

Beta- P distribution with applications to extreme hydrological events

I would like to thank the authors for writing an interesting paper, the content of which is relevant for the readership of the *Environmetrics* journal. The authors describe the use of the Beta- P distribution for characterising extremes of hydrological data, and compare its use to that of more standard modelling techniques from extreme value theory.

Although I do feel that this work has the *potential* to make a useful contribution to the literature, I have some reservations about the paper in its current form. I will list these now, in order of importance.

1. Use of English: grammar and punctuation

The paper requires careful editing before it can be considered for publication. As I read through the article, I started to make a note of places in the text that required attention; after a few pages, I realised that the *entire manuscript* needed attention. In its current form, I *can* understand the aims, methods used and outcomes of the paper; however, poor use of English can, at times, detract from the discussion, making the paper difficult to follow. The authors need to proof-read their manuscript, thinking carefully about how their work is written.

2. Value of the Beta- P distribution

Although I follow the authors' discussion of the Beta- P distribution, and the justification for using it as a model for extremes (Section 3), I'm not sure what is gained by using this distribution over more standard models for extremes from classical extreme value theory (EVT). For example, in the analysis of the Colorado stream flow data (Section 6.1), the authors acknowledge that the "... GEV and beta- P have roughly the same capability for fitting annual maximum streamflow [data]", and that Figure 5 shows that "...the beta- P and GEV distribution fits the data almost equivalently". I agree with the authors here – as we can see from Figure 5, there is hardly any difference in the goodness-of-fit between the beta- P and the GEV. Although the p -values for the various models tried (Table 2) indicate that, for the streamflow data, there is less evidence against the beta- P than there is against the Gumbel/GEV, all three seem plausible here. The authors acknowledge this in the last sentence of Section 6.1 when they say "... the beta- P distribution can describe the above streamflow data **slightly** better than the other distributions listed in Table 1.". Other than a **slightly** better fit, what is gained by using the beta- P ? As the authors discuss, there is even less support for the beta- P when analysing the wave height data than there is for the streamflow data.

Actually, maybe it's not so important to show that the beta- P distribution provides a substantial improvement over the classical EVT models; perhaps the authors' intention is to provide the practitioner with an alternative model to consider at the model-fitting stage. If this is the case, perhaps the authors should make this clear in their aims? As it stands, the paper provides a useful case study of an application of the beta- P model to

hydrological extremes – I just think more focus should be given to the potential usefulness of this model in practice. I’m not sure if a case study showing the application of an existing model to extremes is – in its own right – enough to warrant publication.

Does Figure 6 show that uncertainty in return level estimation is reduced when using the Beta- P ? If so, why might this be the case?

3. Length of article

I think – for what the paper does – it is too long. Indeed, there are various ways in which the authors could reduce the length of the paper. For example,

- the introduction is too long and, in my opinion, could be written much more concisely;
- I’m not sure if it is necessary to explain the checks of model adequacy in quite as much detail as is given in Section 3.1 – for example, the Kolmogorov–Smirnov, Anderson–Darling and AIC are now well-known, commonly-used methods aiding the model selection process. Perhaps a few references here would suffice;
- the summary/discussion section could be shortened, without loss of clarity (in fact, this could *improve* clarity);
- in my opinion, there are too many figures – for example, it is probably not necessary to include both Figures 5 and 8.

4. Other points to consider

- The authors say that Figures 6/7 show predictive return levels. Could the authors explain in what sense these return levels are *predictive*? I am used to the concept of predictive return levels when working within the Bayesian framework, wherein the estimated return level takes into account uncertainty in parameter estimation and randomness in future observations; I’m not entirely sure what the authors mean by “predictive” in their applications, and I think their return levels are probably “estimative”.
- The authors find confidence intervals for return levels using

$$\hat{x}_{T(q)} \pm 1.96\hat{s}e_B,$$

where $\hat{s}e_B$ is the estimated bootstrap sample standard error. Studies have shown that a much better procedure for obtaining confidence intervals from a bootstrap sample is to use Efron’s *bias-corrected and accelerated* intervals (so called “BC_a intervals”), in terms of the coverage provided, especially when the statistic of interest is non-normal (Efron, 1987). Have the authors considered whether or not the coverage of their intervals is close to that intended? If not, they should check this and compare their results to those given by the BC_a intervals.

- The authors mention a 4-parameter beta- P distribution to include a location parameter. I think that including this in the current paper would add significant weight to this work, and might show the true potential of this model.

- In Section 6.1, the authors make several references to the wave height data, which is rather confusing, since the wave height data have their own section (Section 6.2).
- I presume there are no issues of seasonal variability or other forms of non–stationarity in the data used?

5. Minor points

- I think more insight into the wave height data should be given at the start of Section 6.2.
- Section 2 has only one subsection (2.1). Is there any need to label this as a subsection?
- Key words: I think “Extremes” should appear here; also, these words should be listed alphabetically.

References

Efron, B. (1987). Better Bootstrap Confidence Intervals. *J. Am. Stat. Ass.*, **82**(397), pp. 171—185.