P6.5

EVALUATING HAIL DIAGNOSIS TECHNIQUES USING HIGH RESOLUTION VERIFICATION

Kiel L. Ortega[1,2], Angelyn G. Kolodziej[1,2], Joe Young[1,2], Chris J. Wilson[3], Arthur Witt[2] and Travis M. Smith[1,2]
[1]University of Oklahoma/CIMMS [2]NOAA/NSSL [3]NWC REU/Valparaiso University

1. Introduction

Verification of hail diagnosis techniques is extremely difficult given the current resolution and reliability of *Storm Data* (Witt et. al 1998b; Trapp et. al 2006). The difficulty is increased for techniques which are rapidly updating (~1 min) and at high resolution (i.e., a 1 km x 1 km grid) since the resolution of *Storm Data* varies greatly and is of much lower resolution (~1 hour and ~1000 km^2). The Severe Hazards Analysis and Verification Experiment (SHAVE; Smith et al. 2006) has been conducted at the National Severe Storms Laboratory every summer since 2006 with the goal of collecting hail reports at these higher temporal and spatial resolutions (Fig. 1). The dataset is further enhanced by the inclusion of non-severe (hail stone diameter < 19 mm) and “no hail” reports in the vicinity of storm cells.

This manuscript represents a first effort in evaluating the usefulness of high resolution verification data for the evaluation of hail diagnosis techniques.

2. Data

Five storms sampled by SHAVE were analyzed for this study. The storms were selected because the SHAVE data for the storm seems to represent the surface hail fall accurately, the NWS verification was a fair representation for the event (i.e., there was usually more than just one hail report from *Storm Data* and the reported diameters were similar to reports collected by SHAVE), and the storms present challenges to current diagnosis techniques through storm type and/or sampling issues that are a function of the distance of the storm from radar.

Radar data from nearby (within 250 km of the storm) was collected. Several algorithms were run on these data, including:

- an enhanced Hail Detection Algorithm (HDA; Witt et. al 1998a; Marzban and Witt 2001)
- the Storm Cell Identification and Tracking (SCIT; Johnson et al. 1998)
- multiple-radar, grid-based HDA (Stumpf et al. 2004)

The enhanced HDA provides probabilities for 3 hail size classes (coin, golf ball and baseball) while also still providing the original HDA output of the maximum expected size of hail (MESH) and a probability of severe hail (POSH). The enhanced HDA utilizes more environmental data, storm top divergence and storm mid-altitude rotation to produce the probabilities. The SCIT algorithm was used to identify the cells, run the HDA on the cells and track other storm attributes such as storm top divergence, mid-altitude rotation and vertically integrated liquid (VIL). A thorough subjective analysis of the radar data was conducted to identify the timing of features such as weak echo regions (WER), bounded weak echo regions

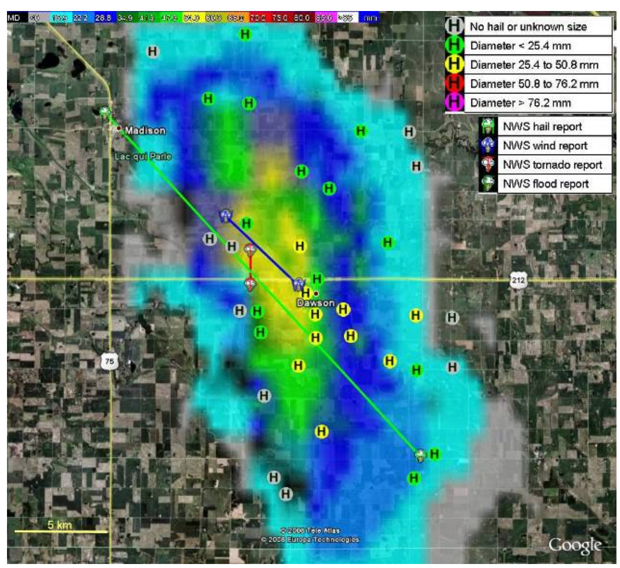

Figure 1: NWS and SHAVE verification for tornadic supercell in Lac qui Parle County, MN on 27 July 2006. Background coloring is multi-radar MESH swath for the event. NWS reports connected by lines are 1 report, with start and end positions.

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

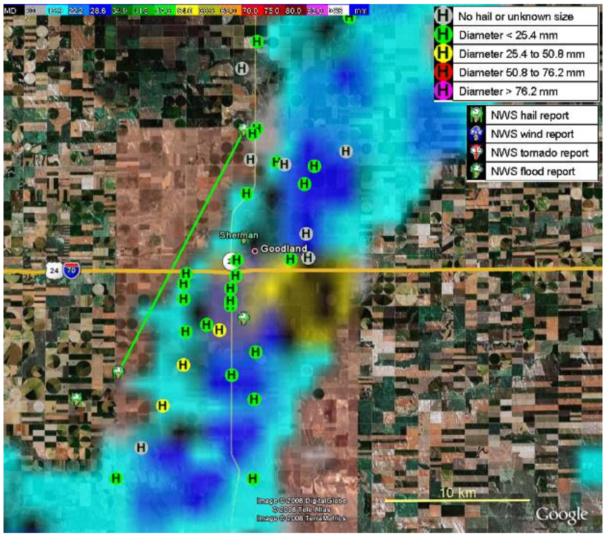

Figure 2: NWS and SHAVE verification for a left moving supercell in Sherman County, KS on 31 May 2006. Background coloring is multi-radar MESH swath

(BWER) and three body scatter spikes (TBSS; Lemon 1998) which are recognized as indicators of a severe storm (i.e., severe-sized hail—diameter greater than or equal to 19 mm).

The 5 cases (Figures 1-5) yielded 230 cell detections, 315 SHAVE reports and 59 NWS reports (of which at least 3 are SHAVE reports which were relayed to the forecast office). A short summary of the cases:

- 31 May 2006: a left moving supercell which entered into the nearest radar's cone of silence with maximum hail stone diameters of 44 mm,
- 27 July 2006: a tornadic, right moving supercell at far distance (130+ km) from the nearest radar with maximum hail stone diameters of 44 mm,
- 13 August 2007: a tornadic, right moving supercell starting at far distance (~180 km) from the nearest radar and then entering the radar's cone of silence with a maximum hail stone diameter of 64 mm,
- 2 June 2008: a low precipitation supercell with maximum hail stone diameter of 100 mm, and
- 10 June 2008: a low precipitation supercell with widespread maximum hail stone diameters exceeding 64 mm.

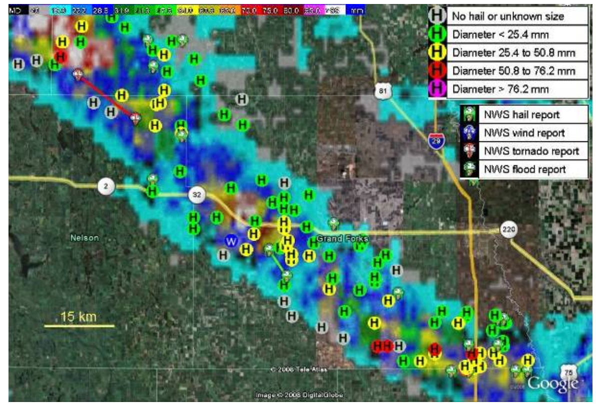

Figure 3: NWS and SHAVE verification for a long-lived, tornadic supercell near Grand Forks, ND on 13 August 2007. Background coloring is multi-radar MESH swath.

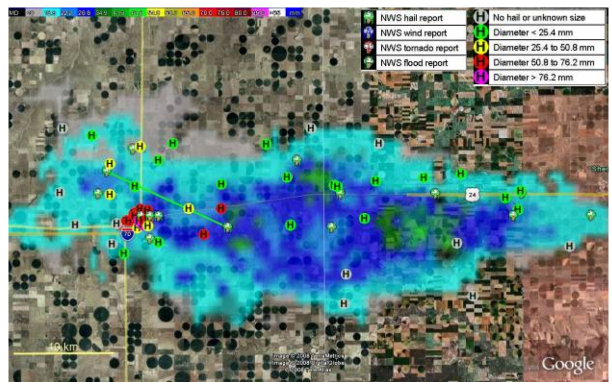

Figure 4: NWS and SHAVE verification for a LP supercell near Burlington, CO on 2 June 2008. Background coloring is multi-radar MESH swath.

3. Methods

The primary goal of the evaluation is to show differences between verification datasets of varying resolution. The secondary goal will be to evaluate whether the different methods yield different hail diagnoses or if a similar conclusion is reached. The techniques used in the evaluation are:

- Storm feature analysis: using single radar data from multiple tilts data and vertical cross-sections to identify features such as WER, BWER and TBSS, and reflectivity heights.
- MESH: from the SCIT HDA and multiple radar, grid-based HDA
- Probabilities: of severe hail (SCIT and grid-based) and hail size classes (SCIT only)

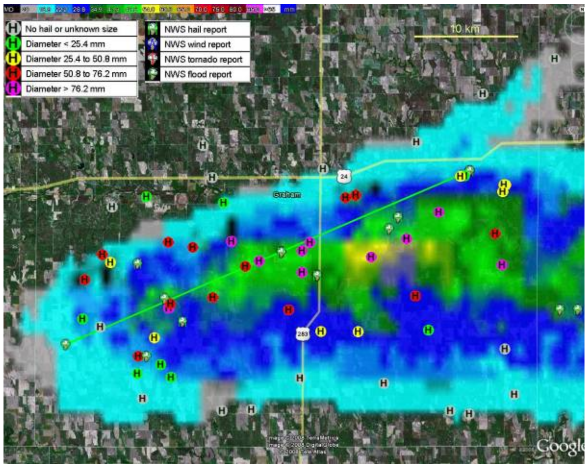

Figure 5: NWS and SHAVE verification from a LP supercell in Graham County, KS on 10 June 2008.

- Reflectivity heights: the height of the top of the elevated 50 dBZ core above the height of 0°C
- Storm attributes derived from SCIT: VIL, storm top divergence and mid-altitude rotation

For SCIT features, a 7.5 km search radius was used to find hail reports near the cell. A large search radius was used because for the larger storms the SCIT-identified storm centroid might be far from where the largest hail was falling at the surface.

The reports were carefully screened so that only reports from the storm of interest were kept. This was done because the report's time was ignored during the evaluation. Time was not taken into consideration because many SHAVE reports are marked as questionable time due to the survey calling 30+ min. after storm passage—so most people could only give coarse estimates of the time hail fall was occurring. Further, most reports from *Storm Data* only have a single time available. For gridded algorithm output, a merged reflectivity composite was used to track the storm and a 5 km search radius from a subjectively determined storm centroid was used to find the maximum of the algorithm output.

MESH	Verification	Bias	RMSE
Cell	SHAVE	-12.5	28.8
Grid	SHAVE	-22.5	40.4
Cell	*Storm Data*	-18.7	39.2
Grid	*Storm Data*	-29.9	46.5

Table 1: Root mean square error (RMSE) and bias (in mm) scores for different hail size diagnosis techniques and verification databases.

4. Discussion

As previously discussed the storms analyzed were storm types which typically present difficulties for hail diagnosis techniques. Table 1 shows the poor performance of a few of the techniques in diagnosing hail size. Given the small sample size of analyzed storms no definite conclusions can be made about the methods; however there may be some significance in the difference between the two verification datasets. Figures 1-5 illustrate a problem that plagues *Storm Data*: inaccuracies in hail size and location. Many of the report swaths available from *Storm Data* do not accurately reflect the surface hail fall depicted by the SHAVE reports. Additionally for the reports depicted by a line between two end points, only 1 hail size is associated with those lines (see Figure 6). These inaccuracies may be leading to the poorer performance of both algorithms when using *Storm Data* as verification as shown in table 1. Figure 6 also shows the increased detail (in this case, higher temporal resolution of the change in maximum hail size) of the surface hail fall when using SHAVE data. Figure 7 shows some further inaccuracies associated with *Storm Data* reports. In this case *Storm Data* reports are consistently smaller than SHAVE reports. This problem may exist if many reports are coming from mobile sources. For instance, a storm chaser might report hail of golf ball size and then leave the area to avoid vehicle damage; meanwhile, baseball-sized hail falls shortly afterwards. SHAVE data avoids this pitfall as SHAVE calls fixed locations, so the maximum diameter for a particular location is known.

Other fields that were investigated (probabilities, VIL and storm velocity characteristics) were found to be extremely noisy or followed similar patterns as is illustrated by MESH in figures 6 and 7. However, the subjective radar analysis was found to be incredibly useful. Features such as BWER or TBSS were present in all cases (as all cases were supercells), however, these features usually preceded any of the other technique's suggestion of severe hail. This was especially true for the LP cases (20080602 and 20080610) where reflectivities were generally low, yet could still be used to identify a BWER (Figure 8).

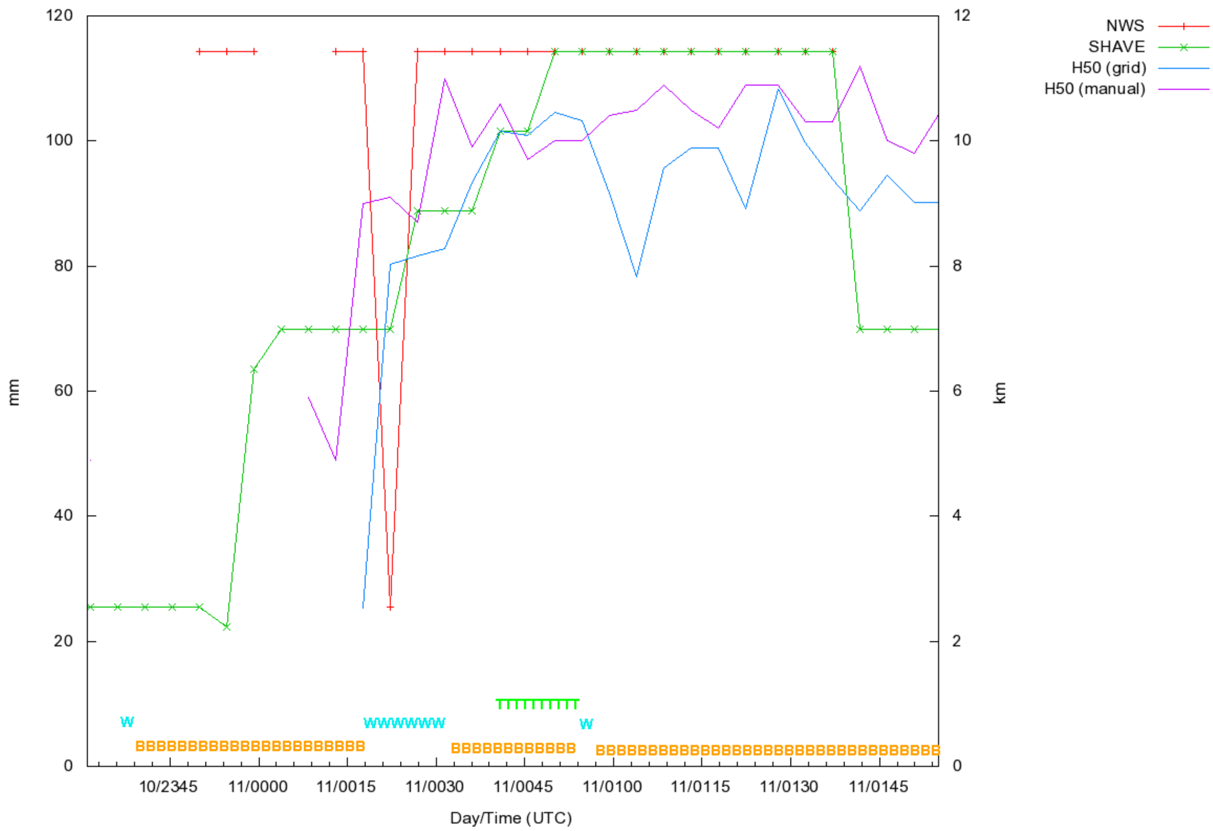

Figure 6: Time series of NWS and SHAVE reports, and interrogation-based and grid-based height of 50 dBZ above 0°C from the storm on 10 June 2008. In this case, SHAVE depicted the increase of hail size to the maximum hail size much better than the reports available from *Storm Data*. 50 dBZ echoes did not start to appear above the freezing level until nearly the same time baseball-sized hail began to fall at the surface. The T's mark where a TBSS was present, the W's a WER and B's a BWER.

5. Future Work

This study limited itself to simple circular searches and using the maximum hail size. However, the SHAVE database lends itself to different manipulation. One potential evaluation method could use storm motion to create sectors in front of storm identifications (Figure 9). This type of searching would: 1) better associate hail sizes to storm attributes, and 2) make for easier lead time calculations. To get past limiting the hail reports only to the maximum, the entire spectrum of reports within a search sector could be used. Can a signature be definitely associated with severe hail if it is associated with only a few severe hail reports? Using the distribution of reported sizes and the high resolution nature of SHAVE reports could help answer that question.

The end goal for this study is to develop a statistically significant evaluation of hail diagnosis techniques. This evaluation would focus on the significance of storm attributes, at varying lead times, with respect to severe hail at the surface. Figure 10 is an illustration of this goal. Figure 10 shows that at 10 minutes lead time, when using multi-radar, grid-based MESH, a 21 mm threshold (maximum Hedike Skill Score) should be used when determining if a storm will produce severe hail. Figure 10 was created using 9 storms yielding 563 (325 severe) SHAVE reports. SHAVE has yielded over 15,000 reports on over 100 storms of varying types and in varying environments. Thus the database should yield a statistically significant evaluation of hail diagnosis techniques.

6. Summary

This manuscript represents a first attempt of using high resolution verification to evaluate and verify hail diagnosis techniques. The cases presented illustrate differences in descriptions of a storm's

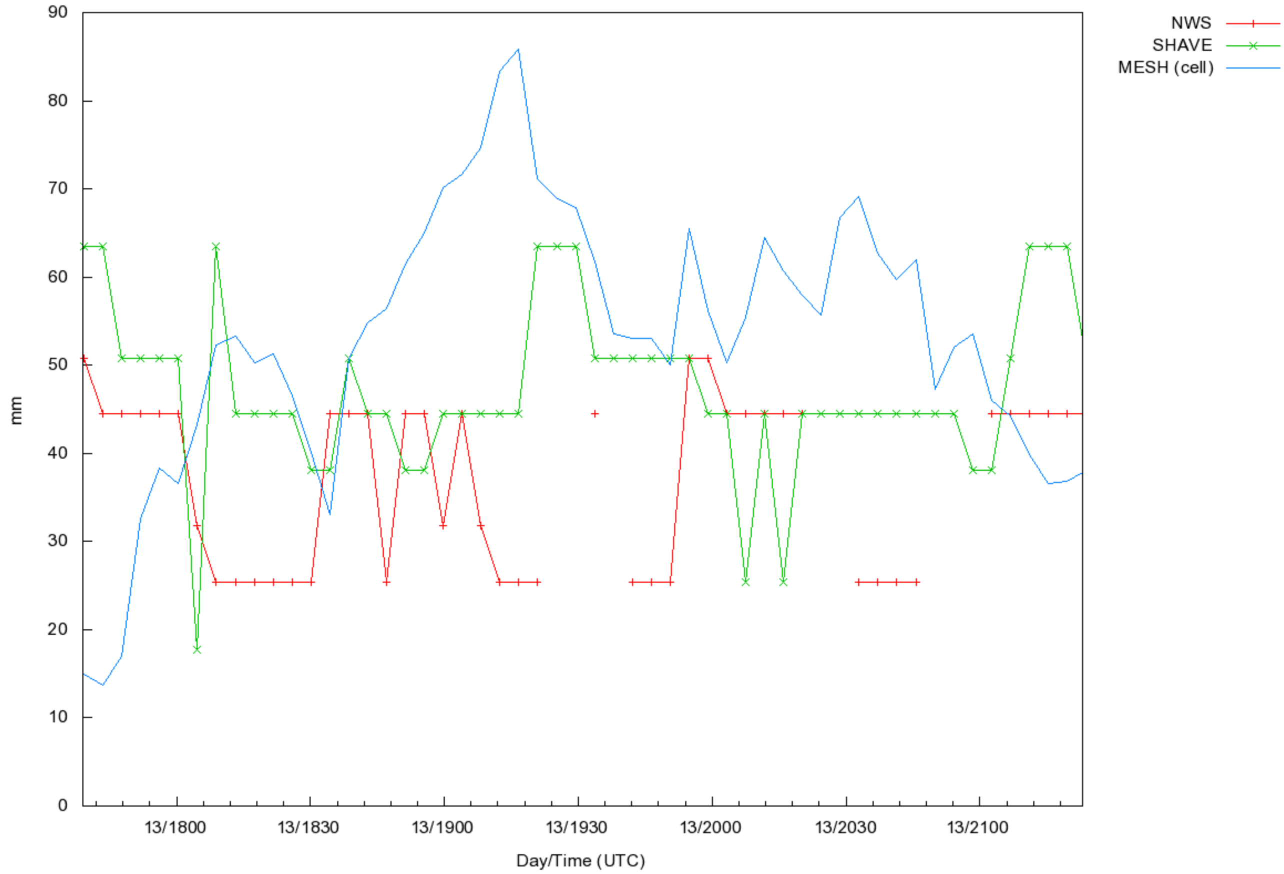

Figure 7: Time series of NWS and SHAVE reports, and cell-based MESH from the storm on 13 August 2008. In this case the NWS reports were consistently of smaller diameter than SHAVE reports

hail path as described by SHAVE and NWS reports (Figures 1-5).

It was found that the differences in how reports from SHAVE or the NWS describe the surface hail fall led to differences in technique evaluation (Figures 6 and 7). The study also discovered that careful radar interrogation matched or exceeded the ability of any of the techniques employed to discriminate severe hail.

Future work will consider using sector searches to better associate attributes and hail sizes. In fact, the SHAVE database, given it's spatial accuracy and resolution, lends itself to be used in a more complex fashion (beyond circular searches) to improve hail diagnosis techniques and the evaluation of those techniques.

Acknowledgements

The authors wish to thank the many students who have worked on SHAVE over the past three years. Their hard work and dedication during summer vacation have made SHAVE a great success. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of the National Severe Storms Laboratory, the National Weather Service Storm Prediction Center or the U.S. Department of Commerce. Funding for this research was provided under NOAA-OU Cooperative Agreement #NA17RJ1227.

References

Johnson, J. T., P. L. MacKeen, A. Witt, E. D. Mitchell, G. J. Stumpf, M. D. Eilts, and K. W. Thomas, 1998: The Storm Cell Identification and Tracking (SCIT) algorithm: An enhanced WSR- 88D algroithm. *Wea. Forecasting*, **13**, 263-276.

Lemon, L.R., 1998: The Radar "Three-Body

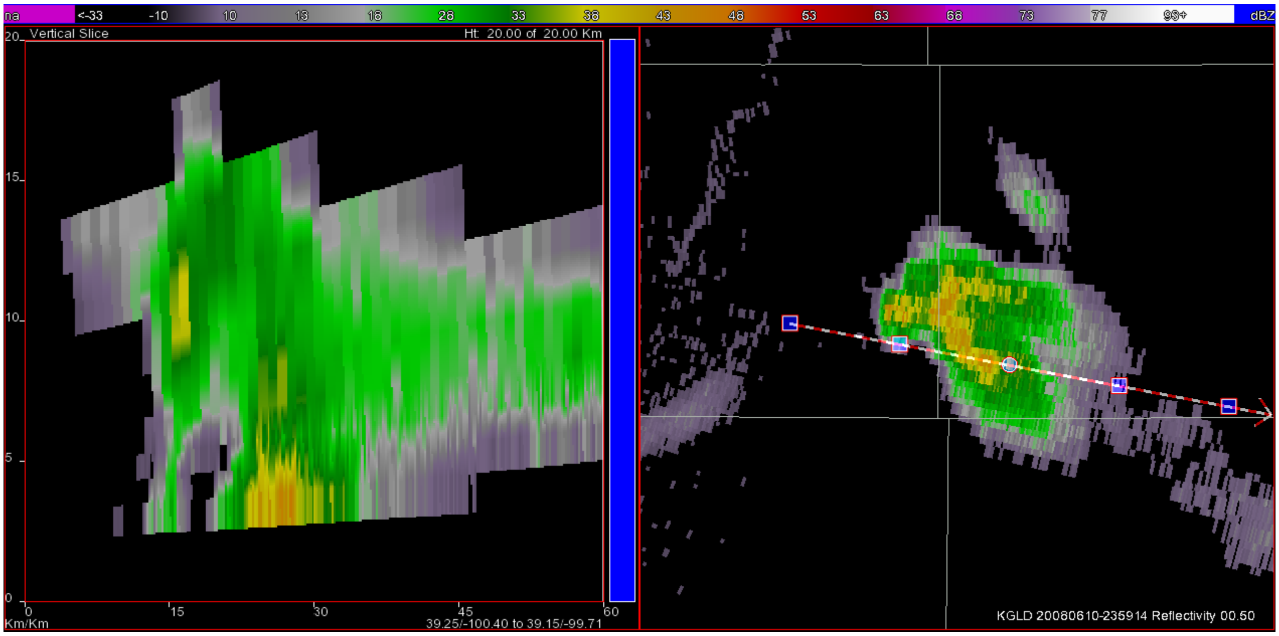

Figure 8: Cross section of LP supercell on 10 June 2008 at 2359 UTC. This was approximately 15 min before baseball-sized hail and larger began falling at the surface. The cross section is 60 km long, starting 112 km from the Goodland, KS WSR-88D.

Scatter Spike": An Operational Large-Hail Signature. *Wea. Forecasting*, **13**, 327–340.

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

Smith, T. M., K. L. Ortega, K. A. Scharfenberg, K.L. Manross, and A.Witt, 2006: The Severe Hail Verification Experiment. *23rd Conference on Severe Local Storms*, St. Louis, MO, Amer. Meteo. Soc., 5.3

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

Trapp, R.J., D.M. Wheatley, N.T. Atkins, R.W. Przybylinski, 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.

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

Witt, A., M. D. Eilts, G. J. Stumpf, E. D.

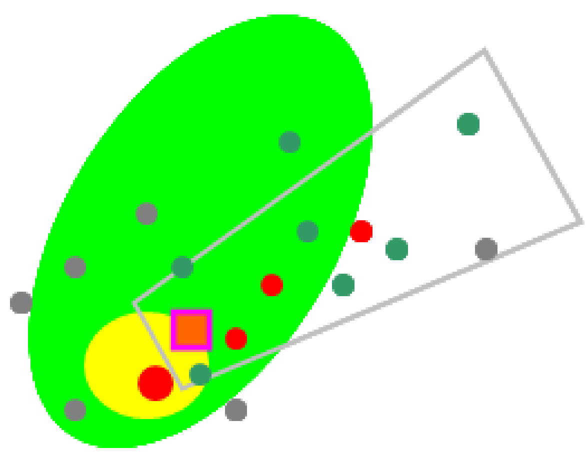

Figure 9: Sector search concept. By using storm motion and cell identifications (orange box), a sector window could be used to evaluate hail fall downstream (no hail reports are represented by gray circles, non-severe by green circles and severe by red circles). The green, yellow, red areas are reflectivity

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

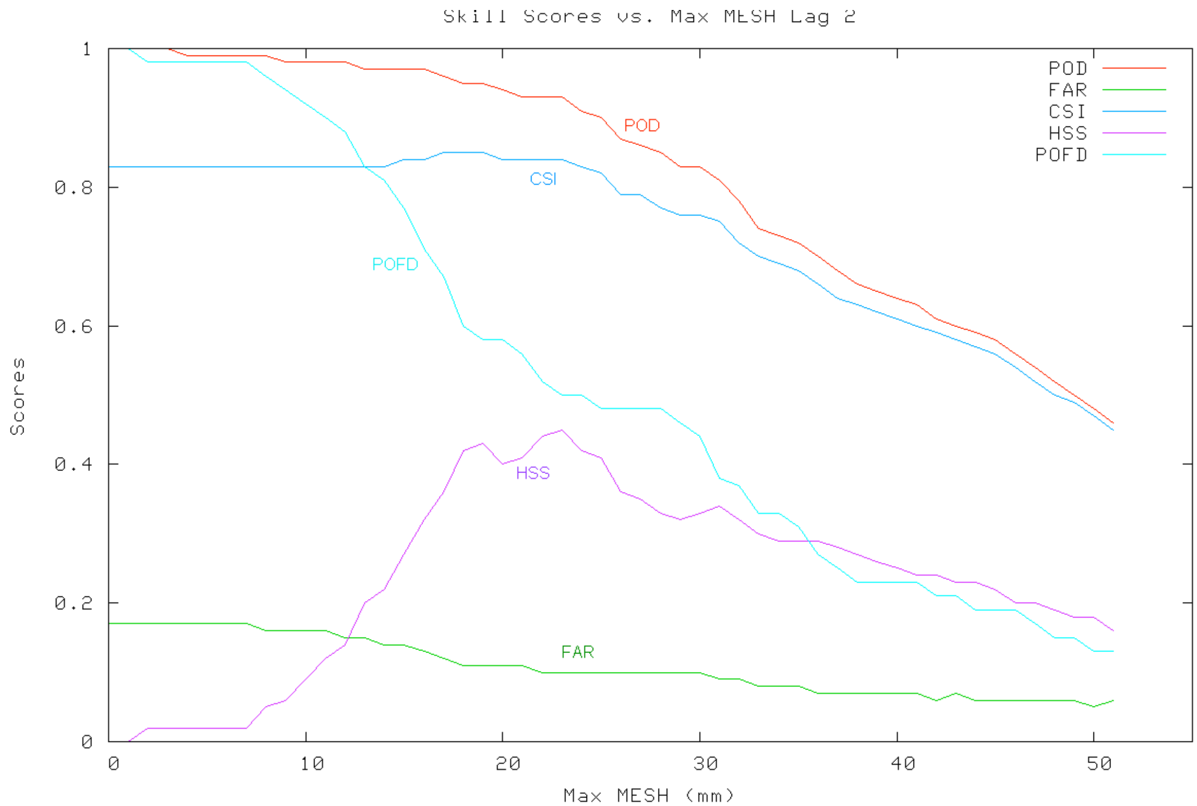

Figure 10: The end goal for the study. Using SHAVE reports different skill scores are computed for different techniques (grid-based MESH in this case) for diagnosing severe hail. However, the skill scores are calculated for certain lead times (in this case, 10 min) to further associate certain attributes or signatures with severe hail.