NHESS-2021-342: Characteristics of hail hazard in South Africa based on satellite detection of convective storms

The authors clearly expended a lot of effort addressing many of my and the other reviewers' concerns. I appreciate their attention to detail. The new phrasing makes it clear an explicit OT temperatures – hail size relationship is not being established herein, which was my main concern in my previous review. Added text, figures, and appendices help clarify the process and provides verification context. I still have some concerns remaining, mainly requests for additional clarification and requests for reader cautioning, but they are not insurmountable to address.

Note: all line numbers herein refer to the author tracked changes document.

Major comments:

My major comments fall into two categories: requests for additional clarification, and points to make to caution the future reader.

Additional clarification:

- Lines 187- 202: What time ERA5 file is used for the insurance claims, considering the insurance data doesn't have a time of occurrence? Give the odd distribution of parameters for the claims data in Figs. 5a, 5b, it would seem a possibly unrepresentative time was chosen.
- Lines 225-228: This additional explanation helps, thank you. That being said, Fig. 7 is still not very clear. Why not separate it into ~3 subfigures over different time intervals, so the three events can be separately shown?
- Lines 241 - 250: I appreciate the added text, but a bit more clarification is still required. Line 243 introduces the phrase "potential hail events", an excellent addition, but further sentences don't use the phase. "Historic events" here could refer to historic hail detections from TRMM/GPM, S. African claims data, or the Australian/European reports.

  I understand that "potential historic hail events" is a mouthful to repeat multiple times; possibly "historic OT events" could be used instead. (And is used, in subsequent sections.)
- Lines 261-264: A Gaussian distribution is fit to each 3° by 5° box, correct? I'm having trouble following what the phrases in parentheses ("mean number of events " and "summer peak) mean.
- Lines 266-270: While I appreciate the extra text, I'm still having trouble following this explanation. I understand a random drawing of the day of year occurrence of N events in a grid box, and from each of the surrounding boxes. But are not all of these events retained? What are "blocks of N ⅓"? Most importantly, what N is chosen (and why?)
- Line 276: No blocks of N⅓ here?

- Lines 316- 317: While I appreciate the addition, the application of the nomenclature to this specific case is still a bit murky. What does the objective function predict? The sample data probabilities of.... a specific length and width?
- Lines 349- 375: While this section is clearer, some additional clarification can be provided. State up front at line 361 what statistical correlation relationships are preserved both in the historical dataset and the model. From Punge et al (2014, P14 hereafter) it seems there is a first step that moves from correlations between length-width-hail size to length-width-OT temp difference before the historical dataset can be constructed, is that correct?

  This explanation will also keep the reader from jumping to conclusions that some sort of OT temp difference and hail size scaling equation exists.

  A very quick recap (or reference to specific section of P14) for how track area is determined would also be helpful.

  I appreciate the authors uploading P14 to Researchgate so I (and future readers) can review it for answers to these questions.

  Finally, why are some of the correlations so different between the historical and modeled length/width and event to storm area ratio? Are these values still within the realm of reasonability?

<u>Cautioning the reader:</u>

- Lines 177-180: This phrasing makes it sound like the extensive calibration of the OT detection algorithm has been for improving its severe weather detection capabilities, but in my opinion has actually been for improving OT detection compared to human ID. The studies cited here found OTs to correlated with severe weather but were not explicitly looking at hail. Punge et al. (2014, 2017) and Jurkovic et al (2015) would be better hail-OT connection citations. It should also be acknowledged here that these sources found OTs near only about 50% of hail events, and Bedka et al (2018) noted that large percentages of OTs do not produce hail. I understand use of OTs is the best option the authors have, but all appropriate caveats need to be acknowledged up front.
- Appendix A: Fantastic addition to the article. It addresses many of the concerns I noted earlier about the relationship between MESH and OT probability.

  I would ask the authors to provide a few cautioning statements for the reader. Fig. A1, while convincingly establishing a link between increased MESH-estimated hail size and increased OT probability, does not establish a relationship with observed hail size at the ground. MESH is not observing hail fall but is instead a proxy for hailfall based on a storm's ability to loft condensate - essentially, updraft strength, much like OT temperature difference. The relationship between a storm's updraft strength and hail size produced at the surface is not linear, and at larger updraft speeds may in fact be <u>inversely</u> proportional (see Fig. 6 of Lin and Kumjian 2022). Readers should be cautioned against assuming Fig. A1 implies a similar connection with observed surface hail size.

- Lines 329- 340: While I appreciate the change in some of the phraseology, the text here still is connecting increased updraft speed with the ability to produce larger hail. While this could be true for smaller hail and/or weaker updrafts, this relationship doesn't hold for stronger updrafts, as Lin and Kumjian (2022) makes clear. (Marion et al. was about tornadoes so is not relevant here.)

  Please caution the reader that updraft strength has been shown to not be directly related with increases in hail size, and for stronger updrafts in particular the relationship potentially reverses. However, given the lack of other available data sauces, OT temperature difference here will be used as an estimate of storm severity, and will be connected to hail size via the reports databases., etc. etc. (I would avoid the term "updraft intensity", as it isn't clear if it means strength, area, or both.)

Minor comments:

- Line 10: Damage is not limited to large hail. Large quantities of small hail can be equally problematic, as can almost any size of windblown hail.
- Line 25: Cf is used for comparison, but only one figure is listed - perhaps e.g. was meant instead? Also note the reference is to their figure.
- Line 25-26: What methods did Smith et al. use to derive their frequency estimate?
- Lines 28- 32: Oddly phrased. What problems did Grieser and Hill (2019) face that leads the authors to conclude that hail pad and hail report data aren't sufficient? I'm assuming the difference is quantity of data in South Africa vs. the U.S., but phrasing could be improved.
- Lines 45- 46: Again, an odd transition. Based on <u>just</u> these sentences, a radar data climatology In S. Africa seems possible. Perhaps adding "but is not available over other large portions of the country" at the end of these sentences.
- Line 47: "for hail" → "for hail detection"
- Line 57-58: would rephrase to"... an appropriate proxy to assess individual severe convective storms (SCSs) and large-scale outbreaks for the potential of hail production. Large-scale hail-producing outbreaks can cause by far..."
- Line 99, Fig. 2b: Determining where the green colors start in Fig. 2b is difficult. Adding a black outline showing anvil detection would be helpful. Also, adding a sentence pointing out the Great Escarpment and the Drakensberg in the topography map would be useful for later references in the text.
- Lines 121- 122: I'd keep the mention of the Sandmæl algorithm but note that South Africa is not continuously covered by visible imagery.
- Line 124: " ... scanned for hailstorms..." → " .... scanned for OTs..."
- Line 72: Typically, 20% is used as a threshold probability in hail detection (e.g., Bang and Cecil 2019, 2021). Why the change here?
- Lines 145- 146: Prein and Holland (2018) focused on comparison of the distribution of hail environments across the globe and hail detection, not hail size

spectra (and they weren't particularly successful in global application.) What publications have focused specifically on *observed* hail size distributions across the globe?

Even if the answer is none, I think arguing from scarcity is reasonable enough, it just should be presented with the necessary caveats.

- Lines 147- 149: Any reason not to include reports from the US? With inclusion of MPing and COCORAHS sources, hail smaller than 2 cm could be included in the spectra calculation.
- Lines 182-184: Is the filter based on environments associated with the insurance claims? These sentences and lines 79-80 make it seem like that is the case, but such connections aren't described in this section.

  Because Punge et al. (2017) is behind a paywall (and Bedka et al (2018) mainly just cites Punge et al.) please provide a brief recap here.
- Lines 203-205, Figs. 5c-e: Adding thicker black lines where the filter threshold were chosen would be helpful.
- Figs. 3,4, 6: I'd prefer grouping these figures all in one figure, to allow for easier, direct comparison.
- Lines 538-539: Which criteria are used in the method described herein, ECS or OT? If OT, why, given that it seems like it misses a lot of hail - producing storms? (Perhaps because of false alarms, which could be indicated in another table column in Table A1.)
- Lines 219-220: What's the temporal resolution of the MSG data used for OT detection? That fact should probably be included in section 2.1.
- Lines 238-239: Probably should note a filter for these erroneous groupings is being developed, in future work. 🙂
- Line 241: What model?
- Line 251: frequency of filtered OTs, correct?
- Line 278: "observed and modeled.... OT events"
- Fig 10: These distributions are for the entire domain shown in Fig. 9, correct? Please note in caption.
- Lines 292-293: But both the "historic" and "modeled" events set here are of (filtered) OT detections, correct? So the chance of missing hail reports on the ground is, in this specific context, irrelevant.
- Line 305: "covered by OTs/hail streaks" → covered by individual OTs"
- Line 306: While I understand your meaning here, since this ratio has an inherent upper bound of 1 the phrasing could be better.
- Line 307: Isn't this product f^2?
- Line 308: " > 105 km², *not shown*"

- Line 328: "Storm's severity" → "tornadic intensity". "Storm severity" is too nebulous a term, since it could be interpreted as meaning "updraft strength", which is not necessarily correlated with severe impacts on the ground.
- Lines 381- 384: Why was this figure removed? A comparison of the frequency of occurrence of hail events (not including size) across the country in the historical vs. stochastic datasets seems of prime importance. If the figure is not retained, then discussion about it should be eliminated.
- Section 3.6: Much improved, and an interesting result when compared to the Smith et al. study. If I am reading correctly, the stochastic event set underestimates the occurrence of hail days in the region, but potentially overestimates severe hail. It's possible the break down in the updraft strength - large hail relationship is causing these large biased "severe" hail numbers (also possible large hail is underreported, as you note.) Any idea what could be causing the overestimation of hail in general?
- Lines 505- 507: While I agree with these statements, I would shift them earlier in the conclusions as they are awkwardly placed here.
- Fig. B1: Great addition. Can these plots be normalized by total number of detections and plotted in the same plot for easier comparison?
- Lines 570- 580: Excellent addition. I'd rearrange the text (or figure) so the subfigures are referenced in order.

Grammatical:

- Line 47: "this" → "these"
- Line 131:" ... east of..." → " .. in east..."
- Line 206: Add "Fig. 6" after Fig. 3, so the comparison has an object.
- Line 510: "95." → "95th"
- line 221: "of the event, *or grouped*, OTs"
- Line 267: "of" → "from"
- Fig 11: years → years'
- Line 417: "hail hazard" → "the hail hazard"