J3.5

Do Weather or Climate Variables and Their Impacts Have Heavy-Tailed Distributions?

Richard W. Katz*
Environmental and Societal Impacts Group
National Center for Atmospheric Research, Boulder, Colorado

1. INTRODUCTION AND SUMMARY

The question of whether the distributions of weather or climate variables and their impacts have "heavy" (i.e., Pareto-like) tails is important for several reasons. For one thing, its neglect can result in a substantial underestimation of the likelihood of extreme events (e.g., in terms of "return periods" or "design values" used in engineering design). For another, the question of whether extreme weather or climate events and their impacts are increasing has been addressed through trend analyses using conventional statistics (in effect, assuming medium-tailed distributions) (e.g., Karl and Knight, 1998). If these distributions were heavy-tailed, then such trend analyses might well be suspect. In particular, the standard errors associated with the estimated slopes of trend lines would be underestimated.

In the present paper, first the methodology necessary to detect and model heavy-tailed distributions is described. This methodology is based on the statistics of extremes (e.g., Reiss and Thomas, 1997). To detect a heavy tail, either the generalized Pareto distribution is fitted to the extreme upper tail of the data or the generalized extreme value distribution is fitted to maximum values extracted from the data (e.g., annual maxima). One convenient way to measure the effects of heavy tails is in terms of differences in estimated design values or return periods.

Next, one example is considered in detail to demonstrate the application of this methodology. This example involves a 100-yr time series of daily precipitation amounts at Fort Collins, CO. With such a long record, the evidence for a heavy-tailed distribution turns out to be relatively strong. Annual cycles in the precipitation parameters are explicitly modeled, reducing the possibility that this heavy tail is indirectly induced in annual maxima.

Last, the evidence of heavy tails in weather or climate variables and their impacts, in general, is reviewed. Among weather or climate variables, precipitation amount has the strongest evidence for a heavy tail. This evidence is not overwhelming when individual sites are analyzed separately, but becomes stronger when either relatively long records are available or when regional analyses are performed (e.g., assuming common shape parameter within the region). For impact variables related to the weather or climate, the evidence of heavy tails is quite a bit stronger (e.g., economic damage from extreme events such as hurricanes). The question is raised about the extent to which any heavy tails in impact variables are attributable to the underlying weather or climate variables, as opposed to an inherent tendency of variables related to income or wealth.

2. METHODOLOGY

2.1 Generalized Pareto Distribution

The concept of a heavy-tailed distribution is formally defined in terms of the Generalized Pareto (GP) distribution. A fundamental theoretical result from the statistics of extremes is that the upper tail of essentially any distribution must be approximately of the GP form (e.g., Smith, 2001).

The GP distribution function is given by:

$$F(x;\, \sigma,\, \gamma) = 1 - [1 + \gamma\, (x/\sigma)]^{-1/\gamma}, \qquad (1)$$

$\sigma > 0$, $1 + \gamma\, (x/\sigma) > 0$ (Reiss and Thomas, 1997, Chapter 1). Here σ is a scale parameter and γ is a shape parameter. If $\gamma > 0$, then the distribution is said to be *heavy tailed*; if $\gamma < 0$, then the upper tail of the distribution is bounded. Taking the limit as $\gamma \to 0$, (1) reduces to the more familiar exponential distribution (which is medium tailed). If $\gamma > 0$, then the moments of the GP distribution are infinite for orders greater than $1/\gamma$ (e.g., the variance is infinite if $\gamma > ½$; the mean is infinite if $\gamma > 1$).

In practice, the two-parameter version of the GP distribution (1) is fitted to the "excess" of the variable over a relatively high threshold. One way to estimate the two parameters, σ and γ, is by the method of maximum likelihood, in this case requiring an iterative numerical algorithm for nonlinear optimization (Smith,

* Corresponding author address: Richard W. Katz, NCAR/ESIG, Box 3000, Boulder, CO 80307; e-mail <rwk@ucar.edu>

2001). The evidence for a heavy tail can be evaluated by the likelihood ratio test for the shape parameter γ of the GP distribution being zero (i.e., test of GP vs. exponential distributions) (Reiss and Thomas, 1997, Chapter 5).

2.2 Generalized Extreme Value Distribution

An alternative approach for detecting a heavy-tailed distribution involves modeling the maximum of a sequence, sometimes termed "block maxima." Another fundamental theoretical result from the statistics of extremes is that any limiting distribution of the maximum must be in the form of the Generalized Extreme Value (GEV) distribution (e.g., Smith, 2001).

The GEV distribution function is given by:

$$F(x;\, \mu, \sigma, \gamma) = \exp\{-[1 + \gamma\,(x - \mu)/\sigma]^{-1/\gamma}\}, \qquad (2)$$

$\sigma > 0$, $1 + \gamma\,(x - \mu)/\sigma > 0$ (Reiss and Thomas, 1997, Chapter 1). Here μ is a location parameter, σ is a scale parameter, and γ is a shape parameter. The shape parameter of the GEV distribution has the same interpretation as that for the GP distribution (i.e., if $\gamma > 0$, then the distribution of the data whose maximum is being taken must be heavy tailed). Taking the limit as $\gamma \to 0$, (2) reduces to the Gumbel Type of extreme value distribution [which corresponds to the data having a medium (or exponential-like) tail].

The three parameters of the GEV distribution, μ, σ, and γ, can be estimated by the method of maximum likelihood, making use of the same type of numerical techniques as required for the GP distribution (Smith, 2001). The evidence for a heavy tail can be evaluated by the likelihood ratio test for the shape parameter γ of the GEV distribution being zero (i.e., test of GEV vs. Gumbel distributions) (Reiss and Thomas, 1997, Chapter 4).

2.3 Poisson–Generalized Pareto Model

The Poisson–GP model, sometimes called the "peaks over threshold" (POT) approach in hydrology, can be viewed as an indirect way of fitting the GEV distribution to maxima (Reiss and Thomas, 1997, Chapter 5). It has the advantage over the block maxima approach that more information about the extreme upper tail of the data can be utilized.

The Poisson–GP model involves the following two components:

(i) Exceedances of a high threshold u are generated by a Poisson process with intensity parameter λ;

(ii) Excesses over the threshold u have a GP distribution with parameters $\sigma^*$and γ^*.

The correspondence between the parameters of the Poisson–GP model (i.e., λ, σ^*, γ^*) and of the GEV distribution (i.e., μ, σ, γ) is given by the following relations:

$$\ln\lambda = -(1/\gamma)\ \ln[1 + \gamma\,(u - \mu)/\sigma], \qquad (3)$$
$$\sigma^* = \sigma + \gamma\,(u - \mu), \quad \gamma^* = \gamma$$

(Smith, 2001). In particular, the two shape parameters are identical, so both approaches involve equivalent measures of heavy-tailed distributions. Taking the limit as $\gamma \to 0$, this model reduces to the Poisson–exponential (i.e., corresponding to the Gumbel distribution for the maximum).

The technique of maximum likelihood can be employed to estimate all the parameters of the Poisson–GP model simultaneously, as is necessary when fitting GEV distributions in which the parameters depend on covariates (e.g., annual cycles). One additional task is to select the threshold u, high enough that the GP approximation is valid but not so high that the number of exceedances is too small (Reiss and Thomas, 1997, Chapter 5).

3. APPLICATION

3.1 Data

A relatively long time series of daily precipitation amount (i.e., 100 yr from 1900–1999) at a single location (Fort Collins, CO) is analyzed. This data set has been of recent interest, because of a flood that occurred on 28 July 1997 (Petersen et al., 1999). It is available at the Colorado Climate Center, Colorado State Univ. (*http://ulysses.atmos.colostate.edu*). Fig. 1 shows a plot of the time series of annual maxima. Despite the recent flood, there is no evidence of a trend.

3.2 Model Fitting

To systematically study the distribution of high precipitation amounts, several models are fitted by the method of maximum likelihood. The GEV distribution is fitted both directly to the annual maxima and indirectly through the Poisson–GP model. Annual cycles in the parameters of the GEV are modeled in the Poisson–GP approach (i.e., sine waves for μ and $\ln\sigma$). For each model, the corresponding constrained model with the shape parameter $\gamma = 0$ is fitted as well for comparative purposes.

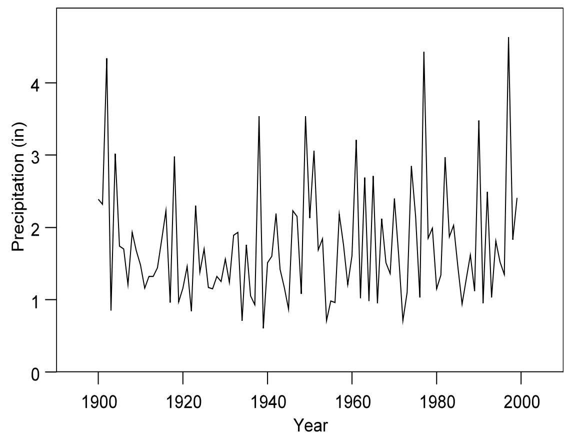

FIG. 1. Time series of annual maxima of daily precipitation amount (in.) at Fort Collins, CO, 1900–1999.

A threshold of u = 0.395 in. is used in the Poisson–GP approach, with quite similar results being obtained for higher threshold values. Given the marked seasonality of precipitation in the region, it is not surprising that the Poisson–GP model with annual cycles fits much better than does the corresponding model without any cycles.

3.3 *Evidence for Heavy Tail*

Table 1 lists the maximum likelihood estimates of the shape parameter γ and *P*-values for the likelihood ratio test of $\gamma = 0$ for the Fort Collins data. The values of the estimated shape parameter are quite similar in all cases, near 0.2, or indicative of a heavy tail. In the block maxima approach, the evidence in support of a heavy tail is somewhat inconclusive. Utilizing more information about the upper tail, the Poisson–GP model indicates overwhelming evidence for a heavy tail, whether or not the annual cycle is explicitly modeled.

TABLE 1. Shape parameter estimates and tests for heavy tails in daily precipitation amounts at Fort Collins, CO.

Model	$\hat{\gamma}$	*P*-value
GEV (no cycles)	0.1736	0.0375
Poisson–GP (no cycles)	0.2119	$< 10^{-11}$
Poisson–GP (cycles)	0.1818	$< 10^{-9}$

Fig. 2 shows how much of a difference allowing for a heavy tail makes in terms of design values for the annual maxima corresponding to a return period of 100 yr for the GEV and Gumbel distributions (derived from the Poisson–GP and Poisson–exponential models with annual cycles). These design values correspond to the annual maximum daily precipitation amount whose probability of occurrence is 0.01. They are termed "effective" because they vary depending on the time of year. If the heavy tail is ignored, a systematic underestimation occurs (ranging from about 1 in. in the winter to 3 in. in the summer).

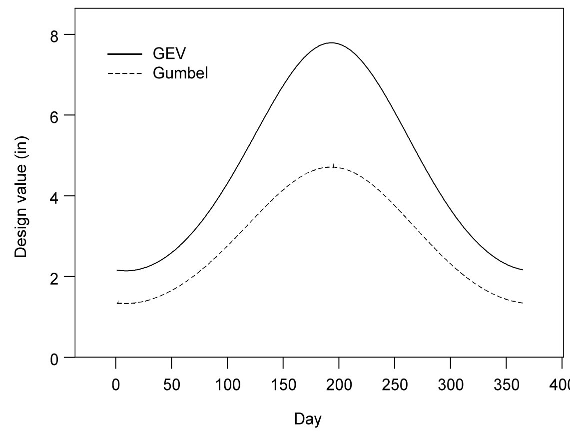

FIG. 2. Effective design value (in.) for 100-yr return period of annual maxima of daily precipitation amount at Fort Collins as function of day within year.

4. DISCUSSION

In general, how strong is the evidence that weather or climate variables and their impacts possess heavy-tailed distributions?

4.1 *Weather or Climate Variables*

Unfortunately, the evidence is somewhat meager because of the lack of systematic study of the extreme tails of weather or climate variables. Still it appears that when one actually looks for heavy tails, they are readily found for certain variables, especially for precipitation amount (consistent with example in Sec. 3).

Precipitation. In the hydrologic literature, streamflow is routinely found to have a heavy-tailed distribution (e.g., Anderson and Meerschaert, 1998). Unless these results are solely attributable to the integrative effects of stream-flow (i.e., being related in a complex, nonlinear manner to precipitation over an entire water basin), they suggest that precipitation amount ought to have a heavy-tailed distribution as well. But this can be difficult to conclusively determine from only a single site, unless the record is relatively long. When precipitation data for several locations are modeled simultaneously in what hydrologists refer to as "regional analysis," the power of "borrowing strength" indicates that precipitation amount does have a heavy-tailed distribution (e.g., Buishand, 1989). Applying the statistics of extremes to the same daily precipitation data set (consisting of a large number of stations with relatively long records) analyzed by Karl and Knight (1998), Smith (2001) concluded that precipitation amount does possess a heavy-tailed distribution (Smith, 2001).

Temperature. For the most part, the evidence points to the temperature having a moderate or bounded upper (and lower) tail. For example, in detailed regional analyses of a large number of sites in the U.S. Midwest and Southeast, Brown and Katz (1995) found that the extreme high (and low) daily temperature has a bounded tail (i.e., $\gamma < 0$).

Wind. The situation for extreme high wind speeds is quite analogous to that described for temperature (Palutikof et al., 1999). Much evidence supports a bounded tail (i.e., $\gamma < 0$), although the use of the Gumbel type is still quite prevalent in engineering design in practice.

4.2 Weather or Climate Impacts

Some evidence exists that the economic impact associated with a given extreme weather or climate event has a heavy-tailed distribution. For example, Dorland et al. (1999) fit a heavy-tailed Pareto distribution to structural damage due to extreme wind gusts, and Rootzén and Tajvidi (1997) showed that the GP distribution (with $\gamma > 0$) fits wind storm damage better than the more traditional (medium-tailed) lognormal. Katz (2001) found evidence that the distribution of economic damage from hurricanes, primarily caused by high winds and flooding, has a heavy tail.

The question remains of whether the heavy-tailed distribution of economic damage is attributable to the underlying weather or climate variable being heavy-tailed. Recalling that the Pareto distribution was originally devised in an attempt to measure economic inequality (Arnold, 1983, Chapter 1), the distribution of income or wealth has a general tendency to be heavy tailed. So, if extreme wind speed does indeed have a distribution with a bounded tail as noted earlier, then the "income" effect alone might well be responsible for the associated damages being heavy tailed. On the other hand, if damage did not have a heavy-tailed distribution, it seems implausible that the underlying climate variable could be heavy tailed.

ACKNOWLEDGMENTS

This work has benefited from collaboration with Philippe Naveau and Marc Parlange. Research was partially supported by NSF Grant DMS-9815344 to the NCAR Geophysical Statistics Project. NCAR is sponsored by the National Science Foundation.

REFERENCES

Anderson, P.L. and M.M. Meerschaert, 1998: Modeling river flows with heavy tails. *Water Resour. Res.*, **34**, 2271–2280.

Arnold, B.C., 1983: *Pareto Distributions*. International Co-operative Publishing House, Fairland, MD.

Brown, B.G., and R.W. Katz, 1995: Regional analysis of temperature extremes: Spatial analog for climate change? *J. Climate*, **8**, 108–119.

Buishand, T.A., 1989: Statistics of extremes in climatology. *Statistica Neerlandica*, **43**, 1–30.

Dorland, C., R.S.J. Tol, and Palutikof, J.P., 1999: Vulnerability of the Netherlands and Northwest Europe to storm damage under climate change. *Climatic Change*, **43**, 513–535.

Karl, T.R., and R.W. Knight, 1998: Secular trends of precipitation amount, frequency, and intensity in the United States. *Bull. Amer. Meteor. Soc.*, **79**, 231–241.

Katz, R.W., 2001: Stochastic modeling of economic impact of extreme events: Application to hurricane damage. Submitted for publication.

Palutikof, J.P., B.B. Brabson, D.H. Lister, and S.T. Adcock, 1999: A review of methods to calculate extreme wind speeds. *Meteorol. Appl.*, **6**, 119–132.

Petersen, W.A. (and 8 others), 1999: Mesoscale and radar observations of the Fort Collins flash flood of 28 July 1997. *Bull. Amer. Meteor. Soc.*, **80**, 191–216.

Reiss, R.-D., and M. Thomas, 1997: *Statistical Analysis of Extreme Values*. Birkhäuser, Basel, 316 pp.

Rootzén, H., and N. Tajvidi, 1997: Extreme value statistics and wind storm losses: A case study. *Scandinavian Actuarial J.*, 70–94.

Smith, R.L., 2001: Extreme value statistics in meteorology and the environment. *Environmental Statistics*, Chapter 8, pp. 300–357. [available at: *http://www.stat.unc.edu/postscript/rs/envstat/env.html*]