Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2018-22-RC2, 2018
© Author(s) 2018. This work is distributed under
the Creative Commons Attribution 4.0 License.

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Local and regional flood frequency analysis based on hierarchical Bayesian model: application to annual maximum streamflow for the Huaihe River basin" *by* Y. Wu et al.

Anonymous Referee #2

Received and published: 23 March 2018

The paper present a hierarchical Bayesian model for flood frequency analysis and compares its performance against some standard methods for a case study in China. The use of Bayesian Hierarchical models for flood frequency analysis has been proposed elsewhere by several authors and in that sense the paper is not very innovative, although it presents an interesting application. The paper is well structured, although it contains some typos and grammar mistakes, so I would recommend to proof-read the manuscript again. I am not entirely convinced by the evidence the authors bring to support their hypothesis that the model proposed in the model would outperform the "good old" RFA approach in practice: they only show the estimated frequency curves of two stations, without quantifying the gain when Bayesian methods are used.

I have a few major comments regarding the manuscript given below:

1. The different methods compared in the paper partially correspond to different sets of priors: Coles and Dixon (1999) discuss how the L-moment estimation enforces some constrains on the value of the shape parameter, while it is known that Bayesian techniques are somewhat equivalent to MLE methods when diffuse priors are used. The fact that 3 different estimation methods are used for the different pooling approaches makes it really hard to understand whether the (claimed) better performance is due to the estimation is due to the estimation procedure (which is tightly linked with the prior used for the shape parameter) or the pooling strategy. One could do at-site Bayesian estimation with similar priors for the each of the site and devise a Bayesian fully-pooled model to make a comparison which is actually based on the pooling strategy. I appreciate the authors are trying to compare the proposed estimation procedure to the most commonly used approaches, but the comparisons are somewhat spurious. Finally the authors claim that the proposed model performs better than any other method, but I find the evidence to support this statement weak - since it is based mostly on the visual inspection of the flood frequency curves of two stations. More formal criterion could be employed - see for example Kobierska et al. (2017).

2. Page 6 - line 16-17: the model the authors suggest is perfectly valid, although looking at Figure 3 it is obvious that both $\log(\mu)$ and $\log(\sigma)$ change with a similar rate as a function of the log(Area). I wander if the authors could have employed a simpler model (which somehow corresponds to the model used in the index flood approach) by taking $\log(\sigma_i) = k \log(\mu_i)$, with $k \sim N(\bar{\kappa}, \tau_\kappa)$ corresponding to a overall "coefficient of variation" term. Further I appreciate that Area does not seem to have an effect on the value of the shape parameter, but the fact that some of the sample values are positive and some are negative begs the question of whether any other catchment property (steepness? permeability?) could be used to classify catchments with upper or lower bound (and by the way we do not know which parametrisation of the GEV is used in the manuscript). By shrinking all the shape parameters towards some overall average

value, we are forcing things towards a Gumbel distribution, which might be the wrong distribution. This goes back to the question of whether the stations in the analysis can be deemed to be a "homogeneous region" , which essentially corresponds to assuming that the shape parameter can be taken to be the same across all stations.

3. Page 7 - line 6: considering the recommendation in Gelman et al (2017) to prefer weakly informative priors to non-informative ones it is strange to see the Gelman et al (2014) citation to justify the use of diffuse priors. In general I would believe that some knowledge is available about proprieties of the GEV parameters - see for example the "geophysical" prior of Martins and Stedinger (2000). The use of vague non-informative priors can be valid, but much of the recent research in applications seem to indicate that some weakly informative priors might be a better reflection of our current understanding of the properties of parameter values.

4. Section 3.2: somewhere in the results it would be good to have some more information on the MCMC chains and reassurance on the convergence of the MCMC procedure (this is mentioned in page 8 line 5-9, if 15000 sample were taken this seems quite a small number for such a complex problem. What is the equivalent sample size?)

5. What do the posterior for α_1 and β_1 look like? Are they centred around 1? Could these be used to derive some scaled version of the flows? In general I would thing that the posterior of the original parameters would be of interest int he paper (although maybe to be placed in a supplementary material section).

6. Page 10 - line 5. Intervals derived from MLE are typically called confidence not credible intervals (credible intervals are obtained in the Bayesian framework). There is little information given on how the confidence intervals for the MLE estimation are built, my guess is the Delta Method on the quantile itself. One option which might maybe avoid the unpleasant behaviour of the interval covering negative values would be to use the profile likelihood approach or to use the delta method for the log(quantile), since the normal approximation will hardly hold at such high return periods as those shown

in the Figures.

7. Page 11 - line 16-18: there seem to be an indication that the RFA assumes homogeneity, while the HB model does not make this assumption. I would argue that assuming that the shape parameter comes from a common normal distribution is a form of homogeneity assumption.

8. Page 11 - line 19: I do not see why using a HB gives a better way to use short records. In the typical RFA setting shorter records get a smaller weight in the estimation of the regional parameter. I can agree that the Bayesian representation is more elegant, but essentially for short records the final estimated value will simply depend more on the priors and on the other stations in the region. Conversely, stations with short records will be given less weight in the estimation, both in HB and RFA.

9. Table 1: are the the negLik, AIC and BIC of just one series? These quantities can not really be calculated by putting together data across different stations... Further, the AIC/BIC is more often used to compare models with different complexity rather than different distributions, for which tests like the Anderson Darling and Kolmogorov-Smirnov Test are more frequently used.

Minor technical points:

is Equation (4) a likelihood or a posterior? The notation of the formula is not very clear (why the subscript H? is i a parameter?)

Page 7 - line 10: what is the β in Λ? I also guess that if the α_* and β_* are included in the vector, one should not have μ and σ

Figure 5 and 6: is the y-axis on the log scale? This is not very common - might be worth making it explicit in the axis.

References

Coles, and Dixon (1999). Likelihood-based inference for extreme value models. Ex-

tremes.

Gelman, A.; Simpson, D.; Betancourt, M. (2017) The Prior Can Often Only Be Understood in the Context of the Likelihood. Entropy, 19, 555.

Kobierska, Engeland, Thorarinsdottir (2017), Evaluation of design flood estimates - a case study for Norway. Hydrology Research; DOI: 10.2166/nh.2017.068

Martins, E. S., and J. R. Stedinger (2000), Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data, Water Resour. Res., 36(3), 737-744, doi:10.1029/1999WR900330.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-22, 2018.