EU COST Action C22: Urban Flood Management
Working Group 1: Flood Probability Assessment
2nd Meeting, Athens, 2-3 June 2006

The underestimation of probability of extreme rainfall and flood by prevailing statistical methodologies and how to avoid it

Demetris Koutsoyiannis
Department of Water Resources, School of Civil Engineering,
National Technical University, Athens, Greece

Why is modelling of extreme rainfall important for urban flood risk estimation?

- Flow measurements are never enough to support flood modelling using only flow data
- Particularly, in urban floods the control points are numerous and the flow gauge sites scarce
- The example of Athens: No flow gauge at all in Kephisos and Ilisos Rivers and in other urban streams
- The obvious alternative: Use of hydrological models with rainfall input data

Brief history of the probability of extreme values

- **1914 (Hazen):** Empirical foundation of hydrological frequency curves known as "duration curves"
- **1922, 1923 (von Bortkiewicz, von Mises):** theoretical foundation of probabilities of extreme values
- **1958 (Gumbel):** convergence of empirical and theoretical approaches
- **Today**: the estimation of hydrological extremes continues to be highly uncertain

  "... the increased mathematisation of hydrological frequency analysis over the past 50 years has not increased the validity of the estimates of frequencies of high extremes and thus has not improved our ability to assess the safety of structures whose design characteristics are based on them. The distribution models used now, though disguised in rigorous mathematical garb, are no more, and quite likely less, valid for estimating the probabilities of rare events than were the extensions 'by eye' of duration curves employed 50 years ago." (Klemeš, 2000)

The notion of distribution of maxima

- Parent variable: Y (e.g. the hourly or daily rainfall depth)
- Parent distribution function: $F(y)$ (with density $f(y)$)
- Variable representing maximum events

$$X := \max\{Y_1, Y_2, \ldots, Y_n\}$$

- Distribution function of maxima: $H_n(x)$
- Exact distribution of maxima for constant n:

$$H_n(x) = [F(x)]^n$$

- Exact distribution of maxima for randomly varying n, following a Poisson process

$$H'_\nu(x) = \exp\{-\nu[1 - F(x)]\}$$

The notion of asymptotic or limiting distribution of maxima

- Asymptotic or limiting distribution for $n \to \infty$ or $v \to \infty$ (Generalised extreme value distribution – GEV; Jenkinson, 1955)

$$H(x) = \exp\{-\,[1 + \kappa(x/\lambda - \psi)]^{-1/\kappa}\} \qquad (\kappa\, x \geq \kappa\lambda(\psi - 1/\kappa)$$

- In hydrology, un upper bound of x is not realistic, so $\kappa \geq 0$
- If $\kappa > 0$, $H(x)$ represents the (three-parameter) extreme value distribution of maxima of type II (EV2)
- In the special case $\kappa = 0$, $H(x)$ represents the extreme value distribution of maxima of type I (EV1 or Gumbel)

$$H(x) = \exp\{-\exp\,[-(x/\lambda - \psi)]\} \qquad (-\infty < x < +\infty)$$

- In the special case where the lower bound is zero ($\kappa\,\psi = 1$), $H(x)$ is two-parameter EV2 (Fréchet distribution)

$$H(x) = \exp\{-[\lambda/(\kappa\, x)]^{1/\kappa}\} \qquad (x \geq 0)$$

When and why does the type of extreme value distribution (EV1 or EV2) matter?

- In typical storm sewer networks, designed on the basis on return periods [$T = 1 / (1 - H)$] of about 5-10 years, the difference of the two distributions may be negligible (in such return periods even interpolation from the empirical distribution would suffice)
- However, for large T (> 50 years), for which extrapolation is required, EV1 results in risk (probability of exceedence of a certain value) significantly lower than EV2
- That is, for large rainfall depths, EV1 yields the lowest possible probability of exceedence (the highest possible T) in comparison to those of EV2 for any value of κ
- For T >~ 1000, the return period estimated by EV1 could be orders of magnitude higher than that of EV2

What is the prevailing model in hydrological practice?

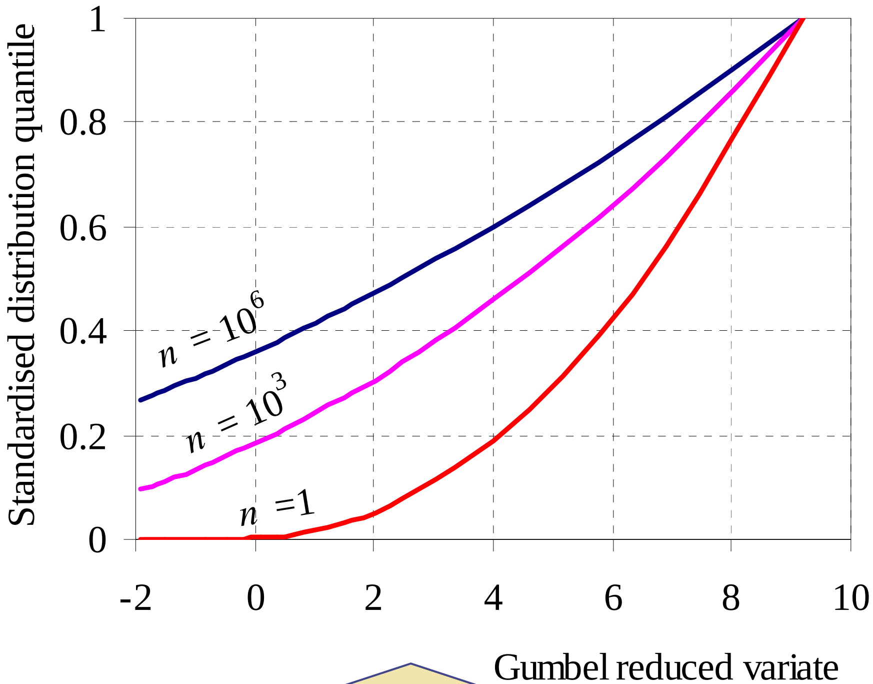
- Definitely, EV1
- For example, most hydrological textbooks do not mention EV2 at all
- Also, in most hydrological studies the adoption of EV1 is “automatic” (especially for extreme rainfall)
- Recently, however, many researchers have expressed scepticism about the appropriateness of EV1

Are there theoretical reasons favouring EV1 against EV2?

- Most types of parent distributions functions used in hydrology, such as **exponential**, **gamma**, **Weibull**, **normal** and **lognormal** belong to the domain of attraction of the Gumbel distribution
- More specifically, rainfall depth at fine time scales (hourly, daily) has been modelled by the gamma or Weibull distributions
- However, the adoption of these distributions is rather empirical, not based on theoretical reasoning
- Thus, the above theoretical argumentation is inconsistent

Assuming that theoretical reasoning supports EV1, what distribution shall I use in my studies?

- **Intuitive answer**: EV1
- **Correct answer**: The exact distribution of maxima, $H_n(x)$ or $H'_\nu(x)$
- The difference of $H_n(x)$ from EV1 may be large
- **Practical answer**: EV2 [it yields good approximation of $H_n(x)$]



Convergence of distribution of maxima for parent distribution Weibull with shape parameter $k = 0.5$

Note: The distribution quantiles have been standardised by $x_{0.9999}$ corresponding to $z_H = 9.21$

What could be a basis of developing theoretical arguments for the selection of the type of the parent distribution?

- The principle of maximum entropy (ME) is a well established mathematical and physical principle that can yield unknown probabilities by theoretical reasoning
- Entropy is defined on grounds of probability theory
- The standard or Boltzmann-Gibbs-Shannon or extensive entropy is

$$\phi := E[-\ln f(Y)] = -\int_{-\infty}^{\infty} f(y) \ln f(y)\, dy$$

- According to a generalized definition, the Tsallis or nonextensive entropy is

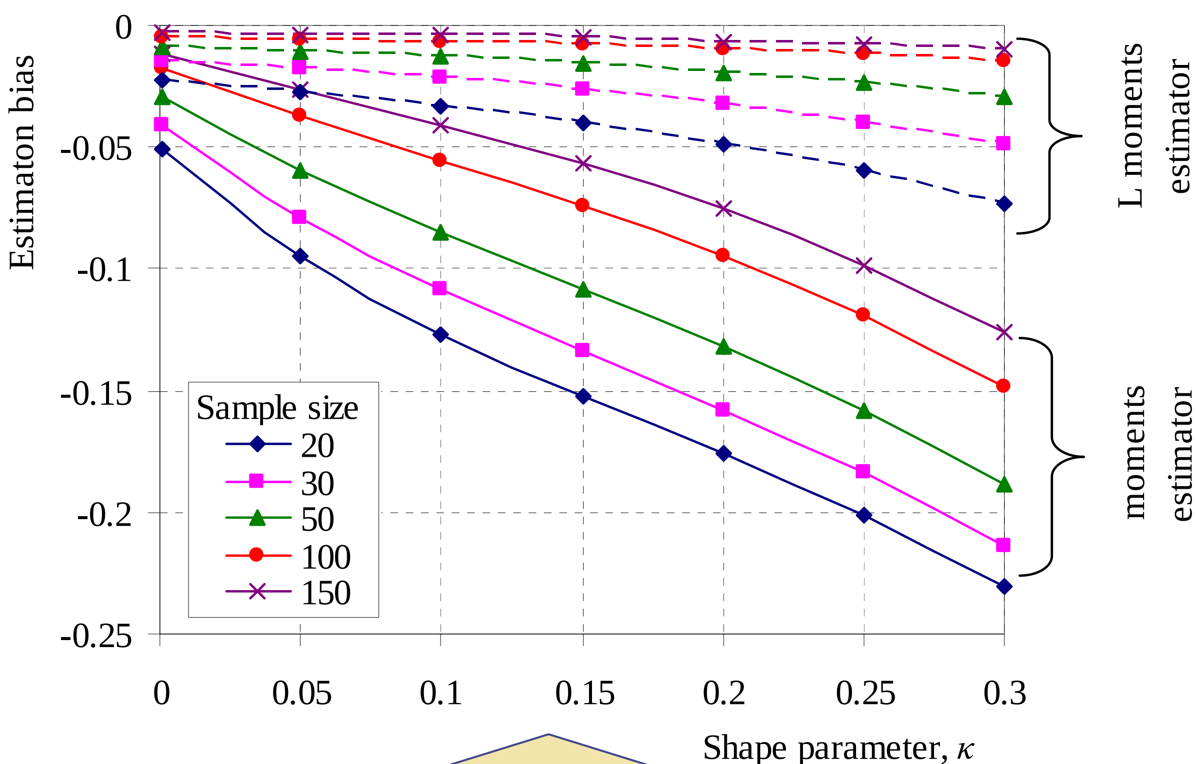
$$\phi_q := \frac{1 - \int_{0}^{\infty} [f(y)]^q\, dx}{q-1}$$

Application of the ME principle

- We assume that Y is continuous and positive, has known mean μ and known variation σ/μ. We can then estimate the distribution function with only this information, applying the ME principle
- Maximization of the standard entropy is possible when variation is relatively low ($\sigma/\mu < 1$) and yields exponential distribution tails (and distribution from normal to exponential)
- However, empirical evidence suggests that at small time scales rainfall exhibits high variation ($\sigma/\mu > 1$)
- In this case we should apply Tsallis entropy, which yields power-type (Pareto) distribution

Why inappropriateness of EV1 has not become evident?

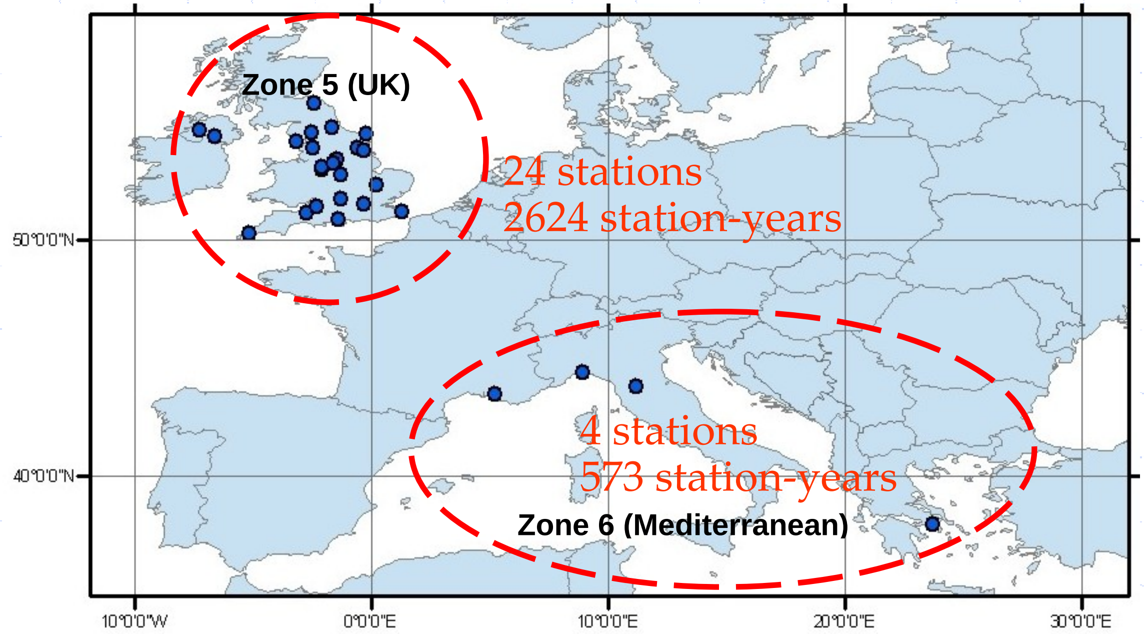
- Even if one is willing to try EV2 as a potential model, it is very likely that one will reject it due to significant bias of estimators
- For small samples, the most common method of moments hides completely EV2
- Even the less biased L-moments method may result in erroneous acceptance of EV1 (e.g. for $\kappa = 0.15$ and $m = 20$ the frequency of not rejecting the EV1 distribution is 80%!)

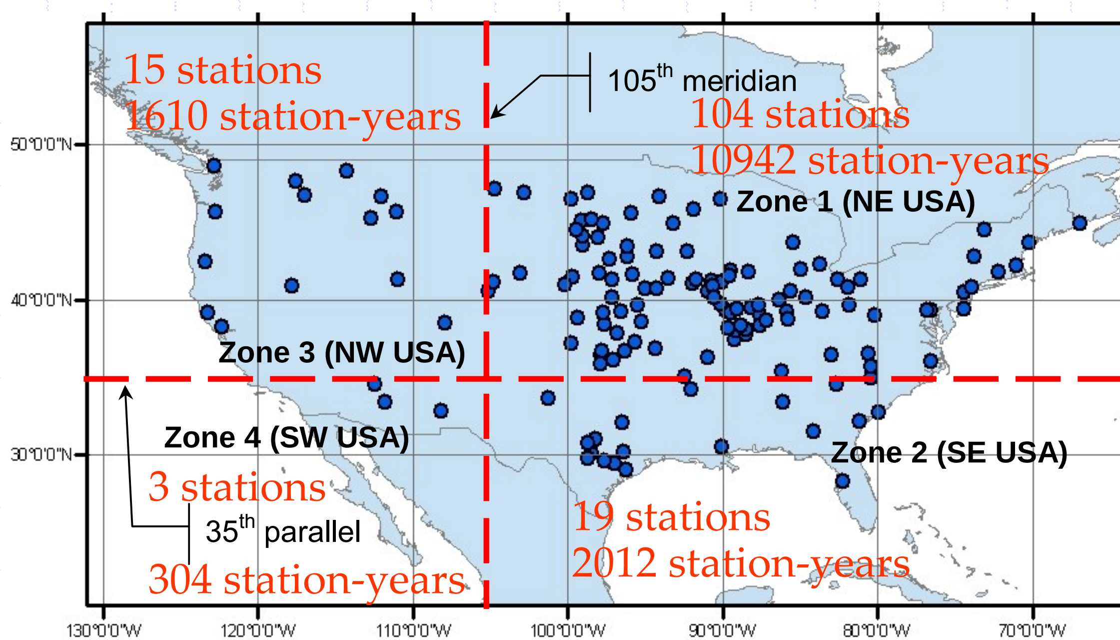


Bias in estimating the shape parameter κ of the GEV distribution
(Obtained by Monte Carlo simulation)

Empirical investigation: Data set

- 169 stations from Europe and North America
- Record lengths 100-154 years
- 18065 station-years in total
- 6 major climatic zones

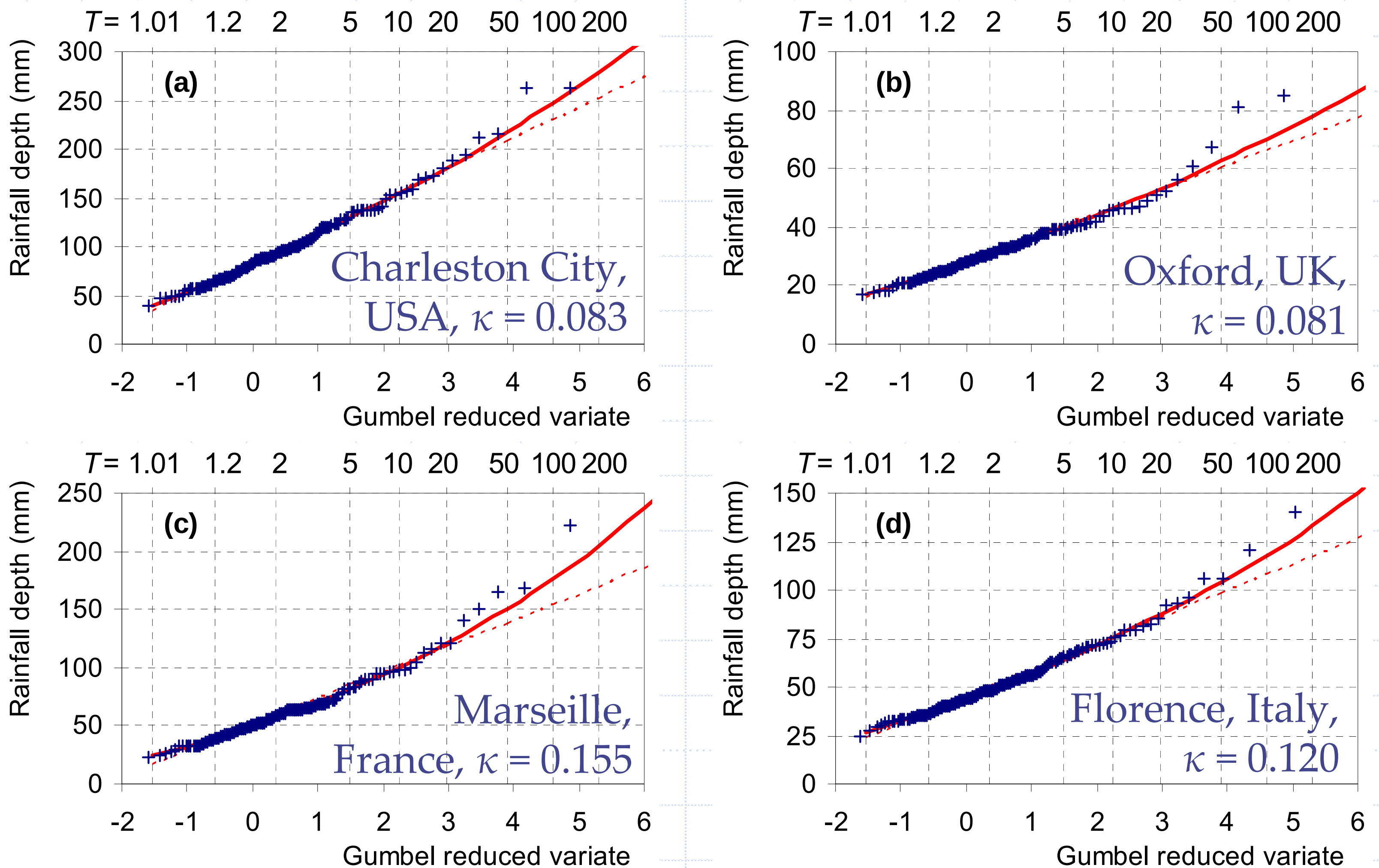




Top ten raingauges (in terms of record length)

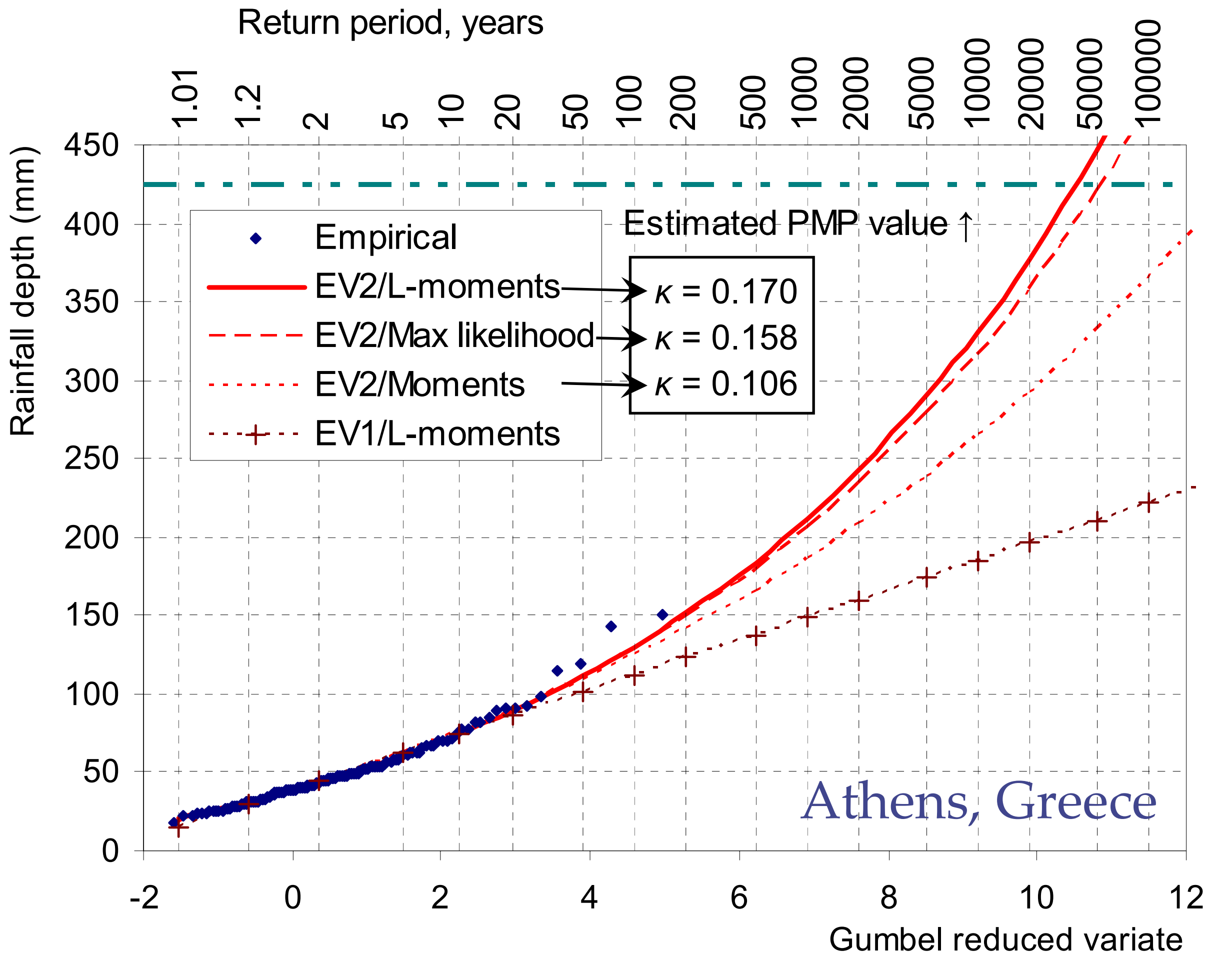
Name	Zone /Country /State	Latitude (°N)	Longitude (°)	Elevation (m)	Record length	Start year	End year	Years with missing values
Florence	6/Italy	43.80	11.20	40	154	1822	1979	4
Genoa	6/Italy	44.40	8.90	21	148	1833	1980	
Athens	6/Greece	37.97	23.78	107	143	1860	2002	
Charleston City	2/USA/SC	32.79	–79.94	3	131	1871	2001	
Oxford	5/UK	51.72	–1.29		130	1853	1993	11
Cheyenne	1/USA/WY	41.16	104.82	1867	130	1871	2001	1
Marseille	6/France	43.45	5.20	6	128	1864	1991	
Armagh	5/UK	54.35	–6.65		128	1866	1993	
Savannah	2/USA/GA	32.14	–81.20	14	128	1871	2001	3
Albany	1/USA/NY	42.76	–73.80	84	128	1874	2001	

Investigation of empirical distributions and comparison with EV2 and EV1 distributions



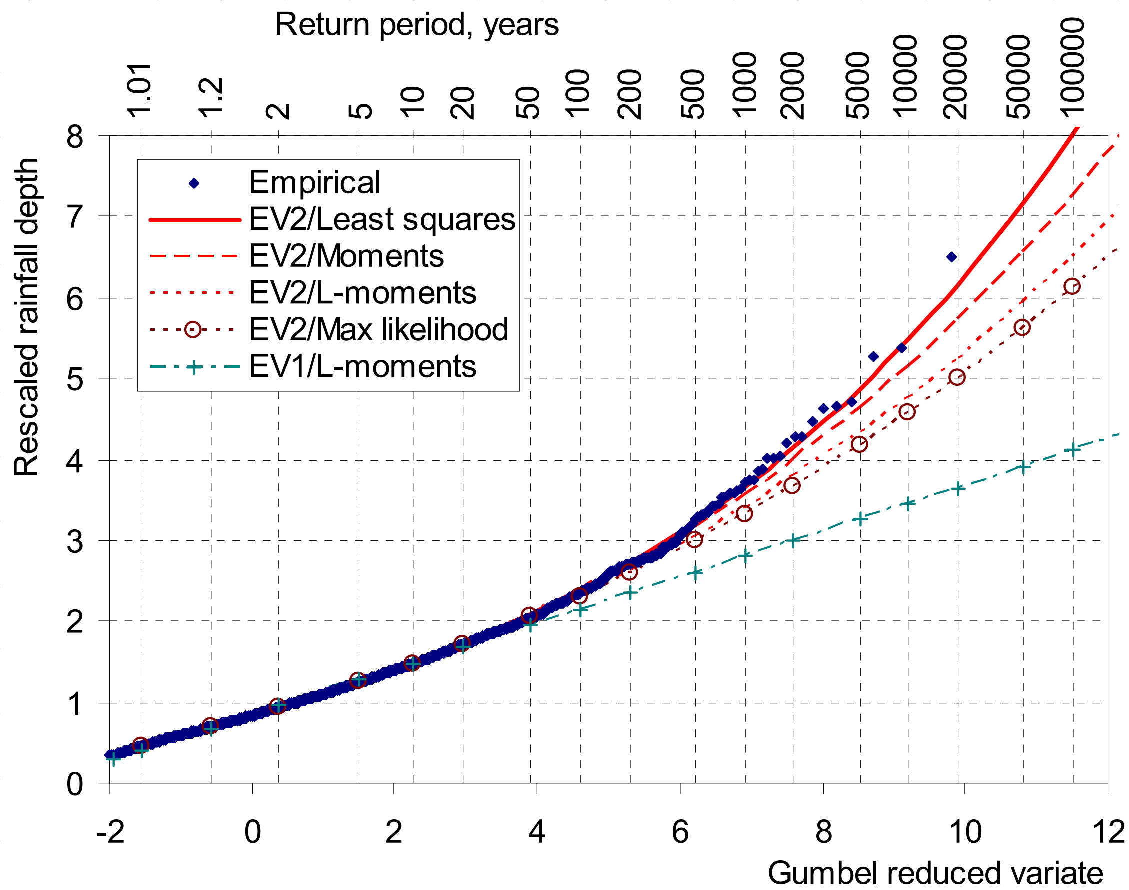
EV2 and EV1 distributions were fitted by the method of L-moments

Demonstration of the differences of EV1 and EV2 estimates of quantiles for high return periods



Standardisation and merging of all records

- Hypothesis of constant dimensionless shape and location parameters (κ, ψ)
- Rescaling of each records by its mean
- Unification of all records (18065 data values)
- Accurate estimation of κ and ψ



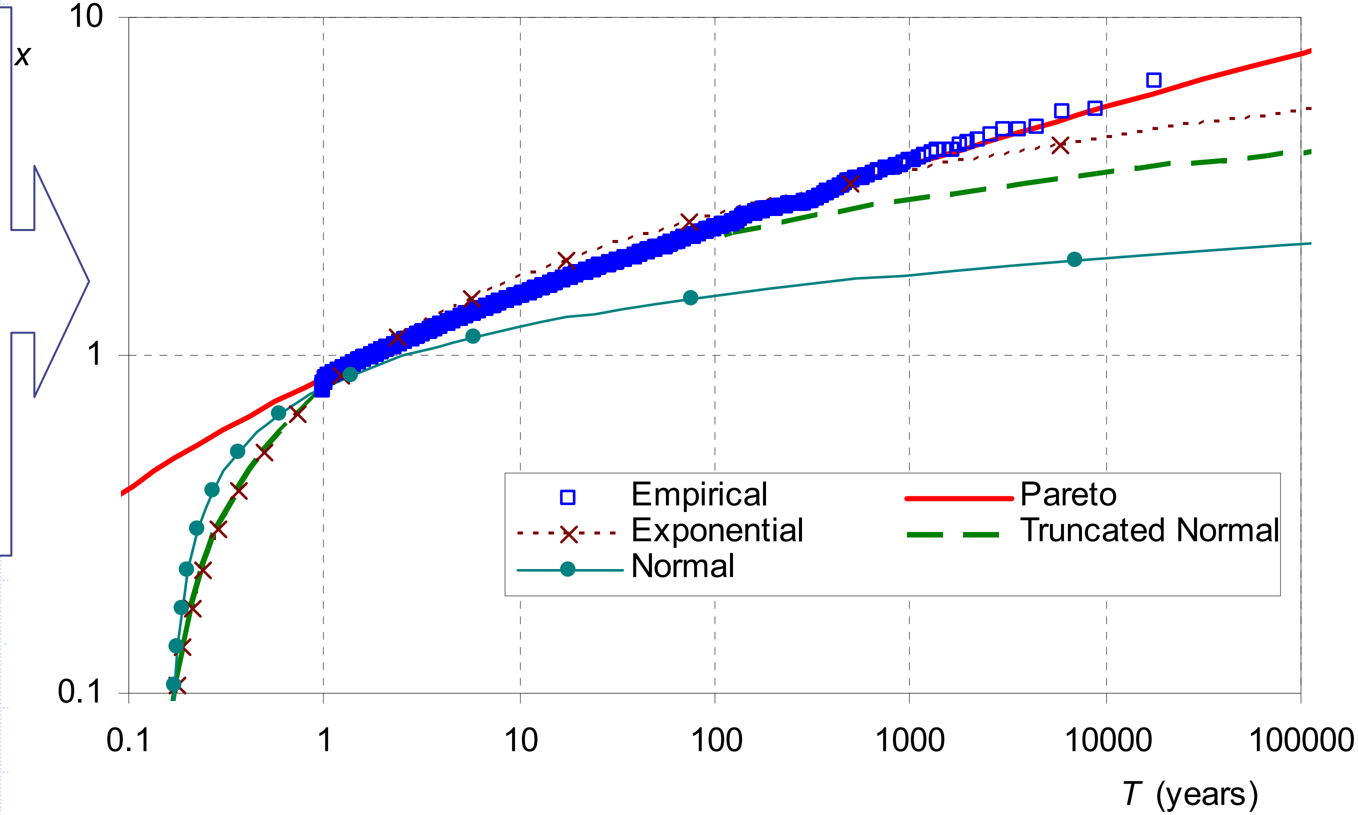
	Estimation method			
Parameter	Max likelihood	Moments	L-moments	Least squares
κ	0.093	0.126	0.104	0.148
λ	0.258	0.248	0.255	0.236
ψ	3.24	3.36	3.28	3.54

Verification of the applicability of the ME principle

Here the parent distribution is analyzed

Data set: Same minus one station (Athens); series above threshold, standardized by mean and unified; period 1822-2002; 17922 station-years of data

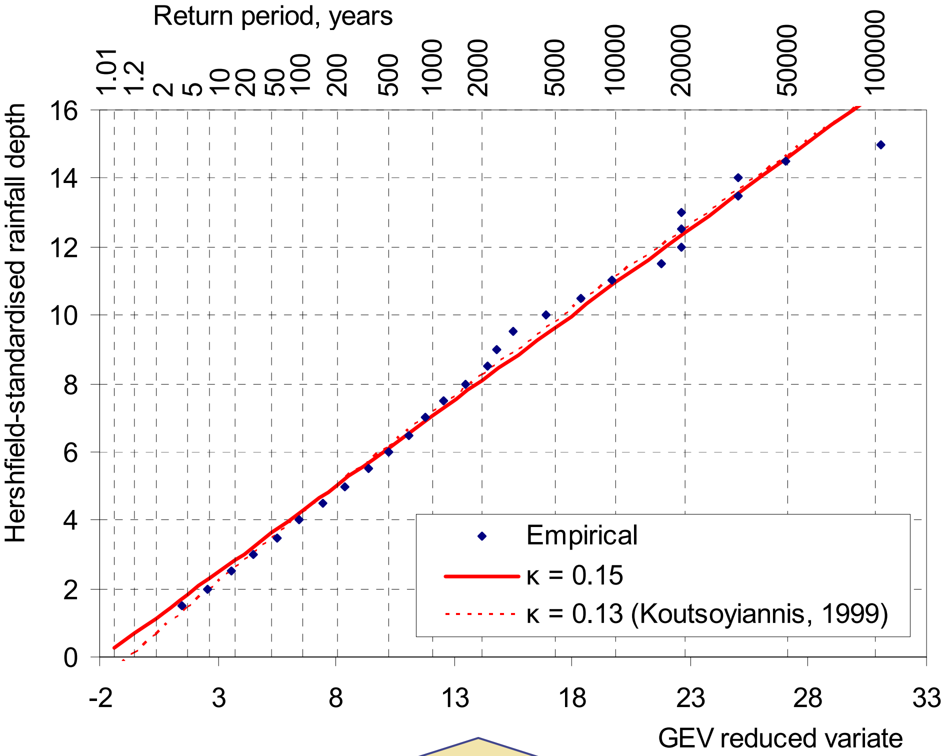
$\mu = 0.28$
(mean minus threshold)
$\sigma/\mu = 1.19 > 1$
ME distribution: Pareto
$\kappa = 0.15$
$\phi_q = 1.16$



Conclusion: Pareto/EV2

Additional support of present findings

- Hershfield's (1961) data set, comprising 95 000 station-years, in a later study (Koutsoyiannis, 1999) was found to have very similar behaviour
- Chaouche (2001) exploited a data base of 200 rainfall series of various time steps (minute-month) from the five continents, each including more than 100 years of data. Using multifractal analyses he showed that
  - a Pareto/EV2 type law describes the rainfall amounts for large return periods
  - the exponent of this law is scale invariant over scales greater than an hour
  - this exponent is almost space invariant



GEV probability plots of the empirical and EV2 distribution functions of standardised rainfall depth *k* for Hershfield's (1961) data set as determined by Koutsoyiannis (1999), and fitted EV2 distributions with $\kappa = 0.13$ (Koutsoyiannis, 1999) and $\kappa = 0.15$

Conclusion and discussion

- Urban flood risk estimation relies mainly on the probability distribution of extreme rainfall
- It can be shown that the distribution tail of flood is of the same type as that of rainfall
- The EV1 distribution, which has been the prevailing distribution in rainfall underestimates risk significantly
- The theoretical and empirical reasons that made the EV1 distribution prevail in hydrology may be not valid
- The principle of maximum entropy and other theoretical arguments support the Pareto/EV2 distribution tails
- Thus to avoid underestimation of risk, a three-parameter EV2 distribution should be used
- The shape parameter κ of EV2 is very hard to estimate on the basis of an individual series, even in series with length 100 years or more
- However, the results of the analysis of 169 long series of rainfall maxima allow the hypothesis that κ is constant ($\kappa = 0.15$) for all examined zones

More information …

- This presentation is available on line at http://www.itia.ntua.gr/e/docinfo/719/
- The full documentation can be found in a couple of papers in *Hydrological Sciences Journal*, August 2004
- References

  - Chaouche K., 2001, *Approche Multifractale de la Modelisation Stochastiqueen Hydrologie*, thèse, Ecole Nationale du Génie Rural, des Eaux et des Forêts, Centre de Paris (http://www.engref.fr/ thesechaouche.htm)
  - Hershfield, D. M., 1961, Estimating the probable maximum precipitation, *Proc. ASCE, J. Hydraul. Div.*, 87(HY5), 99-106
  - Jenkinson, A. F., 1955, The frequency distribution of the annual maximum (or minimum) value of meteorological elements, *Q. J. Royal Meteorol. Soc.*, 81, 158-171
  - Klemeš, V., 2000, Tall tales about tails of hydrological distributions, *J. Hydrol. Engng* **5**(3), 227–231 & 232–239
  - Koutsoyiannis, D., 1999, A probabilistic view of Hershfield's method for estimating probable maximum precipitation, *Water Resources Research*, 35(4), 1313-1322