Comparison of Marginal and Annual-Maximum Methods of IDF Curve Estimation

European Geosciences Union
General Assembly 2007
Vienna, Austria
15 - 20 April 2007

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

(1) Dept. of Civil and Environmental Engineering, MIT, Cambridge, 02139, MA, U.S.A.

(2) Dipartimento di Ingegneria Civile, University of Salerno, Fisciano, SA 84084, Italy

Objectives

We propose IDF curve estimation methods based on the marginal distribution of rainfall intensity and compare the new estimators to standard procedures that use historical annual maxima. The latter procedures assume that the *T*-year rainfall intensity for duration *d*, $I_{max}(d,T)$, is a separable function of *T* and *d*:

$$i_{\max}(d,T) = a(T)\cdot b(d) \qquad (1)$$

where *a* and *b* are suitable functions. The use of marginal rather than annual-maximum information increases the accuracy of the estimators and their robustness against outliers, especially when the rainfall record is only a few years long. If rainfall has multifractal scale invariance, the marginal methods have an especially lean parameterization. We also consider hybrid methods that estimate the IDF curves using both marginal and annual-maximum rainfall information.

A. Marginal and Hybrid Methods

Marginal methods estimate the distribution $F_{I(d)}$ of the average rainfall intensity in *d* and then find the distribution of the annual maximum intensity $I_{max}(d)$ as

$$F_{I_{\max}(d)}(i) = [F_{I(d)}(i)]^{1/d} \qquad (2)$$

where *d* is duration in years. Finally, the IDF value $\hat{\imath}_{max}(d,T)$ is obtained as the *(1-1/T)*-quantile of $F_{Imax(d)}$. Equation 2 makes the simplifying assumptions that (i) the maximum annual rainfall occurs in one of the *1/d* intervals into which the year is partitioned and (ii) rainfall in different *d*-intervals *is* independent. Results based on these assumptions are accurate, especially for long return periods *T*.

For the calculation of $F_{Imax(d)}$, it is important to accurately estimate $F_{I(d)}$ in the upper tail. As illustrated in **Figure 1**, this upper tail has approximately lognormal shape, as in a distribution of the type

$$F_{I(d)}(i) = P_0 + (1-P_0)\Phi\left(\frac{\ln i - m}{\sigma}\right) \qquad (3)$$

where Φ is the standard normal CDF, P_0 is the probability that a *d* interval is dry and *m* and σ are parameters of the log rainfall intensity.

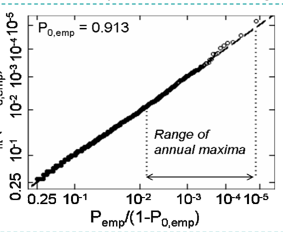

Figure 1 - Tail plot of the lognormal exceedance probability from Eq. 3 against the corresponding empirical exceedance probability for 1-hr intensities (the dashed line indicates a perfect match)

We have found that good results are obtained by estimating the parameters (P_0, *m*, σ) to match the first three moments of *I(d)*. Variants of the method are:

1 - When the parameters are fitted to the empirical moments of *I(d)*, we call the resulting IDF estimation procedure the ***local marginal (LM) method***.

2 - In the case of multifractal rainfall, the moments of *I(d)* can be estimated by fitting straight log-log lines to the empirical moments inside the scaling range. We call this the ***multifractal marginal (MFM) method***. This method produces smoother IDF curves than the LM method.

3 - ***Hybrid versions*** of the LM and MFM methods calibrate the distributions $F_{Imax(d)}$ from Eq. 2 such that their mean value matches the sample average of the annual-maximum value.

Acknowledgements

This work was supported in part by the National Science Foundation under Grant No. EAR-0228835, in part by the Alexander S. Onassis Public Benefit Foundation under Scholarship No. F-ZA 054/2005-2006, and in part by the MIUR ex60% fund of the University of Salerno. The UK Meteorological Office is acknowledged as provider of the Heathrow Data set which was used as part of collaborative work with Christian Onof.

References

Asquith, W.H. (1998) Depth-Duration Frequency of Precipitation for Texas, Water-Resources Investigations Report 98–4044, U.S. Geological Survey, http://pubs.usgs.gov/wri/wri98-4044/pdf/98-4044.pdf.

Bessemoulin 2006, personal communication.

Koutsoyiannis, D., Kozonis, D., and Manetas, A. (1998) A Mathematical Framework for Studying Rainfall Intensity-Duration-Frequency Relationships, J. Hydrol., 206: 118-135.

Veneziano, D., Lepore, C., Langousis, A., and Furcolo, P. (2007) Marginal Methods of IDF Estimation in Scaling and Non-Scaling Rainfall , WRR (submitted).

B. Two Annual-Maximum Methods (Koutsoyiannis et al., 1998)

Semi-parametric Annual Maximum (SPM) Method

The method consists of two steps:

1. Assume a distribution type for the yearly maxima $I_{max}(d)$ and fit the parameters separately for each duration *d*;
2. Estimate a model of the type in Eq. 1 by least-squares fitting the $\hat{\imath}_{max}(d,T)$ estimates from the first step. The fitted model has a parametric *b(d)* function and a nonparametric *a(T)* function.

A popular choice for *b(d)* is the power function:

$$b(d) = 1/(d+\delta)^{\eta} \qquad \delta,\eta > 0 \qquad (5)$$

Completely Parametric Annual-Maximum (CPM) method

Assume that *b(d)* has the form in Eq. 5 and the reduced yearly maxima $Y(d) = I_{max}(d)/b(d)$ have the same distribution for all *d*. Using the two-step procedure of Koutsoyiannis et al. (1998), one finds the parameters of *b(d)* by minimizing the Kruskal-Wallis index for *Y*(*d*) and then obtains *a(T)* by fitting a GEV distribution of the type

$$F(x) = \exp\left\{-\left[1+k\left(\frac{x-\psi}{\lambda}\right)\right]^{-1/k}\right\} \qquad (6)$$

to the combined set of *Y*(*d*) values. In our implementation, the parameters *k*, ψ and λ in Eq. 6 are estimated using the PWM method, which is robust against outliers.

C. IDF Results for Three Historical Records

The methods are compared using historical records from Heathrow Airport (UK), Walnut Gulch (Arizona) and Florence (Italy) of length 51, 49, and 24 yr, respectively. Figure 3 compares results from different methods with the empirical IDF curves. The empirical return period of the *i*th ranked maximum from a series of *n* years is calculated as $T_i=(n+1)/i$.

The three empirical and three lowest model curves are for *T*= [2, 8, 25] for Florence, *T* = [2, 8, 52] for Heathrow, and *T* = [2, 8, 50] for Walnut Gulch. The top three model curves are for *T* = [100, 1000, 10000] yr in all panels.

Due to the separability condition in Eq. 1, the IDF curves estimated by the annual-maximum methods (top panels in Figure 3) are parallel and for long durations *d* tend to be more widely spaced than the empirical curves. By contrast, the marginal and hybrid methods produce non-parallel IDF curves that more closely track the empirical ones.

The range of *d* in Figure 3 generally corresponds to the scaling range. Inside this range, the LM and MFM methods produce similar results. For Walnut Gulch, longer durations outside the scaling range are also shown, to illustrate the local marginal method (green dashed lines) in a non-scaling case.

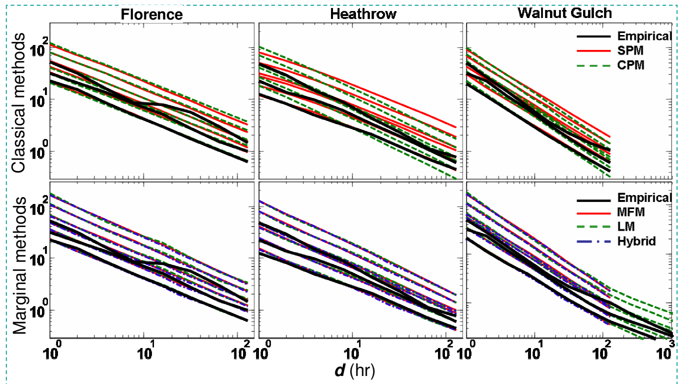

Figure 3 - IDF curves generated by various methods for three historical records.

D. Assessment of Different Methods

1. Separability Condition

Figure 4 shows the variability of the GEV shape parameter *k* with duration *d*. For many data sets, including those considered in this study (dashed lines), *k(d)* is a concave function. The fact that *k* varies with *d* is an indication that the separability assumed by the classical methods does not hold.

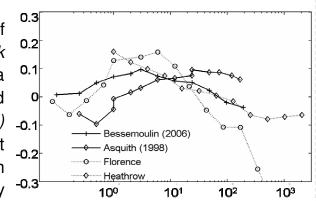

Figure 4 - Variation of the GEV parameter *k* with duration *d*.

2. Sensitivity to Outliers

The Heathrow and Florence records include "outlier years" (1959 and 1970 for Heathrow, 1966 for Florence). Figure 5 shows the sensitivity to outliers for Florence by plotting the ratio of the IDF estimates with and without 1966 against duration *d*, for different estimation methods and return periods *T*. The annual-maximum methods are much more sensitive than the proposed marginal methods, especially for the long durations *d* for which the maximum rainfalls in 1966 are highly anomalous. Similar results were obtained for Heathrow.

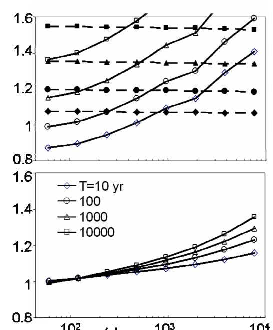

Figure 5 - Florence record. Ratio of the estimated IDF values with and without the outlier year 1966.

3. Bias and Variability for Short Records

Figure 6 shows the bias and variability of the 1-hr $\log_{10}$(IDF) estimates for *T* = 10,100,1000 and 10000 yr, when only 5 years of the empirical record are used [F = Florence, H = Heathrow Airport, W = Walnut Gulch].

The deviations of the 5-yr $\log_{10}$(IDF) estimates from whole-record results are used to estimate the bias *b* and the standard deviation σ and RMS = $(b^2+\sigma^2)^{1/2}$ of the $\log_{10}$(IDF) estimation error. Two hybrid cases are also included in this analysis:

- MFM/H1, when the yearly maxima are assumed available only for the 5-yr segment of the record;
- MFM/H2, when the yearly maxima are assumed available for the entire duration of the record.

The annual-maximum methods perform rather poorly due to high bias, high variance, or both. The main cause of the high variability is that estimation of the GEV parameters is rather erratic and sensitive to outliers. By contrast, the marginal and hybrid methods are nearly unbiased and have moderate variance. As expected, MFM/H2 outperforms MFM/H1. Similar results were obtained for the 24-hr estimates.

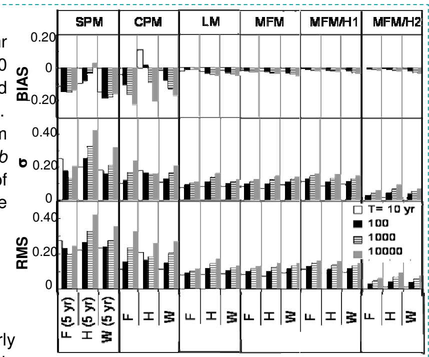

Figure 6 - Bias, standard deviation and RMS error when using 5-yr subsets of the entire records

Conclusions

- Classical IDF curve estimators based on annual maxima are simple but make the inappropriate assumption of separability. Their IDF estimates for long return periods are highly variable and sensitive to outlier rainfall events.
- The marginal MFM and LM methods are statistically more stable, more robust against outliers, and applicable also to short rainfall records. Multifractality reduces the parameterization but this advantage is realized only within the scaling range of *d*.
- The combined use of marginal and annual-maximum information in the hybrid method is advantageous when annual maximum values are available for many years.

For a more detailed account of methods and results, see Veneziano et al. (2007).