Forecaster Performance and Workload: Does Radar Update Time Matter?

KATIE A. WILSON

*School of Meteorology, and Cooperative Institute for Mesoscale Meteorological Studies,
University of Oklahoma, Norman, Oklahoma*

PAMELA L. HEINSELMAN

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

CHARLES M. KUSTER AND DARREL M. KINGFIELD

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

ZIHO KANG

School of Industrial and Systems Engineering, University of Oklahoma, Norman, Oklahoma

ABSTRACT

Impacts of radar update time on forecasters' warning decision processes were analyzed in the 2015 Phased Array Radar Innovative Sensing Experiment. Thirty National Weather Service forecasters worked nine archived phased-array radar (PAR) cases in simulated real time. These cases presented nonsevere, severe hail and/or wind, and tornadic events. Forecasters worked each type of event with approximately 5-min (quarter speed), 2-min (half speed), and 1-min (full speed) PAR updates. Warning performance was analyzed with respect to lead time and verification. Combining all cases, forecasters' median warning lead times when using full-, half-, and quarter-speed PAR updates were 17, 14.5, and 13.6 min, respectively. The use of faster PAR updates also resulted in higher probability of detection and lower false alarm ratio scores. Radar update speed did not impact warning duration or size. Analysis of forecaster performance on a case-by-case basis showed that the impact of PAR update speed varied depending on the situation. This impact was most noticeable during the tornadic cases, where radar update speed positively impacted tornado warning lead time during two supercell events, but not for a short-lived tornado occurring within a bowing line segment. Forecasters' improved ability to correctly discriminate the severe weather threat during a nontornadic supercell event with faster PAR updates was also demonstrated. Forecasters provided subjective assessments of their cognitive workload in all nine cases. On average, forecasters were not cognitively overloaded, but some participants did experience higher levels of cognitive workload at times. A qualitative explanation of these particular instances is provided.

1. Introduction

During convective warning operations, National Weather Service (NWS) forecasters rely primarily on weather radar to monitor storms and make warning decisions. The Weather Surveillance Radar-1988 Doppler (WSR-88D) network currently provides forecasters with volumetric updates every 4–6 min. However, given that phased-array radar (PAR) may become the next generation of weather radar, this technology is being tested and considered for weather applications (Forsyth et al. 2005; Zrnić et al. 2007). Located in Norman, Oklahoma, the National Weather Radar Testbed PAR (hereafter PAR) demonstrates how electronic beam steering can be used to adaptively scan the atmosphere and collect rapid-update (~1 min) volume scans of a 90° azimuthal sector (Heinselman and Torres 2011).

In a continued effort to improve the timeliness and accuracy of warnings, it is vital that the potential impacts of higher temporal resolution radar data on NWS forecasters' warning decision processes are understood. Since

2010, the Phased Array Radar Innovative Sensing Experiment (PARISE) has been addressing a variety of research questions to examine this issue (Heinselman et al. 2012, 2015; Bowden et al. 2015; Bowden and Heinselman 2016). Applications of behavioral science methods (e.g., cognitive task analysis) have resulted in a better understanding of forecasters' thought processes as they interrogate radar data and make warning decisions. This analysis has provided important insight into aspects of forecasters' performance, such as lead time and verification, which have been a consistent focus throughout PARISE. Impacts of 1-min PAR updates on forecasters' performance during a variety of scenarios were assessed in the 2010, 2012, and 2013 PARISEs.

The 2010 PARISE focused on a known challenge within the NWS: being able to provide warning lead time on weak, short-lived tornadoes. Comparing forecasters' decisions when using 43-s versus 4.5-min volumetric PAR updates, this experiment found that participants using faster updates achieved longer tornado warning lead times (Heinselman et al. 2012). However, forecasters using these faster updates also had a higher false alarm ratio (FAR). Because of the small sample size in the first experiment and the concern that faster PAR updates could lead to a higher number of false alarms, the experimental design was modified in the 2012 PARISE, and the number of cases that participants worked was increased (Heinselman et al. 2015). This time, forecasters worked a total of four events (two tornadic and two nontornadic) independently, each with 1-min updates. The participants achieved a median tornado warning lead time of 20 min, which exceeded the EF0/EF1 tornado warning lead time of the participants' respective forecast offices (7 min) and NWS regions (8 min) (Heinselman et al. 2015). All but one forecaster also achieved a probability of false alarm score < 0.5, indicating that warning accuracy was better than chance during this experiment (Heinselman et al. 2015).

Although the 2010 and 2012 PARISE results demonstrated positive impacts of higher temporal resolution radar data on forecasters' warning decisions during weak tornado events, a question that remained was whether the same benefits would be observed during events that only produced severe hail and/or wind. The 2013 PARISE aimed to answer this question using a two-independent-group design, such that half of the participants were assigned to a control group (5-min updates) while the other half were assigned to an experimental group (1-min updates). Performance of the experimental group during these cases was superior to that of the control group, as demonstrated by their statistically significant longer median warning lead time

(21.5 min) compared with that of the control group (17.3 min), and their more accurate warning decisions (Bowden et al. 2015).

Previous PARISE studies have contributed substantially to our understanding of the potential impacts of higher temporal resolution radar data on forecasters' warning decision processes. However, there have been some key limitations preventing the generalizability of our findings about forecasters' performance. The most notable limitation is the sample size; in each PARISE, only 12 forecasters were recruited for participation and only one to four cases were worked. In each experiment, these cases focused on a specific weather threat (i.e., weak tornado or severe hail/wind), and as a result they did not provide the variety of weather events typical in a forecast office. Furthermore, while impacts of 1- and 5-min PAR updates have been explored, we have not assessed how forecasters would perform with 2-min PAR updates. Finally, forecasters' cognitive burden resulting from a greater influx of data was not examined in these previous experiments, and therefore the effects of rapidly updating PAR data on forecasters' cognitive workload was still unknown.

The 2015 PARISE was therefore designed to address these limitations, while continuing to deepen our understanding of forecasters' warning decision processes and target new research questions. Based on findings from previous experiments, we expected forecasters with faster PAR updates to perform better, most notably with respect to warning lead time. We also expected forecasters with faster PAR updates to discriminate between weather threats more successfully. Given that forecaster cognitive workload had not been studied in detail in the literature, we were hopeful that our assessment would provide new insight into forecasters' mental efforts during warning operations. Our expectation was that faster PAR updates would lead to increased cognitive workload, especially during more demanding weather scenarios. In this paper, we provide an overview of the experimental design and methods applied in the 2015 PARISE. We focus our analysis on how forecasters' performance, warning characteristics, and perceived cognitive workload relate to the temporal resolution of radar data and the type of weather threat presented in each case. Finally, we bring together findings from this most recent study and from previous studies to give an overall assessment of what higher temporal resolution radar data will mean for NWS forecasters during warning operations.

2. Methodology

The 2015 PARISE took place over 6 weeks during August and September 2015. Each week, five NWS
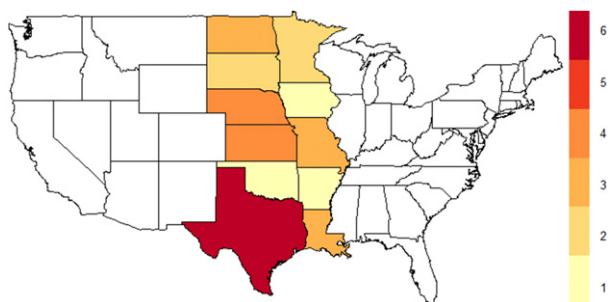
Fig. 1. Forecasters were recruited from the Great Plains region of the United States. The color bar indicates the number of forecasters participating from each of the 10 states.

forecasters visited the NOAA Hazardous Weather Testbed in Norman and completed three experimental components of this study. These components were the traditional experiment, eye-tracking experiment, and focus group. The traditional experiment built directly on earlier PARISE studies, aiming to improve the generalizability of PARISE findings through increased sample sizes of participants and cases worked. Additionally, the traditional experiment explored the concept of cognitive workload for the first time in PARISE. This paper discusses findings from the traditional experiment only.

a. Recruitment

Thirty NWS forecasters were recruited for the 2015 PARISE. Since forecasters would be working archived weather events from central Oklahoma, those most likely to have encountered similar storm types during their own warning operations were targeted. The 30 participating forecasters represented 25 NWS Weather Forecast Offices located across 10 states in the Great Plains (Fig. 1). Of these forecasters, 5 were female and 25 were male, and experience ranged from 1 to 27 yr (mean = 12 yr, standard deviation = 7 yr). Prior to participating in this study, all forecasters completed a multiple-choice survey that comprised 48 questions drawn from forecaster training material designed by the NOAA's Warning Decision Training Division. This survey queried forecasters' knowledge of severe weather definitions and their understanding of conceptual models and weather radar. The purpose of this survey was to obtain a simplistic assessment of forecasters' general knowledge of severe weather warning operations, which when represented as survey scores, could be used as a measure for comparison. The survey scores ranged from 28 to 41 out of a possible 49 points (mean = 36, standard deviation = 3).

b. Experimental design

A goal of the 2015 PARISE was for all forecasters to work a variety of weather events and to be exposed to a variety of temporal resolutions of PAR data. In comparison, each previous PARISE study was confined to a single type of weather (i.e., weak tornado events only or severe hail and wind events only), and forecasters were assigned to work with only 1- or 5-min PAR volumetric updates (Heinselman et al. 2012, 2015; Bowden et al. 2015). This current study continued the assessment of forecaster use of 1- and 5-min PAR volumetric updates, but based on forecasters' suggestions during the 2013 PARISE, also tested forecasters use of 2-min PAR volumetric updates (Bowden and Heinselman 2016).

To examine forecaster use of these three temporal resolutions (full speed, ~1 min; half speed, ~2 min; and quarter speed, ~5 min) for different types of weather events, nine archived PAR cases were selected (see section 3). The chosen experimental design required random assignment of forecasters to three separate groups, and each group comprised 10 forecasters. Group assignment determined the temporal resolution of PAR data that would be used for each case, and all participants were exposed to the full-, half-, and quarter-speed PAR updates for each of the three case types.

c. Methods

1) WORKING EVENTS

The majority of forecasters' participation time was spent on the traditional experiment. Forecasters worked on two–three cases per day, and the nine cases were completed in random order to avoid any order effect. Forecasters were provided with their own AWIPS-2 workstations and worked each case independently. They did not discuss details of the weather events with other participants until the end of the week. First, a practice case was completed to train forecasters on how to set up their cases and to ensure that they were comfortable loading and interrogating PAR data in AWIPS-2. During this initial case, forecasters practiced issuing warnings using the Warning Generation (WarnGen) software, practiced receiving storm reports, and personalized settings in AWIPS-2.

Similar to previous PARISE studies, prior to working each case forecasters viewed a prebriefing video that described the environmental conditions associated with the upcoming case. Mesoscale analysis, sounding information, and satellite and radar data were provided, and forecasters used this information to form and document their expectations for how the event might unfold. When working the case, forecasters were able to view reflectivity, velocity, and spectrum width products in simulated real time. Importantly, forecasters were asked to work the event in their normal forecasting style, and to interrogate the radar data and issue special weather statements, warnings (severe thunderstorm and

TABLE 1. The ISA tool adapted from Kirwan et al. (1997).

Level	Workload	Spare capacity	Description
1	Underutilized	Very much	Nothing to do; rather boring
2	Relaxed	Ample	More time than necessary to complete tasks; time passes slowly
3	Comfortable	Some	The controller has enough work to keep him/herself stimulated; all tasks are under control
4	High	Very little	Certain nonessential tasks are postponed; could not work at this level very long; the controller is working "at the limit"; time passes quickly
5	Excessive	None	Some tasks are not completed; the controller is overloaded and does not feel in control

tornado), and severe weather statements that they deemed necessary. All issued products were recorded in a database for performance analysis.

2) WORKLOAD RATINGS

With an increase in data availability, the impact of higher temporal resolution radar data on forecasters' workload is of interest. Workload is defined as the level of attention resources required to meet the performance criteria and is affected by task demands and past experience (Young and Stanton 2006). Widely used workload assessment methods are the NASA Task Load Index (NASA-TLX; Hart and Staveland 1988) and the Subjective Workload Assessment Technique (SWAT; Reid et al. 1981); however, both methods evaluate workload based on subclassifications such as time demand, effort demand, and stress demand, which can be time consuming and obtrusive when workload needs to be evaluated many times during a prolonged task. Furthermore, given that forecasters' work demand is predominantly cognitive, many of these subclassifications are difficult for forecasters to relate to. Thus, a faster, less obtrusive, and more suitable method was chosen. This method was the instantaneous self-assessment (ISA; Kirwan et al. 1997), which is based on a unidimensional scale and has five qualitative ratings of mental effort, including 1) underutilized, 2) relaxed, 3) comfortable, 4) high, and 5) excessive (Miller 2001). Each level of mental effort was provided with a corresponding description. The ratings can also be thought of in terms of how much spare mental capacity one has (Table 1) (Kirwan et al. 1997). To capture variations in forecasters' mental workload during events, ISA ratings were collected during a video-cued retrospective recall at 5-min intervals. Along with each rating, forecasters provided reasoning for their chosen mental workload level.

3. Radar data

For the 2015 PARISE, the nine cases selected from archived PAR data maximized the variety in storm types, hazard types (e.g., severe hail), and distance from the radar. Each case also met temporal continuity (i.e., no data gaps) and duration criteria, which allowed forecasters ample time to demonstrate their warning decision process in each case. Following these criteria, we selected three null cases, three severe hail and wind cases, and three tornado cases based on storm reports provided by the National Centers for Environmental Information's *Storm Data* publication (NCEI 2016).

Of the three null cases, two (Alpha and Epsilon) were multicell thunderstorms that produced no severe weather reports (Figs. 2a,b; Table 2). The third case (Theta) was considered null with respect to tornadoes. It contained two nontornadic supercells, but the supercell located about 75 km from the radar produced severe hail (Fig. 2c). In all three severe hail and wind events (Delta, Gamma, and Beta), a multicell thunderstorm produced severe weather. In Delta, a storm produced both severe hail and wind, while storms in Gamma produced severe hail only and storms in Beta produced severe wind only (Table 2; Fig. 3). Of the three tornadic cases, Zeta contained a classic supercell that produced two tornadoes (one rated EF1 and the other rated EF2), Iota contained a supercell cluster that produced a tornado rated EF0, and Eta contained a tornadic squall line that produced a tornado rated EF1 (Fig. 4). The supercells in Zeta and Iota also produced severe hail and wind (Table 2).

In all but one of the cases (Alpha), PAR operators collected data using a modified volume coverage pattern (VCP 12; Brown et al. 2005) that included five additional elevation angles above 19.5° (up to 52.9°). For Alpha, a unique VCP with 22 elevation angles between 0.51° and 52.94° was used. The Adaptive Digital Signal Processing Algorithm (ADAPTS; Heinselman and Torres 2011) was also used in all but three cases (Beta, Iota, and Alpha), which resulted in volumetric update times that varied throughout the cases (Table 2).

4. Storm-based warning verification

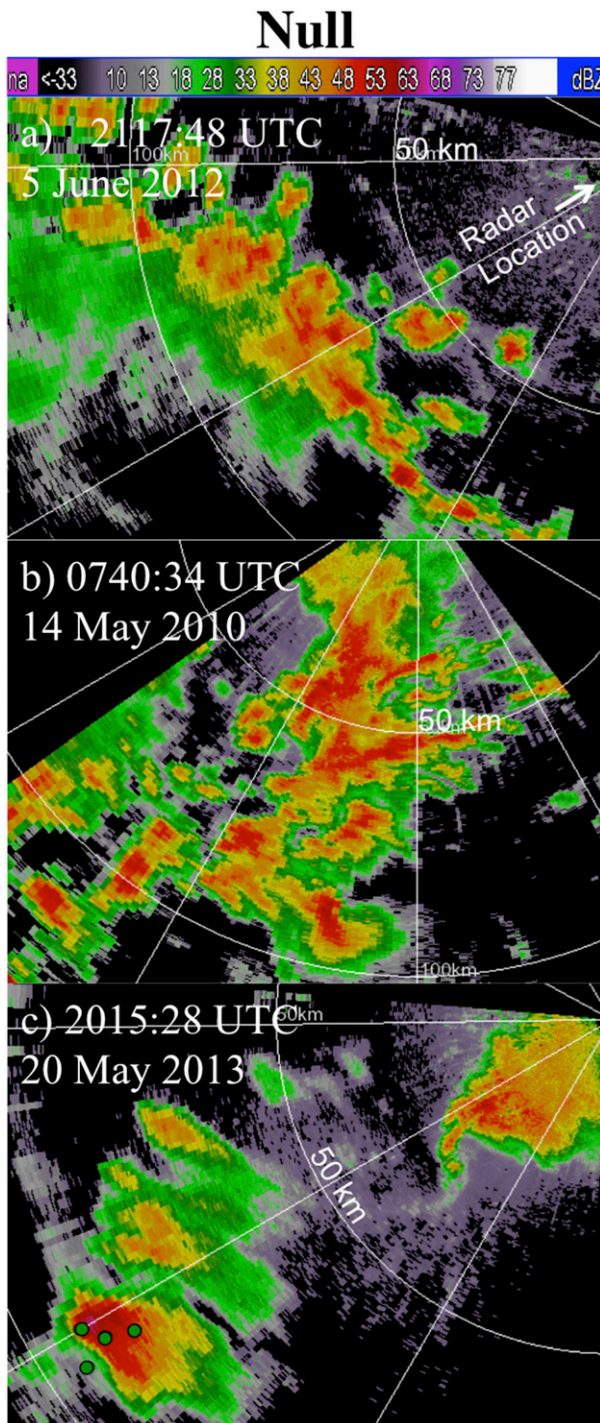Recent PARISE experiments have focused on hazard-specific, storm-based warning verification of

FIG. 2. PAR 0.5° reflectivity for (a) Epsilon, (b) Alpha, and (c) Theta (null tornado case). Green dots in (c) are severe hail reports. Reflectivity (dBZ) color bar is located at the top. White rings are displayed in 50-km increments.

either tornadoes (Heinselman et al. 2012, 2015) or severe hail and winds (Bowden et al. 2015). In the 2015 PARISE, all three hazard types occurred in several of the simulation scenarios, requiring a verification framework for both severe thunderstorm and tornado warnings. As part of NWS Instruction 10-1601 (NWS 2015), two methods are used to verify these convective warnings: event specific and generic (Table 3). In the event-specific verification system, severe thunderstorm warnings are verified only by convective wind or hail events, and tornado warnings are verified only by tornado events. Because these matching hazard-to-warning combinations are only used to calculate hits and lead times, forecasters are neither rewarded nor penalized when an unmatched hazard-to-warning combination occurs. In the generic verification system, any convective hazard occurring in any warning type verifies the warning and allows for a lead time to be calculated for the hazard. Therefore, the generic verification system results in the possibility that a severe hail or wind event can verify a tornado warning and a tornado event can verify a severe thunderstorm warning.

For the above reasons, we decided to develop a hybrid verification system that adds certain components of the generic verification system to the event-specific verification system (Table 3). In this hybrid system, convective wind or hail events occurring within a tornado warning have their lead times calculated and count as a hit, but do not verify the warning. Wind or hail events occurring within a severe thunderstorm warning verify the warning and have event lead times tabulated, as they normally would. Tornado events occurring within severe thunderstorm warnings count as misses and do not verify the warning, with the opposite results occurring within a tornado warning. Our system allows for all events and warnings to be scored for each simulation but is stricter regarding tornado warning issuance and verification. In conjunction with the proposed hybrid verification system, we used the guidance within NWS Instruction 10-1601 (NWS 2015) to calculate the probability of detection (POD), FAR, and lead times for all warnings and hazards.

5. Performance

The expectation that the overall median warning lead time would increase as the update speed increased (became faster) was realized in this study. The use of full-, half-, and quarter-speed PAR data resulted in overall median warning lead times of 17, 14.5, and 13.6 min, respectively. Despite some difference in the median warning lead times, application of the Kruskal–Wallis test (Kruskal and Wallis 1952) showed no statistically significant differences between the three groups (p value = 0.1683).

TABLE 2. Information about selected cases. Three case types are null, severe hail and/or wind, and tornadic. EG is estimated gust, and MG is measured gust.

Case	Time and date	Volume update time (s)	Storm type	Storm report
Null cases				
Epsilon	2059–2139 UTC 5 Jun 2012	55–60	Multicell	None
Alpha	0724–0756 UTC 14 May 2010	65	Multicell	None
Theta[a]	1957–2033 UTC 30 May 2013	67–78	Supercell	1.75-in. hail at 2013 UTC
				1.0-in. hail at 2015 UTC
				1.0-in. hail at 2015 UTC
				1.0-in. hail at 2030 UTC
Severe hail and/or wind cases				
Beta	2120–2200 UTC 12 Aug 2011	72	Multicell	EG 61-kt wind at 2200 UTC
Gamma	2330–2359 UTC 22 Oct 2011	57–62	Multicell	1.50-in. hail at 2358 UTC
Delta	2222–2301 UTC 4 May 2012	32–40	Multicell	1.0-in. hail at 2242 UTC
				1.25-in. hail at 2255 UTC
				MG 59-kt wind
Tornado cases				
Eta	2130–2200 UTC 29 May 2013	61–76	Squall line	EF1 tornado at 2200 UTC
Iota	2209–2301 UTC 30 May 2013	71	Supercell cluster	1.75-in. hail at 2212 UTC
				4.25-in. hail at 2227 UTC
				EG 61-kt wind at 2230 UTC
				EF0 tornado at 2258–2259 UTC
Zeta	2050–2154 UTC 19 May 2013	64–76	Supercell	1.50-in. hail at 2101 UTC
				1.25-in. hail at 2115 UTC
				EG 52-kt wind at 2115 UTC
				1.0-in. hail at 2117 UTC
				1.50-in. hail at 2118 UTC
				EF1 tornado at 2122–2130 UTC
				EF1 tornado at 2133–2134 UTC
				2.60-in. hail at 2137 UTC
				EF2 tornado at 2141–2154 UTC

[a] Theta is a nontornadic hail-producing supercell (null tornado case).

This nonparametric test was chosen because the collected data did not meet normality assumptions. Overall, the POD and FAR scores were similar, with slight improvements as updates became more rapid (Table 4, all cases). Broken down by event type, the greatest differences are found for tornado warning POD and FAR scores (Table 4, all cases). The full-, half-, and quarter-speed POD (FAR) scores were 0.78 (0.29), 0.74 (0.45), and 0.62 (0.44), respectively.

These big-picture findings indicate that of the three update times used, full-speed data were most beneficial to forecasters' ability to issue more accurate warnings with longer lead times. However, of interest is how representative these findings are for each case worked. While examining this question, we found that the results were sensitive to the situation presented. For example, the temporal resolution used during Gamma and Eta had little impact on warning lead times, whereas differences were found in the other severe and tornado cases. Furthermore, in cases containing multiple reports, such as Delta and Zeta, we found the use of faster updates particularly improved warning lead times for the first report of the event. These longer initial warning lead

times are an encouraging result, as warnings verified for the first report of the day tend to be the most challenging (e.g., Andra et al. 2002; Brotzge and Erickson 2009).

Given these situational dependencies, we expected that the overall median warning lead times computed using only first reports from each case, and excluding Gamma and Eta, would show more improvement in warning lead time when using faster updates. Applying these criteria, the median lead times for full, half, and quarter speeds were 14.5, 10.5, and 5.5 min, respectively ($N = 120$ for each update-speed group). In this case, the application of the Kruskal–Wallis test did indicate statistically significant differences between the three groups (p value = 0.0013). A posthoc Wilcoxon–Mann–Whitney rank-sum test (e.g., Wilks 2006) indicates between which groups these statistically significant differences occurred (p value < 0.0170). Again, this nonparametric test was chosen because the data collected did not meet normality assumptions. Comparing the three groups, the full-speed group's median lead time distribution for this subset of the data was most different from that of the quarter-speed group (p value = 0.0003), and provided additional
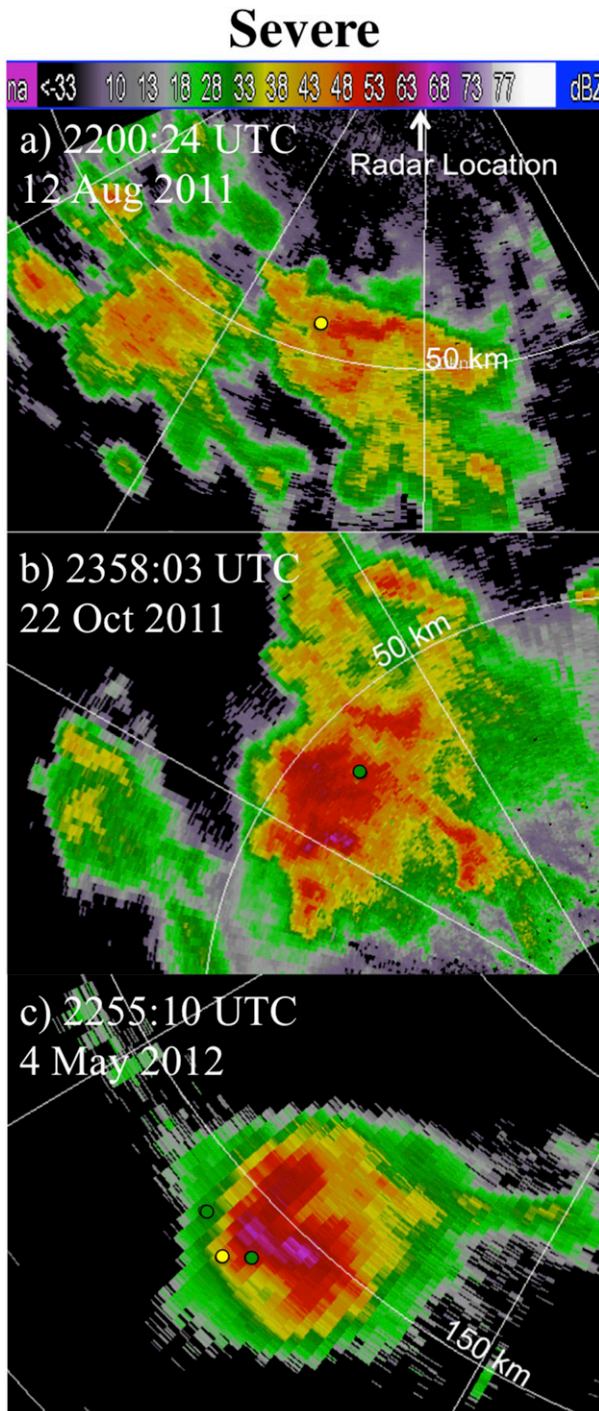
Severe

FIG. 3. PAR 0.5° reflectivity for (a) Beta, (b) Gamma, and (c) Delta. Green dots are severe hail reports and yellow dots are severe wind reports. Reflectivity (dBZ) color bar is located at the top. White rings are displayed in 50-km increments.

confidence that the use of full-speed data did extend warning lead times compared to the use of quarter-speed data, in these cases. Further examination of the first reports by case type revealed that the statistical significance found above was more so due to differences in tornado warning lead times between the three groups (Kruskal–Wallis p value = 0.0380), rather than differences in severe thunderstorm warning lead times (Kruskal–Wallis p value = 0.1162). The remainder of this section discusses the performance results by case type.

a. Performance: Severe cases

The overall severe median warning lead times for the full-, half-, and quarter-speed cases were very similar: 21, 22.5, and 20 min, respectively (N = 150 per group). As noted earlier, the most similar severe warning lead times occurred during Gamma, the hail-only case (Fig. 5a). Hence, this case contributed to the overall similarity in the median severe warning lead times found. To aid qualitative comparison between groups, in each case the median severe warning lead time for the full distribution (N = 30) was computed. For Gamma, the full distribution lies near the 24.5-min median severe warning lead time. All groups achieved a perfect severe POD and FAR score (Table 4, severe cases).

The most dissimilar severe warning lead times between the full-speed group and the quarter-speed group occurred during Beta, the wind-only event (Fig. 5b). Therein, both the full- and half-speed groups achieved severe-warning lead times located mostly near or above the overall 18-min median lead time (N = 30; Fig. 5b). In contrast, more than half of the quarter-speed group achieved severe-warning lead times at least 6 min under the 18-min median. The median severe warning lead times for the full-, half-, and quarter-speed groups were 19.5, 18.0, and 10.5 min, respectively. The quarter-speed group's POD score was slightly lower and the FAR score slightly higher compared with the full-speed group (Table 4). In this wind-only case, the use of half- and full-speed data was overall more advantageous to forecasters' ability to issue warnings with longer lead times than the use of quarter-speed data.

Unlike the other two cases, Delta contained both severe hail and wind reports. Because multiple storm reports were received as forecasters worked the case, warning lead times associated with the first report provided the clearest measure of the impact of temporal resolution on the warning decision process. As in Beta, groups using full- and half-speed data tended to issue warnings earlier (medians of 10 and 11 min, respectively) than the quarter-speed group (median of 6.5 min) (Fig. 5c). However, overall, the half-speed group outperformed the full-speed group, as the
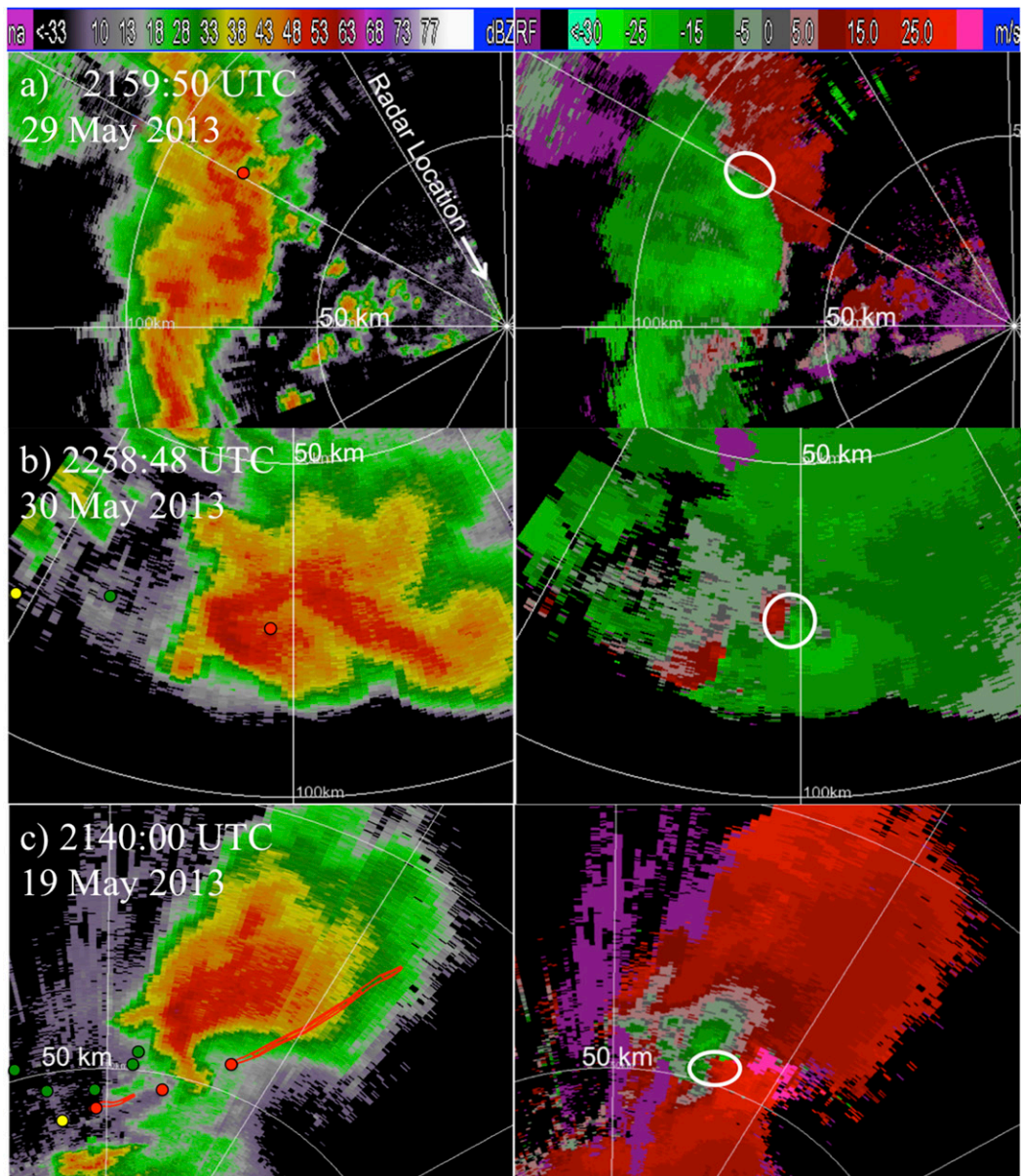
Tornado

FIG. 4. PAR 0.5° (left) reflectivity and (right) velocity for (a) Eta, (b) Iota, and (c) Zeta. Green dots are severe hail reports, and yellow dots are severe wind reports. Red dots are tornado reports (i.e., starting point of tornado path), while red lines in (c) are tornado paths associated with longer-lived tornadoes. Thick white circles show the locations of the couplet of interest. Reflectivity (dBZ) and velocity (m s^{-1}) color bars are located at the top. White rings are displayed in 50-km increments.

former produced the highest number of initial, second, and third severe warning lead times above the overall median warning lead times (10.5, 21.5, and 24.5 min, respectively) (Fig. 5c). One outlier was P29 of the half-speed group, who missed the first hail event; P9 of the quarter-speed group also missed the first event. The use of higher temporal resolution data also resulted in

slightly higher FARs compared to forecasters using quarter-speed data (Table 4).

b. Performance: Tornadic cases

The overall median tornado warning lead times for the full-, half-, and quarter-speed cases were 12.7, 8, and 9 min, respectively ($N = 150$ per group). Like the severe

TABLE 3. The event-specific and generic verification systems used by the NWS and the hybrid verification system employed for PARISE.

	Event-specific verification				Generic verification				PARISE 2015 verification			
	Severe thunderstorm warning		Tornado warning		Severe thunderstorm warning		Tornado warning		Severe thunderstorm warning		Tornado warning	
Hazard	Event hit	Warning verified	Event hit	Warning verified	Event hit	Warning verified	Event hit	Warning verified	Event hit	Warning verified	Event hit	Warning verified
Wind/hail	Yes	Yes	N/A	N/A	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Tornado	N/A	N/A	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes

TABLE 4. The POD and FAR scores across all severe (SVR) and all tornado (TOR) warnings by update speed for all cases; severe cases Gamma, Beta, and Delta; and tornado cases Eta, Iota, and Zeta.

All cases	Full	Half	Quarter
SVR POD	0.90	0.87	0.85
TOR POD	0.78	0.74	0.62
Total POD	0.87	0.85	0.80
SVR FAR	0.37	0.37	0.36
TOR FAR	0.29	0.45	0.44
Total FAR	0.36	0.39	0.38

Severe cases	Speed	SVR POD	SVR FAR
Gamma	Full	1.0	0.0
	Half	1.0	0.0
	Quarter	1.0	0.0
Beta	Full	1.0	0.0
	Half	1.0	0.14
	Quarter	0.9	0.08
Delta	Full	1.0	0.15
	Half	0.97	0.11
	Quarter	0.97	0.0

Tornado cases	Speed	TOR POD	TOR FAR
Eta	Full	0.1	1.0
	Half	0.1	1.0
	Quarter	0.0	1.0
Iota	Full	0.8	0.33
	Half	0.6	0.53
	Quarter	0.1	0.50
Zeta	Full	1.0	0.0
	Half	1.0	0.0
	Quarter	1.0	0.0

cases, performance for tornado cases was determined by the situation presented to forecasters. The most challenging tornado case for all groups was Eta, in which a short-lived EF1-rated tornado was produced at the north end of a bowing line segment approximately 75 km northwest of the PAR (Fig. 4a; Table 2). In this case, only 5 of 30 forecasters decided to issue tornado warnings prior to tornado occurrence: three were in the full-speed group (P11, P14, and P15), and two were in the half-speed group (P22 and P27). Of these five forecasters, tornado warnings verified only for P14, P15, and P22 with associated tornado warning lead times of 0, 2, and 6 min, respectively (Fig. 6a).

Sixteen forecasters decided to issue their first (and only) tornado warning reactively, a few minutes after they received the tornado report. Four of the forecasters were in the full-speed group, whereas six were in the half- and quarter-speed groups. The remaining nine forecasters decided not to issue tornado warnings following the report. As most forecasters issued unverified tornado warnings, the median tornado lead time was
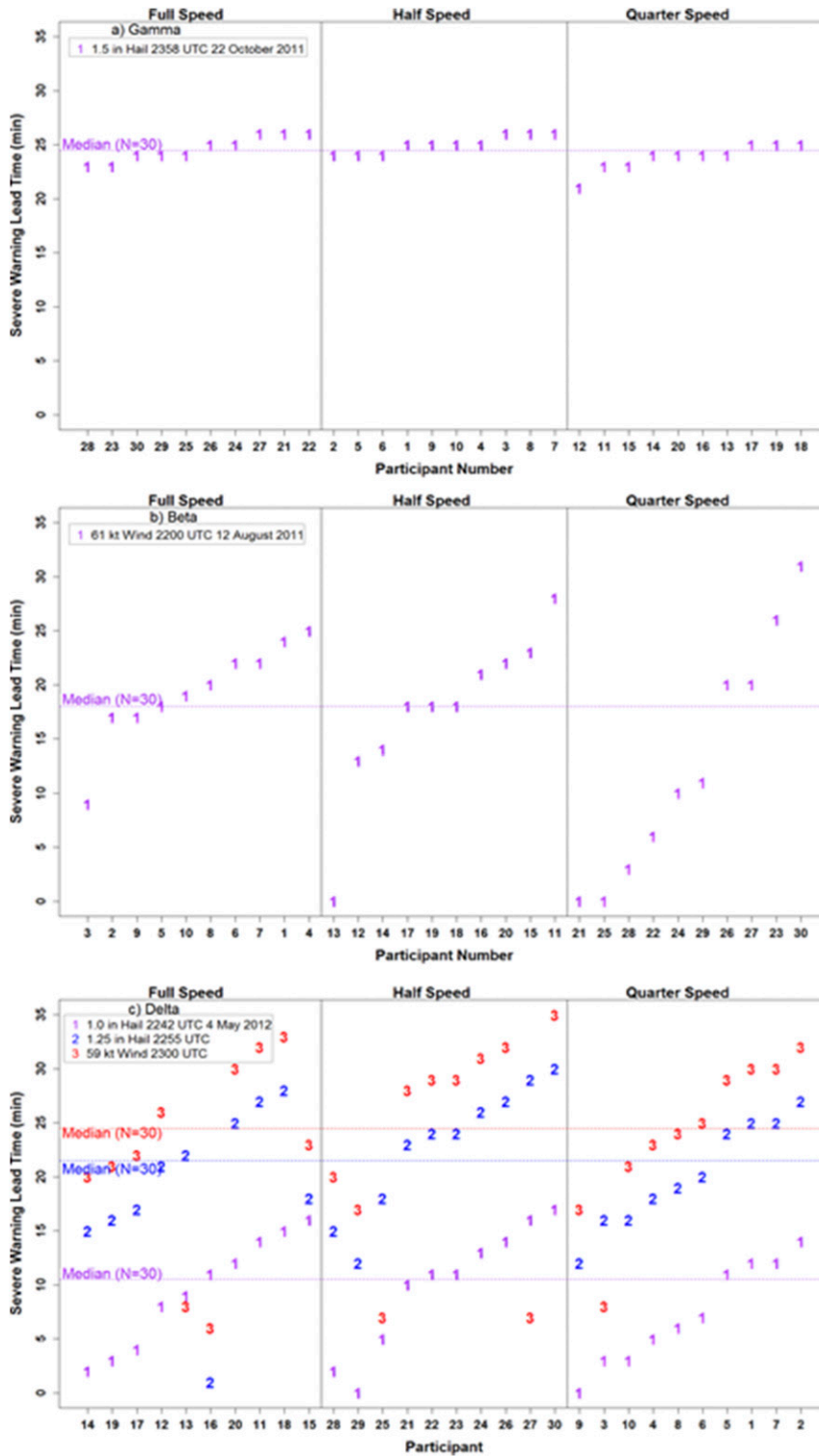
FIG. 5. Distribution of forecasters' severe warning lead times (min) for each case: (a) Gamma, (b) Beta, and (c) Delta, organized by update speed. First, second, and third severe reports are denoted by numbers 1, 2, and 3 (magenta, blue, and red). For each report, the median severe warning lead time (min) for the full distribution is given by a dotted and annotated line (magenta, red, and blue).
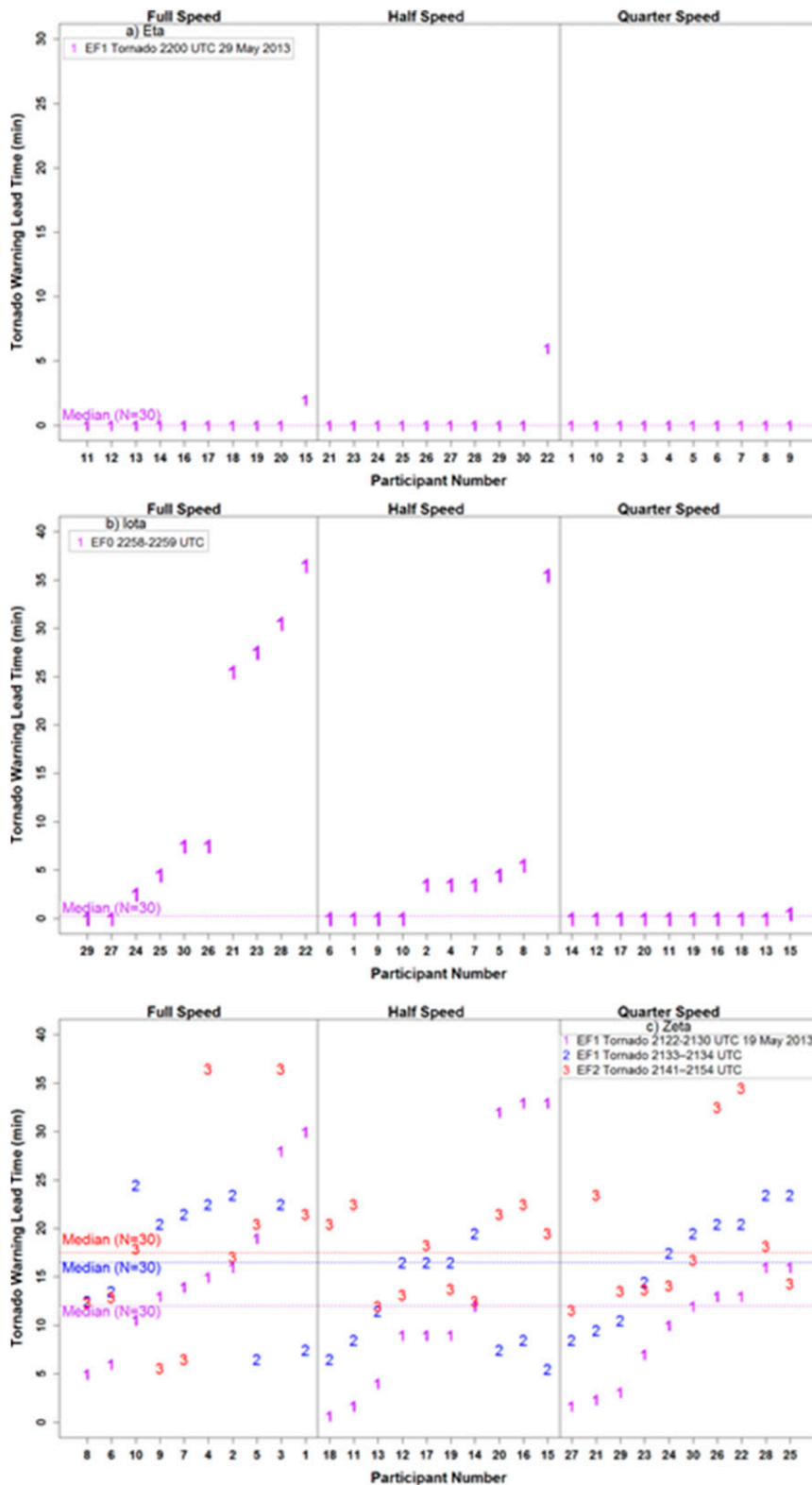
FIG. 6. Distribution of forecasters' tornado warning lead times (min) for each case: (a) Eta, (b) Iota, and (c) Zeta, organized by update speed. First, second, and third severe reports are denoted by numbers 1, 2, and 3 (magenta, blue, and red). For each report, the median severe warning lead time (min) for the full distribution is given by a dotted and annotated line (magenta, red, and blue).

0 min, and the majority of the POD and FAR scores were poor (Table 4, tornadic cases). In this case, radar update speed had little to no discernable impact on the forecasters' performance.

The use of full-speed data was most advantageous during Iota, the case containing a cluster of supercells, one of which produced an EF0-rated tornado (Fig. 4b; Table 2). In this case, the majority of the full-speed groups' tornado warning lead times were longer than the overall median warning lead time of 0.25 min, which is in stark contrast to the quarter-speed group (Fig. 6b). Of the eight in the full-speed group with nonzero tornado warning lead times, half achieved lead times between 25 and 36 min, while the other half achieved lead times under 10 min. Six of 10 participants in the half-speed group achieved nonzero tornado warning lead times; five were 5 min or less, whereas 1 was 35 min. The median tornado warning lead times for full-, half-, and quarter-speed groups were 7.5, 3.5, and 0.0 min, respectively. Besides increasing tornado warning lead time, the use of full-speed data in Iota resulted in fewer tornado misses and false alarms (Table 4). About 30 min prior to Iota's EF0 tornado, 4.5-in. hail and a 61-kt (where $1 \text{ kt} = 0.51 \text{ m s}^{-1}$) wind gust were reported (Table 2). For these reports, the distributions of severe warning lead times between groups were relatively similar, with a tendency for lower lead times for members of the quarter-speed group (not shown).

Unlike the previous two tornado cases, Zeta presented a classic cyclic supercell that produced several tornadoes, including two rated EF1 and one rated EF2 (Table 2). As in the severe case, Delta, of particular interest was whether the use of increasingly rapid updates would enhance the tornado warning lead time for the first tornado occurrence, which in operations tends to be the most difficult to forewarn (e.g., Andra et al. 2002; Brotzge and Erickson 2009). In this case, the full-speed group performed best with about twice as many full-speed participants producing first tornado warning lead times above the overall median of 12 min (median tornado warning lead time = 14.5 min), compared with the half- and quarter-speed groups (median tornado warning lead times of 9 and 11 min, respectively; see Fig. 6c). A few forecasters in the full- and half-speed groups issued tornado warnings with comparatively long lead times ranging from 25 to 35 min (Fig. 6c). These results indicate that the full-speed group and a few forecasters in the half-speed group gained situational awareness unavailable in the 4-min volume updates used by the quarter-speed group. The overall median tornado warning lead times for the second and third tornadoes were similar: 16.5 and 17.5 min,

respectively (Fig. 6c). Also similar were the lead-time distributions associated with these warnings, with a slight tendency for lower lead times for the half-speed group. Regardless of the observed differences in tornado warning lead times between groups, no unverified tornado warnings were issued (Table 4).

c. Performance: Null cases

Epsilon and Alpha presented forecasters with null multicell events (Figs. 2a,b; Table 2). Of the two cases, the results indicate that the use of full-speed data was most advantageous during Epsilon, as only 16 of 30 forecasters decided to issue severe thunderstorm warnings. Of the 16 forecasters who issued warnings, 3 were in the full-speed group, compared with 6 and 7 in the half- and quarter-speed groups, respectively. In contrast, while working Alpha (Fig. 2b), most forecasters (26 of 30) decided to issue severe thunderstorm warnings. Of the four that did not issue severe thunderstorm warnings, one each used full- and half-speed data, while two used quarter-speed data.

During Theta (Fig. 2c; Table 2), the nontornadic supercell case, most forecasters (24 of 30) issued severe thunderstorm warnings, and one-third issued tornado warnings. To assess severe and tornado warning false alarms separately, the FAR was computed with respect to each warning type (Fig. 7). Although the distribution of severe thunderstorm warning FAR scores is fairly similar across update speeds, a few more forecasters achieved severe FAR scores lower than 0.5 using quarter-speed data ($N = 5$) than when using full- or half-speed data ($N = 3$). In contrast, more forecasters using quarter- and half-speed data issued tornado warnings ($N = 5$ and $N = 4$, respectively) than those using full-speed data ($N = 1$). Hence, in this case, the use of full-speed data appeared to be most advantageous in reducing the number of tornado false alarms.

6. Warning polygon size and duration

While analyzing forecaster performance, multiple questions arose about whether warning characteristics (i.e., size and duration) depended on storm mode or radar update speed. We found that the largest differences in warning characteristics were related to each case's storm mode. For example, the largest severe thunderstorm warnings were issued during the squall-line case (Eta; Table 5; Fig. 8a), which is not surprising given that squall lines can stretch over 100 km in length and can produce widespread severe weather (e.g., Funk et al. 1999; Trapp et al. 2005). Various warning strategies employed by 2015 PARISE participants likely resulted in these very large severe thunderstorm warnings. For
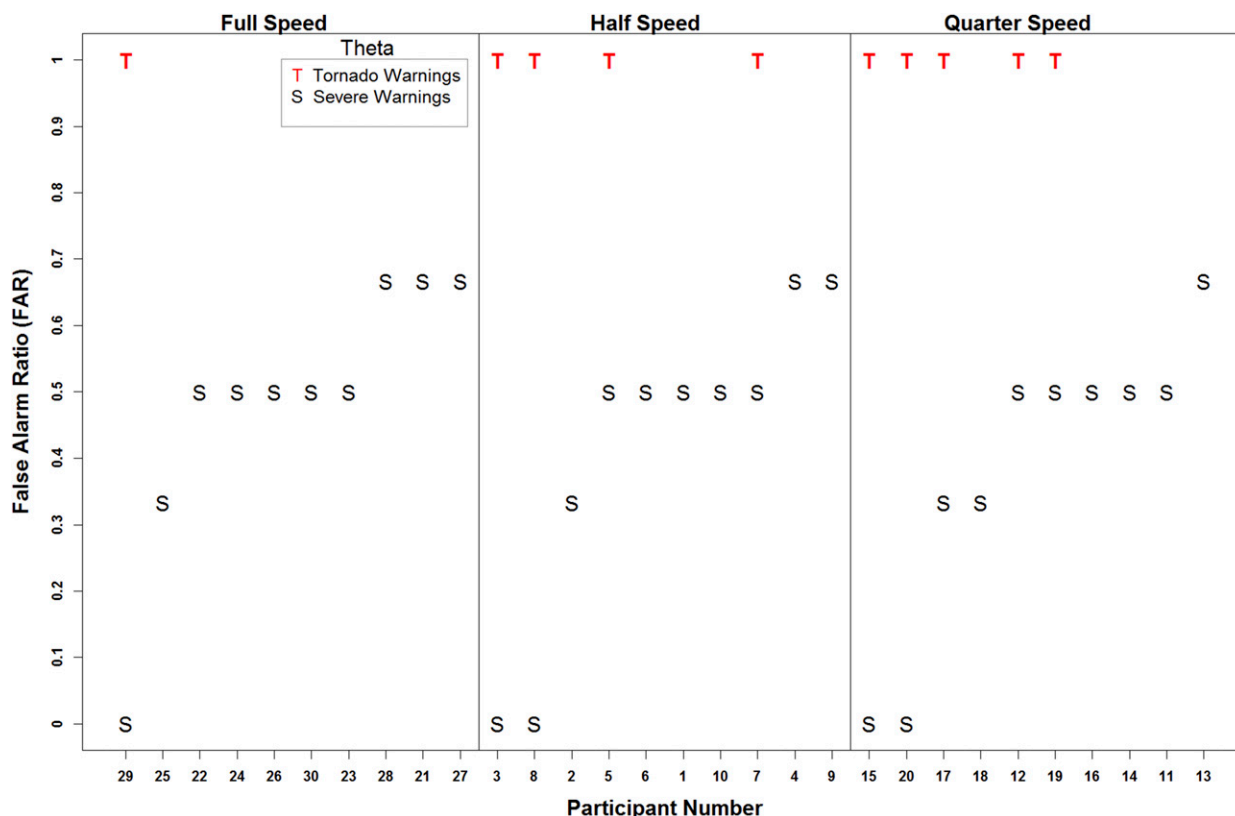
FIG. 7. FARs for severe (black S) and tornado (red T) warnings by participant, plotted by update-speed group.

example, P15 explained the need for a large warning size during Eta. They stated that their main objective was to warn for the deepest reflectivity core, but that the warning should also capture new deep reflectivity cores and potential severe-weather threats that might develop anywhere along the line.

Tornado warning size and duration also varied most based on storm mode. Tornado warnings issued during the squall-line case (Eta) were the largest, but the duration of these warnings was the shortest of the three tornado cases (Table 5; Fig. 9). While working Eta, 12 participants expressed uncertainty in issuing a tornado warning based on radar data alone. In total, 18 participants issued a tornado warning only after receiving a tornado report. Based on performance (section 5), Eta was a challenging case, and the higher uncertainty

TABLE 5. Overall median warning size and duration for all warnings issued in each case. Three case types are null, severe hail and/or wind, and tornadic.

Case	SVR size (km^2)	SVR duration (min)	TOR size (km^2)	TOR duration (min)
Null cases				
Epsilon	1026	36.0	NA	NA
Alpha	936	40.0	NA	NA
Theta[a]	1189	39.0	NA	NA
Severe hail and/or wind cases				
Beta	1158	36.0	NA	NA
Gamma	1169	41.0	NA	NA
Delta	1465	38.5	NA	NA
Tornado cases				
Eta	3887	39.0	848	28
Iota	1402	34.5	637	32
Zeta	1590	42.0	794	33

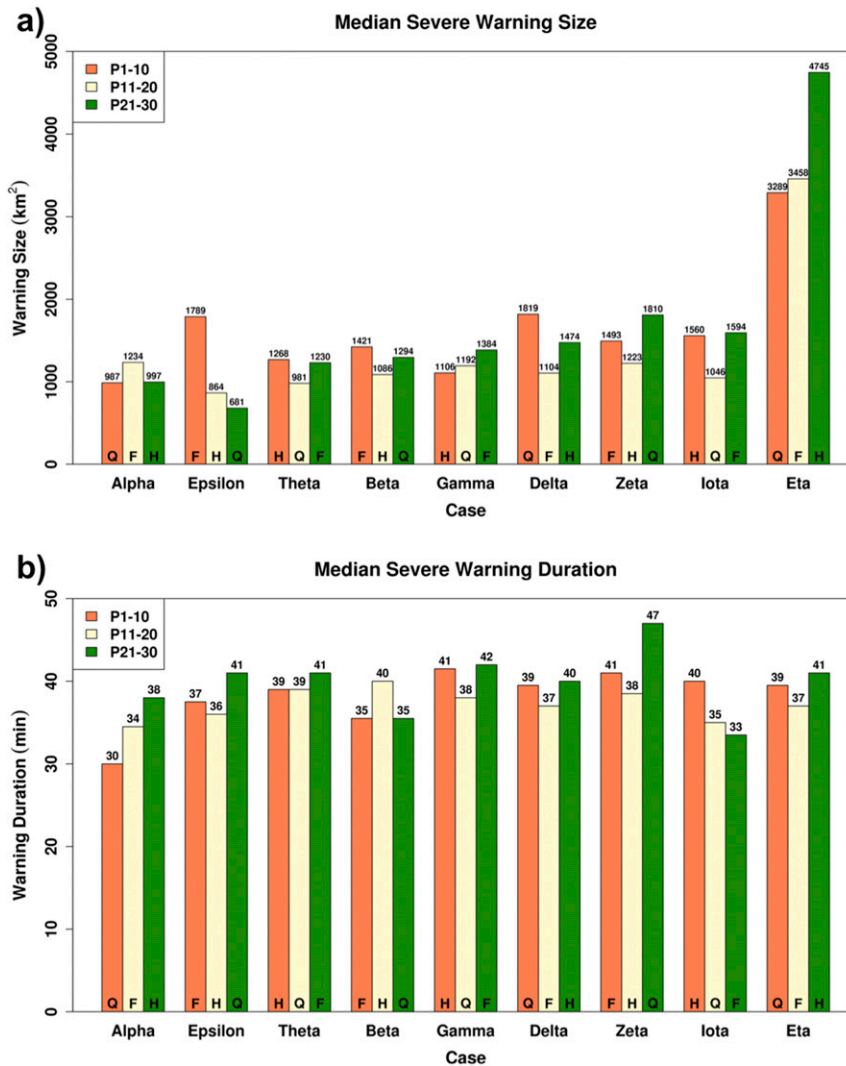[a] Theta is a nontornadic hail-producing supercell (null tornado case).

FIG. 8. Median severe thunderstorm warning (a) size and (b) duration for each participant group. Median values are included near the top of each bar. Radar update speed (F = full, H = half, and Q = quarter) worked by each group for each case is included near the bottom of each bar.

expressed by the participants likely influenced the size and duration of their warnings. In addition, 11 participants explicitly stated that squall-line tornadoes tend to be short lived, which likely resulted in shorter-duration tornado warnings. The participants' perception that squall-line tornadoes tend to be short lived was accurate for this case, as the tornado in Eta lasted 1 min (Table 2). Studies of tornadoes relative to storm mode also align with the participants' perceptions (e.g., Trapp et al. 2005; Davis and Parker 2014). During the other two tornado cases, environmental conditions alerted participants to a heightened potential for strong supercells that can produce long-lived tornadoes, thereby requiring longer warnings. The classic supercell case (Zeta) had the longest tornado warnings,

although these warnings were only 1 min longer than those issued during the supercell cluster case (Iota; Table 5). During Zeta, participants also received multiple tornado reports throughout the case. Knowledge of a confirmed tornado may explain why 17 of the 30 participants issued a second tornado warning that was longer than the first tornado warning.

While differences in warning size and duration were observed for cases with differing storm modes, it is worth noting that these characteristics did not change substantially when radar update speed changed. In addition, when looking at the cases individually, no clear patterns emerged in terms of warning characteristics and radar update speed (Figs. 8 and 9). Since radar update speed did affect lead time (section 5) but not warning characteristics, it is possible that
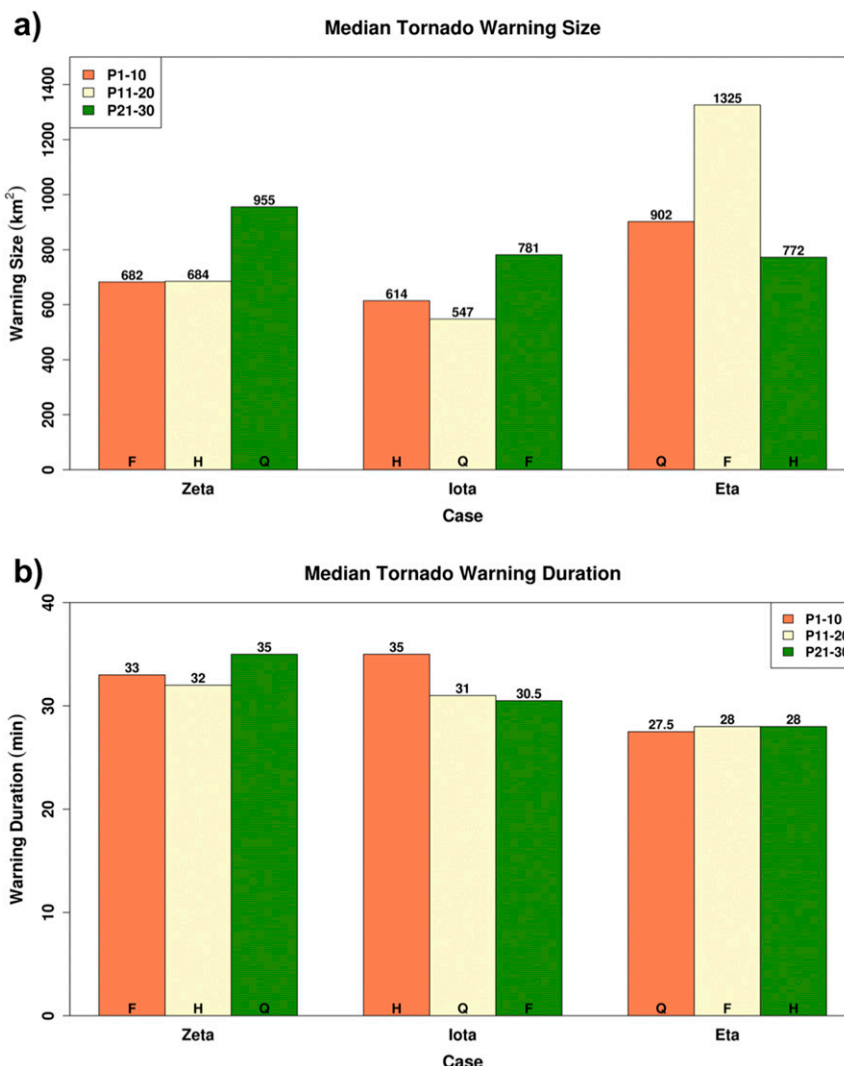
FIG. 9. As in Fig. 8, but for median tornado warning.

7. Cognitive workload

a. Workload distributions and profiles

The ISA workload analysis is based on forecasters' ratings chosen at 5-min intervals during the video-cued retrospective recall. The number of ratings in each case ranged from 6 to 13 depending on case duration. In total, 24 ISA ratings were missed, 8 of which each belonged to the quarter-, half-, and full-speed groups. Over half of these missed ratings occurred during the tornado cases, possibly because of the higher demand of this case type. Given that these

workload reports were incomplete, they were removed from the analysis.

Each group's median 5-min workload rating for the nine cases was either a level 2 or a level 3 (Fig. 10). This result suggests that, on average, forecasters were not cognitively overloaded during this experiment. However, a difference in cognitive workload based on temporal resolution is evident. While the quarter-speed group was on average a level 2 (relaxed) for all of the null and severe hail/wind cases, the full-speed group was a level 3 (comfortable) for half of these (Figs. 10a–f). The half-speed group was a level 3 for only one of these cases (Theta), which although classified as null, presented a nontornadic supercell that produced severe hail. The median workload rating for the tornado cases was a level 3 for all groups (Figs. 10g–i), suggesting that aside from

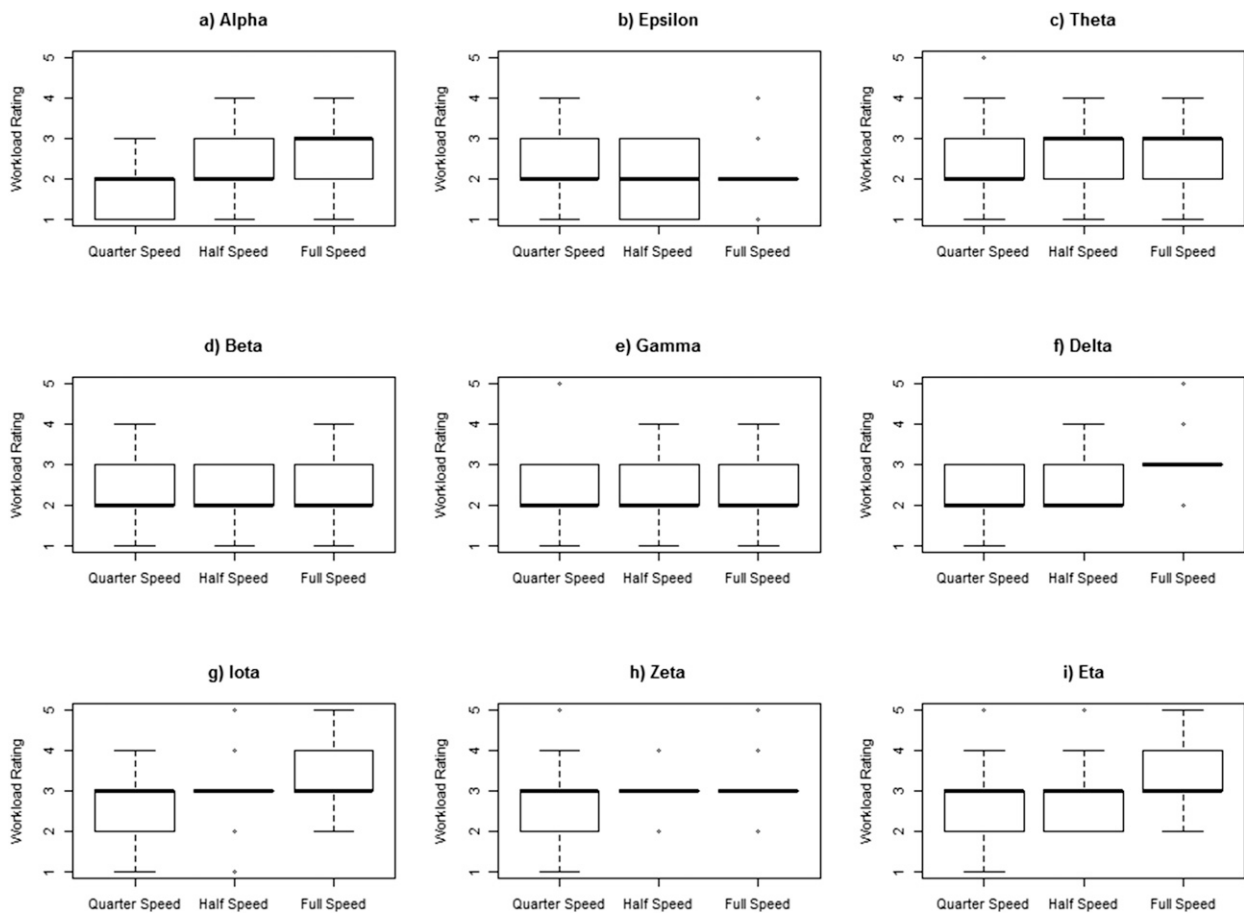changes in radar update speed affects when, not how, a forecaster designs and issues a warning.

FIG. 10. Boxplots of 5-min workload ratings for quarter-, half-, and full-speed groups for the (a)–(c) null, (d)–(f) severe, and (g)–(i) tornadic cases. The solid middle line indicates the median value, and the box edges indicate the lower and upper quartiles (i.e., interquartile range). Minimum and maximum values are identified with whiskers, and outliers are either less than 1.5 times the lower quartile or greater than 1.5 times the upper quartile.

temporal resolution, the increased weather threat contributed to the overall higher levels of workload.

Despite some similarities in the median 5-min workload ratings, a Kruskal–Wallis test (Kruskal and Wallis 1952) showed statistically significant differences in ISA ratings between the three groups in all but two cases (p values < 0.05; Table 6). One of these cases, Gamma, was when forecasters' performance was most similar (Fig. 5a). A posthoc Wilcoxon–Mann–Whitney rank-sum test (e.g., Wilks 2006) indicates between which groups these statistically significant differences occurred (p value < 0.017; Table 6). Comparing the three groups, the quarter-speed group's ISA rating distribution was most different from that of the full-speed group, while the half- and full-speed groups' ISA rating distributions were most similar (Table 6).

Comparisons of ISA rating distributions give an overall impression of the level of cognitive workload experienced within a case. However, given the dynamic nature of weather, the change in workload as cases evolved (i.e., workload profile) was also of interest. We observed that regardless of temporal resolution or case type, 21 of the 30 participants' workload rating patterns were either flat (i.e., little or no change in workload) or fluctuating (i.e., multiple increases and decreases in workload) in the majority of cases worked. Although we did not analyze personality traits during PARISE 2015, these workload behavior tendencies suggest that forecaster personality was also likely an important factor in perceived cognitive workload during the simulations. It is possible that personality traits may have influenced forecasters' coping strategies and approaches to the simulations, thus influencing their ISA ratings. Past studies support this suggestion; personality traits and perceived subjective workload have been found to correlate during vigilance tasks (e.g., Rose et al. 2002; Szalma 2002; Guastello et al. 2015). The influence of personality would also explain differences in the

TABLE 6. Kruskal–Wallis rank-sum test and Wilcoxon–Mann–Whitney rank-sum test p values for differences in cognitive workload distributions across groups with differing temporal resolution.

	Kruskal–Wallis rank-sum test p values	Wilcoxon–Mann–Whitney rank-sum test p values		
Case	Quarter, half, and full speed	Quarter and half speed	Quarter and full speed	Half and full speed
Alpha	<0.001	<0.001	<0.001	<0.001
Epsilon	0.047	0.061	0.015	0.778
Theta	0.049	0.020	0.172	0.181
Beta	0.133	0.050	0.327	0.294
Gamma	0.408	0.967	0.239	0.264
Delta	<0.001	0.081	<0.001	<0.001
Iota	<0.001	<0.001	<0.001	0.740
Zeta	<0.001	<0.001	<0.001	0.777
Eta	<0.001	0.048	<0.001	<0.001

forecasters' level of boredom versus excitement during cases and why some forecasters were more sensitive to changes in task demand than others.

b. Reasoning for higher levels of cognitive workload

1) CATEGORIES

Forecasters' reasoning associated with each ISA rating gives insight into the chosen ratings for perceived cognitive workload. Although the average ISA ratings show that forecasters were generally relaxed and comfortable during the nine cases, many ISA ratings extended to a level 4 (high workload), and there are numerous outliers rated at a level 5 (excessive workload) (Fig. 10). The reasoning provided for all level 4 and level 5 ISA ratings was analyzed ($N = 183$), and six categories were identified. In order of prevalence, these categories are 1) storm characteristics, 2) warnings, 3) case startup, 4) temporal resolution, 5) technical frustrations, and 6) personal (Fig. 11a). Storm characteristics causing higher cognitive workload included the number of storms in the sector, the expected threat, and the evidence of intensification. The warning category is associated with higher cognitive workload as a result of the extra task of issuing products, sacrificing interrogation time, having concerns about polygon placement relative to storms, and the unfortunate realization that warnings were not panning out as expected. Case startup describes the increased workload that was experienced within the first 5–10 min of a case. During this time, higher cognitive workload was experienced because forecasters felt an urgency to load their data, assess the situation, and possibly make warning decisions. The temporal resolution of radar data was associated with a higher workload, such that forecasters felt the need to monitor the data quickly so that they could keep up with trends. Oftentimes forecasters reported higher levels of workload because they did not have enough time to look at all the data and

were not able to pinpoint the important signals. Technical frustrations caused increases in workload typically because WarnGen/AWIPS-2 did not function as it should, which sometimes caused delays in product issuance. Finally, one forecaster reported three ISA ratings of level 5 as result of requiring a bathroom break while monitoring the weather.

2) TEMPORAL RESOLUTION

Forecasters using full-speed PAR data reported approximately twice as many level 4 and 5 ISA ratings than those using quarter- and half-speed PAR data (Fig. 11b). The largest reasoning category for the full-speed group's higher ISA ratings was storm characteristics, followed by temporal resolution (Fig. 11b). In comparison, only a small portion of the half-speed participants reported higher ISA ratings due to temporal resolution, and no quarter-speed participants' reasoning related to temporal resolution (Fig. 11b). Storm characteristics and warning categories accounted for more than half of the reasoning for the quarter- and half-speed groups (Fig. 11b). Technical frustrations also accounted for a large portion of the quarter-speed group's higher ISA ratings, while case startup accounted for a quarter of the half-speed group's responses (Fig. 11b).

Only a small fraction of the higher cognitive workload ratings were a level 5 ($N = 26$). However, these ratings cause most concern because they describe a mental state that is cognitively overloaded. Forecasters using full-speed data gave over half of these ratings ($N = 16$) and related these ratings to every category except for technical frustrations. In comparison, almost all of the level 5 ratings given by quarter-speed participants were due to technical frustrations ($N = 5$ of 7). The remaining level 5 ratings given by quarter- and half-speed participants were associated with case startup and warning reasoning. Excessive workload due to temporal resolution, storm
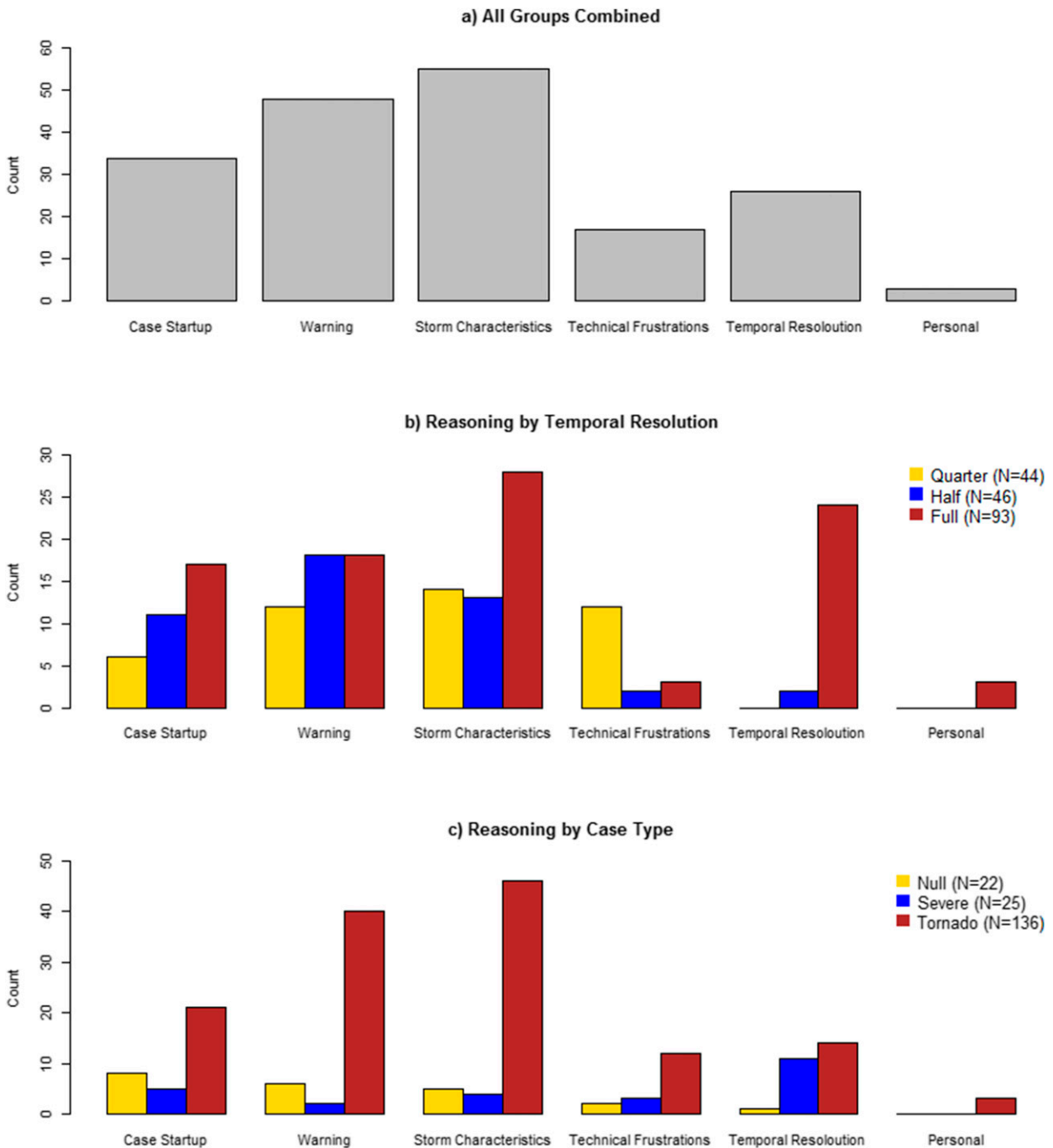
FIG. 11. Reasoning categories for ISA ratings given at levels 4 and 5 for (a) all groups combined, (b) each temporal resolution, and (c) each case type.

characteristics, and personal matters only occurred with full-speed participants.

3) STORM TYPE

Of all the case types, forecasters reporting level 4 and level 5 ISA ratings did so most during the tornado cases (Fig. 11c). Reasoning for this increase in cognitive workload was mostly associated with the storm characteristics and warning categories. Monitoring multiple threats for one supercell, dealing with uncertainty in storm evolution, and feeling overwhelmed with the number of warning products needing to be issued were

all factors leading to these higher levels of experienced cognitive workload. Although temporal resolution was not a large contributor to the higher cognitive workload reported during the tornado cases, it was the largest category for why forecasters reported these higher ISA ratings during the severe hail/wind cases (Fig. 11c). The temporal resolution reasoning was mostly associated with Delta, and occurred as a result of forecasters not being able to examine the data closely as updates were coming in, having difficulty comprehending the structure and evolution of the storm due to the fast updates, and needing to adapt to a different type of interrogation strategy. It is worth noting that update speeds were quickest in Delta compared with the other cases (Table 2). The different reasoning driving level 4 and 5 ISA ratings for tornado and severe hail/wind cases supports the notion that the higher cognitive workload is not only a function of temporal resolution, but also of storm type, as suggested earlier.

8. Discussion

Based on the performance analysis, we found that forecasters' ability to increase severe and tornado warning lead times when using increasingly higher temporal resolution data depended on the weather situation presented. Distributions of positive warning lead times were most comparable during Gamma (Fig. 5a); this result suggests that similar situational awareness was gained by forecasters in all three groups. While working the two other severe cases, the use of increasingly higher temporal resolution data most aided forecasters' ability to issue verified warnings earlier during Beta, the severe wind event (Fig. 5b). A tendency for longer initial warning lead times when using increasingly higher temporal resolution data was also found during Delta, the hail and wind event (Fig. 5c). These findings are consistent with Bowden et al. (2015), who in PARISE 2013 found the use of full-speed PAR data, compared with quarter-speed PAR data, increased median severe thunderstorm warning lead times by 5 min in two severe (large hail and/or damaging wind) cases. In a follow-on study by Bowden and Heinselman (2016), their analyses of forecasters' situational awareness determined that longer severe thunderstorm warning lead times were driven by the forecasters' ability to observe rapid changes in radar-based hail and wind precursors earlier when using 1- versus 5-min radar volume scans. More frequent sampling of specific hail and wind events by PAR was also found to improve the scientific understanding of radar-based severe storm precursors in several case studies, including Heinselman et al. (2008), Emersic et al. (2011), Newman and Heinselman (2012), and Kuster et al. (2016).

The advantage of frequent updates in the analysis of severe storms, and in particular downbursts, has been demonstrated in prior studies using rapid-scan data from other radar platforms (e.g., Roberts and Wilson 1989).

This PARISE was the first in the series of former experiments to explore the ability of forecasters to issue verified tornado warnings with lead time in advance of a short-lived tornado within a bowing line segment. During this event (Eta), the overall lack of verified tornado warnings with positive lead time, especially when using full-speed data, is somewhat discouraging (Fig. 6a). Our expectation for a more positive result was supported by the regional radar climatology of tornadic and nontornadic vortices within nonsupercell storms by Davis and Parker (2014), who found statistically significant differences in their azimuthal shear magnitudes ($0.006 \, s^{-1}$ or higher) when located within 60 km of a WSR-88D. The velocity couplet associated with the Eta tornado was located 15 km outside of this ideal radar range. Davis and Parker (2014) also found the median detection lead time for these nonsupercell tornadic vortices was 10 min, which suggests that the use of 1- or 2-min volume updates has the potential to improve forecasters' detection lead time for such events. While future analyses of participants' retrospective data will provide insight into this finding, anecdotal conversations with NWS forecasters reveal that some forecasters either do not issue tornado warnings during these types of events or wait for confirmation of a first event, owing to the potential for high false alarm rates. Additionally, when bowing lines (like this one) are fast moving, some forecasters discern the impact of the storm's translational motion as a more significant threat than the embedded circulation and, therefore, issue severe thunderstorm warnings instead.

In contrast, for the two tornadic supercell cases (Zeta and Iota), the forecasters' ability to issue verified and timely tornado warnings on the first tornado event improved when using full- and half-speed PAR data (Fig. 6). Zeta, a "classic" tornadic supercell event, appeared to be the more straightforward event since all issued tornado warnings verified. Iota, a tornadic supercell cluster, appeared to be more challenging, as full- or half-speed data were needed to achieve verified tornado warnings with lead time. Additionally, during the nontornadic supercell case (Theta), the use of full-speed data aided the forecasters' ability to discriminate correctly the severe weather threat, resulting in fewer false alarms (Fig. 7). Together these results are consistent with the 2010 and 2012 PARISE findings of Heinselman et al. (2012, 2015), where the use of higher temporal resolution also resulted in longer tornado warning lead times. However, FAR results were mixed, as FAR was impacted negatively in PARISE 2010 and

positively in PARISE 2012 when using faster radar updates (Heinselman et al. 2012, 2015, respectively). The PARISE 2015 FAR results are most consistent with the PARISE 2012 FAR findings. The advantage of frequent updates in the analysis of a potentially tornadic supercell's storm evolution, including specificity of tornado movement, has been demonstrated in prior studies using PAR data (e.g., Kuster et al. 2015), as well as data from other weather radars (e.g., Vasiloff 2001; Wurman et al. 2012; Isom et al. 2013; Pazmany et al. 2013; Kurdzo et al. 2015).

9. Conclusions and future work

The purpose of this paper was to focus on the traditional experiment component of the 2015 PARISE and share performance, warning characteristics, and cognitive workload results. The increased number of participants and cases worked compared with earlier experiments improves the generalizability of our work. The overall finding that median warning lead time increased with increasing update speed is in line with our findings from previous studies. Earlier warnings were provided in two severe hail/wind and two tornado cases, and the use of full-speed data for discriminating the weather threat was particularly useful to forecasters during Theta. However, longer warning lead time with faster update speeds was not observed in all cases, most notably during Eta. This finding suggests that specific training and guidance may be required to fully realize the benefits of full-speed PAR data to forecasters' warning decision processes during more challenging events. Making use of dynamic scanning methods that are already available (e.g., Chrisman 2009, 2014) will be a helpful first step to developing the skills necessary for processing rapidly updating radar data during warning operations.

While the update speed impacted when warnings were issued, it did not influence the size or duration of warning polygons (Figs. 8 and 9). Therefore, further improvements to warning metrics (such as the false alarm area) may require a change in the warning paradigm. This change may be possible through modernization of the current NWS warning system. A move toward probabilistic hazard information via the Forecasting a Continuum of Environmental Threats (FACETs) framework is expected to address multiple aspects of warning characteristics (e.g., Stumpf et al. 2008; Karstens et al. 2015).

Forecasters' subjective assessments of cognitive workload within the PARISE setting suggest that cognitive workload will rarely reach the excessive level, and when it does, it could be due to a variety of reasons that are not necessarily tied to the temporal resolution of the radar data. Our data also suggest that the perceived cognitive workload may relate to the forecasters' personality. Although we have not yet explored this relationship scientifically, investigating this hypothesis would be beneficial to a number of testbed experiments that may also observe effects of individual differences on forecasters' approaches, performance, and perceived workload.

Despite increasing our sample size and the variety of cases worked, we must be mindful of the limitations that still remain in this experiment. In these simulations, forecasters' warning decision processes were isolated to their independent thought; unlike in the forecast office, forecasters did not work in teams and therefore the data collected are not an accurate reflection of what could be expected in real warning operations. Additionally, forecasters' limited access to radar products and the absence of dual-polarization radar data simplified their warning decision processes even further. Considerations of these missing elements and how a future operational PAR system might impact convective warning operations will be addressed in the PARISE 2015 focus group analysis.

REFERENCES

Andra, D. L., Jr., 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., and P. L. Heinselman, 2016: A qualitative analysis of NWS forecasters' use of phased-array radar data during severe hail and wind events. *Wea. Forecasting*, **31**, 43–55, doi:10.1175/WAF-D-15-0089.1.

——, ——, 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.

Brotzge, J., and S. Erickson, 2009: NWS tornado warnings with zero or negative lead times. *Wea. Forecasting*, **24**, 140–154, doi:10.1175/2008WAF2007076.1.

Brown, R. A., R. M. Steadham, B. A. Flickinger, R. R. Lee, D. Sirmans, and V. T. Wood, 2005: New WSR-88D volume coverage pattern 12: Results of field tests. *Wea. Forecasting*, **20**, 385–393, doi:10.1175/WAF848.1.

Chrisman, J. N., 2009: Automated Volume Scan Evaluation and Termination (AVSET): A simple technique to achieve faster volume scan updates. Preprints, *34th Conf. on Radar Meteorology*, Williamsburg, VA, Amer. Meteor. Soc., P4.4. [Available online at https://ams.confex.com/ams/pdfpapers/155324.pdf.]

——, 2014: The continuing evolution of dynamic scanning. *NEXRAD Now*, No. 23, NOAA/NWS/Radar Operations Center, Norman, OK, 8–13. [Available online at http://www.roc.noaa.gov/WSR88D/PublicDocs/NNOW/NNow23a.pdf.]

Davis, J. M., and M. D. Parker, 2014: Radar climatology of tornadic and nontornadic vortices in higher-shear, low-CAPE environments in the mid-Atlantic and southeastern United States. *Wea. Forecasting*, **29**, 828–853, doi:10.1175/WAF-D-13-00127.1.

Emersic, C., P. L. Heinselman, D. R. MacGorman, and E. C. Bruning, 2011: Lightning activity in a hail-producing storm observed with phased-array radar. *Mon. Wea. Rev.*, **139**, 1809–1825, doi:10.1175/2010MWR3574.1.

Forsyth, D. E., and Coauthors, 2005: The National Weather Radar Testbed (phased array). Preprints, *32nd Conf. on Radar Meteorology*, Albuquerque, NM, Amer. Meteor. Soc., 12R.3. [Available online at https://ams.confex.com/ams/32Rad11Meso/techprogram/paper_96377.htm.]

Funk, T. W., K. E. Darmofal, J. D. Kirkpatrick, V. L. DeWald, R. W. Przybylinski, G. K. Schmocker, and Y.-J. Lin, 1999: Storm reflectivity and mesocyclone evolution associated with the 15 April 1994 squall line over Kentucky and southern Indiana. *Wea. Forecasting*, **14**, 976–993, doi:10.1175/1520-0434(1999)014<0976:SRAMEA>2.0.CO;2.

Guastello, S. J., A. Shircel, M. Malon, and P. Timm, 2015: Individual differences in the experience of cognitive workload. *Theor. Issues Ergon. Sci.*, **16**, 20–52, doi:10.1080/1463922X.2013.869371.

Hart, S. G., and L. E. Staveland, 1988: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Hum. Ment. Workload*, **52**, 139–183, doi:10.1016/S0166-4115(08)62386-9.

Heinselman, P. L., and S. M. Torres, 2011: High-temporal-resolution capabilities of the National Weather Radar Testbed Phased-Array Radar. *J. Appl. Meteor. Climatol.*, **50**, 579–593, doi:10.1175/2010JAMC2588.1.

——, 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, doi:10.1175/2008WAF2007071.1.

——, 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.

Isom, B., and Coauthors, 2013: The Atmospheric Imaging Radar: Simultaneous volumetric observations using a phased array weather radar. *J. Atmos. Oceanic Technol.*, **30**, 655–675, doi:10.1175/JTECH-D-12-00063.1.

Karstens, C. D., and Coauthors, 2015: Evaluation of a probabilistic forecasting methodology for severe convective weather in the 2014 Hazardous Weather Testbed. *Wea. Forecasting*, **30**, 1551–1570, doi:10.1175/WAF-D-14-00163.1.

Kirwan, B., A. Evans, L. Donohoe, A. Kilner, T. Lamoureux, T. Atkinson, and H. MacKendrick, 1997: Human factors in the ATM system design life cycle. *FAA/Eurocontrol ATM R&D Seminar*, Paris, France, Federal Aviation Administration and EUROCONTROL, 21 pp. [Available online at http://www.atmseminar.org/seminarContent/seminar1/papers/p_007_CDR.pdf.]

Kruskal, W. H., and W. A. Wallis, 1952: Use of ranks in one-criterion variance analysis. *J. Amer. Stat. Assoc.*, **47**, 583–621, doi:10.1080/01621459.1952.10483441.

Kurdzo, J. M., D. J. Bodine, B. L. Cheong, and R. D. Palmer, 2015: High-temporal resolution polarimetric X-band Doppler radar observations of the 20 May 2013 Moore, Oklahoma, tornado. *Mon. Wea. Rev.*, **143**, 2711–2735, doi:10.1175/MWR-D-14-00357.1.

Kuster, C. M., P. L. Heinselman, and M. Austin, 2015: 31 May 2013 El Reno tornadoes: Advantages of rapid-scan phased-array radar data from a warning forecaster's perspective. *Wea. Forecasting*, **30**, 933–956, doi:10.1175/WAF-D-14-00142.1.

——, ——, and T. J. Schuur, 2016: Rapid-update radar observations of downbursts occurring within an intense multicell thunderstorm on 14 June 2011. *Wea. Forecasting*, **31**, 827–851, doi:10.1175/WAF-D-15-0081.1.

Miller, S., 2001: Workload measures. National Advanced Driving Simulator Rep. N01006, 65 pp. [Available online at http://www.nadssc.uiowa.edu/publicationStorage/200501251347060.N01-006.pdf.]

NCEI, 2016: Storm events database. [Available online at https://www.ncdc.noaa.gov/stormevents/.]

Newman, J. F., and P. L. Heinselman, 2012: Evolution of a quasilinear convective system sampled by phased array radar. *Mon. Wea. Rev.*, **140**, 3467–3486, doi:10.1175/MWR-D-12-00003.1.

NWS, 2015: Verification. National Weather Service Instruction 10-1601, 8 pp. [Available online at http://www.nws.noaa.gov/directives/sym/pd01016001curr.pdf.]

Pazmany, A. L., J. B. Mead, H. B. Bluestein, J. C. Snyder, and J. B. Houser, 2013: A mobile rapid-scanning X-band polarimetric (RaXPol) Doppler radar system. *J. Atmos. Oceanic Technol.*, **30**, 1398–1413, doi:10.1175/JTECH-D-12-00166.1.

Reid, G. B., C. A. Shingledecker, and F. T. Eggemeier, 1981: Application of conjoint measurement to workload scale development. *Proc. Human Factors Society 25th Annual Meeting*, Rochester, NY, Human Factors Society, 522–526.

Roberts, R. D., and J. W. Wilson, 1989: A proposed microburst nowcasting procedure using single-Doppler radar. *J. Appl. Meteor.*, **28**, 285–303, doi:10.1175/1520-0450(1989)028<0285:APMNPU>2.0.CO;2.

Rose, C. L., L. B. Murphy, L. Byard, and K. Nikzad, 2002: The role of the big five personality factors in vigilance performance and workload. *Eur. J. Pers.*, **16**, 185–200, doi:10.1002/per.451.

Stumpf, G. J., T. M. Smith, K. Manross, and D. L. Andra Jr., 2008: The Experimental Warning Program 2008 Spring Experiment at the NOAA Hazardous Weather Testbed. Preprints, *24th Conf. on Severe Local Storms*, Savannah, GA, Amer. Meteor. Soc., 8A.1. [Available online at http://ams.confex.com/ams/pdfpapers/141712.pdf.]

Szalma, J. L., 2002: Individual differences in the stress and workload of sustained attention. *Proc. Human Factors Ergonomics Society 46th Annual Meeting*, Baltimore, MD, Human Factors and Ergonomics Society, 1002–1006, doi:10.1177/154193120204601201.

Trapp, R. J., S. A. Tessendorf, E. S. Godfrey, and H. E. Brooks, 2005: Tornadoes from squall lines and bow echoes. Part I:

Climatological distribution. *Wea. Forecasting*, **20**, 23–34, doi:10.1175/WAF-835.1.

Vasiloff, S. V., 2001: Improving tornado warnings with the Federal Aviation Administration's Terminal Doppler Weather Radar. *Bull. Amer. Meteor. Soc.*, **82**, 861–874, doi:10.1175/1520-0477(2001)082<0861:ITWWTF>2.3.CO;2.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences.* 2nd ed. Academic Press, 467 pp.

Wurman, J., D. Dowell, Y. Richardson, P. Markowski, E. Rasmussen, D. Burgess, L. Wicker, and H. B. Bluestein, 2012: The second Verification of the Origins of Rotation in Tornadoes Experiment: VORTEX2. *Bull. Amer. Meteor. Soc.*, **93**, 1147–1170, doi:10.1175/BAMS-D-11-00010.1.

Young, M. S., and N. A. Stanton, 2006: Mental workload: Theory, measurement, and application. *International Encyclopedia of Ergonomics and Human Factors*, W. Karwowski, Ed., CRC Press, 818–812.

Zrnić, 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.