5.7

EVALUATION OF VARIOUS ALGORITHMS AND DISPLAY CONCEPTS FOR WEATHER FORECASTING

Indra Adrianto[1], Travis M. Smith[2*], Kevin Scharfenberg[2], and Theodore B. Trafalis[1]
[1]School of Industrial Engineering, University of Oklahoma, Norman, OK, [2]CIMMS/NSSL

1. Introduction

During the spring of 2004, a proof-of-concept test of the WDSS-II was conducted at the National Weather Service Forecast Office (NWSFO) in Norman, OK. The goal of this proof-of-concept test was to determine which products may aid forecasters in making more efficient tornado warning and severe thunderstorm warning decisions. This was determined by evaluating the usefulness of WDSS-II multi-sensor applications and display tools in the warning decision-making process. The hands-on evaluation of new products by National Weather Service (NWS) forecasters during the development stage of those products is a critical part of the design process. NSSL is working closely with the NWS's Meteorological Development Laboratory (MDL) to create a pathway into operations for the best of these new tools (Stumpf et al. 2005).

Researchers at the University of Oklahoma and the National Severe Storms Laboratory (NSSL) research scientists have developed various machine-intelligent algorithms and visualization techniques and have implemented them in the Warning Decision Support System - Integrated Information (WDSS-II; Lakshmanan 2002). The WDSS-II is a multi-faceted system that serves many purposes, such as:

- data ingest from multiple sensors such as WSR-88D and TDWR radars, surface observations, numerical models, lightning, and satellites;
- single- and multi-sensor real-time automated data analysis applications;
- a programming interface that allows developers to quickly implement new ideas;
- a four-dimensional data display system that allows users to rapidly interrogate data in a number of different ways.

*Corresponding Author Address: Travis M. Smith, CIMMS/NSSL, 1313 Halley Circle, Norman, OK 73069; email: Travis.Smith@noaa.gov.

New applications and display techniques may be tested in real-time or in post-event playback modes. Because applications can be rapidly prototyped in WDSS-II, it serves as an ideal platform to test new concepts in weather data analysis and warning decision-making before they are migrated into operational systems.

2. WDSS-II Proof-of-Concept Test

The WDSS-II real-time system at the Norman NWSFO was configured to receive data from many sources, including:

- base data and derived products from ten WSR-88D radars in and surrounding the Norman NWSFO County Warning Area,
- a real-time, dynamically-updating three-dimensional merged grid of reflectivity (Lakshmanan 2003) and reflectivity derivatives from these ten WSR-88Ds,
- base data from the Oklahoma City Terminal Doppler Weather Radar (TDWR), and
- polarimetric base data and precipitation classification algorithm output from the NSSL's KOUN research radar.

New single-radar products used by forecasters during the evaluation period included Azimuthal Shear, Divergence, and Rotation Tracks, which are based on the Linear Least Squared Derivatives application described by Smith and Elmore (2004). Multi-radar, multi-sensor products that were evaluated included grid-based fields of hail diagnosis parameters created from the merged 3D reflectivity (Stumpf et al. 2004) and near-storm environment (NSE) data, and a storm identification, tracking, and diagnosis algorithm (Smith and Stumpf 2005). The WDSS-II display evaluation examined: Continuous panning and zooming, b) Radar volume browsing (Lynn and Lakshmanan 2002), and c) 4D radar analysis tools including 2D cross-section and 3D box cross-section display (Hondl 2002; Stumpf et al. 2004).

For each of these components, we evaluate: a) Does the component provide new information that is useful for warning decision making? b) Is the display component easy to use? c) Is the algorithm component easy to interpret and understand? d) What improvements can be made to the component in order to make it more useful?

3. Evaluation Methodology

There are many techniques to evaluate a human-computer interface. Rubin (1994) categorized several techniques including surveys, usability testing, structured walkthroughs, and expert evaluations. Ebling and John (2000) used empirical data such as answers on questionnaires and comments made in protocols to evaluate an interface. In this project, an evaluation study was conducted in order to evaluate and improve the system. Training sessions, real-time observations, and surveys were included in this study.

3.1. Participants

The participants in this study were operational forecasters from the National Weather Service Forecast Office in Norman, Oklahoma. The requirements for the participants were 18 years of age or older, familiar with weather forecasting and with the use of computers and decision-making tools.

3.2. Training Sessions

In the beginning of training sessions, a seminar was given to the forecasters, with eight forecasters in attendance. In this seminar, a research meteorologist explained and demonstrated all the features of the system and the science behind the algorithms. In addition, the ideas on how to use the algorithms for warning decision making or verification were also explained. Two visualization workstations were installed in the National Weather Service Forecast Office in Norman, OK so that the forecasters could use all the system features. The first training session was focused on learning the system in general. We developed a task list for this training that covered all aspects of the system. Each forecaster was asked to operate the system and perform the tasks. The forecasters were encouraged to give feedback or ask questions about the system during this training. This training took about 1 hour for each forecaster.

The second training session, attended by six forecasters, was focused on learning the 4D radar analysis tools. The task list for this training included learning on how to display and analyze the 2D cross-section and 3D box cross-section, as well as use the radar volume browsing. We also developed a questionnaire for this training in order to gain feedback from the forecasters. It took about 90 minutes for each forecaster to complete this training.

3.3. Real-Time Observations

Between May and June 2004, the real-time observations were conducted during ten severe weather events including thunderstorms, tornadoes, flash floods, and hail. These events occurred irregularly. A research meteorologist was available to help forecasters with product interpretation during the event. In this observation, forecasters were not asked to do any task. Only a short questionnaire was given to forecasters who used the visualization workstation shortly after the event. Scharfenberg et al. (2004) explain several severe weather cases during this period from the meteorology perspective.

3.4 Surveys

Surveys in this study were conducted by asking the forecasters to fill out questionnaires. We developed two questionnaires for the second training session and the real-time observation. After the real-time observation period was ended, each forecaster was given another questionnaire. This questionnaire covered overall aspects including the usability of the system. The questions for this questionnaire were adapted from Warning Decision Support System - Integrated Information Proof of Concept Test survey questions (Stumpf 2003) with some additional questions. Perceived Usefulness and Ease of Use scales developed by Davis (1999) were also included in this questionnaire.

4. Results

4.1 Training Sessions Results

During the first training session, the trainer helped the forecasters to perform the tasks and answered the questions. The forecasters were impressed with the responsiveness of the system, continuous panning and zooming, display control panel, and products of the algorithm component in the system. They also provided suggestions related to the mouse control or “knobology”, windows layout and manipulation, data readout, color scheme, 3D navigation, and CAPPI concept.

From the second training session, we collected feedback from the answers on the questionnaire. Most forecasters agreed that the continuous panning and zooming is desired over stepwise re-centering and zooming on AWIPS (Advanced Weather Interactive Processing System). AWIPS is the operational system currently used in the National Weather Service Forecast Office in Norman, OK. The “knobology” of the WDSS-II was considered easy to use but they found that it was different from the one on AWIPS. Some forecasters suggested that the WDSS-II adapts the mouse functions closer to what AWIPS uses. For the 4D radar analysis tools, the forecasters felt that they need more practice to navigate in the 3D mode. On the other hand, creating the 2D cross-section and 3D box cross-section were considered easy for most forecasters. The radar volume browsing concept was also rated easy to use. However, comments or feedback for the usefulness of the 4D radar analysis tools concept were varied. Some forecasters mentioned that they need more exposure during warning operations in order to determine the usefulness of the 4D radar analysis tools for warning decision making. Other forecasters felt that the 4D radar analysis tools concept can be helpful for warning decision making.

4.2 Real-Time Observations Results

During real-time observations, we focused on the utilization of the algorithm products for warning operations. Most forecasters used multiple-radar products such as MESH, POSH, and Reflectivity at -20C to help their warning decisions. Other products in the system were also used by some forecasters during their warning and verification operations. Survey results indicate that they typically did not find any difficulties interpreting and understand the products. Most forecasters agreed the ability to simultaneously view multiple-radars' data was useful for warning decision making. They also mentioned that non-operational sources such as TDWR and KOUN data in the WDSS-II were very useful during warning operations. The Area Weather Update-Warning Decision Update from

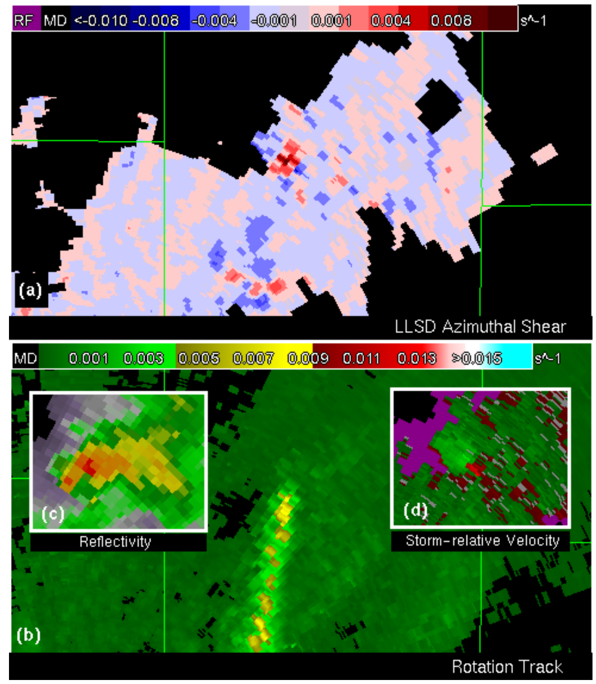

Figure 1: Insets (c) and (d) show conventional displays of 0.5° reflectivity and velocity data from the KTLX WSR-88D from a “mini supercell” observed on June 9, 2004. The image in (a) is a depiction of shear within the storm that is computed from the velocity data shown in (d). The darker red spots indicate locations with high shear. A slow northward movement of the high-shear areas with time is depicted as “rotation tracks” in (b). The graphic in (b) summarizes hours of velocity data into information a warning decision maker can immediately use. (Scharfenberg et al. 2004).

the Norman NWSFO reported the use of the WDSS-II several times during the experiment, verifying that WDSS-II products were utilized during warning operations.

Figure 1 shows an example of the output from the Linear Least Squares Derivatives (LLSD) of radial velocity data that were used by the forecasters for warning and verification of strong circulations. At 2257 UTC on 9 June 2004, a “mini-supercell” thunderstorm was producing a tornado near Chandler, Oklahoma. This particular event occurred in a very tropical climate, where problematic weak, short-lived tornadoes are the norm. The LLSD azimuthal shear product estimated nearly 0.01 s-1 shear in the mesocyclone about at about 700 m MSL (Fig. 1a). Meanwhile, the rotation track map (Fig. 1b) indicated a recent upward trend in mesocyclone intensity and a general north-northeastward motion.

When viewed in real-time, the rotation track maps can provide forecasters a quick summary of the changes in track and intensity of low-level mesocyclones, without the need to step back through multiple volume scans of velocity data over multiple elevation angles. At the end of a tornadic event, these maps are also frequently useful for storm damage surveys. WFO Norman used the rotation track maps to help identify several tornadoes from one supercell thunderstorm on 29-30 May 2004.

4.3 Follow-up Questionnaire Results

The forecasters filled out this questionnaire based on their experience using the system during real-time operations. They were asked to rate the aspects of the system using a rating scale from 1 (the "worst" or least agreement) to 5 (the "best" or strongest agreement) with 3 being a neutral choice, and provide some comments regarding the system. They were also allowed not to answer all questions since they might not have enough exposures with some products.

Overall system including the response time, reliability, and overall features of the system were rated positively since the mean scores for all of these aspects were 4, 3.75, and 3.5 respectively. They also agreed that non-operational sources such as TDWR and KOUN data are useful to help forecasters during warning operations. The mean scores of the usefulness of the system's capability to display merged data from multiple radars, the importance of the multi-sensor concepts to cut down on data overload, the overall skill of multiple-radar products, and the usefulness of the multiple-radar products as warning guidance tools also indicated positive ratings. Most forecasters agreed that the ability to simultaneously view several radars' data was useful. A forecaster said that multiple-radar information is a must for AWIPS implementation.

The skill and usefulness of all multiple-radar Gridded Fields products were rated positively except for the VIL products. The VIL product was rarely used during warning operations since most forecasters preferred the multiple-radar MESH, POSH, and Reflectivity at -20C to diagnose hail potential and severity. Four out of 7 forecasters felt "easy" or "very easy" to interpret and understand the multiple-radar Gridded Fields products, whereas 3 of them had neutral opinions, with a mean score of 3.71. Most forecasters rated positively the importance of the use of "virtual volume" and the 60 second rapid update capability of the multiple-radar Gridded Fields products with the mean scores of 4.2 and 4 respectively. The importance of the ability to integrate data from multiple radars with the Gridded Fields products was rated very high with a mean score of 4.43. The multiple-radar Gridded Fields products have the ability to diagnose storm and hail signatures within poorly sampled regions of single radars which was rated positively with a mean score of 4.17. In summary, 5 out of 6 forecasters agreed their warnings were improved by using these Gridded Fields products.

Only a few forecasters used the multiple-radar SCIT/HDA Cell Icons and Cell Table products in their warning operations. A forecaster rated 5 ("best") for the usefulness of the multiple-radar SCIT/HDA cell icon. On the other hand, two forecasters gave 3 and 5 for the usefulness of the multiple-radar SCIT/HDA cell table. The mean score on how easy or difficult to interpret and understand the multiple-radar SCIT/HDA Cell Icons and Cell Table products (1 is very difficult; 5 is very easy) was 2.5 based on the feedback from 2 forecasters.

The skill and usefulness of all single-radar LLSD products (Azimuthal Shear, Divergence, and Rotation Tracks) were rated positively. However, the mean score on how easy or difficult to interpret and understand the LLSD products was 3 (neutral). The usefulness of the color tables provided for single-radar products were rated positively with a mean score of 3.75.

For the display concept, the usefulness of the continuous panning and zooming and the radar volume browsing were rated positively with the mean scores of 3.71 and 3.6 respectively. The mean score on how easy or difficult to select and view sources and products was 2.5. This low score can be explained that the forecasters were not familiar with the display layout of the system since it is different from their operational system (AWIPS). Forecasters were generally pleased with the 4D base data analysis tool.

For the perceived usefulness and ease of use questionnaire, we use a rating scale from 1 (the “worst” or least agreement) to 7 (the “best” or strongest agreement) with 4 being a neutral choice. Eight forecasters gave their feedback for this questionnaire. The overall mean score of the perceived usefulness of the system was above the neutral scale (4.63). On the other hand, the overall mean score of the perceived ease of use of the system was below the neutral scale (3.60). This result can be explained that some forecasters felt that the knobology of the system is different from the one they use in their operational system (AWIPS). One forecaster mentioned that it was difficult to navigate between two workstations (AWIPS & WDSS-II) that have different knobology in real time operations. Different display layout compared to AWIPS and limited experience with the system during real-time operations were also reasons for this low score.

5. Conclusions

During the development phase of meteorological warning applications, it is very important to gather feedback from those who will eventually be using the applications as part of an operational system. NSSL has worked extensively with the NWS forecasters to develop scientifically sound products that add value to the warning decision-making process. The long-term effects of these interactions will be to reduce false alarms and missed events, and to increase warning lead times. Specifically, there are several feedback items from forecasters that indicate the potential for an improved level of service should they be integrated into NWS operations:

- Multi-sensor applications provide information that is not available from a single source, and fill in data voids that may not be apparent when evaluating a severe storm with a single radar.
- Applications that provide information about the spatial extent of severe weather (primarily tornadoes and hail in this experiment) may provide the key to reducing the area of perceived false alarms.
- Spatial data from the Rotation Track and Hail Track applications developed at NSSL provide an extremely useful verification tool, allowing forecasters to pinpoint areas to focus limited resources when conducting verification phone calls and damage surveys.
- Four-dimensional analysis tools for base data and base data derivatives provide forecasters important new tools for analyzing severe storms. Forecaster feedback during this experiment has resulted in the rapid development of a new, dynamically-updating vertical cross-section tool in the WDSS-II display.

We have learned several important lessons about gaining feedback from forecasters during this and previous proof-of-concept tests of WDSS-II and the original Warning Decision Support System. First, forecasters do not typically like to learn how to use an entirely new demonstration system that sits alongside a well-understood operational system. In the case of WDSS-II, NSSL is working with MDL to incorporate WDSS-II 3D/4D analysis tools into the operational AWIPS system. Additionally, multi-sensor guidance tools will see more use and undergo more thorough evaluation by forecasters when those techniques are integrated into AWIPS. Secondly, 3D/4D analysis tools must be integrated into the NWS warning decision-making program before those tools will become generally accepted by forecasters. Most forecasters have not had experience using those sorts of analysis tools, and have typically relied on a model of a storm’s structure developed by mentally integrating data from multiple radar elevation scans.

The interactions between operational forecasters and application developers are crucial in the development of severe storm analysis tools. In addition to their direct application in the severe weather warning operations, these tools provide information that is useful for new climatology studies as well as providing information to agricultural, insurance, and other interests. The NSSL / NWS partnership encourages the continued development of ideas that improve service to the public and the nation’s economy.

6. Acknowledgements

Funding for this research was provided under the National Science Foundation Grant 0205628. The authors would like to thank David Andra, Michael Foster, and all operational forecasters from the National Weather Service Forecast Office in Norman, OK, who participated in this study. We also thank the CIMMS and NSSL staff who develop and support the WDSS-II, including Jeffrey Brogden, Karen Cooper, Kurt Hondl, Charles Kerr, Valliappa Lakshmanan, Kevin Manross, Lulin Song, Greg Stumpf (CIMMS/NWS/MDL), and Robert Toomey. This manuscript was prepared under CIMMS award NA17RJ1227 from NOAA, U.S. Department of Commerce. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of NOAA or the U.S. Department of Commerce.

References

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*, 319-340.

Ebling, M. R. and B. E., 2000: On the contributions of different empirical data in usability testing. *Proceedings of the conference on Designing Interactive Systems: Processes, Practices, Methods, and techniques*, New York, 289-296.

Hondl, K. D., 2002: Current and planned activities for the Warning Decision Support System – Integrated Information (WDSS-II). *Preprints, 21st Conference on Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., 146-148.

Lakshmanan, V., 2002: WDSS-II: An extensible, multisource meteorological algorithm development interface. *Preprints, 21st Conf. on Severe Local Storms,* San Antonio, TX, Amer. Meteor. Soc., 134-137.

Lakshmanan, V., 2003: Real-time merging of multi-source data. *Preprints, 19th Conf. on IIPS*, Amer. Meteo. Soc., Long Beach, CA.

Lynn, R. J., and V. Lakshmanan, 2002: Virtual radar volumes: creation, algorithm access, and visualization. *Preprints, 21st Conf. on Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., 229-232.

Rubin, J. 1994: *Handbook of Usability Testing*. John Wiley & Sons, New York.

Scharfenberg, K. A., D. J. Miller, D. L. Andra, Jr., and M. J. Foster, 2004: Overview of spring WDSS-II demonstration at WFO Norman. *Preprints, 22nd Conf. on Severe Local Storms*, Hyannis, MA, Amer. Meteor. Soc., CD Preprints.

Smith, T. M., and K. L. Elmore, 2004: The use of radial velocity derivatives to diagnose rotation and divergence. *22nd Conf. on Severe Local Storms*, Hyannis, MA, Amer. Meteor. Soc., CD Preprints.

Smith, T.M. and G.J. Stumpf, 2005: Multi-sensor storm cell identification and analysis. *Preprints, 21st Conf. on IIPS*, San Diego, CA, Amer. Meteor. Soc., CD-ROM.

Stumpf, G.J., S.B. Smith and K. Kelleher, 2005: Collaborative activities of the NWS MDL and NSSL to improve and develop new multiple-sensor severe weather warning guidance applications. *Preprints, 21st Conf. on IIPS*, San Diego, CA, Amer. Meteor. Soc., CD-ROM.

Stumpf, G. J., T. M. Smith, K. L. Manross, and A. E. Gerard, 2003: Warning Decision Support System - Integrated Information (WDSS-II). Part II: Real-time test at Jackson Mississippi NWSFO. *Preprints, 19th Intl. Conf. on Interactive Information and Processing Systems (IIPS) for Meteor., Oceanography, and Hydrology*, Long Beach, CA, Amer. Meteor. Soc., CD preprints.

Stumpf, G. J., T. M. Smith, V. Lakshmanan, K. L. Manross, and K. D. Hondl, 2004a: Status of multiple-sensor severe weather application development at NSSL. *Preprints, 20th Intl. Conf. on Interactive Information Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology,* Seattle, WA, Amer. Meteor. Soc., CD-ROM, 12.3a.

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