European Geosciences Union
General Assembly 2011

Vienna, Austria,
03-08 April 2011

Session: HS7.5/NP6.7
Hydroclimatic stochastics

A worldwide probabilistic analysis of rainfall at multiple time scales based on entropy maximization

Simon-Michael Papalexiou and Demetris Koutsoyiannis

Department of Water Resources and Environmental Engineering,
National Technical University of Athens, Greece

Contact Info: smp@itia.ntua.gr
Available online: http://itia.ntua.gr/1129

Introduction and motivation

- Geophysical processes, e.g., like rainfall or river discharge, can be probabilistically described if treated as random variables obeying a distribution law.
- Of course, choosing a proper probabilistic model is not a trivial task. The common practice is to choose one of the few popular distributions based on empirical considerations e.g., the summary statistics of the data.
- Regarding rainfall, numerous probability models have been suggested as appropriate, depending on the time scale or the study area; however, a theoretically justified and "universally" accepted model does not exist.
- In contrast, the principle of maximum entropy offers a theoretical basis for selecting a distribution law, based on deduction rather than on trial-and-error procedures. Yet, the resulting maximum entropy distribution is not unique as it depends on the maximized entropic form and the imposed constraints.
- Here, we use the principle of maximum entropy and we suggest and justify simple and general constrains that are suitable for positive, highly varying and asymmetric random processes, like rainfall.
- We test the performance of the resulting maximum entropy distributions by studying rainfall worldwide, at various time scales.

Entropy measures

- What is entropy?
  i. Entropy as a concept dates back to the works of Rudolf Clausius in 1850 and of Ludwig Boltzmann around 1870 who gave entropy a statistical meaning and related it to statistical mechanics. Next, the concept of entropy was evolved by J. Willard Gibbs and Von Neumann in quantum mechanics, and was reintroduced in information theory by Claude Shannon in 1948.
  ii. Information entropy is a purely probabilistic concept and is regarded as a measure of the uncertainty related to a random variable (RV).
- In literature there are more than twenty different entropy measures [1], proposed mainly as generalizations of Boltzmann-Gibbs-Shannon (BGS) entropy, which is the most famous and well justified entropy measure. The BGS entropy for a non-negative continuous RV X with density function $f_X(x)$ is defined as

$$S_{\mathrm{BGS}} = -\int_0^\infty f_X(x) \ln f_X(x) \mathrm{d}x \tag{1}$$

- A famous generalization was proposed by Rényi in 1961, while another one, that gained popularity the last decades, is the Havrda-Charvat-Tsallis (HTC) entropy [2,3], defined as

$$S_{\mathrm{HCT}} = \frac{1}{1-q}\left[\int_0^\infty f_X(x)^q \mathrm{d}x - 1\right] \tag{2}$$

which for $q \to 1$ converges to the BGS entropy.

The principle of maximum entropy (POME)

- The principle of maximum entropy, established by Edwin Jaynes [4,5], essentially relies in finding the most suitable probability distribution under the available information. According to Jaynes, the resulted maximum entropy distribution "is the least biased estimate possible on the given information…".
- Mathematically, the given information used in the principle of maximum entropy, is expressed as a set of constraints formed as expectations of functions $g_j(\ )$ of X, i.e.,

$$E\left[g_j(x)\right] = \int_0^\infty g_j(x)\, f_X(x)\mathrm{d}x = c_j, \quad j = 1, \dots, n \tag{3}$$

- The resulting maximum entropy distributions emerge by maximizing the selected form of entropy with constraints c_j, and with the additional constraint (to guarantee the legitimacy of the distribution)

$$\int_0^\infty f_X(x)\, \mathrm{d}x = 1 \tag{4}$$

- The general solution of the maximum entropy distributions resulting from the maximization of BGS entropy and the HCT entropy (accomplished by using the method of Lagrange multipliers) are, respectively,

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

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

where λ_j, with j =0,…, n are the Lagrange multipliers linked to the constraints.

Selecting the constraints

- The choice of the imposed constraints is the most important and determinative part of the method as it defines uniquely the resulting maximum entropy distribution.
- Choosing constraints, however, is not trivial; theoretically, the expectation of any RV function can be used.
- Commonly, entropy maximization is done by assuming known mean and variance, which leads to (a) the Gaussian distribution in the BGS entropy case, and (b) a symmetric bell-shaped distribution with power-type tails in the HCT entropy case.
- So, how should we chose constraints?
  1. Constraints should express our state of knowledge concerning a RV and should summarize all the available information from both observations and theoretical considerations.
  2. We can assume that some coarse features of the RV, e.g., the mean or the variance, are more likely to be preserved in the future than finer features, e.g., the kurtosis coefficient. Therefore, constraints should be simple and express features that are robust to estimate from the sample, and are likely to be preserved in the future.
  3. For some geophysical processes we may know important prior characteristics of the underlying distribution that should be preserved, e.g., a J-shaped or bell-shaped distribution or a heavy- or light-tailed distribution. So, the constraints should provide a distribution consistent with the empirical evidence.

The expectation of ln X

- Here, we aim to define and justify the use of some simple and general constraints, suitable for geophysical RVs, to use with the BGS entropy.
- The geometric mean μ_G given by

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

is measure of central tendency, with the convenient property for geophysical processes to be defined only for positive values. This gives an intuitively meaning to formulate the expectation of ln X as a constraint

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

- Apart from its relationship to the geometric mean and its simplicity, the expectation of ln X, has some desired properties that make it an essential constraint for positively skewed RVs. Samples drawn from positively skewed, or even more, from heavy-tailed distributions, e.g., like those of daily rainfall, exhibit values that act like outliers and consequently strongly influence the sample moments, especially those of higher order. 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. For this reason the logarithmic transformation is probably the most common transformation used in hydrology as it tends to normalize positively skewed data.

The expectations of moments

- The common use of mean and variance as constraints must be attributed to their link with the physical principles of momentum and energy conservation. However, this link is invalid in geophysical processes, e.g., the mean of the rainfall is not its momentum and its variance it is not its energy.
- Theoretical arguments (apart from simplicity and conceptual meaning as measures of central tendency and dispersion) which favor the mean and the variance against, e.g., fractional moments of small order or even negative do not exist. For example, if the second moment is likely to be preserved then probably 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

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

  Thus, we may say that the function x^q for small values of q behaves similar to $\ln x$.
- 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.,

$$m_q = E(X^q) = \int_0^{\infty} x^q\, f_X(x)\mathrm{d}x \qquad (10)$$

The expectations of p-moments

- Many entropy generalizations have emerged to explain empirically detected deviations from exponential type distributions that arise from the BGS entropy using moment constraints. Yet, generalized entropy measures have been criticized for lacking theoretical consistency and for being arbitrary.
- Here, we generalize the important notion of moments [6] inspired by the limiting definition of the exponential function. First, we define the generalized power function

$$x_p^q = \ln(1 + p\,x^q)/p \tag{11}$$

which for $p \to 0$ converges to the familiar power function x^q. Thus, we generalize classical moments by defining the p-moments as

$$m_q^p = E\left(X_p^q\right) = \frac{1}{p}\int_0^\infty \ln(1 + p\,x^q)\; f_X(x)\mathrm{d}x \tag{12}$$

- We believe that there is a strong rationale that supports the use of p-moments, i.e.,
  i. Generalized entropy measures have been successfully used; why not p-moments with the standard definition of entropy?
  ii. Maximization of the BGS entropy using p-moments leads naturally to power-type distributions (including the Pareto and Tsallis distributions for q = 1 and q = 2, respectively).
  iii. p-moments are simple and, for $p = 0$, become identical to classical moments, i.e., $m_q^0 = m_q$.
  iv. They exhibit similar properties with the $\ln x$ function, and thus are suitable for positively skewed RVs; additionally, compared to $E(\ln x)$ they are always positive.

The Generalized Gamma distribution

- 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

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

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 $\Gamma(\ )$ is the Gamma function and $\Gamma(\ ,\ )$ the upper incomplete Gamma functions.
- This distribution, commonly attributed to Stacy [7], is known as the Generalized Gamma distribution (GG). It is a very flexible distribution that includes many well-known distributions as special cases, e.g., the Gamma, the Weibull, the Exponential, or the Chi-square distributions.
- The distribution comprises the scale parameter β, and the shape parameters γ_1 and γ_2. The parameter γ_2 mainly controls the asymptotic behavior of the right tail and the parameter γ_1 that of the left tail. Specifically, the distribution is J-shaped for $0 < \gamma_1 < 1$, bell-shaped for $\gamma_1 > 1$, and equals the Generalized Exponential distribution for $\gamma_1 = 1$.

The Generalized Beta of the second kind distribution

- Maximization of the BGS entropy, given in (1), with constraints (8) and (12) results in the density function

$$f_X(x) = \exp[-\lambda_0 - \lambda_1 \ln x - \lambda_2 \ln(1 + p\, x^q)\, /p] \tag{16}$$

which after algebraic manipulations and parameter renaming can be written as

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

corresponding to the distribution function

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

where $B(\,,)$ and $B_z(\,,)$ denote the Beta and the incomplete Beta functions, respectively.
- This distribution is known as the Generalized Beta of the second kind (GB2) and has been rediscovered many times under different names and parameterizations. Probably, Milke and Johnson [8] were the first who formed it, and proposed it for describing hydrological and meteorological variables.
- The distribution is an extremely flexible four-parameter distribution comprising one scale parameter β, and three shape parameters $\gamma_1, \gamma_2, \gamma_3$, which allow the distribution to form innumerable different shapes. Many of the well-known distributions are special or limiting cases of the GB2 distribution (see e.g. [9,10]).

The Burr type XII distribution

- The GB2 distribution is extremely flexible but not so easy to handle. Yet, one simple three-parameter special case, with analytical distribution function, can be derived by setting $\gamma_1 = 1$ in the density function of GB2. After some trivial algebraic manipulations and parameter renaming the distribution Burr type XII is derived, introduced by Burr in 1942 in the framework of distribution system similar to Pearson's. Its density function is

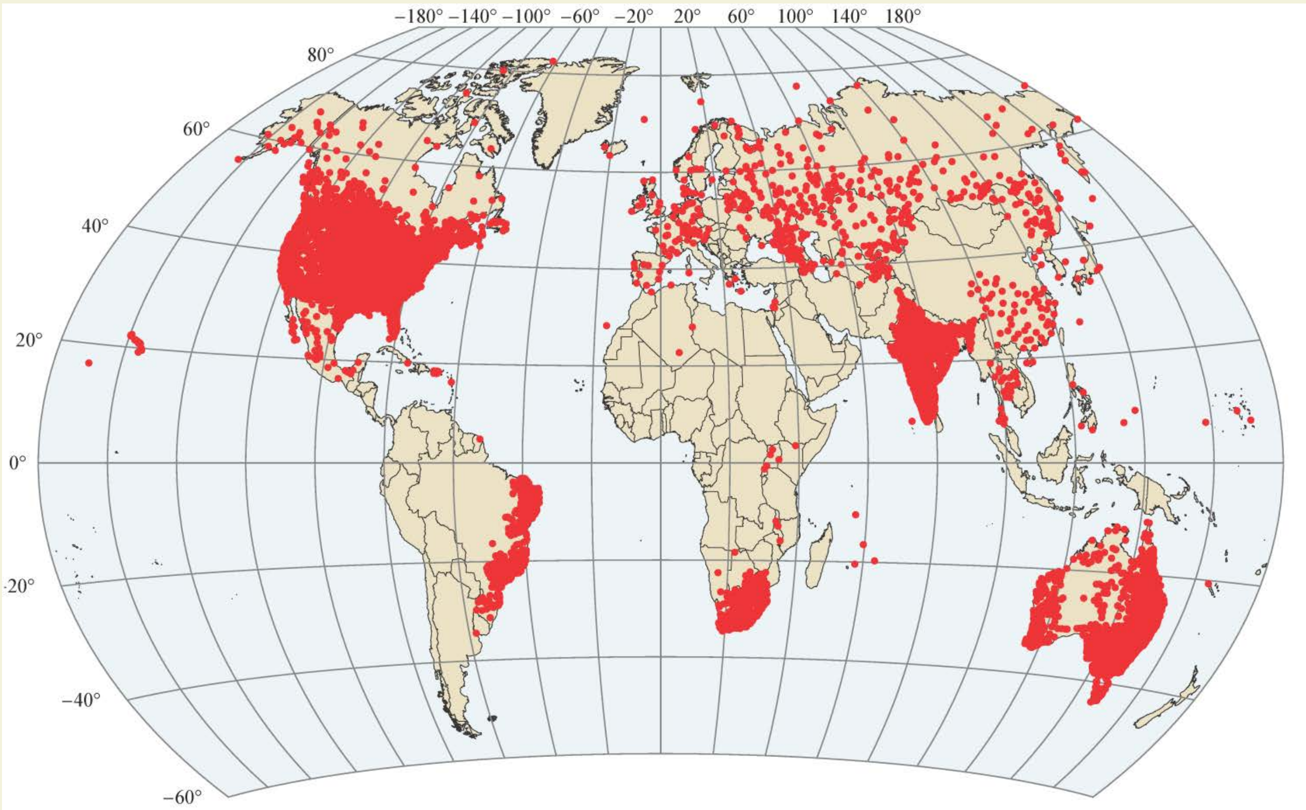
$$f_X(x) = \frac{1}{\beta}\left(\frac{x}{\beta}\right)^{\gamma_1 - 1}\left[1 + \gamma_2\left(\frac{x}{\beta}\right)^{\gamma_1}\right]^{-\frac{1}{\gamma_1\gamma_2} - 1}, x \geq 0 \tag{19}$$

corresponding to the distribution function

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

- This flexible power-type distribution, comprises the scale parameter β and the shape parameters γ_1 and γ_2. In literature, the distribution is given in a different form (see e.g. [11]). However, as given here, the parameter γ_2 completely controls the asymptotic behavior of the right tail and the parameter γ_1 that of the left tail. In detail, the distribution is J-shaped for $0 < \gamma_1 < 1$, bell-shaped for $\gamma_1 > 1$, and for $\gamma_1 = 1$ equals the Pareto type II distribution. Additionally, (19) constitutes a kind of a generalization of the as for $\gamma_2 \to 0$ converges to the Weibull distribution.

Application to rainfall: The dataset



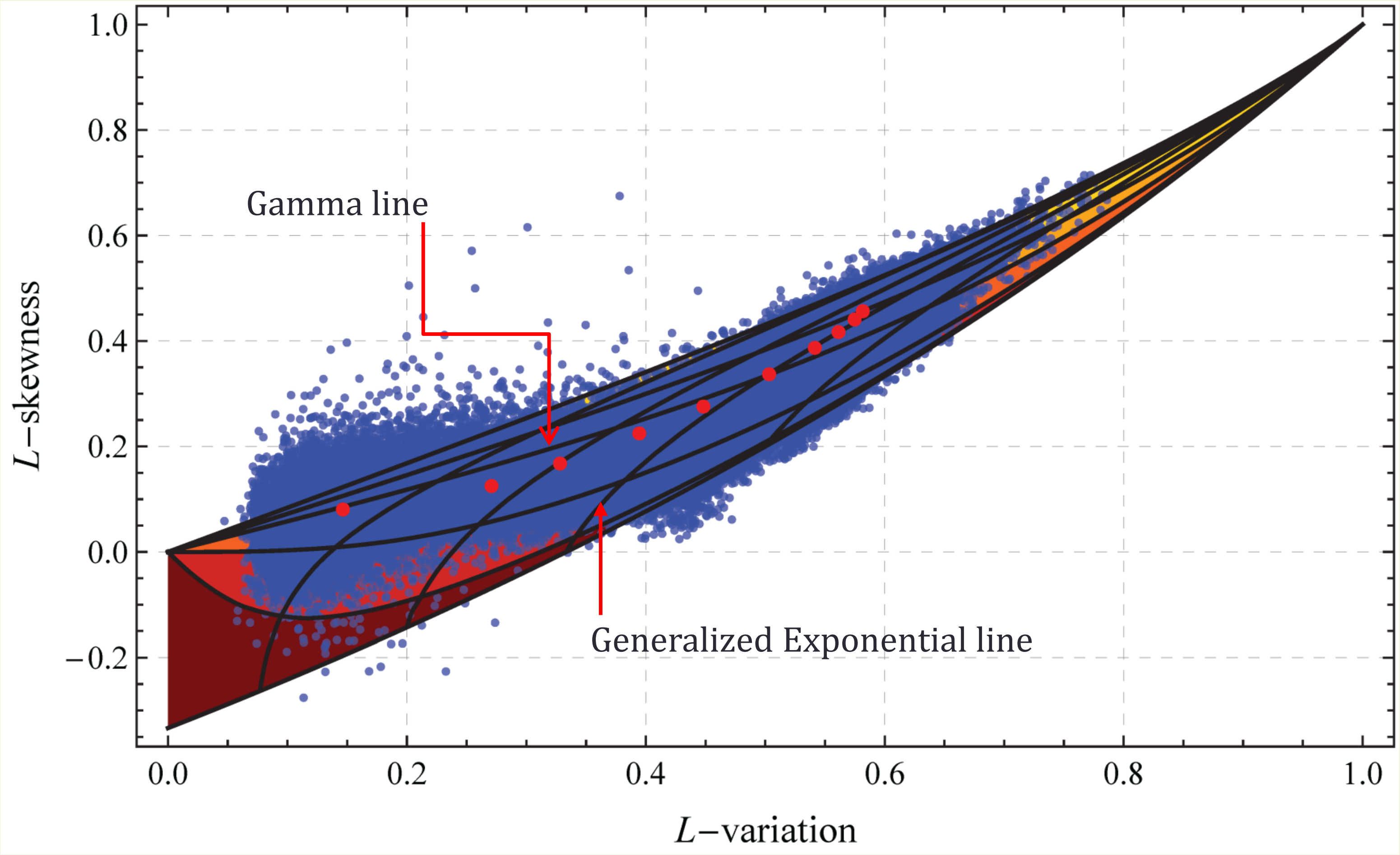
This map depicts the locations of the stations studied. A total of 11 697 daily rainfall records.

The data used here, are rainfall records of the Global Historical Climatology Network-Daily database (http://www.ncdc.noaa.gov/oa/climate/ghcn-daily) which includes data recorded at over 40 000 stations worldwide. Many of those records, however, are too short in length, have missing data, or, contain data suspect in terms of quality. Thus, we selected for analysis only those records fulfilling the following criteria: (a) record length greater or equal than 50 years, (b) missing data less than 10% and, (c) data with quality flags less than 0.1%.

Methodology

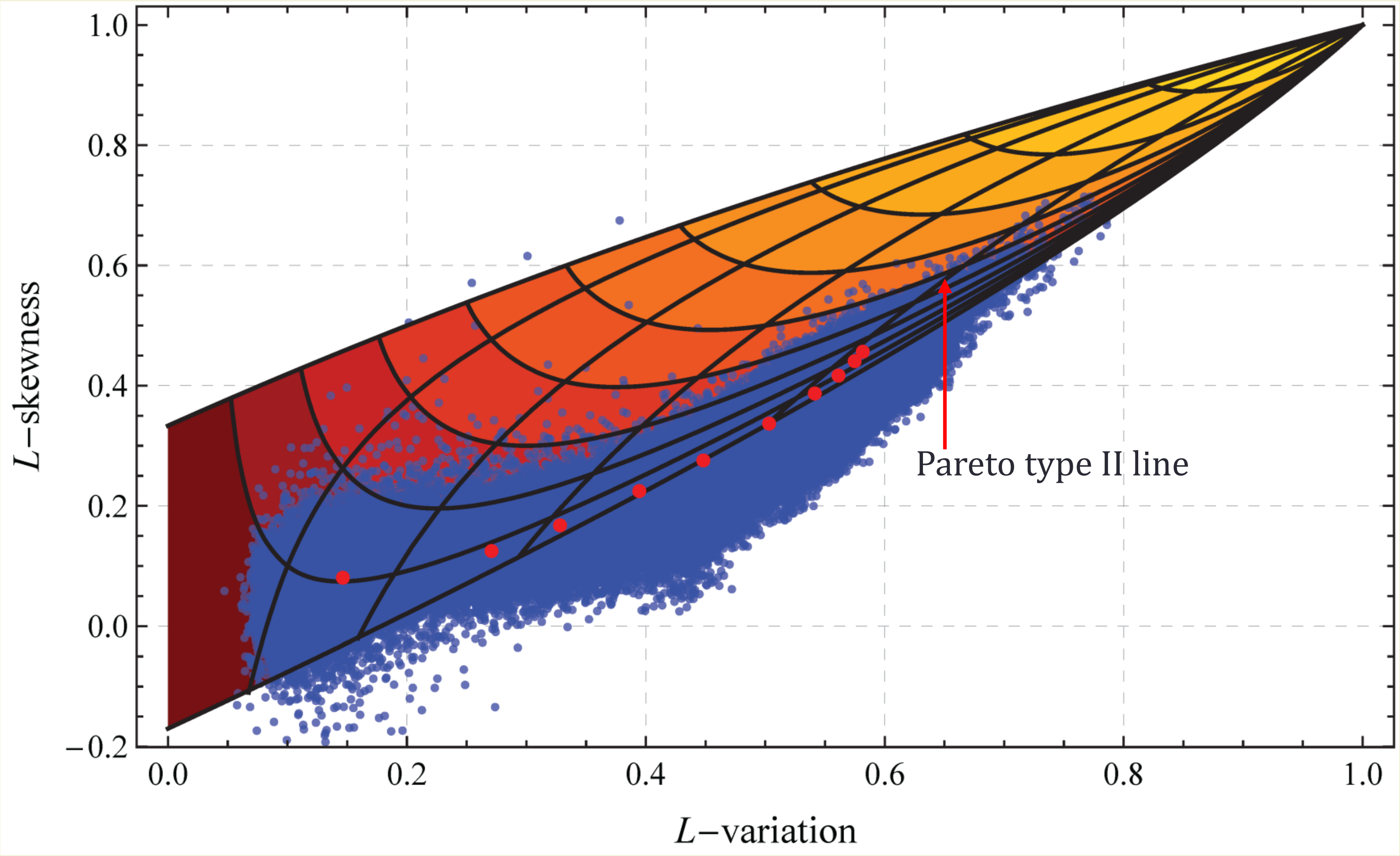
- To test the suitability of the afore-mentioned distributions to describe rainfall at several time scales we proceed as follows:
  1. We aggregated the original time series of each station (given in the daily time scale), at several time scales *k*, by calculating the average of non-overlapping runs of *k* sequential values. The average of a *k*-run was calculated only if the percentage of missing values within the run was less than 15%, e.g., in a 7-run (weekly time scale) this is one day missing.
  2. We selected the positive rainfall values at each time scale and estimated the sample coefficients of *L*-variation (L-C_V) and *L*-skewness (L-C_S) (denoted here as L_2-points).
  3. We formed the theoretical L-C_S *vs.* L-C_V space (denoted here as L_2-space) for the Generalized Gamma and Burr type XII distributions.
  4. We compared the sample L_2-points at each time scale with the theoretical L_2-space of the distributions and estimated the percentage of L_2-points that belong within the theoretical L_2-space.

L_2-space of the GG distribution



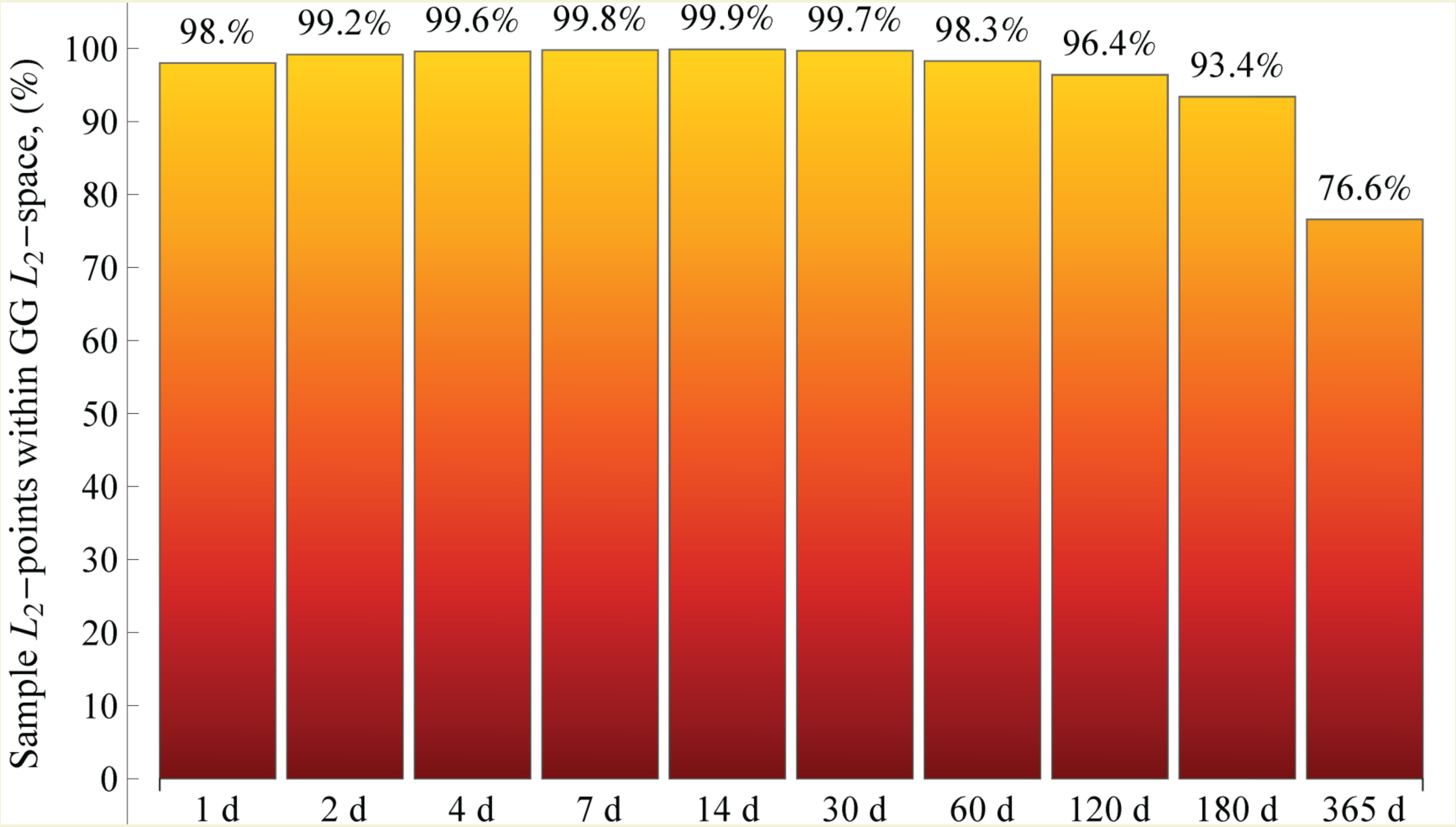
The figure depicts the theoretical L_2-space of the GG distribution and the sample L_2-points (blue dots) of all time scales examined. The red dots show the average L_2-point at each time scale (from left to right the scale gets finer, varying from yearly to daily scale).

L_2-space of the Burr type XII distribution



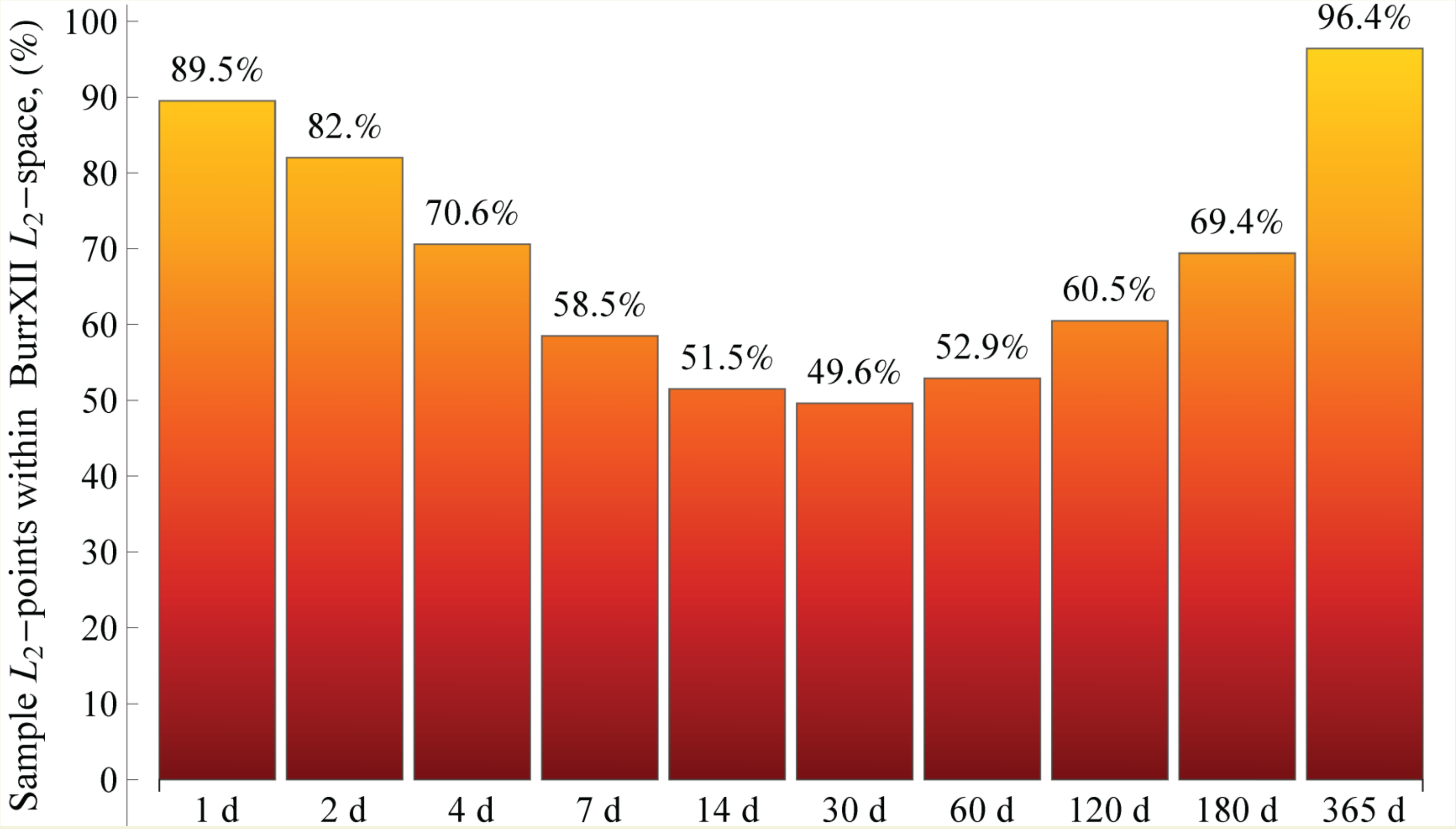
The figure depicts the theoretical L_2-space of the BurrXII distribution and the sample L_2-points (blue dots) of all time scales examined. The red dots show the average L_2-point at each time scale (from left to right the scale gets finer, varying from yearly to daily scale).

Performance of the GG distribution



The figure depicts the percentage (%) of the sample L_2-points, at each time scale, that lie within the theoretical L_2-space of the GG distribution. Almost at all time scales examined the distribution performs exceptionally well.

Performance of the Burr type XII distribution



The figure depicts the percentage (%) of the sample L_2-points, at each time scale, that lie within the theoretical L_2-space of the Burr type XII distribution. The distribution performs well at the daily and the yearly scales with its worst performance at the monthly scale.

Conclusions

- We use the principle of maximum entropy with the BGS entropy to derive suitable distributions for geophysical processes.
- The imposed constraints should, first, summarize all the available information from both observations and theoretical considerations, and second, be simple and express features that are likely to be preserved in the future.
- We propose and justify the use of three simple constraints (a) the $E(\ln X)$, (b) moments of arbitrary order, and (c) p-moments of arbitrary order.
- p-moments, which are a generalization of the classical moments, when they are used with the BGS entropy lead straightforwardly to power-type distributions, thus avoiding generalized entropy measures.
- The combination of the studied constraints produces two very flexible distributions, (a) the Generalized Gamma (exponential-type), and (b) the Generalized Beta of the second kind (power-type).
- The empirical analysis performed to 11 697 rainfall records worldwide, showed that the Generalized Gamma distribution is an exceptional model for rainfall at all time scales, while the Burr type XII distribution (a special case of the GB2 distribution) is a good model for the daily and annual time scales.

References

[1] Esteban MD, Morales D. A summary on entropy statistics. Kybernetika. 1995;31(4):337-346.

[2] Havrda J, Charvát F. Concept of structural a-entropy. Kybernetika. 1967;3:30–35.

[3] Tsallis C. Possible generalization of Boltzmann-Gibbs statistics. Journal of Statistical Physics. 1988;52(1):479-487.

[4] Jaynes ET. Information Theory and Statistical Mechanics. Phys. Rev. 1957;106(4):620.

[5] Jaynes ET. Information Theory and Statistical Mechanics. II. Phys. Rev. 1957;108(2):171.

[6] Papalexiou SM, Koutsoyiannis D. An entropy based model of the probability distribution of daily rainfall [tentative title], Advances in Water Resources [in review] 2011.

[7] Stacy EW. A Generalization of the Gamma Distribution. The Annals of Mathematical Statistics. 1962;33(3):1187-1192.

[8] 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.

[9] Kleiber C, Kotz S. Statistical size distributions in economics and actuarial sciences. Wiley-Interscience; 2003.

[10] McDonald JB, Xu YJ. A generalization of the beta distribution with applications. Journal of Econometrics. 1995;66(1-2):133-152.

[11] Tadikamalla PR. A Look at the Burr and Related Distributions. International Statistical Review / Revue Internationale de Statistique. 1980;48(3):337-344.