Journal: NHESS
Title: **Development and assessment of uni- and multi-variable flood loss models for Emilia-Romagna (Italy)**
Author(s): Carisi et al.
MS No.: NHESS-2017-342
MS Type: Research Article
Iteration: First review

The paper addresses flood loss estimation in Northern Italy, trying to highlight possibilities and limitations. By using flood damages recorded after the flood of the Secchia river in 2014, the authors (i) derive uni- and multi-variable damage models for the study area and compare them with models from the literature (ii) evaluate the transferability of such models to similar contexts and finally (iii) explore the relationship between damage to buildings and damage to contents for the available dataset.

The paper is in the scope of the journal and of interest for the research community working on flood risk; although "local" in the analyses, its results can be generalised to other contexts as well.
The paper is well organised, data are properly described, as well as methods, although some minor integrations/specifications are required with respect to the latter. Likewise, there are some minor imprecisions to be corrected in the whole text. The discussion of results can be improved with respect to some aspects (see below).
In general, the paper is a little bit long. Some suggestions are provided in the following on parts that can be neglected or shortened; nonetheless, the paper can take advantage of an English review aimed at simplifying articulated and (repetitive) sentence.

Major criticisms

Section 1
- The Introduction is too long. I would shorten the first paragraphs on the importance of flood losses and omit the discussion on aleatory and epistemic uncertainty (the following part on specific uncertainties related to damage models is more interesting for the paper).
- Section 1.1 should be re-organised by first declaring the objectives of the research and then the tools/methods. The present form is totally clear only after reading the whole paper.

Section 3.1
- The discussion on the difference between declared and refunded damage can be shortened in my opinion, by neglecting details.
- I agree on the use of declared data (instead of refunded damages) but it is not clear whether implemented damage data above 15.000 euros were verified or not. If this is the case, data below 15.000 euros are less reliable and authors should take this aspect into account in the analysis.
- I do not agree with the use of OMI data for the assessment of buildings value that, as stated by the authors in the Conclusions, "are more an expression of the overall economic well-being of a specific area" rather than of the real value of the buildings. (Re)construction costs are more suitable to the objective in my opinion.

Section 4.1
- The description of the damage models can be shortened by referring to available literature and leaving only the significant information for the paper (i.e. how models have been implemented).
- Authors implement models developed to be applied at the micro-scale (e.g. MCM, Flemo-PS) and models developed to be applied at the meso-scale (e.g. Rhine Atlas, JRCs). I guess whether damage estimation (i.e. models performance) is influenced by the different levels of knowledge/detail of input variables required by the models vs. available data. Did authors explore this aspect?

Section 4.1.1

- How authors converted the absolute curves of MCM in relative curves? MCM curves were developed in 2005 while the flood occurred in 2014; Did authors apply a discount rate to estimated damage?
- Why authors chose to convert absolute curves by mean of the average economic building value in the study area rather than by using different values for the different OMI zones? I would adopt this second option as MCM is a "micro" scale damage model.

Section 4.2
- Which is the formulation of SEMP?

Section 5.1
- From figures 7, 8, 9, it seems that uni-variable local models always estimate a relative damage around 0.1 (independently of the value of the dependent variable). Did authors notice that? How it can be justified?
- How authors justify the bad performance of SVM in estimating the total absolute damage?
- With regard to existing models, I expect that models with the best performance underestimate the total damage (as citizens tend to overestimate damage during declaration). In fact, four of the six best models underestimate. Can authors comment on that?

Section 5.2
- This section could be rewritten and improved to better explain the significance of results. Finding correspondence between authors' considerations and figures/tables is not straightforward at present.
- There is no correspondence between Figure 11 and its description in the text. Check also models acronym. Correspondence between test and figures is often lacking.

Section 5.3
- The link between the performance in estimating damage to buildings and damage to contents is not so evident to me. Why SMV that is the one with the best performance in estimating damage to buildings is quite bad in estimating damage to contents?

Conclusions
- The transferability of local models stated in the last part of the section should be better discussed previously in the paper. Two/three sentences highlighting this point can make conclusions more robust

NB
Pay attention to be consistent in terminology. Authors use damage to "contents" and "content" interchangeably. I guess they are typos. The same can be state for model acronyms (e.g. SMV sometimes becomes MV).

Specific minor comments (which can increase the readability and clarity of the paper)

Section 1

Pg. 2 line 17 "flood risk is the combination of hazard (i.e. the probability of a flood event with a certain intensity to occur in a specific area and in a specific time period) and consequences, providing for instance information on the vulnerability, i.e. the type and number of elements affected by a given flood event, and how well they are able to resist" → from this statement, I understand that consequences and vulnerability are the same "concept", please rephrase

Pg. 2 line 24 "Uncertainty exists in all flood risk components" → do authors mean "in the estimation of" all risk components?

Pg. 2 line 35 "Nevertheless, several authors indicate that damage models still provide an important sources of uncertainty in flood damage estimates" → do authors mean in flood "risk" estimates?

Pg. 3 line 12 "These models were shown to outperform uni-variable loss models, under the condition that sufficiently large and detailed damage data-sets are provided" → for what? Please specify

Pg. 3 line 16 "A further aspect that contributes to the uncertainty is the lack of sufficient, comparable and reliable high quality flood loss data" → for what? Please specify

Pg. 3 line 17 "In the absence of empirical damage data, damage models are either selected from the literature or subjectively and schematically derived by experts using a synthetic approach" → I would move this sentence after line 32, i.e. after the discussion on the importance of data

Pg. 3 line 35 "see e.g. Molinari et al., 2014b, on the transferability of the model developed on the basis of specific flood event data by Luino et al. (2006) and Freni et al. (2010)" → the content of the paper authors refer to is not clear, please rephrase

Pg. 4 line 8 → add the year of the event

Pg. 4 line 11 "The raw data collected by local authorities has been homogenized, geocoded and integrated with other useful information" → while geocoding and integration are fully described in the paper, homogenization is not discussed. Please give more information on this aspect.

Pg. 4 line 18 "Second, we calibrate empirical uni- and multi-variable models to subsections of the study area and validate them using the data observed in different subsections (split-sample validation)" → not clear at this point of the paper, please rephrase

Section 2

Pg. 5 line 7 "the aspect of the area is oriented" → I am not sure "aspect" is the right term here

Pg. 5 line 10-13 → and what about the southern boundary? How was it chosen?

Pg. 5 line 21 "Thanks to several eyewitness accounts, video footage and studies conducted by the scientific committee" → Which scientific committee? Please explain

Section 3

Pg. 5 line 31 "citizens and property owners were asked to fill forms about public properties damages (form A)" → according to my knowledge form A is compiled by public authorities

Pg. 6 line 5 "The database regarded private properties affected by different kinds of potential damages" → why potential?

Pg. 7 line 22 "Focusing on residential buildings, we defined the building's economic values [e/m2] as the average of the values provided for each property in the same OMI zone" → Did authors consider 2014 values?

Section 4

Pg. 9 line 10 "These models associate relative (or absolute) losses with different input variables. The most frequently used models in Europe are uni-variable damage models, i.e. they estimate the amount of relative damages as a function of a single input variable, most commonly water depth, (Merz et al., 2010; Messner et al., 2007; Jongman et al., 2012), differentiated by building use, type, etc. (Gerl et al., 2016). → this aspect has been already discussed in the Introduction and can be omitted here

Pg. 9 line 20 "The damage curve implemented in the Multi-Coloured Manual" → I guess damage "curves"

Pg. 13 line 14 "The variables being randomly permuted presenting a low accuracy are the most important ones in the damage prediction, as their influence in the prediction process is very high" → not clear, please rephrase

Section 5

Pg. 14 line 5 "dummy variable encoding" → please check

Pg. 14 line 8 "The only variables that resulted significantly correlated with the relative loss to buildings were the maximum water depth, building value and structural typology. However, correlations coefficients between these variables and relative damages are low, precisely lower than +/-0.18" → where I can see this value?

Pg. 15 line 13 "Although this values are satisfying in terms of errors, the performance of this models are lower than the ones of the models developed on Secchia's data set (except SEMP model)" → Why authors can state this? A deeper discussion at this point can help the reader also in the following analyses

Pg. 15 line 28 "The reason behind this fact must be attributed to the morphologic and socio-economic context where this models have been drown, that differs considerably from the Secchia ones, in addition to the different criteria adopted to develop them" → which criteria authors refer to? Please comment

Pg. 17 line 31 "damages to contents turned out to be significantly correlated with the building footprint (Spearman correlation coefficient equal to 0.27) instead of the building value. A noteworthy feature of Figure 12 is the very strong and statistically significant positive correlation between damages to buildings and to their content (Spearman correlation coefficient equal to 0.59)" → Where I can see values of the correlation coefficients?

Pg. 18 line 21 "The performance of all considered models, with the exception of the last four in Table 8, show a difference between observed and predicted overall monetary losses to contents that does not exceed e ⊠4 million (except for JRC Belgium that presents a difference value of e 7.2 million)" → These values refer to the bias, i.e. the mean difference between observed and estimated values, not to the difference between observed and predicted overall monetary losses

Section 6

Pg. 19 line 18 "the building values provided by the Italian Revenue Agency (Agenzia delle Entrate - AE) represent the buildings market values at a given time of given building typologies that is more an expression of the overall economic well-being of a specific area rather than the depreciated economic buildings values in case of a flood event " → I agree that OMI values are not representative of buildings value but I do not agree on the need of considering depreciated values. Please comment

Pg.19 line 22 "the multi-variable (MV) forest provide the important advantage to avoid the need to find a parametric function that works with all the data" → not clear, please comment

Pg. 19 line 24 "These can be very useful information for authorities and stakeholders to define preventive measures and/or mitigation strategies " → some examples should increase the robustness of the statement

Figures

Figure 3 - caption "Grey points in the background represent the observed relative loss (buildings only)" → the sentence is repeated

Figure 4 is not recalled in the paper

Figure 6 - caption → please, correct MV with SMV

Figure 12 – caption→ "absolute damage to buildings" is missing from the list of predictive variables

Figure 13 → try to convert in logarithmic axes, the readability can improve

Table 2 → what are "interior and exterior fixtures"

Bibliography

I did not check the bibliography at this stage of the review. I reserve to do this in a second time.