A hierarchical Bayesian model for extreme pesticide residues

I would like to thank the authors for writing an interesting paper that I can imagine is very relevant for the readership of this journal. It is always pleasing to see methodological research in extreme value theory being put into good use, and it is obvious that the hierarchical model for extreme pesticide residues employed in this paper much improves over a standard analysis which uses maximum likelihood to fit lognormal distributions to each dataset independently. I think the paper is extremely well–written and accessible, even to those with a limited knowledge of extreme value theory. Statistically, the methods used in this paper are not new – although I doubt this would be a pre–requisite for publication here. However, I believe the application *is* novel, and I would like to see such modelling procedures used more widely by practitioners. Before this paper is published, I have a few points I would like the authors to consider – I doubt they *all* require action, but rather thought or clarification.

1. In Section 3.2 the authors set up the hierarchical model and introduce the Bayesian paradigm as a framework for making inferences. The authors say that "Uncertainty about the parameters of these distributions is treated using Bayesian updating" (page 7, line 48), without any reference to Bayes' Theorem or the Bayesian paradigm generally. Almost in passing, the authors refer to Markov chain Monte Carlo (page 8, line 12). Generally, would the audience of this journal not benefit from a more detailed discussion of Bayesian methods, including an outline of how MCMC works? perhaps a more detailed discussion of Bayesian methods, with reference to Bayes' Theorem and the link between the posterior, the prior and the likelihood is more important than discussing the nuts–and–bolts of MCMC, but I still feel the audience of this journal might benefit from a more thoughtful treatment here.

2. Page 8, line 24–44: The authors discuss the use of a modified scale parameter $\tilde{\sigma}_j$. I think the discussion about why this is used in place of σ_j (because it gives threshold independence) should have been mentioned earlier in Section 3.1.1 when the authors discuss threshold sensitivity. It is the threshold stability property of the GPD that allows us to assess our initial choice of threshold u by fitting to $u* > u$ and looking for stability in estimates of $(\tilde{\sigma}, \xi)$. However, in the context of the hierarchical model being used in this paper, and working within the Bayesian framework, the threshold stability property of the GPD gives another reason for preferring to work with a modified scale parameter $\tilde{\sigma}$: since σ is dependent on u, use of the GPD(σ, ξ) model may be restrictive since an uninformative prior for σ then becomes informative at higher thresholds. Perhaps this could be mentioned?

3. In Sections 4.1 and 4.2, as far as I can tell, the authors do not give any indication of the speed of convergence or indeed how they checked for convergence of their sampler. For example, perhaps multiple starting points were chosen and convergence to the same stationary distribution after burn–in was observed? How dependent were successive draws from the posterior, and was thinning required? Did the authors make use of any package for fitting, e.g. `WinBugs`? If we cannot expect the reader to be equipped with a knowledge of MCMC methods, would the inclusion of some sample MCMC output (e.g.

trace plots) be helpful? Perhaps these are technicalities that the general readership of this journal will not be interested in?

4. I think the authors should make it clear that their comparisons between the results from the Bayesian hierarchical model and the likelihood approach are not like–for–like. In fact, the increased precision in parameter estimation from the hierarchical model relative to the separate datasets analyses using maximum likelihood is not an artefact of working within the Bayesian framework, but rather because of the pooling of information between datasets. Although the comparisons made are definitely informative, the hierarchical model could also have been fitted using maximum likelihood, possibly with similar precision?

5. Have the authors considered making use of the posterior predictive distribution for high quantiles here? Working within the Bayesian framework has obvious appeal, especially for extreme value analyses: we can account for uncertainty in parameter estimation via the prior distributions for $\log\lambda_j$ and ξ_j, and we have the facility to augment our analysis with carefully chosen forms for the hyper–parameters $a_\lambda, \zeta_\lambda, a_\xi, \zeta_\xi$. However, working within the Bayesian framework, we can also allow for randomness in future observations through the predictive distribution. In my experience, practitioners like an estimate of an extreme quantile, or return level, that has all sources of variability built into it. For example, if an estimated high quantile is to be used as a design requirement for the height of a sea–wall, engineers might rather be provided with a predictive return level estimate than a return level estimate with a measure of variability attached (e.g. posterior mean plus posterior standard deviation, or posterior credible interval).

6. Other minor comments:

   – The precision parameters in Equation (3) are specified by ζ_λ and ζ_ξ, but on page 8 these look more like ς_λ and ς_ξ.

   – Is it really the case that there is no serial correlation between successive threshold exceedances? Usually, temporal clustering above a high threshold is observed.

   – In Section 2, could the authors include graphical summaries of the some of the datasets used?

   – Page 5, line 14 – “informatio” should be “information”.

   – Do the circles in Figures 2 and 4 correspond to posterior means for the 0.999 quantile? If so, this should be stated. How do the posterior means compare with medians? What are the units of measurement?

I look forward to receiving the authors’ responses, and any revisions as a result.