Dr. Andreas Langousis

Sc.D. in Civil and Environmental Engineering
Department of Civil and Environmental Engineering,
Massachusetts Institute of Technology, M.I.T.

MSc in Civil and Environmental Engineering
Department of Civil and Environmental Engineering
Massachusetts Institute of Technology, M.I.T.

Diploma in Civil Engineering
National Technical University of Athens, N.T.U.A.

✉ 65 Thessalonikis st.
Piraeus, 18545, Greece

☎ +30 (210) 4619520
+30 (693) 6898793

andlag@alum.mit.edu
http://alum.mit.edu/www/andlag

A multiple threshold method for fitting the generalized Pareto distribution and a simple representation of the rainfall process

Author: R. Deidda

Comments to Author

Paper summary

Define I_d to be the average rainfall intensity inside duration d and denote by $I_{d,u} = (I_d - u|\ I_d > u)$ the excesses of I_d above a certain threshold u. Since rainfall is an intermittent phenomenon, the cumulative distribution function (CDF) of I_d, $F_{I_d}(i)$, can be modeled as

$$F_{I_d}(i) = \begin{cases} 0 & , \quad i < 0 \\ P_0 & , \quad i = 0 \\ P_0 + (1- P_0)\, F_{I_d|I_d>0}(i) & , \quad i > 0 \end{cases} \tag{1}$$

where P_0 is the probability of zero rainfall and $F_{I_d|I_d>0}$ is a theoretical CDF that adequately fits the non-zero rainfall intensities in the empirical record.

Commonly used distribution models for $(I_d \mid I_d > 0)$ include the Gamma (Pearson III), log-Gamma (log-Pearson III), skewed normal (i.e. a normal distribution fitted to the Box-Cox transformed data), and lognormal; see e.g. Swift and Schreuder (1981) Kedem *et al*, (1990a,b; 1997), Shimizu (1993), Katz (1999), Cheng and Qi (2002), Cho *et al*. (2004), Shoji and Kitaura (2006), Langousis and Veneziano (2007), Langousis *et al*. (2007), Suhaila and Jemain, (2007).

For rainfall excesses $I_{d,u} = (I_d - u|\ I_d > u)$ above a properly selected threshold u, commonly used distribution models include the exponential, Gamma, lognormal, Weibull and generalized Pareto; see e.g. Rosbjerg *et al*. (1992), Onof and Wheater (1994), Cameron *et al.* (1999, 2000), Madsen *et al.* (1997a,b; 2002), Deidda and Puliga (2006, 2009).

The reviewed manuscript presents an approach to model the distribution of I_d assuming that $(I_d \mid I_d > 0)$ has generalized Pareto (GP) distribution with location parameter $e = u = 0$. The scale, $\bar{a}_0$, and shape, $\bar{\xi}_0$, parameters of the GP distribution and the probability of zero rainfall, $\bar{P}_0$, are obtained by following the procedure below.

1) Use different thresholds u_j to fit a GP distribution model to the rainfall excesses I_{d,u_j}
2) for each threshold u_j, estimate the probability $P_{u_j} = P[I_d \leq u_i]$ and the scale, a_{u_j}, and shape, ξ_{u_j}, parameters of the GP distribution,
3) for each j, use theoretical relationships based on properties of the GP distribution to calculate the parameters $(P_{0,j}, a_{0,j}, \xi_{0,j})$ from the parameters $(P_{u_j}, a_{u_j}, \xi_{u_j})$ obtained in step 2,
4) obtain $\bar{a}_0$, $\bar{\xi}_0$ and $\bar{P}_0$ as the median values of the estimates calculated in step 3.

The suggested methodology to model the distribution of I_d was applied to 217 time series of daily rainfalls, each with more than 40 years of complete recordings, collected by the Sardinian Hydrological Survey (Italy). The method showed good performance in reproducing the empirical distributions of daily rainfalls at different locations, even in the case of heavily quantized (i.e. rounded off) data. This is an attractive feature of the suggested approach, since currently used methods to fit theoretical distribution models to rainfall excesses require specification of an optimum threshold that is difficult to determine when the data are heavily quantized.

Contribution and audience

The reviewed manuscript presents a methodology to model the average rainfall intensity inside duration d, I_d, using results obtained by fitting a theoretical distribution model to rainfall excesses $I_{d,u} = (I_d - u \mid I_d > u)$ above a wide range of thresholds u. The subject is of interest to a wide audience of hydrologists and, definitely, within the scope of the Journal.

Technical soundness, organization and style

The manuscript is technically sound and well written. In addition, all Figures and Tables are necessary.

Prior publication

To my knowledge, neither the same nor very similar work has been published elsewhere.

Recommendation

For the reasons mentioned above, it is recommended that the paper is published in *HESS* after minor revisions. A list of comments and suggestions is presented below.

General comment 1:

The reviewed manuscript presents an approach to model the marginal distribution of the average rainfall intensity, I_d, inside duration d. Note that by modeling the marginal distribution of rainfall intensities, one cannot resemble several statistics of the rainfall process, including its autocorrelation function, the alteration of wet and dry intervals and, more in general, any joint distribution function of rainy and/or non rainy intervals in the record. That said, I suggest the author changes the title of the manuscript to: “A multiple threshold method for fitting the generalized Pareto distribution to rainfall data”.

General comment 2:

As stated in the Paper summary, in my opinion, the most attractive feature of the presented approach is that it gives good fits to the empirical distribution of I_d even for heavily quantized data. I suggest the author highlights this feature of the method in both the Abstract and Introduction.

General comment 3:

In a recent study, Veneziano *et al*. (2009) showed that, for multiplicative processes such as rainfall, the excesses, $I_{d,u}$, of I_d above any finite threshold u have not yet converged to a GP distribution. That said, even if a GP distribution provides good fits to the rainfall excesses over different thresholds u, the estimated shape parameter is larger than that calculated asymptotically as $u \to \infty$. I suggest the author mentions the findings of Veneziano *et al*. (2009) in the introduction, since they strengthen his arguments in fitting a GP distribution over a wide range of moderate thresholds, rather than selecting a single high threshold.

General comment 4:

Change: exceedances of thresholds

To: exceedances over thresholds

This comment applies to the whole manuscript.

General comment 5:

The font size used in all figures is small. I suggest the author increases the font size to assure readability of the figures in the final version of the manuscript.

Specific comment 1 (page: 4958; lines: 9-10):

In the suggested approach, the probability of zero rainfall P_0 is estimated theoretically based on the corresponding parameters of a GP distribution fitted over a wide range of thresholds $u > 0$. That said, P_0 can be seen only as a distribution parameter and it is not expected to adequately model the probability of zero rainfall in the empirical record. For this reason, I suggest the author rephrases his statement "… which is able to describe zero and non-zero values of rainfall time series by assuring a perfect overlapping with the GPD…" to something more conservative such as "…which assures a perfect overlapping with the GPD…"

Specific comment 2 (page: 4958; line: 13):

Change: will only

To: is expected to

Specific comment 3 (page: 4958; line: 16):

Change: on the exceedances of a

To: using exceedances over a

Specific comment 4 (page: 4958; lines: 24-26):

Add references for different theoretical distribution models used to describe the marginal distribution of rainfall intensities. A list of references can be found in the Paper summary above.

Specific comment 5 (page: 4959; lines: 10-13):

Add references on weather generators based on Markov chains.

Specific comment 6 (page: 4959; lines: 14-16):

Change: Nevertheless, despite the simple form of Eq. (1) would suggest to fit $F_0(x)$ on all strictly positive rainy observations, a particular care should be taken in this (seemingly very simple) approach.

To: The simple form of Eq. (1) suggests to fit $F_0(x)$ on all strictly positive rainy observations, but a particular care should be taken in this (seemingly very simple) approach.

Specific comment 7 (page: 4959; lines: 23-24):

Change: are empirical evidences

To: is empirical evidence

Specific comment 8 (page: 4959; line: 26):

Change: should be

To: is needed

Specific comment 9 (page: 4959; line: 28):

Change: same and unique distribution

To: same distribution

Specific comment 10 (page: 4960; lines: 12-13):

Change: the shape parameter is expected to be the same, while the other parameters are linked by theoretical relations

To: the shape parameter is expected to be the same asymptotically as the threshold $u \to \infty$, while the other parameters are linked through theoretical relations

Specific comment 11 (page: 4961; lines: 8-9):

Change: optimum one

To: optimum

Specific comment 12 (page: 4964; line: 7):

Change: $1\text{-}\frac{1\text{-}F(x)}{1\text{-}F(u)}$.

To: $1\text{-}\frac{1\text{-}F(x)}{1\text{-}F(u)}$ for $x > u$.

Specific comment 13 (equation 5):

Add: for $x > u$

Specific comment 14 (page: 4965; lines: 18-19):

Add references to methods.

Specific comment 15 (page: 4966; lines: 9-10):

Change: on the exceedances of this threshold

To: using exceedances over this threshold

Specific comment 16 (page: 4966; line: 12):

Change: In doing it

To: In doing so

Specific comment 17 (page: 4967; lines: 8-12):

Please rephrase this paragraph. Equation (10) is obtained by substituting equation (6) into equation (1), while the parameters a_0 and ζ_0 are calculated from equations (8) and (9).

Specific comment 18 (page: 4968; line: 11):

Change: expect invariance

To: expect threshold invariance

Specific comment 19 (page: 4968; line: 21):

Change: correctly recorded

To: recorded

Please do so, also, in other parts of the manuscript where reference is made to rainfall data recorded at 0.2 mm resolution.

Specific comment 20 (page: 4969; lines: 8-9):

Change: calculate the corresponding reparametrizations

To: calculate the parameters

Specific comment 21 (page: 4970; line: 16):

In the light of General Comment 3,

Change: not yet in the domain of attraction of the GPD

To: not fitted by a GPD

Specific comment 22 (page: 4971; lines: 15-17):

I do not understand. Please rephrase or eliminate.

Specific comment 23 (page: 4971; lines: 15-23 and Figure 3):

I suggest the author includes in Figure 3 a plot of a GP distribution fitted using a traditional method. This would illustrate the performance of the new MTM method relative to other, more traditional, methods for fitting a GP distribution to data.

Specific comment 24 (page: 4972; line: 19):

Change: be now

To: now be

Specific comment 25 (page: 4972; line: 28):

Eliminate the exclamation mark.

Specific comment 26 (page: 4973; line: 4):

Change: our

To: the empirical

Specific comment 27 (page: 4973; line: 5):

Change: results very

To: is

Specific comment 28 (page: 4973; line: 6):

Change: for

To: when

Specific comment 29 (page: 4973; lines: 25-26):

Change: Although not deeply checked, the largest ML estimates can be due

To: We investigated the issue to some detail and the larger ML estimates should be due

Specific comment 30 (page: 4975; line: 23):

Change: a comparisons between the empirical survival functions

To: a comparison between the empirical survival function

Specific comment 31 (page: 4976; line: 25):

Change: parameters

To: parameter

Specific comment 32 (page: 4978; line: 25):

Change: not

To: non

References

Cameron, D., K. Beven, and J. Tawn (2000), An Evaluation of three stochastic rainfall models, *J. Hydrol.*, **228**, 130-149.

Cameron, D., K. Beven, J. Tawn, S. Blazkova, and P. Naden (1999), Flood frequency estimation for a gauged upland catchment (with uncertainty), *J. Hydrol.*, **219**, 169-187.

Cheng, M. and Y. Qi (2002) Frontal Rainfall-Rate Distribution and Some Conclusions on the Threshold Method, *J. Appl. Meteor.*, **41**, 1128–1139.

Cho, H.K., K.P. Bowman and G.R. North (2004) A Comparison of Gamma and Lognormal Distributions for Characterizing Satellite Rain Rates from the Tropical Rainfall Measuring Mission, *J. Appl. Meteor.*, **43**, 1586–1597.

Deidda, R. and M. Puliga (2006), Sensitivity of goodness-of-fit statistics to rainfall data rounding off, *Physics and Chemistry of the Earth*, **31**, 1240-1251.

Deidda, R. and M. Puliga (2009) Performances of some parameter estimators of the generalized Pareto distribution over rounded-off samples, *Phys. Chem. Earth*, **34**, 626–634, doi:10.1016/j.pce.2008.12.002.

Katz, R.W. (1999) Extreme Value Theory for Precipitation: Sensitivity analysis for climate change, *Adv. Wat. Resour.*, **23**, 133-139.

Kedem, B., L.S. Chiu and Z. Karni (1990a) An Analysis of the Threshold Method for Measuring Area-Average Rainfall, *J. Appl. Meteor.*, **29**, 3-20.

Kedem, B., L.S. Chiu, and G.R. North (1990b) Estimation of Mean Rain Rate: Application to Satellite Observations, *J. Geophys. Res.*, **95**(D2), 1965–1972.

Kedem, B., R. Pfeiffer and D.A. Short (1997) Variability of Space–Time Mean Rain Rate, *J. Appl. Meteor.*, **36**, 443–451.

Langousis A, D. Veneziano, P. Furcolo, and C. Lepore (2007) Multifractal Rainfall Extremes: Theoretical Analysis and Practical Estimation, *Chaos Solitons and Fractals*, doi:10.1016/j.chaos.2007.06.004.

Langousis, A. and D. Veneziano (2007), Intensity-Duration-Frequency Curves from Scaling Representations of Rainfall, *Wat. Resour. Res.*, **43**, doi: 10.1029/2006WR005245.

Madsen, H., P. S. Mikkelsen, D. Rosbjerg, and P. Harremoes (2002), Regional estimation of rainfall intensity-duration-frequency curves using generalized least squares regression of partial duration series statistics, *Water Resources Research*, **38(**11), 1239, doi:10.1029/2001wr001125.

Madsen, H., P.F. Rasmussen, and D. Rosbjerg (1997a), Comparison of Annual Maximum Series and Partial Duration Series Methods for Modeling Extreme Hydrologic Events 1. At-site Modeling, *Wat. Resour. Res.,* **33**(4), 747-757.

Madsen, H., P.F. Rasmussen, and D. Rosbjerg (1997b), Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events 2. Regional modeling, *Wat. Resour. Res.,* **33**(4), 759-769.

Onof, C. and H.S. Wheater (1994), Improvements to the modeling of British rainfall using a random parameter Barlett-Lewis rectangular pulse model, *J. Hydrol.*, **157**, 177-195.

Rosbjerg, D., H. Madsen, and P.F. Rasmussen (1992), Prediction in Partial Duration Series with Generalized Pareto-Distribute Exceedances, *Wat. Resour. Res.*, **28,** 3001-3010.

Shimizu, K. (1993) A Bivariate Mixed Lognormal Distribution with an Analysis of Rainfall Data, *J. Appl. Meteor.*, **32**, 161–171.

Shoji, T. and H. Kitaura (2006) Statistical and Geostatistical Analysis of Rainfall in Central Japan, *Computers & Geosciences*, **32**, 1007-1024.

Suhaila, J. and A.A. Jemain (2007) Fitting Daily Rainfall Amount in Malaysia Using the Normal Transform Distribution, *J. Appl. Sci*., **7**(14), 1880-1886.

Swift, L.W. and H.T. Schreuder (1981) Fitting Daily Precipitation Amounts Using the S_B Distribution, *Mon. Wea. Rev.*, **109**, 2535–2540.

Veneziano, D., A. Langousis and C. Lepore (2009) New Asymptotic and Pre-Asymptotic Results on Rainfall Maxima from Multifractal Theory, *Wat. Resour. Res.*, **45**, W11421, doi:10.1029/2009WR008257.