Review of the paper “Influences on flood frequency distributions in Irish river catchments” by Ahilan et al.

General comments

This paper explores the relationship between various explanatory variables and the type of distribution of flood data. This topic is evidently of interest for HESS readership. However in the present state, the analyses made by the authors are not fully convincing and could be improved. Moreover, the organization of the paper could also be improved to clarify the insights that can be gained from the authors' analyses. More precisely, I have the following major comments:

- In the “result” section, the place given to previous work is far too important: the authors very shortly describe their own results, and lengthily discuss previous findings. This is very confusing for the reader, since it becomes difficult to know what are the new insights brought by this paper compared to previous work. I have to say that after reading this section, I had the feeling that most conclusions were motivated by results obtained in previous work or by expectations, rather than by the results presented in this particular study. Consequently, I would advise to split this “result” section in two parts: a “result” section solely focusing on the analyses carried out by the authors, and a “discussion” section linking those analyses with previous work, and proposing physical explanations to the obtained results.
- The description of methods could be very significantly shortened: the authors are presenting in full detail standard statistical techniques that could easily be found in textbooks.
- A discussion on the problem of model choice is missing. In particular, the authors seem to take for granted that the flood distribution is a GEV, and that the issue is to choose its type. The fact that type II EV distributions might be preferable is motivated by the fact that distinct hydrogeological or hydraulics processes are activated for small or large floods. But to me, this rather suggests that the flood distribution might be a mixture of several distributions, which can also explain the convexity of empirical quantile curves for some stations. Conversely, it is perfectly possible to observe type II or III EV distributions in the absence of any significant influence of hydrogeology / floodplains etc. I'm not suggesting including other distributions to this study, but at least some discussion would be of value.
- Similarly, a discussion on uncertainties is missing. The whole study is based on point-estimates of parameters, but the purpose of flood frequency analysis is not only to provide a “best estimate” of the quantile curve, but also to quantify related uncertainties. In particular, the uncertainty in the shape parameter is in general very large, so that it could correspond to any of the three EV types. In particular, when a type I distribution cannot be rejected, it doesn't mean that the distribution *is* of type I, and there is therefore a risk in using predictions based on a type I distribution just because one could not reject this hypothesis. Consequently, rather than just choosing one EV type, maybe one should rather quantify this uncertainty and provide predictions that account for it? These issues should be discussed in my opinion.

Specific comments

Would it be possible to recall, from the very beginning of the paper, the correspondence Type I – Gumbel, Type II – Frechet-like (or fat tail) and Type III – Weibull-like (or light tail). I'm personally always getting confused between the three types.

p. 3307 line 8-11: Please provide a reference for this statement. I have to say I'm not keen of such general statements, unless accompanied by more details (e.g. for which distribution? For which target level of precision/accuracy? Etc.).

page 3307, line 17-20: these points are not really convincing: All three types of GEV have closed-form expressions for the pdf, the cdf and useful moments – they just depend on an extra parameter for types II/III.

p. 3309 line 5-10: There is a kind of logical fallacy here: FAI are used to explain the type of GEV distribution, yet to compute a FAI, a flood quantile is required, which requires... choosing a type of GEV! Please comment on this and explain why this may or may not be problematic.

p. 3310 line 14-25: would it be possible to include a few illustrative hydrographs (with corresponding rainfall) from typical and contrasted catchments?

p. 3313 line 17-21: this sentence already appears earlier in the paper, please delete or reword.

Section 3.1. this can be shortened – such summary statistics will be well-known to the reader interested in this paper

p. 3315 line 4: the term “statistical power” is a bit ambiguous since it has a very precise meaning in the context of statistical testing (probability of correctly rejecting the null hypothesis), yet only approach 3 is a test. Please reword.

Section 3.2.1: this can be significantly shortened (virtually removed), all those formula & definitions are standard and can be found in textbooks.

P 3316 line 16-17: I would disagree, it might also be due to the fact that the sample size is too small to see a type II/III behavior.

p. 3317 line 20-23: a bit unclear (what is meant by “a distribution occupying a greater proportion of the measured data”?)

p. 3317 line 25 and 27: even in the absence of bias (in the statistical sense), there would still be sampling uncertainty and therefore scatter in the observed moment ratios. Please reword.

p. 3318 line 2: Wouldn’t it be more logical to use the same number of stations and the same sample sizes than in the analyzed dataset, given the objective behind this Monte-Carlo simulation?

p. 3318 line 7: it could also be from any other (non-GEV) distribution!

p. 3319 line 1-3: this statement is a bit too strong – a negative estimated value of k is not sufficient to exclude a type III distribution, due to possible estimation errors.

P 3319 line 8-16: can be shortened, even removed.

p. 3320 line 1-4: this sentence already appears earlier, it can be deleted.

p. 3320 line 16: How can you compute a BFI in an ungauged catchment? I guess this variable has been mapped or regionalized? Please specify.

P 3322 line 15-30: can be shortened, even removed.

p. 3324 line 4-7: this distinction between OPW and EPA gauges is not really interesting for an international readership.

p. 3324 line 8-10: these numbers should be related to the error level of the test – if all distributions were of type I, would it be surprising to observe 11 decisions “type II” out of 143 stations, given the error level? This point should be at least mentioned and discussed. Specific methods also exist to handle this issue, see e.g. “field significance evaluation” [*Livezey and Chen*, 1983; *Douglas et al.*, 2000] and “false detection rate” [*Benjamini and Hochberg*, 1995; *Wilks*, 2006; *Renard et al.*, 2008]. The authors might wish to implement such methods.

From p. 3324 line 11 to page 3326 line 5: can be significantly shortened, virtually removed, this is mostly textbook material.

p. 3326 lines 6-11: some words are a bit too strong in the context of statistical testing, e.g. “valid” (there will be wrong decisions due to the error level!), “underestimated/overestimated” (the truth is unknown!). Please reword.

p. 3326 line 19-24. I find this clustering quite subjective and not really convincing based on the sole information given in the maps – there are green dots spread out pretty much everywhere in the maps. Unless the authors use information/knowledge that do not appear on the maps, but this should be further explained in this case.

p. 3327 line 2: It's quite difficult to see any FAI details in figure 7c, I'm not sure this GIS-based representation is well suited to this variable. As a consequence, it is difficult for the reader to be convinced by the impact of FAI from the information provided by this sole figure.

p. 3327 line 3-14: should be moved to a discussion section – those are not results from this particular study.

p. 3327 lines 15-16: As previously, this clustering is not convincing: why are two nearby sites excluded from the cluster? How about the three sites on the east coast, where there is apparently no karst? This needs to be justified and discussed in far more depth.

From p. 3327 line 17 to p. 2238 line 6: as previously, this should be moved to a discussion section – those are not results from this particular study.

p. 3328 line 13: “meaningless” is a rather strong word: I don't think estimating a distribution is meaningless in karst regions, but for sure it might require more care! What is disputable is rather to blindly use a given distribution on the sole ground that it works quite well elsewhere!

p. 3329 line 2: maybe something more “extreme” than an annual rainfall might be a better explanatory variable?

p. 3329 line 6: How can the influence of rainfall frequency be deduced from figure 8? The lack of relationship with annual rainfall is not conclusive of the role of frequency. It might be more convincing to define a variable describing the rainfall frequency, and evaluate whether it is indeed linked with the shape parameter.

p. 3329 line 16: I'm sorry but I can't see this increase in Figure 8d.

p. 3329 line 19-22: those are others' results and should rather be placed in a discussion section.

From p. 3329 line 28 to page 3330 line 9: same comment

page 3330 line 14: this is a different issue – in a climate change context, what is questionable is essentially to assume that the flood distribution is stationary (irrespective of the chosen distribution!).

From p. 3330 line 15 to p. 3331 line 7: This discussion on CC does not really belong to a “result” section. Moreover, I think some statements are a bit speculative – for instance, p. 3331 line 1-5: what the authors describe here is a trend, and detecting a trend is not sufficient to demonstrate the impact of CC (see the distinction between *detection* and *attribution* made by the IPCC in the assessment reports).

p. 3332 line 5-8: this sounds a bit speculative to me.

The reference list could be broadened in my opinion: in particular, the topic of identifying the type of distribution of some extreme hydrologic data has a quite long history, which doesn't appear in this list. See e.g. El Adlouni at al. [2008] and references therein.

Tables A1: might possibly be removed.

Tables B1, B2 and B3: Those tables are not useful for an international readership, I doubt many readers will go through those numbers. I think they should be removed. If some of this information is important in the authors' opinion, it should rather be presented as a figure.

Figure 1: this representation is not adequate, the percentages could just directly be given in the text, or alternatively, the authors should rather show the whole distribution of record lengths.

Figure 2: This figure is incomplete and not very useful in the present state: please include a scale, a legend, increase text size, etc.

Fig 3: please increase font size. Moreover, why not using the catchment area in the x-axis, as in left column?

Fig 5: I'm not very familiar with moment diagrams, but I don't understand why the simulated EV1 data are not scattered around the green square representing the EV1 distribution. Please explain.

Fig 8: I think such figures are far easier to read than the previous GIS-based representation, although I understand the latter is of interest for a more detailed investigation (but a zoom is needed, which is not feasible in a article format!). Consequently, I'd encourage the authors to use this type of presentation for their results. In particular:

- Indices could be computed to summarize the variables presented in the GIS (e.g. percentage of lake area, relative FAI area, etc.) and shape parameters could be plotted against these indices.
- I think the authors should also use sample size as a covariate, just to check that it does not spuriously influence the shape parameter values, due to sampling variability.
- Two-sample tests could be used to assess whether the explanatory variables have distinct distributions depending on the EVII/EVIII decision (i.e. create 2 samples (EVII stations) vs. (EVIII

stations) and check whether e.g. Urban area have distinct distributions on these 2 samples)

- A more “continuous” version of the 2-sample test above would be to perform a multiple regression analysis based on the reduced variate Z in equation 8 to assess the impact of the various explanatory variables.

Technical corrections

p. 3318 line 20: “year” missing

p. 3320 line 14: what does FSU mean?

p. 3321 line 2: figure 3a,c,e

p. 2232 line 10-11: “concave upward” and “convex downward” sound like pleonasms, and are actually inverted! (concave means downward, convex upward).

Figure 6: Maybe a log-log scale would be better to zoom in the lower part of the curve? Moreover, the colors used for EV1 and EV3 are difficult to distinguish.

Benjamini, Y., and Y. Hochberg (1995), Controlling the false discovery rate - a practical and powerful approach to multiple testing, *J. R. Stat. Soc. Ser. B-Methodol.*, *57*(1), 289-300.
Douglas, E. M., R. M. Vogel, and C. N. Kroll (2000), Trends in floods and low flows in the United States: impact of spatial correlation, *J. Hydrol.*, *240*(1-2), 90-105.
El Adlouni, S., B. Bobee, and T. Ouarda (2008), On the tails of extreme event distributions in hydrology, *J. Hydrol.*, *355*(1-4), 16-33.
Livezey, R. E., and W. Y. Chen (1983), Statistical field significance and its determination by Monte Carlo techniques., *Monthly Weather Review*, *111*, 46-59.
Renard, B., et al. (2008), Regional methods for trend detection: assessing field significance and trend consistency, *Water Resources Research.*, *44*.
Wilks, D. S. (2006), On "field significance" and the false discovery rate, *Journal of Applied Meteorology and Climatology*, *45*(9), 1181-1189.