Review of "An open-source radar-based hail damage model for buildings and cars" by T. Schmid et al.

General comments

Schmid et al. describe the development of a new open-source hail damage model. All aspects are clearly and thoroughly presented: the building of the hail hazard from radar data, its conversion to damages using insurance company claims and exposure, and evaluation of model performance at smaller (1 km^2) and aggregate spatial scales. Further, the discussion of model limitations is detailed and objective, providing the reader with some understanding of its weaknesses. The information in this article will be of great benefit to researchers and practitioners in this area.

This reviewer thanks the authors for a well-written article containing lots of new and interesting information, and asks that they consider the points below before finalising their manuscript for publication.

Major comments

1. The main application of the model is a near real-time estimate of hail damages to help insurance companies respond appropriately to a hail event, and the model evaluation in Section 5 defines success as modelled total claim numbers/damages within one order of magnitude of observed. However, the difference between 5,000 and 500, or 50,000 claims is very significant to the industry, and more precise information would be valuable for potential users. As a suggestion:

- identify significant events, e.g. those with 500 or more observed or modelled building claims
- compute the ratio of model-to-observed number of claims for each event
- report the mean and std dev of this ratio over all events

2. The hail hazard from radar is described as the main source of uncertainty in modelled losses (in the Abstract, Conclusions and in sections 4.1 and 6). I think other factors should be considered before concluding on the main source of uncertainty in modelled losses.

Values in Table 1 of Kopp et al. (2022) indicate MESHS resolves the 28 June 2021 event as having a very high rank in history, and they describe MESHS return periods above 70 to 100 years. This suggests radar-based hail does resolve high hazard severity, therefore, the low estimated loss for this severe event must be due to the model's impact functions. This counter-evidence is just one event, however, inspection of the top ten observed events (focussing on the top right corners of plots in Figure 8) shows the model has a low bias for these most severe events, hence the model's impact functions could be responsible for those modelled loss errors which are most material to companies (responding to large hail events). If radar-based hail does resolve high hazard severity, then I can think of two reasons why impact functions fail to convert high hazard to severe loss, as now discussed.

First, the horizontal drift of hail from radar levels to the ground means the calibration of damage *per building* cannot be sharply resolved by radar data. For example, high MESHS will often be collocated with buildings experiencing smaller hail at the ground, and vice-versa, leading to a much weaker relation between MESHS and damage ratio than actually exists. This possibility may be worth discussion in the text?

A potential second source of weak relation from hazard to damage is the fact that the damage ratio varies with building attributes. The authors explore some possibilities in lines 462 to 470 and find some weak evidence. Have the authors considered splitting buildings by rural and urban settings, and building separate impact functions? Much higher damage ratios in rural regions are a significant feature in the U.S., though I don't know if this applies to Switzerland too.

To be clear, I ask that these possibilities be considered for discussion in the manuscript, while any further model development is at the authors' discretion.

Minor comments/corrections

1. line 16: could the damages be expressed in EUR, either as a replacement or as an addition to CHF? A significant fraction of readers may have to break away from the manuscript after the first sentence to check the value of a CHF.

2. In Section 2.2.1, could an annual timeseries of number of claims per year, and mean claim size (both normalised) be included? It may help to understand why model validation in Table 2 is slightly poorer in the first ten years – perhaps higher FAR is due to poorer data collection 10-20 years ago?

3. Figure 4, and text: results using the 4 km buffer could be viewed as more appropriate, whereas those for 1 km cells give the reader an indication of how much hail drift can affect the relation between hail at radar levels and damage on the ground. From this viewpoint, the solid and dashed lines could be swapped in the top row of Figure 4, and the text in Section 3 focuses more on the 4 km buffer data? Irrespective of whether the 4 km buffer is viewed as the standard choice, could the HSS for 4 km buffer data be included in Figure 4?

4. Figures 6 and 7, and text: could the CDR be included in these two figures? The CDR is the damage ratio, conditional on damage occurring, and often referred to as mean severity. CDR is expected to rise with increasing hazard severity but it is not clear this is the case from figures 6a and b, or 7a and b.

5. x-axis legend on figure 8d: typo in 'normalized'

6. line 351: any evidence that the latest radar has fewer small patches? Or is there more complete data collection in the present-day? (related to comment 2 above.)

7. Figure 10: the reader may get more benefit from maps of damage ratio values in the 2nd and 3rd columns, with total loss remaining in the title? The damage ratio reflects MESHS values and might be more interesting than loss values which mainly reflect exposure density.

8. Lines 471-474: results in Table 2 show MESHS clearly distinguishes damaging events (> 100 claims) from non-damaging events. The problem for most users will be that the loss model does not resolve the most severe events in history, and as mentioned above in the second Main Comment, MESHS may resolve high hazard severity but they are not translating to high loss.