Hydrol. Earth Syst. Sci. Discuss., 10, C3087–C3091, 2013
www.hydrol-earth-syst-sci-discuss.net/10/C3087/2013/

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Assessment of statistical characteristics of point rainfall in the Onkaparinga catchment in South Australia" *by* M. M. Rashid et al.

Anonymous Referee #2

Received and published: 9 July 2013

The authors present a study of the statistical characteristics of rainfall at 13 raingages over the Onkaparinga catchment. Three single distributions (Weibull, Exponential and Gamma) and one compound (hybrid Gamma-Generalized Pareto) are fitted to daily rainfall over a 51 year period. Rainfall is simulated from each fitted distribution for observed rainy days. The fitness of each distribution is then assessed by computing several statistics for different time-aggregation level (daily, monthly, seasonal and annual). The precipitation concentration index is computed in order to evaluate the concentration of rainfall in time.

GENERAL COMMENTS 1) This work focuses on a specific catchment and since the

application study is limited (as opposed to study which covers several catchments with different climate located accross the world), I think the authors could check more than four distributions for daily rainfall. For instance, as mentioned by the authors, Li et al. 2012 checked 6 distributions. In particular, I suggest evaluating the mixture of 2 exponentials (Wilks 1999) which has been widely used. Also, other hybrid distributions have been proposed for rainfall in the literature. While some of these distributions are not easy to implement, I would suggest to select at least the hybrid exponetial-GP proposed by Li et al. 2012 which can be implemented with MLE. Moreover, this distribution circumvents the need for thresold selection. I think that the problems encountered with the hybrid Gamma-GP of Furrer & Katz 2008 might be due to the fitting procedure. Indeed, the fact that the Gamma is fitted to the entire rainfall range might result in the Gamma being stretched between the need to fit the central and the extreme part. Hence, the Gamma tend to over-estimate lower quantiles and under-estimate the higher quantiles. If the fitting procedure would allow to fit simultaneously the central and the extremal part, that would certainly produce a better fit. Although the hybrid of Li et al. 2012 uses the Exponential rather than the Gamma for the central part, still it would be interesting to see the impact of using a unified fitting procedure.

2) The study and modeling of the spatial dependence structure is a key challenge in current rainfall modeling. Indeed, the evaluation of risk such as the probability of rainfall to exceed a given level in a catchment depends strongly on the spatial dependence structure. I regret that the paper do not address the issue of spatial dependence at all. In Nagongondo et al. 2011 (cited in the paper), the spatial correlation of rainfall is analysed. To complete the study of rainfall characteristics, the authors should, in my opinion, include, for instance, the following spatial statistics : - spatial correlation at several temporal lags - a measure of spatial dependence such as the extremal coefficient or the madogram (see the paper from Vannitsen & Naveau, Spatial dependences among precipitation maxima over Belgium, Nonlinear Processes in Geophysics, vol 14, 2007)

SPECIFIC COMMENTS 1) Introduction : I am curious to know more about the peculiarities of the Onkaparinga catchment, the specificities of rainfall over the catchment and why the study of rainfall is of particular revelance. This is partly address in the data section but I think it should be mentioned right from the start why the authors chose to study rainfall in this particular catchment.

2) Section 3.2 : for completeness, please provide the CDF of the Gamma and the GP which appear in Eq.(7)

3) Section 3.3 : I am not familiar with the index of aggreement Eq. (12) and more details on how it should be interpreted would be welcomed.

4) Section 3.4 : Same kind of comment for PCI. I would like to get more insight into this criterion. The authors could refer to the explanations given in De Luis et al. 2011 (cited in the paper) in the last two paragraphs of the second page.

5) Section 3.5 : I understood that rainfall is simulated only for observed wet days. This is not so obvious and probably should be stated more clearly.

6) Section 4 : the discussion on the fact that the distributions are less skewed and have less kurtosis when the level of temporal aggregation increases is not surprising. this is simply a consequence of the central limit theorem (the sum of i.i.d. variables with finite variances converge to a Normal distribution). This should be mentioned.

7) Section 4.1.1 : I would tend to think that the MLE of the shape and scale parameters of the Gamma might not be independent. Did the authors checked this ? For instance, the scale and shape parameters of the GP are not independent : a smaller scale estimator can be compensated by a larger shape parameter. Dependence in the MLE would affect the discussion on the spatial distribution of the estimators.

8) P.5987 line 10 : the Normal distribution has two parameters, not one, maybe the authors had in mind the Exponential distribution rather than the Normal ?

9) The discussion over the Gamma MLE is not totally clear to me. The skewness

of the Gamma is given by the formula : 2/sqrt(shape) and the excess kurtosis by 6/shape. Since the shape parameter estimate is always around 1 (Fig. 3), I do not see how this could reflect the skewness and kurtosis computed from the observations in Table 3. The Gamma distribution approaches the Normal distribution when the shape parameter is large, about 10. This do not seem to be the case.

10) Section 4.1.2 : more should be said about the issue of threshold selection of the GP distribution. It results in a bias-variance trade-off (higher threshold == better fit of the GP, lower threshold == less variance of the MLE). Choulakian and Stephens 2001 "Goodness-of-fit Tests for the Generalized Pareto Distribution" Technometrics Vol 43 have proposed a Goodness of fit test for the GP which can be used for threshold selection. Dupuis 1998 "Exceedances over High Thresholds: A Guide to Threshold Selection", Extremes Vol.1 has proposed also a sound way to performed threshold selection. Trial and error is not the only option and certainly not the most rigorous one. Also, according to results, it seems that probably the best threshold is not the same for all stations.

11) P. 5988 line 10 "Probability distributions were applied to observed rainfall at different stations and the daily rainfall were estimated. " seems odd. May I suggest : "Probability distributions were fitted to observed rainfall at different stations and the daily rainfall were simulated. "

12) Regarding Table 6 and Fig. 7 : I would like to have global measure of fit such as KS or MAE to evaluate the benefit of the new threshold for the hybrid. Clearly, I am expecting the Gamma and the hybrid to perform exactly the same when we look at percentiles under the threshold. In this regard, I do not think that Table 6 and Fig. 7 are much informative.

13) P. 5989 line 25 : As I said in the general comments, I think that the fitting procedure of the hybrid might explain the poor performance for the annual and seasonal total.

14) P. 5990 line 10 : I think that checking whether the fitted distribution can reproduce

the monthly and annual total is important. However, if one's goal is to use monthly or annual rainfall, then one should fit a distribution to monthly and annual rainfall, not to daily rainfall and then aggregate. The reason for this is that, as mentioned by the authors, thanks to the central limit theorem, aggregated rainfall is easier to model.

15) Fig. 9 : I am not sure what is shown in the figure : monthly standard deviation and skewness of daily rainfall. Are these the statistics of rainfall aggregated at the monthly level ?

16) P. 5990 line 20 : The GEV is a sound model for block maxima such as annual maxima. I do not understand how the authors use the GEV ? For the 99th annual percentile ? How is this justified ? Please provide details since this is an unusual application of the GEV.

17) Figure 10 : I would like to see the results for all 13 stations or at least some comments about them. The same comment is valid for the tables and figures which present only the results for one or few stations.

18) P. 5992 line 10 : "The PCI analysis indicates that the period from December to May is more susceptible to extreme events than the period June to Novemebr". I am not sure how the PCI is informative of the occurrenc of extreme events. I understand, from Figure 12, that a higher PCI from December to May means more concentrated precipitation. I do not see how extremes are linked with this.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 10, 5975, 2013.