14A.3 Exploring Impacts of Rapid-scan Radar Data on NWS Warning Decision Making

Pam Heinselman[1], Daphne LaDue[2], and Heather Lazrus[3,4]

[1]NOAA National Severe Storms Laboratory, Norman, Oklahoma
[2]Center for Analysis and Prediction of Storms, Norman, Oklahoma
[3]Social Science Woven into Meteorology, Cooperative Institute for Mesoscale Meteorological Studies
[4] Current affiliation National Center for Atmospheric Research, Boulder, Colorado

1. INTRODUCTION

The development of rapid-scan capabilities with S-band phased-array radar (PAR) at the National Weather Radar Testbed in Norman, Oklahoma (Zrnić et al. 2007; Heinselman and Torres 2011) presents new opportunities for advancement of weather sensing. An important component of the technology development process is assessment of operational benefit(s) of new radar capabilities. This has been done for previous radar upgrades, only during the technology transfer process (e.g., JDOP (Burgess 1979) and JPOLE (Scharfenberg et al. 2005)).

The need for more frequent scanning of storms to improve operations is supported by several studies, including a VCP-needs survey conducted by the NOAA National Weather Service (NWS) Radar Operations Center (Steadham 2008), a radar-needs assessment conducted by the Office of Federal Coordinator for Meteorology (OFCM 2006), and a strengths and limitations study of operational radar systems (LaDue et al. 2010). Findings from an in-depth 5-yr (2000–2004) study of NWS warning performance (Brotzge and Erickson 2010) suggests that rapid scanning, in part, may improve warning accuracy. They found that approximately 27% of weak (EF0 and EF1) tornadoes were unwarned, compared to 5.3 and 8.6% of EF3 and EF4 tornadoes, respectively. The shorter lifetimes typical of EF0 and EF1 tornadoes present the opportunity to assess improvement in their probably of detection with faster scan times.

The NWRT PAR has demonstrated capability to better sample the rapid evolution of severe weather events compared to the WSR-88D (Heinselman et al. 2008; Newman and Heinselman 2011), but little is known about the spectrum and prevalence of factors impacting NWS forecaster warning decision processes. We are aware of just one study of warning forecasters (Hahn et al. 2003) using cognitive task analysis to capture expertise of warning forecasters. Such information is useful to design of training. Our eventual goal for this larger technology development effort is to understand how new information might impact decision making across the range of users targeted by the technology. In the first step toward this goal, we conducted a pilot study in the spring of 2010 that explored and identified factors impacting NWS forecaster decision making with two different controlled conditions: 1) when radar scan time is similar to the shortest VCP employed on the WSR-88D (VCP 12) and 2) when radar scan time is faster than VCP 12. This pilot study was part of the 2010 Phased Array Radar Innovative Sensing Experiment (PARISE) during the last three weeks of April.

2. PARTICIPANT SELECTION AND DEMOGRAPHICS

NWS forecasters were invited to consider participating in one of three 2010 projects via the NOAA Hazardous Weather Testbed Experimental Warning Program e-mail distributed to National Weather Service Forecast Offices across the nation by the five NWS Regional Offices. The invitation asked respondents to explain in writing their interest in participating in the EWP. The primary applicant pool indicating interest in PARISE contained 34 NWS forecasters; 94% were from offices located in the Central, Eastern, or Southern Regions, and the 12 participants were chosen from these 3 regions.

Twelve participants were selected for the larger PAR project based on the content of their written interest statements, location of their home office, sex, and experience with radar data. Formal recruitment to participate in the day-long experiment was done early in the week

---

**Corresponding author address*: Pam Heinselman, NOAA NSSL, 120 David L. Boren Blvd., Norman, OK, 73072; e-mail: pam.heinselman@noaa.gov

(see below).[2] Forecasters whose interest statements provided evidence of reflective thinking and experience evaluating weather products and display tools were given priority over others. We examined information on years of experience to determine a first guess at forming teams that would balance during each week of the experiment.

The 12 participants included 3 females and 9 males from NWS offices located in 11 different states in or east of the Rocky Mountains (Fig. 1). Diversity in office locations brought together forecasters with experience issuing warnings on storms whose development is impacted by different climatologic conditions and terrain features. Additionally, most participants, 11 of 12, had worked at two or more offices in different geographic regions. The number of years' experience working in the NWS ranged from 5 to 23, with an average of ~12 years of service (Fig. 1). Four of the 12 participants had held positions in private industry 1.5–5 yrs before being employed with the NWS. At the time of the experiment, 11 of the 12 participants were in forecaster positions, and one was a meteorologist in charge. All participants had experience issuing warnings for severe weather, and more than half (7) had several years' experience having consistently worked the warning desk during severe events. The sample of forecasters who participated in the study is neither representative nor broadly generalizable.

3. PAR DATA

NWRT PAR data were used to create two data sets with different update times: one with full-temporal resolution (43s), and the other with simulated WSR-88D-like temporal resolution (~4.5-min volume scan with elevations updating through that time period). This controlled for any real or perceived differences between PAR and WSR-88D data. The 4.5-min update time was chosen to match as closely as possible the sampling time of the Oklahoma City WSR-88D on the case date. The simulated 4.5-min volume scans were constructed by 1) determining 4.5-min update times for each elevation (14 total) over the case duration, 2) matching the nearest-in-time NWRT PAR elevation to these revised times, and 3) assigning the new time stamp to each elevation. Time stamps of simulated data were ±2–20 s different than the full-temporal resolution data. During one 4.5-min VCP 12 volume scan, the 43-s team usually received 5 volumetric updates.

This methodology was applied to a 45-min period (01:13:29–01:59:39 UTC) on 19 August 2007 when the NWRT PAR sampled two supercells as they moved north-northeastward from southwest toward west-central Oklahoma (Fig. 2). These supercells formed a few hours prior to the reintensification of tropical storm Erin in this region (Arndt et al. 2009; Evans et al. 2011). A subsequent damage survey conducted by scientists involved in the Severe Hazards Analysis and Verification Experiment (Ortega et al. 2009) indicated a short-lived (~3 min) tornado occurred during this time. Along its 2.0-km long and 0.036-km wide damage path the tornado removed the roof of a mobile home and snapped branches in the vicinity of Norge, Oklahoma (information online: http://ewp.nssl.noaa.gov/projects/shave/tornsurveys.php#map). An examination of the NWRT PAR data along this damage path indicates the EF1 tornado was produced by the north-most supercell during approximately 0144–0147 UTC 19 August 2007.

These storms were sampled within a 60° sector using a scanning strategy that employed the same elevation angles as VCP 12 (NOAA 2006) and 0.5° overlapped azimuthal sampling (Heinselman and Torres 2011). The implementation of an electronic scanning technique called beam multiplexing (Yu et al. 2007) resulted in 43-s volumetric updates.

The selection of this event provided the opportunity to explore the impact of temporal sampling on the storm structures, evolution, and trends observed by forecasters prior to, during, and after the occurrence of a tornado whose longevity is near the update time of the WSR-88D. Also explored is the weight given to radar data in forecaster decisions to warn and/or not to warn. The forecasters' decision processes provide insight into how rapid-update data may aid decisions to warn or not warn on supercell storms in tropical environments that may produce weak, short-lived tornadoes.

---

[1] The plan for this study was approved by The University of Oklahoma's Office for Human Research Participant Protection (a.k.a. Institutional Review Board).

4. EXPERIMENT DESIGN

This study generally followed a matched-pairs, control-group design (Mertens 2005), though matches were on teams of two and were subjectively determined. Forecasters worked in pairs for two reasons. First, EWP participants have tended to prefer to work in pairs in the NOAA HWT. We chose to exploit that preference to help us to capture their thought processes as they made decisions together. Second, because there are no reliable, objective means to assess radar data interpretation and warning decision making skills, the matched pairs were based upon observations as participants had rotated through partners to work three events earlier in the week. They had gained a sense of each other's knowledge and skills. We suggested groups, explaining that we sought teams that were roughly equivalent in regard to radar data interpretation skills. All were agreeable. The debriefing plans allowed room for them to tell us if they thought any unintended group differences had affected the outcome. None did.

Four NWS forecasters participated each week of the experiment, which ran from midday Tuesday through Friday morning. On Tuesday they were formally introduced to our pilot study and given the option to participate; all consented. Through Wednesday they engaged in activities developed to build forecaster experience using the Warning Decision Support System – Integrated Information (WDSS-II; Lakshmanan et al. 2007) display software and provide valuable feedback on a variety of weather types.

On Thursday, the day of the study, participants worked through the 19 August 2007 case (see section 3) as if they were on the job, issuing weather warnings and updates. At the start of each case they had approximately 20 min to review weather data to gain situational awareness. These weather data were displayed in the weather event simulator (known as WES); included were in situ and remote sensing observations, numerical model output, and products issued by the NOAA NWS. They then wrote a discussion about what they thought would happen in the next hour or so. The case then ran in a displaced real-time mode. (The term case refers to both NWRT PAR data and operational nonradar data). Forecasters were observed and audio recorded; their computer screens were video recorded (all participants provided consent for all of these). Computer files of the discussion text and warning related texts were archived on the computer.

After each case, a debriefing was conducted with each group independently. Forecasters were invited to take notes on critical decision points to use in the Joint Debrief, in which all forecasters from both groups participated at the end of the day to discuss their experience with the data and its impact on their warning decisions. They were asked to rank factors in importance to their decision making and to rate their confidence relative to usual on two continuums: one for the type of event and one for the impact of the radar data. Finally, all participants came together to discuss their experience with the data and its impact on their warning decisions. After the case they were asked for summary information on their work histories that might be relevant to interpreting differences in their warning decisions.

4.1 Limitations

Any approach has limitations. This study had the following. First, the WDSS-II was used in lieu of the Advanced Weather Interactive Processing System (AWIPS) due to the operational software's inherent inability to display data sampled with an update time of 1-min or shorter. We sought to minimize the impact of software differences on the experiment outcomes by upgrading the WDSS-II with WARNGEN functionality similar to AWIPS. Second, we could not objectively match pairs because there is no known fair method to assess warning forecaster expertise. We did our best in absence of such. The team dynamic may have affected warning decision making to a greater extent than what could be gleaned from our data, but it minimized the effect of unfamiliar software as they could help each other manage the added cognitive load. The software options available meant forecasters had additional tools they did not normally have, but the software also could not display data in some ways they were accustomed to.

Some researchers in the decision making field build a strong cause for only studying decision making in natural settings (Schraagen et al. 2008), as Joslyn & Jones (2008) did with Naval weather forecasters creating terminal aerodrome forecasts. We sought to have multiple instances of the same case comparison to analyze, meaning case data had to be used. This approach was due, in part, to the fact that real-time data would be relevant only to NWS operations in Norman, Oklahoma, as only one NWRT PAR exists. Even if used in this capacity, concurrent use of traditional radar data in operations would make it difficult to isolate the impact of NWRT PAR data on decision making. The contrived nature of a simulation, though, tended to mean two things to participants: something was likely to happen, and it would happen in the absence of pressures only live operations have. We simulated several aspects of an operational setting and asked forecasters to attempt to work as closely as they would to normal. They all felt they acted normally, but commented that cases never truly simulate the tension, distractions, and other aspects of operations.

5. ANALYSIS METHODS

Data analysis involved several steps. After a student hire transcribed audio recordings, transcripts were reviewed by the researchers and minor errors were corrected to ensure the data was an accurate record of the experiment (Singleton and Straits 2005). Transcripts were then made more manageable by coding the information to extract meaning (Bernard 2006). Our approach was primarily data-driven, meaning the analytic categories and themes were identified inductively (Boyatzis 1998). Some additional codes flagged text addressing additional interests such as aspects of the data, software, or research design.

The video recordings visually documented interactions of participants with the WDSS-II and WES display, and the specific radar fields and radar signatures, or other observations they were viewing during the case. In coordination with the transcripts, this visual information was used to augment, clarify, and confirm coding of cognitive actions and states of being of participants, as well as issues related to experimental design. The use of video information, for example, illustrated the series of radar moments and radar signatures viewed, interrogated, or interpreted by participants preceding each warning decision.

All three researchers worked together to identify major categories: expressions of state; cognitive actions; and meteorological, environmental, or other data being examined; as well as reflections on the experiment design. For example, forecasters expressed their mental state (e.g., excitement, frustration, surprise, uncertainty); demonstrated cognitive actions (e.g., comparing, deciding, interrogating, interpreting); considered meteorological or other factors in the data (e.g., meteorological factors, time of day, office policies); and reflected on the design of the experiment (e.g., wishes for functionality, software problems, overall design). The resulting coding scheme had several similar codes, such that an intercoder comparison of DL and PH's coding of one case yielded 33% identical codes. For the remaining two thirds, DL modified one third to PH's codes, and PH modified one third to DL's codes. Modifications ranged from changing between similar codes and adding missed codes. In the end, two of us (DL and PH) swapped files frequently to modify and add to the codes while we analyzed the warning decisions, and one of us (HL) focused on nontechnical codes and analysis of those factors.

To further aid our understanding of how each team’s interpretation of storms and warning decision(s) related to the supercell evolution depicted by the 43-s and 4.5-min NWRT PAR data, we performed an analysis of trends of circulation strength over the case study period. Due to the shallow nature of the circulations and the distance of the two supercells from the NWRT PAR, the analysis was limited to the 0.5° velocity field. In this case, the radar sampled circulations at heights of 0.7–0.9 km above radar level (ARL) in the north storm, and at heights of 1.0–1.3 km (ARL) in the south storm; beam widths ranged from 1.8–2.0 km in the north storm, and from 2.1–2.4 km in the south storm. Hence, circulations sampled in north storm were slightly better resolved spatially than those sampled in the south storm, especially at the beginning of the case when their azimuthal positions differed most.

The local, linear least squares derivatives (LLSD) method for calculating azimuthal shear (Smith and Elmore 2004) was used as an objective measure of circulation strength. Though by design LLSD azimuthal shear is more immune to dealiasing errors and noisy velocity data than “peak-to-peak” azimuthal shear, its use of a median filter can also smooth out peaks in the velocity field. Such smoothing can result in underestimates of the true azimuthal shear of a circulation, especially if the circulation is small. Like most radar-derived parameters, LLSD azimuthal shear (hereafter shear) values can also vary according to range, radar angle, and beam width (Smith and Elmore 2004). As a result, it is the *trends* in azimuthal shear, rather than the actual values, that are generally most important. The resulting time series is derived from the maximum LLSD azimuthal shear values observed within each circulation (Figs. 3b, d) and is used as a proxy for understanding the more complex data analysis done by forecasters.

6. TRENDS IN CIRCULATION INTENSITY

The time series of 4.5-min LLSD azimuthal shear values obtained from the north and south supercells contain peaks that suggest three increases in circulation intensity in the north storm, and four increases in circulation intensity in the south storm (Fig. 3d). The lifetimes of each cycle, defined by consecutive upward and downward trends, were about 10 to 20 min. After 01:27:06 UTC, the peak 4.5-min LLSD azimuthal shear values in the north storm exceeded those seen in the south storm (Figs. 3b,d), and were generally at or above values generally associated with mesocyclone-scale circulations (0.01 s^{-1}; Smith and Elmore 2004). As a result, the north storm's circulation looked more organized in the velocity field, exhibiting tighter and stronger circulations than were seen in the south storm. Between 01:40:01 and 01:44:19 UTC, the north storm's circulation underwent a dramatic 0.012 s^{-1} increase in LLSD azimuthal shear, which marked the onset of a short-lived, EF1 tornado (Fig. 3d).

Unsurprisingly, the 43-s LLSD azimuthal shear time series contains spikes of increased and decreased circulation intensity within the broader cycles depicted by the 4.5-min time series (Figs. 3b, d). The short-lived nature of these spikes resulted in undersampling of peaks and bases in the 4.5-min time series. For example, the increase in the south storm's LLSD azimuthal shear seen in the 43-s time series between the 01:40:01 and 01:44:19 UTC, is absent from the 4.5-min time series (Fig 3b, d). Between 01:35:43 and 01:40:01 UTC, in the 43-s time series the north storm's circulation strength diminishes and then begins to reintensify, while in the 4.5-min time series circulation strength appears to have solely diminished (Figs. 3b, d). After 01:44:19 UTC, the second peak in the circulation's intensity (up to 0.02 s^{-1}), is absent from the 4.5-min data. The coarser sampling, however, also reduced the impact of small, seemingly insignificant changes in LLSD azimuthal shear, as seen in the south storm from 01:17:04 through 01:19:56 UTC. Such small changes in LLSD azimuthal shear may be due, in part, to radar sampling issues (Smith and Elmore 2004; Newman et al. 2011).

As in usual warning operations, forecasters assessed the meteorological significance of trends they interpreted through interrogation of the velocity data, in light of circulation structure and correspondence with features like reflectivity notches or appendages in the reflectivity data. The warning decision processes, warning types, and lead times resulting from temporally different data sets follows.

7. OBSERVED DECISION MAKING

During the case, several decisions were made by each team. Some of the most notable weather-related decisions concerned 1) the anticipated severe weather threats, 2) whether to warn or not warn on a storm, 3) the warning type and duration, and 4) when to issue a severe weather statement. Tornadoes were assessed as the primary threat by 5 of the 6 teams, and as a threat secondary to damaging winds by one team. Unlike the other 5 teams, that team did not realize the tropical nature of the event until shortly after they wrote their discussion; they had not noticed and read all the NWS products displayed on the WES. As far as we can tell, however, the additional information about the environment did not alter the team's threat-mindset before starting the radar-portion of the case.

Decisions about which of the two storms to warn on, warning type, and timing varied more than the threat assessments. As shown in Figs. 3a and c, though all teams issued tornado warnings on the north storm, one team (43-s) issued a tornado warning on the south storm. Tornado warnings and one severe warning on the north storm were verified by an EF1-tornado that occurred from approximately 1:44–1:47 UTC. Three of the six teams issued tornado warnings with positive lead time: two 43-s teams with 18.6- and 11.5-min lead times (Fig. 3a), one 4.5-min team with a 4.6-min lead time (Fig. 3c). As discussed later in this section, one 4.5-min team's decision to warn 7.6 min prior to the tornado was delayed due to software issues and then aborted after the 01:40:08 UTC scan arrived. Only the team with wind as their primary threat issued severe thunderstorm warnings: one on each storm. The one issued on the north storm had a 6-min lead time (Fig. 3a). They were also the only 43-s team to issue a tornado warning with negative lead time.

Three warning updates were issued up to 01:47 UTC: two on the southern storm and one on the north storm. Updates on the southern storms were made due to elapsed time rather than desire to convey changes in the storm character. The 43-s team with long lead on the northern storm tornado warning me (18.6 min, Fig. 3a) issued an update during the tornado lifecycle based on meteorological changes; in this case the rapid intensification and tightening of the mesocyclone circulation.

Throughout the following warning decision comparisons data values are those voiced by participants. Video of WDSS-II was used to determine which radar moment and elevation scan they were viewing. The time series of LLSD azimuthal shear in Figs. 3b, d are referenced for comparison when forecasters interpret changes in circulation strength.

7.1 *Warning Decision Comparison A*

In this comparison the dramatic ~19 min difference in warning issue times (Fig. 3a) was due to the 43-s team's low threshold for the persistence and strength of a circulation in a tropical environment and the 4.5-min team's difficulty with the software. In addition, the 43-s team exhibited far more confidence in what they saw prior to the tornadic intensification than the 4.5-min team did.

The 43-s team oriented very quickly to the software and the event. Between 1:18:30 and 1:22:05 UTC, they noted "a bit of rotation" collocated with a notch in reflectivity on the northern storm and began watching for an upward trend in velocity. The velocity trended upward to 25 kts by 1:21:22 UTC (Fig. 3b). They agreed that with this type of system they were not likely to see 50 kts inbound and outbound. When the 0.5° elevation of the 1:22:05 UTC scan came in, Bob noted the southern storm was also trending upward (Fig. 3b). At this point they began to build toward the warning decision. Joe pointed out that the circulation was tightening a bit. Bob agreed and noted it was also a persistent feature. They confirmed the tightening circulation corresponded to the reflectivity notch and Bob said, "And that's where I'm concerned... It doesn't take much." Joe agreed, saying, "I'm okay with starting a tor on that." As they initiated WARNGEN, Joe added, "'Cause I think being aggressive in this environment is a good thing to do." They briefly considered how long the warning should last and proceeded to issue the warning at PAR scan 1:23:31 UTC. They issued the warning just over 2 min later with 18.6 min lead time (Fig. 3a).

Over the next 9 min (1:23:31 – 1:32:07 UTC), their confidence continued. At that point Bob declared, "There is absolutely something going on there [in the northern storm]." They consulted environmental data and monitored the southern storm. By PAR scan 01:35:00 UTC, Bob noted 40 kts of gate-to-gate shear on the northern storm, "pretty good for a tropical system." On PAR scan 01:39:18 UTC, they assessed that the southern storm continued to have weak outbound velocities, but 41 kts inbound, "something that we are going to have to consider," Bob said. Joe agreed, "It seems there is kind of a general area of rotation there." However, by PAR scan 01:40:44 UTC (Fig. 3b), the southern storm had "lost that sharp edge." Approximately 2.5 min before the tornado produced by the northern storm began, PAR scan 01:41:27 UTC showed a tight area of circulation that this team noticed. On the next update, PAR 01:42:10 UTC, they were excited by the strength of the circulation (Fig. 3b). They then initiated WARNGEN to start a warning update based upon meteorological changes in the storm, noting that velocities were now 88 kts inbound and almost 30 outbound, saying, "That's the best we've seen." The Severe Weather Statement was issued just under 2 min later, immediately before the 1:45:02 UTC scan came in.

The 4.5-min team's first warning decision process followed shortly after they detected a brief strengthening, then weakening, of the northern storm using a loop of storm relative motion up to PAR scan 1:26:46 UTC (Fig. 3d). Suspecting the storm may have already produced a very weak tornado, they wondered if they should issue a warning. Interrogating cross sections and velocity up to 1.5°, Allen interpreted that the circulation still looked "kind of big." On PAR scan 01:31:13

UTC Allen pointed to the couplet on the northern storm (Fig. 3d) and said, "I like that."

Confused by the differing effect storm-relative motion (SRM) had on the appearance of the two storms, he said, "That's the southern one, isn't it?" Carl corrected him. Allen read off "26...27 against" at 0.9°. The circulation collocated with the reflectivity notch—weak, but strong enough—along with evidence of the expected recycling, apparently were sufficient to decide to warn. Over the next ~6.1 min, however, Allen struggled to place the dot and set the storm motion. Just as he finished setting a path he liked, PAR updated to 1:40:08 UTC (Figs. 2g and 3) and he said, "Oh, now it looks like crap again," noting it had "lost its kidney" shape. Had this team not encountered software issues, they would likely have issued a warning about 2 min after their decision (the median time to issue a warning during PARISE 2010 was 1:58), with about 6-min lead time. When the 0.5°, 1:44:35 UTC scan (Fig. 2h) came in, they were surprised to see a sudden strengthening in the circulation on the northern storm (Fig 3d). They quickly initiated WARNGEN and hastily issued a tornado warning about 40 s later (Fig. 3c).

Allen started off the 4.5-min team debriefing discussing how frustrations with software can affect warning decisions by stressing and wearing out the forecaster. He did not, however, think it made a difference here. In the end, he thought he probably saw the tight couplet he expected to see if one of these storms produced a tornado, though he saw it too late. During the case, just after the second tightening of the circulation prompted their successful warning, he had reflected, "welcome to tropical stuff."

Right from the start, the 43-s team focused on the northern storm, agreed upon what they saw, and exhibited a team dynamic of building upon each other's thoughts to help them zero in on the correct location and the correct threat. Each member's confidence rankings for the event type were very high (see Fig 4), and although individuals filled out confidence continuums separately, each person's self-ranking fit their performance during this case.

7.2 Warning Decision Comparison B

In this comparison the number and types of warnings issued by the 43-s and 4.5-min teams differed in the threats they perceived and how they incorporated environment. Nonradar cues were used; for the 43-s team these were the apparent desire to continue warnings from prior to the case start time and geographical considerations. The 4.5-min team was mainly concerned with missing "spikes" in the slower-update data. Neither team appeared very confident in the case. Only one person had some experience with tropical tornado events.

The 43-s team decided to issue their first warning early in the case (01:18:05 UTC) after Ken interrogated velocity from 0.5 to 2.4° (4.4–14 kft AGL) and realized there were no warnings in effect north of the Red River [a warning had been issued prior to the start time of the case]. He stated his concern that a velocity maxima the southern storm, with dominant inbound base velocities up to 50 kts, was an “indication of severe wind.” Wind up to 70 mph was the main threat mentioned in the 45-min severe thunderstorm warning; the polygon initially encompassed both supercells. During the following 20 min (1:23:31 – 1:43:31 UTC), Ken continuously interrogated the velocity and reflectivity signatures of both storms and became concerned with the consistent intensity of the north supercell (Fig. 3b). He also noticed that the existing “severe [polygon] is covering [only] half of that northern storm.” These appear to be the deciding factors that lead to issuing the second severe thunderstorm warning at 01:37:09 UTC for the northern supercell (Fig. 3a). Though this severe thunderstorm warning was issued with severe wind threat in mind, it verified with a 6-min lead time (Fig. 3a) due to the development of the EF1 tornado. This team was the only one of the six to issue severe thunderstorm warnings.

Just after a new scan arrived at 01:42:53 UTC (Fig. 3b), Ken’s interrogation of the north storm’s circulation revealed high velocities at 0.5°, “I’ve got 34 kts inbound now on this, 25 outbound." He then examined the reflectivity data and saw correspondence of the circulation with a notch signature. Interrogating through the next three scans, Ken first remarked, “usually at 35 kts you need to start considering tor.” Then further intensification of the circulation convinced Ken to issue a tornado warning. On scan 01:45:02 he read: “50 [kts] outbound. [pause] 41 inbound, 64 out…I don’t think I can ignore that." The tornado warning was issued at 01:47:28 UTC, but with a 3.15-min negative lead time; by then the tornado had likely dissipated.

In contrast, the 4.5-min team determined that the northern storm and short-lived tornadoes were their primary concerns. Their interrogation through the first several scans confirmed that the northern storm was better organized and stronger, though they noted 30 kts inbound on the southern storm in 01:22:19 UTC scan. Interrogating the 01:26:36 UTC scan they saw that the velocities had declined (Fig. 3d), but Lisa voiced concern that they were missing velocity trends due to slower updates by saying, "I bet you they're getting spikes in their data." The 01:31:13 UTC update continued to show circulations, however, and Ron had continued confidence that the northern storm was the one to focus on. In an attempt to seek validation, she asked for storm reports, but none were available. She then expressed her overall discomfort with the case: "I don't see them getting any stronger... they are about the same rotation-wise so the question becomes, do we have a strong enough environment that we are going to issue on these weak radar signatures?" Ron agreed it was a good question, then laughed and said, 'And I wish I had the answer" With interrogation of the 01:31:13 and 01:35:41 UTC scans showing continuity in the north storm's rotation (Fig. 3d), Lisa told Ron she was "anxious to pull [the] trigger because the environment was so good for tornadoes." Together, the radar signatures, environment, slower updates, and content of the Storm Prediction Center's mesoscale discussion led to the team's tornado warning decision. Their tornado warning, issued at about 01:39:40 UTC, verified with an approximate 4.6-min lead time (Fig. 3c).

During the debriefing, the 43-s team's description of their warning decision process conveyed a radar data-driven approach, though our analysis also revealed a nonradar cue for the first severe thunderstorm warning. Ken had stated, "We actually have no warnings in effect north of the Red River, so we need to warn... but so far that's [inbound velocity maxima] the only thing that would indicate [a warning is needed]." No other team expected severe, nontornadic winds given other aspects of analysis (e.g., storms did not produce hail). The second severe thunderstorm warning also had a nonradar component according to Ken: "Geography was 50% of the issue" —the storm had moved half way out of the polygon. Both members of this team reported higher confidence than normal in both their understanding of the event and in what they saw in the PAR data. These confidence rankings are interesting as they were the only team to issue severe thunderstorm warnings.

In contrast to the 43-s team, the 4.5-min team expressed the importance of storm environment, in addition to radar data, in their warning decision process. According to Lisa, though the situation on radar was below her normal warning decision threshold, she chose to warn due to the environment: she was worried they were missing tornadoes in-between sampling times. The confidence rankings for their understanding of the event were the only marks "less than usual" (Fig. 4), yet this was the only 4.5-min team that had positive lead time on a tornado warning. Neither had worked this type of event before. Both forecasters reported near usual confidence in what they saw in the PAR data, which simulated what they see in their normal operations.

7.3 Warning Decision Comparison C

In this comparison the difference between the 4.5-min and 43-s teams, in terms of number of warnings issued and lead time gained on one warning, appeared primarily due to differing tolerance for personal thresholds to be met given the tropical environment. The 43-s team expressed uncertainty at the strength of the early upward trends in the data and chose to warn to gain lead time. The 4.5-min team exhibited more confidence due to their knowledge of this type of event: Steve had a clear idea of the thresholds he wanted to see.

The 43-s team issued two tornado warnings (Fig. 3a), one for each storm. Upon beginning the case, Frank stated that SRM was needed to account for fast storm motion; he set it up. Using the high-temporal aspect of PAR to look for "any kind of feature there kind of rolling around the southwest side of the circulation," as well doing typical interrogation techniques of querying velocity values and examining cross sections, they found evidence for some focus in a circulation on the southern storm. Interested in achieving some lead time, Gina stated she "wouldn't mind" if they issued a tornado warning a bit early. Their tornado warning on the southern storm was issued at 01:23:31 UTC (Fig. 3a). In the following three minutes (through 1:26:23 UTC; Fig. 3b) they saw the southern storm weaken, and both laughed at the irony. They later updated this warning based on elapsed time and their belief the southern storm still posed a threat.

In the northern storm they identified a broad divergent appendage aloft. Frank noted "only 30 kts" at the lowest elevation of the 01:27:49 UTC scan. After changing SRM twice on the 1:29:15 UTC scan, first by tracking the southern storm circulation, then the northern one, they determined the southern storm was weakening and the northern storm was strengthening (Fig. 3b). The inflow region was identified on reflectivity, and a cross-section of SRM was used on both storms to determine both circulations remained "shallow." When the 01:31:24 UTC scan did not alleviate their uncertainty, Gina encouraged issuing another warning given the environment. They issued a tornado warning on the northern storm at 01:32:50 UTC (Fig. 3a). In a loop of reflectivity up to the 01:39:18 UTC scan, they saw a curling motion indicative of a rear flank downdraft, though an examination of a cross-section two scans later maintained their uncertainty: "Still looking the right way. Nothing extra special there." They briefly discussed discontinuing their warnings as the 01:41:27 UTC scan came in, but the next scan at 01:42:10 UTC showed a stronger couplet, increasing one scan later to 86 kts inbound and 26 kts outbound (Fig. 3b).

In contrast, the 4.5-min team immediately determined at the outset of the case that, "If we're going to be warning, we'll be warning on that one." Steve pointed to the northern storm in the velocity display. But that certainty did not come with eagerness to warn. The team examined a cross-section and scans in the lowest few elevations to identify a shallow circulation and divergence. When the 01:31:13 UTC scan came in they noticed it was "getting better as a couplet," with 14 kts inbound collocated with an appendage on reflectivity, about 2,500 ft above the ground (Figs. 2e and 3d). Steve shared his understanding of tropical mesocyclones: that 20 knots gate-to-gate could be sufficient for this case. In a loop of SRM up to 01:31:13 UTC, they saw that the southern storm had "looked interesting" at about 01:22:19 UTC. In the following scans they thought the northern cell was beginning to undergo a cell merger. Both storms seemed to be "having trouble getting their act together." They continued interrogation. When the 01:44:35 UTC scan came in Steve reacted with surprise (Figs 2h and 3d). Wendy laughed and Steve said, "That was quick!" They quickly issued a tornado warning on the northern storm (Fig. 3c).

During the individual team debriefing the 43-s team reported that they issued the first warning because they expected, given the environment, that the broad mesocyclone would tighten; and they specifically sought lead time. They said they were comfortable using the rapid-update data during the case because fluid motions matched their expectation of seeing circulations cycling up and down, and that increased their confidence in their long-lead time warnings. Their confidence was not high to begin with, however, and both marked lower confidence in both the event and in the data than the other team members (Fig. 4).

The 4.5-min team marked high confidence in both the event and data. They also appeared very confident during the case. The signature they eventually saw matched their expectations that it would suddenly appear in one volume scan, and be associated with a shallow circulation. Steve's empirically based threshold had not been met prior to that point. Steve acknowledged the tornado had probably dissipated by the time their warning was issued. During the joint debriefing, when able to see the 43-s data for the first time, Steve thought there were 3–4 scans prior to the shared 01:44:19 UTC scan he thought would have prompted him to warn.

8. DISCUSSION

Similar to this study, others have identified wide variation in individual forecaster decision making that confounded analyses of forecaster behavior (e.g., Hahn et al. 2003; Hoffman et al. 2006; Pliske et al. 1997). We found these variations were due primarily to three factors: forecaster experience, conceptual models, and confidence. Three additional factors identified in this pilot study were tolerance for potentially missing an event, perceived threats, and software issues. Since confounding factors exist in the real world, understanding them is important. It is neither possible nor realistic to reduce them to zero. The identification and understanding of confounding factors will allow us and others working in weather test beds to account for them in future experiments.

8.1 Experience

Forecasters' experience with tropical tornado events varied widely, from many events to none. All four in comparison A had experience with several events; further, one on each team had worked several hurricanes. In comparison B, one had worked a few events, one had worked just one event, and the remaining two had only worked tropical events with heavy rainfall and flooding. In comparison C one forecaster on each team had at least one experience with a tropical tornado event and one did not. Forecasters were directly asked about their experience during debriefings.

Variation in forecaster experience with different types of weather events is common within the NWS, due to both geographic differences in event climatology and tendency for forecasters to change offices for job advancement. Hence, the diversity in experience with tropical storms is likely similar to that found in the field. We expect that variation in experience and the relative lack of training in tropical tornadoes led, in part, to the variations seen in forecaster conceptual models.

8.2 Conceptual Models

Components of forecaster conceptual models were determined through analysis of the full data set: coding of case transcripts, determination of warning decision processes, and review of debriefing transcripts. These components included characteristics of circulations and associated reflectivity features sampled by radar, atmospheric environment, and implications of update time for detection (Table 1).

Concerning circulation characteristics, though all forecasters expected circulations to be shallow and collocated with a notch in reflectivity, their ideas about circulation strength sufficient to warrant a tornado warning differed (Table 1). Three forecasters specifically mentioned threshold inbound/outbound velocities, while the rest did not. Instead, they mentioned circulations would be weaker than those typical of nontropical supercells. There are no specific training requirements, nor guidance from research for tropical cyclone warning decision making (LaDue, J. 2011, personal communications); these may explain the wide variability in circulation strength these forecasters appeared to use in their warning decisions.

During the case, all forecasters with 4.5-min data mentioned they were concerned about missing “spikes” or intensification in circulations indicative of tornado occurrence. This concern was driven by their anticipation that tornadoes forming in a tropical environment would be short-lived. The implication of a longer update time appears to have directly affected two of the 4.5-min teams' warning decisions (aborted warning in comparison A, and comparison B). Since sampling times range from 4–6 min in the NWS, concern felt about missing important storm development is likely a common forecaster experience in similar operational situations. This sampling constraint is also found in LaDue et al. (2010). The experiment design may have amplified the impact of sampling time because teams had examined higher-temporal data for two days prior to this experiment.

Most forecasters stated that environment was a factor in their warning decisions; the exception was the 43-s team in Comparison B (Table 1). These latter forecasters also did not mention the need to use a weaker circulation strength threshold (Table 1). As discussed previously, the environment prompted one 4.5-min (Comparison B) and one 43-s team (Comparison C) to issue tornado warnings though the velocity signatures were not as tight or strong as desired.

8.3 *Confidence in PAR data vs event*

Figure 4 shows how numeric values were assigned to confidence ratings. The average of 43-s team members' confidence scores for what they saw in the rapid scan PAR data was more than usual (+0.69), and slightly higher than their average confidence in their understanding of the event (+0.65). Similarly, but with lower numbers, the 4.5-min team members were also more confident than usual in what they saw in the degraded-scan PAR data (+0.37), and in their understanding of the event (+0.29). The median of 43-s team members' scores were nearly the same, and high, for both the data (+0.78) and the event (+0.76). In contrast, the median score for 4.5-min team members' confidence in what they saw in the data was near usual (+0.07), as would be expected given that degraded PAR simulated a WSR-88D volume scan. The median confidence score for 4.5-min team members' understanding of the event was high (+0.62), but lower than the median for the 43-s team members. These numbers suggest that event type may have had a slightly larger effect on forecaster confidence than the difference in temporal frequency, underscoring event type as an important factor influencing results in this type of study.

8.4 Additional Confounds

The three remaining confounds— tolerance for potentially missing an event, perceived threats, and software issues— have varying implications. The first is the most difficult. We speculate that forecasters are accustomed to sometimes choosing to warn early because they know radar sampling is insufficient for good detection of some events they are concerned about. Our hope is that once forecasters gain confidence in higher-temporal-resolution data, their need to account for missing information will diminish.

The second two can be more easily mitigated. Our pilot was unrealistic in that we allowed forecasters to do their own environmental assessment without any kind of briefing. In a real situation, a forecaster would have been briefed by whomever he or she was relieving for warning operations. Finally, AWIPS-II will be capable of handling PAR data. Experiments like this one will soon be able to use the same software forecasters are accustomed to.

9. CONCLUSIONS

The case examined in this study revealed some of the complexities of NWS forecasters' warning decision processes, including radar and nonradar factors that impacted their decisions. The warning decision comparisons described in section 7 show that though teams examined similar reflectivity and velocity signatures in the two supercell storms, they came to different conclusions about whether and when to warn. Differences appeared due to their experience, conceptual models, confidence, tolerance for potentially missing the event, perceived threats, and software issues.

As a group, the 43-s teams issued 50% more warnings than the 4.5-min teams (Fig. 3a, c). For the 43-s teams, 3 of the 6 warnings verified (2 tornado and 1 severe), 2 warnings were false alarms (tornado and severe on south storm), and one tornado warning was a miss (issued seconds after the tornado likely dissipated). Tornado-warning lead times were 18.6 and 11.5 min, and the severe-warning lead time was 6 min. In contrast, all three tornado warnings issued by the 4.5-min teams verified, though warning times were shorter: 4.6 and 0 min (2 teams). A 0-min lead time is assigned by the NWS to warnings issued with negative lead time, but while the tornado is in progress. Computed lead times for these two cases were -.07 and -1.6-min. If not for a software issue, nonzero tornado lead-time would have been achieved by two 4.5-min teams, rather than one, with values of 4.6 min (Fig. 3a) and 5.6 min (see introduction to Section 7). Although a small sample, these numbers are suggestive of a positive effect of the higher temporal data. They may also suggest that simply being in an experiment could have made forecasters more willing to warn, expecting that something was likely to happen. To more fully understand the variation of how PAR data might impact operations, a larger and more diverse set of cases (including nulls) is needed.

ACKNOWLEDGMENTS

A portion of the funding for this work was provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA08OAR4320904 and NA17RJ1227. This work was made possible by the implementation of advanced signal processing techniques and software development by Ric Adams, Chris Curtis, Eddie Forren, Igor Ivić, David Priegnitz, Sebastian Torres, John Thompson, and David Ward of OU CIMMS. The meteorological software displays and data sets were provided by Charles Kerr and Kevin Manross of OU CIMMS, and Darrel Kingfield and Mark Sessing of the Warning Decision Training Branch. The recruitment of and infrastructure for participants was provided by the Experimental Warning Program leadership of Greg Stumpf and Travis Smith. We thank Kim Klockow, Heather Moser, and Rachel Riley for the meticulous notes they took during the experiment, and our gratitude goes to Rick Hluchan, Jennifer Newman, and Adam Smith for their interactions with visiting forecasters. Following the experiment, Natalie

Daab's careful transcription of the data and James Murnan's improvements to video audio quality were crucial to our analysis. Harold Brooks help with plotting time series images in EXCEL was priceless. We appreciate thought-provoking conversations with Hank Jenkins-Smith and Carol Silva of the Center for Social Science Research at the University of Oklahoma to help us look forward.

REFERENCES

Arndt, D. S., and Co-Authors, 2009: Observations of the overland intensification of tropical storm Erin (2007). *Bull. Amer. Meteor. Soc*., **90**, 1079–1093.

Bernard, R., 2006: *Research Methods in Anthropology: Qualitative and Quantitative Approaches*. 4th ed. Alta Mira Press, 803 pp.

Boyatzis, R.E., 1998: *Transforming Qualitative Information: Thematic Analysis and Code Development.* Sage Publications, 184 pp.

Brotzge, J., S. Erickson, 2010: Tornadoes without NWS warning. *Wea. Forecasting*, **25**, 159–172.

Burgess D., R. J. Donaldson, T. Sieland, J. Hinkelman, D. Sirmans, K. Shreeve, K. Glover, I. Goldman, 1979: Final report on the Joint Doppler operational project (JDOP) 1976–1979. NOAA Tech. Memo. ERL NSSL86, 84 pp.

Evans, C., R. S. Schumacher, T. J. Galarneau, Jr., 2011: Sensitivity in the overland reintensification of tropical cyclone Erin (2007) to near-surface soil moisture characteristics. *Mon. Wea. Rev.*, in press.

Hahn, B. B., E. Rall, and D. W. Klinger, 2003: Cognitive Task Analysis of the Warning Forecaster Task, 15 pp. [Available from Klein Associates Inc, 1750 Commerce Center Blvd., North Fairborn, Ohio 45324-6362.

Heinselman, P. L., D. L. Priegnitz, K. L. Manross, T. M. Smith, and R. W. Adams, 2008: Rapid sampling of severe storms by the National Weather Radar Testbed Phased Array Radar. *Wea. Forecasting*, **23**, 808–824.

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.,* ***5*0**, 579–593

Hoffman, R. R., J. W. Coffey, K. M. Ford, and J. D. Novak, 2006: A method for eliciting, preserving, and sharing the knowledge of forecasters. *Weather and Forecasting*, **21,** 416–428.

Joslyn, S. and D. Jones, 2008: Strategies in Naturalistic Decision Making: A Cognitive Task Analysis of Naval Weather Forecasting. *Naturalistic Decision Making and Macrocognition*, J. M. Schraagen, L. G. Militello, T. Ormerod, and R. Lipshitz, Eds., Ashgate Publishing Co., 183–201.

LaDue, D. S., P. L. Heinselman, 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.

Lakshmanan, V., T. Smith, G. Stumpf, K. Hondl, 2007: The warning decision support system–integrated information. *Wea. Forecasting*, **22**, 596–612.

Mertens, D. M., 2005: *Research and Evaluation in Education and Psychology*. 2nd ed. Sage Publications, 507 pp.

Newman, J. F., and P. L. Heinselman, 2011: Evolution of a quasi-linear convective system sampled by phased-array radar. Extended abstracts, *Interactive Information Processing Systems (IIPS)*, Amer. Meteor. Soc., Seattle, WA, 13B.5.

Newman, J. F.**,** V. Lakshmanan**,** P. L. Heinselman**,** T. M. Smith**,** 2011: Range correction for radar-derived azimuthal shear: applications to a tornado detection algorithm. *Extended Abstracts, 27th Conference on Interactive Information Processing Systems (IIPS)*, Seattle, WA, USA, AMS, 8B.4.

NOAA, 2006: Doppler radar meteorological observations. Part C: WSR-88D products and algorithms, Chapter 5. Federal Meteorological Handbook, FCH-H11C-2006, Office of the Federal Coordinator for Meteorological Services and Supporting Research, Rockville, MD, 23 pp.

OFCM, 2006: Federal research and development needs and priorities for phased array radar. Rep. FMC-R25-2006, Interdepartmental Committee for Meteorological Services and Supporting Research, Committee for Cooperative

Research Joint Action Group for Phased Array Radar Project, 62 pp. [Available online at www.ofcm.gov/r25-mpar/fcm-r25.htm.]

Ortega, K. L., T. M. Smith, K. L. Manross, A. G. Kolodziej, K. A. Scharfenberg, A. Witt, J. J. Gourley, 2009: The Severe hazards analysis and verification experiment. *Bull. Amer. Meteor. Soc.*, **90**, 1519–1530.

Pliske, R., D. W. Klinger, R. Hutton, B. Crandall, B. Knight, and G. Klein, 1997: Understanding Skilled Weather Forecasting: Implications for Training and the Design of Forecasting Tools. Final Contractor Report - January 1996-March 1997 AL/HR-CR-1997-0003, 122 pp.

Schraagen, J. M., L. G. Militello, T. Ormerod, and R. Lipshitz, Eds., 2008: *Naturalistic Decision Making and Macrocognition*. Ashgate Publishing Co., 411 pp.

Scharfenberg, K. A., and Coauthors, 2005: The Joint Polarization Experiment: Polarimetric radar in forecasting and warning decision making. *Wea. Forecasting*, **20**, 775–788.

Singleton, R. A. and B. Straits. 2005. *Approaches to Social Research.* 4th ed. Oxford University Press, 640 pp.

Smith, T. M., and K. L. Elmore, 2004: The use of radial velocity derivatives to diagnose rotation and divergence. Preprints, 11th Conf. on Aviation, Range and Aerospace. Amer. Meteor. Soc., Hyannis, MA, P5.6.

Steadham, R., 2008: 2008 National Weather Service field study. Part 1: Volume coverage pattern usage. Radar Operations Center, Norman, OK, 28 pp. [Available from WSR-88D Radar Operations Center, 1200 Westheimer Dr.,Norman, OK 73069]

Yu T.-Y., M. B. Orescanin, C. D. Curtis, D. S. Zrnić, and D. E. Forsyth, 2007: Beam multiplexing using the phased-array weather radar. *J. Atmos. Oceanic Technol*., **24**, 616–626.

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

Table 1. Components of forecaster conceptual models and percent of forecasters sharing each.

	43-s Team	**4.5-minTeam**
Weaker Couplet Strength	**66%**	**83%**
Trend in Circulation Strength	100%	100%
Update Time Detrimental	**0%**	**100%**
Environment	**66%**	**100%**
Shallow Circulation	100%	100%
Reflectivity Notch	100%	100%

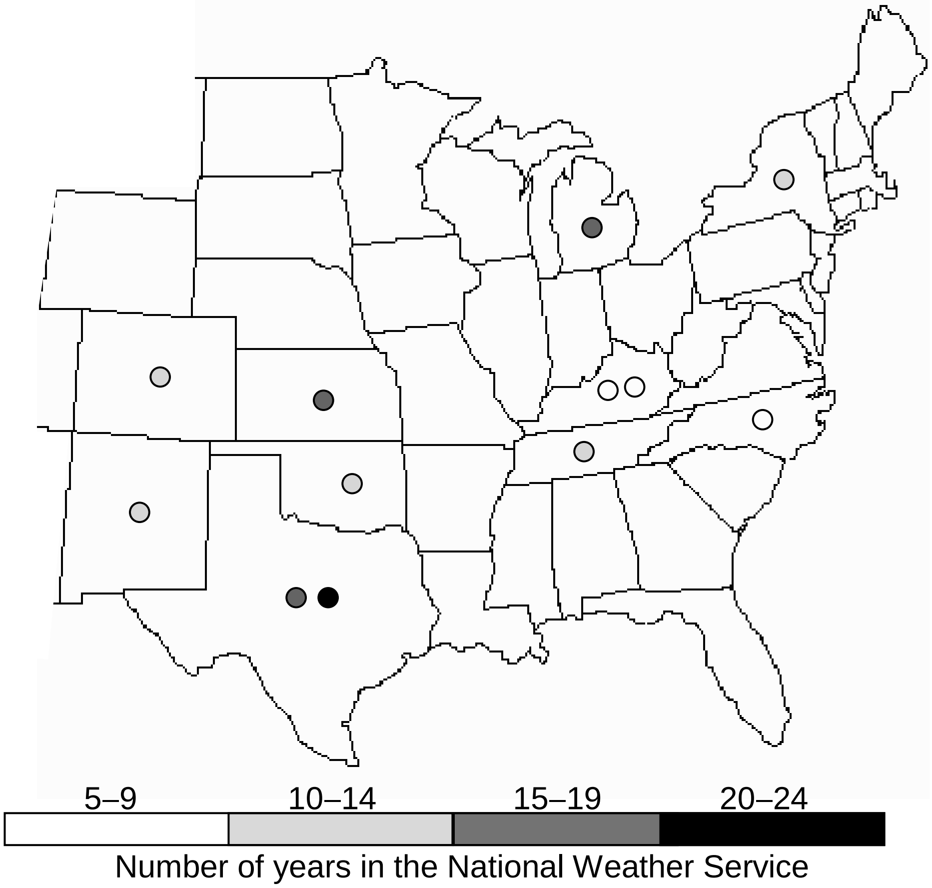

Fig. 1. Circles indicate the states, but not the specific city, where participants' home offices were located. The number of years employed in the NWS is color coded according to the grey scale shown.

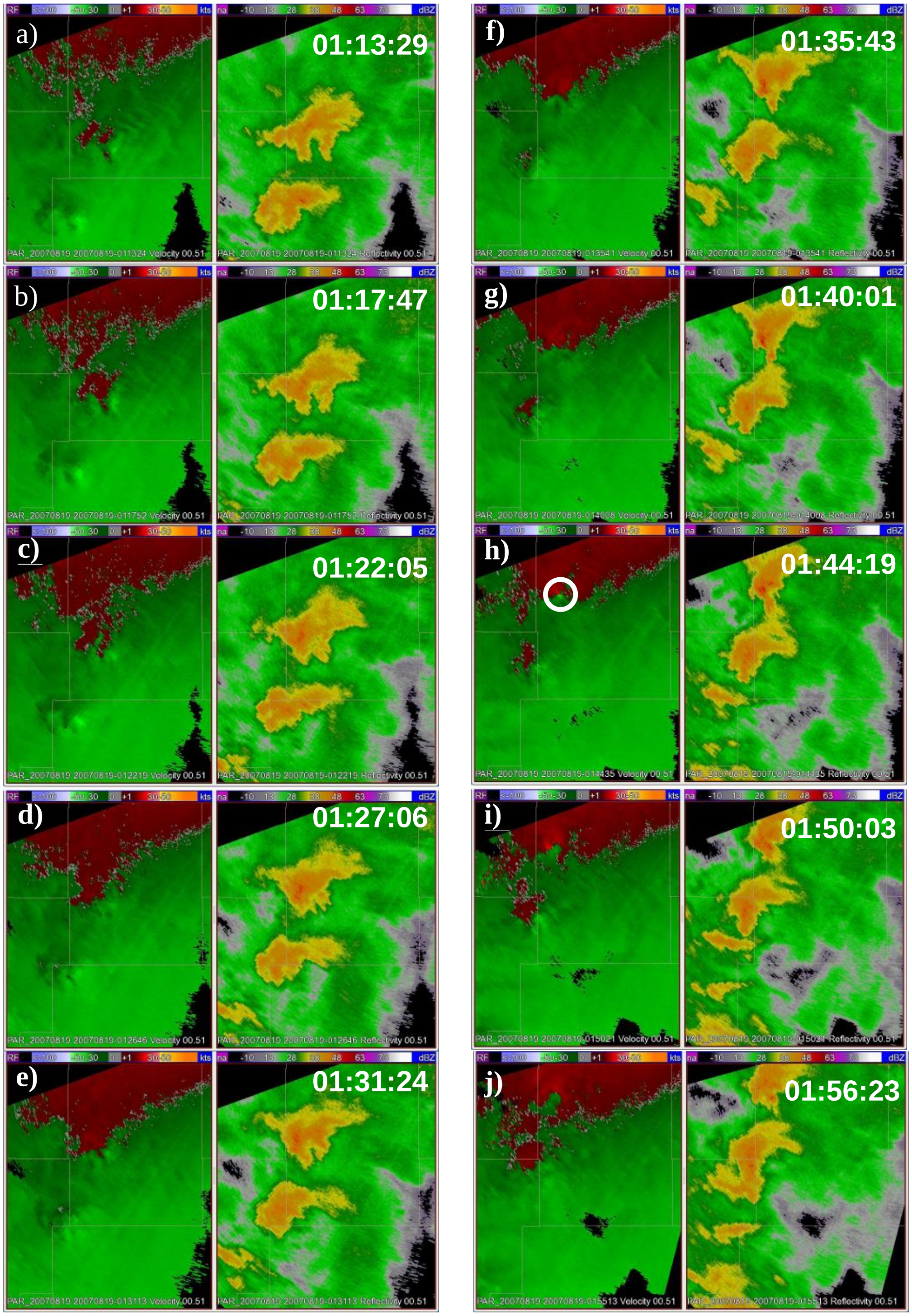

Figure. 2. Time series of 0.5° velocity (left) and reflectivity (right) from 0113–0154 UTC 19 August 2007. The NWRT PAR is located in the direction of the upper-right-hand-corner and the radar range in the lower left-hand-corner is ~113 km. The TVS associated with the EF1 tornado is enclosed by a white circle.

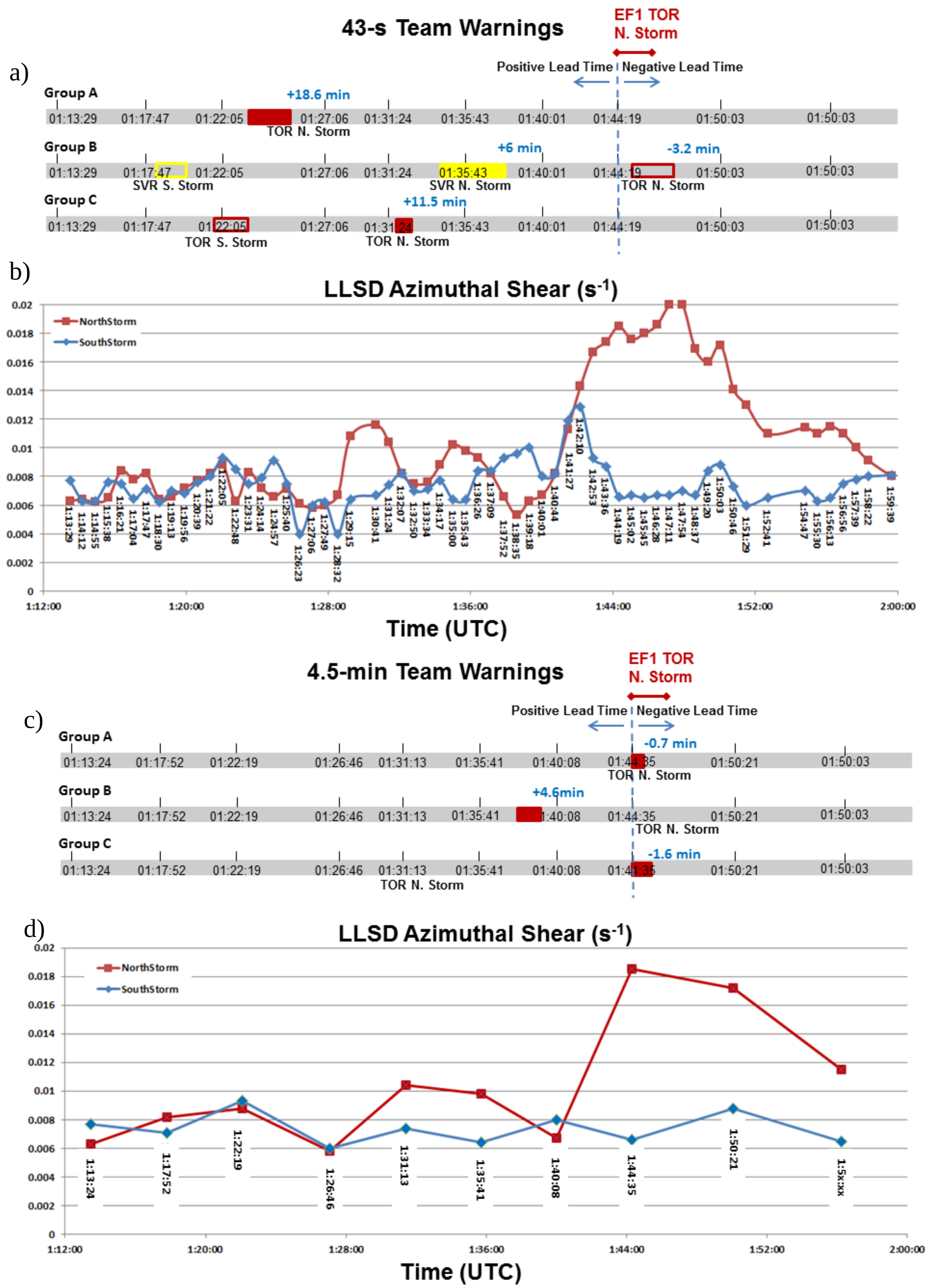

Figure 3. Time series of warning decisions and LLSD azimuthal shear associated with 43-s and 4.5-min teams. Left-side of rectangle is when WarnGen was opened and right-side is when a warning was issued; lead time is relative to tornado occurrence. Warnings issued for tornadoes are red, whereas those for severe thunderstorms are yellow. Unfilled-rectangles indicate unverified warnings. Time stamps on 4.5-min data are aligned with corresponding 43-s times.

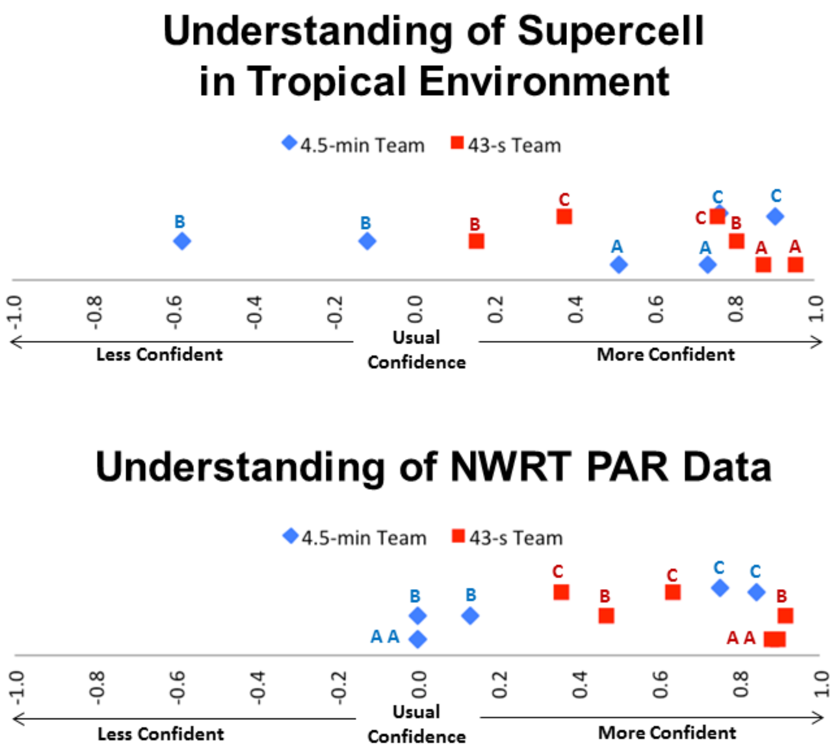

Figure 4. Normalized confidence ratings of forecasters relative to normal operations in their offices. Numeric values were assigned after the fact; forecasters saw "less confident" at the left end, "usual" in the middle, and "more confident" at the right. Top figure is their confidence in how well they understood the event: supercells in a tropical environment; bottom figure is their confidence in how well they understood supercell signatures and evolution using the NWRT PAR data.