European Geosciences Union
General Assembly 2013

Vienna, Austria,
07 – 12 April 2013

Session: HS7.5/NP8.4
Hydroclimatic stochastics

EXTREME RAINFALL DISTRIBUTION TAILS: EXPONENTIAL, SUBEXPONENTIAL OR HYPEREXPONENTIAL?

S. Nerantzaki, S.M. Papalexiou and D. Koutsoyiannis

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

Available online: http://itia.ntua.gr/1333/
Contact Info: sofia_ner@hotmail.com

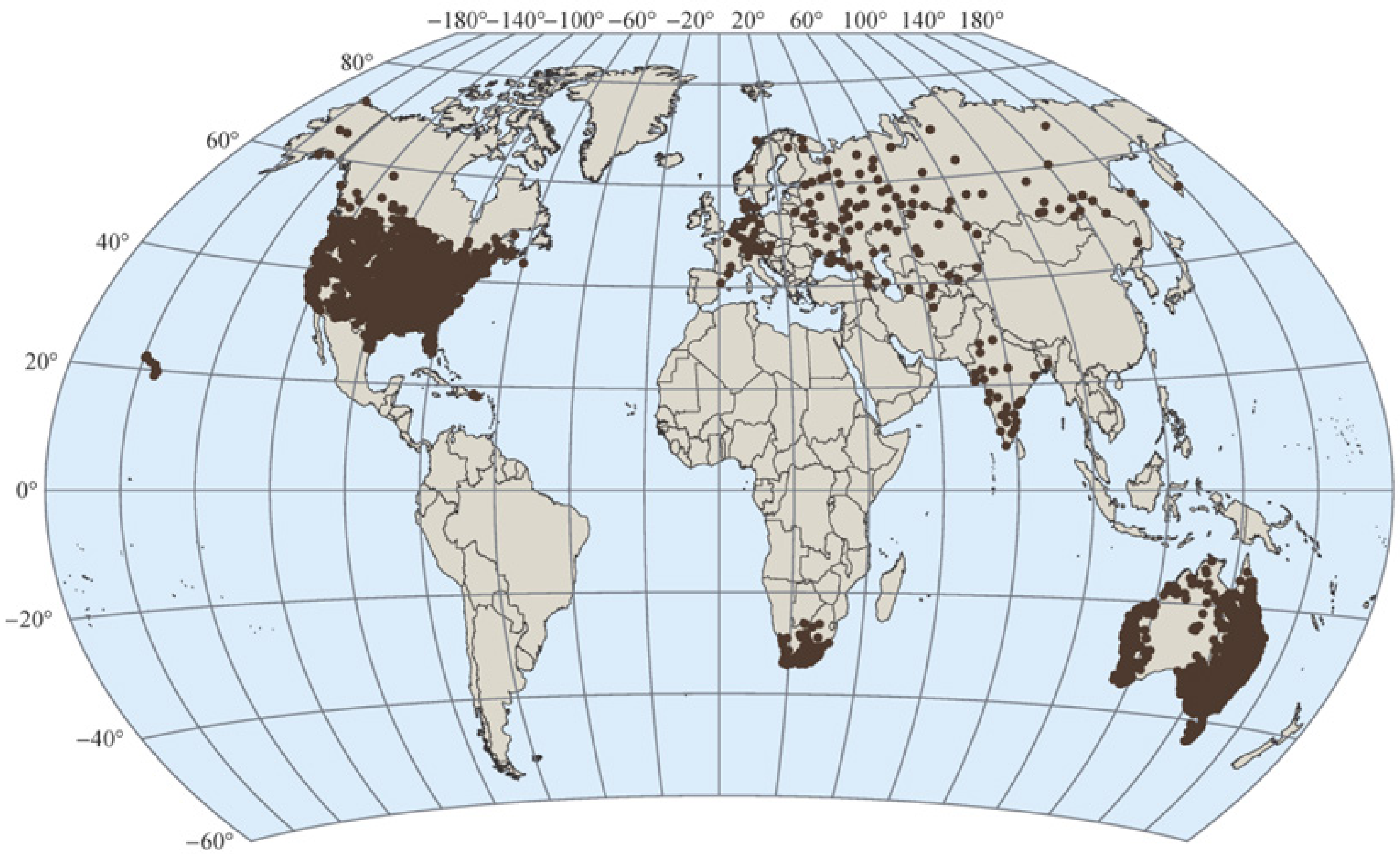
1. Abstract

The upper tail of a probability distribution controls the behavior of both the magnitude and the frequency of extreme events. In general, based on their tail behavior, probability distributions can be categorized into two families (with reference to the exponential distribution): subexponential and hyperexponential. The latter corresponds to milder and less frequent extremes. In order to evaluate the behavior of rainfall extremes, we examine data from 3 477 stations from all over the world with sample size over 100 years. We apply the Mean Excess Function (MEF) which is a common graphical method that results in a zero slope line when applied to exponentially distributed data and in a positive slope in the case of subexponential distributions. To implement the method, we constructed confidence intervals for the slope of the Exponential distribution as functions of the sample size. The validation of the method using Monte Carlo techniques reveals that it performs well especially for large samples. The analysis shows that subexponential distributions are generally in better agreement with rainfall extremes compared to the commonly used exponential ones.

2. Motivation

- The upper tail of a probability distribution controls the behavior of both the magnitude and frequency of extreme events
- Study of the distribution tail of meteorological and hydrological phenomena, and especially rainfall, constitutes important knowledge for the design of hydraulic constructions
- An ill-fitted tail may result in significant errors having major impacts on hydrological design
- Commonly used distribution fitting methods are biased against the distribution tail, as the parameters are estimated for the majority of the data, which by definition do not belong to the tail
- We apply the Mean Excess Function (M.E.F.) to data from 3 477 stations from all over the world aiming to find their tail type: exponential, subexponential or hyperexponential
- The results can be used to investigate the rainfall tail behavior on a global scale

3. Data



- We examined daily rainfall data from the Global Historical Climatology Network which contains data from 40 000 stations http://www.ncdc.noaa.gov/oa/climate/ghcn-daily/

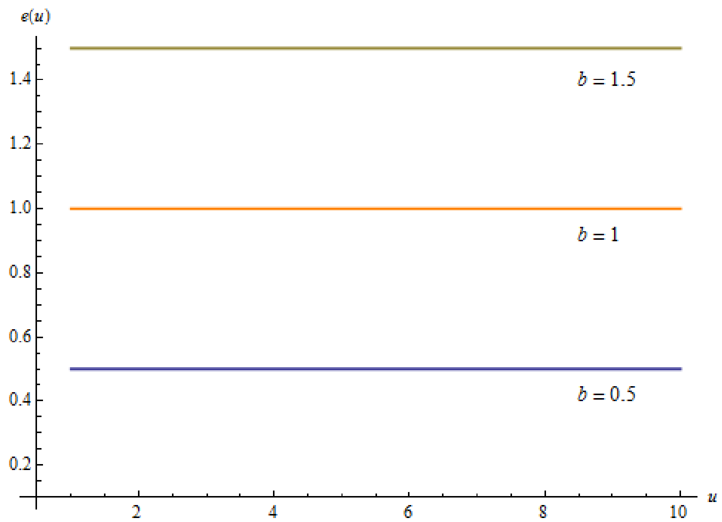
- We only used stations with records of over 100 years and with missing values of less than 20%, for which the “questionable quality” was less than 0.1%

- Taking the above information under consideration, the data used for analysis were finally derived from 3 477 stations
- Their locations are shown on the map

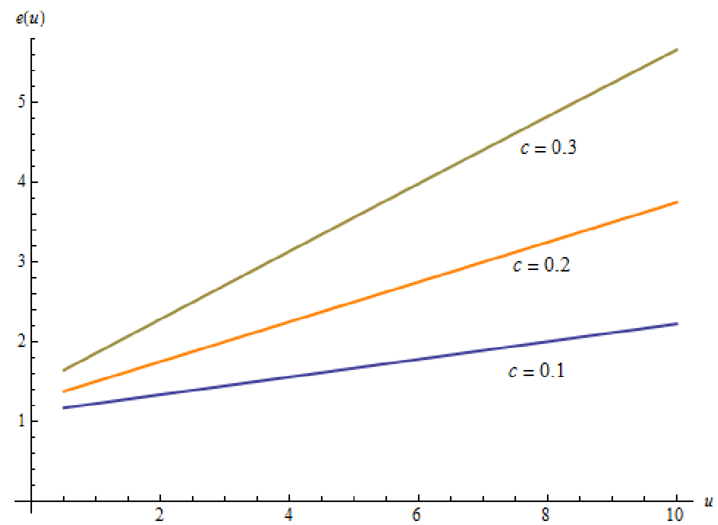
4. Mean Excess Function (MEF)

Mean Excess Function: $e(u) = E(X - u \mid X > u)$ (Benktander & Segerdahl, 1960)

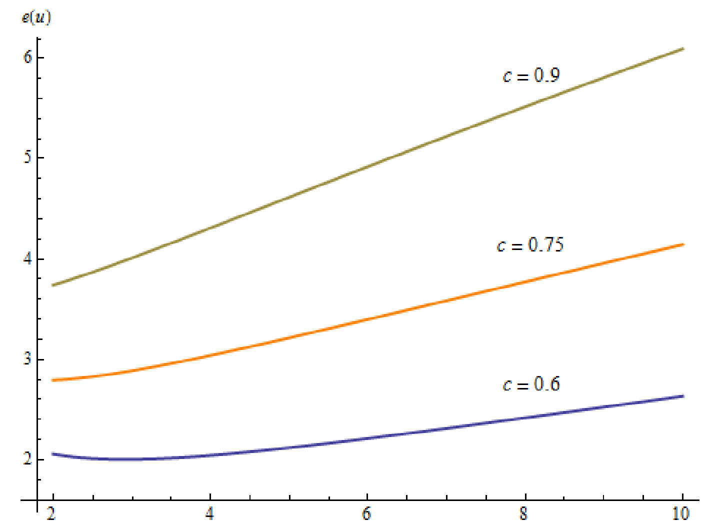
When applied to five known theoretical distributions: $e(u) = \frac{1}{\overline{F}(u)} \int_u^{x_F} \overline{F}(x)\, dx\ ,\quad 0 < u < x_F$ (Embrechts et al., 1997)



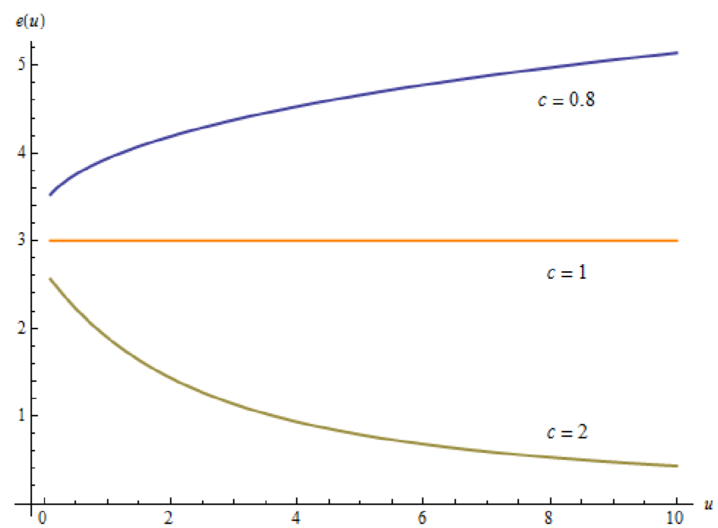
Exponential



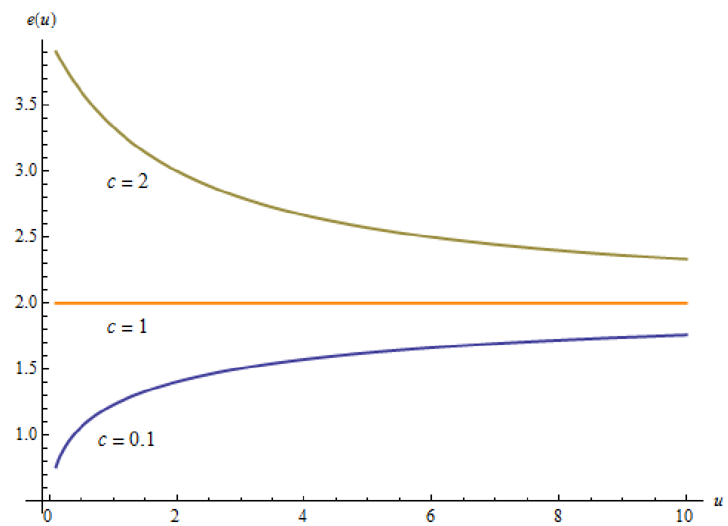
Pareto type II



LogNormal



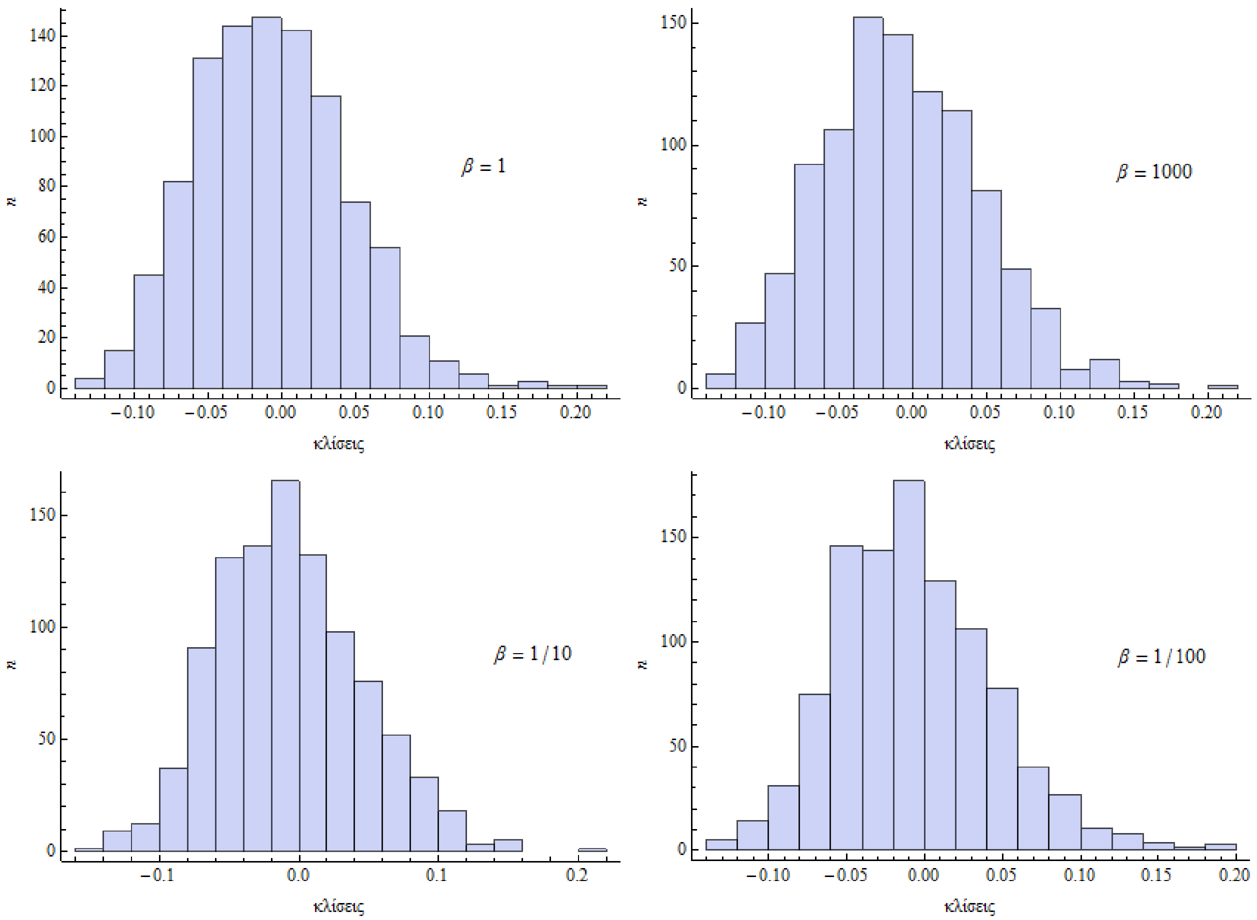
Weibull



Gamma

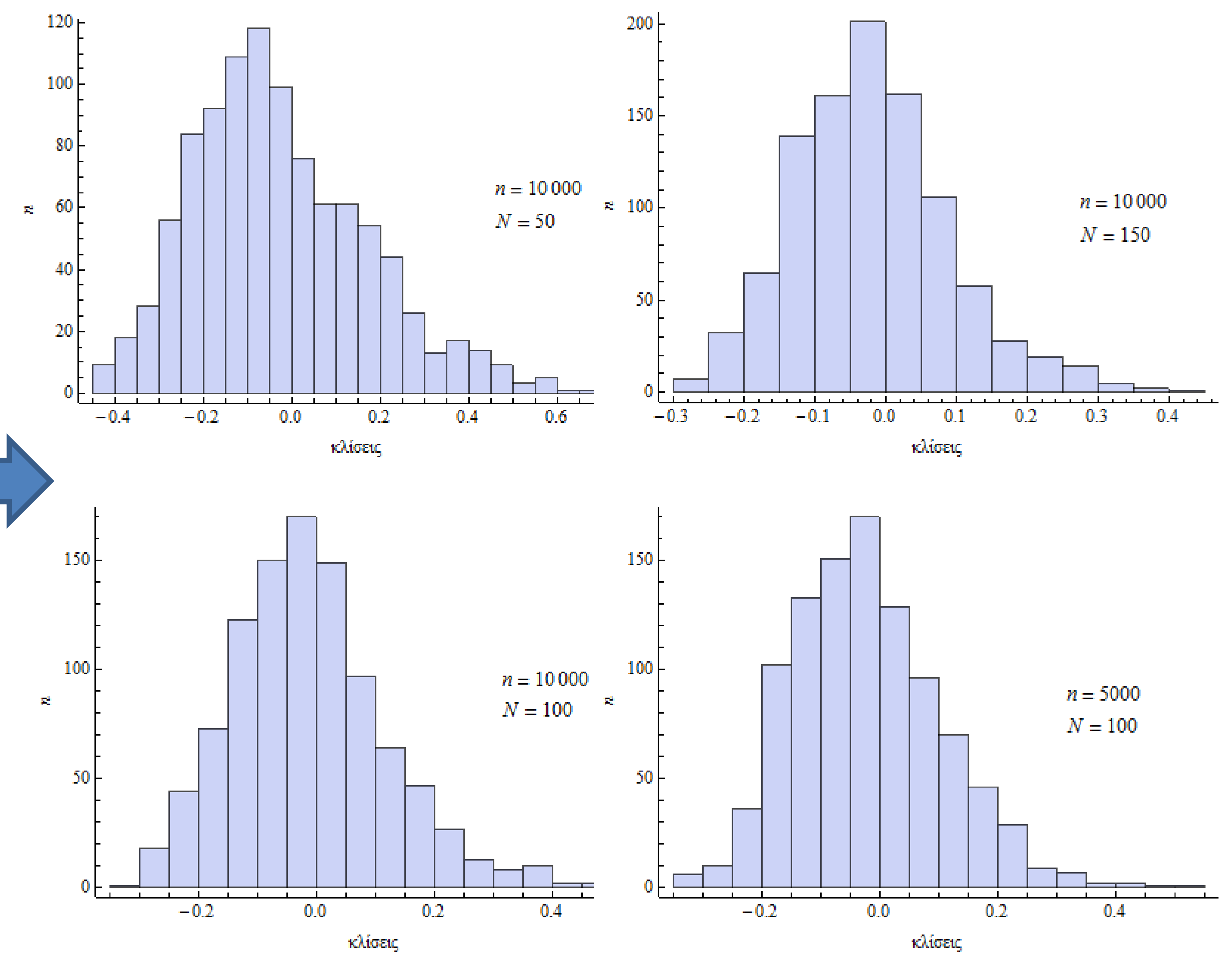
For the exponential distribution (light –tailed), the graphical representation of MEF is constant.
For subexponential distributions (heavy tails), the graph has positive slope (and is linear for the Pareto distribution).
For hyperexponential ones (light tails) the graph has negative slope (El Adlouni et al., 2008)

5. Slope Intervals for Exponential distribution



For each parameter we generated 1000 random samples of Exponential Distribution with sample size 10 000 on which MEF was applied

Dependence of slope intervals on parameter β of the Exponential Distribution:
For all values of the parameter the interval is [-0.2, 0.2]

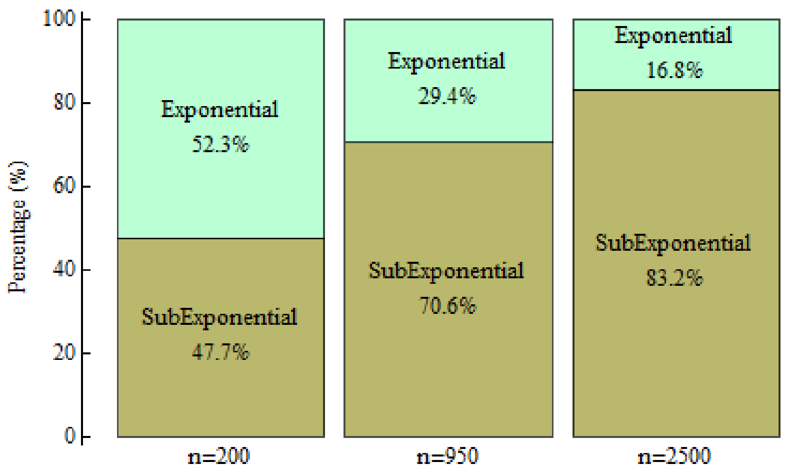


Dependence of slope intervals on sample and tail size of the Exponential Distribution:
The slope interval is independent of the sample size n but varies with tail size N

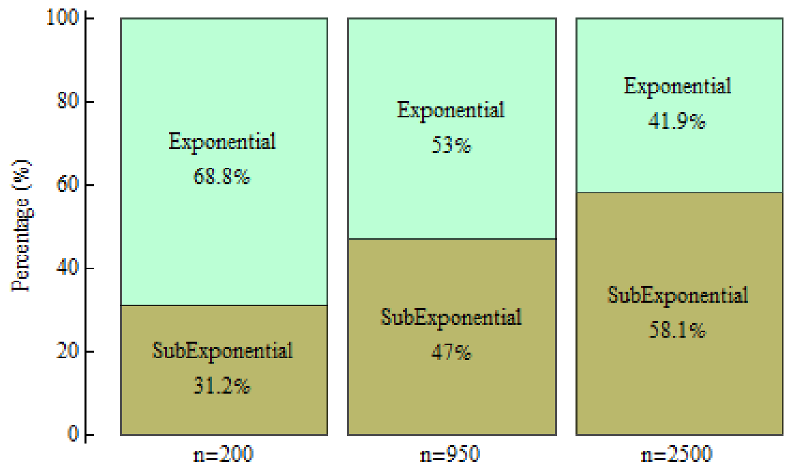
n = 10 000			n = 5 000
N = 50	N = 100	N = 150	N = 100
[−0.35, 0.43]	[−0.24, 0.25]	[−0.20, 0.22]	[−0.24, 0.25]

6. Monte Carlo simulations on theoretical distributions (1)

Weibull

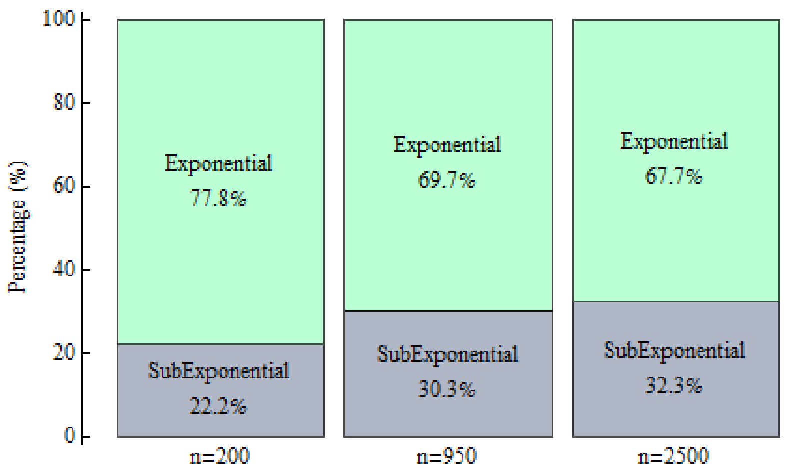


c = 0.6

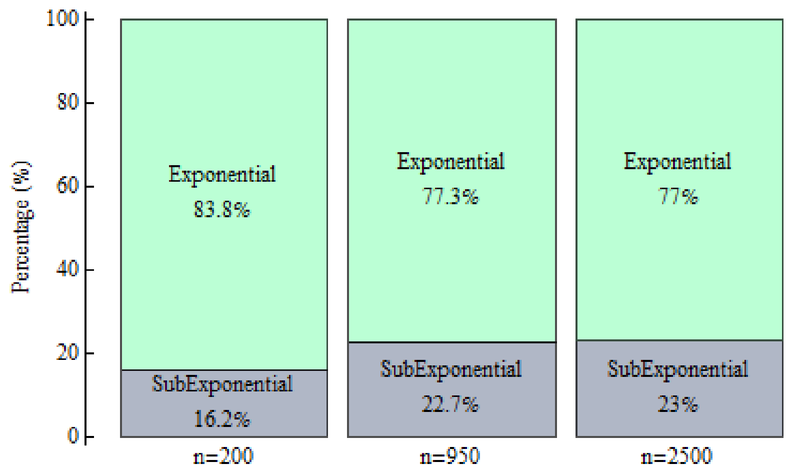


c = 0.8

Gamma



c = 0.2



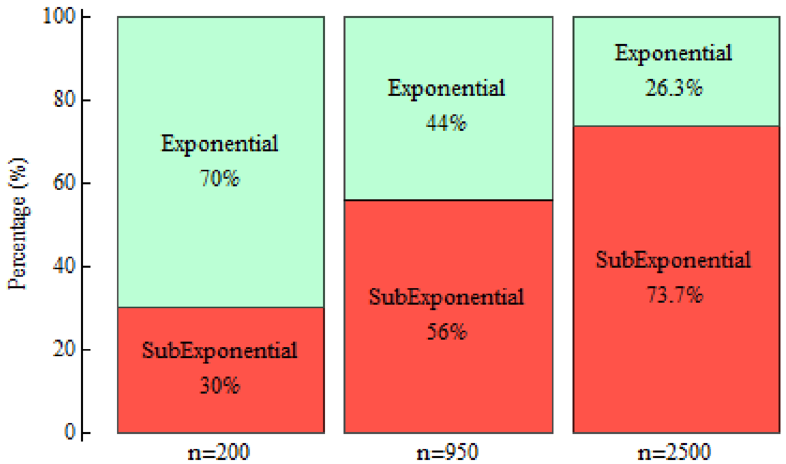
c = 0.3

1000 random samples are generated for various parameter values for the Weibull, Gamma, Pareto type II and LogNormal distributions, *n* being the tail size (considered to be 10% of the sample size). The specific *n* values shown in the graphs were selected in a way so that they are near the min, max and mean values of the tail size of our data

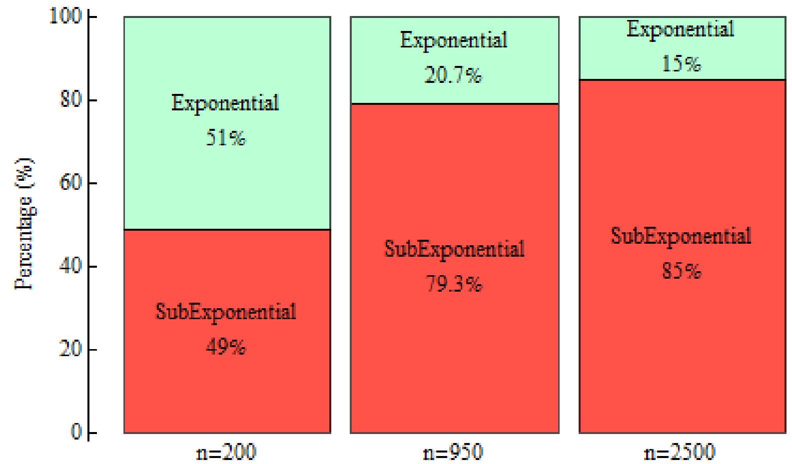
For the Weibull distribution, the increase of the shape parameter *c* implies increase of “exponentiality”. For parameter value equal to 0.8, which is an intermediate state, the results are not clear, as anticipated. For the Gamma distribution the percentage of “exponeniality” was high, which is expected because the behavior of the Gamma distribution tail is close to that of the exponential one.

6. Monte Carlo simulations on theoretical distributions (2)

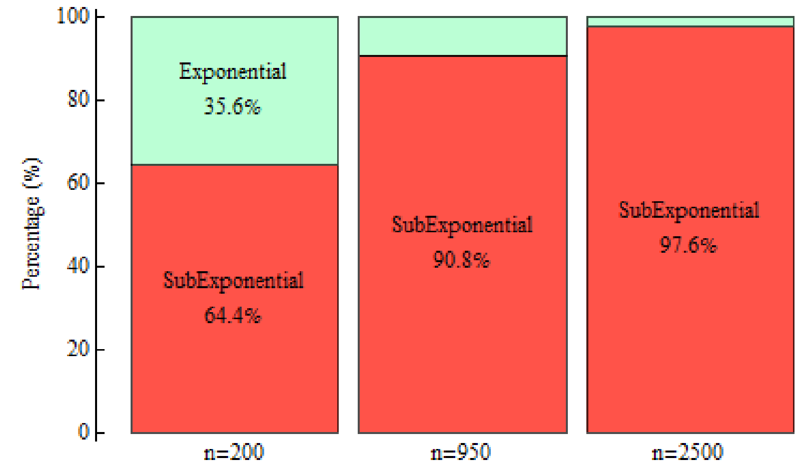
Pareto type II



c = 0.1

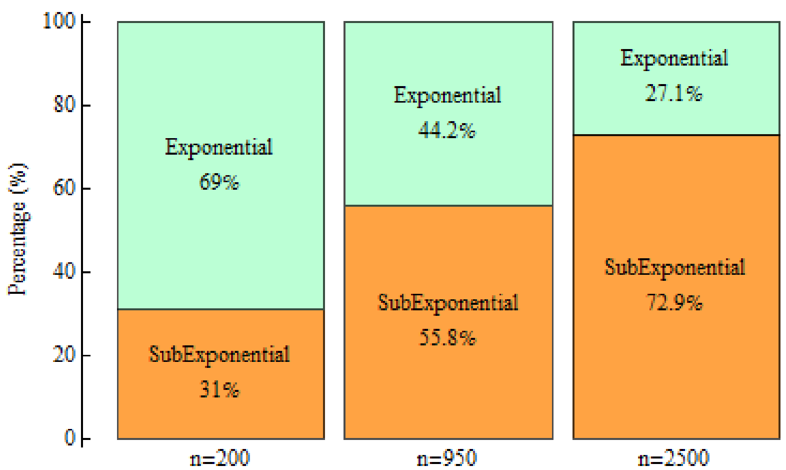


c = 0.15

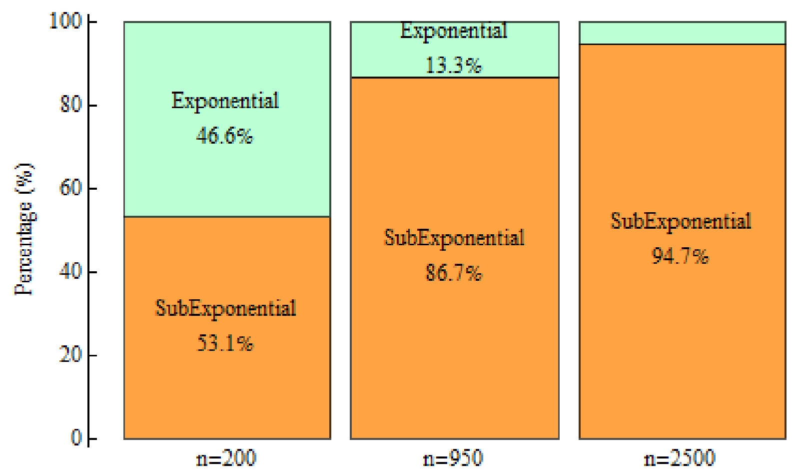


c = 0.2

LogNormal



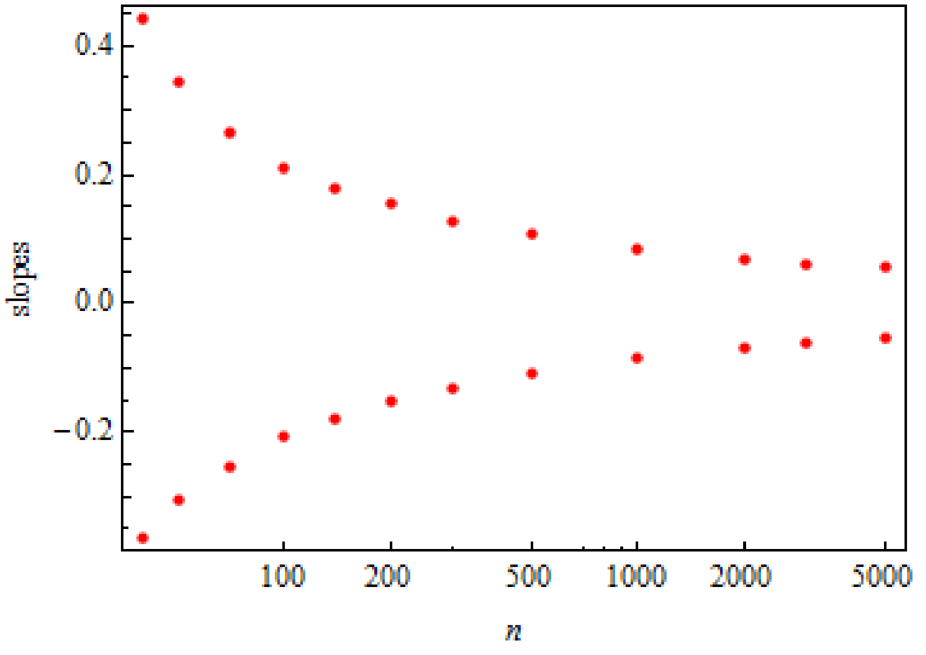
c = 0.6



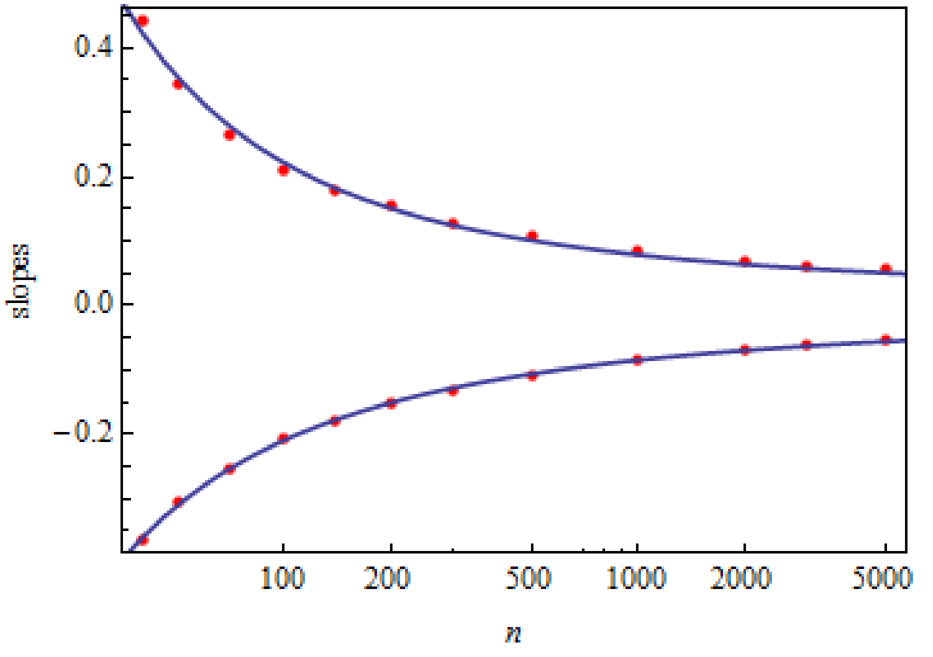
c = 0.8

For a low shape parameter value for the Pareto type II distribution the test is effective for large sample sizes. In our case the 80% of the stations had a record of over 6000 daily values, which corresponds to tail size of 600, and thus the results are satisfactory. The increase of the parameter augments the "subexponentiality" and the results are good even for small tail sizes. The same behavior is also observed for the LogNormal distribution.

8. Construction of MEF Test



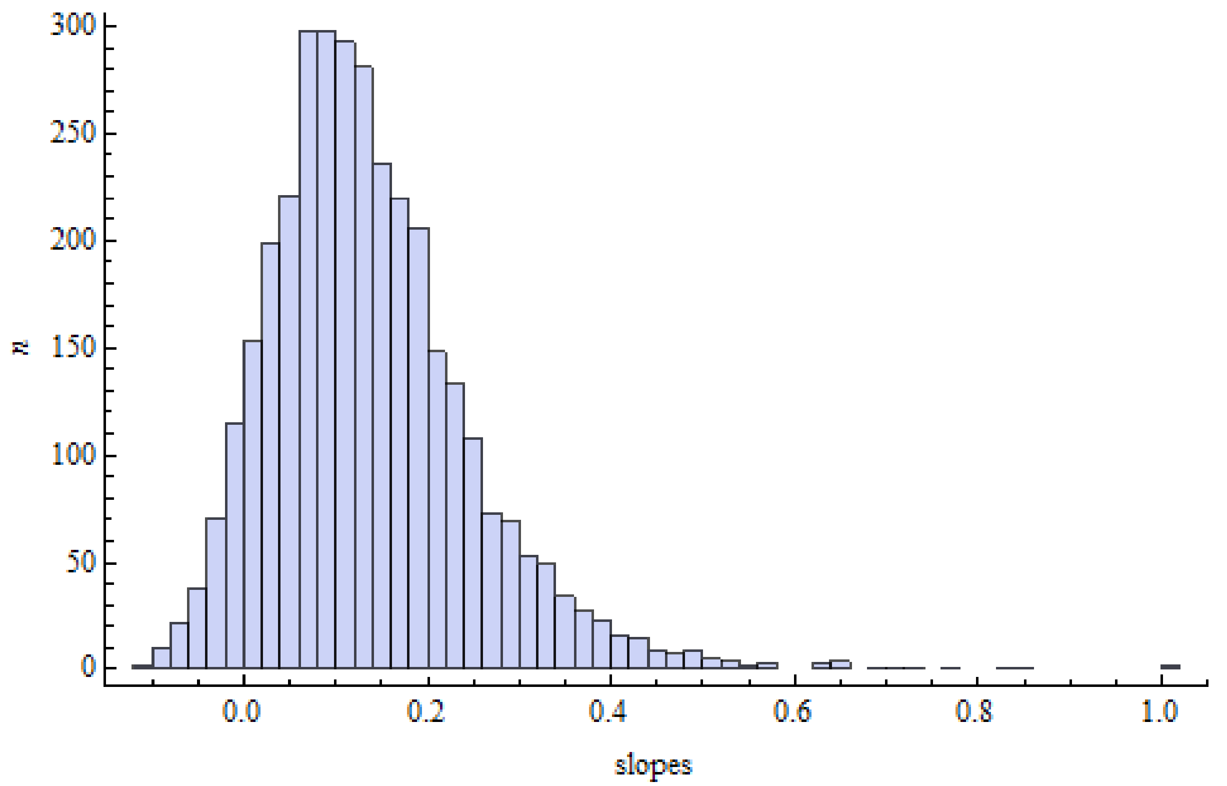
$$y = b\log\left(1+\frac{x}{d}\right)^{-c} \quad (1)$$



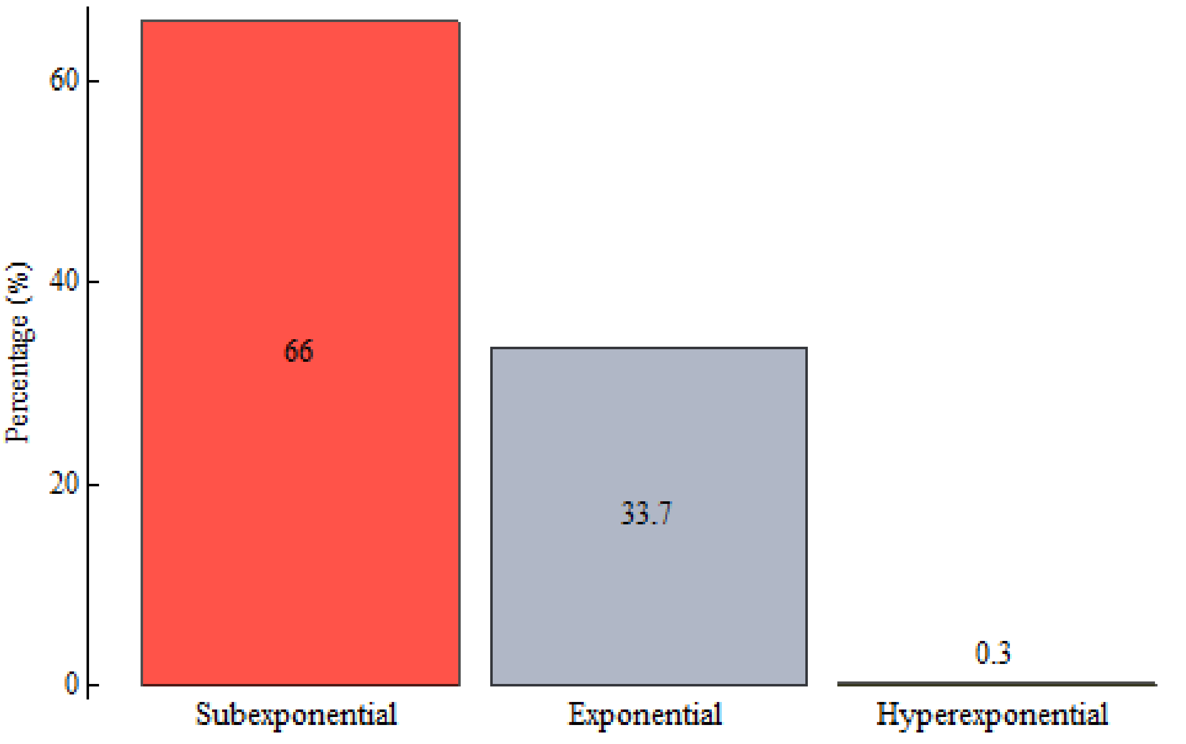
C.I.	Type of limit	*b*	*c*	*d*
90%	upper	0.41	1.25	23.86
	lower	-0.92	1.49	7.14

In order to implement the method on different sizes of records, we created an "Exponentiality Test". We generated 10 000 simulations for various tail sizes. Again, the tail size was considered as the 10% of the sample size. For 90% confidence interval we found the results shown in the first graph. The red dots represent the upper and lower boundaries of possible exponential tail slope values. We then applied a theoretical function of the form of eqn. (1) in order to implement the test on our data, and the values of the coefficients shown in the table were derived.

9. Implementation on data

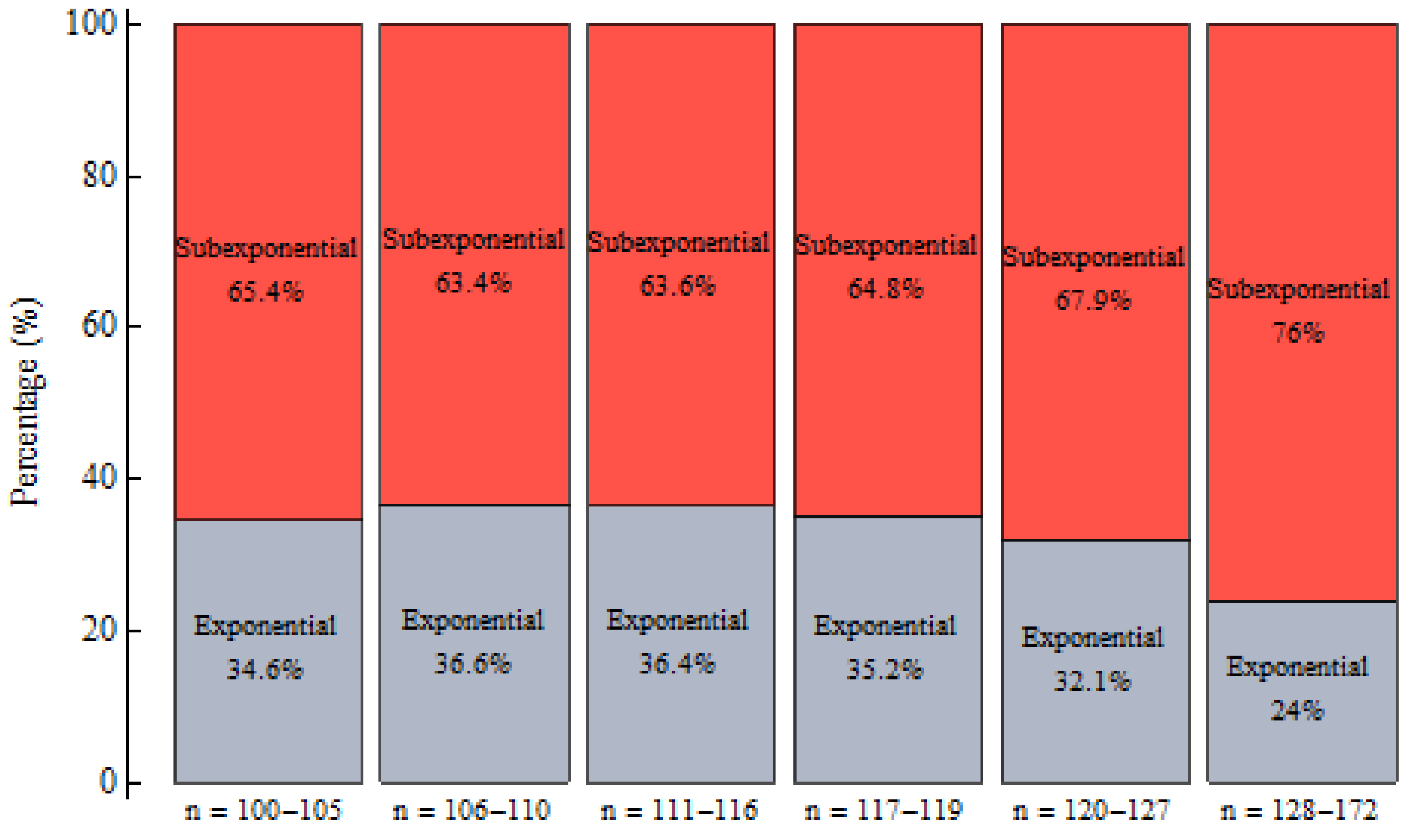


Min	Q1	Median	Q3	Max
- 0.852	0.062	0.121	0.194	2.402



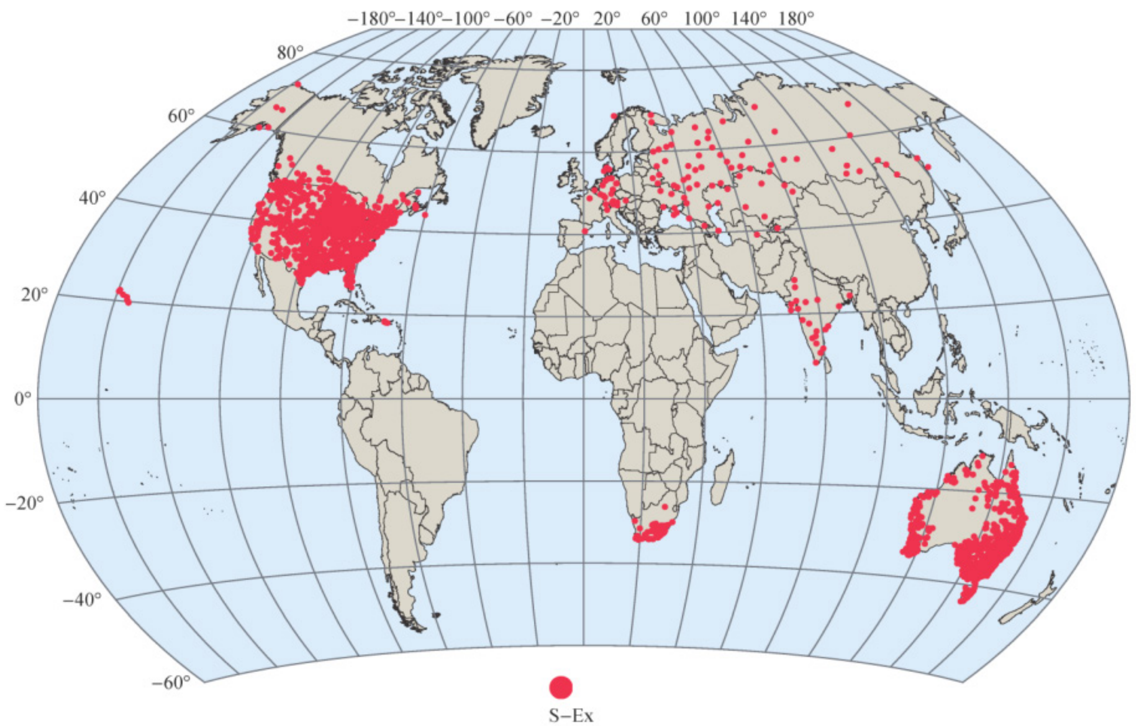
We considered the largest 10% of each record values as tail. The MEF was applied to all stations and the resulting slopes are shown on the left graph while their statistical characteristics are shown on the table. For 90% confidence, the majority of samples (66%) are characterized as subexponential according to the MEF test. In 33.5% of records the hypothesis of an Exponential tail cannot be rejected.

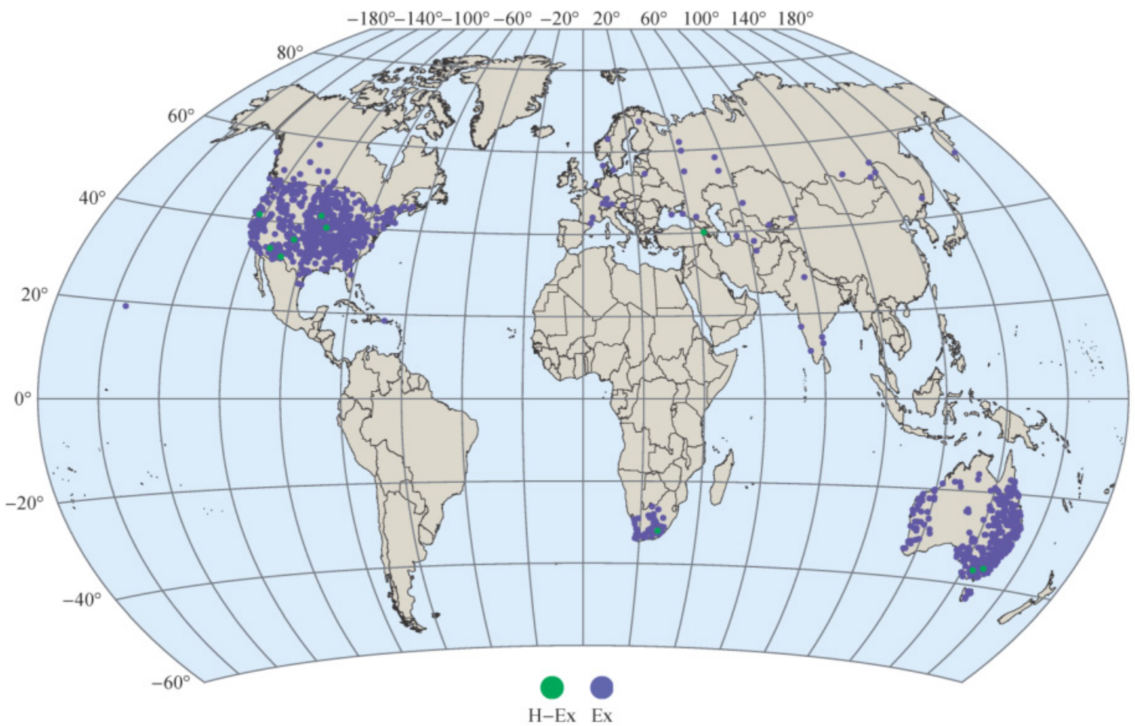
10. Importance of sample size

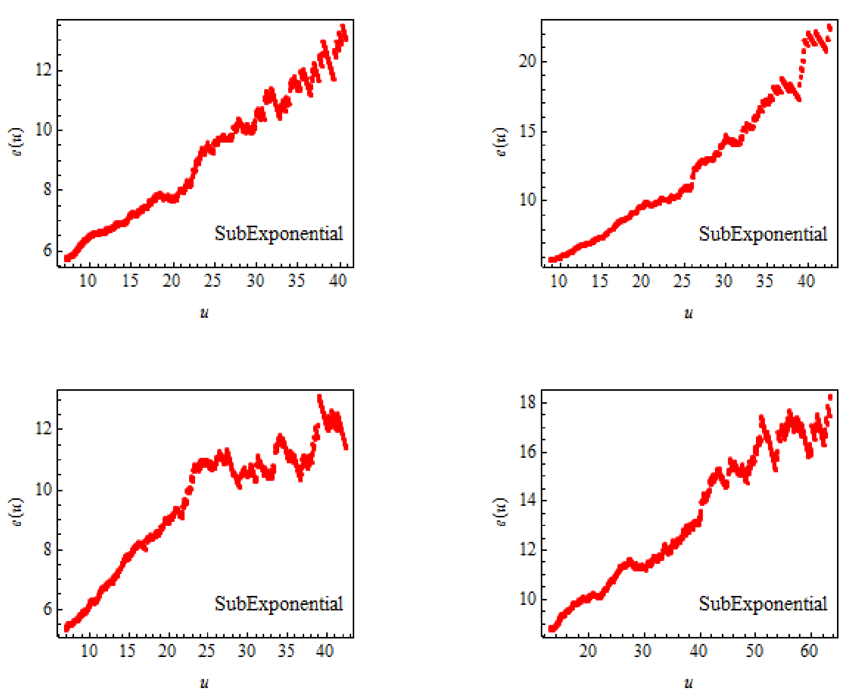


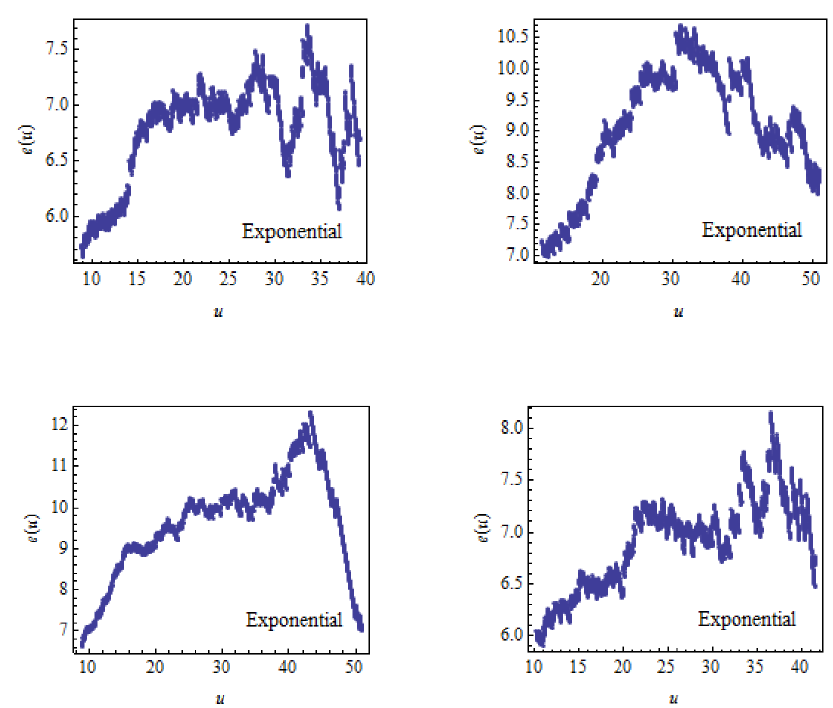
The graph shows the exponential-subexponential rate vs. tail size. The tail sizes were selected in such a way that in every category there were 500-600 stations with tail size in the specific category. For n = 111 years and over, the percentage of "exponential" tails decreases with the increase of tail size.

11. Global Map









12. Conclusions

- The mean excess function can be considered a reliable method for tail classification (Exponential, Subexponential, Hyperexponential) especially for large sample (and tail) sizes.
- According to the method, the majority of stations (66%) appear to have subexponential tails while for the 33.7% the hypothesis of exponential tail cannot be rejected.
- This means that rainfall extremes in their majority are better described by heavy tails (Papalexiou et al., 2012).
- The latest result contradicts the common practice of using light tails for simulating extreme rainfall.

References

- Benktander, G., & Segerdahl, C. 1960. On the analytical representation of claim distributions with special reference to excess of loss reinsurance. *in: XVIth International Congress of Actuaries.* Brussels.
- El Adlouni, S., B. Bobée, and T.B.M.J. Ouarda. 2008. On the Tails of Extreme Event Distributions in Hydrology. *Journal of Hydrology* 355 (1–4) (June 20): 16–33. doi:10.1016/j.jhydrol.2008.02.011.
- Embrechts, Paul, Claudia Klüppelberg, and Thomas Mikosch. 1997. *Modelling Extremal Events: For Insurance and Finance*. Springer.
- Papalexiou, S. M., Koutsoyiannis, D., and Makropoulos, C. 2012. How extreme is extreme? An assessment of daily rainfall distribution tails, *Hydrol. Earth Syst. Sci. Discuss.*, 9, 5757-5778, doi:10.5194/hessd-9-5757-2012.