Radar and Environment-based Hail Damage Estimates using Machine Learning

Luis Ackermann[1], Joshua Soderholm[1], Alain Protat[1], Rhys Whitley[2], Lisa Ye[2], and Nina Ridder[2]

[1]Australian Bureau of Meteorology, Melbourne, Victoria 3001, Australia
[2]Suncorp Group Limited, Brisbane, Queensland 4000, Australia

Correspondence: Luis Ackermann (luis.ackermann@bom.gov.au)

Abstract. Large hail events are typically infrequent, with significant time gaps between occurrences at specific locations. However, when these events do happen, they can cause rapid and substantial economic losses within a matter of minutes. Therefore, it is crucial to have the ability to accurately observe and understand hail phenomena to improve the mitigation of this impact. While in-situ observations are accurate, they are limited in number for an individual storm. Weather radars, on the other hand, provide a larger observation footprint, but current radar-derived hail size estimates exhibit low accuracy due to horizontal advection of hailstones as they fall, the variability of hail size distributions (HSD), complex scattering and attenuation, and mixed hydrometeor types. In this paper, we propose a new radar-derived hail product that is developed using a large dataset of hail damage insurance claims and radar observations. We use these datasets coupled with environmental information to calculate a Hail Damage Estimate (HDE) using a deep neural network approach aiming to quantify hail impact, with a critical success index of 0.88 and a coefficient of determination against observed damage of 0.780.79. Furthermore, we compared HDE to a popular hail size product (MESH), allowing us to identify meteorological conditions that are associated with biases on MESH. Environments with relatively low specific humidity, high CAPE and CIN, low wind speeds aloft and southerly winds at ground are associated with a negative MESH bias, potentially due to differences in HSD, hail hardness, or mixed hydrometeors. In contrast, environments with low CAPE, high CIN, and relatively high specific humidity aloft are associated with a positive MESH bias.

1 Introduction

Hail is a weather phenomenon that can cause substantial damage to crops, infrastructure, buildings, and motor vehicles (Gunturi and Tippett, 2017; Prein and Holland, 2018). It is crucial to accurately quantify and predict hail damage to enable farmers, insurance companies, and government agencies to make informed decisions and minimize the impact of hail events. The spatial coverage of a hailstorm, which can be measured estimated by hail size reports, remote sensing products, and/or the extent of insured damages, is of great importance for assessing the hail risk of an area. Analyzing the environmental characteristics associated with hailstorms has the potential to advance our understanding of hailstorm processes, microphysics, and prediction. By examining these factors during hailstorms, we can gain valuable insights into the dynamics and mechanisms at play, contributing to the broader knowledge in this field.

25   Despite its importance, accurately estimating the size of hail or the severity of hail damage remains a challenge. Currently, there are three main approaches for estimating hail severity: hail measurements at ground level, insurance claims data, and weather radar data. Direct observations of hail can be segregated into two categories, reports and in-situ measurements. Reports have biases related to population location, diurnal sampling bias, and size clustering. In-situ measurements like disdrometers or hail pads, are the most accurate but are generally sparse or deployed across small areas ~~Allen and Tippett (2015)~~(Allen and Tippett, 2015).

30   Insurance data is more widespread than in-situ measurements. However, it has ~~two~~ three main limitations. First, it is restricted to developed or populated areas where insured properties exist. Second, it only provides the cost of damages, which is highly dependent on the value of the property, how vulnerable is the property to hail damage, low-level winds, the Hail Size Distribution (HSD), and the density of hailstones (Giammanco et al., 2015). Third, it might be inaccessible due to policy or privacy concerns.

35   Radar-derived hail products have the advantage of high spatiotemporal resolution, increased homogeneity, and coverage, but ~~two~~ three challenges remain for accurate estimation of hail. First, the size distribution and concentration of hailstones cannot be derived from radar reflectivity alone. Reflectivity is the sum of contributions from individual hydrometeors, and therefore is highly dependent on both its size and concentration (Dennis and Kumjian, 2017b). Similarly, a mixture of liquid and frozen hydrometeors could be present in the volume and result in a size estimate with a positive bias. ~~Second, hailstones are advected~~
40   ~~by environmental and storm-generated winds during their descent, which can result in a mismatch between the radar-estimated hail location from observations aloft and ground observations.~~ The use of polarimetric radars can improve the quality of hail size estimates from radar by providing additional observations related to the size, shape, and orientation of the hydrometeors (Kumjian and Ryzhkov, 2008; Depue et al., 2007; Ortega et al., 2016). However, long term polarimetric radar observations required to create a hail climatology are lacking in most locations; and some areas lack polarimetric radar coverage. The second
45   limitation ~~can be~~ involves a potential mismatch between radar-estimated hail locations and ground observations. Hailstones can be transported by environmental and storm-generated winds during their descent, leading to discrepancies between the radar-estimated hail location based on aloft observations and ground-based reports. This limitation can be mitigated by modeling the trajectories of the hailstones (Brook et al., 2021), assuming three dimensional wind information is available; using only low-level information for hail estimation (Ortega et al., 2016; Depue et al., 2007); or by matching hail size reports or
50   insurance claims to radar-derived hail products within a defined spatiotemporal radius of influence (Warren et al., 2020; Nanni et al., 2000; Cintineo et al., 2012). However, these mitigation strategies have limitations: 3D wind observations are unavailable in most locations, and low-level information is only available close to the radar, which reduces coverage and is more prone to data quality issues such as ground clutter and beam blockage. Additionally, products that estimate hail from reflectivity above the freezing level are often not representative of conditions near the ground. Moreover, this matching requires a sufficiently
55   large sample of hail size reports or insurance claims. The third limitation pertains to the complete lack of information regarding hail hardness, which can have significant effects on the damage generated (Brown-Giammanco et al., 2021) as hailstones with different hardness levels may appear identical from the radar's perspective.

    Despite the limitations of radar-derived hail products, they remain the most effective tool for estimating hail occurrence, calculating hail risk climatologies, and providing situational awareness for operational forecasters. For example, the Aus-

tralian Bureau of Meteorology (BoM) uses the Maximum Expected Size of Hail (MESH) as guidance for issuing thunderstorm warnings (Richter and Deslandes, 2007). MESH was originally calculated by fitting the Severe Hail Index (SHI) to the 75th percentile of 107 maximum hail size reports using a power-law function (Witt et al., 1998). This was later improved by using a larger number of reports, 5897, and fitting to the 75th and 95th percentiles (Murillo and Homeyer, 2019). The SHI is a weighted vertical integration of hail kinetic energy above the melting layer, which is estimated from radar reflectivity (Witt et al., 1998).

In our study, we leverage MESH data from the BoM's national network of weather radars, combined with a 10-year dataset of hail damage insurance claims provided by Suncorp Group Limited (Suncorp) and meteorological data from ERA5 reanalysis (Hersbach et al., 2020), to train a deep neural network capable of predicting hail damage. The ~~manuscript describes the insurance data and applied post-processing (section 2);~~ structure and data flow of the study is:

section 2: Describes the insurance data, applied filtering, and normalization.

section 3: Describes the radar data and calculated products~~(section 3);~~.

section 4: Describes the procedure to match the insurance data to radar observations~~(section 4);~~.

section 5: Describes the development and evaluation of the neural network ~~(section 5); and finally the relationship between~~ driven by the matched insurance and radar data and aided by meteorological data.

section 6: The new hail damage model is applied to the full data archive, no longer limited by insurance exposure, and the relationship between the predicted hail damage, MESH, and meteorology ~~(section 6).~~ is discussed.

2   Insurance Data

~~Building-scale insurance data were provided by Suncorpfor all of Australia~~Suncorp, one of Australia's largest insurance companies, provided building-scale insurance data for the entire country. This dataset included information such as location, date of the event, ~~which included location, event occurred date~~ , sum insured, and incurred loss ~~. The dataset also provided information on the insured property's characteristics.~~ for residential policies. Additionally, it contained details about the insured properties, such as the year of construction, roof material, wall material, and the presence of tree coverage. For instance, a dataset entry might describe a house built in 1975 at a specific latitude and longitude, featuring a tiled roof, wooden walls, and tree coverage, with an insured sum of 700,000 AUD and an incurred loss of 70,000 AUD on January 1, 2010. Data were limited to areas covered by radar observation and encompassed the time period January 2010 - June 2022, which resulted in 311,196 individual damage claims. To ~~prevent any biases towards~~ avoid introducing biases toward more expensive properties, we ~~calculated~~ computed a damage metric ~~as the ratio~~ known as the loss ratio, which is the ratio of incurred loss to ~~insured sum also known as loss ratio, hereafter referred to as simply damage)~~ the insured sum. In this report, we will refer to this metric simply as damage and express it as a percentage. It is important to note that our study only investigates hail damage on residential buildings, while damage on vehicles was not provided due to uncertainty with the damage location.
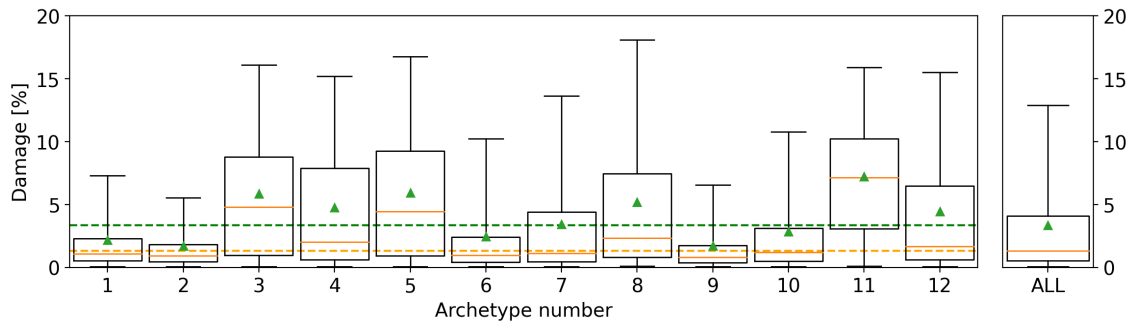
Figure 1. Relative damage distribution for each archetype. The black box shows the extent of the 25th and 75th percentiles of the data; the orange line shows the median, and the green triangle the mean. The whiskers show the 5th and 95th percentiles. The dashed green and orange lines show the mean and median of the full dataset, respectively.

Archetype normalization

We recognize that various property types can exhibit different levels of loss ratio from the same hail size and concentration (Blong, 2007; Brown et al., 2015; Hohl et al., 2002; Mobasher et al., 2022). For example, certain roof types are more susceptible to hail damage than others, and the presence of tree coverage can also influence observed hail damage. Accurate analysis and comparison to the radar estimates requires minimizing the influence of these variables on the observed damage. To achieve this goal, we identified all possible combinations of property characteristics (roof type, wall type, construction year, tree coverage) and selected the most frequently occurring combinations (archetypes), such that at least 85% of policies are represented. This resulted in 12 archetypes with varying levels of vulnerability, 11 archetypes accounted for more than 85% of the policies, with the rest labelled as other (the 12th archetype). The mean damage for each archetype and for the full data was calculated (Figure 1). The results indicated that some archetypes were over three times more vulnerable to hail damage than others. To ensure the overall archetype mean equalled the unscaled mean of the full data, the damage for each archetype was rescaled to match the overall mean (Figure 1, top-right). Note that the archetype details are commercially sensitive and are not shown in this study. The rest of the study uses the normalized damage instead of the original loss ratios, this way the effect from the different vulnerabilities from the various property types can be minimized.

Event identification

Insurance data were limited to hail damage claims but the percentage of days with claims in most radar domains did not accurately represent the true hail frequency, with most days recording at least one claim. The majority of these days recorded less than ten claims each, while a small number of days saw claim spikes in the thousands (Figure 2). To address this issue, we defined intense hail events as days with at least 1000 claims and days with between 500 and 999 claims were classified as medium hail events. A close look at the days surrounding these events show increased claim counts above the baseline starting around one week before and returning to baseline about one week after the peak, potentially due to mislabeling of the date or
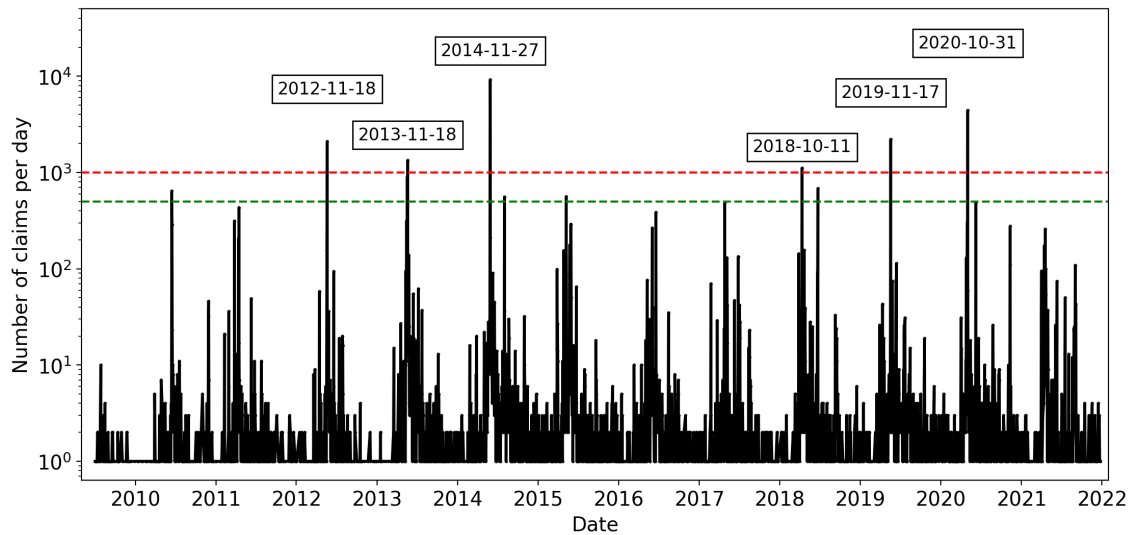
Figure 2. Time series of daily claims for the Brisbane radar domain (Mt Stapylton). Events with more than 999 claims are labelled. A maximum distance of 150 km to the radar was used. The green and red dashed lines show the 500 and 1000 thresholds, respectively.

errors in the reported date of damage. To mitigate this, claims ~~withing~~ within ± 7 days from each peak ~~where~~ were included in the event's dataset.

Claims filtering

To ensure that only regions with sufficient exposure and a substantial amount of damaged properties were included in the analysis, we applied several filters to the insurance dataset~~underwent filtering. To remove areas with insufficient exposure, policies were considered only if they were~~.

1. **Exposure Filter:** Policies were considered valid during an event ~~and~~ only if at least 10 other policies were also valid within a ~~1 radius surrounding the policy (policies within a circle of 1 radius centered at the policy to be evaluated). To identify areas of substantial damage, damage claims were only considered if~~ 1-kilometer radius of the policy in question. This filter helped eliminate areas with insufficient exposure.

2. **Damage Filter:** Damage claims were retained only if they were associated with regions where at least 5% of ~~the~~ policies within a ~~1 radius also reported damage, otherwise the claims were removed. To identify areas of no damage, if~~ 1-kilometer radius reported damage. Otherwise, these claims were excluded. This filter aimed to identify areas with substantial damage.

3. **No-Damage Filter:** If a policy did not report damage~~but~~, it was removed from our analysis if more than 1% of neighboring policies (within a ~~1~~ 1-kilometer radius) reported damage~~, the policy was removed.~~. This step ensured that areas with less than 1% of properties reporting damage were classified as "no-damage" areas. One of the reasons to exclude
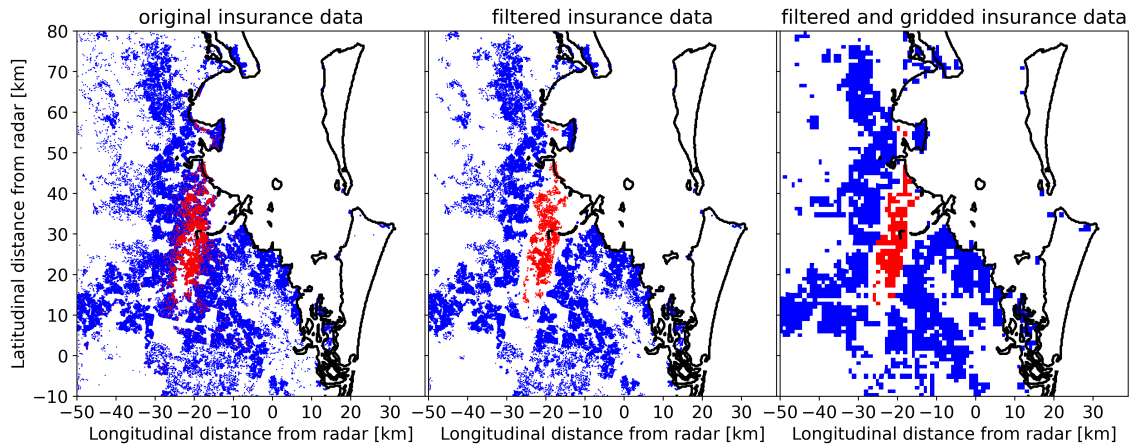
Figure 3. Effect of filtering insurance data; red points show valid policies with reported damage, blue dots show valid policies with no reported damage and white spaces show no valid policies. Left panel shows the raw data, middle panel shows only policies after filtering, and the right panel shows the gridded filtered data for the event.

properties without damage within an area with significant damage is because hailfall can be only a small portion of a 1 km square grid; including undamaged properties would artificially lower the actual damage created by hail when
130   averaged. We do recognize that this method might prevent the identification of certain properties that, due to their low vulnerability, do not record damage even when exposed to hail.

The difference between the two ~~percentages~~ percentage thresholds ($\geq$ 5% for damage areas and $\leq$ 1% for undamaged areas) ~~creates~~ created a "buffer zone" where the occurrence of hail damage was considered uncertain. Once ~~the above filtering was done~~these filters were applied, the insurance data ~~was gridded such that it exactly match the radar's~~ were gridded to match
135   the radars' grid for each domain ~~. This was done by~~ precisely. This involved calculating the mean damage within ~~a borders of each radar's~~ the boundaries of each radars' grid box. ~~Note that this results in~~ It's important to note that grid points with mean damage above zero ~~that is only representative of the~~ only represent damage claims, as undamaged policies ~~would be removed from such ares~~were removed from these areas. These filtering thresholds were ~~identified empirically and can be varied by~~ determined empirically and were found to have minimal impact on the analysis when varied by up to 10%~~without noteworthy~~
140   ~~effect on the analysis~~. Figure 3 ~~illustrates~~ provides a visual representation of the effect of this filtering and the subsequent gridding. ~~Applying all aforementioned filtersresulted in~~ After applying all the filters, our dataset consisted of 18 intense hail events ~~, and~~ 12 medium hail events~~, ;~~ these provided 1,~~557~~ 775 damage grid points, and ~~45,209~~ 76,703 exposed but undamaged grid points. On average, the filtering ~~removed~~ process removed approximately 21.4% of damage claims and 12.3% of exposed but undamaged policies.

Table 1. Australian S-band radar network details.

Radar ID	City	Radar type	Start date	Latitude	Longitude	Beamwidth
2	Melbourne	Meteor1500SDP	07/09/1993	-37.8553	144.7555	1
3	Wollongong	DWSR8502S	22/10/1995	-34.2624	150.8751	1.9
4	Newcastle	DWSR74S	07/09/1999	-32.7298	152.0254	1.9
8	Gympie	DWSR8502S	08/11/1999	-25.9574	152.5768	1.9
23	Gladstone	WSR74S	24/12/1998	-23.855	151.2626	1.9
28	Grafton	WSR74S	23/09/1998	-29.622	152.951	1.9
40	Canberra	DWSR74S	22/11/2002	-35.6614	149.5122	1.9
50	Brisbane	WSR74S	04/11/1994	-27.608	152.539	1.9
64	Adelaide	Meteor1500SDP	27/10/2005	-34.6169	138.4689	1
66	Brisbane	Meteor1500SDP	08/06/2006	-27.7178	153.24	1
69	Namoi	DWSR8502S	02/06/2010	-31.0236	150.1917	1.9
71	Sydney	Meteor1500SDP	15/05/2009	-33.7008	151.2094	1
72	Emerald	DWSR8502S	09/03/2010	-23.5498	148.2392	1.9
75	Mount Isa	DWSR8502S	14/09/2012	-20.7112	139.5552	1.9

3 Radar Data, hail estimate products, and associated meteorology

In this study, data from the Bureau's radar network was utilized (Soderholm et al., 2022), with a focus on selecting S-band radars that are better suited for hail observations compared to C-band radars (Ryzhkov et al., 2013). Although this led to a reduction in the number of radar sites used in the analysis, it still covered a large proportion of the population and claims. The geographical distribution of these radars can be observed in Figure 4, and configuration details can be found in Table 1. Note that only the S-band radars in solid blue circles in Figure 4 were used in relation to the insurance claims as the other S-band radars (dashed lines) did not record hail events within their domains, or their domains were already covered by other S-band radars. Nevertheless, these other S-band radars, in dashed circles in Figure 4, were used later in the study in subsection "HDE - MESH relationship" and in section 6. The radar reflectivity was calibrated following the method outlined in Louf et al. (2019), and gridded at a 1 km horizontal and 500 m vertical resolution utilizing the methodology described in Dahl et al. 2019, appendix A, was implemented, which uses linear interpolation in elevation, and radius of influence on the azimuth-range space. Grid points too far (> 150 km) or too close (< 6 km) to the radar site were excluded from the dataset.

Severe Hail Index (SHI)

The SHI quantifies hail severity using a weighted vertical integration of reflectivities above the environmental freezing level (Witt et al., 1998). The resulting output is a 2D gridded map which is indicative of the severity of the hail event. ~~The~~ In order to
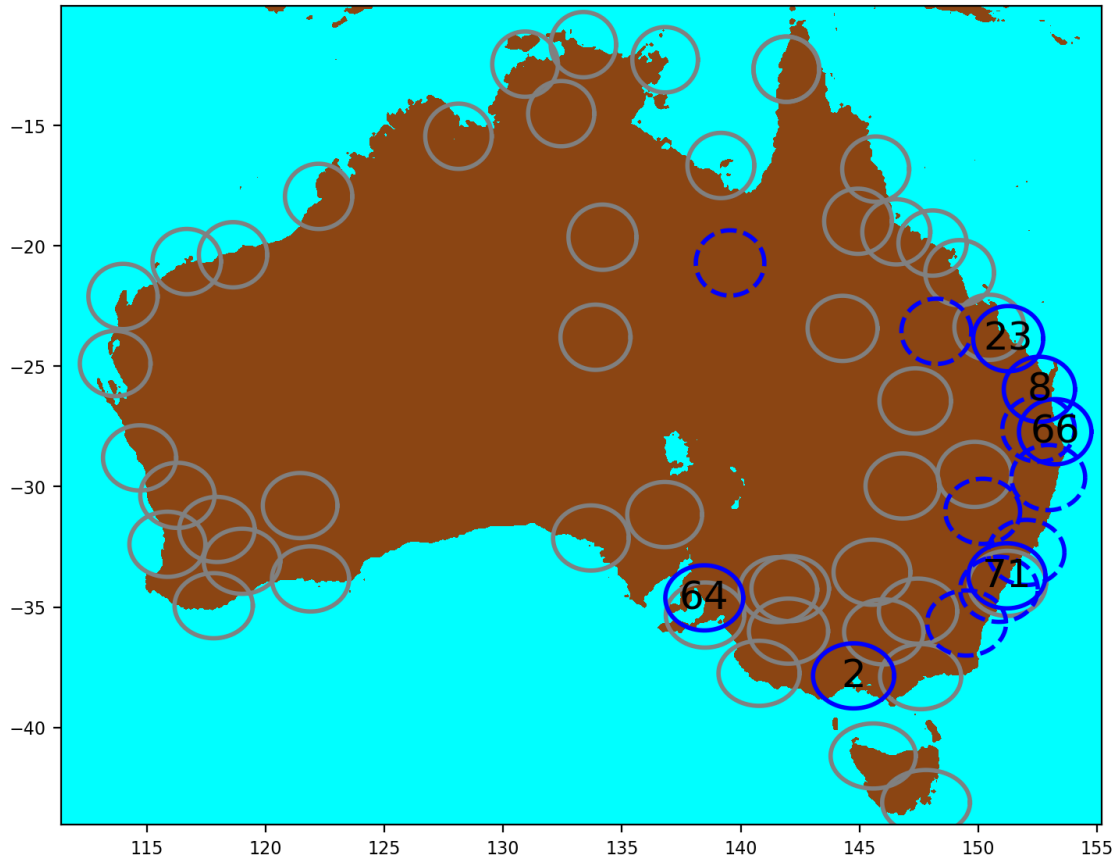
Figure 4. Map of the full Bureau's radar network. The circles ~~show~~ represent the coverage of each radar~~and the number refers~~, with numbers corresponding to the radar ID ~~found~~ as listed in Table 1. ~~S-band~~ C-band radars are shown in grey and S-band radars are blue; with solid blue circles ~~. Blue circles show~~ indicating the radars used in ~~association~~ conjunction with insurance claims. The dashed S-band radars recorded no hail events occurring within their domains and were not used in most of the study except section 6 and the "HDE - MESH relationship" subsection, as these do not involve claims.

160 calculate SHI, knowledge of the height of the $0°$ C and $-20°$ C dry bulb levels ~~were~~ are needed, as the integration of reflectivities is done only between these heights. This information about the temperature profile was retrieved from ERA5 (Hersbach et al., 2020) from the grid point closest in time to the observation and location to the radar location.

Maximum Estimated Size of Hail (MESH)

The Maximum Estimated Size of Hail (MESH) is a quantitative tool that transforms SHI into hail size by fitting SHI to a chosen
165 percentile of maximum observed hail size by using a power curve originally developed by Witt et al. (1998) and improved by

Murillo and Homeyer (2019) with a larger report dataset. For this study we use the 75th fit for the Murillo and Homeyer (2019) dataset. This transformation enables the estimation of hail size from the SHI data.

Calculation of event SHI swath

The event SHI was computed for each grid point within a radar's coverage for the event's day $\pm$ one day to account for possible discrepancies between the reporting time of damage. This , which has at most a daily accuracy, and the time of radar observations. This extra day on both sides of the event's date prevents situations when a hailstorm swath occurs between two days, and therefore be in the next day (or previous day, depending on which day people report most claims), and therefore only part of the swath would be used. Since the maximum SHI for each grid point is being taken, and no two events occur one after the other in the dataset, the addition of the two extra days has no detrimental effect but solves any of the aforementioned timing issues. This computation enabled the production of a storm swath map for each event day (Figure 5). The volume scan period for radars used in this study ranges from 5 to 10 minutes, leading to discontinuous SHI swaths, which are most pronounced for fast-moving storms (Figure 5, left panel). To mitigate this spurious discontinuity, an interpolation algorithm was employed, which uses the estimated field advection from its optical flow to fill in these gaps (Figure 5, right panel). This is done using the scipy python package, specifically the minimize method of the optimize class which applies the Nelder-Mead minimization algorithm (Nelder and Mead, 1965). This algorithm compares two subsequent scans (images) and attempts to minimize the difference between the two images by displacing one and comparing to the other. Once the optimal displacement is computed, a linear interpolation between the two time stamps is calculated. Once the interpolation between scans is computed, the maximum SHI for each grid point is retained from the two original time stamps and the interpolation, along with its corresponding timestamp, allowing for the retrieval of the associated meteorological conditions from ERA5.

Associated meteorological data

The calculation of SHI required information on the 0° and -20° dry bulb levels, which is derived from the ERA5 pressure level dataset for each radar location. In addition to these variables , other meteorological variables Meteorological variables were extracted from ERA5 at ground level and freezing height, these wereincluding: 3D wind components, specific humidity, divergence, vorticity, atmospheric temperature, and atmospheric pressure. CAPE and CIN (most unstable Convective Available Potential Energy and Convective Inhibition, respectively), were also extracted; for more details on the calculation of these variables see Groenemeijer et al. 2019 All these variables were retrieved for each grid point at the time when the maximum SHI of the day occurred. These data served as the input for training the neural network in addition to SHI and the observed damage.

4 Claims and Radar Matching – Virtual Advection

Hail retrievals from radar observations aloft often do not align with observations at ground level, resulting in a common mismatch (Brook et al., 2021). MESH/SHI are calculated from reflectivities above the environmental freezing level, which
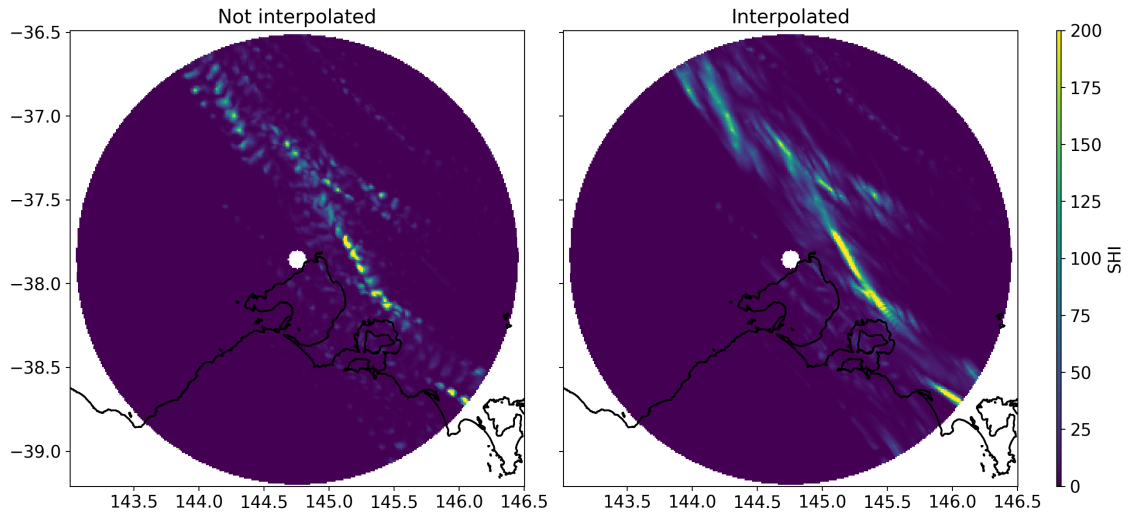
Figure 5. Severe Hail Index (SHI) daily maximum for a sample event. The left panel shows a discontinuous storm track due to the time gap between successive radar scans. The right panel shows the same day with interpolated data based on the optical flow of SHI.

can reside several kilometers above the surface. Depending on the strength of storm-generated and environmental winds, hail descending below this level may be advected several kilometers from ~~is~~ its initial location, leading to the aforementioned mismatch. By design, MESH (and, by extension, SHI) would have greater skill as a predictor of hail damage during events
200   where hail falls with little horizontal movement. To address this issue, we developed a Virtual Advection (VA) algorithm, which matches ground-level observations of hail damage to the appropriate radar observations aloft, mitigating this error. Similar approaches have produced substantial improvement of the correlation between observed damage and radar derived hail estimates (Schiesser, 1990; Hohl et al., 2002; Schmid et al., 1992; Schuster et al., 2006).

VA Algorithm and assumptions

205   To apply the VA algorithm, we made certain assumptions. We assumed that the highest damage within a local area corresponded to the highest observed SHI values aloft, given that differences in vulnerabilities of properties is already mitigated by the archetype normalization. However, this assumption may not hold true in cases where the hailstorm area is not densely covered by insured properties, and large hailstones may fall in uninsured areas or areas without buildings. In such cases, this assumption would systematically lead to a positive bias in SHI values. To address this issue, we filtered the claims data to only include
210   grids with sufficient exposure, as described in Section 2.

To match damage to SHI, we first sorted each events' normalised gridded damage dataset in descending order according to value. Then, we matched the first (highest) damage grid point to the highest SHI grid point within a 4 km Radius-of-Influence (RoI) around the damage grid point. Once we established a pair of damage and SHI grid points, we stored the pair in a new VA dataset with its associated horizontal displacement vector and meteorological variables (from ERA5). We removed

215 this SHI grid point from the available SHI grid and repeated the process until we matched all damage grid points to SHI observations. Once there ~~are~~ were no more damage grids, the average horizontal displacement vector ~~is~~ was calculated. The matching ~~is~~ was restarted but with the average horizontal displacement vector already applied to the 4 km RoI. This double pass approach ~~allows to correct for~~ allowed for correcting for the environmental wind displacement (represented by the average horizontal displacement vector of the first run), and the storm generated winds (done in the second run). The 4 km RoI was

220 selected empirically, with higher values showing only small improvement in the relationship between SHI and damage, and smaller numbers showing poor visual matching of the radar and damage swaths. When both runs ~~are~~ were used, up to 8 km of displacement can be achieved, as both runs could result in displacements in the same direction for a given claim/SHI grid pair. However, this was rare for the studied events, with most events having an average final displacement (after the second run) less than 2 km, and none of the events having average final displacements above 4 km. Note that no local consistency

225 in the displacement vector is not enforced for the second run, which allows convergence and divergence of the SHI field. If all non-zero SHI observations are used before matching all damage grid points, then the closest zero SHI observation is used within the RoI. For grid points with contracts and zero observed damage, we matched the lowest SHI observations (including zero) within the RoI instead of the highest. Grid points with no valid SHI observations within the RoI due to being too far or too close to the radar site were excluded from the VA dataset.

230 We assessed the performance of the VA algorithm using the 2014 Brisbane hailstorm, a well-known event with substantial advection that resulted in a large mismatch between MESH (SHI) observations and reported damage locations (Brook et al., 2021; Warren et al., 2020). The left panel of Figure 6 shows this mismatch. To visualize the algorithm's performance, we calculated the weighted average (by damage) of the horizontal VA displacement vectors (yellow star in the right panel of Figure 6). We then displaced the original MESH grid by this average vector (middle panel of Figure 6), resulting in much better

235 agreement with the observed damage. We refer to this displacement by the average vector as bulk advected, to differentiate it from the data produced by the VA algorithm, which is visualized in the 2D histogram of Figure 6 (right panel). When comparing this bulk advection with the Brook et al. (2022) individual modeling of the hailstone trajectories for this event, good agreement can be observed for the overall swath displacement. In addition, Brook et al. (2022) found that the average motion vector for this event was 2.1 km in a northwesterly cardinal direction, which is very similar to our aforementioned weighted

240 average horizontal displacement (2D histogram in Figure 6).

Performance comparison

The VA algorithm was applied to all hail events, and the resulting MESH–damage dataset is presented in the left panel of Figure 7. The right panel of this figure shows the performance diagram (Roebber, 2009) for various MESH thresholds above which damage is predicted for each advection correction type. Here the 40 mm threshold shows the highest Critical Success

245 Index (CSI) for all correction types; this threshold is shown on the left panel (green dashed line), revealing that most damage occurs for MESH values above this threshold. This diagram illustrates clear improvement for the advection-corrected sets, with the VA algorithm substantially outperforming the bulk advection method. The scatter plot in this panel highlights two events in blue and red, that exhibit distinct MESH–damage relationships. Note that since the contours of individually advected MESH
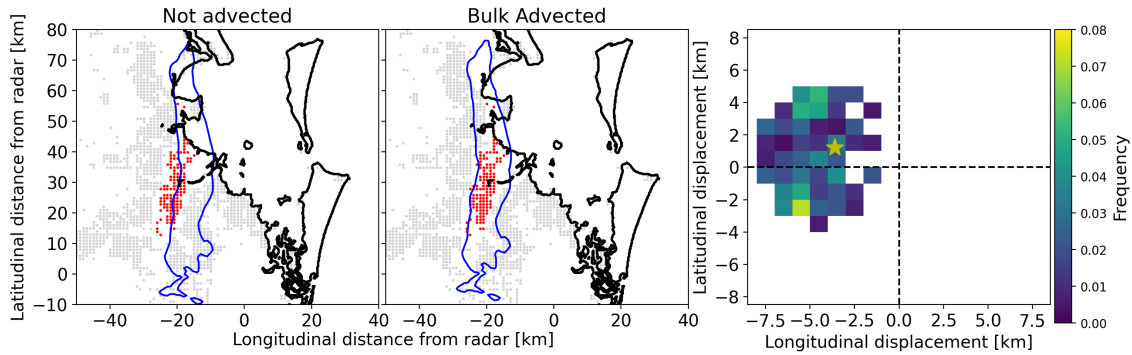
Figure 6. Sample of virtual advection of severe hail event (Brisbane 2014). Left panel shows grid points with exposure that reported no damage in grey, damaged grid points in red, and grid points with insufficient exposure in white. The blue contour line is derived from MESH at 40 mm. The middle panel shows the same but with MESH displaced by the weighted average (by damage) of the virtually advected data. Right panel shows the histogram of the horizontal displacement of MESH to match the damage at ground level, the yellow star shows the weighted average. ~~The dashed range ring is 150 in radius.~~
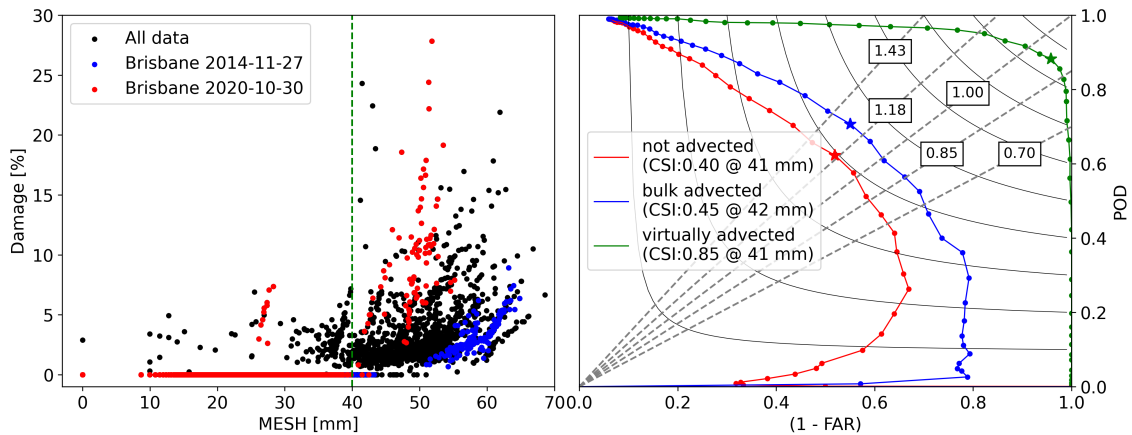
Figure 7. Relationship between MESH and damage. The left panel shows virtually advected MESH against observed damage, with the green dashed line showing the best CSI threshold (for all correction types). The right panel shows the Roebber performance diagram for non-advected MESH, bulk-advected MESH, and individually (virtually) advected MESH. The stars in the right panel show the best CSI for each dataset. The grey dashed lines show the bias with their respective values in boxes.

are not well defined for non-damaging hail and the relationship between the radar hail estimate and observed damage is poor
250 for the bulk-advected dataset, maps of radar derived hail products for all following sections are bulk-advected; while scatter plots comparing these products to damage are individually advected.

5 Hail Damage Estimate (HDE)

In this section, we explore the limitations of using MESH as a predictor of hail damage intensity and present a novel approach for estimating hail damage. While the VA algorithm improves the performance of MESH as a hail damage predictor when above 40 mm, its ability to estimate damage intensity remains inconsistent across different events, as demonstrated in the left panel of Figure 7. This inconsistency suggests that the relationship between MESH and hail damage intensity is dependent on other factors specific to each event, such as the meteorological condition conducive for hail growth, since differences in vulnerabilities due to property factors have already been mitigated (see Archetype normalization subsection). Furthermore, radar reflectivity has inherent limitations since radar reflectivity is the integral of the particle size distribution times the diameter to the 6th power under the Rayleigh approximation, so the same reflectivity could be produced by either a few large hailstones or a high concentration of smaller hailstones, which would result in very different severity of hail damage for the same reflectivity (and therefore SHI). In addition, the presence of other hydrometeors like ice crystals or liquid water can further decouple observed reflectivity from severity of hail damage. To explore the information content of meteorological conditions for improving hail damage estimate, we developed an artificial neural network that incorporates meteorological variables associated with each grid point to improve the SHI–damage relationship. Our approach generates Hail Damage Estimates (HDE) that demonstrate better agreement with the observed damage.

Neural network structure and selection of ~~meteorological data~~input variables

To develop and train the HDE neural network, we utilized TensorFlow and Keras (Abadi et al., 2015).

Initially, a wide range of meteorological variables and radar products were incorporated into the model. We then applied the Shapley Additive Explanations (SHAP, Lundberg and Lee 2017) analysis to identify the most skillful variables, i.e., those with the greatest impact on the prediction (Table 2). This allowed us to maintain the model's accuracy while minimizing its complexity. While initially the network was deeper, it was optimized to maintain accuracy while minimizing computational cost; our final configuration consisted of six layers, each containing 6 (input), 9, 7, 6, 3, and 1 (output) neurons, all of which were densely connected and activated using a rectified linear activation function (excluding the output neuron, which was linear). We linearly normalized the input variables to ensure their values ranged between 0 and 1.

Training and performance

To train our model, we utilized the VA dataset, which contained ~~only~~ the 18 intense ~~hail events (i. e., those with over a thousand claims in a single day).~~ and 12 medium hail events. We had access to ~~46,766~~ 78,486 distinct damage-SHI points that matched our criteria, with ~~18,899~~ 37,145 points indicating no reported damage and zero SHI, ~~26,132~~ 39,558 indicating no damage but SHI > zero, and 1,~~556~~ 775 indicating both reported damage and SHI above zero. Due to the highly unbalanced nature of the data (i.e., no damages greatly outnumber damages), we set the model's initial bias to the natural logarithm of the ratio between the damage count and the no damage count (~~-3.365~~ -3.762). By setting the initial bias to the logarithm of the class ratio, we are effectively providing the model with a starting point that takes into account the class imbalance (He and Garcia, 2009). We

Table 2. Variables used as input for the Hail Damage Estimate (HDE) and their relative importance.

Neural network input variables	Description	Mean relative input importance and standard deviation[%]
SHI	Severe Hail Index	35 ± 5
Q_700hPa	Specific humidity at 700 hPa pressure level	15 ± 3
WS_0C	Wind Speed at 0 C dry bulb level	15 ± 4
V_0M	Meridional wind component at ground level	13 ± 5
CAPE	most unstable Convective Available Potential Energy	12 ± 4
CIN	most unstable Convective Inhibition	10 ± 3

randomly separated the data into two groups: a training dataset (80%) and a validation dataset (20%). To avoid the training and validation dataset from being highly correlated, the separation was done event-wise, meaning that 6 random events were used as validation, and the other 24 as training. The meteorological data associated with non-zero SHI points is well defined, as the time when this SHI occurred is known. This is not the case when a grid point with zero SHI (i.e., outside the storm swath) and any time and therefore associated meteorological data can be assigned. To ensure that the meteorological data used for training was representative of both hail and non-hail atmospheric conditions, we selected random time stamps within ± 1 day of the event for grid points where the SHI was zero. We trained the model ~~650~~ 1000 times for 1000 epochs each, with random separation of the two sets (~~training and validation~~24 training events and 6 validation events) and randomly initialized model weights but with the same initial bias. Figure 8 (left panel) displays a representative training time series, demonstrating that the trained model performs ~~equally~~ well on the training and validation datasets. The vertical dashed line in this panel indicates the epoch when the model reached its optimal training state, before overfitting, and this weight set was then applied for that specific training attempt. To calculate HDE, we utilized an ensemble approach with ~~five members~~ seven members (from the 1000 random training attempts) that maximized CSI and coefficient of determination (R^2). This number of members showed the best balance between computational cost and performance, with additional members showing minimal improvement on the ensemble's skill. The resulting ensemble mean yielded a CSI of 0.88 and an R^2 of ~~0.78~~ 0.79 compared to observations for the full dataset, ~~including the~~ the 18 intense hail events and 12 medium events~~which were completely absent from the training~~. However, the model tends to underestimate large damages ($> 10\%$), as depicted in the right panel of Figure 8, this is likely due to the under-representation of such cases in the dataset. It is important to note that this CSI was achieved at 0.5% damage threshold. From the observed damage data (see figures 7, 8, and 9), it is evident that there are only a few claims between 0% and approximately 0.5%. This is likely because losses below this ratio fall below the policies' deductibles and, therefore, are often not reported by property owners. This apparent discontinuity in the observed damage data was also observed in the unfiltered data (Claims Filtering subsection), indicating that it is not a result of the elimination of uncertain damage areas.
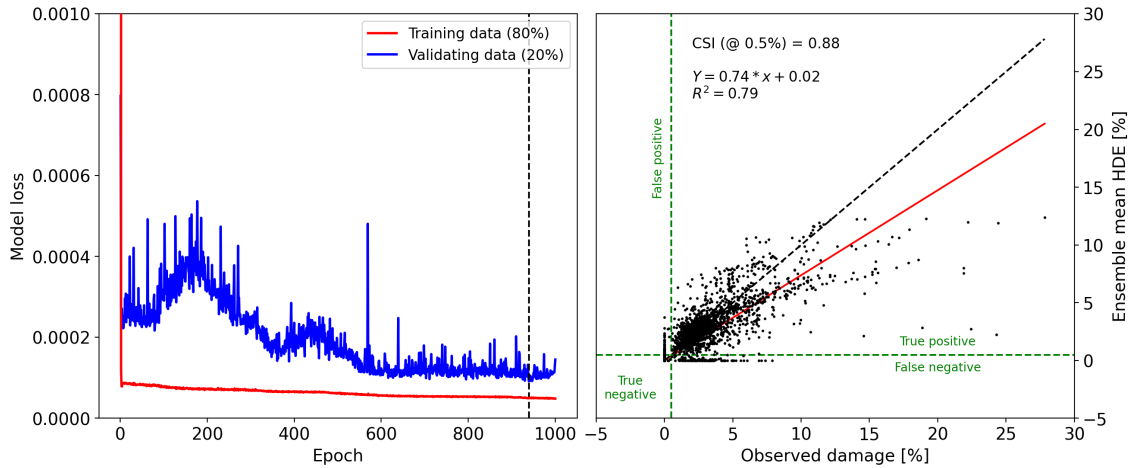
Figure 8. Left panel shows a sample training time series of the model's loss for the training and validation data with the vertical dashed line showing the epoch when the model reached the optimal training state before overfitting. Right panel shows the relationship between observed damage and HDE's ensemble mean. CSI is calculated based on 0.5% damage threshold (green dashed lines). The red line shows the best fit with its equation and coefficient of determination shown on the top left corner. The dashed black line shows the 1:1 relationship.

HDE - MESH relationship

To understand the relationship between HDE and the conventional MESH retrieval, a much larger HDE dataset was required than that available from the VA dataset. Therefore HDE and MESH were calculated for all S-band radars of the Australian operational weather radar archive. In Figure 9, we present the data obtained. A sigmoid function (equation 1) provided the best fit for the data while representing a physically realistic relationship, where damage asymptotically approaches 100% as MESH tends to infinity.

$$HDE = \frac{100}{1 + e^{-A(MESH - B)}} \tag{1}$$

This type of relationship has been used before to relate radar-derived hail estimates to hail damage (Schiesser, 1990). Only points where MESH $>$ 35 mm are shown, as the remaining data points were too numerous to display and damage tends to zero. The fit is relatively good (R^2=0.71), indicating a reasonable correlation between MESH and HDE. Using the inverse of equation 1, we calculated MESH_{HDE}, which allows for an investigation of how the neural network adjusts the radar observations (SHI) using the environmental information. Note that MESH_{HDE} is not intended to replace MESH as a reliable size of hail, but rather as a tool to identify where the model has decreased or increase hail intensity according to the environmental conditions.
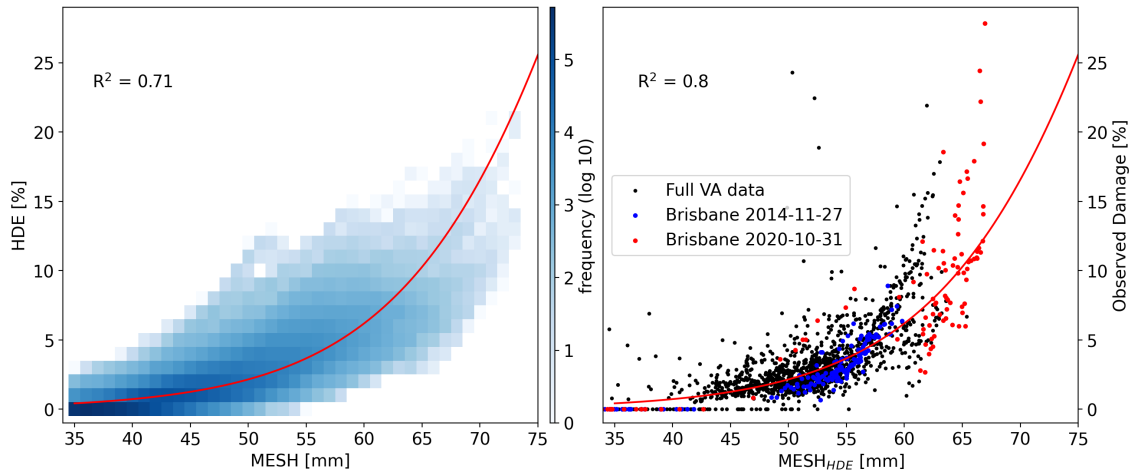
Figure 9. Left panel shows HDE and MESH calculated for the full radar archive as a 2D histogram with 1% and 1mm bins with the best fit of equation 1. Right panel shows the relationship between observed damage and $MESH_{HDE}$, with the previously highlighted events shown in red and blue.

320 We then examined the relationship between $MESH_{HDE}$ and observed claims for the events captured by the damage dataset, as shown in the right panel of Figure 9, with the two Brisbane events previously highlighted in the same colors. The relationship is greatly improved, and both events now show more consistent sizes relative to the observed damage.

HDE case studies

In this section we show the performance and behavior of HDE for the two previously highlighted hail events to analyse how 325 different meteorological conditions drive HDE behaviour.

Brisbane 2014 hail event

On November 27th, 2014, Brisbane experienced a severe hailstorm that caused substantial damage to a densely populated area, resulting in over AUD 1.5 billion losses (normalized to 2017) and more than nine thousand individual building claims (for Suncorp). Giant hailstones, reportedly measuring around 70 mm in size, were observed during this event (Parackal et al., 330 2015). In Figure 10, we present MESH and $MESH_{HDE}$ for this event. We observed that $MESH_{HDE}$ assigns low values of MESH to zero, due to the inability of low MESH values ~~inability~~ to cause any damage to property and therefore resulting in zero HDE. Additionally, within the eastern-most storm cell, where MESH is high and damage occurred, $MESH_{HDE}$ was mostly lower than MESH, as highlighted in the bottom-right panel. This finding shows that $MESH_{HDE}$ more closely aligns with the mean fit (green dashed line in the bottom-right panel) and corrects the observed positive bias of this event when compared 335 with the original VA data. While $MESH_{HDE}$ shows a more consistent fit, it is evident that it is not a good representation of the
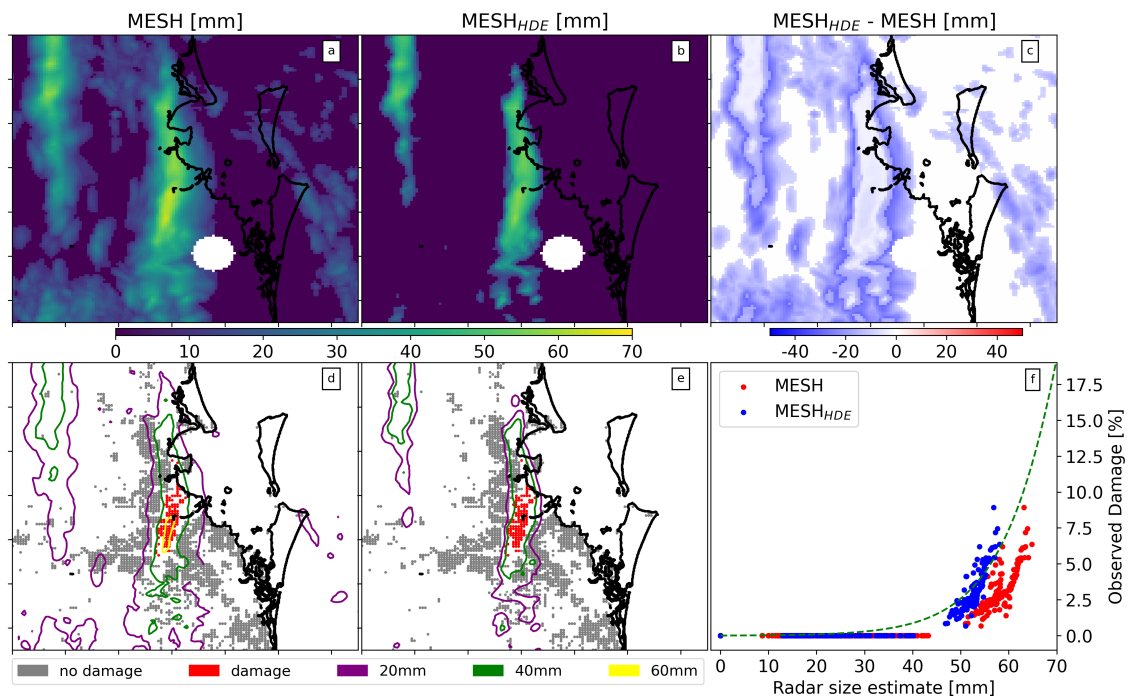
Figure 10. Comparison of MESH, $\mathrm{MESH_{HDE}}$, and observed damages for an intense hail event in Brisbane on 2014-11-27. ~~Top row shows (from left to right)~~ Panels A, B, and C show maps of MESH, $\mathrm{MESH_{HDE}}$ (MESH derived from HDE), and the difference between the two, respectively. ~~Bottom left~~ Panel D and ~~middle~~ E show observed damage with bulk-advected MESH and $\mathrm{MESH_{HDE}}$ contours overlaid. ~~Bottom right panel~~ Panel F shows the relationship between observed damage, and the grid-point advected MESH and $\mathrm{MESH_{HDE}}$.

actual size of expected hail as it is still far below the observed size for this storm. The bottom left and middle panels show the MESH and $\mathrm{MESH_{HDE}}$ contours over the observed damage respectively.

Brisbane 2020 hail event

In contrast to the 2014 event, the 2020 Brisbane hail event showed a negative bias in MESH relative to the full VA dataset. As

340   shown in Figure 11, $\mathrm{MESH_{HDE}}$ was lower than MESH for most of the domain, but considerably higher than MESH where hail damage occurred. It is worth noting that this case exhibited some potentially spurious claims, as highlighted within the red circles in the ~~bottom-right panel . These damage grid points were relatively far from the main storm swaths (indicated with red arrows in the lower middle subplots) and~~ panel F. All these point were tracked and found in a cluster in the map (indicated by the red circle in the panel E) close to the coast and were relatively far from the main storm swaths and could be due to the

345   misclassification of damage cause (i.e., wind or flood damage instead of hail) in the insurance dataset.
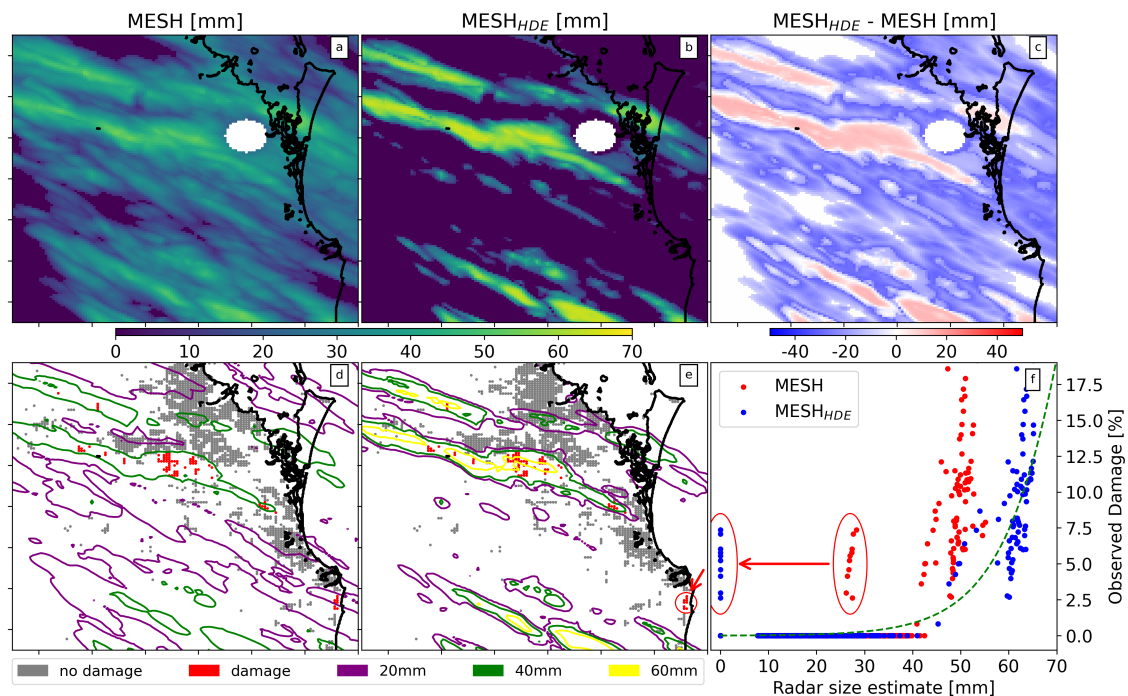
Figure 11. Same as Figure 10 but for an intense hail event in Brisbane during 2020-10-31. The red circles in panel F show potential spurious claims, these were tracked to the map and shown in panel E within a red circle.

6 Hail damage, MESH, and meteorology

The 2020 event was associated with very different meteorological conditions in the ERA5 reanalysis than the 2014 event, with higher CAPE and lower CIN, stronger northerly winds at ground level, lower humidity aloft, and increased winds at the melting layer (Table 3). Looking at the MESH values from these events alone, one would expect the 2014 event to have caused

350 more damage than the 2020 event, as the mean MESH values in the storm cores (here defined as the areas with HDE above zero) were 58.3 mm and 51.6 mm, respectively. Nevertheless, in the observed damage dataset, the 2014 event produced lower damage than the 2020 event, as clearly visible in Figure 7. This discrepancy is not observed in $\mathrm{MESH}_{\mathrm{HDE}}$, with mean values of 51.3 mm and 61.4 mm, for the 2014 and 2020 events respectively.

The cases mentioned highlight the possibility of storms with relatively low MESH values causing severe damage, such as

355 the one that hit Brisbane in 2020, or relatively high MESH values leading to comparatively lower damage, as seen in the 2014 Brisbane storm. The disparity could be attributed to a mixture of hail and other hydrometeor types (liquid water, ice crystals) which would result in a higher MESH value than a volume containing only the hail component. Further, MESH uses the environmental freezing level, but the updraft's freezing level could be well above the environmental freezing level which could involve larger liquid water droplets than conventional supercooled liquid water droplets (Kumjian and Ryzhkov, 2008).

360 The disparity could also be attributed to differences in the HSD in these volumes. Volumes with different HSD but equivalent

Table 3. Mean values of MESH, MESH$_{HDE}$, and environmental parameters for storm cores during the 2014 and 2020 Brisbane events. The error bounds show the standard deviation within the storm cores.

Variables name	2014 event mean	2020 event mean
MESH [mm]	58.3 ± 3.1	51.6 ± 1.3
MESH$_{HDE}$ [mm]	~~51.3~~ $51.9 \pm$ ~~4.5~~ 4.2	~~61.4~~ $59.4 \pm$ ~~5.1~~ 7.3
Q_700hPa [kg $\cdot$ kg^{-1}]	0.0056 ± 0.0001	0.0031 ± 0.0002
CAPE [J $\cdot$ kg^{-1}]	526.9 ± 0.5	3137.8 ± 21.8
CIN [J $\cdot$ kg^{-1}]	282.3 ± 49.3	88.7 ± 28.7
WS_0C [m $\cdot$ s^{-1}]	7.20 ± 0.07	27.16 ± 0.98
V_0M [m $\cdot$ s^{-1}]	1.72 ± 0.48	-10.59 ± 0.16

scattering response would produce similar MESH values but varying ground damage. Since supercooled liquid water (SLW) droplets are considerably small, most of the reflectivity signal is expected to be caused by hail, smaller frozen hydrometeors or large liquid water droplets lofted above the environmental melting layer.

Extending this method, we can utilize the difference between MESH$_{HDE}$ and MESH (hereafter ΔMESH) as an indicator of the skewness of the HSD within a storm, differences in hail hardness, and/or mix hydrometeor volumes. Positive ΔMESH indicates a larger proportion of damaging hailstones, harder/denser hailstones, and/or less proportion of mixed hydrometeors in the volume, while negative ΔMESH indicates the contrary.

Figure 12 shows the relationship between each meteorological parameter and HDE and ΔMESH as box-plots for the full radar archive. Using 40 mm for MESH and 0.5% for HDE as thresholds for damage, and HDE as the truth, the leftmost column (panels A, E, I, M, and Q) represents environments where MESH indicates a false-positive for hail damage while the second column (panels B, F, J, N, and R) shows opposite environments, where MESH indicates a false-negative. The Kolmogorov-Smirnov (Hodges Jr, 1958) test was applied to each meteorological variables for the false-positive and false-negative cases, showing that while all sets showed significant differences (p-values approaching zero), specific humidity aloft showed a substantial difference between the two cases, with a D-statistics of 0.64, which is evident from the corresponding box-plots. A clear threshold was identified based on the analysis (dashed horizontal line in panels A, B, C, and D). It was found that 82% of the false-positive occurrences happened in environments where the specific humidity at 700 hPa was above 0.0053 kg $\cdot$ kg^{-1}. Coincidentally, the same percentage of false-negative occurrences happened in environments where the specific humidity at 700 hPa was below that humidity threshold. By utilizing this threshold, ~~by~~ and applying it only for MESH values between 39 and 41 mm, which account for approximately 50% of the false-positive and false-negative cases, the CSI of MESH can be improved from 0.76 (at 40 mm against HDE) to 0.79. Environments with false-positive results exhibited conditions similar to those in negative ΔMESH bins and low HDE bins. This indicates that most of these false positives resemble environments associated with negative values of ΔMESH and are less likely to have high HDE. False negative cases were associated with low HDE and showed similarities to environments with a positive ΔMESH bias, and these cases had

an average HDE of 0.8% and a maximum of 4.2%. CAPE tends to increase with ΔMESH, whereas CIN is larger for more extreme ΔMESH of either sign. Looking at the relationship to winds, there is a positive correlation with ΔMESH, except for the extreme positive end. In a similar way these extreme values of ΔMESH are associated with northward winds at the ground compared to the rest of the distribution that exhibits no trend.

This analysis indicates that environments with relatively low specific humidity, high CAPE and CIN, low wind speeds aloft and northward winds at ground are likely to have HSD with a higher proportion of large hailstones or a lower proportion of other hydrometeors. Regarding environments associated with extreme values of hail damage, these are associated with high CAPE and CIN values, lower wind speeds at the melting layer, and northward winds at ground. No clear signal with respect to specific humidity aloft is observed. A recent modeling study on hail production (Lin and Kumjian, 2022) found that CAPE acts as a modulator to hail growth, with a non-monotonic relationship with hail size which peaks around 2000–2400 $J \cdot kg^{-1}$. Although we do observe that about 50% of the samples with high values of HDE occur close to this CAPE range (Figure 12, panel G), these HDE bins also show about 25% of samples with CAPE values above 3000 $J \cdot kg^{-1}$ and a clear positive correlation between CAPE and HDE. Regarding the aforementioned HDE relationship with winds, a similar modeling study (Dennis and Kumjian, 2017a) found that increased deep-layer east-west shear increases hailstone mass while increased low-level north-south shear reduces hailstone mass. Here it is important to note that one of the variables included in the initial input layer of the HDE model was absolute wind speed shear between the surface and the melting layer but was discarded as the SHAP analysis showed that it had low influence on the output relative to the other input variables. The zonal or meridional components of the shear vector were not tested.

7 Conclusions

This study has analyzed more than 10 years of insurance and radar data from Australia to investigate the performance of the Maximum Expected Size of Hail (MESH) retrieval in predicting and quantifying hail damage. The results showed that MESH has poor predictive value for damage magnitude estimation, but shows good skill (CSI = 0.88 for MESH larger than 40 mm) as a binary predictor of hail damage when corrected for horizontal advection of hailstones, which is consistent with Brook et al. (2021).

To improve damage magnitude estimation, a neural network was trained with meteorological variables from ERA5 in addition to SHI (from which MESH is derived) against observed hail damage. Using SHAP analysis, the most important meteorological variables were identified as specific humidity at 700 hPa, wind speed at the freezing level, northward winds at ground level, CAPE and CIN. This neural network produced a Hail Damage Estimate (HDE) with a high accuracy (CSI = 0.88 and an R^2 = 0.78) for estimating hail damage occurrence and intensity.

A comparison of HDE with MESH for the full national radar dataset (14 radars with an average of 18.8 years of coverage) revealed a relatively good (R^2 = 0.71) fit between the two using a sigmoid curve. This curve was used to derive $MESH_{HDE}$ from HDE with the goal of identifying environments where MESH shows negative or positive biases, potentially due to differences in hail size distributions (HSD) and/or presence of other hydrometeors along with hail.
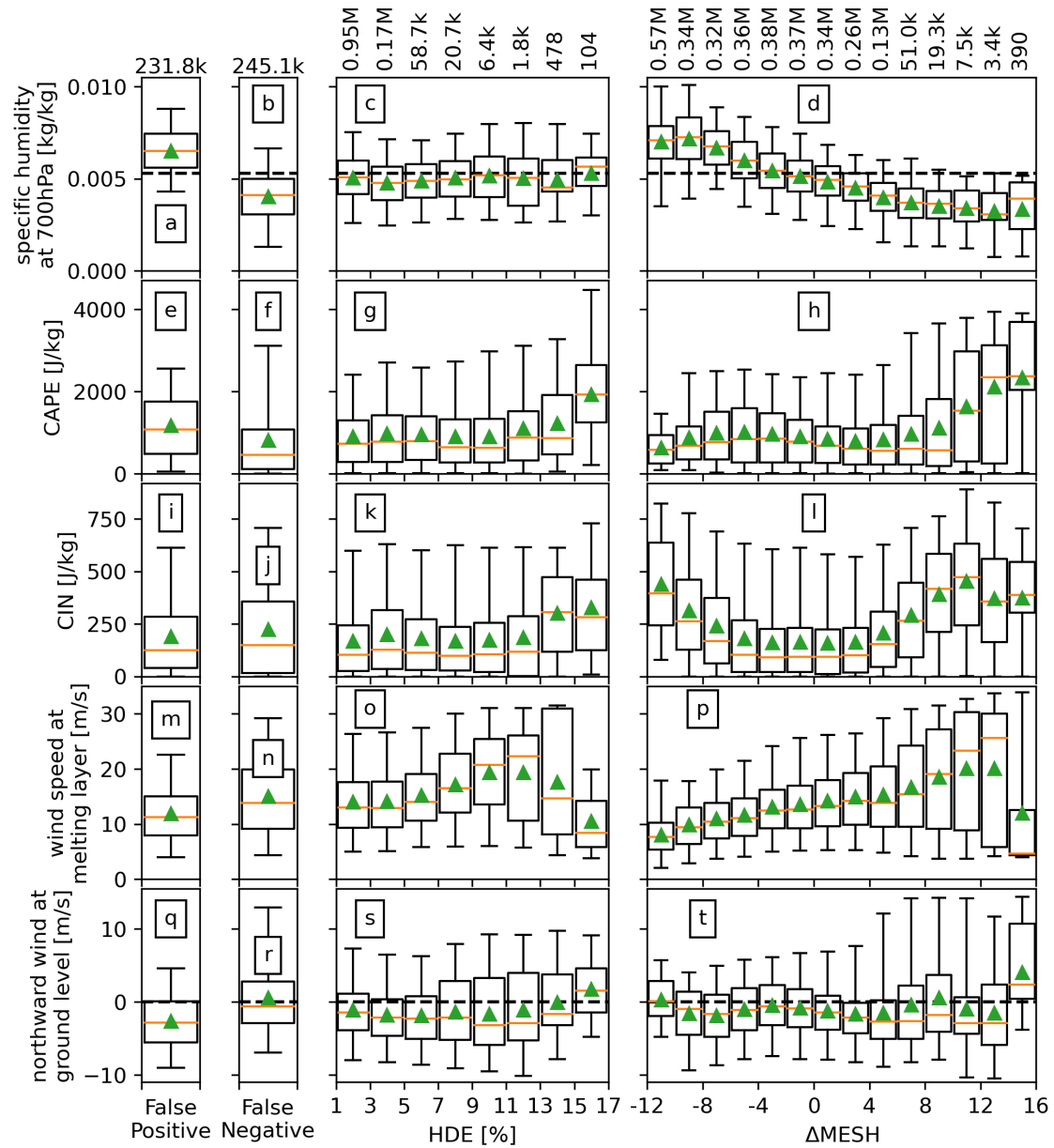
Figure 12. Box-plots of the meteorological variables that drive the HDE model. From left to right columns: (1st) for HDE < 0.5% and MESH ≥ 40mm; (2nd) for HDE ≥ 0.5% and MESH < 40 mm; (3rd) for HDE ≥ 0.5% in 2% bins; (4th) for HDE ≥ 0.5% in 2 mm bins. The 3rd and 4th columns show the distributiion as a function of HDE and ΔMESH, respectively. Sample size per bins are shown on top. The box-plots follow the same style/structure as figure 1, top panel.

Environments with negative bias in radar hail estimates are associated with low CAPE, high CIN, and higher specific humidity aloft and are likely to be non-damaging if MESH is below 50 mm. In contrast, environments with positive bias are associated with high CAPE, high CIN, and lower specific humidity aloft and are likely damaging if MESH is above 30 mm. Extreme hail damage was associated with such positive bias environments that in addition showed low wind speeds aloft and northerly winds at ground.

The study provides important insights into the performance of MESH/SHI for estimating property damage and the potential of using neural networks to improve hail damage estimation and identifying patterns between environmental conditions and a storm's HSD, hail hardness, and/or presence of ~~mix~~ mixed hydrometeor precipitation. It is important to note that these results were developed for Australian storm environments, and might not be representative of global storm environments. This study was limited to S-band radars, and future work should expand this technique to C-band radars. In addition, future work will explore using this novel hail damage estimate for nowcasting applications to provide hail damage warning. Another limitation to our findings is the relatively small sample of individual storms, which might only sample a subset of all environmental conditions that leads to hail storms. Replicating this work on locations with high population density and radar coverage (i.e., Europe or the USA) would be valuable to potentially mitigate this limitation as well as ~~showing~~ determine if these same environmental parameters play dominant roles on HDE.

Data availability. The insurance data provided by Suncorp is not publicly available for privacy reasons. Details regarding the radar data used in this study can be found in found in the AURA database online at https://www.openradar.io (Soderholm et al., 2022). ERA5 data can be accessed at https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5 (Hersbach et al., 2020).

Author contributions. RW and LY prepared the insurance claims data for analysis and contributed on the interpretation and implementation of the archetype normalization. NR validated and tested the HDE product againts internal Suncorp models. JS prepared the radar data and SHI/MESH hail estimate products. JS and AP contributed in the development of section 4, 5, and the interpretation of section 6. LA designed the paper's outline, carried most of the analysis and wrote the paper. All co-authors participated in paper-related discussions and commented on the paper.

Competing interests. The authors have no other competing interest to declare

Acknowledgements. This research was funded by Suncorp and the Australian Bureau of Meteorology.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Good-fellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X.: TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, software available from tensorflow.org, 2015.

Allen, J. T. and Tippett, M. K.: The Characteristics of United States Hail Reports: 1955-2014, E-Journal of Severe Storms Meteorology, 10, 1–31, https://doi.org/10.55599/EJSSM.V10I3.60, 2015.

Blong, R.: Residential building damage and natural perils: Australian examples and issues, https://doi.org/10.1080/0961321042000221007, 32, 379–390, https://doi.org/10.1080/0961321042000221007, 2007.

Brook, J. P., Protat, A., Soderholm, J., Carlin, J. T., McGowan, H., and Warren, R. A.: HailTrack—Improving Radar-Based Hailfall Estimates by Modeling Hail Trajectories, Journal of Applied Meteorology and Climatology, 60, 237–254, https://doi.org/10.1175/JAMC-D-20-0087.1, 2021.

Brook, J. P., Protat, A., Soderholm, J. S., Warren, R. A., and McGowan, H.: A Variational Interpolation Method for Gridding Weather Radar Data, Journal of Atmospheric and Oceanic Technology, 39, 1633–1654, https://doi.org/10.1175/JTECH-D-22-0015.1, 2022.

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.

Brown-Giammanco, T. M., Giammanco, I. M., and Estes, H. E.: New Asphalt Shingle Hail Impact Performance Test Protocol and Damage Assessment, Natural Hazards Review, 22, 04021 050, https://doi.org/10.1061/(ASCE)NH.1527-6996.0000509, 2021.

Cintineo, J. L., Smith, T. M., Lakshmanan, V., Brooks, H. E., and Ortega, K. L.: An Objective High-Resolution Hail Climatology of the Contiguous United States, Weather and Forecasting, 27, 1235–1248, https://doi.org/10.1175/WAF-D-11-00151.1, 2012.

Dahl, N. A., Shapiro, A., Potvin, C. K., Theisen, A., Gebauer, J. G., Schenkman, A. D., and Xue, M.: High-resolution, rapid-scan dual-Doppler retrievals of vertical velocity in a simulated supercell, Journal of Atmospheric and Oceanic Technology, 36, 1477–1500, 2019.

Dennis, E. and Kumjian, M.: The impact of vertical wind shear on hail growth in simulated supercells, Journals of the Atmospheric Sciences, 74, 641–663, https://doi.org/10.1175/JAS-D-16-0066.1, 2017a.

Dennis, E. J. and Kumjian, M. R.: The Impact of Vertical Wind Shear on Hail Growth in Simulated Supercells, Journal of the Atmospheric Sciences, 74, 641–663, https://doi.org/10.1175/JAS-D-16-0066.1, 2017b.

Depue, T. K., Kennedy, P. C., and Rutledge, S. A.: Performance of the Hail Differential Reflectivity (HDR) Polarimetric Radar Hail Indicator, Journal of Applied Meteorology and Climatology, 46, 1290–1301, https://doi.org/10.1175/JAM2529.1, 2007.

Giammanco, I. M., Brown, T. M., Grant, R. G., Dewey, D. L., Hodel, J. D., and Stumpf, R. A.: Evaluating the hardness characteristics of hail through compressive strength measurements, Journal of Atmospheric and Oceanic Technology, 32, 2100–2113, 2015.

Groenemeijer, P., Pucik, T., Tsonevsky, I., and Bechtold, P.: An overview of convective available potential energy and convective inhibition provided by NWP models for operational forecasting, European Centre for Medium-Range Weather Forecasts, 2019.

Gunturi, P. and Tippett, M.: Managing severe thunderstorm risk: Impact of ENSO on US tornado and hail frequencies, Tech. rep., Willis Re Inc, 2017.

He, H. and Garcia, E. A.: Learning from imbalanced data, IEEE Transactions on knowledge and data engineering, 21, 1263–1284, 2009.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G. D., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J. N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/QJ.3803, 2020.

Hodges Jr, J.: The significance probability of the Smirnov two-sample test, Arkiv för matematik, 3, 469–486, 1958.

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/10.1016/S0169-8095(02)00059-5, 2002.

Kumjian, M. R. and Ryzhkov, A. V.: Polarimetric Signatures in Supercell Thunderstorms, Journal of Applied Meteorology and Climatology, 47, 1940–1961, https://doi.org/10.1175/2007JAMC1874.1, 2008.

Lin, Y. and Kumjian, M. R.: Influences of CAPE on Hail Production in Simulated Supercell Storms, Journal of the Atmospheric Sciences, 79, 179 – 204, https://doi.org/https://doi.org/10.1175/JAS-D-21-0054.1, 2022.

Louf, V., Protat, A., Warren, R. A., Collis, S. M., Wolff, D. B., Raunyiar, S., Jakob, C., and Petersen, W. A.: An Integrated Approach to Weather Radar Calibration and Monitoring Using Ground Clutter and Satellite Comparisons, Journal of Atmospheric and Oceanic Technology, 36, 17 – 39, https://doi.org/https://doi.org/10.1175/JTECH-D-18-0007.1, 2019.

Lundberg, S. M. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, in: Advances in Neural Information Processing Systems 30, edited by Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., pp. 4765–4774, Curran Associates, Inc., 2017.

Mobasher, M. E., Adams, B., Mould, J., Hapij, A., and Londono, J. G.: Damage mechanics based analysis of hail impact on metal roofs, Engineering Fracture Mechanics, 272, 108 688, https://doi.org/10.1016/J.ENGFRACMECH.2022.108688, 2022.

Murillo, E. M. and Homeyer, C. R.: Severe Hail Fall and Hailstorm Detection Using Remote Sensing Observations, Journal of Applied Meteorology and Climatology, 58, 947–970, https://doi.org/10.1175/JAMC-D-18-0247.1, 2019.

Nanni, S., Mezzasalma, P., and Alberoni, P. P.: Detection of hail by means of polarimetric radar data and hailpads: results from four storms, Meteorological Applications, 7, 121–128, https://doi.org/10.1017/S135048270000147X, 2000.

Nelder, J. A. and Mead, R.: A Simplex Method for Function Minimization, The Computer Journal, 7, 308–313, https://doi.org/10.1093/comjnl/7.4.308, 1965.

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, https://doi.org/10.1175/JAMC-D-15-0203.1, 2016.

Parackal, K. I., Mason, M. S., Henderson, D. J., Smith, D. J., and Ginger, J. D.: Investigation of damage: Brisbane 27 November 2014 severe storm event, 2015 bushfire and natural hazards CRC and AFAC conference, 2015.

Prein, A. F. and Holland, G. J.: Global estimates of damaging hail hazard, Weather and Climate Extremes, 22, 10–23, https://doi.org/10.1016/J.WACE.2018.10.004, 2018.

Richter, H. and Deslandes, R.: The four large hail assessment techniques in severe thunderstorm warning operationsin Australia, 33rd Conference on Radar Meteorology, Cairns, QLD, Australia, 2007.

Roebber, P. J.: Visualizing multiple measures of forecast quality, Weather and Forecasting, 24, 601–608, 2009.

515   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, https://doi.org/https://doi.org/10.1175/JAMC-D-13-073.1, 2013.

Schiesser, H.: Hailfall: the relationship between radar measurements and crop damage, Atmospheric Research, 25, 559–582, 1990.

Schmid, W., Schiesser, H., and Waldvogel, A.: The kinetic energy of hailfalls. Part IV: Patterns of hailpad and radar data, Journal of Applied
520   Meteorology and Climatology, 31, 1165–1178, 1992.

Schuster, S. S., Blong, R. J., and McAneney, K. J.: Relationship between radar-derived hail kinetic energy and damage to insured buildings for severe hailstorms in Eastern Australia, Atmospheric research, 81, 215–235, 2006.

Soderholm, J., Louf, V., Brook, J., Protat, A., and Warren, R.: Australian Operational Weather Radar Level 2 Dataset, National Computing Infrastructure, https://doi.org/10.25914/JJWZ-0F13, 2022.

525   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/10.1002/QJ.3693, 2020.

Witt, A., Eilts, M. D., Stumpf, G. J., Johnson, J. T., Mitchell, E. D., and Thomas, K. W.: An Enhanced Hail Detection Algorithm for the WSR-88D, Weather and Forecasting, 13, 286–303, 1998.