Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2020-272-RC1, 2020

Natural Hazards and Earth System Sciences
Open Access
Discussions

Interactive comment on "Uncertainty in flood frequency analysis of hydrodynamic model simulations" *by* Xudong Zhou et al.

Anonymous Referee #1

Received and published: 3 September 2020

General comments:

Zhou et al. present a novel and interesting analysis of uncertainties inherent to global hydrodynamic models. Many studies over the past ∼5 years have posited that global flood models produce widely divergent results, yet each of them fails to explain or evidence the cause of this divergence. The authors begin to address some of these previous gaps in the literature and have produced some (though limited) conclusions of genuine interest. However, there are fundamental problems with the methods, analysis, and wider interpretation/discussion of their conclusions that require addressing before publication is to be considered in NHESS. Do not be put off by the volume of comments. I do think the work has great value and is something the field desperately requires. To be the truly impactful contribution this work needs to be, I offer the following (I hope,

Printer-friendly version

Discussion paper

constructive) comments. These broadly relate to:

- Model sensitivity vs. uncertainty analysis
- A single model with often extremely limited geographic scope
- Considered uncertainties in the context of 'unconsidered' uncertainties
- The relevance of hydrologic variable choice
- The contextual relevance of distribution goodness-of-fit
- The need for some benchmark data

****

Specific comments:

The introduction is mostly good, but could contain a richer discussion on why unstitching the uncertainties of global flood models is needed and how past studies essentially have failed to do this. There are also many sweeping or incorrect generalisations, that may simply be a function of imprecise English (for which I am sympathetic and understanding). This is the case throughout the manuscript (e.g. the sentence in line 26-27 p. 2 makes little sense as the same could be said for RFFA; line 3 p. 23 says the studies assess flood risk, when it does not).

The overarching problem in the introduction is that it makes the reader think some quantification of the uncertainties – validation, against observations – is carried out by the authors, and this is not the case. Truly, this analysis is a sensitivity analysis of 1 model. While I appreciate this illustrates the 'uncertainty one should have about conclusions drawn using this model', really it just tells us what the model is sensitive to and by how much. The paper as a whole needs more of a framing as a sensitivity analysis rather than formal uncertainty quantification.

The methodology, if framed as a sensitivity analysis of CaMa-Flood, appears thorough and fit-for-purpose. In general, justification of the use of a single model and subsequent analysis in specific regions (even specific grid cells) is needed. How universal are the conclusions in light of these methodological choices? Of course, these are only uncertainties related to the subjective choice of model tests. It is worth stressing that the reported uncertainties make the (of course, incorrect) assumption that terrain, channel bathymetry, human influence, and model parameterisation are certain.

It is not clear why river depth and water storage are chosen as variables of interest – this needs further explanation, as I can not yet see the significance of doing so. Common application of FFA is to discharge, yet this is not done here. Further discussion of the AIC is needed: what constitutes a 'good' result in this context is not specified. Equally, what is the relevance of this metric in terms of model uncertainty? What is the relevance of a good fitting distribution in the context of the uncertainty in the absolute values themselves? Are the authors saying that a variable with a poor AIC contains no relevant information for FFA? Really, it just shows a suitable distribution has not yet been found. I think section 3.1 fails to recognise the variable of choice is arbitrary and depends on the model used and the question asked. We all know that a 100-year rainfall $\neq$ 100-year streamflow $\neq$ 100-year economic loss. So frame this strand of analysis in the context of why the variable you choose matters and why this is interesting.

I can not see evidence that WAK is the best distribution because of it having 5 parameters. As the authors mentioned, it may just be overfit to the simulation record. The reality is we have no idea which distribution we should extrapolate with – and this is not something the AIC can test.

The section 3.3 analysis of runoff is interesting, but the results are stated in such a way that the authors expected the analysis to produce a 'preferable' runoff product. No feature of the analysis performed could identify such a thing. It is not clear why being a runoff product in 'the middle' is the best place to be: it could be that the lowest estimate types are actually best! It is a problem throughout the paper, where a suitable performance benchmark has not been found. Ensure the results are framed and reported as sensitivities, not as good/bad.

I like the analysis in section 4.1, but I'm not sure why this could not be done for every global grid cell – with normalised results – and presented in the same way. How representative is this grid cell? It may also be interesting here to compare the AIC results to Figure 7c: exploring some of my above comments on why AIC matters more quantitatively (i.e., does high/low AIC [thus, how good the distribution fit is] matter in the context of inundation?)

As for the rest of section 4, the analysis is good. While I appreciate visualising the globe at this scale is difficult, a lot of the calculations could still easily be done globally. It leaves the reader wondering whether different climatologies and geomorphic settings might have different conclusions. Deltas are difficult to model – particularly for models with poor/no representation of coastal boundaries – and so may have distinct features of uncertainty to other areas. I see no reason for the authors not to report findings elsewhere.

I do not see any value to section 4.5. I have little doubt the CaMa-Flood 100-year map is more accurate than the GAR and JRC maps: it is an uninformative comparison, and certainly not "validation". You only have to look at the stripes of JRC's map in Figure 12b to know that that is not a model you should aspire to resemble! I appreciate finding suitable validation data is difficult, but it is difficult to understand the relevance of the authors' conclusions without some. Perhaps running this analysis in the US or western Europe where high-quality models exist and comparing to those would be a good idea.

Section 5 is strong, but will benefit from drawing on some of the above points. Generally, the manuscript is quite long, and so the impact from section 5 is dampened by unnecessarily long analysis in 4.2-4.4. Throughout the paper, I would ensure each test is a worthwhile inclusion for the conclusions drawn. At present, there are many analyses which offer little additional information which I would consider cutting.

Figures are generally good quality, but most need to be larger. I would change the colour scheme of some figures (e.g. 4-6) where colour scales are used for variables which are not ordinal (no reason to go from blue to red, when the distributions are in no order).

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2020-272, 2020.