FROM DIVERSITY TO VOLATILITY: PROBABILITY OF DAILY PRECIPITATION EXTREMES

A. Panorska[1], A. Gershunov[2] and T. Kozubowski[1]

[1]Mathematics Department
University of Nevada, Reno

[2]Scripps Institution of Oceanography
University of California, San Diego

Motivations

- A rigorous view of extreme precipitation events in climate requires a parametric approach.
- Exponential tails may not be adequate to model probabilities of high-frequency precipitation extremes in volatile precipitation regions. We want to check how adequate or inadequate they are.
- Indications are that heavy tailed distributions may be more appropriate for this task.
- However, identifying and fitting heavy tails is difficult
  - » Typical methods involve graphical analysis
  - » There are many heavy-tailed distributions
  - » Identification and parameter estimation needs to be automated

Are precipitation extremes exponentially distributed?

If not, what is a reasonable distribution?

We want the data to show us its tails…

Heavy-tailed modeling of extreme event probabilities

- »Heavy-tailed PDFs (power laws) allow for more extremes than traditional PDFs
- »Arise naturally as limit sums of random variables
- »Random variable X (e.g. precipitation) is "heavy tailed" if the probability (P) that it exceeds a value x is of power order $x^{-\alpha}$ for large x

$$P(X > x) \approx cx^{-\alpha}, \text{ as } x \to \infty, \text{ where } c, \alpha > 0$$

Approach

Peaks over threshold (POT) methodology

Three possible limiting PDFs for exceedances (approximations)

Balkema - de Haan - Pickands theorem (Balkema and de Haan 1974 and Pickands 1975) provides the limiting distribution of exceedances. The theorem says that when the threshold (u) increases, the distribution of the exceedance $X^{[u]}$ converges to a Generalized Pareto (GP) distribution. Any GP distribution has to be one of the following three kinds: exponential, Pareto or beta (finite, not applicable). So: no matter what the original distribution of X is, the exceedance $X^{[u]}$ over any threshold u is (approximately) one of only three distributions (effectively two).

Moreover, if original distribution has exponential tails, then the exceedance pdf will be exponential. If it has heavy tails, then the exceedance pdf will be Pareto.

Using this result, we seek a statistic, a decision rule, to classify observed daily precipitation tails into exponential and heavy

Exponential vs. Heavy Tails

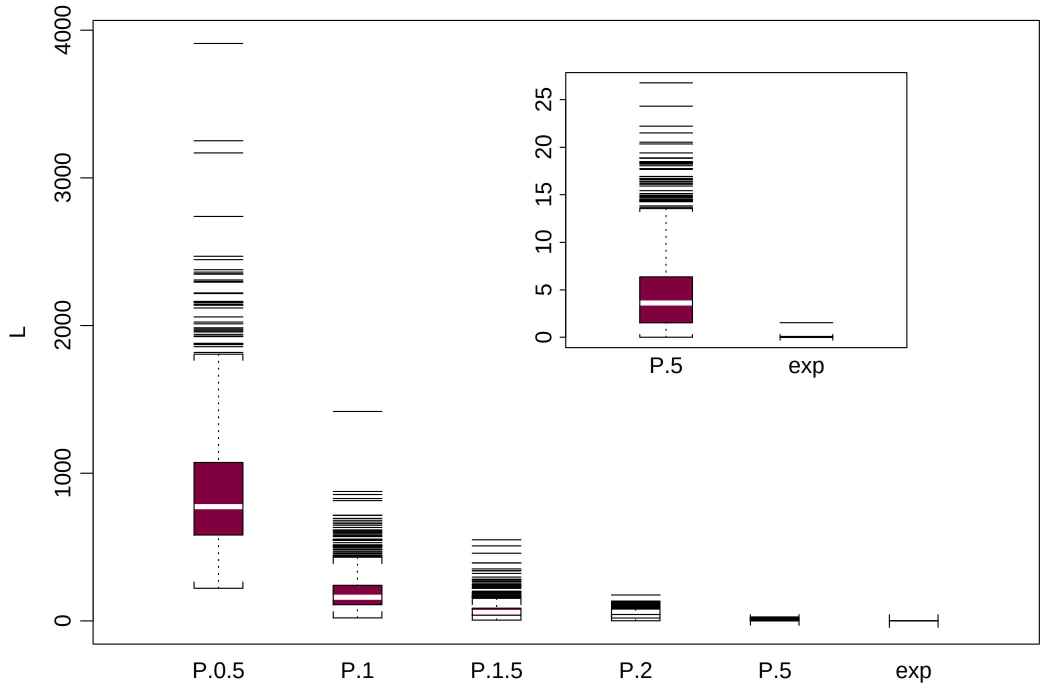
We approached this problem using ideas from the theory of likelihood ratio tests (Lehmann, 1997). The approach is to consider the ratio of the maxima of the likelihoods of the observed sample under the null (Pareto in the numerator) and alternative (exponential in the denominator) models. The logarithm of the likelihood ratio statistic is:

$$L = \log\left(\frac{\sup\limits_{\alpha>0, s>0} L_{Pareto}(\vec{x} \mid \alpha, s)}{\sup\limits_{\sigma>0} L_{\exp}(\vec{x} \mid \sigma)} \right)$$

Where $\vec{X}$ is the observed sample (of excesses); $L_{Pareto}(\vec{x} \mid \alpha, s)$ and $L_{\exp}(\vec{x} \mid \sigma)$ are the likelihood functions of the sample under Pareto and exponential hypotheses, respectively. We use a Pareto distribution with the survival function $S(x) = P(X>x) = (1/(1+1/s\alpha))^{\alpha}$ and exponential distribution with the survival function $S(x) = P(X>x) = \exp(-x/\sigma)$. In the Pareto case, the α parameter determines the thickness of its tail and is of primary importance. The scale parameter s is of secondary importance. In the exponential case σ is the scale parameter.

Statistical Properties of L

Reasonable critical value: L = 1



Boxplots of simulated distributions of L. The first five boxplots were done using 10,000 observations of L from Pareto samples of size 1,000 with α varying from 0.5 (first boxplot) to 5 (second to the last boxplot). The last boxplot corresponds to 10,000 observations of L from exponential samples of size 1,000. The inset blows up the last two boxplots.

Statistical Properties of L

Misclassification Rates of Pareto as Exponential

sample size	α					
	0.5	*1*	*1.5*	*2*	*5*	*10*
200	0%	0%	0%	0%	3%	13%
500	0%	0%	0%	0%	0%	4%
1,000	0%	0%	0%	0%	0%	0%
5,000	0%	0%	0%	0%	0%	0%
10,000	0%	0%	0%	0%	0%	0%

Percentages of Pareto samples, with α's varying in columns and sample sizes varying in rows, which were classified as exponential, that is with $0 \leq L \leq 1$.

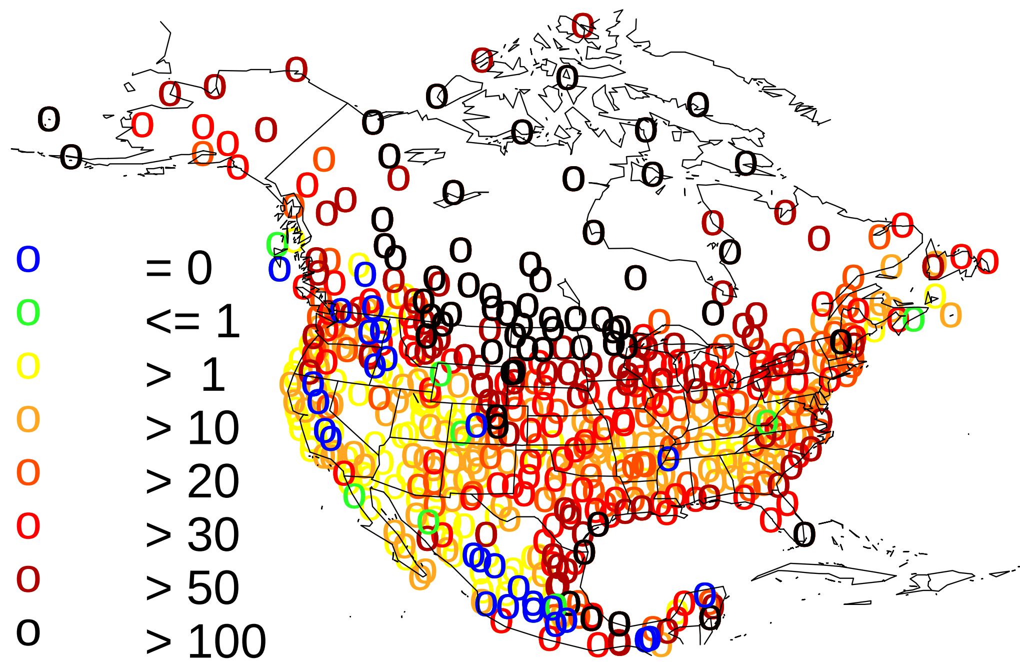
Statistical Properties of L

Misclassification Rates of Exponential as Pareto

Sample size	200	500	1,000	5,000	10,000
Misclassification rate	4%	6%	6%	7%	8%

Percentage of exponential samples (with varying sample sizes) which were classified as Pareto ($L > 1$).

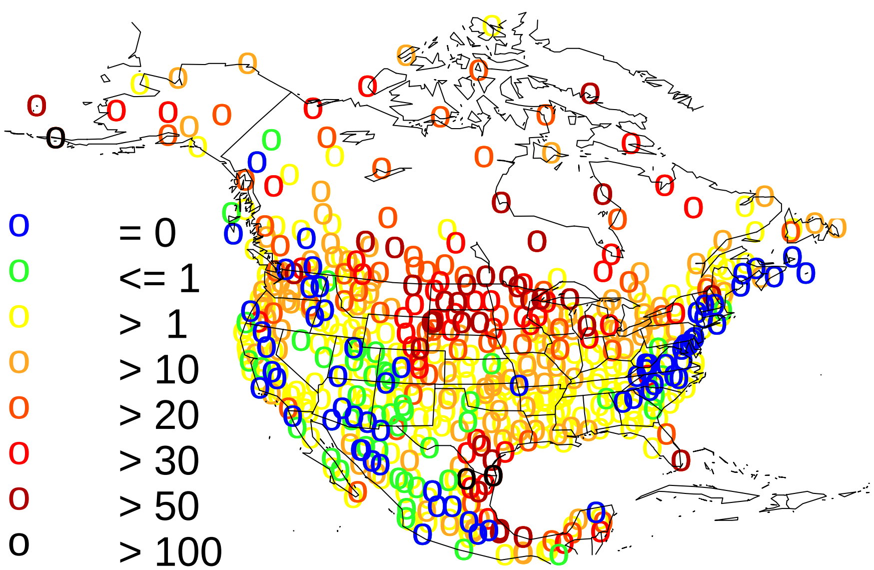
Log Likelihood Ratio (L): all data



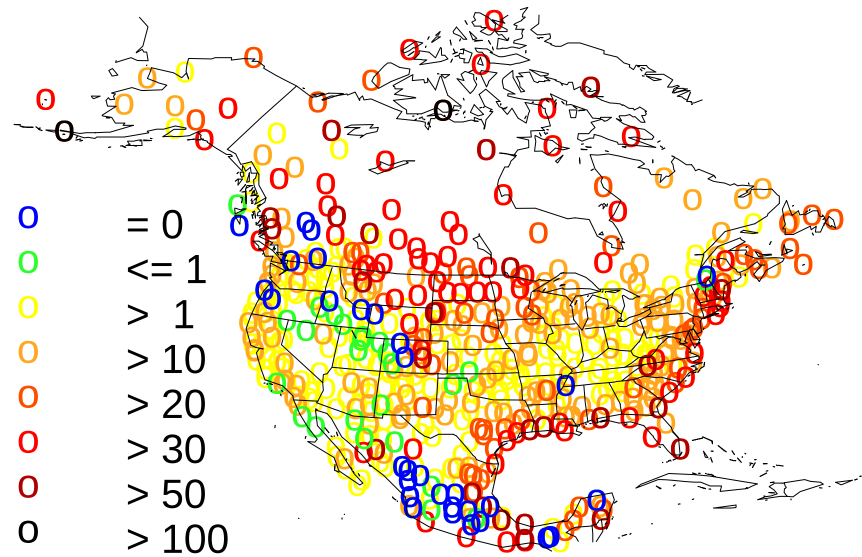
Log likelihood ratio (L) computed for daily **excesses over local median** at each of the 560 stations selected from HCN data for the period 1950 – 2001. L <= 1 (blue and green circles) represent approximately exponential tails, while yellow, red and black circles represent heavier tails.

Log Likelihood Ratio (L): by season

a. Cold season data (November – April)



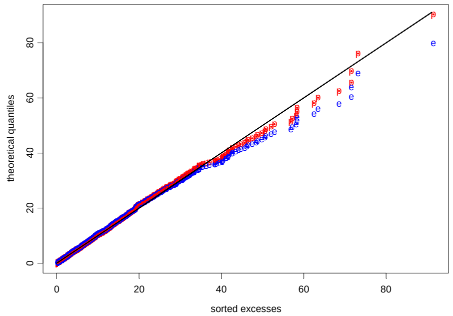
b. Warm season data (May – October)



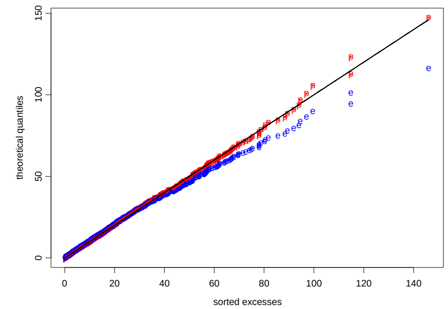
Log likelihood ratio (L). Plates (a) and (b) show L for precipitation in the winter half-year, November – April, and the summer half-year, respectively.

A few local examples

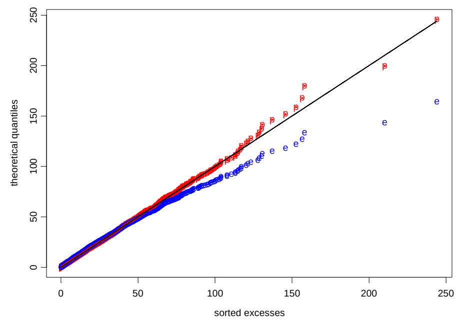
a. Sacramento, L = 1.38



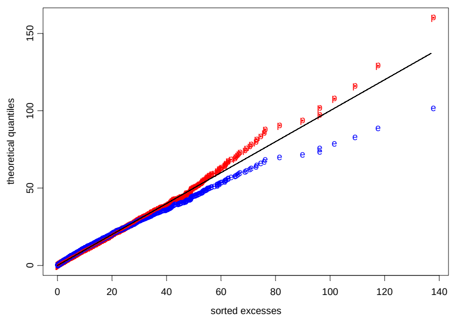
b. Nashville, L = 6.32



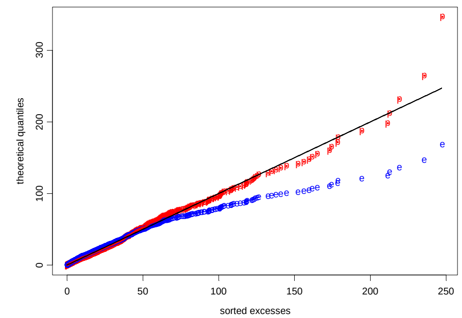
New Orleans, L = 17.4



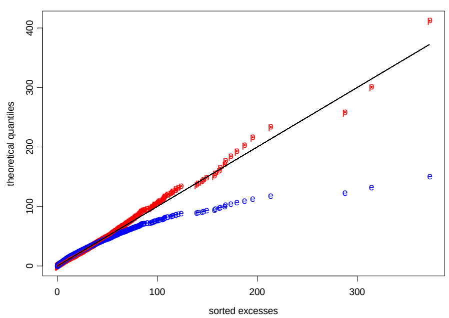
c. St. Louis, L = 20.9



d. Houston Hobby Airport, L = 51.3

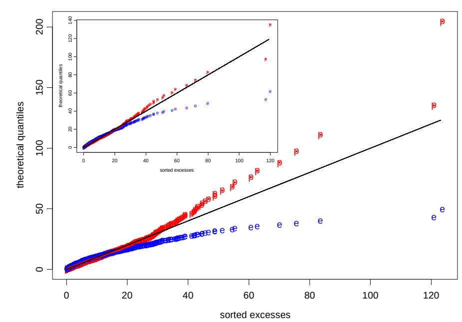


e. Miami Airport, L = 118.8



Probability plots for excesses over threshold (local median, see Table 1) at selected stations. Sorted observed excesses displayed in mm along the x-axis are plotted against the corresponding theoretical quantiles derived from the fitted exponential (blue e's) and Pareto (red p's) models.

f. Williston ND, L = 149.1 [$30.2_{\text{80th \%-ile}}$]



PRECIPITATION STATISTICS AT SELECTED STATIONS

Station	Log likelihood ratio (L)	$P[p > 0]$ (%)	$\text{median}(p_{p>0})$ (mm)	$\text{Max}_{obs}(p)$ (mm)	100-yr event Exp and Pareto (mm)	Pareto $P[p > p_{exp}^{100}]$ (%)
Sacramento	1.38	16	4.3	95	79 and 90	2.3
Nashville	6.32	26	7.1	153	115 and 145	4.7
St. Louis	20.9	30	4.1	142	101 and 158	9.4
Houston	51.3	27	5.1	253	167 and 343	15.7
Miami	118.8	36	4.6	377	149 and 412	28.0
Williston	149.1	26	1.5	125	49 and 200	35.1
New Orleans	17.4	31	6.1	250	165 and 249	7.0

Precipitation statistics at selected stations for the common observational period 1950 – 2001: L, probability of precipitation (i.e. % of days with recorded precipitation), median daily total on days with precipitation, maximum recorded daily total, the estimated 100-year event assuming exponential and Pareto tails, and the Pareto probability of exceeding the exponential 100-year event.

Conclusions

- Exponential tails are inadequate to model daily extremes in most regions of North America, especially in volatile precipitation regions, precisely where extreme events play an important climatic role
- Heavy tailed models are appropriate on theoretical and empirical grounds
- These results can be directly extended to many climatic studies and applications

Climate Change

The choice between heavy tailed and exponentially tailed models is of a qualitative nature. The heavy tailed distributions have much larger high percentiles relative to the rest of the data values than the exponentially tailed ones. That implies that in places where heavy tailed models are appropriate, the future large events may be much larger than those observed up to date and that we need to be prepared for such possibilities. The exponentially tailed models of precipitation will not be able to predict very large (relative to the observed data) events, because their mathematical properties do not allow such extremes.