Characteristics of hail hazard in South Africa based on satellite detection of convective storms

Heinz Jurgen Punge[1], Kristopher M. Bedka[2], Michael Kunz[1], Sarah D. Bang[3], and Kyle F. Itterly[4]

Summary.

This paper reviews and analyses the hail climatology for South Africa and regions within South Africa using satellite detection of convective storms. The paper investigates 14 years of geostationary satellite observations of convective storms and generated a spatiotemporal multivariate stochastic model representing 25000 years. The historic, stochastic and observed insurance exposure and vulnerability data are analysed to identify the expected hail damage for return periods of 200 years.

General Comments

Scientifically the authors have done an excellent job considering the literature review, data collection and modelling that have gone into the research described in this article. One slight drawback is that in the paper itself, the authors attempt to address all the whole multiple complex modelling processes in a relatively short and succinct manner. At times, unfortunately, the description of the process followed does not do the modelling process justice and made it difficult to follow what was done. Some examples will be given below.

The format of the paper also made reading the paper very difficult for me. At times, the figures and tables mentioned in the text were not next to or near the text referencing them. This caused a lot of scrolling up and down in pdf (and eventually I just printed out the text in frustration). As times, figures were even placed to appear to be part of a previous section e.g. Fig. 14 seems to form part of the end of Section 3.3 but is part of Section 3.4 that starts underneath Fig. 14. The format used by the authors may be due to format instructions from the journal. If not, please reconsider the placements of figures and tables to be as close as possible to the relevant text to improve the reading flow of the paper.

Specific comments

- 25. The reference to Grieser and Hill (2019). Did Grieser and Hill focus on hailpad derived metrics for South Africa, another country or just in general?

- 35. "*hailstorm formation is often related to local and meso-scale processes related to, for example*" Perhaps do not use the word related twice in one sentence.

- Figure 1 "*a title: SRTM*" acronym is not defined. Please check whole text for acronyms.

- 95. "*Based on past experience, only OTs detected with a probability >50% and with a surrounding anvil cloud (green and yellow colors in Fig. 1a; the IR anvil detection index, a rating based on an anvil detection model accounting for viewing situations, greater than 10; see also Scarino et al., 2020) are used in this work.*" Make the sentence in brackets a sentence on its own or add to figure description.

- 110 - 115. "*A uniformly distributed random number between -0.5 and 0.5 115 was added to each reported hail diameter to compensate*" I assume the -0.5 and 0.5 is also in mm?

- 155 "*As 9.5% of the OTs occur at a melting level of less than 2 400 m, but ony 3.5% of the microwave hail detections and and 2.5% of the claims, a lower threshold of 2 400 m was introduced for this parameter.*" . Spelling. Sentence seems incomplete.

- 160. "*The latter feature is due to the minimum freezing level condition and remains to be confirmed by independent observation.*" Independent observation from whom?

- 190: "*complemented with hail size information from reporting.*" Reporting? You mean the insurance reporting?

- 200. "*Following Punge et al. (2014), both annual and daily cycles are modeled with Gaussian distributions. For the day of year, domains of 3° × 5° are considered, and depending on the…*" Why Gaussian distribution? What is the statistical justification for it? Not sure how these grid definitions relates to the previously defined rectangular grids of 0.3 x 0.5 mentioned on page 10.

- 205: "*Days are drawn from the boxes distribution for the…*" It is not clear to what the boxes distributions refer to.

- 205: "*…Finally, the day is retained only for N/9 events at random. This procedure has been found empirically to approximate the observed space-time distribution of days in a satisfactory manner.* " Is this procedure self-developed or taken from somewhere? Why n^(1/3) and N/9 - those specific values? What is the proof of empirically proof behind it?

- 210*: "from a region of 10° × 6° around..*" Why the double grid size? Is this to also represent the 8 neighbouring grid cells?

- Paragraphs 205 and 210 can be extended to make the spatial construction more clear. In the current format is it difficult to follow, and relate back to standard spatial weight matrixes using the queen criterion.

- 215: "*Also note the secondary maximum in fall (around days 100–150, i.e. April and May) during nighttime, represented in the model.* " Does this represent a local maximum? How do you see from the graphs it is in the night?

- *215: " It is shifted towards fall over the Southern Ocean." ???* Are you modelling that far away from the shores of the country as well? And will it have any landfall impact?

- 220: "*Time, slightly earlier than Smith et al. (1998, 5–6 pm) but consistent with Olivier (1990) (Fig. 11b). The daily cycle is most pronounced…*" What is the possibility of there being a shift in these times from the 1990's to now? In that case, would the results be comparable?

- Figure 10: I assume the day of the year for 1 to 365 represents 1 Jan to 31 Dec. Perhaps add that to the title to indirectly show the difference in expected hail occurrences for northern and southern hemisphere?

- Figure 11: I'm struggling with what number of days each bar represents. It seems the number 50 falls on the 4th bar?? This will only work if each bar represents 12.5 days?

- 225: "*The distributions are well approximated by the GEVs*". GEV is an extreme distribution that requires a "limit" (e.g. peaks-over-threshold or block-maxima) in the data over which you are modelling events? What was that limit and how was it obtained?

- 230: " *to give unrealistic large values, which is why length and width have been truncated at 1.5 times the largest observed values,*" Why the specific value of 1.5 times the largest observed value?

- 230: "*In addition, the fraction f of the event area (the area of the ellipse spanned by major and minor axis of lengths l and w,*" Remember to write the last l and w in italics.

- 235: "*Table 1 lists the distributions and parameters for these event properties.*" Which method was used to estimate the parameters of the distributions?

- 240: "*We find that most frequently, events have an orientation of around 100°, i.e., propagate eastward to southeastward (Fig. 11f).*" This is for the whole country. But it may be misleading as this is not the typical orientation for a high hail fall region like Gauteng where storms normally originate in Johannesburg and move north-easterly to Pretoria. As seen from the discussion in the next paragraph.

- 280: "*sets of random numbers for each property from a uniform distribution and determine ranks. Then, for each property, we draw values from the actual distribution, sort them, and attribute to events using the pre-determined ranks. "* How? Does this again refer to a previously defined or described methodology?

- 285: "*could be expected, smaller regions show relatively higher variability, but there is strong correlation between the two. This*" Which 2? Smaller regions and the country as a whole?

- 295: This section describes the South African domain in terms of latitude and longitude degrees, subregions etc. Should this description not be done earlier in the paper to set the scene – perhaps where Figure 9 is defined?

- 305: "*50 hail days per year), while in an 305 equivalent sample of subsets from the stochastic event set, the event count ranges from 1 883 to 2 162 events on 671 to 703 days*" Perhaps add the equivalent hail events and days per year for comparison.
- 305-310: " *In the Highveld region, there were 74 days per year…"* These numbers are averages per year. The averages per year for the years defined by the authors and the years defined by Smith et al 1998 are different and it should be considered that several climate changes occurred in the years in between. This includes periods of severe drought in the several regions in the country especially between 2010 -2020
- 310: Can the numbers given in this paragraph be added in a table for easier reference?
- 310: "*However, severe hail (>31 mm in diameter*" Why 31 mm and not a round number like 30mm?