

A Qualitative Analysis of NWS Forecasters' Use of Phased-Array Radar Data during Severe Hail and Wind Events

KATIE A. BOWDEN

Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma

PAMELA L. HEINSELMAN

NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

(Manuscript received 15 July 2015, in final form 5 November 2015)

ABSTRACT

The 2013 Phased Array Radar Innovative Sensing Experiment (PARISE) investigated the impacts of higher-temporal-resolution radar data on National Weather Service forecasters' warning decision processes during severe hail and wind events. In total, 12 forecasters participated in the 2013 PARISE over a 6-week period during the summer of 2013. Participants were assigned to either a control [5-min phased-array radar (PAR) updates] or experimental (1-min PAR updates) group, and worked two cases in simulated real time. This paper focuses on the qualitative retrospective reports of participants' warning decision processes that were collected using the recent case walk-through method. Timelines of participants' warning decision process were created for both cases, which were then thematically coded according to a situational awareness framework. Coded themes included perception, comprehension, and projection. It was found that the experimental group perceived significantly more information during both cases than the control group (case 1 $p = 0.045$ and case 2 $p = 0.041$), which may have improved the quality of their comprehensions and projections. Analysis of timelines reveals that 1-min PAR updates were important to the experimental group's more timely and accurate warning decisions. Not only did the 1-min PAR updates enable experimental participants to perceive precursor signatures earlier than control participants, but through monitoring trends in radar data, the experimental group was able to better detect storm motion, more accurately identify expected weather threats from severe thunderstorms, more easily observe strengthening and diminishing trends in storms, and make more correct tornado-related warning decisions.

1. Introduction

National Weather Service (NWS) forecasters issue severe thunderstorm, tornado, and flash flood warnings to communicate the occurrence of imminent or active weather events to the public. The NWS verifies warnings against observations using performance metrics such as the probability of detection, false alarm ratio, and lead time. While these metrics provide information regarding the accuracy of warning decisions, they do not offer insight into why a warning decision was made. Attaining this insight aids forecasters' ability to develop expertise and is essential to the development of best practices. For this reason, the NWS Warning

Decision Training Branch trains and encourages forecasters to use root-cause analysis to understand the reasons behind their warning decisions (Lindley and Morgan 2004; Quetone et al. 2009).

To gain insight into the NWS warning decision process, Hoium et al. (1997) produced detailed timelines of the information forecasters at the Raleigh, North Carolina, Weather Forecast Office (WFO) used during real-time severe events between 1994 and 1996. From these timelines they constructed a schematic representation of the warning decision process, which evolved from monitoring the situation, to interrogating the radar-based structure of individual storm cells, and then deciding whether or not a severe warning was merited. After issuing a warning, the forecasters sought to verify their warnings via ground truth (i.e., observations of or damage caused by severe weather). As discussed by Andra et al. (2002), the warning decision process can be more complex.

Corresponding author address: Katie Bowden, 120 David L. Boren Blvd., Norman, OK 73072.
E-mail: katie.bowden@noaa.gov

During the 3 May 1999 tornado outbreak in central Oklahoma, forecasters' high levels of situational awareness and successful warning operations at the Norman, Oklahoma, WFO resulted from the integration of scientifically based conceptual models, Doppler radar, ground truth, technology, and the implementation of strategy and expertise. Similarly, in their assessment of the reasoning behind four critical warning decisions during a significant hail-producing supercell event, [Lindley and Morgan \(2004\)](#) identified storm spotter reports, knowledge of both the prestorm and during-storm environments, and sectorized warning operations as major contributors to successful tornado-related warning decisions.

Understanding how new meteorological data, products, and technology might impact forecasters is also crucial to ensuring the success of future warning operations. Researchers conducting experiments in both simulated and real-time settings demonstrate efforts to assess these impacts. For example, studies have investigated forecaster utilization of additional coastal meteorological observations during real-time operations ([Morss and Ralph 2007](#)), output from a convective-permitting ensemble forecast during a high precipitation event ([Evans et al. 2014](#)), and real-time numerical model analyses during severe thunderstorm and tornado events ([Calhoun et al. 2014](#)). During these studies, evaluations of elements such as participants' thought processes, experience, or performance while using this new information was achieved by collecting qualitative data in the form of observations, interviews, surveys, and/or real-time blogging.

[Heinselman et al. \(2012, 2015\)](#) also recognized the importance of obtaining qualitative data during the 2010 and 2012 Phased Array Radar Innovative Sensing Experiments (PARISE). PARISE investigates the impacts of higher-temporal-resolution radar data provided by phased-array radar (PAR) on the warning decision process of NWS forecasters. The work of PARISE is important since PAR is being considered as a future replacement technology to the current Weather Surveillance Radar-1988 Doppler (WSR-88D) network ([Zrnich et al. 2007](#)). In the 2010 and 2012 PARISE programs, [Heinselman et al. \(2012, 2015\)](#) found that low-end [i.e., storms rated at 0 or 1 on the enhanced Fujita scale (EF0/EF1)] tornado warning lead time increased when forecasters used 1-min PAR updates. Specifically, in the 2012 PARISE, 12 participants achieved a median tornado warning lead time of 20 min using 1-min PAR updates, which was higher than the combined southern and central NWS regional median tornado warning lead time of 11 min for EF0/EF1 tornadoes ([Heinselman et al. 2015](#)). A qualitative analysis of participants' warning decision processes revealed that precursors triggering warning

decisions evolved on time scales more effectively captured by the temporal sampling of PAR than by that of the WSR-88D, which in turn aided earlier warning decisions ([Heinselman et al. 2015](#)).

To further the work of PARISE, the 2013 experiment examined the impacts of 1-min PAR updates on the forecaster warning decision process during severe hail and wind events. While a quantitative analysis of participants' performance was reported in [Bowden et al. \(2015\)](#), this paper discusses results from the qualitative data collected. Similar to root-cause analysis ([Quoetone et al. 2009](#)), the qualitative phase of this experiment was designed to ensure a thorough understanding of each participant's warning decision process and obtain the main causes leading to warning decisions. [Bowden et al. \(2015\)](#) reported that during the 2013 PARISE, participants using 1-min PAR updates performed better than those using 5-min PAR updates, as demonstrated through a statistically significant longer median warning lead time of 21.5 min compared to 17.3 min ($p = 0.0252$), and overall had better probability of detection and false alarm ratio scores. Furthermore, a larger number of mastery (i.e., confident and correct) decisions were made by participants utilizing 1-min PAR updates compared to those using 5-min PAR updates. Therefore, the purpose of this paper is to answer the question of what information 1-min PAR data provided forecasters that the 5-min PAR data did not, and how this additional information was used to make better warning decisions. Similar to [Heinselman et al. \(2015\)](#), data were collected using retrospective cognitive task analysis (see [section 2](#)). Herein, a situational awareness framework (e.g., [Endsley 1995](#)) was applied to the qualitative data to measure any differences in participants' number of perceptions, comprehensions, and projections. A comparison of participants' warning decision processes is given to highlight examples of when the temporal resolution of PAR data directly impacted warning decisions. Experimental participants' postexperiment survey results are also shared to provide an understanding for how 1-min PAR updates impacted their forecasting procedures as a whole.

2. Methods

a. Experiment design

To familiarize the reader with key aspects of the experiment design and case selection, we summarize the detailed description found in [Bowden et al. \(2015\)](#), and then describe the qualitative data collection and analysis methods. In previous PARISE efforts, the influence of office culture on forecasters' warning protocols and philosophies was apparent. To reduce this influence in our

findings and allow for a cleaner and fairer comparison between participants, only two WFOs were selected to participate. Twelve forecasters from these two NWS WFOs participated in the 2013 PARISE over a period of 6 weeks, where each week one forecaster from each office visited Norman, Oklahoma. A two-independent-group design was incorporated, such that participants were assigned to either a control (5-min updates) or experimental (1-min updates) group. Both groups received PAR data, though the control group's data were temporally degraded to be similar to the WSR-88D observations. As described in Bowden et al. (2015), matched random assignment was incorporated into the experiment design to ensure a fair comparison between groups by considering participants' experience and knowledge levels.

Participants worked two cases selected from archived PAR data. Both cases presented severe and nonsevere storms, with four areas of interest in case 1 and two areas of interest in case 2 (Fig. 1). Storms verifying severe thunderstorm warnings must be associated with 50-knot (kt ; $1 \text{ kt} = 0.51 \text{ m s}^{-1}$) or higher wind and/or hail of at least 1-in. diameter, whereas storms verifying a tornado warning must be associated with a tornado that occurs within the spatiotemporal limits of the warning polygon (NOAA 2011). Areas of interest were determined by how participants identified storms in space (e.g., the northern versus southern storm cells). Case 1 contained multicell clusters of storms that produced marginally severe hail during 0134–0210 UTC 20 April 2012 (<https://verification/nws.noaa.gov>). Case 2 also contained multicellular storms, but these produced larger hail and downburst-driven severe winds during 2053–2139 UTC 16 July 2009 (<https://verification/nws.noaa.gov>). Participants viewed a weather briefing video prior to working each case. The weather briefing discussed environmental conditions and displayed satellite and radar imagery prior to the case start time. Once participants finished watching the video, they began working the case in simulated real time. Participants were asked to work each case as if they were in their usual forecast office, and instructed that their goal was to decide whether a warning was warranted for the storms encountered during the case. Participants worked cases using the Advanced Weather Interactive Processing System-2 (AWIPS-2), where they could view base velocity, reflectivity, and spectrum width products, and use the Warning Generation (WarnGen) software. Storm reports were communicated to participants verbally.

b. The recent case walk-through

Understanding why participants performed as they did during each case requires an analysis of the full warning

decision process rather than the warning decisions alone. In the 2010 PARISE, Heinselman et al. (2012) attempted to learn about participants' warning decision processes primarily through observation. However, the inference required to interpret participants' thoughts and actions motivated the 2012 PARISE to use a more objective method that would directly elicit participants' cognitive processes (Heinselman et al. 2015).

A method of retrospective reporting was sought, since unlike concurrent reporting, it does not add the cognitive burden and distraction of verbalizing one's thoughts while working on a task (Ericsson and Simon 1993, 1–62). Although a concern of using this method is the accuracy of retrospectively recalled information, comparisons to eye-tracking data (e.g., Guan et al. 2006) and concurrent reports (e.g., Van Gog et al. 2005) validate the quality of retrospective reporting methods. Retrieving information on cognitive processes shortly after the completion of a task has been found to be especially successful when a form of stimulation is provided [e.g., video cue; Van Gog et al. (2005)]. A form of retrospective reporting that was used in both the 2012 and 2013 PARISE is based on Hoffman's (2005) cognitive task analysis methodology template. This template, referred to as the recent case walk-through (RCW), comprises three sweeps that the researcher and participant work through together shortly after a case has been worked (Fig. 2). As noted in Bowden et al. (2015), during the 2013 PARISE each forecaster worked the event and completed the RCW in separate rooms. In sweep 1 we asked participants to recall freely what they were seeing, thinking, and doing as they reviewed a playback video showing only their on-screen activity from the case. During this first sweep, we typed all verbalizations into a timeline and participants paused the video if they wished to elaborate. In the second sweep, participants reviewed the timeline and added or corrected information as needed. The third sweep is the deepening phase of the RCW. Here, participants were asked semi-structured probing questions designed to target our research goals (see the appendix). Questions focused on why warning decisions were made, what information was used to make such decisions, and how confident the forecaster was in their warning decision. Participants were aware prior to working each case that they would complete the RCW procedure; therefore, it is possible that participants altered their predecision process because of a type of reactivity known as justification bias (Ranyard and Svenson 2011). In an effort to minimize justification bias, participants were reminded that the experiment was not a test, that all information would remain anonymous, and to work in a normal manner. Throughout the reporting of this study, control participants are identified as P1–6 and experimental participants are identified as P7–12.

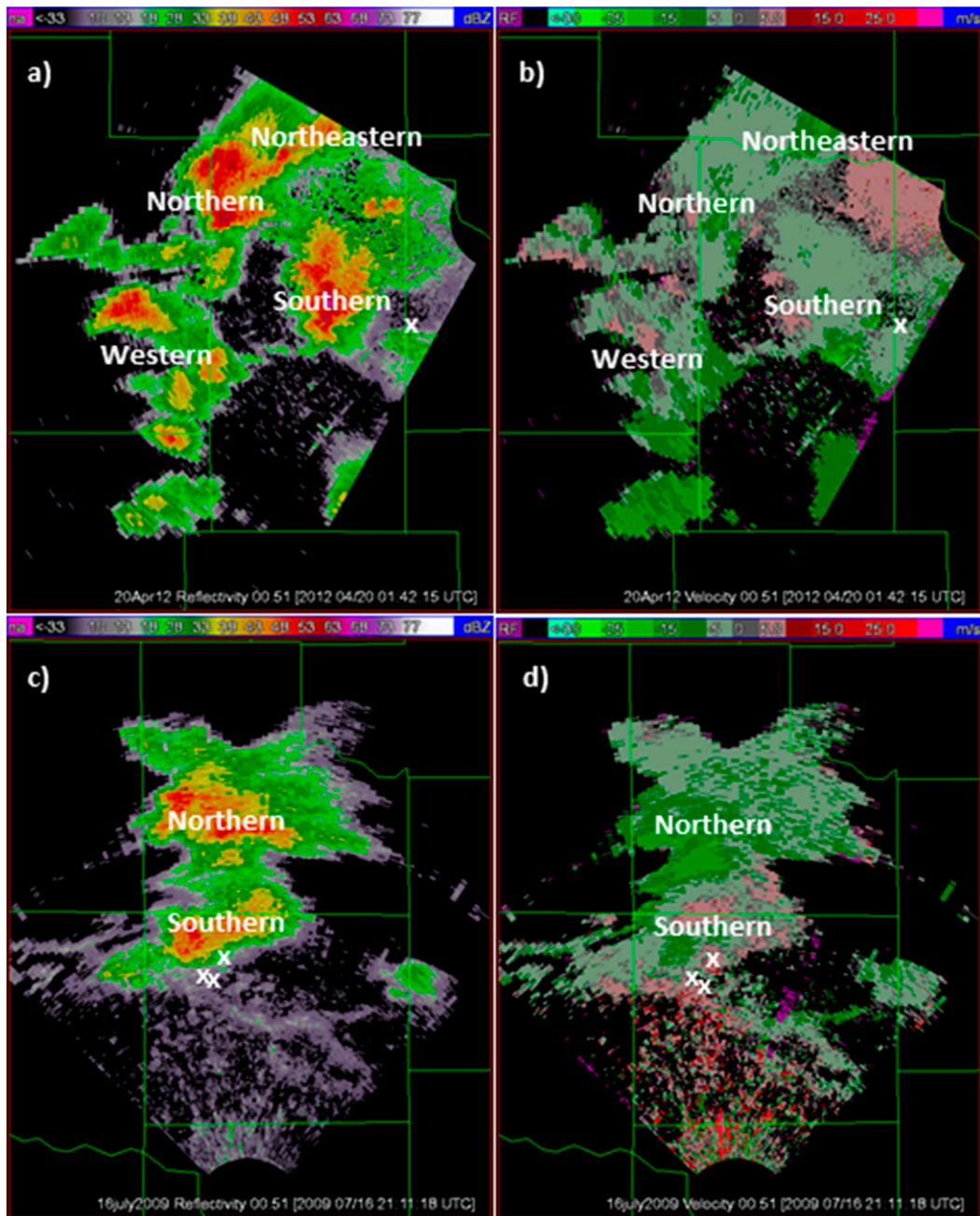


FIG. 1. The 0.51° (a) reflectivity and (b) velocity at 0142 UTC for case 1. Areas of interest are the northern storm, southern storm, western storm cluster, and northeastern storm. The 0.51° (c) reflectivity and (d) velocity at 2111 UTC for case 2. Areas of interest are the northern storm and the southern storm. Radar images shown are prior to when storms produced severe weather. Locations of severe weather reports during case time are indicated by white \times symbols. [Adapted from Bowden et al. (2015).]

3. Situational awareness

To efficiently analyze and understand participants' warning decision processes, a suitable framework was sought, adopted, and applied to the qualitative data (Braun and Clarke 2006). Operational meteorologists

find Endsley's (1995) theoretical framework for situational awareness (SA) during dynamic decision-making well suited to NWS warning operations (e.g., Andra et al. 2002). Endsley (1995) defines this SA framework as "the perception of the elements in the environment within a volume of time and space, the comprehension

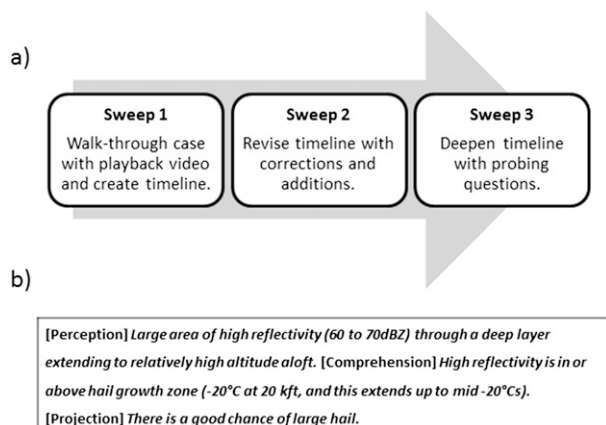


FIG. 2. (a) The recent case walk-through comprises three sweeps designed to elicit rich, insightful, qualitative data regarding one's warning decision process. (b) Situational awareness thematic coding of qualitative data was based on [Endsley's \(1995\)](#) situational awareness framework.

of their meaning, and the projection of their status in the near future." Perception, comprehension, and projection form a hierarchical structure, such that the higher levels are dependent on the success of lower levels. Timelines were thematically coded according to this framework ([Fig. 2b](#)). Thereafter, the numbers of perceptions, comprehensions, and projections made by each participant were counted for each case. Counts of these cognitive actions were summed for the control group and the experimental group and then compared to assess any differences.

The code that was most similar between the two groups was projection, with the experimental group projecting slightly more than the control group only in case 2 ([Figs. 3b and 3d](#)). More instances of comprehension were recalled by the experimental group than the control group in both cases; this comparison is also most notable for case 2 ([Figs. 3b and 3d](#)). The code that was of greatest difference between the two groups was perception ([Figs. 3b and 3d](#)). The Wilcoxon rank sum nonparametric test ([Wilks 2006](#)) was used to test for statistically significant differences between the two group's median counts of perception, comprehension, and projection. Although statistical significance was not established for the difference in median counts of comprehension and projection between the control and experimental groups, statistically significant differences (at the 95% level) were established for the differences between the two group's median counts of perception in both cases ([Table 1](#)).

For both groups, perception was also the code with the greatest spread among participants. One participant in each group (P3 and P10) recalled a relatively low amount of perceptual information compared to other

participants in their same group during both cases ([Figs. 3a and 3c](#)). Regardless of the relatively low number of perceptions, the performance of P10 exceeded that of most other participants, especially P3, in terms of the compound warning decision process POD and FAR statistics, and in terms of the number of mastery decisions made (see [Figs. 4, 5, and 7](#) in [Bowden et al. 2015](#)). Some insight into why these two participants may have recalled fewer perceptions and performed so differently was drawn from the data.

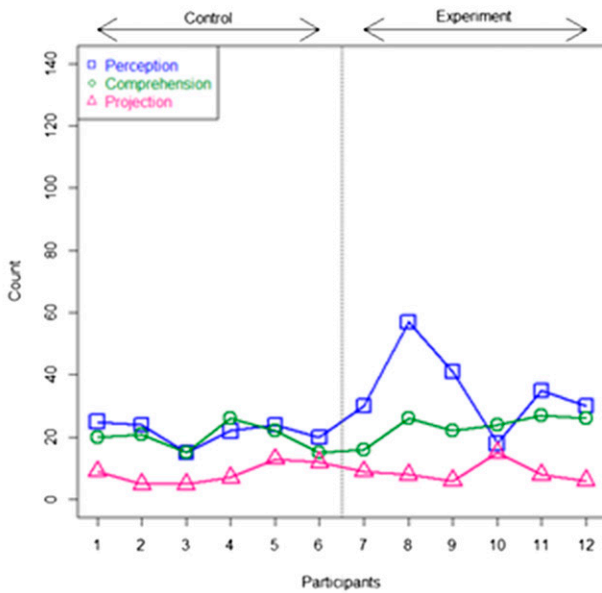
During the RCW, P3 had a difficult time explaining her warning decision process. P3 explained that it was "hard for me to put it into words" since her thought process was "automated." Although automaticity—a subconscious thought process—is advantageous in that it does not require heavy use of cognitive resources, it can be disadvantageous because the consistent mapping rules employed by automatic processing are not always relevant to new and dynamically changing scenarios ([Logan 1988](#)). Evidence of this disadvantage occurred during case 1 when P3 explained that the order of elevation angles used to collect the PAR data was "disorienting" and that P3 had to "get used to the way the data was coming in." It was unexpected for a control participant to feel overwhelmed while using a non-traditional volume coverage pattern strategy that provided 5-min updates. Given this participant's apparent automaticity, it is not surprising that P3 recalled not only the least number of perceptions, but also the least number of comprehensions and projections within the control group ([Figs. 3a and 3c](#)).

In contrast, during the RCW, experimental participant P10 was able to recall his warning decision process with ease. Of the 12 participants, P10 expressed the most value in the use of environmental data in his warning decision process. For example, during case 1, P10 explained that you "can't just use radar data to effectively warn, [you] really need to know [the] storm environment... because expectations are based on environment." Given P10's perspective, it makes sense that his radar data analysis was more focused and contained fewer extraneous details by comparison.

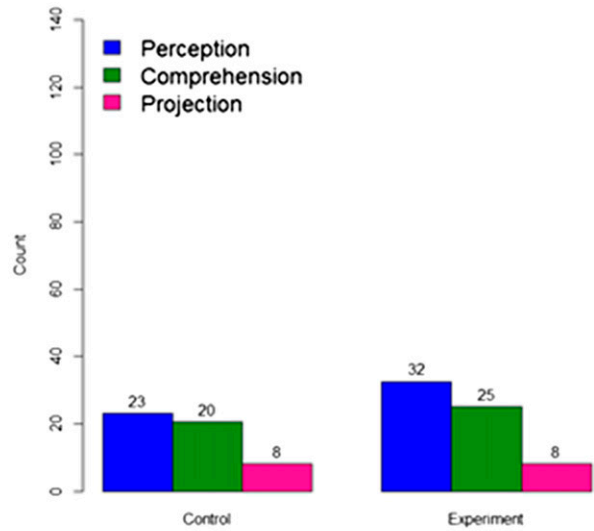
4. Case timelines

Both the situational awareness analysis ([section 3](#)) and the performance analysis of [Bowden et al. \(2015\)](#) found statistically significant differences between data collected on experimental and control participants' decision processes. In addition to the statistically significant higher number of perceptions by experimental participants reported here, [Bowden et al. \(2015\)](#) reported a statistically significant 4.2-min-longer median warning lead time for

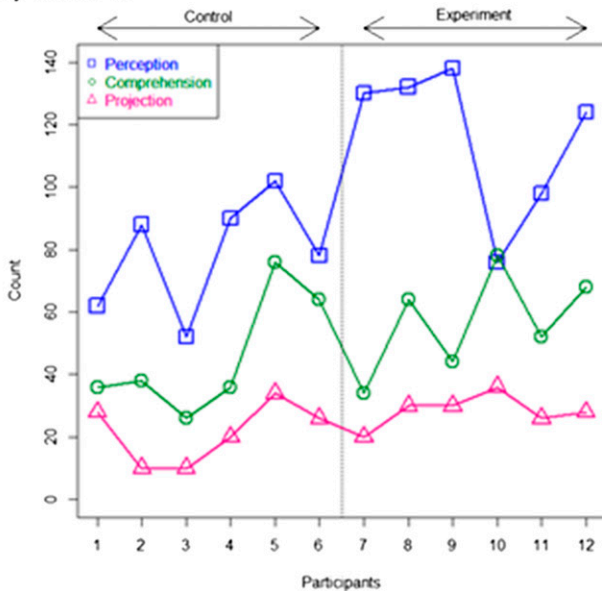
a) Case 1



b) Case 1



c) Case 2



d) Case 2

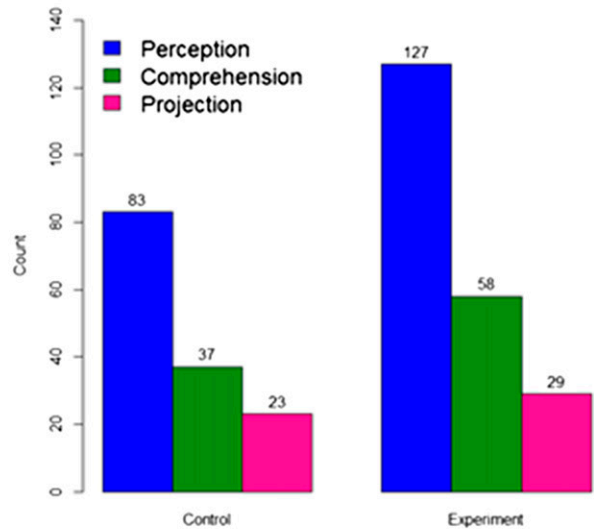


FIG. 3. Individual participant counts for the three levels of situational awareness in (a) case 1 and (c) case 2, and the group median counts for the three levels of situational awareness in (b) case 1 and (d) case 2.

severe weather warnings issued by experimental participants compared to control participants (21.5 vs 17.3 min, respectively). Furthermore, more mastery warning decisions (i.e., correct and confident) were made by experimental participants than by control participants. Together, these findings give us reason to postulate that the use of higher-temporal-resolution PAR data improved the experimental participants' performance. The case timelines provide a means to determine what

(if anything) experimental participants perceived in the rapidly updating PAR data that control participants did not, and how this information aided their warning decision processes compared to control participants.

a. 0134–0210 UTC 20 April 2010 marginal severe hail case

While working the 0134–0210 UTC 20 April 2010 marginal severe hail case, participants' attention focused

TABLE 1. Statistical significance of the difference in the median counts of perception, comprehension, and projection between the control and experimental groups was assessed using the Wilcoxon rank sum nonparametric test. Significant differences (i.e., p value < 0.05) are set in boldface.

	Case 1 P value	Case 2 P value
Perception	0.045	0.041
Comprehension	0.106	0.335
Projection	0.872	0.196

predominantly on the northern and southern storms, and intermittently on the western and northeastern storms (Fig. 1). Most participants' primary concern was the potential of the storms to produce severe-sized hail. All forecasters issued severe warnings on the hail-producing southern storm; all but one forecaster, P11, also issued severe warnings on the nonsevere northern storm (section 2). No warnings were issued on the western or northeastern storms. The detected storm-based precursor that drove the majority of participants' decisions to issue severe warnings was the height and magnitude of the high-reflectivity core above the freezing level, compared to their personal thresholds. Typically, participants warning on the northern storm first sought reflectivity values of approximately 60 dBZ up to 20 kft, whereas participants warning on the southern storm applied more stringent criteria and first sought reflectivity values of approximately 60 dBZ at a higher altitude of 35 kft. The experimental group's overall ability to perceive this precursor earlier enabled their extended median severe-warning lead time on the southern storm compared to the control group (22.3 vs 18.3 min, respectively). Almost all participants were unable to discriminate a difference in hail potential between the northern and southern storms, which is unsurprising given the marginal potential for severe-sized hail. The exception was P11, who correctly determined that the northern storm's reflectivity core was subsevere. P11's successful determination arose from his strict application of a 60 dBZ to 35 kft severe hail criteria.

In addition to extending severe warning lead time, we found that the use of 1-min PAR data improved some experimental participants' decisions during two situations: 1) assessing the southern storm's deviant motion and 2) correctly rejecting a wind threat in a severe thunderstorm warning.

1) DEVIANT MOTION OF THE SOUTHERN STORM

Unlike control participants, experimental participants were able to better detect and correct for deviant motion in the southern storm once it was fully sampled by the PAR at 0143 UTC. The deviant motion was first apparent to P7 at 0151 UTC, when she explained that the southern storm had a "different motion with the full

sector" such that it was moving "east." Earlier in the case, P7 questioned the projected storm track from WarnGen while warning on the southern storm, since she thought the storm would move toward the east rather than the northeast. However, having accepted WarnGen's projected storm motion toward the northeast, P7 now realized that the storm would "move out of [the] polygon soon." Therefore, at 0157 UTC, P7 decided to issue a second warning on the southern storm that now successfully encompassed the upcoming severe hail event. In addition to P7, two other experimental participants also acted upon the southern storm's apparent change in motion. At 0200 UTC, P10 decided to issue a second warning on the southern storm, which better accounted for the storm's motion and encompassed severe hail events that occurred downstream outside of the case time. P8 was the third experimental participant to recognize that the southern storm was tracking toward the east rather than northeast, and acted on this by "fine tun[ing] the direction" of the warning in a severe weather statement during 0200–0205 UTC. Unlike experimental participants P7, P8, and P10, no control participants observed the deviant motion of the southern storm, which was found to be most consequential to P5. Similar to P7, P5 had also issued a warning on the southern storm with the projected northeast track from WarnGen. However, P5 did not rectify the warning because he failed to notice the storm motion trending toward the east. Subsequently, the severe hail report was located on the outer edge of the warning polygon, and P5 missed the event. The position of the remaining participants' southern storm warnings successfully encompassed the severe hail event reported during case time, but did not encompass severe hail reports reported downstream outside of the case time.

2) CORRECT REJECTION OF A WIND THREAT

The northern and southern storms were not associated with any wind reports. While all control participants identified both hail and wind in their warnings, wind was correctly rejected by two experimental participants. For example, when P12 warned on the southern storm, she searched for "some kind of convergence aloft" since it "would indicate the cell is collapsing and that there would be some kind of wind threat." Failing to see this signature in the velocity data, P12 identified hail as the only weather threat in her warning. Later in the case, P10 assessed the potential wind threat of the southern storm. Although P10 saw "higher velocity in the data," he thought that it "did not look too devastating," and decided to identify hail as the only weather threat in his second southern storm warning. With regard to the northern storm, P12 (along with other control and experimental participants)

noticed weakening of the northern storm during 0205–0209 UTC. In particular, P12 described the northern storm as “losing its punch,” and saw that wind signatures indicative of downdrafts or convergence were absent within the storm. For this reason, P12 issued a severe weather statement on the northern storm to correctly reject wind as a threat.

There were no instances during this case where a control participant was confident enough to reject wind in a warning. Since forecasters are not penalized when they are unable to discriminate between severe hail and wind threats, they tend to include both threats in severe thunderstorm warnings. However, P10’s and P12’s successful discrimination of the weather threat was made possible by observing a persistent lack of wind signatures. This observation was simply not as obvious to control participants who received 5-min PAR updates. Therefore, improved accuracy of warning information may be possible if forecasters using 1-min PAR updates carefully consider what type of weather threat the data support.

b. 2053–2159 UTC 16 July 2009 severe hail and wind case

The 2053–2159 UTC 16 July 2009 severe hail and wind event presented two areas of interest: the nonsevere northern storm and the severe southern storm (Fig. 1). When the case began, participants’ attention was immediately drawn to the northern storm because at low levels it was most prominent on radar. Within the first two periods of the case (2053–2103 UTC), five control and four experimental participants were prompted to issue a severe thunderstorm warning on the northern storm after observing features including a midlevel mesocyclone, hook, and a high-reflectivity core of 60 dBZ above 30 kft. Control and experimental participants’ initial decisions to warn on the northern storm were not impacted by the temporal resolution of the PAR data. P7’s and P12’s decisions to warn on the northern storm followed later in the case, at 2110 and 2120 UTC, respectively. Hesitance to warn on the northern storm, based on the belief that it was not severe, resulted in these participants’ comparatively delayed warning decisions. Given that the southern storm was early in its development at case start time, the storm’s evolution was tracked from soon after initiation to when it became severe. For this reason, the southern storm presents an interesting case where control and experimental participants’ observations of both the onset and the development of severe weather precursors can be compared. Analysis of the RCW identifies differences between the two groups’ warning decision processes, where at times, the use of 1-min PAR updates was found to benefit experimental participants. These

differences are highlighted in the following summaries of warning decisions made with regard to 1) inclusion of the southern storm within the northern storm warning, 2) observing diminishing trends in the northern storm, 3) observing strengthening trends in the southern storm, and 4) discerning tornado potential.

1) INCLUSION OF THE SOUTHERN STORM

While most participants’ decisions to warn on the southern storm followed after their decision to warn on the northern storm, one control (P5) and three experimental participants (P7, P10, and P12) decided to include both storms in their first warning. These decisions arose from observations of new storm development or storm growth aloft. P5’s decision to include the southern storm was made after noticing that “despite having a weak representation at the surface, [the southern storm] was quite a strong storm aloft [with] 60 dBZ to 21 kft.” While P5’s warning decision arose from observations of a single volume scan, P10’s decision arose from trends seen in nine 1-min volume scans. He saw that there were “new cells going up” in what he considered to be a “much more favorable environment” because of the storm’s location relative to the front. This knowledge prompted P10 to “account for new development” within the northern storm warning. Since P7’s and P12’s northern storm warning decisions occurred further into the case (2110 and 2120 UTC, respectively), these participants had the opportunity to observe the southern storm grow. Over this time, these participants saw that the “southern storm [was] continuing to develop” (P12), and therefore decided to “cover all bases” (P7) by including the southern storm in the warning polygon. These intentional decisions resulting from observations made by P5, P7, P10, and P12 resulted in verified warnings. Additionally, P1’s northern storm warning verified, though this success was unintentional. P1 explained that he simply “got lucky” by unknowingly positioning his warning polygon such that it later encompassed a severe weather report associated with the southern storm.

2) DIMINISHING TRENDS IN THE NORTHERN STORM

Further into the case (2124–2129 UTC), experimental participants P10 and P8 assessed the northern storm’s intensity and saw that the severe characteristics had diminished. P10 explained that the northern storm “seem[ed] to be struggling to maintain intensity” and that it was now “becoming outflow dominant.” Additionally, P8 “did not see any signs of midlevel convergence” and the “storm-top divergence was only 50 kts.” Unlike P10 and P8, no control participants reported

observations of these diminishing trends at this time. P8's observations led to a correct decision to cancel the northern storm warning at 2127 UTC. Despite P10 also believing that the northern storm was no longer a threat, P10 refrained from canceling the warning because he did not want to cause confusion given that a second warning was in place for the same county.

3) STRENGTHENING TRENDS IN THE SOUTHERN STORM

As a whole, the experimental group made decisions to warn on the southern storm earlier than the control group. This result is evident in their longer median warning lead time of 21 min compared to the control group's median warning lead time of 16.8 min during case 2 (Bowden et al. 2015). The extended lead time obtained by the experimental group is attributed to their ability to detect and observe the evolution of radar precursor signatures indicative of severe weather earlier than the control group. Although participants' primary reason for issuing a warning on the southern storm was their observations of impressive reflectivity values at high altitudes (e.g., values exceeding 60 dBZ at 30 kft), the perceived information between the groups was quite different.

Experimental participants made decisions based on individual storm trends seen in PAR data more frequently than control participants, who instead typically based their decisions on comparative observations between storms (Bowden et al. 2015). Unlike experimental participants, control participants described the development of a weak echo region and elevated high-reflectivity core between scans as "explosive" (P6) and "notable" (P5). Likewise, over the course of three volume scans (from 2119 to 2134 UTC), P2 saw very few pixels of 60-dBZ reflectivity values expand into a large area of 65-dBZ reflectivity values at 1.5° elevation. Although some control participants were able to use the available information to conceptualize how the southern storm was evolving (e.g., P5 understood from his observations that the storm was associated with a strong updraft), not all participants in the control group were able to develop as thorough an understanding. For example, P1 justified his southern storm warning by seeing that the northern storm "went up, looked good on radar, and came down," and therefore expected the southern storm to do the same. Additionally, P3 failed to make an early projection of the southern storm's severity, and was the last participant to decide to warn on the southern storm. P3's decision to warn only surfaced after detecting (as did all other participants) 50-kt outbound radial velocities at the lowest elevation during the 2124–2129 UTC period.

Observing rapid development in the southern storm resulted in both P1 and P5 expressing a need for faster volume updates. P5 wanted faster volume updates to "validate the strength of the southern storm," whereas P1 wanted to ensure that he would capture the occurrence of a downdraft at the ground should it happen. P1 explained that this was because "by the time you see a downdraft hitting the ground, it will mix up and you won't see it as severe if the WSR-88D timing is off."

No experimental participants expressed a need for faster updates. Rather, experimental participants were able to use the 1-min PAR updates to carefully observe increasing trends in the southern storm's development, which was found to be important not only for making a warning decision, but for identifying the magnitude of the threat within the warning and verifying the threat as it occurred. For example, at 2110 UTC, P9 described that the reflectivity profile of the southern storm was "building higher and higher," and saw that the "trend [was] fast; it [was] not going to peter-out." Observing consistent strengthening trends in reflectivity associated with the southern storm prompted P9's decision to warn. P9 then utilized the 1-min updates to monitor this trend while producing the southern storm warning. Upon interrogation of the radar data, P9 saw that the "trend [was] still increasing [in the] core aloft," which aided his accurate assessment of expected hail size. Additionally, between 2113 and 2116 UTC, P8 used 1-min updates to observe that the trend in midaltitude radial convergence was strengthening and "coming closer together," while the storm-top divergence was increasing in magnitude. These trends observed in the 1-min updates prompted P8's decision to warn on the southern storm. Furthermore, during a postexperiment survey (see section 5), P7 described observations of the southern storm's 1-min outflow velocity trends as "indescribable, really amazing." P7 was able to closely track intensifying velocity values at the surface, which P7 felt had a positive impact on her ability to communicate the potential threat to the public.

4) DISCERNING TORNADO POTENTIAL

In case 1, the participants' focus was primarily on a hail and wind threat, whereas in case 2, a number of participants became concerned about the tornado potential of the northern and southern storms. While this concern led to control participants' (P2 and P4) decisions to issue a tornado warning on the northern and southern storms, respectively, no such decision was made by the experimental participants. An analysis of the RCW provides a comparison of a control participant's (P2) and experimental participant's (P12) consideration for a tornado warning on the northern storm,

and an account of the control participants' greater concern for tornadogenesis on the southern storm.

Tornado warning-related decisions for the northern storm were made by control participant P2 and experimental participant P12 during the first 10 min of case 2, during which P2 received two 5-min PAR updates, and P12 received ten 1-min PAR updates. After P2 received her first scan at 2053 UTC, she reported seeing a "kidney bean shape" and midlevel rotation in the northern storm. These features raised P2's concern for tornado potential, and she waited for the next scan. P12's concern for tornadogenesis on the northern storm was also raised after seeing "broad rotation at higher levels" (2053 UTC). P12's observation directed her interrogation toward velocity data, where she noticed "fairly decent rotation at 2.6°" (2054 UTC). P12 conceptualized that the rotation was in an "appropriate location for where you would expect rotation given the storm shape." However, noting that the rotation was at 13 kft, P12 decided to wait and see if it would descend to the surface. While monitoring velocity trends with incoming data, P12 continued to see that the rotation was in the general area of the hook and that it had strengthened aloft. At 2056 UTC, P12 decided to "wait a couple of scans" and track how the rotation was evolving.

P2 received a second scan at 2058 UTC. Compared to the previous scan at 2053 UTC, P2 saw that the circulation had tightened and deepened, which concerned her because it was "approaching [the] boundary." Based on P2's conceptual model, she anticipated that the rotation would "stretch down to the surface." Eager not to wait until the rotation signature was at the lowest level, P2 decided to issue a tornado warning on the northern storm.

In the 2058 UTC scan, P12 also saw that the rotation was "getting closer to the surface." By 2059 UTC, P12 saw "slight rotation at 0.5°," which triggered her decision to mock up a tornado warning while waiting for new data to come in. P12 continued to assess how the rotational signature was trending. By 2101 UTC, though, P12 explained that "whatever was there at the surface has sort of fallen apart and it's a little too far south of the hook." Feeling unsure, P12 watched as the next two scans came in. By 2103 UTC, P12 felt confident that "the rotation I was seeing aloft has diminished and [there was] still no sign of rotation at 0.5°," and decided against issuing a tornado warning. While it is possible that differences between P2's and P12's warning philosophies may have contributed to the decision outcomes, P12's ability to detect and track diminishing trends of low-level velocity signatures in the 1-min PAR data was crucial to her correct rejection of a tornado warning. P2's tornado warning was maintained until she decided to cancel it during 2119–2124 UTC, after seeing that the northern storm had

"completely lost any mesocyclone signature, [and] lost kidney bean structure and inflow notch."

Tornado potential associated with the southern storm was also a concern among the control and experimental participants. Throughout case 2, the southern storm tracked toward the frontal boundary. P4 and P5 found this worrying during 2119–2124 UTC, since they thought that their observations of a convergent signature with subtle rotation could enhance as the storm encountered the boundary. P6 saw that the "updraft [was] strengthening above the low-level mesocyclone," and conceptualized that "stretching" caused by the updraft would "help spin up a tornado." These three control participants acted as follows: P4 mentioned the potential for a tornado in a severe weather statement, P5 chose to "watch for next scan of low-level rotation in case a [tornado warning was] needed," and P6 decided to "instigate the process for a tornado warning." With the next scan at 2124 UTC, P5 and P6 recognized that the current wind threat—as indicated by radar—was straight-line winds. P5 drew this conclusion after seeing a "strong wind core now at the surface" and that the velocity data looked "more like a straightline wind segment than rotation at this point." P6 also came to this conclusion after seeing that a "wall of strong inbounds [had] developed" while the "low-level rotation had at least temporarily weakened." P4, however, became more concerned about a tornado threat because he expected the southern storm to "ingest storm relative helicity" as it intersected with the frontal boundary. Upon receiving the next scan at 2129 UTC, P4 saw both inbound and outbound velocity values, along with rotation in the lowest two tilts. These observations triggered P4's decision to issue a tornado warning on the southern storm (which was a false alarm). While experimental participants P9 and P10 also expressed concern regarding the southern storm's tornado potential during this same time, their concern lasted only a couple of minutes. Analyzing the 1-min updates, these participants were able to observe the evolution of the low-level velocity data and successfully determine the lack of tornado threat quickly. P9 discarded his concern after seeing that the velocity values were "elongated" rather than shaped like a "couplet," while P10 explained that he "never saw any respectable rotation in the lowest slice," and thought that the winds were "more of an outflow surge."

5. Postexperiment survey

A postexperiment survey designed to capture reflections, attitudes toward, and summaries of participants' experiences during the 2013 PARISE was issued

once both cases had been completed. Although questions were issued to both groups, the following synthesis of responses focuses on those from the experimental participants. This focus was chosen since the experimental group's responses inform on benefits, concerns, and other aspects of higher-temporal-resolution radar data use during simulated warning operations that were encountered and are worth future study.

Experimental forecasters were first asked whether they approached each case study with the same forecasting style, and if not what did they change and why. Three participants in this group said that they altered their approach between the two cases. For example, P7 explained that during case 1, she was "behind a lot" because it had "not hit [her] yet that data was coming in every minute." However, as the cases progressed, she described having more control by building awareness for when "new scans were coming in." For P9, he developed the attitude that "it was okay to miss a scan," and that he could go back and look at data as the situation required. Similarly, P11 altered his approach toward 1-min PAR data by "learn[ing] to sit tight and let [the data] come to you" rather than going "back [so] much to previous scans." These responses correspond well to findings from the 2012 PARISE, where over time participants adapted to the pace of 1-min PAR data (Heinselman et al. 2015). As found in Heinselman et al. (2015), rather than interrogating every scan at all elevations, some forecasters reduced the demand on their interrogation strategies by focusing on areas of the storm deemed most important during the evolution process and intermittently surveyed the entire scenario as necessary.

Although only half of the experimental group described a change in their approach to using 1-min PAR data during warning operations, all experimental participants felt that rapidly updating PAR data enhanced their ability to observe the rate at which storms were evolving in finer detail. This finding was apparent after asking participants how the higher-temporal-update time made a difference to their forecasting. The experimental group described that being able to observe "patterns develop more gradually" (P9) enabled them to obtain a sense for how features in the storm were trending. As explained by P11, a common phrase during warning operations is "I'm going to wait one more scan to see what happens." Experimental participants reported, however, that the current 4–6-min delay resulting from the WSR-88D temporal resolution was reduced because trends could be monitored each minute. Furthermore, during times when warning decisions were being made, P8 reported that the use of rapidly updating PAR data "gave [him] the ability to hold off on a warning knowing the next scan was a minute away." He noted that "with the WSR-88D, it is

five minutes. . . by then it may be too late." Similarly, P12 explained that at times when she was unsure of a storm's severity, she "tended to use a draft warning and then as new information came in, [she] made a decision." Such additional information was found to be particularly helpful to P12 while she was considering the tornado potential of the northern storm during case 2. Along with finding value in 1-min PAR updates prior to a warning decision, P9 used incoming data during the generation of warnings, because "before [he] could finish [the warning], things could be changing. [He] could look one last time and make a change if [he] needed to."

Analysis of timelines suggested that being able to monitor the storm evolution in finer detail through the use of 1-min PAR updates meant that experimental participants were able to observe and better match radar observations to their storm conceptual models. This finding was further confirmed when experimental participants were asked what key features or storm evolution patterns they could detect or follow in the storm that they would not have been able to if they were working with the WSR-88D. For example, P10—who is an avid storm chaser—described being able to see "convective development on a scale more similar to visual observations of storms." Experimental participants were particularly reflective of their abilities to view the southern storm's development intimately during case 2. P8 recalled that during case 2, he was able to see "the core descent every 1 kft. . . rather than [from] 10 kft to 2 kft." P12 explained that being able to see the gradual descent of the high-reflectivity core meant that "it [was] not as much of a surprise when intensification happen[ed]." In addition to the core descent, P7 recalled being able to see the "midlevel convergence moving down, hit ground, and splat out. . . [the outflow] was every minute. . . first 30 kts, then 40 kts, then 50 kts, then 60 kts." P7 described these observations as "indescribable, really amazing," as well as "humbling because you can see the impacts on the people." Being able to track the evolution of the storm both prior to and during the onset of severe weather was valuable to P7 because she felt she could convey the potential impacts to the public better.

Two experimental participants also found that 1-min PAR updates were useful for monitoring nonsevere weather. For example, P10's ability to observe the storm evolution each minute provided him with confidence that that he was not missing anything between volume updates, such as brief spinups, as demonstrated during case 2. P11 also found that his decision to not warn on the nonsevere northern storm in case 1 was aided by being able to see that his severe hail criterion was not being met in each scan. Despite seeing that the "northern storm had interesting features with it," P11's continuous monitoring

aloft in the storm reassured him that a warning was not necessary.

Given that all experimental participants reported positive impacts of 1-min PAR data on their warning decision processes, it was of interest whether these participants thought that this temporal resolution was optimal. Therefore, we asked participants how they dealt with the incoming data and whether they would choose to increase or decrease the frequency of updates. Of the six experimental participants, four felt comfortable receiving updates each minute. Of these participants, P8 and P10 thought that this temporal resolution was sufficient for their needs. P8 explained that he “felt like [he] had what [he] needed” and “never felt like [he] needed one more scan to make a decision.” Although P9 and P12 also adapted well to 1-min PAR updates, they expressed concern regarding forecaster fatigue and the potential for falling behind during times when forecasters are busy issuing products. Both participants suggested decreasing their data load by removing unnecessary elevations from a volume coverage pattern depending on the storm phenomena. P12 explained that this could be accomplished by sampling, for example, only lower elevations during a potentially tornadic scenario. Interestingly, although these forecasters felt overwhelmed by faster updates, this sampling technique would result in more frequent updates at lower elevations. The two remaining experimental participants, P7 and P11, were both initially overwhelmed with the 1-min PAR updates. P7 described feeling “like [she] couldn’t evaluate it or keep up,” while P11 thought the data “was like a firehose, couldn’t find the turn-off switch.” However, P7 found that although she felt under pressure to keep up with the rapidly updating data, she did not have the stress of “dying to know what was happening and waiting, worrying” like when working with the 4–6-min WSR-88D updates. Although P7 and P11 became more comfortable with the pace of the data as they worked through the cases, they suggested that forecasters should be introduced to 2-min radar updates first, and that over time they would get used to 1-min radar updates.

6. Conclusions

In this paper, we highlight the importance of delving deeper into forecaster performance beyond lead time and verification statistics to understand, in context, why forecasters make warning decisions. The qualitative data collected during the 2013 PARISE provided us with a wealth of information to analyze what impacts 1-min PAR updates had on forecasters’ warning decision processes while working severe hail and wind events in simulated real time. By applying a situational

awareness framework to participants’ RCW timelines (Endsley 1995; Hoffman 2005), we found that the experimental group perceived significantly more information than the control group in both cases (Table 1). It is possible that this increase in acquired information aided experimental participants’ ability to comprehend the scenario, consequently improving their projections of storm activity, and resulting in a larger number of mastery decisions (Bowden et al. 2015).

Experimental participants demonstrated improved projections of storm activity compared to control participants during both cases. We found that experimental participants not only perceived severe weather precursor signatures earlier than the control group, but they were able to monitor trends in the 1-min PAR data that enabled better projection of storm motion and expected weather threats. Additionally, experimental participants used 1-min PAR updates to observe and confirm with confidence the absence of a precursor signature over successive scans. Such observations led to correct rejections of wind threats in severe thunderstorm warnings, as well as to correct rejections of tornado warnings.

From the postexperiment survey, we considered as a whole how the 1-min PAR updates impacted experimental participants’ warning decision processes. Our findings are highly suggestive that forecasters will respond to and interact with higher-temporal-resolution radar data differently. For this reason, future research should focus efforts on assessing human factor aspects of how forecasters integrate rapidly updating PAR data into their warning procedures. Addressing issues such as forecasters’ mental workload, use of tools and products, and choice of data display in response to a variety of volume update times, will provide further insight into what the optimal temporal resolution of radar data may be during warning operations and how best to deliver that information to the forecaster.

Acknowledgments. Thank you to the 12 NWS forecasters for participating in this study, to the participating WFO’s MICs for supporting recruitment, and to Michael Scotten for participating in the pilot experiment. We also thank AWIPS-2 expert Darrel Kingfield, A/V specialist James Murnan, software expert Eddie Forren, and GIS expert Ami Arthur. Advice from committee members Robert Palmer, David Parsons, and Rick Thomas, along with insightful discussions with Harold Brooks and Lans Rothfusz, aided the development of this study. Finally, thank you to the anonymous reviewers for their helpful feedback. Funding was provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA–University

of Oklahoma Cooperative Agreement NA11OAR4320072, U.S. Department of Commerce.

APPENDIX

Probing Questions Used during the Recent Case Walk-through

- 1) What was the judgment? Why did you make it and what information did you use?
- 2) Is there any other information that would have aided you during this judgment? If so, what?
- 3) Were there any other factors that influenced your judgment, for example, storm reports, population of affected area?
- 4) Why did you rate your judgment confidence at this value?

REFERENCES

- Andra, D. L., E. M. Quoetone, and W. F. Bunting, 2002: Warning decision making: The relative roles of conceptual models, technology, strategy, and forecaster expertise on 3 May 1999. *Wea. Forecasting*, **17**, 559–566, doi:10.1175/1520-0434(2002)017<0559:WDMTRR>2.0.CO;2.
- Bowden, K. A., P. L. Heinselman, D. M. Kingfield, and R. P. Thomas, 2015: Impacts of phased-array radar data on forecaster performance during severe hail and wind events. *Wea. Forecasting*, **30**, 389–404, doi:10.1175/WAF-D-14-00101.1.
- Braun, V., and V. Clarke, 2006: Using thematic analysis in psychology. *Qual. Res. Psychol.*, **3**, 77–101, doi:10.1191/1478088706qp0630a.
- Calhoun, K. M., T. M. Smith, D. M. Kingfield, J. Gao, and D. J. Stensrud, 2014: Forecaster use and evaluation of real-time 3DVAR analyses during severe thunderstorm and tornado warning operations in the Hazardous Weather Testbed. *Wea. Forecasting*, **29**, 601–613, doi:10.1175/WAF-D-13-00107.1.
- Endsley, M. R., 1995: Toward a theory of situation awareness in dynamic systems. *Hum. Factors*, **37**, 32–64, doi:10.1518/001872095779049543.
- Ericsson, K. A., and H. A. Simon, 1993: *Protocol Analysis: Verbal Reports as Data* (rev. ed.). The MIT Press, 496 pp.
- Evans, C., D. F. Van Dyke, and T. Lericos, 2014: How do forecasters utilize output from a convection-permitting ensemble forecast system? Case study of a high-impact precipitation event. *Wea. Forecasting*, **29**, 466–486, doi:10.1175/WAF-D-13-00064.1.
- Guan, Z., S. Lee, E. Cuddihy, and J. Ramey, 2006: The validity of the stimulated retrospective think-aloud method as measured by eye-tracking. *Proc. CHI 2006 Conf. on Human Factors in Computing Systems*, Montreal, QC, Canada, Association for Computing Machinery, 1253–1262, doi:10.1145/1124772.1124961.
- 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, doi:10.1175/WAF-D-11-00145.1.
- , —, D. M. Kingfield, and R. Hoffman, 2015: Tornado warning decisions using phased-array radar data. *Wea. Forecasting*, **30**, 57–78, doi:10.1175/WAF-D-14-00042.1.
- Hoffman, R. R., 2005: *Protocols for cognitive task analysis*. Florida Institute for Human and Machine Cognition, Pensacola, FL, 108 pp. [DTIC ADA475456.]
- Hoium, D. K., A. J. Riordan, J. Monahan, and K. K. Keeter, 1997: Severe thunderstorm and tornado warnings in Raleigh, North Carolina. *Bull. Amer. Meteor. Soc.*, **78**, 2559–2575, doi:10.1175/1520-0477(1997)078<2559:STATWA>2.0.CO;2.
- Lindley, T., and G. Morgan, 2004: The Pecos County, Texas hail storms of 10 May 2002: A null tornado event from a warning decision perspective. *Electron. J. Oper. Meteor.*, **5** (1). [Available online at <http://www.nwas.org/ej/2004-EJ1/>.]
- Logan, G., 1988: Automaticity, resources, and memory: Theoretical controversies and practical implications. *Hum. Factors*, **30**, 583–598, doi:10.1177/001872088803000504.
- Morss, R. E., and F. M. Ralph, 2007: Use of information by National Weather Service forecasters and emergency managers during CALJET and PACJET-2001. *Wea. Forecasting*, **22**, 539–555, doi:10.1175/WAF1001.1.
- NOAA, 2011: Verification. Rep. NWSI 10-51601, 100 pp. [Available online at <http://www.nws.noaa.gov/directives/sym/pd01016001curr.pdf>.]
- Quoetone, E., J. Boettcher, and C. Spannagle, 2009: How did that happen? A look at factors that go into forecaster warning decisions. Preprints, *34th Annual Meeting*, Norfolk, VA, National Weather Association. [Available online at <http://www.nwas.org/meetings/nwa2009/>.]
- Ranyard, R., and O. Svenson, 2011: Verbal data and decision process analysis. *A Handbook of Process Tracing Methods for Decision Research*, M. Schulte-Mecklenbeck, A. Kühberger, and R. Ranyard, Eds., Psychology Press, 115–137.
- Van Gog, T., F. Paas, J. J. G. Van Merriëboer, and P. Witte, 2005: Uncovering the problem solving process: Cued retrospective reporting versus concurrent and retrospective reporting. *J. Exp. Psychol. Appl.*, **11**, 237–244, doi:10.1037/1076-898X.11.4.237.
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Academic Press, 467 pp.
- Zrnich, D. S., and Coauthors, 2007: Agile beam phased array radar for weather observations. *Bull. Amer. Meteor. Soc.*, **88**, 1753–1766, doi:10.1175/BAMS-88-11-1753.