Hydrology and
Earth System
Sciences

Open Access

How extreme is extreme? An assessment of daily rainfall distribution tails

S. M. Papalexiou, D. Koutsoyiannis, and C. Makropoulos

Department of Water Resources, Faculty of Civil Engineering, National Technical University of Athens,
Heroon Polytechneiou 5, 157 80 Zographou, Greece

Correspondence to: S. M. Papalexiou (smp@itia.ntua.gr)

Abstract. The upper part of a probability distribution, usually known as the tail, governs both the magnitude and the frequency of extreme events. The tail behaviour of all probability distributions may be, loosely speaking, categorized into two families: heavy-tailed and light-tailed distributions, with the latter generating "milder" and less frequent extremes compared to the former. This emphasizes how important for hydrological design it is to assess the tail behaviour correctly. Traditionally, the wet-day daily rainfall has been described by light-tailed distributions like the Gamma distribution, although heavier-tailed distributions have also been proposed and used, e.g., the Lognormal, the Pareto, the Kappa, and other distributions. Here we investigate the distribution tails for daily rainfall by comparing the upper part of empirical distributions of thousands of records with four common theoretical tails: those of the Pareto, Lognormal, Weibull and Gamma distributions. Specifically, we use 15 029 daily rainfall records from around the world with record lengths from 50 to 172 yr. The analysis shows that heavier-tailed distributions are in better agreement with the observed rainfall extremes than the more often used lighter tailed distributions. This result has clear implications on extreme event modelling and engineering design.

1 Introduction

Heavy rainfall may induce serious infrastructure failures and may even result in loss of human lives. It is common then to characterize such rainfall with adjectives like "abnormal", "rare" or "extreme". But what can be considered "extreme" rainfall? Behind any discussion on the subjective nature of such pronouncements, there lies the fundamental issue of infrastructure design, and the crucial question of the threshold beyond which events need not be taken into account as they are considered too rare for practical purposes. This question is all the more pertinent in view of the EU Flooding Directive's requirement to consider "extreme (flood) event scenarios" (European Commission, 2007).

Although short-term prediction of rainfall is possible to a degree (and useful for operational purposes), long-term prediction, on which infrastructure design is based, is infeasible in deterministic terms. We thus treat rainfall in a probabilistic manner, i.e., we consider rainfall as a random variable (RV) governed by a distribution law. Such a distribution law enables us to assign a return period to any rainfall amount, so that we can then reasonably argue that a rainfall event, e.g., with return period 1000 yr or more, is indeed an extreme. Yet, which distribution law we should choose is still a matter of debate.

The typical procedure for selecting a distribution law for rainfall is to (a) try some of many, a priori chosen, parametric families of distributions, (b) estimate the parameters according to one of many existing fitting methods, and (c) choose the one best fitted according to some metric or fitting test. Nevertheless, this procedure does not guarantee that the selected distribution will model adequately the tail, which is the upper part of the distribution that controls both the magnitude and frequency of extreme events. On the contrary, as only a very small portion of the empirical data belongs to the tail (unless a very large sample is available), all fitting methods will be "biased" against the tail, since the estimated fitting parameters will point towards the distribution that best describes the largest portion of the data (by definition not

belonging to the tail). Clearly, an ill-fitted tail may result in serious errors in terms of extreme event modelling with potentially severe consequences for hydrological design. For example, in Fig. 1 where four different distributions are fitted to the empirical distribution tail, it can be observed that the predicted magnitude of the 1000-yr event varies significantly.

The distributions can be classified according to the asymptotic behaviour of their tail into two general classes: (a) the subexponential class with tails tending to zero less rapidly than an exponential tail (here the term "exponential tail" is used to describe the tail of the exponential distribution), and (b) the hyperexponential or the superexponential class, with tails approaching zero more rapidly than an exponential tail (Teugels, 1975; Klüppelberg, 1988, 1989). Mathematically, this "intuitive" definition of the subexponential class for a distribution function F is expressed as

$$\lim_{x \to \infty} \frac{1 - F(x)}{\exp(-x/\beta)} = \infty \qquad \forall \beta > 0, \qquad (1)$$

while several equivalent mathematical conditions, in order to classify a distribution as subexponential, have been proposed (see, e.g., Embrechts et al., 1997; Goldie and Klüppelberg, 1998). Furthermore, this is not the only classification, as several other exist (see, e.g., El Adlouni et al., 2008, and references therein). In addition, many different terms have been used in the literature to refer to tails "heavier" than the exponential, e.g., "heavy tails", "fat tails", "thick tails", or, "long tails", that may lead to some ambiguity: see for example the various definitions that exist for the class of heavy-tailed distributions discussed by Werner and Upper (2004). Here, we use the term "heavy tail" in an intuitive and general way, i.e., to refer to tails approaching zero less rapidly than an exponential tail.

The practical implication of a heavy tail is that it predicts more frequent larger magnitude rainfall compared to light tails. Hence, if heavy tails are more suitable for modelling extreme events, the usual approach of adopting light-tailed models (e.g., the Gamma distribution) and fitting them on the whole sample of empirical data would result in a significant underestimation of risk with potential implications for human lives. However, there are significant indications that heavy tailed distributions may be more suitable. For example, in a pioneering study Mielke (1973) proposed the use of the Kappa distribution, a power-type distribution, to describe daily rainfall. Today there are large databases of rainfall records that allow us to investigate the appropriateness of light or heavy tails for modelling extreme events. This is the subject in which this paper aims to contribute.

2 The dataset

The data used in this study are daily rainfall records from the Global Historical Climatology Network-Daily database (version 2.60, www.ncdc.noaa.gov/oa/climate/ghcn-daily),
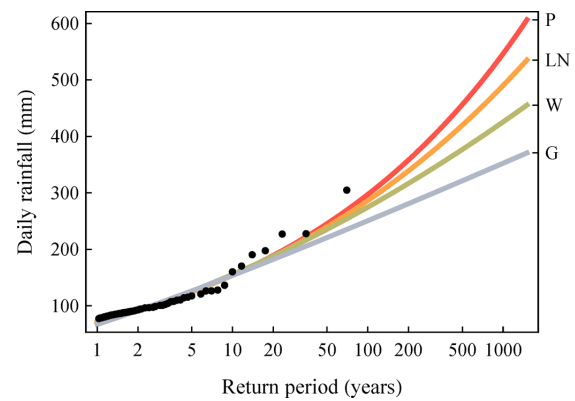

Fig. 1. Four different distribution tails fitted to an empirical tail (P, LN, W and G stands for the Pareto, the Lognormal, the Weibull and the Gamma distribution). A wrong choice may lead to severely underestimated or overestimated rainfall for large return periods.

which includes over 40 000 stations worldwide. Many of the records, however, are too short, have many missing data, or contain data that are suspect in terms of quality (for details regarding the quality flags refer to the Network's website above).

Thus, only records fulfilling the following criteria were selected for the analysis: (a) record length greater or equal than 50 yr, (b) missing data less than 20 %, and (c) data assigned with "quality flags" less than 0.1 %. Among the several different quality flags assigned to measurements, we screened against two: values with quality flags "G" (failed gap check) or "X" (failed bounds check). These were used to flag suspiciously large values, i.e., a sample value that is orders of magnitude larger than the second larger value in the sample. Whenever such a value existed in the records it was deleted (this, however, occurred in only 594 records in total, and in each of these records typically one or two values had to be deleted). Screening with these criteria resulted in 15 137 stations. The locations of these stations as well as their record lengths can be seen in Fig. 2, while Table 1 presents some basic summary statistics of the nonzero daily rainfall of those records.

We note that we did not fill any missing values as we deemed it meaningless for this study, focusing on extreme rainfall, because any regression-type technique would underestimate the real values. Missing values only affect the effective record length and, given the relatively high lower limit of record length we set (50 yr, while much smaller records are often used in hydrology, e.g., 10–30 yr), the resulting problem was not serious. Additionally, the percentage 20 % of missing daily values refers to the worst case and is actually much smaller in the majority of the records; thus, missing values would not alter or modify the conclusions drawn.

Finally, we note that the statistical procedure we describe next failed in a few records, for reasons of algorithmic convergence or time limits. Excluding these records, the total number of records where the analysis was applied is 15 029.
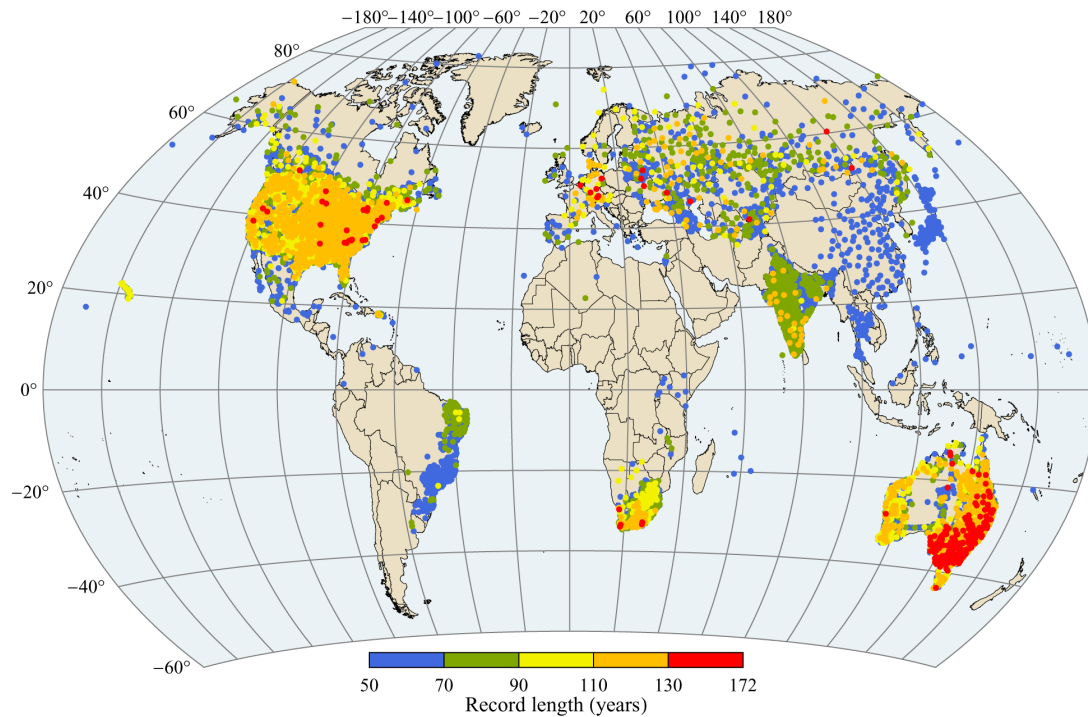
Fig. 2. Locations of the stations studied (a total of 15 137 daily rainfall records with time series length greater than 50 yr). Note that there are overlaps with points corresponding to high record lengths shadowing (being plotted in front of) points of lower record lengths.

3 Defining and fitting the tail

The marginal distribution of rainfall, particularly at small time scales like the daily, belongs to the so-called mixed type distributions, with a discrete part describing the probability of zero rainfall, or the probability dry, and a continuous part expressing the magnitude of the nonzero (wet-day) rainfall. As suggested earlier, studying extreme rainfall requires focusing on the behaviour of the distribution's right tail, which governs the frequency and the magnitude of extremes.

If we denote the rainfall with X, and the nonzero rainfall with $X|X > 0$, then the exceedence probability function (EPF; also known as survival function, complementary distribution function, or tail function) of the nonzero rainfall, using common notation, is defined as

$$P(X > x | X > 0) = \bar{F}_{X|X>0}(x) = 1 - F_{X|X>0}(x), \quad (2)$$

where $F_{X|X>0}(x)$ is any valid probability distribution function chosen to describe nonzero rainfall. It should be clear that the unconditional EPF is easily derived if the probability dry p_0 is known: $\bar{F}_X(x) = (1 - p_0)\bar{F}_{X|X>0}(x)$. Since we focus on the continuous part of the distribution, and more specifically on the right tail, from this point on, for notational simplicity we omit the subscript in $\bar{F}_{X|X>0}(x)$ denoting the conditional EPF function simply as $\bar{F}(x)$. To avoid ambiguity due to the term "tail function" for EPF, we clarify that we

Table 1. Some basic statistics of the 15 137 records of daily rainfall. Apart from probability dry (P_{dry}), these statistics are for the nonzero daily rainfall.

	P_{dry} (%)	No. of nonzero values	Median (mm)	Mean (mm)	SD (mm)	Skew
min	15.11	320	0.40	1.00	1.76	1.37
Q_5	53.92	2121	1.70	3.61	5.01	2.36
Q_{25}	68.55	4038	3.00	6.18	8.28	2.85
Q_{50} (Median)	76.35	5973	4.80	9.27	12.08	3.28
Q_{75}	83.65	8497	6.90	12.65	16.42	3.94
Q_{95}	91.36	13 060	10.20	17.75	24.25	5.38
max	98.25	27 867	25.70	83.96	158.02	26.31
Mean	75.13	6604	5.18	9.77	12.97	3.56
SD	11.46	3508	2.70	4.60	6.20	1.31
Skew	−0.74	1.12	1.03	1.16	1.88	5.58

use the term "tail" to refer only to the upper part of the EPF, i.e., the part that describes the extremes.

At this point, however, we need to define what we consider as the upper part. A common practice is to set a lower threshold value x_L (see, e.g., Cunnane, 1973; Tavares and Da Silva, 1983; Ben-Zvi, 2009) and study the behaviour for values greater than x_L. Yet, there is no universally accepted method to choose this lower value. A commonly accepted method (known as partial duration series method) is to determine the threshold indirectly based on the empirical distribution, in such a way that the number of values above the

threshold equals the number of years N of the record (see, e.g., Cunnane, 1973). The resulting series, defined in this way, is known in the literature as annual exceedance series and is a standard method for studying extremes in hydrology (see, e.g., Chow, 1964; Gupta, 2011).

This may look similar to another common method in which the N annual maxima of the N years are extracted and studied. However, the method of annual maxima, by selecting the maximum value of each year, may distort the tail behaviour (e.g., when the three largest daily values occur within a single year, it only takes into account the largest of them). For this reason, instead of studying the N daily annual maxima, here we focus on the N largest daily values of the record, assuming that these values are representative of the distribution's tail and can provide information for its behaviour. Thus, the method adopted here has the advantage of better representing the exact tail of the parent distribution.

It is worth noting that a common method of studying series above a threshold value is based on the results obtained by Balkema and de Haan (1974) and Pickands III (1975). Loosely speaking, according to these results the conditional distribution above the threshold converges to the Generalized Pareto as the threshold tends to infinity. The latter includes, as a special case, the Exponential distribution. We note, though, that these results are asymptotic results, i.e., valid (or providing a good approximation) if this threshold value tends to infinity (or if it is very large). In the case where the parent distribution is of power type or of exponential type, the theory is applicable even for not so large threshold values because the convergence of the tail is fast. In other cases, e.g., Lognormal or Stretched Exponential distributions, the convergence is very slow. The same applies to the classical extreme value theory (EVT), which predicts that the distribution of maxima converges to one of the three extreme value distributions. For some examples illustrating the slow convergence to the asymptotic distributions of EVT (the same philosophy applies for Balkema–de Haan–Pickands theorem), see, e.g., Papalexiou and Koutsoyiannis (2013) and Koutsoyiannis (2004a).

Given that each station has an N-year record of daily values and a total number n of nonzero values, we define the empirical EPF $\bar{F}_N(x_i)$, conditional on nonzero rainfall, as the empirical probability of exceedance (according to the Weibull plotting position):

$$\bar{F}_N(x_i) = 1 - \frac{r(x_i)}{n+1}, \tag{3}$$

where $r(x_i)$ is the rank of the value x_i, i.e., the position of x_i in the ordered sample $x_{(1)} \leq ... \leq x_{(n)}$ of the nonzero values. Thus, the empirical tail is determined by the N largest nonzero rainfall values of $\bar{F}_N(x_i)$ with $n - N + 1 \leq i \leq n$ (note that $x_L = x_{(n-N+1)}$). Some basic summary statistics of the series of the N largest nonzero rainfall values are presented in Table 2.

Obviously the number of nonzero daily rainfall values is $n = (1 - p_0)n_d N$ where $n_d = 365.25$ is the average number of the days in a year. According to the Weibull plotting position given in Eq. (3), the exceedence probability $\bar{p}(x_L)$ of x_L will be

$$\bar{p}(x_L) = 1 - \frac{n - N + 1}{n + 1} = \frac{N}{(1 - p_0)n_d N + 1} \approx \frac{1}{(1 - p_0)n_d}. \tag{4}$$

This shows that the exceedence probability of the threshold x_L depends only on the probability dry p_0. Interestingly, the average p_0 of the records analysed in this study is approximately 0.75, which implies that the exceedence probability of x_L is on average as low as 0.01, while even for $p_0 = 0.95$ its value is 0.055. We deem that values above this threshold can be assumed to belong to the tail of the distribution. We note that there are studies (see e.g., Beguería et al., 2009) in which the threshold value was chosen to correspond to the 90th percentile, a value much smaller than the one corresponding to our choice of threshold. In Sect. 6 we discuss further the selection of the threshold, also in comparison with different methods of selection.

The fitting method we follow here is straightforward, i.e., we directly fit and compare the performance of different theoretical distribution tails to the empirical tails estimated from the daily rainfall records previously described. The theoretical tails are fitted to the empirical ones by minimizing numerically a modified mean square error (MSE) norm N1 defined as

$$N1 = \frac{1}{N} \sum_{i=n-N+1}^{n} \left(\frac{\bar{F}(x_{(i)})}{\bar{F}_N(x_{(i)})} - 1 \right)^2. \tag{5}$$

A complete verification of the method and a comparison with other norms is presented in Sect. 6. Here we only note that its rationale (and advantage over classical square error norms) is that it properly "weights" each point that contributes in the sum. Namely, it considers the relative error between the theoretical and the empirical values rather than using the x values themselves. For example, if we consider the classical square error, i.e., $(x_i - x_u)^2$, with x_u denoting the quantile value for probability u equal to the empirical probability of the value x_i, then large values would contribute much more to the total error than the smaller ones. This may be a problem especially for rainfall records where the values usually differ more than one order of magnitude, e.g., from 0.1 mm to more than 100 mm. Obviously, the best fitted tail for a specific record is considered to be the one with the smallest MSE.

The proposed approach, which fits the theoretical distribution only to the largest points of each dataset, ensures that the fitted distribution provides the best possible description of the tail and is not affected by lower values. As an example of the fitting method, Fig. 3 depicts the Weibull distribution fitted to an empirical sample (the station was randomly selected and has code IN00121070) by minimizing the norm
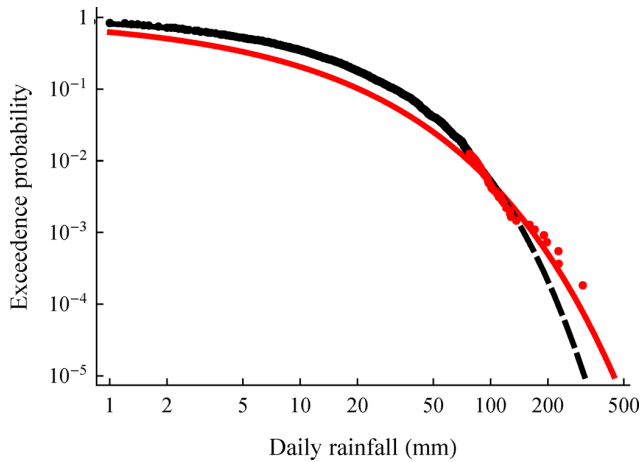
Fig. 3. Explanatory diagram of the fitting approach followed. The dashed line depicts a Weibull distribution fitted to the whole empirical distribution points, while the solid red line depicts the distribution fitted only to the tail points.

given by Eq. (5) in two ways: (a) in all the points of the empirical distribution, and (b) in only the largest N points. It is clear that the first approach (dashed line) does not adequately describe the tail.

It is well known that several other methods have been extensively used to estimate the parameters of candidate distributions, e.g., the lognormal maximum likelihood and the log-probability plot regression (Kroll and Stedinger, 1996), and more recently the log partial probability weighted moments and the partial L-moments (Wang, 1996; Bhattarai, 2004; Moisello, 2007). Yet, the advantage of the proposed method is that any tail can be fitted in the same manner and can be directly compared with other fitted tails since the resulting MSE value can clearly indicate the best fitted; in the aforementioned methods an additional measure has to be estimated in order to compare the performance of the fitted distributions.

4 The fitted distribution tails

It is clear from the previous section that any tail can be fitted to the empirical ones. Nevertheless, in this study we fit and compare the performance of four different and common distribution tails, i.e., the tails of the Pareto type II (PII) the Lognormal (LN), the Weibull (W), and the Gamma (G) distributions. These distributions were chosen for their simplicity, popularity, as well as for being tail-equivalent (or for having similar asymptotic behaviour) with many other more complicated distributions. It is reminded that two distribution functions F and G with support unbounded to the right are called tail-equivalent if $\lim_{x\to\infty} \bar{F}(x)/\bar{G}(x) = c$ with $0 < c < \infty$.

The Pareto and the Lognormal distributions belong to the subexponential class and are considered heavy-tailed

Table 2. Some basic statistics of the 15 137 tail samples defined for an N-year record as the N largest nonzero values.

	No. of tail values	Median (mm)	Mean (mm)	SD (mm)	Max (mm)
min	50	8.90	10.42	3.01	21.50
Q_5	52	28.30	31.71	8.61	68.60
Q_{25}	61	43.55	48.24	13.85	110.00
Q_{50} (Median)	70	62.75	69.12	19.01	152.40
Q_{75}	97	85.30	93.72	27.59	218.40
Q_{95}	122	130.30	144.70	47.48	357.60
max	172	977.00	1041.02	395.96	1750.00
Mean	79	68.78	76.01	22.50	175.06
SD	23	34.84	38.20	13.21	93.42
Skew	0.80	2.73	2.58	3.55	1.79

distributions; the Weibull can belong to both classes, depending on the values of its shape parameter, while the gamma distribution has essentially an exponential tail but not precisely (see below). From a practical point of view, the ordering of these distributions, from heavier to lighter tail, is Pareto, Lognormal, Weibull with shape parameter < 1, Gamma and Weibull with shape parameter > 1 (see, e.g., El Adlouni et al., 2008). Note that Pareto is the only power-type distribution while the other three are of exponential form.

Specifically, the Pareto type II distribution is the simplest power-type distribution defined in $[0,\infty)$. Its probability density function (PDF) and EPF are given, respectively, by

$$f_{\mathrm{PII}}(x) = \frac{1}{\beta}\left(1 + \gamma \frac{x}{\beta}\right)^{-\frac{1}{\gamma}-1} \tag{6}$$

$$\bar{F}_{\mathrm{PII}}(x) = \left(1 + \gamma \frac{x}{\beta}\right)^{-\frac{1}{\gamma}}, \tag{7}$$

and it is defined by the scale parameter $\beta > 0$ and the shape parameter $\gamma \geq 0$ that controls the asymptotic behaviour of the tail. Namely, as the value of γ increases, the tail becomes heavier and consequently extreme values occur more frequently. For $\gamma = 0$ it degenerates to the exponential tail while for $\gamma \geq 0.5$ the distribution has infinite variance. Many other power-type distributions are tail-equivalent, i.e., exhibiting asymptotic behaviour similar to $x^{-1/\gamma}$ with the Pareto type II tail, e.g., the Burr type XII (Burr, 1942; Tadikamalla, 1980), the two- and three-parameter Kappa (Mielke, 1973), the Log-Logistic (e.g., Ahmad et al., 1988) and the Generalized Beta of the second kind (Mielke Jr. and Johnson, 1974).

Another very common distribution used in hydrology is the Lognormal with PDF and EPF, respectively,

$$f_{\mathrm{LN}}(x) = \frac{1}{\sqrt{\pi}\gamma x}\exp\left(-\ln^2\left(\frac{x}{\beta}\right)^{1/\gamma}\right) \tag{8}$$

$$\bar{F}_{\mathrm{LN}}(x) = \frac{1}{2}\mathrm{erfc}\left(\ln\left(\frac{x}{\beta}\right)^{1/\gamma}\right) \tag{9}$$

where $\mathrm{erfc}\,(x) = 2\pi^{-1/2}\int_x^\infty e^{-t^2}\mathrm{d}t$. The distribution comprises the scale parameter $\beta > 0$ and the parameter $\gamma > 0$ that controls the shape and the behaviour of the tail. Lognormal is also considered a heavy-tailed distribution (it belongs to the subexponential family) and can approximate power-law distributions for a large portion of the distribution's body (Mitzenmacher, 2004). Notice that the notation in Eqs. (8) and (9) differs from the common one and illustrates more clearly the kind of the two parameters (scale and shape).

The Weibull distribution, which can be considered as a generalization of the exponential distribution, is another common model in hydrology (Heo et al., 2001a, b) and its PDF and EPF are given, respectively, by

$$f_{\mathrm{W}}(x) = \frac{\gamma}{\beta}\left(\frac{x}{\beta}\right)^{\gamma-1}\exp\left(-\left(\frac{x}{\beta}\right)^\gamma\right) \qquad (10)$$

$$\bar{F}_{\mathrm{W}}(x) = \exp\left(-\left(\frac{x}{\beta}\right)^\gamma\right). \qquad (11)$$

The parameter $\beta > 0$ is a scale parameter, while the shape parameter $\gamma > 0$ governs also the tail's asymptotic behaviour. For $\gamma < 1$ the distribution belongs to the subexponential family with a tail heavier than the exponential one, while for $\gamma > 1$ the distribution is characterized as hyperexponential with a tail thinner than the exponential. Many distributions can be assumed tail-equivalent with the Weibull for a specific value of the parameter γ, e.g., the Generalized Exponential, the Logistic and the Normal.

Finally, one of the most popular models for describing daily rainfall is the Gamma distribution (e.g., Buishand, 1978), which, like the Weibull distribution, belongs to the exponential family. Its PDF and EPF are given, respectively, by

$$f_{\mathrm{G}}(x) = \frac{1}{\beta\Gamma(\gamma)}\left(\frac{x}{\beta}\right)^{\gamma-1}\exp\left(-\frac{x}{\beta}\right) \qquad (12)$$

$$\bar{F}_{\mathrm{G}}(x) = \Gamma\left(\gamma, \frac{x}{\beta}\right)/\Gamma(\gamma) \qquad (13)$$

with $\Gamma(s, x) = \int_x^\infty t^{s-1}e^{-t}\mathrm{d}t$ and $\Gamma(s) = \int_0^\infty t^{s-1}e^{-t}\mathrm{d}t$. Generally, we can assume that the Gamma tail behaves similar to the exponential tail. Yet, this is only approximately correct as the Gamma distribution belongs to a class of distributions (denoted as $S(\gamma)$; see, e.g., Embrechts and Goldie, 1982; Klüppelberg, 1989; Alsmeyer and Sgibnev, 1998) that irrespective of its parameter values cannot be classified as subexponential, while it is not tail-equivalent with the exponential. This can be seen from the fact that the $\lim_{x\to\infty}\bar{F}_{\mathrm{G}}(x)/\bar{G}(x)$ is 0 for $\beta < \beta_{\mathrm{E}}$ and ∞ for $\beta > \beta_{\mathrm{E}}$, where $\bar{G}(x) = \exp(-x/\beta_{\mathrm{E}})$ is the exponential tail. Yet it is noted that if compared with an exponential tail with $\beta = \beta_{\mathrm{E}}$, then

$$\lim_{x\to\infty}\frac{\bar{F}(x)}{\bar{G}(x)} = \begin{cases} 0 & 0 < \gamma < 1 \\ 1 & \gamma = 1 \\ \infty & \gamma > 1 \end{cases}. \qquad (14)$$

Therefore, in this case and practically speaking, for $0 < \gamma < 1$ the Gamma distribution has a "slightly lighter" tail than the exponential tail as it decreases faster, while for $\gamma > 1$ it exhibits a "slightly heavier" tail as it decreases more slowly than the exponential tail.

All four distributions we compare here, and consequently their tails, have similarities in their structure as all have two parameters and specifically one scale parameter and one shape parameter. Nevertheless, among the various distributions with the same parameter structure, inevitably some are more flexible than others. One way to quantify this flexibility is by comparing them in terms of various shape measures (e.g., skewness, kurtosis, etc.). For example, the feasible ranges of skewness for the Pareto, Lognormal, Weibull and Gamma are, respectively, $(2, \infty)$, $(0, \infty)$, $(-1.14, \infty)$ and $(0, \infty)$. Therefore, the Weibull distribution seems to be the most "flexible" distribution among them and the Pareto the least. Yet this argument is not valid when we focus on the tail because the general shape of the tail is basically similar and what differs is the rate at which the tail approaches zero.

5 Results and discussion

The basic statistical results from fitting the four distribution tails, following the methodology described, to the 15 029 daily rainfall records are given in Table 3. In order to assess which tail has the best fit, the four tails were compared in couples in terms of the resulting MSE, i.e., the tail with the smaller MSE is considered better fitted. As shown in Fig. 4, the Pareto tail, when compared with the other three distributions, was better fitted in about 60 % of the stations. Interestingly, the distribution with the heavier tail of each couple, in all cases, was better fitted in a higher percentage of the stations, which implies a rule of thumb of the type "the heavier, the better"!

Another comparison revealing the overall performance of the fitted tails was based on their average rank. That is, the fitted tails in each record were ranked according to their MSE, i.e., the tail with the smaller MSE was ranked as 1 and the one with the largest as 4. Figure 5 depicts the average rank of each tail for all stations. Again, the Pareto performed best, while the most popular model for rainfall, the Gamma distribution, performed worst. The percentages of each distribution tail that was best fitted are 30.7 % for Pareto, 29.8 % for Lognormal, 13.6 % for Weibull and 25.8 % for Gamma. Again, the Pareto distribution is best according to these percentages; interestingly, however, the Gamma distribution has a relatively high percentage, higher than the Weibull. This does not contradict the conclusion derived by the average

Table 3. Summary statistics from the fitting of the four distribution tails into the 15 029 tail-samples of daily rainfall (expressed in mm).

	Pareto			Lognormal		
	MSE	β	γ	MSE	β	γ
Min	0.002	0.42	0.001	0.002	1.22	0.531
Mode*	0.011	7.54	0.134	0.012	8.78	1.060
Mean	0.017	8.80	0.140	0.018	9.46	1.087
Median	0.021	9.51	0.145	0.022	10.59	1.107
Max	0.336	54.79	0.797	0.322	76.74	2.284
SD	0.015	4.92	0.076	0.015	6.44	0.214
Skew	2.910	1.23	0.495	2.755	1.73	0.561

	Weibull			Gamma		
	MSE	β	γ	MSE	β	γ
Min	0.002	0.02	0.230	0.002	3.79	0.010
Mode	0.013	4.33	0.661	0.015	17.50	0.092
Mean	0.019	5.91	0.678	0.023	23.15	0.219
Median	0.022	6.88	0.692	0.032	28.18	0.294
Max	0.298	52.72	1.491	0.482	120.00	2.433
SD	0.015	4.69	0.139	0.034	17.30	0.269
Skew	2.151	1.82	0.668	4.377	1.65	2.567

* The mode was estimated from the empirical density function (histogram) after smoothing.


Fig. 4. Comparison of the fitted tails in couples in terms of the resulting MSE. The heavier tail of each couple is better fitted to the empirical points in a higher percentage of the records.

rank. The explanation is that the Gamma distribution was ranked as best in some cases, but when it was not the best fitted, it was probably the worst fitted.

Figure 6 depicts the empirical distributions of the shape parameters of the fitted tails. It is well-known that the most probable values are the ones around the mode, which for the Pareto shape parameter is 0.134. Interestingly, this value is close to the one determined in a different context by Koutsoyiannis (1999) using Hershfield's (1961) dataset. This implies that power-type distributions, which asymptotically behave like the Pareto, will not have finite power moments of order greater than $1/0.134 \approx 7.5$. Moreover, as the empirical distribution of the Pareto shape parameter in Fig. 6 attests, values around 0.2 are also common, implying non-existence of moments greater than the fifth order. We should thus bear in mind that sample moments of that or higher order (sometimes appearing in research papers) may not exist. Regarding the Weibull tail, the estimated mode of its shape parameter is 0.661, implying a much heavier tail compared to the exponential one. Finally, it is worth noting that the estimated mode of the Gamma shape parameter is as low as 0.092. The shape parameter of the Gamma distribution controls mainly the behaviour of the left tail, resulting in J- or bell-shaped densities (loosely speaking, the right tail is dominated by the exponential function and thus behaves like an exponential tail). A value that low corresponds to an extraordinarily J-shaped density, which would be unrealistic for describing the whole distribution body of daily rainfall. In other words,

a Gamma distribution fitted to the whole set of points would most probably underestimate the behaviour of extremes.

We searched for the existence of any geographical patterns, potentially defining climatic zones, in the best fitted tails, i.e., the existence of zones in the world where the majority of the records were better described by one of the studied distribution tails. The maps in Fig. 7, which depict the locations of the stations where each distribution tail was best fitted, did not unveil any regular patterns in terms of the best fitted distribution but rather seem to follow a random variation.

Another way to investigate for geographical patterns, as the previous map did not reveal any useful information, is to study the fitted tails grouped into two coarser groups: the subexponential group and the exponential-hyperexponential group. The former includes the Pareto, the Lognormal and the Weibull with $\gamma < 1$ tails, while the latter includes the Gamma and the Weibull with $\gamma \geq 1$ tails. Among the 15 029 records, subexponential tails were best fitted in 10 911 cases or in 72.6 % while exponential-hyperexponential tails were best fitted in 4 118 or in 27.4 %. Further, in order to get a clearer picture instead of constructing maps with the locations where the first-group or the second-group tails were best fitted, we studied the percentage of subexponential tails that were best fitted in large regions. Specifically, we constructed a grid covering the entire earth using a latitude difference $\Delta\varphi = 2.5°$ and longitude difference $\Delta\lambda = 5°$. In each grid cell we estimated the percentage of the best fitted subexponential tails simply by counting the number of the best fitted subexponential tails divided by the total number of records within the cell. We present these percentages in the form of a map in Fig. 8, using a colour scale as shown in the map's legend. The cells plotted in the map are those containing at least two records, so that the calculation of percentages have some meaning.
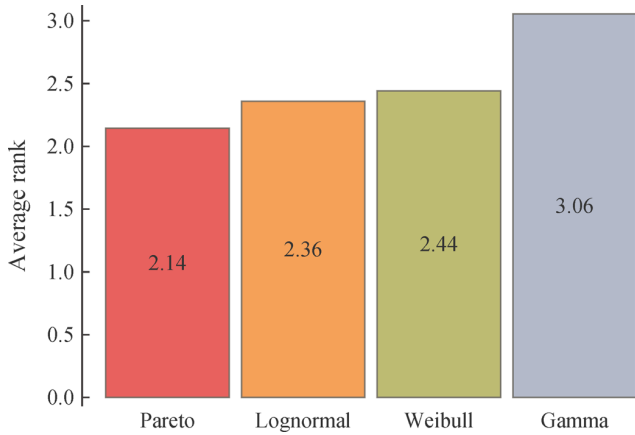
Fig. 5. Mean ranks of the tails for all records. The best-fitted tail is ranked as 1 while the worst-fitted as 4. A lower average rank indicates a better performance.

Fig. 6. Histograms of the shape parameters of the fitted tails.

The map of Fig. 8 clearly shows that in the vast majority of cells subexponential tails dominate (percentage > 60 %). Particularly, out of 532 cells having at least two records, 255 and 163 have percentages of subexponential tails 60–80 % and > 80 %, respectively. In contrast, in only 35 and 79 cells are the percentage values in the ranges 0–40 % and 40–60 %, respectively.

6   Verification of the fitting method

The use of a different norm for fitting the tail into the empirical data could potentially modify the conclusions drawn. Nevertheless, this argument is pointless in the sense that the main concern should be the efficiency of the norm used, i.e., if it possesses desired properties, e.g., if it is unbiased and has lower variance in comparison to other candidates. Usually, the error is expressed in terms of random variable values, e.g., rainfall values, and not in terms of probability. However, a literature search did not reveal or verify that the commonly used norms, e.g., the classical MSE norm, are better than the norm N1 used here (see Eq. 5).

For this reason, we implemented a Monte Carlo scheme, which actually replicates the method we followed, where we evaluate the performance of the norm N1 and also compare it with the more common norms N2 and N3 defined as

$$\text{N2} = \frac{1}{N} \sum_{i=n-N+1}^{n} \left(\frac{x_u}{x_{(i)}} - 1 \right)^2 \tag{15}$$

$$\text{N3} = \frac{1}{N} \sum_{i=n-N+1}^{n} \left(x_u - x_{(i)} \right)^2 . \tag{16}$$

Here, $x_u = Q(u)$ is the value predicted by the quantile function Q of the distribution under study for u equal to the empirical probability of $x_{(i)}$ (the ith element the sample ranked

in ascending order) according to the Weibull plotting position. The norm N2 has the same rationale as the one we used but the error is estimated in terms of rainfall values, rather than in terms of probability, while the norm N3 is the classical and most commonly used MSE norm.

The Monte Carlo scheme we performed can be summarized in the following steps: (a) we generated 1000 random samples from each one of the four distributions we studied with sample size equal to 6600 values, which is approximately the average number of nonzero daily rainfall values per record; (b) we selected the scale and the shape parameter values to be approximately equal with the median values resulted from the analysis of the real world dataset (see Table 3) in order for the generated random samples to be representative of the real data; and (c) we fitted each distribution to its corresponding random sample and estimated the parameters by applying our method for each one of the three norms, while we set N equal to 80 yr, which is approximately the average record length.

The results are presented in Fig. 9. The whiskers of the box plots express the 95 % Monte Carlo confidence interval of the parameters while the dashed lines show the true parameter values. It is clear that the norm N1 we used results in almost unbiased estimation of the parameters while, especially for the Pareto and the Lognormal distributions, it results in markedly smaller variance compared to the classical norm N3. The norm N2 seems to perform very well for the Pareto, Lognormal and Weibull distributions (although somewhat biased) but the results are poor for the Gamma distribution. The classical and the most commonly used norm N3 is by far the worst in term of bias except for the Gamma distribution, for which it performs equally well as N1. In particular, for the subexponential distributions of this simulation, i.e., the Pareto, the Lognormal and the Weibull, the classical norm N3 fails to provide good results. This may point to a more general conclusion, i.e., that the classical MSE, which is inspired based on properties of the normal distribution, is not
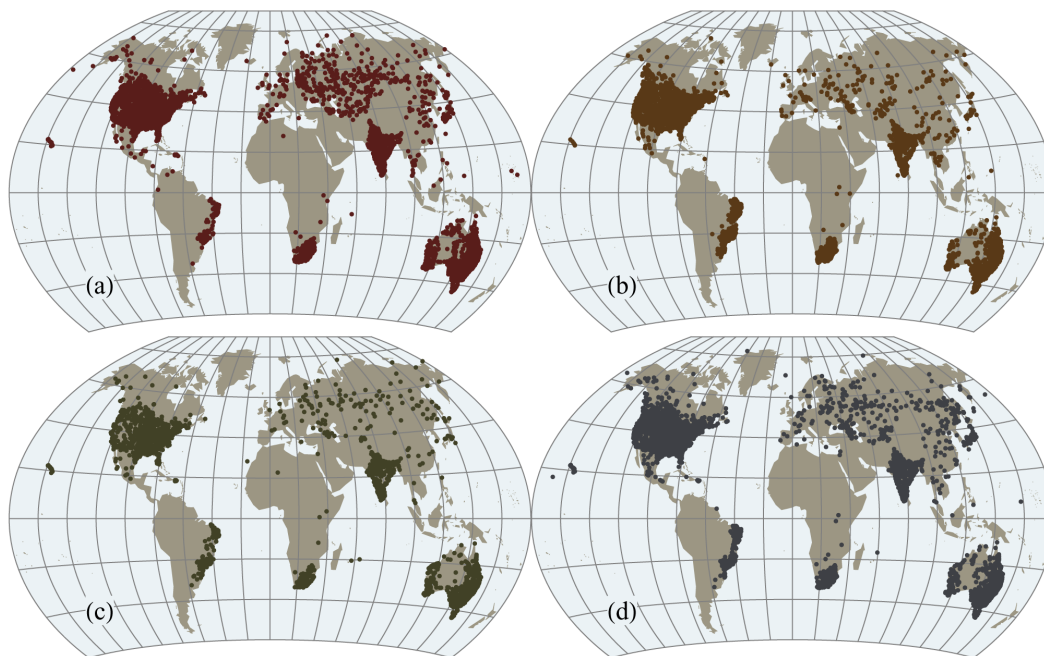
Fig. 7. Geographical depiction of the 15 029 stations where the best fitted tail is **(a)** Pareto in 4621, **(b)** Lognormal in 4486, **(c)** Weibull in 2051, and **(d)** Gamma in 3871.
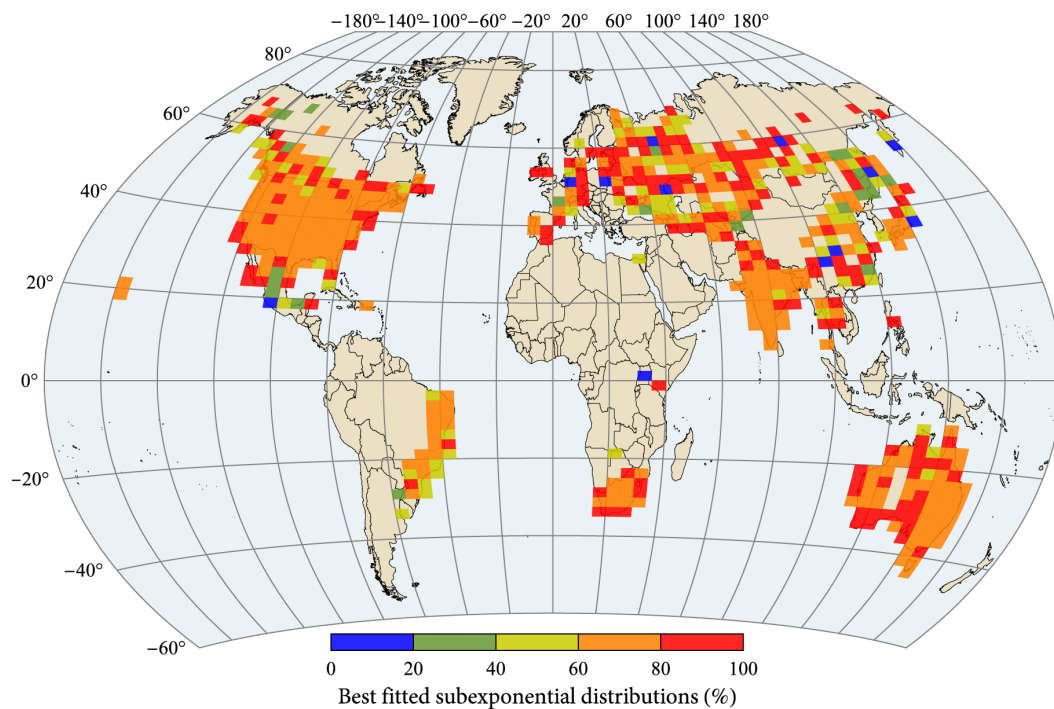
Fig. 8. Geographical variation of the percentage of best fitted subexponential tails in cells defined by latitude difference $\Delta\varphi = 2.5°$ and longitude difference $\Delta\lambda = 5°$. In total, in 72.6 % of the 15 029 records analysed, the subexponential tails were the best fitted.

www.hydrol-earth-syst-sci.net/17/851/2013/
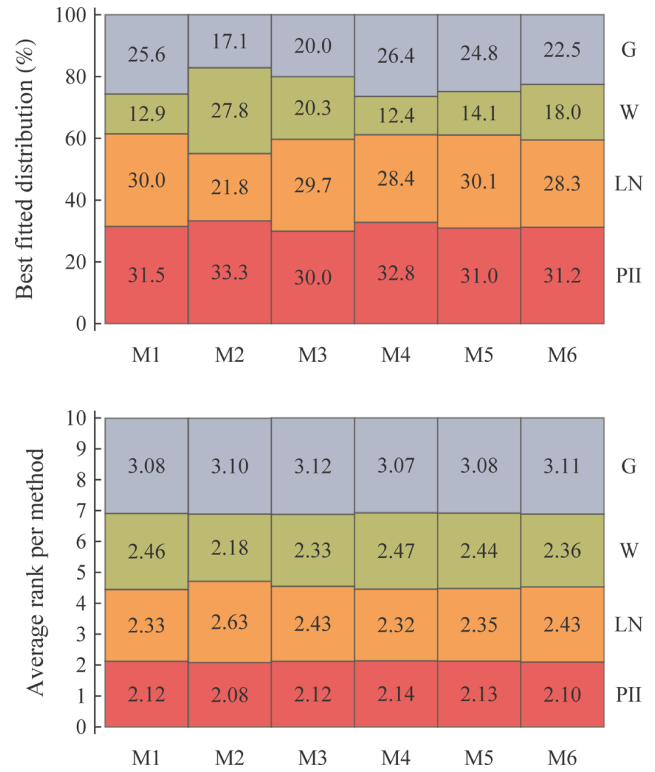
Hydrol. Earth Syst. Sci., 17, 851–862, 2013
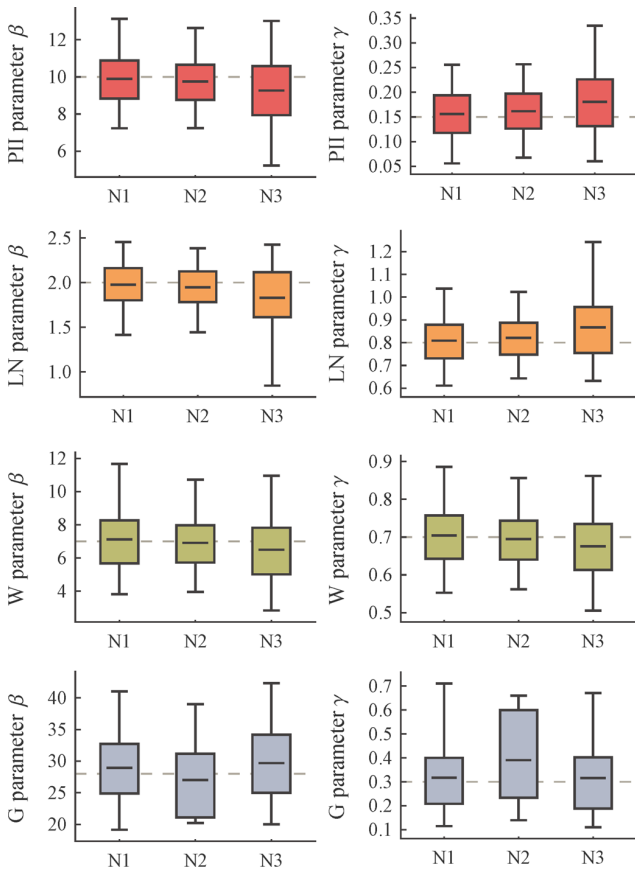
Fig. 9. Results of a Monte Carlo scheme implemented to evaluate the performance of the norm N1 used in fitting of tails in this study, in comparison to commonly used ones (N2, N3).

Fig. 10. Performance results of the four fitted tails in 2000 randomly selected records using six different methods for selecting the sample size: (top panel) percentage of records in which each distribution tail was best fitted; (bottom panel) average ranks of the fitted tails (lower average rank indicates better performance).

a good choice for subexponential distributions. This needs to be further investigated; however, we deem that there is a rationale supporting the following conclusion: subexponential distributions can generate "extremely" extreme values compared to the main "body" of values, and thus, in the classical norm these values will contribute "extremely" to the total error heavily affecting the fitting results.

Another issue of potential concern for the validity of the conclusions drawn is the impact of the sample size, i.e., the number of the largest events N, or equivalently the threshold x_L, for which the four distribution tails are fitted. As mentioned before, we used the annual exceedance series, a standard method in hydrology in which N equals the number of the record's years. Obviously, N can be defined in many different ways, either with reference to record length or as a fixed number for every record studied.

In order to assess the impact of number of events in the performance of the four fitted distribution tails, we selected randomly 2000 records among the 15 029 analysed and we fitted the four distribution tails using six different methods for defining N. The first method (M1) is the one we used for all above analyses, in which N equals the number of the record's years. In the second (M2) and third (M3)

methods we defined the threshold x_L as the 90th- and the 95th-percentiles, respectively, so that N equals the number of events included in the upper 10 % and 5 %, respectively, of the nonzero values. Obviously, in these two methods N varies from record to record depending on the total number of nonzero values, and on the average it equals 667 and 333 values for M2 and M3, respectively. In the rest three methods (M4, M5 and M6) N is defined as a fixed number for every record, i.e., 50, 100 and 200 values, respectively.

The performance results comparing the six methods are summarized in Fig. 10, which depicts (a) the percentage of cases in which each distribution was best fitted and (b) the average rank of each distribution tail. Again the Pareto II tail was best fitted in a higher percentage of records in all cases (M1–M6) with the percentage values varying in a narrow range. The results are essentially the same with those obtained from the analysis of the whole database. The only noticeable difference regards the method M2, in which the Weibull tail sometimes seems to "gain ground" over the Gamma and the Lognormal tails. In general it seems that the Weibull tail improves its performance as N increases. Thus, in M4 where N has the lowest value, i.e., 50 values, it performs the worst, while in M2 where N is maximum (667 values on the average), it performs the best. The average

rank, which is a better measure of the overall performance of the distribution tails, remains essentially the same for each distribution in all methods. An exception is observed again in M2 where the Weibull tail performs better than the Lognormal tail. Apart from this exception the general conclusion is again that the Pareto II performs the best, followed by the Lognormal and the Weibull tails, while the Gamma tail performs the worst in all cases.

7 Summary and conclusions

Daily rainfall records from 15 029 stations are used to investigate the performance of four common tails that correspond to the Pareto, the Weibull, the Lognormal and the Gamma distributions. These theoretical tails were fitted to the empirical tails of the records and their ability to adequately capture the behaviour of extreme events was quantified by comparing the resulting MSE. The ranking from best to worst in terms of their performance is (a) the Pareto, (b) the Lognormal, (c) the Weibull, and (d) the Gamma distributions. The analysis suggests that heavier-tailed distributions in general performed better than their lighter-tailed counterparts. Particularly, in 72.6 % of the records subexponential tails were better fitted while the exponential-hyperexponential tails were better fitted is only 27.4 %. It is instructive that the most popular model used in practice, the Gamma distribution, performed the worst, revealing that the use of this distribution underestimates in general the frequency and the magnitude of extreme events. Nevertheless, we must not neglect the fact that the Gamma distribution was the best fitted in 25.8 % of the records.

Additionally, we note that heavy tails tend to be hidden (see, e.g., Koutsoyiannis, 2004a, b; Papalexiou and Koutsoyiannis, 2013), especially when the sample size is small. Thus, we believe that even in the cases where the Gamma tail performed well, the true underlying distribution tail may be heavier. This leads to the recommendation that heavy-tailed distributions are preferable as a means to model extreme rainfall events worldwide. We also note that the tails studied here are as simple as possible, i.e., only one shape parameter controls their asymptotic behaviour. Yet there are many distributions with more than one shape parameters which may affect their tail behaviour. Particularly, the Generalized Gamma (Stacy, 1962) and the Burr type XII distributions were compared as candidates for the daily rainfall (based on L-moments) in anonther study, using thousands of empirical daily records and the former performed better (Papalexiou and Koutsoyiannis, 2012).

The key implication of this analysis is that the frequency and the magnitude of extreme events have generally been underestimated in the past. Engineering practice needs to acknowledge that extreme events are not as rare as previously thought and to shift toward more heavy-tailed probability distributions.

References

Ahmad, M. I., Sinclair, C. D., and Werritty, A.: Log-logistic flood frequency analysis, J. Hydrol., 98, 205–224, doi:10.1016/0022-1694(88)90015-7, 1988.

Alsmeyer, G. and Sgibnev, M.: On the tail behaviour of the supremum of a random walk defined on a Markov chain, available at: http://kamome.lib.ynu.ac.jp/dspace/handle/10131/5689 (last access: 10 November 2012), 1998.

Balkema, A. A. and De Haan, L.: Residual Life Time at Great Age, Ann. Probab., 2, 792–804, doi:10.1214/aop/1176996548, 1974.

Beguería, S., Vicente-Serrano, S. M., López-Moreno, J. I., and García-Ruiz, J. M.: Annual and seasonal mapping of peak intensity, magnitude and duration of extreme precipitation events across a climatic gradient, northeast Spain, Int. J. Climatol., 29, 1759–1779, 2009.

Ben-Zvi, A.: Rainfall intensity–duration–frequency relationships derived from large partial duration series, J. Hydrol., 367, 104–114, doi:10.1016/j.jhydrol.2009.01.007, 2009.

Bhattarai, K. P.: Partial L-moments for the analysis of censored flood samples, Hydrolog. Sci. J., 49, 855–868, 2004.

Buishand, T. A.: Some remarks on the use of daily rainfall models, J. Hydrol., 36, 295–308, doi:10.1016/0022-1694(78)90150-6, 1978.

Burr, I. W.: Cumulative Frequency Functions, Ann. Math. Stat., 13, 215–232, 1942.

Chow, V. T.: Handbook of applied hydrology: a compendium of water-resources technology, McGraw-Hill, 1964.

Cunnane, C.: A particular comparison of annual maxima and partial duration series methods of flood frequency prediction, J. Hydrol., 18, 257–271, doi:10.1016/0022-1694(73)90051-6, 1973.

El Adlouni, S., Bobée, B., and Ouarda, T. B. M. J.: On the tails of extreme event distributions in hydrology, J. Hydrol., 355, 16–33, doi:10.1016/j.jhydrol.2008.02.011, 2008.

Embrechts, P. and Goldie, C. M.: On convolution tails, Stoch. Proc. Appl., 13, 263–278, doi:10.1016/0304-4149(82)90013-8, 1982.

Embrechts, P., Klüppelberg, C., and Mikosch, T.: Modelling extremal events for insurance and finance, Springer Verlag, Berlin Heidelberg, 1997.

European Commission: Directive 2007/60/EC of the European Parliament and of the Council of 23 October 2007 on the assessment and management of flood risks, Official Journal of the European Communities, L, 288(6.11), 27–34, 2007.

Goldie, C. M. and Klüppelberg, C.: Subexponential distributions, in: A Practical Guide to Heavy Tails: Statistical Techniques and Applications, edited by: Adler, R., Feldman, R., and Taggu, M. S., 435–459, Birkhäuser Boston, 1998.

Gupta, S. K.: Modern Hydrology and Sustainable Water Development, John Wiley & Sons, 2011.

Heo, J. H., Boes, D. C., and Salas, J. D.: Regional flood frequency analysis based on a Weibull model: Part 1. Estimation and asymptotic variances, J. Hydrol., 242, 157–170, 2001a.

Heo, J. H., Salas, J. D., and Boes, D. C.: Regional flood frequency analysis based on a Weibull model: Part 2. Simulations and applications, J. Hydrol., 242, 171–182, 2001b.

Hershfield, D. M.: Estimating the probable maximum precipitation, J. Hydraul. Eng.-ASCE, 87, 99–106, 1961.

Klüppelberg, C.: Subexponential Distributions and Integrated Tails, J. Appl. Probab., 25, 132–141, doi:10.2307/3214240, 1988.

Klüppelberg, C.: Subexponential distributions and characterizations of related classes, Probab. Theory Rel., 82, 259–269, doi:10.1007/BF00354763, 1989.

Koutsoyiannis, D.: A probabilistic view of Hershfield's method for estimating probable maximum precipitation, Water Resour. Res., 35, 1313–1322, 1999.

Koutsoyiannis, D.: Statistics of extremes and estimation of extreme rainfall, 1, Theoretical investigation, Hydrolog. Sci. J., 49, 575–590, 2004a.

Koutsoyiannis, D.: Statistics of extremes and estimation of extreme rainfall, 2, Empirical investigation of long rainfall records, Hydrolog. Sci. J., 49, 591–610, 2004b.

Kroll, C. N. and Stedinger, J. R.: Estimation of moments and quantiles using censored data, Water Resour. Res., 32, 1005–1012, 1996.

Mielke Jr., P. W.: Another Family of Distributions for Describing and Analyzing Precipitation Data, J. Appl. Meteorol., 12, 275–280, 1973.

Mielke Jr., P. W. and Johnson, E. S.: Some generalized beta distributions of the second kind having desirable application features in hydrology and meteorology, Water Resour. Res., 10, 223–226, 1974.

Mitzenmacher, M.: A brief history of generative models for power law and lognormal distributions, Internet Mathematics, 1, 226–251, 2004.

Moisello, U.: On the use of partial probability weighted moments in the analysis of hydrological extremes, Hydrol. Process., 21, 1265–1279, 2007.

Papalexiou, S. M. and Koutsoyiannis, D.: Entropy based derivation of probability distributions: A case study to daily rainfall, Adv. Water Resour., 45, 51–57, doi:10.1016/j.advwatres.2011.11.007, 2012.

Papalexiou, S. M. and Koutsoyiannis, D.: Battle of extreme value distributions: A global survey on extreme daily rainfall, Water Resour. Res., online first, doi:10.1029/2012WR012557, 2013.

Pickands III, J.: Statistical Inference Using Extreme Order Statistics, Ann. Stat., 3, 119–131, 1975.

Stacy, E. W.: A Generalization of the Gamma Distribution, Ann. Math. Stat., 33, 1187–1192, 1962.

Tadikamalla, P. R.: A Look at the Burr and Related Distributions, Int. Stat. Rev., 48, 337–344, 1980.

Tavares, L. V. and Da Silva, J. E.: Partial duration series method revisited, J. Hydrol., 64, 1–14, doi:10.1016/0022-1694(83)90056-2, 1983.

Teugels, J.: Class of subexponential distributions, Ann. Probab., 3, 1000–1011, doi:10.1214/aop/1176996225, 1975.

Wang, Q. J.: Using partial probability weighted moments to fit the extreme value distributions to censored samples, Water Resour. Res., 32, 1767–1771, 1996.

Werner, T. and Upper, C.: Time variation in the tail behavior of Bund future returns, J. Future Markets, 24, 387–398, 2004.