Dear Reviewer,

We appreciate the time and effort that you have dedicated to providing your valuable and constructive feedback on our manuscript. We have considered all comments and outline our response and the changes we propose below. Our response to reviewer comments (black font) are in red font.

General comments and recommendation

The manuscript by Meresa et al. presents an interesting illustration of the uncertainties present and arising in modelling flood frequency and flood magnitude under different climate change projections for four catchments in Ireland.

Thank you for this positive comment.

While it thoroughly covers a wide range of sources of uncertainty along the modelling chain and their interactions by applying an ANOVA, besides the reduced number of catchments analysed, a main drawback of the setup of the study is the lack of a better representation of the uncertainty stemming from the hydrological modelling. The many different preprocessing methods evaluated and also the multiple extreme value distributions are a strength of this study –which is otherwise not applying nor showing something new-, however the robustness of the finding that hydrological model parameter uncertainty is the least important component is very weak. Furthermore I am surprised the authors (apparently?) didn’t expect the results to be different across catchments, as this is the case in some of the studies cited, and not only.

Thank you for this critique. We have responded to the other reviewers about the catchment sample and the aims of our study. We agree that while thorough, as you identify, a limitation is the lack of consideration of hydrological model structure. Our study is not exhaustive and we will further highlight and discuss the limitations of the study in our revisions. For clarity, we are not surprised that different uncertainties vary across catchments. We do however highlight this finding that the dominant sources of uncertainty can be so different even for a small country like Ireland is novel. The default position in the literature is that climate models are the dominant source. This is even the case in the important paper by Addor et al. (2014) that you direct us to. They highlight that “*While there seems to be a general agreement on the dominant contribution of climate models to the uncertainty in discharge projections, different conclusions were drawn about the contribution of the hydrological models, for example*”. Our findings are novel in that they show that uncertainty from bias correction and even extreme value distributions can outweigh the uncertainty in flood quantile projections from a large ensemble of climate models for some catchments. We will be sure to clarify this in our revisions.

We will also integrate further work on parsing the uncertainty in hydrological model parameters, showing how this source of uncertainty is variable, depending on which climate model is selected and the signal of change for different future time periods. For instance initial results are shown in the figure below, whereby the range of change in annual maximum flow from hydrological model parameter uncertainty for 12 CMIP6 climate models are shown for the 2020s, 50s and 80s. This clearly shows that the parameter uncertainty varies depending on the catchment, climate model used and time period considered.

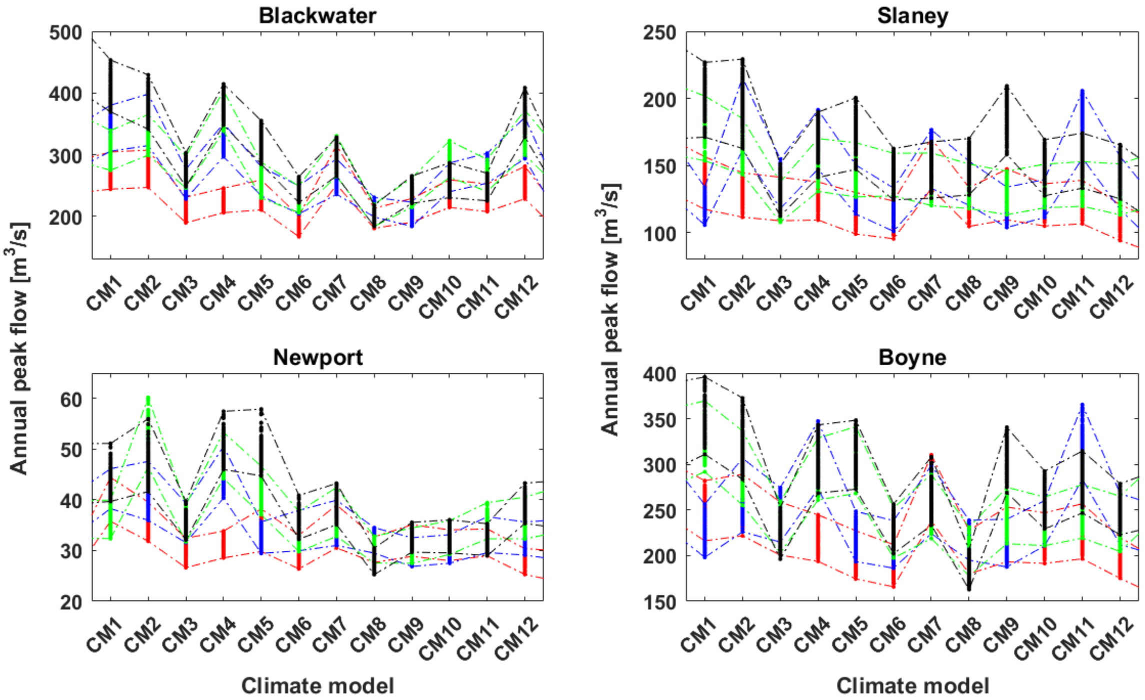


Figure R1. Hydrological parameter uncertainty using the selected behavioural parameter sets for the reference period (1976-2005), clim1 (2020s), clim2 (2050s) and clim3 (2080s). The black vertical line and broken horizontal dotted line represent the clim3, the blue vertical line and broken horizontal dot represent the clim2, the green vertical line and broken horizontal dot represent the clim1, and the red vertical line and broken horizontal dot represents the reference period. Each vertical line represents the hydrological parameter uncertainty with its upper and lower values (horizontal). X-axis presents the list of 12 climate models.

You are running 30’000 simulations and applying GLUE, defining as behavioural parameter sets whose simulation show an NSE >0.5. As the only other thing we know that by doing this you, retain 300 parameter sets for you further analysis. I was wondering about the identifiability of parameters and how do their distributions look like in respect to their likelihood. I think these are some important pieces of information you might add to the Supplement material so that the readers are sure how do these calibrated parameters look like, and/or if they are clustering somewhere.

Thank you for the concern. We will add a supplementary figure to the revised manuscript as requested.

The manuscript is well structured, the methods are generally described in a comprehensible way or supported by relevant sources and/or equations. Even though the discussion provides good points; some important considerations should have been stated already before in the text. While the authors are quite keen on describing findings for each catchment, they don’t really try to explain and justify some of the differences observed, what is an important shortcoming.

This is a fair point and we will do more to outline differences and reasoning as to why in our revisions.

The manuscript generally features high-quality and interesting figures. However the authors might think about possibly reducing the number of figures (are they all relevant, or could be part of the supplement?), and also improving the readability by changing of some colors used.

Reviewer 2 also recommended this. We will remove Fig 6 and associated text and move Fig 8 to supplementary material.

I found some inconsistencies in the equations and an error in a figure. I am reporting all those I found in the technical corrections.

Thank you for such rigor. We will address all of these issues.

Specific comments: specific and technical comments here following

- Introduction:

This is HESS and not NHESS, but still I think using the term flood risk might be misleading for some readers. I think using flood frequency and flood magnitude would be more appropriate. Avoid also to speak about extremes, as the return periods you are looking at here are those usually considered in many countries as the limiting design floods for inhabitated areas.

Common point across all reviews, we will avoid use of ‘risk’. While we agree that they are design floods, we would argue that they are still extremes.

In general there might be more literature out there to cite, but in particular here I miss Addor et al.2014, who did a similar evaluation for several Swiss river catchments, and have actually some common findings.

Thanks for this suggestion, we will integrate all suggested papers. This paper in particular is a very useful suggestion. Addor et al (2014) examined mean flows and highlight the dominance of climate models in catchments, so there are important differences to discuss also, which we will do in revisions.

- Modelling and numerical experiments:

I am not sure numerical experiments is the correct name for what you did.

Fair point, we will change to study design.

In your paper you make very strong statements about the uncertainty related to the hydrological model parameters, but this might be related to the hydrological model used itself – what you also say later in the discussion- however you might already state this here. I would also expect you to actually better justify your choice: why using a single conceptual hydrological model with only 4 parameters? Why completely leaving out a more physically based model (where the assumption of stationary parameters might be relaxed) ?

Thanks you for this point. Yes parameter uncertainty may be related to model structure. Our emphasis here was on integration of bias correction and extreme value distributions, together with the new CMIP6 ensemble. We will highlight that our modelling chain is not complete and give further justification for selection of the hydrological model, taking into account studies that have evaluated model structure uncertainty in assessing future high flows. We can add this point that for future researches, considering hydrological models with different degree of complexity, ranging from simple conceptual models to more sophisticated physically-based models can be an interesting subject for researchers.

P4-L103: in Table 1 with elevation do you mean mean elevation? What is exactly 95% of precipitation value?

Yes mean elevation and we will clarify that we refer to the range between the upper and lower ranges of the precipitation distribution. It refers to range of precipitation (mm) between the respective quantiles of the observed data.

Why do you use only one goodness-of-fit measure –in principle-for selecting the behavioural parameters? Or how do you exactly take PCI into account? This is not clear to me according to your text.

We restrict our selection to NSE given our focus on high flows. That said the selection of objective function is another aspect of the uncertainty cascade that may be influential. We will highlight this in our revisions. We will remove the PCI analysis.

P8-L230: I think here there is a mistake, GEV has three parameters, whereas Log-Normal and Log-Logistic 2? (again pay attention to the consistency of lower/upper cases and the separating dash between text and Equations)

Equation 17: you might write (x-μ)/σ instead of z (for the sake of consistency with the other equations).

P8-L235 k is the shape parameter and not the location parameter, and please define σ and μ too (GEV's scale and location parameters respectively).

Thank you. We will correct each of these.

- Results:

P10 Wouldn't make more sense to have Fig.4 shown and commented before Fig. 3?

It is possible, however we prefer to leave as is, Fig 3 is a general result and Fig 4 gets into more detail. We feel this is better than vice versa.

P10 Fig.6 is a figure you might consider to put in the Supplement

We will remove this figure

P10-L290-295: you could spend a few more words on the deficiency of the model in modelling late summer-autumn, and also explain how the NSE values shown in Fig.7 have been calculated (with daily or monthly data?) It would be also important to spend some few more words also on the performance of the model in the different catchments.

We can of course do this and tease out how NSE objective function may be an issue for later summer/autumn. We note however that given the timing of floods in Ireland this aspect is less problematic. We will discuss.

Fig.7: you are showing the 95% interval, but this is not in line with your text P10-L293, or I am missing something?

Text should read 95%, we will ensure consistency.

P11 first paragraph: I am not sure how “useful” this is. If you want to keep it (and Fig. 8 too), I would suggest you elaborate more on the trend and patterns you are mentioning, on the visible temporary effects, and rather give flow increases as percentages rather than absolute values.

We will refine and move Fig 8 to supplementary material. We agree this figure is of secondary importance.

P11-L3095-306 isn’t there a clear increase with all downscaling methods? Well, RAW data too.

Yes, there is a clear increasing of peak flow in the 2020’s, 2050’s and 2080’s. however, the magnitude is not uniform across the period, bias correction methods, and catchments. Also, it is not a smooth increase from reference period to far future period (2080’s).

P11-L317: you are not showing the reference (i.e. observations) in Fig. 10, so what do you mean by saying ..the smallest changes in flood quantiles..?

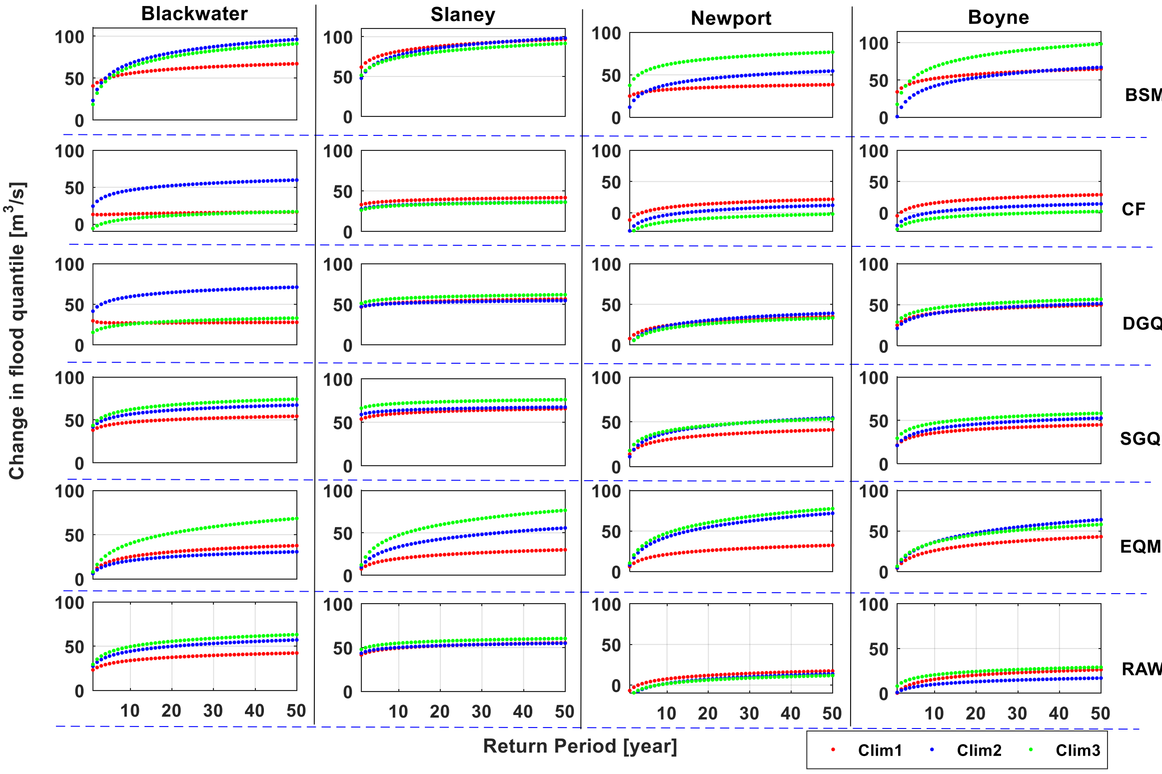
We are comparing model chains here and will clarify this point.

P11-L318: I think here it should be LogN and not LogL?

Good spot, will correct.

Is the y-axes in Fig.11 correct for the Slaney catchments downscaled using EQM? There is a massive difference as compared to the application of the other downscaling methods, and you don’t really mention it in the text…

Thank you for this spot. We will check and revise as necessary.



Revised Figure 11. Percent changes in extreme flow quantiles using ensemble of three distribution types in each catchment for the 2020s (clim1), 2050s (clim2) and 2080s (clim3). Simulated changes are derived using raw (bottom row) and bias corrected (the first four rows) simulations. each line is the mean of changes from 12 climate models and three flood frequency models in the 2020s (red), 2050s (blue) and 2080s (green) period.

P11-L327: aren't the smallest changes for Slaney using BSM ?

It depends on the return period. In the text, we were referencing changes at 20-yrs RT and 100-yrs RT. Nevertheless, you are right if we chose 50-yrs RT. We will clarify the importance of flood magnitude.

P11-L327: aren't the smallest changes for Newport using CF and BSM ?

Again, it depends, see above point and we will clarify.

P12-L332-333: why is that? Is this not contradicting results shown in Fig. 5?

We don't think so but will look further. Fig 5 refers to max annual precipitation. Other aspects are important to changes in flooding and we will expand on this in the revisions.

Figure 13: First of all, some colors are too similar (e.g. BC and BC*DM look almost the same to me), second, it is not so easy to compare the different return periods by eye, as the circles have different diameters, what can be deceptive (e.g. the percentage seems to increase with the return period). It might be helpful to add the actual percentages and write somewhere the return period too. If possible removing the white outline of the percentages might also improve the figure.

Thank you for the comment we will improve presentation in the revised manuscript.

- Discussion:

P 13-L391: Both Bastola et al. a&b?

Yes, will update

P14-L399-400: Across all catchments the uncertainty in future hydrological model parameters .. is wrong. Please correct resp. reformulate this sentence.

Will do

P15-L453: you might want to add a comment on the influence resp. limitation of assuming flood processes to remain stationary within the 30-year windows on your extreme value distribution fits, and if applying instationary fitting would have been a better option.

Thank you, this is an important point that we will discuss further, in addition to extrapolating to 100yr flood.

I think authors really need to be harder on themselves for limiting the study to a single specific hydrological model, and elaborate more on what they would expect to be different by applying different model structures.

We agree and will significantly strengthen this aspect.

- Conclusion:

An important source of uncertainty in any hydrological setup are the discharge data themselves, which are implicitly assumed to be true resp. correct. There is an interesting study by Westerberg et al.2020 on this topic, it might be added as a source we should start considering too when performing sensitivity and uncertainty propagation studies, as an outlook for future work?

Excellent point, we will integrate this into the discussion also. Thanks!

Technical corrections

We will rectify each of the technical corrections. Thanks for being so thorough.

P5-L138: for the sake of consistency add (EQM)

P7 Equation 10: remove the square

P7 Equation 13: check the consistency of lower and upper cases

P8-L213: remove a bracket in the middle in (Equation(14)

P8: Equation 14 vs. text=> check the consistency NQi,p vs. NQin,p

P8 Equation 17 remove ^ in the equation

Fig. 9: the SSPs have wrong numbers, SSP2 should be SSP3 and ssp3 should be SSP5

References

Addor, N., O. Rossler, N. K € oplin, M. € Huss, R. Weingartner, and J. Seibert (2014), Robust changes and sources of uncertainty in the projected hydrological regimes of Swiss catchments, Water Resour. Res., 50, 7541–7562, doi:10.1002/ 2014WR015549.