Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC2
https://doi.org/10.5194/hess-2021-597-RC2, 2022

Comment on hess-2021-597

Anonymous Referee #2

Referee comment on "Use of expert elicitation to assign weights to climate and hydrological models in climate impact studies" by Eva Sebok et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-597-RC2, 2022

Review for Hydrology and Earth System Sciences manuscript ID: hess-2021-597

Title: Use of expert elicitation to assign weights to climate and hydrological models in climate impact studies

Summary:

The paper uses five case study locations in Europe to test an expert elicitation approach to weighting climate and hydrological models. The study uses a structured expert elicitation approach, involving three stages of individual elicitations and consensus building, supported by some initial training material. The study finds that hydrologists are more willing than climate modellers to articulate model weights, with climate modellers preferring model democracy in the absence of further investigation. The shift in approach from in-person to online, due to the covid-19 pandemic, affected the approach followed away from typical elicitation processes.

Overall, I found the paper interesting, well written and clearly structured. The previous review comments seemed to have missed the point that this paper is focused on testing a methodology – i.e., expert elicitation to support model weighting – rather than providing robust scenarios of future climate impacts for the case study locations. As such, I do not share their concerns. Moreover, precisely because "all models are wrong, some are useful", I think the community should welcome efforts to more rigorously include expert judgement in providing actionable information, as relying purely on outputs from models (which we know to be flawed) risks over-confidence in uncertainty estimates.

That being said, the study isn't as conclusive as I'd hoped. The finding that climate

scientists continued to support model democracy – whilst interesting in the context explored – is not surprising. I suspect the finding would have been different if those involved had seen more model validation results – i.e. ability of the simulations to capture observed atmospheric circulation and trends, relevant to precipitation in the locations studied. It is also a shame that the climate modellers and hydrological modellers weren't part of the same expert group as originally envisaged, as this may have yielded some more nuanced views and outcomes for model weighting.

Overall, I think the paper is worthy of publication. It will help advance the use of expert elicitation methods in the climate and hydrology community. I hope my comments and suggestions below help in improving the paper prior to publication.

Specific comments:

Line 94: "with a few exceptions (Mearns et al., 2017)" – add to this the recent study by Grainger et al. 2022 – see references. It would be interesting for the authors to comment on this study and how the methods followed compare, noting that there are very few studies in this space.

Line 102: Also worth citing McSweeney et al. 2015 – see references. This highly cited paper demonstrates a method for first excluding implausible models following model evaluation, and then spanning the uncertainty range of the remaining plausible models.

Line 150: Table 2 shows 7 GCM-RCM model combinations, but the final combination includes two realisations meaning there are 8 simulations considered. Why is this not discussed in the paper, and why were two realisations selected for this model combination?

Line 153: Typo "info3rmation"

Line 162: Typo / technical error in referencing Table 3 in the pdf – check.

Line 177: It is mentioned that the move to virtual elicitation gave an opportunity to explore how this worked. However, the paper doesn't provide much analysis here. I would encourage a short paragraph in the discussion section to reflect more on the pros and cons of this change in approach.

Line 187: How were the 18 experts determined by the "partner institutes of the research

project”? Were there any explicit or implicit considerations – e.g., PhD in a relevant topic, papers published, involvement in CORDEX?

Line 192: It would help demystify things for the reader if you briefly explain why one expert decided to leave the study.

Line 432: I agree obtaining results from models requires time, but the elicitation approach followed is also very time consuming and incurs a cost. I’m not sure saving time is a strong justification for following an elicitation approach.

Line 476: “Climate models often stem from short-term forecast models”. Taken over many decades of model development this is true. However, this is a bit misleading as CMIP5 climate models are quite different from operational numerical weather prediction models. Suggest clarifying what is meant here – yes, climate change is more of a boundary value problem but scientists don’t simply add on elements to a NWP model to simulate future climate – there is a quite a lot more involved.

Line 522: Sentence ending “…are without doubt inappropriate”. This phrasing is too strong given the evidence. Had the climate modellers been provided with compelling evaluation information, I’m sure they would have been open to excluding models. Suggest deleting “without doubt” and rephrasing.

Further thoughts for the discussion section:

- It would be useful to comment further on the uncertainty cascade, referenced in the introduction section. In particular, does having expert opinion included in the articulation of uncertainties add yet another layer of cascading uncertainties? Or does it rather try to address and reduce the cascading uncertainties? It isn’t obvious to me.
- In general, climate models (GCMs and RCMs) are more complicated and have higher dimensionality than hydrological models. Could this be a reason why climate modellers prefer model democracy, especially if they aren’t entirely familiar with all aspects of the models?
- (with particular reference to lines 444 to 448) Another reason why experts won’t assign different weights to the climate models may be because they are all from the same generation – i.e., all RCMs downscaled with CMIP5 models. Might the result be different if comparing CMIP3 vs CMIP6 models for example?
- Sample size is an issue for this study. With only 6 experts in each group, any result cannot be considered robust - i.e., the finding that 6 hydrological modellers were more willing to assign weights compared to 6 climate modellers, is not a robust finding. It would be good to comment on sample size limitations.
- At the end of the conclusions section, you comment on the impact of covid-19 moving to virtual engagement. I suggest moving some of this to the discussion section and elaborating more on the methodological implications and insights that may be relevant

to other studies in the future.

All the best in revising the paper and I look forward to seeing the published article.

References

Grainger, S., Dessai, S., Daron, J., Taylor, A., & Siu, Y. L. (2022). Using expert elicitation to strengthen future regional climate information for climate services. Climate Services, 26, 100278.

McSweeney, C. F., Jones, R. G., Lee, R. W., & Rowell, D. P. (2015). Selecting CMIP5 GCMs for downscaling over multiple r