https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

Retrieval of the hail size number distribution from polarimetric radar data using the double-moment normalization

Matteo Guidicelli[1, 2], Alfonso Ferrone[3], Gionata Ghiggi[1], Marco Gabella[2], Urs Germann[2], and Alexis Berne[1]

[1]Environmental Remote Sensing Laboratory, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
[2]Federal Office of Meteorology and Climatology MeteoSwiss, Locarno-Monti, Switzerland
[3]Data Management and Analytics, Cineca, Bologna, Italy

Correspondence: Alexis Berne (alexis.berne@epfl.ch)

Abstract.

Estimating the distribution of hail sizes is essential for assessing potential damage to infrastructure, vehicles, and agriculture. In this study, we introduce a novel technique for retrieving the hail size number distribution (HSND) from polarimetric C-band radar data. Our approach uses a generalized additive model (GAM) to estimate two empirical moments of the HSND, which is then reconstructed using the double-moment normalization technique, exploiting the relative invariance of the normalized HSND. The model is trained using data from the Swiss automatic hailsensor network (August 2018–September 2025) across three hail-prone regions. Hundreds of polarimetric features are extracted from a high-resolution 3D radar composite combining data from the five operational, dual-pol Swiss radars. Among these, the most predictive features selected by the model include the volume of the region where the cross-correlation coefficient ρ_{HV} falls below 0.97 and the horizontal reflectivity Z_H is above 50 dBZ in a vertical column of 1 km radius, the maximum value of vertical reflectivity Z_V in a column of 1 km radius, the integral of Z_V in a column of 1 km radius. HSND estimates derived from radar show strong agreement with independent hailsensor measurements. Additional validation is performed using photogrammetric drone surveys and crowd-sourced hail reports. The radar-retrieved HSND closely matches the shape of drone-based HSND for the two events, while overestimating the number of hailstones per diameter bin due to melted hailstones prior to drone observations and to the nature of the training data used (a minimum number of 30 hailstones must have been measured by a sensor for a single event to be used). Radar-based percentile diameters of the retrieved HSND exhibit a slightly higher Pearson correlation and lower bias with crowd-sourced hail reports compared to the MESHS (Maximum Expected Severe Hail Size) product operated by MeteoSwiss. The main advantage of the presented technique is that it enables high-resolution (1 km, 5 min) retrievals of the full HSND and related features, such as kinetic energy, potentially providing valuable insights for real-time hail monitoring, nowcasting and long-term statistical assessments of hail features across Switzerland. The proposed model could be easily adapted to other countries, though the invariance of the normalized HSND outside Switzerland should be verified. Because hail is rare, further ground hail observations remain crucial for refining the HSND retrievals, ensuring a more comprehensive evaluation of the proposed approach, and properly assessing the associated uncertainty.

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

1 Introduction

Hailstorms are among the most destructive weather phenomena, causing significant damage to property, vehicles, and agricultural crops (Brown et al., 2015; Púčik et al., 2019; Warren et al., 2020; Kopp et al., 2023b). The severity of hail-related impacts is largely governed by hailstone diameter and number density, as they directly control the associated kinetic energy upon impact. However, obtaining reliable measurements of hail size and number density at large spatial and fine temporal scales remains challenging due to the localized and transient nature of hailstorms (Brimelow, 2018).

Hail features have been widely estimated using weather radar, historically relying on single-polarization systems and, more recently, on dual-polarization radar (Ryzhkov and Zrnić, 2019). Single-polarization radar infers hail presence from reflectivity at the horizontal polarization (Z_H), often in combination with temperature data. Waldvogel et al. (1979) introduced a detection criterion using echo tops and the melting layer height, while other methods, such as the vertically integrated liquid (VIL) (Amburn and Wolf, 1997) and severe hail index (SHI), improve probability estimates of hail occurrence and maximum hail size by integrating reflectivity over the vertical column (Witt et al., 1998; Foote et al., 2005). Additionally, kinetic energy flux has been estimated using empirical relationships between Z_H and hail size distributions measured by hail spectrometers (Waldvogel et al., 1978).

Despite their widespread use in operational applications (Betschart and Hering, 2012; Trefalt et al., 2023), hail climatology studies (Brook et al., 2024; Nisi et al., 2018), and damage assessments (Ackermann et al., 2024; Hohl et al., 2002; Portmann et al., 2024; Schmid et al., 2024), single-polarization radar methods have notable limitations. Distinguishing hail from large raindrops using Z_H alone is challenging because radar reflectivity depends strongly on hydrometeor size and composition. While hailstones are generally larger than raindrops and therefore often produce higher reflectivity, liquid water has a higher dielectric constant than ice, allowing large raindrops to yield Z_H values comparable to those of smaller sized hailstones. During melting, the presence of a liquid water coating further increases the effective dielectric constant of hailstones, enhancing their reflectivity relative to dry hail of the same size (Ryzhkov and Zrnić, 2019; Kaltenboeck and Ryzhkov, 2013).

Dual-polarization radar addresses some of these limitations by providing additional microphysical insights into hydrometeor properties, including shape, size, orientation, density, and concentration. Key polarimetric variables are available in addition to Z_H, such as reflectivity at vertical polarization (Z_V), differential reflectivity (Z_{DR}), copolar cross-correlation coefficient (ρ_{HV}), and specific differential phase (K_{DP}) enable hydrometeor classification (e.g. Besic et al., 2016, 2018), and facilitate the detection of hail on the ground and the differentiation of size classes (Ryzhkov et al., 2013b; Ortega et al., 2016; Forcadell et al., 2024; Klaus and Krause, 2024).

However, polarimetric radar-based hail retrievals still face significant challenges. One major issue is non-Rayleigh scattering, which occurs when the radar wavelength is comparable to the size of hailstones - that is, when hail particles are larger than approximately 2 cm at C-band. This interaction leads to resonance effects, causing increased uncertainty in polarimetric measurements, thus complicating accurate hail detection and size estimation (Ryzhkov et al., 2013a; Kaltenboeck and Ryzhkov, 2013). Longer wavelenghts are less affected, but for hail diameters above approximately 4 cm all, X, C and S-band wavelengths are concerned. Additionally, the irregular shape of hailstones further influences the hailstone's scattering properties (Cecchini

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

et al., 2022; Jiang et al., 2019; Mirkovic et al., 2022). Another challenge is that radars measure hail aloft, while wind advection can cause hailstones to fall far from the locations observed by radars. In addition, hailstones may partially or completely melt below the freezing level, altering their size and phase during descent. This horizontal displacement, along with descent time delays and microphysical evolution, makes direct comparisons between radar and ground measurements difficult.

One extensive source of ground hail data comes from crowd-sourced hail reports, which have been systematically collected in recent years (e.g. Barras et al., 2019, for Switzerland). These reports provide broad spatial coverage, making them valuable for verifying radar-based hail features (Kopp et al., 2024) and training machine-learning models for hail detection (Aregger et al., 2025; Forcadell et al., 2024). However, these reports rely on voluntary participation, and are affected by subjective biases in estimating hailstone diameters. Reporting frequency is also highly dependent on population density and daytime, leading to underrepresentation of rural areas and during night. Furthermore, crowd-sourced reports do not provide reliable information on the hail size distribution (HSD), as people tend to report only a single hail size, typically corresponding the largest observed in the vicinity.

Other traditional ground-based observations rely on hail pads, which have been widely used for studying HSDs over several decades (e.g. Manzato et al., 2020, for northern Italy). In Switzerland, past hail pad campaigns provided valuable insights into hail events (Federer et al., 1986), but no contemporary datasets exist. These deformable materials record hail impacts as indentations, which are later analyzed to derive HSDs. However, hail pads face several limitations, including overlapping indentations in high-density hail events, which introduce uncertainties in size estimates. Additionally, their manual deployment, retrieval, and analysis require substantial logistical effort and do not allow for precise timing of hail impacts.

To address hail pad limitations, automatic hailsensors have been developed to measure hail characteristics in real time. First tested by the Federal Office of Meteorology and Climatology MeteoSwiss in 2015 (Wetzel, 2018) and later expanded in 2018 into a network of 80 sensors distributed in three hail-prone regions, these instruments detect hailstone impacts through vibrations, which are converted into estimates of kinetic energy and diameter. The resulting multi-year dataset provides unprecedented opportunities to study hail size number distributions (HSNDs), a metric that differs from traditional HSDs by representing the direct number of recorded impacts rather than a distribution normalized over area or volume and time (Ferrone et al., 2024). However, due to their small surface area (0.2 m^2), individual hailsensors record a limited number of impacts per event.

Drone-based photogrammetry has recently emerged as a promising method for capturing HSNDs over larger areas of hundreds square meters. By combining photogrammetry with machine-learning models, hailstones can be automatically detected and measured (Soderholm et al., 2020; Lainer et al., 2024). The larger surveyed area allows for significantly more hailstones to be analyzed per event compared to hailsensors. However, logistical challenges, surface conditions, light availability, and the need for specialized expertise still limit its widespread application (Portmann et al., 2025).

The HSND provides essential information for analyzing hailstone diameters, impact counts, and kinetic energy, which are crucial for assessing hail-related damage. However, retrieving a full HSND from radar remains a challenging task. The double-moment normalization technique offers a promising avenue for addressing this challenge. This technique has been extensively applied to rainfall drop size distributions (DSDs) (Lee et al., 2004, 2023; Raupach et al., 2019), demonstrating that normalized

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

DSDs exhibit relative invariance across different events. The double-moment normalized DSD can thus be modeled and used in combination with two empirical moments to reconstruct the distribution. Results of reconstructed DSDs via double-moment normalization are consistent with observed DSDs across varying spatiotemporal scales. One of the key advantages of this approach for liquid precipitation is the established relationship between DSD empirical moments and radar-derived variables, enabling the retrieval of DSDs over large areas. For example, Raupach and Berne (2017) and Shin et al. (2024) successfully applied this technique to rainfall, while Raupach et al. (2019) extended it to drizzle-sized droplets.

Leveraging data from the Swiss network of automatic hailsensors, Ferrone et al. (2024) demonstrated the applicability of double-moment normalization to HSNDs. While the HSND depends on event duration, intensity, and sensor detection area, their study showed that the normalized distribution remains consistent for events in different regions and at different scales. This spatial and temporal invariance of the normalized distribution allowed the use of a constrained generalized gamma function to retrieve the full HSND using just a selected pair of empirical moments.

In this study, we develop a methodology to estimate HSNDs from polarimetric C-band radar data, using ground-based hailsensor measurements as a reference. Our approach employs generalized additive models (GAMs) to estimate two empirical moments of the HSND from polarimetric radar-derived features, which are then used in combination with the fitted generalized gamma function from Ferrone et al. (2024) to reconstruct HSNDs via (inverse) double-moment normalization. We validate our results against independent datasets, including drone surveys and crowd-sourced reports. This method enables the retrieval of HSNDs and associated features (e.g., kinetic energy) at high spatiotemporal resolution (1 km, 5 minutes) across Switzerland, overcoming the spatial limitations of ground-based measurements. While the framework could be adapted to other regions, validating the assumption of HSND invariance beyond Switzerland is necessary.

The presented model clearly shows potential to enhance our understanding of hail events and their impacts, with a high practical relevance, although retrieval robustness, evaluation, and uncertainty estimation remain constrained by the relatively small number of ground observations available. Currently, the available radar-based product for hail size is MESHS, which underpins a range of operational and research applications, including hail monitoring, nowcasting, real-time damage modeling, forecast verification, climatological analyses, and hail damage prevention. Transitioning from MESHS to HSND represents a major advancement, enabling all these applications to be redesigned with substantially more detailed and physically meaningful information.

This manuscript is organized as follows: Sect. 2 introduces the datasets used, while Sect. 3 outlines the data pre-processing steps, the radar feature extraction process and describes the GAM model used to estimate the empirical moments of the HSND. Sect. 4 presents and discusses the results, and finally, Sect. 5 summarizes the key findings and conclusions of the study.

2 Data description

This section describes the instruments and datasets used in the study (see Fig. 1): the Swiss network of polarimetric weather radars and the associated radar-based products (Sect. 2.1), as well as reference ground-based observations, including hailsensor networks, drone data, and crowd-sourced reports (Sect. 2.2).

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

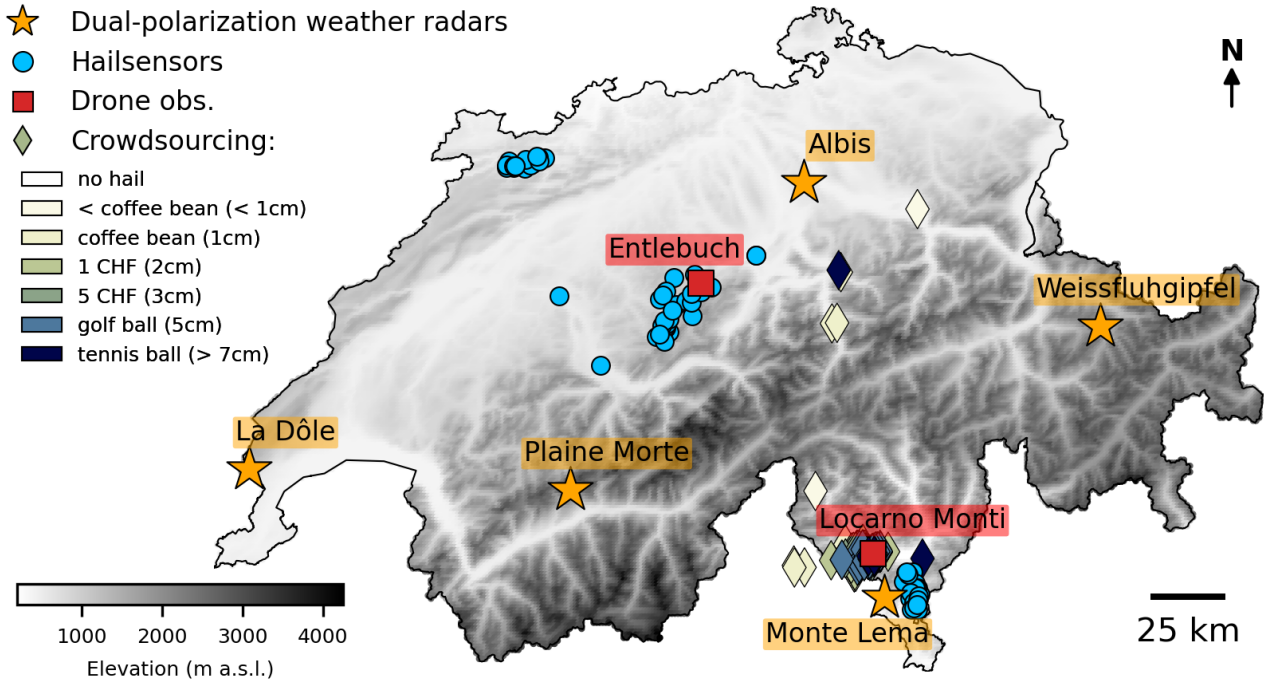

Figure 1. Geographic distribution of the observational data over Switzerland: locations of the five radars from the Rad4Alp network, hailsensor networks in three hail-prone regions (Ticino, Naps, Jura), two drone survey sites (Entlebuch and Locarno-Monti), and crowd-sourced reports collected on 25.08.2023 between 19:15 UTC and 19:35 UTC. The underlying digital elevation model has been provided by Federal Office of Topography Swisstopo (2021).

2.1 Radar-based data

The Swiss radar network, Rad4Alp, operated by MeteoSwiss, consists of five polarimetric C-band weather radars that provide nationwide coverage and high-resolution precipitation measurements (Germann et al., 2022). The network employs volume-scanning techniques, capturing three-dimensional observations of precipitation systems at frequent intervals at 20 different elevation angles (see Fig. A1a). However, deploying and operating weather radars in the complex terrain of the Swiss Alps presents numerous challenges. The rugged topography complicates optimal radar placement, infrastructure construction, and ongoing maintenance, particularly at remote, high-altitude locations. Mountainous terrain also introduces signal disruptions, including beam shielding and ground clutter, which can significantly affect data quality. The radar data utilized in this study comprise quality-checked measurements that have been corrected for ground clutter (Joss and Lee, 1995).

Several radar-derived datasets from the Rad4Alp network are used in this study: polarimetric variables (Sect. 2.1.1), radar-based hail products (Sect. 2.1.2), and a thunderstorm tracking product (Sect. 2.1.3).

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

2.1.1 Polarimetric radar data

Polarimetric radar observations provide valuable insights into the microphysical characteristics of precipitation, aiding in the identification and classification of hydrometeors, including hail (Besic et al., 2016; Kumjian, 2013). The key polarimetric variables analyzed in this study include:

- Reflectivity at horizontal polarization (Z_H): The power returned to the radar from horizontally polarized transmitted pulses, which is influenced by hydrometeor water content, size, density, and type, and can be used in combination with other measurements to infer these properties.
- Reflectivity at vertical polarization (Z_V): The power returned to the radar from vertically polarized transmitted pulses, complementing Z_H in polarimetric analysis.
- Differential reflectivity (Z_{DR}): The log-transformed ratio between horizontal and vertical reflectivity (Z_H/Z_V), indicating particle shape and orientation. High values often indicate large, oblate melting hailstones, while values near zero are indicative of spherical particles, such as small raindrops. Negative values can indicate tumbling dry hailstones, but may also result from differential path-integrated attenuation along the radar beam.
- Specific differential phase shift on propagation (K_{DP}): The phase shift difference between horizontally and vertically polarized waves per unit distance. It is proportional to the concentration and shape of liquid hydrometeors and is useful for distinguishing heavy rain and melting hail. K_{DP} is computed from total differential phase shift (Φ_{DP}) along the radar path using a piecewise least square method, as done in Wolfensberger et al. (2021).
- Copolar cross-correlation coefficient (ρ_{HV}): Describes the correlation between horizontally and vertically polarized returns, helping differentiate between homogeneous precipitation (e.g., rain) and mixed-phase or irregularly shaped particles (e.g., hail). Values range from 0 for uncorrelated polarizations to 1 for fully correlated polarizations in case of homogeneous particle shapes in the radar sampling volume.

2.1.2 Probability of Hail and Maximum Expected Severe Hail Size

To estimate hail occurrence and severity, MeteoSwiss operationally employs two single-polarization radar-based products:

- Probability of Hail (POH; Foote et al., 2005), derived from the Waldvogel criteria (Waldvogel et al., 1979) and subsequently refined using a probabilistic approach (Witt et al., 1998). POH provides an estimate of the probability that hail of any size is reaching the surface.
- Maximum Expected Severe Hail Size (MESHS; Joe et al., 2004; Treloar, 1998; Trefalt et al., 2023), which estimates the largest hailstone diameter at ground level, specifically for hailstones exceeding 2 cm.

Both POH and MESHS are two-dimensional gridded Cartesian products. Their computation is based on the altitude difference between the echo-top height—defined as the highest altitude at which a given radar reflectivity threshold is detected—and

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

the height of the 0°C isotherm. This height difference serves as a proxy for the region where hailstones can grow by accreting supercooled water droplets (Allen et al., 2020). POH is derived using a 45 dBZ echo-top height, whereas MESHS is based on a 50 dBZ echo-top height (see Nisi et al., 2020). The height of the 0°C isotherm is obtained from MeteoSwiss numerical weather prediction model analyses: COSMO-1 until 2024 and ICON-CH1-EPS thereafter (MeteoSwiss, 2024).

In this study, POH is used as a reference for hail detection, while MESHS serves as a benchmark for comparison with diameter percentiles from our radar-retrieved HSND.

2.1.3 Thunderstorm Radar Tracking

The Thunderstorm Radar Tracking (TRT) algorithm (Hering et al., 2004) is an operational product developed by MeteoSwiss to monitor and track convective storms in real time. TRT identifies and follows individual thunderstorm cells based on radar reflectivity and motion estimation.

The algorithm detects convective cells by applying a set of reflectivity-based thresholds and associates them with storm objects. These objects are then tracked over time using an advection scheme that estimates their displacement. TRT also classifies storm severity based on reflectivity and storm-top height, allowing for early warnings of intense thunderstorms and potential hail-producing cells.

For this study, TRT data are used to analyze the temporal and spatial evolution of hailstorms. Furthermore, information on the speed and direction of thunderstorm cells are used to extrapolate hailsensor observations over space prior to the training process (see Appendix C).

2.2 Ground-based observations

Three different ground-based observations are used in this study and are presented hereafter: HSND from automatic hailsensors (Sect. 2.2.1), HSND from drone-based photogrammetry (Sect. 2.2.2), and hail sizes from crowd-sourced reports (Sect. 2.2.3). A summary of the data volume is provided in Table 1, while a more detailed overview of the analyzed hail events is given in Table A1.

2.2.1 Hailsensor HSND measurements and related moments

Between 2018 and 2020, 80 automatic hailsensors were deployed across three hail-prone regions in Switzerland: Jura (15 sensors), Napf (38 sensors), and Ticino (27 sensors). These sensors detect hail impacts on a Makrolon disc (0.2 m^2) and estimate equivalent hailstone diameters by converting oscillations into kinetic energy, assuming a spherical shape and a constant drag coefficient.

The dataset used in this study spans from August 2018, to September 2025. To ensure that measurements correspond to actual storm conditions, we apply a radar reflectivity threshold of > 35 dBZ, following Kopp et al. (2023a). Additionally, we filter out particles smaller than 5 mm in diameter, following Ferrone et al. (2024) and in accordance with the definition of hail provided in the Glossary of Meteorology. Particles in the 1–5 mm range are typically classified as graupel and are not

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

Table 1. Summary of the reference data volumes analyzed in this study. Crowd-sourced observations were retained for the full-day period only for the three events reporting the largest number of hail occurrences (excluding events with hailsensor observations). For the remaining cases, crowd-sourced data were restricted to an approximately one-hour interval centered on the first and last hailstone detections recorded by hail sensors during the event day. For the Locarno-Monti event observed by drone, three experts performed manual annotations that served as training reference for the HSND-retrieval model. Consequently, the three models retrieved different numbers of hailstones. The reported range for the drone data reflects the spread of values across the three models. A detailed description of the considered events is reported in Table A1.

Type of Data	Years	Nr. days	Nr. hailstones
Hailsensors	2018-2025	35	9772
Drone	2021; 2022	2	22298-22631
Crowd-sourcing	2018-2025	37	58209

considered hail in this study. For temporal aggregation, impacts recorded by the same sensor are considered part of a single hail event if they occur within a 20-minute window. Finally, a minimum of 30 impacts during the event must be measured by a hailsensor to be included in the analyzed dataset (for further details on sensor data processing, see Sect. 2.1 in Ferrone et al. (2024)). This results in a dataset composed by 35 different days with HSNDs (see Table 1), typically observed by multiple hailsensors during the same day.

Our analysis begins with the equivalent hailstone diameters (D) recorded by the hailsensors, where $N(D)$ represents the number of impacts measured at a given diameter D. To visualize the HSND, the continuous range of diameter values is discretized into bins to compute $N(D)$. The choice of bin size is arbitrary, as hailstone sizes theoretically vary continuously. Under this assumption, $N(D)dD$ can be interpreted as the number of impacts recorded within an infinitesimally small interval $[D, D + dD]$. Based on this definition, the moment of order p of the distribution is given by:

$$M_p = \int_0^\infty N(D) D^p dD. \tag{1}$$

In practice, for an event with a finite number of impacts, this integral is approximated by summing over all measured diameters, each raised to the power of p:

$$M_p = \sum_{i=1}^{N} D_i^p \tag{2}$$

Based on the analysis presented in Ferrone et al. (2024), moments of order 2 and 4 were selected for double-moment normalization, as they minimize the impact of missing diameters at the extremes of the HSND. Their study demonstrates that low-order moments (e.g., 1 and 2) are primarily affected by the exclusion of small hailstones, while high-order moments (e.g., 5 and 6) are more influenced by the removal of large hailstones, leading to greater bias and variability. Moments 2 and 4 strike

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

an optimal balance, with moment 2 exhibiting lower bias and moment 4 providing better correlation. This choice also aligns with the double-moment normalization guideline of using non-consecutive orders to ensure robust normalization.

In this study, we adopt the same two moment orders (2 and 4) for radar-based retrieval of the HSND. Other moment pairs were tested but did not provide better results in comparison to the validation data.

To enhance comparability with radar-based data and improve spatial representativeness, hailsensor observations have been extrapolated over space, along the transect followed by the thunderstorm cell, as described in Appendix C. The extrapolated hailsensor data serve as target observations for model training (Sect. 3.4.1) and are further utilized to evaluate radar-based estimates of the two moments through a leave-one-event-out cross-validation (Sect. 4.2).

2.2.2 Drone-based HSND retrievals

HSND estimates obtained for two events using drone photogrammetry coupled with machine learning (Lainer et al., 2024; Portmann et al., 2025) serve as an independent reference for testing the HSND retrieved by our radar-based approach.

The first event occurred near Entlebuch (Canton Lucerne, Switzerland) on 20 June 2021 over a field of approximately 750 m^2. The second event was recorded in Locarno (Canton Ticino, Switzerland) on 28 June 2022, covering roughly 194 m^2.

For the first event, the drone survey was conducted approximately nine minutes after the hailstorm onset, and multiple flights were performed to estimate hailstone melting rates, which were found to be approximately 0.5 mm per minute. In contrast, for the second event, adverse weather conditions - strong wind gusts and continued precipitation following the hailstorm - delayed the survey by nearly 20 minutes after the last hailstone impact detected by a nearby automatic hailsensor (see Table A1). Due to sensor malfunctions, only a single flight was completed, and no melting-rate information is available. Further details on this event and on the challenges associated with drone-based hail detection are provided in Portmann et al. (2025).

For both events, we analyzed the minimum and maximum hailstone diameters projected onto the image plane, following the methodology of Lainer et al. (2024). These diameters were used as the axes of an ellipse to compute the equivalent diameter, defined as the diameter of a circle with the same area as the ellipse. The resulting HSNDs, based on equivalent diameters, were compared with the radar-based HSND retrieved by our model (Sect. 4.3).

For the second event, Portmann et al. (2025) produced three independent HSND retrieval models, trained using manual hail-size annotations from three experts. We therefore used all three HSNDs in our comparison with the radar-retrieved HSND.

2.2.3 Crowd-sourced reports

Crowd-sourced hail reports are collected since May 2015 through the MeteoSwiss mobile app, providing an important source of ground-truth data for hail size assessment in Switzerland (Barras et al., 2019). Until September 2017, users reported hail (or the absence of hail) by selecting one of five predefined categories, while afterwards the number of classes was extended to the following seven categories: no hail, $<$1 cm (smaller than a coffee bean), 1 cm (coffee bean), 2 cm (1-franc coin), 3 cm (5-franc coin), 5 cm (golf ball), and $>$7 cm (tennis ball). These real-time reports have already been utilized to verify radar-based hail products such as POH and MESHS (see Kopp et al., 2024).

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

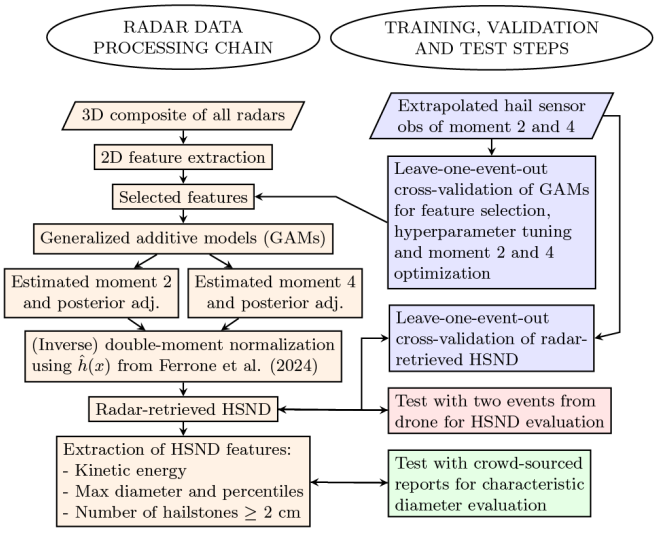

Figure 2. Radar data processing chain (left), and model training, and evaluation workflow (right). Radar-based steps are highlighted in orange, hailsensor-related steps in blue, drone-based steps in red, and crowd-sourcing steps in green.

In order to reduce the impact of the reports uncertainty on our results, we rely on an aggregated classification rather than the seven categories currently available, as similarly done previously by Aregger et al. (2025). Hail is grouped into four size classes and an average diameter (D_{avg}) per class is considered in the analyses: (1) $D \leq 15$ mm ($D_{avg} \simeq 7.5$ mm), potentially including graupel, (2) $15 > D \leq 27$ ($D_{avg} \simeq 22.5$ mm), (3) $27 > D \leq 37$ ($D_{avg} \simeq 32$ mm) and (4) $D > 37$ mm ($D_{avg} \simeq 45$ mm). In addition, hail reports are subjected to several plausibility checks, following the procedures outlined by Barras et al. (2019) and Kopp et al. (2024). Reports are also validated against radar reflectivity: to be considered valid, the radar must detect a minimum reflectivity of 35 dBZ within 4 km of the report location and in the time window within 15 minutes before or after the report.

In this study, crowd-sourced hail reports were used to validate the characteristic diameters (e.g. high percentiles) derived from our radar-based HSND model (Sect. 4.4). We considered all reports from Switzerland on days when at least one hailsensor recorded a minimum of 30 hailstones. In addition, we included the two hail events with the largest number of reported diameters when no hailsensor recorded at least 30 hailstones (28 June 2021 and 12 July 2023). Details of all considered events are provided in Table A1.

3 Methods for radar-based HSND retrieval

This section outlines the methods used in this study (cf. Fig.2), including the merging of the five radars in a unique 3D composite (Sect. 3.1), the radar feature extraction (Sect. 3.2), the pre-processing of hailsensor data prior to model training (Sect. 3.3), the radar-based HSND retrieval (Sect. 3.4), and the definition of error metrics used throughout the analysis (Sect. 3.5).

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

3.1 3D radar composite

In order to provide a more complete and robust representation of the storm structure, we combined overlapping measurements from the five Swiss radars in a unique 3D radar composite.

For each radar, a 3D composite is first created by interpolating its measurements onto a regular 3D grid of 250 m resolution, chosen as a compromise between computational efficiency and high spatial detail that could potentially explain local ground variability in observed HSNDs. The individual radar composites are then merged into a single unified 3D product, retaining for each voxel the measurement from the radar with the highest horizontal reflectivity, which is assumed to be least affected by attenuation. Details on the procedure are reported in Appendix B, where Fig. A2 illustrates the selected radar in the final 3D composite for the event of 25.08.2023 between 19:20 and 19:25 UTC, over the hailsensor located in Locarno-Monti (Ticino). For the same event, examples of vertical cross-sections for other polarimetric variables are reported in Appendix B3.

An overview of the event over its full duration, as recorded by the ground-based hailsensor, is shown in Fig. 3. The figure highlights how the final composite yields a higher 50-dBZ echo-top in horizontal reflectivity than the composite generated from the nearest radar to the hailsensor (Monte Lema radar located at approximately 17 km from the hailsensor). This occurs because, beyond the first part of the storm observed by Monte Lema, other radars are selected (see Fig. A2). Incorporating these alternative viewing angles reduces the impact of attenuation - which can be significant due to intense rainfall and hail - by enabling the storm to be scanned from multiple directions.

3.2 Radar feature Extraction

Hundreds of features are extracted from the unified radar composite (Sect. 3.1), with the final model utilizing only four of them (Sect. 4.1), as those provided the best performance based on cross-validation against hailsensor data (Sect. 3.5).

The features investigated can be grouped into three categories:

1. Echo-top features: these include the altitudes of the 40, 45, and 50 dBZ levels of both horizontal and vertical reflectivity, along with their differences from the altitude of the freezing level, obtained from COSMO-1 until 2024 and ICON-CH1-EPS thereafter.

2. Vertical profile features: for each vertical column, summary statistics (mean, standard deviation, maximum, minimum, percentiles, and interquartile range) of the polarimetric variables Z_H, Z_V, Z_{DR}, ρ_{HV}, and K_{DP} are computed. Various vertical weighting functions were also tested but did not improve model performance and were therefore not retained.

3. Geometrical features: these describe the size and shape of specific regions within the atmosphere. Examples include the volume and maximum horizontal and vertical extent of areas where polarimetric variables exceed or fall below certain thresholds, such as the volume where $\rho_{HV} < 0.97$ (or 0.95, 0.93), and/or where Z_H exceeds 50 dBZ (or 45, 40, 35 dBZ), among others.

All features are computed on a 2D grid with a resolution of 250 m, matching that of the radar composite. To account for the advection of hailstones as observed by the radar, all features are aggregated at multiple spatial scales. Depending on the feature

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

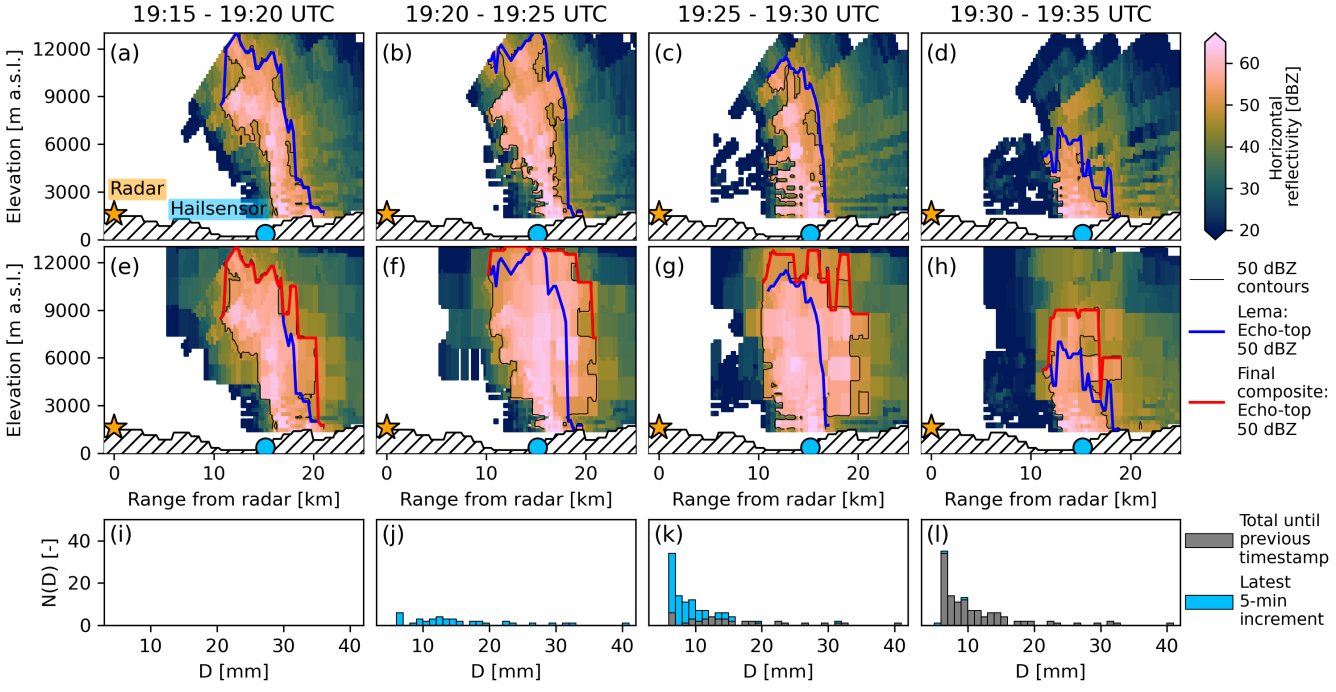

Figure 3. Vertical cross-section of horizontal reflectivity on 25.08.2023 over Locarno area along azimuth 345.87° from Monte Lema radar between 19:15-19:20 UTC (a), 19:20-19:25 (b), 19:25-19:30 (c), 19:30-19:35 (d); similarly for the 3D radar composite combining the five radars of the Swiss network (e, f, g, h); hail size distribution measured by the hailsensor located in Locarno-Monti (i, j, k, l). Only the observed radar data in a 10 km radius from the hailsensor is shown.

type, these features are aggregated by computing the sum, average, maximum, or minimum over spatial windows ranging from 250 m to 9 km. This approach also enables the representation of mesoscale structures, such as the overall volume of a thunderstorm or the horizontal extent of a hail core.

Furthermore, each geometrical and vertical profile feature is also computed within subregions defined by thresholds of Z_H (20, 30, 40, and 50 dBZ), allowing the analysis to focus specifically on areas associated with the potential hail core, where reflectivity values are elevated.

3.3 Hailsensor data pre-processing

The very small sensing area of individual hailsensors (0.2 m^2) limits their direct comparability with radar observations, which sample much larger volumes. Considering a 1° beamwidth and a 500 m range resolution, at 15 km range a radar gate has a cross-section perpendicular to the radar beam of $A_\perp \simeq 5.4 \times 10^4$ m^2 and a sampling volume of $V \simeq 2.7 \times 10^7$ m^3; at 100 km, $A_\perp \simeq 2.4 \times 10^6$ m^2 and $V \simeq 1.2 \times 10^9$ m^3. To mitigate this scale mismatch, hailsensor measurements were spatially extrapolated along the storm trajectory to generate target fields often covering multiple radar grid cells, using information from the TRT product on cell speed and direction. This approach assumes that the storm structure remains approximately stationary in a

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

storm-relative frame over a 5-minute window, allowing the temporal evolution observed at a fixed sensor to represent the spatial variability across the storm segment. The extrapolated fields are then aggregated over the 5-minute radar scan interval, accounting for a potential time offset corresponding to the travel time of hailstones from the cloud to the ground. Details of the procedure are reported in Appendix C.

Key parameters of this procedure—including the time offset, the maximum extrapolation distance, the time window used to aggregate hailsensor observations prior to extrapolation, and the minimum number of hailstones measured per 5-minute interval, were optimized to maximize the Spearman rank correlation between the most predictive radar features (see Sect. 3.2) and the spatially extrapolated empirical moments from the hailsensors. Details of this sensitivity study are provided in Appendix D.

The added value of the spatial extrapolation was found to be limited, with meaningful contributions extending to a maximum distance of roughly 200 m.

3.4 Radar-based retrieval of HSND

In this section, we describe the radar-based generalized additive models (GAMs) used to estimate moments 2 and 4 of the HSND (Sect. 3.4.1), the cross-validation scheme employed to find the optimal set of features and model hyperparameters (Sect. 3.4.2), the a posteriori bias correction of the GAM-estimated moments (Sect. 3.4.3), the combination of the (adjusted) GAMs predicted moments with a (constant) shape function to retrieve the HSND (Sect. 3.4.4), and the subsequent extraction of HSND's features (Sect. 3.4.5).

3.4.1 Estimation of moment 2 and 4

Separate GAMs were trained for the two moments, using the same set of features as predictors (see Sect. 3.4.2). GAMs are flexible, semi-parametric regression models that express the target variable as the sum of smooth functions applied to each predictor, allowing for non-linear relationships while incorporating physically meaningful constraints.

For both moments, the GAM uses four predictor features, each represented by a third-order monotonic cubic spline with five basis functions and a regularization strength (λ) of 0.6 (see Sect. 3.4.2). The final estimate is obtained as the sum of the smooth contributions from all predictors (Eq. 3).

$$\hat{M}_k = \beta_0^{(k)} + \sum_{j=1}^{4} s_j^{(k)}(x_j), \qquad k \in \{2,4\}, \tag{3}$$

where $\hat{M}_k$ denotes the estimated hail size distribution moment of order k, $\beta_0^{(k)}$ is the intercept term, and x_j ($j = 1,\ldots,4$) are the radar-derived predictor features. The functions $s_j^{(k)}(\cdot)$ are smooth monotonic spline functions that model the potentially non-linear contribution of each predictor to the estimated moment. These configurations were selected to balance flexibility and robustness while reflecting the physical behavior of the radar observations.

As a note, we also tested a unique model to mimic the GAM behavior but estimating the two moments together, and optimizing them by minimizing the error between the resulting retrieved HSND and the observed HSND directly during the

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

training process. However, this model setup did not lead to a better performance against the different data sources used for validation in this study.

3.4.2 Cross-validation scheme and feature selection

A leave-one-event-out cross-validation was performed to select features that generalize well to unseen hail events. For each iteration, the model was trained on all hail events except those recorded by hailsensors on the day of the withheld validation event.

This procedure was applied across multiple combinations of feature sets and GAM hyperparameters: regularization strength (0.3, 0.6, 1.0) and number of splines per feature (4, 5, 6, 7, 8). Feature selection began with the predictor showing the highest cross-validated Spearman correlation with the (extrapolated) moment from the hailsensor data. Additional features were then tested sequentially, following their ranking by Spearman correlation with the target variables (moment 2 or 4). A feature was added only if it increased the cross-validated Pearson correlation between model output and observations relative to the previous feature set for at least one among the two moments. When a feature was selected, the same regularization and spline complexity were applied to that feature in both GAMs. The process stopped once no remaining candidate feature improved the cross-validated correlation.

We also tested an alternative, more flexible approach in which each moment's GAM could have different features and different hyperparameters per feature. Although this method produced slightly better cross-validation scores, it was not used since it performed worse on independent test datasets (drone and crowd-sourced hail reports), possibly indicating overfitting.

Because of the limited dataset size, it was not feasible to reserve an additional independent hailsensor test set. Nevertheless, the cross-validation results show that the selected features and hyperparameters are robust when applied to days not included in training.

The following four radar-derived features (in order of inclusion) provided the best performance:

- Volume of the region where $\rho_{HV} < 0.97$ and $Z_H > 50$ dBZ within a 1 km radius
- Sum of Z_V (linear units) in the vertical column within a 1 km radius where $Z_V >$30 dBZ
- Maximum Z_V in the vertical column within a 1 km radius
- Volume of the region where $\rho_{HV} < 0.97$ and $Z_H > 50$ dBZ within a 1 km radius, above the freezing level

These features are discussed in Sect. 4.1.

Note that all selected features are aggregated within a 1 km radius. For this reason, the GAM is finally applied on the same 1 km grid used by other MeteoSwiss products, such as MESHS and POH, without loss of information.

3.4.3 A posteriori bias adjustment

Model predictions of the moments can show systematic biases relative to the reference hailsensor observations. This behavior is common in statistical models trained to minimize a mean error metric (such as the mean squared error used here), which tends

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

to smooth extremes and can therefore introduce bias in the predicted moments. In addition, the limited amount of data for large hail diameters prevents the GAM from constructing well-adapted spline functions for those diameter classes; cross-validation would otherwise tend to result in large errors due to the lack of representative observations.

Thus, we apply bias correction at the level of the predicted moments used to retrieve the HSND in the following step (Sect. 3.4.4). Let $\hat{M}_a$ and $\hat{M}_b$ denote the predicted logarithmic moments of orders a and b, and M_a and M_b the corresponding observed values.

Each predicted moment $\hat{M}_k$ is mapped to the empirical distribution of the observed moment M_k via quantile matching. For each $\hat{M}_k$, we compute its percentile within the predicted distribution and assign the corresponding quantile of the observed distribution:

$$\tilde{M}_k = Q_{M_k}\big(F_{\hat{M}_k}(\hat{M}_k)\big), \tag{4}$$

where $F_{\hat{M}}$ is the empirical cumulative distribution function of the predictions and Q_{M_k} is the quantile function of the observed moment values of order k.

To further stabilize the correction, we fit a linear regression model between predicted moments and their quantile-mapped counterparts, and use this model to obtain adjusted predictions.

The corrected histograms are finally retrieved from the adjusted moments using Eq.5

Another approach, using a log-ratio spline correction applied directly in diameter space was also tested but was proven to be less stable to the small amount of observational data for large diameter classes.

3.4.4 Inverse double moment normalization

Building on the methodology introduced by Lee et al. (2004) and adapted by Raupach et al. (2019), this study reconstructs the HSND using the double moment normalization technique, which relies on two empirical moments, combined with a constant shape function representing the normalized distribution. In this study, we use the fitted normalized distribution, $\hat{h}(x)$, from Ferrone et al. (2024). Thus, the general formula for HSND reconstruction is the following:

$$\hat{N}(D) = \frac{M_i^{(j+1)/(j-i)}}{M_j^{(i+1)/(j-i)}}\hat{h}(x), \tag{5}$$

where $x = \frac{M_i^{1/(j-i)}}{M_j^{1/(j-i)}}D$, and M denotes the observed moment, the GAM-estimated moment ($\hat{M}$), or the a posteriori adjusted moment ($\tilde{M}$). In this study, we recall that moments 2 (M_2) and 4 (M_4) are selected as the orders i and j based on the findings of Ferrone et al. (2024), which identified this pair as the least affected by sampling effects. Furthermore, they encompass the range of moments that are most relevant for practical applications. In fact, moment 4 represents the kinetic energy which is particularly relevant for crop damages. The functional shape of $\hat{h}(x)$ is provided by a generalized gamma distribution with parameters μ and c fixed at 36 and 0.41 respectively, enabling the approximation of HSNDs from events recorded under varying conditions and durations (Ferrone et al., 2024). This approach facilitates direct comparisons across different datasets

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

by leveraging the unitless normalized diameter x and the normalized distribution $\hat{h}(x)$ to derive a consistent representation of HSNDs.

3.4.5 Extracting features from the retrieved HSND

Once the full HSND is retrieved, various hail features can be computed. In this study, we focus on quantities that emphasize the upper tail of the distribution, as larger hailstones are typically responsible for the most significant damages.

The retrieved features include:

- High percentiles of the diameter distribution, which describe the largest hail sizes within the HSND. Specifically, we compute selected percentiles (e.g., the 99.9th (D99.9), 99.99th (D99.99), 99.999th (D99.999), 99.9999th (D99.9999)) to estimate typical sizes of the largest hailstones.
- Maximum diameter, defined as the largest diameter bin of the HSND with at least one hailstone (D_{max}).
- Ratios between consecutive moments of the HSND (e.g., moments 6 and 5).
- Number above a threshold, defined as the total number of hailstones exceeding a given diameter (e.g., 2 cm ($N_{t,\ D>2cm}$)), providing insight into the abundance of potentially damaging hail.
- Total kinetic energy, which quantifies the total mechanical energy carried by all hailstones over the considered area and time interval and serves as an indicator of their destructive potential. The total kinetic energy (TKE) is computed by discretizing the HSND into diameter bins and summing the kinetic energy of the hailstones in each bin:

$$TKE = \sum_i \frac{1}{2} m_i v_i^2 N_i, \quad [\mathrm{J}] \tag{6}$$

where m_i is the mass of a hailstone in diameter bin i, v_i its terminal fall speed, and N_i the number of hailstones in that bin within the considered area and time interval. The mass is calculated assuming spherical hailstones with an ice density of $\rho_{\mathrm{ice}} = 900\ \mathrm{kg\,m^{-3}}$:

$$m_i = \frac{4}{3}\pi \left(\frac{D_i}{2}\right)^3 \rho_{\mathrm{ice}}, \quad [\mathrm{kg}] \tag{7}$$

The terminal fall speed is estimated following Manzato et al. (2020):

$$v_i = \frac{1}{100}\sqrt{\frac{4\mathrm{g}\,\rho_{\mathrm{ice}} D_i}{3 C_D \rho_{\mathrm{dry}}}} \simeq 14 D_i^{0.5}, \quad [\mathrm{m\,s^{-1}}] \tag{8}$$

with $C_D \simeq 0.6$, $\mathrm{g} \simeq 9.81\ \mathrm{m\,s^{-2}}$, $\rho_{\mathrm{dry}} = 1\ \mathrm{kg\,m^{-3}}$, and D_i expressed in meters.

The total kinetic energy TKE represents the energy of all hailstones falling over the area considered at the final GAM spatial resolution (1 $\mathrm{km^2}$) during the five-minute radar observation interval. Unlike the commonly used kinetic energy flux, which is expressed per unit area and time, TKE retains the integral energy over the specified spatial and temporal domain (through N_i) and is therefore directly proportional to both the affected area and event duration.

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

3.5 Performance assessment and error metrics

In this section, we describe the methods used to test the GAM and the related error metrics in comparison with the three reference datasets: hailsensor, drone and crowd-sourced reports.

3.5.1 Validation with hailsensor data

A leave-one-event-out cross-validation was performed and the results were evaluated (i) by means of root mean square error (RMSE), Pearson (PCC) and Spearman (SCC) correlation coefficient against the moments computed from the hailsensor observations, and (ii) in the diameter space by comparing the radar-retrieved HSND with the HSND from hailsensor observations.

3.5.2 Testing the retrieved HSND with drone observations

The radar-retrieved HSND is downscaled to the area covered by the drone survey, assuming a spatially uniform HSND within the radar grid cell. The downscaled HSND is computed as

$$\mathrm{HSND}_{\text{drone area}} = \mathrm{HSND}_{\text{radar area}} \times \frac{\text{drone area}}{10^6\ \mathrm{m}^2}, \tag{9}$$

where $\mathrm{HSND}_{\text{radar area}}$ refers to the HSND in the full radar grid cell (1 km^2), and $\mathrm{HSND}_{\text{drone area}}$ is the corresponding HSND for the drone-surveyed area (expressed in m^2).

For the event observed on 28.06.2022, a hailsensor was located 30 m from the drone survey area. The HSND retrieved via double-moment normalization from hailstone diameters measured by the hailsensor is thus similarly upscaled to the drone-survey area:

$$\mathrm{HSND}_{\text{drone area}} = \mathrm{HSND}_{\text{hs area}} \times \frac{\text{drone area}}{0.2\ \mathrm{m}^2}, \tag{10}$$

where $\mathrm{HSND}_{\text{hs area}}$ refers to the HSND derived from the moments 2 and 4 observed by the hailsensor (0.2 m^2). The two downscaling approaches differ in the original area: 1 km^2 for radar versus 0.2 m^2 for the hailsensor.

The hailsensor recorded 12 hailstones, with a maximum observed diameter of 1.4 cm. This hailsensor-based HSND is used only to support the comparison with the drone observations and was not included in the GAM training dataset, as it contains fewer than 30 hailstones. Consequently, these observations are not reported in Table A1.

Both drone-observed events lasted approximately 10 min (14:28–14:38 UTC on 20.06.2021 and 07:52–07:59 UTC on 28.06.2022). Therefore, the radar-retrieved HSND is accumulated over the two consecutive 5-min radar time steps closest to the storm in time. Since drone surveys are conducted only after the storm has ended, a melting correction assuming a constant rate is applied to the radar-retrieved HSND. Following Lainer et al. (2024), a melting rate of 0.5 mm min^{-1} is adopted, estimated from multiple drone flights over the target area on 20.06.2021. The same rate is applied to the 28.06.2022 event, although it may not fully reflect different ground and atmospheric conditions. The first drone survey (20.06.2021) was conducted approximately 2.5 min after the storm ended, whereas for the second event (28.06.2022) the survey took place about 20 min after the storm.

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

The radar-retrieved HSND, both with and without melting correction, is quantitatively compared to the drone-retrieved HSND in terms of moments 2 and 4. For the event on 28.06.2022, the comparison additionally includes the upscaled hailsensor-derived HSND.

3.5.3 Testing the retrieved percentile diameter with crowd-sourced reports

The selected days for this comparison included those on which at least one hailsensor recorded a minimum of 30 impacts, and the corresponding available timestamps are reported in Table A1. Additionally, for two days with the highest number of crowd-sourced reports - but without corresponding hailsensor data used in model training - all timestamps within the 24-hour period were included in this analysis. One of these two days is further examined in Sect. 4.5.

All reported hail diameters within 10 minutes of the start of the 5-minute radar scan period were considered to account for potential delays in reporting via the app, following a methodology similar to that of Aregger et al. (2025). For each analyzed day, the reported hail diameters were then mapped onto the 2D radar grid and spatially aggregated over multiple radii (3, 5, 7, 9, and 11 km) by computing the 75th and 90th percentiles of the reports for that day. For each aggregation radius, a minimum number of reports was required to reduce the risk that a small sample poorly represents the hail characteristics of the area (3 km: $\geq$6 reports; 5 km: $\geq$10; 7 km: $\geq$14; 9 km: $\geq$18; 11 km: $\geq$22).

Similarly, MESHS and radar-retrieved HSND high-percentile diameters (D99.9, D99.99, D99.999, D99.9999), as well as $N_{\mathrm{t},D>2,\mathrm{cm}}$ and TKE, were aggregated over the same spatial radii. For all features, the maximum value within the aggregation radius was retained.

4 Results and discussion

In the following, we present and discuss results regarding the selected radar features (Sect. 4.1), the evaluation of the GAM model against hailsensor data (Sect. 4.2), the evaluation of the radar-retrieved HSND against drone-retrieved HSND (Sect. 4.3), the comparison of the radar-retrieved characteristic diameters against crowd-sourced reports (Sect. 4.4), with a particular focus on the event of 28.06.2021 (Sect. 4.5). Finally, the challenges encountered due to the limited and different data sources and the subsequent limitations of the study are discussed in Sect. 4.6.

4.1 Selected radar features

As already mentioned, the group of features that yielded the best model performance in cross-validation against the hailsensor data consists of four predictors:

- Volume of the region where $\rho_{HV} < 0.97$ and $Z_H > 50$ dBZ within a 1 km radius.
- Sum of Z_V (linear units) in the vertical column over a 1 km radius.
- Maximum of Z_V in the vertical column over a 1 km radius, where $Z_V > 30$ dBZ.

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

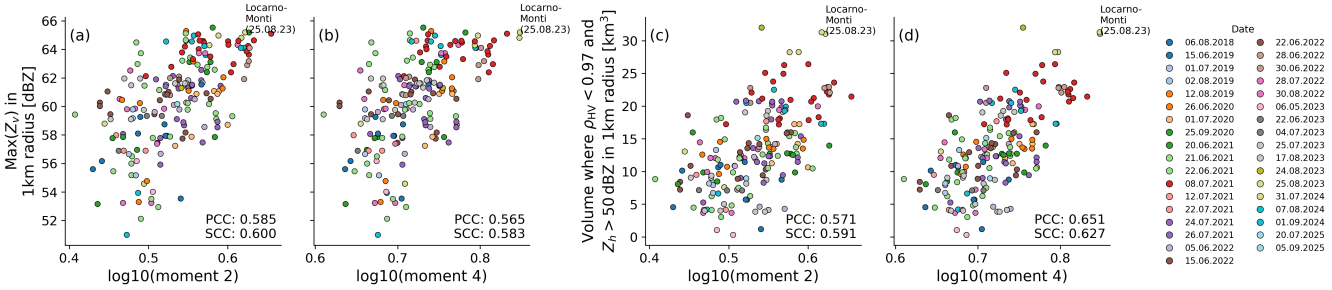

Figure 4. Relationship between the two selected radar-derived features having the highest Spearman correlation with moment 2 and 4, respectively. The hailsensor data were aggregated over 5-minute intervals for all hail events used in the study with 30 s offset with respect to the radar observations (see Appendix D). On the left, echo-top height of 50 dBZ vertical reflectivity (Z_v) as a function of moment 2 (a) and moment 4 (b). On the right, volume where $\rho_{\mathrm{HV}} < 0.97$ and $Z_h > 50\,\mathrm{dBZ}$ as a function of moment 2 (c) and moment 4 (d). The Pearson (PCC) and Spearman (SCC) correlation coefficients are reported for each panel. Values for the Locarno-Monti event of 25.08.2023 illustrated in Fig. 3 are annotated.

– Volume of the region where $\rho_{HV} < 0.97$ and $Z_H > 50$ dBZ within a 1 km radius, above the freezing level.

For details on the incremental improvement obtained by including each feature in the predictor set, see Table A2. These four features show significant positive correlations with moments 2 and 4 derived from the hailsensor data (Fig. 4). Many other extracted features are highly correlated with these four, differing only by the scale of spatial aggregation or the minimum Z_H threshold. Therefore, the selection of these specific features may also reflect the limited amount of reference data from hailsensors, which makes robust cross-validation challenging.

4.1.1 Impact of non-Rayleigh scattering

Non-Rayleigh scattering can complicate direct relationships between polarimetric radar variables and hail size (Ryzhkov et al., 2013a; Kaltenboeck and Ryzhkov, 2013), as resonance effects arise when the radar wavelength is comparable to hailstone diameters (roughly 2 cm or larger at C-band). These effects increase uncertainty in polarimetric measurements and can potentially bias radar-derived hail estimates.

Our approach mitigates these limitations by using features derived from a high-resolution 3D volume of polarimetric variables and by considering moments of the HSND rather than individual diameters. Moment 2 is more sensitive to smaller and more numerous hailstones, whereas moment 4 is dominated by contributions from larger hailstones. As a result, moment 2 is generally less sensitive to non-Rayleigh scattering than moment 4; however, it may still be indirectly affected, as resonance effects can influence the radar observables used to estimate hailstones number. Moment 4, being dominated by large hailstones, is more directly impacted by non-Rayleigh scattering.

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

Among the GAM-selected features, the maximum of Z_V in the vertical column is likely the most affected by resonance. Interestingly, it correlates more strongly with moment 2 (Fig. 4a) than with moment 4 (Fig. 4b), indicating that features reflecting hailstones number are less sensitive to non-Rayleigh scattering than those emphasizing large hailstones.

Our feature set also includes integrated measures, such as the sum of Z_V over the vertical column and the volume of regions where $\rho_{HV} < 0.97$ and $Z_H > 50$ dBZ. The latter is most strongly correlated with moment 4 (Fig. 4d), showing that using volume-based or aggregated variables helps mitigate the effects of non-Rayleigh scattering. By relying on 3D aggregated measures rather than point reflectivity or other dual-pol values, our approach reduces sensitivity to non-Rayleigh scattering while preserving meaningful information about both hailstones size and number.

4.1.2 Physical meaning of the selected features

The importance of reflectivity- and cross-correlation-based features is consistent with previous hail studies. For example, Klaus and Krause (2024) found that the minimum copolar cross-correlation coefficient in the midlevel updraft effectively discriminates between small and severe hail, and that the area of reflectivity $>$ 50 dBZ in the upper updraft indicates the presence of giant hail. Forcadell et al. (2024) used a convolutional neural network for hail detection, where key features included echo-top height, maximum Z_H, and ρ_{HV} below 2000 m a.s.l., with the latter negatively correlated with hail presence. A decrease in ρ_{HV} is expected in the presence of melting hail or hail growing in the wet regime (Ryzhkov and Zrnić, 2019). In our study, negative correlations between ρ_{HV}-based features and HSND moments were also observed, but even stronger positive correlations were found for volumes associated with $\rho_{HV} < 0.97$ (Fig. 4c, d).

Echo-top heights were also found to correlate with hailsensor-observed HSND moments, consistent with the expectation that severe storms produce stronger high-altitude echoes (Trefalt et al., 2023). However, including echo-top features in the GAM did not improve the cross-validation performance against hailsensor data, nor the performance relative to drone-retrieved HSND and crowd-sourced reports. The maximum reflectivity correlates positively with observed moments (Fig. 4a, b), confirming its sensitivity to HSND and association with large hail (Ryzhkov and Zrnić, 2019).

Other polarimetric variables have been identified as relevant in hail detection. For instance, Aregger et al. (2025) developed a hail detection algorithm using Z_{DR} columns. In our study, Z_{DR} columns were computed as the volume of continuous regions exceeding various thresholds in the 3D radar composite. However, these columns could not be reliably identified for most hailsensor-observed events, and including them in the GAM did not improve model performance.

4.2 Evaluation against hailsensor data

A leave-one-event-out cross-validation was applied to assess the robustness of the model in estimating HSND moments (Fig. 5) and in retrieving the full HSND (Fig. 6).

Overall, moment 4 is predicted more accurately than moment 2, as indicated by the higher PCC between predicted and observed moment 4 (Fig. 5e–h) compared to moment 2 (Fig. 5a–d). This difference can be attributed to the inclusion of ρ_{HV}-based features, which are particularly sensitive to the presence of large hailstones and therefore provide stronger constraints

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

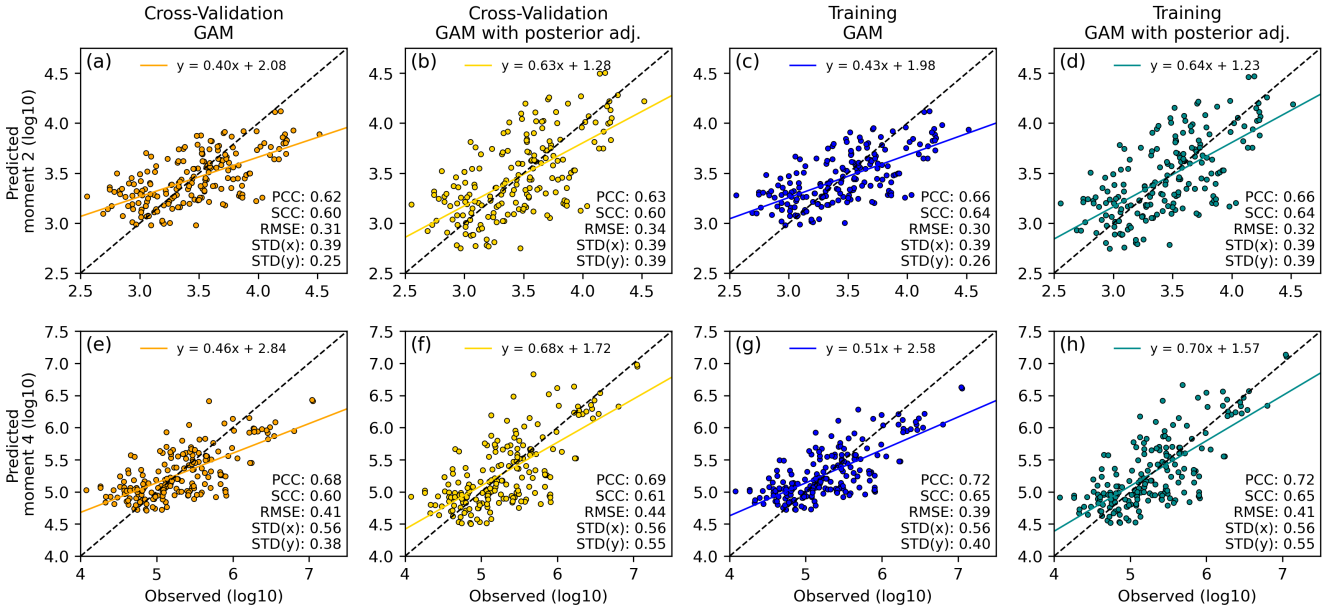

Figure 5. Comparison of observed and GAM predicted log-moment values for moment 2 (a, b, c, d) and moment 4 (e, f, g, h). The presented results refer to the GAM leave-one-event-out cross-validation without a posteriori adjustment (a, e), with a posteriori adjustment (b, f), and the GAM trained with the full dataset and applied on the same dataset without a posteriori adjustment (c, g), with a posteriori adjustment (d, h).

for higher-order moments. Nevertheless, a systematic bias remains, especially for events characterized by large hail diameters, which necessitated an a posteriori adjustment (Sect. 3.4.3).

The limited number of available events (35 in total) constrains the cross-validation, as events with medium to large hailstones (>2 cm) are underrepresented in the training set. As a consequence, the model has limited ability to generalize to these less frequent but high-impact cases. Despite this limitation, the a posteriori adjustment substantially improves the spread of the estimated moments, bringing it closer to the spread observed by the hailsensors. This improvement is achieved while preserving the same PCC as the non-adjusted GAM and with only a slightly increased RMSE. The fact that the GAM trained on the full dataset (Fig. 5c, d, g, h) performs only marginally better than the cross-validated model (Fig. 5a, b, e, f) indicates that the model is unlikely to be affected by overfitting.

Examples from three selected case studies—one per hail - prone region (Ticino: Fig. 6a–c; Napf: Fig. 6d–f; Jura: Fig. 6g–i) — show a tendency of the radar-retrieved HSND to overestimate the number of small hailstones and to underestimate the number of large hailstones, which is explained by a general overestimation of small moment 2 values (Fig. 5). This behavior is particularly evident for the event of 25.08.2023 (Fig. 3), which featured the largest observed hail diameter (4.2 cm; Fig. 3l) and the highest moment 4 value. As an extreme event, it is expected to be especially challenging for the cross-validated model to reproduce the observed HSND accurately. For this event, the model trained on the full dataset and combined with the a

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

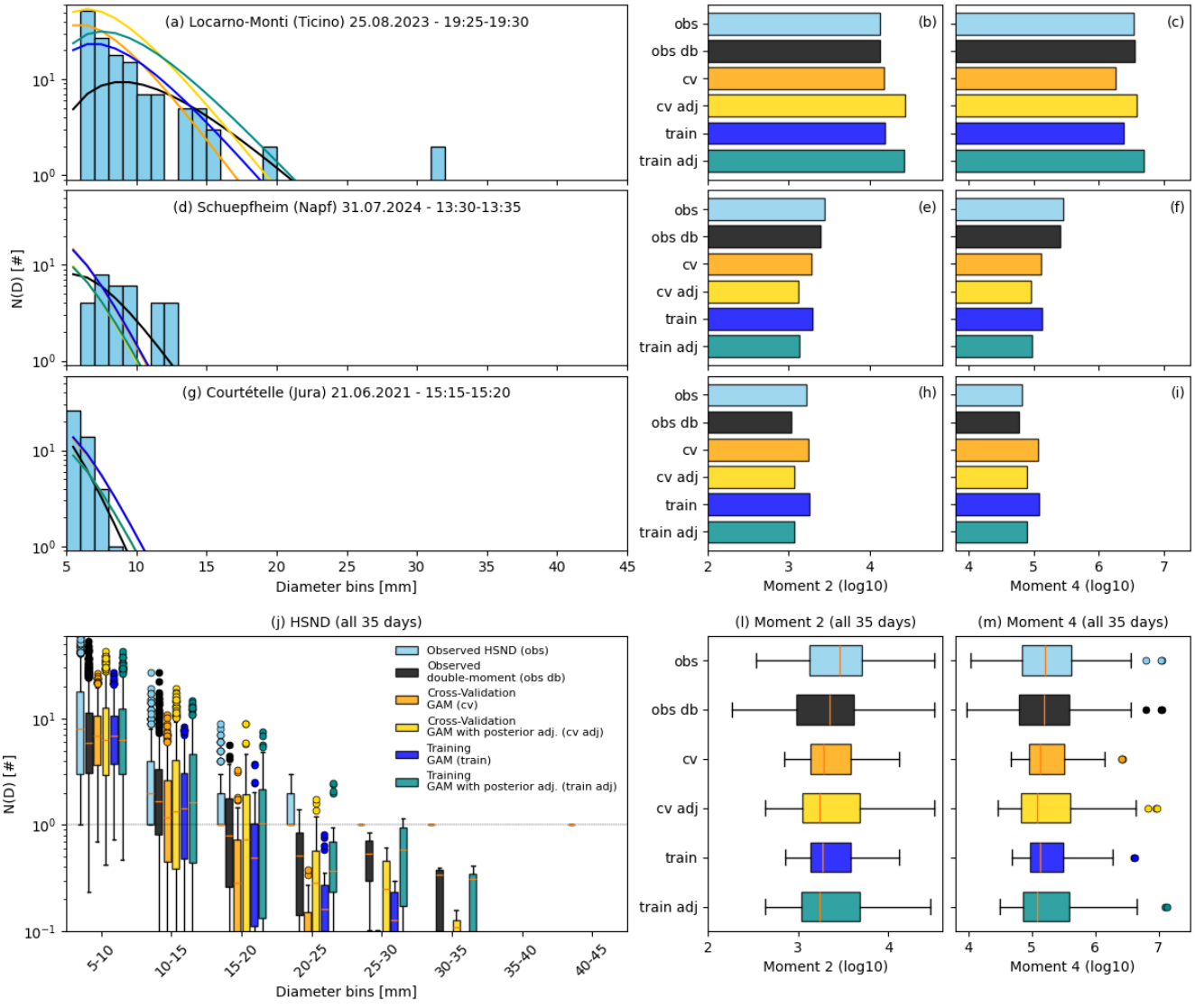

Figure 6. Comparison of observed and radar-retrieved HSND (first column), moment 2 (second column), moment 4 (third column). Three specific hail events are reported (one for each hail-prone region): 25.08.2023 in Locarno-Monti (Ticino) between 19:25-19:30 (radar observational time) (a, b, c), 31.07.2024 in Shüpfheim (Napf) between 13:30-13:35, and 21.06.2021 in Courtételle (Jura) between 15:15-15:20. The last row regroups the data of all hail events occurred in the 35 analyzed days (j, l, m).

posteriori adjustment shows an improved ability to estimate the number of large hailstones, yielding results that are closer to the HSND obtained via double-moment normalization of the hailsensor-observed moments.

Considering all 35 analyzed events (Fig. 6j–m), model performance generally decreases for increasing hail diameters. Nevertheless, the a posteriori bias-correction step effectively reduces underestimation in the upper tail of the distributions, leading

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

to improved agreement with HSNDs derived from double-moment normalization of hailsensor data. While cross-validation results still underestimate large hail, the model trained on the full dataset — used in subsequent comparisons with drone and crowd-sourced observations — exhibits improved calibration across the full diameter range and a spread of estimated moments that is much closer to the observed one.

4.3 Evaluation of radar-retrieved HSND against drone-retrieved HSND

The radar-retrieved HSNDs were compared against distributions obtained from drone-based photogrammetry (Fig. 7). Overall, the retrieved HSND shapes are consistent between the two approaches, although a systematic overestimation of hailstone counts at small diameters is evident in the radar-based retrievals. This overestimation likely originates from the training strategy: only events with at least 30 hailstones detected by hailsensors were included, which biases the model toward hail-rich cases. As a result, the model tends to reproduce HSNDs with higher hailstone counts than average. In addition, small hailstones may be partially missed in the drone-retrieved HSND due to masking by grass, which can further contribute to this discrepancy.

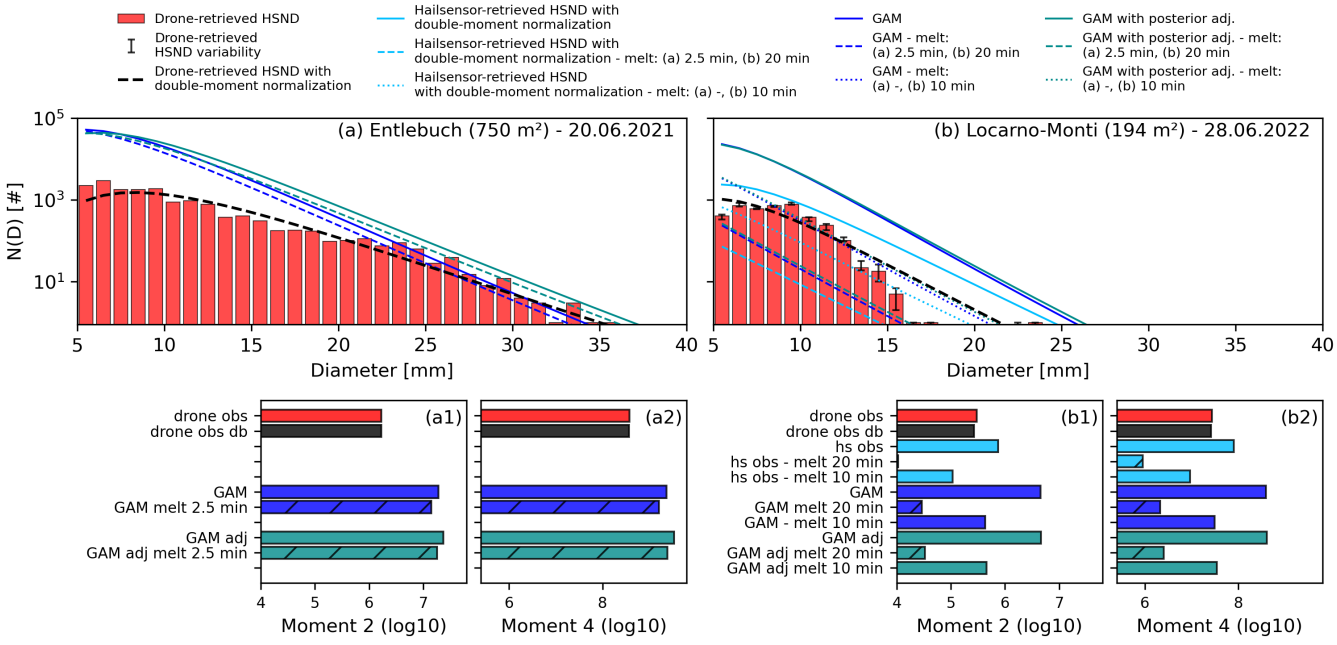

Figure 7. Comparison of radar-retrieved HSND with drone-retrieved HSND over Entlebuch on 20 June 2021 (a) and Locarno-Monti on 28 June 2022 (b). The corresponding moments 2 and 4 are shown in panels (a1, a2) and (b1, b2). Since drone surveys are performed only after the event has ended, a melting correction assuming a constant rate of 0.5 mm/min - estimated by Lainer et al. (2024) for event (a) - is applied to the radar-retrieved HSND. For case (a), the drone survey was conducted 2.5 min after the storm, whereas for case (b), the survey took place approximately 20 min after the storm's end. For illustration, HSND corrected for 10 min of melting is also shown for case (b). For this case, three drone-based retrieval models are presented, corresponding to the three expert annotators whose manual hailstone segmentations were used as training reference data (see Portmann et al. (2025)).

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

The comparison also relies on the assumption that hailstones are uniformly distributed within the radar-based GAM model resolution (1 km^2), whereas drone observations typically cover only a few hundred square meters. Differences between the radar- and drone-based HSNDs may therefore also reflect local spatial variability in hail rather than retrieval errors alone.

Additional uncertainty arises from hailstone melting occurring between the end of the hailfall and the drone survey. This effect is particularly uncertain for the event of 28.06.2022, for which approximately 20 min elapsed before the drone flight. In this case, the applied melting correction is based on the event of 20.06.2021 and may not be fully representative, given potentially different environmental conditions such as ground temperature, humidity, wind, vegetation cover, and hailstone size and density. For the event shown in Fig. 7a (Lainer et al., 2024), melting was observed at ambient temperatures of about 20 °C and relative humidity around 85%, whereas for Fig. 7b (Portmann et al., 2025), conditions were cooler (approximately 17 °C) and more humid (90–100%).

The potential unreliability of the melting-corrected radar-retrieved HSND for Fig. 7b is further supported by the fact that the HSND obtained via double-moment normalization from hailsensor observations agrees more closely with the radar-based HSND than with the drone-retrieved HSND. We therefore also report a radar-retrieved HSND corrected assuming a shorter melting duration of 10 min instead of 20 min (equivalent to an average melting rate of 0.25 mm min^{-1} rather than 0.5 mm min^{-1} over 20 min), which yields great agreement with the drone-based HSND and the HSND retrieved via double-moment normalization from the drone observations.

No pronounced differences are observed between the a posteriori adjusted and non-adjusted GAMs, although minor improvements are evident for large hail diameters in Fig. 7a, leading to better overall consistency between radar- and drone-based HSNDs. Despite remaining scaling differences, the agreement in distribution shape — particularly for the event of 28.06.2022 — is encouraging and suggests that the model captures the underlying HSND reasonably well. Nevertheless, additional drone-based datasets are required to robustly assess and correct absolute biases.

Overall, drone-based HSND retrievals show strong potential for evaluating radar-based hail products, as they provide spatially distributed observations over much larger areas than hailsensors. However, accurate estimation of hailstone melting remains a key challenge, since operational constraints often prevent drone surveys from being conducted immediately after hailfall.

4.4 Comparison of HSND percentile diameters against crowd-sourced reports

Percentile diameters derived from the radar-retrieved HSND were compared with the 75th and 90th percentiles of crowd-sourced hail reports within TRT cells (Sect. 3.5.3). In contrast to MESHS, the HSND-derived diameters exhibit lower bias and stronger correlations with the reported hail sizes (Fig. 8). These results highlight the potential of HSND-based percentile metrics to better represent hail size information from crowd-sourced observations, although the intrinsic uncertainty of the reports limits achievable correlations. Spatial aggregation was applied to mitigate this uncertainty. Aggregation over a radius of approximately 9 km provides the best overall agreement between radar-derived and reported hail sizes. This aggregation helps compensate for reporting inconsistencies, location uncertainty, and the sparse nature of individual observations.

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

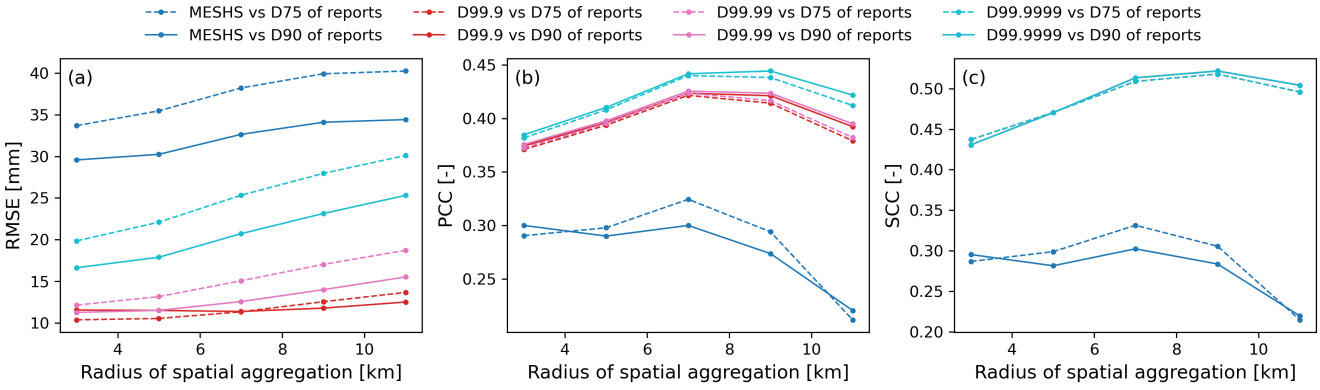

Figure 8. Comparison of crowd-sourced reports, MESHS and percentile diameters of the radar-retrieved HSND. Data were spatially aggregated from 3 to 11 km radius. For MESHS and for percentile diameters of the radar-retrieved HSND, maximum values were compared with the 75th and 90th percentile of the reports in the same aggregation area.

The analysis further shows that higher HSND percentiles tend to yield higher PCC values with the reports, although the differences between percentiles are relatively small. The SCC remains unchanged across percentiles, reflecting the same rank-order agreement. In contrast, the RMSE increases substantially for higher percentiles, indicating increasing uncertainty associated with extreme hail sizes and the intrinsic variability of crowd-sourced reports.

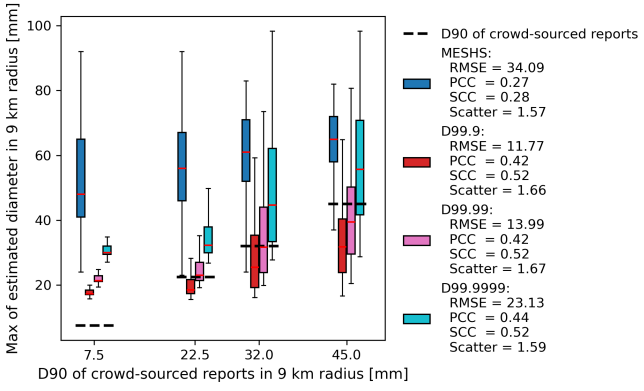

Figure 9. Comparison of crowd-sourced reports, MESHS and percentile diameters of the radar-retrieved HSND on 37 days. Each field was spatially aggregated over 9 km radius (maximum filter for MESHS and radar-based diameters, D90 for the crowd-sourced reports).

It should be noted that MESHS represents the maximum expected hail diameter within a 1 km grid cell, whereas crowd-sourced reports refer to individual hailstone observations, presumably close to the largest stones observed over areas of roughly 10–100 m^2. By contrast, the HSND framework allows the computation of exceedance probabilities for given diameter thresholds, providing a more physically consistent description of hail size. The boxplot summary of all events at 9 km aggrega-

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

tion (Fig. 9) further illustrates the better agreement between high HSND percentiles and crowd-sourced reports compared to MESHS.

4.5 Radar-retrieved HSND properties: case study of 28 June 2021

The added value of the HSND approach is well illustrated by the hailstorm outbreak of 28 June 2021, one of the most heavily reported hail events in Switzerland (Fig. 10). In addition to percentile diameters (Fig. 10c, d), the retrieved HSND enables the derivation of further physically meaningful properties, such as total hailstone counts and total kinetic energy (Fig. 10e, f). These quantities clearly distinguish between two storm tracks, with higher values along the track where the largest hailstones were reported (Fig. 10a) and where hail-related damages were observed (see Schmid et al., 2024, Fig. 10i). In contrast, MESHS (Fig. 10b) assigns similar diameters to both tracks and therefore fails to reproduce the observed spatial variability in ground impact.

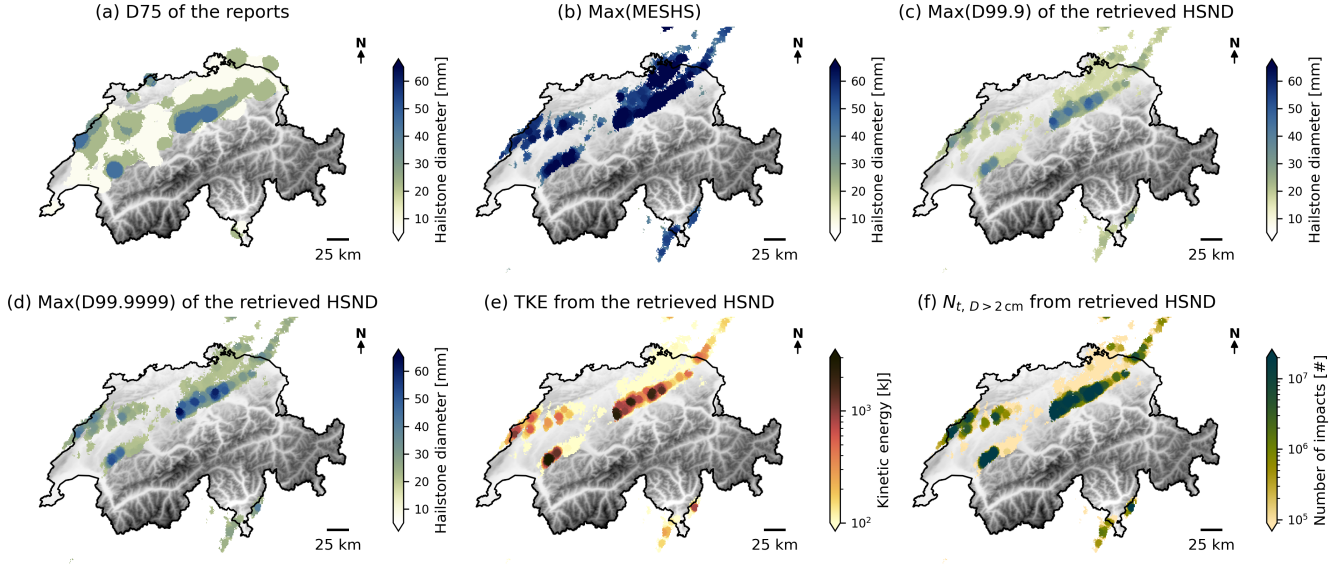

Figure 10. Maps of hail data on 28.06.2021. D75 of crowd-sourced reports (a), MESHS (b), and examples of properties extracted from the radar-retrieved HSND: percentile 99.9999 (c), percentile 99.9 (d), total kinetic energy (e), number of hailstones $>$ 2 cm (f). All features were aggregated over a 9 km radius.

This behavior is confirmed by the quantitative comparison against crowd-sourced reports for this event, which shows substantially higher PCC and SCC, as well as lower RMSE, for the radar-retrieved HSND percentile diameters compared to MESHS (Fig. 11). The best agreement for this day is obtained when aggregating values over an 11 km radius, yielding a PCC of 0.70 between radar-retrieved HSND percentiles and aggregated crowd-sourced reports (Fig. 12).

The improved discrimination achieved by the HSND-based approach is mainly attributable to the ρ_{HV}-based predictor, which exhibits clear contrasts between the damaged and undamaged tracks (Fig. 13c, f). By comparison, echo-top-based

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

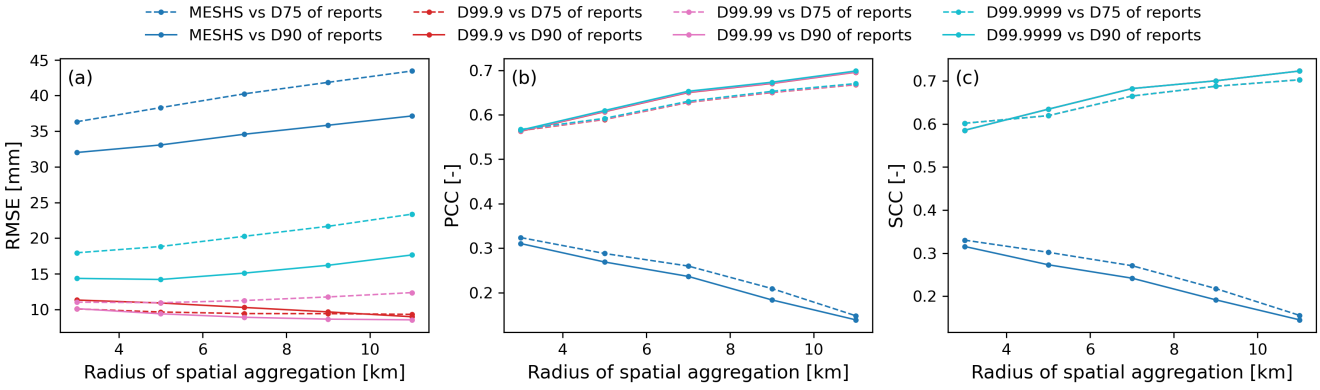

Figure 11. Comparison of crowd-sourced reports, MESHS, and percentile diameters of the radar-retrieved HSND, as in Fig. 8, but specifically for 28.06.2021. Data for 25 events were smoothed with spatial filters from 3 to 11 km. For MESHS and for percentile diameters of the radar-retrieved HSND (D99.9, ...), maximum values were compared with the 90th percentile of the reports in the same aggregation area.

features, which form the basis of MESHS, do not provide such separation (Fig. 13b, e). These results are consistent with the damage analysis presented in Fig. 10 of Schmid et al. (2024) and highlight the added value of radar-retrieved HSND features for identifying hail-related impacts. This added value primarily arises from the inclusion of polarimetric variables beyond reflectivity alone, in particular the ρ_{HV}-based predictor.

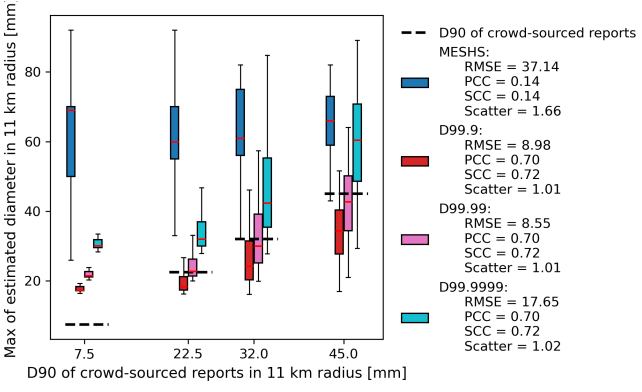

Figure 12. Comparison of crowd-sourced reports, MESHS and percentile diameters of the radar-retrieved HSND, as in Fig. 9, but specifically for 28.06.2021. Each field was smoothed with a spatial filter of 11 km radius (maximum filter for MESHS and radar-based diameters, D90 for the crowd-sourced reports).

4.6 Challenges and model limitations

The different observational datasets used in this study each carry distinct uncertainties, spatial representativeness, and sensitivities to different characteristics of hail. Dual-polarization radar provides volumetric proxies of hail aloft; hailsensors measure

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

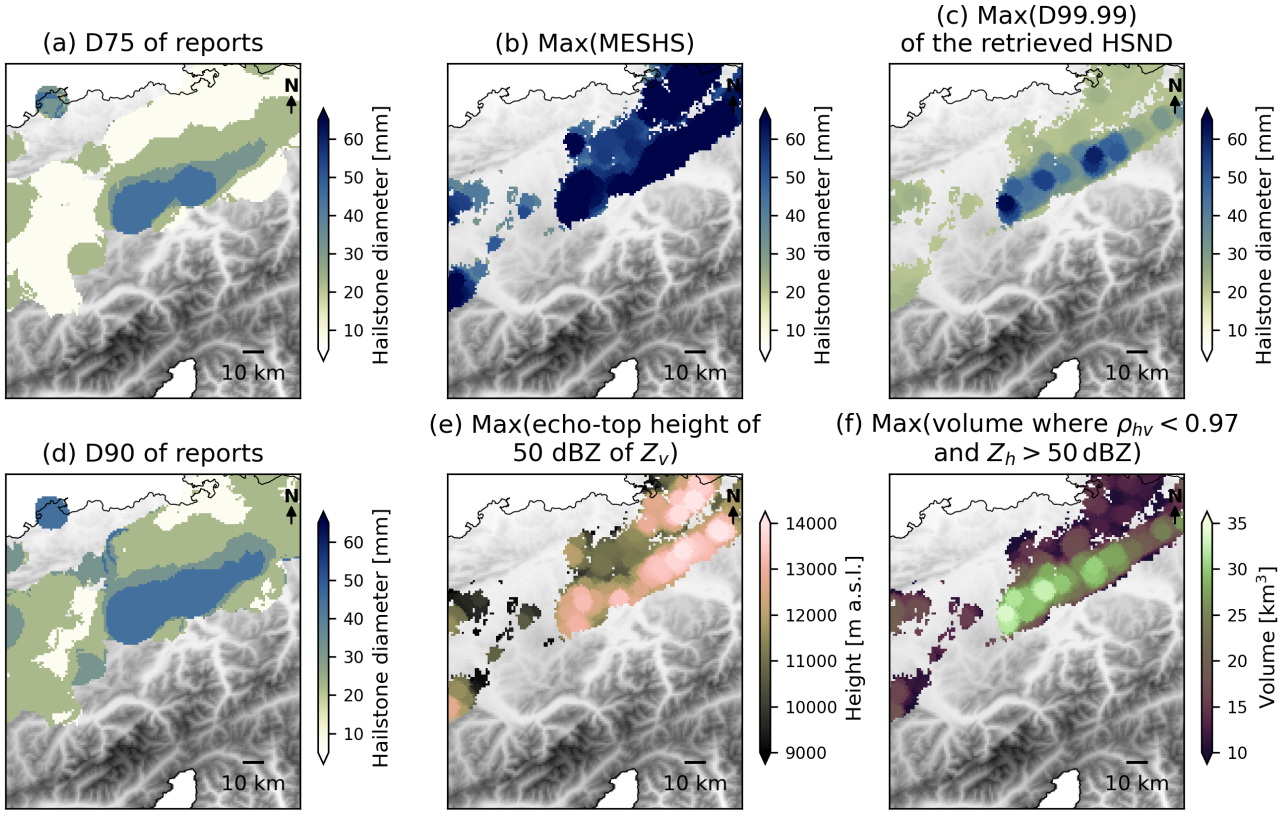

Figure 13. Features extracted from the radar-retrieved HSND on 28.06.2021 over the same area shown in Schmid et al. (2024) for hail damage assessment. All features were aggregated over a 9 km radius.

ground-level hail size distributions but only at discrete locations and over very small areas (0.2 m^2); drone-based HSNDs cover larger, yet still limited spatial extents (typically a few hundred square meters) and are affected by melting and partially hidden hailstones; and crowd-sourced reports are subject to subjectivity and strong sampling biases toward populated areas. Integrating these heterogeneous data sources inevitably introduces uncertainties throughout the model training, validation, and evaluation steps.

An important aspect not addressed in the present study is the explicit quantification of model uncertainty. Future work should therefore explore ensemble-based uncertainty estimation, for example by using the variability between predicted and extrapolated moments to derive a range of plausible HSND realizations rather than a single deterministic retrieval. Such approaches would better reflect uncertainties associated with the hailsensor-based reference moments.

A further challenge arises from the mismatch in spatial scales between the datasets. Radar-based retrievals are representative at the kilometer scale, whereas drone-based measurements and crowd-sourced reports sample much smaller areas. This scale mismatch complicates direct comparisons and likely explains part of the observed discrepancies between datasets. In particular, the assumption of a spatially uniform HSND within a radar grid cell (Sect. 4.3) is likely an oversimplification. However, the

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

lack of observations covering intermediate spatial scales prevents a more detailed analysis of HSND spatial variability and the application of more advanced downscaling or upscaling approaches.

Some of the radar-based features used in this study are inherently influenced by radar sampling geometry. At far ranges, the widening of the radar beam reduces vertical resolution, which can prevent storm tops from being properly captured. Conversely, the minimum elevation angle may not adequately sample the lower storm regions, particularly given the high altitude of Swiss radars and beam blocking by complex terrain. These effects can lead to underestimation of volume-based features. In addition, the increasing radar resolution volume with distance can affect polarimetric variables, which may differ from values observed closer to the radar.

While the training strategy likely leads to a systematic overestimation of hailstone counts for small diameters, the model nevertheless shows strong potential. Specifically, the upper tail of the radar-retrieved HSND agrees well with drone-based estimates, percentile diameters are less sensitive to absolute scaling biases, and integrated quantities such as total kinetic energy provide physically meaningful information related to ground impact. For impact-oriented applications, a consistent bias across events is therefore less critical than the relative differences between storms, which the presented radar-based model is able to capture with an overall good performance.

Additional drone-based HSND observations remain essential to assess whether the observed overestimation of small hailstones is systematic across events and to determine whether such biases can be corrected in a robust manner. Moreover, additional hailsensor data — particularly for extreme events — would allow a more thorough assessment of the generalization capability of the GAM and could support the development of more flexible relationships, for example through additional spline terms or more complex model architectures.

5 Summary and conclusions

This study demonstrated that linking the empirical second and fourth moments of the HSND with radar measurements enables estimation of the full distribution of hail diameters at the ground, using the formula and parameters of $\hat{h}(x)$ defined in Ferrone et al. (2024). This approach is conceptually similar to established retrieval techniques for raindrop and drizzle size distributions (Raupach and Berne, 2017; Shin et al., 2024; Raupach et al., 2019).

A wide range of observational datasets was used, including dual-polarization radars, hailsensors, drone photogrammetry, and crowd-sourced reports. Each source captures different hail characteristics at different scales and with distinct uncertainties, complicating direct comparisons. Nevertheless, the results demonstrate the potential of radar-based HSND retrievals.

Among the radar-based predictors, the GAM model performed best using the following combined features: (i) the volume where $\rho_{HV} < 0.97$ and $Z_H > 50$ dBZ within 1 km, (ii) the column-integrated sum of Z_V (linear units) within 1 km, (iii) the maximum Z_V in the vertical column within 1 km, and (iv) the volume in (i) but above the freezing level. Notably, the volume with $\rho_{HV} < 0.97$ and $Z_H > 50$ dBZ was strongly correlated with the fourth moment (sensitive to larger hailstones), yielding higher cross-validated correlations for moment 4 (0.69) than for moment 2 (0.63).

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

An a posteriori bias adjustment reduced residual underestimation of large-diameter hailstones. Comparisons with two independent drone-based HSNDs showed good agreement in distribution shape but revealed systematic overestimation of small-diameter hailstones. This overestimation likely reflects a combination of general overestimation of small moment 2 values, training bias toward more intense events (only events with >30 hailstones were included), and partially hidden or melted hailstones in the drone data. Additional drone-based observations are needed to better quantify and potentially correct this bias.

Percentile diameters derived from radar-retrieved HSNDs also compared favorably with crowd-sourced reports across multiple spatial scales, showing lower bias and higher correlations than MESHS. The 28 June 2021 case study illustrates the advantages of the approach: radar-derived HSND features such as total kinetic energy, hailstone count, and percentiles clearly distinguished two separate storm tracks. Stronger values aligned with reported ground damage, whereas MESHS failed to discriminate between the tracks. This enhanced discrimination primarily arises from the ρ_{HV}-based predictor in the GAM, while echo-top features provided no clear separation.

In summary, the proposed method enables high-resolution (1 km, 5 min) retrievals of the full HSND and derived features, such as kinetic energy, providing a valuable tool for both real-time hail monitoring and long-term statistical analyses in Switzerland. While the framework could be adapted to other regions, the assumption of HSND invariance outside Switzerland requires further validation. Expanding ground-based observations remains critical for improving retrieval robustness, constraining uncertainties, and enabling a more comprehensive evaluation of this approach.

Appendix A: Event data description

A detail description of the used hailsensor, hail reports and drone data is reported in Table A1.

Appendix B: Merging five radars into a unique 3D radar composite

This appendix details the methodology used to create a 3D radar composite for each radar and subsequently merge them into a unified product and shows some vertical cross-sections of polarimetric variables for the observed by the hailsensor in Locarno-Monti on 25.08.2023. The procedure involves interpolating polar radar data onto a regular 3D Cartesian grid and then selecting the most reliable radar measurement for each voxel.

B1 Procedure:

1. Read radar data: retrieve polar radar measurements, including the coordinates of radar gate centers.
2. Define 3D grid: establish a regular 3D grid for interpolation of the polar radar data.
3. Nearest-neighbor selection: for each voxel in the grid:
   - Identify the three nearest radar gates from the same radar.

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

- Compute distances from the voxel center to each gate.

4. Beamwidth adjustment: to prevent gaps near the radar, the effective half-beamwidth is artificially enlarged at high elevations for gridding purposes only (see Fig. A1), while the physical antenna beamwidth remains unchanged:
   - Elevation $9.5°$: $0.75°$
   - $11.0°$: $1.0°$
   - $13.0°$: $1.5°$
   - $16.0°$: $2.0°$
   - $> 16.0°$: $2.5°$
   - All other elevations: $0.5°$
5. Range filtering: exclude voxels beyond the maximum range allowed by the radar beam and grid resolution:

   $$\text{dist} \leq \sqrt{(r \cdot \sin(\theta_{\text{beam}}))^2 + r_{\text{res}}^2}$$

   where r is the radar range, θ_{beam} is the enlarged half-beamwidth, and $r_{\text{res}} = 500\ \text{m}$ is the range resolution of operational polar data.
6. Coordinate transformation: convert Cartesian grid coordinates $(x_{\text{grid}}, y_{\text{grid}}, z_{\text{grid}})$ to radar coordinates (range r_{grid}, azimuth α_{grid}, and elevation ϵ_{grid}) as detailed in Appendix B2.
7. Voxel assignment: a grid voxel is considered inside a radar gate if:

$$|r_{\text{grid}} - r_{\text{gate}}| \leq \frac{r_{\text{res}}}{2} \tag{B1}$$

$$|\alpha_{\text{grid}} - \alpha_{\text{gate}}| \leq \theta_{\text{beam}} \tag{B2}$$

$$|\epsilon_{\text{grid}} - \epsilon_{\text{gate}}| \leq \theta_{\text{beam, enlarged}} \tag{B3}$$

Once the 3D grid for each radar is constructed, the individual grids are merged into a single 3D composite. For voxels observed by multiple radars, the polarimetric variables are taken from the radar with the highest horizontal reflectivity, which is assumed to be the least attenuated.

Fig. A2 clearly shows an example where using multiple radars allowed to observe higher reflectivity values than using the Monte-Lema radar only.

B2 Transforming Cartesian to radar coordinates

To assign radar volumes to the 3D grid, each voxel's Cartesian coordinates $(x_{\text{grid}}, y_{\text{grid}}, z_{\text{grid}})$ are converted to radar coordinates (range r_{grid}, azimuth α_{grid}, elevation ϵ_{grid}).

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

B2.1 Azimuth:

The azimuth angle is computed as

$$\alpha_{\text{grid}} = \arctan \frac{y_{\text{grid}}}{x_{\text{grid}}} \tag{B4}$$

where x_{grid} and y_{grid} are the coordinates of the voxel relative to the radar location.

B2.2 Range:

The range from the radar to the voxel is determined using the law of cosines, accounting for Earth curvature:

$$r_{\text{grid}} = \sqrt{z_{\text{grid}}^2 + d_{\text{grid}}^2 - 2 z_{\text{grid}} d_{\text{grid}} \sin\left(\frac{s}{2R}\right)} \tag{B5}$$

where d_{grid} is the line-of-sight distance:

$$d_{\text{grid}} = 2(R + z_{\text{grid}}) \sin\left(\frac{s}{2R}\right), \tag{B6}$$

s is the arc length along the Earth's surface ($s = x_{\text{grid}} / \sin \alpha_{\text{grid}}$), and $R = 6371 \cdot 10^3 \cdot k_e$ is the effective Earth radius (where $k_e = 1.25$).

B2.3 Elevation:

Finally, the elevation angle is obtained from

$$\epsilon_{\text{grid}} = \arcsin \frac{r_{\text{grid}}^2 + z_{\text{grid}}^2 - d_{\text{grid}}^2}{2 r_{\text{grid}} z_{\text{grid}}}. \tag{B7}$$

B3 Vertical cross-sections of polarimetric variables on 25.08.2023

In this section, we report vertical-cross sections of polarimetric variables for the 3D radar composite based on the Monte Lema radar only and for the final composite combining the 5 radars between 19:20 and 19:25 UTC (Fig. A3).

Appendix C: Procedure to spatially extrapolate hailsensor observations

The limited sensing area of hailsensors (0.2 m^2) poses challenges when relating localized ground observations to the radar measurements aloft with much larger sampling volumes. Training a model directly on such small-scale observations would yield HSND estimates calibrated to an area much smaller than the typical 2D spatial resolution of radar-based hail products.

To mitigate this scale mismatch, we spatially extrapolated the hailsensor data to generate target fields extending over several kilometers (i.e., multiple radar grid cells). This method relies on the assumption that the temporal variations of HSND measured at a fixed hailsensor are representative of the spatial variability of HSND within the moving storm segment that passes over the sensor. In other words, the storm structure is assumed to remain stationary in the storm-relative reference frame during the 5-minute extrapolation window. Under this assumption, displacing the sensor time series along the storm trajectory reconstructs

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

the spatial distribution of HSND at scales comparable to several 2D radar grid cells. Moreover, a single 2D radar grid cell is associated with multiple target values of HSND, whose variability provides an implicit measure of uncertainty.

Furthermore, it is important to recall that the 3D radar composite combines multiple scans performed at different elevations, ranges, and azimuths over 5 minutes. Considering all scans together, the composite can be interpreted as representing the average storm structure over this interval. To be consistent with this, we cumulate the spatially extrapolated sensor observations over the same 5-minute period, generating transects that reflect the storm segment sampled by the radar product. This procedure can thus be viewed as reconstructing, in space, the segment of the convective cell that produced the sensor impacts in the 5 minutes of radar scans, enabling a direct comparison between radar-derived features and spatially consistent target data.

The different steps of the procedure are detailed hereafter:

1. Temporal grouping: hailstone impacts on each sensor were grouped in 20 s intervals. Each interval contains a set of hailstone diameters H_i and a timestamp t_i. The 20 s interval for grouping hailstone impacts was chosen as a compromise between obtaining a sufficient density of points along the transect, while avoiding excessive redundancy when cumulating over the 5-minute window, and ensuring that each interval contains enough hailstones to provide a robust estimate of HSND.

2. Initial spatial extrapolation: using storm speed and direction (v_x, v_y) from the TRT cell surrounding the hailsensor, each interval was extrapolated backward along the storm trajectory to the start of the 5 minute window preceding the final radar scan at time t_r:

$$x_i^{\text{start}} = x_s - (t_i - t_{\text{start}})\, v_x, \quad y_i^{\text{start}} = y_s - (t_i - t_{\text{start}})\, v_y, \tag{C1}$$

where x_s, y_s are the sensor coordinates, and $t_{\text{start}} = t_r - 5$ min. The minus sign ensures the measurements are moved upstream along the storm vector, reflecting the positions of hailstones earlier in the 5 min period.

3. Forward displacement and accumulation: each interval is then moved forward along the storm vector in steps corresponding to 20 s intervals, covering the remaining time until the radar scan at t_r. The positions along the transect are given by

$$x_{i,j} = x_i^{\text{start}} + j\, v_x\, \Delta t_{\text{step}}, \quad y_{i,j} = y_i^{\text{start}} + j\, v_y\, \Delta t_{\text{step}}, \quad j = 0, \ldots, n_{\text{steps}}, \tag{C2}$$

with $\Delta t_{\text{step}} = 20$ s and n_{steps} chosen to reach t_r.

If multiple extrapolated positions fall on the same position, they are merged by combining the hailstone diameters. This ensures that the accumulated transect provides an estimate of the HSND along the storm segment during the 5 minutes window represented by the 3D radar composite.

From the resulting transect, in the training of the radar-based model (Sect. 3.4.1), we only used as target HSNDs that count at least 10 hailstones and which the maximum time distance from the actual hailsensor observation to the extrapolated time is 30 s, corresponding to a maximum space distance of approximately 200 m (see Appendix D).

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

Appendix D: Sensitivity study to determine the optimal training data

A sensitivity study was performed to determine the following parameters:

1. Time offset between radar observations aloft and hailsensor measurements on the ground
2. Time window used to aggregate sensor data prior to spatial extrapolation
3. Maximum extrapolation time distance
4. Minimum number of samples recorded by a hailsensor within a 5-minute timestamp

All parameter values were determined by optimizing the Spearman correlation (averaged over all folds of a leave-one-event-out cross-validation) between moments 2 and 4 and the two radar features exhibiting the highest Spearman correlation with these moments.

The identified values of parameters (1) and (2) are shown in Fig. A4. A different time window was found for moment 2 (20 s) and moment 4 (30 s); we decided to use the latter since we are generally more interested in large hailstone diameters and therefore prioritize the performance on higher-order moments. The values of parameters (3) and (4) — obtained using the optimal values of (1) and (2) — are shown in Fig. A5.

Appendix E: Feature selection process

Table A2 summarizes the leave-one-event-out cross-validation scores obtained at each step of the incremental inclusion of the four selected features (Sect. 3.4.2).

Code and data availability. The code used in this study is available in a private GitHub repository. The exact version corresponding to this manuscript is tagged as v1.0-paper. Access to the repository can be provided to reviewers upon request. Upon acceptance of the manuscript, the repository will be made publicly available. The figures are produced using scientific colour maps (Crameri, 2018).

The hailsensor measurements are currently owned by La Mobilière. After the hailsensor network ceases operation, ownership will transfer to MeteoSwiss. In the meantime, data requests should be directed to La Mobilière.

Drone data used in this study are publicly available on Zenodo (https://zenodo.org/records/10609730 (Lainer et al., 2024), https://zenodo.org/records/16778719 (Portmann et al., 2025)).

MeteoSwiss is in the process of making its data publicly available under an open-data policy. For more information, see https://www.meteoswiss.admin.ch/services-and-publications/service/open-data.html. For present time, MESHS and POH data are available, while archival functionality will be implemented later. Polar radar data and crowd-source hail reports are not publicly accessible. An example dataset of the radar data used in this study is available from the authors upon request. The crowd-sourced data used in this paper is available on request from MeteoSwiss.

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

Author contributions. MGu, AF, MGa, UG, and AB designed the study. AF worked on the feature extraction and wrote the code for the 3D radar composite with MGu and contribution from GG. MGu performed all other data processing, modeling, analysis of results, and prepared the manuscript, with contributions from AF, GG, MGa, UG, and AB., and supervision from MGa, UG, and AB.

Competing interests. MGu, AF, GG, MGa, and UG declare that no competing interests are present. AB is associate editor for AMT.

820 *Acknowledgements.* We would like to acknowledge La Mobilière for the funding and inNET for the technical implementation of the network of 80 hailsensors. We would also like to thank all the MeteoSwiss and EPFL-LTE collaborators for their support, and in particular Samuel Monhart, Martin Lainer and Jannis Portmann for collecting and providing the drone data and guiding in its processing.

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

References

Ackermann, L., Soderholm, J., Protat, A., Whitley, R., Ye, L., and Ridder, N.: Radar and environment-based hail damage estimates using machine learning, Atmospheric Measurement Techniques, 17, 407–422, https://doi.org/10.5194/amt-17-407-2024, 2024.

Allen, J. T., Giammanco, I. M., Kumjian, M. R., Jurgen Punge, H., Zhang, Q., Groenemeijer, P., Kunz, M., and Ortega, K.: Understanding Hail in the Earth System, Reviews of Geophysics, 58, e2019RG000 665, https://doi.org/https://doi.org/10.1029/2019RG000665, e2019RG000665 10.1029/2019RG000665, 2020.

Amburn, S. A. and Wolf, P. L.: VIL density as a hail indicator, Weather and forecasting, 12, 473–478, 1997.

Aregger, M., Martius, O., Germann, U., and Hering, A.: Differential reflectivity columns and hail: Linking C-band radar-based estimated column characteristics to crowdsourced hail observations in Switzerland, Quarterly Journal of the Royal Meteorological Society, n/a, e5003, https://doi.org/https://doi.org/10.1002/qj.5003, 2025.

Barras, H., Hering, A., Martynov, A., Noti, P.-A., Germann, U., and Martius, O.: Experiences with >50,000 Crowdsourced Hail Reports in Switzerland, Bulletin of the American Meteorological Society, 100, 1429 – 1440, https://doi.org/10.1175/BAMS-D-18-0090.1, 2019.

Besic, N., Figueras i Venturra, J., Grazioli, J., Gabella, M., Germann, U., and Berne, A.: Hydrometeor classification through statistical clustering of polarimetric radar measurements: a semi-supervised approach, Atmospheric Measurement Techniques, 9, 4425–4445, https://doi.org/10.5194/amt-9-4425-2016, 2016.

Besic, N., Gehring, J., Praz, C., Figueras i Ventura, J., Grazioli, J., Gabella, M., Germann, U., and Berne, A.: Unraveling hydrometeor mixtures in polarimetric radar measurements, Atmospheric Measurement Techniques, 11, 4847–4866, https://doi.org/10.5194/amt-11-4847-2018, 2018.

Betschart, M. and Hering, A.: Automatic Hail Detection at MeteoSwiss – Verification of the radar-based hail detection algorithms POH, MESHS and HAI, Arbeitsberichte der MeteoSchweiz, p. 59 pp, 2012.

Brimelow, J.: Hail and Hailstorms, https://doi.org/10.1093/acrefore/9780190228620.013.666, 2018.

Brook, J. P., Soderholm, J. S., Protat, A., McGowan, H., and Warren, R. A.: A Radar-Based Hail Climatology of Australia, Monthly Weather Review, 152, 607 – 628, https://doi.org/10.1175/MWR-D-23-0130.1, 2024.

Brown, T. M., Pogorzelski, W. H., and Giammanco, I. M.: Evaluating Hail Damage Using Property Insurance Claims Data, Weather, Climate, and Society, 7, 197 – 210, https://doi.org/10.1175/WCAS-D-15-0011.1, 2015.

Cecchini, M. A., Heymsfield, A. J., Honeyager, R., Field, P., Machado, L. A. T., and da Silva Dias, M. A. F.: Revisiting the Hail Radar Reflectivity–Kinetic Energy Flux Relation by Combining T-Matrix and Discrete Dipole Approximation Calculations to Size Distribution Observations, Journal of the Atmospheric Sciences, 79, 1927 – 1940, https://doi.org/10.1175/JAS-D-20-0373.1, 2022.

Crameri, F.: Scientific colour maps, Zenodo, https://doi.org/10.5281/zenodo.1243862, 2018.

Federal Office of Topography Swisstopo: The light version of the digital height model of Switzerland, https://www.swisstopo.admin.ch/en/geodata/height/dhm25.html, 2021.

Federer, B., Waldvogel, A., Schmid, W., Schiesser, H. H., Hampel, F., Schweingruber, M., Stahel, W., Bader, J., Mezeix, J. F., Doras, N., D'Aubigny, G., DerMegreditchian, G., and Vento, D.: Main Results of Grossversuch IV, Journal of Applied Meteorology and Climatology, 25, 917 – 957, https://doi.org/10.1175/1520-0450(1986)025<0917:MROGI>2.0.CO;2, 1986.

Ferrone, A., Kopp, J., Lainer, M., Gabella, M., Germann, U., and Berne, A.: Double-moment normalization of hail size number distributions over Switzerland, Atmospheric Measurement Techniques, 17, 7143–7168, https://doi.org/10.5194/amt-17-7143-2024, 2024.

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

Foote, G. B., Krauss, T. W., and Makitov, V.: Hail metrics using conventional radar, in: Proceedings of the 16th Conference on Planned and Inadvertent Weather Modification, San Diego, American Meteorological Society, p. 2791–2796, 2005.

Forcadell, V., Augros, C., Caumont, O., Dedieu, K., Ouradou, M., David, C., Figueras i Ventura, J., Laurantin, O., and Al-Sakka, H.: Severe-hail detection with C-band dual-polarisation radars using convolutional neural networks, Atmospheric Measurement Techniques, 17, 6707–6734, 2024.

Germann, U., Boscacci, M., Clementi, L., Gabella, M., Hering, A., Sartori, M., Sideris, I. V., and Calpini, B.: Weather Radar in Complex Orography, Remote Sensing, 14, https://doi.org/10.3390/rs14030503, 2022.

Hering, A., Morel, C., Galli, G., Senesi, S., Ambrosetti, P., and Boscacci, M.: Nowcasting thunderstorms in the Alpine region using a radar based adaptive thresholding scheme, in: Proceedings of ERAD Conference 2004, Visby, Island of Gotland, Sweden, 6–10 September 2004, Copernicus GmbH, pp. 206–211, https://www.copernicus.org/erad/2004/online/ERAD04_P_206.pdf, 2004.

Hohl, R., Schiesser, H.-H., and Aller, D.: Hailfall: the relationship between radar-derived hail kinetic energy and hail damage to buildings, Atmospheric Research, 63, 177–207, https://doi.org/https://doi.org/10.1016/S0169-8095(02)00059-5, 2002.

Jiang, Z., Kumjian, M. R., Schrom, R. S., Giammanco, I., Brown-Giammanco, T., Estes, H., Maiden, R., and Heymsfield, A. J.: Comparisons of Electromagnetic Scattering Properties of Real Hailstones and Spheroids, Journal of Applied Meteorology and Climatology, 58, 93 – 112, https://doi.org/10.1175/JAMC-D-17-0344.1, 2019.

Joe, P., Burgess, D., Potts, R., Keenan, T., Stumpf, G., and Treloar, A.: The S2K severe weather detection algorithms and their performance, Weather and Forecasting, 19, 43–63, 2004.

Joss, J. and Lee, R.: The Application of Radar-Gauge Comparisons to Operational Precipitation Profile Corrections, Journal of Applied Meteorology and Climatology, 34, 2612 – 2630, https://doi.org/10.1175/1520-0450(1995)034<2612:TAORCT>2.0.CO;2, 1995.

Kaltenboeck, R. and Ryzhkov, A.: Comparison of polarimetric signatures of hail at S and C bands for different hail sizes, Atmospheric Research, 123, 323–336, 2013.

Klaus, V. and Krause, J.: Investigating Hailstorm Updrafts and Nowcasting Hail Size Using a Novel Radar-Based Updraft Detection, Weather and Forecasting, 39, 1795 – 1815, https://doi.org/10.1175/WAF-D-23-0227.1, 2024.

Kopp, J., Manzato, A., Hering, A., Germann, U., and Martius, O.: How observations from automatic hail sensors in Switzerland shed light on local hailfall duration and compare with hailpad measurements, Atmospheric Measurement Techniques, 16, 3487–3503, https://doi.org/10.5194/amt-16-3487-2023, 2023a.

Kopp, J., Schröer, K., Schwierz, C., Hering, A., Germann, U., and Martius, O.: The summer 2021 Switzerland hailstorms: weather situation, major impacts and unique observational data, Weather, 78, 184–191, 2023b.

Kopp, J., Hering, A., Germann, U., and Martius, O.: Verification of weather-radar-based hail metrics with crowdsourced observations from Switzerland, Atmospheric Measurement Techniques, 17, 4529–4552, https://doi.org/10.5194/amt-17-4529-2024, 2024.

Kumjian, M. R.: Principles and Applications of Dual-Polarization Weather Radar. Part I: Description of the Polarimetric Radar Variables., Journal of Operational Meteorology, 1, 2013.

Lainer, M., Brennan, K. P., Hering, A., Kopp, J., Monhart, S., Wolfensberger, D., and Germann, U.: Drone-based photogrammetry combined with deep learning to estimate hail size distributions and melting of hail on the ground, Atmospheric Measurement Techniques, 17, 2539–2557, https://doi.org/10.5194/amt-17-2539-2024, 2024.

Lee, G., Bringi, V., and Thurai, M.: The Retrieval of Drop Size Distribution Parameters Using a Dual-Polarimetric Radar, Remote Sensing, 15, https://doi.org/10.3390/rs15041063, 2023.

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

Lee, G. W., Zawadzki, I., Szyrmer, W., Sempere-Torres, D., and Uijlenhoet, R.: A General Approach to Double-Moment Normalization of Drop Size Distributions, Journal of Applied Meteorology, 43, 264 – 281, https://doi.org/10.1175/1520-0450(2004)043<0264:AGATDN>2.0.CO;2, 2004.

Manzato, A., Riva, V., Tiesi, A., and Marcello Miglietta, M.: Observational analysis and simulations of a severe hailstorm in northeastern Italy, Quarterly Journal of the Royal Meteorological Society, 146, 3587–3611, 2020.

MeteoSwiss: ICON Forecasting Systems – Numerical weather forecasts with the ICON weather forecasting model, https://www.meteoswiss.admin.ch/weather/warning-and-forecasting-systems/icon-forecasting-systems.html, accessed: 2025-11-14, 2024.

Mirkovic, D., Zrnic, D. S., Melnikov, V., and Zhang, P.: Effects of Rough Hail Scattering on Polarimetric Variables, IEEE Transactions on Geoscience and Remote Sensing, 60, 1–14, https://doi.org/10.1109/TGRS.2021.3091907, 2022.

Nisi, L., Hering, A., Germann, U., and Martius, O.: A 15-year hail streak climatology for the Alpine region, Quarterly Journal of the Royal Meteorological Society, 144, 1429–1449, https://doi.org/10.1002/qj.3286, 2018.

Nisi, L., Hering, A., Germann, U., Schroeer, K., Barras, H., Kunz, M., and Martius, O.: Hailstorms in the Alpine region: Diurnal cycle, 4D-characteristics, and the nowcasting potential of lightning properties, Quarterly Journal of the Royal Meteorological Society, 146, 4170–4194, https://doi.org/https://doi.org/10.1002/qj.3897, 2020.

Ortega, K. L., Krause, J. M., and Ryzhkov, A. V.: Polarimetric radar characteristics of melting hail. Part III: Validation of the algorithm for hail size discrimination, Journal of Applied Meteorology and Climatology, 55, 829–848, 2016.

Portmann, J., Lainer, M., Brennan, K. P., Jourdain de Thieulloy, M., Guidicelli, M., and Monhart, S.: Performance assessment of drone-based photogrammetry coupled with machine-learning for the estimation of hail size distributions on the ground, Frontiers in Environmental Science, Volume 13 - 2025, https://doi.org/10.3389/fenvs.2025.1602917, 2025.

Portmann, R., Schmid, T., Villiger, L., Bresch, D. N., and Calanca, P.: Modelling crop hail damage footprints with single-polarization radar: the roles of spatial resolution, hail intensity, and cropland density, Natural Hazards and Earth System Sciences, 24, 2541–2558, https://doi.org/10.5194/nhess-24-2541-2024, 2024.

Púčik, T., Castellano, C., Groenemeijer, P., Kühne, T., Rädler, A. T., Antonescu, B., and Faust, E.: Large Hail Incidence and Its Economic and Societal Impacts across Europe, Monthly Weather Review, 147, 3901 – 3916, https://doi.org/10.1175/MWR-D-19-0204.1, 2019.

Raupach, T. H. and Berne, A.: Retrieval of the raindrop size distribution from polarimetric radar data using double-moment normalisation, Atmospheric Measurement Techniques, 10, 2573–2594, https://doi.org/10.5194/amt-10-2573-2017, 2017.

Raupach, T. H., Thurai, M., Bringi, V. N., and Berne, A.: Reconstructing the Drizzle Mode of the Raindrop Size Distribution Using Double-Moment Normalization, Journal of Applied Meteorology and Climatology, 58, 145 – 164, https://doi.org/10.1175/JAMC-D-18-0156.1, 2019.

Ryzhkov, A. V. and Zrnić, D. S.: Radar polarimetry for weather observations, vol. 486, Springer, 2019.

Ryzhkov, A. V., Kumjian, M. R., Ganson, S. M., and Khain, A. P.: Polarimetric radar characteristics of melting hail. Part I: Theoretical simulations using spectral microphysical modeling, Journal of Applied Meteorology and Climatology, 52, 2849–2870, 2013a.

Ryzhkov, A. V., Kumjian, M. R., Ganson, S. M., and Zhang, P.: Polarimetric radar characteristics of melting hail. Part II: Practical implications, Journal of Applied Meteorology and Climatology, 52, 2871–2886, 2013b.

Schmid, T., Portmann, R., Villiger, L., Schröer, K., and Bresch, D. N.: An open-source radar-based hail damage model for buildings and cars, Natural Hazards and Earth System Sciences, 24, 847–872, https://doi.org/10.5194/nhess-24-847-2024, 2024.

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

Shin, K., Kim, K., Song, J. J., and Lee, G.: Polarimetric Retrieval of Raindrop Size Distribution: Double-Moment Normalization Approach and Machine Learning Techniques, Geophysical Research Letters, 51, e2023GL106 057, https://doi.org/https://doi.org/10.1029/2023GL106057, e2023GL106057 2023GL106057, 2024.

Soderholm, J. S., Kumjian, M. R., McCarthy, N., Maldonado, P., and Wang, M.: Quantifying hail size distributions from the sky – application of drone aerial photogrammetry, Atmospheric Measurement Techniques, 13, 747–754, https://doi.org/10.5194/amt-13-747-2020, 2020.

Trefalt, S., Germann, U., Hering, A., Clementi, L., Boscacci, M., Schroër, K., and Schwierz, C.: Operational radar hail detection algorithms at MeteoSwiss: quality assesssment and improvement, Tech. rep., MeteoSwiss, No. 284, https://doi.org/https://doi.org/10.18751/PMCH TR/284.HailClimateSwitzerland/1.0, 2023.

Treloar, A. B. A.: Vertically integrated radar reflectivity as an indicator of hail size in the Greater Sydney region of Australia, in: Proceedings of the 19th Conference on Severe Local Storms, Minneapolis, American Meteorological Society, pp. 48–51, 1998.

Waldvogel, A., Schmid, W., and Federer, B.: The Kinetic Energy of Hailfalls. Part I: Hailstone Spectra, Journal of Applied Meteorology and Climatology, 17, 515 – 520, https://doi.org/10.1175/1520-0450(1978)017<0515:TKEOHP>2.0.CO;2, 1978.

Waldvogel, A., Federer, B., and Grimm, P.: Criteria for the detection of hail cells, J. Appl. Meteorol., 18, 1521–1525, 1979.

Warren, R. A., Ramsay, H. A., Siems, S. T., Manton, M. J., Peter, J. R., Protat, A., and Pillalamarri, A.: Radar-based climatology of damaging hailstorms in Brisbane and Sydney, Australia, Quarterly Journal of the Royal Meteorological Society, 146, 505–530, https://doi.org/https://doi.org/10.1002/qj.3693, 2020.

Wetzel, E.: Made to measure, https://www.innetag.ch/wp-content/uploads/2020/10/HailSens-MTI_2018_09.pdf, 2018.

Witt, A., Eilts, M. D., Stumpf, G. J., Johnson, J. T., Mitchell, E. D. W., and Thomas, K. W.: An Enhanced Hail Detection Algorithm for the WSR-88D, Weather and Forecasting, 13, 286 – 303, https://doi.org/10.1175/1520-0434(1998)013<0286:AEHDAF>2.0.CO;2, 1998.

Wolfensberger, D., Gabella, M., Boscacci, M., Germann, U., and Berne, A.: RainForest: a random forest algorithm for quantitative precipitation estimation over Switzerland, Atmospheric Measurement Techniques, 14, 3169–3193, https://doi.org/10.5194/amt-14-3169-2021, 2021.

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

Table A1. Detailed description of the hailsensor (hs), hail report (r) and drone (dr) data available for the considered events. The time window T_{hs} represents the period between the first and last observation by any hailsensor during the same day, while the time window T_r is the period of analysis for hail reports, which extends beyond T_{hs}, and T_{dr} is the time window between the first and last image captured by the drone. For each dataset, the number of observations (N) and the minimum, median, and maximum diameters (D) are reported. For the event recorded by the drone on 28.06.2022, three experts performed manual annotations that served as training reference for the HSND-retrieval model. Consequently, the three models retrieved different numbers of hailstones and diameters, represented by the spread of values reported.

Date	T_{hs}	N_{hs}	D_{hs} [mm]	T_r [UTC]	N_r	D_r [mm]	T_{dr} [UTC]	N_{dr}	D_{dr} [mm]
2018.08.06	13:38–13:48	105 (2)	5, 6, 12	12:00–14:55	12	7.5, 7.5, 23	–	–	–
2019.06.15	15:05–17:05	281 (4)	5, 6, 12	14:00–18:55	564	0, 7.5, 32	–	–	–
2019.06.30	23:55–00:30	39 (1)	6, 6, 9	22:00–23:55	13	0, 7.5, 32	–	–	–
2019.08.02	01:16–01:32	63 (2)	5, 6, 8	00:05–02:55	16	7.5, 15, 23	–	–	–
2019.08.12	10:29–10:39	401 (4)	5, 7, 18	10:00–11:30	50	0, 22.5, 45	–	–	–
2020.06.26	12:05–12:13	58 (1)	6, 6, 10	08:15–13:55	0	–	–	–	–
2020.07.01	18:35–18:50	414 (3)	6, 8, 16	16:00–20:55	0	–	–	–	–
2020.09.25	06:13–06:20	846 (5)	5, 7, 14	06:00–08:55	261	0, 7.5, 32	–	–	–
2021.06.20	14:17–14:33	31 (1)	5, 6, 11	01:10–22:00	4977	0, 7.5, 45	14:38–14:41	18212	2, 8, 36
2021.06.21	14:30–15:56	1415 (13)	5, 6, 13	12:00–17:55	5250	0, 7.5, 45	–	–	–
2021.06.22	11:33–11:43	171 (3)	5, 6, 11	09:35–13:55	152	0, 7.5, 45	–	–	–
2021.06.28	–	–	–	00:05–23:55	18613	0, 7.5, 45	–	–	–
2021.07.08	09:40–09:53	1483 (9)	5, 8, 30	07:00–11:55	1971	0, 22.5, 45	–	–	–
2021.07.12	22:43–22:45	82 (1)	5, 6, 12	20:00–23:55	1484	0, 7.5, 45	–	–	–
2021.07.22	23:14–23:32	60 (1)	5, 6, 10	21:00–23:55	41	0, 7.5, 23	–	–	–
2021.07.24	14:04–14:32	579 (7)	5, 7, 17	12:00–16:55	843	0, 22.5, 45	–	–	–
2021.07.26	11:14–19:05	446 (3)	5, 6, 14	09:00–20:55	2655	0, 7.5, 45	–	–	–
2022.06.05	03:20–13:42	322 (4)	5, 7, 12	01:00–14:55	3576	0, 7.5, 45	–	–	–
2022.06.15	18:05–18:22	402 (7)	5, 7, 18	16:00–19:55	1099	0, 22.5, 45	–	–	–
2022.06.22	16:03–16:05	51 (1)	6, 7, 11	14:00–17:55	1450	0, 7.5, 45	–	–	–
2022.06.28	08:00–08:05	524 (3)	5, 7, 23	07:40–08:15	1018	3, 22.5, 45	08:18–08:20	4086-4419	2-3, 8, 22-23
2022.06.30	17:40–17:44	56 (1)	6, 7, 9	16:00–19:55	1557	0, 7.5, 45	–	–	–
2022.07.28	15:39–16:20	473 (5)	6, 7, 15	13:05–17:55	232	0, 7.5, 32	–	–	–
2022.08.30	15:47–15:49	32 (1)	5, 5, 7	14:00–16:55	87	0, 7.5, 32	–	–	–
2023.05.06	17:59–18:01	38 (1)	6, 6, 9	16:00–19:55	925	0, 7.5, 45	–	–	–
2023.06.22	22:30–22:40	38 (1)	6, 7, 12	22:00–22:55	0	–	–	–	–
2023.07.04	22:29–22:33	32 (1)	5, 6, 9	21:00–23:55	393	0, 7.5, 32	–	–	–
2023.07.12	–	–	–	00:00–23:55	6520	0, 7.5, 45	–	–	–
2023.07.25	01:14–01:24	191 (3)	5, 7, 14	00:00–02:55	191	0, 7.5, 45	–	–	–
2023.08.17	14:57–15:17	312 (4)	6, 7, 14	12:00–16:55	155	0, 7.5, 32	–	–	–
2023.08.24	15:53–16:00	44 (1)	6, 8, 20	13:00–17:55	1648	0, 7.5, 45	–	–	–
2023.08.25	19:23–19:31	124 (1)	6, 9, 42	17:00–21:55	1021	0, 32, 45	–	–	–
2024.07.31	13:31–13:50	33 (1)	6, 8, 12	13:00–13:55	421	0, 7.5, 45	–	–	–
2024.08.07	09:03–09:15	482 (4)	6, 7, 17	08:00–09:55	469	0, 7.5, 45	–	–	–
2024.09.01	14:32–14:41	45 (1)	6, 6, 15	14:00–14:55	0	–	–	–	–
2025.07.20	13:38–13:40	59 (1)	6, 7, 9	13:00–13:55	15	2.5, 2.5, 7	–	–	–
2025.09.05	09:16–09:17	40 (1)	6, 7, 15	09:00–09:55	530	2.5, 7.5, 45	–	–	–

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

Table A2. Cross-validation results. Spearman correlation (SCC) is reported between each feature included in the GAM model and the observed moments (M2, M4). Pearson correlation (PCC), root mean square error (RMSE), and standard deviation (STD) are reported between the GAM's estimated moments (log10-transformed, with and without a posteriori adjustment (adj)) and the observed moments (M2, M4). For reference, the STD of the observed moments is 0.39 for M2 and 0.56 for M4. Results are shown for leave-one-event-out cross-validation (cv) and testing on the training dataset (train). The selected features in order of inclusion are: feature 1: vol($\rho_{HV} < 0.97; Z_H > 50$ dBZ, 1 km radius), feature 2: sum($Z_V > 30$ dBZ, 1 km radius, vertical column), feature 3: max(Z_V, 1 km radius, vertical column), feature 4: vol($\rho_{HV} < 0.97; Z_H > 50$ dBZ, 1 km radius, above freezing level).

	Feature 1		Feature 2		Feature 3		Feature 4	
	M2	M4	M2	M4	M2	M4	M2	M4
$SCC_{feat, cv}$	0.59	0.62	0.60	0.60	0.60	0.58	0.55	0.59
$PCC_{GAM, cv}$	0.55	0.64	0.59	0.65	0.61	0.66	0.62	0.68
$PCC_{GAM, adj, cv}$	0.57	0.66	0.60	0.66	0.62	0.68	0.63	0.69
$RMSE_{GAM, cv}$	0.33	0.43	0.32	0.43	0.31	0.42	0.31	0.41
$RMSE_{GAM, adj, cv}$	0.37	0.46	0.35	0.46	0.34	0.45	0.34	0.44
$STD_{GAM, cv}$	0.23	0.36	0.24	0.37	0.25	0.37	0.25	0.38
$STD_{GAM, adj, cv}$	0.40	0.55	0.39	0.56	0.39	0.55	0.39	0.55
$SCC_{feat, train}$	0.59	0.63	0.60	0.59	0.60	0.58	0.55	0.59
$PCC_{GAM, train}$	0.58	0.67	0.62	0.69	0.64	0.70	0.66	0.72
$PCC_{GAM, adj, train}$	0.58	0.67	0.62	0.69	0.64	0.70	0.66	0.72
$RMSE_{GAM, train}$	0.32	0.42	0.31	0.41	0.30	0.40	0.30	0.39
$RMSE_{GAM, adj, train}$	0.36	0.45	0.34	0.44	0.33	0.44	0.32	0.42
$STD_{GAM, train}$	0.23	0.37	0.24	0.38	0.25	0.39	0.26	0.40
$STD_{GAM, adj, train}$	0.39	0.55	0.39	0.55	0.39	0.55	0.39	0.55

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

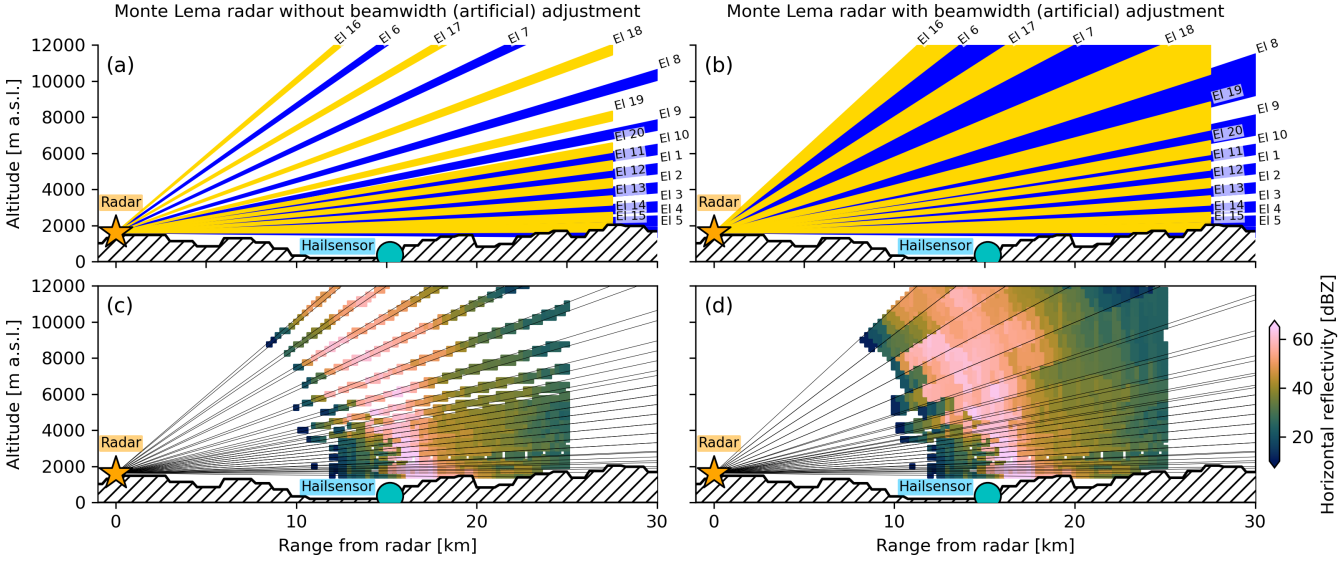

Figure A1. Vertical-cross section indicating the employed scan strategy without (a) and with the artificial half-beamwidth adjustment (b); and the resulting fields of horizontal reflectivity (c, d) for the event of 25.08.2023 between 19:20-19:25 UTC, along the azimuth pointing to the hailsensor located in Locarno-Monti. Radar elevations in blue (1-10) are performed during the first 2.5 min, while the yellow ones that partially overlap the others are performed in the following 2.5 min. Every cycle is thus completed in 5 min with the lowest scan pointing at -0.2° angle, and the highest at 40° angle.

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

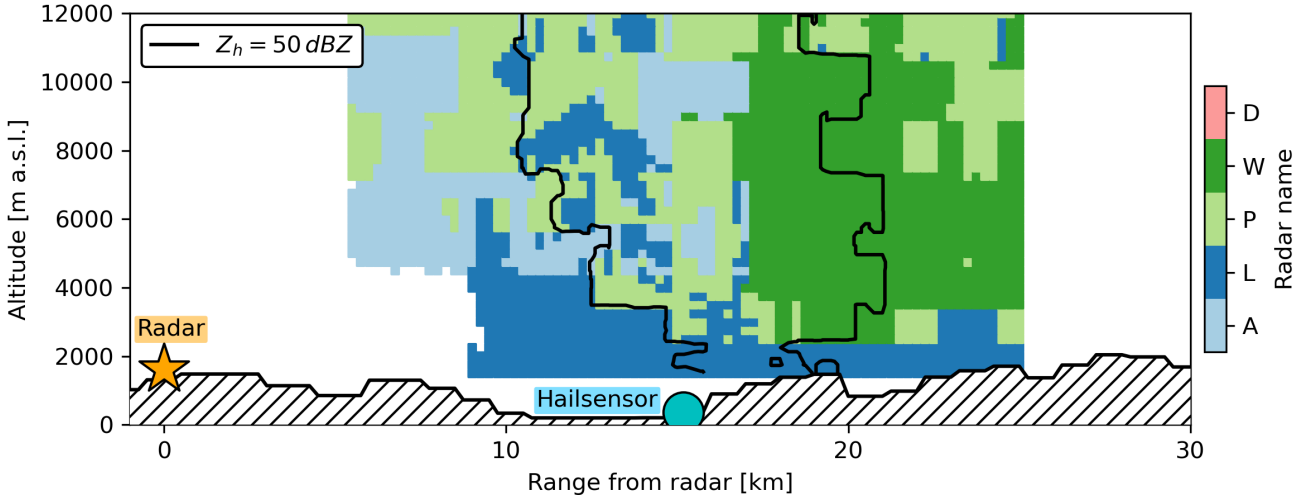

Figure A2. Vertical-cross section indicating the selected radar gates in the final 3D radar composite on 25.08.2023 between 19:20 and 19:25 UTC along the azimuth pointing to the hailsensor located in Locarno-Monti (D: La Dôle, W: Weissfluhgipfel, P: Pointe de la Plaine Morte, L: Monte Lema, A: Albis). The star indicates the position of Monte Lema radar. Observations from the La Dôle radar were not included because the distance to the hailsensor exceeds 160 km, which is the maximum radar range allowed for inclusion in the final composite.

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

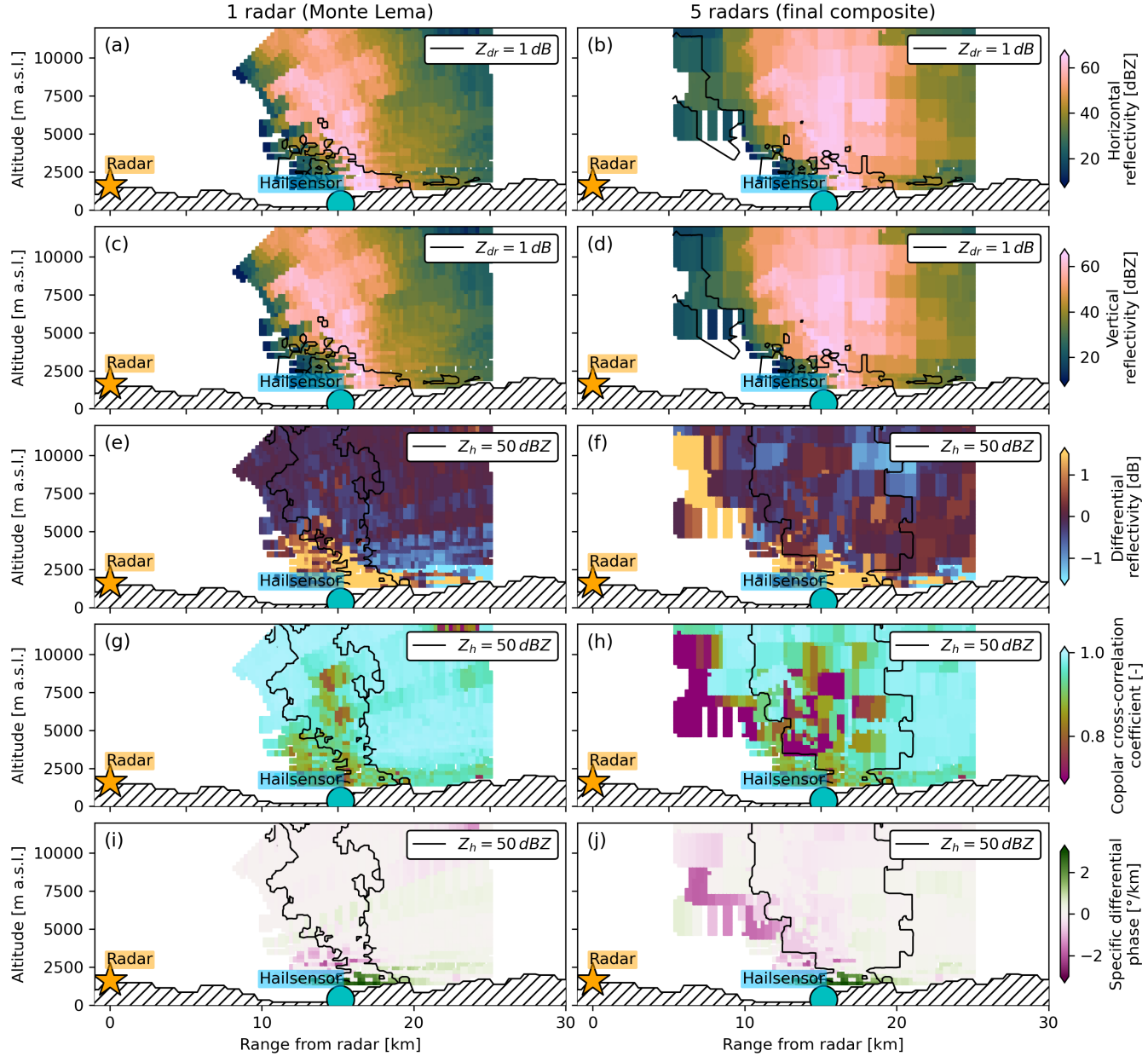

Figure A3. Vertical-cross section of polarimetric variables in the composite based on the Monte Lema radar only (left) and on the final 3D radar composite (right), on 25.08.2023 between 19:20 and 19:25 UTC along the azimuth pointing to the hailsensor located in Locarno-Monti: horizontal reflectivity (a, b), vertical reflectivity (c, d), differential reflectivity (e, f), copolar cross-correlation coefficient (g, h), specific differential phase shift (i, j).

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

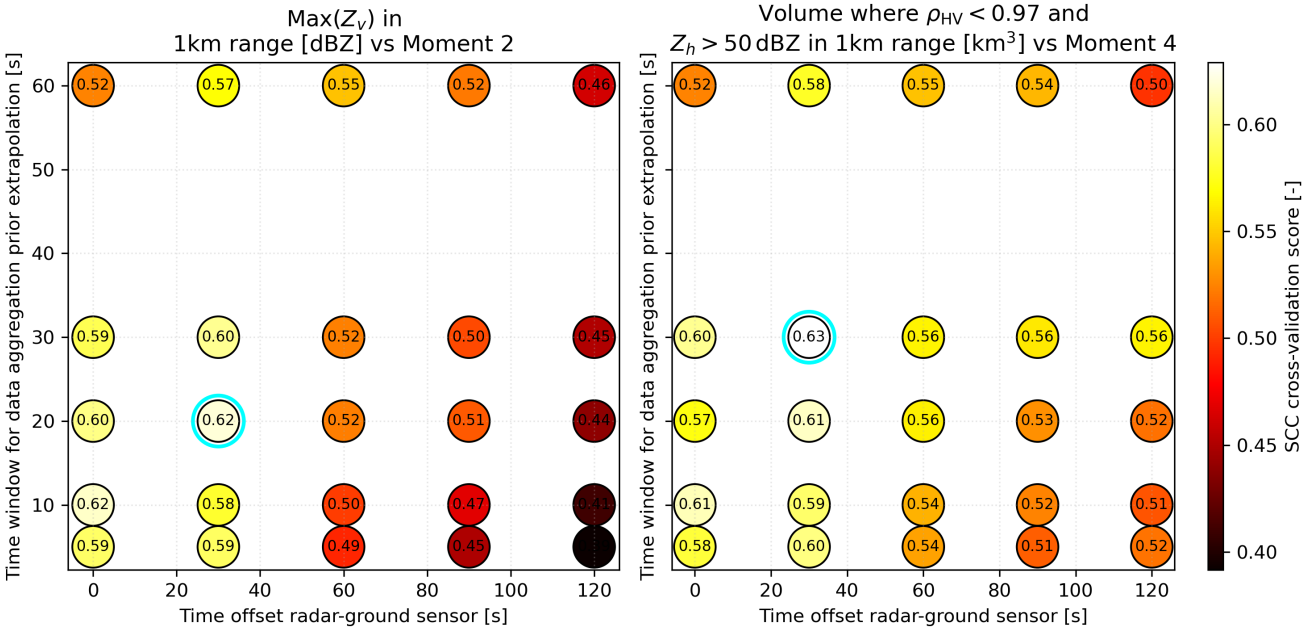

Figure A4. Search for the optimal combination of time offset between the radar observations aloft and the hailsensor observations on the ground, and the time window used to aggregate the sensor data prior spatial extrapolation. A leave-one-event-out cross-validation was performed to identify the combination yielding the highest average correlation across all folds, between the two radar-features having the largest Spearman correlation with moments 2 and 4. The best performance is obtained with a time offset of 30 s and a time window of 30 s for moment 4, and 20 s for moment 2.

https://doi.org/10.5194/egusphere-2026-594
Preprint. Discussion started: 1 July 2026

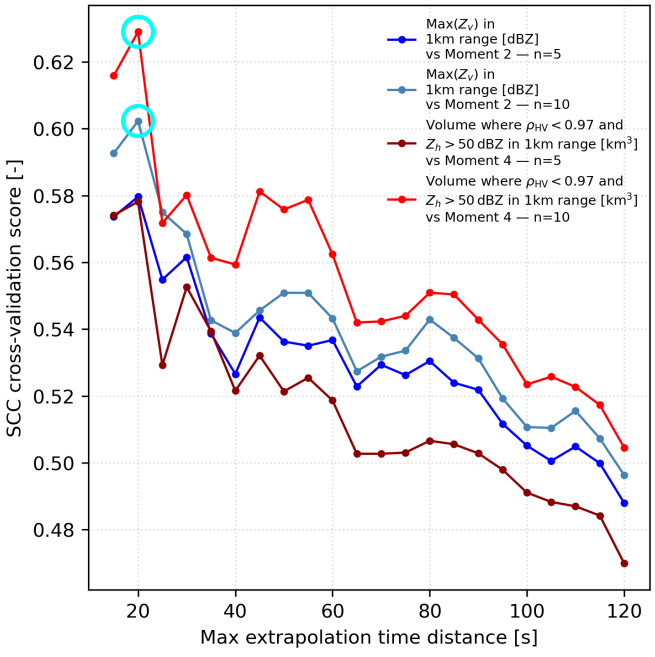

Figure A5. Search for the optimal combination of minimum number of samples (n) observed by a hailsensor for a 5-minutes timestamp, and maximum extrapolation time distance. The best performance is obtained with n = 10 samples and a maximum extrapolation time of 30 s for both features, based on the Spearman correlation (SCC) with moments 2 and 4. A leave-one-event-out cross-validation was performed to identify the combination yielding the highest average correlation across all folds.