Blair, S. F., D. R. Deroche, J. M. Boustead, J. W. Leighton, B. L. Barjenbruch, and W. P. Gargan, 2011: A radar-based assessment of the detectability of giant hail. *Electronic J. Severe Storms Meteor.,* **6** (7), 1–30.

Electronic Journal of
SEVERE STORMS METEOROLOGY

A Radar-Based Assessment of the Detectability of Giant Hail

SCOTT F. BLAIR,* DEREK R. DEROCHE,+ JOSHUA M. BOUSTEAD,#
JARED W. LEIGHTON,* BRIAN L. BARJENBRUCH,* AND WILLIAM P. GARGAN*

**NOAA/NWS, Weather Forecast Office, Topeka, Kansas*
+NOAA/NWS, Weather Forecast Office, Pleasant Hill, Missouri
#NOAA/NWS, Weather Forecast Office, Valley, Nebraska

(Submitted 26 March 2011; in final form 27 November 2011)

ABSTRACT

The occurrence of giant hail, defined as hail ≥102 mm (4.00 in) in diameter, is a relatively rare phenomenon, accounting for <1% of all hail reports in the United States. Despite the infrequent nature of these events, hail of this magnitude has the potential to cause extreme damage to property and a substantial threat to exposed life. The short-term prediction of these events has been challenging. For giant hail since 2005, only 7% of convective warnings and severe-weather statements issued by the National Weather Service (NWS) accurately predicted a maximum hail size ≥102 mm prior to the report, with an average underestimated size error of 55.6 mm (2.19 in).

The objectives of this study are to determine the detectability of giant hail in convective storms and to improve advanced recognition of these events during NWS warning operations. A total of 568 giant-hail reports, gathered over a 15-y period from 1 January 1995 through 31 December 2009 throughout the contiguous United States, served as the primary database for the research. Weather Surveillance Radar-1988 Doppler (WSR-88D) data and North American Regional Reanalysis (NARR) environmental data were collected for each case. Several radar signatures were examined to assess their utility in discriminating storms most favorable for giant hail. It was found that 99% of the storms were supercells with well-organized structure. Giant-hail producing storms were characterized by median values of rotational velocities of 24 m s^{-1} (47 kt), storm-top divergence magnitudes of 72 m s^{-1} (140 kt), and 50-dB*Z* and 60-dB*Z* echo heights of 13 100 m (43 000 ft) and 10 600 m (34 800 ft) respectively. Vertically integrated liquid water (VIL)-based products, maximum reflectivity within the storm, and reflectivity within the preferred hail-growth zone showed little to no skill in discriminating between giant hail and smaller hail sizes.

1. Introduction

One of the biggest challenges to the operational community remains the ability to accurately predict specific maximum hail sizes in real-time warning operations. The National Weather Service (NWS) defines "severe hail" as a hailstone with a diameter ≥25.4 mm (1.00 in) and uses this size as the threshold to issue severe thunderstorm warnings. The term "significant hail" has become synonymous with hail ≥51 mm (2.00 in) in diameter (Hales 1998), and the Storm Prediction Center (SPC) explicitly forecasts the potential for hail of this magnitude in the day-1 convective outlook product. "Giant hail", defined by Knight and Knight (2001) as hail ≥102 mm (4.00 in) in diameter, is a relatively rare phenomenon, but has the potential for extreme economic and societal impacts. One of the more notable examples of the consequences of giant hail was illustrated on 5 May 1995 in

Corresponding author address: Scott F. Blair,
NOAA / National Weather Service,
1116 NE Strait Ave., Topeka, KS 66616.
E-mail: scott.blair@noaa.gov

Fort Worth, TX, when hail >115 mm (4.50 in) struck portions of the city. Over 100 people were treated for injuries, four critically, with damage estimates exceeding $2 billion (NCDC 1995). Giant hail also was responsible for the last known hail fatality in the United States when a young man was struck in the head on 28 March 2000, also in Fort Worth, TX (NCDC 2000).

Radar is the essential observational tool used to detect the presence of hail in convective storms. Several radar-based techniques have been evaluated in the past two decades for their ability to detect severe hail. Witt and Nelson (1991) demonstrated some skill in correlating a storm's upper-level divergent outflow to maximum hail size. Paxton and Shepherd (1993) and Amburn and Wolf (1997) examined the use of various forms of vertically integrated liquid water (VIL) to forecast hail size. In a robust nationwide study of hail sizes compared to VIL values, Edwards and Thompson (1998) concluded that VIL-based methods alone offered little skill in estimating hail size. Lemon (1998) suggested that the three-body scatter signature was a strong indicator of the presence of large hail. More recently, Donavon and Jungbluth (2007) correlated hail size to a relationship between the environmental melting level and the 50-dB*Z* reflectivity height, in weakly to moderately sheared environments. While numerous studies have examined hail detection, and to a lesser degree hail size, very little work has been published with the explicit purpose of detecting storms capable of producing giant hail.

This study evaluates several methods, including radar-based signatures, to aid in the real-time detection of giant hail. These signatures are expected to improve confidence for warning meteorologists when forecasting the maximum hail size in well-organized convection. The paper begins with an overview of the giant-hail database and the radar-based methods used to examine each event. A review of the recent performance of giant-hail detection is discussed in section 3. Section 4 provides a brief climatological overview; with section 5 summarizing the radar-based results found with giant-hail producing storms. An analogous radar analysis then is performed on smaller hail sizes and with high-resolution, ground-truth hail data, which are then compared to the giant-hail results in section 6. Concluding remarks follow in section 7.

2. Data and methodology

A total of 568 giant-hail reports originating from *Storm Data* were incorporated into the study (Fig. 1). An additional 28 reports of giant hail from the Severe Hazards Analysis and Verification Experiment (SHAVE, Ortega et al. 2009) were analyzed for a comparison study. The domain encompassed the entire contiguous United States from 1 January 1995 through 31 December 2009. The central portion of the country, generally east of the Rocky Mountains and west of the Mississippi River, featured the greatest concentration of giant hail.

The 15-y period was selected to benefit from the increased number of hail reports in the severe weather database. Many factors contributed to this reporting increase, including well-trained storm spotter networks, storm chasers, advances in technology, and improved public awareness (Doswell et al. 1999; Brooks and Doswell 2002; Verbout et al. 2006). The presence of the WSR-88D network coincides with the study period, as does an increased emphasis on NWS warning verification, both of which contributed to the upward trend of severe weather reports (Serafin and Wilson 2000; Simmons and Sutter 2005).

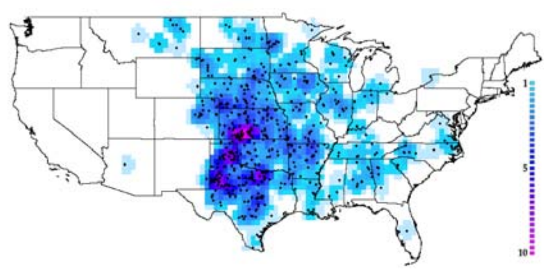

Figure 1: Giant-hail reports (black dots) used in the study from 1 January 1995 through 31 December 2009. The point-density background map output is on a 0.5° × 0.5° gridded display with a 1° radius per report. The maximum value (pink) represents 10 reports. *Click image to enlarge.*

Storm Data reports were used to build the giant-hail database (NCDC 1995–2009). There is an inherent set of limitations concerning hail reports within *Storm Data*. For instance, while the largest hailstone reported in a given storm is used, it cannot be said with confidence that the report represents the largest one to fall from that storm. In many cases, it is suspected that *Storm Data* under-represents the maximum-sized hail of the storm. The largest hailstone, especially of giant-hail size, frequently is not identified by traditional NWS verification practices, which is a result of several factors. During NWS severe weather operations, the available workforce and resources are insufficient for a thorough real-time investigation into the hail-fall character of a storm. Furthermore, there usually is not an emphasis on obtaining the maximum-sized hail in a storm during post-storm verification. Amburn and Wolf (1997) highlighted several of these issues and concluded that verification practices are designed to efficiently verify warnings, not to satisfy scientific studies.

Research efforts examining smaller hail sizes using the low-resolution reports in *Storm Data* are potentially handicapped when drawing conclusions of radar-based signatures due to a wider range of hail sizes that may go unreported. This study mitigated some of these potential size limitations by investigating hail reports at the upper 1% of documented hail sizes in the United States. Additionally, this study used high-resolution hail reports from SHAVE that better identified the hail-fall character in storms. Both *Storm Data* and SHAVE data are used in a comparison study between different hail sizes in section 6.

There is also some uncertainty with the true size of reported softball-sized hail. Jewell and Brimelow (2009) noted that the size of the softball varies in recreational play between the first-ever softball tournament at the 1933 World's Fair in Chicago (114 mm, 4.50 in) and many men's and women's softball leagues today (97 mm, 3.80 in / 89 mm, 3.50 in, respectively). Most NWS storm spotter classes continue to train observers to correlate 114 mm hail to the Chicago-sized softball. For the purpose of this study, errors from inconsistent softball-size reports would account for only 5–13 mm (0.2–0.5 in) below the classification of giant hail.

Lastly, reporting errors by the observer or second-hand relayed reports may contribute to errors within *Storm Data*. A tendency for observers to report estimated hail sizes, such as comparing hail to commonly-sized objects instead of measuring the explicit size of hail, also may result in inaccurate size information in the database (Jewell and Brimelow 2009; Baumgardt 2011).

a. Recent detection performance

Hail-size verification from giant-hail events was performed on SPC convective watches and NWS warnings from 2005–2009 throughout the study domain. Archived watch and warning information was acquired through the SPC online product archive (available via http://www.spc.noaa.gov/products/watch/) and the Iowa Environmental Mesonet application (http://mesonet.agron.iastate.edu/cow/) respectively. Only the watches, severe thunderstorm warnings, and tornado warnings issued prior to a giant-hail report were included in the database. Warnings that did not include a specific hail size were omitted. In cases when updated warning statements increased the forecasted hail size in a storm prior to the hail report, then these data were used in place of the initial warning size.

b. Giant-hail climatology

A monthly and hourly analysis of giant hail across the contiguous United States was conducted from 1995 through 2009. Giant-hail reports in *Storm Data* were separated into unique events. Each unique event was defined as a giant-hail report >100 km from surrounding giant-hail reports and/or reports separated by more than 3 h. This process was meant to mitigate duplicate or multiple hail reports in the database from the same storm, improving the statistical analysis. Following the quality control process, a total of 527 unique events were used in the climatological analysis.

c. Environmental data

North American regional reanalysis (NARR, after Mesinger et al. 2006) data were obtained for each giant-hail event from the NCDC National Operational Model Archive & Distribution System website (http://nomads.ncdc.noaa.gov/). The 3-hourly NARR dataset has a 32-km horizontal grid spacing, with 45 layers in the vertical. NARR data were obtained for the model run closest to the observed hail time, then analyzed using the General Meteorological Package (GEMPAK; desJardines et al. 1991) to obtain the vertical temperature and wind profile for each event. The temperatures with each NARR sounding were interpolated linearly to acquire heights for the temperatures of 0° C, –10° C, –20° C, and –30° C.

d. Radar data

WSR-88D radar data were used to examine each storm near the time of giant-hail reports. The closest radar site to the report location was identified and data were downloaded from the NCDC Hierarchical Data Storage System website interface. Additional radar sites were used in conjunction with the closest site to the report in order to collect mid-storm and near storm-summit data. Inherent limitations exist in acquiring the vertical characteristics of upper-storm information, depending on the selected volume coverage pattern (VCP) scanning strategy (Maddox et al. 1999; Brown et al. 2000). As radar range increases, beam broadening frequently will result in acquiring lower reflectivities at a given elevation AGL due to averaging in the sample volume. Furthermore, there is some ambiguity in the measurement of height sampled by the radar. The height sampled in each elevation scan is the radar beam center, resulting in uncertainty that is usually equal to half of the beam diameter (Howard et al. 1997; Maddox et al. 1999). Therefore, some inherent uncertainty existed with the implied precision of radar height measurements, especially when convection existed at farther ranges from the radar or selected VCP scan strategies allowed for large gaps within the upper storm region. Section 2d-1 discusses the procedures that mitigated some of these errors.

The time and location of each hail report in *Storm Data* were quality controlled. It was found 24% of the report times failed to coincide with any substantial reflectivity, with the storm downstream of the report location. This finding is similar to a previous study by Witt et al. (1998a) that found 29% of hail report times in *Storm Data* did not correlate with a storm. A similar correction scheme to Witt et al. (1998a) was used, by temporally adjusting the report time when the 0.5° reflectivity echo ≥50 dB*Z* aligned with the report location. In cases when the report time or location was irresolvable, the report was eliminated from the radar database.

Radar imagery were examined for each giant-hail report 15 min prior to and 5 min after the quality controlled report time. This provided an average of four full volume scans of radar data, depending on the scanning strategy. Previous radar-based hail studies have incorporated similar time frame schemes (Amburn and Wolf 1997; Witt et al. 1998b; Donavon and Jungbluth 2007).

Each of the 568 giant-hail reports was thoroughly examined using Gibson Ridge Level-II Analyst Edition (GR2AE) software. The software enabled the observation of radar reflectivity and velocity data for all available elevation levels, with cross-sections and three-dimensional capabilities for additional observations of storm structure. Data necessary for a thorough analysis to help discriminate signals unique to giant-hail producing storms were collected on the following:

1) Radar distance

The distance was measured from the WSR-88D Radar Data Acquisition unit to the report location. Reports >185 km from the radar were removed from the database due to limited vertical sampling and beam-broadening issues. Reports <30 km from the radar, when surrounding radars were unavailable for upper-level sampling, also were removed from the database due to limitations in vertical sampling.

2) Reflectivity profile

Reflectivity profiles of each giant-hail producing storm were examined during the 20 min period surrounding the report time. The maximum height AGL of the 50-dB*Z* echo during this period was used as a proxy to identify the time of peak convective intensity. Each slice within the radar volume scan was examined at this time to determine its respective maximum reflectivity and height. Reflectivity values then were interpolated linearly in between each available slice of observed data, in order to provide a more accurate assessment of reflectivity heights from the fixed elevation angles. No extrapolation was conducted below or above the available radar data. The maximum reflectivity heights AGL of 50, 55, 60, 65, 70, and 75 dB*Z* were obtained from the interpolated data. The maximum reflectivity within the entire storm column then was identified. Occasionally, the maximum columnar reflectivity occurred during a different volume scan than the maximum height of the 50-dB*Z* echo. In these instances, the volume scan containing the maximum 50-dB*Z* height was used in order to maintain consistency for comparison.

The NWS upgraded WSR-88Ds to super-resolution data in 2008, which changed the reflectivity data bin size from 1 km × 1° to 0.25 km × 0.5° associated with the split cut elevations—those scans at or below 1.5°. This upgrade increased the likelihood for higher reflectivity values to be identified than with the legacy resolution (Vogel et al. 2009).

3) Reflectivity at significant temperature levels

The peak reflectivity on the 0° C, –10° C, –20° C, and –30° C isosurfaces was obtained by vertically interpolating reflectivity values at the time of maximum convective intensity, as described in section 3b-2. The height of the specific temperature levels was obtained from the interpolated NARR environmental data.

4) Three-body scatter spike

All available radar volumes were analyzed during the 20-min period to identify the presence, or absence, of a three-body scatter spike (TBSS) using reflectivity and spectrum width. A TBSS is a radar microwave scattering artifact defined as a mid-level flare echo consisting of low reflectivities (<20 dB*Z*) and weak velocities approaching the radar, further described by Wilson and Reum (1998) and Lemon (1998).

5) Digital VIL

The maximum value of digital VIL (DVIL) associated with the convection responsible for giant hail during the 20 min period was recorded. DVIL is a radar-derived product that calculates the estimated amount of liquid water in a vertical column, further described by Greene and Clark (1972). DVIL is a 1° × 1 km polar grid product and does not cap reflectivity at 56 dB*Z*, providing an improved resolution and data quality, compared to the legacy 4 km × 4 km Cartesian-grid VIL product. GR2AE was used to display and record DVIL values instead of the NWS display system of Advanced Weather Interactive Processing System (AWIPS). AWIPS caps DVIL values at 80 kg m^{-2}, whereas GR2AE caps the values at 127 kg m^{-2}.

6) VIL density

The maximum value of VIL density (VILD) associated with the convection responsible for giant hail during the 20 min period was recorded. VILD is calculated by dividing the value of the DVIL by the height of the echo top, and is further discussed in Amburn and Wolf (1997). The VILD output for this study was on a 1° × 1 km polar grid with 256 data levels. GR2AE, which caps the values at 13 kg m^{-3}, was used for VILD.

7) Storm motion

The storm motion of the convection producing giant hail was calculated during the 20 min period using the GR2AE manual storm motion tool. Well-identifiable reflectivity structure regions, such as an inflow notch or pendant, at the 0.5° reflectivity level were used as the baselines to compute the average storm motion. These data were used to calculate the storm relative velocity values for each storm.

8) Maximum rotational velocity and height

The maximum rotational velocity (V_r) was identified within the convective column during the 20 min sample period. V_r was defined as:

$$V_r = (|V_{min}| + |V_{max}|) / 2 \quad (1)$$

where the distance between V_{min} and V_{max} was >1 km but did not exceed 15 km. The height AGL was determined from the center of the maximum V_r. Velocity values from suspected low-level tornadic circulations and dealiasing velocity errors were omitted from the sample. Occasionally, velocity data from each radar slice within the storm could not be identified due to range folding, and these cases were removed from the V_r database.

9) Storm type

Storm type was classified into the following categories: isolated supercell, embedded supercell, squall line, or other. The classification for each type was derived from similar works found in Browning (1964), Lemon (1977), Parker and Johnson (2000), Thompson et al. (2003), Trapp et al. (2005), and Hocker and Basara (2008).

An isolated supercell was defined as a convective storm displaying one or more unique radar characteristics such as a bounded weak echo region (BWER), inflow notch, or hook echo. A persistent mesocyclone or meso-anticyclone also must have been present throughout the entire 20-min sample period. The 50-dB*Z* reflectivity contour associated with the storm could not be connected to other convective updrafts.

Embedded supercells were defined similarly to the isolated supercell, with the inherent difference that the supercell shares a 50-dB*Z* reflectivity contour with additional convective updrafts.

Squall line storms were classified as those storms in a nearly continuous chain of reflectivity echoes that created a convective line or complex, shared a common leading edge of outflow, and did not obtain supercell characteristics as previously described.

The classification of 'other' was defined as any storm that failed to meet the criteria previously listed for supercells or squall lines.

10) Rotation orientation

The rotational characteristics of V_r were recorded as either cyclonic or anticyclonic for the updrafts responsible for producing giant hail to the best of the sampling capability of the radars used.

11) Maximum storm-top divergence

The maximum storm-top divergence (STD) at the upper storm levels was examined during the 20 min sample period. STD was defined as:

$$STD = (|V_{min}| + |V_{max}|) \quad (2)$$

with the largest reliable velocity difference used, and differential values measured within 30 km of each other, located near the storm top. Individual data bins with velocity values >15 m s^{-1} of the surrounding data were considered to be erroneous. Velocity values associated with reflectivity <10 dB*Z* were not used due to the potential of velocity errors. Spectrum width also was used to assess the quality of the velocity data. The inbound and outbound values corresponding to the maximum STD value were obtained from either the same or adjacent elevation angles, depending on the storm's proximity to the radar site. Limitations included gaps between elevation angles and beam broadening at long ranges from the radar. The methodology for determining STD is similar to that found in Witt and Nelson (1991).

3. Recent giant-hail forecast performance

Forecasting hail size, especially for giant hail during short-term operations of convective watches and warnings, has been particularly challenging, despite advances over the past decade in recognizing precursors to other severe storm hazards. The SPC issues convective watch products that forecast the maximum diameter of hail expected to occur within an area:

> THE NWS STORM PREDICTION CENTER HAS ISSUED A TORNADO WATCH FOR PORTIONS OF EASTERN KANSAS.
>
> TORNADOES…**HAIL TO 4.0 INCHES IN DIAMETER**…THUNDERSTORM WIND GUSTS TO 70 MPH…AND DANGEROUS LIGHTNING ARE POSSIBLE IN THESE AREAS.

Likewise, the NWS predicts the maximum diameter of hail expected in convective storms through the issuance of warnings and severe-weather statements:

> AT 800 PM CDT…NATIONAL WEATHER SERVICE DOPPLER RADAR INDICATED A SEVERE THUNDERSTORM **CAPABLE OF PRODUCING GRAPEFRUIT SIZE HAIL.**
>
> **LARGE DESTRUCTIVE HAIL UP TO FOUR INCHES** WILL OCCUR WITH THIS STORM. TAKE COVER NOW.
>
> LAT...LON 3966 9579 3966 9601 3977 9598
TIME...MOT...LOC 2315Z 272DEG 14KT
3973 9591 **WIND...HAIL 60MPH 4.00IN**

The forecast-performance review conducted during the 2005–2009 period found that when giant hail occurred, SPC watches (Fig. 2a) underestimated the maximum hail size by an average of 42.3 mm (1.66 in). NWS warnings (Fig. 2b) underestimated the maximum sized hail in giant-hail producing storms by an average of 55.6 mm (2.19 in). Only 8% of SPC watches and 7% of NWS warnings predicted a maximum diameter of hail ≥102 mm (4.00 in) prior to a giant-hail report, illustrating this unique forecasting challenge of high-end hail events. It is particularly surprising that 23% of NWS warnings forecasted penny to quarter-sized hail (19–26 mm; 0.75–1.00 in) during giant-hail occurrences. It is speculated this poor performance might be attributed to 1) a warning forecaster using a default minimum severe hail size when issuing a warning, 2) the lack of timely severe weather statements updating the hail size as a storm intensified, or 3) workload limitations. There was also an apparent tendency for NWS warnings to use a standard-sized diameter of golfball and baseball-sized hail (45 and 70 mm; 1.75 and 2.75 in) to convey a large hail threat, as Fig. 2b illustrates.

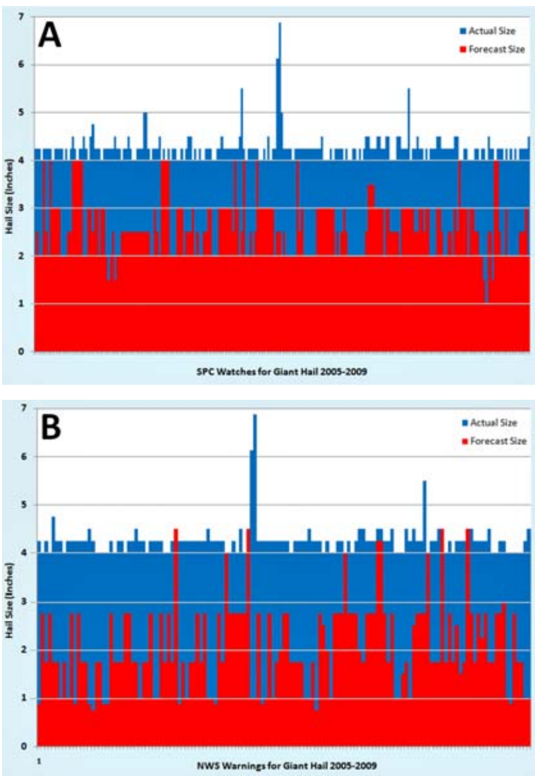

Figure 2: Giant-hail size verification of a) SPC watches (223 cases) and b) NWS warnings (137 cases) during 2005–2009 in the study domain. Forecast hail size (red, in) is compared with the actual hail size (blue, in). *Click images to enlarge.*

Roughly ¼ (26%) of giant-hail events in the database were associated with tornado warnings during the period. This result suggests that either the majority of giant-hail occurrences were not associated with ongoing tornadic events, or that the priority of reporting hail was low during ongoing tornadoes. The vast majority of tornado warnings contained no maximum forecasted hail information. Considering that large hail may be a more widespread and probable threat than a tornado in many situations, the inclusion of maximum hail-size information in tornado warnings could be considered a valuable addition to the warning information.

4. Giant-hail climatology

The monthly distribution of giant-hail reports is similar to other climatological studies performed on supercells, tornadoes, and hail (Brooks et al. 2003; Doswell et al. 2005; Bunkers et al. 2006). Figure 3a shows the monthly distribution of unique giant-hail events within the domain. The four months of April through July accounted for 80% of all giant-hail events, with 54% occurring in May and June alone.

A daily breakdown of unique events by hour is shown in Fig. 3b. The period between 2000 UTC (2 p.m. CST) and 0359 UTC (9:59 p.m. CST) contained 82% of giant-hail events, with a peak between 2200 UTC (4 p.m. CST) and 0159 UTC (7:59 p.m. CST) accounting for 55% of all unique cases. Conversely, the occurrence of giant hail during the morning hours is rare. Giant-hail events from 0600 UTC (12:00 a.m. CST) to 1759 UTC (11:59 a.m. CST) encompass only 9% of the cases.

The monthly and hourly unique event distributions of giant hail closely mirror the climatological period expected for supercell storms within the Great Plains. Previous studies have suggested that storms capable of producing significant hail (≥51 mm, 2 in) are almost exclusively supercells (Rasmussen and Blanchard 1998; Thompson et al. 2003; Doswell et al. 2005; Duda and Gallus 2010). The results herein, in conjunction with previous research, lend some overall guidance to the most probable period for the occurrence of giant-hail events.

5. Results: Radar-based signatures

Convective storms that produced giant hail overwhelmingly were supercells. Isolated and embedded supercells accounted for approximately 99% of all storms in the study. Approximately 88% of these giant-hail producing storms exhibited cyclonic rotation.

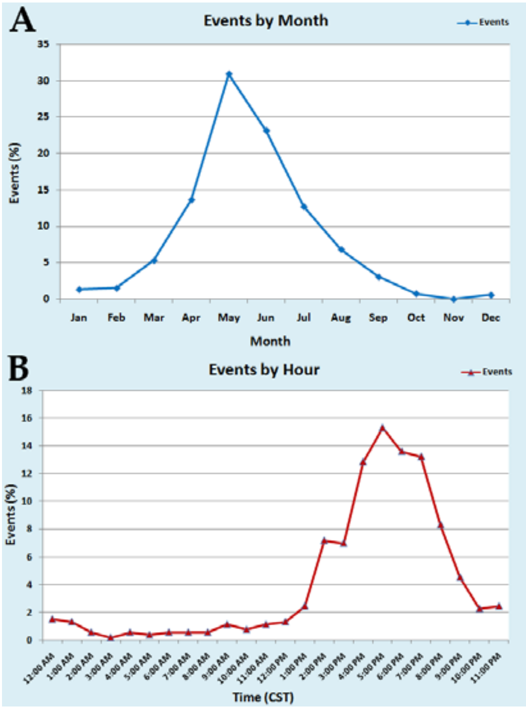

Figure 3: The a) monthly and b) hourly distributions (%) from 527 unique giant-hail events during 1995–2009. *Click images to enlarge.*

The few outliers were characterized as squall lines or other cases of isolated convection with no discernable rotation. Additionally, many giant-hail cases were associated with a BWER structure, similar to previous findings by Lemon (1978). This indicates a wide and strong updraft, typically containing large amounts of supercooled cloud water along the outskirts and BWER summit, supporting large-hail growth (Knight and Knight 2005).

The peak V_r magnitude of the mid-level mesocyclone was found to be 20–29 m s^{-1} (39–56 kt) within the 25th to 75th percentile with a median of 24 m s^{-1} (47 kt) (Fig. 4). The maximum V_r was typically present at a height range of 4500–6500 m (14 700–21 300 ft). Rotunno and Klemp (1985) demonstrated that the effects of storm rotation maintained supercell storm structure and updraft propagation through an inherent vertical pressure perturbation associated within the updraft region. As such, vertical pressure gradient forces induced by rotation enhance vertical accelerations within the updraft, well beyond those accelerations associated with buoyancy alone (Weisman and Klemp 1982; Rotunno and Klemp 1985; Edwards and Thompson 1998; McCaul and Weisman 2001). Miller et al. (1988) showed that giant-hail production was well correlated to the presence of a mid-level mesocyclone, creating a favorable growth trajectory within specific regions of the updraft with preferred vertical motions and fallspeeds. Furthermore, Witt (1998) found a moderate relationship between measured WSR-88D mid-level rotation and maximum hail size. The results herein are consistent with previous studies in that a mid-level mesocyclone with V_r magnitude >20 m s^{-1} should be expected with storms capable of producing giant hail.

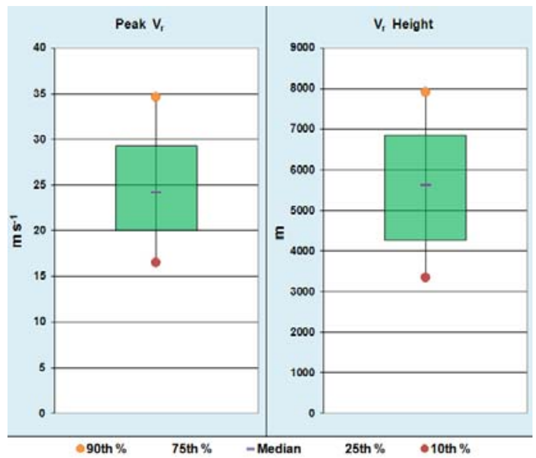

Figure 4: Box and whiskers plot of peak V_r magnitude (m s^{-1}) and height (m AGL) from 514 cases. The shaded box covers the 25th–75th percentiles, the whiskers extend to the 10th and 90th percentiles, and the median values are marked by the heavy horizontal line within each shaded box. *Click image to enlarge.*

The range of STD values for giant-hail cases is shown in Fig. 5. STD magnitudes ranged from 60–88 m s^{-1} (117–171 kt) in the 25th to 75th percentile, with a median value of 72 m s^{-1} (140 kt). While high values of STD suggest a strong updraft, their sole usefulness as predictors of giant hail is questionable. Two previous studies (Witt and Nelson 1991; Boustead 2008) attempted to correlate hail size to STD values. The results of these studies showed some skill in estimating smaller hail sizes (<102 mm); but they only examined a combined total of 6 cases of giant hail. Their results are similar to our findings, with their STD values falling in the upper 15% of our giant-hail cases. Also, these previous studies contained overlapping STD values for giant hail compared to smaller hail sizes within their respective datasets. Our results suggest that extremely high STD values (>100 m s^{-1}, 194 kt), while they may be observed, are not necessary for giant-hail production. STD should be considered a good proxy of updraft strength that may suggest the potential for giant hail.

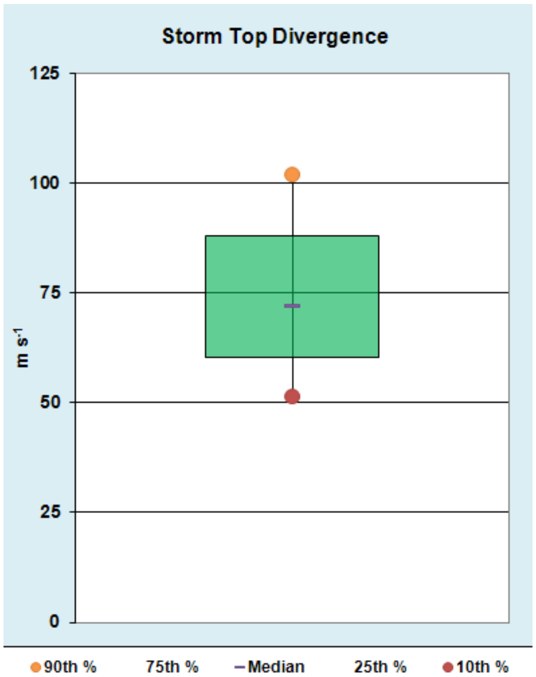

Figure 5: As in Fig. 4 except for storm-top divergence (m s^{-1}) from 437 cases. *Click image to enlarge.*

The maximum reflectivity within the full volume of radar data is depicted in Fig. 6. A median value of 69 dB*Z* was found from all the cases. To reiterate, with the WSR-88Ds being upgraded to super-resolution data in 2008, the likelihood for higher reflectivity values to be identified increased. This change is illustrated in this study; 65% of cases containing a maximum reflectivity ≥75 dB*Z* occurred within the 2 y of super-resolution data (2008–2009; 17% of the database). With the enhanced resolution, maximum reflectivity values within the storm column frequently will be greater than the 15-y average suggested in this study.

The distribution of the maximum 50-, 55-, 60-, 65-, 70-, and 75-dB*Z* echo heights is shown in Fig. 7. A general decrease in height was found as reflectivities increased. The median maximum dB*Z* heights of 50, 55, and 60 were found to be approximately 13 100 m (43 000 ft), 12 100 m (39 700 ft), and 10 600 m (34 800 ft) respectively, indicating a tall updraft with modest reflectivity aloft.

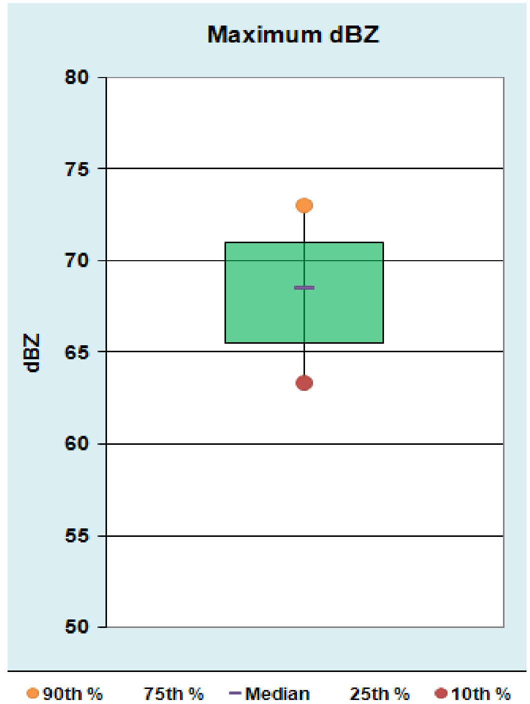

Figure 6: As in Fig. 4 except for maximum reflectivity (dB*Z*) from 568 cases. *Click image to enlarge.*

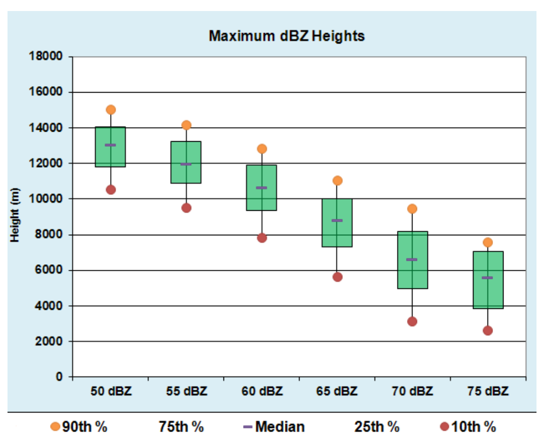

Figure 7: As in Fig. 4 except for maximum reflectivity heights (m AGL) of 50 dB*Z* (487 cases), 55 dB*Z* (532 cases), 60 dB*Z* (552 cases), 65 dB*Z* (449 cases), 70 dB*Z* (220 cases), and 75 dB*Z* (26 cases). *Click image to enlarge.*

The maximum reflectivity at the significant temperature levels of 0° C, –10° C, –20° C, and –30° C is shown in Fig. 8. Approximately 90%

of the cases contained reflectivity ≥60 dB*Z* throughout all four temperature levels. The preferred hail-growth zone generally occurs within the –10° C to –30° C layer, as established by the work from Nelson (1983), Miller et al. (1988), and Knight and Knight (2001). Reflectivities ≥60 dB*Z* were found within the hail-growth region, with a median value of 66 dB*Z* in the column. The upgrade to super resolution may increase reflectivity values within the hail-growth zone, particularly in cases where convection is far from the radar (>130 km), or in cool-season storms when temperature heights are lower.

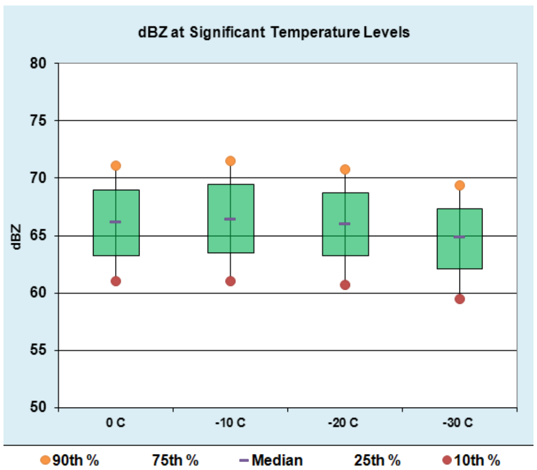

Figure 8: As in Fig. 4 except for the maximum reflectivity at significant temperature levels of 0° C (563 cases), –10° C (568 cases), –20° C (567 cases), and –30° C (564 cases). *Click image to enlarge.*

The relationship between the maximum 50-dB*Z* echo height and the 0° C level was examined. The 50-dB*Z* height can be approximated as a function of updraft strength and the 0° C level serves as a proxy for the potential ice-melting level in the atmosphere. Donavon and Jungbluth (2007) found a statistically significant relationship between these two variables, and used it as a predictor of hail with a diameter near 25 mm. However, they noted that this relationship frequently became unrepresentative in cases with strongly sheared environments and significant hail ≥51 mm, conditions similar to those found with supercells. Figure 9 shows the distribution of giant-hail reports using the maximum 50-dB*Z* echo height and the 0° C level. A favored region was identified between the 50-dB*Z* heights of 11 000 m to 16 000 m and the 0° C level of 3500 m to 4500 m. A gradual increase in the 50-dB*Z* echo height also occurred as the 0° C level increased, although a strong linear relationship was not found. It was uncommon to observe giant hail with 50-dB*Z* echo heights <9000 m or 0° C levels <3000 m. This may suggest that atmospheric conditions below these thresholds are associated with convection of insufficient updraft strength and hail residence time, or too cool of an environment to provide sufficient moisture for giant-hail growth.

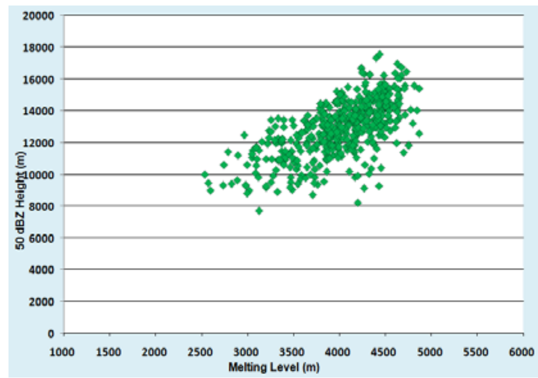

Figure 9: Scatter plot of the relationship between the maximum 50-dB*Z* echo height (m, AGL) and the 0° C level (m AGL) from 487 cases. *Click image to enlarge.*

A direct comparison of the 50-dB*Z* and 0° C height characteristics associated with giant hail to the results found by Donavon and Jungbluth for 19–25 mm (0.75–0.99 in) hail is shown in Fig. 10. In the giant-hail cases, a higher 50-dB*Z* echo height with the same 0° C height frequently produced larger hail. Some overlap between the two different hail sizes also exists, which strongly suggests that other factors, such as storm structure and mid-level rotation, likely contributed to more sizeable stones. Furthermore, inherent differences exist between 19–25 mm and ≥102 mm hail. Rasmussen and Heymsfield (1987) showed that melting effects are less as hail size increases. Giant hail is typically found outside the main liquid precipitation core due to size-sorting associated with the environmental wind shear and terminal fall speeds, whereas smaller hail may be subjected to more liquid precipitation which expedites melting. While the relationship between the 50-dB*Z* height and 0° C level may be effective in gauging the potential for severe hail, using this as a stand-alone predictor of hail size, especially for giant hail, will result in frequently poor hail-size forecasts.

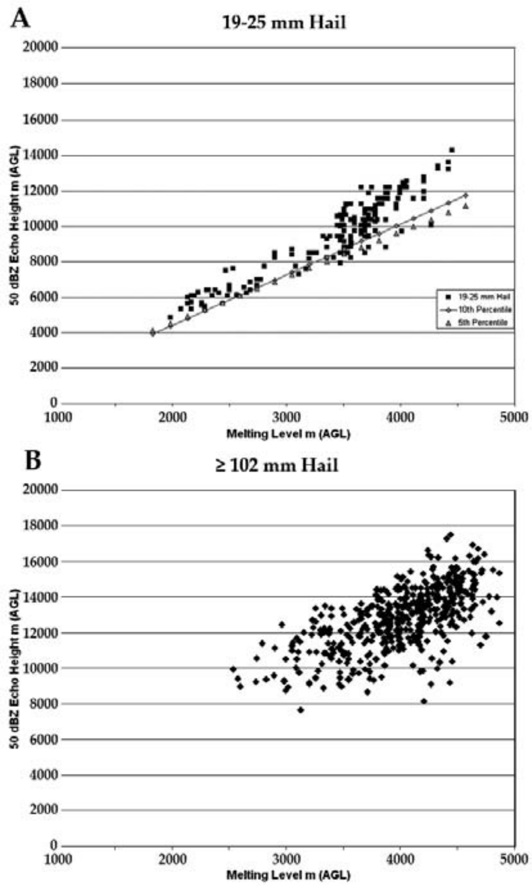

Figure 10: Scatter-plot comparison of the relationship between the maximum 50-dB*Z* echo height (m AGL) and the 0° C level (m AGL) for a) hail 19–25 mm and b) hail ≥102 mm. Figure 10a is from Fig. 2 of Donavon and Jungbluth (2007). The same x-y axis values apply. *Click image to enlarge.*

The relationship between the maximum 60-dB*Z* echo height and the environmental temperature heights of 0° C, –10° C, –20° C, and –30° C is shown in Fig. 11. The large majority of maximum 60-dB*Z* heights occurred between 9000 m to 13 000 m, resulting in approximately 90% of all giant-hail cases residing well above each of the four temperature levels examined. Specifically, the median value of the maximum 60-dB*Z* height extended approximately 6600 m, 5200 m, 3700 m, and 2400 m above the median temperature heights of 0° C, –10° C, –20° C, and –30° C, respectively. It was common to observe 60-dB*Z* heights atop the hail-growth zone, owing to a frequent presence of strong reflectivities within the –10° C to –30° C layer, suggesting large hail formation. This concept is similar to NWS training courses that advise high values of reflectivity (≥60 dB*Z*) at or above the –20° C level signifies a greater hail threat (NWS 2006). Based upon these results, it should be considered a necessary characteristic of giant-hail producing storms for 60 dB*Z* to reach well above the –20° C level, and in most cases, to extend beyond the primary hail-growth zone.

Past studies have reviewed the potential utility of VIL and VILD as hail-size forecast techniques (Amburn and Wolf 1997; Edwards and Thompson 1998). They showed that similar VIL values existed for various hail sizes, and thus VIL was discounted as a credible hail-size predictor. Little work has been done to establish the usefulness of the newer, higher resolution DVIL. DVIL is expected to possess lower values than the 4 km VIL predecessor due to the maximum reflectivities having a lower probability of residing in the same grid space (Stumpf et al. 2004). Figure 12 shows the range of DVIL and VILD for the giant-hail cases. DVIL was >105 kg m^{-2} in the large majority of cases, with a median value at the cap of 127 kg m^{-2}. VILD was >7.5 kg m^{-3} in most of the giant-hail events, with a median value of 10 kg m^{-3}. Higher values of DVIL and VILD likely would have been identified without the respective caps of 127 kg m^{-2} and 13 kg m^{-3} in place.

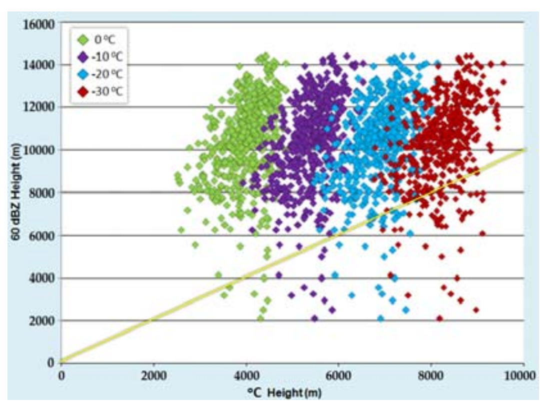

Figure 11: Scatter plot of the relationship between the maximum 60-dB*Z* echo height (m, AGL) and the 0° C (green), –10° C (purple), -20° C (blue), and –30° C (red) temperature heights (m AGL) from 552 cases. An identity line of constant height is in yellow. *Click image to enlarge.*

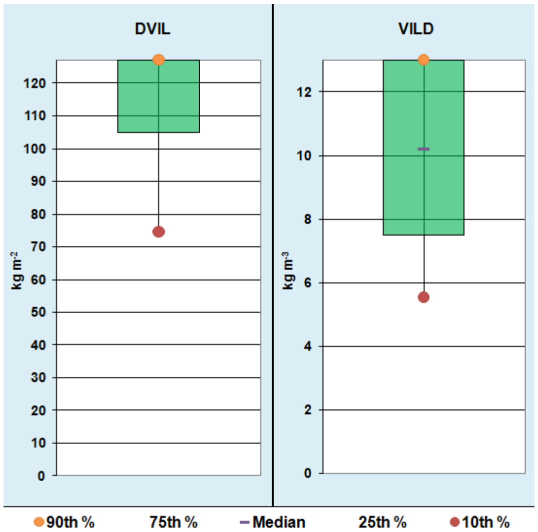

Figure 12: As in Fig. 4 except for digital vertically integrated liquid (kg m^{-2}), and vertically integrated liquid density (kg m^{-3}) from 565 cases. *Click image to enlarge.*

Approximately 43% of storms producing giant hail were found to exhibit a TBSS during the 20-min study period. Downrange echoes frequently masked the TBSS signature, partially explaining the low number of positively identified cases. Our results parallel previous work by Lemon (1998) and NWS (2006) that suggests the TBSS signal is often a sufficient indicator of the occurrence of large hail, but is not a necessary radar signature when determining the presence of large hail.

6. Results: Giant hail vs. large hail

In order to determine whether unique signals were distinguishable for giant-hail events, a comparison study was conducted using smaller hail sizes. Golfball to hen egg-sized hail (45–51 mm, 1.75–2.00 in) reports (hereafter referred to as GBHE) were selected as the comparison size, since a robust updraft and organized storm structure were typically present with hail of this magnitude.

A total of 568 giant-hail and 67 GBHE reports from *Storm Data* were incorporated in the comparison database, in addition to 28 giant-hail and 79 GBHE reports from SHAVE. The SHAVE dataset has a much higher temporal and spatial resolution than NWS collected hail reports published in *Storm Data*, allowing for an increased confidence in the quality of data (Ortega et al. 2009). The *Storm Data* and SHAVE databases were kept as separate subsets to examine whether results were repeatable between the two sources. The same methodology described in section 2d was applied to the additional 174 cases within the comparison study (Fig. 13). GBHE reports were removed from the database if hail >51 mm was reported within 250 km of the storm.

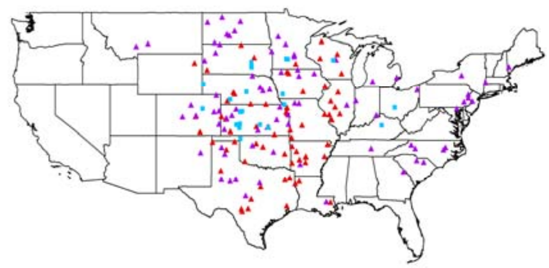

Figure 13: Report locations of giant hail from SHAVE (blue squares), GBHE from *Storm Data* (red triangles), and GBHE from SHAVE (purple triangles) used in the hail-comparison database. *Click image to enlarge.*

Approximately 81% of the GBHE producing storms fit the classification of a supercell, as defined in section 2d-9. The remaining storms were characterized by non- mesocyclonic, strong isolated updrafts, or embedded updrafts within well-organized squall lines. In contrast, supercell storm structure was associated with giant-hail reports from both *Storm Data* and SHAVE in 99% of the events.

The maximum V_r of the mesocyclone for the giant-hail and GBHE events is illustrated in Fig. 14. Giant-hail cases from *Storm Data* were characterized by mesocyclones with V_r values of 20–29 m s^{-1} (39–56 kt) within the 25^{th} to 75^{th} percentile and a median of 24 m s^{-1} (47 kt). Giant-hail reports from SHAVE were similar with V_r values of 22–30 m s^{-1} (43–58 kt) within the 25^{th} to 75^{th} percentile and a median of 26 m s^{-1} (50 kt). The GBHE storms had weaker mid-level mesocyclones, with V_r values of 14–21 m s^{-1} (27–41 kt) within the 25^{th} to 75^{th} percentile and a median of 17 m s^{-1} (33 kt) for the *Storm Data* cases. GBHE V_r values from the SHAVE reports were similar to GBHE *Storm Data* reports with V_r values of 16–23 m s^{-1} (31–45 kt) within the 25^{th} to 75^{th} percentile and a median of 19 m s^{-1} (37 kt).

Peak STD values also were compared for the two different hail-size groups (Fig. 15). Giant-hail cases had substantially larger STD speeds than the GBHE events, with the 25th to 75th percentile ranging from 60–88 m s^{-1} (117–171 kt) for the *Storm Data* reports, and 63–78 m s^{-1} (123–152 kt) for the SHAVE reports. Median STD values for giant hail from *Storm Data* and SHAVE were 72 m s^{-1} and 75 m s^{-1} (140 kt, 146 kt) respectively. In comparison, the 25th to 75th percentile for the GBHE cases for Storm Data and SHAVE reports ranged from 41–62 m s^{-1} (80–121 kt) and 42–65 m s^{-1} (82–126 kt) respectively, with median values of 51 m s^{-1} and 52 m s^{-1} (99 kt, 101 kt).

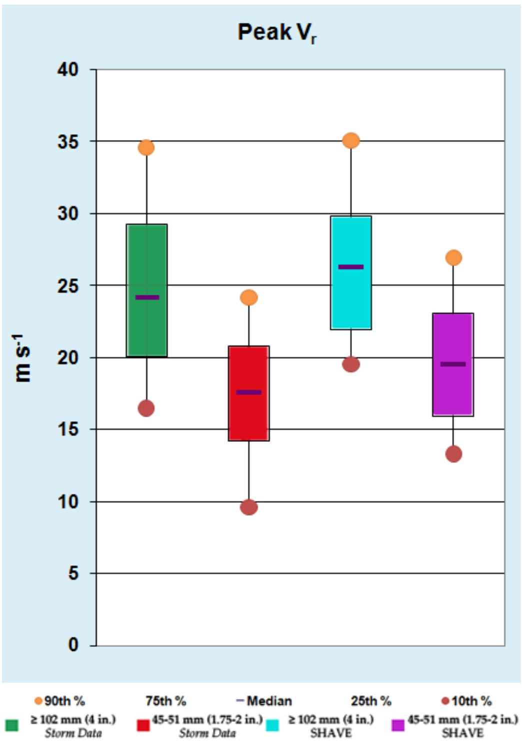

Figure 14: As in Fig. 4 except comparing peak V_r magnitudes (m s^{-1}) from *Storm Data* reports of giant hail (green; 470 cases) and GBHE (red; 65 cases), and from SHAVE reports of giant hail (blue; 28 cases) and GBHE (purple; 70 cases). *Click image to enlarge.*

As shown in the box and whisker diagrams (Figs. 14 and 15), there appears to be minimal overlap between giant-hail and GBHE events for V_r and STD. To determine whether or not the difference between the two hail-size groups is statistically significant, a null hypothesis Student's t-test (Wilks 1995) was conducted. The difference between the V_r values of the giant-hail events and the GBHE values was non-zero at the 99% confidence level. Likewise, the difference between STD values for the two hail-size groups was statistically significant with 99% confidence.

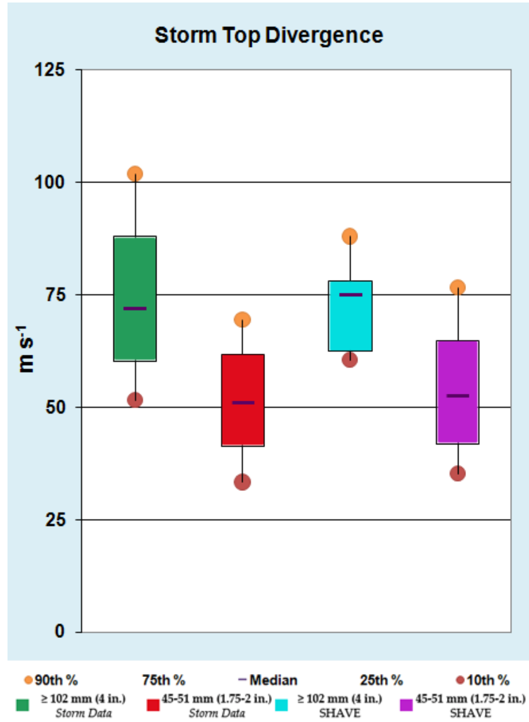

Figure 15: As in Fig. 4 except comparing storm-top divergence magnitudes (m s^{-1}) from *Storm Data* reports of giant hail (green; 409 cases) and GBHE (red; 57 cases), and from SHAVE reports of giant hail (blue; 21 cases) and GBHE (purple; 70 cases). *Click image to enlarge.*

Figure 16 highlights the dissimilarity of maximum reflectivity heights between the giant-hail and GBHE sized groups. The giant-hail events frequently contained greater reflectivity heights of 50, 55, 60, and 65 dB*Z* than GBHE storms. However, these heights were dependent on the environmental conditions for the time of year. This is illustrated through the comparison of maximum reflectivity heights of *Storm Data* reports to SHAVE reports. Reflectivity heights from the SHAVE dataset are substantially greater with GBHE cases overlapping a large portion of *Storm Data* giant-hail heights. The height differences are likely explained by the annual operating period of SHAVE. For instance, all SHAVE reports included in this study occurred during the months of May through August, typically when atmospheric conditions support taller storms. In contrast, only 74% of giant-hail reports and 60% of GBHE reports from *Storm Data* occurred during this four month period. Therefore, differences in height should be expected when comparing the two report sources. While specific maximum reflectivity heights cannot be related directly to the probability of a giant-hail occurrence, giant-hail producing storms frequently were characterized by much higher reflectivity heights than storms that produced GBHE hail sizes, relative to the time of year.

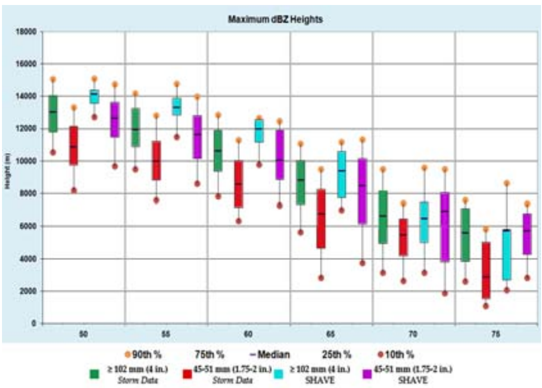

Figure 16: As in Fig. 4 except comparing maximum reflectivity heights (m AGL) of giant hail (*Storm Data* reports green, SHAVE reports blue) and GBHE (*Storm Data* reports red, SHAVE reports purple). Shown are *Storm Data* GBHE cases of 50 dB*Z* (63 cases), 55 dB*Z* (66 cases), 60 dB*Z* (63 cases), 65 dB*Z* (63 cases), 70 dB*Z* (37 cases), and 75 dB*Z* (5 cases); SHAVE giant-hail cases of 50 dB*Z* (26 cases), 55 dB*Z* (28 cases), 60 dB*Z* (28 cases), 65 dB*Z* (26 cases), 70 dB*Z* (18 cases), and 75 dB*Z* (5 cases); SHAVE GBHE cases of 50 dB*Z* (68 cases), 55 dB*Z* (76 cases), 60 dB*Z* (78 cases), 65 dB*Z* (72 cases), 70 dB*Z* (43 cases), and 75 dB*Z* (10 cases). Number of *Storm Data* giant-hail cases same as in Fig. 7. *Click image to enlarge.*

The distribution of maximum reflectivity residing at the temperature levels of 0° C, –10° C, –20° C, and –30° C is shown in Fig. 17. Significant overlap of the reflectivity values between the two hail sizes existed at each temperature level, suggesting that this parameter is a poor discriminator of giant hail. Still, giant-hail producing storms should show high reflectivity at all four temperature levels, depicted by median values of reflectivity ≥65 dB*Z* throughout the preferred hail-growth zone.

Figure 18 illustrates the distribution of both giant and GBHE hail reports using the maximum

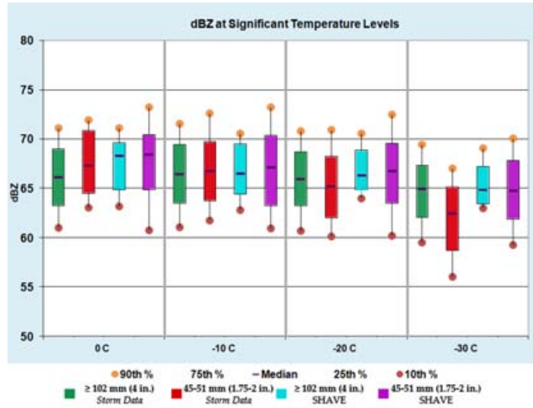

Figure 17: As in Fig. 4 except comparing the maximum reflectivity at significant temperature levels. Shown are *Storm Data* reports of GBHE (red) of 0° C (62 cases), –10° C (67 cases), –20° C (67 cases), and –30° C (67 cases); SHAVE reports of giant hail (blue) of 0° C (26 cases), –10° C (28 cases), –20° C (28 cases), and –30° C (28 cases); and SHAVE reports of GBHE (purple) of 0° C (74 cases), –10° C (78 cases), –20° C (79 cases), and –30° C (78 cases). Number of *Storm Data* giant-hail cases (green) same as in Fig. 8. *Click image to enlarge.*

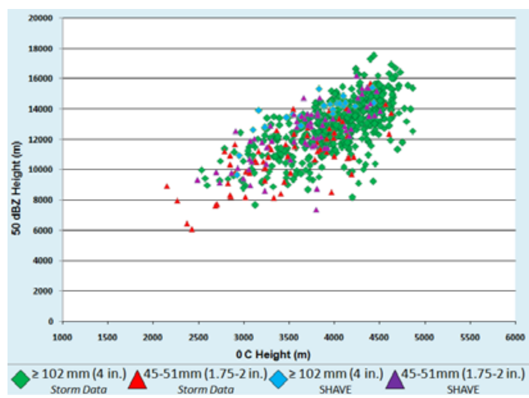

Figure 18: Scatter plot of the relationship between the maximum 50-dB*Z* echo height (m, AGL) and the 0° C level (m AGL), comparing *Storm Data* reports of giant hail (green, 450 cases), and GBHE (red, 63 cases), along with SHAVE reports of giant hail (blue, 26 cases), and GBHE (purple, 68 cases). *Click image to enlarge.*

50-dB*Z* echo height and the 0° C level. Figure 19 shows a similar distribution using the maximum 60-dB*Z* echo height and the –20° C level. There appears to be a minor tendency within the upper threshold of cases where giant hail is preferred compared to GBHE sizes.

However, a modest overlap exists between the two hail-size groups on both the maximum 50 dB*Z* to the 0° C level and the maximum 60 dB*Z* to the –20° C level. The overlap reveals an unreliable relationship using specific maximum-reflectivity heights paired with environmental temperatures for the discrimination of giant-hail sizes.

Other parameters showed little or no skill in discriminating between giant-hail and GBHE hail sizes. A TBSS was present with 44% of giant-hail reports from *Storm Data* and SHAVE, compared to 51% of GBHE events. Generally, these results suggest that no discernable preference for the presence of a TBSS exists between the two hail sizes. Figure 20 reveals the maximum reflectivity within the radar volume for both the 45–51 mm and ≥102 mm cases. Slightly lower values with the *Storm Data* reports of giant hail likely were due to a smaller percentage of events within the dataset occurring after the super-resolution upgrade. Otherwise, reflectivity ≥65 dB*Z* was quite common for both hail sizes. With the substantial overlap of values, maximum reflectivity within the radar volume showed no ability to differentiate giant-hail sizes. Lastly, DVIL and VILD proved ineffective in distinguishing between the two hail-size groups, with a large overlap of values (Fig. 21). Edwards and Thompson (1998) had similar conclusions regarding VIL.

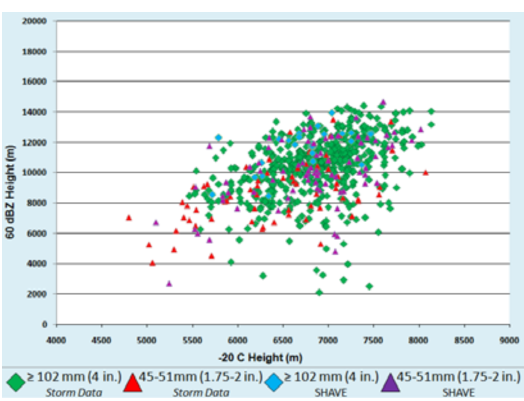

Figure 19: Scatter plot of the relationship between the maximum 60-dB*Z* echo height (m, AGL) and the –20° C level (m AGL), comparing *Storm Data* reports of giant hail (green, 550 cases), and GBHE (red, 66 cases), along with SHAVE reports of giant hail (blue, 28 cases), and GBHE (purple, 78 cases). *Click image to enlarge.*

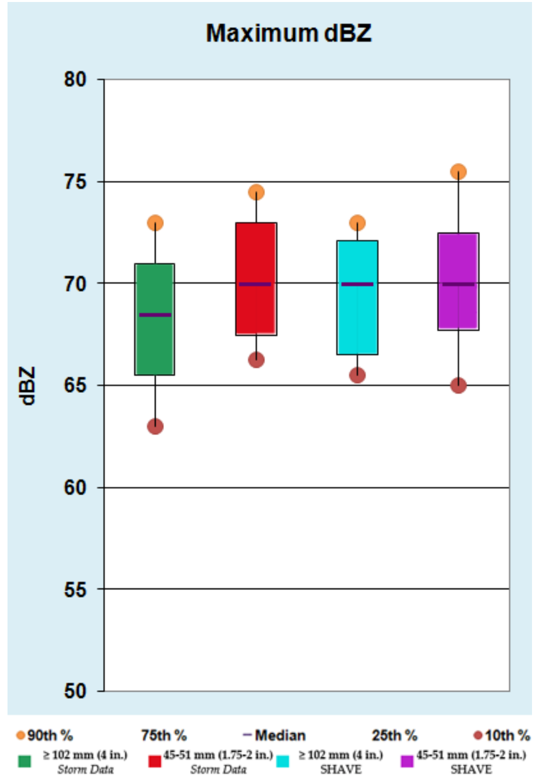

Figure 20: Same as in Fig. 4 except comparing the maximum reflectivity (dB*Z*) of giant hail from *Storm Data* (green, 568 cases) and SHAVE reports (blue, 28 cases), to GBHE from *Storm Data* (red, 67 cases) and SHAVE (purple, 79 cases). *Click image to enlarge.*

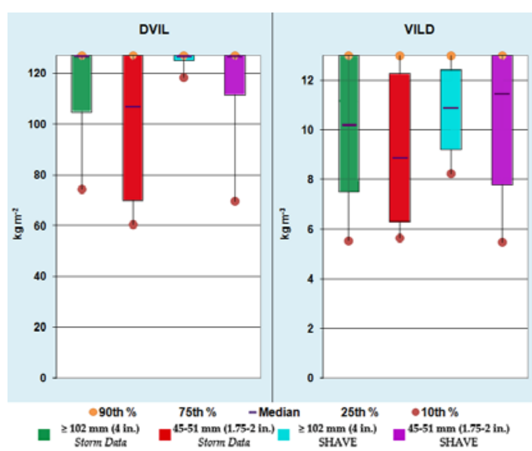

Figure 21: Same as in Fig. 4 except comparing DVIL (kg m^{-2}), and VILD (kg m^{-3}) of giant hail from *Storm Data* (green, 565 cases) and SHAVE reports (blue, 28 cases), to GBHE from *Storm Data* (red, 67 cases) and SHAVE (purple, 78 cases). *Click image to enlarge*

7. Summary

A radar-based assessment of the detection of giant hail using WSR-88D data was conducted for the entire contiguous United States, using a total of 596 *Storm Data* and SHAVE giant-hail reports during 1995–2009. April through July contained ≈80% of all giant-hail events, with 54% in May and June. The majority of events transpired between 2200–0159 UTC. An additional 146 reports of 45–51 mm sized hail were analyzed and compared to the ≥102 mm hail events to determine whether unique characteristics were inherent with giant-hail producing storms. The results from the radar analysis successfully identified operational signals that distinguished storms more favorable for generating giant hail:

- Approximately 99% of the convection was classified as supercellular, with a BWER structure frequently present.
- The peak V_r magnitude of the mid-level mesocyclone typically was 20–29 m s^{-1} (39–56 kt), with a median of 24 m s^{-1} (47 kt).
- The maximum STD magnitude frequently was 60–88 m s^{-1} (117–171 kt), with a median of 72 m s^{-1} (140 kt).

The differences between the V_r and STD values for giant hail versus smaller GBHE hail sizes were statistically significant to the 99% confidence level, and likewise showed the greatest promise in identifying storms capable of hail ≥102 mm.

A much greater overlap in values was found within the maximum reflectivity residing in the preferred hail-growth zone of –10° C to –30° C, which suggests this parameter is a poor discriminator between large and giant hail. Still, median reflectivities ≥65 dB*Z* were present at each temperature level, suggesting that high reflectivity should reside throughout that temperature layer with giant-hail producing storms.

The maximum heights AGL of the 50-, 55-, and 60-dB*Z* reflectivities were regularly higher with giant hail than GBHE. While specific maximum-reflectivity heights that discriminate giant hail cannot be stated due to the variability of environmental conditions, 50- and 60-dB*Z* heights with hail ≥102 mm were commonly 1825 m (6000 ft) higher than those heights with GBHE storms.

Maximum reflectivity within the radar volume, TBSS signatures, and VIL-based products showed little to no skill in discriminating giant-hail sizes. Modest overlap existed between the relationship of the maximum 50-dB*Z* height and the 0° C level, as well as the maximum 60-dB*Z* height and the –20° C level. The TBSS signature is not found to be a reliable indicator of giant hail. In addition, a substantial overlap between hail-size categories was found with the values of DVIL and VILD products. These shortcomings underscore the difficulty of predicting maximum hail sizes with some commonly used radar-based methods, and also suggest that some of these practices may be obsolete as baselines for hail-size forecasting, especially in the presence of well-organized storm structure.

This research serves as a conceptual model that operational meteorologists may use as guidance to the detection of giant hail. Well-organized supercell storm structure, in combination with a moderate to strong mesocyclone, strong storm-top divergence, and high reflectivity values throughout and above the hail-growth zone, should be expected to be present in a storm capable of producing giant hailstones. These methods oversimplify the very complex hail-growth process that requires a number of conditions and favorable trajectories necessary for the production of giant-hail sizes. Some variability in values between storms should be expected, and sufficient ground truth often may not be available to verify giant hail. The advanced detection of giant hail has been and will continue to be challenging, but a few operational signals are identified that may increase advanced recognition and confidence of the potential of giant hail during short-term warning operations.

ACKNOWLEDGMENTS

The authors wish to express their sincere appreciation to George Phillips (NWS Topeka, KS), Jennifer Laflin (NWS Sioux Falls, SD), and Albert Pietrycha (NWS Goodland, KS) for technical editing, as well as Amos Magliocco (University of North Texas), Roger Edwards (SPC), and Ryan McCammon (State of Wyoming) for thorough copyediting. Kiel Ortega (CIMMS/NSSL) is thanked for providing SHAVE hail data. We would also like to thank Ryan Jewell (SPC), Arthur Witt (NSSL), Les Lemon (CIMMS), and Jeff Manion (CRH) for their helpful reviews. The views expressed within this manuscript are those of the authors and do not necessarily represent those of the National Weather Service.

REFERENCES

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

Baumgardt, D., cited 2011: Hail estimation: How good are your spotters? [Available online at http://www.crh.noaa.gov/arx/hail_size_MSP.pdf.]

Boustead, J. M., 2008: Using maximum storm-top divergence and the vertical freezing level to forecast hail size. Preprints, *24th Conf. on Severe Local Storms,* Savannah, GA, Amer. Meteor. Soc., P6.6.

Brooks, H. E., and C. A. Doswell III, 2002: Deaths in the 3 May 1999 Oklahoma City tornado from a historical perspective. *Wea. Forecasting*, **17**, 354–361.

——, ——, and M. P. Kay, 2003: Climatological estimates of local daily tornado probability for the United States. *Wea. Forecasting*, **18**, 626–640.

Brown, R. A., V. T. Wood, and D. Sirmans, 2000: Improved WSR-88D scanning strategies for convective storms. *Wea. Forecasting*, **15**, 208–220.

Browning, K. A., 1964: Airflow and precipitation trajectories within severe local storms which travel to the right of the winds. *J. Atmos. Sci.*, **21**, 634–639.

Bunkers, M. J., M. R. Hjelmfelt, and P. L. Smith, 2006: An observational examination of long-lived supercells. Part I: Characteristics, evolution, and demise. *Wea. Forecasting*, **21**, 673–688.

desJardins, M. L., K. F. Brill, and S. S. Schotz, 1991: GEMPAK5 user's guide. NASA Tech. Memo. 4260, 232 pp. [Available from NASA Code NTT-4, Washington, DC 20546-0001.]

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.

Doswell, C. A. III, A. R. Moller, and H. E. Brooks, 1999: Storm spotting and public awareness since the first tornado forecasts of 1948. *Wea. Forecasting*, **14**, 544–557.

——, 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.

Duda, J. D., and W. A. Gallus, 2010: Spring and summer Midwestern severe weather reports in supercells compared to other morphologies. *Wea. Forecasting*, **25**, 190–206.

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.

Greene, D. R., and R. A. Clark, 1972: Vertically integrated liquid water—A new analysis tool. *Mon. Wea. Rev.*, **100**, 548–552.

Hales, J. E., 1988: Improving the watch/warning system through use of significant event data. Preprints, *15th Conf. on Severe Local Storms*, Baltimore, MD, Amer. Meteor. Soc., 165–168.

Hocker, J. E., and J. B. Basara, 2008: A geographic information systems-based analysis of supercells across Oklahoma from 1994 to 2003. *J. Appl. Meteor. Climatol.*, **47**, 1518–1538.

Howard, K. W., J. J. Gourley, R. A. Maddox, 1997: Uncertainties in WSR-88D measurements and their impacts on monitoring life cycles. *Wea. Forecasting*, **12**, 166–174.

Jewell, R., and J. Brimelow, 2009: Evaluation of Alberta hail growth model using severe hail proximity soundings from the United States. *Wea. Forecasting*, **24**, 1592–1609.

Knight, C. A., and N. C. Knight, 2001: Hailstorms. *Severe Convective Storms, Meteor. Monog.r,* No. 28, Amer. Meteor. Soc., 223–248.

——, and N. C. Knight, 2005: Very large hailstones from Aurora, Nebraska. *Bull. Amer. Meteor. Soc.*, **86**, 1773–1781.

Lemon, L. R., 1977: New severe thunderstorm radar identification techniques and warning criteria: A preliminary report. NOAA Tech. Memo. NWS NSSFC-1, 60 pp. [NTIS PB-273049.]

——, 1978: On the use of storm structure for hail identification. Preprints, *18th Conf. on Radar Meteor.*, Boston, MA, Amer. Meteor. Soc., 203-206.

——, 1998: The radar "three-body scatter spike": An operational large-hail signature. *Wea. Forecasting*, **13**, 327–340.

Maddox, R. A., D. S. Zaras, P. L. MacKeen, J. J. Gourley, R. Rabin, and K. W. Howard, 1999: Echo height measurements with the WSR-88D: Use of data from one versus two radars. *Wea. Forecasting*, **14**, 455–460.

McCaul, E. W., and M. L. Weisman, 2001: The sensitivity of simulated supercell structure and intensity to variations in the shapes of environmental buoyancy and shear profiles. *Mon. Wea. Rev.*, **129**, 664–687.

Mesinger, F., and Coauthors, 2006: North American regional reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360.

Miller, L. J., J. D. Tuttle, and C. A. Knight, 1988: Airflow and hail growth in a severe northern High Plains supercell. *J. Atmos. Sci.*, **45**, 736–762.

NWS Warning Decision Training Branch, 2006: Three body scatter spike. Advanced Warning Operations Course: Storm Interrogation. [Available online at: http://ejssm.org/ojs/public/vol2-3/ICSvr3-V-A_TBSS.pdf.]

NCDC, 1995-2010: *Storm Data*. [Available from National Climatic Data Center, 151 Patton Ave., Asheville, NC 28801-5001.]

Nelson, S. P., 1983: The influence of storm flow structure on hail growth. *J. Atmos. Sci.*, **40**, 1965–1983.

Ortega, K. L., T. M. Smith, K. L. Manross, A. G. Kolodziej, K. A. Scharfenberg, A. Witt, J. J. Gourley, 2009: The Severe Hazards Analysis and Verification Experiment. *Bull. Amer. Meteor. Soc.*, **90**, 1519–1530.

Parker, M. D., and R. H. Johnson, 2000: Organizational modes of midlatitude mesoscale convective systems. *Mon. Wea. Rev.*, **128**, 3413–3436.

Paxton, C. H., and J. M. Shepherd, 1993: Radar diagnostic parameters as indicators of severe weather in central Florida. NOAA Tech. Memo. NWS SR-149, 12 pp. [Available from NWS Southern Region Headquarters, 819 Taylor St., Room 10A26, Fort Worth, TX 76102.]

Rasmussen, E. N., and D. O. Blanchard, 1998: A baseline climatology of sounding-derived supercell and tornado forecast parameters. *Wea. Forecasting*, **13**, 1148–1164.

Rasmussen, R. M., and A. J. Heymsfield, 1987: Melting and shedding of graupel and hail. Part I: Model physics. *J. Atmos. Sci.*, **44**, 2754–2763.

Rotunno, R., and J. Klemp, 1985: On the rotation and propagation of simulated supercell thunderstorms. *J. Atmos. Sci.*, **42**, 271–292.

Serafin, R. J., and J. W. Wilson, 2000: Operational weather radar in the United States: Progress and opportunity. *Bull. Amer. Meteor. Soc.*, **81**, 501–518.

Simmons, K. M., and D. Sutter, 2005: WSR-88D radar, tornado warnings, and tornado casualties. *Wea. Forecasting*, **20**, 301–310.

Stumpf, G. J., T. M. Smith, and J. Hocker, 2004: New hail diagnostic parameters derived by integrating multiple radars and multiple sensors. Preprints, *22nd Conf. on Severe Local Storms,* Hyannis, MA, Amer. Meteor. Soc., P7.8.

Thompson, R. L., R. Edwards, J. A. Hart, K. L. Elmore, and P. Markowski, 2003: Close proximity soundings within supercell environments obtained from the Rapid Update Cycle. *Wea. Forecasting*, **18**, 1243–1261.

Trapp, R. J., S. A. Tessendorf, E. S. Godfrey, and H. E. Brooks, 2005: Tornadoes from squall lines and bow echoes. Part I: Climatological distribution. *Wea. Forecasting*, **20**, 23–34.

Verbout, S. M., H. E. Brooks, L. M. Leslie, and D. M. Schultz, 2006: Evolution of the U.S. tornado database: 1954–2003. *Wea. Forecasting*, **21**, 86–93.

Vogel, J. M., C. D. Payne, C. A. Van Den Broeke, and L. R. Lemon, 2009: Impacts of super-resolution data on National Weather Service warning decision making. Preprints, *9th Annual Student Conf.,* Amer. Meteor. Soc., S19.

Weisman, M. L., and J. B. Klemp, 1982: The Dependence of numerically simulated convective storms on vertical wind shear and buoyancy. *Mon. Wea. Rev.*, **110**, 504–520.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.

Wilson, J. W., and D. Ruem, 1988: The flare echo: Reflectivity and velocity signature. *J. Atmos. Oceanic Technol.*, **5**, 197-205.

Witt, A., 1998: The relationship between WSR-88D measured midaltitude rotation and maximum hail size, Preprints*, 19th Conf. on Severe Local Storms,* Minneapolis, MN, Amer. Meteor. Soc., 740–743.

——, and S. P. Nelson, 1991: the use of single-doppler radar for estimating maximum hailstone size. *J. Appl. Meteor.*, **30**, 425–431.

——, M. D. Eilts, G. J. Stumpf, J. T. Johnson, E. D. Mitchell, and K. W. Thomas, 1998a: An enhanced hail detection algorithm for the WSR-88D. *Wea. Forecasting*, **13**, 286–303.

——, ——, ——, E. D. Mitchell, J. T. Johnson, and K. W. Thomas, 1998b: Evaluating the performance of WSR-88D severe storm detection algorithms. *Wea. Forecasting*, **13**, 513–518.

REVIEWER COMMENTS

[Authors' responses in *blue italics.*]

REVIEWER A (Ryan E. Jewell):

Initial Review:

Recommendation: Revisions required.

General Comments: I appreciate this study as it focuses on extreme hail, which is the most damaging and potentially life-threatening of all. While few people would take any action for a small hail threat, an extreme hail threat can be actionable.

This study is a good start at attempting to delineate between medium to very large sized hail based on a combination of basic environmental parameters and radar data. However, much more work can be done, and it may be too large of a scope to try to address both pre-storm environmental conditions and real time radar data in one study. Importantly, expanding the dataset of cases while adding null cases would allow for quantification of skill.

The pre-storm environmental analysis has been removed from the revised manuscript, as a thorough environmental investigation is beyond the scope of this radar-based study. In addition, the giant-hail database has been expanded to encompass the entire contiguous United States.

We did not investigate any null cases. The authors believe that adding null cases from Storm Data would potentially show unreliable results due to the low-resolution nature of the database, or worse, suggest signals that are in great error. We believe that by using only cases with the positive identification of giant (4+ in) hail, it can be said with modest confidence that very large stones were present. Null cases (<4 in hail), however, may occasionally represent cases where giant hail was present but not recorded in Storm Data. This is discussed below in more detail in response to the following comment.

A theme that I noticed throughout the study was that the Storm Data database is more erroneous for smaller hail reports than for larger hail reports, and that larger hail almost always occurs but is just not reported. If the author feels this strongly, I encourage you to come up with evidence to support this claim. Personally, I have not seen nor suspected any evidence of the sort, except for flakey hail algorithms that consistently overestimate maximum hail size.

The largest hail stone, especially significant hail sizes, is frequently not identified by traditional NWS verification practices, which is a result of several factors.

During NWS severe weather operations, the available workforce and resources are simply not great enough to conduct a real-time thorough investigation into the hail-fall character of a storm. Verification for a severe thunderstorm warning only requires one report of severe hail (1.00+ in) or a 58 mph wind. In our experience working for multiple field offices, one severe hail report is frequently deemed sufficient in many cases during the post-storm verification process. There is usually not an emphasis on obtaining the maximum sized hail in a storm during post-storm verification. Amburn and Wolf (1997) highlighted these verification issues and stated NWS "verification practices are designed to efficiently verify warnings, not to satisfy scientific studies. Verification telephone calls to locate weather events often stop after the first severe weather report is received. Additional calls to determine the largest severe hail observed are not normally made. Due to the method of verification the largest hail to reach the ground is often not found."

Additionally, the amount of smaller hail (1–1.50 in) compared to giant hail (4+ in) is much greater for the smaller variety in the number of stones produced and spatial coverage in convective storms. There is simply a higher probability that the largest stones will go unreported. With many areas characterized by a low population density in the Plains states where giant hail occurs with the greatest frequency, there is an additional inherent lower probability of ground-truth detection. We are not suggesting that the NWS never manages to locate the largest stone in a storm, but it would be inappropriate to believe the largest stone is identified on a regular basis with supercell storms that have a very wide hail-size spectrum, both spatially *and temporally. Perhaps the only reliable high-resolution hail datasets that adequately identify the hail-fall character in storms are the projects SHAVE (http://ewp.nssl.noaa.gov/projects/shave) and HailSTONE (http://www.hailstoneresearch.org). Unfortunately, these projects operate on a limited basis and domain area, but do show these types of size discrepancies.*

To illustrate the limitations of hail size verification in Storm Data, we examined the giant hail reports obtained by SHAVE. Each NWS office has access to these reports following an event. Therefore, one may assume that all giant hail reports from SHAVE would be recorded by the local NWS office in Storm Data, in addition to other giant hail reports collected from the public and spotters. Unfortunately, the majority of NWS offices failed to utilize SHAVE information for giant hail events. An investigation of giant hail reports collected from SHAVE, using 23 unique events (a unique event is defined as a single event, with the maximum hail size utilized and the elimination of multiple reports from the same storm) showed:

- *Only 44% of giant hail unique events in the SHAVE database were listed in Storm Data. (Reports originated directly from SHAVE or other verification sources)*
- *The average maximum diameter hail size in Storm Data was 2.50 in. when giant hail was identified by SHAVE, but not utilized by NWS offices to supplement other reports.*

While this dataset is small, it does indeed quantify the underestimated diameter size of giant hail in Storm Data. However, it is unknown whether these values can be generalized for all occurrences of giant hail across the United States.

[Specific] comments are embedded throughout the electronic copy. I look forward to reviewing the revised paper. Please contact me at any point if you would like to use my hail spreadsheet that contains environmental data derived from surface-modified observed severe hail proximity soundings.

Thank you. We did find the hail spreadsheet useful as a means to verify the performance of NARR-calculated instability parameters, which frequently appeared too low. The 0–6 km wind vector differential data were quite similar between the modified soundings and the NARR datasets.

The "Rasmussen table" below summarizes my evaluation of this study. [Specific] comments follow the table.

Criterion	**Satisfied**	**Deficient, but can be remedied**	**Deficient; *cannot* be remedied by modifying the paper**	**Deficient, *not known* if it can be remedied by modifying the paper**
1. Does the paper fit within the stated scope of the journal?		X		
2. Does the paper 1) identify a gap in scientific knowledge that requires further examination; 2) repeat another study to verify its findings; or 3) add new knowledge to the overall body of scientific understanding?		X		
3 Is the paper free of errors in logic?		X		
4. Do the conclusions follow from the evidence?		X		
5. Are alternative explanations explored as appropriate?		X		
6. Is uncertainty quantified?		X		
7. Is previous work and current understanding represented correctly?		X		
8. Is information conveyed clearly enough to be understood by the typical reader?		X		

How is our research deficient within the stated scope of the EJSSM Journal and how do you propose this to be remedied? The authors believe that the subject matter conforms quite well to the EJSSM policy and scope. (http://ejssm.org/ojs/index.php/ejssm/about/editorialPolicies#focusAndScope)

The authors also question how our research is deficient in the following areas: 1) identify a gap in scientific knowledge that requires further examination; and 3) add new knowledge to the overall body of scientific understanding. We strongly disagree with at least these two subjective assessments.

Otherwise, the majority of minor corrections and comments as suggested by the reviewer embedded within the document were made to the revised manuscript.

[Minor comments omitted...]

Second review:

Recommendation: Revisions required.

General Comments: [re: Nulls] This is all great on the POD side of things, but you need null cases to make the information useful in an operational sense. I was just poking around looking at storms on one night, and found examples of storms with little to no hail but parameters suggesting giant hail. An occasional error is not a good reason to damn the entire database. You just have to work with what you've got and note the limitations.

We were not initially clear on what the reviewer defined as a "null" giant hail case. We assume the most reasonable method to identify a giant hail null case is to examine storms with reported hail <4 in diameter to determine whether similar or unique signatures are present with hail ≥4 in. The POD vs. FAR comparison inherits some earlier concerns we addressed in the previous review utilizing a low-resolution hail database (Storm Data) due to the high uncertainty of sufficient report density to accurately assess the true hail fall character of a storm. The most reliable comparisons must utilize high-resolution hail data (SHAVE, HailSTONE) to accurately assess whether a hail event is actually a null event or from a lack of reports.

Essentially, we have already conducted analysis on null giant hail cases, although we elected not to use the "null" nomenclature but rather use a hail size bin comparison. We contrasted storms producing large (1.75–2 in) versus giant (4"+) hail, utilizing data from both low and high resolution datasets (Storm Data and SHAVE). A statistically significant relationship of the rotational velocity and storm-top divergence parameters were established between the giant hail cases from the large (null) events in both dataset sources. As noted in the paper, other parameters performed quite poorly as discriminators of giant hail. We believe these data from the comparisons of storms producing giant hail and those that did not satisfies the concerns addressed by the reviewer.

Lastly, we must assume the reviewer took some of the giant hail parameters out of context during their impromptu analysis. When analyzing a storm for giant hail potential, it is very important to consider all giant hail signatures concurrently. For instance, basing an analysis on only Z heights without also utilizing the other suggested parameters would yield a substantial FAR. It seems very unlikely that the storms which failed to produce any hail were supercellular with characteristics of moderate/strong Vr and STD values.

Chances are, if hail fell in an area with dense enough population to get a 1 in stone reported, then larger sizes typically are reported as well, regardless of whether a WFO is making calls. Four-inch stones are rare and notable and people like to report them. It's possible maximum hail size reports are exaggerated anyway, which would help to account for under-sampling. All of this is possible.

There are several speculations by the reviewer here. While we would generally agree the probability of giant hail stones being identified is greater in denser populations than in rural areas, there are still typically fewer stones of the 4+ in caliber compared to smaller stone sizes, naturally lending to fewer *people impacted and a lower probability of giant hail reports. We strongly disagree with the notion that "people like to report 4 inch stones". Ongoing research (Blair and Leighton) from an in-person hail survey conducted on the 15 September 2010, Wichita, KS giant hail event revealed that individuals that collected 4+ in hail made zero reports to the NWS, with 6% reporting to the media or law enforcement. Therefore, 94% of individuals with 4+ in hail in their hand made no effort to report these stones.*

Please add the above [NWS verification and SHAVE] information to the paper, also citing the limitations due to few data points.

We have expanded our discussion in Section 2 to summarize the several limitations of Storm Data we mentioned in the first review.

Define "unique event". What kind of separation was required between swaths? 10 feet, 10 miles?

In this brief investigation, we defined a SHAVE unique event of giant hail as a singular event per storm, meaning only the largest diameter hail size identified by SHAVE was compared with the largest size listed in Storm Data. This eliminated multiple reports from the same storm, as only one value per storm was analyzed.

[re: Rasmussen table] They were only deficient within the context of my other comments, fix those and there will be no deficiencies.

[Minor comments omitted...]

REVIEWER B (Arthur Witt):

Initial Review:

Reviewer recommendation: Accept with minor revisions.

Major Comment: The paper is very well written, with a thorough description of the methodology and results. My only significant comment relates to the lack of any mention of the National Severe Storms Lab's Severe Hazards Analysis and Verification Experiment (SHAVE) in the paper:

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

Meyer, T. C., J. M. Erlingis and K. L. Ortega, 2010: Comparing radar signatures to high-resolution hail reports.

Given the list of other projects mentioned in the paper, it would appear appropriate to also include mention of SHAVE.

Thank you. We have incorporated SHAVE data into the revised manuscript for hail size and database comparison.

[Minor comments omitted...]

REVIEWER C (Leslie R. Lemon):

Initial Review:
Reviewer recommendation: Accept with major revisions.

Specific (substantive) points:

Abstract—The authors define "giant hail" as hail >4 in diameter. Historically, beginning with Keith Browning's "giant hail", this term was defined as hail greater than 2 inches in diameter. I recommend that the authors states that their definition, is for the purposes of this paper, 4 in or greater.

Browning's initial definition of "giant hail" as 2 in diameter seemingly has been replaced by the term "significant". Following Hales (1988) classification of significant hail as having a 2 inch diameter threshold, several formal works over the past decade have followed suit (Doswell et al. 2005; Doswell et al. 2006; Davies 2006; Donavon and Jungbluth 2007; Brunner et al. 2007; Horgan et al. 2007; Gallus et al. 2008; Jewell and Brimelow 2009; Bunkers et al. 2010). Knight and Knight (2001) explicitly define 4 in diameter hail as "giant". We believe this sufficiently describes hail of this magnitude for not only the purpose of this paper, but for widespread use in the operational community. We have massaged the introduction to provide a cleaner read regarding the classification of hail sizes.

The authors state: "The short-term prediction of these events has been challenging. Since 2005, only 7% of convective warnings and severe weather statements issued by the National Weather Service accurately predicted a maximum hail size ≥102 mm prior to occurrence, with an average underestimated size error of 54.1 mm (2.13 in)". Does this mean that the average error in all hail warnings from the NWS is 2.13 in? Or is this the error when NWS warnings were issued for "exceptionally" or "extremely large" hail"?

This error was calculated only when giant hail occurred from the warning issued prior to the report. We have reformulated the sentence for clarity.

[Minor comments omitted...]

The authors state: "Several radar-based techniques have been evaluated for their ability to detect severe hail (≥25.4 mm, ≥1.00 in)." This sentence and the paragraph in general seem to imply that we are no longer in the introduction, but that we are now beginning to evaluate detection techniques. Do you mean that for the purposes of this study these techniques are to be evaluated or do you mean that in the literature this has been done in the past? Is this an exhaustive list of techniques to be evaluated within this paper? Please clarify.

We have clarified the sentence to reflect a review of previous radar-based techniques that have attempted to gauge maximum diameter hail size in convective storms. This discussion summarizes some of the more substantial contributions during the WSR-88D era, but is not intended as an exhaustive description for every technique published.

What do we know about giant hail growth and what differentiates a storm that generates modest hail from storms generating giant hail? Do we know? Are there differences in the environment? Don't we need to know these answers in order to develop forecasting skill for giant hail? Why are these 17 states those that most often experience giant hail?

Miller et al. (1998) discusses specific trajectories that lend to giant hail growth. Whether giant hail growth trajectories are unique to each storm or [can be generalized] to others is unknown. We would expect a relatively unique set of conditions that would favor giant hail growth, but those precise conditions remain unknown. The variability of hail growth in a single storm itself illustrates the complex nature of such a process. Ryan Jewell (SPC) has thoroughly investigated different hail size categories from RAOB data, and some environmental parameters show a few discriminating signals. A full examination into the environmental parameters is beyond the radar-based scope of this paper. Instead, this paper attempts to *provide some guidance for real-time warning operations revealing consistent radar-based characteristics associated with giant hail producing storms.*

The authors state: "Hail size verification from giant hail events was performed on convective watches and warnings issued by the SPC (Fig. 1a) and NWS (Fig. 1b) throughout 2005–2009 in 17 states generally between the Continental Divide and Mississippi River Valley region." Please re-write this sentence breaking it up into two or more sentences. How is verification done? I believe that in the introduction the authors need to discuss the problems inherent in hail size estimation, in observers actually observing this hail, and then in reporting the event. The SPC does not issue warnings.

We have expanded our discussion on the inherent limitations of estimating and reporting hail, although this does not belong in the introduction section as suggested. We have restructured the sentence in question and have used different wording to limit confusion. Verification methods are discussed in the methodology section.

The authors state: "It is particularly surprising that 20% of NWS warnings forecasted penny to quarter-sized hail (19–25 mm; 0.75–1.00 in.) during giant hail events. It is speculated this poor performance might be attributed to a radar operator selecting a default minimum severe hail size when issuing a warning, the lack of timely update statements, workload limitations, or the deficiency of recognizing supportive signatures." Poor wording. Warning forecasters issue warnings not "radar operators". Do we know the "supportive signatures" for giant hail? I think not. In fact, if we did, you would not be writing this paper. If we did know, then we might be surprised that warning forecasters did not recognize these supportive signatures when present.

The sentence has been reworded. Operational lingo frequently identifies a person working the radar during warning operations as a "radar operator". Since this wording is relatively inconsequential, we oblige and change the wording to "warning forecaster". As for the deficiency of recognizing supportive signatures, you are correct that no literature has explicitly dealt with investigating the upper threshold of hail sizes. However, in the context of the sentence regarding warning hail forecasts of 0.75–1.00 in., we would argue that at least some anecdotal evidence should have existed to suggest that the storm was capable of producing something larger than a one inch diameter stone. We will adjust the sentence to better reflect a deficiency of the state of the science, rather than a deficiency of preexisting signatures.

Does Fig. 1 deal only with extreme hail size reports and not average hail sizes?

We are not sure what is meant by average hail sizes. The figure states "Giant hail size verification of a) SPC watches and b) NWS warnings during 2005–2009 in the study domain." T his recent performance review of giant hail in watches and warnings focuses solely on the scope of the paper, which is hail ≥101.6 mm (4.00 in) in diameter. The NWS is tasked to accurately warn for the maximum sized hail expected, not the average size of hail in a storm.

The authors should recognize that the warning generation software used by the NWS to facilitate rapid warning issuance includes pre-worded warnings. For example one of these includes 60-mph winds and quarter size hail. Thus, there is a bias toward hail sizes used in pre-worded warning formats. Otherwise, hail sizes actually reported are also often used in warnings and the probability is that these are not the extremes produced by the storms in question.

We contend that this is not generalized to each NWS office, as each office is able to customize the warning generation software (WarnGen) and likewise what parameters are pre-selected to expedite the dissemination process. Choosing specific hail sizes in WarnGen is a simple one-click process, and the warning forecaster should consciously select the appropriate maximum hail size anticipated. Hail reports are frequently utilized in warnings, although they should supplement, not replace, maximum hail size forecast information contained in warnings.

Data and methodology—Are the authors aware of the Severe Hail Verification Experiment (SHAVE) project and the hail data base it has accumulated (Ortega, et al, 2006)? I believe the data from this project could be used to provide ground truth and thereby accurate verification information for a subset of your overall data set.

We are familiar with SHAVE and have incorporated these data into the study.

The implication is that your section 2 or forecast performance is the location in the paper where techniques or warning criteria are discussed. However, here under data methodology, there is a long list of storm or radar characteristics to be examined. Are these 'warning criteria'? They are not introduced as such but my assumption is that these are indeed warning criteria. If so, would they be more appropriately placed under "forecast performance"? On the other hand, should you combine forecast performance and this list of 'indicators' or 'criteria' under your 'results'? Please revisit the organization of this paper.

Section 3 (previously section 2) identifies and quantifies the challenge of detecting giant hail by reviewing past warning performances of this phenomenon. Separately, the long list of storm/radar characteristics under section 2 (previously section 3) belong under the methodology section, as each subset describes the process of how the data were collected and any quality assurance methods. We do not want to mislead the reader by combining and advertising this long list of radar parameters as "warning criteria", since many of these parameters are shown to have little utility to identify giant hail. This comprehensive list was required to investigate the usefulness of multiple interrogation methods, and the results/subsequent sections of the paper reveal the most important 'warning criteria' signals for giant hail.

"Additional radar sites were utilized in conjunction with the closest site to the report in order to collect near storm-summit information." Please note that there are several other radar limitations as well as the actual sounding characteristics that contribute to making radar height estimates to be rather poor in accuracy. Using more than the closest radar site will help improve the accuracy to a limited degree only. Consider the distance between beams and the beam width itself.

We have already accounted for some of these limitations in the paper and would encourage readers to refer to Maddox et al. (1999) and Brown et al. (2000) for a more thorough background on the subject matter. We have added a brief mention of beam broadening limitations that may result in lower reflectivity at far distances from the radar.

Did this study confine data to legacy resolutions or did you also use super-resolution (super res) data? This is important because the super res data has a higher error rate than does legacy data but more importantly reaches higher values than does legacy data. For example, we have observed values reaching 80–82 dB*Z* when with the legacy data we very rarely would reach 78 or 79 dB*Z*. (Later I see that you do mention super res data. This too belongs in the intro or data methodology).

Yes, we did utilize super resolution in the latter two years of the study period. As you alluded, we accounted for these changes in the results section of the paper. We will also mention this in the data and methodology section.

"Reflectivity profiles of each giant hail producing storm were examined during the 20 min period surrounding the report time." With certain assumptions concerning liquid water concentrations hail stones must have long residence times in order to grow to large sizes. For example for hail to grow to ~1 in it must remain in the storm for ~12 min. Hail reaching baseball size (2.75 in) must have a storm residence time of ~35 min. Hail of ≥4 in would suggest a residence time of ~50 min or more. For this reason the authors may wish to increase the 20-min period allotted. Moreover, I am assuming that unsmoothed and unprocessed (by the 3-D algorithm) base data are used.

While we would generally agree these assumptions are decent baselines for hail growth, it is worth noting that a few cases within the dataset produced giant hail within 50 minutes of storm development. Several authors of this paper also have anecdotal evidence from the field that these residence time assumptions are not always valid. Regardless, the purpose of the 20-min window was to capture a snapshot of the radar-based characteristics close to the report time to provide a consistent picture of giant hail producing storms.

We did not attempt to capture the entire residence period of giant hailstones (which would be speculative anyway), but rather the period closest to the confirmed hail-fall of giant hail.

Perhaps special emphasis should be placed on the hail-warning method which uses 60 dB*Z* at the –20º C environmental level. Lemon (1998) included the following in his TBSS paper: English (1973) points out that in the storms she studied and modeled "rain contents of 10 g m^3 (equivalent to a reflectivity of about 62 dB*Z*) were necessary to increase substantially the ... hailstone sizes." For this reason 60 dB*Z* at the –20º C [level] is taught to NWS students (NWS forecasters) at the Warning Decision Training Branch as a hail warning criteria by this reviewer. It seems that hail with these storms is often golf ball or larger. However, it seems your results differ.

I'm not sure how you are interpreting that our results differ from the concept that storms with a modest region of 60 dBZ at –20º C are often associated with 1.75 in hail or larger. More than 90% of the cases of giant hail and GBHE hail sizes are characterized by this precise threshold (see the comparison figure of maximum reflectivity at significant temperature levels). This is simply not a sufficient signal for explicitly detecting giant hail.

Has storm motion been used as a hail warning criteria? I don't seem to recall this. Storm motion has commonly been used as an alerting indicator for a severe storm. Is that what is used here?

We are not intending to imply that storm motion should be used as a hail warning criteria. Rather storm motion was recorded and used to calculate storm relative velocity values for portions of the analysis. We have attempted to clarify its purpose in the study.

It seems that even when a circulation couplet is not present but only shear; this aids in hail production and size.

We agree with the statement that shear zones may assist in the production of hail. In the vast majority of cases in these datasets, well-identifiable mid-level circulations from the velocity data were present with maximum inbound and outbound values.

Note your discussion of "storm type". Lemon (1978) suggested that storm structure be used as a radar hail-warning criteria. Specifically he was looking for the supercell storm structure as defined by Browning and the very similar structure of an organized multicell storm.

Supercell storm structure was present with the vast majority of cases, and should be considered necessary for the production of giant hail.

"The rotational characteristics of V_r were recorded as either cyclonic or anticyclonic for the updrafts responsible for producing giant hail." Note that radar sampling parameters can substantially change circulation size and "rotational" speeds. (See the WDTB AWOC training material [http://www.wdtb.noaa.gov/courses/awoc/ICCore4/precursor-radarSampling/player.html] and "beam sampling").

Agreed, these are inherent limitations of the radar. When these errors were suspected, we did not incorporate those data into the database.

"The maximum storm-top divergence (STD) at the upper storm levels was examined during the 20 min sample period. " Please note that the measured storm summit divergence is a function of the radar's velocity resolution. Normally the velocity resolution of each radar is set at 0.5 m s^{-1}. This limits the velocity measurement capability of the radar to ±62 m s^{-1} or ±123 kt. In order to measure the stronger velocities commonly found in severe storm summit divergence, the radar resolution must be set to 1 m s^{-1}. This then doubles the velocity measurement capability of the radar and should be done when attempting to measure storm summit divergence. In all or nearly all cases the authors examined, the resolution was placed at 0.5 m s^{-1}. For this reason many of the "measured" storm summit divergence magnitudes in this paper may be inaccurate. And this would especially be the case with storms producing giant hail. On the other hand, the authors use GR2Analyst which uses the nearest sounding as input to a robust dealiasing algorithm that is able to 'unfold' the measured velocities beyond the 62 m s^{-1} limit much of the time.

In the majority of cases, WSR-88D velocity measurement increment was set at 0.5 m s^{-1}. It is not possible to change this velocity resolution post-storm. We believe most offices utilize the 0.5 m s^{-1} velocity resolution during severe weather operations, with the exception of extreme weather events such as landfalling hurricanes when velocities frequently will be >62 m s^{-1}. We did not observe base velocities limited to 62 m s^{-1} utilizing GR2Analyst, and the STD values did not reveal any noticeable upper limit as illustrated in the box and whisker plots. This may be partially a function of the GR2Analyst benefiting from a two-dimensional dealiasing algorithm. Therefore, we believe our STD data to be representative of values observed in an operational setting.

Please note that you can check the reliability of data bins by checking the velocity spectrum widths (SW) of these data. If SW values are at or below ~10 kt then these velocity values can be considered reliable. Data quality checks are one of the primary applications of SW.

Spectrum width was used frequently during the analysis to aid in determining the quality of velocity values and other signatures (TBSS). The quality control procedures were also similar to Witt and Nelson (1991) to avoid noisy or unreliable data.

re: Results, "Thermodynamic instability"—Note that the amount of CAPE found in the hail growth zone—between –10º C and –30º C—is also critical for large hail growth. (See http://www.wdtb.noaa.gov/courses/dloc/outline.html#topic7). Was this examined? I see that it was but only later in the paper. Better organization is needed.

The revised manuscript has removed the environmental analysis from the paper as a thorough environmental investigation is beyond the scope of this radar-based study. This is an area that could benefit from future investigations, although its utility is in question as earlier studies (Edwards and Thompson 1998; Jewell and Brimelow 2009) have found that many frequently utilized ingredients-based parameters show little to no skill in discriminating hail sizes due to the complexity of the storm-scale processes that govern hail size.

"It is also noteworthy that very large values of CAPE (4000 J kg-[1] or greater) was not necessary for giant hail, and may be detrimental to its growth." This statement begs the question. Why might this be detrimental?

We retract this statement upon further investigation of NARR-derived instability parameters compared to nearby RAOB observations. The limited resolution (32 km, 3 h) of the NARR database frequently appeared to misrepresent true atmospheric instability, especially in areas with large thermal gradients, where values at a given point may not have indicated conditions across the entire grid box. Modified observed soundings for 116 cases of giant hail revealed 44% of the events were associated with MUCAPE values >4000 J kg^{-1} (R. Jewell 2011, personal communication).

The bounded weak-echo region (BWER) is characterized by the updraft core where cloud water does not grow to detectable sizes until well aloft. Moreover, the BWER is characterized by relatively low reflectivity indicating the absence of "large amounts of supercooled cloud water". The hail growth takes place in the updraft skirts and highly reflective region surrounding the BWER but not within it.

We, along with the referenced work by Knight and Knight (2005), were referring to the BWER as both the low reflectivity region and the outskirts (top and sides) of the structure. We have attempted to clarify our description of the BWER.

In your discussion of STD I didn't see reference to Witt and Nelson (1991) and their table. Their results were rather good, or at least that was my impression.

We mentioned the Witt and Nelson (1991) paper in the methodology section. Their results were certainly a good reference point, although some overlap still existed between hail sizes and STD values. Unfortunately, their study contained only two occurrences of giant hail. We have discussed their results in the updated STD results section.

re: "…several modeling studies that have shown a broad region of moderate updraft strength (20–40 m s^{-1}) can be more favorable than an intense…" This may be true; but as a first approximation in order to grow giant hail, the updraft strength must be comparable to the terminal fall velocity of these giant hailstones. Did you consider the magnitude of those velocities (~50–60 m s^{-1})?

Yes. While that is true, we should also note that some of the hail stone's growth may occur during a period of slower descent through a lower portion of the updraft. It has been suggested some stones may gain a substantial amount of mass during this period of time. Several of the authors have observed this in the field and in photos that show large stones with unusually thick coats of "clear" ice.

How do moderate updrafts prevent stones from descending during growth?

We were referring to broader regions of moderate updraft strength surrounding an intense updraft, as discussed earlier in the paper. We also defined "moderate" up to 40 m s^{-1}, which would be an adequate updraft to suspend stones during a period of substantial growth.

Does the TBSS have little or no value when excluding cases where the signature is masked by other echo? At times the TBSS is more easily seen when using spectrum width. The TBSS is characterized by very broad spectrum width of the spike itself.

We utilized spectrum width in every case when attempting to identify a TBSS.

[Minor comments omitted...]

Reviewer References

English, M., 1973: Alberta hailstorms. Part II: Growth of large hail in the storm. *Alberta Hailstorms, Meteor. Monogr.*, No. 36, Amer. Meteor. Soc ., 37–98.

Lemon, L.R., 1998: The radar "three-body scatter spike": An operational large-hail signature. *Wea. Forecasting*, **13**, 327–340.

Lemon, L. R., 1978: On the use of storm structure for hail identification. Preprints, *18th Conf. on Radar Meteor.*, Boston, MA, Amer. Meteor. Soc., 203–206.

Ortega, K. L., T. M. Smith, and K. A. Scharfenberg, 2006: An analysis of thunderstorm hail fall patterns in the Severe Hail Verification Experiment. Preprints, *23rd Conf.on Severe Local Storms,* Amer. Meteor. Soc., St. Louis, MO.

Second review:

Recommendation: Accept with minor revisions.

Overview: Please find attached my newly marked-up copy of the Blair et al manuscript. Note that parenthetical additions are simply my comments; however, I do make a few changes to wording elsewhere. I believe this paper is an excellent aid to forecasters and I am pleased to be a reviewer. In this go-around I was less critical and let some things go that I might have otherwise mentioned. There remain at least two areas that need attention.

Substantive Comments: First, once again the authors neglect a discussion of all the sources of radar-estimated height errors. This is absolutely critical when considering the plethora of radar estimated heights in this paper as well as operationally. They even add to the sources of error by interpolation. I would require them to include this discussion prior to publication.

We have included a discussion on the inherent uncertainty with the implied precision of radar height measurements in Section 2.

They again resist my suggestion to explicitly list those signatures that they believe are outdated or erroneous. I think this is very appropriate considering that this may be a major contribution of this paper and that they mention this in the summary and conclusions section of the paper. However, it is important that they not suggest abandoning signatures because they are not good discriminators between large hail and giant hail. The ability to judge a storm as a large hail (>1 in) producer is important to the severe thunderstorm warning. But the discriminators they highlight that may help us [to] realize we are dealing with a possible giant-hail producing storm are critical. So I believe the value of their paper may be important to both large hail identification and giant-hail identification.

We certainly do not advocate abandoning radar signatures that proved unreliable for differentiating between large and giant hail for other hail sizes without additional research. Unfortunately, we're not positive we can comment on the relevance of the examined parameters with regards to large (1.75–2 in diameter hail) versus sub-severe hail (<1.00 in) or specific hail size forecasting for all sizes of hail. The scope of the paper was generally to distinguish whether or not unique characteristics existed with giant-hail producing storms and if these radar-based signals could be deciphered from other storms producing large hail. Therefore, the manuscript provides guidance largely to these uses, and the conclusion summarizes the most important giant-hail discriminators. It should be fairly clear to the reader which signatures are outdated to use for hail size forecasting for large versus giant hail, but we cannot speculate whether these parameters are completely useless for smaller versus large hail sizes. It's possible [that] some of the signatures that performed poorly for giant hail might have improved utility in cases with lesser-organized convection, in the absence of storm rotation (e.g. 1 in diameter hail versus 1.75 in).

Additionally, while the paper generally uses references well, I was disappointed that they neglected a few. I found it very interesting that they emphasize a very important "signature" for giant hail and that is the "well-organized" supercell storm structure. A supercell by its very nature is "well-organized". But they do find that when the BWER can be radar-resolved this adds to the possibility of giant hail. That is important. But perhaps of less importance, there has been a long history of the relationship of damaging hail to storms exhibiting the supercell storm structure. In fact, it has been long used in Soviet-bloc countries and others as the indicator to begin using hail suppression techniques on those particular storms. I included some of this discussion and references in my paper emphasizing the importance of storm structure to large hail identification (Lemon, 1978; as well as in my NSSFC tech memos 1 and 3). But alas, this is not important to their conclusions and they are free to omit this.

We have included a brief mention of the Lemon (1978) paper for completeness related to his findings and summary of the WER/BWER as a hail signature.

This will be a very important paper to our WDTB forecaster training. The authors should be proud of their work and contribution.

Third review:

Recommendation: Accept.

Overview: I have read all the attachments you supplied including the final version of Blair et al., and I am very happy to recommend that you publish without further revisions this fine contribution to the science of hail as well as the associated warning and forecast problems.