European Geosciences Union
General Assembly 2009
Vienna, Austria, 19 – 24 April 2009

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

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

(1) Parsons Laboratory for Environmental Science and Engineering, MIT, USA - (2) Dept. Civil Environmental Engineering, MIT, USA

Motivation

In a separate oral communication (Veneziano et al., "Annual Rainfall Maxima: Large-deviation Alternative to Extreme-Value and Extreme-Excess Methods"), we show that, at least for scale-invariant rainfall models, classical extreme value (EV) and excess (EE) theories do not apply to the annual rainfall maxima (AM) under multifractal rainfall; a more comprehensive theoretical setting to study annual maxima is provided by large-deviation (LD) theory. Here we present some practical implications of these theoretical findings for the estimation of the Intensity-Duration-Frequency (IDF) Curves.

According to conventional EV and EE thinking, at least for short averaging durations d,

(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,

(3) the GEV and GP distributions have the same shape parameter k,

(4) k is determined by the upper tail of $I(d)$.

Application of LD theory and multifractal analysis show that the above is incorrect and,

(1) for $d \to 0$, $I_{year}(d)$ has EV2(k) distribution where,
- k depends on the distribution of the scaling factor not the upper tail of $I(d)$,
- k is always higher than the value from EV and EE theories,

(2) for finite d, the distribution of $I_{year}(d)$ between return periods of practical interest is well approximated by an EV2(k) distribution where,
- $k(d)$ depends on d and the MF parameters,
- $k(d)$ is nearly universal.

Estimation of IDF values

Let $i(d,T)$ be the average rainfall intensity in an interval of duration d with return period T. The IDF curves are plots of $I(d,T)$ against d for different fixed T. We compare several estimators of $I(d,T)$. Some are classical and other are new.

Classical Estimators use either Annual Maxima (AM) or Marginal Excesses (ME). The distribution of the annual maxima is assumed to be GEV type

$$GEV \to F_{Iyear}(i) = \exp\left\{-\left[1+k\left(\frac{i-\mu}{\sigma}\right)\right]^{-1/k}\right\} \quad , params: \mu,\sigma,k \qquad (1)$$

and the distribution of the marginal excesses above threshold u is assumed to be GP

$$GP \to F_I(i) = 1-\left[1+k\,\frac{i-u}{\sigma}\right]^{-1/k} \quad , params: \sigma,k \qquad (2)$$

When using Eq.1 $I(d,T)$ is obtained as the upper $1/T$-quantile of $I_{year}(d)$, whereas in the case of Eq. 2 $I(d,T)$ is the $(1/\lambda T)$-upper quantile of I, λ being the rate of which the threshold u is exceeded.

The parameters in Eqs. 1 and 2 are fitted by Maximum Likelihood (ML) or by Probability Weighted Moments (PWM). In Eq. 2, we have fixed u such that the estimator of k has the minimum RMSE.

Alternative Estimators include (i) methods that obtain k from multifractal theory (mf) and (ii) estimators that use different distributional assumptions.

Methods (i) retain the use of GEV for the annual maxima and GP for the marginal excesses. The important difference is that k is derived from the scaling parameters of a beta-lognormal MF model (see above oral presentation). We refer to this estimator as $k_{mf}(d)$. Another possibility is to use a "universal" default value $k_{def}(d)$, obtained as the average of $k_{mf}(d)$ over many sites (see above oral presentation). The other parameters are estimated by ML.

Methods (ii) are based on the assumption that the marginal excess has a truncated lognormal distribution and therefore that the upper tail of the marginal distribution of $I(d)$ above some threshold u has scaled lognormal (SLN) form

$$SLN \to F_I(i) = 1-P_1\cdot\left\{1-\Phi\left(\frac{\ln(i)-\mu}{\sigma}\right)\right\} \quad , params: P_1,\mu,\sigma \qquad (3)$$

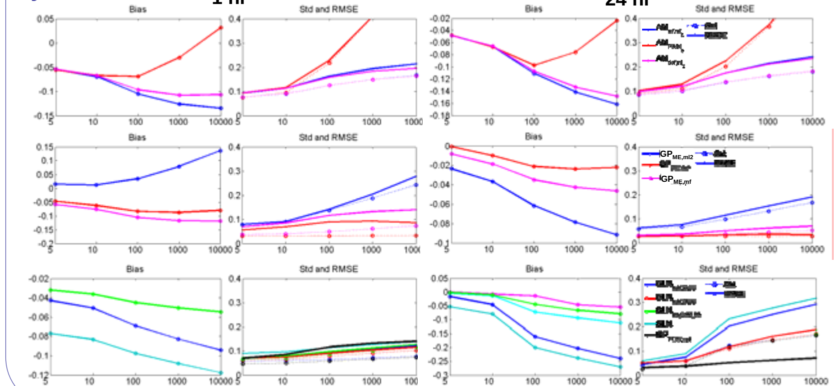
where Φ is the standard normal distribution. This SLN distribution is supported by empirical evidence and asymptotic ($d \to 0$) multifractal analysis. To obtain the distribution of the annual maximum $I_{year}(d)$, the distribution in Eq. 3 is raised to a power $n(d)$ that is either estimated by ML using AM data or fixed to $1/d$ with d expressed in years. The parameters (μ, σ^2) are estimated by ML using marginal excess data and P_1 is either estimated from MF theory as P_{1mf} (like k in Eq. 2) or found by ML together with μ and σ^2.

Legend

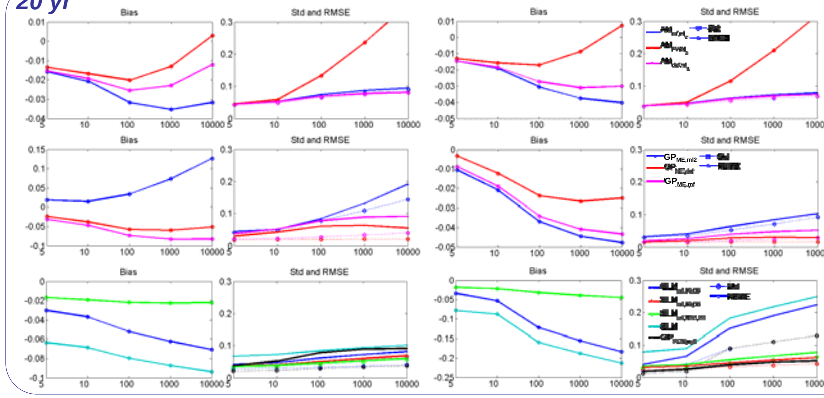
$AM_{mf,ml2}$	– GEV fitted to AM data with k_{mf} and the other 2 parameters estimated by ML.
$AM_{def,ml2}$	– GEV fitted to AM data with k_{def} and the other 2 parameters estimated by ML.
AM_{PWM3}	– GEV fitted to AM data, 3 parameters (including k) are estimated by PWM.
$GP_{ME,ml2}$	– GP fitted to ME data with both parameters estimated by ML.
$GP_{ME,def}$	– GP fitted to ME data with k_{def} and the remaining parameter s estimated by ML.
$GP_{ME,mf}$	– GP fitted to ME data with k_{mf} and the remaining parameter s estimated by ML
$SNL_{n(d),XX,YY}$	– SLN fitted to ME data, all 3 parameters including $n(d)$ fitted to the ME data with threshold equal to the XX quantile for d=1hr and the YY quantile for d=24hr.
SNL	– SLN distribution with $n(d)$ = 1/d, XX=0.8 and YY=0.6.

5 yr records

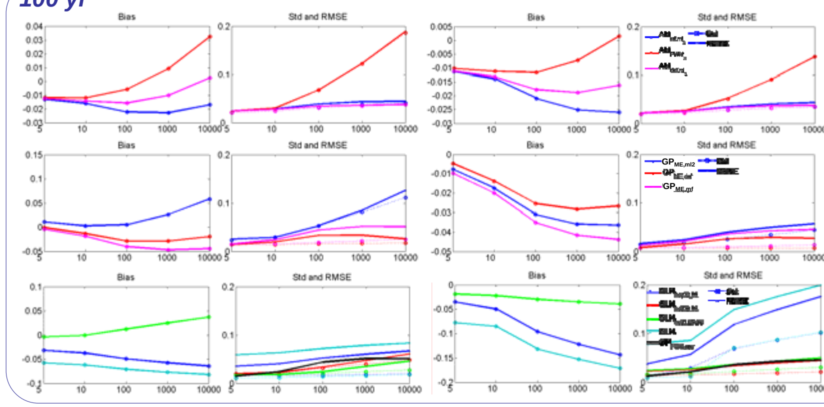
1 hr　　24 hr



20 yr



100 yr



Method for assessing different estimators

- All the estimator of $I(d,T)$ are applied to records simulated as uninterrupted sequences of iid multifractal cascades of the beta-lognormal type with parameters C_β=0.5, C_{LN}=0.04. The duration of each cascade is 15 days.
- The figures show the BIAS, STD and RMSE of $\log_{10}[I(d,T)]$ for d=1 hr and 24 hr and return periods T ranging from 5 to 10000 years (T varies along the horizontal axes).
- Results are shown for record lengths of 5, 20 and 100 years (50 simulations each).
- For each simulated record the $I(d,T)$ estimates are compared with the corresponding exact values obtained analytically by the method of Langousis et *al.* 2007.

Results and Conclusions

Results are displayed in 3 boxes, each for one record length. Within each box, the first row considers estimators that use the AM values and the GEV distribution in Eq. 1, the second row uses the ME values and the GP distribution in Eq.2, and the third row is for estimators that use the SLN distribution of the excesses in Eq. 3.

(See legend of the figures for more details).

The results show that the estimators that either fix k to default values or estimate k using the multifractal method clearly outperform the traditional estimators. This is especially true for long return periods. As expected the relative accuracy of methods based on annual maxima increases as the record length increases. More detailed comments follow.

For short records (5-20 years):

- The estimator based on ME with $k=k_{def}$, $GP_{ME,def}$, performs best, mostly because in this case k_{def} is close to the optimal k. The new estimator that uses $k=k_{mf}$, $GP_{ME,mf}$, is the next high performer, with a generally small bias and even smaller std. The traditional estimator $GP_{ME,ml2}$ (with k and σ estimated by ML using the optimum threshold u as described above) has a relatively small bias but a high variance and performs much worse than $GP_{ME,mf}$.
- AM_{mf+ml2} performs nearly as well as GP_{MEmf} for records of duration ≥ 5 years (d=1 hr) or ≥ 20 years (d=24 hr). This is true especially for long return periods. This result was partly expected as GP_{MEmf} and AM_{mf+ml2} share the same multifractal estimator of k.
- the traditional estimators based on AM data, AM_{pwm3} and AM_{ml3} are highly erratic.
- the estimator based on the SLN model in Eq. 3 performs best when $n(d)$ is estimated from annual maximum data (SLN_{nd} in figures) as opposed to being fixed to 1year/d (SLN). For d=1hr these estimators perform better than the GP estimators, whereas for d=24 hr the performance is comparable.

For long records (>50 years)

- The AM estimator with k_{mf}, $AM_{mf,ml2}$, performs similarly to $GP_{ME,mf}$; it may even outperforrm $GP_{ME,mf}$ for long return periods.

References

1. Veneziano, D., A. Langousis and C. Lepore (2009) Annual Rainfall Maxima: Large-Deviation Alternative to Extreme-Value and Extreme-Excess Methods, European Geosciences Union, Vienna, Austria, April 2009.
2. Langousis A, D. Veneziano, P. Furcolo, and C. Lepore (2007) Multifractal Rainfall Extremes: Theoretical Analysis and Practical Estimation, *Chaos Solitons and Fractals*, doi:10.1016/j.chaos.2007.06.004.

Acknowledgments

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