Hydrol. Earth Syst. Sci. Discuss., 9, C4762–C4768, 2012
www.hydrol-earth-syst-sci-discuss.net/9/C4762/2012/

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

RC: Reviewer's comments
AC: Authors' comments

We thank the reviewer A. Clauset for spending time reviewing our paper. We wish to clarify from the beginning that we disagree with him. We feel that what the reviewer proposes is to follow his approach and methods as described in one of his papers; however we have a different opinion as we explain below in detail.

RC1. *In setting up their analysis, the authors assume that the underlying distribution generating rainfall events is stationary and therefore all events are drawn independently from some unknown underlying distribution. This is a common and reasonable*

assumption, but it also raises the possibility that the heavy-tailed pattern observed is not due to hydrological processes that produce stationary heavy-tailed distributions but rather due to non-stationary light-tailed processes. Testing this hypothesis is an important open question given the authors' results favoring heavy-tailed distributions. However, it may not be necessary to explore this question within this particular publication, but it should at least be discussed as another possible explanation for the observed patterns. Since the data are timestamped, I expect a number of tests of non-stationarity would yield interesting results without much additional work.

AC1. The second author invites the reviewer to read a couple of papers he has written about stationarity and nonstationarity (Koutsoyiannis, 2006, 2011) as well as Lins and Cohn (2011), hoping that perhaps he will agree that these two notions apply to models and not to the real-world processes themselves. In this case, statements like "the heavy-tailed pattern observed is not due to hydrological processes that produce stationary heavy-tailed distributions but rather due to non-stationary light-tailed processes" are not meaningful from a scientific point of view. The fact that heavy-tailed distributions can result by a mixture of light-tailed distributions is known. The reviewer is referred to Koutsoyiannis (2004) who gives an example how this can happen. But this is fine when the distributions and their change of parameters in time are known. If they are not known, then the resulting mixture is better modelled as a stationary distribution.

RC2. *Although the authors do not cast their work within the modern literature on extreme value theory in statistics (a comment made by another referee), I'm not too worried about this. In fact, there must be a physically imposed upper limit on the largest possible rainfall, which means the extreme tail of the distribution must be truncated by finite-size cutoff (exponential tail). The scientifically relevant questions, however, are whether this physical limit is low enough to impact any of the empirical data and what the shape of the distribution is below that cutoff. In this sense, many of the stronger* *results from extreme value theory may not apply and the central question of tail-fitting remains reasonable. Some points, however, do remain relevant, e.g., the classification of general tail structures, and the manuscript would be improved by at least briefly discussing these connections relative to the authors' stated goals.*

AC2. We cannot really understand the link of this comment to our paper. We believe that classic and simple always prevails in the end. We set a simple question, that is, which one of four basic tails performs better. To answer this question we figured out and applied a simple and clear method. Many other methods obviously can be figured out, and obviously they cannot be investigated in a single paper. We emphasize our strong belief that there is no meaning in assuming a "finite-size cutoff" or else an upper bound for rainfall. The philosophical issue of upper bounds in natural quantities was answered several decades ago by one of the giants of probability theory, William Feller (1906-1970). We quote here a paragraph form his celebrated book (Feller, 1971), that is always a joy to read. Regarding a person's life span he writes:

"The question then arises as to which numbers can actually represent the life span of a person. Is there a maximal age beyond which life is impossible, or is any age conceivable? We hesitate to admit that man can grow 1000 years old and yet current actuarial practice admits no bounds to the possible duration of life. According to formulas on which modern mortality tables are based the proportion of men surviving 1000 years is of the order of magnitude of one in $10^{10^{36}}$ a number with 10^{27} billions of zeros. This statement does not make sense from a biological or sociological point of view, but considered exclusively from a statistical standpoint it certainly does not contradict any experience. There are fewer than 10^{10} people born in a century. To test the contention statistically, more than $10^{10^{35}}$ centuries would be required, which is considerably more than $10^{10^{34}}$ lifetimes of the earth. Obviously, such extremely small probabilities are compatible with our notion of impossibility. Their use may appear utterly absurd, but it does no harm and is convenient in simplifying many formulas. MoreÂňover, if we were seriously to discard the possibility of living 1000 years, we should have to accept the

existence of maximum age, and the assumpÂňtion that it should be possible to live x years and impossible to live x years and two seconds is as unappealing as the idea of unlimited life."

The reviewer may also wish to see our earlier works that discuss the fallacy of upper bounds in precipitation (Koutsoyiannis, 1999, 2007; Papalexiou and Koutsoyiannis, 2006).

RC3. *As with many studies of rare events in empirical data, the authors are faced with the question of how to quantitatively identify the value above which the "tail" of the distribution may be modeled separately from its body. In their notation, this is the question of choosing x_L. I agree that there is currently no universally accepted method for choosing x_L; however, there are (more objective) methods with advantages over heuristic of choosing the largest N values that the authors employ. The issue is that choosing x_L too small means including some of the distribution's body in the empirical data, inducing bias in the subsequently estimated tail model parameters if the body follows a different structure than the tail, while choosing it too large means reducing the sample size and the statistical power of any model comparison technique. An arbitrary choice of x_L will lead to an uncontrolled tradeoff between bias and variance, and the resulting conclusions may not be trustworthy. Although there is no single best way to objectively solve this problem, one increasingly popular approach is described in SIAM Review 51(4), 661-703 (2009), which chooses x_L automatically and in a statistically principled manner for each data set.*

AC3. Essentially, here the reviewer proposes to follow his approach as described in his and his colleagues' paper entitled "Power-law distributions in empirical data" (Clauset et al., 2009). In their analysis they estimate the best choice of x_L assuming that that the sample is generated from a power-law distribution. However, this is not the case here. In our study, we do not fit only a power-law tail and we do not assume a priori that the data are generated by a power law distribution. Thus, the x_L choice claimed to be best for power-law distribution may not be best for the others. Additionally, Clauset et al. estimate x_{L} assuming a nonzero lower bound, which apparently is an unrealistic assumption with respect to rainfall whose lower bound is precisely zero. Further, they assume that an exact power law exists beyond the lower limit, so that the tail would be $(x/x_{\mathrm{L}})^{-\gamma}$ (as they write in equation (2.6) of their paper). However, there cannot be any power-law distribution defined for the actual lower bound of rainfall, that is, zero. Thus, a better option is to assume that the tail is proportional to $(1 + x/x_{\mathrm{L}})^{-\gamma}$. Note that in our case the quantity x_{L} is not necessarily regarded as a lower bound. In other words, we did not try to find the best choice for a lower bound x_{L} of a power-law distribution but to determine x_{L} so that the values above it can characterize the tail of any distribution. The exceedence probability of x_{L}, resulting as we defined it, is low. Please see our response to Begueria where we provide some analytical equations that will be incorporated in the revised text.

RC4. *Finally, one choice by the authors did mystify me: why use what is essentially a least squares regression approach to fitting the distributional models when one could instead use the more universally accepted and more statistically principled approach of maximum likelihood? Using maximum likelihood to estimate the model parameters would also allow the comparison of models using powerful techniques like the Vuong likelihood ratio test. This would provide much stronger evidence in favor of one model over another, and would also allow the decision that two or more models are statistically indistinguishable given the current data. One approach to conducting this kind of test is described in the same SIAM Review article mentioned above. For the scientific questions being addressed here, likelihoods seem like a superior methodological approach and I would encourage the authors to consider them. Now, it may be that the authors' existing results would continue to stand under the likelihood approach, but they may not. Either way, the results and conclusions would be placed on more firm methodological footing.*

AC4. We are sorry for mystifying the reviewer. Again we think that here the reviewer proposes to follow his approach in his aforementioned paper. We have performed a Monte Carlo simulation (see the response to F. Laio) to validate our method and concluded that it performs very well resulting essentially in unbiased estimation of the parameters. In the exploratory phase of our research we also tried the maximum likelihood method (in a Monte Carlo framework) but we found poorer results. Finally, the least square method is a well-known and scientifically accepted method. Obviously, many different approaches exist. In any case, our paper is not about comparing estimation methods but about comparing distributions. In our paper we provide the link for the database and the reviewer may feel free to use the method of his own preference to find his own results and hopefully compare them to our results.

References

Clauset, A., Shalizi, C. R. and Newman, M. E. J.: Power-Law Distributions in Empirical Data, SIAM Review, 51(4), 661–703, doi:10.1137/070710111, 2009.

Feller, W.: An introduction to probability theory and its applications, 2nd ed., John Wiley & Sons Inc, New York., 1971.

Koutsoyiannis, D.: A probabilistic view of Hershfield's method for estimating probable maximum precipitation, Water Resources Research, 35(4), 1313–1322, 1999.

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

Koutsoyiannis, D.: Nonstationarity versus scaling in hydrology, Journal of Hydrology, 324, 239–254, 2006.

Koutsoyiannis, D.: A critical review of probability of extreme rainfall: principles and models, in Advances in Urban Flood Management, edited by R. Ashley, S. Garvin, E. Pasche, A. Vassilopoulos, and C. Zevenbergen, pp. 139–166, TaylorFran. [online] Available from: http://www.itia.ntua.gr/en/docinfo/758/ (Accessed 11 October 2012), 2007.

Koutsoyiannis, D.: Hurst-Kolmogorov dynamics and uncertainty, Journal of the American Water

Resources Association, 47(3), 481–495, 2011.

Lins, H. F. and Cohn, T. A.: Stationarity: Wanted Dead or Alive?, JAWRA Journal of the American Water Resources Association, 47(3), 475–480, doi:10.1111/j.1752-1688.2011.00542.x, 2011.

Papalexiou, S. M. and Koutsoyiannis, D.: A probabilistic approach to the concept of probable maximum precipitation, Advances in Geosciences, 7, 51–54, 2006.

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