

EXPERIENCES WITH >50,000 CROWDSOURCED HAIL REPORTS IN SWITZERLAND

HÉLÈNE BARRAS, ALESSANDRO HERING, ANDREY MARTYNOV, PASCAL-ANDREAS NOTI,
URS GERMANN, AND OLIVIA MARTIUS

Fifty-nine thousand crowdsourced hail size reports, gathered in Switzerland since May 2015, are presented, assessed, and compared to two operational radar-based hail detection algorithms.

THE HAIL OBSERVATION GAP. Hail fall in Switzerland at a specific location is infrequent, typically very localized, and characterized by a high spatial variability in hailstone sizes. That said, in the hail hot spots, hail occurs about 2–3 times per square kilometer per year (Nisi et al. 2016; Punge and Kunz 2016). As a consequence, ground observations require a very dense observational network and are therefore very expensive. Similar challenges exist for hail observations worldwide. Since the 1990s, researchers have attempted to fill the gap by involving the general public in gathering weather observations. Examples include the Community Collaborative Rain, Hail

and Snow Network (CoCoRaHS) in North America (Cifelli et al. 2005; Reges et al. 2016), the European Severe Weather Data Base (ESWD; Dotzek et al. 2009), the European Weather Observer application (app) (EWOB; Groenemeijer et al. 2017), and the Mobile Precipitation Identification Near the Ground Project (mPING) mostly in North America (Elmore et al. 2014). Ground observations are essential for developing, verifying, and improving indirect hail detection and hail size estimation algorithms based on remotely sensed data such as weather radar observations.

In Switzerland, two radar-based hail algorithms have been in operation since 2008: the probability of

AFFILIATIONS: BARRAS—Oeschger Centre for Climate Change Research, and Institute of Geography, and Mobiliar Lab for Natural Risks, University of Bern, Bern, and Federal Office of Climatology and Meteorology, Locarno, Switzerland; HERING AND GERMANN—Federal Office of Climatology and Meteorology, Locarno, Switzerland; MARTYNOV—Oeschger Centre for Climate Change Research, and Institute of Geography, University of Bern, Bern, Switzerland; NOTI—Environmental Service, Sion, Switzerland; MARTIUS—Oeschger Centre for Climate Change Research, and Institute of Geography, and Mobiliar Lab for Natural Risks, University of Bern, Bern, Switzerland

CORRESPONDING AUTHOR: Hélène Barras,
helene.barras@giub.unibe.ch

The abstract for this article can be found in this issue, following the table of contents.

DOI:10.1175/BAMS-D-18-0090.1

In final form 18 March 2019

©2019 American Meteorological Society

For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#).

hail (POH) and the maximum expected severe hail size (MESHS). They are used for nowcasting applications and for insurance loss estimates, and they were used to create the first Swiss radar-based hail climatology (Nisi et al. 2016). Recently, the algorithms were used to analyze the initiation and lifetime of hail cells and their swaths in complex topography (Nisi et al. 2018). A first verification of POH in Switzerland by Nisi et al. (2016) is based on insurance car loss data. Insurance loss data primarily provide information on the presence or absence of hail in areas with insured assets; the hail size is estimated from the damage type. However, the claims are often georeferenced to a ZIP code rather than to the actual hail event location. Spatially widespread information regarding the size of hail on the ground has so far been missing in Switzerland.

In May 2015, the Swiss Federal Office of Meteorology and Climatology (MeteoSwiss) started to fill this observational gap by simultaneously launching a pilot network of 11 automatic hail sensors and a hail size crowdsourcing function in the MeteoSwiss app. These hail sensors record the impact of individual hailstones on a Makrolon disk using piezo-electric microphones. The signal correlates positively with the kinetic energy and momentum of the hailstone, and thus, the hailstone diameter can be estimated from these measurements. For more information on the hail sensors, see Löffler-Mang et al. (2011). As of 2018, this pilot network is being extended to include a total of 80 automatic sensors that will measure the kinetic energy and momentum of hailstones for at least 8 years in the three hail hot spot regions of Switzerland (see Nisi et al. 2016).

The crowdsourced reports, the radar-based hail algorithms, and the automatic hail sensor network combine three sources of hail data that are of great complementary value. The radar hail algorithms provide automatic, spatially and temporally continuous estimates of the likelihood and size of hailstones at the ground. Automatic hail sensors have the advantage of measuring hail at the ground in a precise manner, but only at their exact location. The crowdsourced reports are numerous and account for much larger areas than automatic hail sensors, but provide subjective and less precise information of the true size of hail.

Trefalt et al. (2018) combined these hail data sources, as well as a newly developed dual-polarization radar-based hydrometeor classification (Besic et al. 2016, 2018), in a case study of an intense hailstorm in the northern Prealps. This case study showed good agreement between POH, MESHS, and the hailstone sizes sourced from the MeteoSwiss app. Kunz et al.

(2018) and Wapler et al. (2015) emphasized the benefit of combining multiple data sources in similar case studies on hail storms in Germany.

This article will introduce the MeteoSwiss crowdsourced hail reports, demonstrate a strategy to automatically filter them for plausibility, comment on their utility and limitations, and present a comparison to the two radar-based hail algorithms, POH and MESHS.

RADAR AND CROWDSOURCED DATA.

Radar-based hail products. We compare the reports with two operational radar-based hail algorithms, i) POH [Foote et al. (2005) based on Waldvogel et al. (1979)] and ii) MESHS [Joe et al. (2004) based on Treloar (1998)]. POH is a measure for the likelihood of hail occurrence, ranging from 0% to 100%. MESHS estimates the largest expected hail diameter in units of centimeters, starting at 2 cm. In Switzerland, POH and MESHS are used operationally and derived by combining freezing-level height information from the analysis (in real-time applications from the forecast) of the Consortium for Small-Scale Modeling numerical weather prediction model COSMO with the maximum height (echo top or ET) at which a radar reflectivity of at least 45 dBZ for POH (50 dBZ for MESHS) is detected (Donaldson 1961). Both algorithms are described in detail in sections 3.1 and 3.2 in Nisi et al. (2016). MESHS differs from the maximum estimated size of hail (MESH; Witt et al. 1998), a radar-based hail product that is commonly used in North America and that integrates the reflectivity greater than 40 dBZ above the melting layer. The ET information stems from the Swiss radar network which consists of five dual-polarization Doppler C-band radars. The radars scan the atmosphere at 20 elevations from -0.2° to 40° every 5 min (Germann et al. 2015, 2016). POH and MESHS 2D mosaic fields are available in real time every 5 min on a 1×1 km² Cartesian grid covering Switzerland and surrounding areas.

Crowdsourced data. The hail reporting function is part of the app of MeteoSwiss. It is included in the page that shows the radar precipitation fields in animated form, which is one of the most popular pages of the app (Fig. 1a). After passing a simple plausibility check, the hail reports are displayed seconds after they are submitted, overlaid on the radar echoes, and can be animated in time over the past 24 h. Users who observe hail can submit information on the time, location, and size of the hailstones. When a user submits a report, the current time and location of the phone are

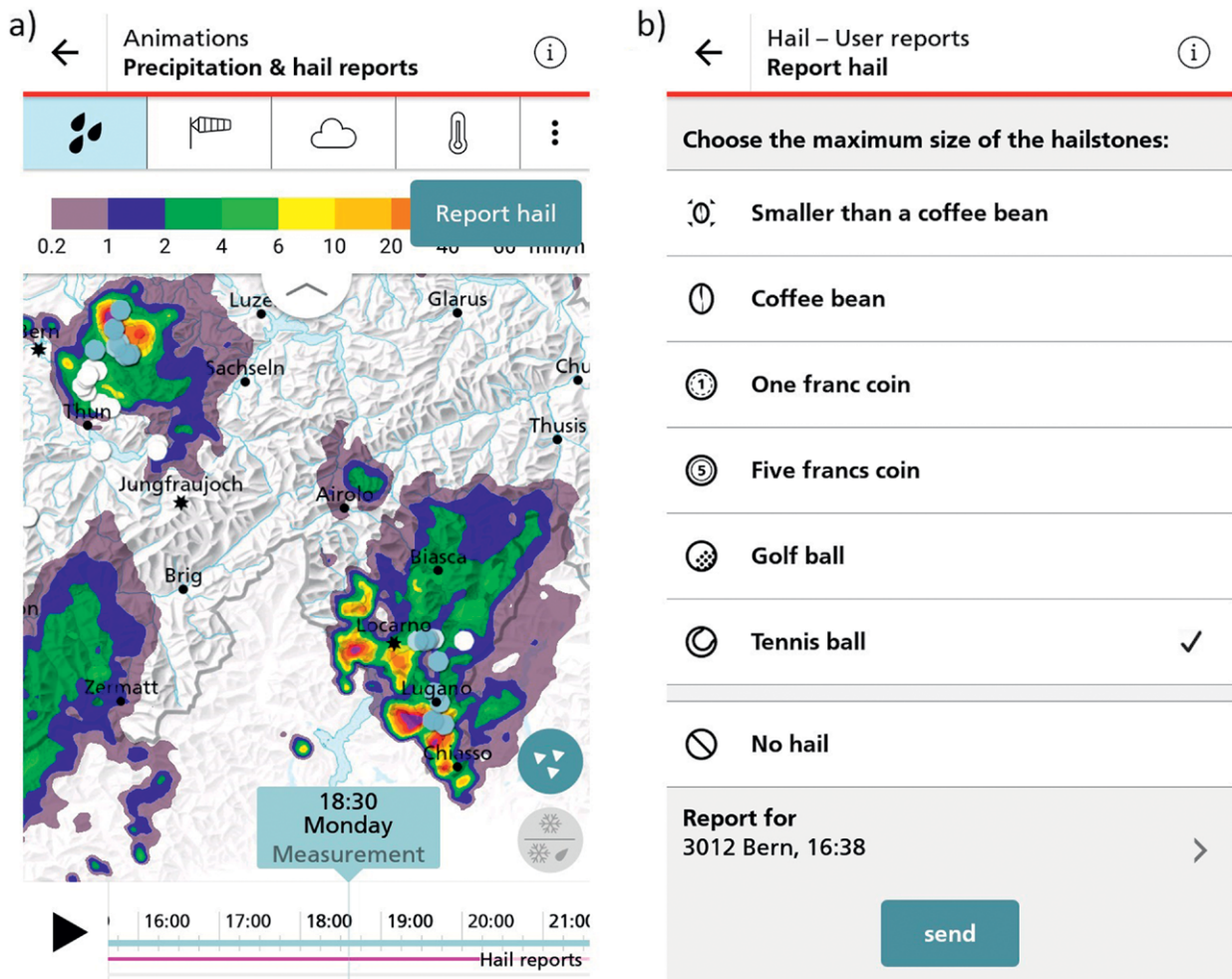


FIG. 1. (a) Screenshot of the animation with radar-based precipitation rates (mm h^{-1} , colors) and the crowd-sourced reports (blue and white dots) on 7 May 2018. The blue dots indicate hail reports for the shown time, white dots indicate past reports. (b) Screenshot of the hail size category scheme in spring 2018.

suggested as default input values, but both parameters can be adapted manually. The location information stems from position tracking by the smartphone. A manual adaptation of the location name and/or ZIP code will reduce the spatial accuracy by several hundred meters (depending on the size of the ZIP code area). The user can manually adapt the time by choosing the minute of the event. Knowing if the location and/or time were reported manually is important for filtering the reports. The user then chooses a size from a predefined hailstone size category scheme (Fig. 1b). Between May 2015 and September 2017, users could choose between the size categories “no hail,” “coffee bean,” “1 Swiss Franc coin (CHF),” “5 CHF,” and “>5 CHF” (see Table 1 for the corresponding diameters in millimeters). This original size category scheme was updated in September 2017 to include a “smaller than coffee bean” category, and the “>5 CHF” size was replaced with two categories, “golf ball” and “tennis

ball” (see Table 1). The “smaller than a coffee bean” category was added to differentiate between graupel ($<5 \text{ mm}$) and hail ($\geq 5 \text{ mm}$). The other two categories extend the range of categories to one that replaces “>5 CHF” and another larger size that mainly serves to catch suspicious reports. In spring 2018, an instruction was added requesting the users to report the largest hailstone size that they see. In addition to the location, the event time [time indicated by the user; in CEST (UTC + 2 h) in summer and CET (UTC + 1 h) in winter], and the hailstone size, the app stores the submission time (time at which the user presses “send”; Fig. 1b) and an anonymous user ID.

Note that users can also report “no hail.” The “no hail” reports provide valuable information in close proximity of a thunderstorm to delineate hail from no-hail areas. However, we do not include the “no hail” reports in this statistical analysis because we cannot use it to count false alarms, since we cannot

dismiss the possibility that hail did occur within the radar grid box ($1 \times 1 \text{ km}^2$) and 5-min time step corresponding to a “no hail” report. To simplify reading the article hereafter, we will refer to the reported categories in terms of hailstone diameter (Table 1). Note that each category spans a wider range of diameters that varies according to the chosen category scheme.

SUCCESSFUL DATA ACQUISITION. From 1 May 2015 to 31 October 2018, 59,020 MeteoSwiss crowdsourced hail reports were submitted by 39,733 different user IDs on 1,203 days over an area of $12,375 \text{ km}^2$ (with at least one report per square kilometer), which corresponds to a quarter of the Swiss territory. The dataset has 17,739 reports in the “no hail” category and 41,281 reports that indicate the presence of hail. More than 10 reports were submitted each day on 718 days, and more than 100 reports were submitted each day on 140 days. These are impressive numbers when compared to the small size and population of Switzerland. Crowdsourcing hail with the MeteoSwiss mobile app has been successful for several reasons. First, hail is a rare natural phenomenon that fascinates many people, is easy to recognize, and often interrupts people’s activities. Second, the crowdsourcing function is embedded in the radar animation of the MeteoSwiss weather app. This app is widely used, with an average of about 500,000 active users per day (of a population of approximately 8 million). Third, the MeteoSwiss mobile app has been downloaded more than 8 million times and is therefore the most popular weather app in Switzerland. The high number of users provides an unprecedented spatial and temporal observational coverage that could not be acquired differently given today’s observational methods, knowledge, and monetary restrictions. Last, MeteoSwiss deliberately publishes blog posts about the reporting function in the spring, at the beginning of the hail season, to encourage usage. The blog is part of the app and is popular.

An example of the crowdsourced data submitted on 31 May 2017 in the region of Thun is shown in Fig. 2. The 86 reports were mainly submitted from densely populated areas. Almost all of the 86 reports are located inside the POH $>80\%$ area. There are some reports of the largest size category collocated with MESH values $>60 \text{ mm}$. On this day, several reports were submitted within the same 1-km^2 grid box. From one grid box (see coordinate $614.5/178.5$ in Fig. 2), 17 reports were submitted within 26 min; the maximum number of reports per individual grid box ever recorded within 1 h. There is quite some variability in the reported hailstone sizes among the

TABLE 1. Original and current crowdsourcing hail report size category scheme, the corresponding approximate diameters, and range of diameters they cover.

Size category	Diameter (mm)	Diameter range (mm)
Original		
Coffee bean	5–8	$>0\text{--}15$
One Swiss franc coin	23	15–27
Five Swiss francs coin	32	27–32
Larger than five Swiss francs coin	>32	>32
Current		
Smaller than a coffee bean	$>0\text{--}5$	$>0\text{--}5$
Coffee bean	5–8	5–15
One Swiss franc coin	23	15–27
Five Swiss francs coin	32	27–37
Golf ball	43	37–55
Tennis ball	68	>55

17 reports, which points to subgrid-scale variability in the hailstone size, as well as to the uncertainty of the size estimates submitted by the app users.

CROWDSOURCED DATA ACQUISITION USING A GOVERNMENT APP VERSUS A CUSTOM APP.

While the wide distribution of the MeteoSwiss app is a huge advantage for the dissemination of the app and hence the number of reports, working with the government weather app has implications for the hail reporting options. The app is one main warning channel for the Swiss authorities and, therefore, the stability of the app has precedence over the reporting function. Every additional function imperils the stability and has to meet strict requirements. Hence, working with a custom app (e.g., mPING, EWOB) dedicated solely to gathering information on hail (or thunderstorms) would have the advantage of a substantial extension of the reporting options. The MeteoSwiss app is continuously being updated and improved, and one next step will be to provide an official Internet page informing on the hail reporting function. Suggestions for expanding the hail reporting function include the option of submitting photos and reporting the hail cover thickness, hail shape, hail size distributions, hail density, hailstone temperatures, event duration, or the damage caused. Such information would be very valuable; for example, Brimelow and Taylor (2017) verified the MESH algorithm with hail sizes estimated from photos posted in social media. In addition, quality control measures could be included in

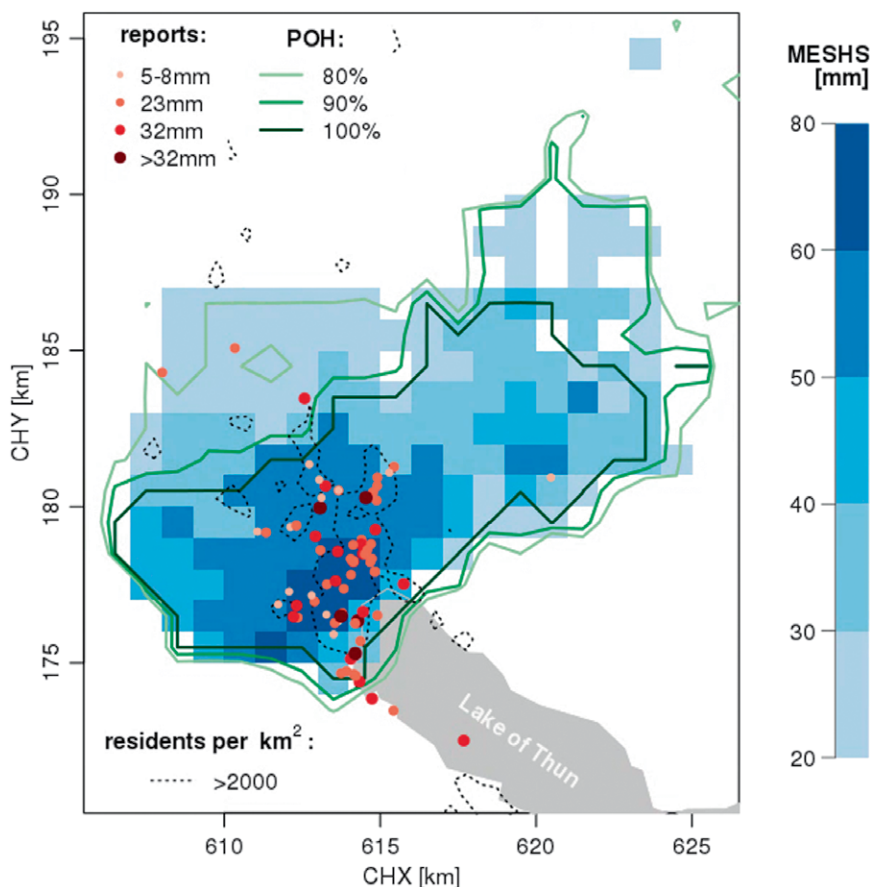


FIG. 2. Maximum values of MESHS (blue grid boxes, mm), POH (green contours, %), and the crowdsourced hail reports (red dots, see also Table 1) for 31 May 2017 in the region of Thun, Switzerland. The dashed dark gray lines contour the areas with more than 2,000 residents per 1 km² in the year 2017 (Federal Statistical Office of Switzerland 2017). The light gray area shows the Lake of Thun. The 32-mm report on the lake was probably submitted from a boat. The axes indicate the Swiss Coordinate System (km, tick marks every 5 km). The image is centered at approximately 46.802°N, 7.655°E.

a custom app, such as the option to submit an email address to contact people who submit reports for later verification.

QUALITY CONTROL OF THE CROWD-SOURCED REPORTS. *Plausibility filters.* The crowdsourced reports are influenced by human perception and sense of humor. This is why the crowdsourced data need to be quality controlled. Particularly for the comparison to the radar algorithms, erroneous reports need to be removed. We apply a multistep procedure to the 41,281 MeteoSwiss crowdsourced hail reports (excluding “no hail” reports) that is applicable in real time. First, we only keep the reports within an area that includes Switzerland and approximates the area that is well covered by the Swiss radar network (between 45.5°N, 5.6°E and 47.9°N, 10.7°E). This removes 479 (1%) reports.

Second, any duplicate of the same anonymous ID, time (rounded to 5 min), coordinate (rounded to 1 km), and size is removed, in case the same user repeats the same report within a few seconds. This criterion accounts for 724 (2%) reports.

We then apply a time filter and discard reports with more than 30 min difference between the submission time and event time. The reasoning behind this filter being that when people report hail hours after the event happened, they might not remember the size of the hailstones and/or the time of the hail event very accurately. This is also one of the reasons why the app suggests a size category scheme rather than allowing people to directly estimate the size in centimeters. This removes 3,195 (8%) reports.

Next, reports that are implausible due to the meteorological conditions are removed. This reflectivity filter requires a minimum radar reflectivity of 35 dBZ,

that is, a convective cell, to be located in the neighborhood of the report. The neighborhood method follows the so-called single observation neighborhood forecast verification (Ebert 2008) and filters the reports as follows: we consider all radar grid boxes, for all time steps between 15 min before and 15 min after the reported time, whose centers are within a radius of 4 km from the exact report location. In most cases, this temporal space includes six time steps. Figure 3 shows an example for two reports, using a radius of 2 km and for two time steps. Depending on the location of a report within its grid box, the spatial radius will include a different number of neighborhood grid boxes. The neighborhood accounts for the up to 2–4-km wind drift of hailstones (Schiesser 1990; Schmid et al. 1992; Hohl et al. 2002) and for a margin of error in the reporting time. This filter is based on radar information, as are the POH and

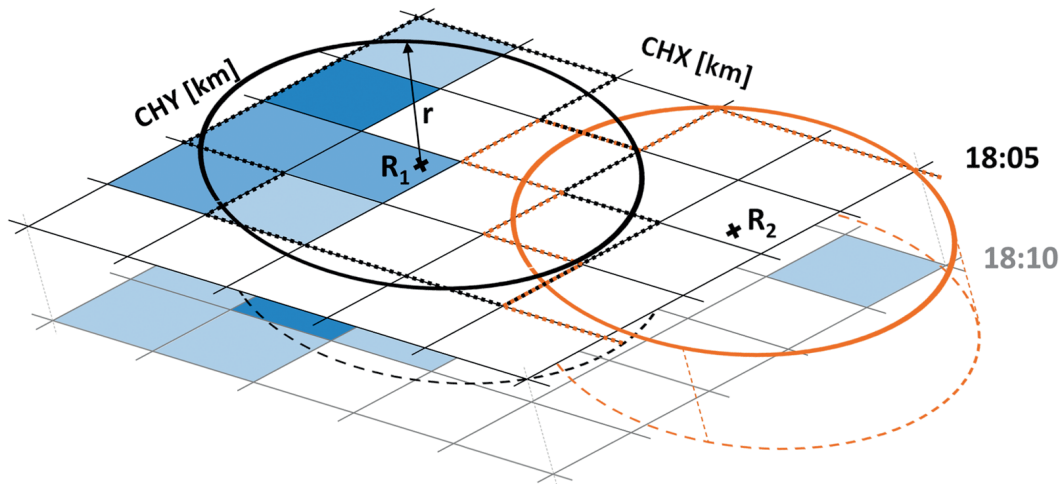


FIG. 3. Example radar grids at two time steps with two crowdsourced reports (R_1 , R_2). The blue grid boxes indicate nonzero radar values. The circles delimit the areas that are within a radius of 2 km around each report and the dotted lines show which grid boxes are within the neighborhoods defined by the circles.

MESHS products, and hence not fully independent. Unfortunately, there is no fully independent validation information available in Switzerland. However, the 35-dBZ threshold is smaller than the thresholds used to define POH and MESHS. This filter removes 16,892 reports, that is, 41% of the reports.

Next, reports by individual users with an unusual reporting pattern are removed. This includes reports from users with at least three reports of at least three different sizes, including the largest size category, within an hour. Furthermore, we filter reports if a

user submits more than three reports on the same day and chooses a different, manually adapted location for each report. The last filter removes reports in which the same user submitted <5–8- or 5–8-mm reports and the largest size category within 2 min. These filters remove 327 (1%) reports.

Note that the quality control is not based on the number of reports for an individual event. Indeed, there are many cases in which single reports from lightly populated, remote areas are plausible and even confirmed by independent reports. There are also

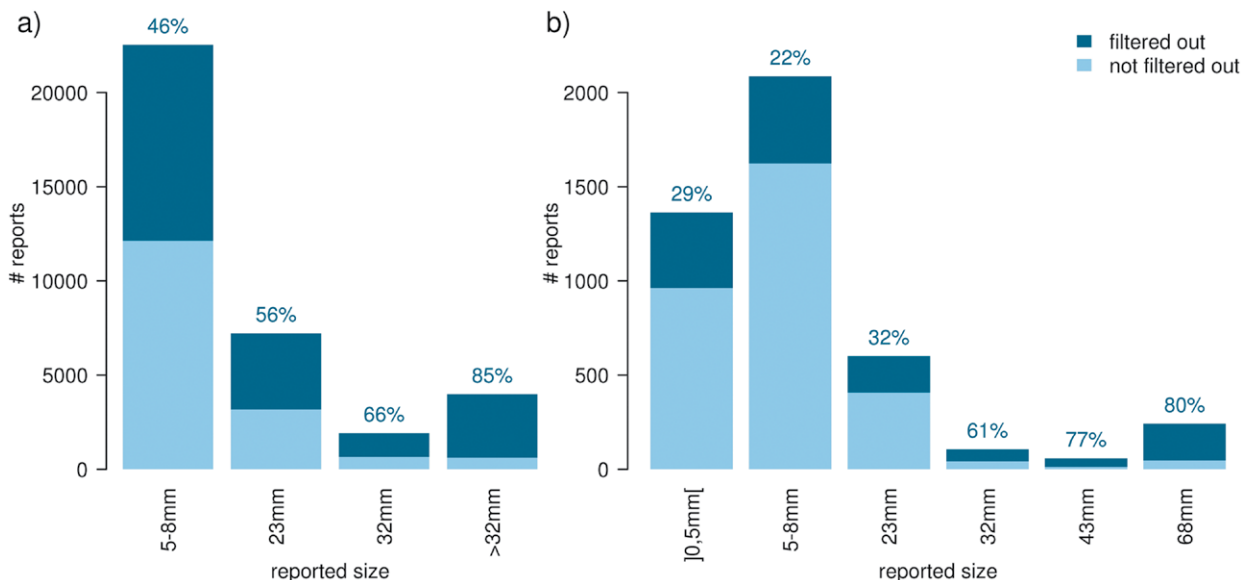


FIG. 4. Number of reports per size category with (a) the original size category scheme that were valid from May 2015 to Aug 2017 and (b) the current size category scheme (valid since Sep 2017). “No hail” reports are excluded. The fraction of reports that were filtered out are given as percentages above the bars and indicated by the dark blue color. Note the different y-axis ranges.

TABLE 2. Number of matches between the filtered reports and POH and MESHS for each reported category, considering the radar gridbox value containing the report (A) and a 2-km and 5-min neighborhood window around the report (B). Numbers in rows 2–5 are absolute numbers (percentage of filtered reports).

	5–8 mm	23 mm	32 mm	>32 mm	Total
Number of filtered reports	12,136	3,171	653	610	16,570
Matches with POH (A)	4,506 (37%)	1,598 (50%)	263 (40%)	60 (10%)	6,427 (38%)
Matches with POH (B)	6,593 (54%)	1,971 (62%)	317 (49%)	101 (17%)	8,982 (54%)
Matches with MESHS (A)	1,139 (9%)	719 (23%)	157 (24%)	29 (5%)	2,044 (12%)
Matches with MESHS (B)	2,717 (22%)	1,306 (41%)	231 (35%)	49 (8%)	4,303 (26%)

several cases in which the reflectivity filter identifies implausible reports clustered in a populated area.

The effect of the filters on the number of reports for each size category is shown in Fig. 4. Considering both size category schemes, 19,664 or 48% of all reports remain after filtering. For the original size category scheme (Fig. 4a), 16,570 or 40% of the reports remain. We refer to the remaining reports collected with the original size category scheme as the filtered reports. Of the 16,570 filtered reports, 12,136 (73%) are 5–8-mm, 3,171 (19%) are 23-mm, 653 (4%) are 32-mm, and 610 (4%) are >32-mm reports (see Table 2, top row, and Fig. 4a). The filters mainly reduce the number of >32-mm reports. This is expected, as the largest size category (until September 2017, >32 mm) might be chosen as a joke. Figure 4b shows the filter effects for reports submitted using the new size category scheme. Since the sample size is small and because of the change in category scheme, we do not compare the two histograms any further. The effects of altering hail reporting thresholds are discussed by Allen and Tippett (2015). The large fraction of filtered reports for the large size categories (43 and 68 mm) suggests that the filters are efficient, particularly since these reports were mostly submitted during the winter half year, when such large hailstones are almost impossible. However, more tennis ball (68 mm) reports than golf ball (43 mm) reports remain in the sample after filtering, which indicates that the filters do not remove all untrustworthy reports.

Almost 81% (13,420) of the filtered reports were submitted on 100 hail days.

The number of reports is greatest in the late afternoon and evening (Fig. 5), which reflects the typical thunderstorm diurnal cycle (e.g., Mandapaka et al. 2013; Nisi et al. 2018). The radar-based hail climatology (Nisi et al. 2016) indicates a second hail maximum at night that likely develops through down-valley winds and thunderstorm outflows converging in moist and unstable pre-Alpine air masses. This second maximum is not visible in the number of reports, most probably due to the general population being indoors. The spatial report density (not shown) primarily reflects the population density rather than the spatial hail frequency (see also Fig. 2) and therefore reflects the locations at which people and/or assets may be exposed, which is an advantage for hail risk studies.

Comparison of the MeteoSwiss crowdsourced hail reports with independent hail information. The network of hail sensors under construction already captured five events with graupel or very small hailstones and three events with maximum hail diameters >20 mm.

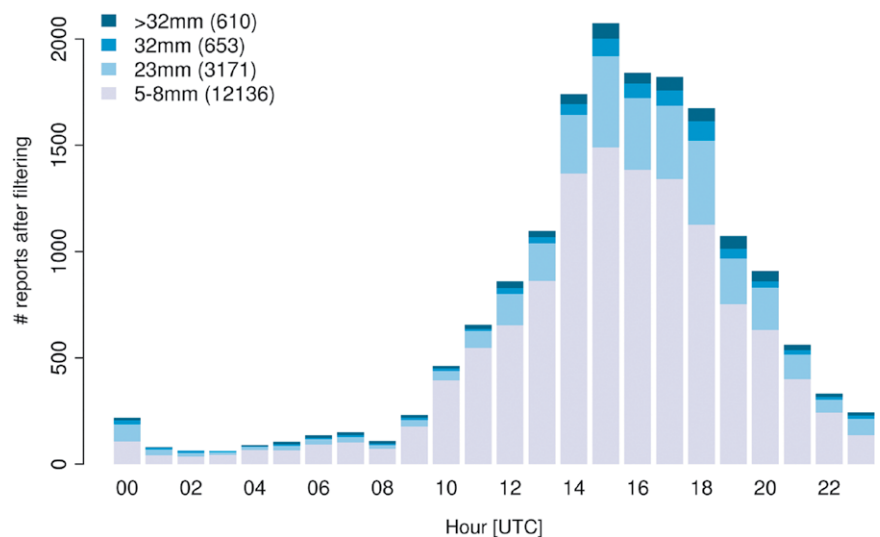


FIG. 5. Number of crowdsourced hail reports per hour of the day and per size category after filtering. The legend also shows the number of reports per size category.

During three of the five graupel/very small hail events, 1, 9, and 16 coffee bean reports (no reports of larger sizes) were recorded within a 2-km radius around the hail sensors. The MeteoSwiss crowdsourced hail reports submitted during the three hail events with hail diameters >20 mm, within 2 km of the sensors, were mostly equal to or larger than the diameters measured by automatic hail sensors. More events are needed to make a quantitative comparison.

Between May 2015 and July 2018, 110 filtered MeteoSwiss crowdsourced reports could be matched with 25 ESWD reports. For 21 cases, we found at least one MeteoSwiss report of the same size as the ESWD reports within the time uncertainty given by ESWD and within 2 km of the ESWD report. The MeteoSwiss crowdsourced reports for the remaining four cases indicated smaller hail diameters than the ESWD reports. In two cases, MeteoSwiss crowdsourced reports suggested larger hail sizes than indicated by ESWD. While the number of compared reports is too small for a conclusive statement, the results point toward the filtered MeteoSwiss crowdsourced reports being in good agreement with independent crowdsourced data.

COMPARISON WITH RADAR-BASED HAIL ALGORITHMS. *Matching the reports to POH and MESHS.* We match the filtered reports to nonzero POH and MESHS values. We use the 16,570 filtered reports received with the original size category scheme, as they constitute 84% of the sample. Again, we use a space and time neighborhood to match the reports with radar fields. Aside from the horizontal drift of hailstones, arguments can be made to allow for a margin of error in reporting time. Users might need to move themselves, a car, or flowers to safety before they report the hail fall, or they might not remember the time of the hail event. In addition, hail remains on the ground for some time before it melts and users might report hail on the ground rather than the hail fall. It therefore might be important to consider a spatial and temporal neighborhood to match the reports to the radar fields. To illustrate the sensitivity of the match to the chosen spatial and temporal neighborhoods, two methods are used to match the reports to the radar-based fields.

Method A assumes no spatial drift, an accurate reporting time, and uses the POH and MESHS values of the grid box and the 5-min

time step closest to the reporting location and time. Method B uses the maximum POH or MESHS value within a spatial neighborhood radius of 2 km and a temporal neighborhood of 5 min centered around the exact reporting location and time. This neighborhood method is identical to the method applied in the reflectivity filter, but with a different neighborhood size (Fig. 3). Note that with the neighborhood method several reports might be matched with the same radar value. Of all filtered reports (including both size category schemes and excluding “no hail” reports), 86% (16,815 out of 19,664) are single reports within the respective grid box and 5-min time step. In 61% of cases with more than one report within the same grid box and 5-min time step, one unique size category was reported. Most of the cases (72%) where at least two sizes were reported within the same grid box and 5-min time step are combinations of <5–8-, 5–8-, and/or 23-mm reports. Repeating the analysis with only the maximum reported sizes does not significantly alter the results, which is why we conducted the analysis with all reports and not just the maxima. Table 2 shows the number of matches with POH and MESHS per size category for both methods. As expected, method B produces more matches than method A. The reports matched with method B but not with method A include cases in which hail drifted.

A sensitivity study that considered neighborhoods ranging between 2 and 6 km and between 5 and 30 min revealed little sensitivity of the results. The largest changes in the results occur when going from no neighborhood (method A) to a small neighborhood (method B; Table 2). Compared to the number of additional radar grid boxes that are considered with a larger neighborhood size, the increase in fraction of matches is relatively small (Table 3).

Only 9% of the filtered 5–8-mm reports are matched with MESHS using method A (Table 2). Since MESHS includes only hailstones ≥ 2 cm, a very low number of MESHS matches is expected for the smallest size class. POH estimates the probability of

TABLE 3. Fraction of matches between the filtered reports and POH and MESHS considering different neighborhood sizes and the median number of grid boxes per neighborhood.

Neighborhood radii	2 km and 5 min (method B)	4 km and 5 min	4 km and 15 min
Matches with POH	54%	60%	67%
Matches with MESHS	26%	33%	41%
Median No. of grid boxes within the neighborhood	26	100	300

hail for hailstones of all sizes. Using method A, 37% of the 5–8-mm reports are matched with a POH signal, and using method B, 54% are matched. For the 23-mm category, 23% of the reports are matched with MESHS using method A, and 24% are matched for the larger 32-mm class (41% and 35%, respectively, with method B). Interestingly, the fraction of matched reports decreases substantially for the largest size class (5%, method A and MESHS). This fraction is also very low for POH and when using method B, which suggests that there is still a significant number of reports in this category that are likely “joke” reports.

Between 46% (method B) and 72% (method A) of the filtered reports cannot be matched with POH larger than zero (74% and 88% for MESHS). There are several possible explanations for this. First, the neighborhood used to match the reports (i.e., 2 km and 5 min) is much more restrictive than the one used to filter the reports (i.e., 4 km and 15 min). However, increasing the matching neighborhood size does not greatly increase the number of matched reports (see Table 3). If the spatial neighborhood radius was doubled to 4 km, which quadruples the number of considered grid boxes, the total fraction of matched reports increases from 54% to 60% for POH and from 26% to 33% for MESHS. If the temporal neighborhood radius is additionally increased from 5 to 15 min, the fractions further increase to 67% (POH) and 41% (MESHS), which still leaves 33% (POH) and 59% (MESHS) unmatched filtered reports.

Second, recall that POH and MESHS are defined using reflectivity thresholds (45 dBZ for POH and 50 dBZ for MESHS). There is therefore a 10-dBZ difference between the minimum reflectivity of the

filter (35 dBZ) and the required reflectivity for a POH signal. For 43% of the filtered reports that were not matched with POH using method B (39% for MESHS), the maximum reflectivity in the neighborhood was below the 45-dBZ threshold (50 dBZ for MESHS; not shown). It is therefore likely that hail (or graupel) can develop in Switzerland even if the radar reflectivity does not reach the threshold values of 45 or 50 dBZ. Third, the freezing-level height derived from the model influences the POH signal. The model may simulate a locally high freezing-level height stemming from the diabatic heating in a simulated thunderstorm cell. As a consequence, POH would be smaller or zero, since the distance between the freezing-level height and the maximum height with 45 dBZ would decrease. The same applies analogously to MESHS.

Fourth, the radar algorithms were fitted for convective thunderstorms happening during the summer season, and may miss events with graupel and/or small hail in the winter half year. The fraction of unmatched reports is much higher between October and April (88% for POH, 98% for MESHS with method B) than between May and September (38%, 71%). Another reason for the large fraction of unmatched reports in the winter half year may be that users mistakenly report sleet. Finally, despite the filters, we likely still have an unknown number of erroneous reports in our sample (see Fig. 6).

Evaluation of the MeteoSwiss crowdsourced hail reports.

The POH values of the matched reports increase with increasing reported size (Fig. 6a). Note that POH is not intended to provide any hailstone size information. In an ideal setting, POH would be independent

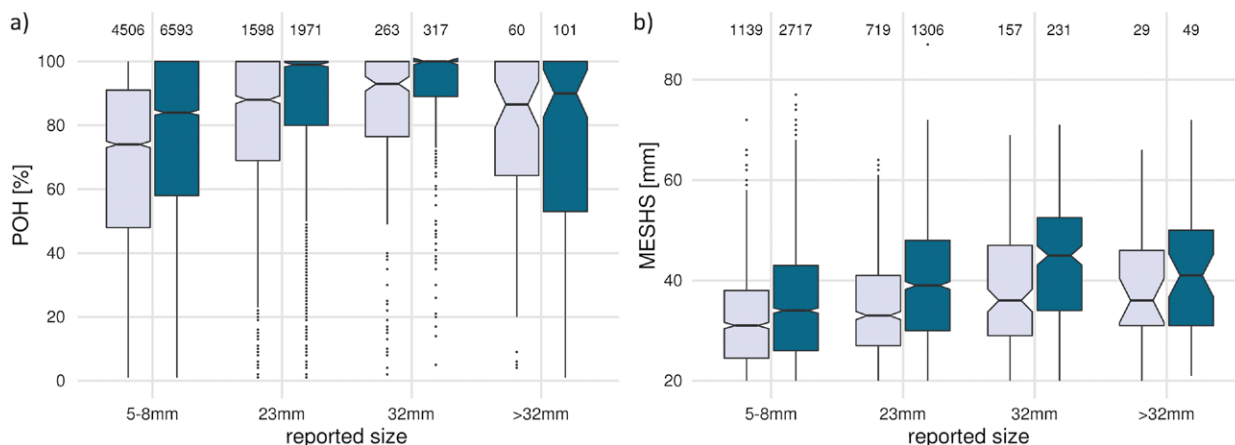


FIG. 6. Boxplots of (a) POH vs reported size and (b) MESHS vs reported size considering no neighborhood (light) and a neighborhood radius of 2 km and 5 min (dark). The numbers at the top of the plots indicate the number of values (matches) that contribute to each boxplot. The boxplots are Tukey-style whiskers with notches showing the 95% confidence interval for the median m , given by $m \pm 1.58 \times \text{IQR}/\sqrt{n}$ (McGill et al. 1978). IQR is the interquartile range and n is the sample size (see also Krzywinski and Altman 2014).

of the hail size. However, given how POH is defined, we expect POH to be higher for large hail sizes and lower for smaller hail sizes. Figure 4b suggests that the original 5–8-mm category includes reports of graupel. In the original scheme, “coffee bean” (5–8 mm) was the smallest available size category; the large fraction of “smaller than a coffee bean” reports in the current category scheme strongly suggests the presence of graupel in the original “coffee bean” category. This is consistent with the POH values for this category being significantly lower than the POH values for the larger hailstone size categories (Fig. 6a). Since the notches of the POH boxplots do not overlap when comparing the 5–8-, 23-, and 32-mm size categories, the increase in median POH with the reported size is significant. There is a significant difference between the medians of the method A and method B POH values when comparing the 5–8-, 23-, and 32-mm categories (non-overlapping notches and Mann–Whitney U test with p value of 0.05, not shown). Only a small fraction of reports are matched with small POH values. Using method A, only a quarter of the matched POH values are below 70% for the 23- and 32-mm reports. Using method B, only a quarter of the values are below 80%. Last, more than 50% of the matched POH values for the 23- and 32-mm reports and for method B have values >98%. The large interquartile and notch range of the POH values matched with >32-mm reports reflect the much smaller sample size and might indicate that this sample potentially still contains some incorrect reports despite the filtering.

MESHS values increase with increasing reported size (see Fig. 6b). This increase in the median is significant (Mann–Whitney U test, p value of 0.05) in all cases except when comparing the medians of 23 and 32 mm with >32 mm (for both methods A and B). This increase in median values (except for >32 mm) shows that the MESHS correctly recognizes the relative maximum expected size of hail above 2 cm. The interquartile ranges (IQRs) of MESHS span 1.5–2 cm. They approximate the size range that would be assigned to the reporting categories using the nearest neighbor (see Table 1). The constant IQRs suggest that the variance in MESHS is constant throughout the reported sizes.

When considering the 23- and 32-mm reports, MESHS is roughly 10–15 mm larger than the reported size, depending on whether method A or B is used for matching. The >32 mm and method B matching boxplot has lower quartile values than the boxplot of the MESHS values matched with 32-mm reports. As previously discussed, we assume that the matched sample likely still contains reports in which users

exaggerated the reported size. However, the lack of additional fully independent data prohibits a definitive statement if the users systematically overestimate the hail size of the largest reporting category or MESHS systematically underestimates the size. Since the sample of matched reports for >32 mm is very small in comparison with the other reporting categories, the incorrect reports have a larger influence. We therefore expect the quartiles to be larger once the sample size has reached several hundred reports. Once more 32-, 43-, and 68-mm reports are gathered, the IQRs for 32 mm and these larger categories can be meaningfully compared.

SUMMARY AND CONCLUSIONS. The crowdsourced hail reports gathered with the MeteoSwiss app constitute an extremely valuable observational dataset on the presence and approximate size of hail in Switzerland. This dataset has the advantage of unprecedented spatial and temporal coverage, and the automatic real-time processing and visualization is very convenient for nowcasting applications. Beside the scientific value of the dataset, we hope that the crowdsourcing function serves as a bridge between the general population and the world of research. This requires feedback from the scientists to the app users, which is currently provided through blog posts linked to the app and in newspaper articles. It will be extended in the future to include information on a dedicated website.

The reported hailstone sizes indicate that hail with a size close to the size of coffee beans is most abundant (note that this size category likely contains also reports of graupel). The number of reports follows the typical diurnal cycle of thunderstorm activity, with most reports being submitted in the early evening and evening. The spatial distribution of the reports primarily reflects the population density.

While the crowdsourced dataset dramatically increases the number of hail observations, they need to be quality controlled. Our reflectivity filter requires reports to be close to a radar reflectivity area of at least 35 dBZ. Overall, the plausibility filters remove approximately half of the reports in the dataset.

Our analyses suggest that except in the largest size category, enough false reports are filtered out for them to not substantially influence statistical analyses. The dense spatial and temporal coverage of the filtered reports allowed us to carry out a systematic comparison to the two operational, single-polarization radar-based hail algorithms, probability of hail (POH) and maximum expected severe hail size (MESHS). The fraction of unmatched reports

between May and September (38% for POH and 71% for MESHS; using method B) suggest that POH and MESHS are too restrictive in identifying hail areas. Of these unmatched reports, 43% (39% for MESHS) were submitted in an area with a maximum reflectivity between 35 and 45 (or 50 for MESHS) dBZ. Using a lower reflectivity threshold in the algorithms may therefore improve their quality. However, adapting the radar-based algorithms should entail a quantification of the false alarm rate, which cannot be achieved with the crowdsourced reports alone.

The positive correlation between reported sizes and the values of POH and MESHS suggest that the filters adequately separate plausible reports from improbable reports, except for the largest hail size category. Furthermore, the comparison of MESHS with the reported size shows that MESHS can be used as an estimate of the maximum size of hail >2 cm in terms of relative comparisons. Absolute MESHS values matched with the 23- and 32-mm categories exceed the reported hailstone size on average by 1.5 cm when a spatial neighborhood is considered to match the crowdsourced reports with MESHS values (method B). This difference merits further investigation using data from the hail sensor network. If the measurement campaign with the 80 new automatic hail sensors is successful, we will be able to test this conclusion and further improve the hail algorithms.

ACKNOWLEDGMENTS. The implementation of the crowdsourcing function in the app was enthusiastically supported by Markus Aebischer and Bertrand Calpini (MeteoSwiss) and was funded by the Mobiliar Lab for Natural Risks of the University of Bern thanks to the help of Matthias Künzler. Pascal-Andreas Noti and Andrey Martynov carried out a pilot study of the MeteoSwiss crowdsourced hail reports and the comparison to the radar algorithms in the framework of Pascal Noti's master's thesis (<http://occrdata.unibe.ch/students/theses/msc/I92.pdf>). We also thank Mattia Brughelli and Veronika Roethlisberger for their advice on the population data. Last but not least, we thank the three reviewers for their encouraging comments and their great suggestions for improving our article.

REFERENCES

- Allen, J. T., and M. K. Tippett, 2015: The characteristics of United States hail reports: 1955–2014. *Electron. J. Severe Storms Meteor.*, **10** (3), www.ejssm.org/ojs/index.php/ejssm/article/view/149/104.
- Besic, N., J. Figueras i Ventura, J. Grazioli, M. Gabella, U. Germann, and A. Berne, 2016: Hydrometeor classification through statistical clustering of polarimetric radar measurements: A semi-supervised approach. *Atmos. Meas. Tech.*, **9**, 4425–4445, <https://doi.org/10.5194/amt-9-4425-2016>.
- , J. Gehring, C. Praz, J. Figueras i Ventura, J. Grazioli, M. Gabella, U. Germann, and A. Berne, 2018: Unraveling hydrometeor mixtures in polarimetric radar measurements. *Atmos. Meas. Tech.*, **11**, 4847–4866, <https://doi.org/10.5194/amt-11-4847-2018>.
- Brimelow, J. C., and N. M. Taylor, 2017: 240 Verification of the MESH product over the Canadian Prairies using a high-quality surface hail report dataset sourced from social media. *38th Conf. on Radar Meteorology*, Chicago, IL, Amer. Meteor. Soc., 240, <https://ams.confex.com/ams/38RADAR/webprogram/Paper321272.html>.
- Cifelli, R., N. Doesken, P. Kennedy, L. D. Carey, S. A. Rutledge, C. Gimmestad, and T. Depue, 2005: The community collaborative rain, hail, and snow network: Informal education for scientists and citizens. *Bull. Amer. Meteor. Soc.*, **86**, 1069–1078, <https://doi.org/10.1175/BAMS-86-8-1069>.
- Donaldson, R. J., 1961: Radar reflectivity profiles in thunderstorms. *J. Meteor.*, **18**, 292–305, [https://doi.org/10.1175/1520-0469\(1961\)018<0292:RRPIT>2.CO;2](https://doi.org/10.1175/1520-0469(1961)018<0292:RRPIT>2.CO;2).
- Dotzek, N., P. Groenemeijer, B. Feuerstein, and A. M. Holzer, 2009: Overview of ESSL's severe convective storms research using the European Severe Weather Database ESWD. *Atmos. Res.*, **93**, 575–586, <https://doi.org/10.1016/j.atmosres.2008.10.020>.
- Ebert, E. E., 2008: Fuzzy verification of high-resolution gridded forecasts: A review and proposed framework. *Meteor. Appl.*, **15**, 51–64, <https://doi.org/10.1002/met.25>.
- Elmore, K. L., Z. L. Flamig, V. Lakshmanan, B. T. Kaney, V. Farmer, H. D. Reeves, and L. P. Rothfus, 2014: MPING: Crowd-sourcing weather reports for research. *Bull. Amer. Meteor. Soc.*, **95**, 1335–1342, <https://doi.org/10.1175/BAMS-D-13-00014.1>.
- Federal Statistical Office of Switzerland, 2017: Statistics of buildings and housing (StatBL). Subset: Total Population in the year 2017 (STATPOP2017). Federal Statistical Office of Switzerland, accessed 4 October 2018, www.bfs.admin.ch/bfsstatic/dam/assets/6027943/master.
- Foote, G. B., T. W. Krauss, and V. Makitov, 2005: Hail metrics using convectional radar. *Proc. 16th Conf. on Planned and Inadvertent Weather Modification*, San Diego, CA, Amer. Meteor. Soc., 1.5, https://ams.confex.com/ams/Annual2005/techprogram/paper_86773.htm.
- Germann, U., M. Boscacci, M. Gabella, and M. Sartori, 2015: Radar design for prediction in the Swiss Alps.

- Meteorological Technology International*, April, UKIP Media & Events, 42–45, www.ukimediaevents.com/publication/574f8129/44.
- , J. Figueras i Ventura, M. Gabella, A. Hering, I. Sideris, and B. Calpini, 2016: Triggering innovation: The latest MeteoSwiss Alpine weather radar network. *Meteorological Technology International*, April, UKIP Media & Events, 62–65, www.ukimediaevents.com/publication/2d183b22/64.
- Groenemeijer, P., and Coauthors, 2017: Severe convective storms in Europe: Ten years of research and education at the European Severe Storms Laboratory. *Bull. Amer. Meteor. Soc.*, **98**, 2641–2651, <https://doi.org/10.1175/BAMS-D-16-0067.1>.
- Hohl, R., H. H. Schiesser, and I. Knepper, 2002: The use of weather radars to estimate hail damage to automobiles: An exploratory study in Switzerland. *Atmos. Res.*, **61**, 215–238, [https://doi.org/10.1016/S0169-8095\(01\)00134-X](https://doi.org/10.1016/S0169-8095(01)00134-X).
- Joe, P., D. Burgess, R. Potts, T. Keenan, G. Stumpf, and A. Treloar, 2004: The S2K severe weather detection algorithms and their performance. *Wea. Forecasting*, **19**, 43–63, [https://doi.org/10.1175/1520-0434\(2004\)019<0043:TSSWDA>2.0.CO;2](https://doi.org/10.1175/1520-0434(2004)019<0043:TSSWDA>2.0.CO;2).
- Krzywinski, M., and N. Altman, 2014: Visualizing samples with box plots. *Nat. Methods*, **11**, 119–120, <https://doi.org/10.1038/nmeth.2813>.
- Kunz, M., U. Blahak, J. Handwerker, M. Schmidberger, H. J. Punge, S. Mohr, E. Fluck, and K. M. Bedka, 2018: The severe hailstorm in southwest Germany on 28 July 2013: Characteristics, impacts and meteorological conditions. *Quart. J. Roy. Meteor. Soc.*, **144**, 231–250, <https://doi.org/10.1002/qj.3197>.
- Löffler-Mang, M., D. Schön, and M. Landry, 2011: Characteristics of a new automatic hail recorder. *Atmos. Res.*, **100**, 439–446, <https://doi.org/10.1016/j.atmosres.2010.10.026>.
- Mandapaka, P. V., U. Germann, and L. Panziera, 2013: Diurnal cycle of precipitation over complex Alpine orography: Inferences from high-resolution radar observations. *Quart. J. Roy. Meteor. Soc.*, **139**, 1025–1046, <https://doi.org/10.1002/qj.2013>.
- McGill, R., J. W. Tukey, and W. A. Larsen, 1978: Variations of box plots. *Amer. Stat.*, **32**, 12–16, <https://doi.org/10.1080/00031305.1978.10479236>.
- Nisi, L., O. Martius, A. Hering, M. Kunz, and U. Germann, 2016: Spatial and temporal distribution of hailstorms in the Alpine region: A long-term, high resolution, radar-based analysis. *Quart. J. Roy. Meteor. Soc.*, **142**, 1590–1604, <https://doi.org/10.1002/qj.2771>.
- , A. Hering, U. Germann, and O. Martius, 2018: A 15-year hail streak climatology for the Alpine region. *Quart. J. Roy. Meteor. Soc.*, **144**, 1429–1449, <https://doi.org/10.1002/qj.3286>.
- Punge, H. J., and M. Kunz, 2016: Hail observations and hailstorm characteristics in Europe: A review. *Atmos. Res.*, **176–177**, 159–184, <https://doi.org/10.1016/j.atmosres.2016.02.012>.
- Reges, H. W., N. Doesken, J. Turner, N. Newman, A. Bergantino, and Z. Schwalbe, 2016: CoCoRaHS: The evolution and accomplishments of a volunteer rain gauge network. *Bull. Amer. Meteor. Soc.*, **97**, 1831–1846, <https://doi.org/10.1175/BAMS-D-14-00213.1>.
- Schiesser, H. H., 1990: Hailfall: The relationship between radar measurements and crop damage. *Atmos. Res.*, **25**, 559–582, [https://doi.org/10.1016/0169-8095\(90\)90038-E](https://doi.org/10.1016/0169-8095(90)90038-E).
- Schmid, W., H. H. Schiesser, and A. Waldvogel, 1992: The kinetic energy of hailfalls. Part IV: Patterns of hail-pad and radar data. *J. Appl. Meteor.*, **31**, 1165–1178, [https://doi.org/10.1175/1520-0450\(1992\)031<1165:TKEOHP>2.0.CO;2](https://doi.org/10.1175/1520-0450(1992)031<1165:TKEOHP>2.0.CO;2).
- Trefalt, S., and Coauthors, 2018: A severe hail storm in complex topography in Switzerland—Observations and processes. *Atmos. Res.*, **209**, 76–94, <https://doi.org/10.1016/j.atmosres.2018.03.007>.
- Treloar, A. B., 1998: Vertically integrated radar reflectivity as an indicator of hail size in the greater Sydney region of Australia. Preprints, *19th Conf. on Severe Local Storms*, Minneapolis, MN, Amer. Meteor. Soc., 48–51.
- Waldvogel, A., B. Federer, and P. Grimm, 1979: Criteria for the detection of hail cells. *J. Appl. Meteor.*, **18**, 1521–1525, [https://doi.org/10.1175/1520-0450\(1979\)018<1521:CFTDOH>2.0.CO;2](https://doi.org/10.1175/1520-0450(1979)018<1521:CFTDOH>2.0.CO;2).
- Wapler, K., F. Harnisch, T. Pardowitz, and F. Senf, 2015: Characterisation and predictability of a strong and a weak forcing severe convective event—A multi-data approach. *Meteor. Z.*, **24**, 393–410, <https://doi.org/10.1127/metz/2015/0625>.
- Witt, A., M. D. Eilts, G. J. Stumpf, J. T. Johnson, E. D. Mitchell, and K. W. Thomas, 1998: An enhanced hail detection algorithm for the WSR-88D. *Wea. Forecasting*, **13**, 286–303, [https://doi.org/10.1175/1520-0434\(1998\)013<0286:AEHDAF>2.0.CO;2](https://doi.org/10.1175/1520-0434(1998)013<0286:AEHDAF>2.0.CO;2).