Reviewer response to manuscript nhess-2023-22 'Return levels of extreme European windstorms, their dependency on the NAO, and potential future risks' by Matthew D. K. Priestley, David B. Stephenson, Adam A. Scaife, Daniel Bannister, Christopher J. T. Allen, and David Wilkie

Reviewer 2

Thank you for your considered responses to my comments, and for revising the paper in line with these. I am largely happy with the revisions, but still have a couple of minor points that if addressed would help to clarify the approaches taken.
We thank the reviewer for their comments and further consideration of our manuscript. Below we detail our response to each of the reviewers points independently.

These are:
- I think the motivation/question you are trying to address in Section 3.3 could be better stated. I still find the relatively short 10 year validation period approach a bit confusing. You mentioned in your response that this is related to the length of a reinsurance contract (i.e. I assume you mean you want to understand how many years long a catalogue needs to be to minimise the MSE when estimating the 1 in 200 year return level for the next 10 years, as this is the typical length of the future contract and hence situation in the reinsurance sector), but you don't mention this as motivation in Section 3.3. I think this sounds like an interesting motivation and context and stating this in the paper would help to set the context of this choice of method. If this is indeed the aim here, please could this be added as explanation.
We agree with the reviewer and have added further justification at the start of section 3.3. We now state '*One factor that can contribute to the uncertainty in the estimation of the 200-year return level is the length of the historical catalogue. "This is especially important for re-insurers and their need to understand risk in the next 10 years, as this is a time horizon for business planning, and 10 years is also the typical maximum service life of a catastrophe model. The varying length of catalogue has a substantial impact on the average footprint (Fig. 1b-d) and when applied to the statistical model at a return period of 200 years, these differences are likely to be amplified. Therefore, the understanding of future risk is likely to differ with these different length catalogues.'*

- Related to this, it is interesting that your conclusions did change very slightly when using the 10 year validation period (5-15 to 10-15 years required to minimise the MSE). I wonder if the results vary if a 15, 20 etc. year validation period is used. I understand that this is then a trade off re. how many years are left for the training data in the cross validation, but if this sensitivity study has been carried out maybe the author could comment (no need to show the results) e.g., does the 10-15 year conclusion remain for other validation period lengths?
We have performed further tests using validation periods of 15 and 20 years (Fig. R1 and R2). As can be seen in Figs R1 and R2 the same pattern of MSE curve is present and the same conclusions can be drawn. The MSE reduces rapidly beyond a catalogue length of 1-5 years and then is more stable after 10-15 years. The values of MSE are different as with a longer validation period there will be less variability in the 200-year return level. However,

validation lengths longer that 10 years are not used in the re-insurance industry and therefore we elect to retain the original figure and discussion in the manuscript. We expand the text at the end of section 3.3 to state that '*Validation periods of different lengths (i.e. 5, 20 years) were also tested with conclusions being insensitive to this change (not shown).*

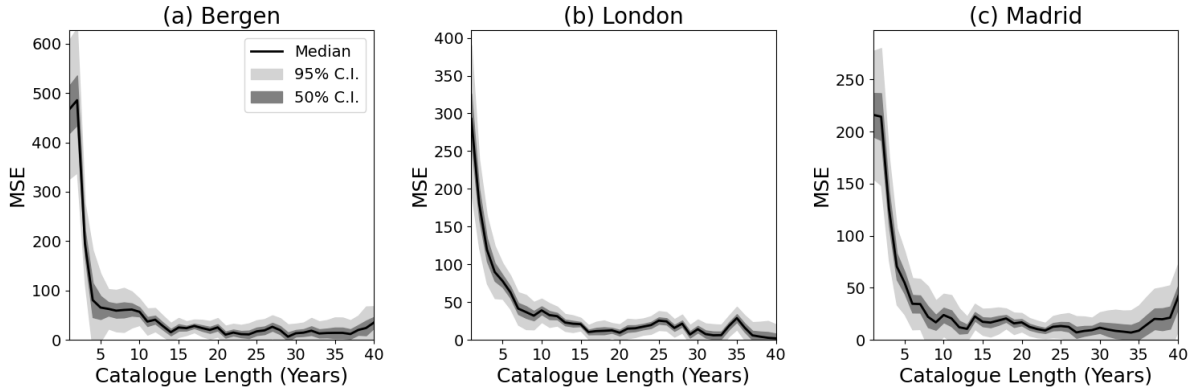

Fig R1. *Mean square error of the 200-year return level estimation of different length historical catalogues against a subsequent 15-year period from the WISC catalogue for (a) Bergen, (b) London, and (c) Madrid. Solid black line shows the median mean squared error when using all possible periods. The dark and light gray areas represent the 50% and 95% confidence interval on the standard error respectively.*

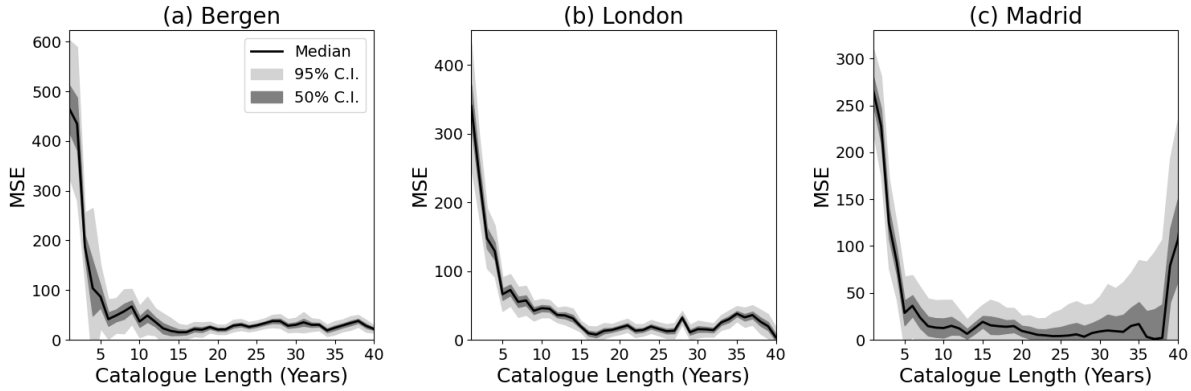

Fig R2. *As Fig. R1 but using a 20-year validation period*

- I like that you have set out a few distinct research questions in the introduction and I can see how the various sections can be related back to these questions, but I think this could be made clearer. I think the manuscript would be strengthened by explicitly referring back to these questions in each section of the paper, and then again in the conclusion to explicitly answer them one by one.

We thank the reviewer for their comment. We now refer back to the questions in each section of the paper. In the conclusions we state *'We raised several questions in section 1, and the key conclusions that answer these are as follows:'*

- Throughout the paper, where there are confidence intervals in plots, please can you add a description of how they are estimated/produced? E.g. in Fig 7 are these made up of all of the combinations of years with a given catalogue length?

We have updated the discussion of how confidence intervals are calculated in all of the relevant figures.

- In line 120 I think you are missing a conditional on u in the second Pr(Y>y|S)
We thank the reviewer for their comment. However, the expression for Pr(Y>y|S) is correct as written as it can be logically deduced from Eqns (2) and (3) and the definition p(u)=Pr(Y>u|S).