Uncertainty, entropy, scaling and hydrological stochastics

1. Marginal distributional properties of hydrological processes and state scaling

Demetris Koutsoyiannis

Department of Water Resources, Faculty of Civil Engineering, National Technical University of Athens, Heroon Polytechneiou 5, GR-157 80 Zographou, Greece (dk@itia.ntua.gr)

Abstract The well-established physical and mathematical principle of maximum entropy (ME), is used to explain the distributional and autocorrelation properties of hydrological processes, including the scaling behaviour both in state and in time. In this context, maximum entropy is interpreted as maximum uncertainty. The conditions used for the maximization of entropy are as simple as possible, i.e. that hydrological processes are non-negative with specified coefficients of variation (CV) and lag one autocorrelation. In this first part of the study, the marginal distributional properties of hydrological variables and the state scaling behaviour are investigated. Application of the ME principle under these very simple conditions results in the truncated normal distribution for small values of CV and in a nonexponential type (Pareto) distribution for high values of CV. In addition, the normal and the exponential distributions appear as limiting cases of these two distributions. Testing of these theoretical results with numerous hydrological data sets on several scales validates the applicability of the ME principle, thus emphasizing the dominance of uncertainty in hydrological processes. Both theoretical and empirical results show that the state scaling is only an approximation for the high return periods, which is merely valid when processes have high variation on small time scales. In other cases the normal distributional behaviour, which does not have state scaling properties, is a more appropriate approximation. Interestingly however, as discussed in the second part of the study, the normal distribution combined with positive autocorrelation of a process, results in time scaling behaviour due to the ME principle.

Keywords entropy; hydrological design; hydrological extremes; hydrological statistics; power laws; risk; scaling; uncertainty.

Incertitude, entropie, graduation et propriétés stochastiques hydrologiques

1. Propriétés distributionnelles marginales et d'échelle d'état des processus hydrologiques

Résumé Le principe bien établi à la fois physique et mathématique de l'entropie maximum (ME), est employé pour expliquer les propriétés distributionnelles et d'autocorrélation des processus hydrologiques, y compris le comportement d'échelle dans l'état et dans le temps. Dans ce contexte, l'entropie maximum est interprétée en tant qu'incertitude maximum. Les conditions utilisées pour la maximisation de l'entropie sont le plus simples possible, c.-à-d. que les processus hydrologiques sont non négatifs avec coefficients de variation (*CV*) et d'autocorrélation fixés. Dans la présente première partie de l'étude, les propriétés distributionnelles marginales des variables hydrologiques et le comportement d'échelle d'état sont étudiés. L'application du principe de ME dans ces conditions très simples aboutit à une distribution normale tronquée pour de petites valeurs de *CV* et à une distribution non exponentielle (du type Pareto) pour les grandes valeurs de *CV*. En outre, les distributions normales et exponentielles apparaissent comme cas limite de ces deux distributions. L'essai de ces résultats théoriques sur de nombreux échantillons hydrologiques sur plusieurs échelles valide l'applicabilité du principe de ME, de ce fait soulignant la dominance de l'incertitude dans des processus hydrologiques. Les résultats théoriques et empiriques prouvent que la graduation d'état est seulement une approximation pour les périodes de retour élevées, qui est simplement valide quand les processus ont une variabilité élevée sur de petites échelles de temps. Dans d'autres cas, le comportement distributionnel normal, qui n'a pas des propriétés d'échelle d'état, est une approximation plus appropriée. Intéressant cependant, comme discuté dans la deuxième partie de l'étude, la distribution normale combinée avec l'autocorrelation positif d'un processus, a comme conséquence le comportement d'échelle de temps pour cause au principe ME.

Mots clefs entropie; conception hydrologique; extrêmes hydrologiques; statistiques hydrologiques; lois de puissance; risque; mise en échelle; incertitude.

First draft, September 2004, Revision January 2005

INTRODUCTION

Ἕν οἶδα, ὅτι οὐδέν οἶδα.

Σωκράτης

(One I know, that I know nothing.

Socrates)

Major questions of hydrological statistics

Hydrological statistics uses a rich repertoire of statistical distributions (e.g. Kite, 1988), such as normal, lognormal, Pearson, log-Peason, Weibull, Extreme Value of maxima and minima of types I, II, and II (EV1, EV2, EV3) and fits them to several hydrological quantities such as rainfall and runoff. For each case examined, the problem is to choose the most appropriate theoretical distribution from the repertoire, based on comparisons with empirical distributions. Generally, hydrological statistics asks questions of the type "which?" and avoids questions of the type "why?". However, the latter would provide an explanation of the appropriateness or inappropriateness of a certain distribution and thus would also help choose the most appropriate distribution. In few cases, explanations exist. Thus, in a region where the number of rain days in a year is high, the distribution of the annual rainfall depth tends to follow the normal distribution and this is explained by the Central Limit Theorem. But, when the time scale is finer than annual, e.g. monthly, daily, or hourly, the Central Limit Theorem does not help. As the time scales become finer and finer, the distribution of rainfall becomes more and more asymmetric and J-shaped. This is the case for streamflow, as well, but not for every process of interest. For example, temperature keeps the symmetric bell-shaped pattern regardless of the scale. These observations imply a lot of "why's" which have not been answered and have been studied rarely (e.g. Eagleson, 1972).

The situation is similar in applications of stochastic processes to hydrology, where again the typical problem is to find "which" model (e.g. from the ARMA(p, q) family) is the most appropriate. Here the term "hydrological stochastics" is used to describe both typical hydrological statistics as well as application of stochastic processes to hydrology. (The term "stochastics" has been used recently by mathematicians for a joint description of probability theory, statistics and stochastic processes, e.g. Barndorff-Nielsen et al., 2001).

The importance of "why's" becomes more evident when studying extreme events. Obviously, observations of extreme events cannot be as numerous as those of regular events. Therefore, a theoretical reasoning of the appropriateness of a statistical distribution, in addition to the empirical study of the data, would help for a more justified and correct choice. In the study of extreme events theoretical reasoning results in three possible asymptotic behaviours producing the EV1, EV2 and EV3 distributions. These correspond to a behaviour in the tail of the parent distribution that is respectively exponential type, power (Pareto) type bounded from below and power type bounded from above. The questions, "which of them" and "why", are very important in this case because the difference of the three distributions is very substantial at large return periods and simultaneously invisible at small return periods corresponding to the time span of typical hydrological record lengths. For a long time, EV1 was regarded as the most appropriate. More recent research provides evidence in favour of EV2 (Koutsoyiannis, 2004a, b), whose importance in hydrology was initially underlined by Bernier (1964). However, these questions have not been answered in a theoretically sound manner so far.

The forms of scaling

The tendency to shift from the exponential type tails of parent distributions to the power type goes with the flow in other disciplines. It is known that power type distributions have been observed in geophysics, in economics and even in humanities. The power law behaviour of a distribution can be expressed formally by either of the following power type equations

$$x = (\lambda/\kappa)\,(T/\delta)^{\kappa}, \quad f(x) = \frac{1}{\lambda}\left(\frac{\lambda}{\kappa\,x}\right)^{1 + 1/\kappa}, \quad F^*(x) = \left(\frac{\lambda}{\kappa\,x}\right)^{1/\kappa}, \quad F^*(l\,x) = l^{-1/\kappa}\,F^*(x) \qquad (1)$$

where X denotes a random variable that represents the quantity of interest; x a value of this variable ($x > \lambda/\kappa$); T the return period of value x; δ is the mean interarrival time of an event that is represented by the variable X (e.g. if X represents annual values, $\delta = 1$ year); $f(x)$ the probability density function of X; $F^*(x)$ the survival function of X; λ and κ positive distributional parameters (scale and shape parameters respectively); and l any positive number. The quantities T, $F^*(x)$, $f(x)$ and the probability distribution function $F(x)$ of X are related to each other by

$$F^*(x) := P\{X > x\} =: 1 - F(x), \quad T := \frac{\delta}{F^*(x)}, \quad f(x) := \frac{dF(x)}{\partial x} = -\frac{dF^*(x)}{dx} \tag{2}$$

where $P\{\ \}$ stands for the probability of an event.

All four equations of set (1) are equivalent to each other and express a single distribution function. This function, (sometimes termed a fractal distribution), is a special case of the generalized Pareto distribution. The last of the set of equations (1) emphasizes the scaling property of the distribution, i.e. the fact that the mathematical expression describing the survival function does not change with a scaling of the amount x by a real number l except for a multiplicative factor which is a power law of the scaling factor l. To describe this property the term *scaling in state*, or the simpler *state scaling*, will be used. This is to distinguish from another type of scaling, which is essentially different, the *time scaling*. The latter behaviour, which is nothing different from a mathematical description (again involving several power laws) of the well-known Hurst phenomenon, will be studied in the second part of the study.

The wide presence of power laws and scaling, both in state and in time, in observed time series of almost every discipline have amazed many scientists and led them to write papers and books related to them. From the large family of similar books, three are mentioned here, which have the eloquent titles "Ubiquity" (Buchanan, 2000), "Fractals, Chaos, Power Laws" (Schroeder, 1991), and "How Nature Works" (Bak, 1996). In the last one, an explanation of this behaviour is given based on the principle of "self organized criticality".

The enrolment of the entropy concept

One may parallel the wide presence of scaling and power laws to the alike of the normal distribution. The fact that this distribution was found to describe several variables in nature and even in the society may have initially seemed a mysterious amazing puzzle. However, the Central Limit Theorem, whose formulation and proof passed several phases (from the early 19^{th} century with Laplace to the mid-20^{th} century with Lindeberg, Feller and Lévy), removed the mystery, as it explains the emerging of this distribution and describes the conditions under which this distribution is anticipated to emerge.

Here it should be noted that the normal distribution is contrary to the power-law distribution in (1). Thus, there must be different conditions that favour the emerging of the one or the other distribution. Can these conditions be unified in a single measure which can determine whether the normal, the power type, or another type of distribution will emerge in a specific case? This is one of the questions considered in this paper; the answer to this question replies also some of the "why" type questions mentioned earlier. And this unification is sought upon the principle of maximum entropy (ME), which is a well established principle in physics and mathematics. It is known (e.g. Papoulis, 1991, p. 573; see also next section) that the ME principle, under certain conditions, produces the normal distribution independently of the Central Limit Theorem. Thus, it may be worth searching whether the same principle can produce the state scaling behaviour or another behaviour, under different conditions. It may be also worth searching whether the same principle can produce the time scaling behaviour thus giving an explanation of the Hurst phenomenon.

These issues are investigated in this study which is separated into two parts. The first deals with marginal distributional properties and the state scaling and the second with joint distributional properties and the time scaling. The notion of entropy, which is used throughout both papers, has many meanings and interpretations in physics and mathematics. The standard interpretation in the theory of stochastic processes (also used here) is that entropy is a measure of uncertainty or ignorance (e.g. Papoulis, 1991). In this respect, maximization of entropy is equivalent to maximization of uncertainty; thus, if the application of ME principle proves to be successful in a certain natural phenomenon, i.e. proves to give results complying with observation, then this can be interpreted as the dominance of uncertainty in this natural phenomenon. In other disciplines, entropy may be regarded as a measure of complexity, order or disorder, information and discrimination (e.g. Georgii, 2003).

The entropy concept has already been used in hydrology and water resources in a lot of cases, which are comprehensively summarized by Singh (1997). Typical applications in hydrology include derivation of distributions (for example, Singh and Fiorentino (1992) contains 15 distributions derived by the ME principle, including most distributions frequently used in hydrology), parameter estimation, flow forecasting, characterization of basin morphology, design of hydrological networks, and parameter estimation of aquifers (Singh, 1997). In this study, it is attempted to use the entropy concept in a somewhat different context. For example, to refer to the already mentioned Pareto distribution, Sing and Guo (1995a, b) found which specific constraints give rise to this distribution and, subsequently, what the estimates of the parameters of the distribution would be. In contrast, here it is endeavoured to study the question whether this distribution (and the implied scaling behaviour) or another one (e.g. the normal distribution) can arise from unified conditions, which are as simple as possible. In this respect, the scope of the present study is exploratory and explanatory. Thus, the tools suitable to this context are graphical depictions and comparisons emphasizing the general 'shapes' of distributional and dependence behaviours. On the other hand, no emphasis has been given to precise calculations and computational estimation methods, for which the interested reader is referenced to the above mentioned studies and those cited within them.

The simple unified conditions are for the case of marginal distributional properties the hypotheses that a hydrological variable is non-negative and possesses a certain variability, expressed by the coefficient of variation (CV). For the case of time scaling, simple additional hypotheses are used as described in the second part of the study.

THE ENTROPY CONCEPT

The origin of entropy in thermodynamics

The concept of entropy (in Greek εντροπία, etymologized from τροπή, i.e. change, turn, drift, as in *trop*osphere) originated in the middle of the 19th century in the work of Clausius, and was fundamental to formulate the second law of thermodynamics. In the late 19th and early 20th century, Boltzmann, then complemented by Gibbs, gave it a statistical mechanical content, showing that entropy of a macroscopical stationary state is proportional to the logarithm of the number w of possible microscopical states that correspond to this macroscopical state. Also, Planck recorded the explicit relationship between entropy and probability. In the middle of the 20th century, Shannon generalized the mathematical form of entropy and also explored it further. At the same time, Kolmogorov founded the concept on more mathematical grounds on the basis of the measure theory (see e.g. Papoulis, 1991, p. 535; Müller, 2003a; Keane, 2003; Tsallis, 2004.)

The entropy in discrete state variables

For a discrete random variable X taking the values x_j ($j = 1,\ldots, w$) with probabilities $p_j \equiv p(x_j)$ such that

$$\sum_{j=1}^{w} p_j = 1 \tag{3}$$

the entropy, or more precisely the Shannon or extensive entropy, is by definition (e.g. Papoulis, 1991, p. 558) the quantity

$$\varphi := E[-\ln p(X)] = -\sum_{j=1}^{w} p_j \ln p_j \tag{4}$$

where $E[\ ]$ denotes expected value. In physics texts, entropy is usually denoted by the letter S whereas in statistics texts the symbol H is commonly used to avoid confusion with standard deviation. Here the symbol H is reserved for the Hurst coefficient, so the Greek letter φ was preferred, which has been used aforetime for entropy.

It is easily seen that the ME principle, i.e. the maximization of (4) under constraint (3), results in equal probabilities $p_j = 1/w$. For instance, the ME principle yields equal probabilities of 1/6 for each outcome in a die, a result that is also produced by Jakob Bernoulli's "principle of insufficient reason". Conceptually, the two principles are equivalent (Jaynes, 1957; Papoulis, 1991) but the ME principle is more effective in problems involving asymmetric constraints, as in the following paragraphs. It is easily seen that in the equiprobability case the entropy is $\varphi = \ln w$, which is Boltzmann's formula of entropy. Furthermore, it is seen that as w tends to infinity φ does the same. So the case $w = \infty$ does not have a physical meaning.

However, it regains physical meaning if a constraint additional to (3) is considered. For example, it could be postulated that the mean of the random variable is a given quantity μ:

$$\sum_{j=1}^{\infty} x_j\, p_j = \mu \tag{5}$$

Incorporating constraints (3) and (5) into (4) by means of Lagrange multipliers λ_0 and λ_1, taking the derivative of φ with respect to the unknown p_j and equating to zero, so as to maximize φ, it is easily obtained that,

$$p_j = \exp(-\lambda_0 - \lambda_1 x_j) \tag{6}$$

where λ_0 and λ_1 are easily calculated from (3) and (5). The resulting maximum entropy is

$$\varphi = \lambda_0 + \lambda_1 \mu \tag{7}$$

The entropy in continuous state variables

For a continuous random variable X that takes values x with probability density $f(x)$ satisfying

$$\int_{-\infty}^{\infty} f(x)\, dx = 1 \tag{8}$$

the (Shannon or extensive) entropy is by definition (e.g. Papoulis, 1991, p. 559)

$$\varphi := E[-\ln f(X)] = -\int_{-\infty}^{\infty} f(x) \ln f(x)\, dx \tag{9}$$

Here it should be noted that definition (9) for a continuous variable is not fully consistent with definition (4) for a discrete variable. That is, if one partitions the continuous x axis into intervals of length δ, thus forming a discrete representation of the random variable, applies entropy definition (4) and takes the limit as δ tends to zero, one will find that (4) results in infinite entropy. To overcome it, the entropy of a continuous variable is not the limit of φ of the discrete case, but rather the limit of the quantity $\varphi + \ln \delta$, which takes a finite value as δ tends to zero. This implies some differences in the properties of entropy in the two cases, among which are the following:

(1) the entropy of a discrete variable is always positive whereas that of a continuous variable may be positive, zero or negative;

(2) in the discrete case the entropy of any one-to-one transformation $Y = g(X)$ of the random variable X is exactly φ, i.e. identifies with that of X, but in the continuous case it is $\varphi + E[\ln|g'(X)|]$, where $g'(\ )$ is the derivative of $g(\ )$.

The last property implies that in the discrete case the value of entropy in a certain phenomenon does not depend on the metric X (or Y) that is used to quantify the description of the phenomenon. It depends, however, on the metric X (or Y) in the continuous case.

If the density $f(x)$ is defined in the interval (a, b) then application of the ME principle results in the uniform distribution in (a, b). If any of a and b tends to $\pm\infty$, the ME principle cannot be applied unless additional constraints are imposed. The most common ones, which will also be used in this paper, are the requirements for finite first and second moments, i.e.

$$E[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx = \mu_1 \equiv \mu \tag{10}$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)\, dx = \mu_2 \tag{11}$$

In physics, constraints similar to (8), (10) and (11) are used to maximize the entropy of a system so as to find the distribution of microstates of a system (Müller, 2003a, b; Papoulis, 1991, p. 573). Interestingly, if X denotes the momentum of molecules or atoms in a gas volume and $f(x)\, dx$ is the number of molecules or atoms with momenta between x and $x + dx$, then the left sides of constraints (8), (10) and (11) are respectively proportional to the macroscopic mass, momentum, and energy of the gas.

Application of the ME principle with constraints (8), (10) and (11) results in (e.g. Papoulis, 1991, p. 571; Dowson & Wagg, 1973; Tagliani, 1993, 2002a, b)

$$f(x) = \exp(-\lambda_0 - \lambda_1 x - \lambda_2 x^2) \tag{12}$$

where λ_0, λ_1 and λ_2 are Lagrange multipliers that can be calculated combining (12) and the constraints. The resulting maximum entropy is

$$\varphi = \lambda_0 + \lambda_1 \mu_1 + \lambda_2 \mu_2 \tag{13}$$

Careful inspection of (12) shows that $f(x)$ is the normal distribution; specifically, after algebraic manipulations and use of the obvious relationship $\mu_2 = \mu^2 + \sigma^2$ (where σ is the standard deviation), it is obtained that

$$\lambda_0 = \ln(\sigma\sqrt{2\pi}) + \frac{\mu^2}{2\sigma^2},\ \ \lambda_1 = -\frac{\mu}{\sigma^2},\ \ \lambda_2 = \frac{1}{2\sigma^2},\ \ \varphi = \ln(\sigma\sqrt{2\pi e}) \tag{14}$$

Substituting these values to (12), the well-known formula of the normal curve is obtained. Interestingly, (14) shows that the entropy of a normally distributed variable depends only on its standard deviation, not on its mean.

Hydrological variables are non-negative. If it is assumed that $x > 0$, in addition to constraints (8), (10) and (11), then (12) is still valid but for $x > 0$ (Papoulis, 1991, p. 571), thus representing the truncated normal distribution; the entropy is still given by (13). Specific formulas for the calculations associated with this distribution are given in Koutsoyiannis (2005). Obviously, in this case, (14) is not valid and the entropy does depend on both μ and σ.

An interesting limiting case is when $\lambda_2 = 0$ (it cannot be negative) and at the same time $x > 0$. The same limiting case emerges if in the application of the ME principle constraint (11) is omitted. Careful inspection of (12) in this case shows that $f(x)$ is the exponential distribution; specifically, after algebraic manipulations, it is obtained that

$$\lambda_0 = \ln \mu,\ \ \lambda_1 = 1/\mu,\ \ f(x) = (1/\mu)\exp(-x/\mu),\ \ F^*(x) = 1 - F(x) = \exp(-x/\mu),\ \ \varphi = 1 + \ln \mu \tag{15}$$

Generalization of entropic form

Tsallis (1988; 2004) heuristically generalized the Shannon entropy by postulating the entropic form

$$\varphi_q := \frac{1 - \sum_{i=0}^{w} p_i^q}{q - 1} \tag{16}$$

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 Shannon entropy, some of which will be discussed below. It can be straightforwardly checked that the limit for $q \to 1$ precisely reproduces the Shannon entropy in (4), i.e., $\varphi_1 \equiv \varphi$. As in the case of φ, φ_q achieves its maximum value at equiprobability ($p_i = 1/w$; this happens for $q > 0$ whereas for $q < 0$ this value corresponds to a minimum). Tsallis entropy can be extended to continuous random variables, in which case, assuming a non-negative random variable,

$$\varphi_q = \frac{1 - \int_0^\infty [f(x)]^q \, dx}{q - 1} \tag{17}$$

(Tsallis, 2004, p. 24). Other generalized entropic forms have been also proposed, i.e. the Rényi entropy (Rényi, 1970) and the normalized entropy (Landsberg and Verdal, 1998; Rajagopal and Abe, 1999); these can be expressed as monotonic functions of Tsallis φ_q. As Tsallis entropy has been fruitfully used in several scientific fields including physics, chemistry, biology, economics, medicine, computer sciences and social sciences (Tsallis, 2004), it will be also used here, in addition to Shannon entropy.

In contrast to Shannon entropy whose maximization, with typical constraints such as (8), (10) and (11), results in exponential-type distributions, the maximization of φ_q with these constraints yields a power-type distribution, i.e.,

$$f(x) = [1 + \kappa\,(\lambda_0 + \lambda_1\, x + \lambda_2\, x^2)]^{-1 - 1/\kappa} \tag{18}$$

$$\varphi_q = (\kappa + 1)\,(\lambda_0 + \lambda_1\, \mu_1 + \lambda_2\, \mu_2) \tag{19}$$

where $\kappa := (1 - q)/q$ (see derivation in Koutsoyiannis, 2005). It is easily seen that for $\kappa \to 0$ ($q \to 1$), (18) and (19) become identical to (12) and (13) respectively which correspond to Shannon entropy. The Lagrange multipliers λ_i can be estimated combining (18) and constraints (8), (10) and (11). This case, however, is difficult to handle analytically, because complicated formulas appear.

Again, a simple limiting case emerges when $\lambda_2 = 0$. Combining (18) and (8) it is obtained that $\lambda_1 = (1 + \kappa\, \lambda_0)^{-1/\kappa}$; furthermore, setting $\lambda = \lambda_1^{\;-1-\kappa}$, (18) yields

$$f(x) = \frac{1}{\lambda}\left(1 + \frac{\kappa\, x}{\lambda}\right)^{-1 - 1/\kappa}, \quad F^*(x) = 1 - F(x) = \left(1 + \frac{\kappa\, x}{\lambda}\right)^{-1/\kappa} \tag{20}$$

which is the Pareto distribution for $x > 0$. If ones sets $Y = X + \lambda/\kappa$, so that $y > \lambda/\kappa$, then one obtains the Pareto distribution of the form (1). Furthermore, if $\kappa = 0$, (20) identifies with the exponential distribution (15).

THE APPLICATION OF THE ME PRINCIPLE FOR HYDROLOGICAL VARIABLES

General considerations

In this section the ME principle will be applied to hydrological variables. It is assumed that a hydrological variable can be represented as a random variable X, which possesses a positive mean μ and standard deviation σ (so that the second moment about the origin is $\mu_2 = \mu^2 + \sigma^2$). It should be noted that in the proposed framework the origin should be identical to the lower bound of the variable examined; this concerns especially the mean μ. If a variable of interest has a lower bound different from zero, then a shift of origin (to make it identical to the lower bound) is necessary for the available time series before estimation of μ. Without loss of generality, the variable X can be standardized by dividing it by μ, so that it have mean 1 and standard deviation σ/μ, which is the CV of the original X. Here, the value of σ/μ is not regarded as merely a statistic that can be used, for instance, in the parameter estimation of a given distribution. Rather, it will be used to determine which is the appropriate distribution of the studied variable. This will be done seeking the distribution that maximizes the entropy for the given σ/μ, i.e. applying the ME principle, which, in the author's view, corresponds to the hypothesis that nature behaves in a manner that makes uncertainty as high as possible.

It can be easily shown that the entropies φ' and φ'_q of the standardized variable $X' = X/\mu$ are related to the respective ones of the original X by

$$\varphi' = \varphi - \ln \mu, \quad \varphi'_q = \mu^{q-1}\, \varphi_q + \frac{1 - \mu^{q-1}}{q - 1} \tag{21}$$

Furthermore, φ' and φ'_q will be called standard entropies and for convenience the primes in the notation will be omitted unless there is risk of confusion.

In seeking the distribution that maximizes uncertainty, the extensive entropy φ will be used, unless this results in infeasible situations, whence the nonextensive entropy φ_q will be used instead. When the variation σ/μ is low, extensive entropy is sufficient; thus the cases of low and high variation will be distinguished.

Variables with low variation

Application of the ME principle for the conditions set results in (12), which is the truncated normal distribution. If σ/μ is very low, the truncation is negligible, so the resulting distribution is the symmetric, bell-shaped normal distribution. As σ/μ increases, the truncation becomes important and at the critical value $\sigma/\mu = \sqrt{\pi/2 - 1} = 0.75551$ (whence $\lambda_1 = 0$) the mode of the distribution becomes zero, so the distribution becomes (inverse) J-shaped. It continues to be J-shaped up to $\sigma/\mu = 1$, whence the truncated normal identifies with the exponential distribution ($\lambda_2 = 0$). In the exponential case, as obtained from (21) and (15), the standard Shannon entropy takes its highest possible value which is unity.

Variables with high variation – Extensive entropy approach

For CV even higher ($\sigma/\mu > 1$) the ME distribution, in terms of extensive entropy, does not exist (Dowson & Wragg, 1973). To clarify the non-existence, it will be temporarily assumed that the domain of X is the finite interval $[0, b)$, with $b > 0$. In this case, the ME distribution exists and has the form (12) with $\lambda_2 < 0$, so that the density function is U-shaped. As the domain of the variable becomes wider, i.e. as $b \to \infty$, $\lambda_2 \to 0$, so the distribution tends to the exponential, which however has $\sigma/\mu = 1$, not equal to the required value. Thus, the exponential distribution is the limiting case for the extensive entropy concept, and its extensive standard entropy $\varphi = 1$ must be regarded as the upper limit for any distribution.

The non-existence of the ME distribution for the case examined does not preclude comparison of specified alternative distributions to detect which of them results in the maximum, among them, entropy and thus uncertainty. To this aim, a systematic parametric comparison is done, which is based on the following probability density function, devised for this purpose:

$$f(x) = (1 + \kappa\, \lambda_0 + \kappa\, \lambda_1\, x^{\nu_1})^{-1 - 1/\kappa}\, x^{\nu_2 - 1} \tag{22}$$

This comprises four parameters, the scale parameter λ_1 and the shape parameters ν_1, ν_2 and κ; λ_0 should not be regarded a parameter but a constant having a value that assures satisfaction of (8) (conservation of mass). It can be observed that the variable X^{ν_1} has Beta Prime (also known as Beta of the second kind) distribution (Evans et al., 2000); thus the distribution of X will be called Power-transformed Beta Prime (PBP). This distribution merges several exponential-type and power-type distributions. As seen in Table 1, several of the common distributions, such as the exponential-type Gamma and Weibull and the power-type Pareto, are special cases of PBP. The Beta Prime distribution, more rarely used in hydrology (Yevjevich, 1972; Koutsoyiannis, 2004a) is another special case. Two additional limiting special cases, denoted as PBP/L1 and PBP/L2, emerge when the shape parameter κ takes its extreme values, i.e. when $\kappa \to 0$ and $\kappa \to \infty$, respectively. The normal distribution can be regarded, too, as a special case of this generalized distribution, as it is known that the Gamma distribution for high values of shape parameter tends to the normal distribution. Thus, given that many of the distributions used in hydrology are special cases of (22), this generalized density can serve as a basis for an intercomparison of them in terms of the resulting entropy.

An extended numerical investigation was done based on (22). The interested reader may find details (including equations for the handling the PBP and its special cases) in Koutsoyiannis (2005); here a single example is provided (Figure 1), for $\mu = 1$ and $\sigma/\mu = 1.5$. The two parameter special cases Gamma and Weibull (corresponding to $\kappa = 0$) as well as the Pareto case ($\kappa > 1$) are easy to fit given μ and σ/μ. The resulting entropies of these three cases are marked in Figure 1. Clearly, the Pareto case yields entropy significantly higher than the exponential type Gamma and Weibull. In addition, Figure 1 shows the resulting entropy of the generalized four-parameter form of (22). Specifically, for each specified value of κ, the value of ν_1 that maximizes Shannon entropy was calculated and the maximized entropy has been plotted in Figure 1 versus κ. It is observed that the maximized entropy increases slightly with the increase of κ, without departing significantly from the Pareto case. The limiting exponential-type case PBP/L1 (for $\kappa \to 0$) gives entropy higher than in the Gamma and Weibull cases and lower than in the Pareto case. The limiting power-type case PBP/L2 (for $\kappa \to \infty$) gives entropy even higher than in the Pareto case.

However, a more careful inspection of the limiting PBP/L2 distribution shows that it has problems related to its observability. Its probability plot in comparison with that of the exponential distribution (see Koutsoyiannis, 2005) shows that the two are indistinguishable except for very high return periods. So, even if a hydrological process followed the PBP/L2, its observation, based on a typical record with length, say, 100 years or less, would indicate that the distribution is exponential and the CV is virtually unity. Thus, a sample estimate of CV greater than unity prompts for rejection of PBP/L2. Such problems do not appear with the Pareto distribution, which differs from the exponential distribution in the entire domain.

This must be regarded as a deficiency of the extensive entropy approach. A second deficiency is the fact that for growing σ/μ above 1, which intuitively signifies increasing variability and increasing uncertainty, the entropy does not increase. This is depicted in Figure 2, where the maximized standard entropy is plotted versus σ/μ. For $\sigma/\mu \leq 1$ the method provides a clear result that the ME distribution is the truncated normal (becoming exponential at the limit $\sigma/\mu = 1$) and the entropy is an increasing function of σ/μ. For $\sigma/\mu > 1$ the Shannon entropy of the Pareto distribution, also plotted in Figure 2, is slightly decreasing function of σ/μ, which is contrary to the notion of increasing uncertainty and rather unreasonable.

In conclusion this heuristic approach, is insufficient to yield a clear result of a ME distribution and incorporates an inconsistency. However, the approach was very helpful to demonstrate that (a) the exponential-type distributions like Gamma and Weibull do not comply with the ME postulate and (b) a power-type distribution at least for high return periods should be expected based on the ME principle.

Variables with high variation – Nonextensive entropy approach

The deficiencies of the extensive entropy approach are overcome using the generalized nonextensive entropy approach. In this case, there is no theoretical upper limit for the value of σ/μ. Once the value q is specified, application of the ME principle with constraints (8), (10) and (11) results in (18). However, q, the generalized entropy qualifier, should not be chosen in an arbitrary manner; its selection is a difficult task as described in Tsallis (2004, p. 23). Here a simplified approach was followed, suitable to the exploratory character of the study. To establish consistency with the extensive entropy approach discussed above, it was assumed that q should be as close to unity as possible (equivalently, κ should be as close to zero as possible). In this case, if $\sigma/\mu \leq 1$, κ will be zero, and the ME distribution will be the truncated normal. If $\sigma/\mu > 1$, κ will be positive but as small as possible. Thus, its value can be estimated assuming that the quadratic term in (18) vanishes ($\lambda_2 = 0$), whence the Pareto distribution (20) emerges. This assumption makes the exponential distribution the common limiting distribution of both the truncated normal distribution, when $\sigma/\mu \to 1$ from below, and the Pareto distribution when $\sigma/\mu \to 1$ from above. This simplified approach incorporates a weakness, i.e. the dependence of q on the time scale on which the process is studied: at larger time scales a process exhibits smaller CV, so the approach will yield higher q. An improved approach is currently studied and will be reported in the near future.

Final remarks

In conclusion, the results of this extended discussion gives the following picture of what one should expect if the ME principle applies to nature, in our case to hydrological processes:

1. When the variation σ/μ is smaller than 1, the truncated normal distribution applies and the uncertainty is described by the extensive entropy; this distribution is bell-shaped for small σ/μ and becomes J-shaped for σ/μ greater than about 0.75.
2. When σ/μ is greater than 1, the Pareto distribution applies and the uncertainty is described by the nonextensive entropy; this distribution is J-shaped.
3. The limiting cases $\sigma/\mu \to 0$ and $\sigma/\mu = 1$ are described by the normal and exponential distributions, respectively.

In this manner, the inconsistency of the extensive entropy approach between variability and entropy for high CV was eliminated. This is shown in Figure 2, where in addition to the extensive entropy, the non-extensive standard entropy of the Pareto case is plotted, which is an increasing function of σ/μ.

The resulting ME distributions for a wide range of CV are depicted in Figure 3 in terms of probability density functions and in Figure 4 in terms of distribution functions or, more precisely, plots of the random variate vs. return period T (semi-logarithmic plot, on which the exponential distribution appears as a straight line). All plots correspond to $\mu = 1$. Particularly, Figure 3a emphasizes the wide spectra of shapes of densities, from exponential-type bell-shaped to power-type J-shaped, which are all obtained by the unique ME principle and from a single index σ/μ. One may think that the differences in densities in Figure 3a are due to different standard deviations, which vary from 0.1 to 100. Therefore, the typical standardised variate $(x - \mu)/\sigma$, which has zero mean and unit standard deviation, was also used to construct another form of the plot, shown in Figure 3b. Again, the differences in shapes of ME densities are obvious. Here, it should be underlined that the latter kind of standardization, which has been of wide use in hydrological statistics, may be a good practice for variables with normal distribution but turns to be a bad practice for most hydrological variables, as it hides the important index σ/μ, which combined with ME determines the shape of the distribution.

CASE STUDIES

In this section, the theoretical results are tested based on real world data. Several long records of rainfall, runoff and temperature at scales hourly to annual were examined; among these eight are presented as listed in Table 2 (some additional are given in Koutsoyiannis, 2005). The empirical distribution functions (based on Weibull plotting positions) are graphically compared with the ME theoretical distributions. The graphical comparison is consistent with the exploratory character of this study. All distribution plots are constructed in the form of Figure 4, so that the shape of the distribution can be easily recognized visually and related to that of the limiting exponential distribution and eventually to the theoretical results shown in Figure 4. The exponential as well as the normal distribution are also plotted in all cases, as sort of 'benchmark' distributions.

Figure 5 depicts the distribution of hourly rainfall (nonzero values) in Athens, Greece, for the month of January. Athens is a dry place with mean annual rainfall around 400 mm, but January is the wettest month. The empirical CV is 1.47 > 1; thus the ME distribution is Pareto. Clearly, the figure shows that, indeed, the Pareto distribution fits perfectly the empirical data and departs significantly from both the exponential and normal distribution. "Imperfection" of the Pareto fitting to three (out of 2919) points, those around T = 20 years and beyond, reflects just the high uncertainty of empirical estimation of large T. (It is assumed here that the reader is familiar with probability plots and with the dramatic increase of estimation uncertainty for high T, so this will be not discussed further).

Figure 6 depicts the distribution of annual rainfall depth in Aliartos, Greece. Aliartos is less than a hundred kilometres to the north of Athens but its rainfall is 60% higher. The available rainfall record here is longer than that of Athens. Now the empirical σ/μ is 0.24 << 1; thus the ME distribution is the truncated normal, which due to low σ/μ is almost indistinguishable from the normal distribution. The figure shows that this distribution is very close to the empirical one and that the shape of both is extremely divergent from that of Figure 5. Again some imperfections appear for very high T. Apart from estimation uncertainty, in this case there may be another explanation. As in small scales (as in the case of Figure 5) the Pareto behaviour dominates, it is anticipated that the most extreme values of rainfall must be better described by a power-type law. Thus, the truncated normal distribution must be regarded as an appropriate model for typical return periods but not for the very high ones. Further research on this, i.e. the effect of a power-type distribution on small scales to the tail of the distribution on aggregate scales is currently ongoing and will be reported in the near future.

The next example is the most interesting one due to its practical significance. This example comes from an earlier study (Koutsoyiannis, 2004b) related to extreme daily rainfall maxima worldwide. In this study, daily rainfall records from 169 stations worldwide were analysed and it was found that several dimensionless statistics, including CV, of the annual maximum series are virtually constant worldwide, except for an error that can be attributed to a pure statistical sampling effect. This enabled the formation of a compound series-above-threshold for 168 of the stations. To this aim, all series were standardized by their mean and merged in one sample with length 17922 station-years. This sample is re-examined here in light of the ME principle. Its empirical distribution is depicted in Figure 7, where values lower than 0.79 are not shown, as this number is the lowest value of the merged series-above-threshold.

Before studying the sample, a few observations should be done about handling of the distributions when a series-above-threshold is available. To this aim, the variable $Y = X - c$ is introduced, where c denotes a threshold, and the survival function of Y conditional on being positive, i.e. $\overset{*}{F}_Y(y|Y > 0)$, is studied. If $\overset{*}{F}_X(x)$ denotes the (unconditional) survival function of X, it is easily obtained that

$$\overset{*}{F}_Y(y|Y > 0) = \frac{\overset{*}{F}_X(y + c)}{\overset{*}{F}_X(c)} \tag{23}$$

Based on (23) it can be shown that if $F_X(x)$ is truncated normal, then $F_Y(y|Y > 0)$ is again truncated normal with same λ_2. After algebraic manipulations, it can be seen that as c increases, both the conditional mean (μ_c) and standard deviation (σ_c) decrease and the conditional CV (σ_c/μ_c) increases, approaching unity as $c \to\infty$. In other words, as threshold increases, the conditional distribution of Y tends to become exponential.

In the case of the Pareto distribution the behaviour is different. From (23) and (20) it is obtained that, if $F_X(x)$ is Pareto, then $F_Y(y|Y > 0)$ is again Pareto, which can be expressed as in (20) with same shape parameter κ and with scale parameter λ replaced by $\lambda + \kappa\, c$. Based on these observations it can be seen that as c increases, both μ_c and σ_c increase (see Figure 8) while σ_c/μ_c remains constant (it depends on κ only). In other words, as threshold increases, the same Pareto distribution applies except for a change in the scale parameter.

Equation (23) and the observations of the previous two paragraphs enable the fitting of the distribution of the variable of interest X if a sample of Y is available. Coming again to the case study with the sample of 17922 station-years with threshold $c = 0.79$, the CV of Y is 0.95, i.e. very close to, but lower than, unity. This value implies a truncated normal distribution, close to the exponential. However, Figure 7, where the empirical distribution of this sample is depicted and compared to this and other theoretical distributions does not validate the appropriateness of the truncated normal distribution**.**

Here one must suspect the value 0.95 of σ_c/μ_c, which may contain an estimation error, and investigate it further. Indeed, Figure 8 shows that a slight increase of the threshold c results in σ_c/μ_c higher than unity. Above a threshold $c = 0.9$, σ_c/μ_c becomes virtually constant, with average value 1.19. This clearly indicates a Pareto ME distribution and this is verified both in Figure 7 and in Figure 8. In the latter it is observed that, as c increases, both σ_c and μ_c follow the Pareto pattern (increasing functions) rather than the truncated normal pattern (decreasing functions). The ME value of κ for $\sigma/\mu = 1.19$, is 0.15, i.e. equal to that estimated in Koutsoyiannis (2004b) using a weighted least-squares method.

The next variable examined here is temperature, which is of interest to hydrology due to its direct link to evaporation. To apply the ME theoretical framework to temperature, one may think that the non-negativity constraint does not apply to this variable. In this case, as explained earlier, the ME principle results in the normal distribution. However, more precisely, it should be observed that temperature is bounded from below by the absolute zero (zero kelvin), so that it is again a non-negative variable if expressed in kelvins. The second option (that with the non-negative constraint) was preferred here as it is more precise and in accord to the formalism used in all other cases; as seen in the next paragraph the result is again the normal distribution, i.e. both options are practically equivalent.

Daily temperature records for Athens were used in this case, whose distributions for the months of January (the coldest), August (the hottest, together with July) and April are depicted on Figure 9. The very small values of CV in this case (0.01, 0.0075 and 0.0095), imply normal ME distribution (truncated normal is indistinguishable from normal). So, even at a scale as fine as the daily, ME does not imply a power-type law but an exponential-type law. This is validated in Figure 9 that depicts comparison of the empirical and normal distribution. This is also the case for larger time scales such as annual. To make an example for the annual scale, one of the longest available temperature records worldwide was used, that of Geneva (228 years). The CV in this case is even lower, 0.0024, which again implies a normal distribution. This is validated in Figure 10 that depicts comparison of the empirical and normal distribution.

The last variable examined is runoff, in which the application of the ME principle is more difficult. In a detailed approach, one should distinguish in runoff the base flow component and the flood component, and apply the ME principle separately for each component. In the exploratory approach followed here such a separation will be not done. The data used is for the Boeoticos Kephisos river basin, the basin where Aliartos raingauge (mentioned earlier) is located. A significant part of the 2048 km^2 area of this basin lies on karst, which leads to a significant base flow component. This implies a non-negligible smoothing of the rainfall input for small and medium rainfall, simultaneously amplifying the variability for high rainfall, when floods occur. The river is almost in natural condition, as no large constructions such as dams have been built; however, the flow regime is artificially modified in summer months, in which the natural flows are minimal, as farmers exploit river flow for irrigation and almost no water reaches the river outlet. This will be taken into account in the application of the ME principle as discussed below.

Measurements in the river outlet have been performed since 1907, which makes the resulting runoff time series the longest available in Greece (96 years). Unfortunately, the current digitized archive is for monthly scale only. For the daily scale only 23 years runoff is available in the archive (see Table 2). The empirical distribution of all available daily data values are depicted in Figure 12. For the reasons already explained, the interest here is focused on the higher values of discharge, which are not affected by the base flow mechanism and the summer regulations. The CV of the data set is 1.29 and if the artificially induced zero values are excluded it becomes 1.05. This implies a Pareto distribution but the proximity of the value to 1 calls for a more careful analysis. An analysis similar to that of the extreme rainfall case study discussed earlier was performed, which is consistent with the focus on the high values. Application of a threshold to the series gradually increases the CV as shown in Figure 11 up to a value 1.19 for a threshold equal to the median of the original series (0.245 mm); no further increase is observed above this value (expect for some irregular fluctuations). Based on the value $\sigma/\mu = 1.19$ the Pareto distribution was fitted ($\kappa = 0.15$), as shown in Figure 12. It is observed that the Pareto distribution is in good agreement with the empirical distribution, whereas exponential and normal distributions depart significantly. Although the fitting was based on a threshold equal to the median, the empirical distribution plot in Figure 12 includes also the lower values. It is observed that, even below the threshold, i.e., in values governed by the base flow mechanism, this Pareto distribution can be a good approximation. On larger scales up to annual, analyses with the complete 96 year sample yielded similar results as in the rainfall case; for example, the annual CV is 0.41 (much lower than in the daily time series yet higher than that of the annual rainfall which is 0.24 as discussed above) which implies a truncated normal distribution type; this was verified in a probability plot (similar to that of Figure 6, not included in the paper for brevity).

SYNOPSIS, CONCLUSION AND DISCUSSION

The principle of maximum entropy (ME) is used to explain the statistical distributions followed by hydrological variables. In this context, maximum entropy is interpreted as maximum uncertainty. When feasible, the extensive or Shannon entropy is used as a measure of uncertainty; otherwise the generalized nonextensive or Tsallis entropy is used instead. The only conditions used for the maximization of entropy are that a hydrological variable is non-negative and possesses a certain variability, expressed by the coefficient of variation σ/μ.

The theoretical analyses show that two alternative distribution types emerge by the application of the ME principle. Specifically, when the variation σ/μ is smaller than 1, the uncertainty can be described by the extensive entropy, which results in the truncated normal distribution; this distribution is bell-shaped for small σ/μ and becomes J-shaped for σ/μ greater than about 0.75. When σ/μ is greater than 1, the uncertainty is described by the nonextensive entropy, which results in J-shaped power-type (Pareto) distributions. In addition, the normal distribution appears as the limit of the truncated normal when $\sigma/\mu \to 0$. Besides, the exponential distribution appears as the limit of both the truncated normal and the Pareto distribution when $\sigma/\mu = 1$.

Testing of these theoretical results with numerous hydrological data sets validates the applicability of the ME principle. Specifically, given that rainfall possesses high variation on small time scales (e.g. hourly or daily), a Pareto distribution is expected and this is validated by the data. As time scale becomes larger (e.g. annual) the variation decreases and the truncated normal distribution applies; this is also validated by the data. Statistics of runoff exhibit more or less similar behaviour with those of rainfall. In contrast, absolute temperature exhibits very low variation even in the finest time scales and thus the normal distribution applies on all times scales; again this is validated by the data.

Several issues related to the application of the ME principle to hydrological processes should be addressed with further research; some that are investigated in ongoing research have already been indicated earlier. The empirical investigation of applicability of the ME principle in a wide range of hydrological series worldwide will provide additional insights and potentially enhance the confidence on the method or trace its limitations. The theoretical and empirical investigation of the distributional behaviour of a variable at many time scales simultaneously is another important issue. This is related to a more rational choice of the nonextensive entropy qualifier q and is especially useful for the behaviour in the tails when the variation on large scales is low and on small scales high (> 1, thus implying a power-type distributional behaviour). The uncertainty of the empirical estimation of CV from historical data is also a significant issue given the important role of this indicator in the resulting ME distribution. Potential physical explanations of the observed values of σ/μ in different processes and potential empirical investigation of the geographical variation or invariance of this coefficient worldwide (as in the case for rainfall extremes, where, as mentioned above, it was found to be almost invariant worldwide) will enhance the applicability of ME principle.

In conclusion, the ME principle seems to be fundamental in hydrology, as it explains the distributional behaviour of hydrological variables. The dominance of uncertainty, implied by the applicability of the ME principle and the agreement of theoretical results and observations, need not be regarded as a disappointing attribute for scientists and engineers. In contrast, the knowledge that uncertainty dominates is a very important and helpful knowledge, as it can be used to establish the probabilistic law of a certain phenomenon and to design structures more safely. Thus, knowing that, as a result of the ME principle, the distribution of rainfall on a small time scale is power- rather than exponential-type will lead to the use of EV2, rather than EV1, distribution of maximum rainfall. This, in turn, will help avoid underestimation of design rainfall, which unfortunately was the rule for several tens of years (Koutsoyiannis, 2004a, b). In this respect, maximization of uncertainty can be paralleled Socrates' view, quoted in the beginning of the Introduction, who regarded that his knowledge of ignorance made him the wisest man, as it is learned from his Apology by Plato.

In terms of the state scaling law in (1), which has been discussed by many as applicable to hydrology, both the theoretical analyses and the empirical evidence do not validate it precisely. State scaling is only an approximation, merely valid when processes have high variation and when return period is high. A better approximation, produced by the ME principle, is the Pareto distribution in the form (20), equally simple yet more accurate than (1). In cases of low variation, state scaling disappears and the normal distributional behaviour, which does not have scaling properties, emerges. Interestingly however, as discussed in the second part of the study, the normal distribution combined with positive autocorrelation of a process, results in time scaling behaviour due to the ME principle.

Acknowledgements The data for Athens became available courtesy of the National Observatory of Athens; thanks are due to the director and members of its Institute of Atmospheric Research M .Petrakis, D. Founda and B. Kotroni. The data sets of the Boeoticos Kephisos basin were formed by the Itia research group of the National Technical University based on raw data by the Ministry of Environment, Physical Planning and Public Works and the Athens Water Supply and Sewerage Company. The temperature data of Geneva are made available online by the National Climatic Data Center (http://www1.ncdc.noaa.gov/pub/data/paleo/climate1500ad). Calculations were facilitated by numerical methods developed and made freely available by the Foxes team (http://digilander.libero.it/foxes/index.htm). Thanks are due to Aldo Tagliani for his explanations on some questions and to two anonymous reviewers for their comments that helped improve the presentation and prevented some overstatements due to author's enthusiasm.

REFERENCES

Bak, P. (1996) *How Nature Works, The Science of Self-Organized Criticality*. Copernicus, Springer-Verlag, New York, USA.

Barndorff-Nielsen, O. E., Gill, R. D., & Jupp, P. E. (2001) Quantum information, in *Mathematics Unlimited – 2001 and Beyond* (edited by B. Engquist and W. Schmid), Springer, Berlin.

Bernier, J. (1964) La prévision statistique des bas débits. In: *Symposium on Surface Waters. General Assembly of Berkeley*, IAHS Publ. no 63:340 351.

Buchanan, M., (2000) *Ubiquity, The Science of History, or Why the World is Simpler Than We Think,* Weidenfeld & Nicolson, London.

Dowson, D. C. & Wragg, A. (1973) Maximum-entropy distributions having prescribed first and second moments, *IEEE Transactions on Information Theory*, IT-19, 689-693.

Eagleson, P.S. (1972) Dynamics of Flood Frequency, *Water Resources Research*, 8(4):878 898.

Evans, M., Hastings, N., & Peacock, B (2000) *Statistical Distributions*, Wiley, New York, NY.

Georgii, H.-O. (2003) Probabilistic aspects of entropy, in *Entropy* (edited by A. Greven, G. Keller & G. Warnecke), Princeton University Press, Princeton, NJ.

Jaynes, E.T. (1957) Information theory and statistical mechanics, *Physical Review* 106(4):620-630.

Keane, M. (2003) Entropy in ergodic theory, in *Entropy* (edited by A. Greven, G. Keller & G. Warnecke), Princeton University Press, Princeton, NJ.

Kite, G. W. (1988) *Frequency and Risk Analyses in Hydrology*, Water Resources Publications, Littleton, CO.

Koutsoyiannis, D. (2004a) Statistics of extremes and estimation of extreme rainfall, 1, Theoretical investigation, *Hydrological Sciences Journal*, 49(4), 575-590, August 2004.

Koutsoyiannis, D. (2004b) Statistics of extremes and estimation of extreme rainfall, 2, Empirical investigation of long rainfall records, *Hydrological Sciences Journal*, 49(4), 591-610, August 2004.

Koutsoyiannius, D. (2005) Internal report (http://www.itia.ntua.gr/e/docinfo/641/).

Landsberg, P. T. & Vedral, P. (1998) Distributions and channel capacities in generalized statistical mechanics, *Phys. Lett. A*, 247, 211-217.

Müller, I. (2003a) Entropy: a subtle concept in thermodynamics, in *Entropy* (edited by A. Greven, G. Keller & G. Warnecke), Princeton University Press, Princeton, NJ.

Müller, I. (2003b) Entropy in nonequilibrium, in *Entropy* (edited by A. Greven, G. Keller & G. Warnecke), Princeton University Press, Princeton, NJ.

Papoulis, A. (1991) *Probability, Random Variables, and Stochastic Processes* (third edn), 57. McGraw-Hill, New York, USA.

Rajagopal, A. K., & Abe, S., Implications of form invariance to the structure of nonextensive entropies, Phys. Rev. Lett. 83. 1711-1714.

Rényi, A. (1970) Probability Theory, North-Holland, Amsterdam.

Singh, V. P. (1997) The use of entropy in hydrology and water resources, *Hydrological Processes*, 11, 587-626.

Singh, V.P. & Guo, H. (1995a) Parameter estimation for 2-parameter Pareto distribution by POME, *Water Resources Management*, 9, 81-93.

Singh,V.P. & Fiorentino, M. (eds.) (1992) *Entropy and Energy Dissipation in Water Resources*, Kluwer ,Dordrecht.

Singh, V.P. & Guo, H. (1995b) Parameter estimation for 3-parameter generalized Pareto distribution by the principle of maximum entropy (POME), *Hydrological Sciences Journal*, 40(2), 165-181.

Schroeder, M. R. (1991) *Fractals, Chaos, Power Laws: Minutes From an Infinite Paradise*, FreeMan, New York, NY.

Tagliani, A. (1993) On the application of maximum entropy to the moments problem, *J. Math. Phys.*, 34(1), 326-336.

Tagliani, A. (2002a) Entropy estimate of probability densities having assigned moments: Stieltjes case, *Applied Mathematics and Computation*, 130, 201-211.

Tagliani, A. (2002b) Maximum entropy solutions and moment problem in unbounded domains, *Applied Mathematics Letters*, 16, 519-524.

Tsallis, C. (1988) Possible generalization of Boltzmann-Gibbs statistics, *J. Stat. Phys.* 52, 479-487.

Tsallis, C. (2004) Nonextensive statistical mechanics: construction and physical interpretation, in *Nonextensive Entropy, Interdisciplinary Applications* (edited by M. Gell-Mann & C. Tsallis), Oxford University Press, New York, NY.

Yevjevich, V. (1972) *Probability and Statistics in Hydrology*. Water Resources Publications, Fort Collins, Colorado, USA.

Table 1. Special cases of the PBP distribution.

Limit conditions	Density function	Reference
$\kappa \to 0$	$f(x) = \exp(-\lambda_0 - \lambda_1 x^{\nu_1})\, x^{\nu_2 - 1}$	PBP/L1
$\kappa \to 0, \nu_1 = 1$	$f(x) = \exp(-\lambda_0 - \lambda_1 x)\, x^{\nu_2 - 1}$	Gamma
$\kappa \to 0, \nu_2 = \nu_1$	$f(x) = \exp(-\lambda_0 - \lambda_1 x^{\nu_1})\, x^{\nu_1 - 1}$	Weibull
$\nu_2 = \nu_1 = 1$	$f(x) = (1 + \kappa\lambda_0 + \kappa\lambda_1 x)^{-1-1/\kappa}$	Pareto
$\nu_1 = 1$	$f(x) = (1 + \kappa\lambda_0 + \kappa\lambda_1 x)^{-1-1/\kappa}\, x^{\nu_2 - 1}$	Beta prime
$\kappa \to \infty, \kappa\lambda_0 \to \kappa_0, \kappa\lambda_1 \to \kappa_1$	$f(x) = \dfrac{x^{\nu_2 - 1}}{1 + \kappa_0 + \kappa_1 x^{\nu_1}}$	PBP/L2

Table 2. List of the case studies and respective data sets and their characteristics.

Case #	Process[a] & time scale[b]	Station	Data period	Period of year	Record length[c] (years)	Number of data values[d]	Units	Mean	CV & ME distribution	MSE[e]
1	P/H	Athens	1927-1996	Jan.	70 (65)	48397 (2919)	mm	1.07	1.47/P	1.32
2	P/A	Aliartos	1907-2003	Year	96	96	mm	658.4	0.24/TN	-0.008
3a[f]	P/D	168 stations worldwide	1822-2002	Year	17922[g]	17922	-	0.34[h]	0.95/TN	0.998
3b[f]								0.28[h]	1.19/P	1.160
4	T/D	Athens	1930-2003	Jan.	74	2294	K	282.2	0.01/N	-3.19
5	T/D	Athens	1930-2003	Apr.	74	2220	K	388.4	0.0095/N	-3.22
6	T/D	Athens	1930-2003	Aug.	74	2294	K	300.2	0.0075/N	-3.47
7	T/A	Geneva	1753-1980	Year	228	228	K	282.8	0.0024/N	-4.62
8	R/D	Boeoticos Kephisos	1978-2003	Year	25 (23)	8402	mm	0.38 0.45[h]	1.29/P 1.19/P	1.160

Notes: [a] P: precipitation (rainfall); T: temperature; R: runoff. [b] H: hourly; D: daily; A: annual. [c] In parentheses the equivalent years if missing data are not counted. [d] In parentheses the number of values without counting the zero values. [e] Maximized standard entropy. [f] Time series above threshold, standardized; 3a and 3b correspond to different values of threshold (see text). [g] Number of station-years of the unified record. [h] Mean, conditional on being greater than threshold, minus threshold. (see text).

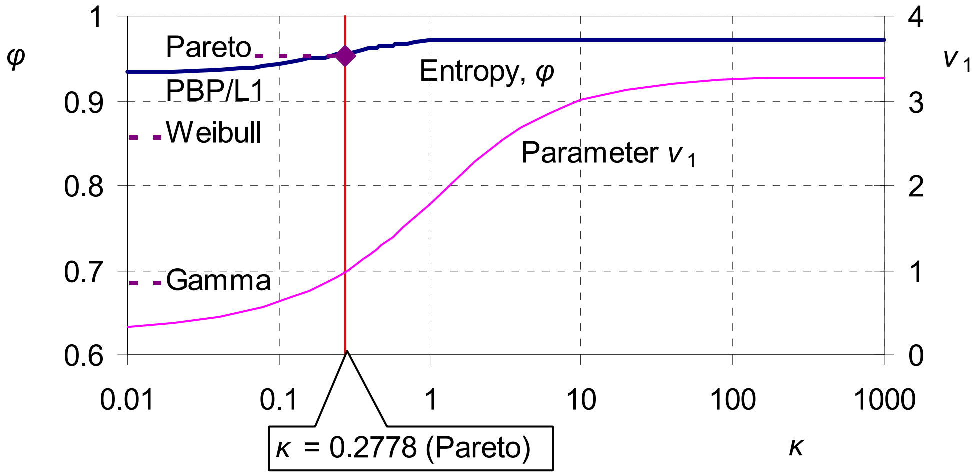

Figure 1. Maximized, over varying v_1 and specified κ, entropy of the PBP distribution with μ = 1, σ/μ = 1.5, and optimal value of parameter v_1 as functions of the parameter κ.

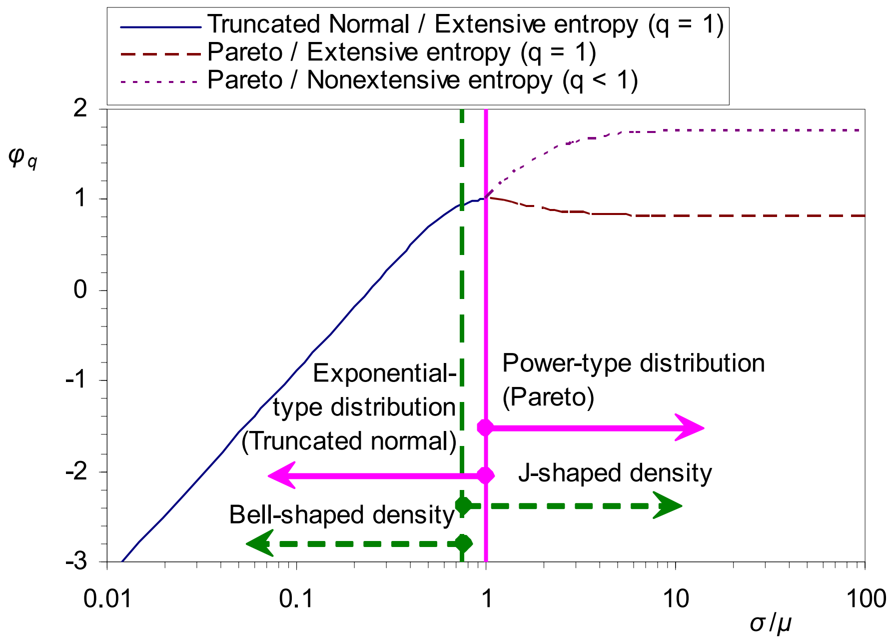

Figure 2. Maximized standard entropy and maximising distribution versus the coefficient of variation σ/μ..

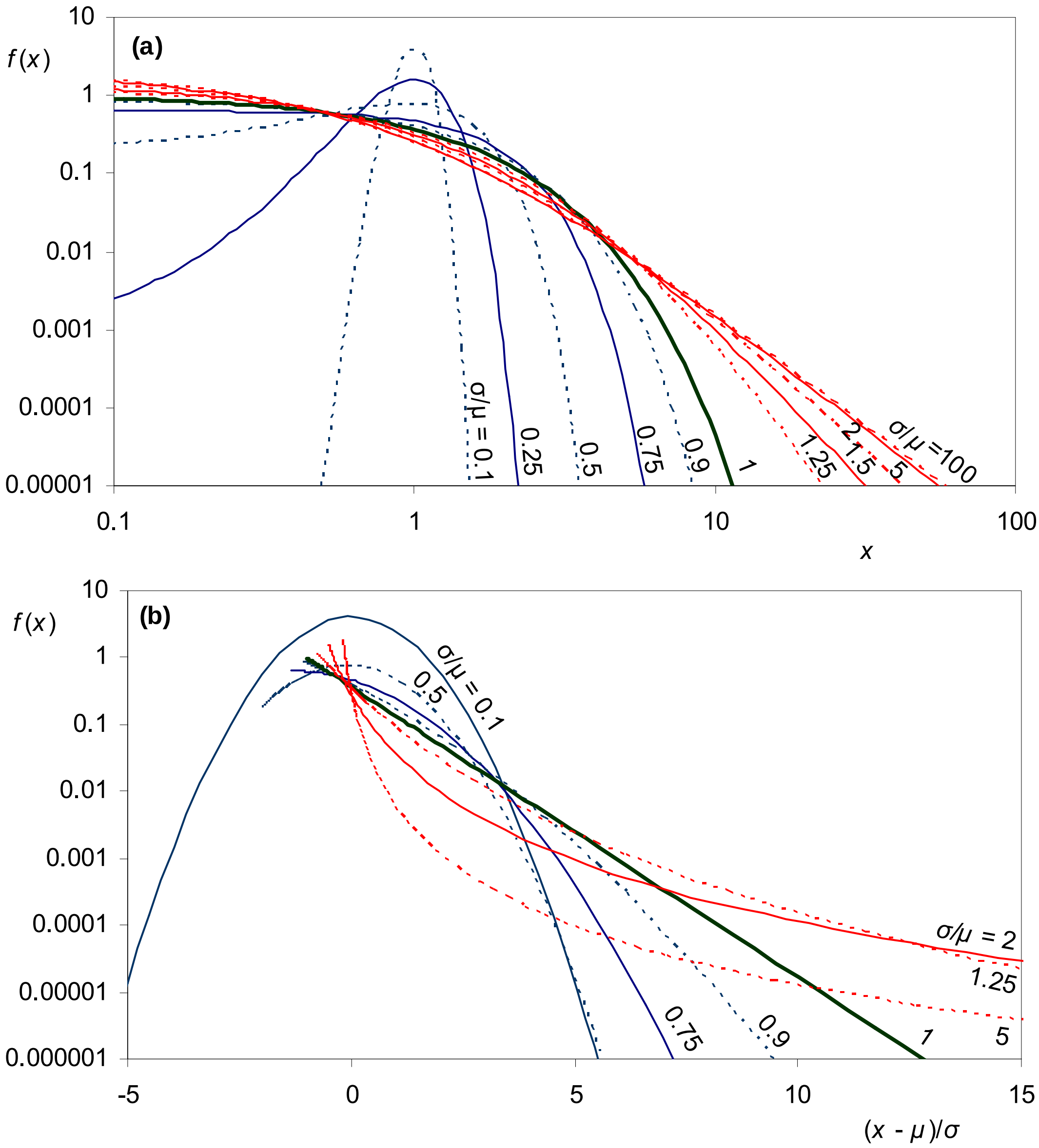

Figure 3. Plots of the ME probability density functions versus (a) the random variate x, and (b) the standardized variate $(x - \mu)/\sigma$, for $\mu = 1$ and for several values of the coefficient of variation σ/μ. The ME densities are truncated normal for $\sigma/\mu < 1$, exponential for $\sigma/\mu = 1$ and Pareto for $\sigma/\mu > 1$.

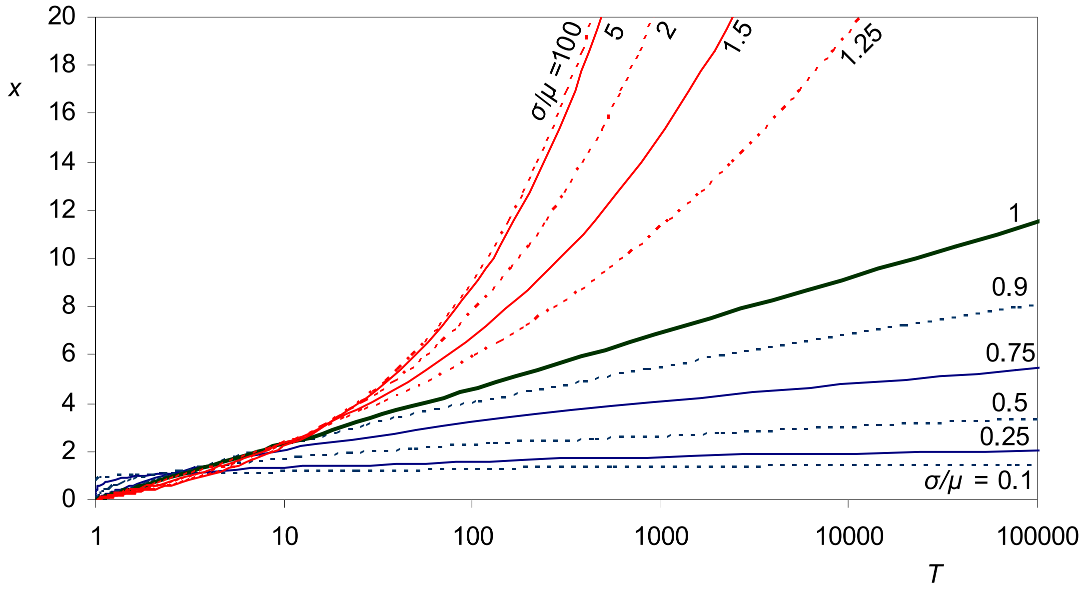

Figure 4. Graphs of the ME distribution functions (shown as plots of the random variate versus return period) for $\mu = 1$ and for several values of the coefficient of variation σ/μ. The ME distributions are truncated normal for $\sigma/\mu < 1$, exponential for $\sigma/\mu = 1$ and Pareto for $\sigma/\mu > 1$.

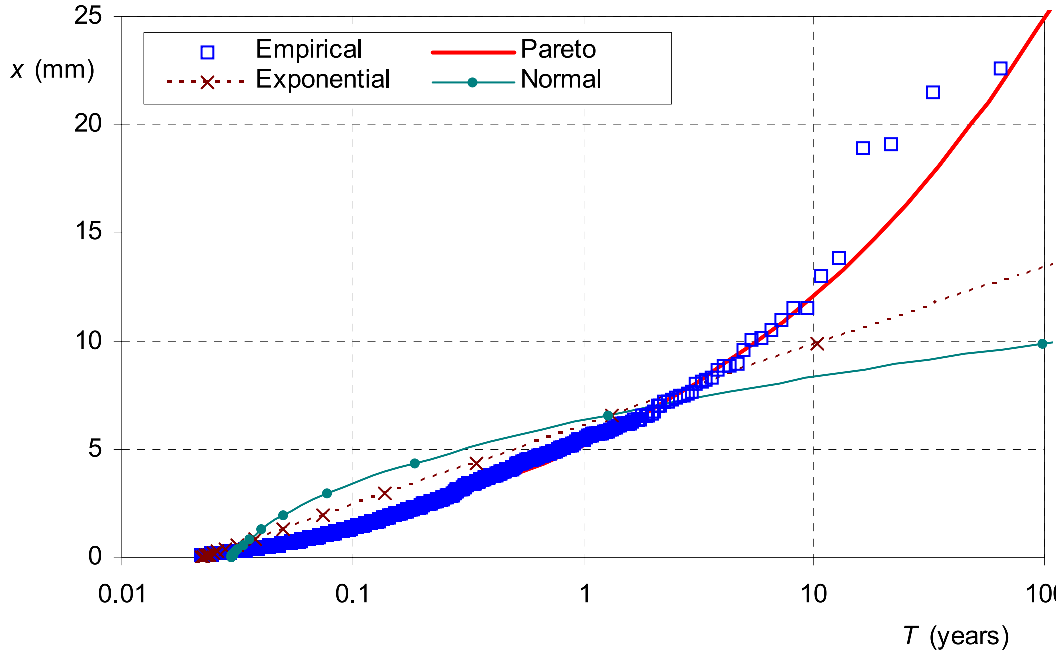

Figure 5. Plot of historical hourly rainfall depth in Athens for the month of January versus return period, in comparison to Pareto, exponential and normal distributions (due to the high density of empirical points for return periods smaller than 5 years, the curve of the Pareto distribution is not distinguished because it lies just above the empirical points).

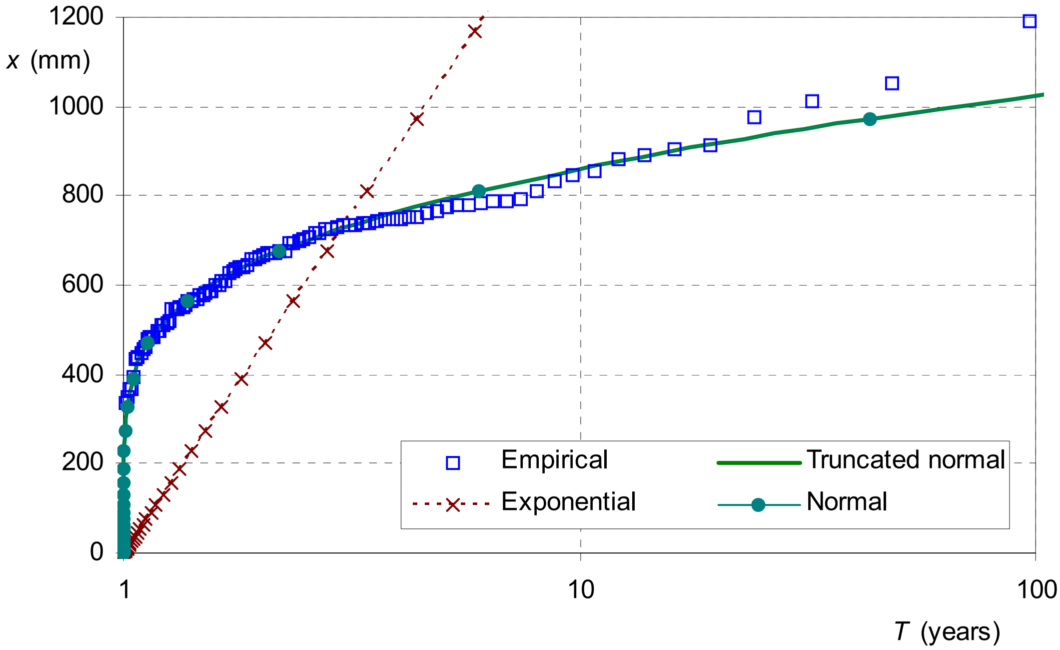

Figure 6. Plot of historical annual rainfall depth in Aliartos versus return period, in comparison to truncated normal, exponential and normal (almost indistinguishable from truncated normal) distributions.

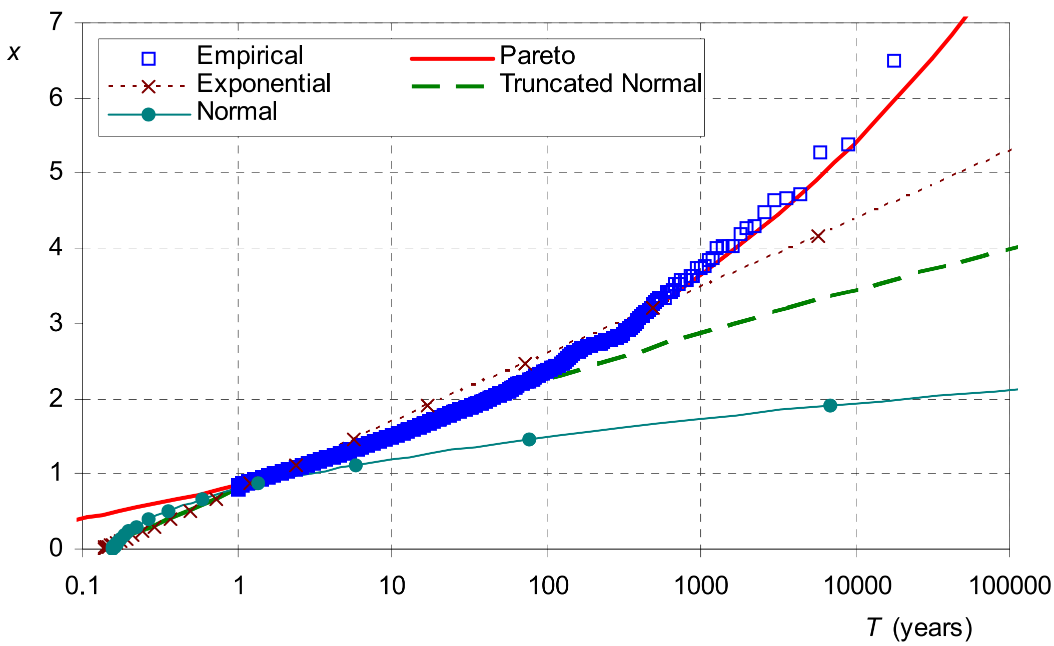

Figure 7. Plot of daily rainfall depth from the unified standardized sample above threshold, formed from data of 168 stations worldwide, versus return period, in comparison to Pareto, exponential, truncated normal and normal distributions. Exponential-type and Pareto distributions were fitted assuming conditional coefficient of variation 0.95 and 1.19, respectively.

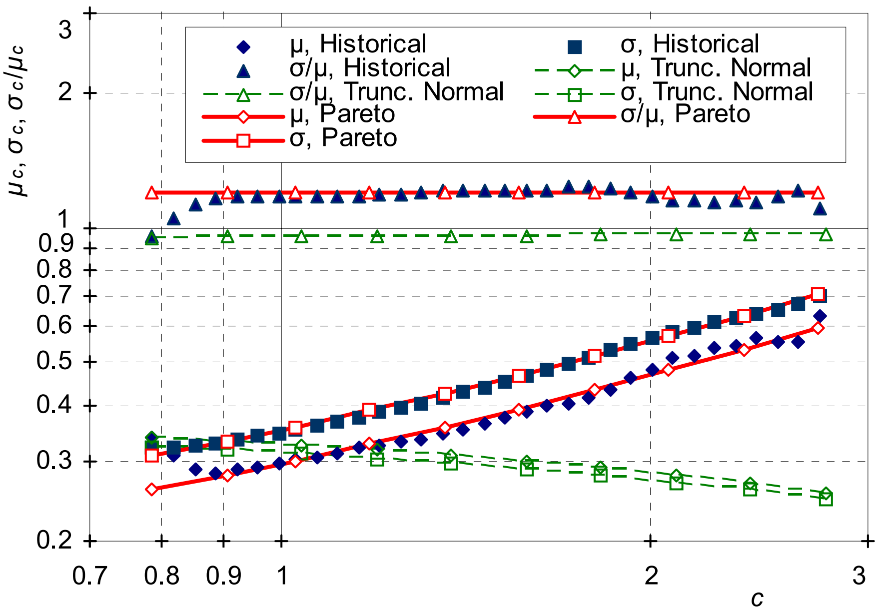

Figure 8. Conditional mean (μ_c), standard deviation (σ_c) and coefficient of variation (σ_c/μ_c), as functions of the threshold c, as calculated from the empirical 168-station rainfall sample of Figure 7, also compared to theoretical curves for the truncated normal and Pareto distributions.

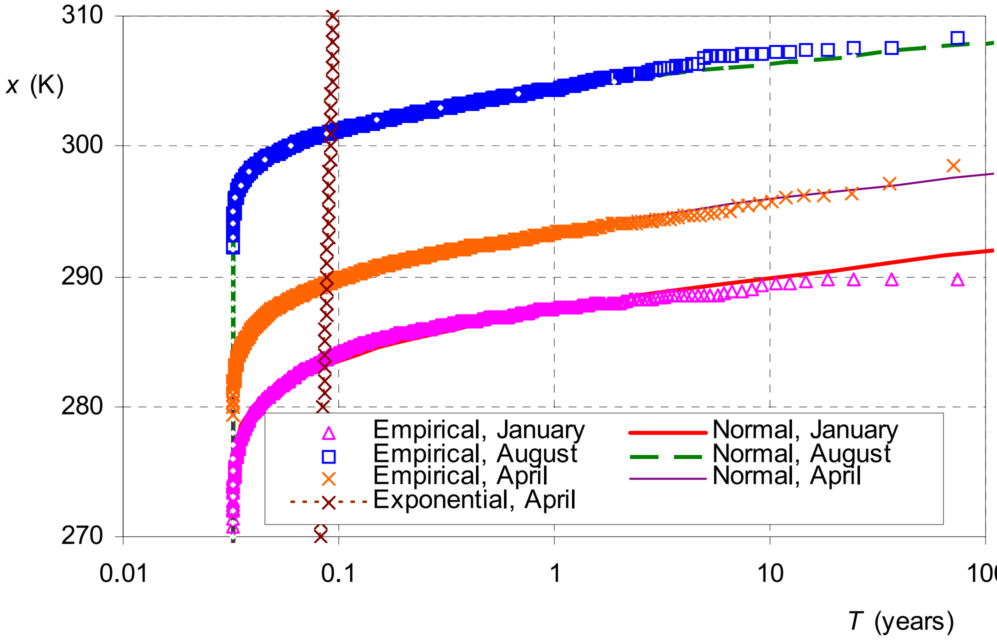

Figure 9. Plot of historical mean daily temperature in Athens for three months of the year versus return period, in comparison to the normal and exponential distributions (the exponential distributions for months January and August are almost indistinguishable from that of April).

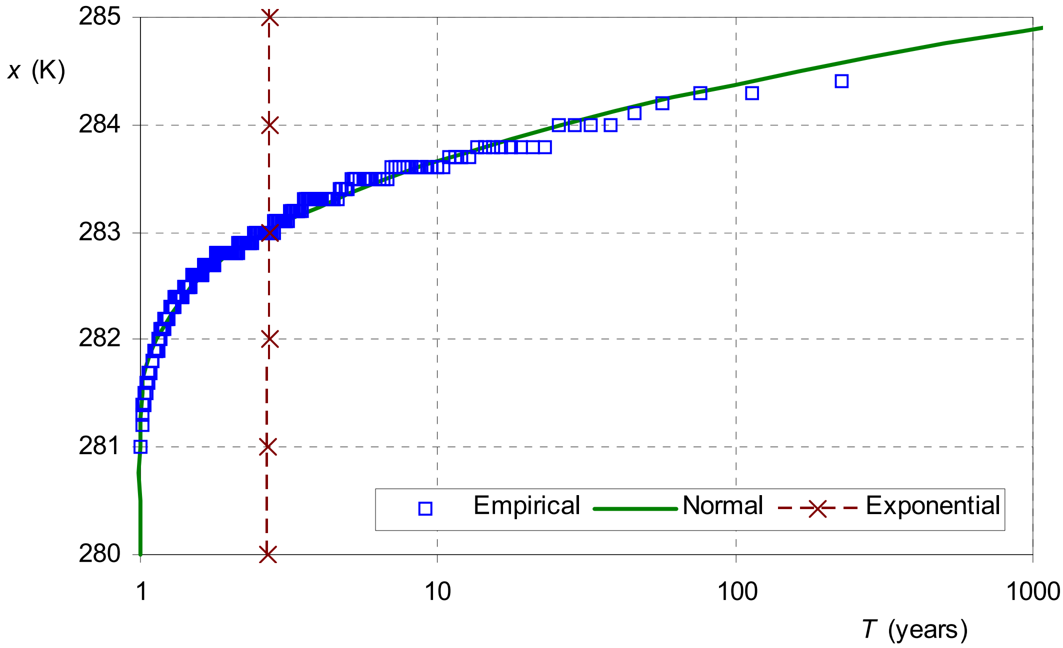

Figure 10. Plot of historical mean annual temperature in Geneva versus return period, in comparison to the normal and exponential distributions.

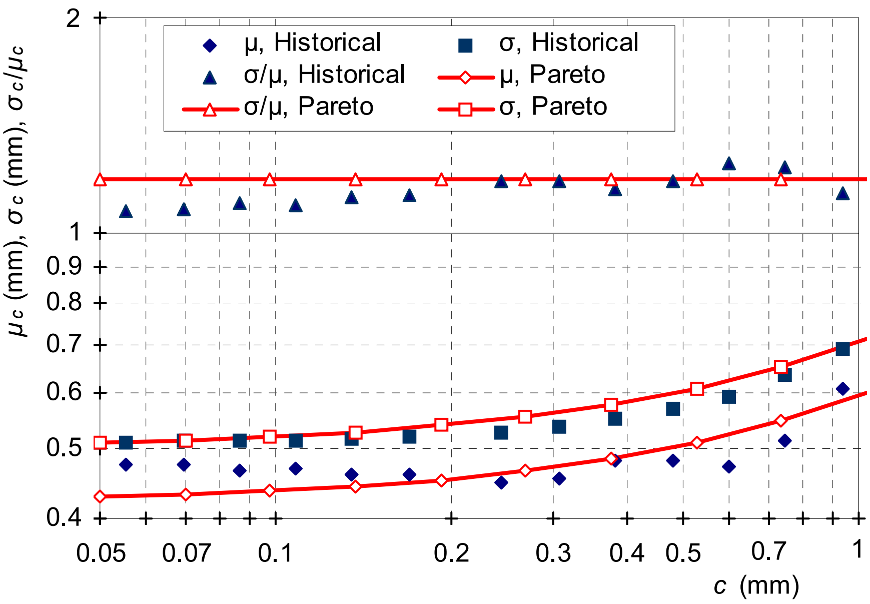

Figure 11. Conditional mean (μ_c), standard deviation (σ_c) and coefficient of variation (σ_c/μ_c), as functions of the threshold c, as calculated from the Boeoticos Kephisos daily runoff data, also compared to theoretical curves for the Pareto distribution (1 mm corresponds to 23.7 m^3/s).

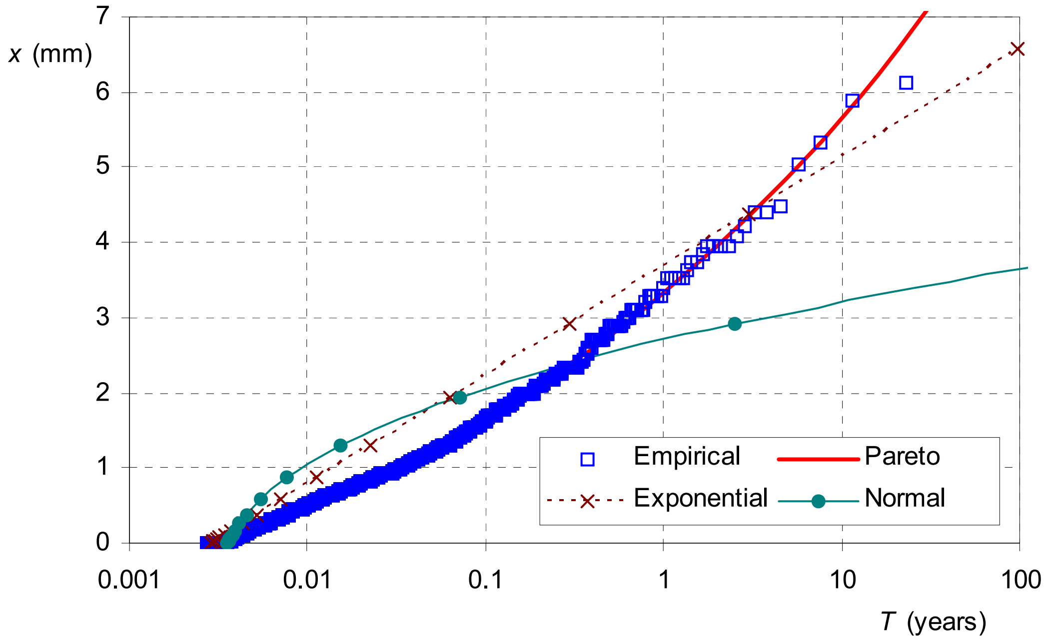

Figure 12. Plot of daily runoff of Boeoticos Kephisos versus return period, in comparison to Pareto, exponential and normal distributions (1 mm corresponds to 23.7 m^3/s; due to the high density of empirical points for return periods smaller than 1 year, the curve of the Pareto distribution is not distinguished because it lies just over the empirical points).