Review of EGUsphere-2024-729

In this paper, the authors exploit a rich dataset of crowdsourced hail reports to examine the validity of radar-based hail metrics employed by Switzerland's National Weather Service (MeteoSwiss), namely the probability of hail at the ground (POH) and the maximum expected severe hailstone size (MESHS), which are crucial for assessing hail presence and estimating its size. Through a new spatiotemporal clustering method (ST-DBSCAN) coupled with radar reflectivity to filter the reports, they conduct a meticulous analysis using various metrics such as Hit Rate, False Alarms Ratio, Critical Success Index, and Heidke Skill Score. They identify shortcomings in the current calibration of POH and suggest a recalibration based on the filtered reports.

The paper is logically structured and very well readable, although the scientificity of the language can be improved here and there. It does not describe groundbreaking science, nor does it scratch on the current state-of-the -art regarding hail detection (e.g., by dual-pol). Nevertheless, in the context of operational meteorology, it contains important information to improve the quality of the hail detection and warning at many meteorological services worldwide. Being in the field myself, I recognize the fact that this kind of publications is unfortunately largely lacking in peer-reviewed scientific literature. Therefore, I strongly recommend publication of the current paper, although some important revisions outlined below are required.

General comments

The main issues I see with the paper are listed below. Please provide a solution to all objections and indicate where the text has been adapted accordingly.

- The title is too long. Modify it to two lines maximum. Choose the main message you want to convey and formulate your title accordingly.

- Cited literature is a bit scarce and focused on Swiss publications, some of them being technical reports. I understand the topic is strongly operational, but I suggest widening the literature study a bit in the introduction and conclusion. E.g.: is single-pol hail detection still relevant in times of dual-pol radars being the standard?

- Heidke Skill Score (HSS) is introduced as one of the metrics, but it is not really integrated in the discussion. Either remove HSS from the paper completely, or integrate it better in the discussion.

- “Jumping cells” induced by the finite 5-min radar sampling and the potential high velocity of the hail-producing cells are not mentioned in the paper. Nevertheless, they can jeopardize hail statistics considerably. Or is there an advection correction applied on the radar images? Add some sentences on how this effect can influence your validation results. I suggest adding a figure of the POH 24h overview for the case you describe (June 20, 2021), to have a visual assessment of this effect (if any).

- Appendix C does not fit into the “story” the paper and reads like a separate study. For the reader there is no incentive to read it while reading the main paper. An appendix should always

support the main text, and is not intended to contain a small side-study. I recommend eliminating Appendix C and publish these results elsewhere.

- The authors make several logical choices to limit the wrong false alarms. The scientific motivation for these choices and adopted thresholds is in some cases quite weak or even absent. For example: one retains only the reports between 06:00 and 21:15 UTC, but no justification is given. For this particular choice, one could construct a histogram of all reports versus reporting time and motivate this choice better on the basis of this histogram.

- Similar comment for the definition of the ZRH region. The Swiss100 region is confined with clear criteria based on population density, while the ZRH region is quite arbitrary chosen, in fact without any link with population density. As such, some unpopulated areas, like the Zürichsee, are included in the ZRH region. The authors should either motivate this choice better, or, as an alternative, redo the statistics for the ZRH region with taking a population density criterion to confine the region (similar methodology as the Swiss100 region).

- For POH, a cubic fit is traditionally taken. Given the quite large uncertainty on the exact POH relation (see e.g. Fig. 16), I wonder if such a cubic fit is not overfitting the data points. A goodness-of-fit criterion could be added, and the results for a second order polynomial could be given accordingly. Is the cubic fit justified?

Specific comments

L34. Explain “hail hotspots”.

L65. Add the total number of inhabitants of Switzerland for reference.

L70. “suspicious reporting patterns”: are you referring to the preprocessing explained in the beginning of Section 2.5? If so, add “detailed further in the paper”.

L100. Not clear whether the POH is calculated for the individual radars first, and then composed together, *or* that the POH calculation is done on a three-dimensional composite. Please explain briefly on how the POH field from the different radars is calculated.

L104. Spatial and temporal resolution of the COSMO model used? If the temporal resolution of the model is 1h, was there a temporal interpolation to match the radar timestamps? Please explain in more detail.

L125. An additional uncertainty is introduced since a user is only able to report discrete sizes: what should a user report if the hail stone diameter he/she measures is 1.5 mm?

L132. “no one is around to report”: gives the impression that the authors assume that everyone in Switzerland has the MeteoSwiss app and will report hail for sure. Hence reformulate.

L134. Does MeteoSwiss send out location-based notifications when hail is approaching to a certain user? Or is this feature not available?

L172. In EU, it is not allowed to retain the complete history of a user ID due to GDPR regulations. Can this be done in Switzerland then?

L174. Why can’t a user send more than 4 reports per day? This can be just a very engaged user.

L186. Why 5? Did you do any sensitivity study by varying this number?

L204. 35dBZ could be lacking in the radar data by beam blocking or by “jumping” cells (see above). I miss some nuance here: there can be gaps in the radar data too.

L266. I miss a description in the main text of row 4 of Fig. 7. The “fraction of matches” is not explained in the main text. Please add a short description in the main text, and refer to the bottom row of Fig. 7.

L295. In the introduction, a maximum wind drift of 3 km is cited from literature (L48). Nevertheless 4 km is chosen here. Why? Motivate.

L334. Did you consider eliminating the highest category completely? Hail of this size is extremely rare and the reporting of such sizes even rarer. To me, the majority of these reports seems fake.

L378. Here 2 km is chosen to account for the wind drift: again another choice! Motivate.

L389. The interval (-3km – 12km) is again a quite arbitrary choice. On the contrary, you can look for both the lowest and highest ET45-H0 for which you received a hail report. Please extract these two numbers from your available data and compare it to the current choice.

L405. When there is no echotop 45, I would say ET45 is undefined and not equal to 0! Hence, the argument on L405 seems wrong to me.

L415. So POH maps based on this recalibration should be made not on radar-pixel resolution (1km), but should be provided on 2x2 km^2 maps, right?

L442. “first complete assessment of the skill of MESHS” → does this contradict L39-L40 of the introduction?

Technical comments

- Typography should be checked and improved throughout the paper, e.g., avoid writing physical units in math font.

L26. Rephrase “and they” … “and they” (split up in several sentences).

L86. Remove “and illustrate the spatio-temporal clustering method in section 2.5.1”. Keep the outline at the end of Section 1 limited to the subsections and don’t mention subsubsections.

L158-L160. Split in two sentences.

L182. “Noise, it can discover clusters of …” “Noise. It can discover clusters of …” (make two sentences).

L201. (Fig. 6)b) → (Fig. 6b)

L380. $pFAR_{prob}$ → FAR_{prob}

L394. “Red” and “green” curves? I see orange and blue…

Caption Fig. 14. “red” and “green”?