Investigating hail remote detection accuracy: A comprehensive verification of radar metrics with 150'000 crowdsourced observations over Switzerland

Jérôme Kopp[1], Alessandro Hering[2], Urs Germann[2], Olivia Martius[1]

[1]University of Bern, Switzerland
[2]Federal Office of Climatology and Meteorology MeteoSwiss, Switzerland

Hail detection and sizing using radar is a common practice and radar-based algorithms have been developed and operationally deployed in several countries. Switzerland National Weather Service (MeteoSwiss) uses two radar hail metrics: the probability of hail at the ground (POH) to assess the presence of hail, and the maximum expected severe hailstone size (MESHS) to estimate the largest hailstone diameter. Radar-based hail metrics have the advantage of extended spatial coverage and high resolution, however they don't measure hail directly on the ground. Therefore, they need to be calibrated and further verified with ground-based observations. Switzerland benefits from a large dataset of crowdsourced hail observations gathered through the reporting function of the MeteoSwiss app. Crowdsourced observations can contain wrong reports, both intended (jokes) or unintended (misuse), and have to be filtered before being used. Radar reflectivity is often used to remove reports where the maximum reflectivity is below a usual storm environment. However, this filtering method renders the observations dependent on the same radar signal used to compute hail metrics. Therefore, we test a spatio-temporal clustering method (ST-DBSCAN) based solely on the data to remove implausible reports. We then use the filtered dataset to make an extended verification of POH and MESHS in terms of Probability of Detection (POD), False Alarms Ratio (FAR), Critical Success Index (CSI) and Heidke Skill Score (HSS). We estimate the most skillful POH threshold to predict the presence of hail. We investigate the conditions leading to POH false alarms (radar signal without observation) and misses (observations without radar signal). We assess how good MESHS is compared to POH in discriminating > 2 cm hailstones, and how good MESHS is in estimating the maximum hail size on the ground for thresholds of 3 cm, 4 cm, and 6 cm. We found that POH has a good skill for hail detection with HSS reaching 0.8 (FAR < 0.2), but that MESHS struggles in estimating sizes above 3 cm (FAR > 0.5).