Reviewer 2

Dear authors:

The authors built a Hail Damage Estimate method using Radar and ERA datasets. Some technological methods were adapted in generating data samples such as the filtering of claims and the virtual advection of SHI. These methods makes the sample more reliable. However one question confuse me: Nether the bulk advection or virtually advected were needed, but the advection algorithm needs the damage dataset as input, was this a conflict?

We apologise for the confusion, the virtual advection of MESH was absolutely needed for the analysis, as it provided the dataset used to match damage at ground with MESH/SHI aloft. It formed the basis for the training dataset (in addition to meteorological parameters) of the neural network. The bulk advection is not used directly in the neural network, but is used to qualitatively evaluate the performance of the virtual advection method.

The last paragraph in the introduction section was changed such that the way the data was used is more clear. The paragraph reads as below:

“

In our study, we leverage MESH data from the BoM’s national network of weather radars, combined with a 10-year dataset of hail damage insurance claims provided by Suncorp Group Limited (Suncorp) and meteorological data from ERA5 reanalysis (Hersbach et al., 2020), to train a deep neural network capable of predicting hail damage. The structure and data flow of the study is:

> section 2: Describes the insurance data, applied filtering, and normalization.
>
> section 3: Describes the radar data and calculated products.
>
> section 4: Describes the procedure to match the insurance data to radar observations.
>
> section 5: Describes the development and evaluation of the neural network driven by these matched insurance and radar data and aided by meteorological data.
>
> section 6: The new hail damage model is applied to the full data archive, no longer limited by insurance exposure, and the relationship between the predicted hail damage, MESH, and meteorology is discussed.

”

Major Comments:

1. In Figure 1, relative damage distribution of 12 archetypes were showed, I did not see this kind of archetype were used in the following research, this was only used for normalization? The different archetype in different areas may impact the final result.

The archetype normalization was indeed used in the following research. The following sentence was added to the end of the archetype normalization subsection to clarify that this is the case.

“The rest of the study uses the normalized damage instead of the original loss ratios, this way the effect from the different vulnerabilities from the various property types can be minimized.”

2. The relative damage precentage need further explained, was 0% damage in or not in the damage sample? Lots of zero damage were scattered in the following discussions such as figure 7, 8, 9,10 and 11.

The definition of damage was clarified as shown below in the Insurance Data section:

“To avoid introducing biases toward more expensive properties, we computed a damage metric known as the loss ratio, which is the ratio of incurred loss to the insured sum.
In this report, we will refer to this metric simply as damage and expressed as a percentage.”

Indeed, 0% damage was included in the study. The section that describes the Claims filtering was modified to clearly state this.

“After applying all the filters, our dataset consisted of 18 intense hail events and 12 medium hail events; these provided 1,775 damage grid points, and 76,703 exposed but undamaged grid points.”

3. In the claims filtering part, an obvious white cycle ring surrounding the hail area were created after filtering(Figure 2) . The discontinuous scatter round zero precentage damage (the left panel of figure 7 and bottom right in figure 10) need an explanation, was it caused by the 0% damage samples or the white cycle ring after claims filtering? The impact to CSI, POD and Far should also be discussed.

Figure 2 shows the time series of hail damage claims for Brisbane as a sample, and does not refer to filtering. We assume the reviewer is referring to figure 3. We are not certain what white cycle ring the reviewer is referring, but we speculate it might be what we refer in line 102 as a “buffer zone” between the damaged areas and non-damaged areas. Grid points in these “buffer” areas were not used in the analysis as hail damage is uncertain in this transition zones. We wanted to create a dataset with data that was as clean and certain as possible in order to minimize the uncertainty in the training dataset of the neural network.

The claims filtering section was modified to make it easier to read.

Regarding the discontinuous scatter between zero damage and about 0.5% the following paragraph was added to the manuscript in the Training and performance subsection.

“It is important to note that this CSI was achieved at 0.5\% damage threshold.
From the observed damage data (see figures 7, 8, and 9), it is evident that there are only a few claims between 0\% and approximately 0.5\%.
This is likely because losses below this ratio fall below the policies' deductibles and, therefore, are often not reported by property owners.
This apparent discontinuity in the observed damage data was also observed in the unfiltered data (Claims Filtering subsection), indicating that it is not a result of the elimination of uncertain damage areas.“

4. Besides SHI, why not involve Vertical Integrated Liquid water, echo top and composite reflectivity in your machine learning model ?

Vertical integrated liquid, echo top height, and maximum column reflectivity were all included in the initial iterations of the neural network, but were eliminated from the input early on during the SHAP analysis, likely due to the high correlation to SHI. The subsection describing the neural network structure and selection of input data was modified to clearly state this.

5. What’s the time interval of the new developed MSEH? In daily? Can this method be applied to instant radar datasets, which I believe more meaningful in hail warning.

It is only limited by the radar scan frequency. This is currently being developed for use as a hail nowcasting product. This possible application was added in the conclusions.

“In addition, future work will use this novel hail damage estimate for nowcasting applications to provide hail warning.“

6. Some composite reflectivitymap should be showed to see if the red cycle were spurious claims in figure 11.

This was done but we believe that the MESH daily max contours shown in figure 11 demonstrate this as the lowest contour is 20mm and is far from these claims.

The following sentences were added to better guide the reader:

"It is worth noting that this case exhibited some potentially spurious claims, as highlighted within the red circles in the panel F. All these point were tracked and found in a cluster in the map (indicated by the red circle in the panel E) close to the coast and were relatively far from the main storm swaths and could be due to the misclassification of damage cause (i.e., wind or flood damage instead of hail) in the insurance dataset."

Minor Comments:

1. (a) (b) (c) (d), should be labeled in panel figures.

Corrected figures 10, 11, and 12, thanks

2. Figure 6, what's the meaning of 'The dashed range ring is 150 km in radius'?

This was needed in a previous version of the image which was zoomed out, removed. Thanks

Note on the training of the neural network:

We identified that our initial approach to splitting the training and validation datasets could potentially lead to model overfitting due to the high correlation between the two sets. To address this concern, we adjusted our approach by performing event-wise splitting of these sets. This modification helped reduce the possibility of high correlation in meteorological conditions between the training and validation datasets. The resulting model maintained nearly the same level of performance as before, albeit now requiring 7 ensemble members instead of the initial 5. We have updated the model training section accordingly to reflect this adjustment.