THE SEVERE HAZARDS ANALYSIS AND VERIFICATION EXPERIMENT

BY KIEL L. ORTEGA, TRAVIS M. SMITH, KEVIN L. MANROSS, KEVIN A. SCHARFENBERG, ARTHUR WITT, ANGELYN G. KOLODZIEJ, AND JONATHAN J. GOURLEY

Active collection of severe weather reports via telephone interviews enables students to build a high-resolution dataset useful for radar ground truth and other purposes.

In general, severe storm verification from the National Weather Service (NWS) has spatial and temporal scales similar to associated severe weather warnings (Hales and Kelly 1985), on the order of 1,000 km² and tens of minutes. Although this spacing may be useful in verifying warnings, the spacing is too coarse to verify new severe weather applications that have high temporal (on the order of 1–10 min) and spatial (on the order of 1 km) resolution. Reports of nonsevere weather by the NWS (e.g., hail less than 1.9 cm in diameter; NWS 2005) and confirmed reports of no severe weather near a severe thunderstorm are not widely available, yet are critical with respect to evaluating the performance of high-resolution applications and models. Furthermore, Witt et al. (1998b) and Trapp et al. (2006) have cited inconsistencies in the reported times, locations, and severe weather magnitudes (e.g., hail size) in the U.S. official climate record of severe weather events, the National Climate Data Center’s Storm Data publication (NWS 2007).

The Warning Decision Support System—Integrated Information (WDSS-II; Lakshmanan et al. 2007; Hondl 2002) generates 3D grids of specialized experimental severe weather products by integrating data from multiple radars and other sensors. With the onset of Weather Surveillance Radar-1988 Doppler (WSR-88D) base-data distribution over the Internet (Droegemeier 2002) and inexpensive fast computing, it is possible to run WDSS-II for the entire contiguous United States (CONUS) at a resolution of approximately 1 km or better, both horizontally and vertically, and on the order of 1–5 min (Lakshmanan et al. 2006). Such a system is collocated at the National Severe Storms Laboratory (NSSL) and Storm Prediction Center (SPC), with experimental products shared online (at http://wdssii.nssl.noaa.gov; Smith and...
When these high-resolution data are coupled with geographic information, it becomes possible to make a detailed assessment of when and where hazardous, severe weather events may have occurred. Evaluation of these research products requires an observational dataset more detailed than what is currently available in Storm Data.

The Severe Hazards Analysis and Verification Experiment (SHAVE), originally called the Severe Hail Verification Experiment (Smith et al. 2006), was conducted by NSSL each May-August of 2006–08. SHAVE was designed to take advantage of the ability to blend high-resolution radar data with geographical information systems (GISs) to accomplish the primary goal of collecting high temporal and spatial resolution hail reports in hail swaths from thunderstorms (Figs. 1 and 2). These reports were collected via verification phone calls to locations along and near the storm’s path immediately following the storm’s passage. Meteorology undergraduate students enrolled at the University of Oklahoma (OU) were hired to make the phone calls. Scientists from NSSL and OU’s Cooperative Institute for Mesoscale Meteorological Studies provided guidance on regions to focus phone calls.

The initial scientific objectives of SHAVE were as follows:

- use the high-resolution verification data in the development of techniques for probabilistic hazard grids for severe thunderstorms (Ortega 2008);
- evaluate the performance of multisensor, multiradar hail algorithms (Ortega et al. 2006);
- associate changes in hail size distribution with storm evolution; and
- enhance climatological information about hail in the United States.

Although the primary goal of SHAVE was to collect high-resolution hail reports, each year the goals of the project were expanded. In 2007, wind and tornado damage were added to the verification efforts. In 2008, flash flooding was added to the verification efforts. These additional objectives often required different data collection strategies. Wind verification, for example, occurred the day following an event. Preliminary reports of wind damage, compiled by the SPC, were overlaid in a GIS viewer and phone surveys were conducted in the areas surrounding the area of the reported storm.

**DATA COLLECTION FACILITIES.** SHAVE was conducted within the National Oceanic and Atmospheric Administration’s, Hazardous Weather Testbed (HWT; in Norman, Oklahoma) as part of the Experimental Warning Program (Stumpf et al. 2008). SHAVE operations take place in NSSL’s Development Laboratory located in the National Weather Center in Norman, Oklahoma. The major facilities required in SHAVE are as follows:

- Computing cluster [described in detail by Lakshmanan et al. (2006)]: The computer cluster

![Fig. 1. SHAVE and Storm Data report data overlaid in Google Earth on an NSSL radar-based hail swath from a storm on 27 Jul 2006 in Lac qui Parle County, MN. Where NWS reports are attached by a line, it is one report in Storm Data, with a start and end location.](image1)

![Fig. 2. SHAVE and Storm Data overlaid in Google Earth on a NSSL radar-based hail swath for a storm on 16 Jul 2008 in Yates County, NY.](image2)
enabled the processing of radar data and derived products in a format suitable for displaying in a GIS viewer. Examples of this type of data are available online (at http://wdssii.nssl.noaa.gov).

- **Display machines:** The NSSL Development Laboratory (Fig. 3) provides room for up to 15 computers with adequate display, processing, and memory capabilities to run Google Earth Pro\(^1\) and an Internet browser. The computers were used by the students during operations to connect to the computing cluster, display appropriate data, and enter reports into a common database. The operations coordinator’s (described in the following section) computer also had the capability to run utilities to help parse and send phone number information to the student callers.

- **Collaboration displays:** Four 1.27-m (50 in.) plasma screen monitors were used to display radar information relative to geographic features of the storm being investigated, a Web page listing recent phone calls, and an overview of the day’s operations that could be viewed by visitors.

- **Telephone lines:** A telephone and headset was available for each student caller; and

- **Geographic information:** The geographic information used in the project comes from several sources. Google Earth (Fig. 4) provided street overlays, satellite views, and address and phone number information for locations such as businesses. Residential address and phone number data were obtained from two different sources: rural plat maps, which contain property ownership records and county directory information, and Delorme Street Atlas USA\(^2\), which contains an address and phone number database for the CONUS. Plat maps provide nondigitized area street maps, the relative positions of residences on the streets, and their phone numbers. The digital Delorme phone number database provided a more convenient method for searching and disseminating phone numbers to the student callers when compared to using the plat maps (Fig. 4). The following two major problems were found to occur with phone numbers: disconnected phone lines and inaccurate geolocations (latitude/longitude associated with an address).

**SHAVE OPERATIONS.** There were three primary roles for SHAVE participants: project scientist, operations coordinator, and student caller. The project scientists developed the survey scripts, operations plan, data collection infrastructure, and scientific objectives for the project. The operations coordinator was responsible for determining which storms to interrogate on a daily basis, collecting the phone numbers from plat maps and Delorme’s database, and directing the student callers. The student callers were responsible for making the survey phone calls.

A typical day during the project began with the decision by project scientists whether or not to commence operations. On most days this coincided with the SPC’s issuance of the 1630 UTC Day 1 Convective Outlook. Several other factors were also considered when deciding to commence operations, including the following:

- whether hail producing storms would develop in areas that can be easily surveyed;
- whether widespread severe winds or wind damage had been reported the day before;
- whether flash flooding had been reported or if rainfall accumulations had exceeded flash-flood guidance values;
- whether storms would form before 2130 LT (end of the experiment day);
- anticipated storm modes; and
- staff availability and fatigue.

Notification of the operational status of the project was posted on the project’s Web site and distributed to participants via e-mail each morning. Participants reported to the operations area at the time specified

---

\(^1\) Google Earth Pro was adopted by SHAVE; however, any mapping service is suitable. SHAVE uses Google Earth because of its support of Keyhole Markup Language (KML). The use of Google Earth by NSSL is not an endorsement.

\(^2\) Any digital address and phone number database is suitable. The use of Delorme’s software by NSSL is not an endorsement.
in the notification message and operations began shortly thereafter.

The data collection effort proceeded with at least three participants to fill the three primary roles (student callers ranged in number from one to five, depending on weather and availability). The project scientist was either on duty or on call each day to determine the status of operations and to help solve any technical problems. The operations coordinator made decisions on the type of storms to investigate, the exact storms to examine, and finding the disseminating geographically referenced phone numbers to the student callers. The operations coordinator occasionally took part in making phone calls, if the workload allowed for it. The rest of the team made phone calls and collected data from the public about the particular threat under investigation.

The operations coordinator called an end to operations based on several conditions, including the following:

- storm activity had declined and was no longer anticipated to increase;
- storms were still active but it was past 2130 LT in the area of concern;
- storms were located in unpopulated areas or areas with inaccurate phone number geolocations;
- all possible locations for wind and flash-flood impacts had been surveyed; and
- staff fatigue levels and anticipated level of operations for the following day(s).

At the close of operations, the operations coordinator ran a program that archived the data and created and sent summary reports to the project participants and affected NWS offices.

**DATA COLLECTION.** Phone surveys were developed for hail, wind, and flash-flood events. Phone number collection was done under the philosophy of “more is better; all is best” (Fig. 2). The original intent for hail surveys was to collect phone numbers in such a way that cross sections could be completed at evenly spaced intervals along the storm’s path. However, it was quickly learned that this would not lead to evenly spaced high spatial resolution data for two primary reasons: a low successful call rate was achieved—that is, 33.3% of the phone calls resulted in a successful data point (Table 1), and calling all numbers along a cross section was time consuming, leading to poor spatial resolution down the path of the storm.

Phone numbers of businesses and residences were collected using several sources. Rural county plat maps proved to be invaluable sources during SHAVE 2006. These maps essentially provided a county phone book with maps of area roads and the location of each residence relative to the roads. These plat maps, combined with high-resolution satellite photographs viewable in Google Earth, led to high confidence in the exact location of many SHAVE reports. During SHAVE 2006, phone numbers were also found using features within Google Earth (Fig. 4). One feature was the “Layers” menu (bottom left in Fig. 4). This menu allowed for the student callers to overlay many maps and search and overlay different layers (such as businesses). The red telephone icons on the map indicate where SHAVE has found phone numbers, either through plat maps or Delorme’s phone number database. For privacy reasons, contact information has been changed in generic values, such as phone numbers to 555-555-5555.

![FIG. 4. Image showing phone number information overlaid in Google Earth. (left) The two menus for searching and overlaying different layers (such as businesses). The red telephone icons on the map indicate where SHAVE has found phone numbers, either through plat maps or Delorme’s phone number database. For privacy reasons, contact information has been changed in generic values, such as phone numbers to 555-555-5555.](image)

3 A successful call was one in which the survey was marked as "good" (meaning that all the information was acceptable) or "questionable time" (meaning that all the information, except for the reported times, was acceptable). Student callers subjectively determined whether information was acceptable or not.
different features, including businesses. If a student caller clicked on these features, contact information appeared in a pop-up window. Another feature was the “Search” menu (top left in Fig. 4), which allowed for the student callers to find additional business listings not available through the Layers menu. These two features allowed for rapid, automatic placement of many phone numbers using very little time by the student callers; however, there were a few concerns about the use of these numbers. One concern was the inaccurate geolocations associated with these points. The other concern was that nearly all of the search results and all of the features available are businesses. On many occasions, phone surveys with businesses required too much time because SHAVE student callers would need to either talk to the management of the business (many of whom were usually not readily available) or the business would be reluctant to reveal the severity of the weather they received or information about any damage that may have occurred.

During SHAVE 2007 and 2008, most phone numbers were residential phone numbers found using Delorme Street Atlas. Delorme Street Atlas contains a phone number database, giving SHAVE participants access to addresses and phone numbers in the United States and Canada. The operations coordinator could search for phone numbers in Delorme, by either clicking on individual streets or using a search (e.g., a ZIP code). The operations coordinator could then export that data to Google Earth for georeferencing (Fig. 4). The operations coordinator would save the data into a KML file and send it to the student callers. The digital phone number database and automatic georeferencing did not completely replace plat maps because some rural locations still have poor addressing grids, resulting in inaccurate georeferencing. However, the database did allow for many more numbers to become available to the student callers, and it provided a much easier method to procure numbers.

The Delorme database had some influence on the character of the database as a result of the large number of phone numbers made available. The first effect was a reduction in the number of “questionable hail” reports (Table 1). The reason for this is that if a survey took place and the reported hail size was put in question by the student caller, then a “next door” neighbor survey would be completed. If the questionable report was indeed questionable/inaccurate, the survey was eventually marked as not completed instead of as questionable hail. This was done to ensure that inaccurate data were not accidentally used in any postprocessing of the reports. The second effect was an increase in “questionable time” reports (Table 1). The increase was also due to the adoption of a “30-min plus” rule during the 2007 experiment: if a survey occurred more than 30 min after hail fall, the survey was marked as questionable time. The 30-min plus rule was adopted after comments from student callers described a decreased quality of how people described the period during which the hail was falling. For example, instead of getting descriptions such as “the hail fell from 7:10 to 7:17,” student callers

<table>
<thead>
<tr>
<th>Table 1. Summary of phone surveys (number of calls) completed during SHAVE.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey completed information accepted (%)</td>
</tr>
<tr>
<td>All information accepted</td>
</tr>
<tr>
<td>Questionable time (did not know or survey occurred 30-plus min after storm passage)</td>
</tr>
<tr>
<td>Questionable hail (reported size did not correlate to nearby reports)</td>
</tr>
<tr>
<td>Questionable location (inaccurate geolocation of address)</td>
</tr>
<tr>
<td>Wrong location (phone number at different location)</td>
</tr>
<tr>
<td>Survey completed information rejected (%)</td>
</tr>
<tr>
<td>Busy/intercept operator</td>
</tr>
<tr>
<td>No answer</td>
</tr>
<tr>
<td>Do not call/disconnected</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>Survey not completed (%)</td>
</tr>
</tbody>
</table>

3 The use of Delorme’s software by NSSL is not an endorsement.
hours after or the day following a widespread wind event. In this mode of operation, local storm reports issued by the NWS were used to find areas to survey and reports of no wind damage were recorded. During wind surveys, the degree of damage to structures or trees and the time the wind damage occurred was the primary information that the survey was intending to collect; if wind speed estimates or measurements were volunteered, then the information was recorded. However, the survey only specifically asked for wind damage and the time the wind damage occurred. The students were to ask for specific details of the damage. For example, an initial comment of "tree fell" would be furthered to include the type and size of tree, whether it was uprooted or snapped, and the general condition of the tree. Figure 5 shows SHAVE and Storm Data verification from a bow echo that produced widespread damaging winds near Omaha, Nebraska, on 27 June 2008.

**Flash flooding.** Flash flood surveys were conducted if any of the following three criteria were met: 1) 6-h precipitation accumulation exceeded flash-flood guidance (FFG) values less than 24 h ago; 2) a flash-flood warning or urban/small stream advisory was issued by an NWS forecast office during the past 24 h; or 3) a survey for a different severe weather threat indicated flash flooding was a problem. The ratio product that compared precipitation to FFG values was generated locally at NSSL by acquiring the latest CONUS mosaic of FFG values from the NWS's National Precipitation Verification Unit and comparing it to rainfall estimates from NSSL's National Mosaic and quantitative precipitation estimation (QPE; NMQ; www.nmq.noaa.gov/) system. It was noted that the stitched CONUS grid of FFG values exhibited discontinuities at River Forecast Center (RFC) boundaries. These boundaries were discovered to be a result of the different methods used in the NWS to derive FFG and now gridded FFG (GFFG) values. The 6-h ratio product was the primary tool used to guide the location of the phone calls; however, there were occasions when NWS-issued warnings, advisories, and ratio products at 1- and 3-h duration were relied upon. The objective of the survey questions was to find information about flooding that resulted from rain that fell less than 6 h prior to the onset of flooding. To avoid flooding reports resulting from levee or dam breaches, or from flooding of main river stems, student callers cross-referenced a static flow accumulation map derived from a digital elevation model to check the size of the contributing basin areas. The flow accumulation map was color coded to highlight regions with flow accumulations less than 260 km$^2$, which is generally accepted as the threshold basin area for rainfall-induced flash floods (see, e.g., Davis 1998; Reed et al. 2007). Satellite views of the river channels available in Google Earth were also used to avoid large-scale fluvial flooding events.

The survey questions about flash flooding initially came from NWS Instruction 10-1605, Storm Data preparation (NWS 2007), but they were modified and supplemented based on conversations with forecasters at the NWS Arkansas–Red Basin RFC and researchers at the Office of Hydrology. The questions were focused on the location, nature, magnitude, frequency, and times flooding began and ended. In addition, any evacuations or rescues were included.

---

Fig. 5. SHAVE and Storm Data reports overlaid in Google Earth on a NSSL radar-based hail swath for a storm on 27 Jun 2008 near Omaha, NE. Where NWS reports are attached by a line, it is one report in Storm Data, with a start and end location.
in the reports. The caller first determined if the resident had experienced any type of flooding problems. If the answer was no, then the report was recorded as a null event and the call was terminated. If the response was positive, then the caller determined if the impacts of flooding were inundation or damage to property, including croplands, road or bridge closures, rivers or creeks out of banks, or backups due to insufficient drainage. Any or all of the previous impacts could have been recorded in the report by the caller. The Web page also enabled the callers to type in comments as needed.

The next question attempted to determine the magnitude of flash flooding through the depth and/or horizontal extent of the floodwaters. Caution was exercised in the survey to ensure residents did not approach moving or standing waters to estimate their depth. Several reference values of reported water depth were listed relative to car tires, mailboxes, or windows and then later converted to heights in metric units. In the case of flooded creeks or streams, questions focused on the lateral extent water was out of its banks. Questions about the frequency of flooding were found to be particularly informative. The caller would ask if the type of reported flooding occurs every time it rains, only when it rains hard, about once a year, about once every 5, 10, 50, or 100 yr, or “never seen it like this before.” Finally, the caller attempted to record the beginning and end times when flash flooding occurred or if it was still occurring. Figure 6 shows an example of flash-flood verification done by SHAVE and compared to reports available in Storm Data for an event in central Indiana on 7 June 2008.

**Dataset Uses and Results.** As shown in Table 2, SHAVE provides a unique database mostly populated with reports (nonsevere and confirmations of “null” events) not readily available from Storm...
Fig. 7. Distribution of reported maximum hail sizes for different MESH values. Of the 3921 SHAVE hail reports used, 1539 were severe (>19 mm in diameter).

Data. These data are important for algorithm evaluation work and provide a fuller account of an event than just using “severe” reports. The data collected during SHAVE have several possible uses because of the variety of questions asked during the surveys. Although maximum hail size and flooding depth have obvious applications in algorithm evaluation, questions such as, how much of the ground was covered by hail and how often the observed flooding occurs, could have very unique applications. Work continues into applications of the wind damage verification data. Two difficulties arise in the current method SHAVE student callers use to collect wind damage verification. One difficulty is in pinpointing areas in which to make verification phone calls for wind damage; further, a member of the public being surveyed may not know the full extent of wind damage in their area. The other difficulty is the wind damage data itself. Currently, it is more qualitative (description of damage) than quantitative (wind speed). One possible solution to this problem is to use the enhanced Fujita scale (LaDue and Mahoney 2006), which is currently being evaluated by SHAVE project scientists.

Hail data collected by SHAVE can be applied to a number of different research goals. One use is to evaluate gridded, multisensor hail diagnosis techniques (Stumpf et al. 2004). Wilson et al. (2009) used SHAVE hail data to investigate the skill in using gridded MESH as a synthetic verification tool and evaluated a number of different techniques for their skill in severe hail prediction. Preliminary results found limited skill in both evaluations, which may be a result of the wide range of hail sizes for a given MESH value, as revealed by SHAVE hail data (Fig. 7). Another, more unique, use of SHAVE hail data is to evaluate how well the public reports on hail sizes. Figure 8 shows a comparison between the estimated hail size and the measured hail size. The overall bias was –2 mm with a root mean square error of 12 mm, suggesting those who provided both estimates and measurements of maximum hail size slightly underestimated the measured hail size, but their estimates were within ±12 mm of the measured hail size. Further, as Fig. 9 shows, an evaluation of how frequently the public estimates

Fig. 8. Estimated hail stone diameter compared to measured hail sizes. Estimated hail size equals measured hail size along the diagonal line.
Skill in predicting flash floods. The flash-flood data also have applications beyond algorithm evaluation. The flash-flood survey asks for both the impact of the flooding (Fig. 10) and the frequency of the observed flooding (Fig. 11); these data could be combined to identify locations that are prone to flooding. For example, if a local SHAVE-type experiment occurred and found that a pond or stream was prone to severe flooding every time it rained, permanent warning signs could be placed near the flood-prone area.

**SUMMARY.** SHAVE presents a framework for conducting high-resolution remote verification of severe storms. The data collected during SHAVE provide a unique dataset that contribute to current research to more accurately and confidently (as a result of the high resolution) relate radar data to ground truth. The dataset is unique, not only for its spatial accuracy and resolution, but also for the types of reports contained within the dataset because a majority of the reports are “no hail/wind/flood” (Table 2). Also, some of the data obtained through SHAVE surveys has uses beyond meteorological uses, such as the economical results of flash flooding due to the frequency and effects of the flooding.

The project is unique in the fact that, except for the building of the project’s infrastructure and initial

---

7 Severe flooding was considered to be standing water greater than 0.91 m in depth, or moving water greater than 0.15 m in depth.
training, the project is largely student led and run. Opportunities to learn about severe storm forecasting, radar interrogation of storms, and severe storm hazards provide the students with an extraordinary learning experience and potential research topics for course work. Further, the students must apply critical thinking and teamwork during the verification process; combined with invaluable face time with project scientists, the project also provides participating students with great professional development experience.

The project Web site—which includes the project plan, facilities summary, example cases, and a data archive—is available online (http://ewp.nssl.noaa.gov/projects/shave/). SHAVE is expected to continue beyond 2009, with a focus on dual-polarization radar data and a possible expansion into winter weather verification.

ACKNOWLEDGMENTS. The authors would like to acknowledge the student callers—Jenifer Bowen, Chad Echols, Jessica Erlingis, Margaret Frey, Steve Irwin, Carey Legett, Tiffany Meyer, James Miller, Kelsey Mulder, Christa Riley, Brett Roberts, and Rachael Sigler—who made possible the collection of this unique dataset. Thanks to Jennifer Palucki, Rick Smith, Alan Stewart, and many others for their comments on the hail portion of the survey script. John Schmidt, Ernie Wells, and Seann Reed supplied useful comments on the flash-flooding script. Robert Coggins, Karen Cooper, Vicki Farmer, Zac Flamig, Doug Kennedy, and Joe Young provided essential computer support for the project. Thanks to Don Burgess, Greg Stumpf, and three anonymous reviewers, whose comments helped improve the quality of the paper. Funding was provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA17RJ1227 and NOAA’s High Performance Computing and Communications Office. Mentions of specific software in this manuscript are not endorsements of said software by the funding sources.

REFERENCES


Hales, J. E., Jr., and D. L. Kelly, 1985: The relationship between the collection of severe thunderstorm reports and warning verification. Preprints, 14th Conf. on Severe Local Storms, Indianapolis, IN, Amer. Meteor. Soc., 13–16.


Lakshmanan, V., T. Smith, K. Cooper, J. Levit, K. Hondl, G. Stumpf, and D. Bright, 2006: High-resolution radar data and products over the continental United States. Preprints, 22nd Int. Conf. on Interactive Information Processing Systems for Meteorology, Oceanography,


Educators, students, and weather enthusiasts! A glossary of over 3000 terms on weather and climate designed specifically for a general audience! Produced under the Project ATMOSPHERE initiative, the development of The Glossary of Weather and Climate was inspired by increasing contemporary interest in the atmosphere and global change. The objective of the glossary is to provide a readily understandable, up-to-date reference for terms that are frequently used in discussions or descriptions of meteorological and climatological phenomena. In addition, the glossary includes definitions of related oceanic and hydrologic terms.

©1996 American Meteorological Society. Available in both hardcover and softcover, B&W, 272 pages, $26.95 list/$21.00 member (softcover); $34.95 (hardcover). Prices include shipping and handling. Please send prepaid orders to Order Department, American Meteorological Society, 45 Beacon St., Boston, MA 02108-3693.