Hydrol. Earth Syst. Sci. Discuss., 9, C2524–C2530, 2012
www.hydrol-earth-syst-sci-discuss.net/9/C2524/2012/

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.

R. Deidda (Referee)

rdeidda@unica.it

Received and published: 26 June 2012

General comments.

The manuscript compares the performances of four probability distributions (Pareto type II, Lognormal, Weibull, Gamma) in fitting the largest records of daily rainfall collected by more than 15'000 rain gauges worldwide, with recording lengths ranging from 50 to 160 years. The outcome of the analysis is certainly of great interest for many scientists and, although not highlighted enough in the abstract and conclusions (see comments below), the results suggest that large rainfall values are better fitted by a (Generalized) Pareto distribution. I believe the Authors should better clarify the objective of the paper: Are they catching the heaviest tails ("the heavier, the better",

see my comment below) or the distribution that best fits extremes? I try to clarify this point in my specific comments below and, apart from: 1) some minor errors due to incorrect interpretations in the meaning of the shape parameters for two out of four distribution models used (see technical corrections), and 2) some other minor, but in my opinion relevant, issues I would like the Authors to address in a revised version of the manuscript, my judgment for the paper is positive and I encourage publication after minor revisions.

Specific comments.

1) "the heavier, the better" (page 5766, line 15) and other similar sentences (in several parts of the manuscript) can be misleading. Indeed, distributions fitted to the highest rainfall values can be characterized by a wide range of shape parameters and degrees of skewness. Thus, in my opinion, it is not that important to catch the highest tails, but the distribution that reliably fits heavy tailed as well as exponentially distributed records, as it is the case in this and other studies. Using rainfall records restricted to a limited region, I made some analyses on the tails of daily rainfalls (see reference below) and found that distributions are often heavy tailed but, in some cases, they can also be exponential. I have also found that the shape parameters can display patterns depending on orography (manuscript in preparation). Wilson and Toumi (2005) used the stretched exponential distribution (i.e. the Weibull in eq.(6) provided by the Authors) on a worldwide daily rainfall database and showed a dependence of the shape parameter on the geographic location. In some cases, the shape parameter of the stretched exponential was equal to 1, indicating an exponential tail. Thus, if the region of interest lies where the shape parameter is very close to 1, I would say "the exponential, the better"! In conclusion, I strongly suggest to reformulate some sentences in the manuscript to avoid misleading emphasis and, also, better convey the message "rainfall can display a wide range of more or less tailed extremes, the XYZ distribution can fit better whatever the shape parameter is".

2) Why not Generalized Pareto distribution? On Page 5764, lines 7-11, the four distri-

butions used for comparisons are classified as follows: “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”. Why the Authors avoid using the Generalized Pareto distribution (GPD), which also includes the Pareto type II distribution used by the Authors? The Generalized Pareto distribution has the advantage to describe heavy tailed (sub-exponential) distributions for positive values of the shape parameter, the exponential distribution (shape parameter equal to zero) and hyper-exponential distributions (negative shape parameter values). Recent studies by Begueria (2005), Deidda and Puliga (2006), and Begueria et al. (2009) reported strong evidence [based on L-moment ratio diagrams (Hosking, 1990)] that GPD is the best candidate for daily rainfall series. In addition, there are theoretical arguments to substantiate the use of GPD in fitting the excesses above proper thresholds (see e.g. Coles, 2001, Deidda, 2010 and references therein), while the adoption of Lognormal, Weibull or Gamma distribution models is not supported by extreme value theory. Concluding, I strongly suggest the use of a GPD rather than a Pareto type II model. The latter is included in the GPD family, corresponding to positive values of the shape parameter.

3) The first subplot in Figure 6 (top left), displays the shape parameter estimates for the Pareto type II distribution model. The irregularity in the unexpected large number of records filling the first bin, should be due to the use of a Pareto type II distribution model. More precisely, I suppose that the numerical algorithms for parameter estimation were bounded to provide ONLY positive shape parameter values and to avoid degeneracy of eq.(6)) when the shape parameter approach zero (exponential tail). From Table 1, one concludes that the lower bound set by the Authors should be equal to 0.001. Following my suggestion to fit a Generalized Pareto rather than a Pareto type II distribution model (see above), the Authors should also find shape parameter values equal to zero (exponential distribution) or smaller than zero (hyper-exponential distributions). I am quite confident that this new result will be coherent with those presented

in the bottom subplots of the same Figure for the Weibull and the Gamma distribution models. For those models, shape parameters larger than one are associated with hyper-exponential distributions. How to interpret these (few) estimates characterizing hyper-exponential distributions? On the basis of my modest opinion, after a visual inspection of the survival function (see next point for plotting suggestions), in most of the cases negative shape parameter values should be due to statistical variability (i.e. estimation variance) and the data could be reliably described using an exponential distribution model. However, this is just my personal opinion, based on experience from rainfall data originating from a limited geographical region and, hence, I do not expect to be generally accepted. Anyway, using the Generalized Pareto distribution, a negative shape parameter characterizes an upper bounded distribution.

4) Plotting of survival functions. A very useful diagnostic plot to identify different tail behaviors is that of the logarithm of the survival function versus recorded values. Linear behavior corresponds to an exponential tail, convexity characterizes sub-exponential (heavy-tailed) distributions, whereas concavity characterizes hyper-exponential (bounded) distributions. Since characterization of distribution tails is one of the main scopes of the paper (see e.g. the sentence/definition on page 5760 lines 3-4, which I like a lot: “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”), I suggest to use this kind of plots.

5) Threshold selection. This is still an unresolved issue. There are methods to cope with the uncertainty in determining exactly the optimum threshold (see e.g. Deidda 2010), but the authors skipped this issue and decided to consider a number of highest values equal to the number of years of observation. It is my opinion that an optimum threshold would allow inclusion of more values with consequent reduction of estimation variance, but I understand the Authors’ choice since they analyze a large amount of stations. In such a way they are almost sure the distribution of the excesses belongs to the domain of attraction of the Generalized Pareto distribution, but not necessary a

Pareto type II distribution.

6) Page 5763 lines 9-12: “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”. This sentence is theoretically incorrect: please remove or reformulate this sentence. Weights, such as that introduced in eq.(3), are sometimes applied to goodness of fit statistics for tails: indeed the CDF is usually S-shaped thus even a very small difference between empirical and fitted CDFs would imply a large error in the quantile. Anyway, the only theoretically consistent approach to treat each data point equally, is by building a norm on quantiles, as suggested by another reviewer.

7) Last but not least, as already suggested, Authors should make it clear, especially in the abstract, introduction and conclusion which are the objectives of the paper. I believe that the paper can contribute to identify the distribution which can best fit a wide range of tail types as those observed in rainfall time series (including exponential). I would remark that we cannot simply say “we need a heavy tailed distribution”, since, for instance, if we are performing a regional analysis we cannot apply a Lognormal in a station and a Pareto Distribution in a close station: usually we have to make a choice and use the same distribution over the study area.

Technical corrections.

8) Page 5764, line 18: “For $\gamma \rightarrow \infty$ it degenerates to the exponential tail ....”. This sentence is incorrect: eq.(4) tends to be an exponential distribution when γ approaches zero.

9) Page 5766, line 5: “for $\gamma > 1$ the distribution is sub-exponential and form and for $\gamma < 1$ hyper exponential.” Apart from errors in English usage, this sentence is also incorrect: the Gamma distribution is sub-exponential (heavy tailed) when the shape parameter γ is smaller than 1. Fixing these two errors in shape parameters interpretation and introducing the GPD will make all the subplots in Figure 6 coherent each other and will

help the readers to correctly draw their picture.

10) There are several errors in English usage, but I cannot be of any additional help, since my English is not that good.

References

Begueria, S.: Uncertainties in partial duration series modelling of extremes related to the choice of the threshold value, J. Hydrol., 303, 215–230, 2005.

Begueria, S., Vicente-Serrano, S. M., Lopez-Moreno, J. I., and Garcia-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, doi:10.1002/joc.1808, 2009.

Coles, S.: An introduction to statistical modeling of extreme values, Springer-Verlag, London, 2001.

Deidda, R. (2010), A multiple threshold method for fitting the generalized Pareto distribution to rainfall time series, Hydrology and Earth System Sciences, 14, 2559-2575, doi:10.5194/hess-14-2559-2010

Deidda, R., and M. Puliga (2009), Performances of some parameter estimators of the Generalized Pareto Distribution over rounded-off samples, Physics and Chemistry of the Earth, 34, 626–634, doi:10.1016/j.pce.2008.12.002

Deidda, R. (2007), An efficient rounding-off rule estimator: Application to daily rainfall time series, Water Resources Research, 43, W12405, doi:10.1029/2006WR005409.

Deidda, R., Puliga, M. (2006), "Sensitivity of goodness-of-fit statistics to rainfall data rounding off", Physics and Chemistry of the Earth, 31 (18) , 1240-1251 , doi:10.1016/j.pce.2006.04.041.3

Hosking, J. R. M.: L-moments: Analysis and estimation of distributions using linear combinations of order statistics, J. Roy. Stat. Soc. B Met., 52, 105–124, 1990.

Wilson, P. S., and R. Toumi (2005), A fundamental probability distribution for heavy rainfall, Geophys. Res. Lett., 32, L14812, doi:10.1029/2005GL022465.

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