A diagnostic framework for understanding climatology of tails of hourly precipitation extremes in the United States

Simon Michael Papalexiou1, Amir AghaKouchak1, and Efi Foufoula-Georgiou1

1Department 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 U.S. 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.

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 non-linear increase in the heaviness of the tail with elevation

Keywords: hourly precipitation extremes, regionalization of distribution tails, heavy tails, predictive uncertainty, climatology of tails

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2018WR022732
1. **Introduction and motivation**

The National Weather Service (NWS) of the United States reports a recent 30-year average flood loss of $7.96 billion/year and 82 fatalities/year (Downton et al., 2005; see also, Pielke et al., 2002), naming flash floods 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 sub-daily timescales, such as hourly. Moreover, the output of rainfall-runoff models typically forced with hourly or sub-hourly precipitation series for decision support and operation, depend on the quality of the precipitation input (e.g., Beven, 2011; Moradkhani & Sorooshian, 2009).

Over the recent years, predictive studies on future trends of hourly precipitation have mainly focused on the Clausius-Clapeyron (C-C) relation and how hourly precipitation scales with air temperature, suggesting in some cases higher increases than the theoretically predicted C-C value of 7%/°C (Berg et al., 2009; Formayer & Fritz, 2017; Lenderink et al., 2011; Lenderink & Meijgaard, 2008; Lenderink & Meijgaard, 2010; Mishra et al., 2012; Shaw et al., 2011). Trends in hourly precipitation extremes have been investigated in only a few studies over the U.S., e.g., Muschinski and Katz (2013) analyzed a limited number of U.S. 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 (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 non-zero 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 modelling that can provide more reliable climate information, e.g., 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., Bárdoossy & Pegram, 2011; Benestad, 2010; Faticchi 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 modelling, however, make feasible to reproduce any desired marginal distribution and correlation structure (including intermittency), and generate thus, realistic precipitation at any timescale including hourly (Papalexopoulos, 2018).

Thus, 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 U.S. exploring the behavior of hourly extremes. Our aim is three-fold to: (a) outline a Bayesian-adjustment approach understanding and diagnosing the climatology of tail including the type of distribution that better describes hourly precipitation extremes, (b) 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) investigate the spatial pattern of tail behavior across the U.S. trying to reveal their “climatology”.

2. Hourly precipitation data
We use the largest database of hourly precipitation records, available from NOAA, comprising 7,127 U.S. 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 year 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 Fig. 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 1684 records which do not show the aforementioned quality issues and the results were almost identical. Therefore, we used the whole dataset to obtain a more robust representation of the spatial pattern of the tails.

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

3.1 Types and definition of tail

The tail function is typically defined as $F_X(x) = 1 - F_X(x)$ which is the complementary cumulative distribution function (CCDF), also known as the survival function (SF) (e.g., Klüppelberg, 1988). The term “tail”, however, is linked with extremes or large values and refers to CCDF’s upper part, i.e., the tail 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, $F(x)$ 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 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 sub- or hyper-exponentials, respectively. Also, many tails can be identified as equivalent, i.e., the tails $F(x)$ and $G(x)$ are equivalent if $\lim_{x \to \infty} F(x)/G(x) = c$, where $c$ is a constant. Here we use and compare two major types of tails based on their vast popularity across many scientific fields, i.e., power-type, or else Pareto tails (e.g., Newman, 2005; Schroeder, 2012) 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

$$F_{PII}(x) = \left(1 + \frac{x}{\beta}\right)^{-\frac{1}{\gamma}}$$

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, i.e., $x^{-1/\gamma}$. Many well-known distributions are tail equivalent to the PII distribution, e.g., 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

$$F_W(x) = \exp\left(-\left(\frac{x}{\beta}\right)^\gamma\right)$$

with scale parameter $\beta > 0$ and tail index $\gamma > 0$. The W tail is sub-exponential (or else stretched-exponential), exponential, and hyper-exponential, for tail index $\gamma > 1$, $\gamma = 1$, and $\gamma < 1$, respectively. Weibull tails are popular as the W distribution is extensively used in statistical and engineering literature as well as in other fields, e.g., life sciences.
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 (pdf), i.e., $f_G(x) = x^{y-1} \exp(-x^y)$ and $f_{GG}(x) = x^{y+1} \exp(-x^{y+2})$ reveals their structural similarity with the Weibull pdf $f_W(x) = x^{y-1} \exp(-x^y)$; note also that the exponential term quickly dominates the polynomial part.

The empirical tail cannot be uniquely defined and different approaches exist, e.g., peak above threshold definitions using a percentage of the largest sample values, or using a fixed number of peaks, e.g., $m$ 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

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

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

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

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 $N$ the sample size of all nonzero values. Note that the PRMSE norm, which is a function of the parameters $\beta$ and $\gamma$ 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, e.g., 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-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 1000 random samples from each distribution with sample size, scale and shape parameters, varying randomly according to the observed empirical distributions, e.g., using the distributions of tail indices given as box plots in Fig. 5 (Section 4.1 describes the details of results shown in Fig. 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 Monte Carlo simulation to be as realistic as possible we studied the bivariate distribution of the estimated scale parameter and tail index, e.g., Fig. 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 with the red colored points marking 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 implies that the \((\beta, \gamma)\) points should be randomly sampled from a bivariate distribution. Here we use the 90\% of “central” points (red-colored in Fig. 2) to construct a smoothed non-parametric bivariate distribution that is used for random sampling. The same process, i.e., of forming the non-parametric 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, ten 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 \((\hat{\beta}, \hat{\gamma})\) point estimates will generate a more “dispersed” distribution than the sampling one.

![Fig. 2. Estimated \((\hat{\beta}, \hat{\gamma})\) parameters for the PII 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 MC simulations.](image)

We present the MC results in Table 1 and in Fig. 3 depicting success rates of each distribution tail. 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, i.e., from \( S_{10\%} \) to \( S_{0.5\%} \) the precision decreases for both tails.

**Table 1.** Success and failure probabilities of the true tail to be identified correctly and erroneously, respectively, based on Monte Carlo simulations. True tails are in the first column.

<table>
<thead>
<tr>
<th>( p% = )</th>
<th>10</th>
<th>5</th>
<th>2</th>
<th>1</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>PII</td>
<td>88.9</td>
<td>80.6</td>
<td>75.5</td>
<td>70.1</td>
<td>66.5</td>
</tr>
<tr>
<td>W</td>
<td>12.4</td>
<td>25.4</td>
<td>34.4</td>
<td>45.5</td>
<td>52.9</td>
</tr>
<tr>
<td>PII</td>
<td>11.1</td>
<td>19.4</td>
<td>24.5</td>
<td>29.9</td>
<td>33.5</td>
</tr>
<tr>
<td>W</td>
<td>87.6</td>
<td>74.6</td>
<td>65.6</td>
<td>54.5</td>
<td>47.1</td>
</tr>
</tbody>
</table>
Fig. 3. Success and failure probability for a tail to be identified correctly.

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

$$\Pr(A|B) = \frac{\Pr(A) \Pr(B|A)}{\Pr(B)} \quad (5)$$

The results presented in Fig. 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 $\hat{p}_{\text{PII}} = \Pr(\hat{\text{PII}})$ and $\hat{p}_W = \Pr(\hat{W})$, e.g., for the $S_{10\%}$ sample $\hat{p}_{\text{PII}} = 0.326$ and $\hat{p}_W = 0.263$. Let us denote the true probabilities that we are trying to estimate with $p_{\text{PII}} = \Pr(\text{PII})$ and $p_W = \Pr(W)$. Also, we denote the results of the MC simulation given in Table 1 with the standard conditional notation, i.e.,

$$p_{11} = \Pr(\hat{\text{PII}}|\text{PII}) \quad p_{12} = \Pr(\hat{W}|\text{PII})$$
$$p_{21} = \Pr(\hat{\text{PII}}|W) \quad p_{22} = \Pr(\hat{W}|W) \quad (6)$$

so that, e.g., $p_{11} = \Pr(\hat{\text{PII}}|\text{PII})$ 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

$$\Pr(\text{PII}|\hat{\text{PII}}) + \Pr(W|\hat{\text{PII}}) = 1$$
$$\Pr(\text{PII}|\hat{W}) + \Pr(W|\hat{W}) = 1 \quad (7)$$

as for any given outcome (either $\hat{\text{PII}}$ or $\hat{W}$) 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 Eqs. (7) using Bayes formula and the definitions given in Eqs. (6) we get

$$\frac{p_{\text{PII}}}{\hat{p}_{\text{PII}}} \frac{p_{11}}{\hat{p}_{\text{PII}}} + \frac{p_W}{\hat{p}_W} \frac{p_{12}}{\hat{p}_W} = 1$$
$$\frac{p_{\text{PII}}}{\hat{p}_{\text{PII}}} \frac{p_{21}}{\hat{p}_{\text{PII}}} + \frac{p_W}{\hat{p}_W} \frac{p_{22}}{\hat{p}_W} = 1 \quad (8)$$

Solving this linear system, we obtain the equations
which are used to adjust the initial estimates given in Fig. 4A.

4. Results

4.1 Fitting results

We compare the fitted tails in terms of the resulting PRMSE, i.e., 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 processes for the five tail levels we study; see Fig. 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 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 behaviour, 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 rates of a method may significantly alter the results. In Monte Carlo simulations we know a priori the true tail, i.e., 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 Monte Carlo simulations and we adjust the empirical results based on the Bayesian framework described. We find that the adjusted results, given in Fig. 4B, provide a different picture than the results in Fig. 4A. Particularly, the adjusted results show that as we consider a lower tail level the percentage of tail samples that the PII tail performed better than the W tail actually decreases, resulting in almost complete domination of the W tail at the 0.5% level. Note, that estimating 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, e.g., 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 Fig. 5C) supporting both PII and W tails as good tail choices.
Fig. 4. (A) Percentage of better-fitted tails to hourly precipitation extremes for various empirical tail samples; (B) Bayesian adjusted values.

These results are enhanced by studying the estimated tail indices presented in the box plots of Fig. 5. As we move from $S_{10\%}$ to $S_{0.5\%}$ the PII shape parameters constantly decrease, e.g., the median values decrease from 0.24 to 0.19 (see Fig. 5A). This implies that the Pareto tail is not a robust choice for hourly precipitation (at least for this dataset), 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 Fig. 5B), and a very narrow interquartile range. The upper limit of the 95% empirical confidence intervals, expressed by the outer box plots fences, is less than 0.9, implying that almost all samples analysed have a stretched-exponential tail, i.e., heavier than the exponential.
Fig. 5. Box plots of the estimated PII and W tail indices as well as the resulting fitting error. The whiskers define the 95% empirical confidence interval.

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 (Fig. 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 Fig. 5A). Both maps (created by applying, first, ordinary kriging with spherical semivariogram and search radius fine points, and second, Gaussian smoothing taking into account the eight neighbouring pixels) show the same spatial pattern revealing strong spatial coherence but also large variability. Three major regions emerge: the first, comprising “heaviest” tails ($0.26 \leq \gamma_{\text{PII}} < 0.33$ and $0.38 \leq \gamma_{W} < 0.50$), is spotted at the mountainous regions of U.S., i.e., 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 ($0.10 \leq \gamma_{\text{PII}} < 0.18$ and $0.60 \leq \gamma_{W} < 0.74$), is located in the Southern U.S., 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 spotted in most of the Pacific Coast and all of California. The third region, with relatively “moderate” tails ($0.18 \leq \gamma_{\text{PII}} < 0.26$ and $0.50 \leq \gamma_{W} < 0.60$), 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 tail's heaviness seems to reveal a relation with elevation. Heavy tails in high elevations, however, does not indicate that elevation is a determinant of the tails of the distribution as many other factors (e.g., synoptic conditions, vertical stability, moisture convergence) contribute to the intensity of rainfall. Relatively “thin” tail is observed in southeastern United States where rainfall is mainly convective (short-term intense rain over a relatively smaller spatial extent compared to a typical stratiform event). However, different regions receive various types of storms throughout the year. For example, 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 type of storm type or a regional physical process. The tail parameter represents the overall behaviour of extremes considering samples from different storms over a long period of time. A discussion on extreme rainfall and flood peaks mechanisms over the U.S. can be found, respectively, in Barebero et al. (2018) and Villarini & Smith (2010).

Fig. 6. Spatial variation of Pareto II (PII) and Weibull (W) tail indices.

It is noteworthy that a comparison between the spatial variation of the tail indices (Fig. 6) with 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 U.S. 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, e.g., regions classified in the Snow class over the U.S. 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 $\rho$ and Kendall’s $\tau$ correlation coefficients between elevation $H$ and $\gamma_{\text{PII}}$ or $\gamma_W$ (calculated using all 4703 $(H, \gamma_{\text{PII}})$ and $(H, \gamma_W)$ points) show low correlation, i.e., $\rho(H, \gamma_{\text{PII}}) = 0.30$, $\tau(H, \gamma_{\text{PII}}) = 0.22$, and $\rho(H, \gamma_W) = -0.27$, $\tau(H, \gamma_W) = -0.21$. 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 $(H, \gamma_{\text{PII}})$ and $(H, \gamma_W)$ 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 $\bar{H}$ and mean $\bar{\gamma}_{\text{PII}}$ or $\bar{\gamma}_W$. For example, for $n = 100$, forty blocks are formed resulting in 40 $(\bar{H}, \bar{\gamma}_{\text{PII}})$ or $(\bar{H}, \bar{\gamma}_W)$ points (see Fig. 7). This averaging process increases vastly the correlation coefficient values, as the block size increases (see Fig. 8), from low to very high values above 0.9.

Fig. 7. PII and W tail indices vs. elevation. The grey shaped areas result from ensemble curves; the coloured curves within the grey areas are the “median” curves given in Eqs. (10); the cloud of points is for block size $n = 100$. 

© 2018 American Geophysical Union. All rights reserved.
This increase in correlation (especially in Pearson’s $\rho$) does not necessarily imply a linear relationship between tail indices and elevation. Actually, the cloud of points (see Fig. 7A,B depicting the points for block size $n = 100$) indicates concave increase and convex decrease, respectively, for $\gamma_{\text{PII}}$ and $\gamma_{\text{W}}$ over elevation. We deem a function $g(x)$ interpolating these points should be real at $\gamma = 0$ and tend to a constant as $\gamma \to \infty$ (very large values) given that an infinite increase of the tail index with elevation is physically inconsistent. One function with these properties is $g(\gamma) = a + b \exp(-c\gamma)$, 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 $\lim_{\gamma \to \infty} g(\gamma) = a$.

We estimate the parameters $a, b, c$ by least square fitting, yet for different block sizes different clouds of ($\hat{\gamma}_{\text{PII}}$, $\hat{\gamma}_{\text{W}}$) and ($\hat{\gamma}_{\text{W}}$, $\hat{\gamma}_{\text{PII}}$) 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 $n = \{10, 20, ..., 400\}$, i.e., a total of 40 different clouds. We show the ensemble of fitted curves in Fig. 7A,B as grey shaded areas (remarkably narrow) revealing that block size does not affect significantly the estimated parameters. We provide characteristic curves for the tail indices $\gamma_{\text{PII}}$ and $\gamma_{\text{W}}$ vs. elevation by calculating the median values of the 40 estimated $a, b, c$, parameters, resulting in:

\[
\begin{align*}
\gamma_{\text{PII}}(\gamma) &= 0.30 + 0.14 \exp(-0.0009\gamma) \\
\gamma_{\text{W}}(\gamma) &= 0.46 - 0.16 \exp(-0.0011\gamma)
\end{align*}
\]

where elevation $\gamma$ is in meters. The curves are shown in Fig. 7A,B and for $\gamma = 0$ m the tail-index values are $\gamma_{\text{PII}} = 0.16$ and $\gamma_{\text{W}} = 0.62$ while their maxima for $\gamma \to \infty$ are $\gamma_{\text{PII}} = 0.30$ and $\gamma_{\text{W}} = 0.46$.

![Fig. 8. Pearson’s and Kendall’s correlation coefficients of PII and W tail indices vs. elevation for various block sizes.](image)

As an example, we use a 50-year hourly record from Lake Charles regional Airport in Louisiana to fit the Gamma, the BrXII (three-parameter generalization of the Pareto II) and the GG (three-parameter generalization of Weibull or Gamma) distributions with fixed PII and W tail indices for the latter two, respectively, selected from the maps of Fig. 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 mm 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 Fig. 9A) predict for $T = 100$ years, respectively, 75% and 25% larger precipitation than the fitted G distribution, while for $T = 1000$ 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 to predicted amounts by a light-tailed Gamma distribution (Fig. 9B). Particularly, we consider a characteristic hourly precipitation sample with coefficient of variation $C_V = \sigma/\mu \approx 1.8$ and probability of dry hours $p_0 = 0.90$. Assuming an arbitrary mean value $\mu$ (results do not change with different $\mu$ as long as $C_V$ remains the same), we fit the Gamma, the Generalized Gamma, and Burr type XII distributions by preserving the mean and standard deviation (obviously $\sigma = \mu \times C_V$) and using as tail index for the Generalized Gamma and Burr type XII the corresponding median values i.e., 0.56 and 0.19, respectively. Specifically, we show (Fig. 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.

![Image](image.png)

**Fig. 9.** (A) Fitted Gamma (G), Burr type XII (BrXII) and Generalized Gamma (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 Fig. 6. (B) Prediction difference of the GG and the BrXII distributions compared to a G distribution for a typical hourly precipitation record in USA.

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 here 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, which however are estimated without explicitly taking into account the behavior of extremes, and certainly, the information summarized in statistics like the mean and variance is not sufficient to characterize either the general shape or the tail of a distribution. 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).

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 behaviour 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 Monte Carlo experiment, i.e., we generate samples from a W distribution with tail index $\gamma = 0.56$ and we calculate the probability of a posteriori estimating $\gamma$ 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 (Fig. 10), e.g., 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, e.g., use the method of moments in order 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).

![Fig. 10. Distribution of the tail index estimated from random samples from a W 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 $\gamma$ with error less than 5% based on single “station” estimates or on regional values by averaging estimates in 2, 5 and 10 “stations”.

© 2018 American Geophysical Union. All rights reserved.
5. **Summary and Conclusions**

We analyzed the largest available database of hourly precipitation records in the U.S. 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 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, i.e., for $S_{1\%}$ and $S_{0.5\%}$ (see Fig. 4A and Fig. 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 U.S. The Weibull tail index $\gamma_\text{W}$ is remarkably stable across all tail samples studied with a median value equal to 0.56 (Fig. 5B). The Pareto II tail index $\gamma_\text{PII}$ is affected by the empirical tail sample, with a median value of 0.24 at $S_{10\%}$ constantly decreasing to 0.19 at $S_{0.5\%}$ (Fig. 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 (sub-exponential) 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 Fig. 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 Eqs. (10) and Fig. 7). However, we stress that elevation cannot be the sole determinant of tail’s heaviness as regional precipitation characteristics and physical mechanisms affect also extremes' behaviour. These functions as well as the spatial variation maps in Fig. 6 can be used as guidelines, or as a first choice, in selecting tail index values for probabilistic and stochastic modelling 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.

**Acknowledgement Samples, and Data:** 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**


© 2018 American Geophysical Union. All rights reserved.


