

Entropy based derivation of probability distributions: A case study to daily rainfall

Simon Michael Papalexiou, Demetris Koutsoyiannis

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

Abstract

The principle of maximum entropy, along with empirical considerations, can provide consistent basis for constructing a consistent probability distribution model for highly varying geophysical processes. Here we examine the potential of using this principle with the Boltzmann-Gibbs-Shannon entropy definition in the probabilistic modelling of rainfall in different areas worldwide. We define and theoretically justify specific simple and general entropy maximization constraints which lead to two flexible distributions, i.e., the three-parameter Generalized Gamma (GG) and the four-parameter Generalized Beta of the second kind (GB2), with the former being a particular limiting case of the latter. We test the theoretical results in 11 519 daily rainfall records across the globe. The GB2 distribution seems to be able to describe all empirical records while two of its specific three-parameter cases, the GG and the Burr Type XII distributions perform very well by describing the 97.6% and 87.7% of the empirical records, respectively.

Keywords: maximum entropy; daily rainfall; Generalized Gamma distribution; Generalized Beta distribution; Burr distribution

1. Introduction

Even though long-term predictions of rainfall are not possible in deterministic terms (e.g., weather forecasts are skillful for no more than a week ahead), in probabilistic terms it is possible to assign a stochastic model or a probabilistic law and to any rainfall amount assign a return period or a probability of exceedence. Actually, most infrastructures affected by rainfall and flood are designed this way. Rainfall is generally characterized as an intermittent stochastic process (for fine timescales), with a mixed-type marginal distribution, partly discrete and partly continuous. The discrete part is concentrated at zero and defines the probability dry, while the rest is continuously spread over the positive real axis and determines the nonzero rainfall distribution. The discrete part of the rainfall distribution can be easily estimated as the ratio of the number of dry days to total number of days. On the contrary, the continuous part of the distribution cannot be easily assessed.

Rainfall is usually studied in many different timescales, e.g., from sub-hourly to yearly, yet, the daily timescale is one of the most convenient and important in hydrological design. Specifically, it is the smallest timescale for which thousands of records exist with some of them being more than a century long. Nevertheless, and although daily rainfall has been extensively studied over the years, a search in the literature reveals that a universally accepted model for the wet-day daily rainfall distribution does not exist. On the contrary, many distributions have been proposed in specific studies for specific locations of the world including, e.g., the two-parameter Gamma, which is probably the prevailing model, the two- and three-parameter Lognormal, the Generalized Logistic, the Pearson Type III, the Pareto and the Generalized Pareto, the three- and four-parameter Kappa distributions, and many more.

The common method to construct an appropriate probability distribution model for describing one or more samples is to try a variety of different models and choose the best fitted using a particular mathematical norm, e.g., a least square error or a likelihood norm. Nevertheless, this approach is rather naïve and laborious; first, there are (at least theoretically) infinitely many different models to try, and second, this method does not offer any theoretical

justification for the final choice, thus making it an ad hoc empirical choice. This practice explains why so numerous models have been proposed. Here, we use the principle of maximum entropy as a solid theoretical background for constructing an appropriate probability distribution for rainfall and for geophysical processes in general. Our study is both theoretical, in terms of using the principle of maximum entropy and seeking the appropriate constraints for entropy maximization, and also empirical, since we test the theoretical results using 11 519 daily rainfall samples across the world. Our main target is to assess whether a single generalized model could be appropriate for all rainfall records worldwide.

2. The entropic framework

2.1 Entropy measures

The concept of entropy dates back to the works of Rudolf Clausius in 1850, yet, it was Ludwig Boltzmann around 1870 who gave entropy a statistical meaning and related it to statistical mechanics. The concept of entropy was advanced later in the works of J. Willard Gibbs in thermodynamics and Von Neumann in quantum mechanics, and was reintroduced in information theory by Claude Shannon [17] in 1948, who showed that entropy is a purely probabilistic concept, a measure of the uncertainty related to a random variable (RV).

The most famous and well justified measure of entropy for continuous RVs, is the Boltzmann-Gibbs-Shannon (BGS) entropy, which for a non-negative RV X is

$$S_X = -\int_0^{\infty} f_X(x) \ln f_X(x) dx \quad (1)$$

where $f_X(x)$ is the probability density function of X . The BGS entropy is not the only entropy measure. A search in the literature reveals that more than twenty different entropy measures have been proposed, mainly generalizations of BGS entropy (for a summary of entropy measures see [2]). Among those measures, it is worth noting the Rényi entropy, introduced by the Hungarian mathematician Alfréd Rényi in 1961, which have been used in many different disciplines, e.g., ecology and statistics. It is also worth noting another entropy measure that has gained much popularity in the last decade, the Havrda-Charvat-Tsallis

(HTC) entropy. It was initially proposed by Havrda and Charvat [4], and was reintroduced and applied to physics by Tsallis [21]. Apart from its use in physics, the HTC entropy has also been used more recently in hydrology [11,18] as it gives rise to power-type distributions. The HTC entropy is a generalization of the BGS entropy given by

$$S_X(q) = \frac{1 - \int_0^{\infty} [f_X(x)]^q dx}{q-1} \quad (2)$$

It is easy to verify that for $q = 1$ it becomes identical to the BGS entropy.

2.2 The principle of maximum entropy

The principle of maximum entropy was established, as a tool for inference under uncertainty, by Edwin Jaynes [6,7]. In essence, the principle of maximum entropy relies in finding the most suitable probability distribution under the available information. As Jaynes [6] expressed it, the resulted maximum entropy distribution “is the least biased estimate possible on the given information; i.e., it is maximally noncommittal with regard to missing information”.

In a mathematical frame, the given information used in the principle of maximum entropy, is expressed as a set of constraints formed as expectations of functions $g_j(\cdot)$ of X , i.e.,

$$E[g_j(X)] = \int_0^{\infty} g_j(x) f_X(x) dx = c_j \quad j = 1, \dots, n \quad (3)$$

The resulting maximum entropy distributions emerge by maximizing the selected form of entropy with constraints (3), and with the obvious additional constraint

$$\int_0^{\infty} f_X(x) dx = 1 \quad (4)$$

The maximization is accomplished by using calculus of variation and the method of Lagrange multipliers. Particularity, the general solution of the maximum entropy distributions resulting from the maximization of BGS entropy and the HCT entropy, assuming arbitrary constraints are, respectively,

$$f_X(x) = \exp \left[-\lambda_0 - \sum_{j=1}^n \lambda_j g_j(x) \right] \quad (5)$$

$$f_X(x) = \left\{ 1 + (1-q) \left[\lambda_0 + \sum_{j=1}^n \lambda_j g_j(x) \right] \right\}^{-\frac{1}{1-q}} \quad (6)$$

where λ_j , with $j = 1, \dots, n$ are the Lagrange multipliers linked to the constraints (3) and λ_0 is the multiplier linked to the constraint (4), i.e., λ_0 guarantees the legitimacy of the distribution.

2.3 Justification of the constraints

It becomes clear from the above discourse that the resulting maximum entropy distribution is uniquely defined by the choice of the imposed constraints. This implies that this choice is the most important and determinative part of the method. Constraints express our state of knowledge concerning a RV and should summarize all the available information from observations or from theoretical considerations. Nevertheless, choosing constraints is not trivial; they are introduced as expectations of RV functions without any intrinsic limitation on the form of those functions.

So, how should we choose the appropriate constraints among an infinite number of choices? In classical statistical mechanics, these constraints are imposed by physical principles such as the mass, momentum and energy conservation. However, in complex geophysical processes, these principles cannot help. In geophysical processes, the standard procedure to assign a probability law is to study the available observations and infer the underlying distribution without entropy considerations. However, whatever we infer in this way, is in fact based on a small portion of the past (the available record), which may (or may not) change in the future. Nevertheless, we can reasonably assume that some RV features may be more likely to be approximately preserved in the future than others, e.g., coarse features like the mean and the variance are less likely to change in the future [8] than finer features based on higher moments (e.g., it is well known that the kurtosis coefficient is extremely sensitive to observations and additional observations may radically alter it). Therefore, as a first rule, constraints should be simple and express those features that are likely to be preserved in the future.

The previous rule is rather subjective in the sense that it is difficult to distinguish between simple and not simple constraints or to foresee what RV quantities will be preserved. Furthermore, the use of a particular set of “simple” constraints may lead to a distribution that is not supported by the empirical data. Obviously, it is difficult to reject or verify the detailed shape features of a distribution based on a small sample which apparently does not provide the sufficient amount of information needed. Nonetheless, many geophysical processes, even if long records do not exist for particular regions, are extensively recorded worldwide e.g., thousands of stations record precipitation, temperature, etc. Thus, the study of this massive amount of information may lead in determining some important prior characteristics of the underlying distribution that should be preserved, e.g., a J- or bell-shaped distribution or a heavy- or light-tailed distribution. Therefore, constraints should be chosen not only based on simplicity, but also on the appropriateness of the resulting distribution given the empirical evidence.

Commonly used constraints in maximizing entropy assume known mean and variance, i.e., known first and second moments, which are clearly two very simple constraints. Particularly, entropy maximization assuming known first two moments leads: (a) to the celebrated normal distribution in the BGS entropy case, or, to the truncated normal if the mandatory constraint of non-negativity for geophysical processes is imposed, and (b) to a symmetric bell-shaped distribution with power-type tails in the HCT entropy case, or, its truncated version for a non-negative RV. The distribution arising in the HCT case for zero mean is now known as the Tsallis distribution. For non-zero mean the resulting distribution is the Pearson type VII introduced by Pearson in 1916, whose special case is the Tsallis distribution. Both these distributions are symmetric bell-shaped, in which asymmetry can only emerge by truncation at zero. As a consequence, those distributions may likely fail to describe sufficiently many geophysical processes that exhibit a rich pattern of asymmetries (e.g., it is well known that the rainfall in small time scales is heavily skewed and likely heavy tailed).

Accordingly, in this study we aim to define some simple and general constraints alternative to those of the first two moments that lead to suitable probability distributions for geophysical processes, particularly for rainfall. Additionally, we aim to only use the BGS

entropy, which is theoretically justified and widely accepted, avoiding the use of generalized entropy measures.

The mean is one of the most commonly used constraints, as it is a classical measure of central tendency. Another useful measure of central tendency, exhibiting the convenient property for geophysical processes to be defined only for positive values, is the geometric mean μ_G . An estimate of this, from a sample of size n , is given by

$$\hat{\mu}_G = \left(\prod_{i=1}^n x_i \right)^{1/n} = \exp\left(\frac{1}{n} \sum_{i=1}^n \ln x_i \right) = \exp(\overline{\ln x}) \quad (7)$$

where the overbar stands for the sample average. The sample geometric mean (also referred as a constraint in [9]) is smaller than the arithmetic mean. Intuitively, this leads us to formulate the following constraint for entropy maximization

$$E[\ln X] = \ln \mu_G \quad (8)$$

The expectation of $\ln X$, apart from its relationship to the geometric mean and its simplicity, makes an essential constraint for positively skewed RVs. To clarify, samples drawn from positively skewed distributions and, even more so, drawn from heavy-tailed distributions, exhibit values located on the right area very far from the mean value; in a sense, those values act like outliers and consequently strongly influence the sample moments, especially those of higher order. Therefore, it is not rational to assume that sample moments, especially based on samples drawn from heavy-tailed distributions, are likely to be preserved. On the contrary, the function $\ln x$ applied to this kind of samples eliminates the influence of those “extreme” values and offers a very robust measure that is more likely to be preserved than the estimated sample moments. Essentially for this reason, the logarithmic transformation is probably the most common transformation used in hydrology as it tends to normalize positively skewed data.

As stated above, the link of the mean and variance with the physical principles of momentum and energy conservation is invalid in geophysical processes. For example, the mean of the rainfall is not its momentum and its variance it is not its energy. Even in these

processes, mean and variance (as measures of central tendency and dispersion) provide useful information, which can at least explain general behaviours and shapes of probability density functions [11]. However, this information is good only for explanatory purposes and does not enable detailed and accurate modelling. For, there do not exist theoretical arguments (apart from simplicity and conceptual meaning as measures of central tendency and dispersion) which to favor mean and variance against, e.g., fractional moments of small order or even negative. For example, if the second moment is likely to be preserved, then one could think that the square root moment is more likely to be preserved as it is more robust in outliers. Additionally, we can relate low order fractional moments with the $\ln x$ function, as it is well known that

$$\lim_{q \rightarrow 0} \frac{x^q - 1}{q} = \ln x \quad (9)$$

Thus, we may say that the function x^q for small values of q behaves similar to $\ln x$, thus exhibiting properties similar to those of the logarithmic function described above.

Based on this reasoning we deem that, instead of choosing the order of moments a priori, it is better to let the order unspecified, so that any value can be a posteriori chosen, including small fractional values. This leads in imposing as a constraint any moment m_q of order q , i.e.,

$$m_q = E[X^q] = \int_0^{\infty} x^q f_X(x) dx \quad (10)$$

One reason that many entropy generalizations have emerged was to explain many empirically detected deviations from exponential type distributions that arise from the BGS entropy using standard moment constraints. Yet, generalized entropy measures have been criticized for lacking theoretical consistency and for being arbitrary, a reasonable argument considering the large number of entropy generalizations available in the literature. Here, instead of using generalized entropy measures that may result in power-law distributions, we generalize the important notion of moments inspired by the limiting definition of the exponential function, i.e., $\exp(x^q) = \lim_{p \rightarrow 0} (1 + p x^q)^{1/p}$. We first define the function x_p^q as

$$x_p^q := \ln(1 + p x^q) / p \quad (11)$$

which for $p = 0$ becomes the familiar power function x^q as $x_0^q = \lim_{p \rightarrow 0} \ln(1 + p x^q) / p = x^q$. Thus, we can define a generalization of the classical moments, for which we use the name p -moments, by

$$m_p^q = E[X_p^q] = \frac{1}{p} \int_0^\infty \ln(1 + p x^q) f_X(x) dx \quad (12)$$

Arguably, this generalization is arbitrary and many other moment generalizations can be (and in fact are) constructed. Nonetheless, we believe that there is a rationale that supports the use of p -moments, which can be summarized as follows: (a) if generalized entropy measures, considered by many as arbitrary, have been successfully used, then there is no reason to avoid using generalized moments; (b) maximization of the BGS entropy using p -moments leads, as will become apparent in the next section, to flexible power-type distributions (including the Pareto and Tsallis distributions for $q = 1$ and $q = 2$, respectively); (c) p -moments are simple and, for $p = 0$, become identical to the ordinary moments; and (d) they are based on the x_p^q function that exhibits all the desired properties, like those of the $\ln x$ function described above, and thus are suitable for positively skewed RVs; additionally, compared to $E[\ln X]$ they are always positive.

2.4 The resulting entropy distributions

Entropy optimization can be accomplished in many different combinations of the previously defined constraints; however, here we use two simple combinations of the aforementioned constraints based on the type and the generality of the distributions that emerge. We combine the $E[\ln X]$ constraint, first, with classical moments, and second, with p -moments, letting in both cases the moment order arbitrary.

In the first case, the maximization of the BGS entropy, given in (1), with constraints (8) and (10) results in the density function

$$f_X(x) = \exp(-\lambda_0 - \lambda_1 \ln x - \lambda_2 x^q) \quad (13)$$

which after algebraic manipulations and parameter renaming can be written as

$$f_x(x) = \frac{\gamma_2}{\beta \Gamma(\gamma_1 / \gamma_2)} \left(\frac{x}{\beta}\right)^{\gamma_1 - 1} \exp\left[-\left(\frac{x}{\beta}\right)^{\gamma_2}\right], \quad x \geq 0 \quad (14)$$

corresponding to the distribution function

$$F_x(x) = 1 - \Gamma\left[\frac{\gamma_1}{\gamma_2}, \left(\frac{x}{\beta}\right)^{\gamma_2}\right] / \Gamma\left(\frac{\gamma_1}{\gamma_2}\right) \quad (15)$$

where $\Gamma(\cdot)$ is the Gamma function and $\Gamma(\cdot, \cdot)$ is the upper incomplete Gamma function.

This distribution, commonly attributed to Stacy [19] appeared much earlier in the literature in the works of Amoroso around 1920 [10], and seems to have been rediscovered many times under different forms (see e.g., [10]). Here, we use a slightly different form from that proposed by Stacy. Essentially, it is a generalization of the Gamma distribution and will be denoted by $GG(\beta, \gamma_1, \gamma_2)$, or simply GG. It is a very flexible distribution that includes many other well-known distributions as particular cases, e.g., the Gamma, the Weibull, the Exponential, or even the Chi-squared distributions and others.

The distribution includes the scale parameter $\beta > 0$, and the shape parameters $\gamma_1 > 0$ and $\gamma_2 > 0$. The parameter γ_1 controls the behavior of the left tail, i.e., if $0 < \gamma_1 < 1$ the density function is J-shaped and for $x \rightarrow 0$, $f_x(x) \rightarrow \infty$; if $\gamma_1 > 1$ the density function is bell-shaped and mainly positively skewed; yet, for certain values of γ_1 and γ_2 it can be symmetric or even negatively skewed, and for $x = 0$, $f_x(x) = 0$; finally, for $\gamma_1 = 1$ the distribution degenerates to a generalized exponential function and for $x = 0$, $f_x(0) < \infty$. The parameter γ_2 is very important as for fixed γ_1 it controls the behavior of the right tail, i.e., it determines the frequency and the magnitude of the extreme events. Generally and loosely speaking, for $\gamma_2 < 1$ the distribution can be characterized as sub-exponential or heavy-tailed, and for $\gamma_2 > 1$ as hyper-exponential or light-tailed (for a classification of distribution tails see [3]).

Notably, the distribution is also valid if the shape parameters are simultaneously negative (a generalized inverse Gamma distribution); however, the distribution loses some important shape characteristics and seems not suitable for geophysical RV like rainfall. Thus,

here the distribution is only considered for positive shape parameters.

In the second case, the maximization of the BGS entropy with constraints (8) and (12) results in the density function

$$f_X(x) = \exp\left[-\lambda_0 - \lambda_1 \ln x - \lambda_2 \ln(1 + p x^q) / p\right] \quad (16)$$

which after algebraic manipulations and parameter renaming can be written as

$$f_X(x) = \frac{\gamma_3}{\beta B(\gamma_1, \gamma_2)} \left(\frac{x}{\beta}\right)^{\gamma_1 \gamma_3 - 1} \left[1 + \left(\frac{x}{\beta}\right)^{\gamma_3}\right]^{-(\gamma_1 + \gamma_2)}, \quad x \geq 0 \quad (17)$$

corresponding to the distribution function

$$F_X(x) = B_z(\gamma_1, \gamma_2) / B(\gamma_1, \gamma_2), \quad \text{where } z = \left[1 + (x/\beta)^{-\gamma_3}\right]^{-1} \quad (18)$$

where $B(\cdot, \cdot)$ is the Beta function and $B_z(\cdot, \cdot)$ is the incomplete Beta function.

This distribution has not been formed earlier on a similar rationale, yet, a search in the literature reveals that it has been rediscovered many times under different names and parameterizations. It is most commonly known as the Generalized Beta of the second kind—hereafter denoted as $GB2(\beta, \gamma_1, \gamma_2, \gamma_3)$, or simply GB2. It seems that Milke and Johnson [14] were the first that formed this distribution, and proposed it for describing hydrological and meteorological variables. It has also been used in different disciplines, e.g., McDonald [12] used the GB2 as an income distribution. Nevertheless, the distribution can be considered as a simple generalization of many well-known and much earlier introduced distributions, e.g., the F -distribution or the Pearson type VI of the celebrated Pearson system.

The GB2 distribution is a very flexible four-parameter distribution with $\beta > 0$ being the scale parameter, and $\gamma_1 > 0$, $\gamma_2 > 0$ and $\gamma_3 > 0$ being the three shape parameters, allowing the distribution to form very many different shapes. The GB2 distribution includes as special or limiting cases many of the well-known distributions, e.g., the Beta of the second kind, the Pareto type II, the Loglogistic, the Burr type XII, even the Generalized Gamma (for a complete account see [10,13]).

Obviously, the flexibility of the GB2 distribution makes it a good model for describing

rainfall—we have already used the GB2, under the name JH distribution, to describe the rainfall in a large range of timescales [15] and to construct theoretically consistent IDF curves [16]. Nonetheless, as a general rule based on the principle of parsimony, a three-parameter model is preferable than a four-parameter model, provided that the simpler model describes the data adequately. Additionally, it is not reasonable to compare the performance of the GG distribution, which is a three-parameter model, with GB2, which is a four-parameter model. Thus, a simpler form of the GB2 distribution is selected based on its flexibility and its simple analytical expression of the distribution function, and consequently, of the quantile function.

A simple three-parameter form of GB2 is derived by setting $\gamma_1 = 1$ in equation (17). By renaming the parameters and after algebraic manipulations we obtain a distribution known as the Burr type XII [1] (denoted hereafter as BurrXII), which was introduced by Burr in 1942 in the framework of a distribution system similar to Pearson's. Its probability density function is

$$f_X(x) = \frac{1}{\beta} \left(\frac{x}{\beta} \right)^{\gamma_1 - 1} \left[1 + \gamma_2 \left(\frac{x}{\beta} \right)^{\gamma_1} \right]^{-\frac{1}{\gamma_2 \gamma_1}}, \quad x \geq 0 \quad (19)$$

and its distribution function is

$$F_X(x) = 1 - \left[1 + \gamma_2 \left(\frac{x}{\beta} \right)^{\gamma_1} \right]^{-\frac{1}{\gamma_2 \gamma_1}} \quad (20)$$

The BurrXII distribution is a flexible power-type distribution that comprises the scale parameter $\beta > 0$ and the shape parameters $\gamma_1 > 0$ and $\gamma_2 \geq 0$.

The form of the BurrXII distribution we use here is not the one found in the literature (see e.g. [20]). We preferred the expression (19) because it is suggestive of a generalization of the familiar Weibull distribution (for $\gamma_2 \rightarrow 0$) and also because the asymptotic behavior of the right tail is solely controlled by the parameter γ_2 (for large values of X , $P\{X > x\} \approx \gamma_2 \beta^{1/\gamma_2} x^{-1/\gamma_2}$). The distribution has a finite variance distribution for $0 \leq \gamma_2 < 0.5$ and finite mean for $0 \leq \gamma_2 < 1$. Finally, the shape parameter γ_1 controls the left tail as for $0 < \gamma_1 < 1$ the distribution is J-shaped, for $\gamma_1 > 1$ bell-shaped and for $\gamma_1 = 1$ degenerates to the familiar Pareto type II distribution.

3. Application

To test the applicability of the above theoretical framework, a large data set of daily rainfall observations was used. This is a subset of the Global Historical Climatology Network – Daily database (<http://www.ncdc.noaa.gov/oa/climate/ghcn-daily>) that includes data recorded at over 43 000 stations. A total of 11 519 stations were selected from this database with the following criteria: (a) record length of over 50 years; (b) percentage of missing values less than 1%; and (c) percentage of flags for suspect quality less than 1% (see the above web site for the details about flags). The locations of the selected stations are shown in Fig. 1.

To each of the 11 519 daily rainfall samples we tested the suitability of the above described, entropy derived distributions using evaluation tools based on the L -moments ratio plots. A typical approach to identify a suitable distribution for one or more samples is to compare empirically derived L -ratios with the respective theoretical quantities. The latter are plotted on a graph as a point, line or area (depending on the number and the type of the distributional parameters). The most common plot of this type is the L -kurtosis vs. L -skewness plot. Nevertheless, here we prefer the L -skewness vs. L -variation plot because, as we deal with positive RVs, the L -variation is meaningful (nonzero and nonnegative, with values ranging from 0 to 1), and evidently its estimation is more robust than that of the L -kurtosis.

The most general distribution described above is the four parameter GB2 distribution, whose graphical depiction in terms of L -ratio plots is not feasible as different parameter sets may correspond to the same point on an L -ratio plot. Therefore, we made plots for the three-parameter special cases, i.e. the GG and the BurrXII distributions, which have one scale parameter and two shape parameters. Evidently, the suitability of one of those two entails also the suitability of the more general GB2 distribution. Each of the GG and the BurrXII distributions forms an area on the L -ratio plot. The area was depicted by drawing several theoretical lines that correspond to specific values of the tail parameter γ_2 (with varying parameter γ_1). The theoretical points of L -variation and L -skewness were calculated numerically based on the integral definition of the L -moments [5].

Fig. 2 and Fig. 3 depict the theoretical area of the GG and BurrXII distributions, respectively, in the L -skewness vs. L -variation plot, as well as the empirical points of the corresponding sample statistics of the 11 519 records. In Fig. 2, 97.6% of the empirical points lie in the theoretical area of the GG distribution for a tail parameter $0.1 \leq \gamma_2 \leq 2$. It is worth noting that the vast majority of the empirical points are located above the line of the Gamma distribution ($\gamma_2 = 1$) resulting in an average tail parameter $\gamma_2 = 0.53$. This indicates that the Gamma distribution, which has probably been the most common model for daily rainfall, should not be uncritically adopted as it would seriously underestimate the frequency and the magnitude of extreme events. Specifically, for $\gamma_2 < 1$ (as observed in the vast majority of cases of Fig. 2), the GG distribution exhibits a tail heavier than that of the Gamma distribution. In addition, in points where the GG distribution is unsuitable, a heavier type of distribution tail is needed.

Regarding the BurrXII distribution in Fig. 3, 87.7% of the empirical points lie in a space corresponding to a tail parameter $0 \leq \gamma_2 \leq 0.7$. The average value of the tail parameter is $\gamma_2 = 0.19$ meaning that approximately only moments of up to order $1/0.19 \approx 5$ exist. Notably, the empirical points that the BurrXII fails to describe can be described by the GG distribution and vice versa. Particularly, the empirical points in Fig. 3 that lie above the line $\gamma_2 = 0.5$ (more or less the ones that GG distribution fails to describe) correspond to BurrXII distributions with infinite variance and thus, it is hard to describe by an exponential-type distribution whose all moments are finite.

An additional comment that regards both distributions and also empirically recognized, (see for example the empirical densities in Fig. 2), is that a large percentage of the empirical points of daily rainfall are better described by bell-shaped distributions (highly positively skewed though) rather than J-shaped. Thus, a general model for rainfall should comprehend both J-shaped and bell-shaped densities. This excludes simple models like the Exponential, the Pareto, and the Lognormal distributions (commonly used as rainfall models) because the first two can only be J-shaped and the third only bell-shaped. Finally, as mentioned before, both distributions are special cases of the GB2 distribution and thus the GB2 distribution describes 100% of the empirical points, being thus, a suitable model for rainfall, not only for

the daily scale examined here, but also for finer (e.g. sub-hourly) and coarser (e.g. annual) scales [15].

4. Summary and conclusions

In order to derive statistical distributions suitable for geophysical processes, and particularly for rainfall, we propose a rationale for defining and selecting constraints within a BGS entropy maximization framework. Entropy maximization offers a solid theoretical basis for identifying a probabilistic law based on the available information, in contrast to the common technique of choosing a distribution from a repertoire based on trial-and-error methods. This rationale is based on the premises that the constraints should be as few and simple as possible and incorporate prior information on the process of interest. This prior information may concern the general shapes of densities and could be obtained by studying the process worldwide. We studied and justified conceptually three particular constraints, related to the logarithmic and power functions, which are suitable for positive, highly varying and asymmetric RVs. Namely, the constraints are the expected values of (a) $\ln x$; (b) x^q ; and (c) $\ln(1 + p x^q) / p$. The last constraint generalizes the classical moments and naturally leads to power-type distributions avoiding generalized entropy measures.

The BGS entropy maximization under two combinations of these constraints leads to two flexible distributions, i.e., a three-parameter exponential type, known as the Generalized Gamma (GG), and, a four-parameter power type, known as the Generalized Beta of the second kind (GB2)—the former is a particular limiting case of the latter. Another three-parameter model, known as the Burr type XII (power type), easily derived from the GB2, proves to be also useful. In order to evaluate the performance of the three-parameter entropy derived distributions, we used a very large database with thousands of daily rainfall records across the world. We formed the theoretical area of those distributions in an L -skewness vs. L -variation plot and compared it with the corresponding sample statistics of 11 519 daily rainfall records. Both the GG and BurrXII distributions performed very well by describing 97.6% and 87.7%, respectively, of the empirical points. Notably, the two distributions are complementary in the sense that empirical points that cannot be described by one can be

described by the other. Consequently, as both distributions are special cases of the GB2, the latter can describe 100% of the empirical points being thus a model suitable for all daily rainfall records.

Both the empirical analysis of this massive number of records as well as the distributions tested, lead to two useful conclusions regarding the shape characteristics of the daily rainfall distribution. First, a suitable distribution for daily rainfall must be able to form both J- and bell-shaped densities, with J-shaped densities having also the property for $x \rightarrow 0$, $f_x(x) \rightarrow \infty$. This excludes commonly used models like the Exponential, the Pareto or the Lognormal distributions and many more. Of course, these distributions may be suitable for particular cases or for specific rainfall ranges (above certain thresholds) and actually are, but cannot be proposed as general models for daily rainfall. Second, regarding the right distribution tail (a very important feature as it controls the behavior of extremes) the analysis showed that heavy-tailed distributions describe better the vast majority of the empirical points compared to light-tailed distributions. Consequently, the Gamma distribution (probably the most common daily rainfall model) is rejected as a general model for daily rainfall, because in the majority of rainfall records would seriously underestimate the extreme events.

Evidently, rainfall in different places of Earth is influenced by local characteristics such as climate, topography, distance from the sea and many more. The diversity of such characteristics produces different rainfall patterns across the globe. This does not contradict our finding that a single flexible probabilistic law (the GB2 distribution) or simpler special cases thereof (the GG and BurrXII distributions) can model rainfall over all examined cases. The diversity of characteristics is rather reflected in the diversity of shapes that the GB2 distribution can produce, as well as in the wide range of feasible parameter values.

References

- [1] Burr IW. Cumulative Frequency Functions. *The Annals of Mathematical Statistics*. 1942;13(2):215-232.
- [2] Esteban MD, Morales D. A summary on entropy statistics. *Kybernetika*. 1995;31(4):337-346.

- [3] Goldie CM, Klüppelberg C. Subexponential distributions. *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. 1998:435–459.
- [4] Havrda J, Charvát F. Concept of structural α -entropy. *Kybernetika*. 1967;3:30–35.
- [5] Hosking JR. L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1990;52(1):105–124.
- [6] Jaynes ET. Information Theory and Statistical Mechanics. *Phys. Rev.* 1957;106(4):620.
- [7] Jaynes ET. Information Theory and Statistical Mechanics. II. *Phys. Rev.* 1957;108(2):171.
- [8] Jaynes ET. *Probability: The logic of science*. Cambridge University Press; 2003.
- [9] Kapur JN. *Maximum-entropy models in science and engineering*. John Wiley & Sons; 1989.
- [10] Kleiber C, Kotz S. *Statistical size distributions in economics and actuarial sciences*. Wiley-Interscience; 2003.
- [11] Koutsoyiannis D. Uncertainty, entropy, scaling and hydrological stochastics, 1, Marginal distributional properties of hydrological processes and state scaling. *Hydrological Sciences Journal*. 2005;50(3):381-404.
- [12] McDonald JB. Some Generalized Functions for the Size Distribution of Income. *Econometrica*. 1984;52(3):647-663.
- [13] McDonald JB, Xu YJ. A generalization of the beta distribution with applications. *Journal of Econometrics*. 1995;66(1-2):133-152.
- [14] Mielke Jr PW, Johnson ES. Some generalized beta distributions of the second kind having desirable application features in hydrology and meteorology. *Water Resources Research*. 1974;10(2):223–226.
- [15] Papalexiou SM, Koutsoyiannis D. Probabilistic description of rainfall intensity at multiple time scales. In: *IHP 2008 Capri Symposium: “The Role of Hydrology in Water Resources Management”*.; 2008. Available at: <http://www.itia.ntua.gr/en/docinfo/884/> [Accessed October 6, 2010].
- [16] Papalexiou SM, Koutsoyiannis D. Ombrian curves in a maximum entropy framework.

In: European Geosciences Union General Assembly 2008.; 2008:00702. Available at:
<http://www.itia.ntua.gr/en/docinfo/851/> [Accessed October 6, 2010].

- [17] Shannon CE. The mathematical theory of communication. *Bell System Technical Journal*. 1948;27:379–423.
- [18] Singh VP. Entropy theory for derivation of infiltration equations. *Water Resour. Res.* 2010;46:20.
- [19] Stacy EW. A Generalization of the Gamma Distribution. *The Annals of Mathematical Statistics*. 1962;33(3):1187-1192.
- [20] Tadikamalla PR. A Look at the Burr and Related Distributions. *International Statistical Review / Revue Internationale de Statistique*. 1980;48(3):337-344.
- [21] Tsallis C. Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*. 1988;52(1):479-487.

Figures

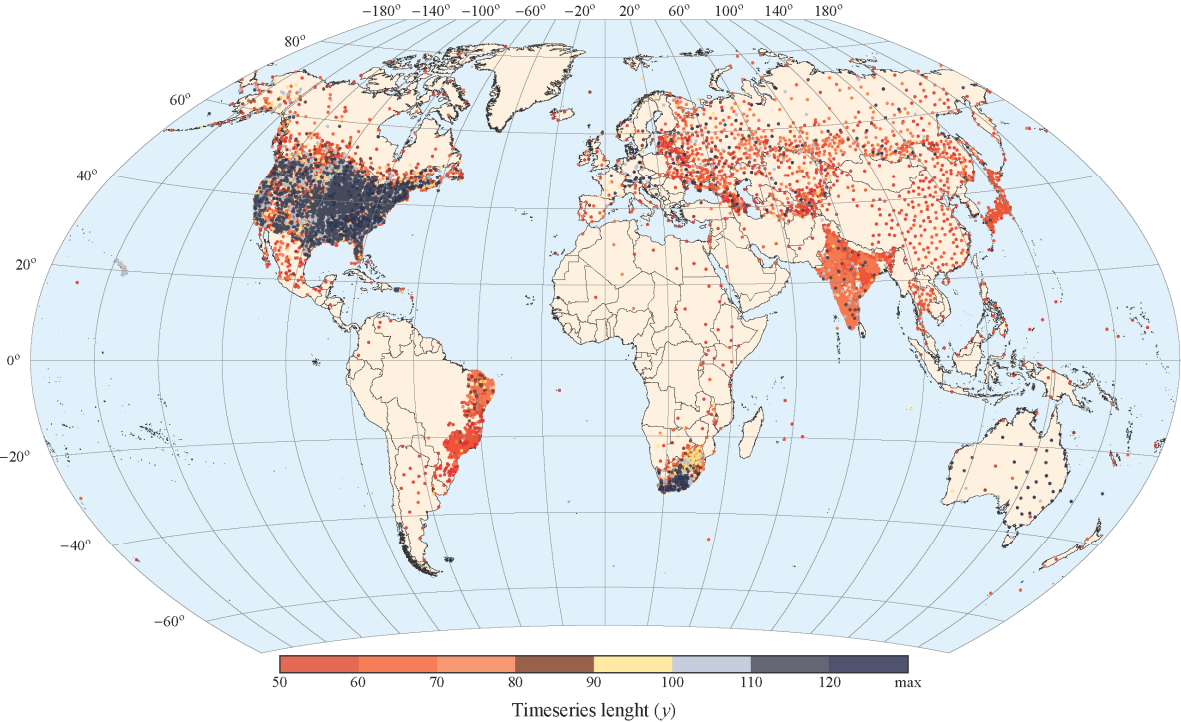


Fig. 1 Locations of the rain gauge stations of the study, which are a subset of the Global Historical Climatology Network-Daily database containing those stations with daily rainfall record length of over 50 years (a total of 11 519 stations with very few missing values).

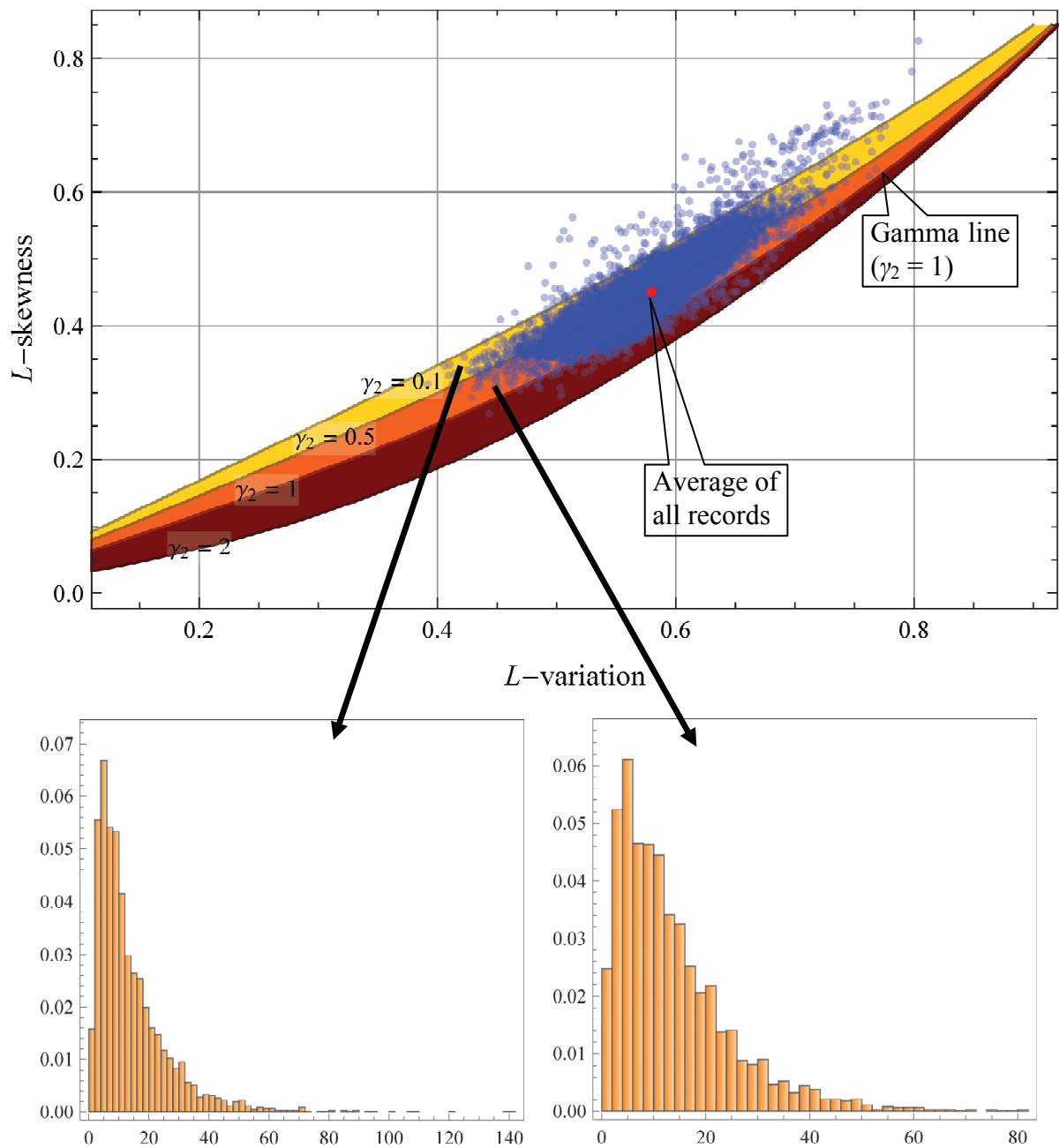


Fig. 2 Theoretical relationships of L -skewness vs. L -variation of the Generalized Gamma distribution and empirical points of the corresponding sample statistics of the 11 519 records. 97.6% of the empirical points lie in the space of the GG for $0.1 < \gamma_2 < 2$. For two of the records with the indicated positions on the L -moments diagram the empirical probability density functions are also shown.

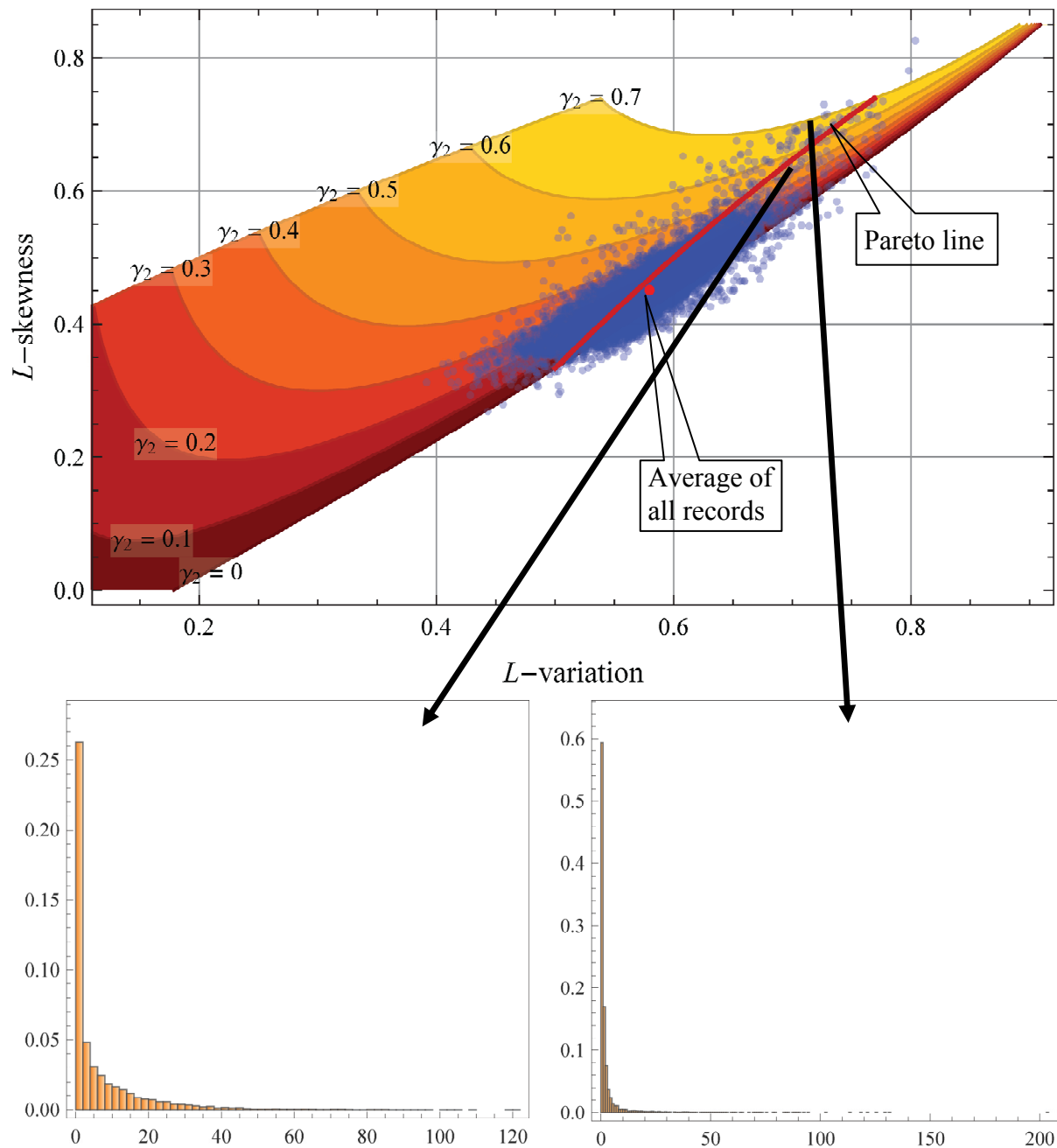


Fig. 3 Theoretical relationships of L -skewness vs. L -variation of the Burr type XII distribution and empirical points of the corresponding sample statistics of the 11 519 records. 87.7% of the empirical points lie in the space of the BurrXII for $0 < \gamma_2 < 0.7$. For two of the records with the indicated positions on the L -moments diagram the empirical probability density functions are also shown.