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

The authors have assembled a novel methodology for estimating hailfall in South Africa, a region with frequent hailfall but sparse observations. Such a hailfall climatology in South Africa is clearly needed, particularly with the possibility of shifting or increasing hailfall frequency with a changing climate. From this hail event climatology, they have additionally assembled a statistical model that includes estimations of hail size, hail swath shape and orientation, and frequency of occurrence over a much longer period. The creation of all these products is ambitious, but I feel the authors have overreached what is scientifically defensible though the necessary chain of assumptions. I don't doubt that there is a strong operational need for extended hail climatology products like these in this region, but if choosing to publish the work the assumptions must be reasonably defended. In sum, I recommend rejection in the paper's current form, but would welcome reviewing a resubmission on a narrow, better-grounded portion of the work.

<u>Major comments/fatal flaws:</u>
The work performed here was obviously extensive, and I appreciate the effort to scientifically ground an operational product. I've broken down my view of the chain of reasoning presented in the paper, along with my opinion of how well each step is grounded in the article.

1. *Hail occurrence can be estimated using the Khlopenkov et al. (2021) OT detections in GOES data over CONUS.* This step is well-grounded, given Khlopenkov et al. and Cooney et al. (2021) results discussed in the introduction, although a quick sentence or two discussing the skill level of that algorithm with the severe hail report database used in those studies would be useful to add.
2. *The Khlopenkov et al. OT algorithm can be applied to MSG SEVIRI data over S. Africa with similar success as GOES data over CONUS, with the additional environmental filtering applied.* This claim is generally supported by the results in the paper (c.f., Figs. 3 and 4), but needs a fuller explanation. The geographic hotspots are similar in Figs. 3 and 4, but is the frequency of potential hail occurrences reasonable? Comparison of OTs, GPM/TRMM detections, and radar-based detections over CONUS could confirm the relative change in frequency between OT and GPM/TRMM detections over S. Africa is reasonable. Comparisons should also be made to climatologies made over the region from other methods, such as those discussed in the introduction (Admirat et al. 1985; Prein and Holland 2018; Kunz et al. 2020; Dyson et al. 2020).
3. *The hail grouping methodology into events reasonably represents hail swaths from a single storm system.* While the description of the methodology (lines 176-177) is intuitive and simple, the results of the grouping methodology in Fig. 7 don't seem to follow that description. Why are there multiple events occurring at a single place and time? Once the methodology itself is cleaned up, a few example applications of this methodology in an area with radar data would show its value in establishing hail events and their duration and speed. Right now, the results of the methodology are only briefly compared in text to two other radar-based studies of severe convective storms (not limited to hailstorms) in the literature.
4. *The created hail event climatology shows reasonable distributions of hail event frequency by time of year and time of day.* No comparison of these distributions is made to the observational or GPM/TRMM datasets. While they are admittedly sparse, they should at least be able to confirm general seasonality. Comparisons should also be made to the other climatology datasets mentioned in point 2 above.
5. *The statistical method established in lines 202-213 can be used to produce similar hail event daily and seasonal hail event variations established by points 1-4 above (assuming points 1-4 are successful at representing actual hailfall).* The annual and daily

distributions produced by the model do appear similar – I'd prefer a difference plot instead of a side-by-side comparison, given the relatively large magnitudes involved. However, the description of the statistical method is not clear, and only one reference is sited. How common are methods like these? The steps involved in its description are very specific, making one wonder if the model is being over-fit to its underlying dataset. How similar is the methodology used here to Punze et al. (2014, unfortunately behind a paywall), what changes were made, and why?

6. *The statistical method in lines 225-238 can be used to produce similar hail event length, width, area, and orientation as the event climatology produced in point 3 above (again, assuming point 3 is valid).* These results do seem reasonable as presented in Fig. 11, but no point of comparison is provided. How well do other statistical methods perform? What is expected behavior?
7. *Hail size can be estimated using the OT climatology product produced in point 2 (I don't think the event climatology from point 4 is being used here, but text isn't clear).* This claim is (currently) indefensible.
    - Marion et al. (2019) suggested a relationship between OT *area*, not strength, with updraft width and hence potential tornadic intensity. That's a not insignificant difference. Hail size, particularly as one reaches larger hail sizes, is more related to updraft width than updraft strength (e.g., Nelson 1983, Foote 1984; Kumjian et al. 2021). I am concerned that by relating hail size to an updraft strength metric, an erroneous hail size distribution will be produced.
    - Khlopenkov et al. (2021) connected OT detection probability with hail *occurrence* and did not try to distinguish among hail sizes.
    - Figure 2 appears to represent original work from the authors (sentence is oddly phrased, making it seem like it is sourced from Murillo and Homeyer 2019). While I do appreciate the correlation shown, I am concerned the MESH95 dataset is being used, and not actual hail reports. Per Murillo and Homeyer, the MESH95 dataset has a significant large bias, with 40 mm being most skillful at determining 25 mm hail, and 64 mm being most skillful at determining 50 mm hail. That bias does not appear to be accounted for in Fig. 2. Further, while Murillo and Homeyer (2019) did not specifically examine the skill of tropospheric-OT temperature difference in differentiating among hail sizes, they did examine the distribution of minimum GOES IR Brightness and GOES OT Area (see their Figs. 6a, b, 8a, b), and did *not* find a strong relationship between those fields and observed hail size.
    - In my opinion, this claim cannot be supported given the current literature, and hail sizes should be removed from the database (or only provided to customers with a strong caution about their use, and not published in the literature).

Given these issues above, I cannot recommend the article for acceptance. I would be happy to review an article focusing on points 1-4 above, after addressing the issues I've described. A companion paper focusing on points 5-6, after points 1-4 are successfully established, would also be interesting. I cannot support an article including point 7 at the current time.

<u>Other major comments:</u>

- Section 3 would be much easier to follow if the observations were presented first (i.e., annual, daily cycles) with comparison to other parts of the world, and only then the development of the statistical method discussed. In my opinion the main scientific point of the article should be the new climatology and establishing those results should be given more weight than the statistical developments (which rely on the new climatology being accurate). I would prefer statistic results being shifted to a new paper or at the least a new section.
- Event grouping:

    - Lines 173-174: This definition seems straightforward and reasonable. However, Fig. 7 doesn't use that definition, instead identifying multiple events that exist within a larger event. For the climatology to be useful (e.g., for Fig. 8 to convey valuable information),
      explicitly restricting each OT to a single event is necessary.
    - Line 197-198: Why not an equirectangular grid? Why this resolution? Lines 197-198: Shouldn't this just be calculated as the number of events total in each box, divided by the total number of years? Introducing the number of overshooting tops seems unnecessarily complicated. Are the authors retaining their criteria that OTs must last longer than a single time step (lines 181-182)?
    - Line 200: Wouldn't using number of events instead of OTs better address this issue? To further constrain the issue of multiple events in the same location, events could be capped at one per day.
- Temporal distributions
    - Lines 203- 209: What made the authors think a Gaussian distribution was best? Why was this large a grid needed- can the authors discuss additional sizes that were tried?
    - Lines 205- 209: Can the authors point to other studies or methods even in another field, that use a similar technique? These values appear very tuned to these specific observations/year groups and hence potentially not broadly applicable. Finally, the explanation itself is confusing. "The boxes distribution in the N events in this box and 8 surrounding boxes" - do you mean a distribution in time or in space? All blocks of $N^{1/3}$ (not clear what that means) are then randomly assigned to the same event?

<u>Minor comments:</u>

- Lines 33-35: Given this article is about S. Africa, it would be worthwhile to point out that HAILCAST originated there with Poolman (1992).
- Lines 33-35: I would also note retrospective dynamical downscaling with NWP models can be used to generate climatologies that include the local and mesoscale processes listed here, particularly if convective permitting resolution is used.
- Line 40: The results of the Ayob (2019) study should be included for comparison with results from this study.
- Figs. 1 and 2 need to be switched to the correct order.
- Fig. 1: Given that, from my understanding, only the IR 10.8 brightness temperature is used in OT detection, it isn't clear how those OTs in Fig. 1a are identified as they don’t align with local minima. Are they supposed to be northwest of every cold cloud peak?
- Line 96: Is Fig. 1a also supposed to be showing the IR anvil detection index?
- Fig. 6: With the two figures several pages apart, it is hard to see the difference between Figs. 6 and 3. Could they be combined into one figure with two side-by-side sub figures?
- Lines 111-113: What will the European and Australia reports be used for? Or are they reports of hail in S. Africa? Confusing.
- Line 111: Prein and Holland (2018) was not a study of observed data.