

Nat. Hazards Earth Syst. Sci. Discuss., referee comment RC1
<https://doi.org/10.5194/nhess-2022-227-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on nhess-2022-227

Zoran Stojadinovic (Referee)

Referee comment on "Development of a Seismic Loss Prediction Model for Residential Buildings using Machine Learning - Christchurch, New Zealand" by Samuel Roeslin et al., Nat. Hazards Earth Syst. Sci. Discuss., <https://doi.org/10.5194/nhess-2022-227-RC1>, 2022

1. The overall quality of the preprint (general comments)

The overall quality of the preprint is good. The topic of mapping a building representation directly to monetary compensation is exciting and significant for the science community. The research is well structured and explained.

But, since the dataset is substantially large and the prediction task relatively easy (3 broad classes), this reviewer expected slightly better results in the confusion matrix. Therefore, it seems reasonable that the model could improve with some adjustments.

2. Individual scientific questions/issues (specific comments)

Here are some suggestions to improve the model and preprint:

a) Dataset.

For better prediction results, the authors should preserve (and demonstrate) the original data distribution from the initial dataset when merging and filtering instances. For the same reasons, this reviewer believes that it is necessary to include the class of undamaged buildings in the dataset (with 0\$ compensations), unavoidable when mapping damage states.

b) Building representation.

For a better presentation of the mapping problem, the authors should show the data distribution for more features (like in Figure 6 for construction type). Surprisingly, the height of buildings isn't included in the building representation even though it captures the dynamics of the building. A feature like "number of floors" could possibly be informative.

c) More discussion on the confusion matrix would be helpful. For example, how do the authors explain the accuracy difference between b) and c) in Figure 12? Does it have to do with PGA ranges of earthquakes? The worse prediction seems to be "Predicted Medium / Actual OverCap". How to explain this?

d) It would be interesting to evaluate the prediction accuracy for the sum of compensations for all buildings. It is reasonable to expect good prediction accuracy for total cost since errors would cancel out each other. But it is difficult to perform without precise "OverCap" values.

e) Finally, how do the authors evaluate the usefulness of the research and model implementation for new earthquakes? Namely, what about the changing value of money over time and frequent changes in market prices? How to implement the model without the class of undamaged buildings (this version could work just if damaged buildings were pre-selected)?

3. Technical corrections

There is a significant number of needed technical corrections. Some examples are

highlighted in the attached file. The authors should carefully check the paper.

Please also note the supplement to this comment:

<https://nhess.copernicus.org/preprints/nhess-2022-227/nhess-2022-227-RC1-supplement.pdf>