3.2 AN EVALUATION OF TORNADO INTENSITY USING VELOCITY AND STRENGTH ATTRIBUTES FROM THE WSR-88D MESOCYCLONE DETECTION ALGORITHM

Darrel M. Kingfield[*,1,2], James G. LaDue[3], and Kiel L. Ortega[1,2]
[1]Cooperative Institute for Mesoscale Meteorological Studies (CIMMS), Norman, OK
[2]NOAA/OAR/National Severe Storms Laboratory (NSSL), Norman, OK
[3]NOAA/NWS/OCWWS/Warning Decision Training Branch (WDTB), Norman, OK

1. INTRODUCTION

National Weather Service (NWS) service assessments following the 2011 April 27 southeast US tornado outbreak, and the 2011 May 22 Joplin tornado found that the population failed to personalize the potential severity of the tornadoes for which they were warned. The lack of specific information contained within the warnings has been cited as a major reason why more action was not taken (NOAA, 2011a, 2011b). As a result, a number of meetings amongst government, the private and the academic sectors have begun to formulate a plan to improve impact-based warning services. In fact, this is the number one goal of the Weather Ready Nation initiative within the NWS (NOAA, 2012). This path implies that the NWS needs to include some information on the severity of a tornado within the content of future tornado warnings. Currently, there is no guidance to assist forecasters in determining tornado intensity.

There are already many methods currently being executed to evaluate a relationship between common radar attributes and tornado intensity. A study by Toth et al. (2011) is comparing maximum tornado intensity to low-level velocity difference (LLDV) values as seen by both the Weather Service Radar 1988-Doppler (WSR-88D) and mobile, near-range radar. LaDue et al. (2012) is investigating how high-resolution damage survey information coincides with WSR-88D vortex parameters on a scan-by-scan basis. Smith et al. (2012) is performing a manual analysis comparing maximum tornado strength against both near storm environmental parameters on a 40km grid and LLDV values calculated using WSR-88D data.

Our motivation was to determine what relationship could be found between tornado intensity and WSR-88D algorithms already in an NWS forecast office. For the extent of this study, we utilized the Mesocyclone Detection Algorithm (MDA; Stumpf et al. 1998) due to the variety of strength and velocity parameters calculated. Previous studies have utilized the MDA in investigating tornado occurrence and tornado-producing mesocyclones (Jones et al. 2004; Trapp et al. 2005). But the operational MDA has undergone several improvements since these studies, including the integration of super-resolution data. Super-resolution data (Torres and Curtis, 2007) at the split-cut elevations reduces the effective beamwidth to 1.02° from the legacy 1.38°, allowing for vortex signatures to be resolved at longer ranges from the radar (Brown et al. 2002; Brown et al. 2005). In this paper, we investigate how the MDA output correlates to maximum intensity found in a tornado path.

* *Corresponding Author Address:* Darrel M. Kingfield, NOAA/OAR/NSSL, 120 David L. Boren Blvd., Norman, OK, 73072; e-mail: darrel.kingfield@noaa.gov

2. DATA & METHODS

2.1 Acquisition & Playback

The initial tornado track dataset was composed of all tornado events logged into Storm Data from 01 January 2009 to 31 December 2011. In order for the tornado event to be retained for radar processing, it had to be surveyed by an NWS employee or reference a damage survey in the event narrative using the Enhanced Fujita Scale (EF-Scale; WSEC 2006).

For each retained event, the recorded start and end latitude/longitude coordinates were extracted. All Next-Generation Radar (NEXRAD) sites within 300km of either coordinate pair had their Level-II data retrieved from the National Climatic Data Center (NCDC). The retrieval times were +/- 30 minutes from the event times on record.

In order to generate the WSR-88D MDA products, these data were played back at real-time speed using the NWS Operational Build 12.2 Radar Product Generator (RPG) software. In total, 19,755 hours of data were played through this system, resulting in 327,676 MDA detections for 3923 tornado events.

2.2 Quality Control

In order to classify the detections against the tornado events, tornado tracks had to be drawn. With Storm Data only recording the start/end points on record, a line was drawn connecting these points to form each tornado track. In addition, a minimum time threshold had to be defined to give the radar adequate time to sample the storm during the tornado. The authors proposed a minimum of three volume scans, one before the recorded event time, at least one scan during the event, and one after the recorded event time. We did this in an attempt to

offset temporal deviations between what occurred in reality versus what was recorded in Storm Data (Witt et al. 1998; Stumpf et al. 1998 ; Trapp et al. 2005). From the original 3923 events, 2121 events were on the ground less than four minutes, not meeting our three scan minimum, and were removed from the dataset.

A 5-km search radius was defined as the threshold of whether a MDA detection would be paired with the event. This radius was chosen to minimize the number of nearby non-tornadic mesocyclones being counted in the dataset. To estimate the location of the tornado in the event window, intermediate points along the tornado track were derived by dividing the track length by the event duration. The result is an estimated position at each minute the tornado was reported to be on the ground. These points served as the basis for a floating search window to classify the MDA detections with only one window active at a time. In order for a detection to be retained, it had to fall within this 5-km spatial window and have a detection time with the 1-min temporal window (Figure 1).

With each Storm Data record, there is only one documented EF-rating for the entire path. The absence of high-resolution ground truth in a large majority of these events makes it difficult to assess multiple strengths. As a result, the authors assume the strength on record is the highest rating confidently observed during a survey. To ensure the greatest chance of matching a detection to the strongest segment of the tornado, only the peak values of the various MDA parameters were extracted. Since we are only looking at the peak values, we added another rule that there had to be at least two detections from a radar in order for that information to be retained for analysis. Otherwise, a single detection would by default contain all of the maximum values. Of the remaining 1802 events, 452 events had zero or a single MDA detection associated with it and were subsequently removed from the dataset.

While an appropriately sized buffer/time window will minimize contamination from non-tornadic sources, it does not alleviate the issue of a single detection being classified within two (or more) tornado tracks at the same time. To counter this, the EF-ratings for each tornado track are compared. If one track was rated stronger, the detection will be placed with this track. If both tracks were of equal strength, the detection was paired with the closest track.

After the initial quality control, we were left with 7,297 MDA detections from 0-300km from the radar associated with 962 tornado events. Transforming the detections over a single radar space (Figure 2) showed a sharp drop off in the number of detections beyond 200 km, with no detections between 220-250km and beyond 290km. As a result, the authors decided to focus our analysis space to detections between 0-200km from the radar. By removing the observations beyond 200km, our analysis totals were lowered to 905 tornado events and 5839 MDA detections (Figure 3). These data were separated into 20km bins to investigate the influence of radar range on data quality.

3. ANALYSIS & OBSERVATIONS

3.1 MDA Parameter Breakdown

There are five velocity and strength variables currently produced by the operational MDA, defined in Stumpf et al. 1998: Low-Level Rotational Velocity (LLVR), Low-Level Delta Velocity (LLDV), Maximum Rotational Velocity (MXRV), Mesocyclone Strength Rank (SR), and Mesocyclone Strength Index (MSI). One of the realms to explore was if there was a single parameter or combination of multiple parameters that best discriminates between different classes of tornado intensity. To do this, we looked at the classification accuracy between 2 groups, weak (EF0-1) and strong (EF2+), through both a Linear Discriminant Analysis (LDA) and a Quadratic Discriminant Analysis (QDA) generated model (Wilks, 2006).

The dataset was split up in a training set (70% of the data), validation set (15%), and testing set (15%) for all 20km range combinations (e.g. 20-60km, 40-140km, etc.). For each range group, a LDA and QDA model was built using the training set and its predictive hit rate was extracted when the trained model was applied to the validation dataset. This cross-validation process was bootstrapped through 2000 iterations with the resulting median retained for comparison. This was first executed for each of the five individual MDA parameters with the best performing parameter retained. We then extended the model, adding more MDA parameters to determine if a combination of parameters would perform better than the top performing single parameter model.

There was no significant improvement in discriminating between weak versus strong tornadoes when using a Mutli-MDA parameter model vs. a single parameter model. From the single parameter models, we found that the low-level velocity parameters were able to discriminate the best between the two groups only slightly better than the maximum-velocity and strength parameters when combining the results from all range permutations (Figure 4). As expected, when looking at different range groups, the low-level velocity models perform worse at long ranges from the radar (>120km) and close to the radar (<40km). At these ranges, the max SR model provided the best classification rate (Figure 5).

SR requires that the base of the first 2D circulation feature be below 5km above radar level (ARL) with a minimum vertical depth of 3km in half-beamwidth depth (Stumpf et al. 1998). At long distances from the radar, the lowest angle may be

around 4km ARL. With a tornado on the ground, relying on information at 4km and above may not provide enough confidence to discriminate a tornado intensity, especially if we start looking at a finer classification system (e.g. +/- 1 EF-rating). With max LLDV providing the best overall model fit in terms of highest median and smallest confidence interval, we will investigate this parameter further in the next sections.

3.2 Separation of Parameter Means by Intensity

We focus on LLDV to further examine its ability to discriminate between individual and groups of EF-scale ratings. Continuing to look at all range combinations, we evaluated this using several sets of comparison groups:

1) EF0-1 vs. EF2-5
2) EF0-1 vs. EF2-3, EF2-3 vs. EF4-5
3) EF0 vs. EF1, EF1 vs. EF2, EF2 vs. EF3, EF3 vs. EF4-5

The first set is a simple comparison of weak (EF0-1) vs. strong (EF2+) tornadoes. The second set breaks up the "strong" classification into significant (EF2-3) and violent (EF4-5). The third set compares EF ratings to their next strongest EF-rating. Due to the small sample size of EF5 events, these values were grouped with the EF4 class.

Maximum LLDV values for each EF-rating were extracted from the radar range being examined and placed into their comparison group defined above. For example, if we were evaluating the first comparison set, all peak LLDV values corresponding to an EF0 or EF1 tornado were placed in one group while all other peak LLDV values matched to an EF2 or stronger tornado were placed in the opposing group. Each group distribution had their bootstrapped means calculated through 5000 iterations and their 95% confidence intervals extracted. The separation between these confidence intervals was calculated and this value was placed in each 20km range bin. For example, if we were looking at 80-120km and the separation between two groups (e.g. EF2 vs. EF3) was 8 ms^{-1}, the 80km and 100km bins would have 8 added to their respective totals. Figure 6 shows the output for each of the three comparison tests.

Similar to what was seen in the model output, detections at mid-ranges from the radar (~40-140km) showed the most separation between the comparison groups. As we break down the number of EF-ratings in each comparison group, the separation between the means diminishes. In the second comparison test, there was more separation when comparing EF0-1 vs. EF2-3 (Figure 6b) than when comparing EF2-3 vs. EF4-5 (Figure 6c). Looking at individual EF-rating comparisons (Figure 6d-g), there is negative separation when comparing EF0 vs. EF1 and the EF3 vs. EF4-5. Looking at the EF1 vs. EF2 and EF2 vs. EF3 groups, we did see some positive separation between the confidence intervals and, like the previous comparison sets, the peak division was found at mid-ranges from the radar.

3.3 Distribution Comparisons

While the differences in the confidence intervals of central tendency may look promising, the distributions of LLDV vs. EF-scale ratings show significant overlap (Figure 7). The maximum LLDV range for EF0 tornadoes is completely contained within the EF1 distribution. The same is also seen with EF4-5 detections within the EF3 distribution.

Even the most simple comparison of weak (EF0-1) vs. strong (EF2+) tornadoes, there is substantial overall in the middle 50% of the peak LLDV distribution (Figure 8). From the composition of these ranges, a peak LLDV of 40 ms^{-1} falls within the 62nd percentile for EF0-1 and the 30th percentile for EF2 and stronger events. With this amount of overlap, an extremely large or small value is necessary to make a confident estimate on the strength of the tornado. While these figures only show the breakdown from 40-140km from the radar, the same trends are seen at all other range combinations.

3.4 Skill Scores

We wanted to assess the quality of a tornado strength prediction using peak LLDV from the MDA. We accomplished this through the utilization of a two-by-two contingency matrix (Wilks 2006). The values of these matrix variables are calculated at pre-defined LLDV thresholds. In this study, we computed the matrix at peak values of LLDV from 10-75 ms^{-1} with the step interval of 0.5 ms^{-1} and the forecast was whether a strong tornado (EF2+) occurred. Table 1 shows an overview of this process:

		OBSERVED	
		YES	NO
FORECAST	YES	EF2+ LLDV ≥ T Hit (X)	EF0-1 LLDV ≥ T False Alarm (Z)
	NO	EF2+ LLDV < T Miss (Y)	EF0-1 LLDV < T Correct Null (W)

Table 1: The two-by-two contingency table for forecasting a strong tornado with a variable threshold (T).

For example, let's define our LLDV threshold to be 20 ms^{-1}. All EF2+ events that had a LLDV greater than or equal to this threshold would count as a hit (X). All EF0-1 events with a maximum LLDV greater than or equal to this threshold would count as a false alarm (Z). All EF2+ events less than this threshold would be classified as a miss (Y). All EF0-1 events with a

maximum LLDV less than this threshold would be classified as a correct null (W). Utilizing this matrix, these four parameters were resampled via bootstrapping through 2500 iterations with the Heidke Skill Score (HSS) calculated each time. The HSS was the preferred technique as it utilizes all information in the contingency matrix (Wilks 2006).

$$HSS = \frac{2\,(XW - YZ)}{(X+Y)(Y+W) + (X+Z)(Z+W)}$$

Evaluating the bootstrapped HSS from 40-140km from the radar (Figure 9), we see overall positive skill with a peak median HSS of 0.34 around 40 ms^{-1}. We wanted to see how this location and peak HSS compared to other peak HSS values at all other radar range combinations (Figure 10). In many cases, especially at mid-ranges from the radar (e.g. 60-120km and as seen in Figure 9), the peak skill varies by a few hundredths or even thousandths of a decimal place around 35-50 ms^{-1}. This trend is much less prominent when solely looking at longer ranges from the radar (>140km). In these cases, peak HSS coincides with a larger LLDV with a more well-defined peak. At ranges closer to the radar (<60km), we saw higher HSS peaks at lower LLDV values, but the size of the bootstrapped confidence intervals at these ranges indicates poor sampling could be a contributing factor.

4. DISCUSSION

Under current conditions, relating vortex attributes from a single WSR-88D is not sufficient at estimating tornado intensity. Confident discrimination between even two groups of tornado ratings (EF0-1 vs. EF2-5) is difficult to do with this dataset. Both the LDA and QDA techniques had at least a 30% misclassification rate when evaluating single or multi-parameter models at various radar ranges. There are several contributing factors to why this is occurring.

Surveying with the EF-scale, there is uncertainty in degrees of damage versus wind speed, lack of damage indicators in rural areas (e.g. vegetation), and the quality of surveys being employed (e.g. Doswell and Burgess 1988; Doswell et al. 2009; Edwards et al. 2010). While tornadoes are more high-profile events than other severe reports (wind, hail, etc.), there is a possibility of inaccuracy when recording event time and location to Storm Data (Witt et al., 1998; Trapp et al., 2006). While the RPG does the best possible job of dealiasing the velocity data, it is unknown how much of an impact these failures had on the quality of our dataset. Even with super-resolution data at lower levels, beam offset, registration, and aspect ratio can all negatively impact the quality of the data (Wood and Brown 1997).

5. CONCLUSIONS

A three year climatology of all surveyed tornado events were compared to peak values from the WSR-88D MDA. Of the original 3923 events and 327,676 MDA detections, 905 tornado events and 5839 detections remained after quality control processing. First- and second-order models were used to determine which combination of MDA parameters could discriminate the best between a 2 classes of EF ratings, weak vs. strong. The single parameter models performed equal to or better than the multi-parameter models with the low-level velocity parameters slightly edging out the peak velocity and strength parameters. Looking at MDA-derived LLDV, while there are some significant departures when comparing the bootstrapped means between some classes (e.g. EF2 vs. EF3, EF0-1 vs. EF2+), there was substantial overlap when comparing all distributions. Calculating the HSS when discriminating between weak vs. strong events, we found values between 0.30 – 0.40 depending on range from the radar. Evaluating the HSS values at mid-ranges from the radar where we would expect well-sampled LLDV values, the peak value is not well-defined. This makes it difficult to provide a single value that serves as a threshold between classifying weak vs. strong tornado.

5.1 Future Work

The process of generating tornado strength guidance using current WSR-88D algorithms may not be a mature technique now, but there are several avenues to consider for a better solution. The two major error sources that need to be mitigated are quality damage surveys and velocity data.

Addressing the damage survey issue, we could look at population density and only evaluate tornadoes that occurred in areas with a high urban footprint. Optimally, it would be better for there to be more high-quality damage surveys. When evaluating high-resolution damage surveys, LaDue et al. (2012) found a promising association when looking at both tornado strength and width but was hindered by the high number of aliasing failures in the WSR-88D dataset.

Addressing the dealiasing failures, utilizing a more optimal dealiasing technique such as the Linear Least Squares Derivative (LLSD) (Smith et al. 2004) for a single radar and/or merging surrounding radars to improve sampling is another avenue to explore (Lakshmanan et al. 2006) in detail. An exploratory investigation of tornado intensity through case studies with these two techniques was done by Labriola et al. (2013). Their results were also inconclusive, unless width of the shear was included.

Acknowledgements We thank Bob Lee, Rich Murnan, Jessica Schultz, and Dave Zittel from the Radar Operations Center Applications Branch, and Kimberly Elmore from NSSL for their contributions to this research. This work was supported under the Cooperative Agreement NA17RJ1227 of the NOAA-University of Oklahoma/Cooperative Institute for Mesoscale Meteorological Studies. The views expressed in this paper are those of the authors and do not necessarily represent those of the NOAA, NSSL, or CIMMS.

REFERENCES

Brown, R. A., V. T. Wood, and D. Sirmans, 2002: Improved Tornado Detection using Simulated and Actual WSR-88D Data with Enhanced Resolution. *J. Atmos. Oceanic Technol.*, **19**, 1759-1771.

_____, B. A. Flickinger, E. Forren, D. M. Schultz, D. Sirmans, P. L. Spencer, V. T. Wood, and C. L. Ziegler, 2005: Improved Detection of Severe Storms Using Experimental Fine-Resolution WSR-88D Measurements. *Wea. Forecasting*, **20**, 3–14.

Doswell, C. A. III, and D. W. Burgess, 1988: On Some Issues of United States Tornado Climatology. *Mon. Wea. Rev.*, **116**, 495-501.

_____, Brooks, H.E., Dotzek, N., 2009. On the implementation of the Enhanced Fujita scale in the USA. Atmos. Res. 93, 564–574.

Edwards, R., J. G. LaDue, J. T. Ferree, K. A. Scharfenberg, C. Maier, and W. L. Coulbourne, 2010: The Enhanced Fujita Scale: Past, Present and Future. Preprints, 25th Conf. of Severe Local Storms, Denver, CO. Amer. Meteor. Soc., 4A.1

Jones, T. A., K. M. McGrath, and J. T. Snow, 2004: Association between NSSL Mesocyclone Detection Algorithm-Detected Vortices and Tornadoes. *Wea. Forecasting*, **19**, 872–890.

Labriola J., K. L. Ortega, D. M. Kingfield, and M. L. Miller. 2013: Investigating the Relationship of Multi-Radar Multi-Sensor Parameters to Tornado Intensity, 12th Annual Student Conference, Austin, TX, Amer. Meteor. Soc., S118.

LaDue, J. G., K. Ortega, B. Smith, G. Stumpf, and D. M. Kingfield, 2012: A Comparison of High Resolution Tornado Surveys to Doppler Radar Observed Vortex Parameters: 2011-2012 Case Studies, Preprints, 26th Conf. on Severe Local Storms, Nashville, TN, Amer. Meteor. Soc., 6.3.

Lakshmanan, V., T. Smith, K. Hondl, G. J. Stumpf, A. Witt, 2006: A Real-Time, Three-Dimensional, Rapidly Updating, Heterogeneous Radar Merger Technique for Reflectivity, Velocity, and Derived Products. *Wea. Forecasting*. **21**, 802-823.

National Oceanic and Atmospheric Administration (NOAA), 2011a: NWS Service Assessment: The historic tornadoes of April 2011. 76 pp.

National Oceanic and Atmospheric Administration (NOAA), 2011b: May 22, 2011 Joplin Tornado (Regional Service Assessment). 40 pp.

National Oceanic and Atmospheric Administration (NOAA), 2012: The NWS Weather Ready Nation Road Map, 69 pp.

Smith, B.T., R. L. Thompson, H. E. Brooke, A. R. Dean, and K. L. Elmore. 2012: Diagnosis of Conditional Maximum Tornado Damage Probabilities, Preprints, 26th Conf. on Severe Local Storms, Nashville, TN, Amer. Meteor. Soc.,P2.20.

Smith, T. M., and K. L. Elmore, 2004: The Use of Radial Velocity Derivatives to Diagnose Rotation and Divergence. Preprints. 11th Conf. on Aviation, Range, and Aerospace, Hyannis, MA. Amer. Meteor. Soc., P5.6

Stumpf, G. J., A. Witt, E. D. Mitchell, P. L. Spencer, J. T. Johnson, M. D. Eilts, K. W. Thomas, and D. W. Burgess, 1998: The National Severe Storms Laboratory Mesocyclone Detection Algorithm for the WSR-88D. *Wea. Forecasting*, **13**, 304–326.

Torres, S. M., and C. D. Curtis, 2007: Initial implementation of super-resolution data on the NEXRAD network, Preprints. 23rd Conf. on Information Processing Systems, San Antonio, TX, Amer. Meteor. Soc., 5B.10.

Toth, M., R. J. Trapp, K. A. Kosiba, and J. Wurman: 2012: Improving Tornado Intensity Estimates Using Operational Doppler Weather Radar. 26th Conf. on Severe Local Storms, Nashville, TN, Amer. Meteor. Soc.,3.3.

Trapp, R. J., G. J. Stumpf, and K.L. Manross, 2005: A Reassessment of the Percentage of Tornadic Mesocyclones. *Wea. Forecasting,* 20, 680–687.

_____, D. M. Wheatley, N. T. Atkins, R. W. Przybylinksi, and R. Wolf, 2006: Buyer Beware: Some Words of Caution on the Use of Severe Wind Reports

in Postevent Assessment and Research. *Wea. Forecasting*, **21**, 408-415.

Warning Decision Training Branch, 2010: RPG Build 12.0 Training Guide. National Weather Service, 20pp.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 627 pp.

Witt, A., M.D. Eilts, G.J. Stumpf, E.D. Mitchell, J.T. Johnson, and K.W. Thomas, 1998: Evaluating the performance of WSR-88D Severe Storm Detection Algorithms. *Wea. Forecasting,* **13**, 513

Wind Science and Engineering Center, 2006: A Recommendation for an Enhanced Fujita Scale (EF-scale). Wind Science and Engineering Center, Texas Tech University, Lubbock, Texas, 95 pp.

Wood, V. T., and R. A. Brown, 1997: Effects of Radar Sampling on Single-Doppler Velocity Signatures of Mesocyclones and Tornadoes, *Wea. Forecasting,* **12**, 928-938.

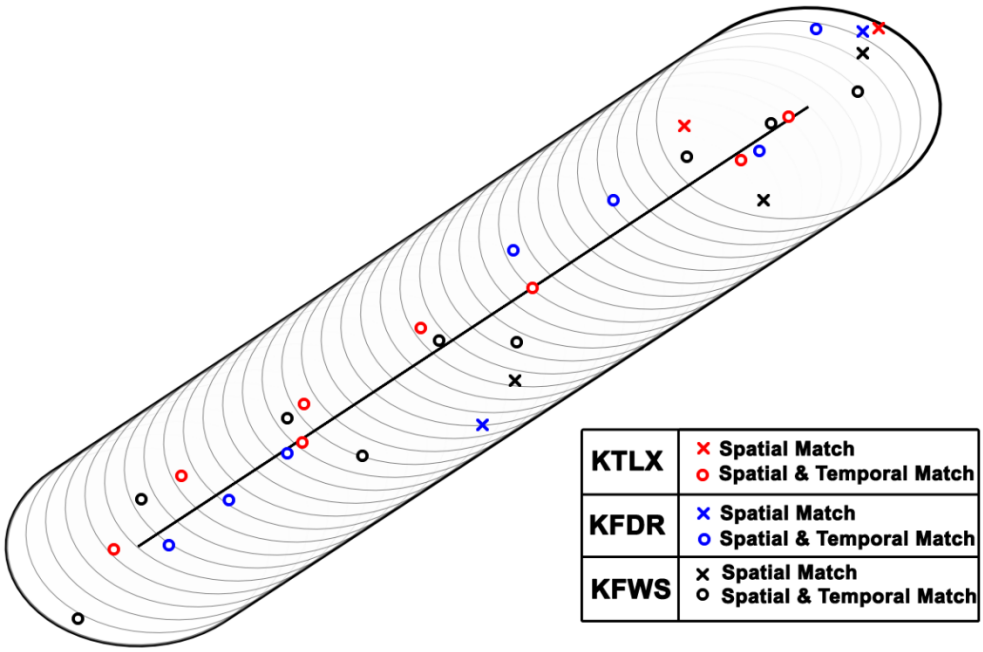

Figure 1: Spatiotemporal pattern matching for the 11 February 2009 Lone Grove, OK EF4 tornado. The tornado was on the ground for 34 minutes so there were 34 5km temporal domains utilized for matching. There were 3 radars scanning within 200km of the tornado track. Detections that fell within the 5km circle (spatial match) but did not occur inside the temporal window are marked with an 'X' and were removed from the dataset. Detections matching in both space and time are marked with an 'O'.

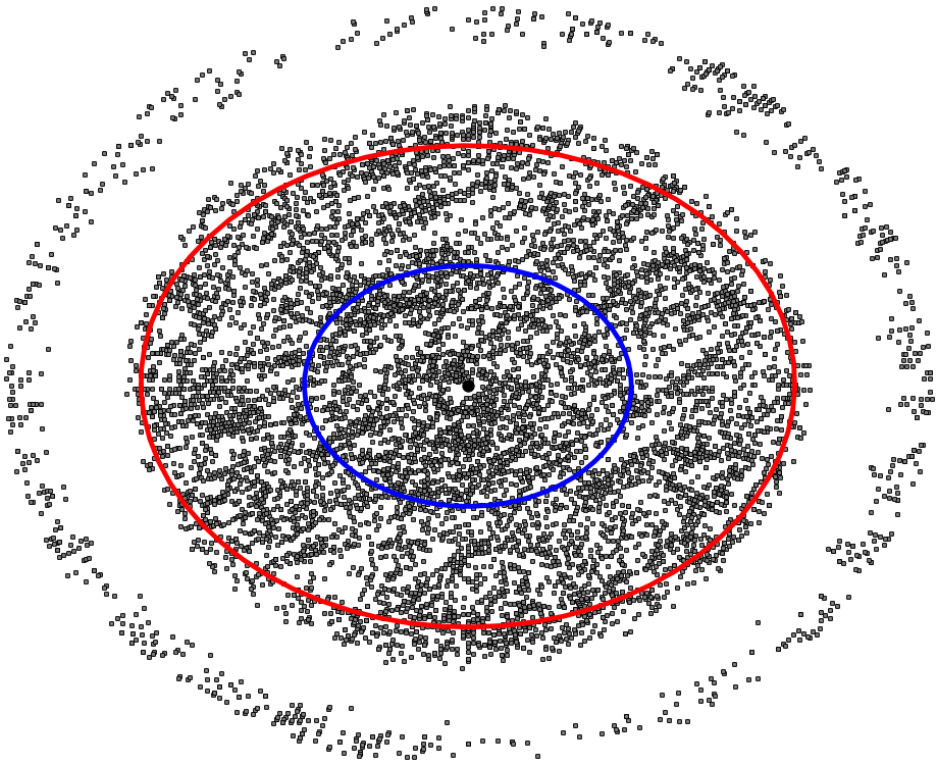

Figure 2: Range/Azimuth values for all matching MDA detections transformed over a single radar source point. The blue ring indicates a range of 100km from the radar. The red ring indicates a range of 200km from the radar.

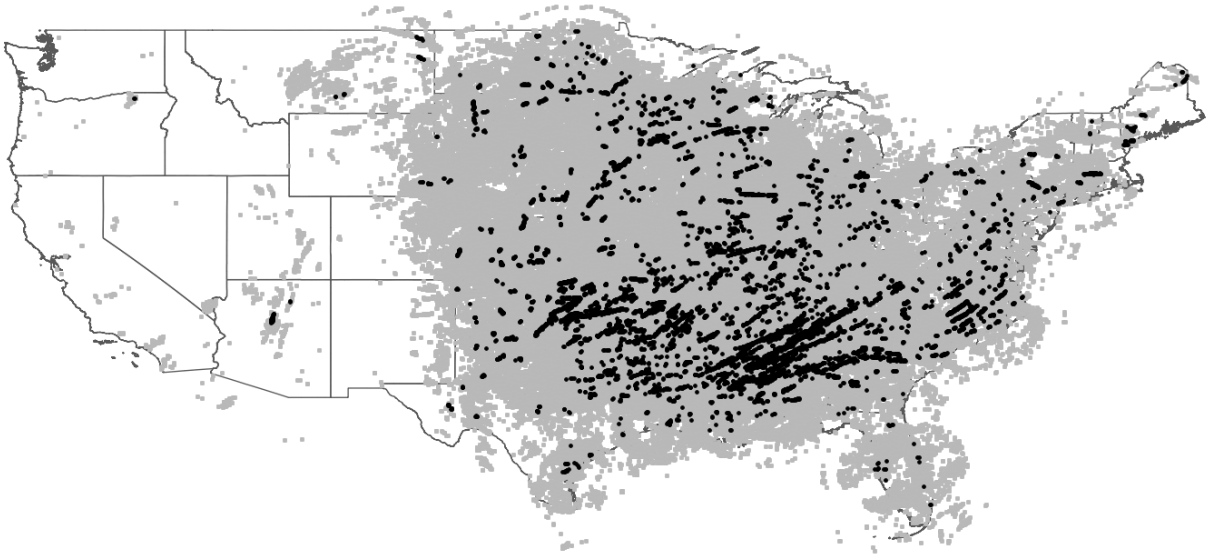

Figure 3: Geographical distribution of all MDA detections utilized in this study. The gray squares are detections that failed the quality control process. The black circles are detections that passed the quality control checks and were retained for analysis.

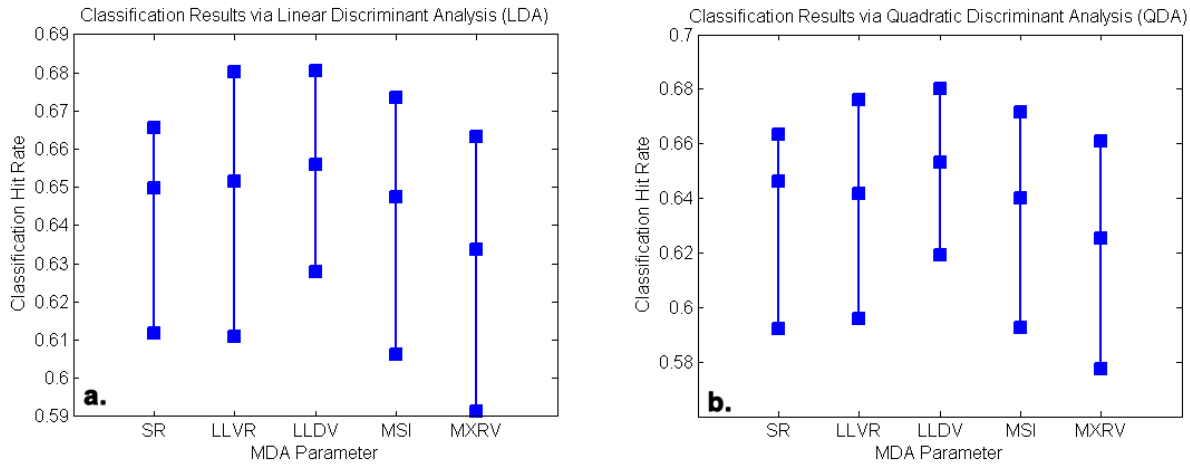

Figure 4: 95% confidence interval of bootstrapped hit rates from the LDA (a) and the QDA (b) summarized at all radar range combinations from 0-200km. The lowest box is the 2.5th percentile, the middle box is the 50th percentile, and the top box is the 97.5th percentile.

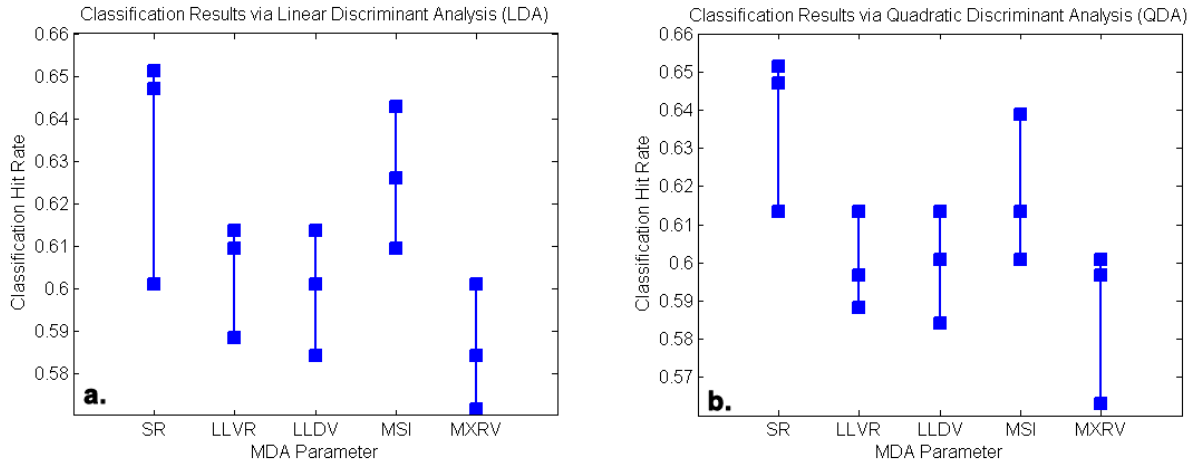

Figure 5: 95% confidence interval of bootstrapped hit rates from the LDA (a) and the QDA (b) for all tornadoes/detections at 120-200km from the radar. The lowest box is the 2.5th percentile, the middle box is the 50th percentile, and the top box is the 97.5th percentile.

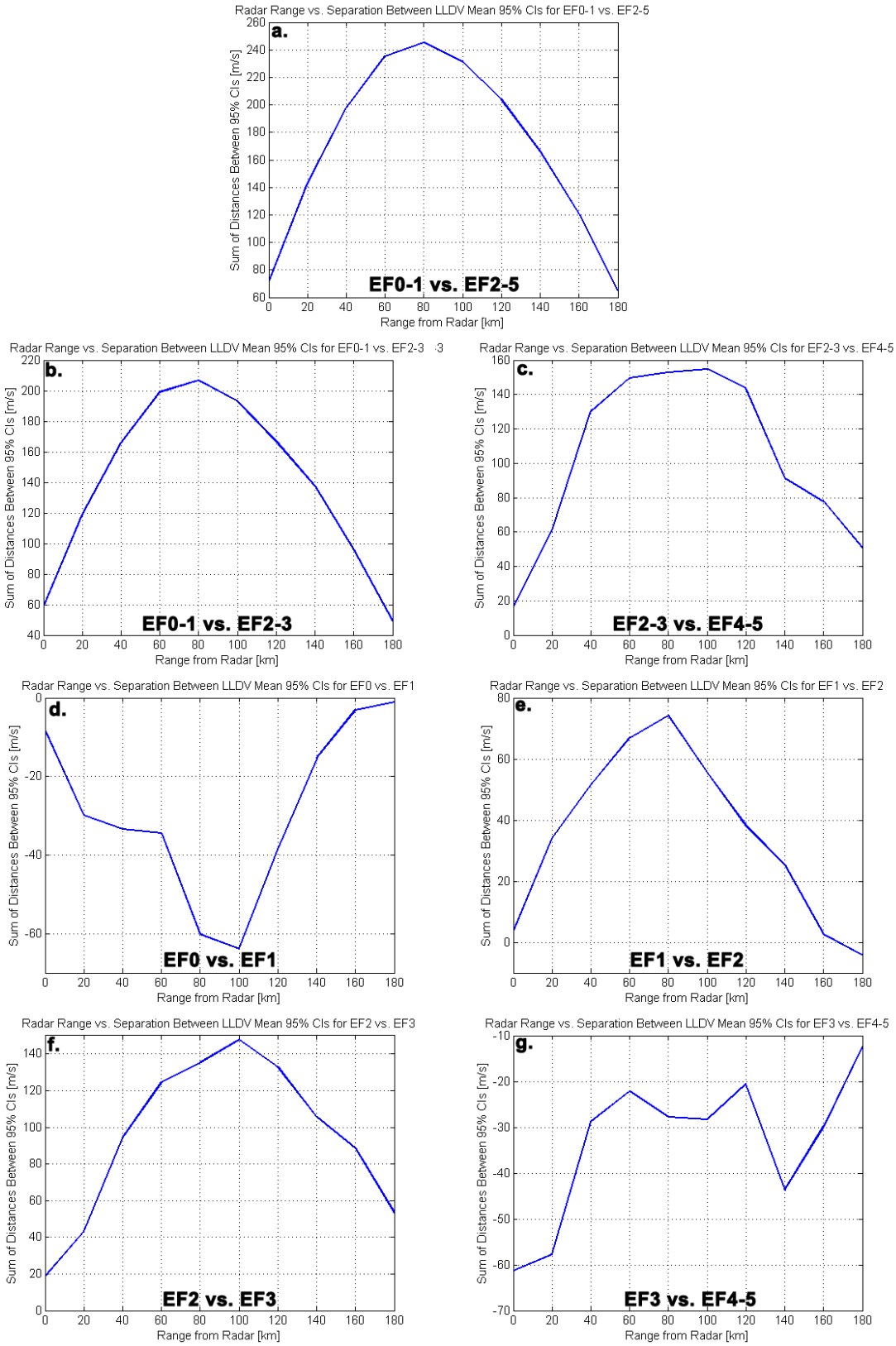

Figure 6: Summary of distances separating the 95% confidence intervals between two EF-rating groups at various ranges from the radar.

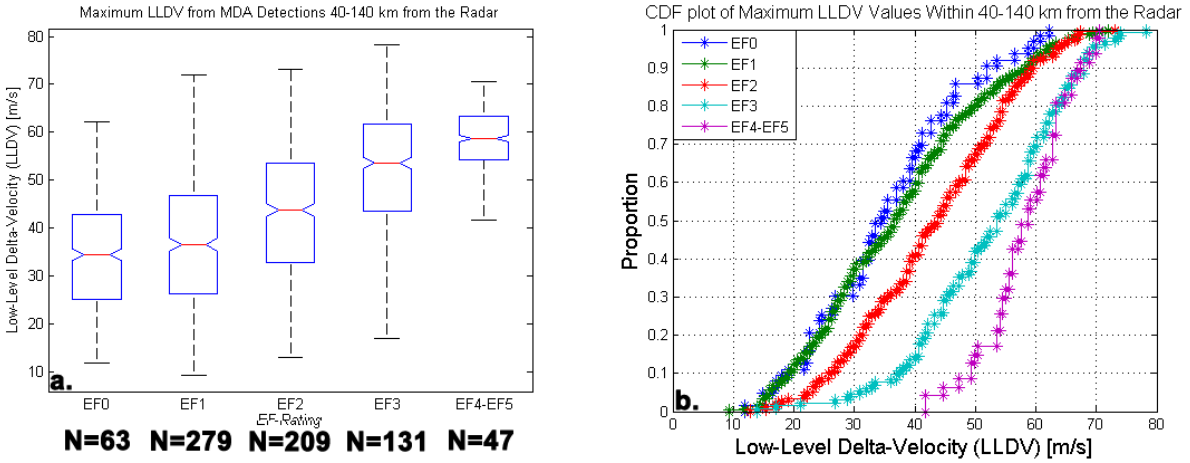

Figure 7: Boxplot (a) and cumulative distribution plot (b) of maximum MDA-derived LLDV for each EF-rating (EF4 and EF5 combined) at 40-140km from the radar. The whiskers extend +/- 2.7σ from the distribution median. The notches in each box indicate the boundaries for whether or not two medians are significantly different at the 5% significance level. Non-overlapping boxes signify the medians are statistically significant.

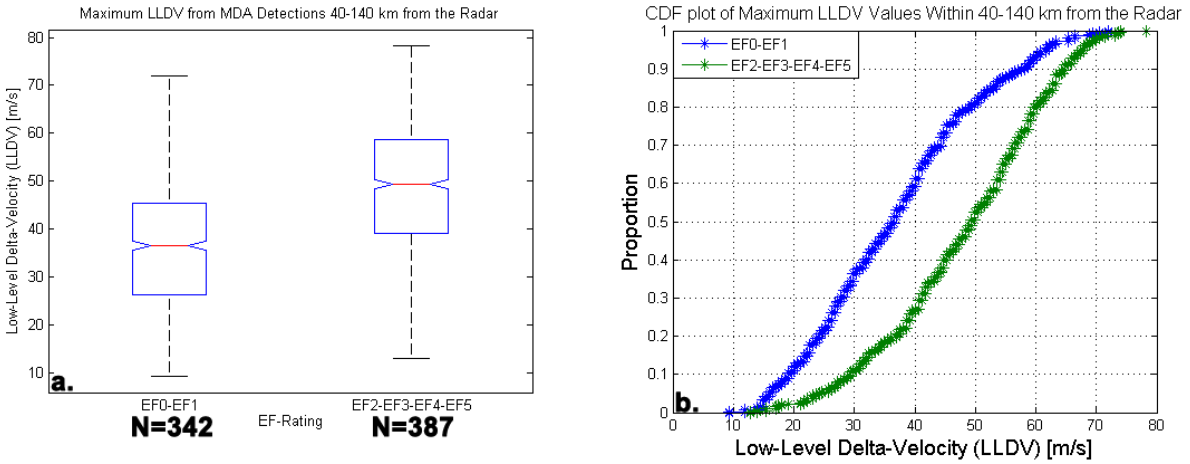

Figure 8: Boxplot (a) and cumulative distribution plot (b) of maximum MDA-derived LLDV for weak (EF0-1) and strong (EF2+) tornadoes at 40-140km from the radar. The whiskers extend +/- 2.7σ from the distribution median. The notches in each box indicate the boundaries for whether or not two medians are significantly different at the 5% significance level. Non-overlapping boxes signify the medians are statistically significant.

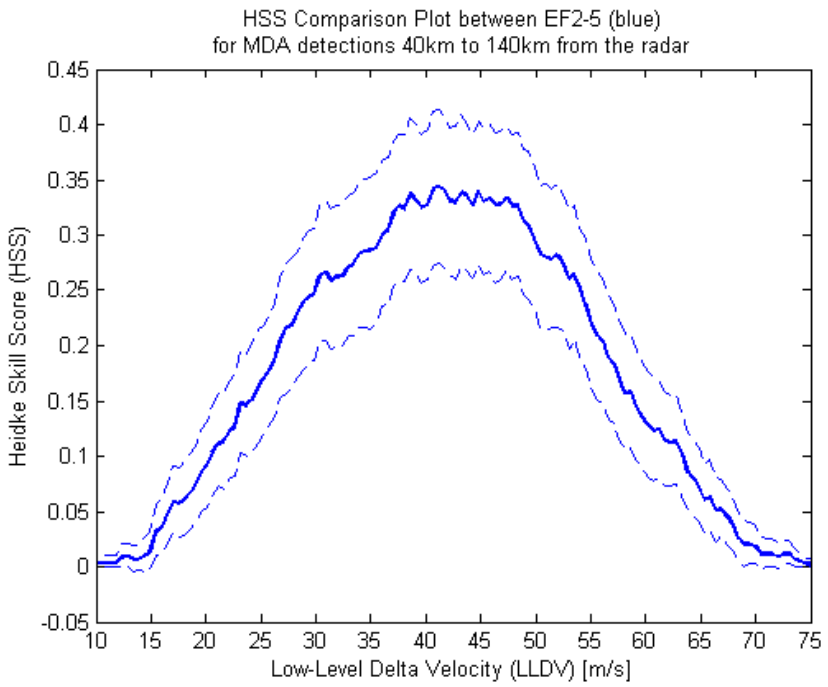

Figure 9: Bootstrapped HSS values of LLDV for all detections 40-140km from the radar.

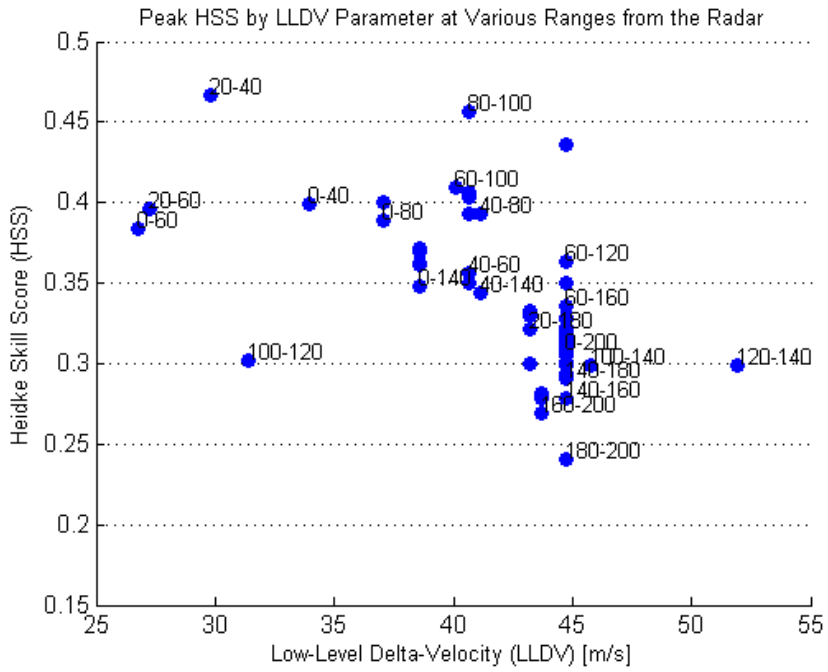

Figure 10: Peak HSS values and the LLDV threshold where this occurred at various ranges from the radar.