

Severe Hail Fall and Hailstorm Detection Using Remote Sensing Observations

ELISA M. MURILLO AND CAMERON R. HOMEYER

School of Meteorology, University of Oklahoma, Norman, Oklahoma

(Manuscript received 18 September 2018, in final form 14 February 2019)

ABSTRACT

Severe hail days account for the vast majority of severe weather-induced property losses in the United States each year. In the United States, real-time detection of severe storms is largely conducted using ground-based radar observations, mostly using the operational Next Generation Weather Radar network (NEXRAD), which provides three-dimensional information on the physics and dynamics of storms at ~5-min time intervals. Recent NEXRAD upgrades to higher resolution and to dual-polarization capabilities have provided improved hydrometeor discrimination in real time. New geostationary satellite platforms have also led to significant changes in observing capabilities over the United States beginning in 2016, with spatio-temporal resolution that is comparable to that of NEXRAD. Given these recent improvements, a thorough assessment of their ability to identify hailstorms and hail occurrence and to discriminate between hail sizes is needed. This study provides a comprehensive comparative analysis of existing observational radar and satellite products from more than 10 000 storms objectively identified via radar echo-top tracking and nearly 6000 hail reports during 30 recent severe weather days (2013–present). It is found that radar observations provide the most skillful discrimination between severe and nonsevere hailstorms and identification of individual hail occurrence. Single-polarization and dual-polarization radar observations perform similarly at these tasks, with the greatest skill found from combining both single- and dual-polarization metrics. In addition, revisions to the “maximum expected size of hail” (MESH) metric are proposed and are shown to improve spatiotemporal comparisons between reported hail sizes and radar-based estimates for the cases studied.

1. Introduction

Hail damage produces the greatest annual property loss—approximately 60% of the total caused by severe winds, tornadoes, and hail (i.e., all severe weather) combined (Gunturi and Tippett 2017). Between 2013 and 2017 alone, hail damage accounted for over \$9 billion of property and crop damage, which is ~56% of the total severe weather damage (NOAA Severe Weather Database). As a result of its economic impact, hail occurrence has been studied across the globe since the mid-1970s [e.g., see early work by Mather et al. (1976) and Waldvogel et al. (1979)].

Numerous objective methods to identify hail occurrence and size have been developed and documented in the literature, some of which are in use operationally and predate the Next Generation Weather Radar (NEXRAD) network in the United States (e.g., Cook 1958; Petrocchi 1982; Cheng et al. 1985). Radar-based methods generally fall into two groups: those based on

single-polarization (SP) radar observations and those based on dual-polarization (DP) radar observations. Notable SP radar methods that have been shown to perform well are the vertically integrated liquid water (VIL) density (Amburn and Wolf 1997) and the maximum expected size of hail (MESH; Witt et al. 1998a), both of which are based on vertical integration of radar reflectivity at horizontal polarization Z_H . These and other similar Z_H -based methods have been cross evaluated and/or verified against hail reports in multiple studies during the past two decades (e.g., Edwards and Thompson 1998; Holleman et al. 2000; Marzban and Witt 2001; Ortega et al. 2005; San Ambrosio et al. 2007; Donavon and Jungbluth 2007; López and Sanchez 2009; Liu and Heckman 2010; Saltikoff et al. 2010; Cintineo et al. 2012; Skripniková and Řezáčová 2014; Nisi et al. 2016; Lukach et al. 2017; Capozzi et al. 2018; Wang et al. 2018; Ortega 2018).

While many studies have found SP radar parameters to be useful for (and skillful at) identifying severe hail-producing storms [those that produce ≥ 1 -in. hail (1 in. = 2.54 cm) in the largest dimension], methods used for

Corresponding author: Elisa M. Murillo, murillem@ou.edu

DOI: 10.1175/JAMC-D-18-0247.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Unauthenticated | Downloaded 12/29/21 03:24 PM UTC

performance evaluation and the sample sizes used for verification have varied considerably. Few studies exist that evaluate these parameters for large sample sizes (>1000 severe hail reports), broad regions, or long time periods. Additional limitations exist when considering the challenge of hail size discrimination, for which many SP parameters have shown limited utility. For example, [Edwards and Thompson \(1998\)](#) found that VIL density was not useful for operational hail size discrimination, and [Picca and Ryzhkov \(2012\)](#) found that MESH overestimated hail size. Providing reliable estimates of hail size based on radar observations continues to be a significant challenge.

DP radar observations have provided unique information for identifying hail occurrence because of their inclusion of measurements that represent hydrometeor shape, such as differential radar reflectivity Z_{DR} . For S-band observations, large and approximately spherical hail is typically characterized by $Z_H \geq 45$ dBZ and $Z_{DR} \approx 0$ dB, whereas large raindrops are often characterized by similar Z_H and $Z_{DR} \geq 2$ dB (e.g., [Heinselman and Ryzhkov 2006](#)). There have been a number of approaches used to leverage these scattering characteristics for hail identification. One approach has led to the creation of the so-called hail differential reflectivity (H_{DR} ; [Aydin et al. 1986](#); [Depue et al. 2007](#)), which uses DP observations to identify deviations from the expected relationship between Z_H and Z_{DR} for rain. Another approach has been to use fuzzy logic to determine the most likely hydrometeor type based on its scattering characteristics using multiple DP radar variables (e.g., [Vivekanandan et al. 1999](#); [Straka et al. 2000](#); [Heinselman and Ryzhkov 2006](#); [Elmore 2011](#); [Mahale et al. 2014](#)). This technique is commonly referred to as a hydrometeor classification algorithm (HCA). [Park et al. \(2009\)](#) outline the operational HCA used in the NEXRAD network, which features a rain–hail mixture classification as its largest size hydrometeor category. A recent improvement to this operational HCA has been made to distinguish between three hail size categories, which was shown to outperform the current operational SP hail detection algorithm (referred to as the hail size discrimination algorithm; [Ryzhkov et al. 2013](#); [Ortega et al. 2016](#)).

While radar observations have been ubiquitous in objective hail identification studies and have been leveraged for nowcasting purposes, satellite observations have received relatively little attention. This limited use is largely due to geostationary satellite imagery being traditionally available at coarse spatial and temporal scales relative to radar observations. Despite these limitations, many satellite signatures have been identified and linked to severe weather occurrence including

rapid cloud-top cooling during storm initiation ([Cintineo et al. 2013](#)), overshooting tops (OTs; cloud tops extending above the equilibrium level of deep convection) and the “enhanced V” in infrared (IR) imagery, now understood to represent the occurrence of above-anvil cirrus plumes (AACP—clouds injected into the lower stratosphere via gravity-wave breaking in the vicinity of OTs; [McCann 1983](#); [Brunner et al. 2007](#); [Setvák et al. 2010](#); [Dworak et al. 2012](#); [Homeyer 2014](#); [Punge et al. 2014](#); [Bedka et al. 2015](#); [Homeyer et al. 2017](#)). [Bedka et al. \(2018\)](#) recently found that AACPs are strong binary satellite indicators of significant severe hail (≥ 2 in. in the largest dimension)—producing storms. Alternative approaches from satellite using passive microwave imagery have also been developed, but these data are not available from geostationary platforms (e.g., [Cecil and Blankenship 2012](#); [Ferraro et al. 2015](#); [Mroz et al. 2017](#); [Ni et al. 2017](#); [Mroz et al. 2018](#)).

The Geostationary Operational Environmental Satellite (GOES) systems have been the primary source of satellite imagery over the United States for operational applications during the past several decades. Prior to the launch of the GOES-R satellite in 2017 (now operationally known as *GOES-16*), *GOES-13*, *GOES-14*, and *GOES-15* were operational and featured visible imagery at 1-km horizontal resolution and IR imagery at 4-km horizontal resolution, with image updates every 5–15 min ([Menzel and Purdom 1994](#)). Beginning in 2012, *GOES-14* (serving as a reserve system at the time) was operated in Super Rapid Scan Operations for GOES-R (SRSOR) mode, collecting imagery at 1-min intervals in preparation for the increased temporal resolution of the GOES-R series satellite. GOES-R features the Advanced Baseline Imager, which has 0.5-km horizontal resolution for visible imagery, 2 km for IR imagery, and an operational “flex mode” for image temporal resolution of 30 s to 1 min ([Schmit et al. 2005, 2013](#)). Several recent studies have assessed the potential utility of higher-resolution GOES imagery for severe weather operations ([Cintineo et al. 2013, 2014](#); [Gravelle et al. 2016](#); [Line et al. 2016](#); [Bedka and Khlopenkov 2016](#)). While these studies documented the benefits of *GOES-14* SRSOR observations for severe storm prediction, there remains a lack of focus on severe hail detection. Even with the addition of hail-specific studies using *GOES-14* SRSOR observations, the utility/benefit of *GOES-16* (and similar systems) for enhancing severe storm identification remains mostly unknown because of its novelty.

Lightning characteristics have also been used to focus on the broad severe weather problem. In particular, lightning flash rates have been used to identify strong convective updrafts, which are often the source of severe

weather events. For example, a phenomenon referred to as a “lightning jump,” representing a rapid increase in total lightning, has been correlated with severe weather occurrence in many studies (e.g., Williams et al. 1999; Deierling and Petersen 2008; Schultz et al. 2009; Liu and Heckman 2010). Schultz et al. (2017) found lightning jumps useful for identifying storms with a greater likelihood of producing severe hail. Farnell et al. (2018) also found that lightning jumps provide lead times of ~ 55 min prior to the first severe hail occurrence in a storm and often identify storms that produce large (which they define as >0.79 -in. or 2-cm diameter) hail.

The development and evaluation of methods for objective hail detection is not only motivated by an operational need of such products, but also for high-confidence verification of numerical models and climate studies. Hail report databases are plagued by serious limitations for such applications, including biases in reported sizes, time of occurrence, and location (Kelly et al. 1985; Bardsley 1990; Witt et al. 1998b, 2018; Fraile et al. 1992; Schaefer et al. 2004; Doswell et al. 2005; Trapp et al. 2006; Ortega et al. 2009; Allen and Tippett 2015; Blair et al. 2017; Allen et al. 2017). Hail is often reported well after it falls, which commonly biases the time of hail occurrence. In addition, spatial and temporal distributions of hail reports are influenced by population, road networks, storm chasers and time of day (Allen and Tippett 2015). For example, a strong storm could produce hail in a low-population region and result in few to no reports. On the other hand, an equivalent storm in a region with a larger population or heavily trafficked road network is likely to result in more frequent and representative hail reports than the previous scenario. Even with sufficient witnesses, hazards perceived as highly dangerous tend to be reported more frequently at the expense of those perceived as less dangerous (Kelly et al. 1985; Witt et al. 2018). It is less likely for hail to be reported when a tornado is occurring simultaneously, a result of people 1) taking shelter from the tornado and therefore unable to observe hail and/or 2) reporting what they perceive as the more threatening hazard. One of the greatest limitations with this database, however, is the error associated with hail sizes. Reporters often use reference objects to estimate hail size, resulting in a clustering of reports at a few unique sizes. This produces a discretized hail size distribution that is likely not representative of the true distribution (Allen et al. 2017).

This study presents a comprehensive analysis of SP radar, DP radar, satellite, and lightning products to determine the best indicators for identifying hail occurrence and hail-producing storms. Thorough assessment

of the ability of established metrics to identify hail occurrence and size from multiple platforms has been uncommon. This study seeks to answer the following questions: 1) What are the best measurements for identifying hail occurrence? 2) What are the best measurements for distinguishing between storms that produce severe hail and those that do not? 3) Given the historically limited availability of DP radar observations, can SP radar estimates of hail size be improved for future studies? We use a statistical approach to compare the most commonly used objective hail identification methods and less common approaches, focusing mostly on quantitative metrics from both satellite and radar observations over the contiguous United States (CONUS).

2. Data and methods

The 30 severe weather days analyzed here were chosen based on 1) data availability, 2) environmental set up, 3) dominant convective mode, and 4) severe weather frequency/magnitude, resulting in a variety of convective scenarios. Seventeen of the 30 days included were chosen because of the availability of 1-min GOES satellite imagery and lightning observations. The study areas of the 30 days are clustered mainly in the central United States but extend throughout most of the east as well (Fig. 1). While the majority of the cases are dominated by severe discrete or semidiscrete convection, 7 of the 30 days are dominated by severe mesoscale convective systems. Table 1 details the locations, storm totals, and report totals for all 30 days (storm totals result from objective storm tracking as outlined in section 2f).

a. Environmental data

Numerical model analyses are required to calculate quantities based on the remotely sensed datasets, for which we employ hourly analyses from the Rapid Refresh (RAP) model operated by the National Centers for Environmental Prediction (NOAA/NCEP/ESRL 2012). Analyses are available at 13-km horizontal resolution and 50 vertical levels. Environmental parameters used for analysis include the tropopause altitude and altitudes of multiple isotherms. Based on comparisons with collocated radiosonde observations in several of the cases analyzed here (not shown), errors in these parameters are minimal and, if considered, have a negligible impact on the radar- and satellite-derived products.

b. Radar observations

The NEXRAD network consists of more than 140 S-band (10–11-cm wavelength) Weather Surveillance

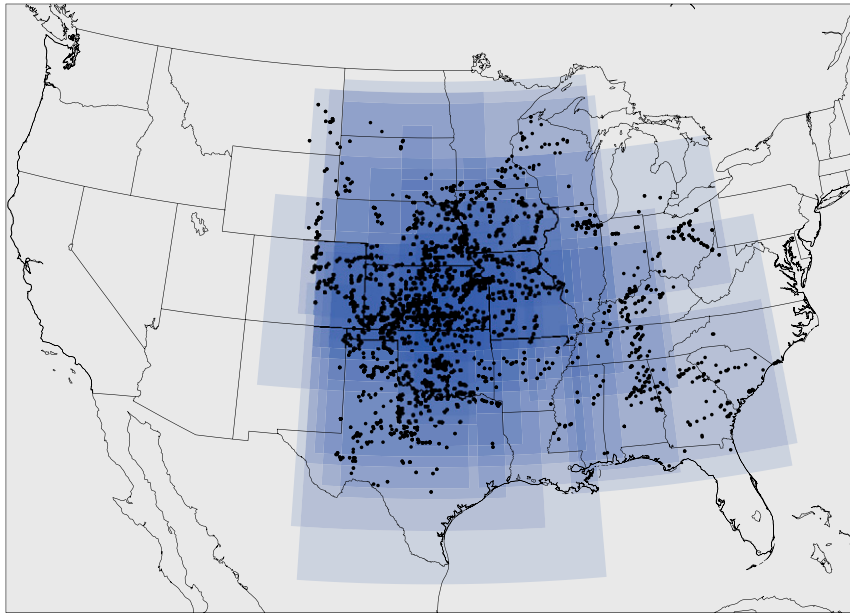


FIG. 1. Map of all hail reports analyzed in this study (black dots) and the geographical domains (blue color fill) of the corresponding 30 severe weather days listed in Table 1. Darker blues indicate greater overlapping of case domains.

Radar-1988 Doppler (WSR-88D) systems over the CONUS (Crum and Alberty 1993). WSR-88D systems observe clouds and precipitation on a polar grid in range, azimuth, and elevation relative to the location of the radar. Each WSR-88D provides data volumes at ~ 5 -min intervals when sampling convection. All NEXRAD observations used in this study were obtained at a range resolution of 250 m, an azimuthal resolution of 0.5° for the lowest 3–5 elevations and 1.0° otherwise, and typically at 14 elevations per volume. The variables measured and used in this study include Z_H , the radial Doppler velocity V_R , Z_{DR} , and the copolar correlation coefficient ρ_{HV} . All NEXRAD data were obtained from the National Centers for Environmental Information (NCEI; NOAA/NESDIS/NCEI 1991).

We merge NEXRAD data into a large-area, four-dimensional space–time composite for analysis using the methods described in Homeyer and Bowman (2017) and references therein, referred to as GridRad. Briefly, GridRad uses a space- and time-weighted binning technique to merge individual radar volumes onto a common grid. Observations at short ranges from a radar and times close to that of the analysis are given the most weight, such that the highest resolution (and polarimetric quality) observations are retained in the output volumes. These volumes feature 5-min temporal resolution, ~ 2 -km horizontal resolution, 0.5-km vertical resolution in the first 7 km MSL and 1-km resolution

above that level. Because of complexities involved in merging V_R , its azimuthal derivative (azimuthal shear) and radial derivative (radial divergence) are instead used. The weighted space and time averaging employed in the GridRad algorithm reduces uncertainty in the radar variables, allowing for prominent and spatially consistent features to be retained. While some detail at scales near and below that of the GridRad data that is present in the native NEXRAD volumes and potentially relevant to hail identification is lost as a result of the space and time averaging, such losses are expected to have a minimal impact on the analyses performed in this study because of the large sample size. An example comparison between GridRad at 1.5-km constant altitude (MSL) and the lowest elevation scan from a single radar can be seen in Fig. 2.

Prior to merging of the radar variables using the GridRad algorithm, several quality-control techniques are employed. First, observations with $\rho_{HV} < 0.5$ are not binned in order to limit the impact of non-meteorological scatterers on the resulting GridRad volumes. Second, Z_{DR} observations from individual radars undergo objective bias correction using a “natural scatterer” approach that targets snow aggregates at altitudes above (temperatures below) 0°C (e.g., Ryzhkov et al. 2005a). For more information involving these techniques and justification for their use, see the extensive review in Homeyer and Kumjian (2015).

TABLE 1. Dates, number of storms, number of hailstorms, number of hail reports, and the latitude–longitude extent of the analysis domain for all 30 cases. Boldface entries represents days for which 1-min satellite imagery (and asterisk for *GOES-14*; a dagger for *GOES-16*) and ENTLN data were available (4 June 2015 was an exception on which ENTLN data were *not* obtained).

Date	No. storms (No. severe storms)	No. Hail Storms (No. hail reports)	Analysis domain coordinates
20 May 2013	246 (47)	40 (131)	31.5°–40.0°N, 93.0°–99.0°W
31 May 2013	391 (40)	34 (128)	34.5°–40.5°N, 87.0°–99.0°W
12 Jun 2013*	555 (132)	36 (380)	38.0°–45.0°N, 80.0°–96.0°W
27 Apr 2014	223 (34)	32 (110)	34.0°–42.0°N, 91.5°–99.0°W
10 May 2014*	112 (37)	35 (363)	36.0°–43.0°N, 90.0°–99.0°W
11 May 2014*	330 (58)	48 (398)	36.0°–44.5°N, 92.0°–102.0°W
21 May 2014*	54 (9)	8 (79)	37.5°–41.0°N, 101.0°–106.0°W
16 Jun 2014	406 (36)	31 (150)	41.0°–44.0°N, 89.0°–100.0°W
17 Jun 2014	155 (20)	17 (98)	41.5°–48.0°N, 94.5°–106.0°W
18 Jun 2014	79 (6)	2 (9)	43.5°–46.5°N, 98.0°–100.0°W
13 Oct 2014	707 (610)	11 (55)	29.5°–40.5°N, 84.5°–95.5°W
6 May 2015	202 (44)	32 (115)	32.5°–41.5°N, 95.5°–100.0°W
19 May 2015*	329 (37)	22 (121)	29.0°–37.0°N, 94.0°–103.0°W
24 May 2015*	123 (11)	11 (98)	36.0°–41.0°N, 97.0°–105.0°W
25 May 2015*	669 (111)	30 (159)	25.0°–41.0°N, 89.0°–105.0°W
27 May 2015*	387 (48)	43 (392)	29.5°–41.5°N, 96.0°–104.0°W
4 Jun 2015*	290 (39)	35 (208)	34.0°–43.0°N, 93.0°–108.0°W
23 Dec 2015	137 (31)	25 (121)	33.5°–42.0°N, 84.0°–92.5°W
15 Apr 2016	160 (27)	25 (131)	34.5°–40.5°N, 99.0°–104.0°W
9 May 2016	199 (50)	49 (235)	33.0°–41.5°N, 94.0°–100.0°W
24 May 2016	150 (21)	18 (70)	35.5°–41.0°N, 97.0°–104.0°W
25 May 2016	17 (5)	5 (17)	35.5°–40.0°N, 95.0°–99.5°W
28 Mar 2017†	458 (49)	30 (249)	27.5°–39.0°N, 92.0°–105.5°W
3 Apr 2017†	499 (74)	12 (58)	29.0°–37.0°N, 78.5°–90.5°W
5 Apr 2017†	496 (119)	78 (776)	30.0°–41.0°N, 78.0°–90.0°W
16 May 2017†	947 (138)	88 (472)	30.5°–48.0°N, 87.5°–105.0°W
18 May 2017†	691 (152)	70 (364)	31.0°–40.5°N, 89.5°–104.5°W
28 Jun 2017†	604 (82)	37 (133)	36.5°–47.5°N, 87.0°–104.0°W
29 Jun 2017†	520 (68)	44 (300)	35.5°–45.0°N, 90.0°–105.0°W
19 Jul 2017†	514 (77)	8 (34)	40.5°–48.5°N, 87.5°–104.0°W
Total	10 650 (2212)	956 (5954)	—

Additional techniques are applied to the output GridRad volumes prior to analysis. Low-confidence echoes, which consist of GridRad volumes observed only at distant ranges from contributing radars (circa 180 km or more) and/or sampled infrequently by the NEXRAD network are removed (Homeyer and Bowman 2017). To reduce analysis of observations resulting from nonmeteorological scatterers retained in the GridRad volumes, echoes are removed when $Z_H < 40$ dBZ and $\rho_{HV} < 0.9$. All echoes with $Z_H > 40$ dBZ (regardless of ρ_{HV} value) are retained for analysis.

When V_R exceeds the maximum detectable value at a particular operating mode (i.e., the Nyquist velocity), it becomes aliased or “folded.” Dealiasing is completed prior to computing the V_R derivatives by implementing the open-source routine “dealias_region_based” from the Python ARM Radar Toolkit (Py-ART; Helmus and Collis 2016). Once the V_R fields are dealiased, a 3×3 median filter and a 5-gate running-mean range filter

are applied to reduce random noise. The azimuthal and radial derivatives are then computed using first-order centered differencing, where the spatial resolution of the derivatives varies with range and elevation to account for changes in resolution associated with the native polar grid of the NEXRAD observations (i.e., we compute azimuthal shear at approximately equivalent spatial resolution at all ranges from the radar by changing the width of the derivative in azimuth). All data quality measures outlined above ensure that GridRad volumes contain high-confidence observations of storms.

We focus our analysis on quantitative parameters rather than qualitative (such as the HCA) as their binary nature limits the parameter space to evaluate. The parameters evaluated provide information on physical and dynamical characteristics of storms (Table 2). For example, Z_H is proportional to the sixth moment of hydrometeor size distribution and to the first moment of hydrometeor concentration. Z_{DR} is the ratio between

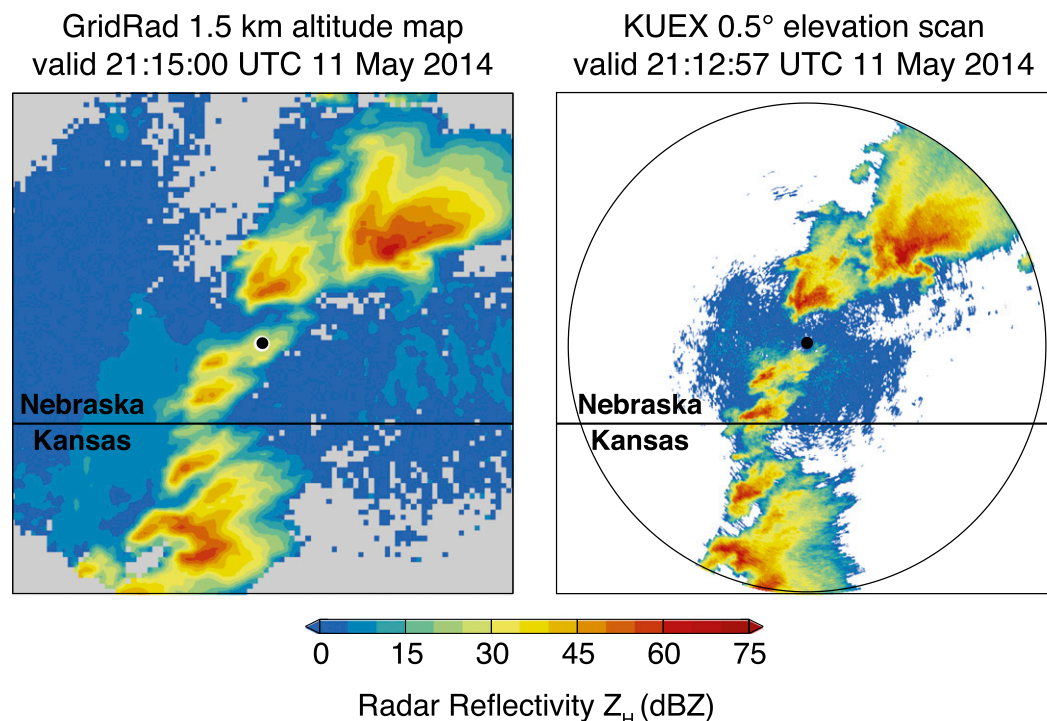


FIG. 2. Maps of (left) GridRad radar reflectivity at a constant altitude of 1.5 km MSL and (right) radar reflectivity from the 0.5° elevation scan of the Hastings, Nebraska, WSR-88D (International Civil Aviation Organization code KUEX) for a cluster of storms occurring on 11 May 2014, including a severe hail-producing supercell to the north and east of the KUEX radar. The low-level KUEX scan is within 3 min of the GridRad analysis. The black dot in each plot indicates the location of the KUEX radar.

reflectivity factors at horizontal and vertical polarizations, providing a size and/or concentration-weighted measure of particle shape. Azimuthal shear is approximately half the total vertical vorticity (assuming circular vortices), and radial divergence is approximately half of the total horizontal divergence (except at high elevations where the beam is less horizontal). More information involving the physical and dynamical interpretation of these radar observations can be found in several textbooks and review papers (e.g., [Herzogh and Jameson 1992](#); [Doviak and Zrnić 1993](#); [Hubbert et al. 1998](#); [Zrnić and Ryzhkov 1999](#); [Straka et al. 2000](#); [Bringi and Chandrasekar 2001](#); [Ryzhkov et al. 2005b](#); [Kumjian 2013a,b,c](#)).

To avoid noise within weak echo that is irrelevant for diagnosing hail characteristics, dynamical variables were only analyzed within $Z_H \geq 30$ dBZ. Computation of existing SP and DP radar parameters of hail detection was performed as follows. H_{DR} , from [Aydin et al. \(1986\)](#) and [Depue et al. \(2007\)](#), combines the polarimetric variables Z_H and Z_{DR} to identify deviations in scattering characteristics of hydrometeors from that expected of rain using a piecewise linear Z_H – Z_{DR} relationship:

$$H_{DR} = Z_H - f(Z_{DR}), \quad \text{where}$$

$$f(Z_{DR}) = \begin{cases} 27 & \text{for } (Z_{DR} \leq 0 \text{ dB}) \\ 19Z_{DR} + 27 & \text{for } (0 \leq Z_{DR} \leq 1.74 \text{ dB}) \\ 60 & \text{for } (Z_{DR} > 1.74 \text{ dB}). \end{cases} \quad (1)$$

MESH, from [Witt et al. \(1998a\)](#), is a commonly used metric to estimate hail occurrence and size. It is based on a power-law relationship to the severe

TABLE 2. All variables analyzed throughout the study, categorized by their source and representation.

Radar	Satellite	Lightning
Half rotation	<i>Dynamical</i>	—
Half divergence	mAMV CTV	
	mAMV CTD	
	<i>Physical</i>	
Z_H , Z_{DR} , ρ_{HV} , and H_{DR}	OT probability	Total flash density
MESH	OT IR brightness temperature	In-cloud flash density
VIL density	Visible texture rating	Cloud-to-ground flash density
Echo-top altitude		

hail index (SHI), which is defined by the following relationships:

$$\begin{aligned} W(Z) &= 0 & \text{for } (Z \leq Z_L) \\ &= \frac{Z - Z_L}{Z_U - Z_L} & \text{for } (Z_L < Z < Z_U) \\ &= 1 & \text{for } (Z \geq Z_U), \end{aligned} \quad (2)$$

$$\begin{aligned} W_T(H) &= 0 & \text{for } (H \leq H_0) \\ &= \frac{H - H_0}{H_{m20} - H_0} & \text{for } (H_0 < H < H_{m20}) \\ &= 1 & \text{for } (H \geq H_{m20}), \end{aligned} \quad (3)$$

$$\dot{E} = 5 \times 10^{-6} \times W(Z) \times 10^{0.084Z}, \quad \text{and} \quad (4)$$

$$\text{SHI} = 0.1 \int_{H_0}^{H_T} W_T(H) \dot{E} dH, \quad (5)$$

where Z is Z_H , Z_L is 40 dBZ, Z_U is 50 dBZ, H is the height above ground level (AGL) of the radar observation, H_0 is the collocated height AGL of the temperature $T = 0^\circ\text{C}$ level, and H_{m20} is the height AGL of the $T = -20^\circ\text{C}$ level. A simple power-law relationship between SHI and MESH was empirically developed to fit the 75th percentile of 147 hail reports, resulting in the commonly used relation:

$$\text{MESH} = 2.54(\text{SHI})^{0.5}. \quad (6)$$

VIL, adapted from Amburn and Wolf (1997), is defined as

$$\text{VIL} = 3.44 \times 10^{-6} \int_{H_{\text{bottom}}}^{H_{\text{top}}} Z^{4/7} dH, \quad (7)$$

where Z_{bottom} and Z_{top} are Z_H at the bottom and top of a layer ($\text{mm}^6 \text{m}^{-3}$) and dH is the layer depth (m). VIL density can be computed multiple ways; the most common method used by prior studies is described in Amburn and Wolf (1997) and divides VIL by the $Z_H = 18.5$ dBZ echo-top height. However, this study computes VIL density by dividing by the echo depth (the total height where $Z_H \geq 18.5$ dBZ). This approach helps to normalize all radar observations by their three-dimensional coverage. For large regions observed by NEXRAD radars, measurements are not available at lower altitudes and gaps in radar coverage may exist aloft. Thus, VIL density calculated using only the echo top can be biased low in cases where coverage is limited or reduced.

While MESH and VIL density vertically integrate Z_H to estimate the amount of large particles in a volume,

other methods use physical parameters at one level to infer updraft characteristics and hail potential. For example, echo-top heights, the maximum altitude where Z_H exceeds a specified threshold value, are correlated with the intensity of upward motion within a storm and were computed for analysis in this study. While echo-top heights for multiple thresholds were analyzed ($Z_H = 10, 20, 30$, and 40 dBZ), only results from 40-dBZ echo-top altitudes are discussed in this paper. Figure 3 shows example GridRad maps of many of the variables analyzed in this study.

c. Geostationary satellite imagery

GOES imagery (from *GOES-13*, *GOES-14*, and *GOES-16*) was retrieved from both NCEI (NOAA/NCEI 2013) and the University of Wisconsin–Madison Space Science and Engineering Center (University of Wisconsin–Madison Space Science and Engineering Center 2011). Beginning in March 2017, preliminary data from *GOES-16* were available until the system was declared operational in December 2017. While *GOES-16* can operate in 30-s mode, 1-min imagery was the finest that was analyzed in this study. For 17 of the 30 severe weather days (boldface type in Table 1), 1-min imagery was available from either *GOES-14* (denoted by an asterisk in Table 1) or *GOES-16* (denoted by a dagger in Table 1). *GOES-13* was used for the remaining days, where temporal resolution of imagery ranged from 5 to 15 min.

GOES-13 and *GOES-14* data include 1-km-per-pixel spatial resolution for imagery from the $0.65\text{-}\mu\text{m}$ visible channel and 4-km-per-pixel spatial resolution for imagery from the $10.7\text{-}\mu\text{m}$ IR channel. *GOES-16* data include 0.5-km-per-pixel spatial resolution for $0.65\text{-}\mu\text{m}$ visible channel imagery and 2-km-per-pixel spatial resolution for $10.3\text{-}\mu\text{m}$ IR channel imagery (Schmit et al. 2005, 2013). For 29–30 June 2017, $10.3\text{-}\mu\text{m}$ channel data were unavailable, so the $11.2\text{-}\mu\text{m}$ IR channel was used as a substitute. The resulting satellite parameters were not biased by this difference, because the $11.2\text{-}\mu\text{m}$ channel is only slightly more sensitive to water vapor (Schmit et al. 2005).

Using these GOES data, products based on both visible and IR imagery were derived for analysis, including visible texture and thermal properties of OTs and mesoscale atmospheric motion vectors (mAMVs). OTs exhibit more texture than the surrounding anvil and greater texture is often indicative of stronger vertical motion. Bedka and Khlopenkov (2016) developed an objective OT identification/probability and a product quantifying OT texture in visible imagery (named “visible texture rating”), which we employ for analysis in this study. Dynamical products based on mAMVs were derived through satellite feature tracking in 1-min visible

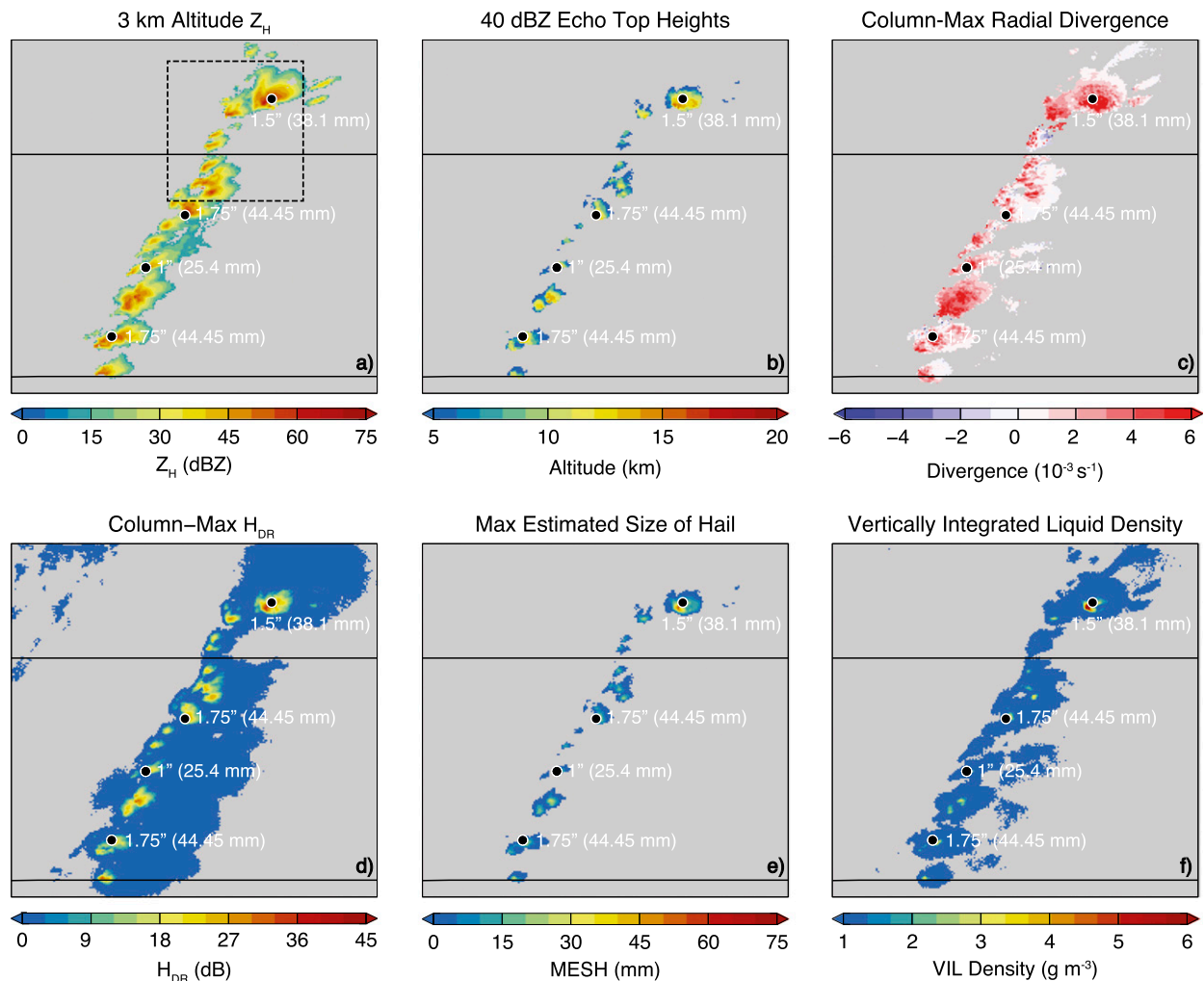


FIG. 3. GridRad maps of (a) radar reflectivity at an altitude of 3 km MSL, (b) 40-dBZ echo-top heights, (c) column-maximum radial divergence, (d) column-maximum hail differential reflectivity, (e) maximum expected size of hail, and (f) vertically integrated liquid density, valid at 2115 UTC 11 May 2014. Black circles in each panel indicate valid hail reports at this time, with corresponding hail sizes noted in white text. The black box in (a) indicates the subregion that is pictured in Fig. 2.

and IR imagery via the algorithms outlined in section 2 of Bedka and Mecikalski (2005), section 2a of Bedka et al. (2009), and section 3c of Apke et al. (2018). Cloud-top vorticity (CTV) and cloud-top divergence (CTD) based on the mAMVs are retained for analysis when 1-min-temporal-resolution imagery is available. mAMV products are unavailable for cases using *GOES-16* data since the algorithms have not yet been adapted to the finer spatial resolution of *GOES-16* imagery.

d. Lightning observations

Lightning data from the Earth Networks Total Lightning Network (ENTLN) were included (for all 1-min GOES cases except 4 June 2015) given its utility to identify strong updrafts (as discussed in section 1). ENTLN detects lightning using pulses in vertical electric field

measurements from parts of the 1–12-MHz frequency range from over 700 sites across the CONUS (Liu and Heckman 2010). Individual pulses are located in space and time by statistically solving overdetermined electrical signal time-of-arrival equations using measurements from at least five stations. Sources close together in space and time are grouped into flashes (consisting of both cloud-to-ground and intracloud flashes), which were binned into $0.08^\circ \times 0.08^\circ$ longitude–latitude flash density grids for analysis (an area of $\sim 64 \text{ km}^2$) to emulate what could be provided by the *GOES-16* and *GOES-17* Geostationary Lightning Mapper (Goodman et al. 2013).

e. Hail reports

The NCEI severe weather data inventory (SWDI) storm event archive contains the time, location, size,

and source of all confirmed U.S. hail reports that are ≥ 0.75 in. in diameter (NOAA/NCEI 2014). We employ these reports and a smaller number of reports from trusted sources logged by the NOAA Storm Prediction Center (SPC; NOAA/NWS/SPC 1955) for analysis. SPC reports also provide the time, location, and size of severe hail reports, which is defined as those ≥ 1 in. in diameter (all cases analyzed in this study occurred after the severe hail criteria change from ≥ 0.75 to ≥ 1 in. in 2010).

f. Storm tracking

All datasets outlined in the previous sections were analyzed on an individual storm (i.e., “cell”) basis in this study. Storm tracks were developed using objective radar echo-top tracking based on the methods outlined in Homeyer et al. (2017) and Sandmæl (2017). Briefly, storm tracking is accomplished through time linking of 40-dBZ echo-top maxima within echoes classified as convection by the “storm labeling in 3 dimensions” (SL3D) algorithm (Starzec et al. 2017). Objectively tracked storms were reviewed for each case to manually identify and merge discontinuous tracks that correspond to the same storm, resulting in a total of 10 650 storms from the 30 study days. These quality-controlled 5-min tracks were linearly interpolated in time to 1-min intervals to match the finest resolution of the available datasets. Observations and environmental parameters were extracted along the tracks within a 10-km radius of the storm location for analysis. For collocation with GOES data, the tracks were corrected for parallax based on the 10-dBZ echo-top height. Hail reports were linked to the closest storm within a maximum allowed distance of 10 km from the report location, similar to Ortega (2018). Storm areal averages, medians, maxima (or minima), and areas greater than given thresholds for all metrics in Table 2 were also extracted within 10 km.

Storm tracks were then separated into two groups: the “hail” population and the “no hail” population. The hail population consisted of storms with at least one hail report during their lifetime. The no-hail population has lower confidence than its hail counterpart, as one cannot definitively confirm a lack of hail occurrence (i.e., hail may have occurred but was simply not reported). To account for population-driven biases of the no-hail group, storms that occurred in regions of low population were excluded from the analysis. Population density from the Center for International Earth Science Information Network for 2015, the most recent available year, was extracted along each storm track. The distance between storm location and the nearest city was calculated and similarly linked to the storm tracks, but the results using population density were

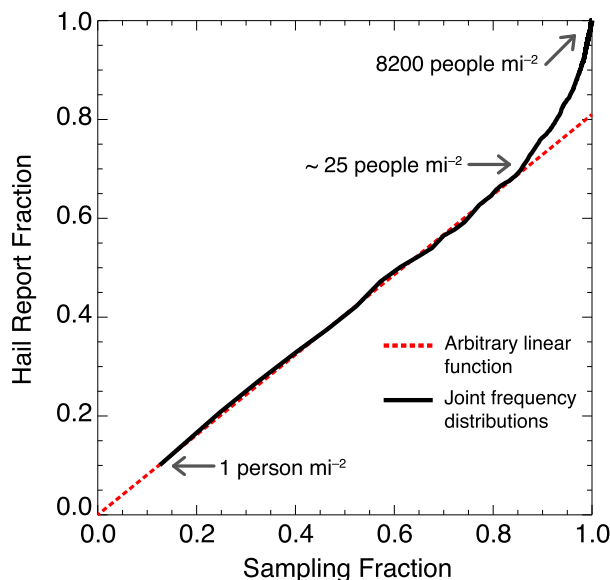


FIG. 4. Comparison of cumulative frequency distributions as a function of population density for all radar observations (sampling) and hail reports analyzed in this study. The thick black line is the joint cumulative frequency distribution relationship, and the red line is an arbitrary linear function to demonstrate the deviation of the observed relationship from linearity. Values of population density along the joint cumulative frequency distribution are labeled at several points within the plot.

clearer. Requiring a lower limit of population density or an upper limit of distance to nearest city aims to retain no-hail storms (or portions of such storms) for analysis with greater confidence (i.e., a sufficient number of potential witnesses are in the path of the storm). By comparing cumulative frequency distributions of radar observations (each representing a potential hail occurrence) and hail reports, both as a function of population density, it is possible to identify if/when the relationship between hail report frequency and population density is not one to one (or at least linear). Figure 4 provides this comparison for all observations analyzed. This comparison reveals that hail reports and population density are linearly correlated below a population density of approximately 25 people per square mile. Beyond this threshold, hail reporting and population density are no longer linearly correlated, and therefore reporting is expected to be less biased. Thus, we imposed a lower limit of 25 people per square mile population density to the no-hail population, minimizing the analyzed observations to segments of no-hail storms with population density higher than this value. Only the filtered sample of no-hail storm observations was used for the rest of this study, which retained approximately 14% of the original sample (see Table 3).

TABLE 3. Resulting storm data used for the first and second types of analysis. The complete no-hail sample is listed for reference, but only the population-filtered no-hail observations were utilized during this study.

Category	No. unique 1-min obs	No. hail reports
Reports ≥ 1 and < 2 in.	3919	4405
Reports ≥ 2 in.	835	995
Filtered no-hail sample	45 589	—
Original no-hail sample	341 348	—

g. Performance evaluation

When testing the skill to objectively distinguish between hail-producing and non-hail-producing storms, storms were labeled as hail producers when a given parameter threshold was met. Probability of detection (POD), false-alarm rate (FAR), and critical success index (CSI) were calculated by testing thresholds for each metric as indicators, as follows:

$$\text{POD} = \frac{\text{No. correctly identified hail storms}}{\text{No. obs hail storms}}, \quad (8)$$

$$\text{FAR} = \frac{\text{No. storms misidentified as hail storms}}{\text{No. identified hail storms}}, \quad \text{and} \quad (9)$$

$$\text{CSI} = \left(\frac{1}{1 - \text{FAR}} + \frac{1}{\text{POD}} - 1 \right)^{-1}. \quad (10)$$

Additionally, parameters were combined into multivariable indicators such that all parameters must exceed their corresponding threshold to indicate a hail-producing storm (e.g., using MESH of 40 mm and VIL density of 7 g m^{-3}).

h. Statistical significance testing

Statistical significance for hail and no-hail population comparisons (discussed in section 3a) is determined using the two-sample Kolmogorov–Smirnov test, as follows:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(X_i) \quad \text{and} \quad (11)$$

$$D_n = \sup_x |F_n(x) - F(x)|, \quad (12)$$

where X_i represents each observation, $F_n(x)$ is the cumulative distribution function calculated for each sample, and D_n is the critical value (commonly referred to as a p value). The null hypothesis (that the two samples originated from the same distribution) is rejected when the critical value is less than the desired confidence interval α . For this evaluation, we set $\alpha = 0.01$ (confidence

of 99%, such that differences are significant if the p value is less than or equal to 0.01).

Statistical significance of the relative skill of discriminating severe hail-producing storms (discussed in section 3b) was assessed using a bootstrapping approach, where 80% of the analyzed storm population was randomly selected 10 000 times to compute independent estimates of skill (CSI as a function of POD). The resulting uncertainty (standard deviation or σ) in skill estimates was used to identify significant differences, if they existed. Differences were deemed significant if they were larger than the 3σ uncertainties of considered metrics (analogous to a confidence interval of 99%).

3. Results

a. Objective hail-occurrence identification

To evaluate the potential for hail identification using the datasets available, metrics detailed in Table 2 within 5 min of a hail report were divided into size categories and compared with metrics at all instances of the no-hail population (see Table 3 for a list of categories and number of observations). A 5-min time window was chosen to mitigate errors resulting from 1) the relatively coarse time sampling of the radar data (~ 5 min for the single-radar volumes and exactly 5 min for the analyzed GridRad data) and 2) expected errors in report timing. Though exact errors in report timing are unknown, we expect most to be no longer than ± 5 min. Several metrics were able to identify hail occurrence well, with the greatest separation between hail and no-hail observations found in storm-maximum MESH, VIL density, and H_{DR} above and below the melting level (AML and BML, respectively, Fig. 5). For these four variables, the no-hail distribution was significantly different from the two hail distribution categories, and each hail category was significantly different from one another (see Table 4 for p values). Joint frequency distributions of multiple parameters were also evaluated, but nearly all provided similar separation to that found using single variables. One combination provided some improved separation— H_{DR} BML and MESH—with MESH allowing for more reliable discrimination in the overlapping tails of the H_{DR} BML hail and no-hail populations (not shown).

In comparison, several metrics that have shown some utility at discriminating between severe and nonsevere storms (or hail and nonhail storms) in previous studies either did not show significant separation between hail and no-hail populations near the time of hail occurrence, or their differences, although significant, were not of practical use because of broad overlap between populations. Satellite and lightning parameters associated

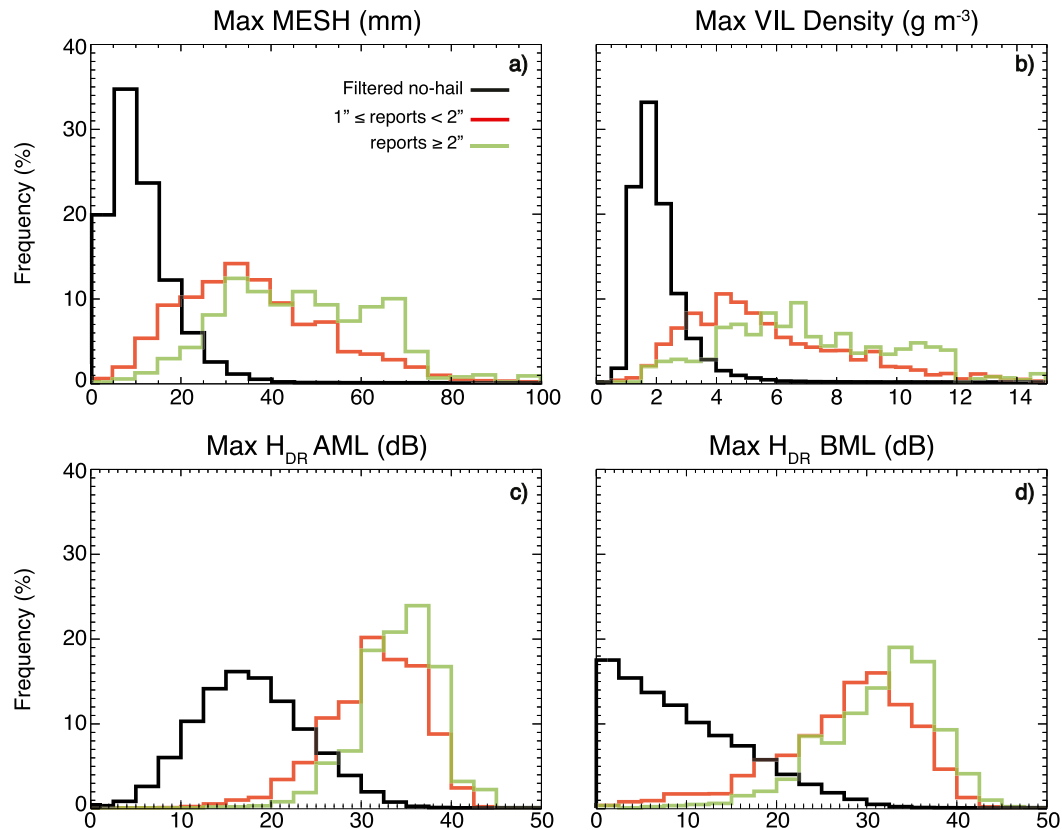


FIG. 5. Frequency distributions of storm-maximum (a) MESH, (b) VIL density, (c) H_{DR} above the environmental melting (0°C) level, and (d) H_{DR} below the environmental melting level within ± 5 min of a hail report (red curves for hail reports of 1–2 in. in diameter, and green curves for ≥ 2 in. in diameter) and within population-filtered segments of no-hail storms (black curves).

with severe weather, such as those in Fig. 6, displayed overlap between all distributions (see Table 4 for p values). The maximum mAMV CTD showed some reasonable skill, displaying the best separation of all satellite products. However, mAMVs are not routinely available because of their reliance on 1-min imagery, and, as a result, these distributions only included data from 8 of the 30 severe weather days analyzed here

(denoted by asterisks in Table 1). Alternative radar parameters such as 40-dBZ echo-top altitudes and column-maximum radial divergence (Figs. 6e,f), were not as skillful as other radar variables previously mentioned. These results demonstrate that, while many products may be useful for severe weather prediction, or in specific scenarios, objective hail identification is best accomplished using a few select SP and DP radar parameters.

TABLE 4. The p values from Kolmogorov–Smirnov testing for all of the variable distributions presented in this study. Values that are statistically significant are in boldface type. Here, A indicates $1 \leq \text{hail} < 2$ in., B indicates hail ≥ 2 in., and C indicates no hail.

	A vs C	B vs C	A vs B
Max H_{DR} BML	0.000 00	0.000 00	$1.370\,87 \times 10^{-32}$
Max H_{DR} AML	0.000 00	0.000 00	$7.289\,28 \times 10^{-32}$
Max MESH	0.000 00	0.000 00	$3.740\,35 \times 10^{-41}$
Max VIL density	0.000 00	0.000 00	$2.561\,57 \times 10^{-42}$
Min GOES IR brightness	0.000 00	$7.170\,52 \times 10^{-26}$	0.000 09
GOES OT area	0.000 00	$3.172\,18 \times 10^{-34}$	0.066 51
ENTLN total lightning flash density	0.000 00	0.000 00	0.050 35
Max GOES mAMV cloud-top divergence	0.000 00	0.000 00	$2.887\,49 \times 10^{-26}$
40-dBZ echo-top altitude	0.000 00	0.000 00	$7.597\,45 \times 10^{-40}$
Max radar divergence column	0.000 00	0.000 00	$9.740\,05 \times 10^{-23}$

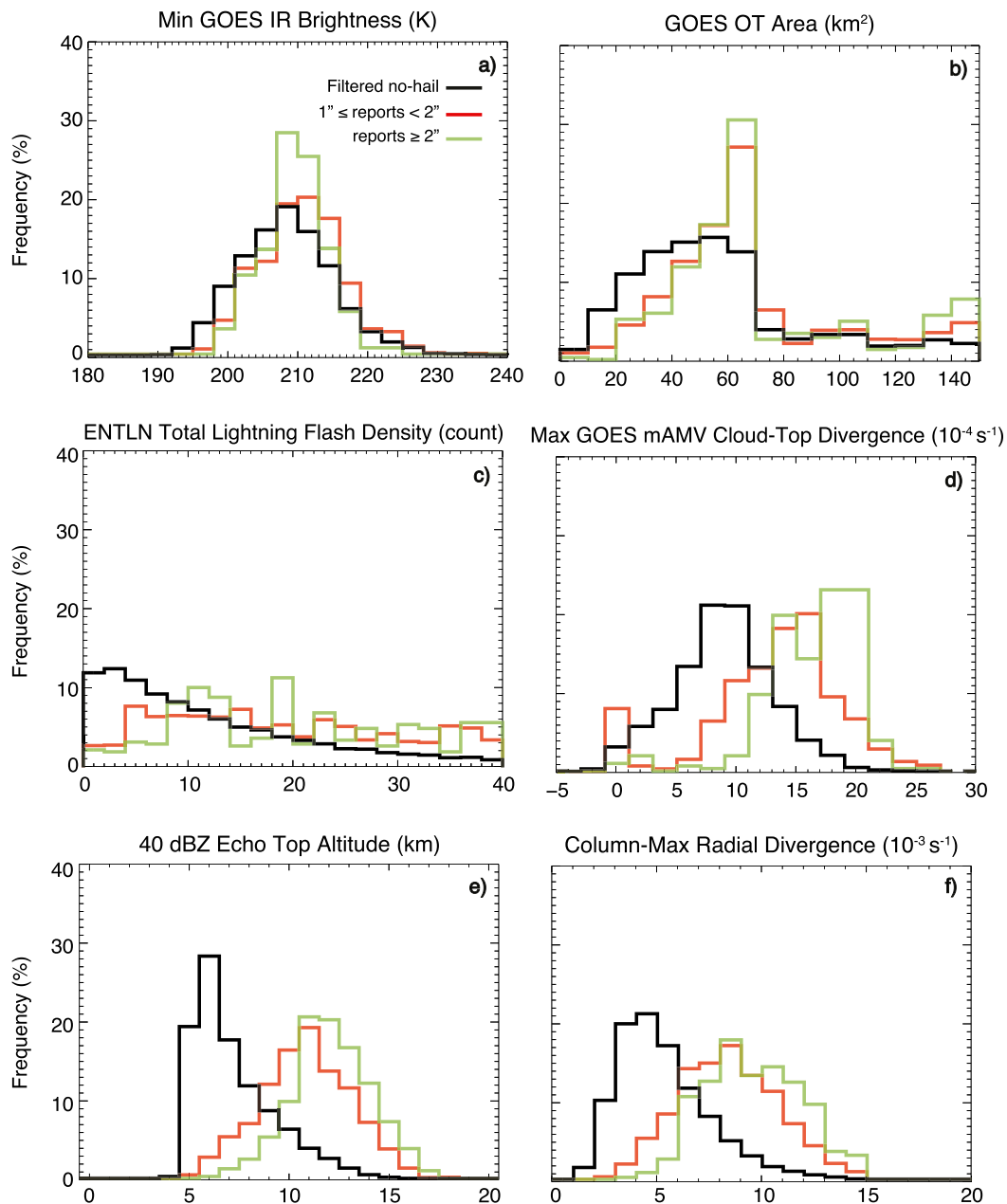


FIG. 6. As in Fig. 5, but for (a) storm-minimum GOES IR brightness temperature, (b) GOES overshooting-top area, (c) ENTLN total lightning flash density, (d) storm-maximum GOES mAMV cloud-top divergence, (e) 40-dBZ echo-top altitude, and (f) storm-maximum radar divergence.

Similar comparisons were completed at the exact time of reports to determine whether or not the previous findings were a result of considering extremes in these metrics over a 10-min period (Figs. 7 and 8). These comparisons are consistent with the time window results, but show more overlap between the hail and no-hail populations. Despite increased overlap, differences between populations remain significant. Given the consistent results between these two approaches and the increased separation in

the time window analysis, all remaining analyses use the ± 5 -min extremes of the metrics for reference.

b. Objective severe hail-producing storm identification

While we found that several parameters are skillful at identifying hail occurrence, verification for any objective hail identifier remains a challenge because of the numerous reporting biases. To fully assess the utility of

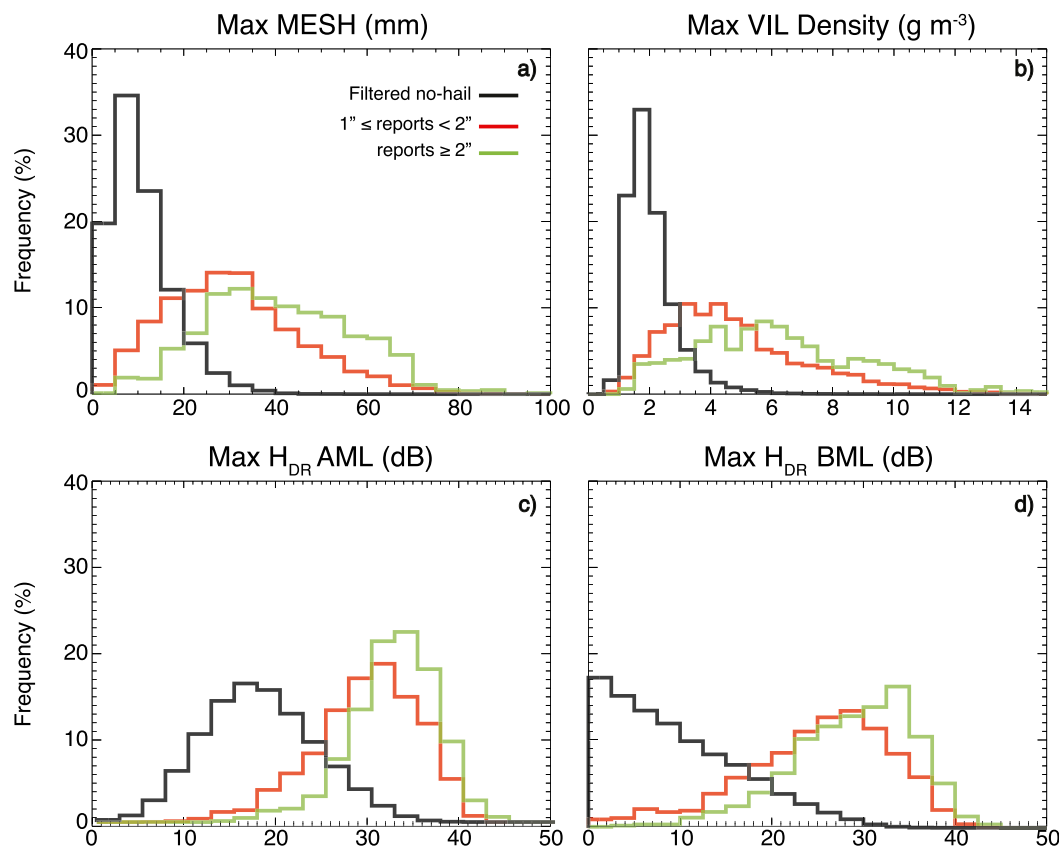


FIG. 7. As in Fig. 5, but at the time of hail reports.

objective identifiers, we must approach hail identification from both an individual report perspective and a storm perspective. Identifying severe hail-producing storms rather than occurrence allows us to better evaluate the performance of these metrics and build confidence in the results from section 3a. There are 805 severe hailstorms in our dataset (those with at least one ≥ 1 -in. report). When evaluating the no-hail storm population, only time elements of storms that occur in population-dense regions of at least 25 people per square mile ($1 \text{ mi}^2 = 259 \text{ ha}$) are retained, resulting in data from 3823 no-hail storms.

POD, FAR, and CSI for ranges of VIL density, MESH, and H_{DR} BML, the three best discriminators for hail occurrence, are evaluated to identify the most skillful (i.e., greatest CSI with a corresponding high POD) thresholds for each metric (Fig. 9). Each variable resulted in similar performance, with a slightly broader peak and higher maximum CSI for H_{DR} BML that is displayed better when comparing CSI as a function of POD (left panel of Fig. 10). However, these differences are not significant given the overlapping $\pm 3\sigma$ uncertainty from bootstrapping. Even so, larger increases in skill are possible when using a multivariable approach, such that two metrics

must meet the specified criteria. Namely, using either SP variable with H_{DR} BML (right panel of Fig. 10) provides an increase of ~ 5 percentage points in CSI relative to H_{DR} BML alone. While the $\pm 3\sigma$ uncertainties still overlap, greater separation in the mean performance exists between the multivariable and single-variable indicators. Thus, using both SP and DP metrics appears to provide practical increases in skill. It is also worth noting that the peak CSI tends to shift toward higher POD in the multivariable approach. This performance evaluation, however, is sensitive to the sizes of each population. It is possible that the true skill is slightly lower than that depicted here, considering that we are minimizing the no-hail distribution to population-dense regions. Nonetheless, the relative performance between a single-variable and multivariable approach is expected to be independent of these sample size limitations.

When evaluating the performance for identifying significant severe hailstorms, CSI is lower and more variable across both the individual parameter and joint parameter space (Figs. 11 and 12). This is likely a product of decreased sample size for the significant severe hailstorm population and a relative increase in sample size

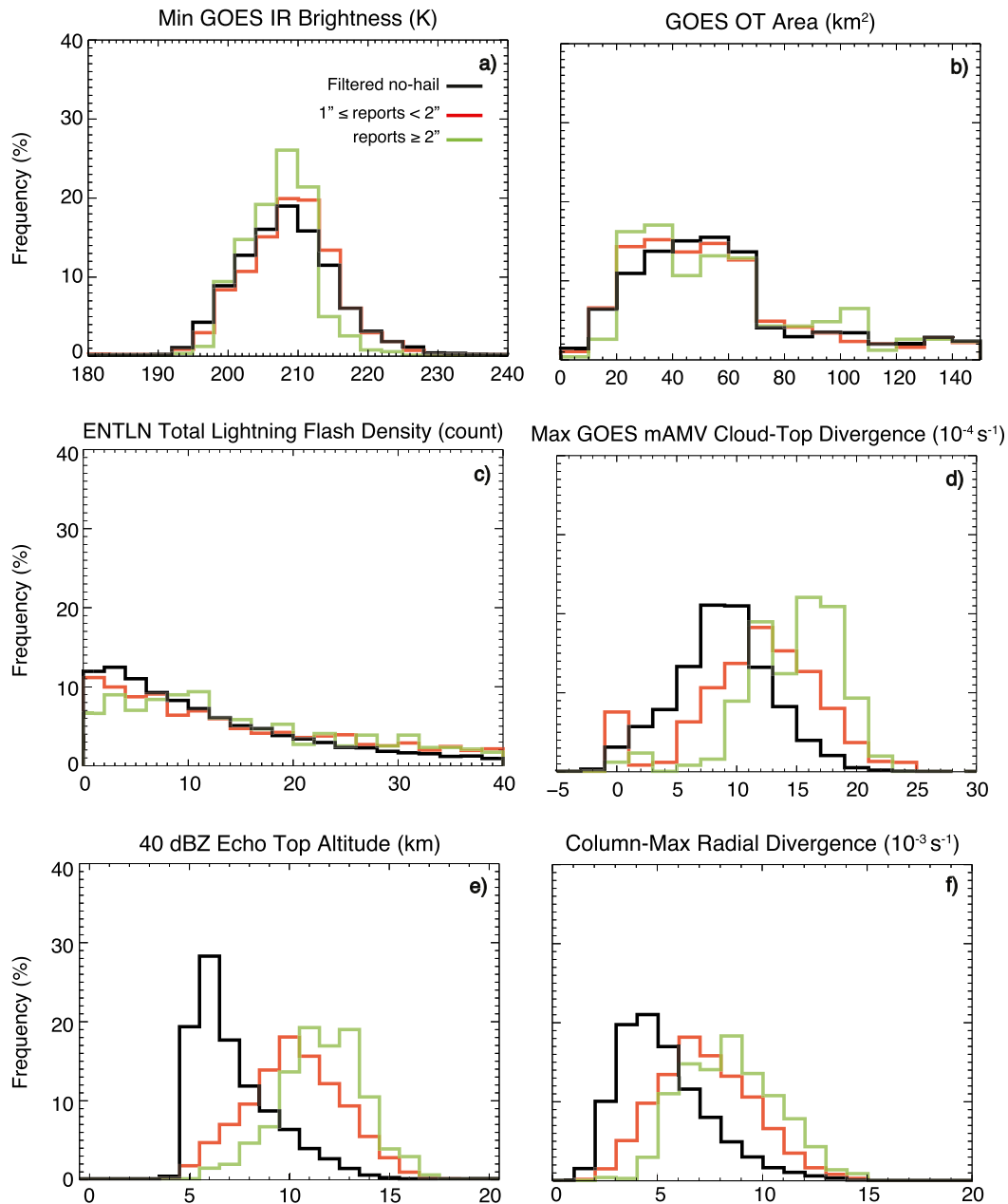


FIG. 8. As in Fig. 6, but at the time of hail reports.

of the null storm population, now consisting of storms that produced smaller severe hail and the no-hail storms. Peak skill shifts to higher values for each metric, indicating that each metric provides some hail size discrimination (as inferred from the analysis in section 3a). In contrast to the performance evaluation for all severe hailstorms, the SP radar metrics slightly outperform H_{DR} BML for identification of significant severe hailstorms. When comparing the single and multivariable identifiers, overall increases and a shift in peak skill to higher

POD are again found when using multivariable (SP and DP metrics) methods, which is especially true for the combination of MESH and H_{DR} BML.

c. Objective hail size identification

Traditionally, radar observations are collocated with hail reports to establish the best hail size relationships, which was not a primary goal of this study. Since we instead retain the extreme values of each metric within a storm, we can assess the characteristics of observed hail

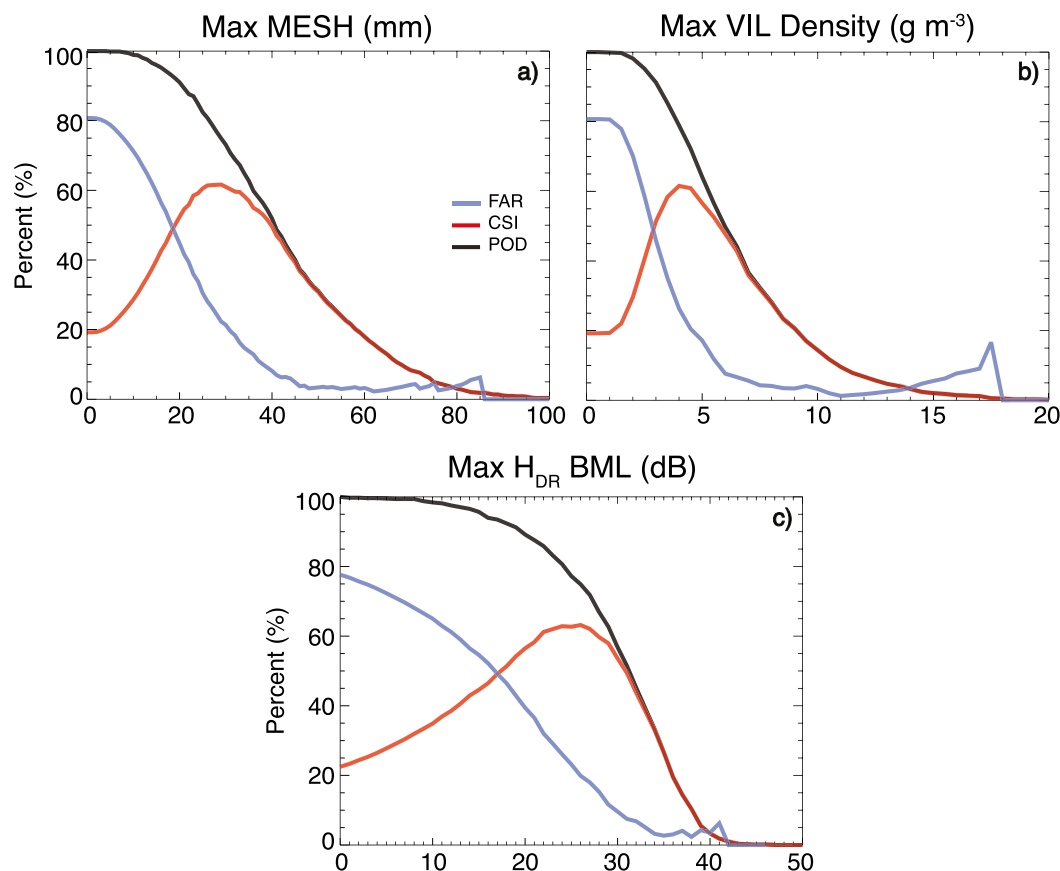


FIG. 9. The POD, FAR, and CSI of threshold-based hail-storm identification using storm-maximum (a) MESH, (b) VIL density, and (c) H_{DR} below the environmental melting level.

distributions as a function of the metrics for broad size discrimination and maximum hail size potential. Based on the analyses in sections 3a and 3b, we only evaluate MESH, VIL density, and H_{DR} below the melting level. The size discrimination performance provided by these three metrics is similar (not shown) so, for brevity and since SP metrics can be readily applied globally and to existing long-term datasets, we focus our attention on MESH for the remainder of this analysis. Since the reported hail size distribution is biased via clustering to reference objects, uniform random noise was added to the report observations following methods outlined in Allen et al. (2017) as follows:

$$y_{\text{new}} = y_{\text{old}} + y_{\text{noise}} \quad \text{and} \quad (13)$$

$$y_{\text{noise}} = \pm(0.247y_{\text{old}} + 0.0279), \quad (14)$$

where y_{noise} is a random value distributed around 0 with a standard deviation of $0.247y_{\text{old}} + 0.0279$, designed to avoid introducing large bias to small reported sizes, and is not allowed to exceed ± 0.5 in. for any report. This

creates a more continuous, but similarly shaped, hail size distribution that is expected to be more representative of what occurs in nature. Hereinafter, the report sizes with uniform random noise added are referred to as smoothed hail sizes.

Broad, near-linear relationships were found between MESH and both raw and smoothed hail sizes, but MESH based on the 75th-percentile hail size equation in Witt et al. (1998a) was biased. Namely, the original MESH equation resulted in an underestimate of smaller hail sizes and an overestimate of larger hail sizes relative to the 75th percentile of the total hail report distribution (Fig. 13). It is important to note that the power-law relationship between SHI and MESH from Witt et al. (1998a) was empirically derived using a much smaller dataset than that here (147 vs 5954 reports, respectively). Thus, a revised power-law relationship seems appropriate given that MESH is routinely used in operational and climatological studies. Thus, four revised MESH equations are presented in Fig. 13 based on power-law fits to the 75th- and 95th-percentile hail sizes in the raw and smoothed report datasets. When utilizing the raw hail

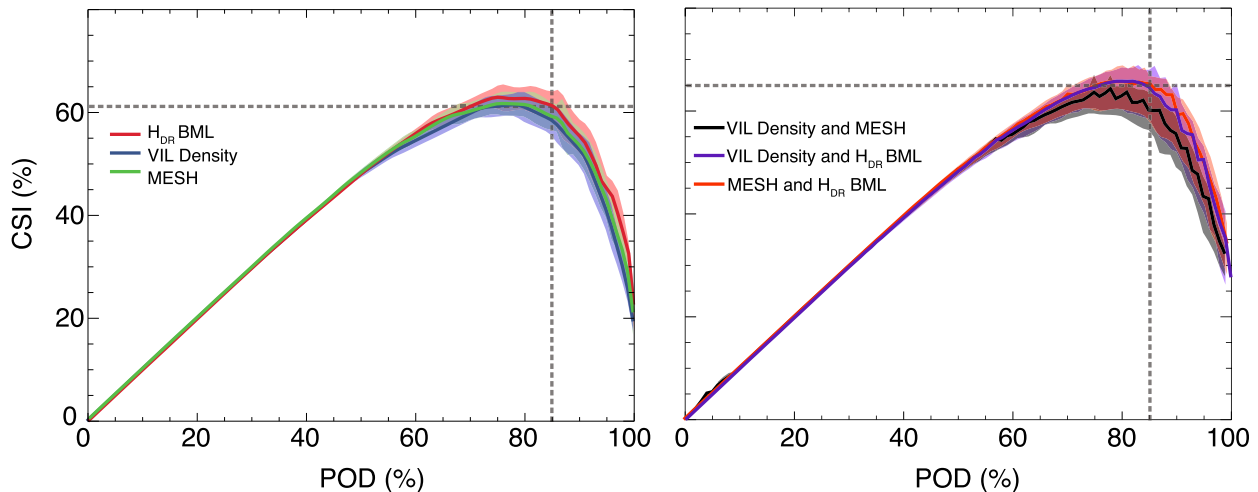


FIG. 10. (left) Mean CSI as a function of POD resulting from bootstrapping for H_{DR} below the environmental melting level (red), VIL density (blue), and MESH (green) for the range of threshold values covered in Fig. 9. (right) Maximum mean CSI as a function of POD resulting from bootstrapping for multivariable hail-storm identification using VIL density and MESH (black), VIL density and H_{DR} below the environmental melting level (purple), and MESH and H_{DR} below the environmental melting level (orange). Color-filled areas indicate the $\pm 3\sigma$ uncertainty in average skill ($\alpha = 0.01$) for each variable. Vertical dashed lines in each panel indicate a POD of 85%, and horizontal dashed lines indicate the highest corresponding CSI.

reports, percentile bin values are constrained by the discretized reference object sizes, which has a minor effect on the resulting power-law fits. Binned values using the smoothed report dataset are not as constrained to the report sizes and are more consistent with the shape of the resulting power-law fits.

The vast majority of hail reports coincided with SHI values in a manner consistent with a power-law relationship, such that an increased envelope of reported sizes was associated with increased SHI. However, 59 one-minute reports of 4.5-in. hail fell outside the envelope covered by nearly all of the remaining observations (red diamonds in Fig. 13a). It was found that 53 of these reports were associated with a single storm over an area in the Texas Panhandle with low population density during the 27 May 2015 case, from an SWDI report with a 53-min time window valid from 2301 to 2353 UTC. Cursor evaluation of radar data during this period revealed that the storm was a discrete supercell that was intensifying up to the end time of the report window, where the maximum Z_H exceeded 50 dBZ for the first time. Based on this evolution, it is probable that the report was only valid near the end of the 53-min window. Including these likely dubious 53 reports skews the 95th-percentile calculations and the smoothed report dataset away from the distribution formed from the remaining 5897 reports. Therefore, for simplicity, all 4.5-in. hail reports were removed when introducing random noise and computing the MESH–SHI power-law fits. Based on the results using the smoothed report dataset, the new MESH–SHI

relationships for the 75th and 95th percentiles are as follows:

$$\text{MESH}_{75} = 16.566(\text{SHI})^{0.181} \quad \text{and} \quad (15)$$

$$\text{MESH}_{95} = 17.270(\text{SHI})^{0.272}. \quad (16)$$

While other aspects of the Witt et al. (1998a) hail detection algorithm have been revised in previous work, including probability of severe hail (e.g., Mahale et al. 2014), this is the first time (to the authors' knowledge) that the MESH–SHI relationship has been revisited.

To evaluate the practical utility of MESH_{75} and MESH_{95} in comparison with the original Witt et al. (1998a) relationship ($\text{MESH}_{\text{WittEtAl}}$ hereinafter), we evaluated time-accumulated maps of MESH extremes (commonly referred to as “hail swaths”) and coincident hail reports for several days and provide two representative examples in Fig. 14. The two cases shown are 4 April 2017 over the southeastern United States and 16 May 2017 over the central U.S. Great Plains. The 4 April 2017 case features severe hail fall in regions with more than 25 people per square mile (our threshold for statistical analysis), and the 16 May 2017 case features severe hail fall over regions with low population density. These comparisons demonstrate that the new MESH_{75} and MESH_{95} relationships provide improved spatial coverage of areas where severe hail reports occurred and greater agreement between the largest predicted and observed hailstones when compared with the $\text{MESH}_{\text{WittEtAl}}$ relationship,

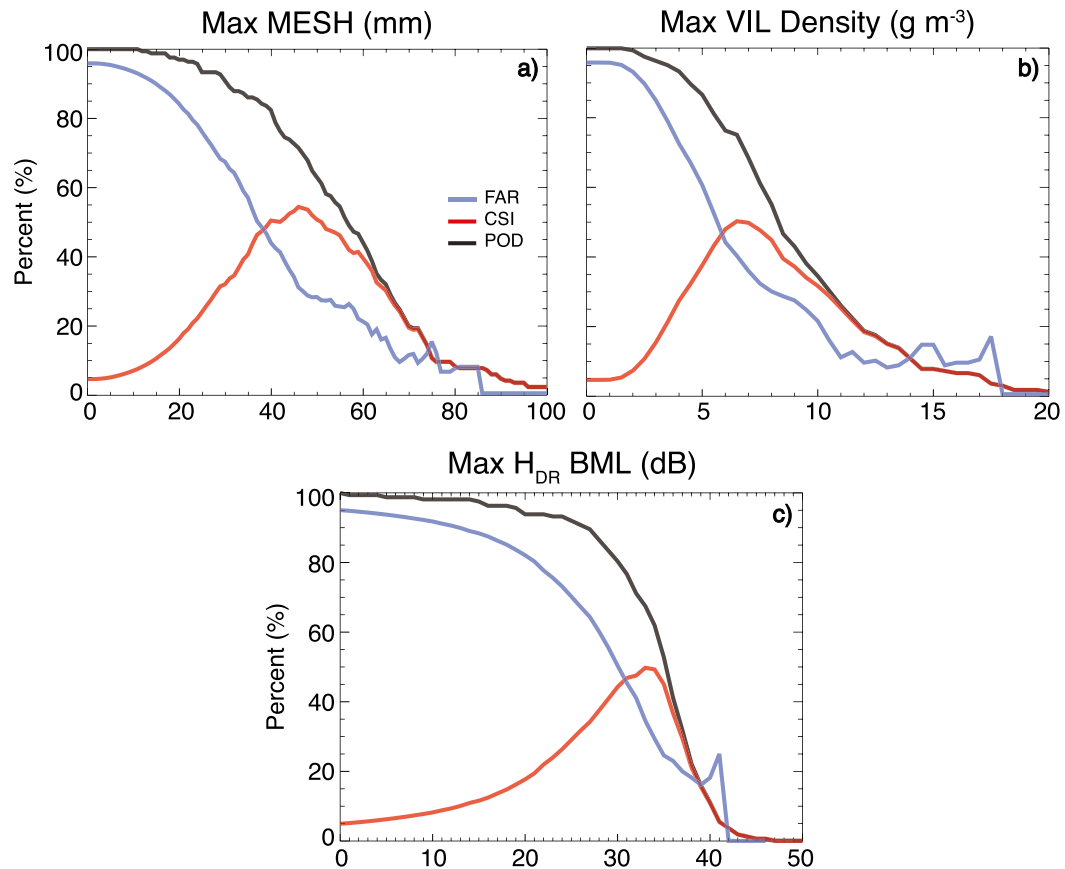


FIG. 11. As in Fig. 9, but for significant hail-producing storms.

with MESH₉₅ providing the best overall estimate of maximum hail size. Though overforecasting may be possible from MESH₉₅, previous studies have shown that hail reports tend to underestimate the true maximum size (Bardsley 1990; Blair and Leighton 2012;

Blair et al. 2017; Allen et al. 2017, and references therein). Furthermore, difficulty remains in providing adequate validations of radar-estimated hail fall given the broader limitations of reporting outlined in the Introduction. A cursory evaluation of MESH₇₅ and

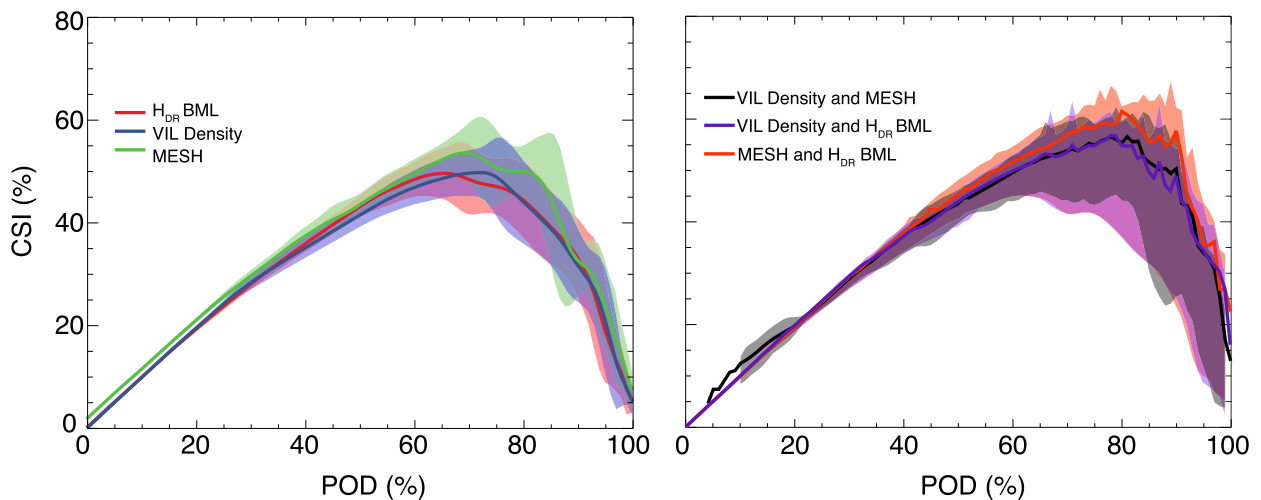


FIG. 12. As in Fig. 10, but for significant hail-producing storms.

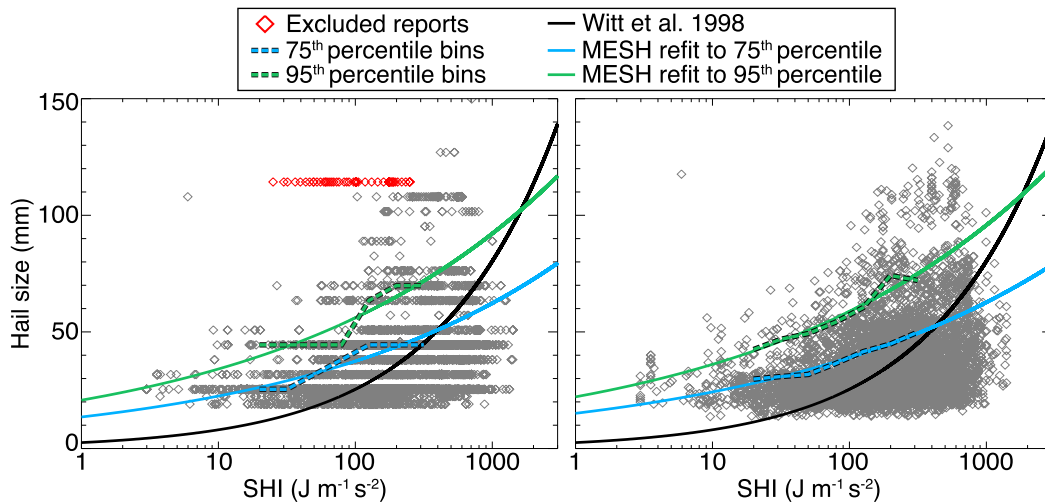


FIG. 13. Scatterplot of SHI for (left) all raw reported hail sizes and (right) smoothed hail sizes via random noise. Superimposed are the original MESH fit to the 75th percentile of hail size from Witt et al. (1998a) (black line) and two revised MESH function fits to the 75th and 95th percentiles of hail size based on the data analyzed in this study (blue and light-green lines, respectively). The log-scale bin values used to constrain the new MESH fits are shown by the dashed lines. Reports that are colored red were excluded from the fit calculation.

MESH₉₅ using GridRad volumes from randomly selected days between 2004 and 2016 (not shown) reveals similar agreement between predicted and reported hail sizes.

It is worth noting that we repeated the performance evaluations from section 3b with MESH₇₅ and MESH₉₅ and found no significant differences apart from changes in the MESH thresholds at which peak skill was achieved, with an increase from 29 mm to 40 and 64 mm, respectively. This result is expected given that the only difference between MESH calculations is the coefficients of the power-law fit.

4. Discussion and conclusions

This study completed a comprehensive analysis assessing the ability of quantitative SP radar, DP radar, satellite, and lightning products to identify hail occurrence, hail-producing storms, and hail size. Our conclusions are as follows:

- 1) A population density threshold of 25 people per square mile along a storm's path appears to be a helpful criterion to establish a less biased (and therefore more confident) null population for hailstorm analyses.
- 2) MESH, VIL density, and H_{DR} BML are the best discriminators for hail occurrence.
- 3) The H_{DR} BML performs best for single-parameter severe hailstorm (produced ≥ 1 -in. report) detection, while MESH performs best for significant severe hailstorm (produced ≥ 2 -in. report) detection.
- 4) A combination of SP and DP radar metrics (i.e., a multiparameter approach) provides noticeably improved performance for hailstorm detection relative to any single-parameter approach.
- 5) Satellite and lightning parameters displayed the greatest overlap between hail and no-hail populations, resulting in reduced practical utility as discriminators.
- 6) Radar parameters that perform well at hail occurrence and severe hailstorm identification scale roughly linearly with observed hail size, but no one parameter shows superior performance. However, revising the original MESH formulation from Witt et al. (1998a) to a relationship associated with the 95th percentile of observed hail size improves the spatiotemporal comparisons between hail reports and radar-based estimates.

The skill seen using MESH and VIL density to distinguish severe hail occurrence is likely a result of two factors. First, these parameters are computed by vertical integration of Z_H within a storm. Larger integrated values represent greater storm depths, which correspond to deeper/stronger convection that is more favorable for hail production. Second, Z_H is most sensitive to particle size, such that the presence of large hydrometeors results in additional increases in integrated values compared to similarly deep storms without such particles. Thus, higher MESH and VIL density values represent both deeper/stronger storms and a greater likelihood of larger particles. H_{DR} BML is also skillful at hail identification because of the

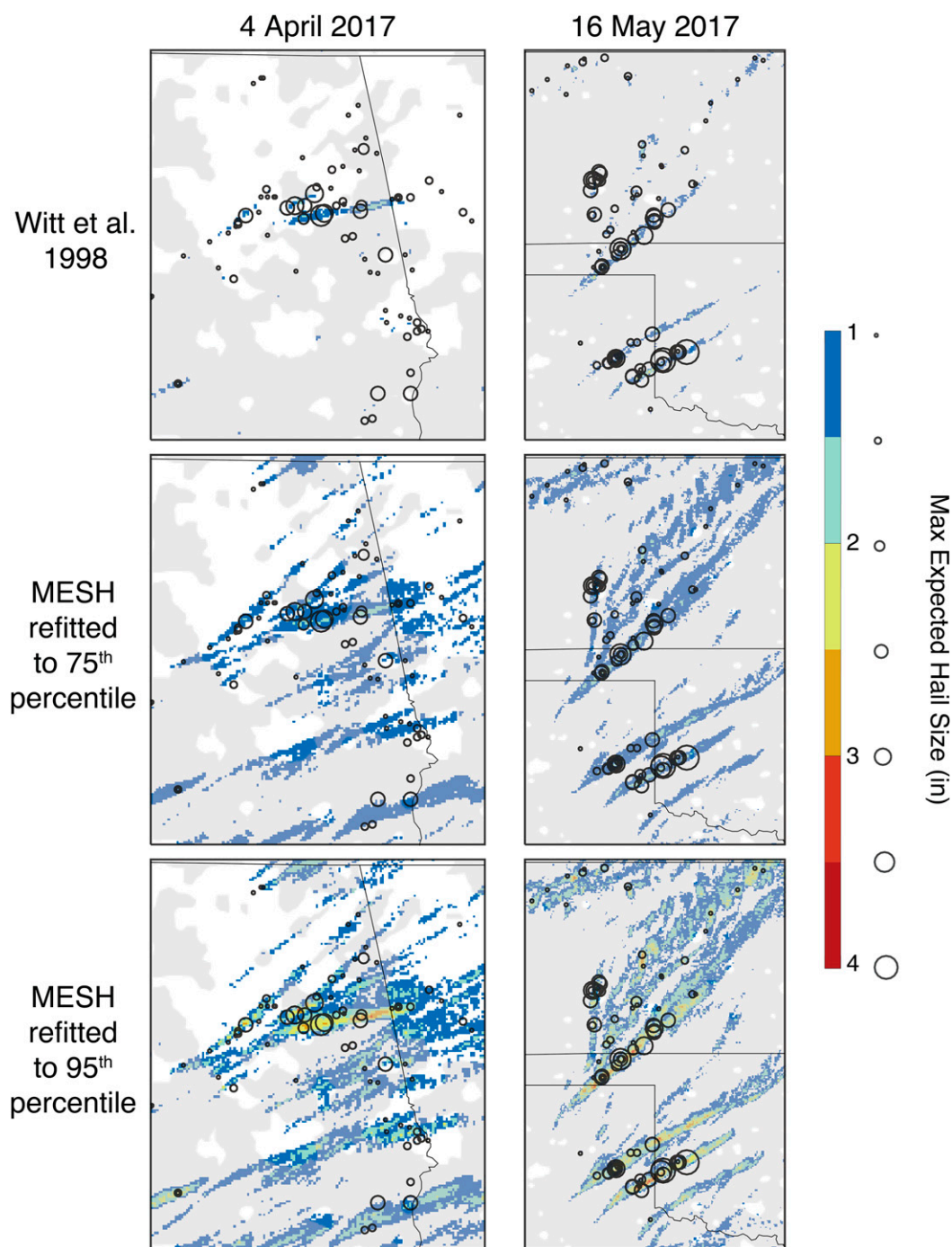


FIG. 14. Time-accumulated maximum MESH values (or "swaths") over portions of the domains in two cases, (left) 4 Apr 2017 in the southeastern United States and (right) 16 May 2017 in the central Great Plains, for (top) the original Witt et al. (1998a) MESH equation and the revised fits to the (middle) 75th and (bottom) 95th percentiles of observed hail sizes analyzed in this study. Black open circles indicate hail reports, which increase in diameter with increasing hail size (illustrated adjacent to the color bar). The superimposed dark gray mask indicates regions that are less than 25 people per square mile.

combined utility of Z_H and Z_{DR} that enables representation of physical characteristics of hydrometeors. Large hydrometeors that are observed by the radar as relatively spherical (e.g., large, tumbling hail) have different scattering characteristics than large, oblate rain drops. Therefore, combining Z_H and Z_{DR} results in skillful discrimination between hail and no-hail storms because it is able to reveal volumes where the scattering characteristics are dominated by hailstones. Conversely, those parameters that did not provide skillful discrimination (including those listed in Table 2 but not shown) either did not resolve the dynamical characteristics well enough to indicate major differences between the storm populations or did not adequately identify the physical characteristics of hydrometeors.

While previous studies utilizing satellite and/or lightning data have been successful at discriminating between severe and nonsevere storms (e.g., Cintineo et al. 2013, 2014; Gravelle et al. 2016; Line et al. 2016; Bedka and Khlopenkov 2016; Schultz et al. 2017; Farnell et al. 2018), it is not surprising that this study did not provide similar results. Satellite signatures, which tend to be weak dynamical indicators, are often observed throughout a storm's life cycle, and lightning signatures tend to occur earlier. Additionally, both are commonly binary (OTs, AACPs, lightning jumps, etc.—which are not evaluated here). Thus, these indicators can provide skill in short-term prediction of severe storms but less skill for the detection of a specific hazard. Severe storms on average appear more intense (colder cloud tops, increased texture in visible and near-IR imagery, and more frequent lightning) than nonsevere storms from satellite observations, but it is extremely difficult to accurately discriminate between individual severe hazards using satellite data alone. Our results demonstrate that radar observations outperform quantitative metrics from satellite and lightning data for severe hail detection.

It is possible that the resulting performance of radar metrics may change when evaluating single-radar observations. The multiradar composite GridRad data are the result of distance- and time-weighted binning of single-radar observations (i.e., averaging), which accounts for the time offset between elevation scans. However, operational single-radar observations are not adjusted for the resulting artificial storm tilt, which is a challenge for operational applications. The binning and averaging processes also produces slightly lower extreme values of the radar variables than in single-radar observations. Since all data are similarly impacted by this averaging, the relative performance of single-radar analyses is expected to be similar, with the threshold values at peak skill expected to increase in magnitude. Greater uncertainty is expected for metrics based on DP variables, since single-radar data

are often impacted by systematic biases in Z_{DR} from poor radar calibration, which is another challenge for operational applications. Bias correction, such as the objective approach used when creating GridRad analyses, must be completed before any DP identification method is used. To fully assess the potential for operational applications, future studies must evaluate the identification skill for metrics that performed well in this study, with a similarly large dataset, but utilizing single-radar observations.

We evaluated the performance of extreme values of hail metrics within radar-tracked storms over a 10-min period rather than analyzing values coincident with the hail reports, which is often the approach taken in previous work (e.g., Ortega 2018). This choice was made to avoid the spatiotemporal limitations of the hail reports and the observations. While this approach could result in unintended biases and negatively impacted performance, the results for several metrics agree with past studies using the traditional space–time matching approach [including H_{DR} from Depue et al. (2007)]. More sophisticated methods (e.g., linear discriminant analysis) could also improve identification skill. However, given the large dataset used, the simpler single- and multiparameter threshold discrimination approach here likely produces similar results to a more complex technique.

While these analyses leverage operational datasets for hailstorm discrimination, further work is needed to assess the operational utility of these findings. Nowcasting was not a goal of this study, but it is possible to identify lead times associated with the three variables we used for discrimination between storms that do and do not produce severe hail. For example, the first time at which a threshold is exceeded could be compared to the time of the first severe hail report associated with a storm. However, given the limitations of the hail reports, it is not clear that such a result would be useful (or even accurate). Data from hail detection arrays or a large set of improved hail reports (i.e., without discretization), when available, should be used in future studies to assess the forecasting/nowcasting capabilities of the metrics analyzed here and to pursue improved hail size discrimination. An evaluation of the performance of the revised MESH₇₅ and MESH₉₅ relationships with such a dataset is also warranted.

The major findings in this study can be utilized to provide improved radar-based hail fall and hailstorm detection in the future. Potential applications include verification of model forecasts, hail climatologies using large-area radar data archives such as GridRad, the Multi-Radar Multi-Sensor (MRMS) system or its historical counterpart the Multi-Year Reanalysis of Remotely Sensed Storms (MYRORSS), and operational implementations (including real-time MRMS and potentially

single-radar observations). Continued evaluation of these radar observations and data from additional remote sensing platforms (e.g., *GOES-16* and *GOES-17*), especially using higher quality hail observations, is needed to determine their potential uses for hailstorm discrimination. High-performing radar metrics, such as those identified here, may also be useful for evaluating alternative datasets in greater detail rather than relying solely on incomplete hail reports.

Acknowledgments. This work was supported by the National Aeronautics and Space Administration (NASA) under Award NNX15AV81G. We thank Kristopher Bedka of NASA Langley and John Mecikalski and Jason Apke of the University of Alabama in Huntsville for providing feedback and access to several derived satellite datasets, and we thank Chris Sloop and Stan Heckman of Earth Networks for access to the ENTLN data. Multidataset storm-track files produced during this effort are available from the authors upon request.

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), <http://www.ejssm.org/ojs/index.php/ejssm/article/view/149/104>.
- , —, 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>.
- Amurn, S. A., and P. L. Wolf, 1997: VIL density as a hail indicator. *Wea. Forecasting*, **12**, 473–478, [https://doi.org/10.1175/1520-0434\(1997\)012<0473:VDAHI>2.0.CO;2](https://doi.org/10.1175/1520-0434(1997)012<0473:VDAHI>2.0.CO;2).
- Apke, J. M., J. R. Mecikalski, K. Bedka, E. W. McCaul, C. R. Homeyer, and C. P. Jewett, 2018: Relationships between deep convection updraft characteristics and satellite-based super rapid scan mesoscale atmospheric motion vector-derived flow. *Mon. Wea. Rev.*, **146**, 3461–3480, <https://doi.org/10.1175/MWR-D-18-0119.1>.
- Aydin, K., T. A. Seliga, and V. Balaji, 1986: Remote sensing of hail with a dual linear polarization radar. *J. Climate Appl. Meteor.*, **25**, 1475–1484, [https://doi.org/10.1175/1520-0450\(1986\)025<1475:RSOHW>2.0.CO;2](https://doi.org/10.1175/1520-0450(1986)025<1475:RSOHW>2.0.CO;2).
- Bardsley, W. E., 1990: On the maximum observed hailstone size. *J. Appl. Meteor.*, **29**, 1185–1187, [https://doi.org/10.1175/1520-0450\(1990\)029<1185:OTMOHS>2.0.CO;2](https://doi.org/10.1175/1520-0450(1990)029<1185:OTMOHS>2.0.CO;2).
- Bedka, K. M., and J. R. Mecikalski, 2005: Application of satellite-derived atmospheric motion vectors for estimating mesoscale flows. *J. Appl. Meteor.*, **44**, 1761–1772, <https://doi.org/10.1175/JAM2264.1>.
- , and K. Khlopenkov, 2016: A probabilistic multispectral pattern recognition method for detection of overshooting cloud tops using passive satellite imager observations. *J. Appl. Meteor. Climatol.*, **55**, 1983–2005, <https://doi.org/10.1175/JAMC-D-15-0249.1>.
- , C. S. Velden, R. A. Petersen, W. F. Feltz, and J. R. Mecikalski, 2009: Comparisons of satellite-derived atmospheric motion vectors, rawinsondes, and NOAA wind profiler observations. *J. Appl. Meteor. Climatol.*, **48**, 1542–1561, <https://doi.org/10.1175/2009JAMC1867.1>.
- , C. Wang, R. Rogers, L. D. Carey, W. Feltz, and J. Kanak, 2015: Examining deep convective cloud evolution using total lightning, WSR-88D, and *GOES-14* super rapid scan datasets. *Wea. Forecasting*, **30**, 571–590, <https://doi.org/10.1175/WAF-D-14-00062.1>.
- , E. M. Murillo, C. R. Homeyer, B. Scarino, and H. Mersiovsky, 2018: The above anvil cirrus plume: An important severe weather indicator in visible and infrared satellite imagery. *Wea. Forecasting*, **33**, 1159–1181, <https://doi.org/10.1175/WAF-D-18-0040.1>.
- Blair, S. F., and J. W. Leighton, 2012: Creating high-resolution hail datasets using social media and post-storm ground surveys. *Electron. J. Severe Storms Meteor.*, **13**, 32–45.
- , 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>.
- Bringi, V. N., and V. Chandrasekar, 2001: *Polarimetric Doppler Weather Radar*. Cambridge University Press, 636 pp.
- Brunner, J. C., S. A. Ackerman, A. S. Bachmeier, and R. M. Rabin, 2007: A quantitative analysis of the enhanced-V feature in relation to severe weather. *Wea. Forecasting*, **22**, 853–872, <https://doi.org/10.1175/WAF1022.1>.
- Capozzi, V., E. Picciotti, V. Mazzarella, F. S. Marzano, and G. Budillon, 2018: Fuzzy-logic detection and probability of hail exploiting short-range X-band weather radar. *Atmos. Res.*, **201**, 17–33, <https://doi.org/10.1016/j.atmosres.2017.10.006>.
- Cecil, D. J., and C. B. Blankenship, 2012: Toward a global climatology of severe hailstorms as estimated by satellite passive microwave imagers. *J. Climate*, **25**, 687–703, <https://doi.org/10.1175/JCLI-D-11-00130.1>.
- Cheng, L., M. English, and R. Wong, 1985: Hailstone size distributions and their relationship to storm thermodynamics. *J. Climate Appl. Meteor.*, **24**, 1059–1067, [https://doi.org/10.1175/1520-0450\(1985\)024<1059:HSDATR>2.0.CO;2](https://doi.org/10.1175/1520-0450(1985)024<1059:HSDATR>2.0.CO;2).
- 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>.
- , M. J. Pavolonis, J. M. Sieglaff, and A. K. Heidinger, 2013: Evolution of severe and nonsevere convection inferred from GOES-derived cloud properties. *J. Appl. Meteor. Climatol.*, **52**, 2009–2023, <https://doi.org/10.1175/JAMC-D-12-0330.1>.
- , —, —, and D. T. Lindsey, 2014: An empirical model for assessing the severe weather potential of developing convection. *Wea. Forecasting*, **29**, 639–653, <https://doi.org/10.1175/WAF-D-13-00113.1>.
- Cook, B., 1958: Hail determination by radar analysis. *Mon. Wea. Rev.*, **86**, 435–438, [https://doi.org/10.1175/1520-0493\(1958\)086<0435:HDBRA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1958)086<0435:HDBRA>2.0.CO;2).
- Crum, T. D., and R. L. Alberty, 1993: The WSR-88D and the WSR-88D operational support facility. *Bull. Amer. Meteor. Soc.*, **74**, 1669–1688, [https://doi.org/10.1175/1520-0477\(1993\)074<1669:TWATWO>2.0.CO;2](https://doi.org/10.1175/1520-0477(1993)074<1669:TWATWO>2.0.CO;2).
- Deierling, W., and W. A. Petersen, 2008: Total lightning activity as an indicator of updraft characteristics. *J. Geophys. Res.*, **113**, D16210, <https://doi.org/10.1029/2007JD009598>.
- Depue, T. K., P. C. Kennedy, and S. A. Rutledge, 2007: Performance of the hail differential reflectivity (H_{DR}) polarimetric radar hail indicator. *J. Appl. Meteor. Climatol.*, **46**, 1290–1301, <https://doi.org/10.1175/JAM2529.1>.
- Donavon, R. A., and K. A. Jungbluth, 2007: Evaluation of a technique for radar identification of large hail across the upper

- Midwest and central plains of the United States. *Wea. Forecasting*, **22**, 244–254, <https://doi.org/10.1175/WAF1008.1>.
- Doswell, C. A. I., H. E. Brooks, and M. P. Kay, 2005: Climatological estimates of daily local nontornadic severe thunderstorm probability for the United States. *Wea. Forecasting*, **20**, 577–595, <https://doi.org/10.1175/WAF866.1>.
- Doviak, R. J., and D. S. Zrnić, 1993: *Doppler Radar and Weather Observations*. 2nd ed. Dover Publications, 562 pp.
- Dworak, R., K. Bedka, J. Brunner, and W. Feltz, 2012: Comparison between GOES-12 overshooting-top detections, WSR-88D radar reflectivity, and severe storm reports. *Wea. Forecasting*, **27**, 684–699, <https://doi.org/10.1175/WAF-D-11-00070.1>.
- 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](https://doi.org/10.1175/1520-0434(1998)013<0277:NCOHSW>2.0.CO;2).
- Elmore, K. L., 2011: The NSSL hydrometeor classification algorithm in winter surface precipitation: Evaluation and future development. *Wea. Forecasting*, **26**, 756–765, <https://doi.org/10.1175/WAF-D-10-05011.1>.
- 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>.
- Ferraro, R., J. Beauchamp, D. Cecil, and G. Heymsfield, 2015: A prototype hail detection algorithm and hail climatology developed with the Advanced Microwave Sounding Unit (AMSU). *Atmos. Res.*, **163**, 24–35, <https://doi.org/10.1016/j.atmosres.2014.08.010>.
- Fraile, R., A. Castro, and J. Sánchez, 1992: Analysis of hailstone size distributions from a hailpad network. *Atmos. Res.*, **28**, 311–326, [https://doi.org/10.1016/0169-8095\(92\)90015-3](https://doi.org/10.1016/0169-8095(92)90015-3).
- Goodman, S. J., and Coauthors, 2013: The GOES-R Geostationary Lightning Mapper (GLM). *Atmos. Res.*, **125–126**, 34–49, <https://doi.org/10.1016/j.atmosres.2013.01.006>.
- Gravelle, C. M., J. R. Mecikalski, W. E. Line, K. M. Bedka, R. A. Petersen, J. M. Sieglaff, G. T. Stano, and S. J. Goodman, 2016: Demonstration of a GOES-R satellite convective toolkit to “bridge the gap” between severe weather watches and warnings: An example from the 20 May 2013 Moore, Oklahoma, tornado outbreak. *Bull. Amer. Meteor. Soc.*, **97**, 69–84, <https://doi.org/10.1175/BAMS-D-14-00054.1>.
- 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., [http://www.willisre.com/Media_Room/Press_Releases_\(Browse_All\)/2017/WillisRe_Impact_of_ENSO_on_US_Tornado_and_Hail_frequencies_Final.pdf](http://www.willisre.com/Media_Room/Press_Releases_(Browse_All)/2017/WillisRe_Impact_of_ENSO_on_US_Tornado_and_Hail_frequencies_Final.pdf).
- 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>.
- Helmus, J., and S. Collis, 2016: The Python ARM Radar Toolkit (Py-ART), a library for working with weather radar data in the Python programming language. *J. Open Res. Software*, **4**, e25, <https://doi.org/10.5334/jors.119>.
- Herzogh, P. H., and A. R. Jameson, 1992: Observing precipitation through dual-polarization radar measurements. *Bull. Amer. Meteor. Soc.*, **73**, 1365–1376, [https://doi.org/10.1175/1520-0477\(1992\)073<1365:OPTDPR>2.0.CO;2](https://doi.org/10.1175/1520-0477(1992)073<1365:OPTDPR>2.0.CO;2).
- Holleman, I., H. Wessels, J. Onvlee, and S. Barlag, 2000: Development of a hail-detection-product: S10: Deep convection. *Phys. Chem. Earth*, **25B**, 1293–1297, [https://doi.org/10.1016/S1464-1909\(00\)00197-0](https://doi.org/10.1016/S1464-1909(00)00197-0).
- Homeyer, C. R., 2014: Formation of the enhanced-V infrared cloud-top feature from high-resolution three-dimensional radar observations. *J. Atmos. Sci.*, **71**, 332–348, <https://doi.org/10.1175/JAS-D-13-079.1>.
- , and M. R. Kumjian, 2015: Microphysical characteristics of overshooting convection from polarimetric radar observations. *J. Atmos. Sci.*, **72**, 870–891, <https://doi.org/10.1175/JAS-D-13-0388.1>.
- , and K. P. Bowman, 2017: Algorithm description document for version 3.1 of the three-dimensional gridded NEXRAD WSR-88D radar (GridRad) dataset. Tech. Rep., 23 pp., <http://gridrad.org/pdf/GridRad-v3.1-Algorithm-Description.pdf>.
- , J. D. McAuliffe, and K. M. Bedka, 2017: On the development of above-anvil cirrus plumes in extratropical convection. *J. Atmos. Sci.*, **74**, 1617–1633, <https://doi.org/10.1175/JAS-D-16-0269.1>.
- Hubbert, J., V. Bringi, L. Carey, and S. Bolen, 1998: CSU-CHILL polarimetric radar measurements from a severe hail storm in eastern Colorado. *J. Appl. Meteor.*, **37**, 749–775, [https://doi.org/10.1175/1520-0450\(1998\)037<0749:CCPRMF>2.0.CO;2](https://doi.org/10.1175/1520-0450(1998)037<0749:CCPRMF>2.0.CO;2).
- Kelly, D. L., J. T. Schaefer, and C. A. Doswell, 1985: Climatology of nontornadic severe thunderstorm events in the United States. *Mon. Wea. Rev.*, **113**, 1997–2014, [https://doi.org/10.1175/1520-0493\(1985\)113<1997:CONSTE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1985)113<1997:CONSTE>2.0.CO;2).
- Kumjian, M. R., 2013a: Principles and applications of dual-polarization weather radar. Part I: Description of the polarimetric radar variables. *J. Oper. Meteor.*, **1**, 226–242, <https://doi.org/10.15191/nwajom.2013.0119>.
- , 2013b: Principles and applications of dual-polarization weather radar. Part II: Warm and cold season applications. *J. Oper. Meteor.*, **1**, 243–264, <https://doi.org/10.15191/nwajom.2013.0120>.
- , 2013c: Principles and applications of dual-polarization weather radar. Part III: Artifacts. *J. Oper. Meteor.*, **1**, 265–274, <https://doi.org/10.15191/nwajom.2013.0121>.
- Line, W. E., T. J. Schmit, D. T. Lindsey, and S. J. Goodman, 2016: Use of geostationary super rapid scan satellite imagery by the Storm Prediction Center. *Wea. Forecasting*, **31**, 483–494, <https://doi.org/10.1175/WAF-D-15-0135.1>.
- Liu, C., and S. Heckman, 2010: The application of total lightning detection and cell tracking for severe weather prediction. *WMO Technical Conf. on Meteorology and Environmental Instruments and Methods of Observation*, World Meteorological Organization, https://www.wmo.int/pages/prog/www/IMOP/publications/IOM-104_TECO-2010/P2_7_Heckman_USA.pdf.
- López, L., and J. Sanchez, 2009: Discriminant methods for radar detection of hail. *Atmos. Res.*, **93**, 358–368, <https://doi.org/10.1016/j.atmosres.2008.09.028>.
- Lukach, M., L. Foresti, O. Giot, and L. Delobbe, 2017: Estimating the occurrence and severity of hail based on 10 years of observations from weather radar in Belgium. *Meteor. Appl.*, **24**, 250–259, <https://doi.org/10.1002/met.1623>.
- Mahale, V. N., G. Zhang, and M. Xue, 2014: Fuzzy logic classification of s-band polarimetric radar echoes to identify three-body scattering and improve data quality. *J. Appl. Meteor.*, **53**, 2017–2033, <https://doi.org/10.1175/JAMC-D-13-0358.1>.
- Marzban, C., and A. Witt, 2001: A Bayesian neural network for severe-hail size prediction. *Wea. Forecasting*, **16**, 600–610, [https://doi.org/10.1175/1520-0434\(2001\)016<0600:ABNFS>2.0.CO;2](https://doi.org/10.1175/1520-0434(2001)016<0600:ABNFS>2.0.CO;2).
- Mather, G. K., D. Treddenick, and R. Parsons, 1976: An observed relationship between the height of the 45 dBZ contours in storm profiles and surface hail reports. *J. Appl. Meteor.*, **15**,

- 1336–1340, [https://doi.org/10.1175/1520-0450\(1976\)015<1336:AORBTH>2.0.CO;2](https://doi.org/10.1175/1520-0450(1976)015<1336:AORBTH>2.0.CO;2).
- McCann, D. W., 1983: The enhanced-V: A satellite observable severe storm signature. *Mon. Wea. Rev.*, **111**, 887–894, [https://doi.org/10.1175/1520-0493\(1983\)111<0887:TEVASO>2.0.CO;2](https://doi.org/10.1175/1520-0493(1983)111<0887:TEVASO>2.0.CO;2).
- Menzel, W. P., and J. F. W. Purdom, 1994: Introducing GOES-I: The first of a new generation of geostationary operational environmental satellites. *Bull. Amer. Meteor. Soc.*, **75**, 757–782, [https://doi.org/10.1175/1520-0477\(1994\)075<0757:IGITFO>2.0.CO;2](https://doi.org/10.1175/1520-0477(1994)075<0757:IGITFO>2.0.CO;2).
- Mroz, K., A. Battaglia, T. J. Lang, D. J. Cecil, S. Tanelli, and F. Tridon, 2017: Hail-detection algorithm for the GPM Core Observatory satellite sensors. *J. Appl. Meteor. Climatol.*, **56**, 1939–1957, <https://doi.org/10.1175/JAMC-D-16-0368.1>.
- , —, —, S. Tanelli, and G. F. Sacco, 2018: Global precipitation measuring dual-frequency precipitation radar observations of hailstorm vertical structure: Current capabilities and drawbacks. *J. Appl. Meteor. Climatol.*, **57**, 2161–2178, <https://doi.org/10.1175/JAMC-D-18-0020.1>.
- Ni, X., C. Liu, D. J. Cecil, and Q. Zhang, 2017: On the detection of hail using satellite passive microwave radiometers and precipitation radar. *J. Appl. Meteor. Climatol.*, **56**, 2693–2709, <https://doi.org/10.1175/JAMC-D-17-0065.1>.
- 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/NCEI, 2013: Satellite data. NOAA/NCEI, accessed August 2016–July 2018, <https://www.ncdc.noaa.gov/data-access/satellite-data>.
- , 2014: Storm events database, version 3. NOAA/NCEI, accessed August 2016–July 2018, <https://www.ncdc.noaa.gov/stormevents/>.
- NOAA/NCEP/ESRL, 2012: Rapid Refresh (RAP). NOAA/NCEP/ESRL, accessed August 2016–July 2018, <https://rapidrefresh.noaa.gov/>.
- NOAA/NESDIS/NCEI, 1991: NOAA Next Generation Radar (NEXRAD) Level 2 Base Data. NOAA/NESDIS/NCEI, accessed August 2016–July 2018, <https://doi.org/10.7289/V5W9574V>.
- NOAA/NWS/SPC, 1955: Severe Weather Database. NOAA/NWS/SPC, accessed August 2016–July 2018, <https://www.spc.noaa.gov/wcm/#data>.
- Ortega, K. L., 2018: Evaluating multi-radar, multi-sensor products for surface hailfall diagnosis. *Electron. J. Severe Storms Meteor.*, **13**, <http://www.ejssm.org/ojs/index.php/ejssm/article/view/163/113>.
- , T. M. Smith, G. J. Stumpf, J. Hocker, and L. Lopez, 2005: A comparison of multi-sensor hail diagnosis techniques. *21st Conf. on Interactive Information Processing Systems*, San Diego, CA, Amer. Meteor. Soc., P1.11, <http://ams.confex.com/ams/pdfpapers/87640.pdf>.
- , —, 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>.
- Park, H. S., A. V. Ryzhkov, D. S. Zrnić, and K.-E. Kim, 2009: The hydrometeor classification algorithm for the polarimetric WSR-88D: Description and application to an MCS. *Wea. Forecasting*, **24**, 730–748, <https://doi.org/10.1175/2008WAF2222205.1>.
- Petrocchi, P. J., 1982: Automatic detection of hail by radar. Air Force Geophysics Laboratory Tech. Rep. AFGLTR-82-0277 (Environ. Res. Paper 796), 33 pp., <https://apps.dtic.mil/dtic/tr/fulltext/u2/a130078.pdf>.
- Picca, J., and A. Ryzhkov, 2012: A dual-wavelength polarimetric analysis of the 16 May 2010 Oklahoma City extreme hailstorm. *Mon. Wea. Rev.*, **140**, 1385–1403, <https://doi.org/10.1175/MWR-D-11-00112.1>.
- Punge, H. J., K. M. Bedka, M. Kunz, and A. Werner, 2014: A new physically based stochastic event catalog for hail in Europe. *Nat. Hazards*, **73**, 1625–1645, <https://doi.org/10.1007/s11069-014-1161-0>.
- Ryzhkov, A. V., S. E. Giangrande, V. Melnikov, and T. Schuur, 2005a: Calibration issues of dual-polarization radar measurements. *J. Atmos. Oceanic Technol.*, **22**, 1138–1155, <https://doi.org/10.1175/JTECH1772.1>.
- , T. J. Schuur, D. W. Burgess, P. L. Heinselman, S. E. Giangrande, and D. S. Zrnić, 2005b: The Joint Polarization Experiment: Polarimetric rainfall measurements and hydrometeor classification. *J. Atmos. Oceanic Technol.*, **86**, 809–824, <https://doi.org/10.1175/BAMS-86-6-809>.
- , M. R. Kumjian, S. M. Ganson, and P. Zhang, 2013: 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>.
- Saltikoff, E., J.-P. Tuovinen, J. Kotro, T. Kuitunen, and H. Hohti, 2010: A climatological comparison of radar and ground observations of hail in Finland. *J. Appl. Meteor. Climatol.*, **49**, 101–114, <https://doi.org/10.1175/2009JAMC2116.1>.
- San Ambrosio, I., F. Martín, and F. Elizaga, 2007: Development and behaviour of a radar-based operational tool for hailstorms identification. *Atmos. Res.*, **83**, 306–314, <https://doi.org/10.1016/j.atmosres.2005.08.012>.
- Sandmæl, T. N., 2017: An evaluation of radar- and satellite-data based products to discriminate between tornadic and non-tornadic storms. M.S. thesis, School of Meteorology, University of Oklahoma, 98 pp., <https://hdl.handle.net/11244/52775>.
- Schaefer, J., J. J. Levit, S. J. Weiss, and D. W. McCarthy, 2004: The frequency of large hail over the contiguous United States. *14th Conf. on Applied Climatology*, Seattle, WA, Amer. Meteor. Soc., 3.3, <https://ams.confex.com/ams/pdfpapers/69834.pdf>.
- Schmit, T. J., and Coauthors, 2013: GOES-14 super rapid scan operations to prepare for GOES-R. *J. Appl. Remote Sens.*, **7**, <https://doi.org/10.1117/1.JRS.7.073462>.
- , M. M. Gunshor, W. P. Menzel, J. J. Gurka, J. Li, and A. S. Bachmeier, 2005: Introducing the Next-Generation Advanced Baseline Imager on GOES-R. *Bull. Amer. Meteor. Soc.*, **86**, 1079–1096, <https://doi.org/10.1175/BAMS-86-8-1079>.
- Schultz, C. J., W. A. Petersen, and L. D. Carey, 2009: Preliminary development and evaluation of lightning jump algorithms for the real-time detection of severe weather. *J. Appl. Meteor. Climatol.*, **48**, 2543–2563, <https://doi.org/10.1175/2009JAMC2237.1>.
- , L. D. Carey, E. V. Schultz, and R. J. Blakeslee, 2017: Kinematic and microphysical significance of lightning jump versus nonjump increases in total flash rate. *Wea. Forecasting*, **32**, 275–288, <https://doi.org/10.1175/WAF-D-15-0175.1>.
- Setvák, M., and Coauthors, 2010: Satellite-observed cold-ring-shaped features atop deep convective clouds. *Atmos. Res.*, **97**, 80–96, <https://doi.org/10.1016/j.atmosres.2010.03.009>.

- Skripniková, K., and D. Řezáčová, 2014: Radar-based hail detection. *Atmos. Res.*, **144**, 175–185, <https://doi.org/10.1016/j.atmosres.2013.06.002>.
- Starzec, M., C. R. Homeyer, and G. L. Mullendore, 2017: Storm labeling in three dimensions (SL3D): A volumetric radar echo and dual-polarization updraft classification algorithm. *Mon. Wea. Rev.*, **145**, 1127–1145, <https://doi.org/10.1175/MWR-D-16-0089.1>.
- Straka, J. M., D. S. Zrnić, and A. V. Ryzhkov, 2000: Bulk hydrometeor classification and quantification using polarimetric radar data: Synthesis of relations. *J. Appl. Meteor.*, **39**, 1341–1372, [https://doi.org/10.1175/1520-0450\(2000\)039<1341:BHCAQU>2.0.CO;2](https://doi.org/10.1175/1520-0450(2000)039<1341:BHCAQU>2.0.CO;2).
- Trapp, R. J., D. M. Wheatley, N. T. Atkins, R. W. Przybylinski, and R. Wolf, 2006: Buyer beware: Some words of caution on the use of severe wind reports in postevent assessment and research. *Wea. Forecasting*, **21**, 408–415, <https://doi.org/10.1175/WAF925.1>.
- University of Wisconsin–Madison Space Science and Engineering Center, 2011: Space weather data. Accessed August 2016–July 2018, <https://www.ssec.wisc.edu/data/geo-list>.
- Vivekanandan, J., D. S. Zrnić, S. M. Ellis, R. Oye, A. V. Ryzhkov, and J. Straka, 1999: Cloud microphysics retrieval using S-band dual-polarization radar measurements. *Bull. Amer. Meteor. Soc.*, **80**, 381–388, [https://doi.org/10.1175/1520-0477\(1999\)080<0381:CMRUSB>2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080<0381:CMRUSB>2.0.CO;2).
- Waldvogel, A., B. Federer, and P. Grimm, 1979: Criteria for the detection of hail cells. *J. Appl. Meteor.*, **18**, 1521–1525, [https://doi.org/10.1175/1520-0450\(1979\)018<1521:CFTDOH>2.0.CO;2](https://doi.org/10.1175/1520-0450(1979)018<1521:CFTDOH>2.0.CO;2).
- Wang, P., J. Shi, J. Hou, and Y. Hu, 2018: The identification of hail storms in the early stage using time series analysis. *J. Geophys. Res. Atmos.*, **123**, 929–947, <https://doi.org/10.1002/2017JD027449>.
- Williams, E., and Coauthors, 1999: The behavior of total lightning activity in severe Florida thunderstorms. *Atmos. Res.*, **51**, 245–265, [https://doi.org/10.1016/S0169-8095\(99\)00011-3](https://doi.org/10.1016/S0169-8095(99)00011-3).
- Witt, A., M. D. Eilts, G. J. Stumpf, J. T. Johnson, E. D. W. Mitchell, and K. W. Thomas, 1998a: An enhanced hail detection algorithm for the WSR-88D. *Wea. Forecasting*, **13**, 286–303, [https://doi.org/10.1175/1520-0434\(1998\)013<0286:AEHDAF>2.0.CO;2](https://doi.org/10.1175/1520-0434(1998)013<0286:AEHDAF>2.0.CO;2).
- , —, —, E. D. W. Mitchell, J. T. Johnson, and K. W. Thomas, 1998b: Evaluating the performance of WSR-88D severe storm detection algorithms. *Wea. Forecasting*, **13**, 513–518, [https://doi.org/10.1175/1520-0434\(1998\)013<0513:ETPOWS>2.0.CO;2](https://doi.org/10.1175/1520-0434(1998)013<0513:ETPOWS>2.0.CO;2).
- , D. Burgess, A. Seimon, J. T. Allen, J. Snyder, and H. 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>.
- Zrnić, D. S., and A. V. Ryzhkov, 1999: Polarimetry for weather surveillance radars. *Wea. Forecasting*, **80**, 389–406, [https://doi.org/10.1175/1520-0477\(1999\)080<0389:PFWSR>2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080<0389:PFWSR>2.0.CO;2).

CORRIGENDUM

ELISA M. MURILLO^a AND CAMERON R. HOMEYER^a

^a*School of Meteorology, University of Oklahoma, Norman, Oklahoma*

(Manuscript received 11 December 2020, in final form 18 February 2021)

Murillo and Homeyer (2019) introduced revised fits between the radar maximum expected size of hail (MESH) parameter and the severe hail index (SHI). However, the coefficients included in the text for these fits were incorrect. All of the analyses presented therein used the correct coefficients. The incorrect equations as given in section 3c of Murillo and Homeyer (2019) are

$$\text{MESH}_{75} = 16.566(\text{SHI})^{0.181} \quad \text{and} \quad (15)$$

$$\text{MESH}_{95} = 17.270(\text{SHI})^{0.272}. \quad (16)$$

The corrected equations are

$$\text{MESH}_{75} = 15.096(\text{SHI})^{0.206} \quad \text{and} \quad (15)$$

$$\text{MESH}_{95} = 22.157(\text{SHI})^{0.212}. \quad (16)$$

Acknowledgments. Thanks are given to Nathan Wendt from the NOAA Storm Prediction Center for bringing this error to our attention.

REFERENCE

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>.

Corresponding author: Elisa Murillo, murillem@ou.edu

DOI: 10.1175/JAMC-D-20-0271.1

© 2021 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) (www.ametsoc.org/PUBSReuseLicenses).