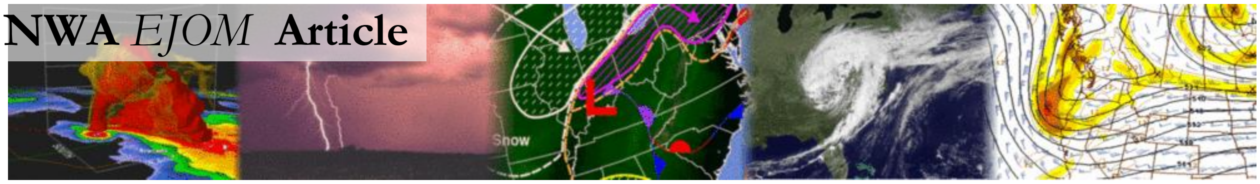
Cavanaugh, D. E., and J. A. Schultz, 2012: WSR-88D signatures associated with one inch hail in the Southern Plains. *Electronic J. Operational Meteor.*, **13** (1), 1–14.



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

DENNIS E. CAVANAUGH AND JESSICA A. SCHULTZ*
National Weather Service, Fort Worth, TX
*Current Affiliation: Radar Operations Center, Norman, OK

(Manuscript received 9 June 2011; in final form 22 February 2012)

ABSTRACT

Four hail detection methods utilizing base data from National Weather Service Weather Surveillance Radar-1988 Doppler are evaluated for use in Southern Plains severe weather operations. Signal Detection Theory is used to evaluate these methods in order to identify the method that best discriminates between severe and non-severe hail producing thunderstorms. The altitude of the 50-dBZ reflectivity echo relative to the melting level is identified as the most effective hail detection method of those methods tested in this study. Signal Detection Theory is then applied to this method to determine which warning decision threshold will maximize Probability of Detection while minimizing false alarms.

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 five 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).

The goal of this study was 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 investigated 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

Corresponding author address: Dennis E. Cavanaugh, 3401 Northern Cross Blvd, Fort Worth, TX

- The maximum dBZ at the -20°C level
- The maximum dBZ at the -30°C level

Severe hail warning criteria were 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 were calculated for each threshold. Signal detection theory was then applied using these scores to compare the effectiveness of the hail detection methods used in this study. Signal detection theory was also used to determine the optimal hail detection threshold which can be used by operational meteorologists to aid in the warning decision process. Results were compared to recommended warning criteria as specified in Donavon and Jungbluth (2007, hereafter DJ07). Lemon (1980) suggested a 50-dBZ reflectivity echo above 8230 m (27000 ft) was an indicator of a severe storm capable of producing large hail and DJ07 introduced the comparison of the 50-dBZ reflectivity echo altitude relative to the melting layer depth 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 were used as a basis for recommending how base data hail detection algorithms can be best applied to severe convective storm warning operations.

2. Data and Methodology

The domain chosen for this research was across the Southern Plains of the United States and included WFOs Fort Worth (FWD), San Angelo (SJT), Midland (MAF), Lubbock (LUB), Amarillo (AMA) and Norman (OUN) (Fig. 1). To establish a training dataset to create warning criterion thresholds, all one inch hail reports across the domain during 2008 were collected 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 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 near storm environment. Any storm associated with a hail report (of any size) was included in the scoring phase of the study. The authors also included additional storms with 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.

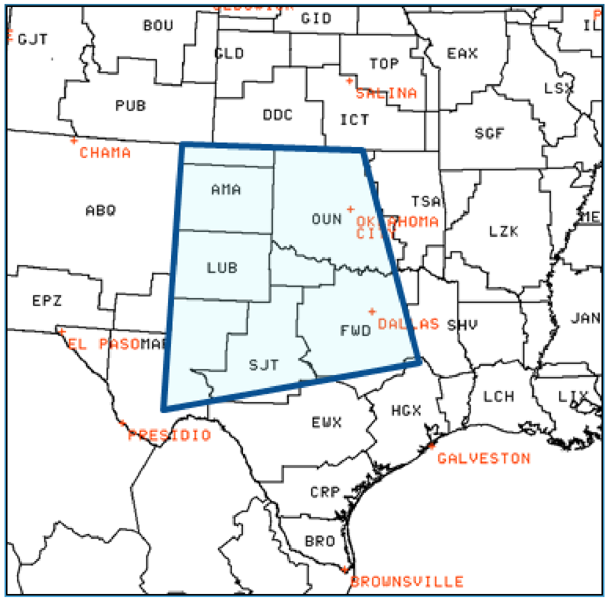


Figure 1. The domain of the study. The blue box represents the area where data for this study were collected. This includes 6 WFOs in the Southern Plains.

Environmental data for the study were collected from the National Centers for Environmental Prediction North American Regional Reanalysis data set (NARR; Mesinger et al. 2006). 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 et al. 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 were 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 due to relatively low values of convectively available potential energy (CAPE) during the cool season, and relatively low environmental shear values combined with high melting level depths over the Southern Plains during the summer months. These factors favor environments where either weaker or shorter lived convection commonly occurs, limiting the opportunity for hail growth in a convective updraft or allowing hydrometeors to melt before reaching the ground.

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. The maximum dBZ at a particular level was calculated by recording the maximum reflectivity bin as long as that bin was less than 5 dBZ stronger than adjacent reflectivity bins. It is worth noting that GR radar software uses the actual radar dish angle to calculate heights AGL. This will cause a difference at times in the calculation of heights AGL on GR software and those displayed on workstations at NWS WFOs. The difference can be as much as 150 m (500 ft) and 300 m (1000 ft) depending on the degree of departure from standard elevation angles.

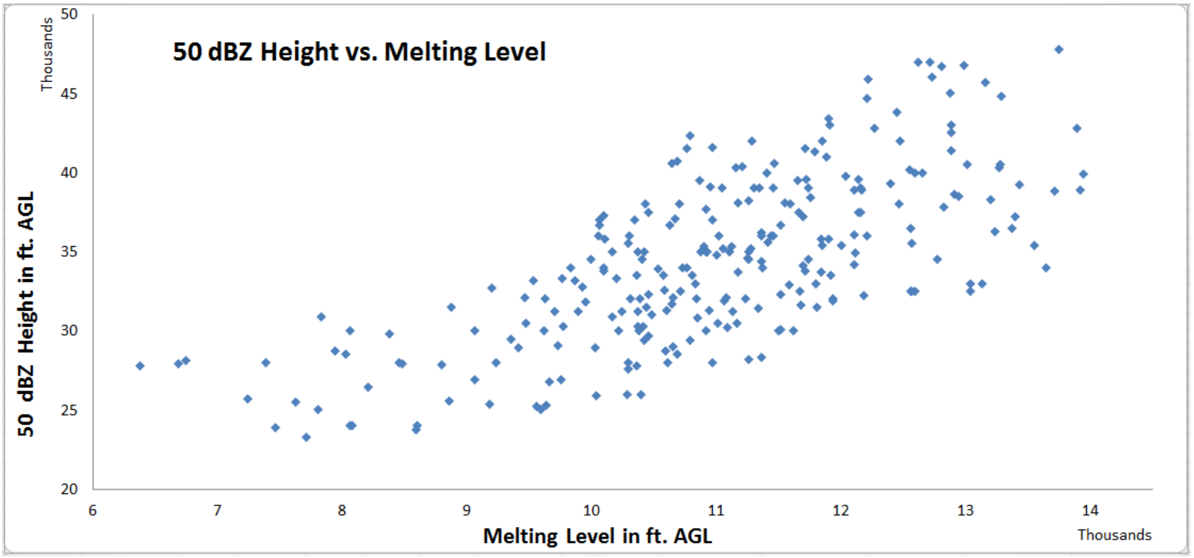


Figure 2. The melting level plotted as a function of the altitude of the 50-dBZ reflectivity echo in feet AGL. The plotted data are the 260 storms that make up the training database from 2008. The majority of the data fall between the 8,000 and 13,500 ft melting levels.

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:

$$Y = 2.4848x + 7092.9 \text{ with a coefficient of determination of } 0.71 \quad (1)$$

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

$$Y = 2.348x + 597.57 \text{ with a coefficient of determination of } 0.65 \quad (2)$$

These regression equations support a moderate to strong linear relationship between the depth of the melting layer (“x “in the equations) and the altitude of the 50- or 60-dBZ reflectivity echo (“y” in the equations). Visual inspection of the data (Fig. 2) indicates that the modeling of these data may be improved by segmenting the data into two bins, with a break point consistent with the 10500 ft. melting level. These data were then analyzed in bins using the methods identified in DJ07. Once the base regression segments were calculated, quantile regression was then applied to the data in order to create equations for the 90th, 75th, 50th, 25th, 10th and 5th percentiles. Quantile regression calculates a regression that places a percentile of the data above or below a given threshold (DJ07). An example of 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.

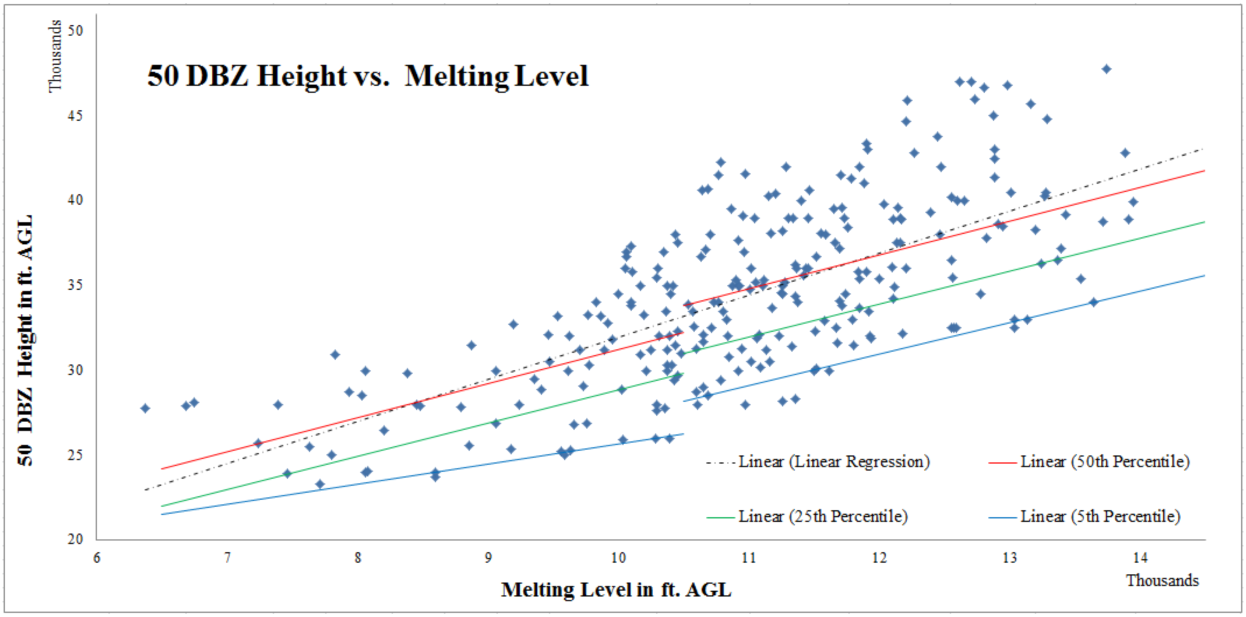


Figure 3. The 50-dBZ height hail detection method with the 50th, 25th, and 5th percentile linear regression lines plotted.

a. 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 is determining how many correct forecasts of non-severe storms there are in any particular severe weather event. Storms were added to the evaluation database in an effort to improve the modeling of this value.

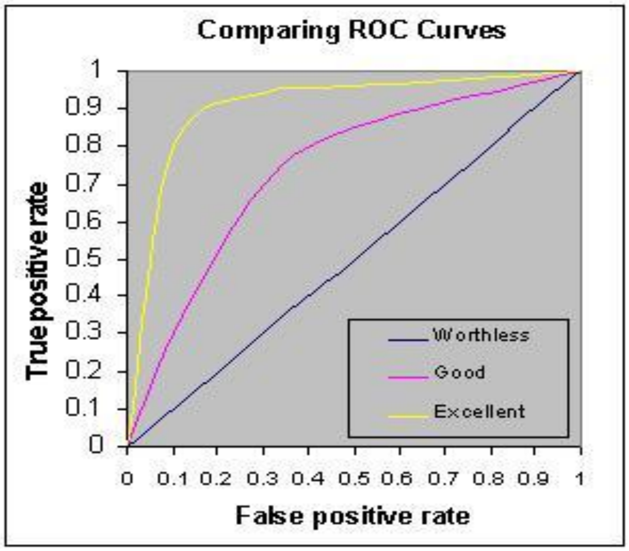
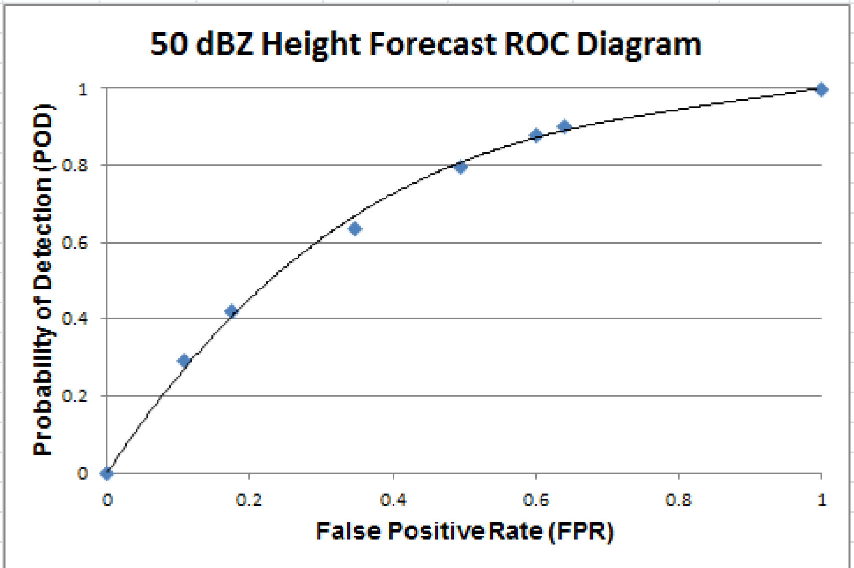


Figure 4. The graph represents what ROC curves would look like for a diagnostic test whose ability to discriminate between 2 classes can be described as "worthless", "good" and "excellent". Note that a diagonal line represents a test with no skill and a curve closer to the upper left hand corner represents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

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 were calculated which represents one point on the ROC diagram. After all thresholds are scored, a ROC curve was 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 to compare the relative effectiveness of each method (Fig. 6). The area under the curve statistic used in this manner is not meant to be an evaluation of any hail detection method in general. To use the area under the curve statistic to evaluate any diagnostic test, it is assumed that the distribution of yes and no events associated with the parameter being tested is Gaussian in nature (Swets 1988). Because thunderstorms are not equally sampled to determine if they contain hail or not, the data collected in this study do not represent a fair Gaussian distribution of 'yes' and 'no' outcomes associated with correlating radar signatures to hail reports. As such, the area under the curve is only being used to compare methods to one another and not to justify any individual method as a useful severe hail diagnostic technique.

Based on the ROC analysis of these methods, the altitude of the 50-dBZ reflectivity echo was the most effective hail detection method tested (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.



90th %	Is Svr	Is not	75th %	Is Svr	Is not	50th %	Is Svr	Is not
Forecast Svr	51	8	Forecast Svr	73	13	Forecast Svr	111	26
Forecast Not	123	67	Forecast Not	101	62	Forecast Not	63	49
POD:	0.293		**POD:**	0.420		**POD:**	0.638	
FPR:	0.107		**FPR:**	0.173		**FPR:**	0.347	

25th %	Is Svr	Is not	10th %	Is Svr	Is not	5th %	Is Svr	Is not
Forecast Svr	139	37	Forecast Svr	153	45	Forecast Svr	157	48
Forecast Not	35	38	Forecast Not	21	30	Forecast Not	17	27
POD:	0.799		**POD:**	0.879		**POD:**	0.902	
FPR:	0.493		**FPR:**	0.600		**FPR:**	0.640	

Figure 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.

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 with the greatest 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 were included in Fig. 7.

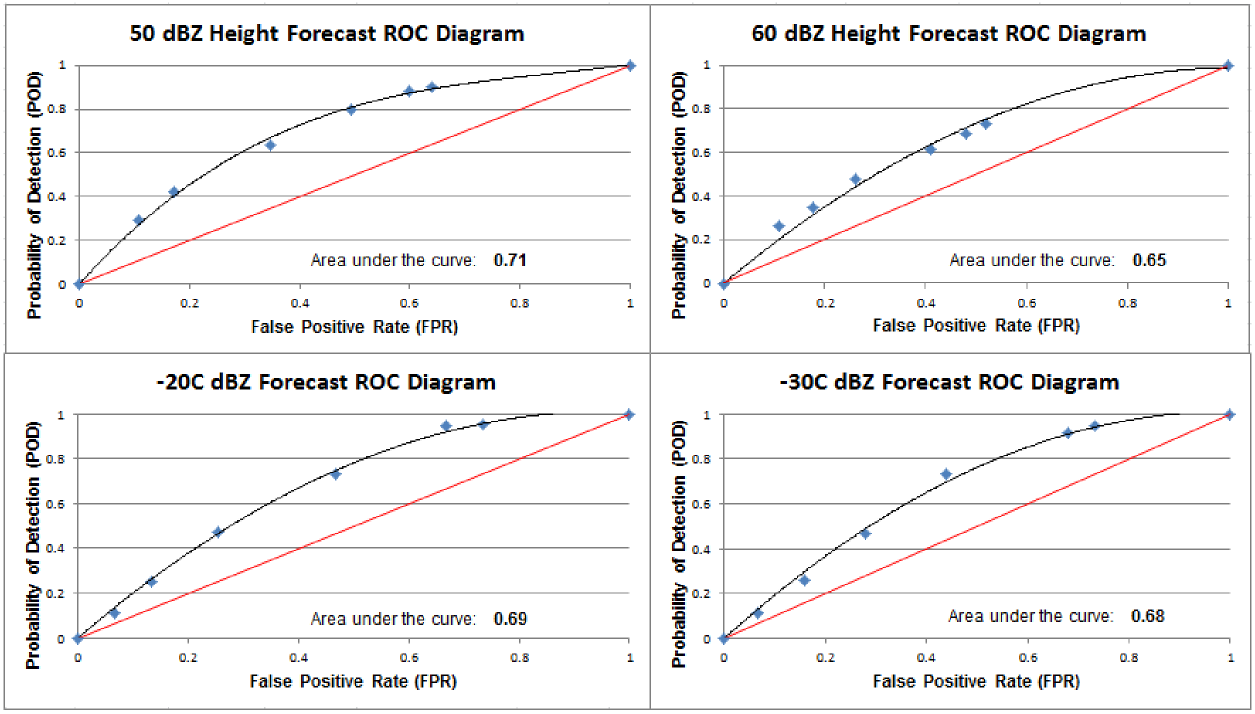
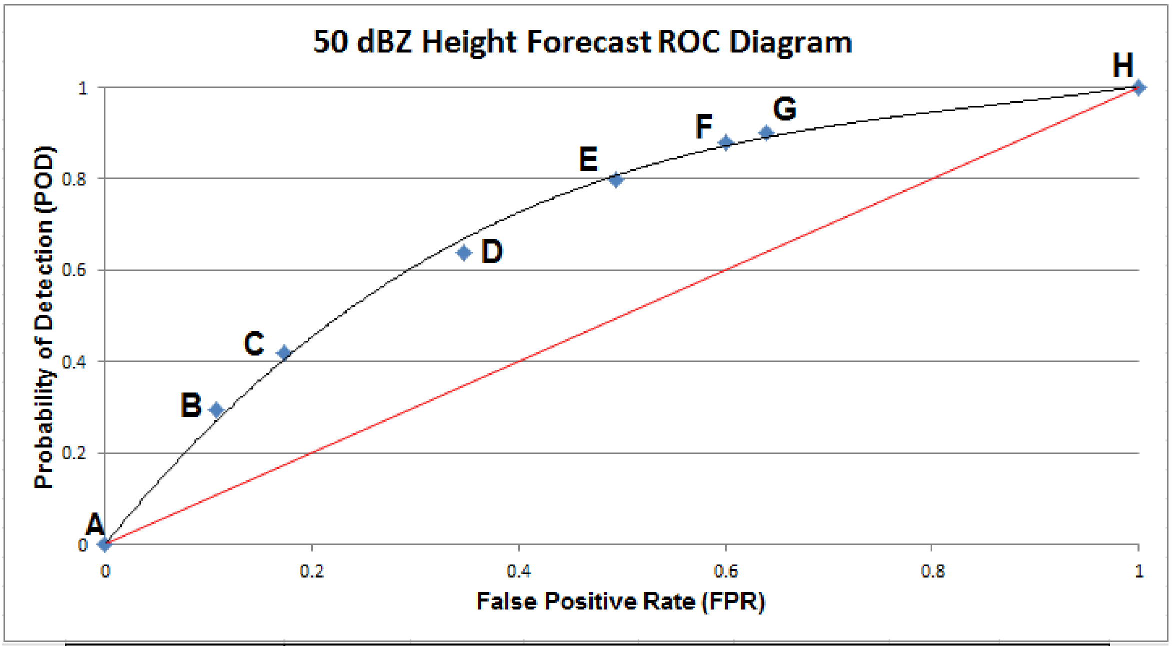


Figure 6. The ROC curves for each hail detection method with the area under the curve displayed in the bottom right corner of each diagram. The red line in each diagram represents the ROC curve corresponding to a hail detection method with zero skill at discriminating severe hail producing storms.

For the 50-dBZ height the slope fell below 1 in the E-F interval which corresponds with the 25th and 10th percentiles, respectively (Fig. 7). This indicated that the 25th percentile should offer the best values for discriminating between severe hail producing storms and non-severe hail producing storms. This threshold had a POD of 0.80 with a FPR of 0.49. In an unbiased warning system using only this test to detect severe hail, these scores indicated that the probability of a warning being issued for a storm containing severe hail was almost twice as high as the probability of a warning being issued when severe hail does not occur. Brooks (2004) points 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 is high, then the 10th percentile may be a better choice as a warning decision threshold. The POD was higher at the 10th percentile with a score of 0.88, but the FPR increased to 0.60 which allowed for a large increase in false alarm probability. The 5th percentile is not recommended as an alternate warning criterion over the 10th percentile because an analysis of the ROC curve for this method implies a nearly 2 to 1 increase in FPR over POD (i.e. a slope of 0.58 in associated with line segment F-G in Fig. 7). A table including both the 25th percentile and the 10th percentile values for the 50-dBZ height method were included as Table 1 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 the 10th percentile values based on a subjective value of risk associated with a potential missed event.



50 dBZ Height							
Interval	A-B	B-C	C-D	D-E	E-F	F-G	G-H
Slope	2.74	1.92	1.25	1.1	0.75	0.58	0.27
-20 °C dBZ							
Slope	1.64	2.16	1.82	1.21	1.09	0.09	0.17

Figure 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.

For the ROC curve representing maximum dBZ at -20°C level, the slope fell below 1 in the F-G interval which corresponded with the 10th and 5th percentiles (lettering convention from Fig. 7). This indicated that the 10th percentile should offer the best value in discriminating between severe hail producing storms and non-severe hail producing storms using this method. 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, which suggests that the 10th percentile is a superior threshold. While the 10th percentile POD was very high, the FPR was also high. The ratio between the two was 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 indicated that the -20°C method would be better used in conjunction with the 50-dBZ height method as opposed to a stand-alone warning decision method. This was also suggested in general by the ROC curves which identified the 50-dBZ height method as a better diagnostic test than the -20°C method. The recommended warning criteria associated with the 10th percentile was also included in Table 1. Results for the 60-dBZ height and the max dBZ at the -30C level methods suggest that these parameters are not as useful detecting severe sized hail. The 60-dBZ height and max dBZ at the -30 C level are included in Table 2 for reference.

When a 60-dBZ echo is detected in a thunderstorm, Table 2 can provide guidance for warning decisions for radar meteorologists.

Table 1. 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 the 10th percentile; however the dataset average and 25th percentile are also listed. The unbiased recommended severe hail warning criterion is the 25th percentile from the 50-dBZ height method, while the 10th percentile may be used if costs associated with missed events are expected to be high.

Melting Level	50 dBZ height 25th Percentile	10th Percentile
6500	22000	21600
7000	23000	22300
7500	24000	23000
8000	24900	23700
8500	25900	24400
9000	26900	25100
9500	27900	25800
10000	28800	26600
10500	29800	27300
11000	31900	29800
11500	32900	30600
12000	33900	31500
12500	34900	32400
13000	35800	33300
13500	36800	34200
14000	37800	35000
14500	38800	35900
Max dBZ at -20°C	**Max dBZ at -20°C**	**Max dBZ at -20°C**
Mean dBZ: 63	25th Percentile: 60	10th Percentile: 56

b. DJ07 comparison

The suggested 50-dBZ levels in the DJ07 study (their Table 1) were scored on performance in the 2009 evaluation database. Donavon's (2010, hereafter D10) updated study for detecting one inch hail in storms was also scored in the 2009 evaluation database. The DJ07 and D10 recommended warning decision thresholds were compared to this study's 10th and 25th percentile thresholds for 50-dBZ altitude, and these results are displayed as Table 3. Both DJ07 and D10 scored with a better POD than the suggested (25th percentile) unbiased decision threshold in this study; however both DJ07 and D10 had a significantly higher FPR. DJ07 and D10 scored most comparably with the 10^{th} percentile threshold in this study, with the 10^{th} percentile scoring a slightly improved (lower) FPR. These comparisons indicate that the 10^{th} percentile threshold can be a useful warning decision threshold if POD is valued higher than FPR (or alternatively if there is great risk to property or public safety associated with a missed event).

Table 2. The results of the 60-dBZ height vs. the melting level and the max dBZ at the -30°C level are displayed below as a reference. These results are not recommended for use as severe hail warning criteria due to the improved performance of other methods evaluated in this study. The 10^{th} percentile results are displayed for the 60-dBZ height method and can also be used if the cost associated with a missed event is thought to be high. The same information displayed for the max dBZ at the -20°C level are displayed for the max dBZ at the -30°C level.

	60 dBZ height	
Melting Level	25th Percentile	10th Percentile
6500	12700	9600
7000	13900	10700
7500	15100	11700
8000	16200	12800
8500	17400	13800
9000	18500	14800
9500	19700	15900
10000	20800	16900
10500	22000	18000
11000	23200	19000
11500	24300	20100
12000	25500	21100
12500	26600	22200
13000	27800	23200
13500	29000	24300
14000	30100	25300
14500	31300	26400
Max dBZ at -30°C	Max dBZ at -30°C	Max dBZ at -30°C
Mean dBZ: 60	25th Percentile: 57	10th Percentile: 53

Table 3. The results of scoring the DJ07 and D10 recommended 50-dBZ heights compared to the 50-dBZ heights derived from the 10th and 25th percentiles in this study.

D10	Is Svr	Is not	50 dBZ - 10th	Is Svr	Is not
Forecast Svr	152	51	Forecast Svr	153	45
Forecast Not	22	24	Forecast Not	21	30
POD:	0.874		POD:	0.879	
FPR:	0.68		FPR:	0.6	
DJ07	Is Svr	Is not	50 dBZ - 25th	Is Svr	Is not
Forecast Svr	152	57	Forecast Svr	139	37
Forecast Not	22	18	Forecast Not	35	38
POD:	0.874		POD:	0.799	
FPR:	0.76		FPR:	0.4933	

c. 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 prevented 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 of 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 were 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 and D10 recommendations and it was found to have an improved FPR. The 25th percentile threshold was found to score the best in an unbiased warning system, while the 10th percentile is recommended if POD is valued more than FPR. These results are 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. If a 60-dBZ echo is detected in a storm, the results provided in Table 2 can be used to raise confidence in a warning decision.

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 study.

Acknowledgements. The authors would like to extend thanks to Greg Patrick, Science and Operations Officer at the NWS in Fort Worth for his review of this work and for helpful guidance throughout this study. We thank Dr. Charles Doswell III, Senior Research Scientist at the Cooperative Institute for Mesoscale Meteorological Studies, 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 Rodney Donavon, meteorologist at WFO Des Moines, IA, for his careful review of this work and for his guidance in improving the methodology and results of this study. We thank Mike Johnson, meteorologist at WFO San Angelo, TX, for his help in collecting operational data from forecast operations while at WFO Amarillo, TX. We also thank the operational forecasters at WFOs Amarillo and Fort Worth 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. *Bull. Amer. Meteor. Soc.*, **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., 2010: Updating the Donavon and Jungbluth 2007 Hail Criteria to One Inch. 14th Annual Northern Plains Convective Conference, Sioux Falls, SD.

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. *Wea. Forecasting,* **22,** 244-254.

Lemon, L. R., 1980: Severe thunderstorm radar identification techniques and warning criteria. NOAA Tech. Memo. NWS NSSFC-3, 60 pp. [Available from NOAA Central Library, 1315 East–West Highway, Silver Spring, MD 20910.]

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

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

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

Swets, J.A., 1988: Measuring the accuracy of diagnostic systems. *Science*, **240,** 1285-1293.

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