European Geosciences Union
General Assembly 2012

Vienna, Austria,
22-27 April 2012

Session: HS7.5/NP8.3
Hydroclimatic stochastics

A global survey on the distribution of annual maxima of daily rainfall: Gumbel or Fréchet?

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/1207

The limiting laws

If a random variable (RV) X follows the distribution $F_X(x)$ then the distribution function of the maximum of n independent and identically distributed RV's, i.e., $Y_n = \max(X_1,\ldots,X_n)$ will be,

$$G_{Y_n}(x) = (F_X(x))^n \quad (1)$$

Now, if $n \to \infty$ there are three limiting laws, the type I or Gumbel (G), the type II or Fréchet (F) and the type III or reversed Weibull (RW) with distribution functions respectively given by

$$G_{\mathrm{G}}(x) = \exp\left(-\exp\left(-\frac{x-\alpha}{\beta}\right)\right) \qquad x \in \mathbb{R} \quad (2)$$

$$G_{\mathrm{F}}(x) = \exp\left(-\left(\frac{x-\alpha}{\beta}\right)^{-\gamma}\right) \qquad x \geq \alpha \quad (3)$$

$$G_{\mathrm{RW}}(x) = \exp\left(-\left(-\frac{x-\alpha}{\beta}\right)^{\gamma}\right) \qquad x \leq \alpha \quad (4)$$

These distributions comprise a location parameter $\alpha \in \mathbb{R}$ and a scale parameter $\beta > 0$, with the Fréchet and the reversed Weibull distributions having the additional shape parameter $\gamma > 0$.

These three distributions can be unified into a single expression known as the Generalized Extreme Value (GEV) distribution (also known as the Fisher-Tippet) with distribution function given by

$$G_{\mathrm{GEV}}(x) = \exp\left(-\left(1+\gamma\frac{x-\alpha}{\beta}\right)^{-1/\gamma}\right) \qquad 1+\gamma\frac{x-\alpha}{\beta} \geq 0 \quad (5)$$

This simple reparameterization exploits the limiting definition of the exponential function so that the Gumbel distribution can also emerges for $\gamma \to 0$.

Convergence to the limiting laws

The distribution of the maximum value, given in Eq. (1), converges to one of the three liming laws given that the maximum value is selected from a sample with size tending to infinity.

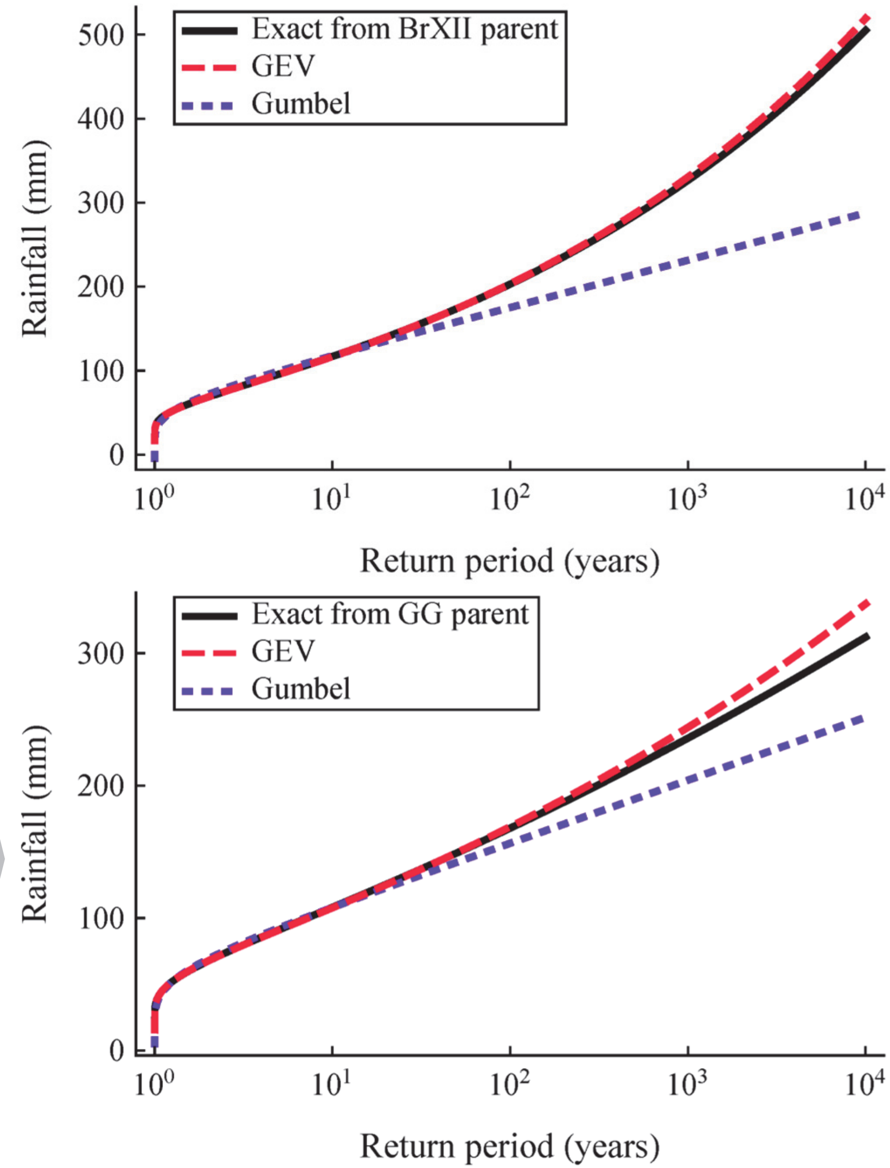
Infinity in real world does not exist, and particularly for daily rainfall, in the best case the sample size would be equal with the number of the year's days, i.e., 366 values. Obviously the actual number of rainy days is much lower and also varies from year to year.

So, the convergence to the limiting laws should not be taken for granted.

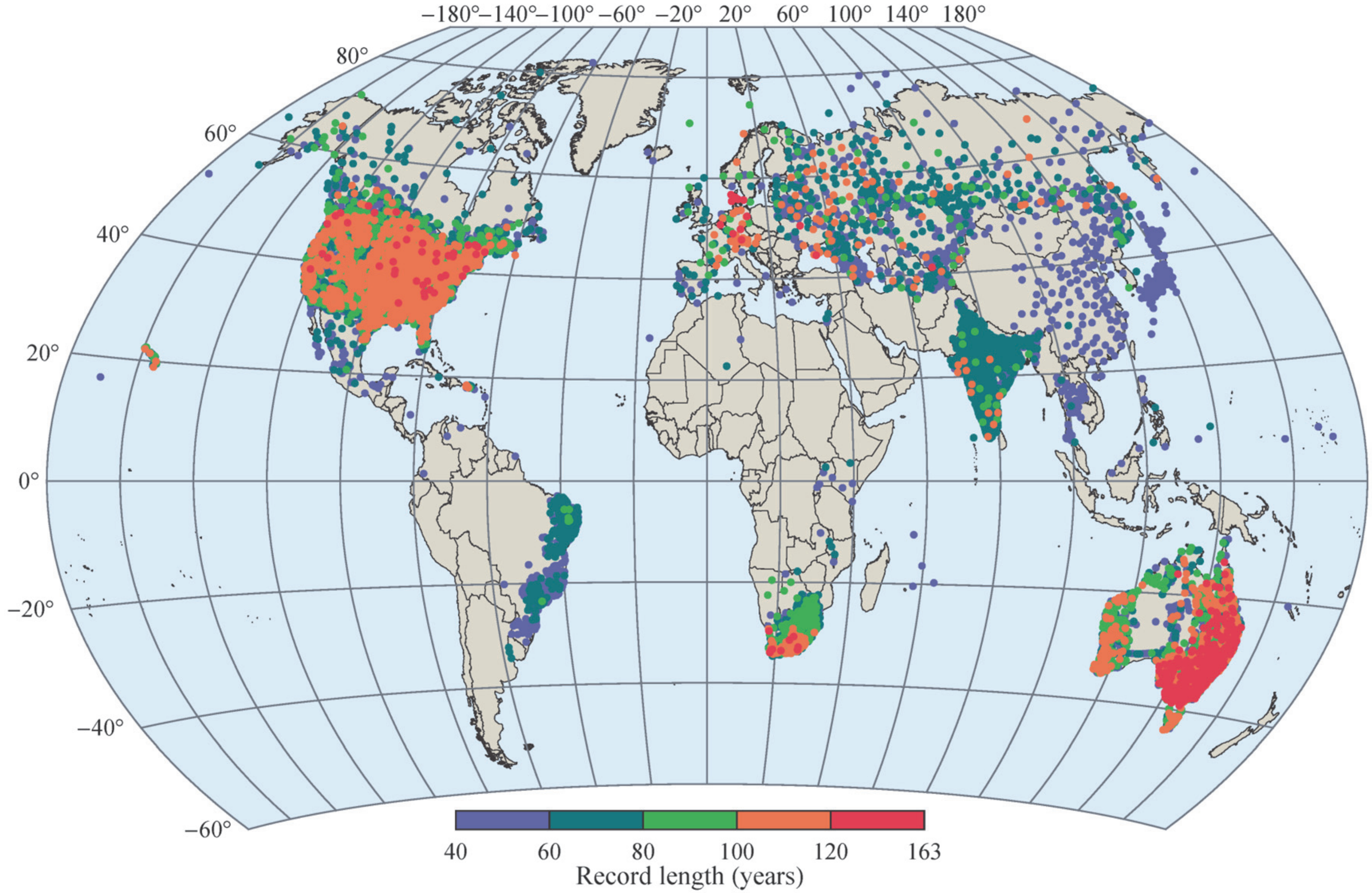
The figure on the right demonstrates the issue on convergence. In a previous study* we found that the Burr XII and the Generalized Gamma distribution are both good models for describing daily rainfall.

So, using representative Burr XII and GG distributions as parent we estimated (a) the exact distribution of the annual maximum for a representative number of rainy days, i.e., $n = 90$, (b) calculated their L-moments, and (c) fitted a GEV and a Gumbel distribution using the L-moments estimated in step (b).

*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



The original dataset



The data used here, are daily rainfall records from the Global Historical Climatology Network-Daily database. Among the records comprising the database, only those fulfilling the following criteria were selected: (a) record length greater or equal than 50 years, (b) missing data less than 20% and, (c) data assigned with "quality flags" less than 0.1%. The map depicts the locations of the stations, a total of 15 137 records.

Extraction of the annual maxima

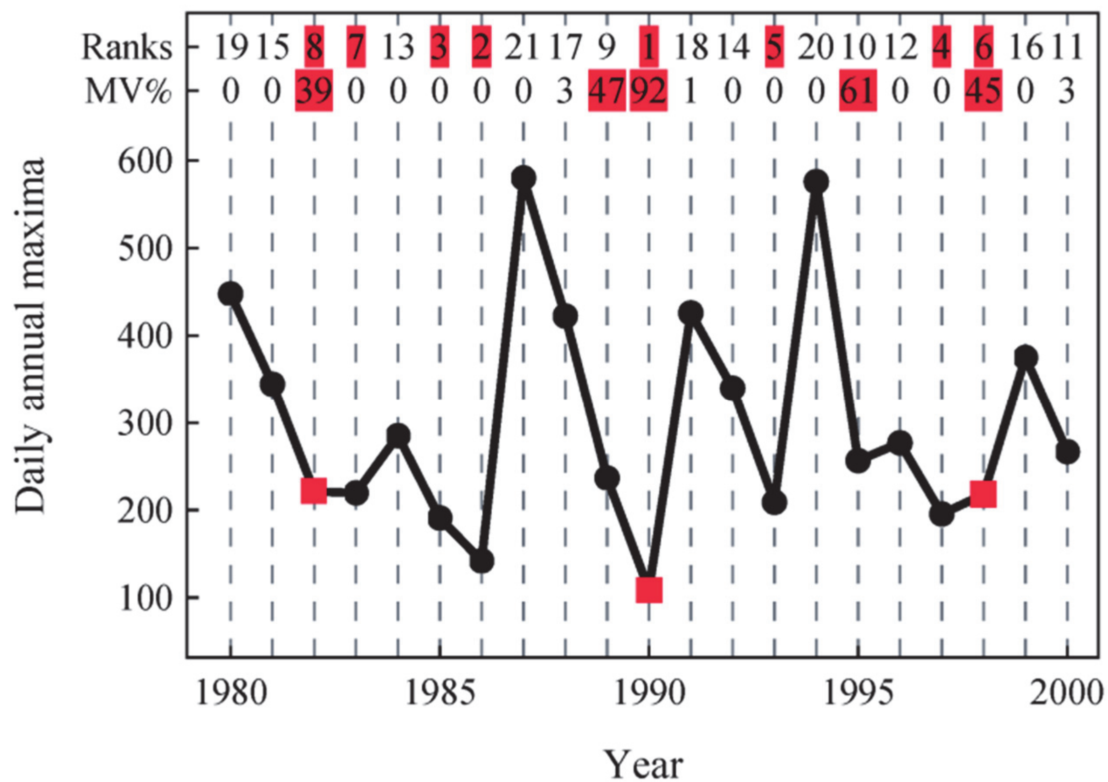
If daily records contain missing values forming the annual maxima series is not trivial. For example the maximum value of an incomplete year may not be the real one, as it is likely for a larger value to have occurred in those days that rainfall was not recorded.

Basically, there are three different methods for extracting the annual maxima series:

(a) in first method (M1), specific criteria are used to assess the validity of the annual maxima, (e.g., the annual maximum value is considered valid only if the missing-values percentage is small),

(b) in the second method (M2), only the maxima of complete years are accepted as valid while the those of incomplete years are assumed unknown,

(c) in the third method (M3), the annual maxima of every year are extracted irrespective of the years' missing-values percentage.



Explanatory plot of the maxima extraction method. The annual maximum daily rainfall is considered unknown (red rectancles) if its rank in the sample is in the smaller 40% of ranks (red shaded ranks) and the missing-value percentage (MV%) of the year it belongs is larger than 1/3 (red shaded percentages).

Validation of the maxima extraction method

The criteria defined previously where not selected unjustifiably, but rather emerged after intensive Monte Carlo simulations

The Monte Carlo scheme could be summarized in four basic steps:

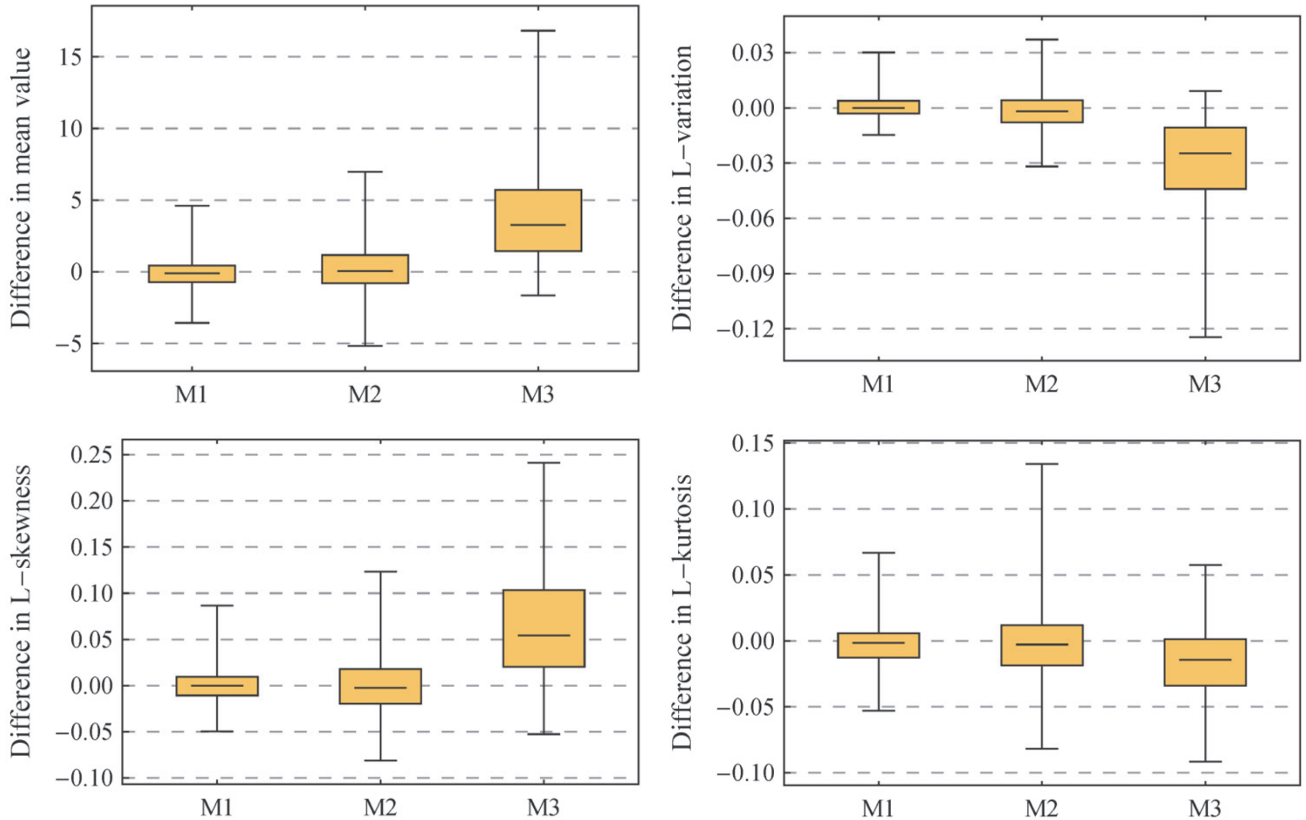
(a) a subset of complete daily records is selected and the annual maxima series are created,

(b) this daily-records subset is modified to contain missing values,

(c) annual maxima series are extracted from the modified daily-records subset by utilizing the maxima extraction method for various criteria values, and

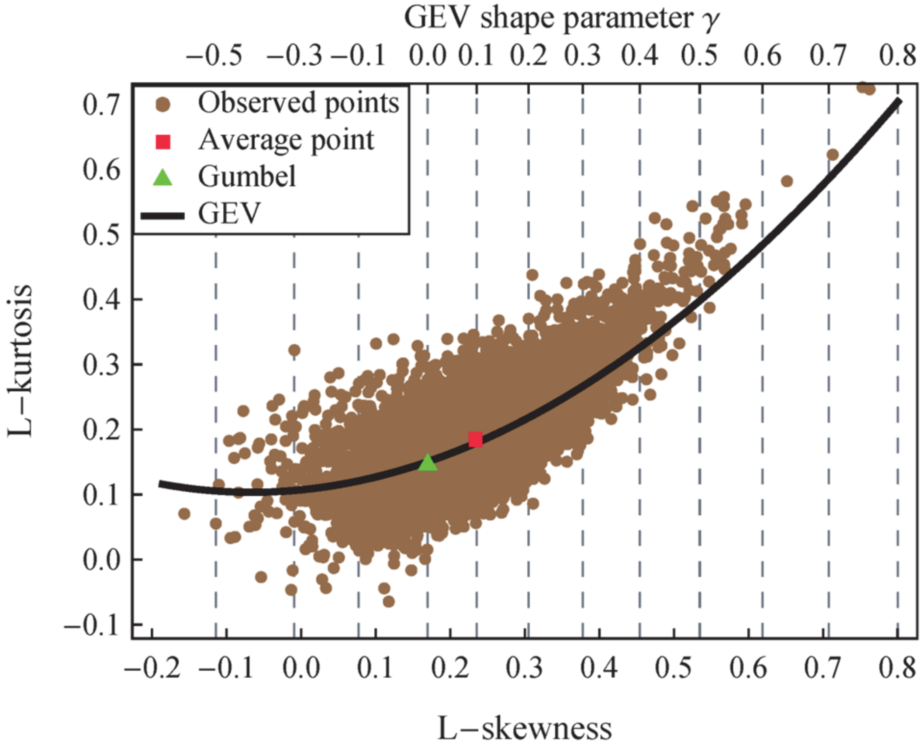
(d) the real maxima series created in step (a) are compared with those created in step (c).

The idea is finding, those criteria resulting in maxima series with statistical characteristics similar to the real ones.



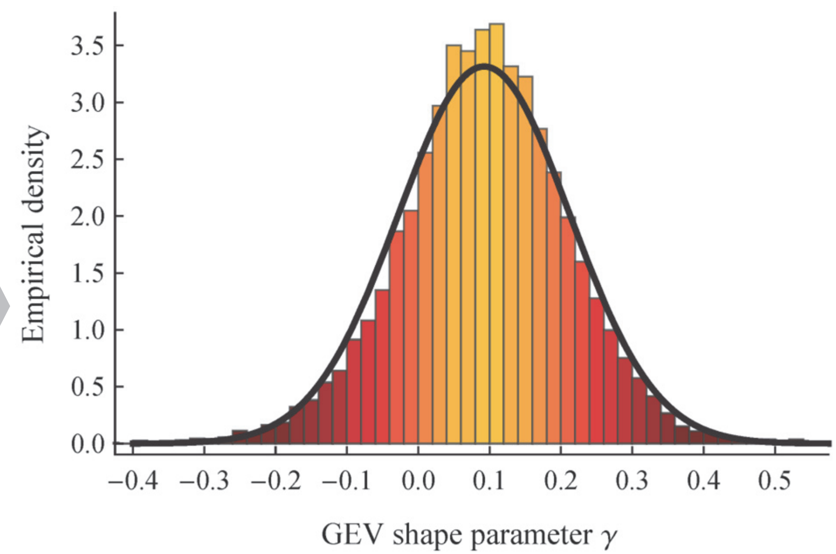
The box plots depict the resulting differences of various sample statistics between the real annual maxima series and the ones created from the incomplete daily series. The advantage of the first method compared to the others is clearly seen by the smaller range of the box plots. The box plots lower and upper fences represent, respectively, the sample quantiles Q_1 and Q_{99}.

Fitting the GEV distribution



L-kurtosis *vs.* L-skewness plot the 15 137 observed points. Interestedly, only 20% of points lies on the left of the Gumbel distribution, corresponding thus to a GEV distribution with $\gamma < 0$ (reversed Weibull law), while 80% of points lies on the right corresponding to a GEV distribution with $\gamma > 0$ (Fréchet law). The average point lies almost exactly on the GEV line and corresponds to $\gamma \approx 0.09$. The figure may not reveal the percentage of points that could be described by a Gumbel distribution, yet, it offers a clear indication that the Fréchet law prevails.

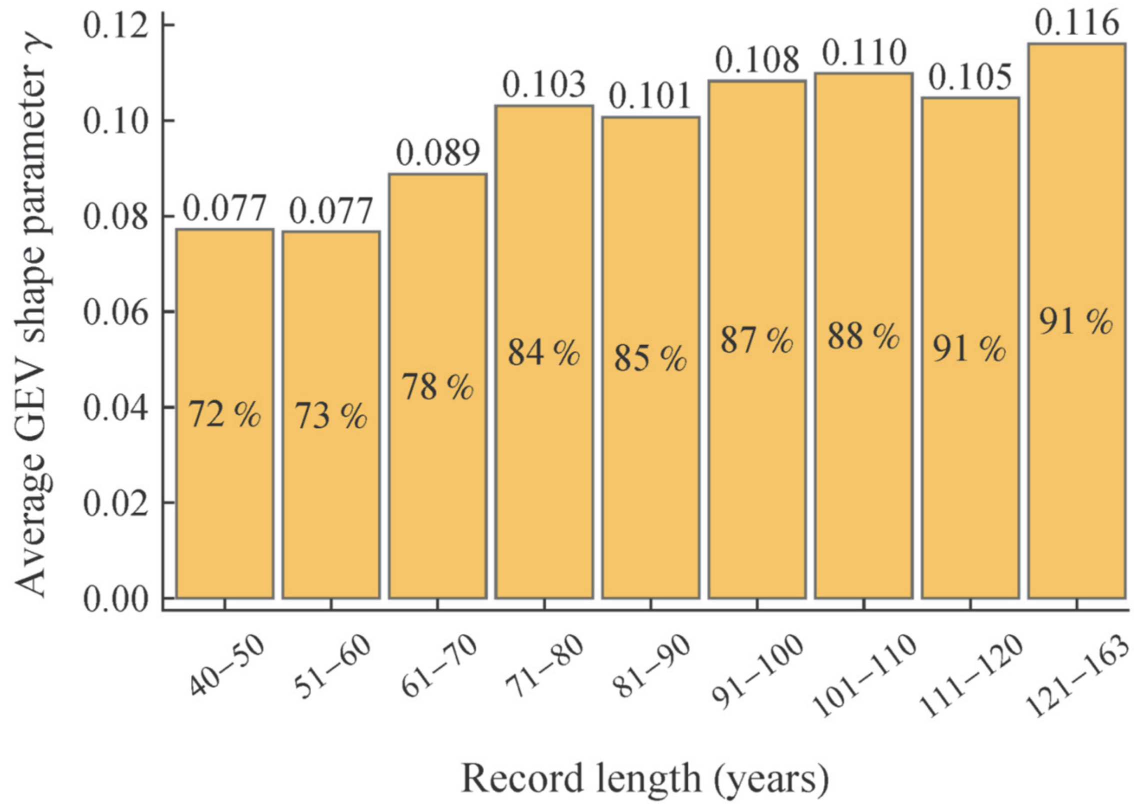
The figure depicts the empirical distribution of the GEV shape parameter as well as a fitted normal distribution with mean 0.093 and standard deviation 0.12.
It is worth noting the large variation of the estimated GEV shape parameter as it ranges from −0.59 to 0.76 with mean value 0.093; the 90% empirical confidence interval is much smaller, i.e., from −0.11 to 0.28 .



GEV shape parameter *vs*. record length I

Larger samples offer more accurate estimates; in this direction we study the estimated GEV shape parameter in relationship with the record length, as the records studied here vary in length from 40 to 163 years.

We grouped the 15 137 estimated shape parameter values into nine groups based on the length of the record that were estimated; and second, we estimated various statistics for each group.



Clearly, the figure indicates an upward "trend" over record length both in the mean shape parameter value and in the percentage of records having positive shape parameter.

The figure depicts the mean value of the GEV shape parameter for various ranges of record length. While the number in the boxes indicates the percentage of records with positive shape parameter value.

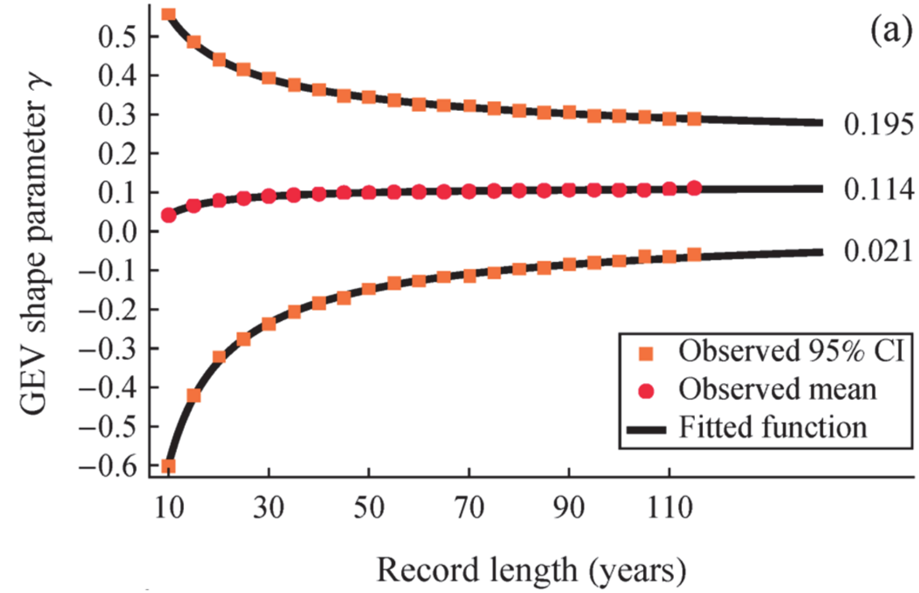
GEV shape parameter *vs.* record length II

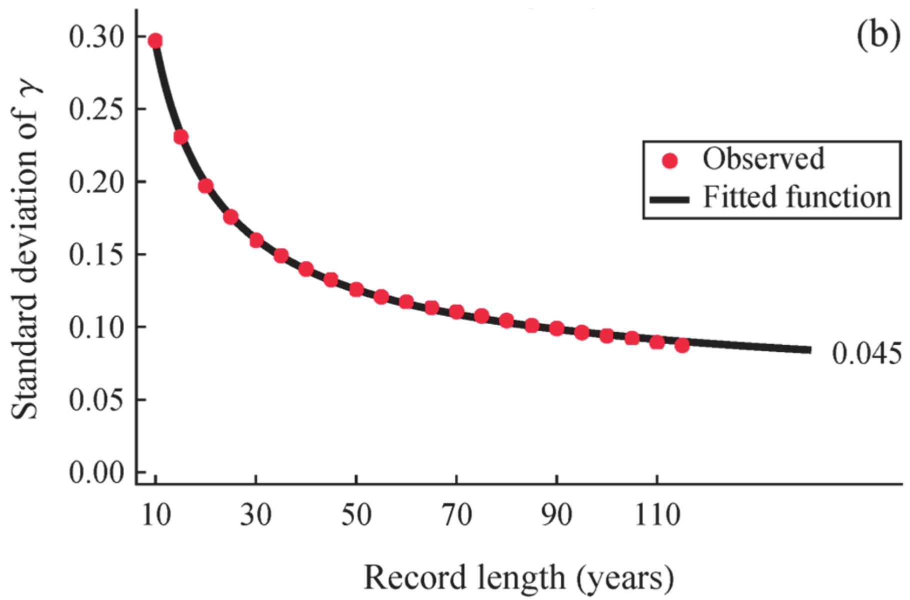
The previous analysis indicated a relationship between the GEV shape parameter and the record length but it did not reveal a precise law. So, we partitioned each record into many sub-records of smaller length and we analyzed them. The figures present these results.

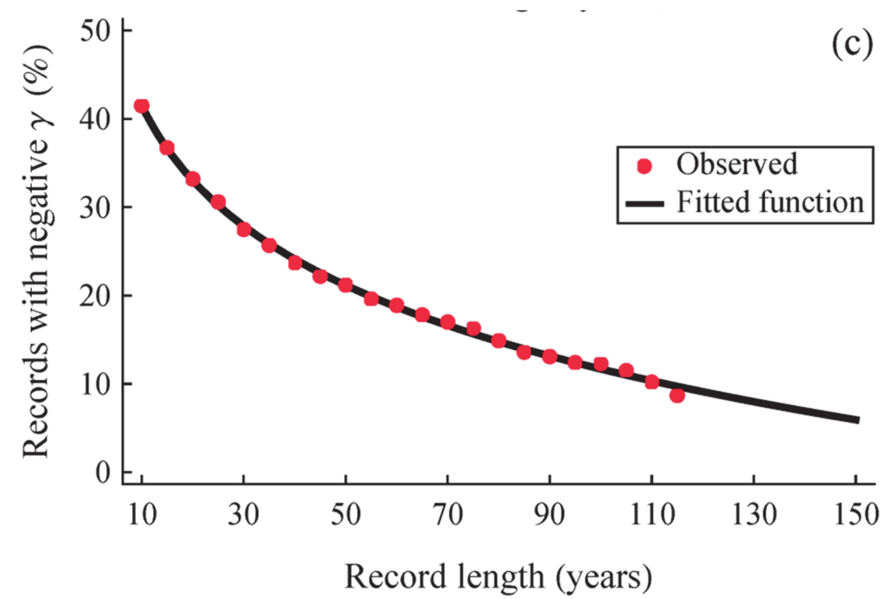
(a) Mean, Q_5 and Q_{95} observed points *vs.* record length, and the estimated asymptotic values of the fitted curves.

(b) Standard deviation vs. record length,

(c) Percentage of record with negative shape parameter vs. record length.





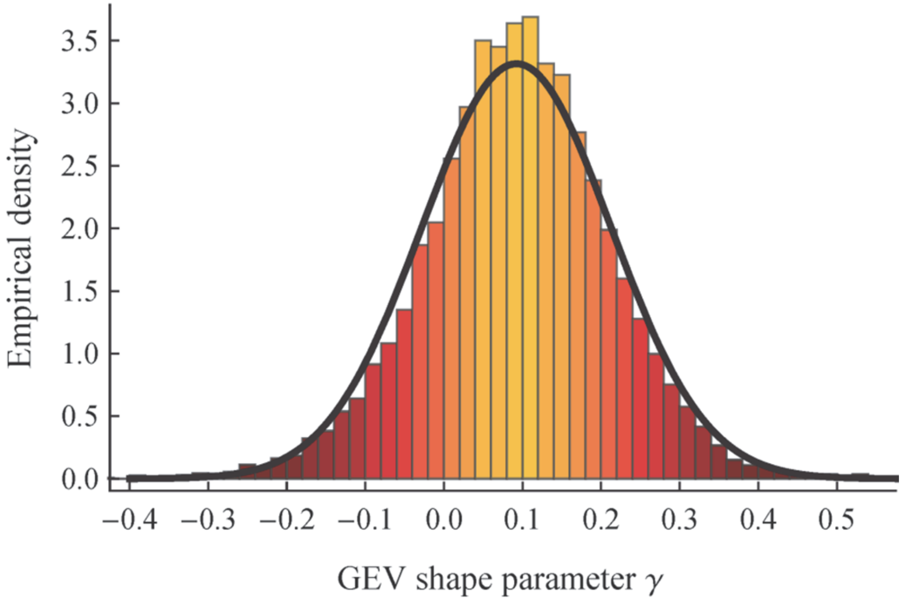


A Monte Carlo verification

In order to verify that the true distribution of the GEV shape parameter is a Normal with mean value 0.114 and standard deviation 0.045 we performed a Monte Carlo simulation summarized in the following steps:

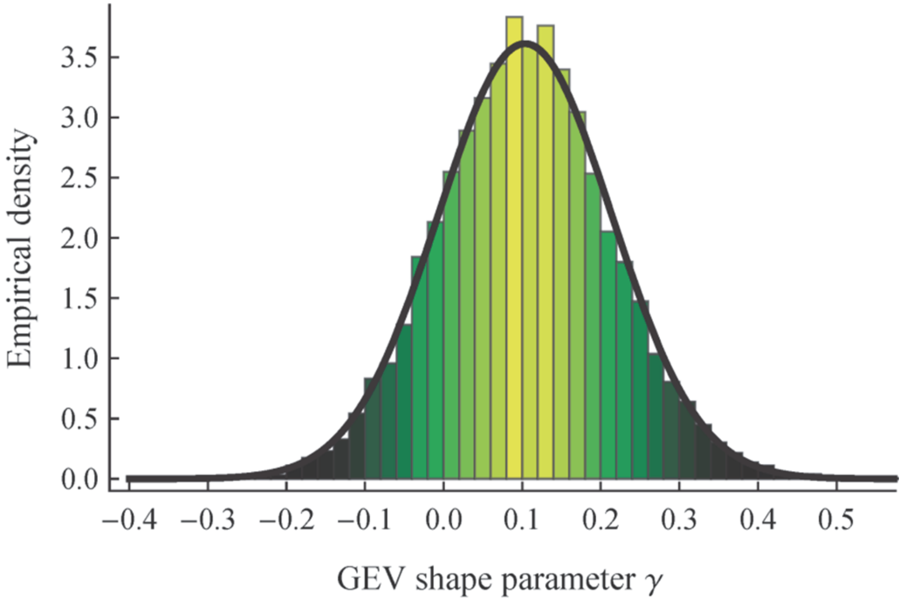
(a) We generated 15 137 random samples, with sizes exactly equal with the original records lengths, from a GEV distribution with the shape parameter being randomly generated from the anticipated normal distribution and with the location and scale parameter fixed to their mean values.

(b) We estimated the shape parameter values of those samples and we formed the empirical distribution and compared it with the original one.

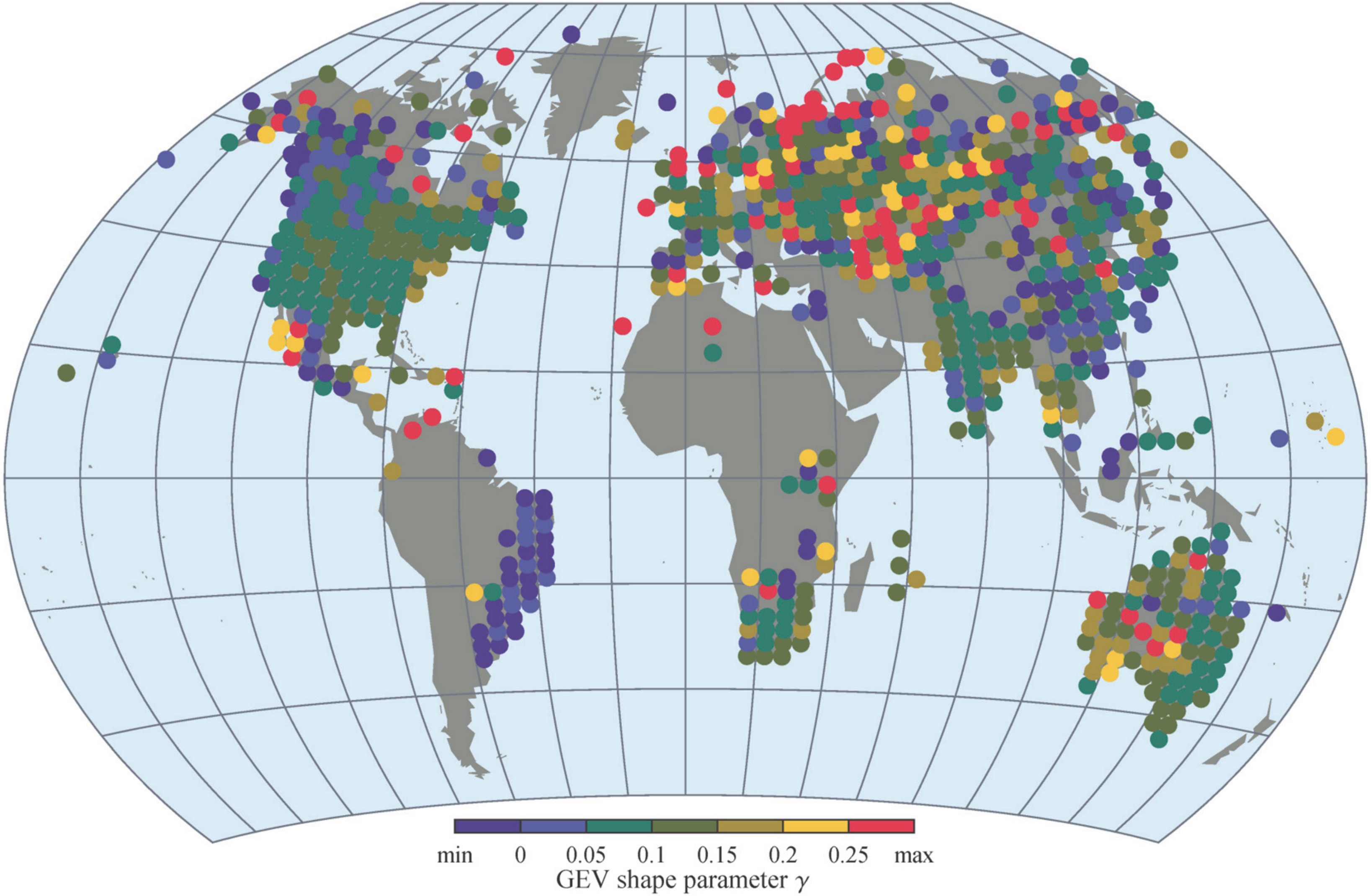


Empirical distribution of the GEV shape parameter resulted from the Monte Carlo simulation. The solid line depicts the fitted normal.

While the prior distribution of γ was a normal with mean 0.114 and SD 0.045, the estimated posterior has mean 0.104 and SD 0.11, values very close with those of the empirical distribution of γ emerged from the real records (upper figure on the right) which respectively are 0.092 and 0.12.

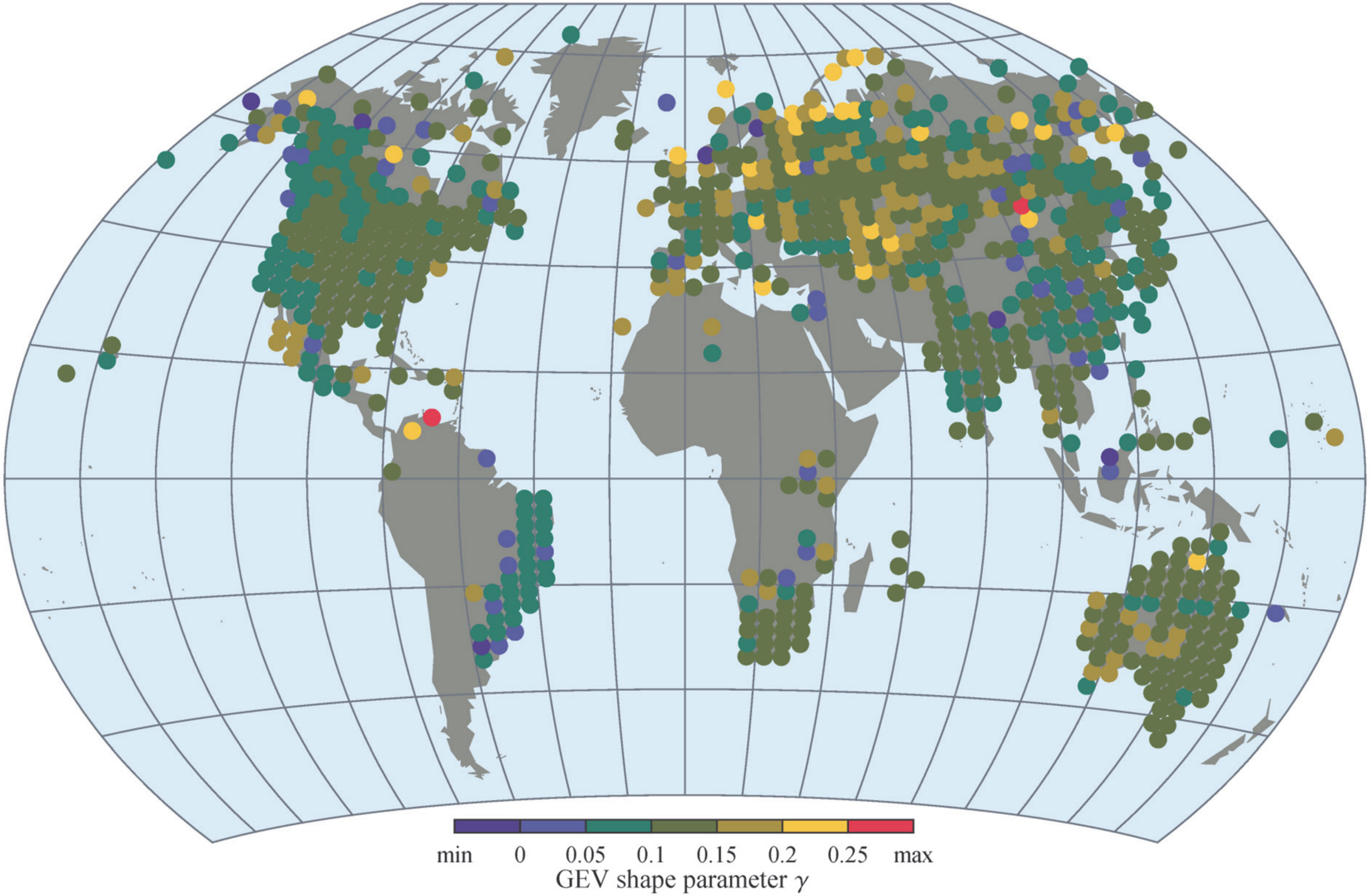


Geographical variation of the GEV shape parameter I



Geographical distribution of the mean value of the GEV shape parameter in regions of latitude difference $\Delta\varphi = 2.5°$ and longitude difference $\Delta\lambda = 5°$.

Geographical variation of the GEV shape parameter II



Geographical distribution of the corrected mean value of the GEV shape parameter in regions of latitude difference $\Delta\varphi = 2.5°$ and longitude difference $\Delta\lambda = 5°$.

Conclusions

We analyzed 15137 records of daily rainfall annual maxima from all over the world by fitting the GEV distribution and focusing on its shape parameter values. The analysis revealed that

- *"Φύσις κρύπτεσθαι φιλεῖ…"* There is a clear relationship of the GEV shape parameter value with the record length implying that only very large samples can reveal its true distribution or the true behaviour of the extreme rainfall.
- The fitted functions suggest that the true distribution of the GEV shape parameter is approximately normal with mean value 0.114 and standard deviation 0.045.
- The percentage of records with negative shape parameter rapidly decreases over sample size while the fitted function indicates that for record length greater than 226 years this percentage would be zero. Interestedly, none of the 16 records available with length greater than 140 years resulted in negative γ.
- The probability for a negative shape parameter to occur, according to the distribution fitted, is only 0.005 and combined with the previous finding suggests that a GEV with negative shape parameter is inappropriate for rainfall.
- We should expect the GEV shape parameter to belong in a narrow range as the 99% CI ranges approximately from 0 to 0.23.
- There is a clear geographical variation of the GEV shape parameter with different areas of the world exhibiting different behaviour in extremes.
- A final proposal: in the case where data suggest a GEV distribution with negative shape parameter it should not be used. Instead it is more reasonable to use a Gumbel or for more safety a GEV distribution with a shape parameter value equal to 0.114.

References

Balkema, A. A., and L. D. Haan (1972), On R. Von Mises' Condition for the Domain of Attraction of exp(-e-x)1, *The Annals of Mathematical Statistics*, *43*(4), 1352–1354.

Barndorff-Nielsen, O. (1963), On the Limit Behaviour of Extreme Order Statistics, *The Annals of Mathematical Statistics*, *34*(3), 992–1002.

Berman, S. M. (1964), Limit Theorems for the Maximum Term in Stationary Sequences, *The Annals of Mathematical Statistics*, *35*(2), 502–516.

von Bortkiewicz, L. (1922), Variationsbreite und mittlerer Fehler, *Sitzungsber. Berli. Math. Ges.*, (21), 3–11.

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

Fisher, R. A., and L. H. C. Tippett (1928), Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample, *Mathematical Proceedings of the Cambridge Philosophical Society*, *24*(02), 180–190, doi:10.1017/S0305004100015681.

Fréchet, M. (1927), Sur la loi de probabilité de l'écart maximum, *Ann. soc. polon. math*, *6*, 93.

Galambos, J. (1972), On the Distribution of the Maximum of Random Variables, *The Annals of Mathematical Statistics*, *43*(2), 516–521.

Gnedenko, B. (1943), Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire, *The Annals of Mathematics*, *44*(3), 423–453, doi:10.2307/1968974.

Gumbel, E. J. (1958), *Statistics of Extremes*, Columbia University Press.

De Haan, L. (1971), A form of regular variation and its application to the domain of attraction of the double exponential, *Z. Wahrsch. Geb.*, (17), 241–258.

Hosking, J. R. M. (1984), Testing Whether the Shape Parameter is Zero in the Generalized Extreme- Value Distribution, *Biometrika*, *71*(2), 367–374, doi:10.2307/2336254.

Hosking, J. R. M. (1990), L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics, *Journal of the Royal Statistical Society. Series B (Methodological)*, *52*(1), 105–124.

Hosking, J. R. M. (1992), Moments or L Moments? An Example Comparing Two Measures of Distributional Shape, *The American Statistician*, *46*(3), 186–189, doi:10.2307/2685210.

Hosking, J. R. M., and J. R. Wallis (1993), Some statistics useful in regional frequency analysis, *Water Resour. Res.*, *29*(2), PP. 271–281, doi:199310.1029/92WR01980.

Hosking, J. R. M., J. R. Wallis, and E. F. Wood (1985), Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments, *Technometrics*, *27*(3), 251–261, doi:10.2307/1269706.

References

Jenkinson, A. F. (1955), The frequency distribution of the annual maximum (or minimum) values of meteorological elements, *Quarterly Journal of the Royal Meteorological Society*, *81*(348), 158–171, doi:10.1002/qj.49708134804.

Juncosa, M. L. (1949), The asymptotic behavior of the minimum in a sequence of random variables, *Duke Mathematical Journal*, *16*(4), 609–618, doi:10.1215/S0012-7094-49-01658-0.

Kotz, S., and S. Nadarajah (2000), *Extreme value distributions: theory and applications*, Imperial College Press.

Koutsoyiannis, D. (2004a), Statistics of extremes and estimation of extreme rainfall, 1, Theoretical investigation, Hydrological Sciences Journal, 49(4), 575–590.

Koutsoyiannis, D. (2004b), Statistics of extremes and estimation of extreme rainfall, 2, Empirical investigation of long rainfall records, Hydrological Sciences Journal, 49(4), 591–610.

von Mises, R. (1936), La distribution de la plus grande de n valeurs, *Rev. math. Union interbalcanique*, *1*(1).

Papalexiou, S. M., and D. Koutsoyiannis (2011), Entropy based derivation of probability distributions: A case study to daily rainfall, *Advances in Water Resources*, (0), doi:10.1016/j.advwatres.2011.11.007. [online] Available from: http://www.sciencedirect.com/science/article/pii/S0309170811002193 (Accessed 5 April 2012)

Papalexiou, S. M., D. Koutsoyiannis, and C. Makropoulos (2012), How extreme is extreme? An assessment of daily rainfall distribution tails, *Hydrology and Earth System Sciences Discussions*, *9*(5), 5757–5778, doi:10.5194/hessd-9-5757-2012.

Peel, M. C., Q. Wang, R. M. Vogel, and T. A. McMahon (2001), The utility of L-moment ratio diagrams for selecting a regional probability distribution, *Hydrological sciences journal*, *46*(1), 147–156.

Pickands III, J. (1975), Statistical Inference Using Extreme Order Statistics, *The Annals of Statistics*, *3*(1), 119–131.

Reiss, R. D., and M. Thomas (2007), *Statistical analysis of extreme values: with applications to insurance, finance, hydrology and other fields*, Birkhauser.

Smirnov, N. V. (1949), Limit distributions for the terms of a variational series, *Trudy Matematicheskogo Instituta im. VA Steklova*, *25*, 3–60.

Vogel, R. M., and N. M. Fennessey (1993), L moment diagrams should replace product moment diagrams, *Water Resources Research*, *29*(6), 1745–1752.

Watson, G. S. (1954), Extreme Values in Samples from m-Dependent Stationary Stochastic Processes, *The Annals of Mathematical Statistics*, *25*(4), 798–800.