Review of EGUsphere-2024-729

This is a review of "A comprehensive verification of the weather radar-based hail metrics POH and MESHS and a recalibration of POH using dense crowdsourced observations from Switzerland" by Kopp et al.

The study presents some useful new findings on hail detection from radar, taking advantage of Switzerland's world-leading observational network for both radar and crowdsourced reports of hail. The important findings include verifications of the often used metrics probability of hail (POH) and maximum expected severe hailstone size (MESHS). A report clustering method that uses only report data and can be used instead of radar-based filtering shows promising results for quality control of dense hail report data. MESHS shows high false alarm rates. An analysis of POH shows that the original version of POH does not correspond with a probability in Switzerland, and the authors show a new calibration of POH.

This is an excellent paper. The results are sound and will be of use to the community who use MESHS, POH and related radar products to infer hail occurrence and/or hail size at the ground. These radar-based metrics are empirical and without well-defined uncertainties. This work, which helps to quantify uncertainties on these measurements in the European domain, is therefore valuable. The paper is highly relevant to AMT.

I suggest that only minor changes and some further explanation of thresholds is required before the paper will be ready for publication.

General comments

1. **Threshold choices**: there are thresholds used in the paper (e.g. "populated areas" are those with a population of at least 100 people per km^2, reports from a user submitting more than four reports per day are removed, five reports are required for a cluster to be defined, etc) that are given without explanation or justification. I understand the difficulty in choosing a reasonable threshold but it would be helpful to include a discussion of how these thresholds were chosen and whether or not results in the article are sensitive to these choices. I note that the authors have done a great job of justifying and testing other thresholds (such as for EpsD, EpsT, and POH and MESHS thresholds in the verification, for example), so what is required for these other more minor thresholds is simply some reasoning for why the specified values

were chosen.

Specific comments

1. Line 33: "because of the nature of hail" - the authors could explicitly name here what the difficult nature is (hailstorms are relatively small and rare, essentially).

2. Line 48: Regarding advection of falling hailstones, a relevant reference may be Ackermann et al 2024 (DOI: 10.5194/amt-17-407-2024).

3. Line 104-105: It would be helpful to mention whether the COSMO runs used here were done for this study or whether operational MeteoSwiss runs were used.

4. Line 117: The definition of maximum column reflectivity (CZC) is missing the fact that the CZC is defined for a given time period (e.g. one radar set of elevation scans or one scan cycle).

5. Figure 4 caption: the Natural Earth populated areas dataset requires a reference.

6. Line 150: I read this as meaning that for a given day and a given point, if the daily maximum radar reflectivity occurred outside 6-22 UTC then the entire day is considered hail free at that point. If this is the case could some valid hailstorms be removed (if for example there was hail during the day but a larger reflectivity was recorded overnight). The authors should consider stating how often grid cells were removed in this way to show any possible impact on the results.

7. Line 161-162 and 167: "hail < 5 mm" - since the WMO classifies hail as being at least 5 mm by definition I suggest changing this to "ice particles < 5 mm in diameter".

8. Line 262: What the authors refer to here as the POH is also called the success ratio (see e.g. Roebber 2019, DOI: 10.1175/2008WAF2222159.1.)

9. Line 273: CHF is used without a definition here and should be replaced by "Swiss franc".

10. Line 281: It is not really explained why the authors focus on the Zurich region and then present the results for all Switzerland in the appendix. The authors should explain this choice.

11. Line 415: To be picky, "this probability" here refers to a matching distance of 2 km as defined by Equations 2 and 3. I suggest replacing "This probability" with "The probability of observing hail in the neighbourhood of a given point", or similar.

12. Figures 14, 15 and 16: It was not initially clear to me how the probability on the y axis of these plots was calculated. Reading back I see it is $1-FAR_{prob}$. The authors could consider making this more clear, for example by including $1-FAR_{prob}$ in the y axis labels.

Technical corrections

1. In general I think "weather radar-based" should be written "weather-radar-based".

2. Lines 108-110: The units do not need to be italicised here. This is true elsewhere in the text also, such as on line 119. There is some inconsistency in the writing of units. Units should at all times be non-italicised and with a space between the numeral and the unit.

3. Some in-line citations still have brackets (e.g. on line 21 and in the caption for Figure 2).

4. Line 180: "HRC" is introduced here without definition (the definition comes a few lines later). Acronyms should be defined when first used.

5. Table 2: There is a mix of capitalisations in the column of seasons. Also, "year" is not a season so perhaps the title should be "time period" or similar.

6. Line 384: "restraint" should be "restrained".

7. Figure 14: 'red' and 'green' curves appear as yellow and blue on my display (also affects the text around Line 394).