Geophysical Research Abstracts,
Vol. 11, EGU2009-6357-1, 2009
EGU General Assembly 2009

Annual Rainfall Maxima: Practical Estimation Based on Large-Deviation Results

C. Lepore (1,2), D. Veneziano (2), and A. Langousis (2)

(1) Massachusetts Institute of Technology - Parsons Laboratory for Environmental Science and Engineering – MIT, Cambridge MA 02139, U.S.A. (chlepore@MIT.EDU), (2) Massachusetts Institute of Technology - Department of Civil and Environmental Engineering, MIT, Cambridge MA 02139, U.S.A.

In a separate communication (Veneziano *et al.*, "Annual Rainfall Maxima: Large-deviation Alternative to Extreme-Value and Extreme-Excess Methods," EGU 2009), we show that, at least for scale-invariant rainfall models, classical extreme value analysis based on Gumbel's extreme value (EV) theory and peak-over-threshold (PoT) analysis based on Pickands' extreme excess (EE) theory do not apply to annual rainfall maxima (AM). A more appropriate theoretical setting is provided by large-deviation (LD) theory. This paper delves with some practical implications of these findings.

All above theories predict that, as the averaging durations $d \to 0$, (1) the annual maximum rainfall intensity in d, $I_{year}(d)$, has generalized extreme value (GEV) distribution, (2) the excess of the average intensity in d, $I(d)$, above a level u on the order of the annual maximum has generalized Pareto (GP) distribution, and (3) the GEV and GP distributions have the same shape parameter k. The value of k depends on the theory used. According to EV and EE theories, k is determined by the upper tail of $I(d)$, whereas LD theory shows that k is determined by less extreme regions of the distribution of $I(d)$. The LD parameter k_{LD} is always in the EV2 range and is larger than the value $k_{EV/EE}$ predicted by EV and EE theories.

Since all theories predict that the annual maxima have GEV distribution and the corresponding excesses have GP distribution, methods that directly fit GEV and GP distributions to the data without reference to its asymptotic value should not be affected by which theory is correct. However, the theoretical results have other significant practical implications:

1. Accurate estimation of k from at-site data is difficult. For this reason, k is often estimated regionally. The estimate of k from LD theory is much more robust than that from EV and EE theories and relies on the scaling of the moments of rainfall of order 2.5-3.5. This scaling is nearly universal for rainfall, providing a good "prior" value of k (around 0.3-0.4), which can be used also at un-gauged sites.

2. The shift of focus to regions of the marginal distribution of $I(d)$ below the extreme upper tail, and the recognition that in practice one needs extreme rainfall estimates over a range of finite durations $d$for which $I_{year}(d)$ does not have GEV distribution make non-asymptotic methods more attractive. These methods fit marginal distributions to the order statistics of $I(d)$ or to PoT values above thresholds not much below the level of the annual maxima and estimate the distribution of $I_{year}(d)$ as

$$P[I_{year}(d) > x] \approx \{P[I(d) > x]\}^{n(d)} \qquad (1)$$

$$P[I_{year}(d) > x] \approx e^{-\lambda_{d,u} P[I_{PoT}(d;u) > x-u]} \qquad (2)$$

where $n(d)$ is a parameter that gives the effective number of independent $I(d)$ variables in one year, $\lambda_{d,u}$ is the annual rate at which $I(d)$ up-crosses level u, and $I_{PoT}(d;u)$ is the PoT intensity for averaging duration d and threshold u.

We have implemented procedures based on these non-asymptotic approaches, with the following specific characteristics:

1. The distributions of $I(d)$ (in the upper region) and $I_{PoT}(d;u)$ are taken to have scaled lognormal shape, with 3 parameters (the mean value m, the variance σ^2, and a scaling factor c >0 on the probability density). This choice of distribution is based on both empirical evidence and asymptotic multifractal results;
2. The unknown parameters $\{m$, σ^2, c, $n(d)\}$ or $\{m$, σ^2, c, $\lambda_{d,u}\}$ are estimated simultaneously from marginal or PoT and AM data (the latter data mainly constrain $n(d)$ and $\lambda_{d,u}$) using maximum likelihood.
3. The upper region for $I(d)$ is chosen such that the predicted AM distribution from Eq. 1 closely matches the empirical AM distribution.

Application to several actual and simulated rainfall records shows that this approach is superior in accuracy and robustness to conventional AM and PoT methods.

This work is funded by project RISK (High Speed Rail) of the M.I.T.-Portugal Program.