Review "Estimating flood damage in Italy: Empirical vs expert-based modelling approach"

The manuscript "Estimating flood damage in Italy: Empirical vs expert-based modelling approach" validates different types of flood damage models for Italy and discusses the advantages and disadvantages of these models. This is a very interesting paper and the most extensive comparison of flood damage models for a specific area I have seen so far. I therefore believe this paper is a useful contribution to the scientific literature. I do however have some comments/questions regarding the setup of the study and some discussion points to be considered.

More important points:

- Currently the data-driven models developed in this study have been produced with data points from the same event it is validated on, hence no model transfer of the data-driven models is included. In practice a model transfer from one event to another is always required for flood risk studies, it would therefore be fairer to always train the models on 2 events and validate it on the third event. Such an approach is also carried out in Schröter et al., (2014) and Wagenaar et al., (2018) and both studies show that multi-variable models typically have more difficulties in such a transfer setting.
- I think the data-driven UVMs wouldn't perform so well in a transfer setting because the main advantage of MVMs seems their transferability (Wagenaar et al., 2018). In the current setup this advantage of MVMs isn't used. Also if the model setup is changed some discussion is required on how significant the model transfer is between the events and whether a MVM is required or whether the events are so similar that a UVM would do.
- For the wider applicability of the results of this research some more discussion is required on to what extend the good performing literature models are tailored to the specific flood event and setting. These expert-based models seem to be made for Italy and for similar flood events to the one seen in this study. Are these models for example also made for the same region, did the developers have access to the damage data of these events or did they carry out surveys in the region? Point here is to help the reader identify when you can take a model from the literature and when you can't and for this we need more information about the good performing literature models.

Minor points:

- The abstract currently mostly summarizes the method, as a reader I would be very curious about the findings (what works better). Could you summarize these in the abstract.
- Page 2 line 16-18: Can you clarify this sentence, it is unclear and seems very crucial for the story so I wouldn't want to look up the references to get this clarification.
- Page 3, line 32: You mention 1000 flood events in 45 years, that seems way too much, what do you mean here by the word "events"?
- Page 6, line 27: You choose to use relative flood damages rather than absolute flood damages. This is a common choice but I think not an obvious one, can you motivate this decision?
- Section 3.2 introduction: Nice overview on UVMs and MVMs but I think this needs something on the transferability advantage of MVMs (see above).
- Section 3.2.1: Can you make a heading for each literature model.

- Section 3.2.1: Huizinga got his damage curves from the literature also, could you reference to the study that Huizinga got his damage function from.
- Page 9, line 2: Change "observation" in "observed"
- The Random Forests and ANN both have all sorts of tuning parameters. Like number of neurons (ANN), minimum number of observations per leaf (RF), learning rate (ANN) and more. Could you describe how you determined these settings?
- On page 11, from line 20. You describe something about the setup of the study. I think this should be somewhere else in the manuscript as this probably applies to all data-driven models (that would be most fair to do this the same for all data-driven models). If not why did you do that differently for the other models?
- Sometimes you use the word "water velocity", sometimes "flow velocity" and sometimes "water flow velocity", I think commonly the word "flow velocity" is used. Can you unify this throughout the paper.
- Page 16, line 14. Not all these citations fit a root function to data they just all have damage curves that have the shape of a root function. So please rephrase the sentence before the citation (message can be the same).
- In this study a limited number of variables was available for the MVMs. If more variables had been available the models might have performed better. Can you make this point somewhere.

References:

Schröter, K., Kreibich, H., Vogel, K., Riggelsen, C., Scherbaum, F. and Merz, B.(2014), How useful are complex flood damage models? Water Resour. Res. 50, 3378–3395. doi:10.1002/2013WR014396, 2014.

Wagenaar, D., Lüdtke, S., Schröter, K., Bouwer, L., Kreibich, H., 2018. Regional and Temporal Transferability of Multivariable Flood Damage Models. Water Resources Research. Volume 54, Issue 1. https://doi.org/10.1029/2017WR022233