EVALUATION OF NWS STORM-BASED WARNINGS USING GRIDDED PRODUCTS

Todd R. Ferebee[1, 2], Kiel L. Ortega[3], and Kevin A. Scharfenberg[4]

[1]National Weather Center Research Experiences for Undergraduates Program
and
[2]North Carolina State University, Department of Marine, Earth and Atmospheric Sciences

[3]Cooperative Institute for Mesoscale Meteorological Studies,
University of Oklahoma, and NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

[4]National Weather Service, Office of Climate, Water, and Weather Services, Norman, Oklahoma

ABSTRACT

In 2008, the National Weather Service began issuing storm-based polygon warnings instead of county warnings. Only one severe hail, wind, or tornado report is needed to verify an entire warning polygon. Few severe weather reports in the warning, and in turn for the storm which prompted the warning, makes it difficult to determine the spatial extent of severe weather for a particular storm. Since 2006, the Severe Hazards Analysis and Verification Experiment (SHAVE) has been collecting severe weather reports at temporal and spatial resolutions much higher than those available in *Storm Data.* The National Severe Storms Laboratory (NSSL) produces several severe weather products, such as reflectivities at different isotherms and estimated hail size, on a grid for the entire contiguous United States. These grids could provide for synthetic verification of severe weather especially for the spatial extent of severe weather. This study will investigate how well the grids perform in determining where severe hail fell by using high resolution SHAVE reports. Discussion for applications of such grids for warning verification and improvement will also be included.

1. INTRODUCTION[1]

While there may be as many as 20,000 hail reports that are submitted to the National Weather Service (NWS) each year, the spatial and temporal accuracy and resolution of these reports is rather problematic. This is the main reason why the Severe Hazards Analysis and Verification Experiment (SHAVE; Ortega et al. 2009) group was formed in 2006. Utilizing Google Maps, digital telephone number directories, and other various computer software programs, up to 10 University of Oklahoma Meteorology undergraduate students work within the National Severe Storms Laboratory (NSSL) making phone calls with the intent of collecting high-resolution data describing severe hazards such as wind, hail, and flooding reports. It is these high-resolution, ground truth datasets that can then be used to improve the NSSL's prototype multi-sensor and multi-radar algorithms. However, current severe weather warnings are not verified using these high resolution (1 square km approximately) datasets provided by SHAVE. Instead, lower resolution (100 square km approximately) *Storm Data* datasets are used which is problematic because it often becomes difficult to describe the total area impacted by the severe weather event at such a low resolution of reports as seen in Figure 1. In addition, the temporal spacing between individual SHAVE hail reports is lower than in the *Storm Data* hail reports; thus, for any given storm day, SHAVE receives far more hail reports than *Storm Data*. SHAVE has also been shown to be far more accurate and reliable than *Storm Data* which has reports that are far too sparse for use in many studies (Witt et al. 1998, Trapp et al. 2006).

Throughout the course of this study, 22 different multi-sensor and multi-radar severe

[1] *Corresponding author address:*
Todd Ferebee, North Carolina State University, 2300 Avent Ferry Rd., Apartment B3, Raleigh, NC 27606
E-mail: trferebe@ncsu.edu

weather products were tested for eight different cases in which there were numerous SHAVE hail reports to verify the size of the hail that fell in each of the storms (Tables 1 and 2). An experimental real-time severe weather data application, the Warning Decision Support System of Integrated Information (WDSSII; Lakshamanan et al. 2007), was used to perform these tests. One of the many advantages of using WDSSII is that it merges single WSR-88D radar data onto a Cartesian grid which gives the user the ability to receive faster radar updates than with conventional radar. In addition, WDSSII gives the user the ability to create temporal and spatial swaths of various severe weather products. By investigating these 22 different algorithms for both non-severe and severe cases, the goal of this research was to:

- evaluate the performance of these NSSL algorithms
- provide forecasters with the knowledge of which algorithms performed the best with the intended goal of improved warning verification
- give forecasters the ability to determine new ways of verifying warnings, based on swaths of data rather than just a few scattered reports

2. DATA AND METHODOLOGY

2.1 *THE CASES*

As previously mentioned, eight cases were included in our analysis of the different severe weather algorithms used by the NSSL. Four of the cases were multicellular-type storms and the other four of them were supercellular as seen in Table 1. Supercellular storms contained a sustained mesocyclone based on radar examination. The other storms showed more typical multicellular characteristics. The locations of the storms were quite diverse, with some of them in the Midwest and others in the Eastern US. As would be expected based on climatology, most of the multicellular-type storms were concentrated in the Eastern US and most of the supercellular-type storms were concentrated in the Midwest. Most of the cases used in this study were chosen because they had a large number of SHAVE reports associated with them, which is important for the verification part of this study, as a low number of ground-truth observations would lead to incomplete information about the true nature of the hail sizes produced by the storm. However, it should be noted that for the multicellular storm type, a lower number of

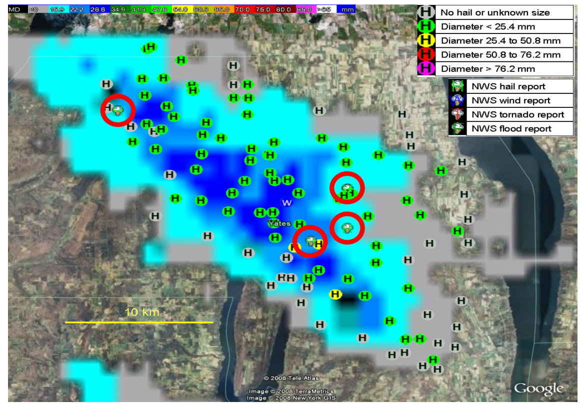


Figure 1: This figure is taken from the July 16, 2008 case in West Central New York. Notice the resolution difference in SHAVE hail reports vs. *StormData* hail reports. Note: Red circles denote *StormData* hail reports.

Storm Date	***Storm Location***	***Storm Type***	***Number of reports***
20070620	South-Central MN	Supercell	84
20080610	Northern KS	Supercell	54
20080714	Southeast NC	Multi-cell	27
20080716	West Central NY	Supercell	196
20080722	Northeastern OH	Multi-cell	31
20090624	Northeastern KS	Multi-cell	44
20090714	Northwestern IA	Multi-cell	77
20090716	Central OK	Supercell	140

Table 1: This table contains a description of the 8 cases used in the study including the storm date, location, type, and number of SHAVE reports included in our analysis.

FIELDS UNDER INVESTIGATION
Average Reflectivity Below the Wet Bulb Freezing Level
Azimuthal Shear at the Lowest Altitude
Average 0-3 km Azimuthal Shear
Average 3-6 km Azimuthal Shear
18 DBZ EchoTop
30 DBZ EchoTop
45 DBZ EchoTop
50 DBZ EchoTop
55 DBZ EchoTop
60 DBZ EchoTop
65 DBZ EchoTop
Height of 50 DBZ Reflectivity above the Freezing Level
0-3 km AGL Merged Azimuthal Shear
3-6 km AGL Merged Azimuthal Shear
MESH (Maximum Estimated Size of Hail)
POSH (Probability of Severe Hail)
0 Degrees Celsius Reflectivity
-10 Degrees Celsius Reflectivity
-20 Degrees Celsius Reflectivity

Reflectivity at the Lowest Altitude
SHI (Severe Hail Index)
VIL (Vertically Integrated Liquid)

Table 2: A table containing all 22 NSSL algorithms being investigated in this study.

observations were provided due to the type of storm and a variety of other factors.

2.2 *THE PROCESS*

The first step in conducting this research was to create the temporal and spatial swaths of the 22 different severe weather products within the WDSSII command line. Once the swaths were created, the SHAVE hail reports were then added to the swaths for comparison which were also created in the WDSSII command line as seen in Figure 2.

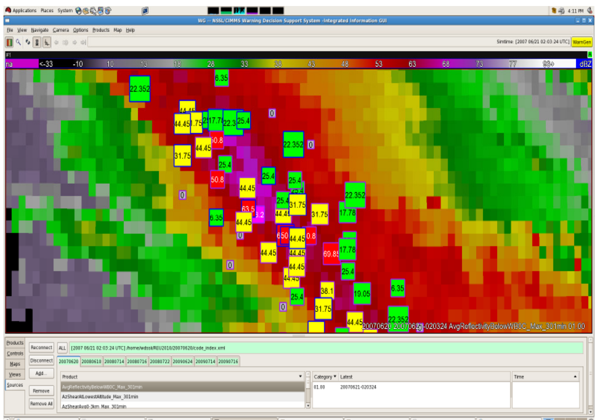


Figure 2: The Average Reflectivity Below the Wet Bulb Freezing Level product overlayed with SHAVE hail reports (mm) for the June 20, 2007 case in south-central Minnesota.

Next, all the SHAVE and swath data was exported into an Excel spreadsheet in order to prepare the data for data analysis. Before discussing the process of data analysis, it becomes necessary for a description to be given about how each individual SHAVE hail report was matched up with one unique value of the product analyzed at its corresponding grid point. In WDSSII, an algorithm called "PointMatch" matches up the corresponding geolocation of the SHAVE hail report, called "Truth", with the exact value of the product at that specific time and geolocation within the corresponding grid cell. It is this algorithm that made the data analysis process possible for this study. However, it should be mentioned that some uncertainty exists in the exact reported hail times and the exact time when the radar image was updated. Figure 3 provides an excellent illustration of how this process works.

In the data analysis process, over 700 individual box-and-whisker plots were generated for each of the eight individual cases containing all 22 different severe weather algorithms for four different hail size categories. Also, an additional 88 individual boxplots were produced for each of the algorithms with the combined data from all eight cases contained within them for the purpose of evaluating the performance of each individual algorithm.

The hail size categories that were used in this study include the following:

- "No hail": No hail was reported
- "Non-Severe Hail": All hail sizes less than 1 inch (25.4 mm) were included
- "Severe Hail": All hail sizes greater than 1 inch (25.4 mm) were included
- "Significant Severe Hail": All hail sizes greater than 2 inches (50.8 mm) were included

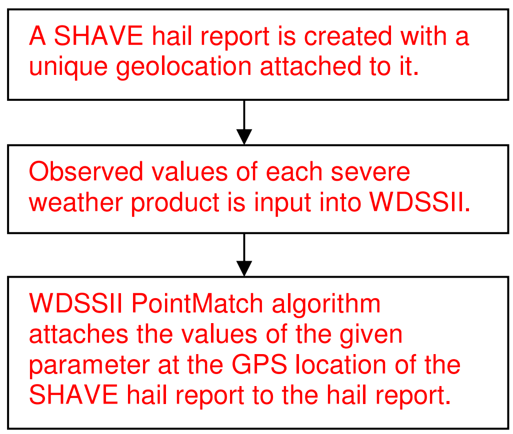


Figure 3: An illustration of how the "PointMatch" algorithm associates the value of a given product to a SHAVE hail report.

If more time were given, the severe hail category would have only included hail sizes in the one to two inch range to prevent overlapping severe hail and significant severe hail size bins in the generated boxplots. However, given the general scope of this project and the limited time available, this hail size range still provides the scientific community with a general idea of how well the given parameters performed in this study.

3. *RESULTS AND DISCUSSION*

3.1 *THE RESULTS*

As mentioned in the introduction section, one of the main goals of this study was to evaluate the performance of 22 well known NSSL algorithms for a total of eight individual cases. To test these algorithms, individual boxplots based on storm type and algorithm types were produced in the statistical program “R”. The performance of the algorithms was based on how well the given algorithm was able to delineate between the various hail size categories included in the study. Approximately one-half of the algorithms that were tested performed reasonably well based on subjective visual examination of the boxplots. A listing of those algorithms which performed the best is given in Table 3.

A list of (11) Best-Performing NSSL Algorithms
30 DBZ EchoTop
45 DBZ EchoTop
50 DBZ EchoTop
55 DBZ EchoTop
60 DBZ EchoTop
Height of 50 DBZ Reflectivity Above the Freezing Level
VIL (Vertically Integrated Liquid)
MESH (Maximum Estimated Size of Hail)
SHI (Severe Hail Index)
POSH (Probability of Severe Hail)
-20 Degree Celsius Reflectivity

Table 3: These 11 parameters showed the best skill upon evaluation.

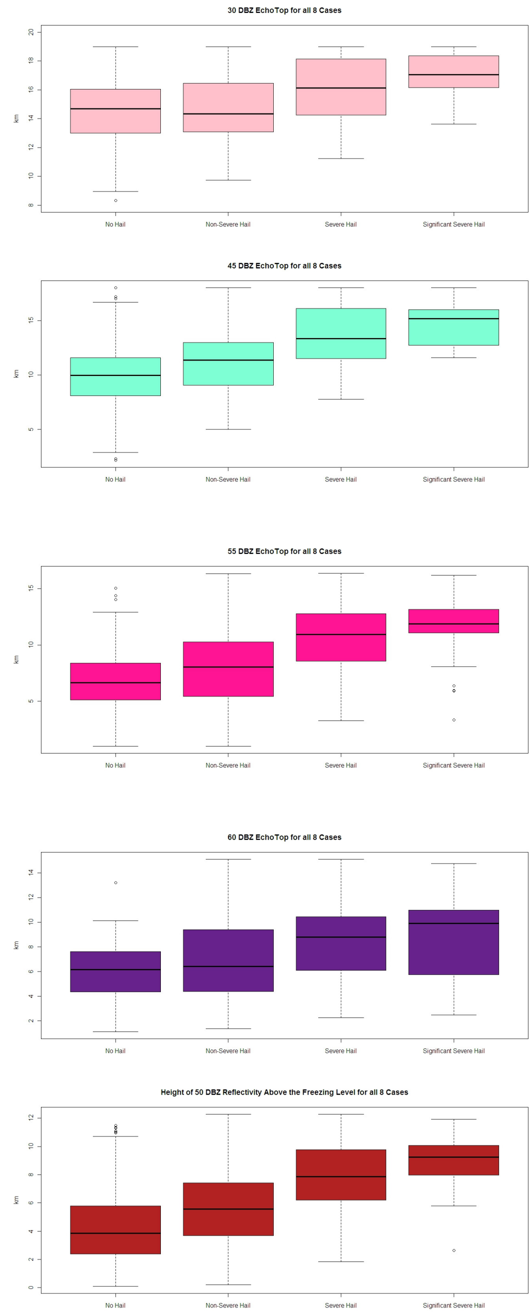
3.2 *BOXPLOTS AND DISCUSSION*

In describing the generated boxplots, the results were broken down by storm type, hail size category, and by various hail size categories to see if any trends could be noted in the dataset. After the data analysis process was complete for all the algorithms and all the cases, there were a couple of noteworthy trends in the dataset for the eleven chosen algorithms that performed the best:

1) When the boxplots for the non-severe hail reports were compared with the severe hail reports, a clear distinction could be made amongst the various hail size categories

2) In comparing the boxplots for the non-supercellular hail cases with those of supercellular hail cases, there was a drastic reduction in the performance of many of the algorithms evaluated

In Figure 4, a compilation of 11 boxplots containing data from all eight cases is given for the best performing NSSL algorithms. Notice how clearly the distinction can be made amongst the non-severe versus severe-sized hail categories.



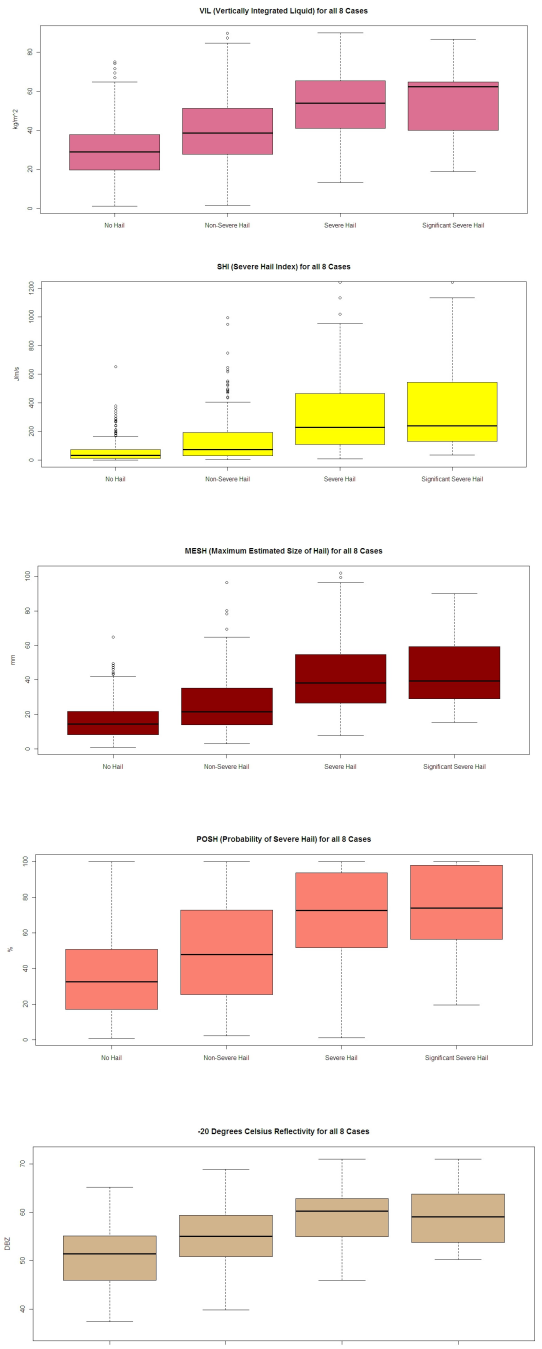


Figure 4: These boxplots show the distribution of WDSSII algorithm output for various hail size categories from SHAVE data. Boxes represent the 25th to 75th percentile and the whiskers represent 1.5 times the interquartile range, with the heavy horizontal black line indicating the median.

Furthermore, Figure 5 illustrates an example of a comparison between a non-supercell case and a supercell case. Notice how there appears to be a drastic reduction in the performance of the given algorithms when switching from supercellular type thunderstorms to multi-cellular type thunderstorms. However, this result was not seen in all the various tested algorithms.

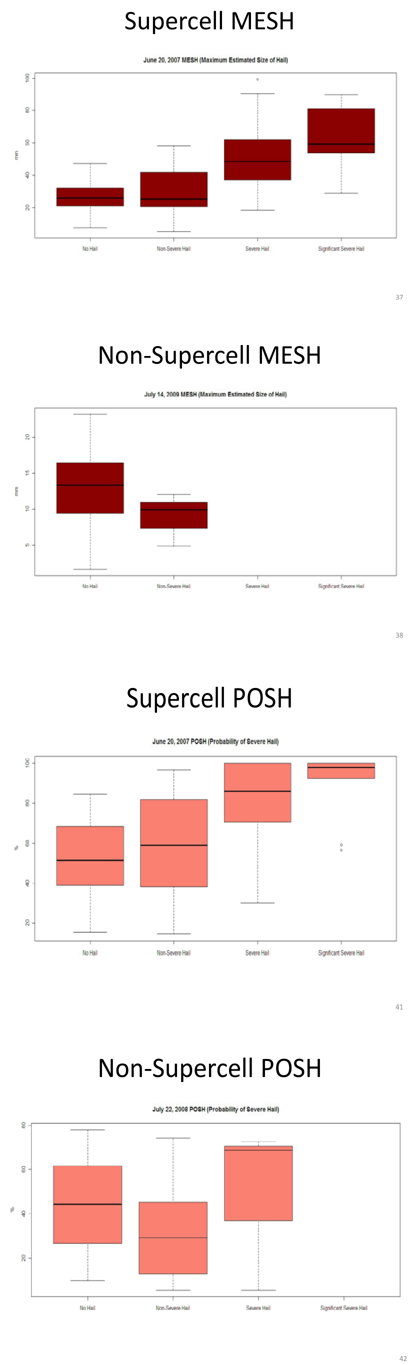


Figure 5: As in Figure 4, except comparing a supercell case to a non-supercell case for two WDSSII algorithms.

4. CONCLUSIONS

This study has shown how high-resolution, ground truth datasets can be used to improve the algorithms used by forecasters in determining hail size. Throughout the course of this study, 22 different multi-sensor and multi-radar severe weather products were evaluated for eight individual cases and, after subjective visual analysis, 11 were found to be relatively skillful in delineating the various hail size categories from one another. A few selected non-supercell cases and supercell cases were also compared and it was shown that some of the available experimental NSSL algorithms appeared to have more skill in supercell cases than in non-supercell cases. This study has provided forecasters with the knowledge of which algorithms to use when forecasting for hail with the ultimate hope that in the future, these results can be used by forecasters to improve warning verification.

5. ACKNOWLEDGMENTS

A special thanks to Kevin Manross for all his help in the beginning stages of this project.

This material is based upon work supported by the National Science Foundation under Grant No. ATM-0648566. This work was prepared by the authors with funding provided by National Science Foundation Grant No. ATM-0648566, and NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA17RJ1227, U.S. Department of Commerce. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, NOAA, or the U.S. Department of Commerce.

6. REFERENCES

Lakshmanan, V., T. Smith, G. Stumpf, and K. Hondl, 2007: The Warning Decision Support System - Integrated Information. *Wea. and Forecasting,* **22,** 596-612.

Ortega, K. L., T. M. Smith, K. L. Manross, A. G. Kolodziej, K. A. Scharfenberg, A. Witt, and J. J. Gourley, 2009: The Severe Hazards Analysis and Verification Experiment. *Bull. Amer. Meteor. Soc.,* **90,** 1519-1530.

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. and Forecasting,* **21,** 408-415.

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. and Forecasting,* **13,** 513-518.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Hail reports available through NWS Storm Data are temporally and spatially sparse, making it difficult to describe the entire event

Only one report of large hail is needed to verify an entire Severe Thunderstorm Warning area

Data from multiple radars can be merged together to create time-accumulated algorithm output, such as the maximum expected hail size "swaths"

WHAT THIS STUDY ADDS

Dense ground reports of hail (and no hail) are matched up with gridded radar-based hail algorithm output for evaluation

Several radar-based algorithms are identified as good performers to aid in hail warning operations and event evaluation

The radar-based hail algorithms perform better for storms identified as supercells when compared to multicells