https://doi.org/10.1029/2012GL052790

Mitzenmacher, M. (2004). A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, *1*(2), 226–251. https://doi.org/10.1080/15427951.2004.10129088

Moradkhani, H., & Sorooshian, S. (2009). General review of rainfall-runoff modeling: Model calibration, data assimilation, and uncertainty analysis. In S. Sorooshian, K. L. Hsu, E. Coppola, B. Tomassetti, M. Verdecchia, & G. Visconti (Eds.), *Hydrological modelling and the water cycle, Water Science and Technology Library* (Vol. 63). Berlin: Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-77843-1_1

Muschinski, T., & Katz, J. I. (2013). Trends in hourly rainfall statistics in the United States under a warming climate. *Nature Climate Change*, *3*(6), 577–580. https://doi.org/10.1038/nclimate1828

Newman, M. E. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, *46*(5), 323–351. https://doi.org/10.1080/00107510500052444

Onof, C., Townend, J., & Kee, R. (2005). Comparison of two hourly to 5-min rainfall disaggregators. *Atmospheric Research*, *77*(1–4), 176–187. https://doi.org/10.1016/j.atmosres.2004.10.022

Onof, C., & Wheater, H. S. (1993). Modelling of British rainfall using a random parameter Bartlett-Lewis rectangular pulse model. *Journal of Hydrology*, *149*(1-4), 67–95. https://doi.org/10.1016/0022-1694(93)90100-N

Papalexiou, S. M. (2018). Unified theory for stochastic modelling of hydroclimatic processes: Preserving marginal distributions, correlation structures, and intermittency. *Advances in Water Resources*, *115*, 234–252. https://doi.org/10.1016/j.advwatres.2018.02.013

Papalexiou, S. M., & Koutsoyiannis, D. (2012). Entropy based derivation of probability distributions: A case study to daily rainfall. *Advances in Water Resources*, *45*, 51–57. https://doi.org/10.1016/j.advwatres.2011.11.007

Papalexiou, S. M., & Koutsoyiannis, D. (2016). A global survey on the seasonal variation of the marginal distribution of daily precipitation. *Advances in Water Resources*, *94*, 131–145. https://doi.org/10.1016/j.advwatres.2016.05.005

Papalexiou, S. M., Koutsoyiannis, D., & Makropoulos, C. (2013). How extreme is extreme? An assessment of daily rainfall distribution tails. *Hydrology and Earth System Sciences*, *17*(2), 851–862. https://doi.org/10.5194/hess-17-851-2013

Papalexiou, S. M., Markonis, Y., Lombardo, F., AghaKouchak, A., & Foufoula-Georgiou, E. (2018). Precise temporal Disaggregation Preserving Marginals and Correlations (DiPMaC) for stationary and non-stationary processes. *Water Resources Research*, *54*. https://doi.org/10.1029/2018WR022726

Perica, S., & Foufoula-Georgiou, E. (1996). Model for multiscale disaggregation of spatial rainfall based on coupling meteorological and scaling descriptions. *Journal of Geophysical Research*, *101*(D21), 26,347–26,361. https://doi.org/10.1029/96JD01870

Pielke, R. A., Downton, M. W., & Miller, J. B. (2002). *Flood damage in the United States, 1926–2000: A reanalysis of National Weather Service estimates*. CO: University Corporation for Atmospheric Research Boulder. Retrieved from http://sciencepolicy.colorado.edu/flooddamagedata/flooddamagedata.pdf

Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., et al. (2015). A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Reviews of Geophysics*, *53*(2), 323–361. https://doi.org/10.1002/2014RG000475

Rodriguez-Iturbe, I., Cox, D. R., & Isham, V. (1987). Some models for rainfall based on stochastic point processes. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *410*(1839), 269–288. https://doi.org/10.1098/rspa.1987.0039

Schroeder, M. R. (1991). *Fractals, chaos, power laws: Minutes from an infinite paradise*. New York: W.H. Freeman.

Shaw, S. B., Royem, A. A., & Riha, S. J. (2011). The relationship between extreme hourly precipitation and surface temperature in different hydroclimatic regions of the United States. *Journal of Hydrometeorology*, *12*(2), 319–325. https://doi.org/10.1175/2011JHM1364.1

Stacy, E. W. (1962). A generalization of the gamma distribution. *The Annals of Mathematical Statistics*, *33*(3), 1187–1192. https://doi.org/10.1214/aoms/1177704481

Tadikamalla, P. R. (1980). A look at the Burr and Related distributions. *International Statistical Review/Revue Internationale de Statistique*, *48*(3), 337–344.

Venugopal, V., Foufoula-Georgiou, E., & Sapozhnikov, V. (1999). A space-time downscaling model for rainfall. *Journal of Geophysical Research*, *104*(D16), 19,705–19,721. https://doi.org/10.1029/1999JD900338

Verhoest, N., Troch, P. A., & De Troch, F. P. (1997). On the applicability of Bartlett–Lewis rectangular pulses models in the modeling of design storms at a point. *Journal of Hydrology*, *202*(1–4), 108–120. https://doi.org/10.1016/S0022-1694(97)00060-7

Villarini, G., & Smith, J. A. (2010). Flood peak distributions for the eastern United States. *Water Resources Research*, *46*(6), W06504. https://doi.org/10.1029/2009WR008395

Vrac, M., & Naveau, P. (2007). Stochastic downscaling of precipitation: From dry events to heavy rainfalls. *Water Resources Research*, *43*(7), W07402. https://doi.org/10.1029/2006WR005308

Werner, T., & Upper, C. (2004). Time variation in the tail behavior of Bund future returns. *Journal of Futures Markets*, *24*(4), 387–398. https://doi.org/10.1002/fut.10120

Wheater, H. S., Chandler, R. E., Onof, C. J., Isham, V. S., Bellone, E., Yang, C., et al. (2005). Spatial-temporal rainfall modelling for flood risk estimation. *Stochastic Environmental Research and Risk Assessment*, *19*(6), 403–416. https://doi.org/10.1007/s00477-005-0011-8

Wilby, R. L., & Wigley, T. M. L. (1997). Downscaling general circulation model output: A review of methods and limitations. *Progress in Physical Geography*, *21*(4), 530–548. https://doi.org/10.1177/030913339702100403

Wilby, R. L., Wigley, T. M. L., Conway, D., Jones, P. D., Hewitson, B. C., Main, J., & Wilks, D. S. (1998). Statistical downscaling of general circulation model output: A comparison of methods. *Water Resources Research*, *34*(11), 2995–3008. https://doi.org/10.1029/98WR02577

Yu, L., Zhong, S., Pei, L., Bian, X., & Heilman, W. E. (2016). Contribution of large-scale circulation anomalies to changes in extreme precipitation frequency in the United States. *Environmental Research Letters*, *11*(4), 044003. https://doi.org/10.1088/1748-9326/11/4/044003