

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

Anonymous Referee #2

Received and published: 27 August 2020

Dear Editor / Authors,

Thank you very much for the opportunity to act as a review on your very interesting paper which I believe is very much worthy of publication. The paper is really addressing two questions:

1. Can a Machine Learning Technique provide good predictions / estimates of flood damage using Open Data in the same location? 2. Can the results of the open data model then be transferred to another location, and how good are the predictions / estimates?

These are both very interesting questions and the paper addresses it well.

I am not an expert on Machine Learning Algorithms such as Random Forests, so I

C1

found myself having to do some background reading to understand the methodology. I suspect this may be a problem for general readers such as myself. If I look at Figure 2, this presents the steps well, but I wonder if there is a possibility for a clearer explanation in a few lines of the purpose of the Random Forest approach. Almost along the lines of "Random Forests are able to make predictions of flood loss by creating numerous decision trees, based on the random selection of decision nodes"?

The methodology and analysis is well described and the figures are clear and well labelled.

The only reservation is that the conclusion and abstract could be strengthened because I think it's an interesting paper. You write in the abstract that "However, our results show that using numerical spatial measures derived from OpenStreetMap building geometries does not resolve all problems of model transfer." You say the models are useful, but I don't get a sense from the abstract or conclusion that you are very confident in this. Similarly, if I were to jump to the conclusions, I don't get a clear sense of how well the open data models work, first of all, in the same location, and when you transfer them to different location, without having to go back into the results and discussion. I feel the conclusion should be clearer here to state what was the real value in using OpenStreetMap data.

I have some grammar / typo suggestions.

Page 2 Line 11 - "Modeling". In the rest of the paper, you use modelling – please be consistent with the spelling, except in the references where titles are quoted directly. Line 16 – advance, not advancement Line 32 – "Tree-based". I would use a hyphen here Page 3 Line 11 "It was shown that particularly geometric information about buildings as for instance building area and height are useful variables to describe building characteristics relevant for estimating flood losses (Schröter et al., 2018)." I think this sentence could be simplified – "It was shown that geometric information such as building area and height are useful ..." Line 14 – "building footprint geometry" – the word

C2

footprint or geometry can be removed Line 23 – “most of civil and common uses” – of can be deleted. Line 33 – modelsi – please correct the typo Page 7 Figure 1 - could you present the locations all on one map of Germany? I appreciate this would mean overlaying Dresden and the Elbe, but three maps seems unnecessary. Line 10 – The spatial measures are described in a table – I think the paragraph can be eliminated as the table repeats the information. Page 9 – Please correct transver in the figure. Page 20 Line 30 - Please correct OpenStreeMap.

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