Natural Hazards and Earth System Sciences Discussions

Nat. Hazards Earth Syst. Sci. Discuss., referee comment RC1
https://doi.org/10.5194/nhess-2022-188-RC1, 2022

Comment on nhess-2022-188

Seth Bryant (Referee)

Referee comment on "Assessing the spatial spread–skill of ensemble flood maps with remote-sensing observations" by Helen Hooker et al., Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2022-188-RC1, 2022

The authors present a method for evaluating the accuracy of ensemble flood forecasts which may advance our ability to predict inundation more accurately. This provides a nice advancement from the recently published Hooker et al., (2022) (10.1016/j.jhydrol.2022.128170). However, the manuscript is difficult to read, impeding a full evaluation of the work. I would be grateful if the authors would consider the following:

Two pages are copied verbatim from Hooker et al., (2022). Instead, these should be summarized, and the reader directed to this other publication.

There are numerous grammatical issues, redundant sentences/phrases, imprecise/inaccurate vocabulary, and a confusing overall sequence/structure which make the manuscript difficult to follow. The authors should consider the perspective of the reader, striving to be as concise and logical as possible.

While I'm unfamiliar with the details of flood forecasts, I can imagine and appreciate the motivation for such a metric. However, I'm skeptical the method proposed is appropriate for application against a simulation-library like Flood Foresight. For example, if each inundation raster within the library is monotonically nested (i.e., cells become progressively more flooded), a neighborhood approach seems unnecessary. More information on the Flood Foresight simulation implemented in this study is needed to evaluate this properly.

Similarly, additional details of the application of the permanent water body layer (in both

the SAR-derived layer and the Flood Foresight layer) are necessary to evaluate the utility of the proposed method to the case study. For example, if the same source layer pre-filter is implemented in both the 'observed' and the 'simulation' data, rewarding the simulation for accuracy in these cells seems inappropriate.

To demonstrate the utility of the metric, the authors should consider comparing against some alternative. When is the proposed two-phased sophisticated method more appropriate than existing simple methods?

Additional synthesis of the results would be helpful to demonstrate the utility of the proposed method. For example, the authors suggest the metric can provide some 'link to physical processes', but no discussion of this is provided for the case study. How can the metric help us understand the role of dynamic morphology and levee performance in ensemble accuracy? How should we use Figure 9? For emergency response?

The accuracy of the derived SAR layer should be evaluated carefully, and its quality demonstrated to the reader. If this 'observed' layer is poor, the case study is not useful.

More specific and detailed comments are provided in the attached pdf.

I thank the authors for their contribution, and I look forward to the revised manuscript.

Please also note the supplement to this comment:
https://nhess.copernicus.org/preprints/nhess-2022-188/nhess-2022-188-RC1-supplement.pdf