Hydrol. Earth Syst. Sci. Discuss., 7, C55–C59, 2010
www.hydrol-earth-syst-sci-discuss.net/7/C55/2010/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Introducing a rainfall compound distribution model based on weather patterns sub-sampling" *by* F. Garavaglia et al.

Anonymous Referee #1

Received and published: 22 February 2010

The manuscript proposes to infer separately the statistical daily rainfall distributions of predefined classes of rain events based on meteorological synoptic situations for a better estimation of daily rainfall quantiles. The idea is not new, but examples of its application are seldom presented in scientific papers. It is therefore to my opinion an interesting contribution to the literature on statistical inference applied to hydrological data. The manuscript is short, clear and makes appropriate acknowledgements to the existing literature on the topic.

Nevertheless, the content of the manuscript and especially the arguments put forward to convince the readers of the benefits of the proposed approach if compared to a standard approach have nevertheless some clear weaknesses that must corrected before publication:

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

- Nuances should be introduced in the theoretical justification of the approach. The fact that the sub-samples are “statistically homogeneous” can not be demonstrated. It should be clearly presented as a hypothesis. Moreover, even if so, there is no good reason to think that the statistical distribution of each isolated sub-sample will better fit the predefined simplistic statistical models – if possible exponential - than the distribution of the compound sample. Extreme value theory is not a question here (see detailed comments).

- The efficiency or robustness of the proposed approach can not be demonstrated on a single example. If the paper is based on one single example of application, judgements on efficiency and robustness should be removed.

- The example is not particularly well chosen: the highest observed rainfall values seem to belong essentially to one class (pattern 4) and one season (autumn). Why are so many efforts done to isolate these extremes when the standard approach consisting in selecting annual maxima or POT values would provide more or less the same sample? Can clearer examples be selected where the sample of the largest values has a clearly non-exponential distribution – which is not the case here – and where the distribution of sub-samples can be well approximated by exponential distributions?

- Confidence bounds are terribly missing, especially if robustness is an issue. Is there not possible to use Bayesian-MCMC inference techniques to assess inference uncertainties, especially for the proposed approach which is based on a larger set of data but involves also more parameters: the parameters of the sub-samples distributions but also the occurrence of each class! This increase of the number of parameters is a priori not a factor of robustness. . .

Moreover, I would appreciate more objectivity in the presentation of the method and results. The approach is interesting even if it partly fails. It is not necessary to demonstrate at all cost that it is revolutionary.

- The comparison in fig. 9 is not fair. There is no reason to fit a GEV distribution to the

HESSD

7, C55–C59, 2010

Interactive Comment

whole data set and exponential distributions to the sub-samples according to the mean residuals life plots (fig. 5) and to the empirical distribution (fig. 9).

- The compound distribution is only defined for intensities greater than 18 mm/day. This should be better explained and indicated on the figures. It does not reproduce the “bend of the empirical distribution in the range of observable frequencies” in the present case. The interpolated part of the distributions should be removed.

Detailed comments:

1)In the introduction, a large emphasis is put on extreme value theory as if it was obvious that the distributions of the largest observed hydrological data, which according to the sampling methods and the depth of observation are mainly composed of “frequent” data, were necessarily well approximated by one of the extreme value distributions. Of course it is not the case. The extreme value theory is an asymptotic theory and data set heterogeneity is only one of the possible reasons for quantile underestimations, the first cited by Coles et al. (2003) being the ignorance of statistical model itself. Even if statistically homogeneous, there is no reason why the distribution of data sub-samples should be better approximated by a GEV distribution than the distribution of the whole sample. The justification of the approach based on extreme value theory is absolutely not convincing and really naïve to take the words of Coles et al. (2003). This introduction has to be reworded in a more conditional form: this work is a test of a method that may provide more accurate estimates for this and this reason. . .

2)Line 30: how is it possible to evaluate that the sub-samples are homogeneous from a statistical point of view? It is, I think, not possible!

3)Line 47: replace Lyons by Lyon also in the rest of the manuscript.

4)Line 60: the classification of weather patterns used is oriented towards the detection of heavy precipitation situations unlike what is said in the manuscript since it has been calibrated to provide quantitative precipitation estimates. The comments on lines 114-

HESSD

7, C55–C59, 2010

Interactive Comment

120 are very welcome.

5)Part 3.1: the whole presentation is disturbing me (see comment 1). Line 167: the selected observations, annual maxima for instance, can not be considered as “extreme” values, at least not in the sense of the extreme value theory.

6)Line 183: replace “true” by “real”.

7)Line 186: if heavy rains “most likely occur during fall”, then a simple extraction of maximum values will predominantly contain fall season events and the calibrated extreme distribution will be controlled by these fall season events. What is then added value of a seasonal or by event type analysis (see also next comments)?

8)Line 192: if we are interested in the extreme values, the separation in seasons must rather be based on the higher bound of the boxplot. The highest are observed between May or June and November. The lowest are not observed between May and August but in winter: please correct this.

9)Line 221: “according to the asymptotic theory. . .. can be fitted with a GP distribution”. I disagree totally with this statement. What element can be put forward to show that the asymptote is close???

10)Part 4.2: The conclusions drawn from fig 5 are completely subjective! The authors conclude what they want to conclude. The residual life plot is much more “horizontal” for the whole data set than for the sub-sets. It is particularly hard to consider it as horizontal for WP4, which is the data set which controls the shape of the distribution for the higher quantiles. . . An exponential model could also have been chosen for the complete set to be consistent.

11)Line 249: the fact that the sub-samples are more “homogeneous”, from a statistical point of view I suppose, is absolutely not demonstrated!

12)Line 250: the probabilistic model is more parsimonious is not linked to the proposed method but to the arbitrary choice of a shape parameter equal to 1.

HESSD

7, C55–C59, 2010

13)Line 251: Âń To provide a continuous probabilistic description . . . the CDF of each sub-sample is extended below its threshold by a linear interpolation of empirical quantiles Âż. It is very important to have mentioned this fact. Strictly speaking, the fitted distribution does only hold for intensities exceeding 18 mm/day. It would have been preferable for the soundness of the paper to draw only this distribution on figures 8 and 9.

14)Line 270: “Being the WP. . .. “. This holds for autumn but not for the summer which is not too far from the autumn season (fig 8). Is there not a problem in the proportions indicated in table 3? Or is a distribution fitted for each WP and each season, which would reduce significantly the usefulness and meaning of WPs?

15)Line 295: the bend is significant for return periods significantly lower than 1 year, generally not taken into account in statistical adjustments and neither in this case since the real threshold is 18 mm/day (return period of about 1 year). Since this bend corresponds to the range of values for which “a linear interpolation of empirical quantiles” is used, the proposed approach does not really represent it!

16)Line 307: The proposed adjustments should be compared with the adjustment of an exponential distribution of the maximum annual values or POT values exceeding 18 mm/day. The “robustness” – I do not like this ambiguous term- of the approach is not linked to the approach but to the constraint introduced in fixing the shape parameter value of the GEV equal to 1 (i.e. imposing an exponential model). The four adjustments can not directly be compared.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 7, 313, 2010.