

Interactive comment on “Are flood damage models converging to reality? Lessons learnt from a blind test” by Daniela Molinari et al.

Anonymous Referee #3

Received and published: 19 April 2020

Review: Are flood damage models converging to reality? Lessons learnt from a blind test

D Molinari et al.

The paper addresses the characteristics, performance and transferability of various flood damage models. It compares the modelling results of nine models by means of a ‘blind validation test’. This is a very interesting, innovative and challenging approach which has successfully been mastered by the authors revealing some interesting results. Overall the paper is well-structured and - with only few exceptions - well-written.

Within this very positive general picture of the paper there are some few points I would like to mention, which represent some kind of limitations, most of them of methodolog-

C1

ical type. In my opinion, a main drawback of the whole study lies in the approach to ‘validate’ the results. The authors claim to assess the models’ reliability through a comparison with observed damages of a real flood event (chapter 3.3) and show respective results (for example Table 4 and 5). The following chap 3.4 is then dedicated to explaining why the models results differ so strongly from the observed results. Crucial reasons for this discrepancy found are then assigned to inconsistencies in the damage claims, that is, the validation data. This seems for me like an odd approach. If the damage claim dataset was meant to be used as a validation set, more emphasis should have been put on clarifying inconsistencies and maybe on further filtering the dataset down to a set of reliable data on damage of a reduced number of buildings. The authors only declare having had some informal conversations with experts but the explanations about low damage values observed remain fuzzy. For the approach of this paper a more thorough survey of affected people would have been necessary. The statement in line 471ff “According to our interpretation, inconsistency between expected and declared damage can be attributed to the fact that what is declared by citizens does not correspond to the actual money required to replace or reconstruct the whole physical damage suffered by the building” cannot satisfy and actually triggers a the more philosophical question, whether the ‘damage’ targeted by the model represents the damage felt by the people affected. I suggest shifting those parts of the chapter 3.4 with the explanation about how the validation dataset was generated to an earlier part of the paper as background information about the approach (where also the various models are described).

Additionally, I am not fully convinced by the way some of the results are analysed and presented, see my more detailed comments below.

Below some additional more specific comments:

- Line 81: “Reality is hardly reproduced by observed data” – I would suggest not to use the term ‘reality’ since there is no univocal damage value as you prove later on yourself
- Line 86f: “comparative studies over a broad range of test cases are essential for

C2

acquiring more confidence in the reliability of modelling tools” – after having read your paper I would not say that the test case increased the confidence in reliability, I would put the emphasis more towards understanding in detail how certain model results come about - Line 101: “the focus of this study lies in this specific set of models” – what does this mean? - Line 149: “was not uniform, as only some of the owners justified costs for fixing the damage by means of invoices” – if not earlier, here the reader should suspect that the quality of the ‘validation’ dataset is questionable. I would propose to already link here to further explanations about this ‘observed’ damage data. - Line 186: consider to call it variation rather than ‘difference’ - Line 195: unclear for me, until here I thought the observation data was derived from damage claims made after the flood event. How can you use ‘official claims’ to explain inconsistencies between estimations and observations? Or is the officially claimed data different from the claims mentioned earlier? Then you could have used the official ones for validation? - Line 207: does only this model use stage-damage curves? Or why is that here mentioned explicitly?

Comments on presentation of the model results: - The detailed statistical comparison of the model results with the overall average does in my opinion not really add value to the result analysis (this relates particularly to Table 3 and partly to Table 4, it is more adequately visualized in Figure 5!). Firstly, because the number of models is so small. Secondly, and more important, since it is a bit like comparing apples and pears as you say yourself in the interpretation. The large differences in the model results derive from the different types of models. Therefore, I would not list in detail the variation of the models to the average but only describe and explain the differences of those models with similar approaches. The analysis could be done in a more qualitative way since numbers such as average or variation do not make so much sense when the overall number of models is so small. - I consider some of the explanations about calculation findings as being far too long because too obvious when the conclusion is that the differences can be traced back to the different model approaches (for example Lin 332 ff)

C3

Continuation of specific other comments: - Line 453 ff: it is not clear if the percentages refer to the amount of building or the outlier value (I suppose the former but that needs to be clarified) - Line 581: “Consultations of experts with local knowledge can ensure the correct interpretation and use of observed damage data” – I would not agree with that, it may help but does not ensure. . . . -

Text parts with language issues: - Line 95: “being them unknown” – unclear, please consider reformulation - Line 442: something went wrong with “in, as.”

1. Does the paper address relevant scientific and/or technical questions within the scope of NHESS? yes 2. Does the paper present new data and/or novel concepts, ideas, tools, methods or results? yes 3. Are these up to international standards? yes 4. Are the scientific methods and assumptions valid and outlined clearly? Yes, with some limitations (see above) 5. Are the results sufficient to support the interpretations and the conclusions? Yes, with some limitations (see above) 6. Does the author reach substantial conclusions? yes 7. Is the description of the data used, the methods used, the experiments and calculations made, and the results obtained sufficiently complete and accurate to allow their reproduction by fellow scientists (traceability of results)? yes 8. Does the title clearly and unambiguously reflect the contents of the paper? yes 9. Does the abstract provide a concise, complete and unambiguous summary of the work done and the results obtained? Yes 10. Are the title and the abstract pertinent, and easy to understand to a wide and diversified audience? yes 11. Are mathematical formulae, symbols, abbreviations and units correctly defined and used? If the formulae, symbols or abbreviations are numerous, are there tables or appendixes listing them? yes 12. Is the size, quality and readability of each figure adequate to the type and quantity of data presented? Yes, with some limitations 13. Does the author give proper credit to previous and/or related work, and does he/she indicate clearly his/her own contribution? yes 14. Are the number and quality of the references appropriate? yes 15. Are the references accessible by fellow scientists? yes 16. Is the overall presentation well structured, clear and easy to understand by a wide and general audience?

C4

yes 17. Is the length of the paper adequate, too long or too short? Adequate, some parts are too long 18. Is there any part of the paper (title, abstract, main text, formulae, symbols, figures and their captions, tables, list of references, appendixes) that needs to be clarified, reduced, added, combined, or eliminated? See my detailed comments 19. Is the technical language precise and understandable by fellow scientists? yes 20. Is the English language of good quality, fluent, simple and easy to read and understand by a wide and diversified audience? Yes with some limitations (see my comments) 21. Is the amount and quality of supplementary material (if any) appropriate? N.a.

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., <https://doi.org/10.5194/nhess-2020-40>, 2020.