P4.8 WSR-88D SIGNATURES ASSOCIATED WITH ONE INCH HAIL IN THE SOUTHERN PLAINS

Dennis Cavanaugh[1♦] and Jessica A. Schultz[1*]

[1]NOAA/NWS Fort Worth, TX
[*]Current Affiliation: NOAA/NWS Radar Operations Center, Norman, OK

1. INTRODUCTION

On January 5th, 2010 the National Weather Service changed the minimum hail size criterion for severe thunderstorms from 19 mm (0.75 in) to 25 mm (1.0 in). Central Region Weather Forecast Offices (WFOs) serving counties in the state of Kansas have been using one inch hail as a severe thunderstorm warning criterion since 2005 while participating in a four year service assessment experiment (George Phillips, personal communication). Employees of these offices have developed and used various techniques for identifying one inch hail in thunderstorms utilizing base data from Weather Surveillance Radar-1988 Doppler (WSR-88D) data.

The goal of this study is to identify which of these methods perform best when applied to identifying the occurrence of severe hail in the Southern Plains region. Only methods that utilized base WSR-88D data were considered in an effort to gain lead time on volumetric based hail detection algorithms. This study investigates four methods for assessing potential for one inch hail in thunderstorms. These methods are as follows:

- The altitude of the 50 dBZ reflectivity echo relative to the melting level
- The altitude of the 60 dBZ reflectivity echo relative to the melting level
- The maximum dBZ at the -20°C level
- The maximum dBZ at the -30°C level

Severe hail warning criteria are developed for each of these methods on a spectrum of statistical thresholds. Each of these thresholds was tested using severe hail events in the Southern Plains in 2009, and the results were scored in 2x2 contingency tables. From these tables, skill scores are calculated for each threshold. Signal detection theory is then applied using these scores in order to compare the effectiveness of the hail detection methods used in this study. Signal detection theory is also used to determine the optimal hail detection threshold which can be used by operational meteorologists to aid in the warning decision process. Results are compared to recommended warning criteria as specified in Donavon and Jungbluth (2007, hereafter DJ07) which introduced the altitude of the 50 dBZ reflectivity echo as a useful hail detection method. These methods were also tested operationally at WFOs Fort Worth, TX and Amarillo, TX in early 2010. The results of the real-time testing are used to gain insight as to how hail detection methods are used operationally to recommend how base data hail detection algorithms can be best applied operationally.

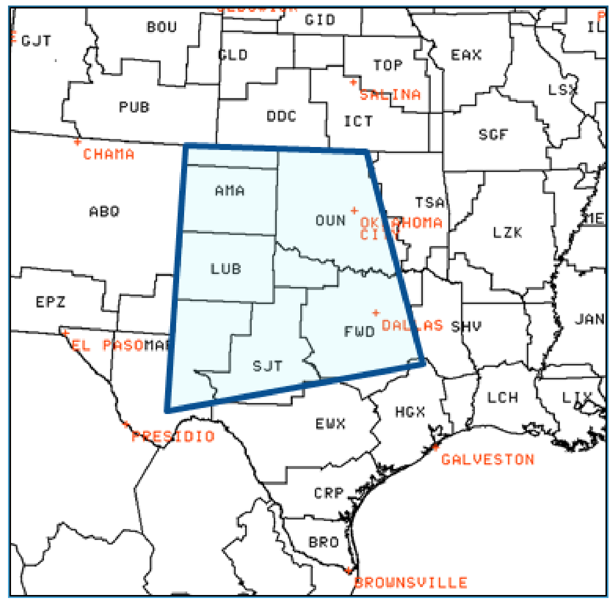


Fig. 1 - The domain of the study. The blue box represents the area where data for this study were collected. This includes 6 WFOs in the Southern Plains.

2. DATA AND METHODOLOGY

The domain used in this study to describe the Southern Plains is shown in Fig. 1. From this domain, data from WFOs Fort Worth (FWD), San Angelo (SJT), Midland (MAF), Lubbock (LUB), Amarillo (AMA) and Norman (OUN) were included. To establish a training dataset to create warning criterion thresholds, all one inch hail reports from 2008 were collected from these WFOs from the National Climatic Data Center's (NCDC) Storm Data database, yielding approximately 470 instances of one inch hail. For each event, level II radar data were collected from the NCDC WSR-88D Data Inventory website. Hail reports were manually

[1♦] *Corresponding author address:* Dennis Cavanaugh, 3401 Northern Cross Blvd., Fort Worth, TX 76137; email: Dennis.Cavanaugh@noaa.gov

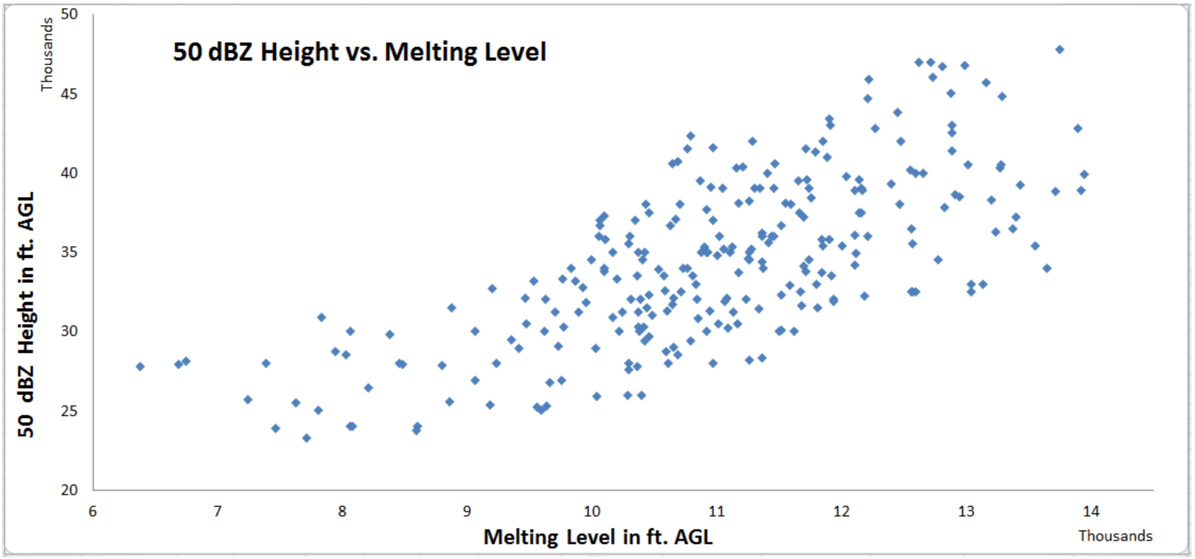


Fig. 2 – The melting level plotted as a function of the altitude of the 50 dBZ reflectivity echo in feet AGL. The plotted data are the 260 storms that make up the training database from 2008. The majority of the data fall between the 8,000 and 13,500 ft melting levels.

compared with radar data to quality control the one inch hail data. Quality control steps were employed based on criteria applied in the DJ07 study. Most notably, a storm had to have been within 5 miles of a hail report no more than 15 minutes before the hail report time to be included in the study. Multiple one inch hail reports from the same storm were included as one report to avoid biasing the data towards one particular storm. After applying these filters, the training database consisted of 260 one inch hail reports.

The evaluation database was comprised of 249 storms from 7 severe weather episodes that occurred in the Southern Plains in 2009. Events were chosen in the spring, summer and autumn in an effort to avoid bias towards a particular time of the year. Any storm associated with a hail report (of any size) was included in the scoring phase of the study. The authors also subjectively included additional storms that had high reflectivity profiles which were similar in intensity to storms that produced severe sized hail but were not associated with any hail reports. This was done in an effort to include most reflectivity signatures on radar for which a warning decision for large hail could reasonably be characterized as difficult.

Environmental data for the study were collected from the National Centers for Environmental Prediction North American Regional Reanalysis data set. The melting level, -20°C and -30°C levels above ground level (AGL) were recorded utilizing the NSHARP utility in the General Meteorology Package (Gempak, desJardins, 1991). Data were collected as AGL following the convention used in DJ07. The environmental data were recorded at the latitude and longitude of each hail event. The range of melting levels collected expanded from approximately 1830 m (6000 ft) to 4400 m (14500 ft), with the majority of data falling in the 2400 m (8000 ft) to 4100 m (13500 ft) range (Fig. 2). The relative lack of reports on the high and low ends of this spectrum is theorized to be largely due to relatively low CAPE values during the cool season, and relatively low environmental shear values over the Southern Plains during the summer months. Both of these factors favor environments where either weaker or shorter lived convection commonly occurs, limiting the opportunity for hail growth in a convective updraft.

Most radar data were viewed using the Gibson Ridge level-II radar software package (GRLevel2). To mitigate vertical gaps in the data, multiple radars were utilized when possible and linear interpolation was performed between elevation angles when necessary.

3. RESULTS

When creating warning criterion thresholds for the methods tested in this study, the general approach to the analysis followed that used in DJ07. For the altitude of the 50 and 60 dBZ height methods, a linear regression was performed for each method. The linear regression equation for the altitude of the 50 dBZ reflectivity echo is:

$$Y = 2.49x + 7090$$

The regression equation for the altitude of the 60 dBZ reflectivity echo is:

$$Y = 2.35x + 598$$

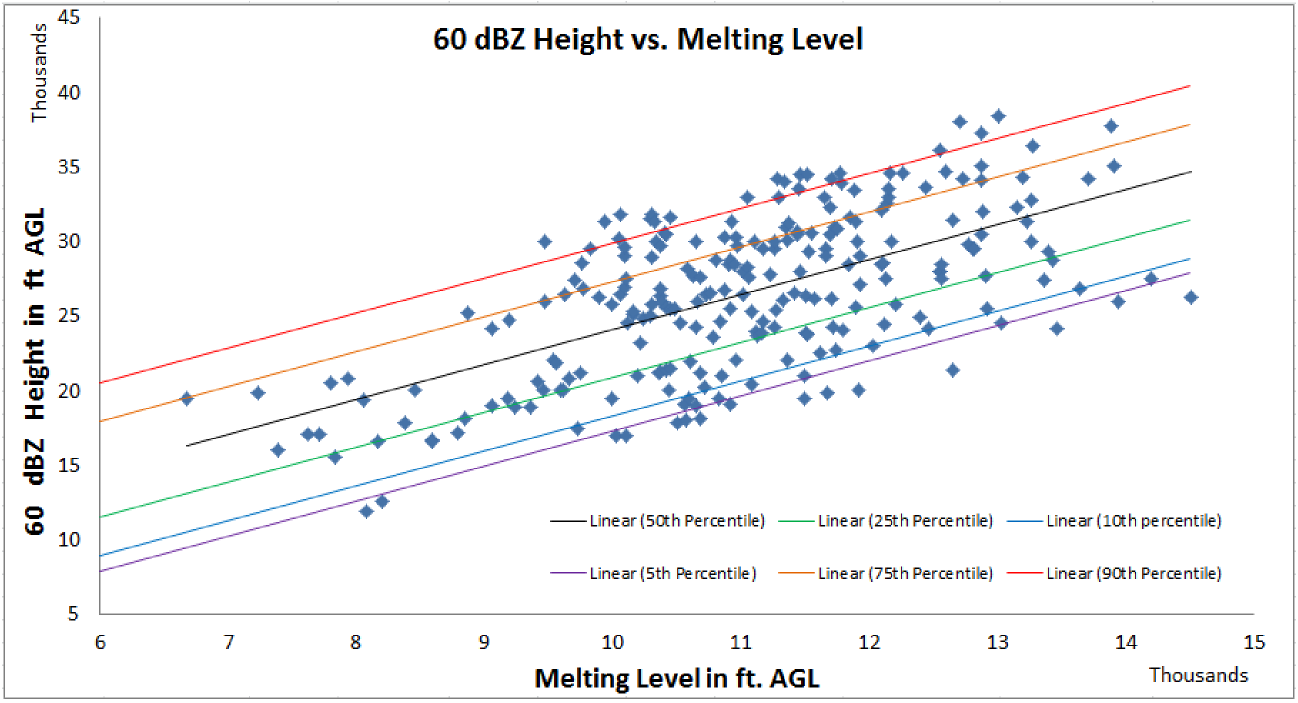


Fig. 3 - The 60 dBZ height hail detection method with the 90th, 75th, 50th, 25th, 10th, and 5th percentile linear regression lines plotted.

Quantile regression was then applied to the data in order to create equations for thresholds at the 90th, 75th, 25th, 10th and 5th percentiles. Quantile regression calculates a linear regression that places a percentile of the data above or below a given threshold (DJ07). An example the quantile regression is plotted in Fig. 3. For the maximum dBZ at the -20°C and -30°C methods, the process for calculating the thresholds was more straightforward as these data do not have any linear dependency; therefore, simple percentiles of the data were calculated. These thresholds were chosen to get a sufficient spread of data for analysis utilizing signal detection theory.

Signal Detection Theory Application

Signal detection theory provides a method to evaluate a warning system where the conflicting goals of increasing probability of detection (POD) and decreasing false alarm ratio (FAR) are sought (Brooks 2004). Receiver Operating Characteristic (ROC) analysis is a tool in the field of signal detection theory which provides a method to analyze the relative usefulness of a test where a "yes" or "no" outcome can be determined (Mason, 1982). A ROC diagram plots the false positive rate (FPR) as a function of the POD for various thresholds associated with a particular diagnostic test. The resulting ROC curve can be used as a general assessment of the diagnostic test (Fig. 4).

The FPR is the number of false positives divided by the total number of non-events in a 2x2 contingency table. The FAR is the number of false positives divided by the total number of positive detections. A fundamental difficulty in calculating the FPR in this study is determining how many correct forecasts of non-severe storms there are in any particular severe weather event. Storms were subjectively added to the evaluation database in an effort to improve the modeling of this value.

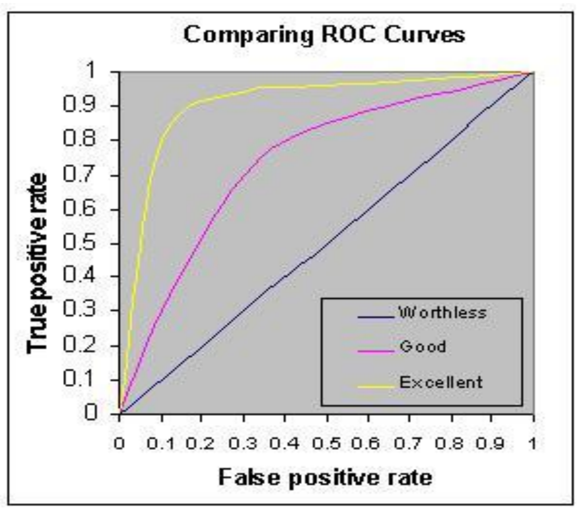
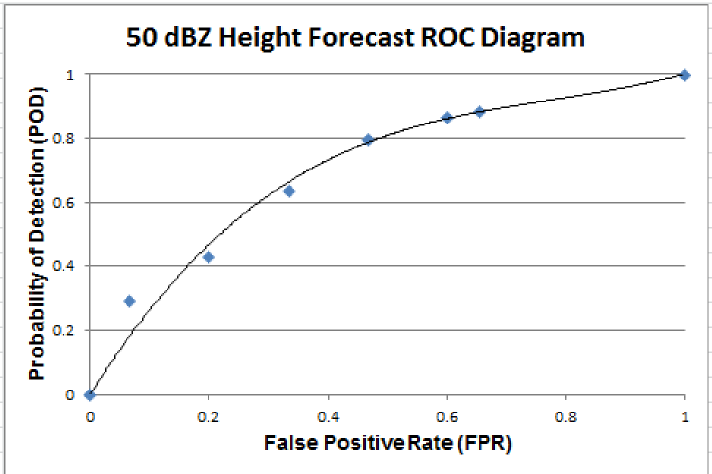


Fig. 4 - The graph represents what ROC curves would look like for a diagnostic test whose ability to discriminate between 2 classes can be described as "worthless", "good" and "excellent". Note that a diagonal line represents a test with no skill and a curve closer to the upper left hand corner represents a nearly perfect test. (From Tape, 2001)



90th %	Is Svr	Is not	75th %	Is Svr	Is not	50th %	Is Svr	Is not
Forecast Svr	51	5	Forecast Svr	75	15	Forecast Svr	111	25
Forecast Not	123	70	Forecast Not	99	60	Forecast Not	63	50
POD:	0.293		**POD:**	0.431		**POD:**	0.638	
FPR:	0.067		**FPR:**	0.200		**FPR:**	0.333	

25th %	Is Svr	Is not	10th %	Is Svr	Is not	5th %	Is Svr	Is not
Forecast Svr	139	35	Forecast Svr	151	45	Forecast Svr	154	49
Forecast Not	35	40	Forecast Not	23	30	Forecast Not	20	26
POD:	0.799		**POD:**	0.868		**POD:**	0.885	
FPR:	0.467		**FPR:**	0.600		**FPR:**	0.653	

Fig. 5 - An example of the creation of a ROC diagram using scoring from the 50 dBZ height hail detection method. Each threshold is labeled in the table in red text.

To generate a ROC curve for each hail detection method, each individual threshold had its performance in the 2009 evaluation dataset scored on a 2x2 contingency table. For each table the POD and the FPR are calculated which represents one point on the ROC diagram. After all thresholds are scored, a ROC curve can then be generated by fitting a curve to the plotted points (Fig. 5). Marzban and Witt (2001) note that the area under the ROC curve is often used as a scalar measurement of performance with an area of 0.5 representing no skill and an area of 1.0 representing perfect skill. After the ROC curves were calculated for each hail detection method, the area under each curve was calculated in order to compare the relative effectiveness of each method (Fig. 6).

Based on the ROC analysis of these methods, the altitude of the 50 dBZ reflectivity echo is the most effective hail detection method tested in this study (Fig. 6). The maximum dBZ at the -20°C and -30°C altitudes were nearly equivalent to one another in terms of performance and were only slightly less effective than the 50 dBZ method. The altitude of the 60 dBZ reflectivity echo was the worst performer. This is likely due to several storms in the evaluation database that were associated with one inch hail but did not contain any elevated 60 dBZ reflectivity core.

ROC curves were then used to determine the best warning decision threshold for the two best performing hail detection methods. Choi (1998) identified a method utilizing properties of the slope of successive line segments between points on a ROC diagram for choosing a threshold that has the best diagnostic value. Beginning at the origin, he calculates the slope of successive line segments on the ROC curve. When the slope of line segments between successive points becomes less than 1, further progression along the ROC curve is of little diagnostic value. This is because further progression on the curve yields a greater increase in FPR than is gained in POD. The threshold at which this transition occurs is then identified as having the best diagnostic value on the ROC curve. The results of these calculations are included in Fig. 7.

For the 50 dBZ height the slope falls below 1 in the E-F interval which corresponds with the 25th and 10th percentiles, respectively (see Fig. 7). This indicates that the 25th percentile should offer the best values to discriminate between severe hail producing storms and non-severe hail producing storms. This threshold had a POD of 0.80 with a FPR of 0.47. In an unbiased warning system using only this test to detect severe hail, these scores indicate that the probability of a warning being issued for a storm containing severe hail is almost twice as high as the probability of a warning being issued when severe hail does not occur. Brooks (2004) pointed out that an unbiased forecast is not always desirable depending on the relative costs of a missed event or a false alarm. If the cost of a missed hail event

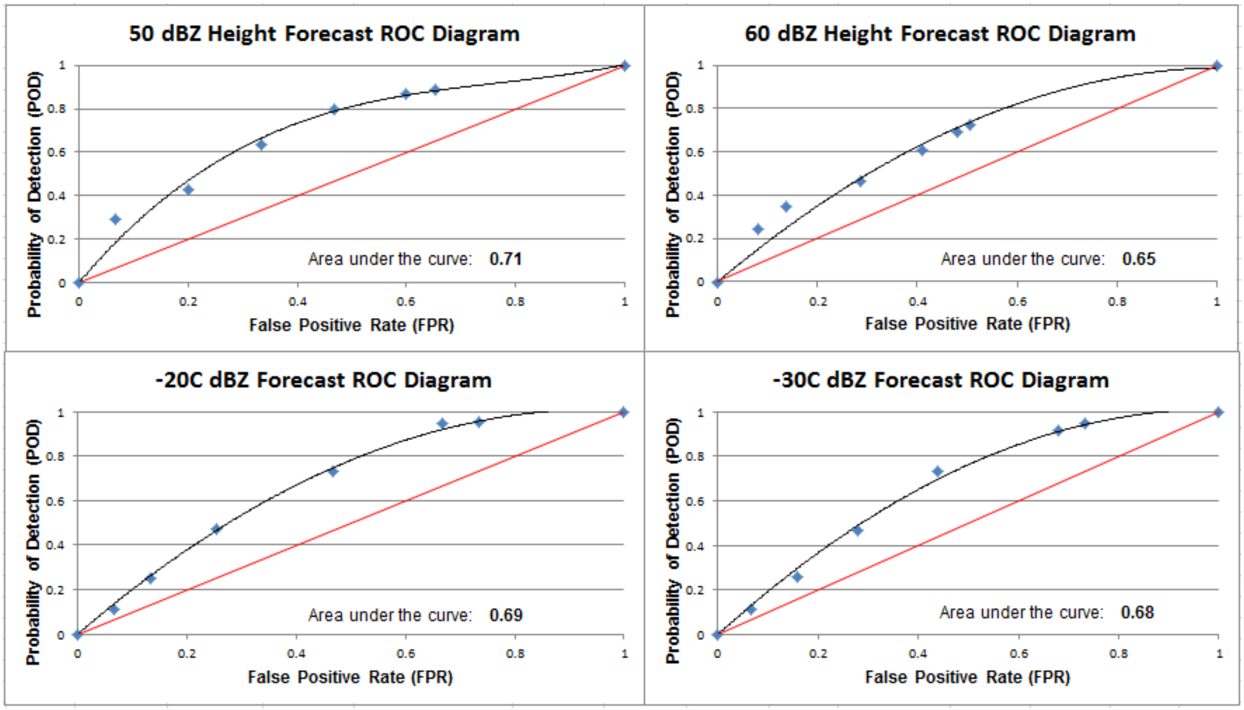


Fig. 6 – The ROC curves for each hail detection method with the area under the curve displayed in the bottom right corner of each diagram. The red line in each diagram represents the ROC curve corresponding to a hail detection method with zero skill at discriminating severe hail producing storms.

is high, then the 10th percentile may be a better choice as a warning decision threshold. The POD is higher at the 10th percentile with a score of 0.87, but the FPR increases to 0.60 which allows for a large increase in false alarm probability. It is beyond the scope of this study to assess the cost of a missed hail event. A table including both the 25th percentile and the 10th percentile values for the 50 dBZ height method are included as Table 2 at the Appendix of this paper as recommended warning thresholds for severe hail in the Southern Plains. The authors leave it up to operational forecasters to decide whether to use the unbiased recommended values in the 25th percentile column or to favor the 10th percentile values which will allow for fewer missed events but more false alarms.

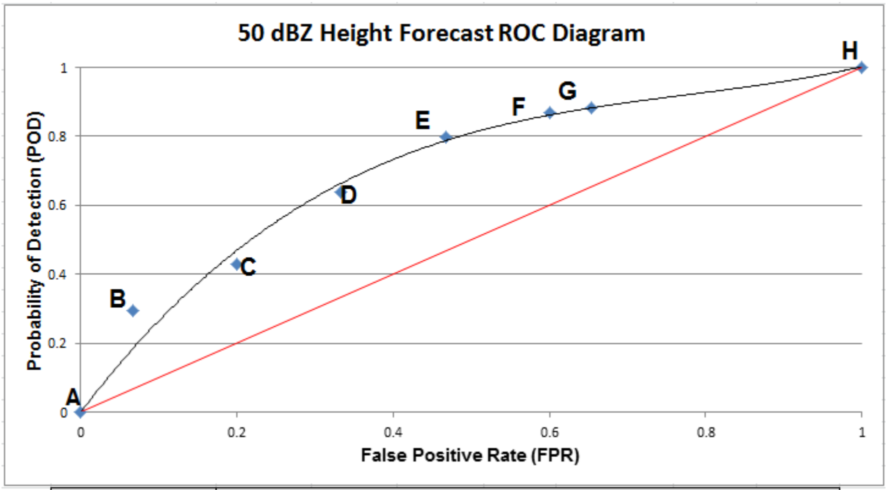
For the ROC curve representing maximum dBZ at -20°C level, the slope falls below 1 in the F-G interval which corresponds with the 10th and 5th percentiles (lettering convention from Fig. 7). This indicates that the 10th percentile should offer the best value to discriminate between severe hail producing storms and non-severe hail producing storms. This threshold had a POD of 0.95 and a FPR of 0.67. In this case the next threshold, the 5th percentile, also has a POD of 0.95 but FPR of 0.73, so there is no doubt that the 10th percentile is a superior threshold. While the 10th percentile POD is very high, it is concerning that the FPR is also high. The ratio between the two is approximately 1.5:1, meaning an unbiased warning system based on this method would be 1.5 times more likely to properly identify a severe storm than to forecast a false alarm. These results seem to indicate that the -20°C method may be better used in conjunction with the 50 dBZ height method as opposed to a stand-alone warning decision method. This is also suggested in general by the ROC curves which identify the 50 dBZ height method as a better diagnostic test than the -20°C method. The recommended warning criteria associated with the 10th percentile is also included in Table 2.

Table 1 – The results of scoring the DJ07 recommended 50 dBZ heights compared to the 50 dBZ heights derived from the 10th percentile in this study.

DJ07	Is Svr	Is not	50 dBZ - 10th	Is Svr	Is not
Forecast Svr	152	57	Forecast Svr	151	45
Forecast Not	22	18	Forecast Not	23	30
POD:	0.874		POD:	0.868	
FPR:	0.76		FPR:	0.6	

DJ07 Comparison

The suggested 50 dBZ levels in the DJ07 study (their Table 1) were also scored on performance in the 2009 evaluation database. The DJ07 recommended warning decision thresholds were compared to this study's 10th percentile threshold since DJ07 used the same percentile to develop their recommended warning criterion (Table 1). While the PODs of the two methods



50 dBZ Height							
Interval	A-B	B-C	C-D	D-E	E-F	F-G	G-H
Slope	4.4	1.03	1.56	1.21	0.52	0.32	0.37
-20 °C dBZ							
Slope	1.64	2.16	1.82	1.21	1.09	0.09	0.17

Fig. 7 – The results of the calculation of slope intervals for each method are listed in the table. Each point on the ROC diagram is labeled A-H beginning at the origin. The calculated slopes represent the slopes of each line segment between successive points.

are nearly identical, the DJ07 study had a significantly higher FPR. This is likely due to the training database in DJ07 being taken from North Dakota and Iowa as opposed to the Southern Plains. These results demonstrate that there is value in developing regional 50 dBZ altitude studies for use in warning operations following the methodology defined in DJ07.

Operational Testing

To collect data from operational forecasters, a questionnaire was developed. Radar operators were asked to record which hail detection method they were using to make a warning decision, and then to write down the event tracking number (ETN) of the severe thunderstorm warning they issued as a result. Radar operators were also asked to write down their subjective thoughts on how their hail detection method of choice performed and how it was used in their warning decision process. Twenty one individual responses were received which accounted for 38 severe thunderstorm warnings issued. The limited number of responses prevents objective comparison of operational testing in 2010 to performance metrics in the 2009 evaluation database used in this study. The subjective responses were useful to gain insight as to how these methods were used in actual warning operations.

Forecasters noted that they used the hail detection method of their choice in support their warning decision process while interrogating thunderstorm structure. If a storm surpassed suggested severe hail criteria forecasters were more apt to issue a severe thunderstorm warning in the presence of other supporting radar data that suggest an organized thunderstorm (i.e. the presence of mid-level rotation, or a weak echo region). Respondents also noted that in high CAPE environments warnings were sometimes issued before the storm exceeded the severe hail criteria in order to improve lead time and avoid missed events due to rapid hail growth in stronger updrafts. The respondents seemed to be most comfortable using the 50 dBZ height method, and most noted that they felt this method performed well and helped in the warning decision process for severe hail.

4. CONCLUSIONS

Four hail detection methods that utilize only base WSR-88D data were applied to a database of Southern Plains thunderstorms and then evaluated to compare their relative effectiveness. The altitude of the 50 dBZ reflectivity echo incorporated with the melting level was found to be the most effective hail detection method tested. This method was found to perform well when tested in several Southern Plains severe thunderstorm events. This hail detection method was compared to the DJ07 recommendations and it was found to have an improved FPR. These results suggest that there may be value in applying the DJ07 methodology to local or regional training databases to improve the detection of severe hail when utilizing base WSR-88D data. This hail detection method is most likely to have a positive impact on warning operations when used in the Southern Plains in conjunction with other radar interrogation techniques. The maximum dBZ at the -20°C hail detection method may be used to increase confidence in making a warning decision when used with the 50 dBZ hail detection method.

For future work, there will likely be added value in incorporating data from additional thunderstorms into the training database. This would be especially beneficial for modeling severe hail producing storms in the cool season and during the late summer months. Once the WSR-88D network in the Southern Plains has been upgraded to collect Dual-Polarization data, some of these moments of data may be useful in discriminating severe hail producing storms and should be incorporated into the database.

5. Acknowledgements

The authors would like to extend thanks to Greg Patrick, Science and Operations Officer at the NWS in FWD for his review of this work and for helpful guidance throughout this study. We thank Dr. Charles Doswell III, Senior Research Scientist at CIMMS for suggesting the use of Signal Detection Theory in this study and for providing guidance on incorporating this theory in a meaningful way. We thank Mike Johnson, now a Senior Forecaster at WFO SJT, for his help in collecting operational data from forecast operations at WFO AMA. We also thank the operational forecasters at WFOs AMA and FWD who took the time to record their thoughts on the hail detection methods during busy warning operations.

References

Brooks, H.E., 2004: TORNADO-WARNING PERFORMANCE IN THE PAST AND FUTURE A Perspective from Signal Detection Theory. *Bulletin of the American Meteorological Society, **85,** 837-843.*

Choi, B.C.K., 1998: Slopes of a Receiver Operating Curve and Likelihood Ratios for a Diagnostic Test. *American Journal of Epidemiology*, **148**, 1127-1132.

desJardins M.L., K.F. Brill, and S. S. Schotz, 1991: GEMPAK 5. Part I-GEMPAK 5 programmer's guide. National Aeronautics and Space Administration, 176 pp. [Available from Scientific and Technical Information Devision, Goddard Space Flight Center, Greenbelt, MD 20771.].

Donavon, R.A. and Jungbluth, K.A., 2007: Evaluation of a Technique for Radar Identification of Large Hail across the Upper Midwest and Central Plains of the United States. *Weather and Forecasting*, **22**, 244-254.

Marzban, C. and Witt, A, 2001: A Bayesian Neural Network for Severe-Hail Size Prediction. *Weather and Forecasting*, **16**, 600-610.

Mason, I., 1982: A model for assessment of weather forecasts. *Australian Meteorological Magazine*, **30**, 291-303.

Tape, T. G., 2001: *Interpreting diagnostic tests*: *The area under a ROC curve.* Retrieved from http://darwin.unmc.edu/dxtests/ROC3.htm

Appendix

Table 2 - Recommendations for warning criterion based on the altitude of the 50 dBZ reflectivity echo relative to the melting level. All values listed are in feet above ground level (AGL). The recommended warning criterion for the maximum dBZ at the -20°C level is also listed, with units of dBZ. The unbiased recommended severe hail warning criterion is the 25th percentile while the 10th percentile may be used if costs associated with missed events are expected to be high.

	50 dBZ height	
Melting Level	25th Percentile	10th Percentile
6500	20500	18200
7000	21700	19500
7500	23000	20700
8000	24200	22000
8500	25400	23200
9000	26700	24400
9500	27900	25700
10000	29200	26900
10500	30400	28200
11000	31700	29400
11500	32900	30700
12000	34100	31900
12500	35400	33100
13000	36600	34400
13500	37900	35600
14000	39100	36900
14500	40400	38100

Max dBZ at -20°C
56