Uncertainty, entropy, scaling and hydrological stochastics

Additional information for the study of marginal distributional properties of hydrological processes

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 This report contains additional information for the paper *Uncertainty, entropy, scaling and hydrological stochastics, 1, Marginal distributional properties of hydrological processes and state scaling*, including mathematical derivations and complete sets of equations, complete descriptions of numerical investigations and additional case studies.

1. Derivation of ME distribution for non-extensive entropy

For a non-negative continuous random variable Tsallis entropy is written

$$\varphi_q = \frac{1 - \int_0^\infty [f(x)]^q \, dx}{q - 1}. \tag{A1}$$

The constraints considered for entropy maximization, apart from the non-negativity one are (a) basic property of $f(x)$,

$$\int_0^\infty f(x) \, dx = 1, \tag{A2}$$

(b) finite first moment,

$$E[X] = \int_0^\infty x \, f(x) \, dx = \mu_1 \equiv \mu, \tag{A3}$$

and (c) finite second moment,

$$E[X^2] = \int_0^\infty x^2 f(x) \, dx = \mu_2 \tag{A4}$$

By virtue of (A2), (A1) can be written alternatively in the form

$$\varphi_q = \int_0^\infty f(x)\, \frac{1 - [f(x)]^{q-1}}{q-1}\, dx \tag{A5}$$

To maximize φ_q assuming $q > 0$, constraints (A2), (A3) and (A4) are incorporated into (A5) with appropriate Lagrange multipliers λ'_0, λ'_1 and λ'_2, so that

$$\varphi_q = \int_0^\infty f(x)\, \frac{1 - [f(x)]^{q-1}}{q-1}\, dx - \lambda'_0 \left(\int_0^\infty f(x) dx - 1 \right) - \lambda'_1 \left(\int_0^\infty x\, f(x) dx - \mu_1 \right) - \lambda'_2 \left(\int_0^\infty x^2\, f(x) dx - \mu_2 \right) \tag{A6}$$

or

$$\varphi_q = \int_0^\infty f(x) \left\{ \frac{1 - [f(x)]^{q-1}}{q-1} - \lambda'_0 - \lambda'_1 x - \lambda'_2 x^2 \right\} dx + \lambda'_0 + \lambda'_1 \mu_1 + \lambda'_2 \mu_2 \tag{A7}$$

Now, it can be observed that if the bracketed term in (A7) is by identity zero, then functional variation with respect to the unknown $f(x)$ (i.e., $\partial\varphi_q / \partial f(x)$) will be zero, which ensures achieving the maximum. Thus, equating the bracketed term to zero, after algebraic manipulations, it is obtained,

$$f(x) = [1 + (1 - q)\, (\lambda'_0 + \lambda'_1 x + \lambda'_2 x^2)]^{-1/(1-q)} \tag{A8}$$

and

$$\varphi_q = \lambda'_0 + \lambda'_1 \mu_1 + \lambda'_2 \mu_2 \tag{A9}$$

Setting $\kappa = (1 - q)/q$ and $\lambda_i = \lambda'_i\, q$, (A8) and (A9) can be written respectively as

$$f(x) = [1 + \kappa\, (\lambda_0 + \lambda_1 x + \lambda_2 x^2)]^{-1-1/\kappa} \tag{A10}$$

and

$$\varphi_q = (\kappa + 1)\, (\lambda_0 + \lambda_1 \mu_1 + \lambda_2 \mu_2) \tag{A11}$$

2. Handling of the truncated normal distribution

The equations required for the handling of the truncated normal distribution are gathered in Table A1.

Table A1. Equations of the truncated normal distribution**.**

Description	Equation	Reference
Density function	$f(x) = \exp(-\lambda_0 - \lambda_1 x - \lambda_2 x^2)$	(T1.1)
Constraints	$x \geq 0, \lambda_2 \geq 0$	(T1.2)
Distribution function	$F(x) = 1 - \dfrac{1 - \mathrm{erf}(\xi + \sqrt{\lambda_2}\, x)}{1 - \mathrm{erf}(\xi)}, \quad \xi := \dfrac{\lambda_1}{2\sqrt{\lambda_2}}$	(T1.3)
First moment	$\mu_1 = \dfrac{\exp(-\xi^2)}{\sqrt{\pi\, \lambda_2}\, [1 - \mathrm{erf}(\xi)]} - \dfrac{\xi}{\sqrt{\lambda_2}}$	(T1.4)
Second moment	$\mu_2 = \dfrac{1 + 2\xi^2}{2\lambda_2} - \dfrac{\xi \exp(-\xi^2)}{\sqrt{\pi\, \lambda_2}\, [1 - \mathrm{erf}(\xi)]}$	(T1.5)
Entropy	$\varphi = \lambda_0 + \lambda_1 \mu_1 + \lambda_2 \mu_2$	(T1.6)
Determination of constant λ_0	$\dfrac{\sqrt{\pi} \exp(\xi^2 - \lambda_0)\, [1 - \mathrm{erf}(\xi)]}{2\sqrt{\lambda_2}} = 1$	(T1.7)
Determination of λ_2	From (T1.4)	
Determination of ξ	$\dfrac{\mu_2}{\mu_1^2} = \dfrac{\sqrt{\pi} \exp(\xi^2)\, [1 - \mathrm{erf}(\xi)]\, \{\sqrt{\pi}\, (1 + 2\xi^2) \exp(\xi^2)\, [1 - \mathrm{erf}(\xi)] - 2\xi\}}{2\{\sqrt{\pi}\, \xi \exp(\xi^2)\, [1 - \mathrm{erf}(\xi)] - 1\}^2}$	(T1.8)
Definition of $\mathrm{erf}(z)$ [(a)]	$\mathrm{erf}(z) := \dfrac{2}{\sqrt{\pi}} \int_0^z \exp(-t^2)\, dt$	(T1.9)

[(a)] erf( ) is the error function.

3. Handling of the Power-transformed Beta Prime distribution

The density of the Power-transformed Beta Prime (PBP) distribution is given by

$$f(x) = (1 + \kappa\, \lambda_0 + \kappa\, \lambda_1\, x^{\nu_1})^{-1 - 1/\kappa}\, x^{\nu_2 - 1} \tag{A12}$$

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 (A2). Table A2 gathers all required equations for the handling of this distribution. This distribution merges several exponential-type and power-type distributions, such as the exponential-type Gamma and Weibull, the power-type Pareto, and the Beta Prime distribution. Two additional limiting special cases, denoted as PBP/L1 and PBP/L2, which could be named respectively the enhanced exponential-type and the enhanced power-type distributions, emerge when the shape parameter κ takes its extreme values, i.e. when $\kappa \to 0$ and $\kappa \to \infty$, respectively. Equations for the handling of these cases are gathered in Table A3 and Table A4, respectively.

Table A2. Equations of the PBP distribution**.**

Description	Equation	Reference
Density function	$f(x) = (1 + \kappa\lambda_0 + \kappa\lambda_1 x^{\nu_1})^{-1-1/\kappa} x^{\nu_2 - 1}$	(T2.1)
Constraints	$x \geq 0, \kappa \geq 0, \lambda_0 \geq -1/\kappa, \lambda_1 \geq 0, \nu_1 \geq 0, 0 \leq \nu_2 \leq \nu_1(1 + 1/\kappa)$ For the existence of qth moment: $\nu_2 \leq \nu_1(1 + 1/\kappa) - q$	(T2. 2)
Distribution function [(a)]	$F(x) = \dfrac{B_{y(x)/[1+y(x)]}(\nu_2/\nu_1, 1 + 1/\kappa - \nu_2/\nu_1)}{B(\nu_2/\nu_1, 1 + 1/\kappa - \nu_2/\nu_1)}$, $y(x) := \xi x^{\nu_1}, \ \xi := \kappa\lambda_1 / (1 + \kappa\lambda_0)$	(T2.3)
qth moment	$\mu_q = \xi^{-q/\nu_1} \dfrac{B[(\nu_2+q)/\nu_1, 1 + 1/\kappa - (\nu_2+q)/\nu_1]}{B(\nu_2/\nu_1, 1 + 1/\kappa - \nu_2/\nu_1)}$	(T2.4)
Entropy	$\varphi = (1 + 1/\kappa)\,[\kappa + \ln(1 + \kappa\lambda_0) + \psi(1/\kappa) - \psi(1 + 1/\kappa - \nu_2/\nu_1)] +$ $[(\nu_2 - 1)/\nu_1]\,[\ln\xi - \psi(\nu_2/\nu_1) + \psi(1 + 1/\kappa - \nu_2/\nu_1)]$	(T2.5)
Determination of constant λ_0	$(1/\nu_1)\,(1 + \kappa\lambda_0)^{-1-1/\kappa} \xi^{-\nu_2/\nu_1} B(\nu_2/\nu_1, 1 + 1/\kappa - \nu_2/\nu_1) = 1$	(T2.6)
Determination of ξ	From (T2.4) for $q = 1$	
Relationship of shape parameters	$\dfrac{\mu_2}{\mu_1^2} = \dfrac{B[(\nu_2+2)/\nu_1, 1 + 1/\kappa - (\nu_2+2)/\nu_1]\, B(\nu_2/\nu_1, 1 + 1/\kappa - \nu_2/\nu_1)}{B^2[(\nu_2+1)/\nu_1, 1 + 1/\kappa - (\nu_2+1)/\nu_1]}$	(T2.7)
Definition of $B(\theta, \tau)$ and $B_z(\theta, \tau)$ [(b)]	$B(\theta, \tau) := \int_0^1 t^{\theta-1}(1-t)^{\tau-1}\,dt, B_z(\theta, \tau) := \int_0^z t^{\theta-1}(1-t)^{\tau-1}\,dt$	(T2.8)
Definition of $\psi(\theta)$ and $\Gamma(\theta)$ [(c)]	$\psi(\theta) := \Gamma'(\theta) / \Gamma(\theta), \quad \Gamma(\theta) := \int_0^\infty t^{\theta-1} e^{-t}\,dt$	(T2.9)

Notes: [(a)] The random variable $y(X)$ in (T2.3) follows the beta prime distribution (also known as beta of the second kind or inverted beta). [(b)] $B(\theta, \tau)$ and $B_z(\theta, \tau)$ are respectively the complete and incomplete beta function. [(c)] $\psi(\theta)$ and $\Gamma(\theta)$ are respectively the digamma and the gamma function.

Table A3. Equations of the special case PBP/L1 of the PBP distribution.

Description	Equation	Reference
Density function	$f(x) = \exp(-\lambda_0 - \lambda_1 x^{v_1})\, x^{v_2 - 1}$	(T3.1)
Constraints	$x \geq 0,\ \lambda_1 \geq 0,\ v_1 \geq 0,\ v_2 \geq 0$	(T3.2)
Distribution function [(a)]	$F(x) = 1 - \dfrac{\Gamma_{y(x)}(v_2/v_1)}{\Gamma(v_2/v_1)},\quad y(x) := \lambda_1 x^{v_1}$	(T3.3)
qth moment	$\mu_q = \lambda_1^{-q/v_1} \dfrac{\Gamma[(v_2+q)/v_1]}{\Gamma(v_2/v_1)}$	(T3.4)
Entropy	$\varphi = \lambda_0 + v_2/v_1 + [(v_2 - 1)/v_1]\,[\ln \lambda_1 - \psi(v_2/v_1)]$	(T3.5)
Determination of constant λ_0	$(1/v_1) \exp(-\lambda_0)\, \lambda_1^{-v_2/v_1}\, \Gamma(v_2/v_1) = 1$	(T3.6)
Determination of λ_1	From (T3.4) for $q = 1$	
Relationship of shape parameters	$\dfrac{\mu_2}{\mu_1^2} = \dfrac{\Gamma[(v_2+2)/v_1]\,\Gamma(v_2/v_1)}{\Gamma^2[(v_2+1)/v_1]}$	(T3.7)
Definition of $\Gamma_z(\theta)$ [(b)]	$\Gamma_z(\theta) := \int_z^{\infty} t^{\theta - 1}\, e^{-t}\, dt$	(T3.8)

Notes: [(a)] The random variable $y(X)$ in (T3.3) follows the gamma distribution. [(b)] $\Gamma_z(\theta)$ is the incomplete gamma function whereas $\Gamma(\theta)$ is the gamma function defined in equation (T2.9).

Table A4. Equations of the special case PBP/L2 of the PBP distribution.

Description	Equation	Reference
Density function	$f(x) = \dfrac{x^{v_2 - 1}}{1 + \kappa_0 + \kappa_1 x^{v_1}}$	(T4.1)
Constraints	$x \geq 0,\ \kappa_0 \geq 0,\ \kappa_1 \geq 0,\ v_1 \geq 0,\ 0 \leq v_2 \leq v_1$ For the existence of qth moment: $v_2 \leq v_1 - q$	(T4.2)
Distribution function[(a)]	$F(x) = \dfrac{B_{y(x)/[1 + y(x)]}(v_2/v_1,\ 1 - v_2/v_1)}{B(v_2/v_1,\ 1 - v_2/v_1)},\ y(x) := \xi x^{v_1},\ \xi := \kappa_1 / (1 + \kappa_0)$	(T4.3)
qth moment	$\mu_q = \xi^{-q/v_1} \sin(\pi\, v_2/v_1) / \sin[\pi(v_2+q)/v_1]$	(T4.4)
Entropy	$\varphi = -\gamma + \ln(1 + \kappa_0) - \psi(1 - v_2/v_1) +$ $[(v_2 - 1)/v_1]\,[\ln \xi - \psi(v_2/v_1) + \psi(1 - v_2/v_1)]$	(T4.5)
Determination of constant κ_0	$\pi\, \xi^{-v_2/v_1} / [(1 + \kappa_0)\, v_1 \sin(\pi\, v_2/v_1)] = 1$	(T4.6)
Determination of ξ	From (T4.4) for $q = 1$	(T4.7)
Relationship of shape parameters	$\dfrac{\mu_2}{\mu_1^2} = \dfrac{\sin^2[\pi\,(v_2+1)/v_1]}{\sin[\pi(v_2+2)/v_1]\, \sin(\pi\, v_2/v_1)}$	(T4.8)

Notes: [(a)] The random variable $y(X)$ follows the beta prime distribution; $B(\theta, \tau)$ and $B_z(\theta, \tau)$ are respectively the complete and incomplete beta function defined in equation (T2.8).

4. Additional information for the numerical investigation of variables with high variation – Extensive entropy approach

To apply the PBP distribution assuming $\mu = 1$ and some $\sigma/\mu > 1$, it is observed that PBP has two free parameters; let them be κ and ν_1. If these are fixed to some values, the other parameters λ_1 and ν_2 and the constant λ_0 can be estimated from the equations (T2.7) and (T2.4) of Table A2 so that $\mu_1 = \mu$ and $\mu_2 = \mu^2[(\sigma/\mu)^2 + 1]$. Then, the extensive entropy can be calculated from equation (T2.5) of Table A2. An example of results of such calculations for $\sigma/\mu = 1.5$ is given graphically in Figure 1 in the form of entropy contours of the PBP distribution with respect to the variation of parameters κ and ν_1. It can be observed in this figure that (a) there is an infeasible area of values (κ, ν_1) for which no solution exists; (b) on the boundary of the infeasible area, a single solution (λ_1, ν_2) exists; (c) in all other cases, two solutions (λ_1, ν_2) exist, which correspond to different values of entropy φ. The two sub-areas, in which the solution with the lower or the higher value of ν_2 dominate, i.e. yield higher entropy, have been specified in Figure 1. Apart for entropy contours in the feasible area, the exponential-type ($\kappa = 0$) special cases Gamma, Weibull and PBP/L1, as well as the Pareto case have been marked in Figure 1. It is observed that the exponential-type cases yield low entropy, with the Gamma and the PBP/L1 being the cases with lowest and highest values of entropy among the three. The Pareto case yields entropy significantly higher than Gamma and Weibull. However, there are sub-areas in the feasible area yielding entropy even higher than that in the Pareto distribution. In general, for a specified κ, the maximum entropy appears at a point close to the boundary of the feasible area but not exactly on this boundary.

Similar behaviour was verified for other values $\sigma/\mu > 1$. Graphical comparisons of the resulting Pareto and PBP/L2 distributions for $\mu = 1$ and for values of $\sigma/\mu = 1$ (whence both distributions identify with the exponential distribution), 1.25, 2 and 5 are shown in Figure 2. The distributions are given as double logarithmic plots of the random variate versus return period. The following may be observed from this figure for the curves corresponding to $\kappa > 1$: (a) The linearity of plots of both Pareto and PBP/L2 distributions for high return periods (e.g. $T > 100$) indicates a power-type behaviour. (b) The Pareto distribution differs from the exponential distribution in the entire domain, but the PBP/L2 distribution is indistinguishable from the exponential distribution for low return periods. This complies with the initial discussion that, according to the ME principle the exponential distribution is the limiting case for the extensive entropy concept, so to achieve maximum entropy for $\sigma/\mu > 1$, the distribution must be very close to the exponential, except for very high return periods where a departure is

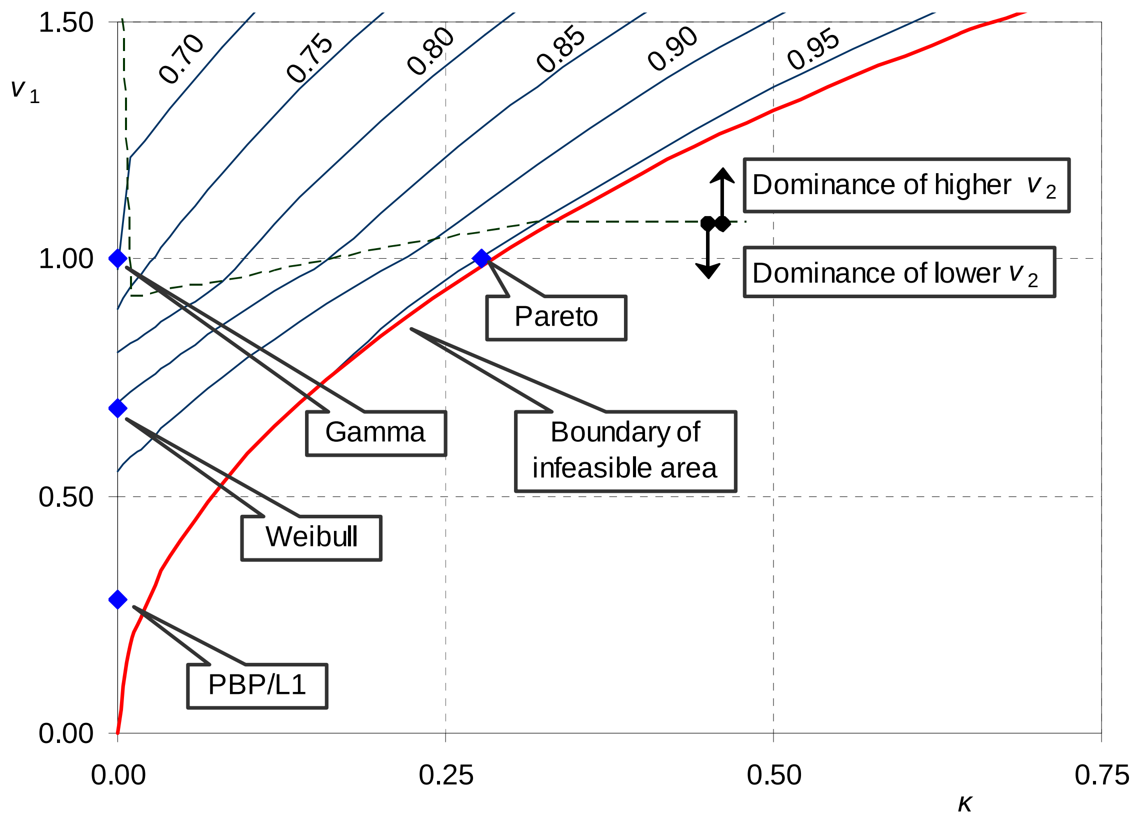

Figure 1. Entropy contours of the PBP distribution with $\mu = 1$ and $\sigma/\mu = 1.5$ with respect to the variation of parameters κ and v_1.

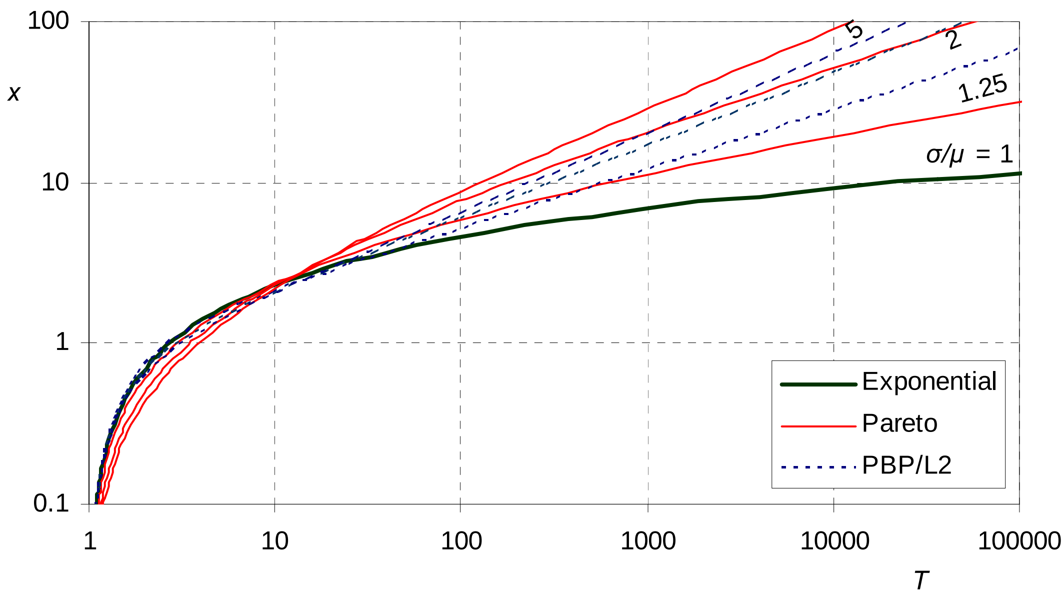

Figure 2. Graphical comparisons of Pareto and PBP/L2 distributions (shown as double logarithmic plots of the random variate versus return period) for $\mu = 1$ and for several values of $\sigma/\mu > 1$. For $\sigma/\mu = 1$ both distributions identify with the exponential distribution.

necessary to attain the required σ/μ. In addition, one must expect that there exist other distributions even closer to the exponential that would yield extensive standard entropy even higher than that of PBP/L2, i.e. even closer to the upper bound 1. Here, it should be noted that, in fact, such a distribution would not be "observable". For, if it were very close to the exponential distribution (except for very infrequent events with extremely large return period), then a typical sample of this distribution would yield an estimate of σ/μ very close to 1 and not close to the theoretical σ/μ.

5. Additional case studies

In this section, some case studies additional to those described in the main paper are presented, whose details are listed in Table 5.

Table 5. List of the case studies and respective data sets and their characteristics.

Case #	A1	A2	A3	A4	A5
Process	Rainfall	Rainfall	Rainfall	Runoff	Runoff
Time scale	Daily	Monthly	Monthly	Daily	Monthly
Station	Athens	Aliartos	Aliartos	Boeoticos Kephisos	Boeoticos Kephisos
Data period	1927-1996	1907-2003	1907-2003	1978-2003	1907-2003
Period of year	January	January	August	Year	January
Record length[(a)] (years)	70 (65)	96	96	25 (23)	96
Number of data values[(b)]	1998 (613)	96	96 (71)	8402	96
Units	mm	mm	mm	mm	mm
Mean	5.05	94.2	17	0.38 0.45[(d,e)]	25.7
Coefficient of variation	1.39	0.56	1.67	1.29 1.19[(e)]	0.52
Distribution maximizing entropy	Pareto	Truncated Normal	Pareto	Pareto	Truncated Normal
Maximized standard entropy	1.291	0.784	1.401	1.16	0.727

Notes: [(a)] In parentheses the equivalent years if missing data are not counted. [(b)] In parentheses the number of values without counting the zero values. [(d)] Mean minus threshold. [(e)] Value conditional on being greater than threshold.

In the main paper a case referring to the distribution of hourly rainfall in Athens, Greece, for the month of January (wettest month) is discussed. The distribution of the daily rainfall (nonzero depths) at the same station and month is depicted here in Figure 3. The coefficient of variation is 1.39 > 1 and the Pareto distribution is the closest to the empirical one.

Figure 4 depicts the distribution of monthly rainfall depth in Aliartos, Greece, for the month of January (wettest month). The empirical coefficient of variation is 0.56 < 1; thus the ME distribution is the truncated normal. Indeed, the figure shows that this distribution is the closest to the empirical one. A similar plot is given in Figure 5 but for the month of August. August is the driest month and there are years with no rainfall in this month. As in the earlier cases, only nonzero rainfall depths were modelled. The coefficient of variation is 1.67 > 1 and thus the Pareto distribution is the ME distribution. Indeed, the figure shows that this is the closest to the empirical distribution.

In the main paper, the runoff of the Boeoticos Kephisos river was examined at the daily scale. Here, in addition, the monthly and annual scales are examined. Figure 6 and Figure 7 depict the distribution of runoff at respectively the monthly (January) and annual scales. As in the case of rainfall, the ME distribution, which is the truncated normal, fits well the empirical one except for the tail of distribution. As a result, the Gamma distribution seems to give a better approximation of the tails in this case. However, it must be anticipated that a model based on ME principle on multiple time scales, as described in the annual rainfall case, could give a better representation of the distribution tails.

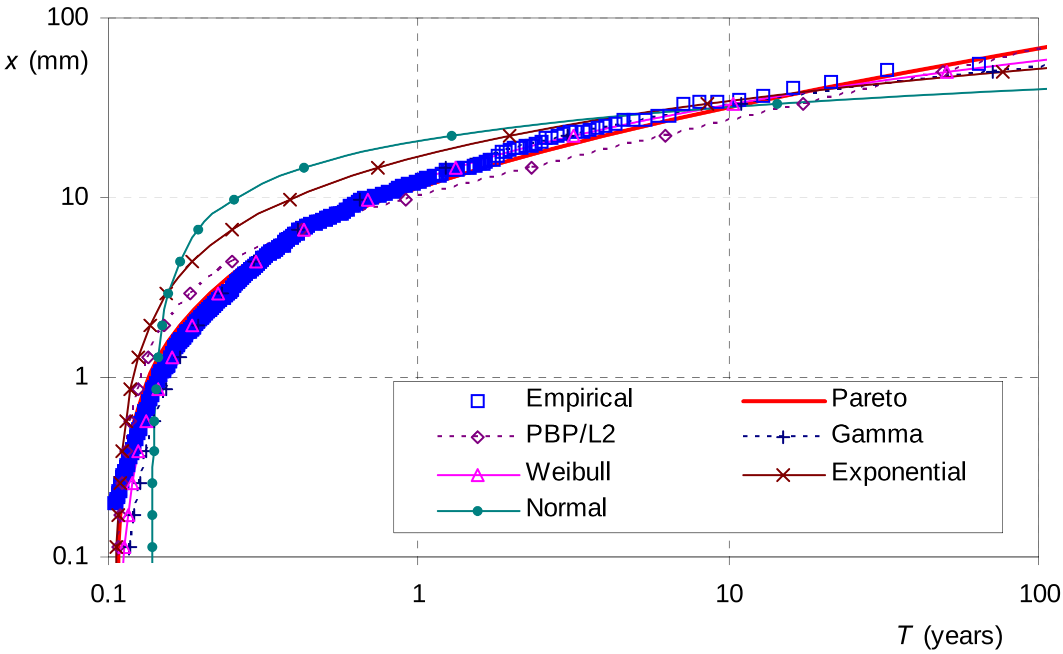

Figure 3. Plot of historical daily rainfall depth in Athens for the month of January versus return period, in comparison to several theoretical distributions.

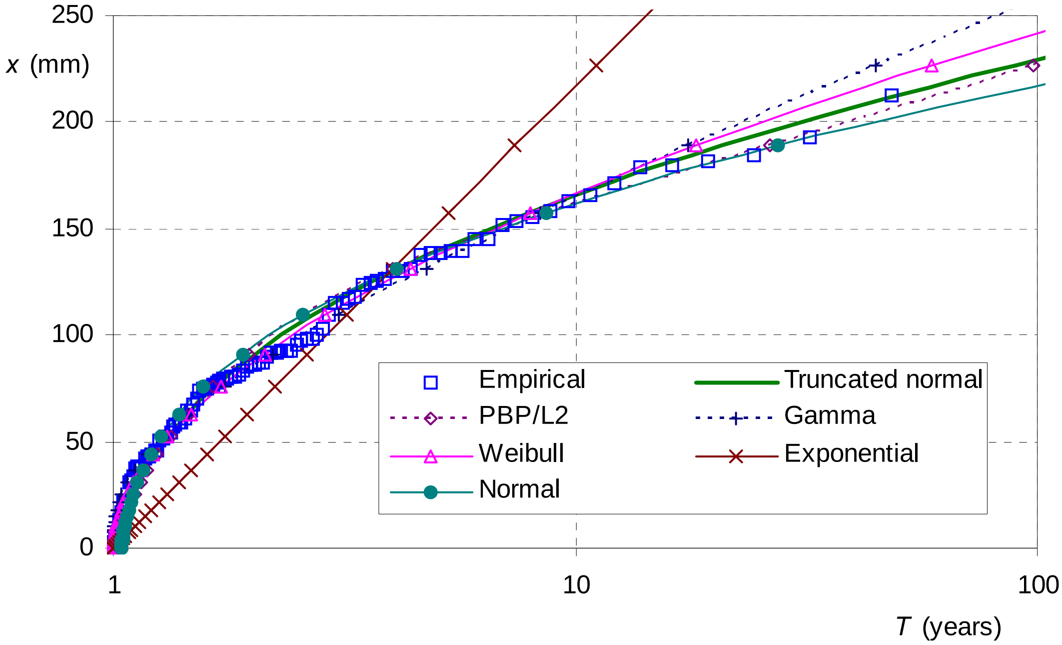

Figure 4. Plot of historical monthly rainfall depth in Aliartos for the month of January versus return period, in comparison to several theoretical distributions.

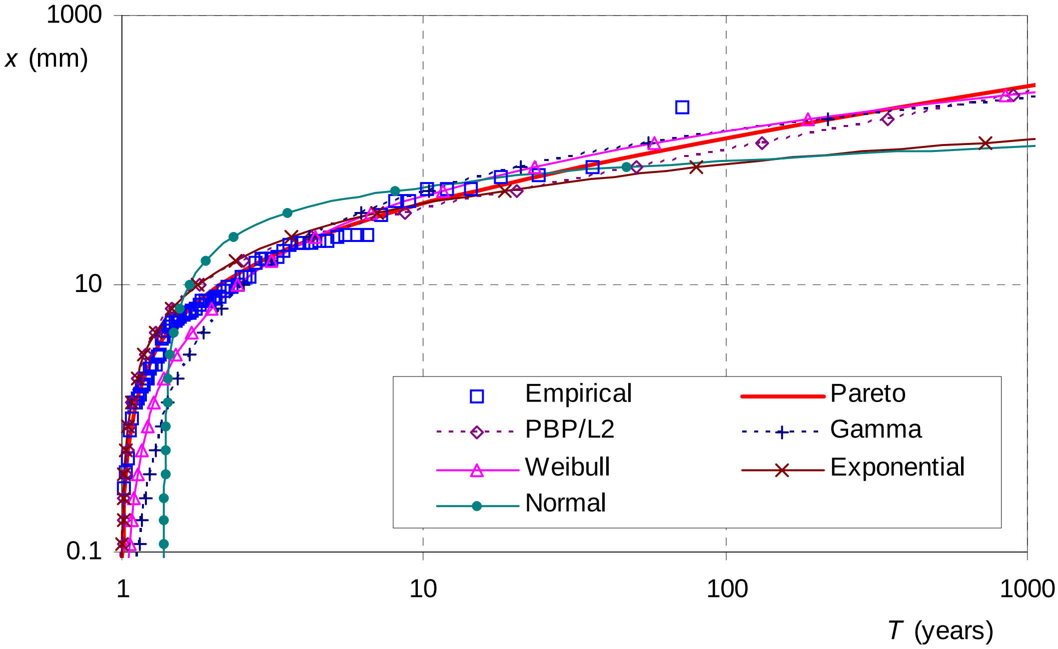

Figure 5. Plot of historical monthly rainfall depth in Aliartos for the month of August versus return period, in comparison to several theoretical distributions.

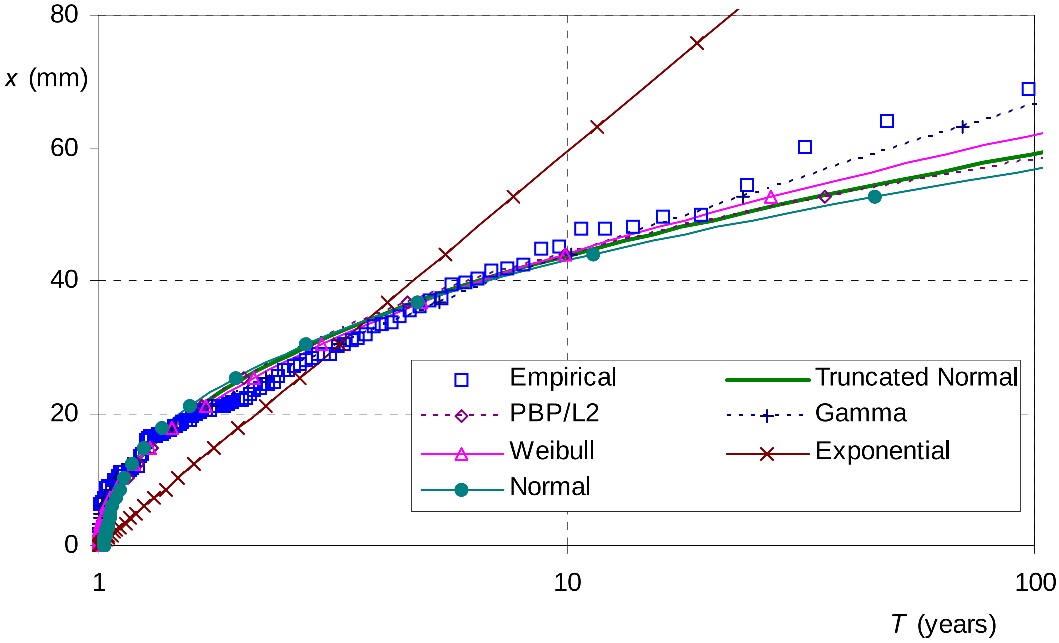

Figure 6. Plot of historical monthly runoff in Boeoticos Kephisos for the month of January versus return period, in comparison to several theoretical distributions.

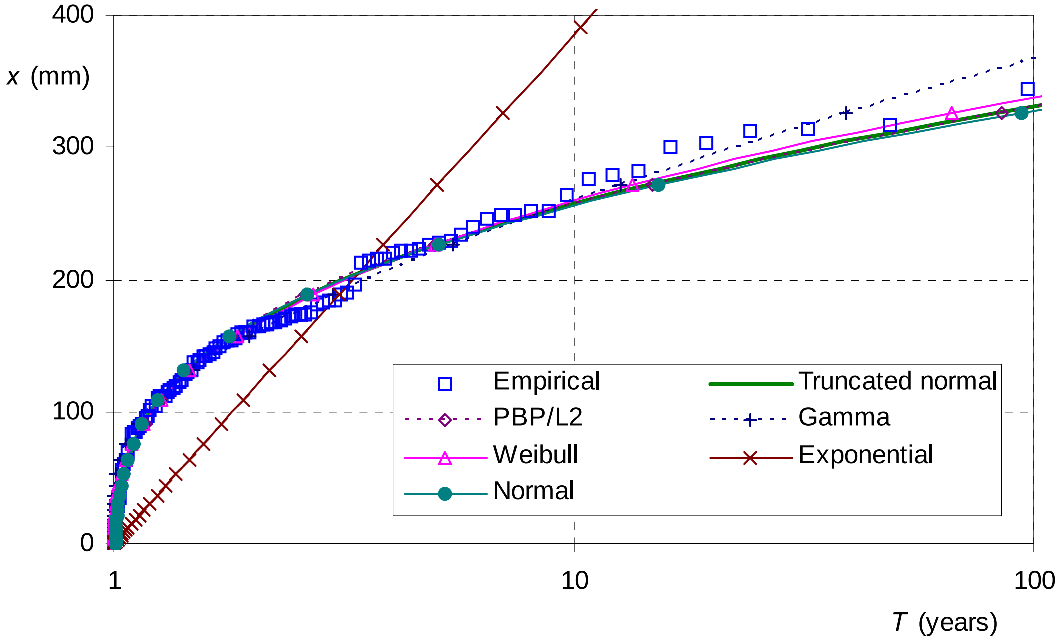

Figure 7. Plot of historical annual runoff in Boeoticos Kephisos versus return period, in comparison to several theoretical distributions.