https://hess.copernicus.org/preprints/hess-2022-401/

Review comment posted 20 Feb 2023 by Kolbjorn Engeland, with responses from the authors.

The paper addresses the challenge of non-stationary flood frequency modelling and how to make such models useful for decision makers. In general, the paper is well written and could be published (provided it is different enough from Faulkner et al., 2020).

Thank you very much for taking the time to comment on our paper. We are very grateful for your insights.

The paper is rather short, and have a lot of questions after reading the paper. I therefore think there is room for several types of improvements. Below are some suggestions.

1: I think the numbering of sections might be improved. In particular, the heading on line 55 does not belong to the introduction section since the following paragraphs describe the methods applied in this paper. One solution is to create one section 'Methods' that contains lines 56 – 176.

Thank you for pointing this out. We will improve the numbering.

2: No results are shown for the GLO distribution, so it is not necessary to include it at all in the paper.

Agreed, and the first reviewer agrees too. We will remove it.

3: I am not completely convinced of the usefulness of Model 3 where de-trended physical covariates are included.

We agree that the justification for Model 3 needs to be strengthened. We see it more of a stepping stone, an aid to understanding, rather than an end in itself. We propose to add some text to make clearer what the value of Model 3 is, namely that it can help to disentangle the effect of any long-term trend from shorter-term cycles or fluctuations in the physical covariates. Comparing the various models helps us to assess if it is the specifics of the physical covariate or simply that it is may be approximately linear over time which is causing the physical covariate to appear statistically significant.

Model 3 turns out to be the best-fitting out of 1-4 at many gauges, and we could hypothesise (and will test this) that these tend to be gauges without significant long-term trends in peak flows. There may still be cyclical non-stationarity at some of these gauges that is being described by cycles in the covariates. However, as we have written, it would seem odd if year-to-year variations were captured by a detrended physical covariate without the longer-term changes in that covariate also being important, so we will add Model 5 into the comparison of Models 1-4.

Another possibility at some sites is that there is little trend present in the covariate and so Model 3 is similar to Model 5. As you have suggested, we will examine the extent of trends in the covariates.

We address some of your follow-up parts to this comment briefly below.

Firstly, I cannot see how the covariates are detrended: for which period was the trend calculated, and did you use a linear trend?

We will add to the paper: it was a linear trend over the whole period of record.

Secondly, how many of the physical covariates had a significant (and substantial trend) where model 3 and 5 were different ?

This is something that we could check and add to the paper.

Thirdly, using a de-trended covariate indicate that there is an interaction between time and the physical covariate or potentially interactions with other physical covariate that you have not included.

Agreed (we have said something similar in the paper) – and this seems a reason to avoid preferring model (3), even if it does show some good fits.

Finally, using a detrended covariate will make it difficult to apply the model for a climate in the coming decades.

This is correct. However we do not necessarily recommend using any of these models to estimate future flood frequency, for reasons explained in the discussion section.

This choice of using de-trended covariate need a better explanation and discussion. I think it might be helpful to add Model 6 and the methods described in Appendix B. An alternative solution to using independent covariates is to use regularization methods similar to Lasso regression.

We would have liked to add Model 6 but reluctantly decided we would need to leave that for future work, having already spent far longer than we planned in writing this paper!

Thank you for the Lasso suggestion, which could be a promising way of fitting more complex models without having to try large numbers of combinations of covariates. We will mention regularisation methods in the discussion. We anticipate that Lasso could still have problems picking between casual and non-casual covariates when they are all strongly co-linear.

4: You included trends in both location and scale parameters of the GEV distribution. Did you systematically evaluate all combination of trends in scale and location parameters?

Yes, we included all combinations of trends in both parameters.

Could different physical covariates be selected for the location and scale parameters?

No, we only included one physical covariate at a time. We do not think it would be sensible to have different covariates in the scale that are not in the location, given the theoretical way the these two parameters link to the properties of the core underlying distribution of the hourly river flow data. We will make this clear.

Are the results in Figure 2, Figure 3 and Figure 4 based on the best fitting covariate for the location or scale parameters? Some more details are needed.

The goodness of fit is judged using the BIC which measures the model as a whole rather than the individual parameters.

5: What are the signs of the detected trends (or regression coefficients)?

We agree this is a question that could interest readers and will add this information.

6: I think more results similar to those shown in Figure 5 could be produced. It could be good to show one more plot where floods for the non-stationary model is smaller than the floods for the stationary model. It could also be interesting to see this plot for a model where time is included as a covariate and a model where time is excluded as a covariate.

We can add and discuss some more example plots.

7: Could it be useful to see the results in Figure 6 on a map in order to highlight where the ratio is smaller and larger than 1? Have all records the same year as the last year in the records? If not, how much might the results be influenced by the end year of the record?

This type of information is given in Faulkner et al. (2020) and we will point readers to it.

8: Is it possible to detect more results from the data used to produce Figure 6? In particular, I would like to know for which models or catchments the ratios are far from one. Is it in catchments where the selected models include time as a covariate or are the specific catchment properties or geographical locations that might explain the differences in ratios? Another possibility is that the estimate of the shape parameter are different in the stationary and non-stationary models.

We can add some discussion of this to the paper.

The maximum likelihood estimator is know to results in shape parameter estimates that are not robust, in particular for short record lengths. Could the results you get depend on record length? A penalized maximum likelihood estimator is often recommended for the GEV distribution. Alternatively, a Bayesian approach with a prior on the shape parameter could have been used. I think the robustness of the ML estimator should be discussed.

This is a good point. We did try a penalised ML estimator and found that this approach, at least as implemented in the extRemes package, was unexpectedly counterproductive, leading to some large increases in the shape parameter. We will comment on this and possible future steps to improvement.

9: Uncertainties in the estimates are not discussed. I think this is important since the inflation in uncertainty is a drawback of non-stationary modelling. Figure 5 shows that he uncertainty in the non-stationary model is high, and even if the difference in integrated flow estimate is substantial, the difference is not necessarily significant. Figure 6 shows that a large par of the ratios are close to 1. Is it possible to use the results from all 375 flow records to summarize in how many cases the non-stationary integrated flow estimates are outside the confidence intervals for the stationary model?

We agree that this inflation in uncertainty is important to acknowledge. Thank you for the suggestion for analysis of confidence intervals, which we will look into.

10: The implications for practical applications could be discussed more. Would you recommend to always use the non-stationary model? What are the recommendations if the non-stationary model results ins smaller design floods or in substantially higher design floods (up to 5 times higher than for a stationary model.) ?

We will add some discussion of this, which was also requested by the first reviewer.