Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2020-206-RC1, 2020


Natural Hazards and Earth System Sciences
Discussions

# *Interactive comment on* "Are new open building data useful for flood vulnerability modelling?" *by* Marco Cerri et al.

**Anonymous Referee #1**

Received and published: 20 July 2020

In this paper the Authors analyzed the possible contribution of using Open Street Map (OSM) data for enhancing the predictive performance and transferability in space of multi-variable flood damage models for the residential sector. To this purpose, they built a dataset by combining empirical observations from historical flood events in Germany and data derived from OSM, with the latter essentially related to building footprint geometry. Random forest regression models (RFM) were then learned on this dataset using regional sub-sets and were tested for predicting flood losses in other regions. The manuscript is overall well written and presented and the topic perfectly fits the scope of NHESS, following the path of similar papers published in the Journal in recent years. However, in my opinion, the study suffers from a main methodological criticality, i.e. the representativity of the new additional parameters in correctly characterizing



the building vulnerability to floods. Indeed, the nine selected parameters derived from OSM used for learning RFM were only related to the shape and extension of the building footprint area (with an obvious high correlation among them), neglecting instead other fundamental vulnerability variables, e.g. building material and type, presence of a basement, etc. As it is well known and understandable, footprint geometry has a high influence in determining flood losses; however, as shown in previous studies, the observed damage variability depends on many (hazard and) vulnerability factors, which should not be neglected for a comprehensive modelling of flood damages. This becomes even more important when we consider the problem of the spatial transferability of empirical damage models. For instance, we may have two regions which have similar characteristics in terms of footprint geometry, but very different construction types: in this case, an OSM-based multi-variable model would be totally unreliable. For this reason, the main question that the Authors asked in the title "Are new open building data useful for flood vulnerability modelling?" is a bit pretentious, given that the answer is quite obvious if they limit their analysis on including only the nine additional variables listed in Table 2. For the same reason, also the results shown in Section 4 are expected; moreover, these indicated that the consideration of all the new footprint parameters does not actually greatly improve model performances (Table 4). Also the variable importance shown in Figure 5 is only partly informative: it basically says that water depth is more important than building shape and extension, but this is already known (and also shown in similar studies, e.g. Wagenaar et al. 2017, Amadio et al. 2019, both published in NHESS). The Authors are right in saying that information on building attributes in the OSM database are scarce and not useful for the kind of analysis they performed in their study. However, they could have exploited other public databases existing in Germany (e.g. cadastral, city planning maps, etc.) for building a more complete dataset. Therefore, I would suggest to the Authors to consider this possibility and repeat the same analysis in order to have more interesting results for improving our knowledge on flood damage modelling.

Specific comments: - P1.L13-15 and L16-18: based on previous general comments,

I find these sentences potentially dangerous. - P1.L17: what do you mean with “consistent”? - P3.L14-16 and L29-30: you said that one of the main aims of the paper is to understand which building variables are useful to characterize building vulnerability, but you actually investigated only footprint-related indicators, which only capture part of the overall building vulnerability. - P3.L32: typo “modelsi”. - Figure 2. Acronyms shown in the figure are defined in the text of the paper, but it would be better to report them also in the figure caption. - P10.L19: “this analyses” -> “these analyses”. - P10.L31: missing parenthesis after “Table 2”. - P12.L4: please rewrite this sentence. - P12.L8: remove comma after “reasoning is” - P12.L5-14: this part should be moved to the previous section. - P12.L18-19: please rewrite this sentence. - P13.L21-26: as discussed in general comments, this result is expected and only partly informative, because you neglected other important vulnerability variables. - P14.L1-2: this is also expected and due to the selected variables. - P14.L11: missing parenthesis after “Table 4”. - P15.L15: you finally chose the models with 6 and 8 variables (as the best performing ones). This is fine, but, actually, the variability in the performance indicators is very small (this is also due to the used variables), and probably you could have opt for the simpler models. - Figure A2 should be moved to the main text (and not in the Appendix) and discussed in more detail for the interpretation of the results. - P19.L6: I think this point deserves more discussion and analysis (see also my general comments). You just mention it. - P20.L17-22: as in the abstract, these are potentially dangerous statements. - P21.L7-8: as in the abstract, these are potentially dangerous statements. - P12. L9-10: I agree and this is what I suggest you to do (you can use information from other public databases to be merged with data coming from OSM). Otherwise, at present, this study provides partial (and potentially misleading) insights for flood damage modelling.

---

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2020-206, 2020.