International Conference On Hydrology: Science & Practice for the 21st Century
Theme 2: Hydrology of extremes
British Hydrological Society, London, 12-16 July 2004

Exploration of long records of extreme rainfall and design rainfall inferences

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

Introduction

- **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 daily rainfall depth)
- Parent distribution function: $F(y)$
- 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)

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

Why does the type of extreme value distribution (EV1 or EV2) matter?

- EV1 results in risk significantly higher than EV2 for engineering structures
- That is, for small probabilities of exceedence $(1 - H)$, or large return periods $[T = 1 / (1 - H)]$, EV1 yields the smallest possible quantiles x_H in comparison to those of EV2 for any value of κ
- For $T = 10^4$-10^6 (used e.g. in the design of major hydraulic structures), the design value estimated by EV1 could be half that of EV2 or less

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

Why EV1 is so prevailing in hydrology?

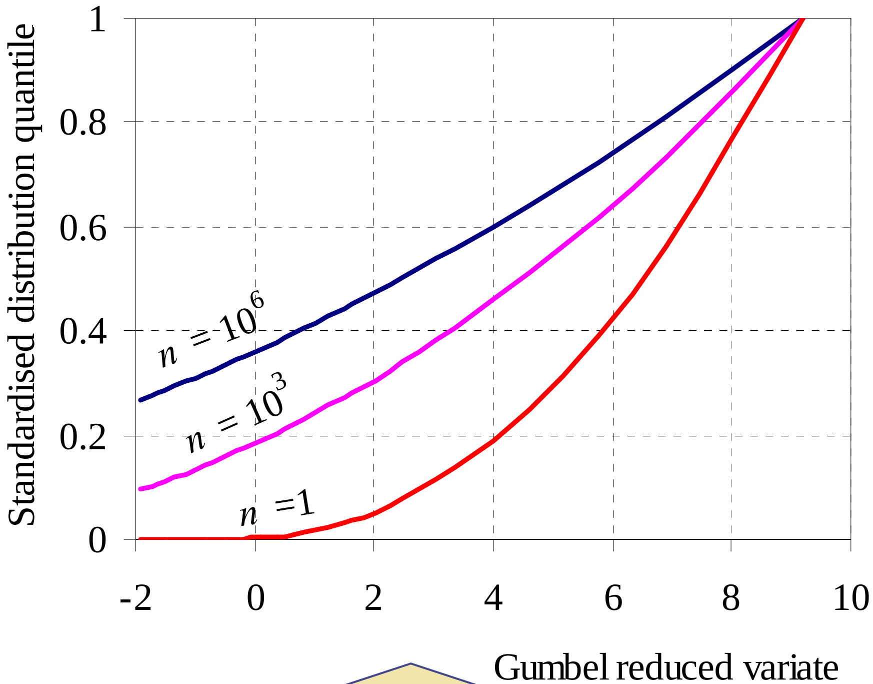
- **Theoretical reasons** (To be discussed later)
- **Simplicity:** The mathematical handling of the two-parameter EV1 is much simpler than that of the three-parameter EV2
- **Accuracy of estimated parameters:** Two parameters are more accurately estimated than three
- **Practical reasons:** EV1 offers a **linear probability plot** (Gumbel probability plot) of observed x_H vs. observed $z_H := -\ln(-\ln H)$ (Gumbel reduced variate); in contrast, a linear probability plot is not possible for the EV2, unless one parameter is fixed
- **Institutional reasons:** Many institutions suggest, or even require, the use 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
- More recent studies advocate a shift from these distributions to Pareto type distributions, which belong to the domain of attraction of EV2
- No concrete conclusions have been drawn to date

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

- **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 $z_H = 9.21$

How stable is EV1 if distributional parameters change?

- Let us assume that rainfall depth at fine time scales (hourly, daily) follows a distribution belonging to the domain of attraction of EV1 (e.g. gamma or Weibull)
- However, the parameters of distribution are not constant all the time but vary due to:
  - seasonal effects
  - overyear (large scale) fluctuations
- Parameter variations may change the domain of attraction to EV2

Theoretical example of the shift of domain of attraction

Assumptions:

Probability density function of Y_i conditional on α_i:

$$f_i(y \mid \alpha_i) = \alpha_i^{\theta}\, y^{\theta - 1}\, \mathrm{e}^{-\alpha_i y} / \Gamma(\theta)$$

Probability density function of α_i:

$$g(\alpha_i) = \beta^{\tau}\, \alpha_i^{\tau - 1}\, \mathrm{e}^{-\beta \alpha_i} / \Gamma(\tau)$$

Results:

Unconditional density function of Y:

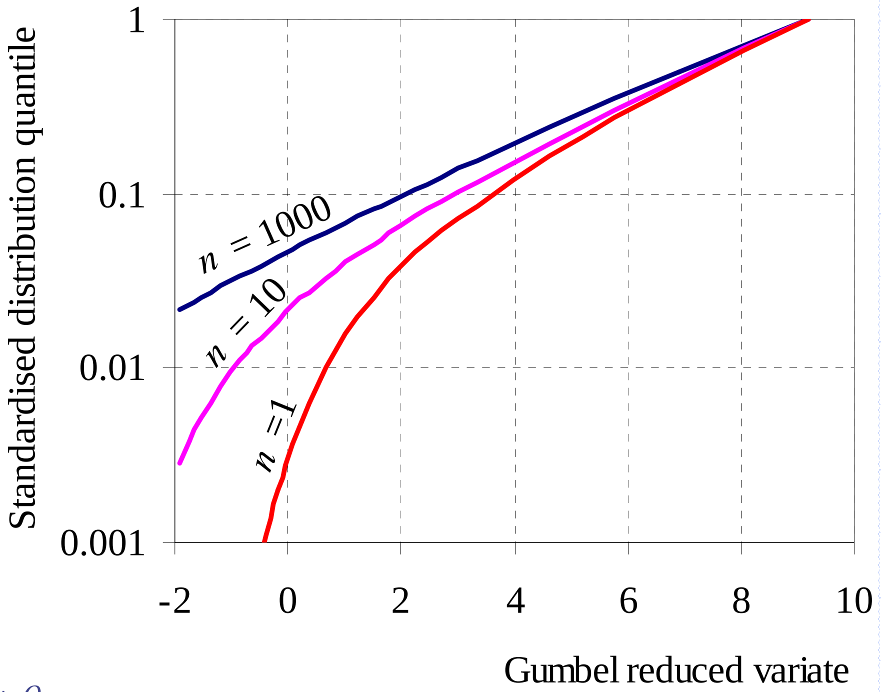
$$f(y) = [1/ \beta\, B(\theta + \tau)]\, (y/\beta)^{\theta - 1} / (1 + y/\beta)^{\tau + \theta}$$

(Beta distribution of the second kind)

Exact distribution of maxima for constant n:

$$H_n(x) = [B_{x/(x + \beta)}\, (\theta, \tau) / B\, (\theta, \tau)]^n$$

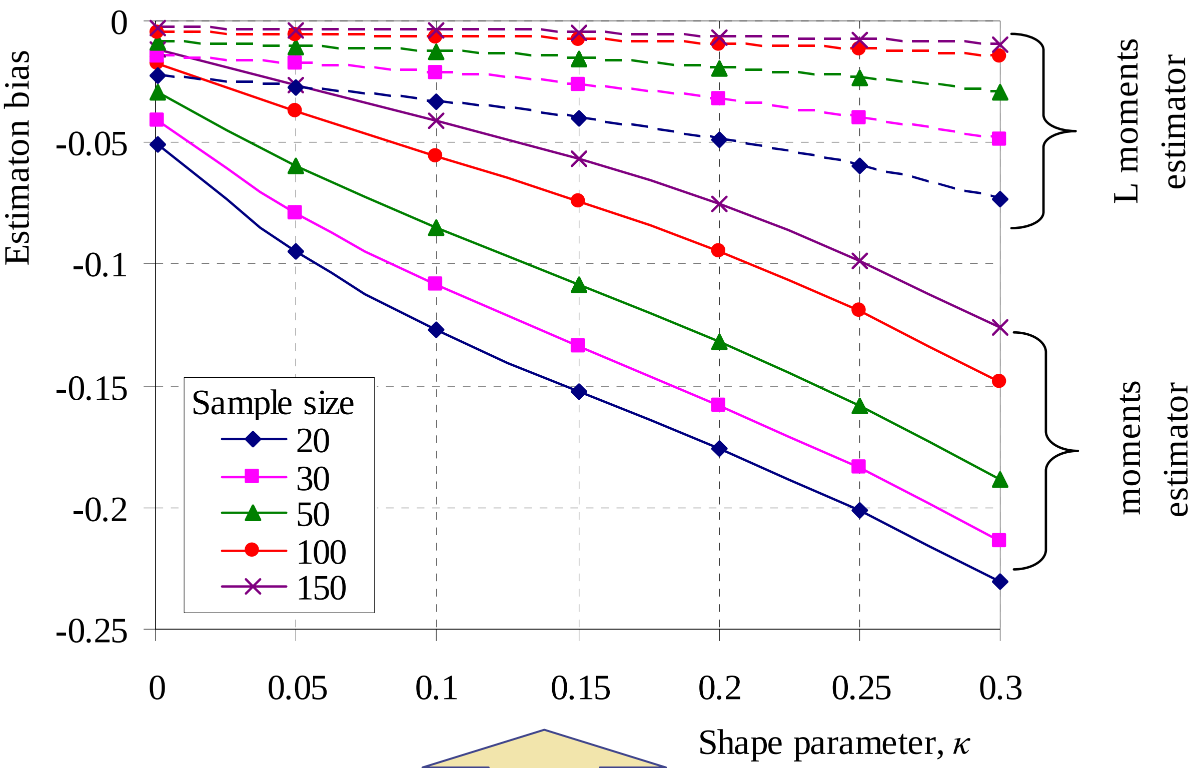
Asymptote:
EV2 (Fréchet distribution)



Convergence of distribution of maxima for gamma parent distribution with shape parameter $\theta = 0.5$ and scale parameter randomly varying following a gamma distribution with $\tau = 3$ and $\beta = 1$ (Fréchet probability plot)

Why inappropriateness of EV1 has not become evident?

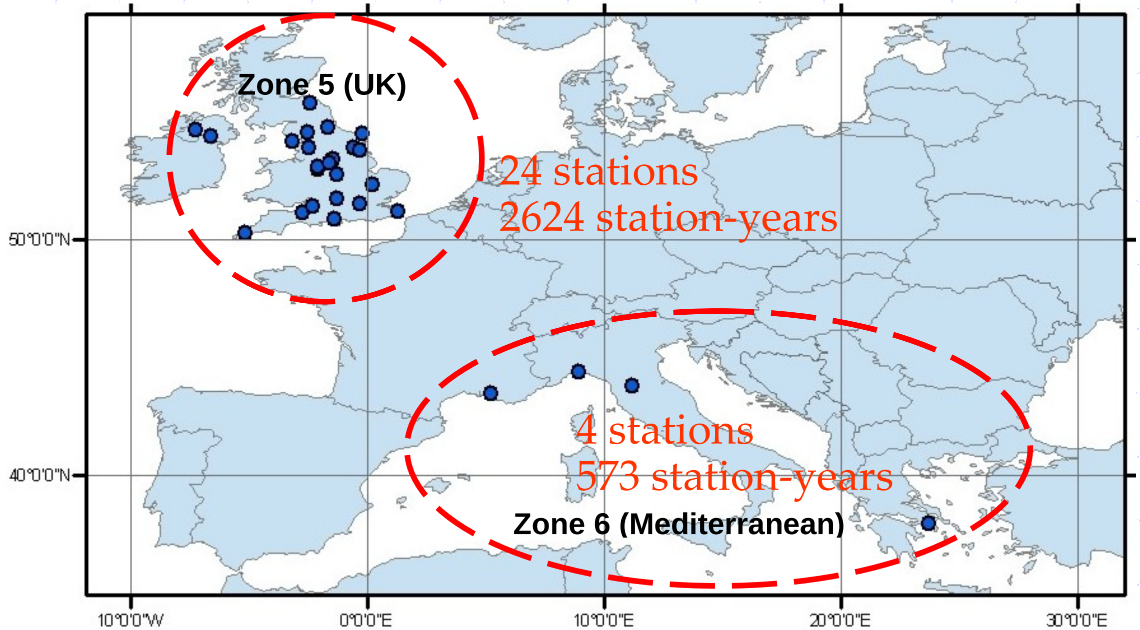
- Even if one is willing to try EV2 as a potential model, it is very likely that he/she will reject it due to significant bias of estimators
- For small samples, the most common method of moments hides completely EV1
- Even the less biased L-moments method may result in erroneous acceptance of EV1 (e.g.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

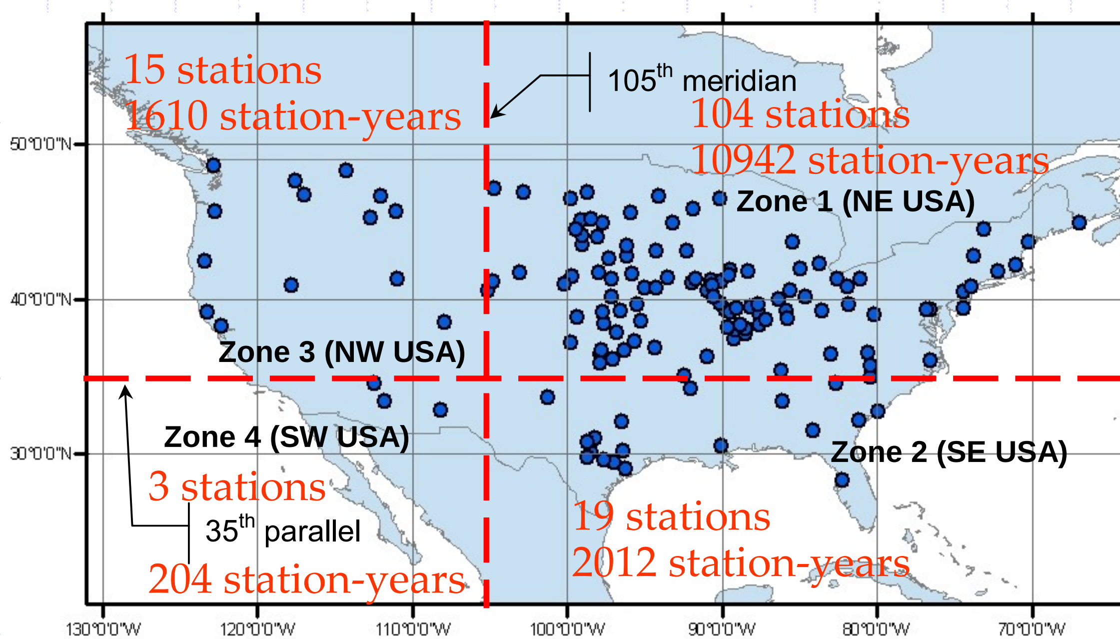


Bias in estimating the shape parameter κ of the GEV distribution
(Obtained by Monte Carlo simulation)

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

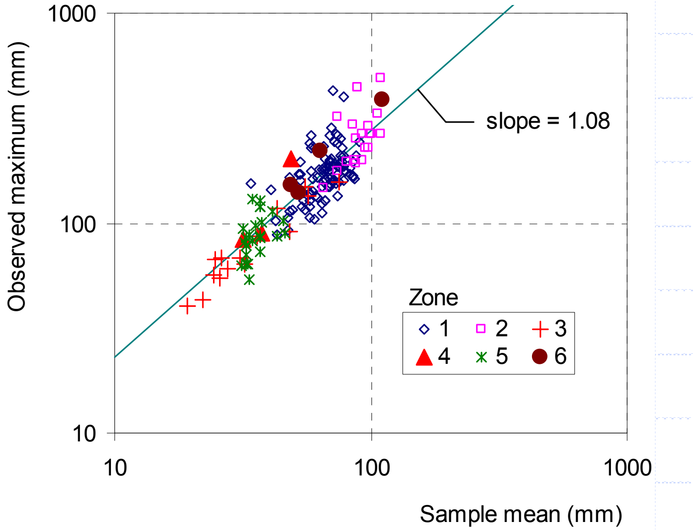


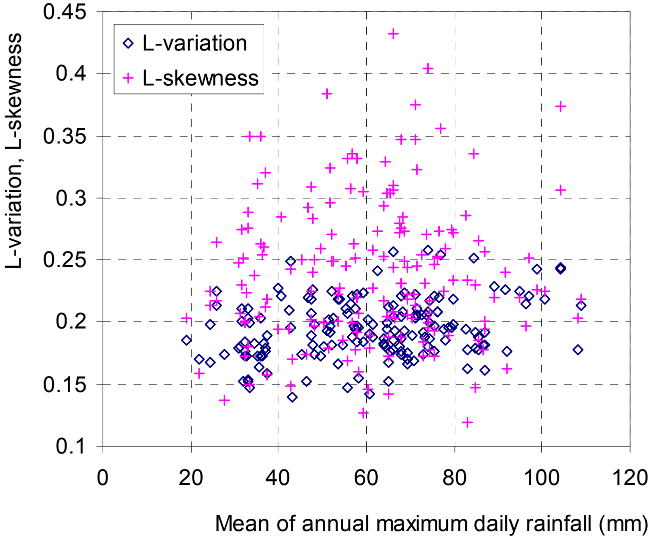


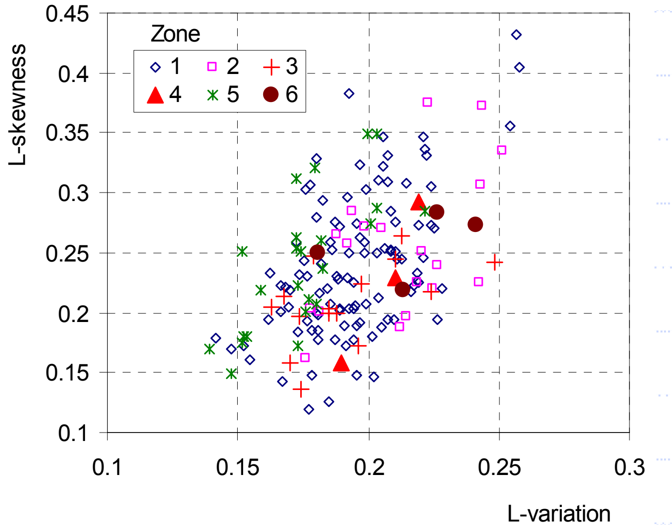
Top ten raingauges (in terms of record length)

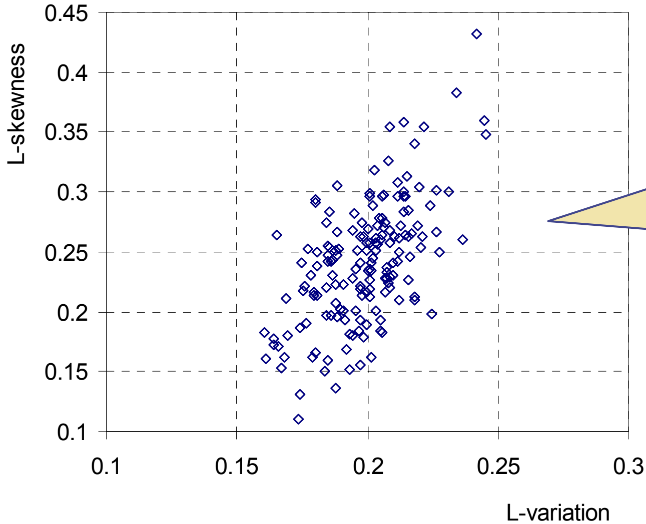
Name	Zone /Country /State	Latitude (ºN)	Longitude (º)	Elevation (m)	Record length	Start year	End year	Years with missing values
Florence	6/Italy	43.80	11.20	40	154	1822	1979	4
Genoa	6/Italy	44.40	8.90	21	148	1833	1980	
Athens	6/Greece	37.97	23.78	107	143	1860	2002	
Charleston City	2/USA/SC	32.79	–79.94	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	

Preliminary investigation of statistics of annual maximum daily rainfall



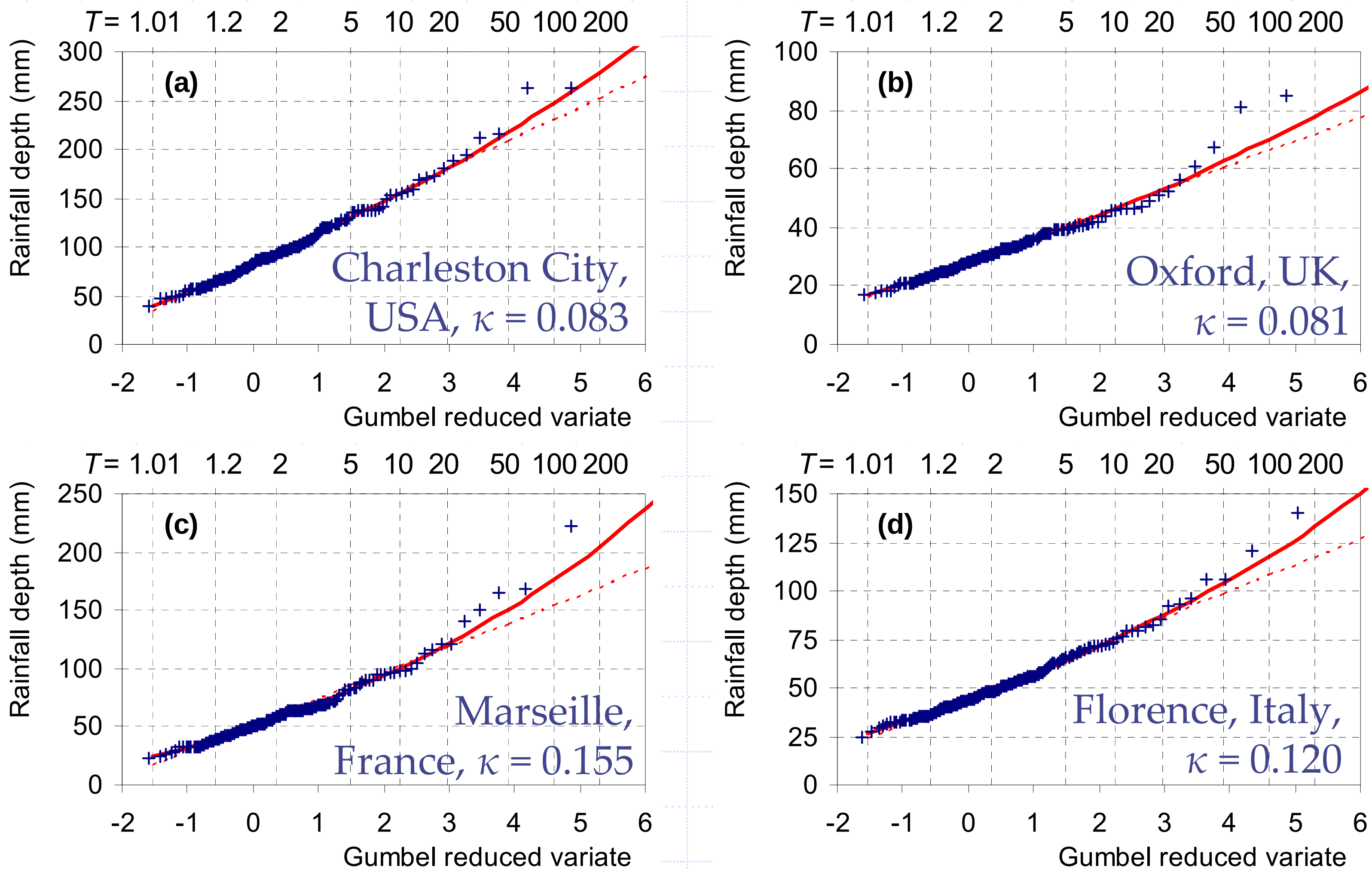






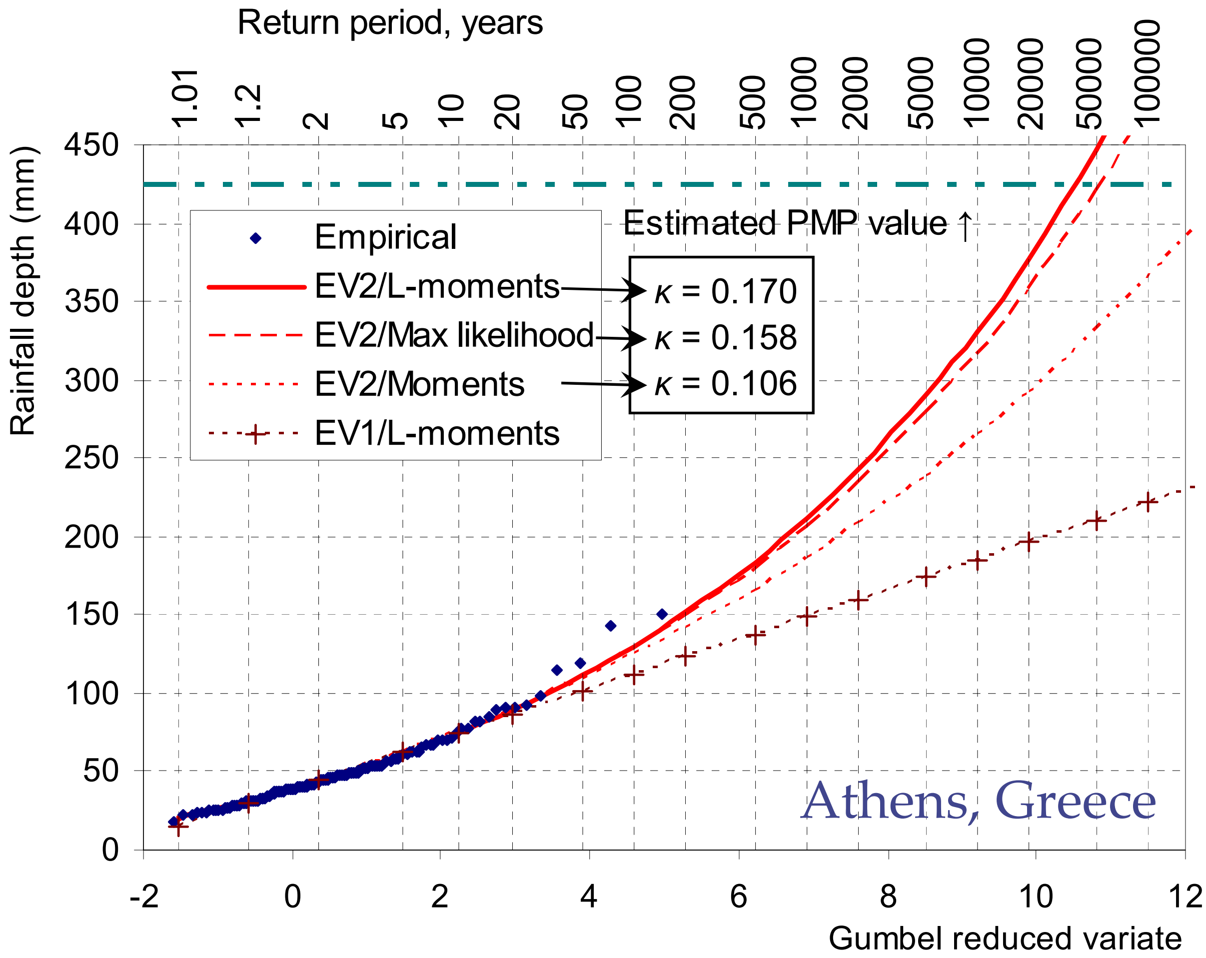
Results from 169 synthetic samples from the EV2 distribution

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



Averages over all raingauges and dispersion characteristics of the parameters of the GEV distribution

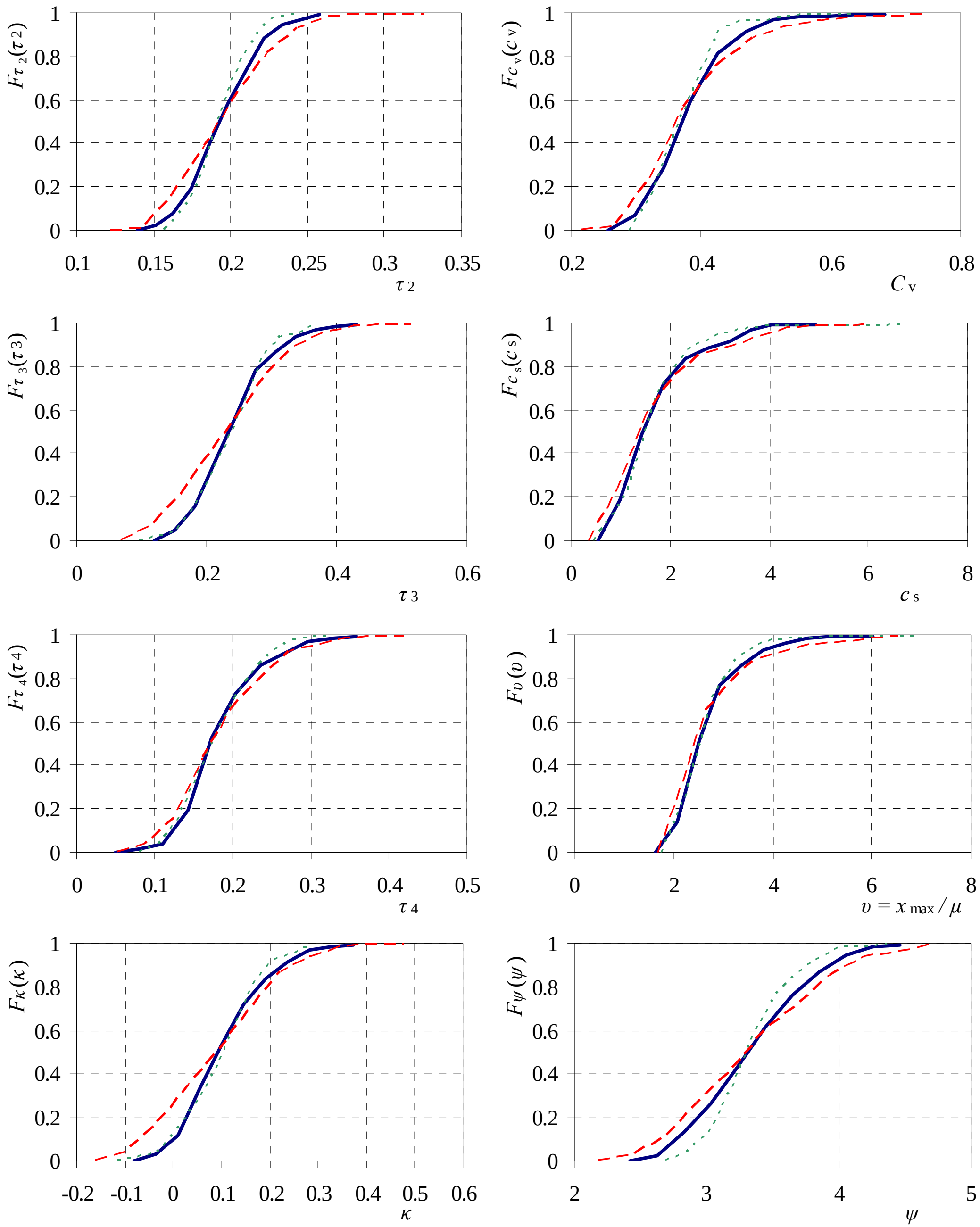
Parameter		Value
shape parameter, κ	Mean	0.103
	Standard deviation	0.085
	Min	–0.080
	Max	0.373
	Percent positive	92%
scale parameter, λ (mm^{-1})	Mean	15.52
	Standard deviation	5.81
	Min	4.86
	Max	32.13
location parameter, ψ	Mean	3.34
	Standard deviation	0.43
	Min	2.42
	Max	4.47

Estimation method: L-Moments

Empirical distributions of dimensionless sample statistics

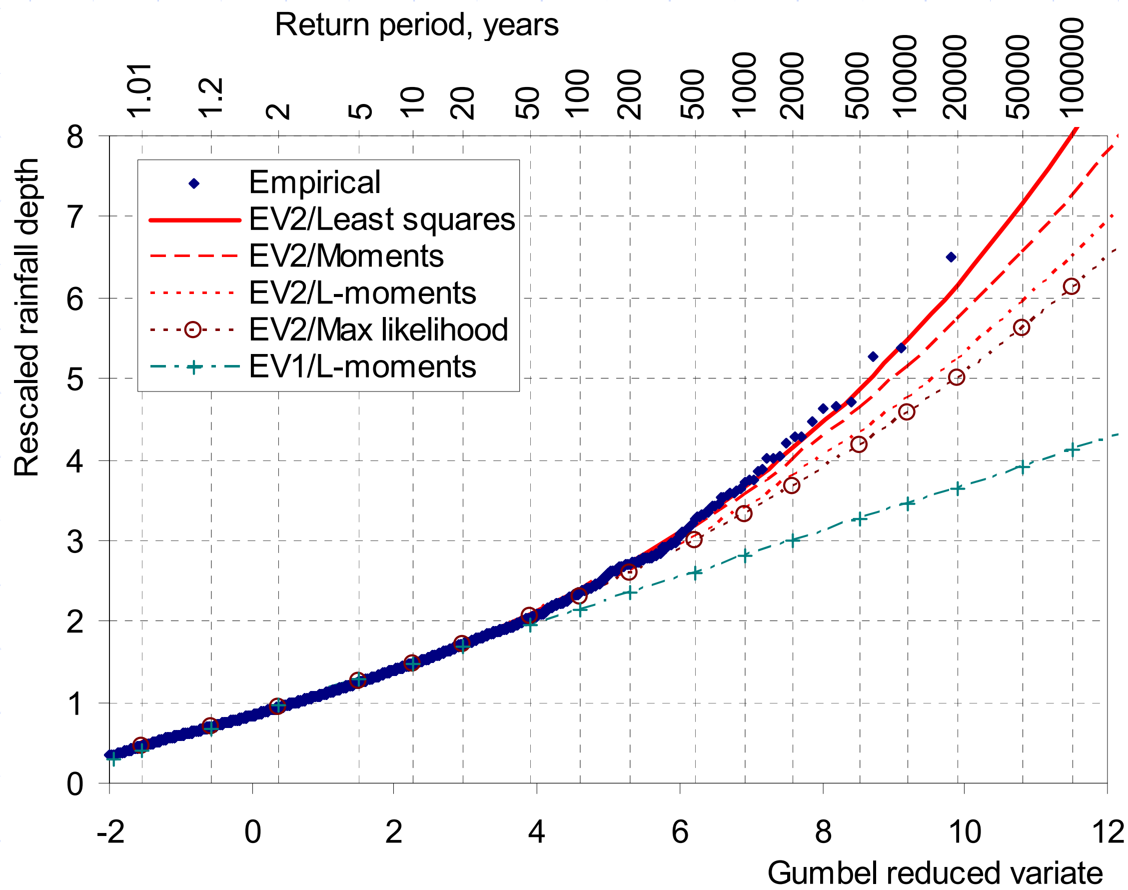
Empirical distribution functions computed from:

- the 169 historical annual maximum daily rainfall series
- 169 synthetic samples with lengths and means equal to those of historical series generated from the GEV distribution with constant $\kappa = 0.103$ and $\psi = 3.34$
- 169 synthetic samples with lengths and means equal to those of historical series generated from the GEV distribution with κ and ψ randomly varying following uniform distributions



Hypothesis of constant dimensionless parameters (shape κ and location ψ)

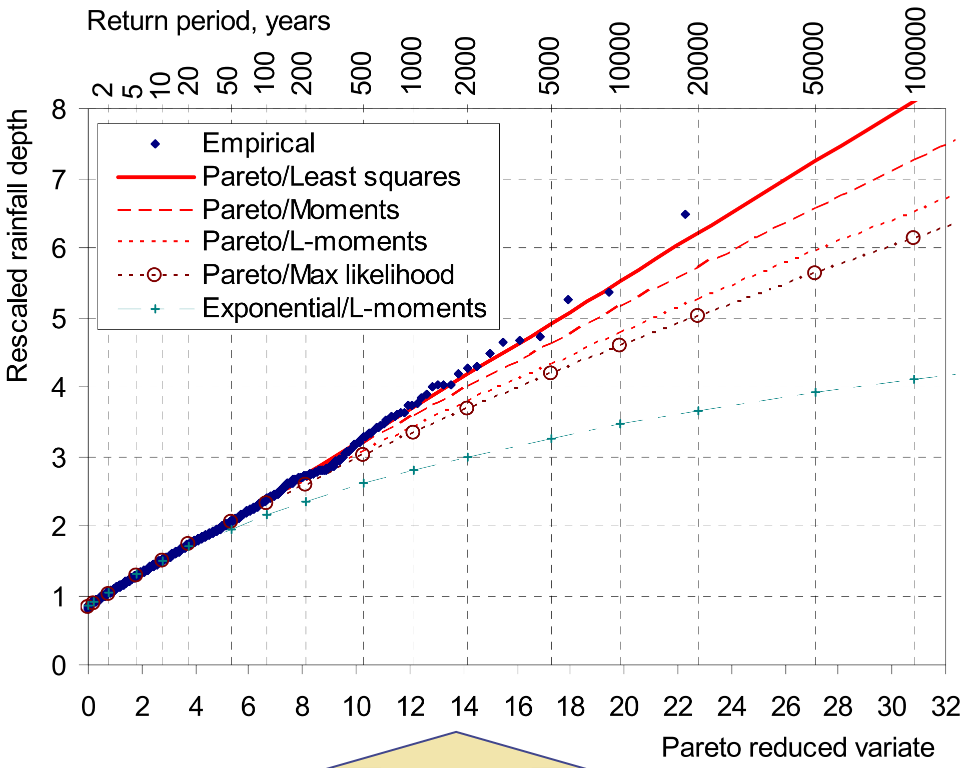
- Rescaling of each records by its mean
- Unification of all records (18065 data values)
- Accurate estimation of κ and ψ



	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

A note on series above threshold (partial duration series)

- If the maximum annual series follows EV2, i.e.,
  $H(x) = \exp\{-[1 + \kappa(x/\lambda - \psi)]^{-1/\kappa}\}$
- Then the partial duration series follows the generalised Pareto,
  $G(x) = 1 - [1 + \kappa(x/\lambda - \psi)]^{-1/\kappa}\}$
  with same parameter values
- This is absolutely validated with the data set of this study
- Thus either of the two series (annual maxima, partial duration) can be used interchangeably

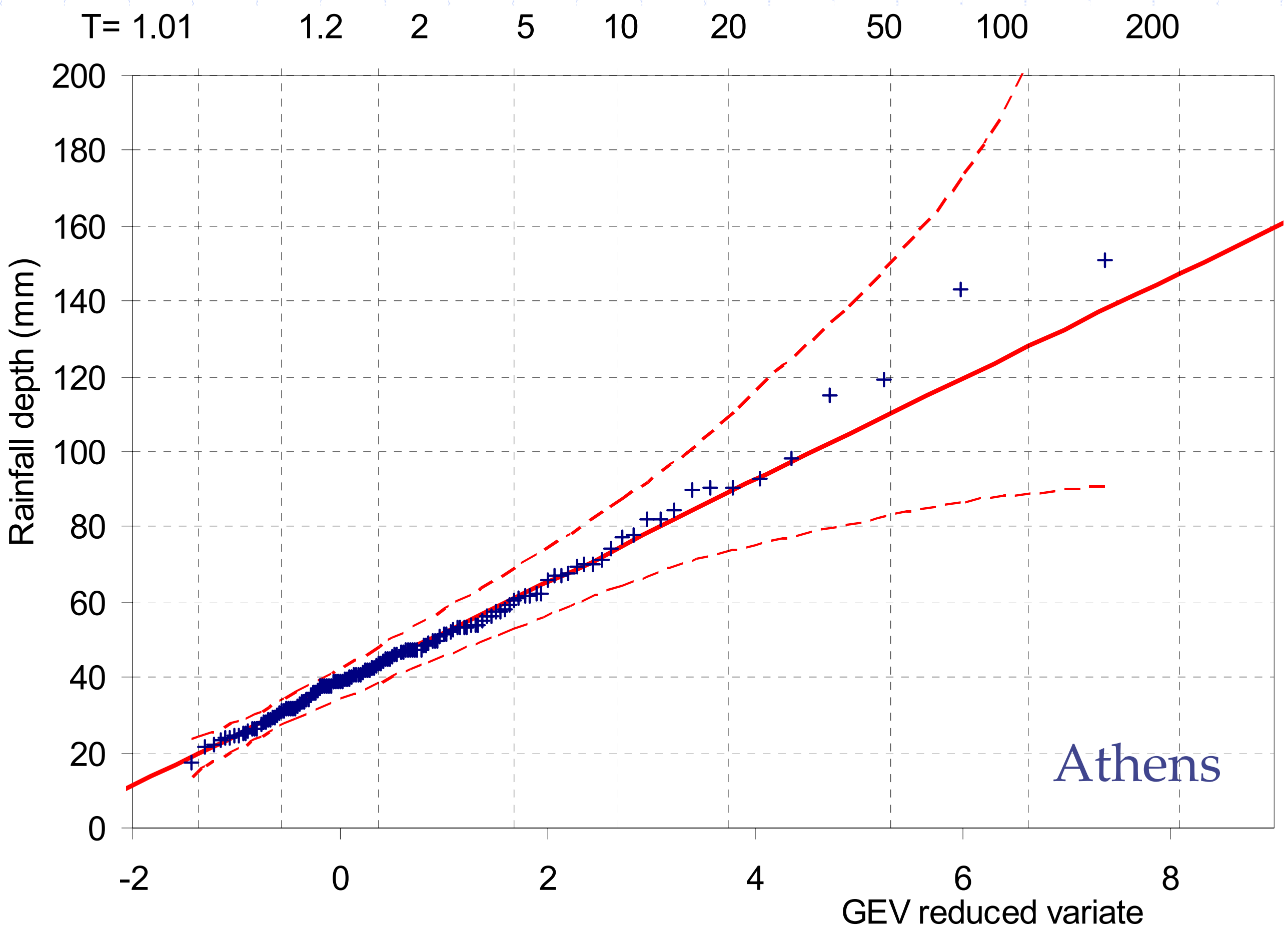


Empirical distribution of the union of the 168 series of rescaled rainfall depths over threshold (17922 station-years) in comparison with the Pareto distribution with parameters estimated from the unified series of annual maxima
(Pareto probability plot with $\kappa = 0.15$)

From EV1 to EV2: Practical considerations

	General formula	EV1	EV2, $\kappa = 0.15$	EV2, general case
Calculation of quantile	$x_H = \lambda\,(z_H + \psi)$	$z_H = -\ln(-\ln H)$	$z_H = \dfrac{[(-\ln H)^{-0.15} - 1]}{0.15}$	$z_H = \dfrac{[(-\ln H)^{-\kappa} - 1]}{\kappa}$
Construction of linear probability plot	Plot x_H against z_H			(Not possible for unknown κ)
Estimation of λ, moments method	$\lambda = c_1\,\sigma$	$c_1 = 0.78$	$c_1 = 0.61$	$c_1 = \kappa\,[(\Gamma(1 - 2\,\kappa) - \Gamma^2(1 - \kappa))]^{-0.5}$
Estimation of λ, L-moments method	$\lambda = c_2\,\lambda_2$	$c_2 = 1.443$	$c_2 = 1.23$	$c_2 = \kappa\,/\,[\Gamma(1 - \kappa)\,(2\kappa - 1)]$
Estimation of ψ	$\psi = \mu/\lambda - c_3$	$c_3 = 0.577$	$c_3 = 0.75$	$c_3 = [\Gamma(1-\kappa) - 1]/\kappa$

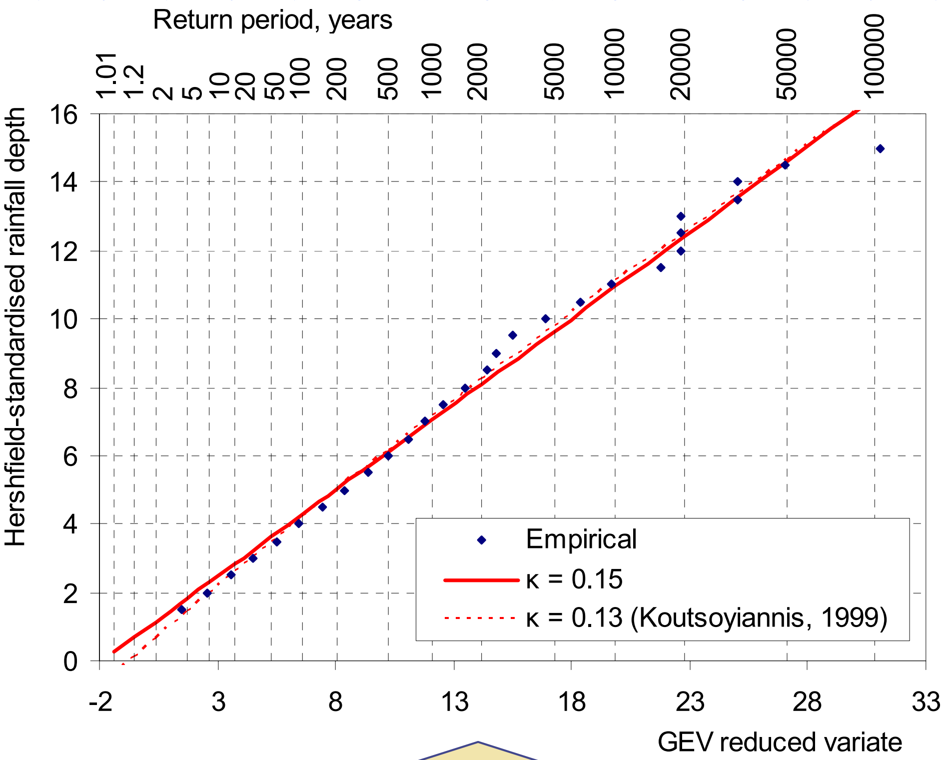
An example of GEV linear probability plot



+ Empirical distribution
EV2 distribution
95% Monte Carlo prediction limits for the empirical distribution

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 $\kappa = 0.15$

Conclusions

- The EV1 distribution should be avoided when studying hydrological extremes
- The theoretical and empirical reasons that made the EV1 distribution prevail in hydrology may be not valid
- The three-parameter EV2 distribution is a better alternative
- 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
- The location parameter ψ of EV2 is fairly constant (average $\psi = 3.54$, coefficient of variation 0.13). However, there is no need to regard it as a fixed constant as it can be estimated with relative accuracy on the basis of an individual series
- The scale parameter λ of EV2 varies with the station location. There is no need to seek a generalised law for it as it can be estimated with relative accuracy on the basis of an individual series
- In engineering practice, the handling of EV2 can be as easy as that of EV1 if the shape parameter of the former is fixed to the value $\kappa = 0.15$

More information …

- This presentation is available on line at http://www.itia.ntua.gr/e/docinfo/624/
- 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