Investigating Hailstorm Updrafts and Nowcasting Hail Size Using a Novel Radar-Based Updraft Detection

VINZENT KLAUS[a] AND JOHN KRAUSE[b,c]

[a] *Institute of Meteorology and Climatology, BOKU University, Vienna, Austria*
[b] *Cooperative Institute for Severe and High-Impact Weather Research and Operations, The University of Oklahoma, Norman, Oklahoma*
[c] *NOAA/OAR National Severe Storms Laboratory, Norman, Oklahoma*

(Manuscript received 20 December 2023, in final form 28 July 2024, accepted 17 September 2024)

ABSTRACT: Nowcasting hail size poses a major challenge in operational practice due to physical limitations of weather radar technology once hailstones are sufficiently large to enter the resonance scattering regime. Numerous radar-based hail size proxies have been derived in recent decades, but their performance is generally poor in identifying giant hail (≥10 cm). Using a novel thunderstorm updraft detection method, we examine the updraft characteristics of hailstorms in the U.S. Great Plains based on a NEXRAD dataset of 114 hail events between 2013 and 2023. We find that some radar-derived variables within the detected updraft are well suited for discriminating between small hail (1.0–3.0 cm) and severe hail (≥3.5 cm), e.g., minimum copolar cross-correlation coefficient in the midlevel updraft, whereas other radar metrics such as the area of reflectivity > 50 dBZ in the upper portion of the updraft suggest the presence of giant hail. However, the statistical distributions of each variable overlap for different hail sizes, and there is no single metric which performs well across the entire hail size spectrum. Therefore, we trained a random forest model to nowcast hail size categories using a multitude of these radar metrics. The model shows promising performance for discriminating hail sizes > 5 cm but requires further refinement for smaller hail. We showcase the model's capabilities for a set of hailstorms in the Great Plains.

SIGNIFICANCE STATEMENT: Large hail, which may reach sizes up to 20 cm in extreme cases, is a severe threat in thunderstorms and responsible for significant damage in thunderstorm-prone areas around the world. Weather radars are generally used for issuing warnings for the threat of large hail, but it is surprisingly difficult to apply their data to distinguish between hail sizes. We utilize a new approach that detects the thunderstorm updraft region, which is where hailstones grow, and analyze the radar data therein. Based on these data, we then develop a model for assessing hail sizes that could be used for warnings and illustrate the model results for six hailstorms.

KEYWORDS: Hail; Supercells; Thunderstorms; Radars/Radar observations; Nowcasting; Machine learning

1. Introduction

Large hail is one of the most damaging threats associated with deep moist convection, causing annual losses to property, agriculture, and livestock on the order of 10 billion USD in the United States alone (Gunturi and Tippett 2017; Ward et al. 2020). As the impact kinetic energy and damage potential of hailstones strongly correlate with size (Heymsfield et al. 2014; Heymsfield et al. 2018), an accurate assessment of maximum expected hailstone size is critical in operational nowcasting. However, high-end hail events are particularly challenging to forecast and nowcast, and large hail sizes are regularly underestimated in operational warnings (Blair et al. 2017).

Weather radars facilitate the nowcasting of hailstorms and the timely issuing of warnings. Their data are used to identify the occurrence of hail in thunderstorms, and numerous empirical relationships between radar (and radar-derived) variables and hail size have been described in the literature. One of the most widely used hail size proxies, the maximum expected size of hail (MESH) introduced by Witt et al. (1998), is derived from the vertical integration of radar reflectivity factor Z_H. Based on a comprehensive observation dataset from the United States, Ortega (2018) showed that MESH was suited for finding locations of severe hailfall but did not align well with specific hail sizes observed. Using a significantly higher number of hail reports than Witt et al. (1998), Murillo and Homeyer (2019) revised the empirical fit of MESH, but the underlying large variation of observed hail sizes for any given MESH value remained. Other recent attempts include polarimetric radar variables, generally differential reflectivity Z_{DR} and copolar cross-correlation coefficient ρ_{HV} (Heinselman and Ryzhkov 2006; Ryzhkov et al. 2013b; Ortega et al. 2016). However, all these algorithms show limited skill differentiating between hailstone sizes once the hail is large enough to enter the Mie backscattering regime, which for the S band applies to hailstones larger than approximately 5-cm diameter and to even smaller hailstones in the C and X bands (Ryzhkov et al. 2013a). Hail size discrimination is further complicated by the wide variety of hailstone shapes and textures as well as their complex falling behavior. Differing properties of dry and water-coated hailstones result in a wide range of Z_H, Z_{DR}, and ρ_{HV} associated with hail (Jiang et al. 2019). Several previous studies have leveraged indirect, storm intensity-based radar features to assess maximum hail size, for example, storm top divergence (Witt and Nelson 1991), scale and intensity of the bounded weak echo region (Gutierrez and Kumjian 2021), or midlevel rotation (Blair et al. 2011;

Corresponding author: Vinzent Klaus, vinzent.klaus@boku.ac.at

DOI: 10.1175/WAF-D-23-0227.1

TABLE 1. Convention for hail sizing used throughout this study and number of filtered hail cells available in each category.

Class	General naming convention of hail sizes (cm)	Definition of hail sizes for our cell classification (cm)	Number of cells available after filtering
Small	≥1.0	1.0–3.0	138
Severe	≥3.5	3.5–5.5	130
Significantly severe	≥6.0	6.0–9.0	84
Giant	≥10.0	≥10.0	32

Witt et al. 2022). Operational use of these methods is limited by sampling strategies that prioritize low-level scans (Cho et al. 2022; Kingfield and French 2022) and derived empirical functions based on relatively small datasets, tailored to specific geographic regions (Allen et al. 2019).

Here, we try to reconcile current direct radar estimates of hail size with indirect storm-intensity-based proxies. While we fully acknowledge that giant hail size assessment might be close to impossible based on gate-by-gate fuzzy logic methods or by a single kinematic proxy such as storm top divergence, we hypothesize that hail sizes in the Mie scattering regime can be inferred through the combination of multiple radar variables characterizing the dynamical and microphysical updraft properties. We present and evaluate a new object-based approach that specifically targets the updraft area, since this is the part of the storm where hydrometeors grow and hailstones can be suspended. Our updraft detection is based on the radar signature of Z_{DR} columns (Kumjian and Ryzhkov 2008; Snyder et al. 2017; Kuster et al. 2020), which we use to extract data that potentially relate to the presence of hail or the storm's potential to grow large hail. In addition to established hail markers, we explore a suite of new radar metrics aimed at differentiating between hail sizes. We evaluate them on a dataset of 384 hailstorms in the United States compiled by linking radar data with hail reports from the NOAA National Centers for Environmental Information (NCEI) storm event database. By using rigid quality control algorithms, we try to overcome the challenges caused by the scarcity, unreliability, and lack of "true negatives" (reports of no hail) of the reports in the storm events database. In contrast to Homeyer et al. (2023) where the three-dimensional polarimetric structure of hailstorms was analyzed based on a composite mean technique, we limit our evaluation to hailstorm updraft characteristics that could be exploited in nowcasting hail sizes.

Based on the hailstorm dataset, we then train a model using machine learning (ML), in our case a random forest (RF; Breiman 2001), to predict multiple hail size categories up to giant hail with more than 10-cm diameter (see Table 1). Since we aim to identify the hail growth process, we expect that short lead times on the order of 10–30 min before hail reaches the surfaces might be possible to achieve, allowing the method to be used in nowcasting (compare Picca and Ryzhkov 2010).

In section 2, we specify the input data and outline our methods for analyzing potential hail size-dependent radar signatures. Furthermore, we describe the steps to train and verify an RF hail size nowcasting model. In section 3, we present the analysis of selected hail (precursor) signals. In section 4, we evaluate the performance of the RF hail nowcasting model in comparison with MESH and show nowcasting examples for six hailstorms. In section 5, we summarize our findings and provide an outlook for future developments.

2. Data and methods

The data processing for evaluating different radar metrics and identifying their significance in hail nowcasting can be divided into six major steps which we describe in detail as follows: 1) collection and preprocessing of input data, 2) updraft detection and tracking, 3) linking of hail reports to storm cells, 4) retrieval of potential hail indicators and precursors, 5) RF model training and tuning, and 6) RF model evaluation.

a. Collection and preprocessing of input data

We collected data for 114 hail events in a study area roughly corresponding to the Great Plains (28°–48°N and 105°–90°E), the area most prone to severe hail within the United States (Cintineo et al. 2012; Allen et al. 2017). All radar data used in this study are S-band data from the WSR-88D NEXRAD network after its upgrade to dual-polarization capability (NOAA NWS Radar Operations Center 1991), resulting in a study period of approximately 10 years from 2013 to 2023. Each event can contain multiple individual hailstorms and is uniquely characterized by the identification (ID) of the nearest NEXRAD radar (whose data are used) and the start day in UTC.

While events from 2013 to 2015 were manually selected based on the availability of hail report data from the Severe Hazards Analysis and Verification Experiment (SHAVE) project (Ortega et al. 2009), all events after 2015 were automatically selected by an algorithm scanning through 54 624 hail reports found in the NOAA National Weather Service (NWS) storm events database (NOAA NWS 2023). This algorithm was tuned to select periods with abundant large hail reports in the vicinity of NEXRAD radar sites. We considered only events with at least 20 hail reports ≥ 1 cm diameter or 10 reports ≥ 5 cm or two reports ≥ 10 cm that occurred within 5 h and at a range between 20 and 90 km to the nearest NEXRAD site in the study area. The distance criterion was designed to ensure sufficient midlevel radar coverage, since the strongest hail growth is believed to occur between isotherm heights of −10° and −30°C (Nelson 1983; Allen et al. 2019). The procedure resulted in a total of 114 events distributed over 102 different hail days.

Figure 1 shows the study area, the radars, and all hail reports used for our study. The complete event list is shown in Table A1 in appendix A.

For each event, publicly available NEXRAD level II data retrieved from Amazon Web Services (AWS) were interpolated

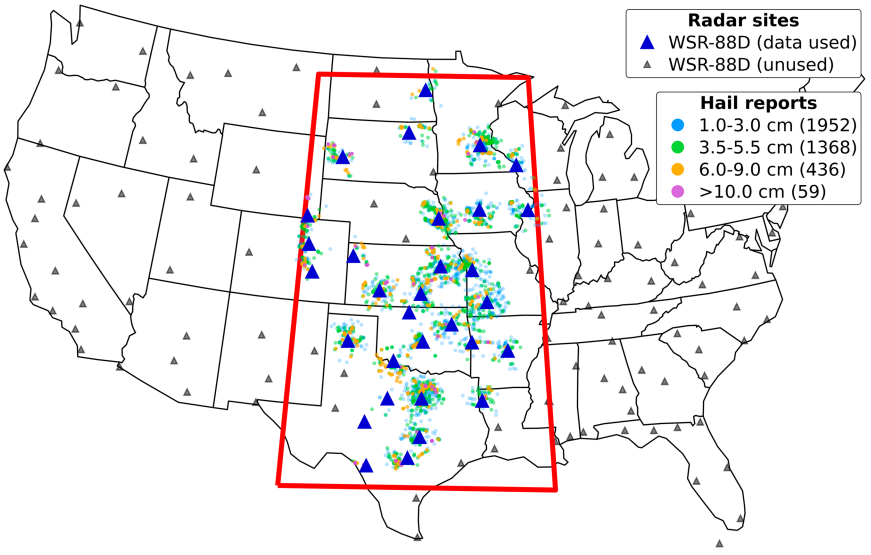

FIG. 1. Map of the contiguous United States with our study area outlined in red. Radars used for at least one event are indicated by large blue triangles, and unused radars are indicated by small black triangles. Hail reports used in the study are shown as small dots, colored by hail class.

onto a 1 km × 1 km Cartesian grid at isotherm heights of −10°, −20°, −30°, and −40°C (CAPPI) using an inverse distance weighting function. The total extent of the grid was $-200\text{ km} < X < 200\text{ km}$ and $-200\text{ km} < Y < 200\text{ km}$ centered on the respective radar site. The isotherm heights were uniform across the grid and determined using Rapid Refresh (RAP; Benjamin et al. 2016) model soundings at the grid point closest to the radar location. These data were retrieved from model data message (MDM) files provided with the NEXRAD data on AWS. Geopotential heights of the model soundings were adjusted to heights above ground level, and then, the temperature data were linearly interpolated to find the heights of the respective isothermal level. If no model sounding for a particular day and location was available in the level II database—this was the case for 38 events, most of them before 2017—we used the nearest observation sounding before the start of the event. In 20 of these cases, radiosonde observations were available within 60 km of the radar site, and the overall maximum distance between radar and radiosonde launch location was 225 km. The temporal offset between radiosonde launch and event begin was confined to 12 h. We tried to estimate the error caused by this temporal offset by analyzing the 12-h trends of sounding-derived −10°C isothermal heights for these 38 cases and found that they changed more than ±500 m in only 3 cases.

For all events, we also prepared additional radar variables later required for the updraft analysis. Most notably, we calculated vertically integrated liquid (VIL) from Z_H (Greene and Clark 1972), azimuthal shear from radial velocity (Mahalik et al. 2019), and the depolarization ratio (DR) proxy (Ryzhkov et al. 2017) using Z_{DR} and ρ_{HV}:

$$\mathrm{DR} = \frac{1 + Z_{\mathrm{DR}}^{-1} - 2\rho_{\mathrm{HV}} Z_{\mathrm{DR}}^{-0.5}}{1 + Z_{\mathrm{DR}}^{-1} + 2\rho_{\mathrm{HV}} Z_{\mathrm{DR}}^{-0.5}}.$$

The DR value is maximized for combinations of low Z_{DR} and low ρ_{HV} and can be used as a filter for nonmeteorological echoes, e.g., hail spikes (Kilambi et al. 2018), but has also been suggested as an indicator for hail. Another derived variable is logCC, which is a transformation of ρ_{HV} to a Gaussian distribution proposed by Keat et al. (2016):

$$\mathrm{logCC} = -\log_{10}(1 - \rho_{\mathrm{HV}}).$$

By using logCC, we aim to facilitate statistical interpretation of copolar cross correlation, e.g., when calculating average and standard deviation within the updraft region.

b. *Updraft detection and tracking*

One of the most crucial aspects of our study is the accurate detection and tracking of storm updraft areas. Updraft detection was accomplished using the "Z_{DR} hotspot technique" by Krause and Klaus (2024). In a nutshell, the method focuses on identifying positive anomalies of Z_{DR} at the −10°C isotherm height, a level typically reached by Z_{DR} columns, which are widely recognized as suitable proxies for thunderstorm updraft location (e.g., Kumjian et al. 2014; Snyder et al. 2017; Ilotoviz et al. 2018; Chase et al. 2024). For each point of the gridded Z_{DR} field, the hotspot method calculates the median Z_{DR} in an inner 3 km × 3 km grid box, centered on the grid point in question, and in an outer 7 km × 7 km box, excluding the points from the inner box. The subtraction of median Z_{DR} in the outer box from median Z_{DR} in the inner box yields the "hotspot value." Due to the hotspot technique's pixel-to-pixel comparison, it is not affected by Z_{DR} miscalibration, which is a common problem with the widely used algorithm by Snyder et al. (2015). In contrast to the "modified Thunderstorm Risk Estimation from Nowcasting Development via Size Sorting" (mTRENDSS) algorithm described in French and Kingfield

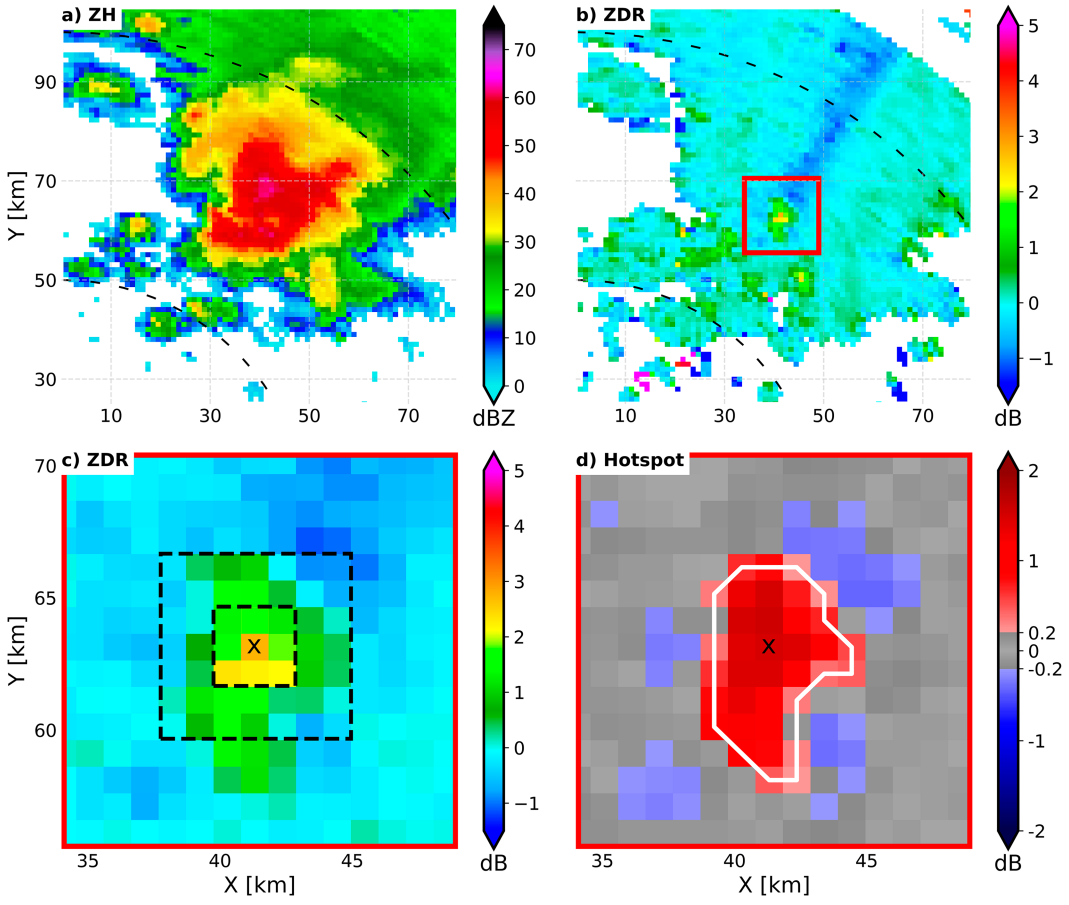

FIG. 2. Example output of the Z_{DR} hotspot technique for the NEXRAD site at Fort Worth, Texas (KFWS), at 2234 UTC 11 Jun 2016. Storm-wide (a) Z_H and (b) Z_{DR} CAPPIs at the height of the −10°C isotherm and, within the red inset rectangle in (b), (c) Z_{DR} with inner and outer box used for hotspot calculation and (d) hotspot field.

(2021)—which is also insensitive to Z_{DR} miscalibration—the search window for background Z_{DR} values is much smaller with the hotspot technique, limiting the impact of differential attenuation. As a threshold for marking potential updraft areas, we used a hotspot value of 0.2 dB, corresponding roughly to one standard deviation of the hotspot value during a subset of 11 events. Figure 2 shows an example of a "Z_{DR} hotspot" in a mature storm, and Fig. 2c illustrates the two boxes used for deriving the hotspot value.

To merge adjacent updraft detections, we dilated and subsequently eroded the updraft objects by one pixel (equaling 1 km), a process known as "closing" in mathematical morphology. Next, we required a minimum updraft size of 5 km^2, and at least one pixel within the updraft where $Z_H > 25$ dBZ, which yielded the best results based on a manual inspection of selected cases.

This procedure still left spurious updraft detections at the fringes of storm cells, and the resulting objects were difficult to track individually due to their small spatial scale and occasional intermittent appearance over time. In an attempt to resolve both issues, we coupled the updraft detections to an existing cell tracking algorithm, the "multicell identification and tracking" (MCIT) system (Rosenfeld 1987; Hu et al. 2019). Although untested by us, other storm cell tracking methods could replace MCIT as part of this algorithm. The linking of updraft objects to MCIT cells and the removal of erroneous updraft objects are described in detail in appendix B. An example of MCIT cells with their assigned updraft areas is shown in Fig. 3. After assigning the updrafts, we filtered the MCIT cells to ensure a high-quality dataset. We kept MCIT cells when the cell center was within 120 km of the radar range and discarded all MCIT cells that were tracked for less than six time steps (~30 min).

c. *Linking hail reports to storm cells*

Hail reports were matched with MCIT cells. We used the time information of the hail report and extended it by a period of 15 min before the report time window began and 2 min after the report ended, as hail reports often appeared to lag behind the passing of a thunderstorm. Within this interval (−15, +2 min), we assigned the hail report to the MCIT cell that had its VIL maximum closest to the report location, as long as the distance between report location and cell center was <15 km and the MCIT cell had a maximum VIL > 25 kg m^{-2}, because weaker cells are very unlikely

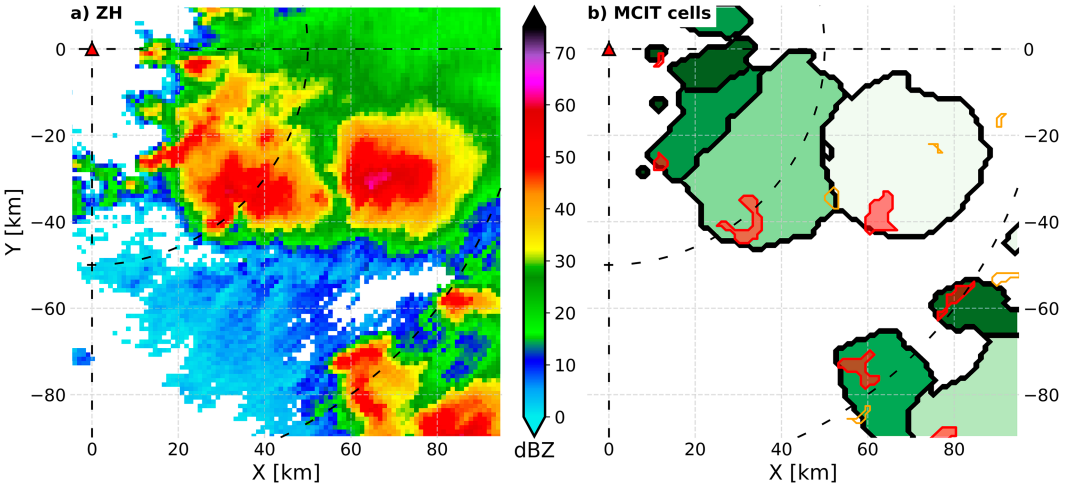

FIG. 3. Example of MCIT cells with assigned updrafts as seen from the NEXRAD radar at Frederick, Oklahoma (KFDR), at 0035 UTC 23 May 2020. (a) CAPPI of Z_H at the −10°C isotherm height, and (b) different MCIT cells are shaded greens with assigned (red; filled) and discarded (orange; contours only) updrafts.

to produce hail (Edwards and Thompson 1998). When a hail report was assigned to an MCIT cell, the hail report time was changed to the radar scan time when the MCIT cell center (location of VIL maximum) was closest to the report location. In total, 753 cells distributed over the 114 events were linked to at least one hail report.

There are many limitations in the hail reports we used for this study; there is a lack of observations in sparsely populated areas, the hail size reported is often categorical and estimated rather than continuous and measured, and the exact timing of the hailfall at the ground contains major uncertainties (Allen and Tippett 2015). Our assignment process was designed to yield the earliest plausible time of hail at the ground. This is particularly relevant as the overarching goal of this study is to examine potential nowcasting applications, which are characterized by short lead times. In addition to timing uncertainties, the reported hail size may also not be representative of the maximum hail size of a given storm at that time (Ortega et al. 2009; Blair et al. 2017). While we cannot resolve this problem when there is a lack of reports, we always match the largest reported hail size if there are multiple reports tagged to the same MCIT time step.

In preparation for the following steps, we then categorized the storm cells by the maximum size of hail they produced using the definitions from Table 1. We expected that the wide size classes would be better suited to account for inaccuracies of the maximum hail size produced by a storm. Note that we also intentionally left a buffer between the hail size classes analogous to Homeyer et al. (2023), and hail reports falling between the classes were not used. From the 753 categorized hail cells, we then removed those that were tracked for less than five time steps before the first hail report fell into their respective class, yielding a total of 384 cells.

d. Retrieval of potential hail indicators and precursors

To analyze the different updraft characteristics of storms classified by maximum hailstone size and to generate the training data for the RF model, we extracted radar data from the updraft areas at multiple levels. In addition to the area identified at the −10°C isotherm height, we also extracted radar data from the same area at the −20°, −30°, and −40°C isotherm heights, since we expect that hail growth processes would also occur at these heights. To account for tilted updrafts and to include information from the immediate surroundings of the Z_{DR} column (such as very low Z_{DR} adjacent to the column, possibly indicating the presence of large hail), we dilated the updraft area by 3 pixels (=3 km) on each level. We then collected the radar metrics that we expected to be related to the storm's potential to grow hail or to directly indicate the growth and presence of hail. We evaluated these metrics for each hail category from the previous step.

For the initial analysis, we selected time steps shortly before the first hail report falling into the respective hail class definition. We then calculated the statistical distributions of the hail metrics for each hail size class. Among the metrics are not only basic aggregations such as average ρ_{HV} or Z_{DR} within the updraft but also more complex combinations such as maximum DR where $Z_H > 50$ dBZ or area-based metrics like the updraft area with $Z_H > 50$ dBZ. In total, we calculated 60 different radar metrics potentially related to hail formation or presence of hail at each available isotherm height. Only four metrics were not exclusively collected from the detected updraft area but taken from the MCIT output: total cell size, maximum VIL, MESH (version by Witt et al. 1998) within the storm cell, and the distance between updraft center and MCIT cell center. A list with short descriptions of all the metrics we tested is given in Table A2 in appendix A.

To investigate whether there were any consistent precursor signals indicating imminent hailfall, we also analyzed the temporal evolution of radar metrics relative to the first hail observation of the respective hail size class. Due to differing scan rates of the radars, we linearly interpolated available tracking data to 2.5-min increments for the period from 45 min before to 15 min after the first report of given size and then calculated

the median for each time step across all available cells in each hail size category.

e. RF model training and tuning

Careful data selection, preprocessing, and model choice are critical for the success of any ML-based forecast system, even more so in our case considering the limitations of hail observations on which the model will be trained on. As previously stated, we chose to classify storm cells into broad hail size categories (Table 1) and consequently opted for a classification (discrete output) rather than a regressor (continuous) model. We considered multiple ML methods but ultimately decided to use an RF due to its simplicity, easy parallelization and hyperparameter tuning, and its ability to model highly nonlinear systems (Gagne et al. 2014). RFs are an ensemble of decision trees suited for regression and classification tasks and have already been successfully applied to a range of problems in the field of radar meteorology. For instance, Wolfensberger et al. (2021) used an RF for quantitative precipitation estimation (QPE) within the Swiss C-band radar network, achieving higher skill than the operational QPE system. Medina et al. (2019) used an RF to nowcast downbursts with dual-polarization radar data. Czernecki et al. (2019) showed the ability of RFs to predict the occurrence of hail > 2 cm using a mix of numerical weather prediction (NWP) and radar data but did not focus on hail size discrimination. RFs also offer the possibility to examine the importance of individual predictors—"features" in the context of ML—for their classification decision, allowing some interpretation of the role of predictors for the process in question (McGovern et al. 2019).

Note that all the time steps within this window are treated equally when training the model, without consideration of the time difference relative to the observation. Although some radar signals indicating hail may vary considerably within the 35-min window, this approach substantially increases the size of the training dataset, which is critical for training any RF. Furthermore, the timing of signals is already obscured by the error of hail report times. Consequently, initial experiments with shorter time windows showed a reduction in model skill. In addition to the storm cell classification based on maximum hail size, we also introduced a "no hail" category so we could train the RF model to discern between hailstorms and storms not producing hail. For this, we used particularly rigorous quality checks: First, we only considered MCIT cells without tagged hail reports, excluding all cells that are found in one of the hail classes. Afterward, we checked the remaining cells for any hail report within 50 km for a time window of 40 min before and after each time step. If these conditions were satisfied, the cell was added to the no-hail dataset, ultimately containing 740 tracked cells. We explicitly did not add any radar-based quality checks due to the risk that the RF would be trained to pick up any self-chosen thresholds for the hail/no-hail discrimination.

For tuning and training the RF, we used Python 3.11 with the machine learning library scikit-learn 1.3.1 (Pedregosa et al. 2011). As the primary score for assessing model performance, we used the Peirce skill score (PSS; Peirce 1884; Manzato 2007) which is widely used in forecast verification and in the context of ML evaluation (e.g., Jergensen et al. 2020; Kamangir et al. 2020). After initial tests with a single RF setup, we decided to use a stacked model consisting of two separate RFs due to improved skill. The first RF model (hereafter "hail detection model") was designed for the binary decision of whether a cell will grow hail > 1 cm or not. If the hail detection RF indicates hail, the second RF model (hereafter "hail size model") is applied and classifies the hail size into the four categories detailed in Table 1.

After selecting and preprocessing the input data, we reduced the number of input variables by eliminating highly correlated radar metrics that would not improve the ML model but complicate the interpretation of feature importance (e.g., Gregorutti et al. 2017). For this, we used hierarchical clustering on the Spearman rank–order correlations and kept only a single feature from each cluster, manually defined by a scalar threshold of 0.8 after visual inspection of the cluster dendrogram (not shown). This resulted in a significant reduction in variables from 243 to 34 as many of them were highly correlated, e.g., most of the metrics related to azimuthal shear on all levels from $-10°$ to $-40°$C. We further use the same set of 34 variables for both RFs in our model stack, listed in Table 2. All variables extracted from the hotspot area are denoted with the prescript "HS" followed by the isotherm height of the respective CAPPI level.

Before training the model, we split the available data into a training set containing approximately 80% of all data and a heldout test set used for evaluation with the remaining roughly 20%. In creating these subsets, we paid attention not to split up tracking data of the same MCIT cell to avoid artificially increased skill scores due to autocorrelation.

In the next step, we used a grid search with fivefold cross validation on the training data to tune the RF hyperparameters for both models separately ("GridSearchCV" in the scikit-learn package), with PSS serving as evaluation score for the tuning. The combination of hyperparameters that emerged as maximizing the skill in the cross-validation grid search is given in Table 3.

Next, we trained the two RF models using the variables from Table 2 and the hyperparameters from Table 3. Due to the strong imbalance of hail classes in the training set—8% of all time steps belonged to the highest hail size class compared to 31% and 36% for the lowest two categories, respectively—we decided to add a calibration step to the hail size RF in addition to weighing rare hail classes higher during training. For calibration, we used logistic regression to map the probabilistic output of the classifier to the observed event probability given a certain class is predicted. This approach, although mainly designed for probability estimates, turned out to improve the deterministic classification particularly for the rare categories of significantly severe and giant hail.

f. RF model evaluation

After training the models, we assessed the importance of individual features using two different approaches: first, the impurity importance, calculated directly in the RF training process as the decrease in an impurity score—in our case Shannon entropy—achieved by splits in the decision tree involving a certain predictor (Breiman 2001; McGovern et al. 2019); second, the permutation

TABLE 2. Final list of variables used for training two RFs for hail nowcasting. All variables with preceding HS in their name are extracted from the detected updraft area.

Name	Description
MESH	Maximum expected size of hail (Witt et al. 1998); maximum value within the MCIT storm cell
MCIT_HS-10Cdist	Distance from MCIT cell core to updraft core
HS-10C_AreaRef40	Area $Z_H > 40$ dBZ (−10°C)
HS-10C_AveAZS	Average azimuthal shear (−10°C)
HS-10C_AveDR	Average DR proxy (−10°C)
HS-10C_AveHS	Average hotspot value (−10°C)
HS-10C_AveHSLogCC	Average hotspot value of logCC; average HS (logCC) (−10°C)
HS-10C_AveRho	Average cross correlation (−10°C)
HS-10C_AveZdr	Average Z_{DR} (−10°C)
HS-10C_CellArea	Hotspot area (−10°C)
HS-10C_DRAveRef40	Average DR where $Z_H > 40$ dBZ (−10°C)
HS-10C_Drcount14Ref40	Area of DR > −14 where $Z_H > 40$ dBZ (−10°C)
HS-10C_DiffAZS	Differential azimuthal shear; minimum azimuthal shear subtracted from maximum azimuthal shear (−10°C)
HS-10C_MaxHSLogCC	Maximum hotspot of logCC; maximum HS (logCC) (−10°C)
HS-10C_MaxLogCC	Maximum of logCC (−10°C)
HS-10C_MinHS	Minimum hotspot value (−10°C)
HS-10C_MinHSLogCC	Minimum hotspot of logCC; minimum HS (logCC) (−10°C)
HS-10C_MinRef	Minimum Z_H (−10°C)
HS-10C_SigColdHS	"Cold spots" of Z_{DR}; area of HS < −0.2 (−10°C)
HS-10C_SigColdHSLogCC	Cold spots of logCC; area of HS (logCC) < −0.2 (−10°C)
HS-20C_AveHS	Average hotspot value (−20°C)
HS-20C_Drcount14Ref40	Area of DR > −14 where $Z_H > 40$ dBZ (−20°C)
HS-20C_SigColdHS	Cold spots of Z_{DR}; area of HS < −0.2 (−20°C)
HS-20C_SigColdHSRef40	Area of HS < −0.2 where $Z_H > 40$ dBZ (−20°C)
HS-30C_AveHS	Average hotspot value (−30°C)
HS-30C_AveHSLogCC	Average hotspot value where logCC < 1.3 (−30°C)
HS-30C_DR17count	Area of DR > −17 (−30°C)
HS-30C_DRAveRef50	Average DR where $Z_H > 50$ dBZ (−30°C)
HS-30C_Drcount14Ref40	Area of DR > −14 where $Z_H > 40$ dBZ (−30°C)
HS-30C_MaxHS	Maximum hotspot value (−30°C)
HS-30C_MaxHSLogCC	Maximum hotspot of logCC; maximum HS (logCC) (−30°C)
HS-30C_MinHS	Minimum hotspot value (−30°C)
HS-30C_MinHSLogCC	Minimum hotspot of logCC; minimum HS (logCC) (−30°C)
HS-40C_Drcount14Ref40	Area of DR > −14 where $Z_H > 40$ dBZ (−40°C)

importance, which is calculated on the test data by randomly shuffling values of one feature and observing the decrease in PSS compared to the unshuffled input data. The random shuffling is repeated 50 times to calculate the median decrease in the skill score. If the decrease in skill when shuffling one feature is high, this feature can be assumed to be important for the model's decision. We note that the importance of correlated features may be underestimated using this method, but our variable reduction from section 2e limits such error.

For the nowcasting verification, we first use the hail detection RF for the hail/no-hail nowcast. If this model indicates hail, the hail size RF is used to forecast the expected hail size class. Using the heldout test dataset, we verified this hail classification and evaluated at which lead time between 40 and 10 min the nowcast achieved the highest skill scores, using the result of this evaluation to generate six nowcasting examples from cells in the independent data (see section 4).

3. Evaluation of hail precursor signals

We begin by showcasing a subset of the metrics explored as potential hail indicators. In Fig. 4, we show statistical distributions

TABLE 3. Hyperparameters used for training the two RF models.

Hyperparameter	RF hail detection	RF hail size
Maximum depth of individual trees	15	15
Minimum samples per leaf	1	2
Minimum samples per split	2	6
Number of estimators (trees)	500	200
Bootstrapping	False	True

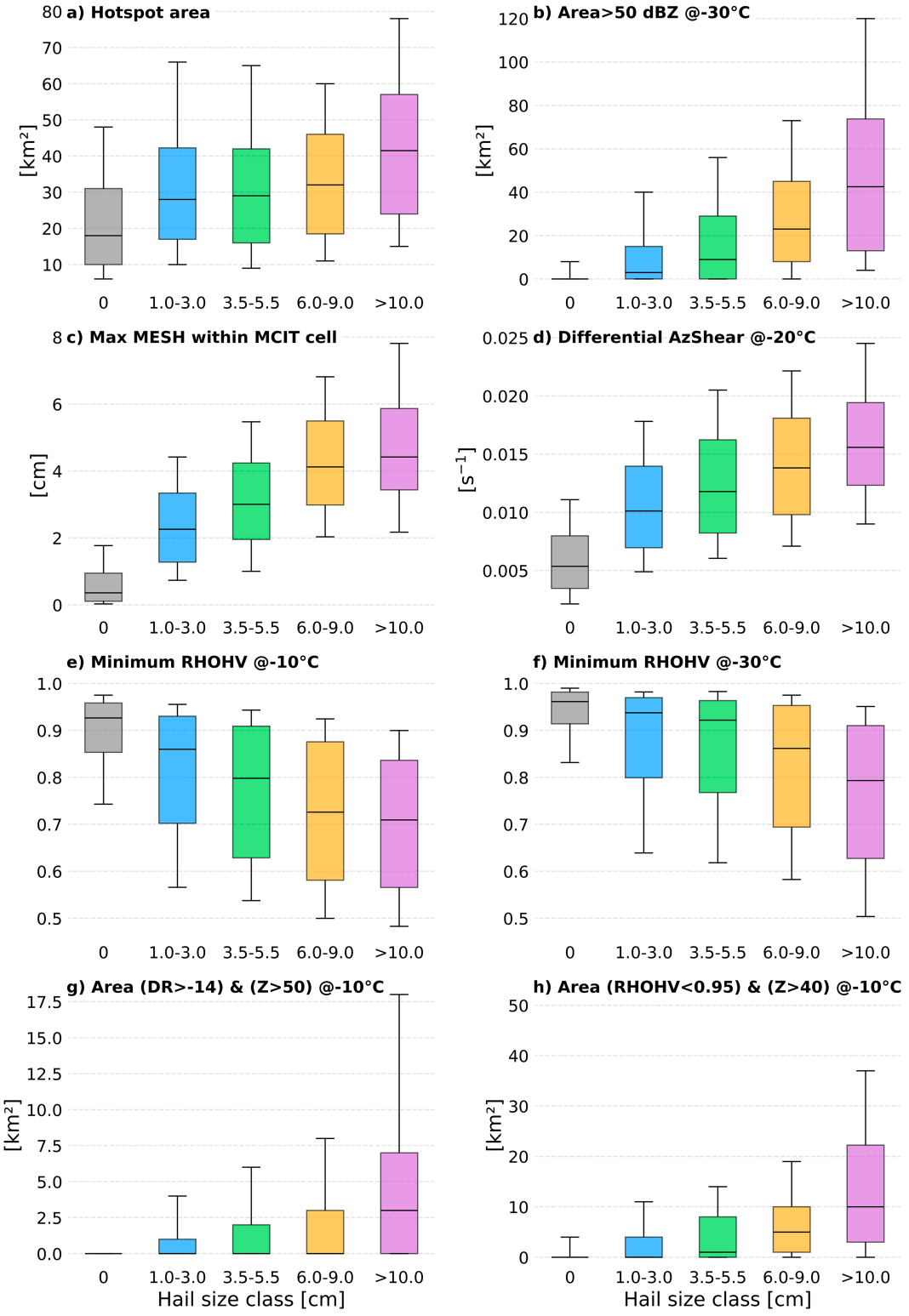

FIG. 4. Boxplots of different radar metrics prior to hailfall, ordered by hail size class. The whiskers indicate the 10th and 90th percentiles of the distribution, respectively.

of eight metrics stratified by hailstorm classes, sampled from 40 to 10 min prior to the first hail report of the respective hail class. This time window was selected to analyze early signals for the respective hail size.

Interestingly, a feature for which we expected a clear signal related to hail size, the total detected updraft area, increases only marginally from small to significantly severe hail and slightly more toward giant hail ≥ 10 cm (Fig. 4a). This result is somewhat contradictory to recent studies which have emphasized the importance of updraft size for large hail growth, by providing either more favorable hail growth trajectories or higher liquid water content due to less entrainment (Kumjian and Lombardo 2020; Peters et al. 2020; Gutierrez and Kumjian 2021). We note that our Z_{DR}-based method may underestimate total updraft size particularly in the presence of large hail due to its intrinsic low Z_{DR} at the S band. Additionally, sensitivity tests conducted for one of the events revealed that the hotspot size is rather sensitive to changes in the −10°C isothermal height (median hotspot size changed up to 10 km^2 by shifting the −10°C isothermal height 500 m downward or upward). In contrast, aggregate metrics sampled within the detected hotspots remained largely unaffected by moderate changes in isothermal height.

Moving to the upper-level reflectivity core, we observe a stronger hail-size-related signal as shown by distributions of the updraft area with $Z_H > 50$ dBZ at the −30°C isotherm (Fig. 4b). The upper-level reflectivity core seems particularly well suited to detect giant hail. At an isotherm height of −40°C, only cells producing hail ≥ 6 cm have a 50-dBZ area within the updraft that is significantly above zero with approximately 10 km^2 and cells with hail ≥ 10 cm have a median 50-dBZ area of 35 km^2 (not shown). However, the interquartile ranges still overlap largely between these categories. Another Z_H-related variable, MESH (Fig. 4c), is the only metric shown that is not sampled from within the updraft area, but the whole storm. It aligns well with the reported hail size up to the "large" category, while hail ≥ 6 cm is strongly underestimated, a finding in line with previous evaluations (Ortega et al. 2009; Witt et al. 2018; Murillo and Homeyer 2019). Differential azimuthal shear (Fig. 4d) is a storm intensity metric indicating cell rotation and differs strongly between nonhail and hail cells, but the increase levels off toward larger hail sizes. This confirms that storm rotation is an important prerequisite for large hail growth, as well documented in the literature (e.g., Taszarek et al. 2017; Allen et al. 2019; Kumjian and Lombardo 2020). Nevertheless, the statistical overlap of azimuthal shear between the hail size classes limits its use as a hail size nowcasting tool. Figures 4e and 4f show the minimum ρ_{HV} within the updraft at −10° and −30°C, respectively. Again, there is a significant overlap in the statistical distributions of this variable between the hail size classes. At both levels, ρ_{HV} decreases with hail size, but at different rates. At −10°C, ρ_{HV} decreases most strongly in the lower end of the hail spectrum and with little difference among hail sizes ≥ 6 cm. In contrast, at −30°C, there is only a slight decrease in ρ_{HV} between the smaller hail categories, but a sharp dip toward giant hail. We speculate that this is caused by the stronger and taller updrafts that support giant hail growth, with mixed-phase processes and hail growth occurring at higher levels. A similar observation was made by Witt et al. (2022) who showed a strong negative correlation between minimum ρ_{HV} at the −30°C isotherm height and hail sizes above 12 cm and a weaker correlation at the −10°C isotherm height. Examining this phenomenon with higher radar frequencies could reveal larger statistical separations between the hail size classes, as previous research shows that the ρ_{HV} reduction in hail is generally more pronounced at the C and X bands (Kaltenboeck and Ryzhkov 2013).

Last, we highlight two area-based metrics that we deem promising as early signals of giant hail. In Fig. 4g, we show the number of updraft grid points fulfilling the value combination of DR > −14 dB and $Z_H > 50$ dBZ at the −10°C isotherm height. DR primarily depends on particle shape and phase composition and has been shown to strongly increase with hail size (Ryzhkov et al. 2014). We tested multiple thresholds of DR and found −14 dB to distinguish most strongly between significantly severe hail and giant hail. This threshold was combined with a requirement for $Z_H > 50$ dBZ to ensure that large meteorological scatterers were present at the same location, because high DR can also be caused by nonmeteorological scatterers (Kilambi et al. 2018). Only cells that contain giant hail have a median area > 0 meeting the combined DR–Z_H condition. The small area even for the giant hail category (median of 3 km^2) might imply that this metric is closely related to the actual area of hail of this magnitude rather than being an indirect proxy through storm intensity. In contrast, much larger areas satisfy the condition $\rho_{HV} < 0.95$ and $Z_H > 40$ dBZ at the −10°C isotherm (Fig. 4h). This is likely because the thresholds are less restrictive and areas with smaller hail and mixed-phase processes are included. Still, there is a strong increase in this area for significantly severe and giant hail cells. We attribute this to two likely causes: The updraft area occupied by hail and mixed-phase processes generally increases with hail diameter, and/or ρ_{HV} decreases downradial of hail due to nonuniform beamfilling, which is regularly observed in intense storms (Ryzhkov 2007). We purposely selected 40 dBZ as the threshold in this case as this reduced the interquartile range for most hail categories.

In Fig. 5, we show the temporal evolutions of median and 25th and 75th percentiles for the same eight metrics prior to and shortly after the first hail report of a given size. These data have to be examined with caution, as hail report timing errors may obfuscate underlying time-dependent signals.

As seen before, the hotspot area does not diverge significantly between the hail classes from 1 to 9 cm. Interestingly, the hotspot area of giant hail cells differs most strongly from the smaller-sized hail classes more than thirty minutes before the hail reports and then decreases until being almost equal to the other classes shortly before estimated hailfall. There is a similar decrease in hotspot area in the significant severe hail class, although less pronounced and with later onset ($t = -20$ min). We speculate that in both cases the decrease is caused by the low Z_{DR} of large hail, reducing the detected updraft area. Updraft rotation, estimated by azimuthal shear in Fig. 5d, on the other hand, is stable over time, with not only slight differences between the classes but also large overlap of the distributions.

The upper-level reflectivity core (Fig. 5b) provides robust differences between the hail classes, especially toward larger

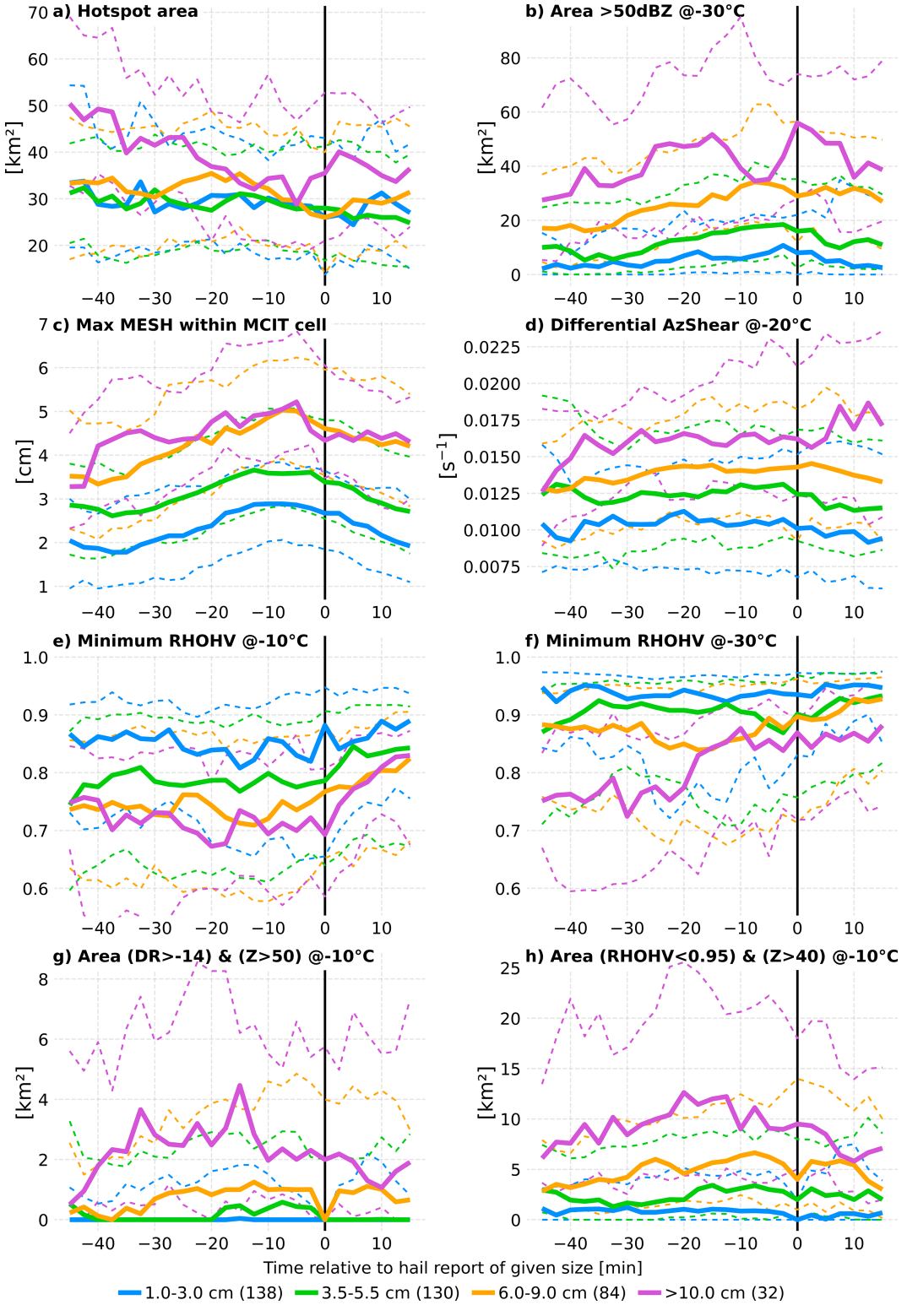

FIG. 5. Median time series of selected metrics relative to the first hail report of given size. The dotted lines correspond to the 25th and 75th percentiles. Numbers next to the hail class definitions indicate the number of individual storm cells contributing to the time series.

hail. In contrast, MESH in Fig. 5c only shows minor differences between the two highest hail classes but performs well for other categories. All MESH medians reach their maximum around -10 min $< t < -5$ min, increasing the confidence that our assigned report time serves as a good estimate of hailfall at the ground, as it takes some minutes for hailstones to descend to the surface.

The minimum ρ_{HV} at $-10°$ and -30°C shown in Figs. 5e and 5f indicates that hail growth and mixed-phase processes in severe, significantly severe, and giant hail cells are ongoing for an extended period as ρ_{HV} is low (minimum $\rho_{HV} < 0.9$ at -10°C isotherm height) already 30–45 min before the report time, while the area of low ρ_{HV} overlapping with high Z_H at the -10°C isotherm (Fig. 5h) is maximized at 10–20 min before the report time. We note that it is likely that the storms already produce significant hail before $t = 0$, as we take the time of the first report exceeding a certain hail size threshold as reference. Therefore, giant hail cells, for example, may have been producing severe hail for a longer time before reaching the giant hail class for the first time—which appears likely given the long lifetime of supercells that are capable of producing such large hail.

Our conclusion from this analysis is that no single metric from our arsenal of potential hail (growth) indicators performs well across the entire spectrum of hail sizes. We found that some radar metrics discriminate well between hail/no hailstorms, such as differential azimuthal shear. Other metrics, for example, MESH, discriminate well between hail up to the significantly severe class but do not capture giant hail. Yet other metrics, namely, minimum ρ_{HV} in the upper part of the updraft or area-based metrics using a Z_H-ρ_{HV} threshold, showed the strongest signal at the upper end of the hail size spectrum but were generally not suited for distinguishing between hail sizes < 5 cm.

Therefore, our analysis finds that a nonlinear modeling approach using a variety of input features is most promising for hail size nowcasting as we illustrate below.

4. Hail nowcasting using the random forest

a. Feature importance

Before verifying model predictions, we assess the value of individual model features (predictors) using the two feature importance methods outlined in section 2f. A selection of the 10 most important features emerging from each method is shown for the hail detection (Figs. 6a,b) and the hail size model (Figs. 6c,d), respectively.

The two variants of feature importance agree on MESH as the most important predictor for both hail detection and hail size classification. While MESH is leading by far in hail detection, the next-ranking features follow MESH much more closely in the hail size model. This reflects the decreasing value of Z_H—and indirectly MESH—once hailstone sizes enter the Mie resonance regime and is in line with the results from section 3. However, MESH is valuable for discerning between smaller-sized hail and therefore still ranks as the most important feature in the hail size model. For the hail detection model, differential azimuthal shear at -10°C isotherm height ranks as the second most important feature, indicating the relevance of a rotating updraft for hail development. Radar proxies for storm rotation are also found in the top five most important features of the hail size model but are slightly surpassed by polarimetric variables or combinations thereof. One of these combination metrics featured prominently in both hail detection and size classification is average DR where $Z_H > 50$ dBZ within the hotspot area at the height of the -30°C isotherm, underlining the potential of DR in hail detection. We note that this variable also implicitly contains the information about whether a storm cell can reach >50 dBZ at -30°C.

As expected, based on the evaluation in section 3, the total size of the detected hotspot is not among the highest-ranking features. Interestingly, PSS even increases when the total area is randomly shuffled, showing that the model improves when updraft size is not considered. Either other area-based metrics such as the size of the reflectivity core in the updraft area (HS-10C_AreaRef40) take up this role, or the Z_{DR} hotspot size is not a reliable proxy for the updraft size in the presence of large hail.

b. Model results

First, we evaluate the lead time at which the nowcasting model achieves the highest PSS compared to hail observations in the test data. As we trained the model using equally weighted lead times between 15 and 45 min, further investigations are necessary to find the optimal model lead time. To do so, we calculate the PSS in dependence on the lead time, for which we use a moving window of 5 min to increase the statistical robustness of the PSS. We calculate PSS for the stacked model as a whole and additionally for the hail size RF alone, as the expected large number of "correct negatives"—no hail predicted and none observed—may skew the PSS of the stacked model. The resulting PSS curves as functions of lead time (Fig. 7) show relatively stable PSS between 30- and 15-min lead time and a sharp drop in skill of the hail size model for lead times > 32 min. As expected, the PSS of the hail size RF alone is lower than the PSS of the combined model, which includes the correct negatives without any hail. Based on the combined inspection of both stacked model and hail size model, we conclude that lead times of 15–20 min are a sensible choice if applying the presented model for nowcasting. The stability of PSS from 32- to 15-min lead times may reflect the error margins of the hail report times and shows the challenge of correctly predicting the onset of large hail based on the data available. It is worth noting that skillful lead times could be considerably reduced in rapidly growing storms considering the typical 5-min lags between consecutive radar scans and modeling studies suggesting large hail growth times of only 10–15 min (Kumjian and Lombardo 2020; Kumjian et al. 2021).

We then evaluate the hail classes predicted by the stacked model on the test dataset against the observed hail sizes in a contingency table and perform a similar evaluation using MESH to put the new model in perspective to an existing hail

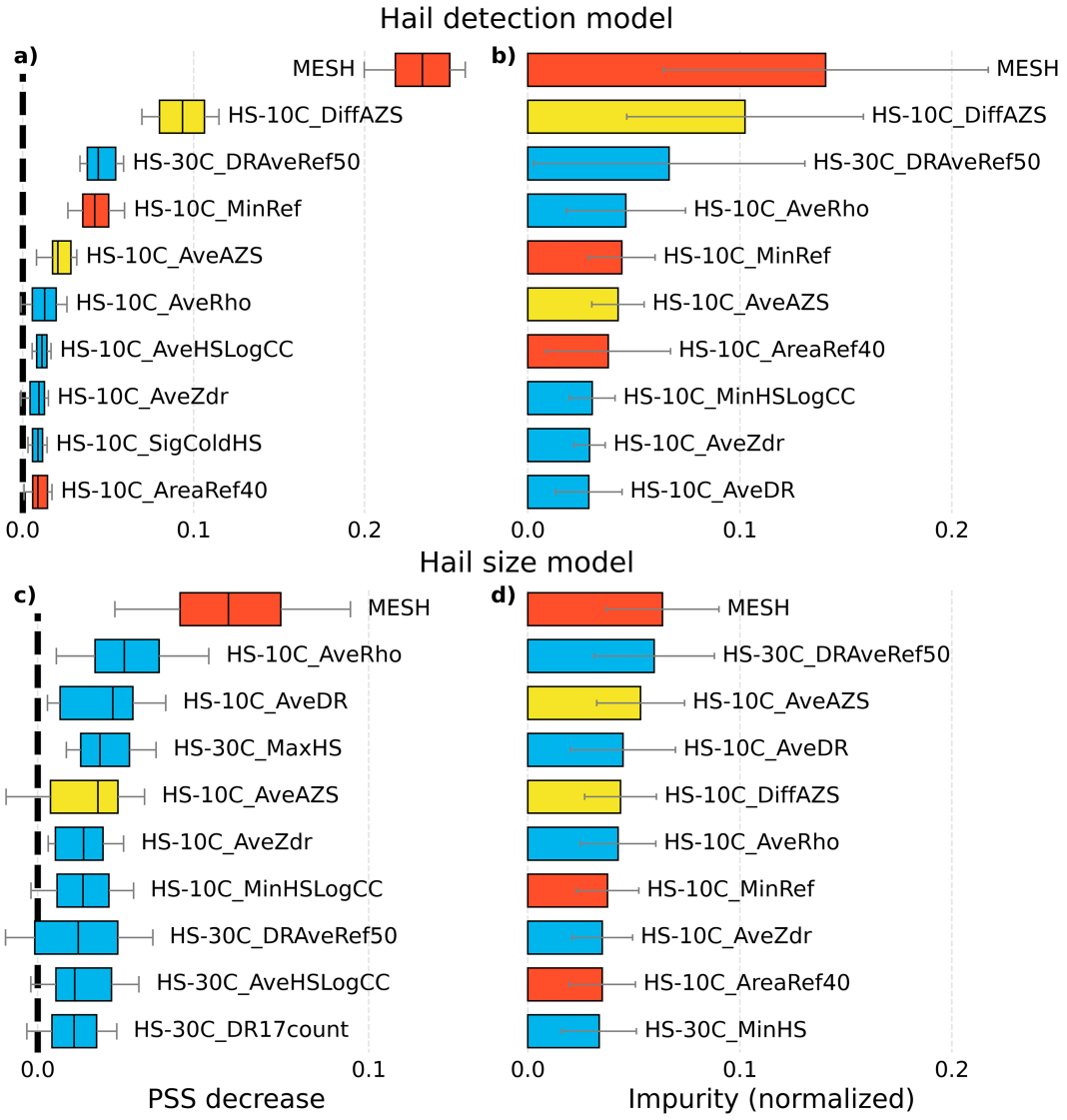

FIG. 6. (left) Permutation feature importance and (right) impurity importance for (top) the hail detection model and (bottom) the hail size model. The feature importance is given as boxplots with whiskers marking the 10th and 90th percentiles of PSS decrease. The impurity importance shown is the average impurity decrease per variable, with whiskers indicating the standard deviation. Reflectivity-based variables are in orange, azimuthal shear is in yellow, and polarimetric variables are in light blue.

size model (Fig. 8). For the RF nowcast model, we select only lead times between 30 and 10 min prior to the hail report and compare this to a "MESH nowcast" for the same lead times and, additionally, to a "MESH analysis" in a time window of ±10 min around the report time.

We note that the chosen time window of 20 min often yields more than one hail size estimate for one observation since multiple tracking time steps often fall into this period. This is done deliberately because we assume that hailfall of any given size is ongoing for at least some minutes. If MESH values fall into the gaps that are left as buffers in the RF model (e.g., between 3.0 and 3.5 cm), they are rounded up or down to the closest category. For the no-hail category, we utilized the respective cell dataset containing only cells with no associated hail reports throughout their lifetime (compare section 2e). We used all tracking time steps from these cells for building the contingency table; therefore, MESH nowcast and MESH analysis share the same first column in Fig. 8.

An important consideration when comparing RF nowcasts to MESH is that time steps without updraft detections could not be used in the RF contingency table; the RF relies on updraft data and cannot handle missing values. This is a serious limitation: 46% of the tracking time steps with subsequent hail observations of any size were without updraft detection, 24% of the time steps prior to at least significantly severe hail, and 22% of giant hail tracking time steps. In other words,

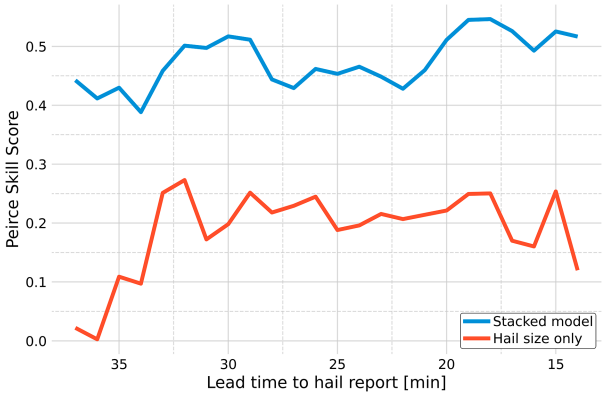

FIG. 7. Five-minute running mean of PSS in dependence on lead time.

successful updraft detections and assignments are increasingly likely with increasing hail size, but far from always available. As a result, the number of observations is significantly higher in the MESH contingency tables.

The contingency table for the RF model (Fig. 8a) shows reasonable agreement between observations and forecasts among larger hail. The model often nowcasts giant hail accurately, with the forecasted category being either matched or only one class off. The hail/no-hail discrimination is also successful in most cases (285 hits in Fig. 8a). Model skill is lowest for the small hail class, which is often observed when the next-higher class is predicted. The overall PSS is 0.44, but this value is strongly influenced by the correct identification of nonhailstorms. Considering only observations of hail $\geq$ 1.0 cm, the PSS decreases to 0.19 but slightly increases to 0.25 for observed hail sizes $\geq$ 3.5 cm.

The MESH contingency tables (Figs. 8b,c) show good agreement around the severe hail class (3.5–5.5 cm) but reflect MESH's weakness to detect giant hail. There is also a substantial number of misclassifications between no hail and the smallest hail category as MESH is not optimized for small hail $<$ 2.5 cm (Witt et al. 1998). Interestingly, the MESH nowcast has slightly higher skill than the analysis, potentially because of the delay between hail formation in the cloud and fallout at the surface. The nowcast achieves an overall PSS of 0.41 that decreases to 0.20 for hail observations $\geq$ 1.0 cm and to 0.03 for hail $\geq$ 3.5 cm.

c. *Nowcasting example*

After evaluating general model performance, we turn to illustrative examples for nowcasting, aiming to answer the questions: 1) How does the RF model perform in a real-world nowcasting scenario of a severe hailstorm? 2) Under which conditions does it fail? We try to address these questions by applying the developed model to cells from the test dataset and comparing the forecasted hail size class to hail reports and maximum MESH within the storm cell area (Fig. 9). We selected the five hailstorms accountable for hail $>$ 10 cm for which most hail reports were available and the overall longest-tracked hailstorm, lasting almost 4 h and producing reported hail up to 8 cm. Based on the RF model evaluation in section 4b, we presumed a lead time of 20 min, and therefore, there are no nowcasts available for the early cell lifetime.

One obvious takeaway from Fig. 9 is the challenge of comparing radar data and model results to hail reports, as the reports have large gaps when MESH and RF nowcasts still suggest hailfall. However, if there are hail reports available, the RF model generally also indicates hail. RF nowcasts often move in concert with MESH, although with a time delay corresponding to the nowcast lead time, particularly evident by the temporal evolution of MESH and RF nowcasts in Fig. 9d between 0130 and 0200 UTC. Rapid intensification or decay of hailstorms is therefore rarely detected in time and sometimes seems to require a signal in MESH first (e.g., the weakening of KSHV20200425 in Fig. 9b after 0330 UTC). Still, the relationship between RF nowcasts and MESH is clearly nonlinear as shown by instances when significant increases or decreases in MESH prompt no change in the RF forecasted hail class, seen, for example, in the KTWX20150910 case (Fig. 9e).

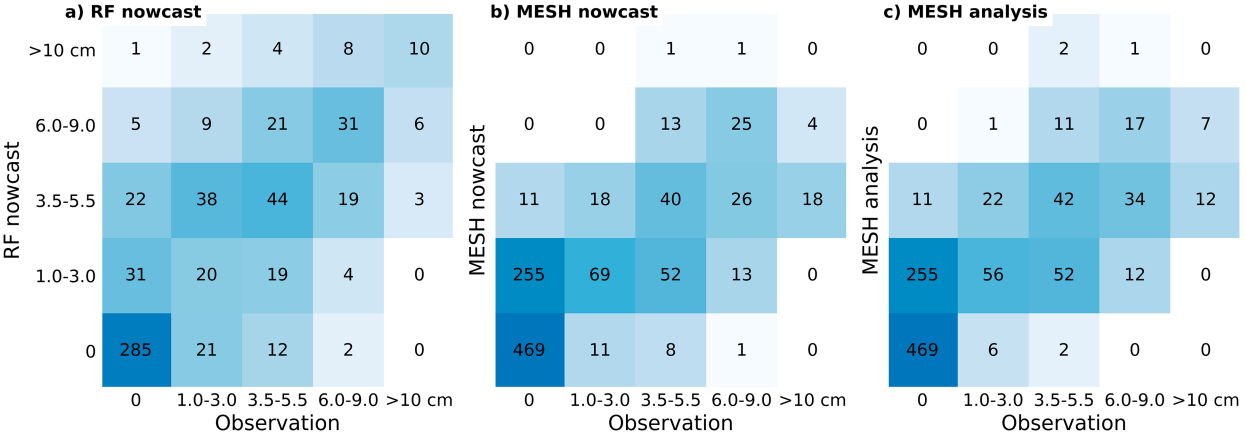

FIG. 8. Contingency table of the (a) RF model with lead times from 10 to 30 min, (b) MESH with lead times from 10 to 30 min, and (c) MESH from 10 min before to 10 min after hail report time. The number in each box indicates the amount of observation/forecast pairs falling into this bin, and the background colors are log scaled based on this number.

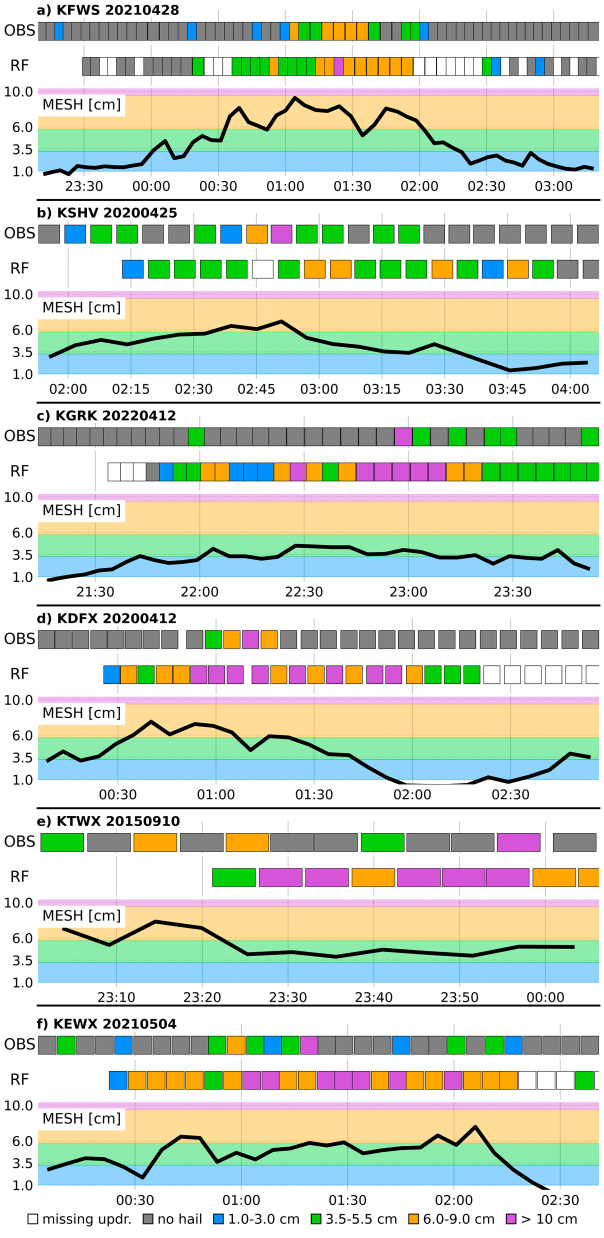

FIG. 9. Six example nowcasts of the RF model compared to hail observations (OBS) and MESH. OBS on the first line and time-aligned model (RF) 20-min forecasts of hail size on the second line, and colors indicate the hail size category. The box width is set to 5 min. Background shading of the MESH curve corresponds to the hail class color coding. Event names are composed of NEXRAD radar site ID followed by year, month, and day.

Despite no information from previous time steps being used in the RF nowcast, predictions are surprisingly stable over time and rarely change more than one class from one time step to the next apart from early cell lifetime. Furthermore, in all events except KFWS20210428 (Fig. 9a), no updraft assignments are missing during critical periods of the thunderstorm lifetime, i.e., when large hail can be expected. We manually inspected radar data from KFWS20210428 to explore the reasons for missed updraft assignments and found that there was no pronounced Z_{DR} column over large periods of the cell's lifetime, in stark contrast to obvious markers of storm strength such as very high $Z_H > 65$ dBZ at the height of the −10°C isotherm (another cell on the same day produced a new Texas hail size record of 16.3 cm). During peak intensity, approximately from 0030 to 0130 UTC, a small Z_{DR} column could be found, but was partly obscured by strong differential attenuation, most likely from the presence of large hail.

In another case, KDFX20200412 in Fig. 9d, the RF model seemingly does not predict the decay of the hailstorm. Significant severe hail and even giant hail are still predicted after 0130 UTC when MESH is already <3.5 cm, and no-hail reports are available. Upon inspection of the radar data from this day, we found that the hailstorm persisted in its distinct supercell appearance after 0130 UTC but moved directly toward the radar site and entered the inner 20-km radius around the radar at 0140 UTC, subsequently reducing MESH due to missing upper-level samples. Despite the precipitation core being largely located within the cone of silence and MESH < 1 cm, nowcasts generated based on the 0200 UTC radar scan still indicate hail potential, apparently only based on data from the updraft just to the southwest of the cone of silence and in contrast to low MESH.

There are also a few instances of underforecasting, for example, for KSHV20200425 in Fig. 9b. This is one of the rare cases when the RF model predicts small hail < 3 cm, while MESH indicates larger hail between 3.5 and 5.5 cm at the time of model initiation. We also revisited this case and found that this supercell storm was split up into two cells by the storm-tracking system at 0220 UTC, complicating updraft detection and assignment, the latter of which even failed for the 0226 UTC time step, resulting in the missing nowcast (white box in Fig. 9b) for 20 min later at 0246 UTC. Storm tracking resumes to correctly identify the cell at 0238 UTC, which is shortly before the peak hail activity at 0245 UTC, when multiple hail reports were submitted and pronounced differential attenuation is visible in the Z_{DR} field. This case shows not only the need for highly reliable storm tracking and updraft assignment algorithms in our feature-based approach but also the added value of the updraft information.

5. Conclusions and outlook

To study the updraft characteristics of storms producing differently sized hail and to find radar metrics differing among them, we analyzed a large sample of hailstorms in the Great Plains region of the United States based on a novel updraft detection system and ground observations of hail.

We found statistical differences between the hail size classes for several radar variables: Azimuthal shear allowed for the discrimination between hail/no-hail cells, MESH performed well for hail < 9 cm, and upper-level Z_H and ρ_{HV} in the updraft showed strong giant hail signals. However, no single variable was suitable for distinguishing between all hail sizes. Consequently, we aimed to combine a large number of

radar metrics in a feature-based hail nowcasting model using a random forest machine learning approach, trained on ground observations of hail. The model feature importance showed that MESH was the single most valuable predictor, followed by variables extracted from the updraft regions such as storm rotation (azimuthal shear) for identifying hail-producing storms and ρ_{HV}-based variables for assessing hail size. We evaluated the model and found that it showed promising performance for detecting very large and giant hail and for discerning cells without hail, but encountered difficulties discriminating between hail sizes below 5.5 cm. The skill of model forecasts is highest for lead times around 20 min, but different nowcasting examples showed that sudden temporal changes in storm intensity are often not accurately identified. Additionally, in the current model version, no nowcasts can be generated if no updraft has been detected and assigned to the storm cell, which can be the case if there is no pronounced Z_{DR} column or if the storm tracking provides erroneous output.

Based on the results of this study, we believe in the potential of the feature-based approach as it combines information about overall storm strength with radar data extracted from the updraft area of the storm. However, there are various aspects that require improvement before considering an operational use. First, storm tracking and updraft assignment are critical parts of the nowcasting system that need further fine-tuning in order to work reliably for all modes of severe convection, from small, isolated cells to large multicell clusters and linear convective systems. Second, we used hail report data from the Storm Prediction Center (SPC) database which comes with known limitations regarding the accuracy of hail time and sizing. While we attempted to enhance data quality by time-correcting hail reports and using broad hail size classes, the remaining uncertainties may still impede more time-specific nowcasts. Third, we focused on the detection of very large hail, which is reflected in the choice of events. We need to extend the available training dataset and include more small hail events for better model performance in classifying hail < 5.5 cm.

The Z_{DR}-based updraft detection has been one of the core elements of this study but is also its main physical limitation. Once large hail is present in the updraft, the Z_{DR} column can be (partly) obscured by the low Z_{DR} of these hailstones, yielding updraft detections that are significantly smaller than the actual convective updraft. In extreme cases, no updraft assignment is possible when giant hail is falling. In the future, we would like to experiment with other approaches so that hail size nowcasts will not depend on detections of Z_{DR} columns alone—either by utilizing ML approaches that handle missing data in their input or by training another RF using only storm-wide data. A comparison with the nowcasting model from this study would show whether the additional steps to identify the updraft and extract data there from are necessary to derive a valuable hail size model.

In the next model version, we would also like to include additional radar variables such as specific differential phase K_{DP}, which could aid in the discrimination of small hail sizes (e.g., Kumjian et al. 2019; Homeyer et al. 2023) and storm top divergence as another updraft strength proxy. Adding other data sources such as lightning data could also be considered, as lightning jumps have been shown to precede storm intensification and hailfall (Nisi et al. 2016; Farnell et al. 2018).

We are also interested in investigating the model performance on other datasets, for example, on C-band data and in complex terrain. Since our model has been trained solely on S-band data from the Great Plains, we expect that model performance will deteriorate in regions typically subject to different synoptic and microphysical conditions and for other weather radar frequency bands with different backscattering behavior, particularly when encountering large hail. However, the large dataset required for training ML models—both radar data and hail observations—is an obstacle for extending our methods to other domains, as, for example, high-quality long-term polarimetric radar data are often not freely available outside of the United States.

Last, we note that there are other scientific questions that would be interesting to address with our dataset, e.g., the discrimination between tornadic and nontornadic supercells (compare Loeffler et al. 2020) or the assimilation of detected updrafts in NWP (Carlin et al. 2017). Furthermore, the hotspot technique could be extended to find anomalies of other polarimetric variables such as ρ_{HV} or K_{DP} and explore their role in hail growth.

Acknowledgments. The authors would like to express their gratitude to Dr. Alexander Ryzhkov, Dr. Jeff Snyder, Dr. Harald Rieder, and Dr. Rudolf Kaltenboeck for their encouragement, scientific advice, and proofreading of the manuscript. Special thanks go to Dr. Jeff Snyder for introducing the first author to the art of storm chasing in Oklahoma. Vinzent Klaus' research stay at the Cooperative Institute for Severe and High Impact Weather Research and Operations (CIWRO) in Norman, Oklahoma, was financed by a "Marietta Blau scholarship" of the Austrian Agency for Education and Internationalisation (OeAD). Funding for John Krause was provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement NA21OAR4320204, U.S. Department of Commerce.

Data availability statement. Radar data and most hail reports used in this study are publicly available from the National Centers for Environmental Information (NCEI) at https://www.ncei.noaa.gov/products/radar. MCIT cell tracking data coupled with updraft detections and NOAA hail reports that we used for training the RF model are accessible at https://doi.org/10.5281/zenodo.10401247. We aim to make the code used for tracking storms and assigning updraft detections publicly available in the future.

APPENDIX A

Tables of All Events and Radar Variables

First, Table A1 lists all events with target time periods and the NEXRAD sites in use.

Second, Table A2 details all radar variables tested in this study.

TABLE A1. Full event list of hail days used in this study.

Event ID	Radar ID	Radar location	Time
KICT20130723	KICT	Wichita, Kansas	2200–0400 UTC 23 Jul 2013
KOAX20140603	KOAX	Omaha, Nebraska	1800–0300 UTC 3 Jun 2014
KSJT20140612	KSJT	San Angelo, Texas	2000–0300 UTC 12 Jun 2014
KSGF20150324	KSGF	Springfield, Missouri	2100–0300 UTC 24 Mar 2015
KINX20150325	KINX	Tulsa, Oklahoma	1900–0300 UTC 25 Mar 2015
KLZK20150419	KLZK	Little Rock, Arkansas	2100–0300 UTC 19 Apr 2015
KDYX20150426	KDYX	Dyess AFB, Texas	0000–0300 UTC 27 Apr 2015
KUDX20150619	KUDX	Rapid City, South Dakota	1900–0600 UTC 19 Jun 2015
KDMX20150624	KDMX	Des Moines, Iowa	2200–0500 UTC 24 Jun 2015
KARX20150713	KARX	La Crosse, Wisconsin	1900–0300 UTC 13 Jul 2015
KABR20150717	KABR	Aberdeen, South Dakota	2000–0300 UTC 17 Jul 2015
KTWX20150910	KTWX	Topeka, Kansas	2200–0400 UTC 10 Sep 2015
KFWS20160324	KFWS	Fort Worth, Texas	0000–0600 UTC 24 Mar 2016
KFWS20160411	KFWS	Fort Worth, Texas	2000–0200 UTC 11 Apr 2016
KEWX20160413	KEWX	Austin, Texas	0100–0600 UTC 13 Apr 2016
KEWX20160425	KEWX	Austin, Texas	2100–0300 UTC 25 Apr 2016
KTWX20160426	KTWX	Topeka, Kansas	1900–0000 UTC 26 Apr 2016
KEAX20160426	KEAX	Kansas City, Missouri	1100–1600 UTC 26 Apr 2016
KVNX20160509	KVNX	Vance AFB, Oklahoma	1800–0300 UTC 9 May 2016
KAMA20160516	KAMA	Amarillo, Texas	2300–0600 UTC 16 May 2016
KICT20160526	KICT	Wichita, Kansas	1700–0200 UTC 26 May 2016
KEAX20160526	KEAX	Kansas City, Missouri	1300–2100 UTC 26 May 2016
KICT20160615	KICT	Wichita, Kansas	2100–0500 UTC 15 Jun 2016
KGLD20160707	KGLD	Goodland, Kansas	2300–0600 UTC 7 Jul 2016
KPUX20160708	KPUX	Pueblo, Colorado	0100–0800 UTC 8 Jul 2016
KPUX20160729	KPUX	Pueblo, Colorado	0200–0800 UTC 29 Jul 2016
KSHV20170121	KSHV	Shreveport, Louisiana	2100–0400 UTC 21 Jan 2017
KDVN20170228	KDVN	Davenport, Louisiana	2000–0400 UTC 28 Feb 2017
KEAX20170301	KEAX	Kansas City, Missouri	0100–0700 UTC 1 Mar 2017
KEAX20170306	KEAX	Kansas City, Missouri	2300–0600 UTC 6 Mar 2017
KSGF20170309	KSGF	Springfield, Missouri	2000–0300 UTC 9 Mar 2017
KFWS20170326	KFWS	Fort Worth, Texas	2100–0500 UTC 26 Mar 2017
KINX20170404	KINX	Tulsa, Oklahoma	2000–0200 UTC 4 Apr 2017
KOAX20170415	KOAX	Omaha, Nebraska	2000–0400 UTC 15 Apr 2017
KAMA20170416	KAMA	Amarillo, Texas	2100–0500 UTC 16 Apr 2017
KAMA20170515	KAMA	Amarillo, Texas	2200–0300 UTC 15 May 2017
KDDC20170516	KDDC	Dodge City, Kansas	1900–0300 UTC 16 May 2017
KEAX20170518	KEAX	Kansas City, Missouri	2300–0500 UTC 18 May 2017
KTWX20170518	KTWX	Topeka, Kansas	2000–0300 UTC 18 May 2017
KINX20170527	KINX	Tulsa, Oklahoma	2200–0600 UTC 27 May 2017
KMPX20170611	KMPX	Minneapolis, Minnesota	1200–1600 UTC 11 Jun 2017
KCYS20170612	KCYS	Cheyenne, Wyoming	1900–0200 UTC 12 Jun 2017
KARX20170616	KARX	La Crosse, Wisconsin	2100–0400 UTC 16 Jun 2017
KDDC20170617	KDDC	Dodge City, Kansas	2200–0500 UTC 17 Jun 2017
KOAX20170629	KOAX	Omaha, Nebraska	2200–0600 UTC 29 Jun 2017
KMVX20170704	KMVX	Grand Forks, Dakota	2300–0600 UTC 4 Jul 2017
KMPX20170710	KMPX	Minneapolis, Minnesota	0000–0500 UTC 10 Jul 2017
KUDX20170814	KUDX	Rapid City, South Dakota	2000–0400 UTC 14 Aug 2017
KDDC20171006	KDDC	Dodge City, Kansas	1900–0200 UTC 6 Oct 2017
KGRK20180318	KGRK	Granger, Texas	2000–0000 UTC 18 Mar 2018
KFWS20180406	KFWS	Fort Worth, Texas	2100–0300 UTC 6 Apr 2018
KFWS20180413	KFWS	Fort Worth, Texas	1900–0000 UTC 13 Apr 2018
KEAX20180413	KEAX	Kansas City, Missouri	2000–0300 UTC 13 Apr 2018

TABLE A1. (*Continued*)

Event ID	Radar ID	Radar location	Time
KFWS20180421	KFWS	Fort Worth, Texas	2000–0300 UTC 21 Apr 2018
KICT20180514	KICT	Wichita, Kansas	1900–0300 UTC 14 May 2018
KTWX20180514	KTWX	Topeka, Kansas	2100–0300 UTC 15 May 2018
KDDC20180529	KDDC	Dodge City, Kansas	2000–0200 UTC 29 May 2018
KPUX20180613	KPUX	Pueblo, Colorado	0500–0900 UTC 13 Jun 2018
KFTG20180619	KFTG	Denver, Colorado	1700–0000 UTC 19 Jun 2018
KPUX20180806	KPUX	Pueblo, Colorado	1900–0300 UTC 6 Aug 2018
KAMA20190322	KAMA	Amarillo, Texas	2100–0400 UTC 22 Mar 2019
KSRX20190324	KSRX	Fort Smith, Arkansas	2100–0200 UTC 24 Mar 2019
KFWS20190324	KFWS	Fort Worth, Texas	2100–0400 UTC 24 Mar 2019
KAMA20190507	KAMA	Amarillo, Texas	1700–0000 UTC 7 May 2019
KEAX20190601	KEAX	Kansas City, Missouri	1900–0000 UTC 1 Jun 2019
KDYX20190620	KDYX	Dyess AFB, Texas	0000–0300 UTC 20 Jun 2019
KFTG20190704	KFTG	Denver, Colorado	2000–0300 UTC 4 Jul 2019
KMPX20190726	KMPX	Minneapolis, Minnesota	2100–0400 UTC 26 Jul 2019
KMPX20190805	KMPX	Minneapolis, Minnesota	1600–0000 UTC 5 Aug 2019
KEAX20190927	KEAX	Kansas City, Missouri	2200–0400 UTC 27 Sep 2019
KDVN20200407	KDVN	Davenport, Iowa	2200–0600 UTC 7 Apr 2020
KTWX20200411	KTWX	Topeka, Kansas	2200–0700 UTC 11 Apr 2020
KDFX20200412	KDFX	Laughlin, Texas	0000–0300 UTC 12 Apr 2020
KSHV20200424	KSHV	Shreveport, Louisiana	2200–0500 UTC 24 Apr 2020
KFWS20200428	KFWS	Fort Worth, Texas	0600–1200 UTC 28 Apr 2020
KINX20200428	KINX	Tulsa, Oklahoma	2100–0300 UTC 28 Apr 2020
KTWX20200504	KTWX	Topeka, Kansas	1000–1600 UTC 4 May 2020
KFDR20200507	KFDR	Fredrick, Oklahoma	2200–0400 UTC 7 May 2020
KFDR20200522	KFDR	Fredrick, Oklahoma	2100–0500 UTC 22 May 2020
KEWX20200527	KEWX	Austin, Texas	2300–0500 UTC 27 May 2020
KUDX20200710	KUDX	Rapid City, South Dakota	2100–0000 UTC 10 Jul 2020
KTWX20200711	KTWX	Topeka, Kansas	1700–2200 UTC 11 Jul 2020
KUDX20200808	KUDX	Rapid City, South Dakota	2000–0200 UTC 8 Aug 2020
KSGF20210317	KSGF	Springfield, Missouri	2100–0300 UTC 17 Mar 2021
KFWS20210324	KFWS	Fort Worth, Texas	2100–0500 UTC 24 Mar 2021
KFWS20210409	KFWS	Fort Worth, Texas	2200–0400 UTC 9 Apr 2021
KGRK20210415	KGRK	Granger, Texas	1900–0000 UTC 15 Apr 2021
KFDR20210423	KFDR	Fredrick, Oklahoma	2100–0300 UTC 23 Apr 2021
KFWS20210428	KFWS	Fort Worth, Texas	2300–0400 UTC 28 Apr 2021
KEWX20210503	KEWX	Austin, Texas	2300–0500 UTC 3 May 2021
KFWS20210503	KFWS	Fort Worth, Texas	2000–0400 UTC 3 May 2021
KFTG20210613	KFTG	Denver, Colorado	2100–0300 UTC 13 Jun 2021
KMPX20210617	KMPX	Minneapolis, Minnesota	2200–0300 UTC 17 Jun 2021
KDMX20210619	KDMX	Des Moines, Iowa	0200–0700 UTC 19 Jun 2021
KLZK20220411	KLZK	Little Rock, Arkansas	2200–0600 UTC 11 Apr 2022
KSRX20220411	KSRX	Fort Smith, Arkansas	2000–0400 UTC 11 Apr 2022
KGRK20220412	KGRK	Granger, Texas	2100–0400 UTC 12 Apr 2022
KMPX20220412	KMPX	Minneapolis, Minnesota	2300–0400 UTC 12 Apr 2022
KMPX20220509	KMPX	Minneapolis, Minnesota	1200–1600 UTC 9 May 2022
KFDR20220513	KFDR	Fredrick, Oklahoma	2200–0300 UTC 13 May 2022
KSGF20220514	KSGF	Springfield, Missouri	1900–0000 UTC 14 May 2022
KMPX20220519	KMPX	Minneapolis, Minnesota	1900–0000 UTC 19 May 2022
KICT20220531	KICT	Wichita, Kansas	2000–0400 UTC 31 May 2022
KDDC20220607	KDDC	Dodge City, Kansas	0000–0400 UTC 7 Jun 2022
KUDX20220612	KUDX	Rapid City, South Dakota	2300–0300 UTC 12 Jun 2022
KTLX20230227	KTLX	Oklahoma City, Oklahoma	0000–0600 UTC 27 Feb 2023
KFWS20230316	KFWS	Fort Worth, Texas	1900–0300 UTC 16 Mar 2023
KDMX20230331	KDMX	Des Moines, Iowa	1600–2300 UTC 31 Mar 2023
KDMX20230404	KDMX	Des Moines, Iowa	2100–0400 UTC 4 Apr 2023
KSGF20230415	KSGF	Springfield, Missouri	1800–0000 UTC 15 Apr 2023
KTWX20230419	KTWX	Topeka, Kansas	2200–0600 UTC 19 Apr 2023
KOAX20230419	KOAX	Omaha, Nebraska	2100–0400 UTC 19 Apr 2023
KTLX20230419	KTLX	Oklahoma City, Oklahoma	2100–0200 UTC 19 Apr 2023
KGRK20230428	KGRK	Granger, Texas	1900–0200 UTC 28 Apr 2023

TABLE A2. All radar metrics used for the evaluation of hail (precursor) signals. All variables extracted from the detected updraft begin with HS. Each of these variables is calculated at isotherm heights of −10°, −20°, −30°, and −40°C, which we indicate by "*x*°C" Hotspot values always refer to hotspots of ZDR unless stated otherwise.

Name	Description
MCIT-HS-xC_Dist	Distance from MCIT cell core to updraft core
MCIT-MESH	Maximum expected size of hail (Witt et al. 1998); maximum value within the MCIT storm cell
MCIT-MaxLog10Vil	Maximum log (VIL) within the MCIT storm cell
MCIT-size	Total size of MCIT cell
HS-xC_AreaRef40	Area $Z_H > 40$ dBZ (*x*°C)
HS-xC_AreaRef50	Area $Z_H > 50$ dBZ (*x*°C)
HS-xC_AveAZS	Average azimuthal shear (*x*°C)
HS-xC_AveDR	Average DR (*x*°C)
HS-xC_AveHS	Average hotspot value (*x*°C)
HS-xC_AveHSLogCC	Average hotspot value of logCC; average HS (logCC) (*x*°C)
HS-xC_AveLogCC	Average logCC (*x*°C)
HS-xC_AveRef	Average Z_H (*x*°C)
HS-xC_AveRho	Average ρ_{HV} (*x*°C)
HS-xC_AveZdr	Average Z_{DR} (*x*°C)
HS-xC_CellArea	Hotspot area (*x*°C)
HS-xC_DR17count	Area of DR > −17 (*x*°C)
HS-xC_DR17percent	Area fraction of DR > −17 within the hotspot (*x*°C)
HS-xC_DRAveRef40	Average DR where $Z_H > 50$ dBZ (*x*°C)
HS-xC_DRAveRef50	Average DR where $Z_H > 50$ dBZ (*x*°C)
HS-xC_DRcount14Ref40	Area of DR > −14 where $Z_H > 40$ dBZ (*x*°C)
HS-xC_DRcount14Ref50	Area of DR > −14 where $Z_H > 50$ dBZ (*x*°C)
HS-xC_DRcount17Ref40	Area of DR > −17 where $Z_H > 40$ dBZ (*x*°C)
HS-xC_DRcount17Ref50	Area of DR > −17 where $Z_H > 50$ dBZ (*x*°C)
HS-xC_DRcount20Ref40	Area of DR > −20 where $Z_H > 40$ dBZ (*x*°C)
HS-xC_DRcount20Ref50	Area of DR > −20 where $Z_H > 50$ dBZ (*x*°C)
HS-xC_DiffAZS	Differential azimuthal shear; minimum azimuthal shear subtracted from maximum azimuthal shear (*x*°C)
HS-xC_DilArea	Dilated hotspot area; hotspot area dilated by 3 pixels in each direction (3 km)
HS-xC_LogCCRef40	Average logCC where $Z_H > 40$ dBZ (*x*°C)
HS-xC_LogCCRef50	Average logCC where $Z_H > 50$ dBZ (*x*°C)
HS-xC_MaxAZS	Maximum azimuthal shear (*x*°C)
HS-xC_MaxDR	Maximum DR (*x*°C)
HS-xC_MaxHS	Maximum hotspot value (*x*°C)
HS-xC_MaxHSLogCC	Maximum hotspot value of logCC; maximum HS (logCC) (*x*°C)
HS-xC_MaxLogCC	Maximum logCC (*x*°C)
HS-xC_MaxRef	Maximum Z_H (*x*°C)
HS-xC_MaxRho	Maximum ρ_{HV} (*x*°C)
HS-xC_MaxZdr	Maximum Z_{DR} (*x*°C)
HS-xC_MinAZS	Minimum azimuthal shear (*x*°C)
HS-xC_MinDR	Minimum DR (*x*°C)
HS-xC_MinHS	Minimum hotspot value (*x*°C)
HS-xC_MinHSLogCC	Minimum hotspot value of logCC; minimum HS (logCC) (*x*°C)
HS-xC_MinLogCC	Minimum logCC (*x*°C)
HS-xC_MinRef	Minimum Z_H (*x*°C)
HS-xC_MinRho	Minimum ρ_{HV} (*x*°C)
HS-xC_MinZdr	Minimum Z_{DR} (*x*°C)
HS-xC_PercentNegHSLogCC	Area fraction of HS (logCC) < 0 (*x*°C)
HS-xC_STDAZS	Standard deviation of azimuthal shear (*x*°C)
HS-xC_STDDR	Standard deviation of DR (*x*°C)
HS-xC_STDHS	Standard deviation of hotspot value (*x*°C)
HS-xC_STDHSLogCC	Standard deviation of HS (logCC) (*x*°C)
HS-xC_STDLogCC	Standard deviation of logCC (*x*°C)
HS-xC_STDRef	Standard deviation of Z_H (*x*°C)
HS-xC_STDRho	Standard deviation of ρ_{HV} (*x*°C)
HS-xC_STDZdr	Standard deviation of Z_{DR} (*x*°C)
HS-xC_SigColdHS	"Cold spots of Z_{DR}"; area of HS (Z_{DR}) < −0.2 (*x*°C)

TABLE A2. (*Continued*)

Name	Description
HS-xC_SigColdHSLogCC	Cold spots of logCC; area of HS (logCC) < -0.2 (x°C)
HS-xC_SigColdHSRef40	Area of HS(Z_{DR}) < -0.2 where $Z_H > 40$ dBZ (x°C)
HS-xC_SigColdHSRef50	Area of HS(Z_{DR}) < -0.2 where $Z_H > 50$ dBZ (x°C)
HS-xC_SigWarmHS	Area of HS(Z_{DR}) > 0.2 (x°C)
HS-xC_SigWarmHSLogCC	Area of HS(logCC) > 0.2 (x°C)
HS-xC_TH0ColdHSRef40	Area of HS(Z_{DR}) < 0 where $Z_H > 40$ dBZ (x°C)
HS-xC_TH0ColdHSRef50	Area of HS(Z_{DR}) < 0 where $Z_H > 50$ dBZ (x°C)
HS-xC_maxDRRef40	Maximum DR where ZH > 40 dBZ (x°C)
HS-xC_maxDRRef50	Maximum DR where ZH > 50 dBZ (x°C)

APPENDIX B

Algorithm for Assigning Updraft Detections to MCIT Cells

1) Sort MCIT cells and updraft objects by size in descending order.
2) Beginning with the largest MCIT cell, go through the list of sorted updraft objects and assign the first updraft that overlaps the MCIT cell and is within a cell-size-dependent distance of the MCIT center location of the previous time step. If no previous time step is available, take the current location. The rationale for taking the previous cell location is that updrafts are usually found trailing the precipitation maximum and the maximum VIL of the previous time step is therefore closer to the current updraft location, given the data are available at a common 5-min scan interval. The maximum allowed distance is given by max{10 (km), [cell area (km^2)]/}, the latter being the radius of a circle with the same area as the storm cell. Once an updraft is assigned, it is removed from the list of available updraft objects.
3) For the same MCIT cell, continue through the list of updraft objects. If there is another overlap, add the updraft object to the cell if the center of the updraft is within 12 km of the first assigned updraft. This is to avoid adding updrafts on different sides of the cell. Again, if an updraft is assigned, it is removed from the list of available updrafts.
4) Continue with the next-largest MCIT cell and repeat steps 2 and 3.
5) After the first loop going through all MCIT cells, there is a second, final pass looping through MCIT cells that have no attached updraft yet. We now assign any remaining updraft object to an MCIT cell which overlaps it.
6) Updrafts that have not been assigned to MCIT cells are discarded.

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), https://ejssm.com/ojs/index.php/site/article/view/60.

——, ——, Y. Kaheil, A. H. Sobel, C. Lepore, S. Nong, and A. Muehlbauer, 2017: An extreme value model for U.S. hail size. *Mon. Wea. Rev.*, **145**, 4501–4519, https://doi.org/10.1175/MWR-D-17-0119.1.

——, I. M. Giammanco, M. R. Kumjian, H. Jurgen Punge, Q. Zhang, P. Groenemeijer, M. Kunz, and K. Ortega, 2019: Understanding hail in the Earth system. *Rev. Geophys.*, **58**, e2019RG000665, https://doi.org/10.1029/2019RG000665.

Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The rapid refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, https://doi.org/10.1175/MWR-D-15-0242.1.

Blair, S. F., D. R. Deroche, J. M. Boustead, J. W. Leighton, B. L. Barjenbruch, and W. P. Gargan, 2011: Radar-based assessment of the detectability of giant hail. *Electron. J. Severe Storms Meteor.*, **6** (7), https://ejssm.com/ojs/index.php/site/article/view/34.

——, and Coauthors, 2017: High-resolution hail observations: Implications for NWS warning operations. *Wea. Forecasting*, **32**, 1101–1119, https://doi.org/10.1175/WAF-D-16-0203.1.

Breiman, L., 2001: Random forests. *Mach. Learn.*, **45**, 5–32, https://doi.org/10.1023/A:1010933404324.

Carlin, J. T., J. Gao, J. C. Snyder, and A. V. Ryzhkov, 2017: Assimilation of Z_{DR} columns for improving the spinup and forecast of convective storms in storm-scale models: Proof-of-concept experiments. *Mon. Wea. Rev.*, **145**, 5033–5057, https://doi.org/10.1175/MWR-D-17-0103.1.

Chase, R. J., A. McGovern, C. R. Homeyer, P. J. Marinescu, and C. K. Potvin, 2024: Machine learning estimation of maximum vertical velocity from radar. *Artif. Intell. Earth Syst.*, **3**, 230095, https://doi.org/10.1175/AIES-D-23-0095.1.

Cho, J. Y. N., J. M. Kurdzo, B. J. Bennett, M. E. Weber, J. W. Dellicarpini, A. Loconto, and H. Frank, 2022: Impact of WSR-88D intra-volume low-level scans on severe weather warning performance. *Wea. Forecasting*, **37**, 1169–1189, https://doi.org/10.1175/WAF-D-21-0152.1.

Cintineo, J. L., T. M. Smith, V. Lakshmanan, H. E. Brooks, and K. L. Ortega, 2012: An objective high-resolution hail climatology of the contiguous United States. *Wea. Forecasting*, **27**, 1235–1248, https://doi.org/10.1175/WAF-D-11-00151.1.

Czernecki, B., M. Taszarek, M. Marosz, M. Półrolniczak, L. Kolendowicz, A. Wyszogrodzki, and J. Szturc, 2019: Application of machine learning to large hail prediction—The importance of radar reflectivity, lightning occurrence and convective parameters derived from ERA5. *Atmos. Res.*, **227**, 249–262, https://doi.org/10.1016/j.atmosres.2019.05.010.

Edwards, R., and R. L. Thompson, 1998: Nationwide comparisons of hail size with WSR-88D vertically integrated liquid water and derived thermodynamic sounding data. *Wea. Forecasting*,

13, 277–285, https://doi.org/10.1175/1520-0434(1998)013<0277:NCOHSW>2.0.CO;2.

Farnell, C., T. Rigo, and N. Pineda, 2018: Exploring radar and lightning variables associated with the lightning jump. Can we predict the size of the hail? *Atmos. Res.*, **202**, 175–186, https://doi.org/10.1016/j.atmosres.2017.11.019.

French, M. M., and D. M. Kingfield, 2021: Tornado formation and intensity prediction using polarimetric radar estimates of updraft area. *Wea. Forecasting.*, **36**, 2211–2231, https://doi.org/10.1175/WAF-D-21-0087.1.

Gagne, D. J., II, A. McGovern, and M. Xue, 2014: Machine learning enhancement of storm-scale ensemble probabilistic quantitative precipitation forecasts. *Wea. Forecasting*, **29**, 1024–1043, https://doi.org/10.1175/WAF-D-13-00108.1.

Greene, D. R., and R. A. Clark, 1972: Vertically integrated liquid water—A new analysis tool. *Mon. Wea. Rev.*, **100**, 548–552, https://doi.org/10.1175/1520-0493(1972)100<0548:VILWNA>2.3.CO;2.

Gregorutti, B., B. Michel, and P. Saint-Pierre, 2017: Correlation and variable importance in random forests. *Stat. Comput.*, **27**, 659–678, https://doi.org/10.1007/s11222-016-9646-1.

Gunturi, P., and M. K. Tippett, 2017: Managing severe thunderstorm risk: Impact of ENSO on U.S. tornado and hail frequencies. *WillisRe Tech. Rep.*, **5** pp., https://www.columbia.edu/~mkt14/files/WillisRe_Impact_of_ENSO_on_US_Tornado_and_Hail_frequencies_Final.pdf.

Gutierrez, R. E., and M. R. Kumjian, 2021: Environmental and radar characteristics of gargantuan hail-producing storms. *Mon. Wea. Rev.*, **149**, 2523–2538, https://doi.org/10.1175/MWR-D-20-0298.1.

Heinselman, P. L., and A. V. Ryzhkov, 2006: Validation of polarimetric hail detection. *Wea. Forecasting*, **21**, 839–850, https://doi.org/10.1175/WAF956.1.

Heymsfield, A. J., I. M. Giammanco, and R. Wright, 2014: Terminal velocities and kinetic energies of natural hailstones. *Geophys. Res. Lett.*, **41**, 8666–8672, https://doi.org/10.1002/2014GL062324.

——, M. Szakáll, A. Jost, I. Giammanco, and R. Wright, 2018: A comprehensive observational study of Graupel and hail terminal velocity, mass flux, and kinetic energy. *J. Atmos. Sci.*, **75**, 3861–3885, https://doi.org/10.1175/JAS-D-18-0035.1.

Homeyer, C. R., E. M. Murillo, and M. R. Kumjian, 2023: Relationships between 10 years of radar-observed supercell characteristics and hail potential. *Mon. Wea. Rev.*, **151**, 2609–2632, https://doi.org/10.1175/MWR-D-23-0019.1.

Hu, J., and Coauthors, 2019: Tracking and characterization of convective cells through their maturation into stratiform storm elements using polarimetric radar and lightning detection. *Atmos. Res.*, **226**, 192–207, https://doi.org/10.1016/j.atmosres.2019.04.015.

Ilotoviz, E., A. Khain, A. V. Ryzhkov, and J. C. Snyder, 2018: Relationship between aerosols, hail microphysics, and Z_{DR} columns. *J. Atmos. Sci.*, **75**, 1755–1781, https://doi.org/10.1175/JAS-D-17-0127.1.

Jergensen, G. E., A. McGovern, R. Lagerquist, and T. Smith, 2020: Classifying convective storms using machine learning. *Wea. Forecasting*, **35**, 537–559, https://doi.org/10.1175/WAF-D-19-0170.1.

Jiang, Z., M. R. Kumjian, R. S. Schrom, I. Giammanco, T. Brown-Giammanco, H. Estes, R. Maiden, and A. J. Heymsfield, 2019: Comparisons of electromagnetic scattering properties of real hailstones and spheroids. *J. Appl. Meteor. Climatol.*, **58**, 93–112, https://doi.org/10.1175/JAMC-D-17-0344.1.

Kaltenboeck, R., and A. Ryzhkov, 2013: Comparison of polarimetric signatures of hail at S and C bands for different hail sizes. *Atmos. Res.*, **123**, 323–336, https://doi.org/10.1016/j.atmosres.2012.05.013.

Kamangir, H., W. Collins, P. Tissot, and S. A. King, 2020: A deep-learning model to predict thunderstorms within 400 km^2 South Texas domains. *Meteor. Appl.*, **27**, e1905, https://doi.org/10.1002/met.1905.

Keat, W. J., C. D. Westbrook, and A. J. Illingworth, 2016: High-precision measurements of the copolar correlation coefficient: Non-Gaussian errors and retrieval of the dispersion parameter μ in rainfall. *J. Appl. Meteor. Climatol.*, **55**, 1615–1632, https://doi.org/10.1175/JAMC-D-15-0272.1.

Kilambi, A., F. Fabry, and V. Meunier, 2018: A simple and effective method for separating meteorological from nonmeteorological targets using dual-polarization data. *J. Atmos. Oceanic Technol.*, **35**, 1415–1424, https://doi.org/10.1175/JTECH-D-17-0175.1.

Kingfield, D. M., and M. M. French, 2022: The influence of WSR-88D intra-volume scanning strategies on thunderstorm observations and warnings in the dual-polarization radar era: 2011–20. *Wea. Forecasting*, **37**, 283–301, https://doi.org/10.1175/WAF-D-21-0127.1.

Krause, J., and V. Klaus, 2024: Identifying Z_{DR} columns in radar data with the hotspot technique. *Wea. Forecasting*, **39**, 581–595, https://doi.org/10.1175/WAF-D-23-0146.1.

Kumjian, M. R., and A. V. Ryzhkov, 2008: Polarimetric signatures in supercell thunderstorms. *J. Appl. Meteor. Climatol.*, **47**, 1940–1961, https://doi.org/10.1175/2007JAMC1874.1.

——, and K. Lombardo, 2020: A hail growth trajectory model for exploring the environmental controls on hail size: Model physics and idealized tests. *J. Atmos. Sci.*, **77**, 2765–2791, https://doi.org/10.1175/JAS-D-20-0016.1.

——, A. P. Khain, N. Benmoshe, E. Ilotoviz, A. V. Ryzhkov, and V. T. J. Phillips, 2014: The anatomy and physics of Z_{DR} columns: Investigating a polarimetric radar signature with a spectral bin microphysical model. *J. Appl. Meteor. Climatol.*, **53**, 1820–1843, https://doi.org/10.1175/JAMC-D-13-0354.1.

——, Z. J. Lebo, and A. M. Ward, 2019: Storms producing large accumulations of small hail. *J. Appl. Meteor. Climatol.*, **58**, 341–364, https://doi.org/10.1175/JAMC-D-18-0073.1.

——, K. Lombardo, and S. Loeffler, 2021: The evolution of hail production in simulated supercell storms. *J. Atmos. Sci.*, **78**, 3417–3440, https://doi.org/10.1175/JAS-D-21-0034.1.

Kuster, C. M., T. J. Schuur, T. T. Lindley, and J. C. Snyder, 2020: Using Z_{DR} columns in forecaster conceptual models and warning decision-making. *Wea. Forecasting*, **35**, 2507–2522, https://doi.org/10.1175/WAF-D-20-0083.1.

Loeffler, S. D., M. R. Kumjian, M. Jurewicz, and M. M. French, 2020: Differentiating between tornadic and nontornadic supercells using polarimetric radar signatures of hydrometeor size sorting. *Geophys. Res. Lett.*, **47**, e2020GL088242, https://doi.org/10.1029/2020GL088242.

Mahalik, M. C., B. R. Smith, K. L. Elmore, D. M. Kingfield, K. L. Ortega, and T. M. Smith, 2019: Estimates of gradients in radar moments using a linear least squares derivative technique. *Wea. Forecasting*, **34**, 415–434, https://doi.org/10.1175/WAF-D-18-0095.1.

Manzato, A., 2007: A note on the maximum Peirce skill score. *Wea. Forecasting*, **22**, 1148–1154, https://doi.org/10.1175/WAF1041.1.

McGovern, A., R. Lagerquist, D. John Gagne II, G. E. Jergensen, K. L. Elmore, C. R. Homeyer, and T. Smith, 2019: Making the black box more transparent: Understanding the physical implications of machine learning. *Bull. Amer. Meteor. Soc.*, **100**, 2175–2199, https://doi.org/10.1175/BAMS-D-18-0195.1.

Medina, B. L., L. D. Carey, C. G. Amiot, R. M. Mecikalski, W. P. Roeder, T. M. McNamara, and R. J. Blakeslee, 2019: A random forest method to forecast downbursts based on dual-polarization radar signatures. *Remote Sens.*, **11**, 826, https://doi.org/10.3390/rs11070826.

Murillo, E. M., and C. R. Homeyer, 2019: Severe hail fall and hailstorm detection using remote sensing observations. *J. Appl. Meteor. Climatol.*, **58**, 947–970, https://doi.org/10.1175/JAMC-D-18-0247.1.

Nelson, S. P., 1983: The influence of storm flow structure on hail growth. *J. Atmos. Sci.*, **40**, 1965–1983, https://doi.org/10.1175/1520-0469(1983)040<1965:TIOSFS>2.0.CO;2.

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.

NOAA NWS, 2023: Storm events database. NOAA National Weather Service, accessed 3 February 2023, https://www.ncdc.noaa.gov/stormevents/.

NOAA NWS Radar Operations Center, 1991: NOAA Next Generation Radar (NEXRAD) Level II Base Data, https://doi.org/10.7289/V5W9574V.

Ortega, K. L., 2018: Evaluating multi-radar, multi-sensor products for surface hailfall diagnosis. *Electron. J. Severe Storms Meteor.*, **13** (1), https://ejssm.org/archives/2018/vol-13-1-2018/.

——, T. M. Smith, K. L. Manross, K. A. Scharfenberg, A. Witt, A. G. Kolodziej, and J. J. Gourley, 2009: The severe hazards analysis and verification experiment. *Bull. Amer. Meteor. Soc.*, **90**, 1519–1530, https://doi.org/10.1175/2009BAMS2815.1.

——, J. M. Krause, and A. V. Ryzhkov, 2016: Polarimetric radar characteristics of melting hail. Part III: Validation of the algorithm for hail size discrimination. *J. Appl. Meteor. Climatol.*, **55**, 829–848, https://doi.org/10.1175/JAMC-D-15-0203.1.

Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.*, **12**, 2825–2830.

Peirce, C. S., 1884: The numerical measure of the success of predictions. *Science*, **4**, 453–454, https://doi.org/10.1126/science.ns-4.93.453.b.

Peters, J. M., H. Morrison, C. J. Nowotarski, J. P. Mulholland, and R. L. Thompson, 2020: A formula for the maximum vertical velocity in supercell updrafts. *J. Atmos. Sci.*, **77**, 3747–3757, https://doi.org/10.1175/JAS-D-20-0103.1.

Picca, J. C., and A. Ryzhkov, 2010: Polarimetric signatures of melting hail at S and C bands: Detection and short-term forecast. *26th Conf. on Interactive Information and Processing Systems*, Atlanta, GA, Amer. Meteor. Soc., 10B.4, https://ams.confex.com/ams/90annual/techprogram/paper_161240.htm.

Rosenfeld, D., 1987: Objective method for analysis and tracking of convective cells as seen by radar. *J. Atmos. Oceanic Technol.*, **4**, 422–434, https://doi.org/10.1175/1520-0426(1987)004<0422:OMFAAT>2.0.CO;2.

Ryzhkov, A. V., 2007: The impact of beam broadening on the quality of radar polarimetric data. *J. Atmos. Oceanic Technol.*, **24**, 729–744, https://doi.org/10.1175/JTECH2003.1.

——, M. R. Kumjian, S. M. Ganson, and A. P. Khain, 2013a: Polarimetric radar characteristics of melting hail. Part I: Theoretical simulations using spectral microphysical modeling. *J. Appl. Meteor. Climatol.*, **52**, 2849–2870, https://doi.org/10.1175/JAMC-D-13-073.1.

——, ——, ——, and P. Zhang, 2013b: Polarimetric radar characteristics of melting hail. Part II: Practical implications. *J. Appl. Meteor. Climatol.*, **52**, 2871–2886, https://doi.org/10.1175/JAMC-D-13-074.1.

——, P. Zhang, Q. Cao, S. Matrosov, V. Melnikov, and M. Knight, 2014: Measurements of circular depolarization ratio with the radar with simultaneous transmission/reception. *Eighth European Conf. on Radar in Meteorology and Hydrology*, Garmisch-Partenkirchen, Germany, ERAD, 11.3.

——, and Coauthors, 2017: Estimation of depolarization ratio using weather radars with simultaneous transmission/reception. *J. Appl. Meteor. Climatol.*, **56**, 1797–1816, https://doi.org/10.1175/JAMC-D-16-0098.1.

Snyder, J. C., A. V. Ryzhkov, M. R. Kumjian, A. P. Khain, and J. Picca, 2015: A Z_{DR} column detection algorithm to examine convective storm updrafts. *Wea. Forecasting*, **30**, 1819–1844, https://doi.org/10.1175/WAF-D-15-0068.1.

——, H. B. Bluestein, D. T. Dawson II, and Y. Jung, 2017: Simulations of polarimetric, X-band radar signatures in supercells. Part II: Z_{DR} columns and rings and K_{DP} columns. *J. Appl. Meteor. Climatol.*, **56**, 2001–2026, https://doi.org/10.1175/JAMC-D-16-0139.1.

Taszarek, M., H. E. Brooks, and B. Czernecki, 2017: Sounding-derived parameters associated with convective hazards in Europe. *Mon. Wea. Rev.*, **145**, 1511–1528, https://doi.org/10.1175/MWR-D-16-0384.1.

Ward, P. J., and Coauthors, 2020: Review article: Natural hazard risk assessments at the global scale. *Nat. Hazards Earth Syst. Sci.*, **20**, 1069–1096, https://doi.org/10.5194/nhess-20-1069-2020.

Witt, A., and S. P. Nelson, 1991: The use of single-Doppler radar for estimating maximum hailstone size. *J. Appl. Meteor.*, **30**, 425–431, https://doi.org/10.1175/1520-0450(1991)030<0425:TUOSDR>2.0.CO;2.

——, M. D. Eilts, G. J. Stumpf, J. T. Johnson, E. D. W. 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.

——, D. W. Burgess, A. Seimon, J. T. Allen, J. C. Snyder, and H. B. Bluestein, 2018: Rapid-scan radar observations of an Oklahoma tornadic hailstorm producing giant hail. *Wea. Forecasting*, **33**, 1263–1282, https://doi.org/10.1175/WAF-D-18-0003.1.

——, J. Snyder, and J. M. Krause, 2022: Radar observations of hailstorms producing extremely large hail. *Extended abstracts of the 30th Conf. on Severe Local Storms*, Santa Fe, NM, Amer. Meteor. Soc., 91, https://ams.confex.com/ams/40RADAR/meetingapp.cgi/Paper/426083.

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