Hydrol. Earth Syst. Sci. Discuss., 9, C4783–C4789, 2012
www.hydrol-earth-syst-sci-discuss.net/9/C4783/2012/

Hydrology and Earth System Sciences Discussions

Interactive comment on "How extreme is extreme? An assessment of daily rainfall distribution tails" *by* S. M. Papalexiou et al.

S. M. Papalexiou et al.

smp@itia.ntua.gr

Received and published: 14 October 2012

CC: Commenter's comments
AC: Authors' comments

We honestly thank P. Willems for his comments. We note that several times we refer him to our replies to other reviewers. This we did in order to be more concise, because some of his comments have been already answered in detail in our previous replies.

CC1. *The comment by Referee Clauset on the possibility that the heavy-tailed pattern observed may be due to non-stationary light-tailed processes is valid. In my paper (Willems, 2000) I have shown that POT extremes of rainfall intensities follow a mixed or two-component exponential distribution. However, when these extremes are studied*

per storm type or limited to a season where one storm type dominates, one-component exponential distributions were found. Combining rainfall extremes from different seasons and/or storm types may lead to the wrong conclusion that the distribution is heavy tailed. Calibration of a "wrong" heavy tailed distribution may in that case lead to a close match of the calibrated theoretical distribution and the empirical quantiles or exceedance probabilities, but may not lead to reliable extrapolations beyond the range of the empirical data.

AC1. The Commenter is referred to our reply to A. Clauset where we have answered this issue in detail. We also note that in an ongoing study we performed a seasonal analysis of nonzero daily rainfall (approximately in 200 000 time series of daily rainfall analysed in a monthly basis) based on L-moments diagrams and we have found no evidence in support of the Commenter's claim. We further note that light tailed and heavy tailed distributions are models of reality; we never claimed it is the reality itself. Mixtures thereof are again models of reality and are again heavy or light tailed distributions. We did not include mixtures of light tailed distributions and we are not going to include them in our study, because they contradict the principle of parsimony which should be a strong desideratum in such studies: the Commenter may see that we use distributions with only two parameters for the tails. Otherwise, if one does not care about parsimony, even a power type distribution can be approximated by a mixture of some exponential distributions for the range of distribution function which can be seen based on observations.

CC2. *To meet the concern raised in previous comment, next to matching the empirical quantiles or exceedance probabilities, reliable representation of the tail's shape and the asymptotic distribution properties towards higher return periods is of equal importance. This is certainly the case when the objective of the extreme value analysis is extrapolation of the distribution beyond the range of empirical return periods (as is the case in many engineering applications, and which is also the main focus of this* *paper). Note that in the statistical literature several methods have been proposed to directly estimate the distribution's shape parameter; see e.g. Beirlant et al. (1996), Kratz and Resnick (1996). Other methods are based on the analysis of asymptotic distribution properties in quantile plots (Willems et al., 2007). For heavy tailed datasets, the distribution's tail appears asymptotically linear towards the higher quantiles or return periods in a Pareto quantile plot (plot of the logarithmically transformed rainfall intensity versus logarithmically transformed exceedance probability). The asymptotic linear slope equals the (inverse of the) shape parameter. For datasets with exponential tails, asymptotic linear tail behaviour is observed in an exponential quantile plot (same as the Pareto quantile plot, but no logarithmic transformation applied to the rainfall intensity in ordinate). See also the similar comment by Referee Deidda (his comment 4).*

AC2. We would like to stress once again (as we did in our replies to other Reviewers and Commenders) that we have not made any kind of assumption before performing our analysis. We directly fitted and compared four different very common tails. Other methods and others approaches obviously exist or can be devised. We note that we are aware of the log-log plots and of many other graphical tools. Yet, generating and presenting more than 15000 graphs would be useless. In general the problem to quantify convexity or concavity into these graphs in order to draw conclusions is not trivial and certainly is out of the scope of this paper. However, we note that in an ongoing study we have "algorithmically translated" mean excess function plots and we will invite the Commenter to review that study, when completed, if he is interested.

CC3. *As Referee Laio, I was surprised to read that the performance of the different theoretical distribution tails was evaluated based on the error on the exceedance probability. In engineering design applications, quantiles are indeed estimated for given exceedance probabilities or return periods rather than exceedance probabilities estimated for given rainfall intensities. In extreme value analysis based on the analysis of*

the tail behaviour in quantile plots (e.g. Csörgo et al., 1985; Beirlant et al., 1996), it is common to apply weighting factors to the MSE computation (e.g. Willems et al., 2007). Most common are the weighting factors by Hill (1975).

AC3. The Commenter is referred to our reply to F. Laio where he can find the results of a Monte Carlo study that verifies our approach. Common practice is not always the best choice and we believe that in research we should not be surprised to see new/different methods sometimes.

CC4. *Rather than prior fixing the number of extremes or the POT threshold, the threshold could be optimized by minimizing the MSE. The MSE will increase for the smaller exceedance probabilities due to the increased variance when the parameter estimation is based on a lower number of observations (increased statistical uncertainty). When more extremes are considered, the bias in the asymptotic distribution's tail may increase and consequently the MSE may increase. In the intermediate range, the optimal threshold can be selected at the threshold with minimum MSE. Statistically principled determination of the threshold was also proposed by Referee Clauset.*

AC4. The Commenter is referred to our reply to A. Clauset (comment AC3) and also to our reply to S. Begueria (comment AC2).

CC5. *Rather than separating distribution tails in two categories, heavy and light tails, it is more common to use three classes of tails: heavy, normal and light. The shape parameter γ, also called 'extreme value index', is positive for heavy tails, zero for normal tails, negative for light tails. According to the sign of the extreme value index, the following three classes are traditionally considered for extreme value distributions: class I (for $\gamma > 0$), class II (for $\gamma = 0$), and class III (for $\gamma < 0$, having upper bound). The Generalized Pareto Distribution (GPD) (for PDS/POT extremes) but also the Generalized Extreme Value (GEV) distribution explicitly considers these three types for the same distribution. These types correspond with the three distribution families defined by the* *authors on p.5765 lines 14-15: sub-exponential, exponential and hyper-exponential.*

AC5. This study is about the tails of the parent distribution and not about annual maxima, to which GEV applies. The classification of tails we use is, thus, in accord to this fact and is not new (see e.g., Goldie and Klüppelberg, 1998).We will explain this better in the revised paper. Regarding the GEV distribution, which is another thing, the Commenter is referred to Papalexiou and Koutsoyiannis (2012a) (predecessor presentation of a study under review) to see a worldwide analysis (more than 15000 stations).

CC6. *I agree with Referee Deidda that when the shape parameter is close to 1, the most parsimonious model can be preferred because of the reduced variance in the parameter estimation.*

AC6. We also believe in parsimony as explained above (in AC1). If in the majority of stations the distribution of nonzero rainfall was close to exponential, then exponential distribution would be a nice model. However this is not the case (see e.g., Papalexiou and Koutsoyiannis, 2012b). Also, there are theoretical reasons disfavouring the exponential distribution (Koutsoyiannis, 2005). Furthermore, from an engineering point of view, the exponential distribution is in most cases dangerous as it produces too low values for high return periods.

CC7. *It is indeed surprising that the GPD distribution was not considered by the authors given that the distribution of excess values over a threshold (PDS/POT extremes) converges to the GPD, as was shown by Pickands (1975). This distribution includes the Pareto type II distribution (heavy tailed for $\gamma > 0$) used by the authors, the exponential distribution ($\gamma = 0$, normal tailed) and light tailed distribution ($\gamma < 0$). Same comment was made be other referees or commenters.*

AC7. The Commenter is referred to our detailed reply to R. Deidda (comment AC2 and AC3).

CC8. *I am not sure that lognormal distributions are heavy tailed (p. 5765 line 5). As also indicated by Referee Laio in his comment 1, the lognormal distribution has an exponentially decaying tail.*

AC8. The Lognormal is clearly a subexponential distribution (see e.g., Embrechts et al., 1997; Goldie and Klüppelberg, 1998; Mitzenmacher, 2004).

CC9. *I think the authors made a mistake on p.5764 line 18. When the shape parameter or extreme value index converges to zero, the Pareto type II distribution's tail degenerates to the exponential tail, and not for towards infinity as the authors write.*

AC9. The Commenter is right and we thank him (and R. Deidda) for spotting this typing error, which we failed to spot this in the proof corrections.

References

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

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

Koutsoyiannis, D.: Uncertainty, entropy, scaling and hydrological stochastics, 1, Marginal distributional properties of hydrological processes and state scaling, Hydrological Sciences Journal, 50(3), 381–404, 2005.

Mitzenmacher, M.: A brief history of generative models for power law and lognormal distributions, Internet mathematics, 1(2), 226–251, 2004.

Papalexiou, S. M. and Koutsoyiannis, D.: A global survey on the distribution of annual maxima of daily rainfall: Gumbel or Fréchet?, in European Geosciences Union General Assembly 2012, p. 10563. [online] Available from: http://www.itia.ntua.gr/en/docinfo/1207/ (Accessed 3 October 2012a), 2012.

Papalexiou, S. M. and Koutsoyiannis, D.: Entropy based derivation of probability distributions: A case study to daily rainfall, Advances in Water Resources, 45(0), 51–57, doi:10.1016/j.advwatres.2011.11.007, 2012b.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 9, 5757, 2012.