European Geosciences Union General Assembly 2005 Vienna, Austria, 24 - 29 April 2005 Session HS26: Regional frequency analysis of hydrological extremes

SIMILARITIES AND SCALING OF EXTREME RAINFALL WORLDWIDE

D. Koutsoyiannis, Department of Water Resources, National Technical University of Athens

1. Abstract

Long records of annual maximum daily rainfall from 169 stations from Europe and the USA, with lengths exceeding 100 years, are statistically analyzed. It is observed that several dimensionless statistics of the annual maximum series are virtually constant worldwide, except for an error that can be attributed to a pure statistical sampling effect. Thus, if all series are standardized by their mean, they can be described by practically the same statistical law. From the study of the compound series from all stations with length 17922 station-years it becomes clear that this extreme value law is of type II rather than type I (Gumbel) as thought before. This implies a power type (Pareto) parent distribution, which has scaling properties for low probabilities of exceedence. Two major questions arise from this research: (1) Why the statistical law of standardized extreme rainfall is virtually the same over a wide range of geographical areas and climates? (2) Why is this law power-type? The second question is answered using the principle of maximum entropy. Specifically, it is shown that this principle, which corresponds to maximum uncertainty, results in a Pareto type distribution, if the coefficient of variation is high, and also predicts the scaling exponent, which is verified by the historical data.

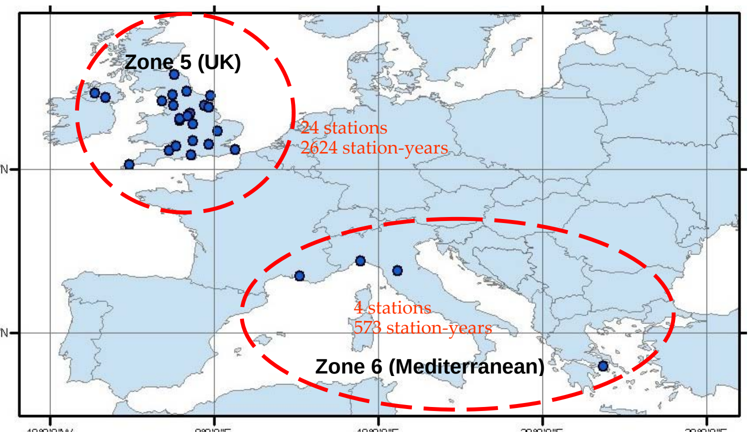
2. Data set

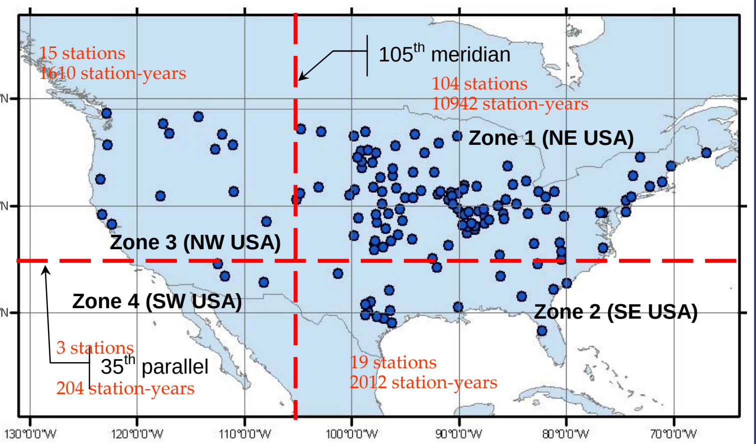
169 stations from Europe and North America

Record lengths 100-154 years

18065 station-years in total

6 major climatic zones (Note: see details in Koutsoyiannis, 2004)





3. 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	

4. Initial exploration

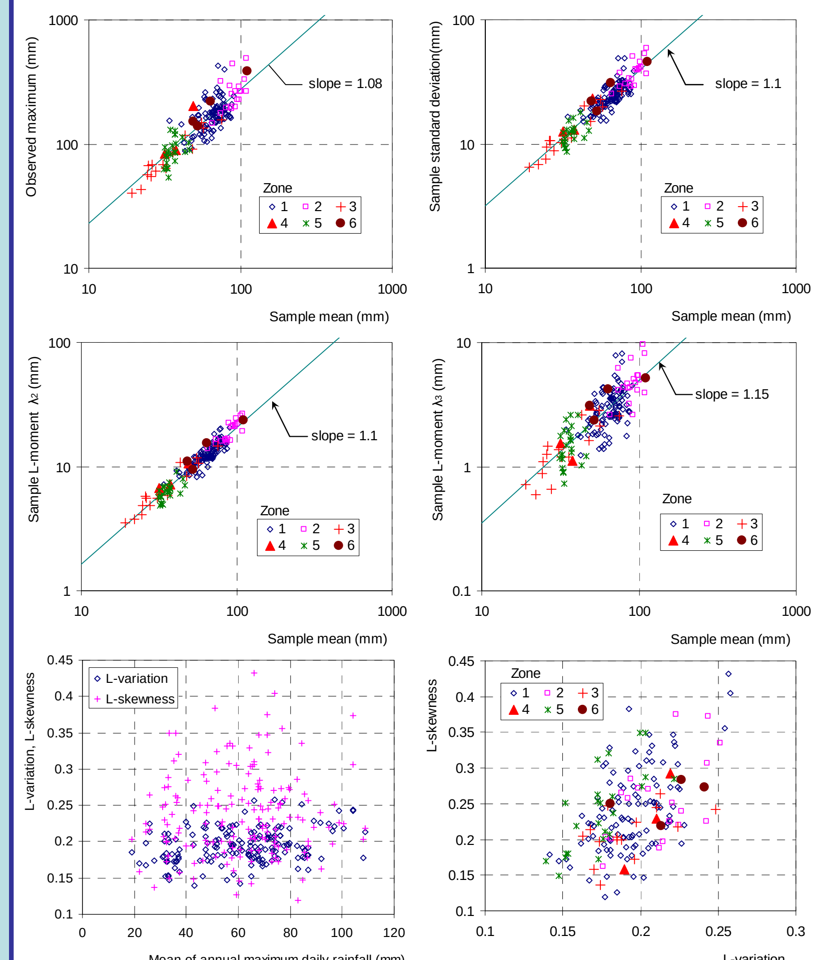
Standard deviation (σ) and L moments (λ_2 and λ_3) as well as observed maximum values per station are very well correlated with sample mean (μ).

The correlations do not change in different climate zones.

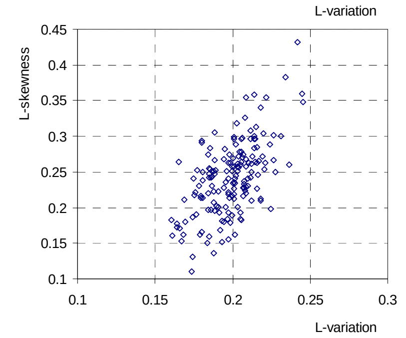
These statistics, whose dimensions are same with the sample mean (mm), are approximately proportional to sample mean.

Dimensionless statistics, such as L coefficients of variation (τ_2) and skewness (τ_3) are not correlated with sample mean.

The L coefficients of variation and skewness are correlated with each other. This correlation, however, does not contain a physical meaning and is a statistical effect.



Results from 169 synthetic samples from the EV2 distribution



5. Fitting of the Generalized Extreme Value (GEV) distribution

GEV equation

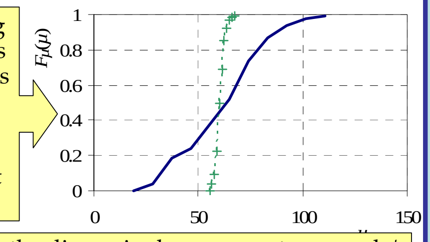
$$F(x) = \exp\left\{-\left[1 + \kappa\left(\frac{x}{\lambda} - \psi\right)\right]^{-1/\kappa}\right\}$$

Averages over all raingauges and dispersion characteristics of the parameters of the GEV distribution. Estimation method: L-Moments

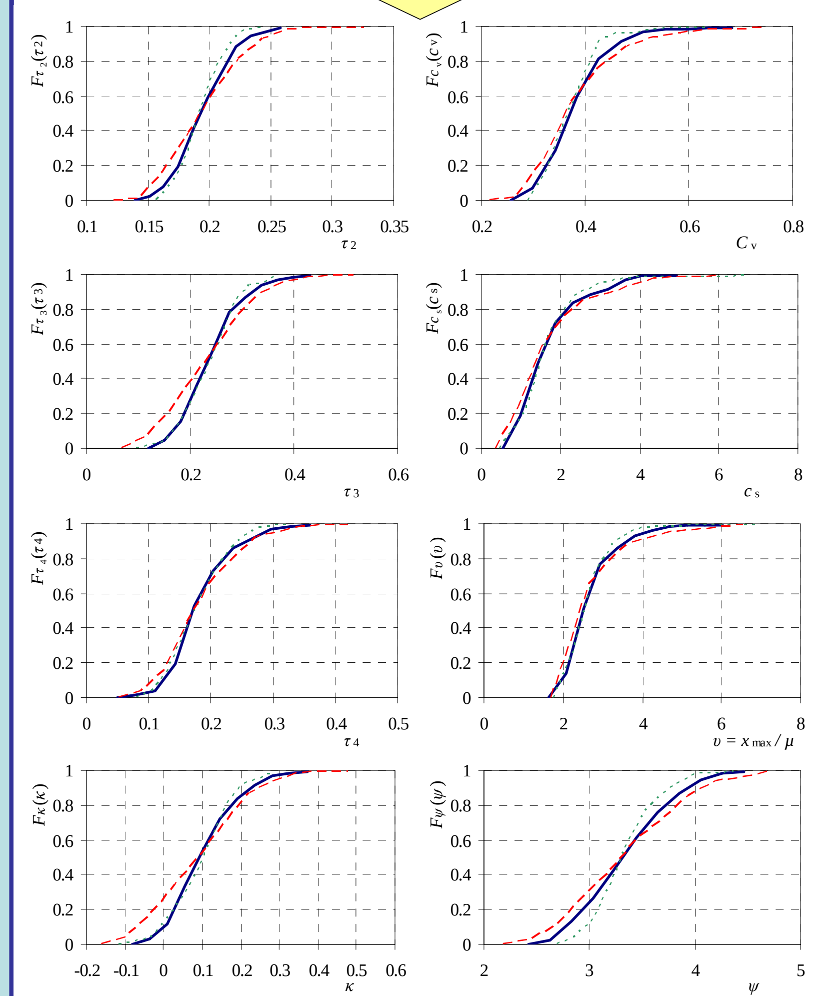
Parameter		Value
Shape parameter, κ	Mean	0.103
	Standard deviation	0.085
	Min	–0.080
	Max	0.373
	Percent positive	92%
Scale parameter, λ (mm^{-1})	Mean	15.52
	Standard deviation	5.81
	Min	4.86
	Max	32.13
Location parameter, ψ	Mean	3.34
	Standard deviation	0.43
	Min	2.42
	Max	4.47

6. Study of the variation of parameters

A Monte Carlo simulation assuming GEV distribution and all parameters constant, equal to the average values in panel 5, shows that simulated variation of the mean (crosses) is much lower than the actual one (continuous line). This suggests that variation reflects a climatic effect.



A Monte Carlo simulation assuming the dimensionless parameters κ and ψ constant, equal to the average values in panel 5, and the scale parameter varying, shows that simulated variation of several dimensionless parameters almost equals the actual one. This suggests that the variation reflects a statistical sampling effect rather than a climatic effect.



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 κ and ψ;
- 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.

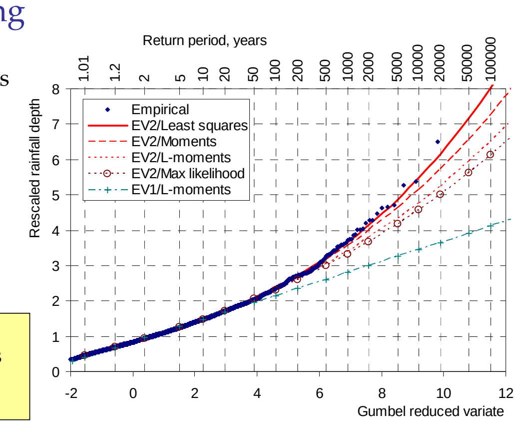
7. Final model fitting

Adoption of the GEV distribution and hypothesis of constant dimensionless parameters (κ and ψ)

Standardization of each record by its mean

Unification of all records (18065 data values)

Estimation of κ and ψ



Main conclusion: The distribution of extremes is Extreme Value Type II (EV2, not EV1).

	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

8. Additional support of Type II behaviour

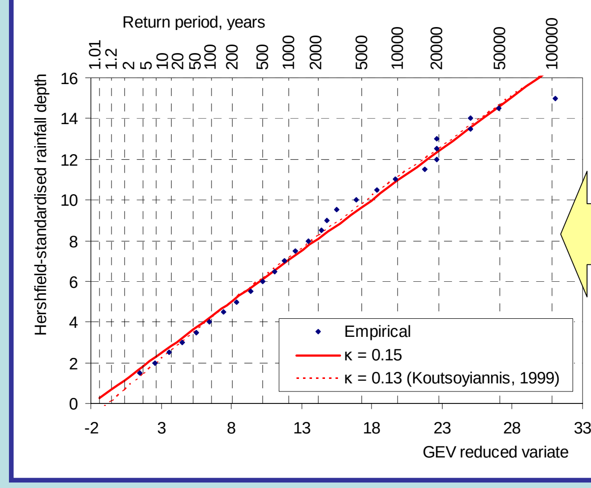
1. Chaouche (2001) exploited a data base of 200 rainfall series of various time steps (minute-month) from the five continents, each including more than 100 years of data. Using multifractal analyses he showed that:

   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.

2. Hershfield's (1961) data set, comprising 95 000 station-years, in a later study (Koutsoyiannis, 1999) was found to have very similar behaviour.



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 κ = 0.13 (Koutsoyiannis, 1999) and κ = 0.15.

9. Major theoretical questions

(1) Why the statistical law of standardized extreme rainfall is virtually the same over a wide range of geographical areas and climates?

The answer is difficult and not addressed here.

(2) Why is this law power-type?

The answer is sought upon the entropy concept and the principle of maximum entropy (ME).

10. The entropy concept

For a discrete random variable X taking values x_j with probability mass function $p_j \equiv p(x_j)$, the Boltzmann-Gibbs-Shannon (or extensive) entropy is defined as

$$\phi := E[-\ln p(X)] = -\sum_{j=1}^{w} p_j \ln p_j, \quad \text{where} \quad \sum_{j=1}^{w} p_j = 1$$

For a continuous random variable X with probability density function $f(x)$, the entropy is defined as

$$\phi := E[-\ln f(X)] = -\int_{-\infty}^{\infty} f(x) \ln f(x)\, dx, \quad \text{where} \quad \int_{-\infty}^{\infty} f(x)\, dx = 1$$

In both cases the entropy ϕ is a measure of **uncertainty** about X and equals the information gained when X is observed (Papoulis, 1991).

In other disciplines (statistical mechanics, thermodynamics, dynamical systems, fluid mechanics), entropy is regarded as a measure of **order** or **disorder** and **complexity**.

11. Generalization of entropy

Tsallis (1988) heuristically generalized the Boltzmann-Gibbs-Shannon entropy by postulating the entropic form

$$\phi_q := \frac{1 - \sum_{i=0}^{w} p_i^q}{q-1} \text{ (for discrete } x\text{)}, \qquad \phi_q := \frac{1 - \int_0^{\infty} [f(x)]^q\, dx}{q-1} \text{ (for continuous } x\text{)}$$

where q is any real number. This has been called Tsallis entropy or nonextensive entropy and remedies disabilities or inconsistencies in the use of the classical entropy. For $q \to 1$ this precisely reproduces the Boltzmann-Gibbs-Shannon entropy, i.e., $\phi_1 \equiv \phi$.

12. The principle of maximum entropy (ME)

In a probabilistic context, the ME principle was introduced by Janes (1957) as a generalization of the "principle of insufficient reason" (PIR) attributed to Bernoulli or to Laplace.

The ME principle is used to infer unknown probabilities from known information.

It is related to the homonymous physical principle that determines thermodynamical states.

It postulates that the entropy of a random variable should be at maximum, under some conditions, formulated as constraints, which incorporate the information that is given about this variable.

13. Application of the ME principle to hydrological variables with high variation

We use the following four constraints, which involve two parameters, the mean μ and the standard deviation σ, estimated from the sample.

Mass, Mean/Momentum

$$\int_{-\infty}^{\infty} f(x)\, dx = 1, \quad E[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx = \mu$$

Variance/Energy, Non-negativity

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)\, dx = \sigma^2 + \mu^2, \quad x \ge 0$$

The non-negativity constraint is essential for hydrological variables.

Maximization of the Boltzmann-Gibbs-Shannon entropy with these constraints results in the truncated (for $x \ge 0$) normal distribution. This tends to the normal distribution as $\sigma/\mu \to 0$ and to the exponential distribution as $\sigma/\mu \to 1$. For $\sigma/\mu > 1$ the Boltzmann-Gibbs-Shannon ME distribution does not exist. In this case the Tsallis entropy can be used which results in:

$$f(x) = [1 + \kappa\,(\lambda_0 + \lambda_1 x + \lambda_2 x^2)]^{-1-1/\kappa}, \quad \phi_q = (\kappa + 1)\,(\lambda_0 + \lambda_1 \mu_1 + \lambda_2 \mu_2)$$

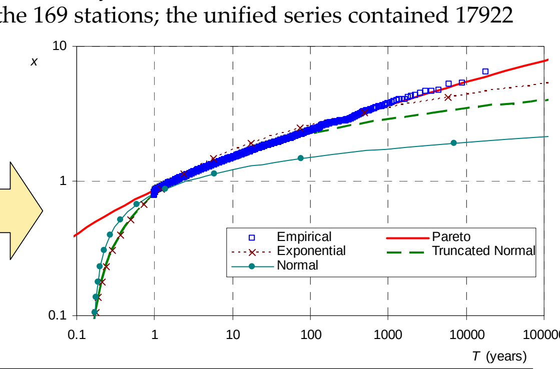
where $\kappa := (1 - q)/q$ and λ_i Lagrange multipliers. In the absence of any information for κ or q, we can set $\lambda_2 = 0$ (Koutsoyiannis, 2005) and obtain the Pareto distribution, which for large x exhibits scaling properties:

$$f(x) = (1/\lambda)\,(1 + \kappa x / \lambda)^{-1-1/\kappa} \quad F^*(x) = (1 + \kappa x / \lambda)^{-1/\kappa}$$

14. Application to extreme daily rainfall

Data set: the same as previously. Series above threshold (rather than series of annual maxima), standardized by mean, were formed and used (this was possible for the 168 of the 169 stations; the unified series contained 17922 station-years).

μ = 0.28 (mean minus threshold)
σ/μ = 1.19 > 1
ME distribution: Pareto
κ = 0.15 (as in earlier analysis)
ϕ_q = 1.16



Conclusion: Parent distribution: Pareto ⟶ Distribution of maxima: EV2

15. Conclusions

1. Long records of daily rainfall from 169 stations worldwide, indicate impressive similarities in extreme rainfall over all climates.
2. Specifically, several dimensionless statistics are virtually constant worldwide, except for an error attributed to a statistical sampling effect.
3. Thus, if all series are standardized by their mean, they can be described by virtually the same statistical law.
4. Clearly, this law is not exponential/EV1 (Gumbel) as thought before, but Pareto/EV2 and has scaling properties for low probabilities of exceedence.
5. The emergence of the Pareto law is explained by the principle of maximum entropy, which corresponds to maximum uncertainty.

References

Chaouche K. (2001), Approche Multifractale de la Modélisation Stochastique en Hydrologie, thèse, École Nationale du Génie Rural, des Eaux et des Forêts, Centre de Paris.
Hershfield, D. M. (1961), Estimating the probable maximum precipitation, *Proc. ASCE, J. Hydraul. Div.*, 87(HY5), 99-106.
Jaynes, E.T. (1957), Information theory and statistical mechanics, *Physical Review*, 106(4), 620-630.
Koutsoyiannis, D. (1999), A probabilistic view of Hershfield's method for estimating probable maximum precipitation, *Wat. Resour. Res.*, 35(4), 1313-1322.
Koutsoyiannis, D. (2004), Statistics of extremes and estimation of extreme rainfall, 2, Empirical investigation of long rainfall records, *Hydrol. Sci. J.*, 49(4), 591-610.
Koutsoyiannis, D. (2005), Uncertainty, entropy, scaling and hydrological stochastics, 1, Marginal distributional properties of hydrological processes and state scaling, *Hydrol. Sci. J.*, (in press).
Tsallis, C. (1988), Possible generalization of Boltzmann-Gibbs statistics, *J. Stat. Phys.* 52, 479-487.

Poster available on line at http://www.itia.ntua.gr/e/docinfo/653/