Hydrol. Earth Syst. Sci. Discuss., 9, 5757–5778, 2012
www.hydrol-earth-syst-sci-discuss.net/9/5757/2012/
doi:10.5194/hessd-9-5757-2012

Hydrology and Earth System Sciences Discussions

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, 15780 Zographou, Greece

Received: 6 April 2012 – Accepted: 20 April 2012 – Published: 2 May 2012

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

Published by Copernicus Publications on behalf of the European Geosciences Union.

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 in two families: heavy-tailed and light-tailed distributions, with the latter generating more "mild" and infrequent extremes compared to the former. This emphasizes how important for hydrological design is to assess correctly the tail behaviour. Traditionally, the wet-day daily rainfall has been described by light-tailed distributions like the Gamma, although heavier-tailed distributions have also been proposed and used, e.g. the Lognormal, the Pareto, the Kappa, and others. Here, we investigate the issue of 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 163 yr. The analysis shows that heavier-tailed distributions are in better agreement with the observed rainfall extremes than the more often used lighter tailed distributios, with 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" (EC, 2007).

Rainfall is a geophysical process, and although short term prediction is possible to a degree (and useful for operational purposes), long term prediction, that would be valuable for infrastructure design, is not. 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 (a) to select a priori some of the many parametric families of distributions, (b) estimate the parameters according to one of the many 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 the asymptotic behaviour of their tail in two general classes, the sub-exponential class with probability densities tending to zero less rapidly than the exponential density, and the hyper-exponential class, with densities having tails approaching zero more rapidly than the exponential (Teugels, 1975; Klüppelberg, 1989, 1988). Yet, a unique classification does not exist (see e.g. El Adlouni et al., 2008 and references therein), while many 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 exponential tails.

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 the adoption of light-tailed models (e.g. the Gamma distribution), fitted on the whole sample of empirical data for design purposes (of for example flood protection schemes) 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 Milke (1973) proposed the use 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 Data

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) which includes data recorded at over 40 000 stations worldwide. Many of these records, however, are too short in length, have many missing data, or, contain data 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) which are 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.

Finally, the statistical procedure, described next, failed to be applied in some stations, for reasons of algorithmic convergence or time limits, resulting in their exclusion from the analysis. The final dataset analysed comprises 15 029 stations globally. The locations of the stations can be seen in Fig. 2.

3 Methods

3.1 Statistical approach

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 non-zero or the 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 non-zero rainfall with $X|X>0$, then the exceedence probability function (EPF; also known as survival function, complementary distribution function, or, tail function) of the non-zero rainfall, using common notation, is defined as

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

where $F_{X|X>0}(x)$ is any valid probability distribution function chosen to describe non-zero rainfall. It should be clear that the unconditional EPF is easily defined in terms of probability dry p_0 as $\bar{F}_X(x)=(1-p_0)\bar{F}_{X|X>0}(x)$. 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, we clarify that we 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 define this lower value. A commonly accepted method (known as partial duration series method) is to define 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).

In comparison to the other common method, in which the annual maxima are studied, the method adopted has the advantage of better representing the exact tail of the parent distribution. In contrast, the annual maxima method 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, the annual maxima method only takes into account the largest of them). For this reason, instead of studying the N daily annual maxima, we focus on the N largest daily values. It is then assumed that these values are representative of the distribution's tail and can provide information for its behaviour.

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

$$\bar{F}_N(x_i)=1-\frac{r(x_i)}{n+1} \tag{2}$$

where $r(x_i)$ is the rank of the value x_i, i.e. the position of x_i in the ordered sample $x_{(1)}\leq\ldots\leq x_{(n)}$ of the non-zero values. Thus the empirical tail is determined by the N largest non-zero rainfall values of $\bar{F}_N(x_i)$ with $n-N+1\leq i\leq n$.

We then 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 defined as

$$\mathrm{MSE} = \frac{1}{N} \sum_{i=n-N+1}^{n} \left[\frac{\bar{F}(x_i)}{\bar{F}_N(x_i)} - 1 \right]^2 . \tag{3}$$

The need for this modified MSE arises from the fact that in the classical MSE norm, in which the error depends on the difference between the theoretical and the empirical values, i.e. $\mathrm{SE} = \left[\bar{F}(x_i) - \bar{F}_N(x_i)\right]^2$, the contribution in the sum of the most important values for our analysis (those with smallest $\bar{F}(x_i)$) is too small. On the contrary, the norm given in Eq. (3) treats each data point equally as it considers the relative error between the theoretical and the empirical values which is independent of the absolute values. 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 N larger 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 given by Eq. (3) 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 noted that several other methods have been used extensively in extreme value estimation, 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 log partial probability weighted moments and partial L-moments (Wang, 1996; Bhattarai, 2004; Moisello, 2007). Yet, the advantage is that the tails fitted by the method proposed here can be directly compared since the resulting MSE

can clearly indicate the best fitted, while in the afore-mentioned methods an additional measure has to be estimated in order to compare the performance of the fitted distributions.

3.2 Statistical approach

In this study we estimate the performance of four different common tail types, the Pareto, the Lognormal, the Weibull and, the Gamma tails, whose asymptotic behaviour may be similar with many other distributions. The Pareto and the Lognormal distributions belong to the sub-exponential class and are considered heavy-tailed distributions. The Weibull and the Gamma distribution, depending on the values of the shape parameter can belong to both classes, but in general their tails are lighter than the Pareto or the Lognormal.

Specifically, the Pareto type II distribution is the simplest power-type distribution defined in $[0, \infty)$. Its EPF is given by

$$\bar{F}(x) = \left(1 + \gamma \frac{x}{\beta}\right)^{-\frac{1}{\gamma}} \tag{4}$$

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, e.g. as the value of γ increases the tail becomes heavier and consequently extreme values are assumed to occur more frequent. For $\gamma \to \infty$ 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 (Tadikamalla, 1980) the two- and three-parameter Kappa (Mielke, 1973) the Log-Logistic (e.g. Ahmad et al., 1988) or the Generalized Beta of the second kind (Mielke and Johnson, 1974).

Another very common distribution used in hydrology is the Lognormal with EPF

$$\bar{F}(x) = \frac{1}{2}\mathrm{erfc}\left(\frac{\ln x - \beta}{\sqrt{2}\gamma}\right) \tag{5}$$

where $\mathrm{erfc}(x) = 2\pi^{-1/2}\int_x^\infty \mathrm{e}^{-t^2}\,\mathrm{d}t$. The distribution comprises the scale parameter $\beta \in \mathbb{R}$ 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 sub-exponential family) and can approximate power-law distributions for a large portion of the distribution's body (Mitzenmacher, 2004)

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

$$\bar{F}(x) = \exp\left(-\frac{x^\gamma}{\beta}\right). \tag{6}$$

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 sub-exponential family with a tail that is heavier than the exponential one, while for $\gamma > 1$ the distribution is characterized as hyper-exponential with a tail thinner that the exponential. Many distributions can be assumed tail-equivalent with the Weibull for a specific value of the parameter γ, e.g. the exponential, 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 EPF is given by

$$\bar{F}(x) = \Gamma(\gamma, \frac{x}{\beta})/\Gamma(\gamma) \tag{7}$$

with $\Gamma(s,x) = \int_x^\infty t^{s-1}\,\mathrm{e}^{-t}\,\mathrm{d}t$ and $\Gamma(s) = \int_0^\infty t^{s-1}\,\mathrm{e}^{-t}\,\mathrm{d}t$. Generally, we can assume that the Gamma tail behaves similar to the exponential tail. Yet, this is only approximately correct as the exponential tail is not asymptotically equivalent with the Gamma tail, i.e. its precise asymptotic behaviour of the right tail depends on the value of the parameter γ, so, for $\gamma > 1$ the distribution is sub-exponential and form and for $\gamma < 1$ hyper exponential.

4 Results

We fit the four previously given distribution tails, following the methodology described, to 15 029 daily rainfall records. The basic statistical results from the fitting are given in Table 1. 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 was better fitted in approximately more than 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 undertaken in terms of 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 the worst. The percentages of best fit distribution tails 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 rank. The explanation is that the Gamma distribution was ranked as best in some cases, but when it was not best, it was probably the worst.

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), data set. 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 Gamma 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), and a value that low corresponds to an extraordinarily J-shaped density that 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.

Finally, 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, that depict the station locations 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.

5 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 Log-Normal and the Gamma distributions. These theoretical tails were fitted to the empirical tails of the records and their ability to capture adequately 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. It is instructive that the most popular model used in practice, the Gamma distribution, performed the worst, revealing that the use if this distribution underestimates the frequency and magnitude of extreme events.

Additionally, we note that heavy tails tend to be hidden (see e.g. Koutsoyiannis, 2004a,b) 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 shape parameters than one that may affect the tail behaviour, e.g. the Generalized Gamma distribution (Stacy, 1962). Particularly, the Generalized Gamma (a non-power type distribution) and the Burr type XII distributions were compared as candidates for the daily rainfall (based on L-moments) in an earlier study, using thousands of empirical daily records and the former performed better (Papalexiou and Koutsoyiannis, 2011).

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 we had thought – and shift to heavy-tailed distributions for their analysis.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

References

Ahmad, M. I., Sinclair, C. D., and Werritty, A.: Log-logistic flood frequency analysis, J. Hydrol., 98, 205–224, 1988.

Ben-Zvi, A.: Rainfall intensity-duration-frequency relationships derived from large partial duration series, J. Hydrol., 367, 104–114, 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, 1978.

Cunnane, C.: A particular comparison of annual maxima and partial duration series methods of flood frequency prediction, J. Hydrol., 18, 257–271, 1973.

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 November 2007, p. 27, 2007.

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, Proc. ASCE, J. Hydraul. Div., 87, 99–106, 1961.

Klüppelberg, C.: Subexponential distributions and integrated tails, J. Appl. Probab., 25, 132–141, 1988.

Klüppelberg, C.: Subexponential distributions and characterizations of related classes, Probab. Theory Rel., 82, 259–269, 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, 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 Math., 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., doi:10.1016/j.advwatres.2011.11.007, in press, 2011.

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, 1983.

Teugels, J.: Class of subexponential distributions, Ann. Probab., 3, 1000–1011, 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. Futures Markets, 24, 387–398, 2004.

Table 1. Summary statistics from the fitting of the four distribution tails into the 15 029 daily rainfall records.

	Pareto			Lognormal		
	MSE	β	γ	MSE	β	γ
Min	0.002	0.42	0.001	0.002	0.20	0.376
Mode*	0.011	7.54	0.134	0.012	2.31	0.750
Median	0.017	8.80	0.140	0.018	2.25	0.768
Mean	0.021	9.51	0.145	0.022	2.18	0.783
Max	0.336	54.79	0.797	0.322	4.34	1.615
SD	0.015	4.92	0.076	0.015	0.63	0.151
Skewness	2.910	1.23	0.495	2.755	-0.43	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
Median	0.019	5.91	0.678	0.023	23.15	0.219
Mean	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
Skewness	2.151	1.82	0.668	4.377	1.65	2.567

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

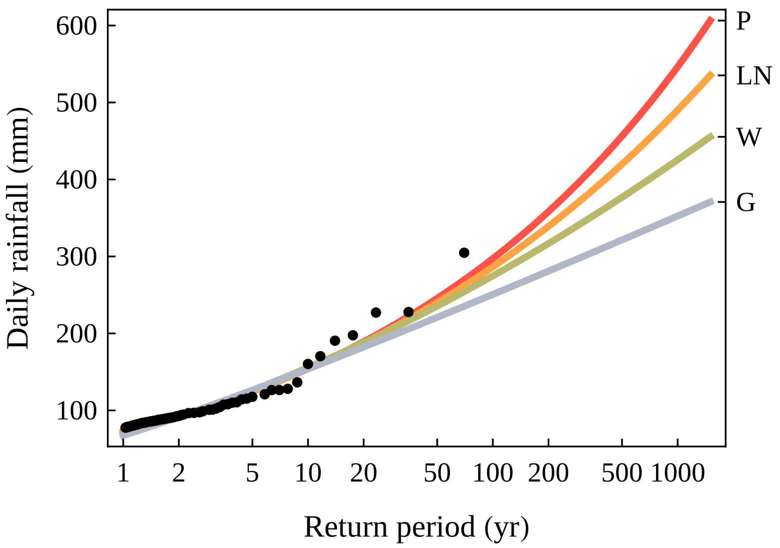

Fig. 1. Four different distribution tails fitted to an empirical tail (P, LN, W and G stands for the Pareto, the Lognormal, the Weibull and the Gamma distribution). A wrong choice may lead to severely underestimated or overestimated rainfall for large return periods.

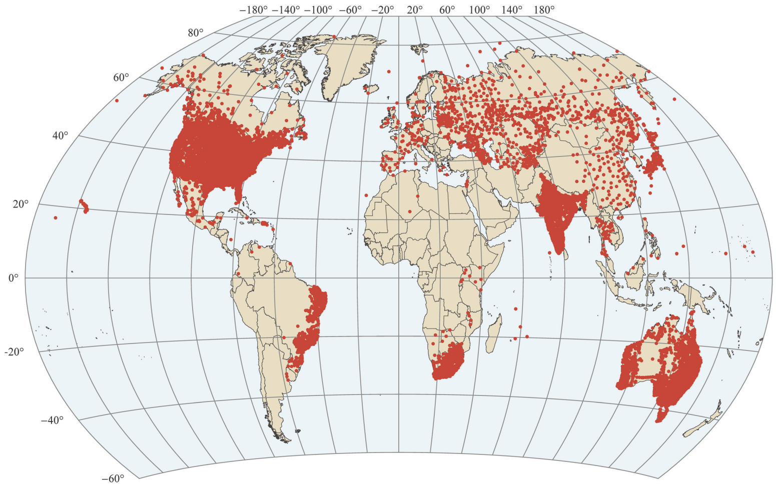

Fig. 2. Locations of the stations studied (a total of 15 029 daily rainfall records with time series length greater than 50 yr).

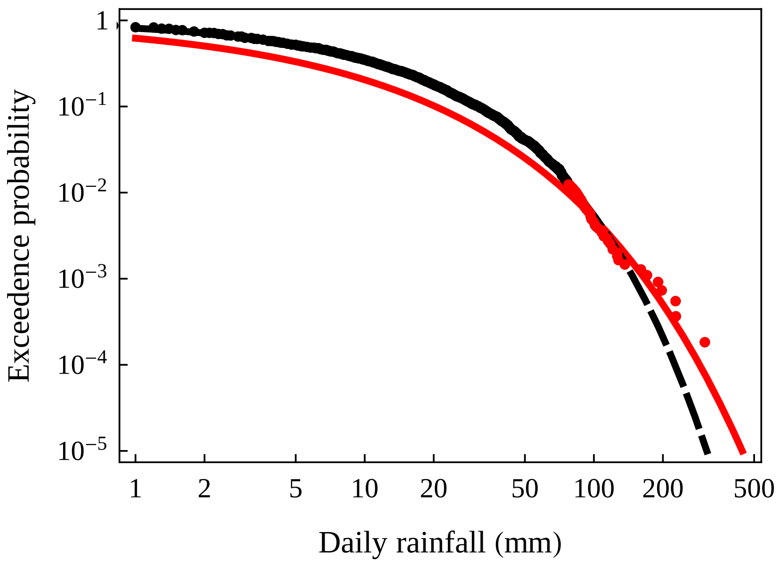

Fig. 3. Explanatory diagram of the fitting approach followed. The dashed line depicts a Weibull distribution fitted to the whole empirical distribution points while the solid line depicts the distribution fitted only to the tail points.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

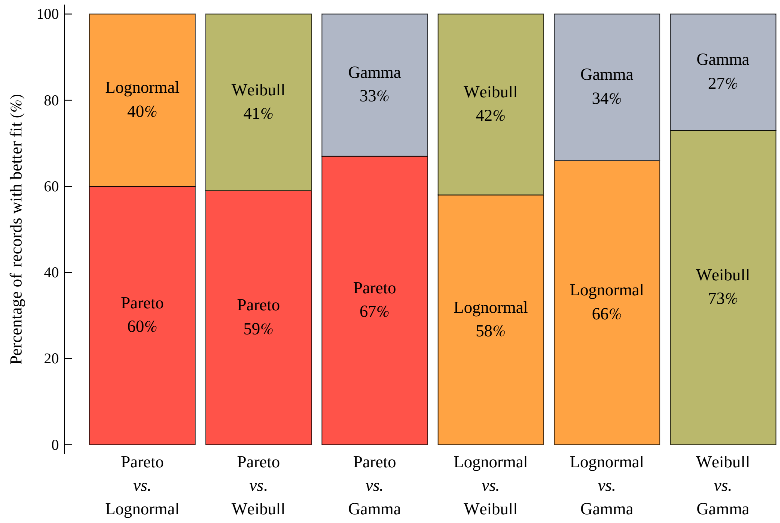

Fig. 4. Comparison of the fitted tails in couples in terms of the resulting MSE. The heavier tail of each couple is better fitted to the empirical points in a higher percentage of the records.

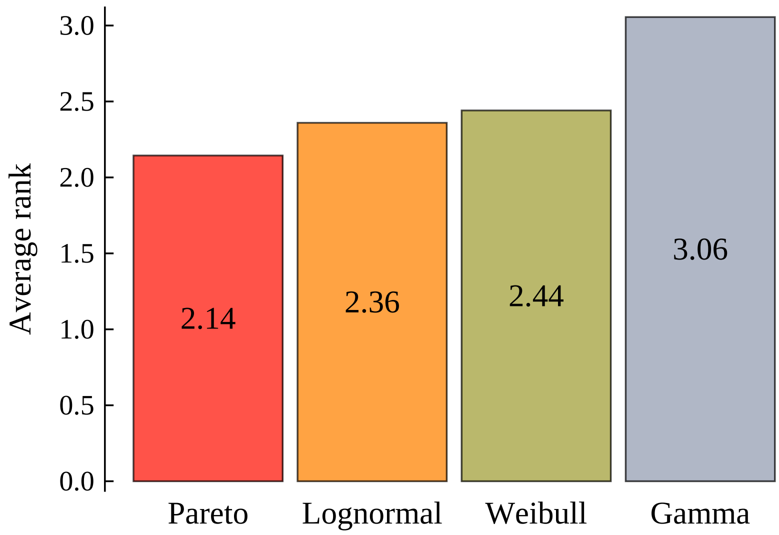

Fig. 5. Mean ranks of the tails for all records. The best-fitted tail was ranked as 1 while the worst-fitted as 4. A lower average rank indicates a better performance.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

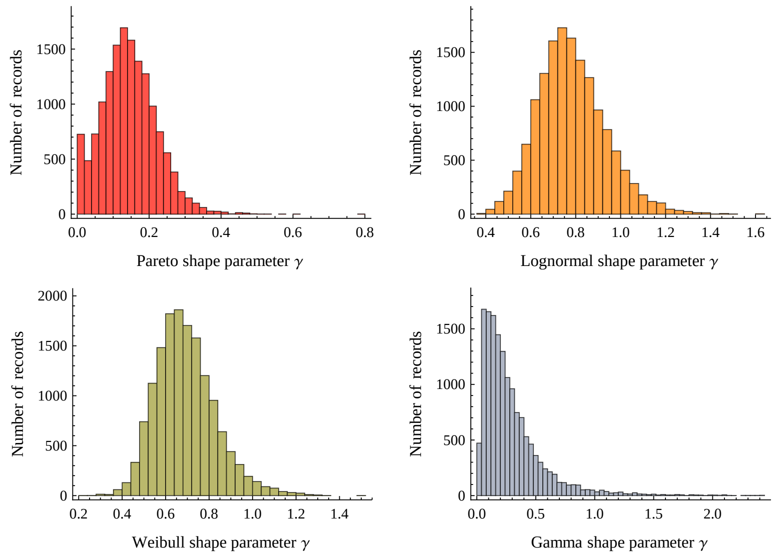

Fig. 6. Histograms of the shape parameters of the fitted tails.

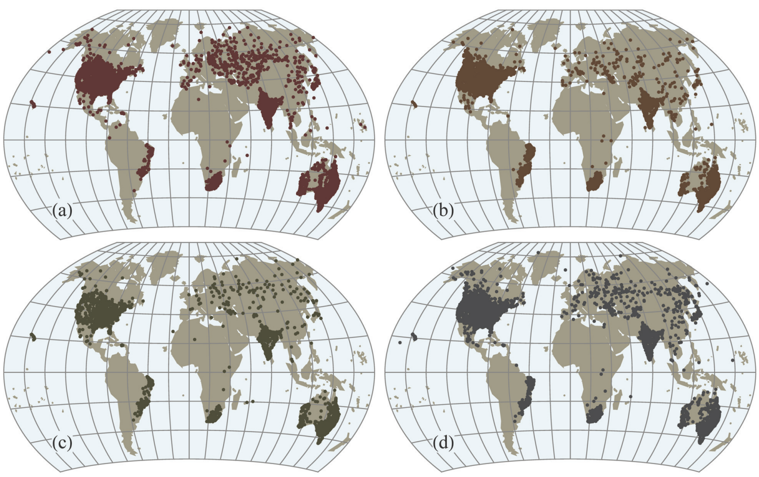

Fig. 7. Geographical depiction of stations where the best fitted tail is **(a)** Pareto, in 4621, **(b)** Lognormal, in 4486, **(c)** Weibull, in 2051 and, **(d)** Gamma, in 3871.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |