Nat. Hazards Earth Syst. Sci. Discuss., 3, C2045–C2048, 2015
www.nat-hazards-earth-syst-sci-discuss.net/3/C2045/2015/

Natural Hazards and Earth System Sciences
Discussions

Interactive comment on "Using open building data in the development of exposure datasets for catastrophe risk modelling" *by* R. Figueiredo and M. Martina

V. Roezer (Referee)

vroezer@gfz-potsdam.de

Received and published: 15 October 2015

General comments

The presented method for disaggregating census data for exposure datasets takes an innovative approach and has some potential with more and more building vector data being openly available. The method itself is outlined in a clear and understandable way, additionally supported by a flowchart visualizing the steps taken. However the manuscript is not convincing in the critical assessment of the model results, as well as the limitations of the presented method. The validation of the aggregation results is

currently limited to a relative comparison of a not further described method. I would also recommend to improve the description of the data used in the manuscript, as it lacks some important information such as the year of the building vector dataset.

Specific comments

Accuracy assessment:
Currently the performance of the presented method is evaluated by comparing the disaggregated floor area per storey for reinforced concrete buildings with the results of a disaggregation approach using the 2011 GEOSTAT dataset. In my opinion this approach has several weaknesses:

- It does not quantify the improved accuracy over the population distribution method
- It does not validate the accuracy in regard of real world data
- It does not quantify the uncertainties of the presented method

Therefore I would recommend taking data from field surveys or other real world data into account for validating the presented method. Validating disaggregation results with field data or other real world data is frequently treated in literature (e.g. Thieken et al. 2006, Stevens et al. 2015). According to your height classification method using field survey data, I would assume that there is already field data available.

Sample areas:
Regarding the sample areas, I would recommend to take a third grid cell from a more densely populated area of Pavia into account, or give more detailed reasons why you have decided to choose GC1 and GC2 as sample areas (see Fig. 7 in the manuscript). Currently one could speculate that the presented method only out-performs existing approaches in areas with low urbanization rates and therefore very uneven population distributions.

Time shift between data sets:
The four data sets used in this study (census data, building vector data, CORINE land cover data and GEOSTAT population dataset) are all from different time steps, spanning gaps of over 10 years. Although urban structures are not changing very dynamically, data sets with differences of 10 years and more can be an important source of uncertainty in model results. Therefore I would recommend to include this issue in the discussion, regarding limitations and uncertainties of the presented method.

Technical corrections

5047-L9: “in order” instead of “in other”
5047-L19: split sentence instead of semi-colon
5047-L23: “high spatial resolution” instead of “high level of spatial resolution”
5047-L27ff: The length and the structure of this sentence makes it very difficult to read.
5048-L17ff: Sentence too long. Better split into two separate sentences.
5049-L10: “for now” instead of the “time being”
5049-L12: “will increase” instead of “will continue increasing”
5050-L1: “the aforementioned type of data”: I would recommend to be more specific here, as you mention at least three different types of data before.
5051-L9ff: Sentence is very long and difficult to read.
5052-L1ff: Please add source and year of the data. I would also recommend to give more details about the building vector dataset.
5052-L15: “known” instead of “know”
5052-L21: “large number of buildings”: I would recommend to be more specific about the number of buildings, to maintain the traceability of results.
5052-L25ff: I would recommend to add a formula here, to be consistent with the following steps and improve readability.
5053-L6f: Sentence structure unclear. Please rephrase.
5053-L20: Please define the variable “s” to make clear that “s” stands for “number of

storeys”
5055-L26f: Please add a source or a more detailed description about the exposure model “derived from the 2011 GEOSTAT population dataset”
5060-L5: “Fig. 15” instead of “Fig. 16”
5070: Please add description of “OT”, “RC” and “URM” in the caption to improve readability. The color-coding of the bars in Figure 1 are misleading, as they suggest a causal relationship between the bars in the two graphs.
5071: I would recommend to add all information coming from the census data or at least the variables used in the study such as type and year of the building

References

Stevens FR, Gaughan AE, Linard C, Tatem AJ (2015) Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. PLoS ONE 10(2): e0107042.

Thieken, A. H., Müller, M., Kleist L., Seifert I., Borst, D and Werner, U.: Regionalisation of asset values for risk analyses. Natural Hazards and Earth System Science, 2006, 6 (2), pp.167-178.

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., 3, 5045, 2015.