7B.4   **EVALUATING A HAIL SIZE DISCRIMINATION ALGORITHM FOR DUAL-POLARIZED WSR-88Ds USING HIGH RESOLUTION REPORTS AND FORECASTER FEEDBACK**

Kiel L. Ortega[1], Alexander V. Ryzhkov[1], John Krause[1], Pengfei Zhang[1] and Matthew R. Kumjian[2]
[1]University of Oklahoma/Cooperative Institute for Mesoscale Meteorological Studies and NOAA/OAR/National Severe Storms Laboratory, Norman, OK
[2]Advanced Studies Program, National Center for Atmospheric Research, Boulder, CO

1. Introduction

Identification of hail, especially hail which is considered dangerous and potentially damaging—or "severe", within thunderstorms is of the utmost importance for operational meteorologists. For U.S. National Weather Service (NWS) meteorologists "severe" hail is defined as having a diameter exceeding or equal to 2.5 cm (1.0 inch), while hail with diameters exceeding or equal to 5.0 cm (2.0 inch) is considered especially dangerous and considered "significantly severe."

Conventional detection of hail using dual-polarization data is based on the idea that large hail tumbles, leading to a differential reflectivity (Z_{DR}) near 0 dB. Regions of a storm that have large reflectivity factor (Z_H) and near-zero Z_{DR} are then assumed to contain hail. Additionally, when the radar probing volume is filled with a mixture of hydrometeor types (e.g., rain and hail), the measured copolar cross-correlation coefficieint (ρ_{HV}) will be decreased.

The problem with the conventional method described above is its oversimplification. The melting process leads to nonuniform scattering characteristics of hailstones across the size spectrum; these differences significantly affect the measured polarimetric variables. Ryzhkov et al. (2013a) investigated this issue using theoretical simulations and used the results of that modeling effort to implement a Hail Size Discrimination Algorithm (HSDA; Ryzhkov et al. 2013b). The algorithm will be described in the next section.

Corresponding author address: Kiel Ortega, 120 David L. Boren Blvd, Rm 3925 Norman, OK 73072; email: kiel.ortega@noaa.gov

Validation for the HSDA was accomplished through the Severe Hazards Analysis and Verification Experiment (SHAVE; Ortega et al. 2009). SHAVE uses publically available phone number information and a Google Maps-based data entry system, which allows students to overlay phone numbers along with radar data, satellite products and NWS watch and warning products. SHAVE collects data at higher resolution that *Storm Data* and collects hail reports of all sizes, including 'no hail' and non-severe hail. Two example cases from SHAVE operations are shown in Figure 1.

Since the HSDA is meant to be implemented for NWS use, the algorithm was tested in a quasi-operational setting within NOAA's Hazardous
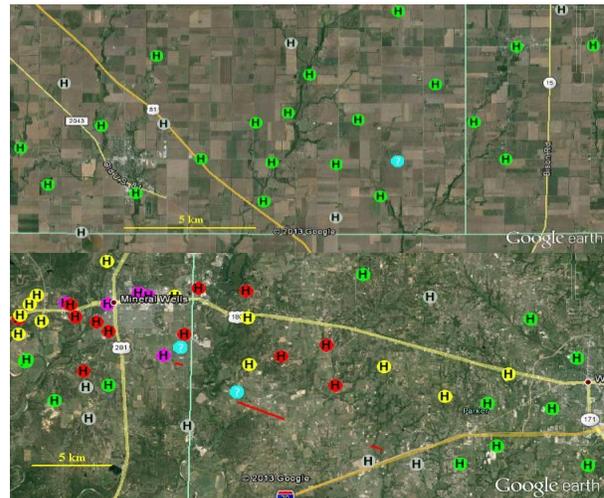
Figure 1: SHAVE reports from rural Kansas on 29 March 2012 for a marginally severe thunderstorm (top). SHAVE reports for a supercell in north central Texas on 15 May 2013 which produced hail up 11.4 cm; red lines denote tornado tracks from the storm (bottom). Gray icons: no hail; Green: up to 2.5 cm diameter; Yellow: 2.5 to 5.0 cm diameter; Red: 5.0 cm to 7.6 cm diameter; Magenta: greater than 7.6 cm diameter.

Weather Testbed's Experimental Warning Program. The algorithm was coded into the Open Radar Products Generator, which allowed for the HSDA to be displayed directly with WSR-88D within AWIPS-II, the forthcoming graphical user interface for NWS forecast and warning operations. The HWT/EWP activities allowed for researchers developing the algorithm to gather feedback directly from operational meteorologists to help focus future verification efforts and algorithm development/modification.

## 2. The Hail Size Discrimination Algorithm (HSDA)

A complete discussion on the HSDA can be found in Section 2 of Ryzhkov et al. (2013b).

The first step of the HSDA is to run the operational Hydrometeor Classification Algorithm (HCA; Park et al. 2009). When the HCA identifies 'Rain/Hail', the HSDA is then triggered to run and return 1 of 3 hail classifications: Small (dia. < 2.5 cm); Large (2.5 cm < dia. < 5.0 cm); Giant (dia. > 5.0 cm). The HCA identifications of 'Rain/Hail' are then replaced by the HSDA size classifications.

The HSDA takes into account the different stages of melting hail by binning the height of the radar beam relative to the height of the wet bulb temperature (T_w) equal to 0°C and -25°C. There are 6 bins:

1) $H > H(T_{w\_-25C})$
2) $H(T_{w\_0C}) < H < H(T_{w\_-25C})$
3) $H(T_{w\_0C}) - 1$ km $< H < (T_{w\_0C})$
4) $H(T_{w\_0C}) - 2$ km $< H < H(T_{w\_0C}) - 1$ km
5) $H(T_{w\_0C}) - 3$ km $< H < H(T_{w\_0C}) - 2$ km
6) $H < H(T_{w\_0C}) - 3$ km

Each height bin has different membership functions for Z_H, Z_{DR} and ρ_{HV} and are shown in Fig. 2. For each hail classification the sum of the membership functions is found and the hail classification with the highest sum is the designated class.
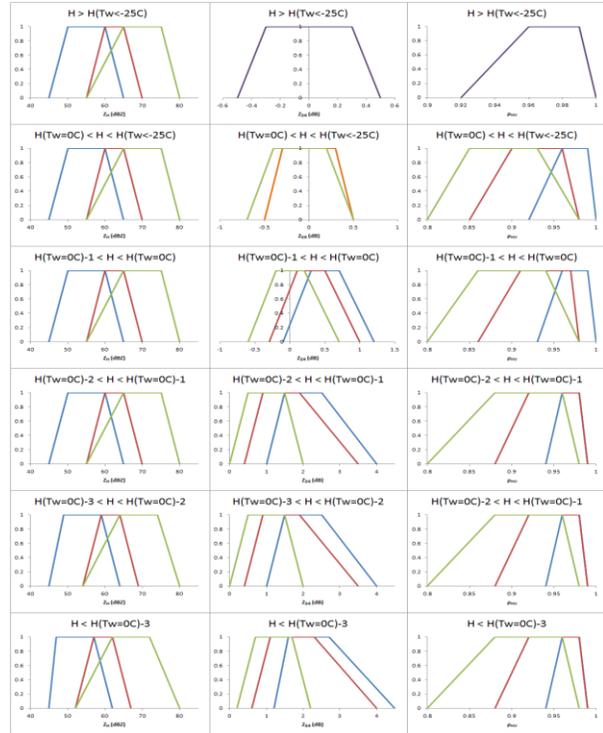
Figure 2: Membership functions for the different radar parameters and height bins. Lines: blue-small hail; red-large hail; green-giant hail; orange-large and small hail; purple-all hail.

3. Dataset

Thirty-nine cases yielding 1,806 SHAVE reports were used for this study. Five hundred eighty-seven reports were of 'no hail', 665 were small hail, 459 were large hail and 95 were giant hail. All reports were within 120 km of the radar site. SHAVE reports were first matched to the radar data by finding radar volume which had the highest Z_H at the lowest tilt (0.5 degrees) at the reports location. A 2 km by 2 degree window was then centered on the report and radar data from all tilts was collected from inside the search window. Additional information such as the beam height and height of T_w equal to 0°C and -25°C, was appended to the search window data.

In an attempt to better relate radar data aloft to surface hail reports, 14 storms with very good SHAVE report coverage were tracked, yielding 117 volumes for analysis. A larger window (6 km by 6 degrees) was centered over a subjectively determined storm location, usually near the Z_{DR}

Matching	POD	FAR	CSI
Point	0.243	0.670	0.163
Window	0.205	0.658	0.150
Window (hail only)	0.766	0.644	0.305

Table 1: Summary of HSDA skill scores for different matching techniques.

column, bounded weak echo region or highest reflectivity aloft. The time from the current storm location to each report was calculated using the location and storm motion. The largest report in each 5 minute time bin was then found.

4. Algorithm Performance

The HCA and HSDA identifications were compared in several ways to the SHAVE hail reports. The simplest comparison was to take the lowest tilt available and compare the pixel value at the report location to the maximum hail size reported. This type of matching leads to very poor skill scores with the probability of detection (POD) equal to 0.243, the false alarm ratio (FAR) equal to 0.670, a critical success index (CSI) of 0.163 and a Heidke Skill Score (HSS) equal to -0.228 (Doswell et al. 1990). While there were a large number of false alarms (435), there were even more misses (666). Thus the major issue revealed through this comparison is that the HCA is not identifying 'rain/hail' but instead some type of rain (usually 'heavy rain') for many locations receiving hail of any size.

In an attempt to reduce the false alarms and misses, the most common HCA detection within the 2 degree by 2 km window was found. Detections of hail were first considered without the HSDA classifications; if hail was the most common HCA detection in the window, then the most common HSDA classification was found. The result of this matching led to a reduction in false alarms (357), however the misses increased (723). The resulting POD equaled 0.205, FAR equaled 0.658 and the CSI equaled 0.150.

In order to evaluate the skill of the HSDA, only reports with hail detections within the window were considered and the most common HSDA classification was used for comparison. This

reduced the number of reports from 1,806 to 1,455. The HSDA detections were scored strictly, thus false alarms were HSDA detections larger than the reported hail and misses were HSDA detections small than the reported hail. The POD equaled 0.766, the FAR equaled 0.644 and the CSI equaled 0.305. The large number of false alarms (877) could be from only a single pixel, however using a minimum of 3 pixels led to a similar FAR (0.660).

The tendency for the HSDA to overestimate hail size was observed by both researchers and NWS forecasters during the HWT EWP. Giant hail detections were very common, even though most storms were not producing giant hail. Since the NWS does not have a tiered severe thunderstorm warning, forecasters noted that the high false alarm rate was not necessarily a negative aspect of the algorithm as the giant hail detections gave them more confidence in the presence of at least severe hail.

It should be noted that even with the window scoring, the algorithm was scored very strictly. Thus while a window might have contained identifications large hail, if there were more pixels of giant hail the window was scored as giant hail. Thus the FAR may be a result of the strict scoring and matching methodologies and not necessarily incorrect identifications. However the current method and resulting scores suggest that areas of identifications of larger hail size categories may be too large.

5. Tuning the Algorithm

Comparisons of the SHAVE hail reports were made to the distribution of values within the different search windows. For the lower levels, the 2 degree by 2 km window centered on each report was used. The distributions were created by first finding the median of each parameter. All values from the median towards the direction typical for hail (i.e., lower Z_{DR}, higher Z_H, lower ρ_{HV}). All of these values were then combined for each hail size category. The resulting distributions were then compared to the trapezoidal membership functions.

3
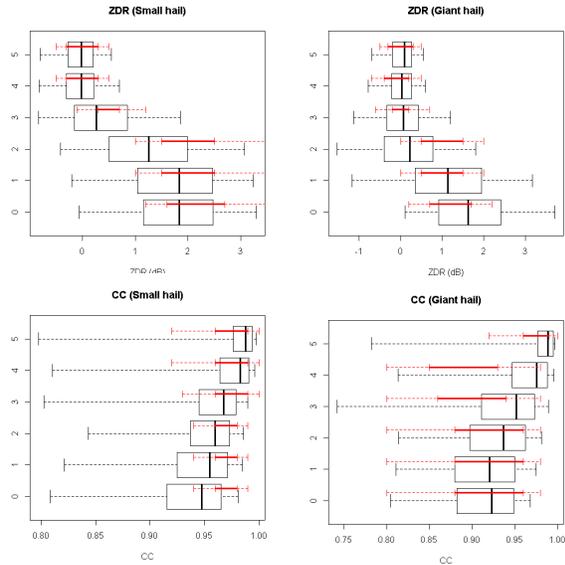
Figure 3: Box and whisker plot (whiskers: 5/95th percentiles; box: 25/75th percentiles; black line: median) for Z_{DR} (top) and ρ_{HV} (bottom) for small (left) and giant (right) hail. Current HSDA membership trapezoid functions are shown in red (where the function equals 1, the line is solid).

Two-dimensional histograms of Z_H-Z_{DR} and Z_H-ρ_{HV} were created to further investigate the parameter space for each hail size category. This was done using both the 2 km by 2 degree window centered on each report (useful for low altitude comparisons) and the 6 km by 6 degree window centered on the subjectively determined storm center (helpful for comparisons to radar echo aloft). Comparisons of these histograms might also reveal the heights in which the two different techniques are better for comparisons.

a. Distributions

Vertical profiles of the distributions for the different parameters show mixed agreement with the theoretically derived membership functions. However, both sets of vertical profiles generally show the same characteristics (Fig. 3) and differences could be due to the matching methodology used. For instance, the increase of Z_{DR} for all hail sizes below the melting layer and the smaller decrease for larger sizes is present in both the modeled and observed distributions.
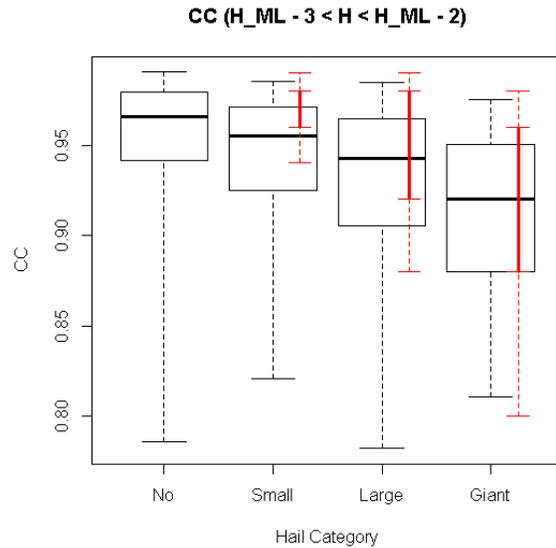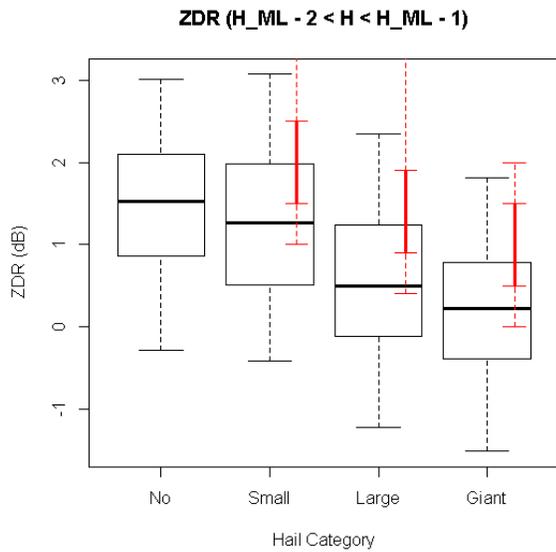

Figure 5: As in Fig. 3, except for ρ_{HV} for the layer between 2 and 3 km below the melting layer.
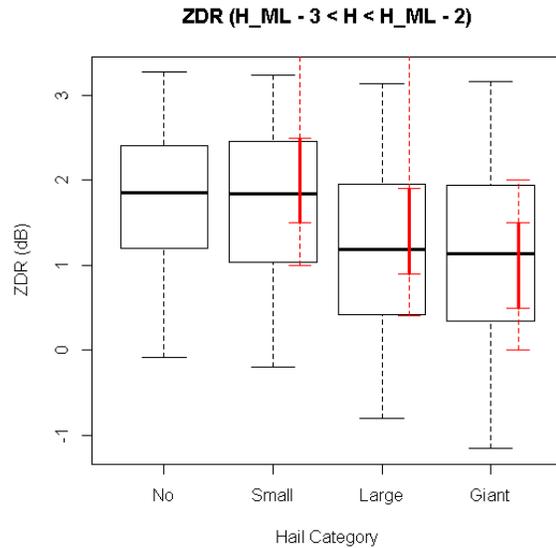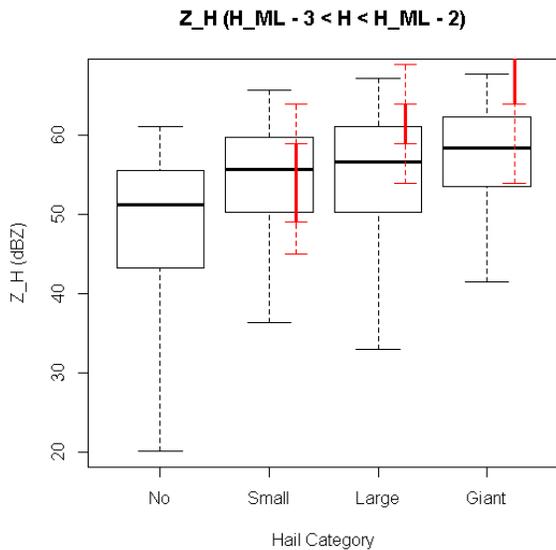

Figure 4: As in Fig. 3, except for Z_{DR} for the layer 2 and 3 km below the melting layer.

The distributions for ρ_{HV} revealed fairly good agreement with the current trapezoidal membership functions, especially for giant hail (Fig. 4). The agreement begins to break down once the altitude of the radar beam approaches the melting layer (Fig. 3). However, the breakdown of the agreement may not be a result from actual inaccuracies, but instead a breakdown

4

ZDR (H_ML - 2 < H < H_ML - 1)

Figure 6: As in Fig. 3, except for Z_{DR} for the layer between 1 and 2 km below the melting layer.

Z_H (H_ML - 3 < H < H_ML - 2)

Figure 7: As in Fig. 3, except for Z_H for the layer between 2 and 3 km below the melting layer.

of the matching of reports at the surface to valid radar echo aloft.

Z_{DR} distributions (Fig. 5) show fairly good agreement except for the long tails towards larger values of Z_{DR}. At the lowest altitude bin (3 km below the melting layer), there's a shift within the giant hail distribution towards larger Z_{DR} values. For the layer between 1 and 2 km below the melting layer the distributions suggest the current
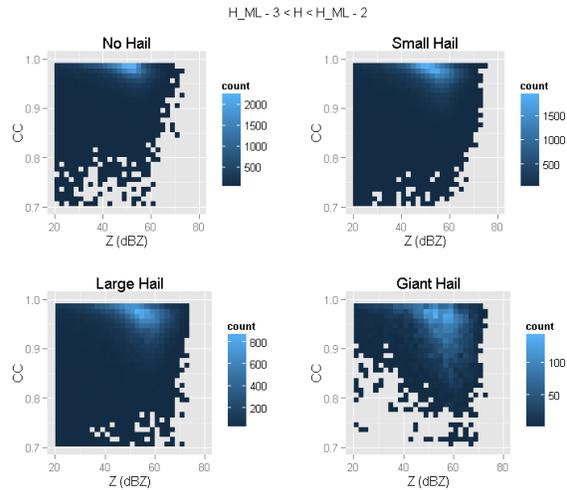

Figure 8: 2D histogram for Z_H and ρ_{HV} for the layer 2 and 3 km below the melting layer. The count is the number of pixels within the 2km by 2degree window which had the Z_H-ρ_{HV} relationship defined by the bin.

membership functions do not accurately cover the observed distribution of Z_{DR} values and downward shift in Z_{DR} values is needed (Fig. 6). However, this may be from the matching of the report at the surface to radar echo aloft.

Z_H distributions reveal that the trapezoids for large and giant hail cover values which are much higher than is observed (Fig. 7). Also, the shifts towards higher Z_H are very subtle, which suggests difficulty in discriminating between the hail size classes using only Z_H.

b. 2D Histograms (Report-based)

The 2D histograms constructed from the data within the 2km by 2 degree window around each report reaffirm the relationships observed from the boxplots. Z_H is only subtly shifted towards higher values for larger hail sizes and giant hail is associated with low ρ_{HV} values, though the distribution of those values doesn't fall within a favored space, as is observed for the smaller hail size categories (Fig. 8). The Z_H-Z_{DR} 2D histograms show clustering of Z_{DR} values below 1 dB for large and giant hail. While the 2D histograms confirm the relationships observed in the boxplots, they also show significant overlap of the parameters for different hail size categories.
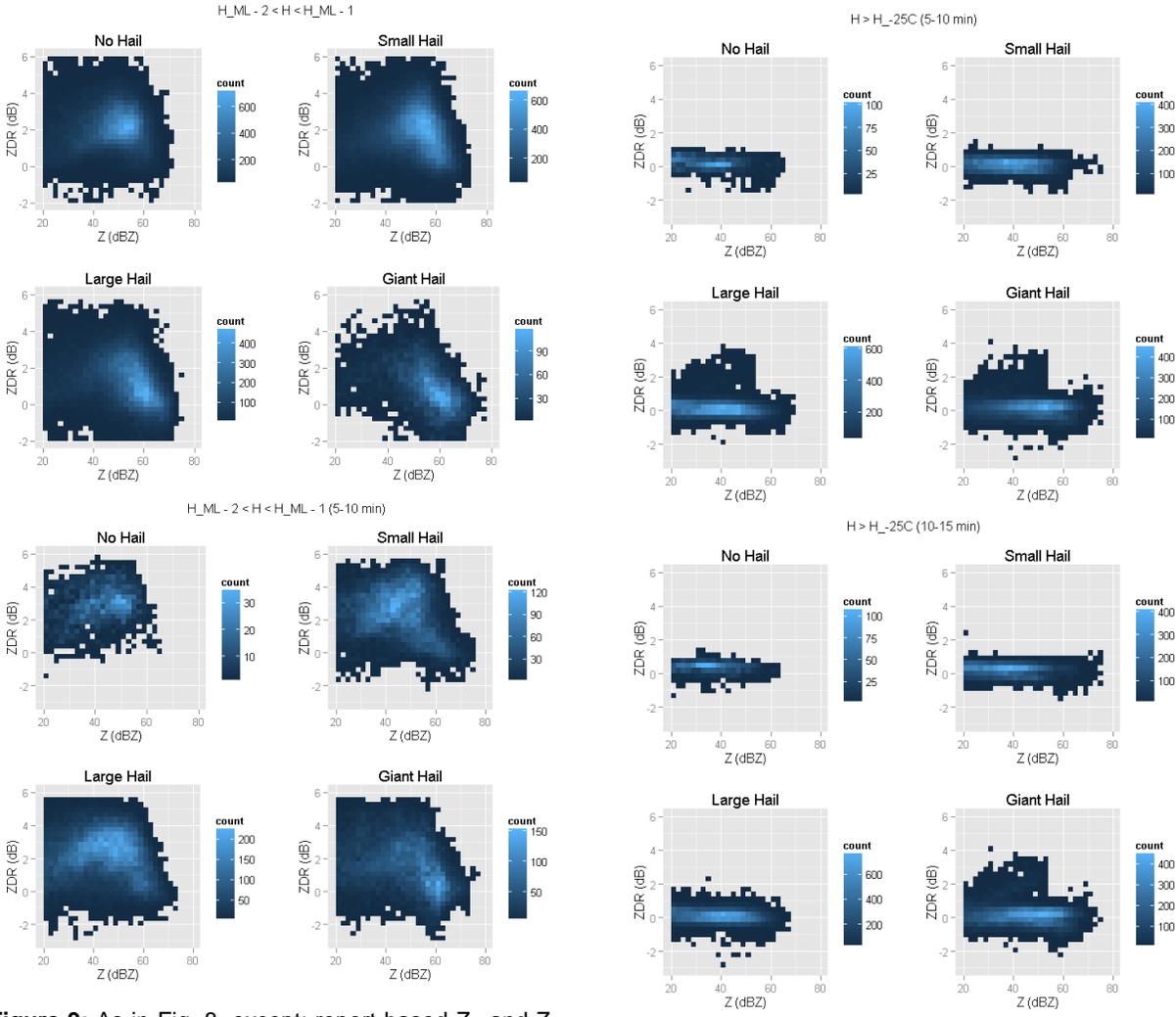
Figure 9: As in Fig. 8, except: report-based Z_H and Z_{DR} for the layer between 1 and 2 km below the melting layer (top) and the storm-based search window for the 5-10 minute lead time bin (bottom).

This may be the result of using a 2 km by 2 degree window and neighboring windows overlapping with different hail sizes for each window.

c. 2D Histograms (Storm-based)

The 6 km by 6 degree storm-based histograms do reveal differences between the 2 km by 2 degree report-based histograms, mostly for Z_H and Z_{DR} pairings. This is readily observed when comparing the Z_H and Z_{DR} relationship for the layer between 1 and 2 km below the melting layer (Fig. 9). The storm-based histograms also confirm the earlier observations that there are only subtle shifts towards higher Z_H values for larger hail.

Figure 10: As in Fig. 8, except for storm-based search window for Z_H and Z_{DR} for the layer above height of the wet bulb temperature equal to -25°C for the 5-10 min (top) and 10-15 (bottom) lead time bin.

The histograms for giant hail do reveal that for the layer between the melting layer and height of the wet bulb temperature equal to -25°C lower values of ρ_{HV} are slightly more common than for smaller (not shown). Also for the 5-10 and 10-15 minute lead time bins (even the 15-20 minute; not shown) the presence of Z_{DR} column extending above the -25°C level is present for giant hail but not for other hail size categories (Fig. 10).

6. Conclusions

This study evaluated the HSDA using high resolution reports. Also, the parameter space of Z_H, Z_{DR} and ρ_{HV} for hail-producing storms was explored in order to later tune the algorithm for improved performance.

The biggest factor for poor algorithm performance was due to the HCA not first identifying 'rain/hail' so that the HSDA could be used. When comparing hail reports to HSDA only detections within a 2 km by 2 degree window the FAR was quite high. The high FAR could be the result of the matching methodologies and future work could score the algorithm in a less strict manner; for instance, instead of matching to the most common HSDA classification in a window, the window could be classified to the correct HSDA detection for the given hail size if any pixel within the window is so classified.

The distributions and 2D histograms confirm the main purpose of the theoretical modeling which was that the parameter spaces which define hail for polarimetric variables do indeed shift depending on the state of melting hail, as well as the size of the hail. Thus conventional rules for identifying hail are most likely not valid at all heights since they may not account for melting.

Distributions created from search windows centered on each report revealed that the ρ_{HV} relationships derived from the theoretical modeling were closest to what was observed. Z_H was shown to be much less than the modeling results, while Z_{DR} was highly dependent on the height under consideration.

Two-dimensional histograms created from the search window centered on reports confirmed the distributions. Comparisons of report-based histograms to those created from a subjectively determined storm center revealed that care must be taken when comparing ground-based reports to radar echo aloft. The storm-based histograms also revealed that the Z_{DR} column extending to or above the -25°C wet bulb temperature height was only present for giant hail cases.

Future work might take into account "near-misses" as forecasters in the HWT did not see the high FAR as a necessarily bad characteristic of the algorithm. Thus evaluations of when the algorithm was near the observed hail size, but not necessarily classified size, the algorithm could still potentially score a hit. This might also help evaluate when the algorithm is completely off track with regards to the hail sizes. Future work needs to include an evaluation of the impact of search window size on the resulting parameter distributions. An evaluation of finer layers above the melting layer (e.g., melting layer to height of -5°C, -5°C to -10°C) should be explored in order to evaluate if the HSDA could be improved for detections aloft and to also explore more subjective signature detections, such as the height of the Z_{DR} column should reach a certain height prior to small (or large, or giant) hail fall.

References

Doswell, C. A., R. Davies-Jones and D. L. Keller, 1990: On summary skill measures of skill in rare event forecasting based on contingency tables. *Wea. Forecasting*, **5**, 576-585.

Ortega, K. L., T. M. Smith, K. L. Manross, A. G. Kolodziej, K. A. Scharfenberg, A. Witt and J. J. Gourley, 2009: The severe hazards

analysis and verification experiment. *Bull. Amer. Meteor. Soc.*, **90**, 1519-1530.

Ryzhkov, A. V., M. R. Kumjian, S. M. Ganson and A. P. Khain, 2013a: Polarimetric radar characteristics of melting hail. Part I: Theoretical simulations using spectral microphysical modeling. *J. Appl. Meteor., in press.*

Ryzhkov, A. V., M. R. Kumjian, S. M. Ganson and Pengfei Zhang, 2013b: Polarimetric radar characteristics of melting hail. Part II: Practical implications. *J. Appl. Meteor., in press*.