6.2 **THE PHASED ARRAY RADAR INNOVATIVE SENSING EXPERIMENT 2013**

*Katie Bowden, Pamela Heinselman, Darrel Kingfield, and Rick Thomas

1. INTRODUCTION

National Weather Service (NWS) forecasters currently utilize the WSR-88D network during warning operations. However, with the WSR-88D nearing the end of its projected 20-year lifecycle, Phased Array Radar (PAR) is being considered as a potential future replacement technology (Zrnić et al. 2007). With electronic beam steering capabilities, the PAR is capable of obtaining volumetric updates in less than 1-min (Heinselman and Torres 2011). This is considerably faster than the current volumetric updates provided every 4-6 min by the WSR-88D, and may prove beneficial to forecasters who struggle to observe rapidly evolving weather (LaDue et al. 2010).

The Phased Array Radar Innovative Sensing Experiment (PARISE) addresses the question of the potential impacts of rapid radar data on the warning decision process of NWS forecasters. Two earlier PARISEs took place in 2010 and 2012 (Heinselman et al. 2012; Heinselman et al. 2013) and explored the use of PAR data during low-end EF0 and EF1 tornado events. Both experiments reported promising results, with a key finding being improved tornado lead time during the use of rapid radar data in simulated real time.

A common question, however, has been whether the benefits of rapid radar data observed during tornadic events are also apparent during other types of severe weather. To expand on the previous work of PARISE, the 2013 experiment switched the focus to investigate the impact of higher-temporal resolution radar data during severe hail and wind events.

2. METHODOLOGY

2.1 Experimental Design

PARISE 2013 recruited a total of twelve NWS forecasters from 2 WFOs situated in the official southern and eastern NWS regions. The experiment took place over six weeks during the summer of 2013, with one participant from both offices visiting each week. The experiment followed a two-independent group design incorporating matched random assignment. The independent variable was volumetric update time, with the control group receiving temporally degraded PAR data simulating 5–min updates, and the experiment group receiving full-PAR data with 1–min updates. The participants' responses to a pre-experiment survey allowed matching the groups on several important individual difference variables (i.e., experience and knowledge) via a matched random assignment procedure.

2.2 Case Studies

Although participants worked three cases during PARISE 2013, the results presented in this paper focus on data collected from the first two. Case selection was based on criteria that required sufficient longevity, continuity, and coverage of storms collected by the PAR. Case 1 presented a marginally severe hail event that took place on 20 April 2010. Case 2 presented a downburst event that occurred on 16 July 2009 and produced both severe hail and wind. Case duration was 35-min and 44-min, respectively. Storm reports were obtained from the official NWS verification database, Storm Data (https://verification.nws.noaa.gov/).

	Time Period (UTC)	Number of Elevations	Vertical Range Coverage
20 April 2012	0134 – 0210	19	0.51° – 52.90°
16 July 2009	2050 – 2053	14	0.51° – 15.50°
	2054 – 2138	14	0.51° – 38.80°

Table 1. Scanning strategy characteristics.

The scanning strategies used during cases 1 and 2 are described in Table 1. For case 1, the PAR collected data using an enhanced volume coverage pattern (VCP) 12 strategy comprised of 19 elevation angles ranging from 0.51° to 52.90°. Case 2 used two

* *Corresponding author address*: Katie Bowden, OU School of Meteorology/ OU CIMMS, 120 David L. Boren Blvd., Norman, OK, 73072; email: Katie.bowden@noaa.gov

VCP strategies to collect data. The first three volume scans (2050–2052 UTC) comprised 14 elevation angles ranging from 0.51° and 15.50°. With the storms approaching the PAR, the vertical coverage of the VCP increased at the expense of low-level dense sampling. From 2053 UTC through the end of the case, the PAR scanned 14 elevation angles ranging between 0.51° and 38.80°.

2.3 Working the Cases

Prior to working the case studies, Darrel Kingfield led a 30–min session for familiarization of the Advanced Weather Interactive Processing System-2 (AWIPS-2). Currently, WFOs across the U.S. use AWIPS-1 to synthesize weather data, provide forecasts, and issue warnings in an efficient manner. AWIPS-2 is planned to replace AWIPS-1 in the near future, and though it will offer some new functionalities, the base features of AWIPS-1 will still be available. Throughout the experiment, AWIPS-2 was utilized to play back cases in displaced real time, where participants were provided base velocity, reflectivity, and spectrum width products. While familiarizing themselves with AWIPS-2, participants practiced navigating the system and loading available products, as well as issuing warnings using the operational NWS warning tool Warning Generation (WarnGen). Personal procedures were loaded to the designated computers of each participant to further enhance familiarity within their work environment. Participants did not move on to the case study phase until they were comfortable using AWIPS-2, though this typically did not take very long. A benefit of participants being at ease with using AWIPS-2 as their primary forecasting tool is the minimization of software distraction, allowing for their time and effort to be focused on the job at hand.

Following the AWIPS training session, participants were assigned to separate rooms and were accompanied by one researcher. Each case was worked in simulated real time, where participants were asked to perform their forecast roles as they would under normal warning operations. Similar to PARISE 2012, participants watched a video briefing prior to working each case. Provided by Jim Ladue of the Weather Decision Training Branch (WDTB), briefings presented meteorological information that enabled participants to form expectations and familiarize themselves with the environment that they would be working with.

2.4 Case Walk–Through

Assessing the impacts of higher-temporal resolution radar data on the warning decision process can be accomplished using quantitative methods (e.g. lead time and verification statistics), but these results alone don't tell the story behind why we observe what we do. Gaining insight into the *why* and *how* of decisions requires the use of qualitative data techniques specifically designed to probe cognitive activity (e.g., subjective evaluations). PARISE 2013 dedicated a portion of the experiment to collecting this information.

Ultimately, information regarding the goals, knowledge, and thoughts of a participant whom was engaged in the cases presented during PARISE 2013 was sought through a cognitive task analysis (CTA) approach. Protocols for CTA (Hoffman 2005) provides a template as guidance to CTA methodology, which was appropriately modified to meet the goals and questions of PARISE 2013. Hoffman (2005) describes a sequence of three sweeps that the participant and researcher work through together. The first sweep asks participants to walk through the recent task, recalling what they did step-by-step while reviewing a playback video of their on-screen activity during the case. Sweep two involves a revision of the timeline. The purpose of this sweep is to provide the participant with an opportunity to correct or add information. During the third and final sweep, the researcher plays a more active role in eliciting information from the subject. This sweep is referred to as "deepening" and utilizes a set of probing questions to target research goals and questions.

4. VERIFICATION

4.1 Compound Warning Decision Model

While observing participants, it quickly became apparent that the warning decision process was not a one-step procedure but was rather a compound decision. Comprised of multiple elements, the compound warning decision model describes three important parts: detection, identification, and re-identification. First, participants will *detect* the potential of severe weather in some region of the area they are monitoring. Upon this decision, the WarnGen software and an appropriate warning is prepared. At this time, participants *identify* the weather threats that are associated with the region that is being warned. For example, a severe thunderstorm warning (SVR) requires the identification of hail and/or wind. Once the threat expectation has been chosen, participants issue the warning. Through reassessment of prior radar data and interrogation of new incoming radar data, participants are able to track the evolution of a storm and *re-identify* the weather threat. Re-identification may involve the maintenance of a threat, a change in magnitude of threat, or a change in threat type expected. Through the issuance of a severe weather statement (SVS), participants are able to communicate updated information to the public.

4.2 POD and FAR Scores

The probability of detection (POD) and false alarm ratio (FAR) scores were calculated for all three levels of the compound warning decision model for all warning decisions made during cases 1 and 2. For detection, verification was based on whether a severe event occurred within a warning both spatially and temporally. Severe events that were not included in warnings were considered a miss. For identification and re-identification, individual weather threats (e.g. hail and wind for a SVR warning) were verified and then combined to compute overall POD and FAR scores.

To obtain a group comparison, the mean POD and FAR scores were calculated for both control and experiment groups by averaging the scores of participants in the same group. The results show that, in both cases 1 and 2, the experiment group achieved superior POD and FAR scores to the control group for all three levels of the compound warning decision model with the exception of two instances (Table 2). These instances include 1) equal POD scores for identification in case 2 and 2) a lower re-identification POD score for the experiment group.

4.3 LEAD TIME

Storm reports were treated as instantaneous events. Lead time was calculated as the event time minus the time of the warning issued. A lead time of zero was assigned to events that were either not warned on or warned on after the event occurred. Case 1 included one event, and case 2 included three events. To compare between groups, the mean lead time was calculated as the average lead time across all participants within the same group.

Individual lead time results (Figs. 1 and 2) show a general shift towards longer lead times for participants within the experiment group. This is reflected in the group mean lead times, whereby the control group achieved a mean lead time of 16.4-min in both cases 1 and 2, and the experiment group achieved a mean lead time of 22.0-min in case 1 and 21.8-min in case 2. Combining the lead time results from both cases, we find that the experiment group exceeded the mean lead time of the control group by 5.5-min.

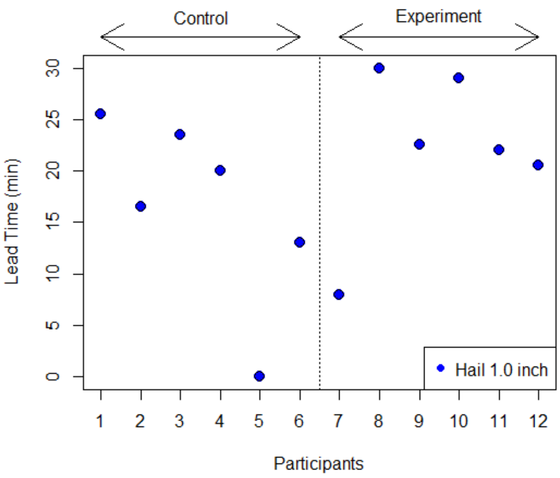

Figure 1. Participant lead time in case 1.

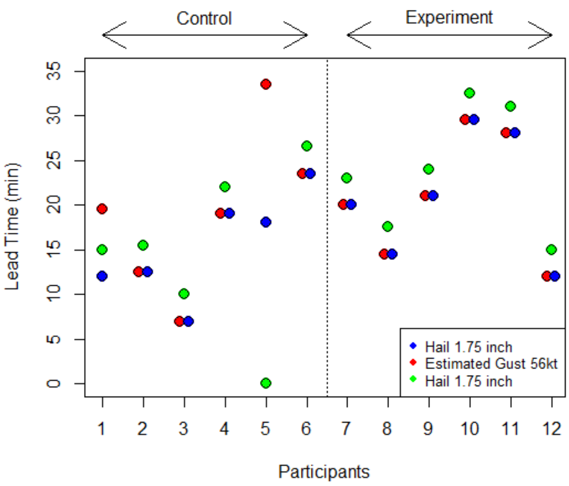

Figure 2. Participant lead time in case 2.

5. DECISION TYPES

It is a common assumption that with more radar data, forecasters will be more confident and will therefore make better decisions. To investigate this, the relationship between confidence and correctness was assessed by employing the Confidence-Based Assessment (CBA) behavioral model (Adams and

Ewen 2004). A professor of education, Dr. Bruno, developed this model to assess the types of decisions students were making during multiple choice tests with the goal of understanding the knowledge base behind their choices. By instructing students to assign a level of confidence to answers, the decision can be classified into one of four categories (Fig. 3).

Uninformed decisions are both incorrect and made without confidence, whereby the decision maker is aware that they do not have a suitable knowledge base to make an informed decision. Doubtful decisions are correct decisions that are made hesitantly due to a lack in confidence. Misinformed decisions are perhaps the most risky decisions, whereby the decision maker is confident but incorrect in their knowledge. The best types of decisions are mastery decisions. Both confident and informed, these smart decisions are the most desirable.

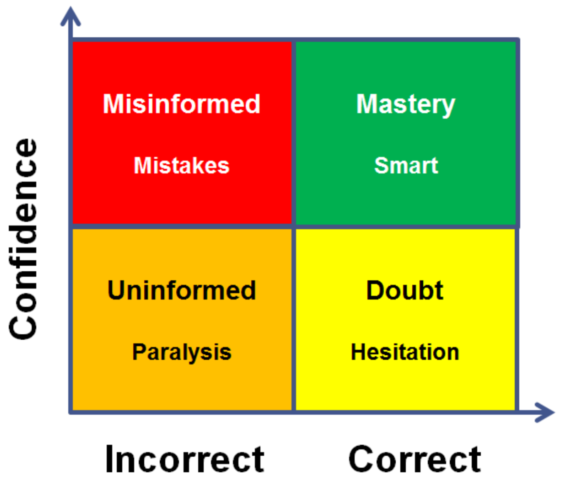

Figure 3. Decision type categories based on the relationship between confidence and correctness. Adapted from Adams and Ewen (2004).

The key decisions in the experiment group (N=54) and control group (N=53) were assessed across both cases. Key decisions were defined as SVRs or SVSs, and were considered correct for hits/correct rejections of active/absent severe weather. Participants assessed their confidence on a scale ranging from not sure (0%), to partially sure (50%), to sure (100%). Given participants tended to have different baselines for judging self-confidence, the results were normalized onto a scale of 1-7. A value of ≥ 5 was considered confident since this rating corresponded closer to sure than partially sure. Each key decision was assessed in terms of confidence and correctness and was classified accordingly. Overall, the experiment group made better decisions compared to the control group (Fig. 4), with more mastery decisions and consequently less doubtful, uninformed, and - most importantly - misinformed decisions being made. This finding demonstrates that in these cases, the use of rapid data did result in better decisions being made.

6. Conclusions

The 2013 experiment broadened the focus of PARISE by considering severe weather other than tornadoes. The use of higher-temporal resolution radar data was found to positively impact the warning decision process of NWS forecasters during severe hail and wind events. The compound warning decision process was used to assess verification on a more intricate scale. Group mean POD and FAR scores were superior for the experiment group in almost all instances of detection, identification, and re-identification. Lead time was also benefited, with the experiment group providing an additional mean lead time of 5.5-min compared to the control group. Furthermore, assessment of decision confidence and correctness found that rapid data led to better decision making. With more mastery decision, the experiment group made less undesirable decisions than the control group.

There are still plenty of questions that remain and will form a basis for future research. In particular, investigating whether there is an ideal update time is important since participants have only been exposed to either 1-min or the traditional 4–6-min volumetric updates during PARISE. Indeed, it may be that there is not one ideal temporal resolution if forecasters respond differently to rapid data. A within subjects comparison would also be useful for eliminating individual differences that are difficult to control for. Additionally, from a quantitative aspect, a larger sample needs to be obtained for assessment of statistical significance.

Acknowledgments. We thank all twelve participating forecasters for your enthusiasm and focus during this experiment. Support from participating WFO MICs, Keith Stellman (PTC) and Darin Figurskey (RAH), was much appreciated. Additionally, we thank the following colleagues for contributions to this experiment. Norman NWS forecaster Michael Scotten; audio and video specialist James Murnan; GIS expert Ami Arthur; software expert Eddie Forren. Lastly, thanks to Lans Rothfusz and Harold Brooks *for advice and guidance during the design and preparation of this experiment.*

REFERENCES

Adams, T. M. and G. W. Ewen, 2009: The importance of confidence in improving educational outcomes. *35th Annual Conf. Distance Teaching and Learning*, Madison, WI, University of Wisconsin-Madison, 5 pp.

Heinselman, P. L., and S. M. Torres, 2011: High-temporal resolution capabilities of the National Weather Radar Testbed phased-array radar. *J. Appl. Meteor. Climatol.,* **50**, 579–593.

Heinselman, P. L., D. S. LaDue, and H. Lazrus, 2012: Exploring impacts of rapid-scan radar data on NWS warning decisions, *Wea. Forecasting*, **27**, 1031–1044.

Heinselman, P. L., D. S. LaDue, 2013: Supercell storm evolution observed by forecasters using PAR data. Extended Abstract, *36th Conf. on Radar Meteorology*, Breckenridge, CO, Amer. Meteor. Soc., 3B.4.

Hoffman, R. R., 2005: Protocols for Cognitive Task Analysis. DTIC No. ADA475456, 108 pp.

LaDue, D. S., P. L. Heinselman, and J. F. Newman, 2010: Strengths and limitations of current radar systems for two stakeholder groups in the southern plains. *Bull. Amer. Meteor. Soc.,* **91**, 899–910.

Zrnić, D. S., and Coauthors, 2007: Agile beam phased array radar for weather observations. *Bull. Amer. Meteor. Soc.,* **88**, 1753–1766.

		Case 1		Case 2	
		Mean Control	Mean Experiment	Mean Control	Mean Experiment
Detection	POD	0.83	1.00	0.95	1.00
	FAR	0.58	0.45	0.33	0.25
Identification (Overall Threat)	POD	0.83	1.00	0.88	0.88
	FAR	0.79	0.70	0.36	0.22
Re-Identification (Overall Threat)	POD	0.60	0.83	1.00	0.90
	FAR	0.87	0.70	0.23	0.19

Table 2. Mean POD and FAR scores for control and experiment groups. Statistics are calculated for both cases for all three levels of the compound warning decision process.

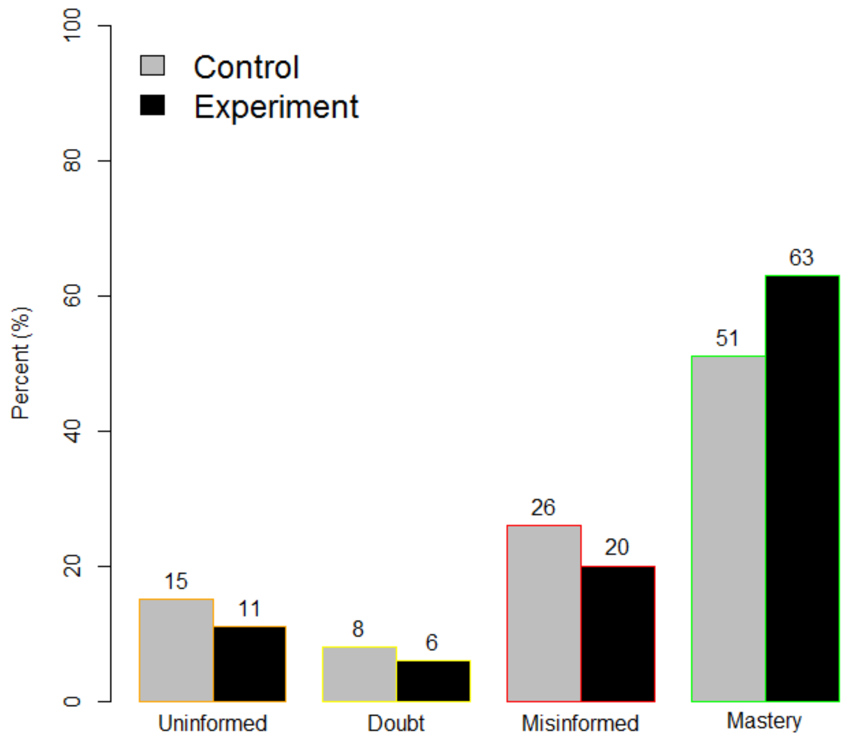

Figure 4. Decision types from both cases 1 and 2 for the control and experiment groups. Colors correspond to Fig. 3, such that orange, yellow, red, and green represent uninformed, doubt, misinformed, and mastery decisions, respectively.