The Creation of a Research Television Studio to Test Probabilistic Hazard Information with Broadcast Meteorologists in NOAA's Hazardous Weather Testbed

HOLLY B. OBERMEIER,[a,b] KODI L. BERRY,[b] KIMBERLY E. KLOCKOW-MCCLAIN,[a,b] ADRIAN CAMPBELL,[a,b] CAROLINE CARITHERS,[c] ALAN GERARD,[b] AND JOSEPH E. TRUJILLO-FALCÓN[a,b]

[a] *Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma, Norman, Oklahoma*
[b] *NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma*
[c] *WRKG-TV, Mobile, Alabama*

(Manuscript received 19 November 2021, in final form 3 June 2022)

ABSTRACT: Broadcast meteorologists play an essential role in communicating severe weather information from the National Weather Service to the public. Because of their importance, researchers incorporated broadcast meteorologists in the development of probabilistic hazard information (PHI) in NOAA's Hazardous Weather Testbed. As part of Forecasting a Continuum of Environmental Threats (FACETs), PHI is meant to bring additional context to severe weather warnings through the inclusion of probability information. Since this information represents a shift in the current paradigm of solely deterministic NWS warnings, understanding end user needs is paramount to create usable and accessible products that result in their intended outcome to serve the public. This paper outlines the establishment of "K-Probabilistic Hazard Information Television" (KPHI-TV), a research infrastructure under the Hazardous Weather Testbed created to study broadcast meteorologists and PHI. A description of the design of KPHI-TV and methods used by researchers are presented, including displaced real-time cases and semistructured interviews. Researchers completed an analysis of the 2018 experiment, using a quantitative analysis of television coverage decisions with PHI, and a thematic analysis of semistructured interviews. Results indicate that no clear probabilistic decision thresholds for PHI emerged among the participants. Other themes arose, including the relationship between PHI and the warning polygon, and communication challenges. Overall, broadcast participants preferred a system that includes PHI over the warning polygon alone, but raised other concerns, suggesting iterative research in the design and implementation of PHI should continue.

SIGNIFICANCE STATEMENT: Broadcast meteorologists are the primary source of severe weather warning information for the U.S. public. As a result, researchers at NOAA's National Severe Storms Laboratory and the Cooperative Institute for Mesoscale Meteorological Studies developed a mock television studio to allow broadcast meteorologists to use and communicate experimental products "on air" as part of the research-and-development process. Feedback provided by broadcasters is incorporated into products through an iterative process. Since 2016, 18 broadcasters have tested probabilistic hazard information at the warning time scale (0–1 h) for severe wind and hail, tornadoes, and lightning.

KEYWORDS: Severe storms; Tornadoes; Probability forecasts/models/distribution; Broadcasting; Communications/decision making; Societal impacts

1. Introduction

For decades, broadcast meteorologists have added value to NWS severe weather information and are crucial intermediaries between NWS forecasters and the public. Despite their importance, the role of these important practitioners in weather communication is rarely researched (Wilson 2008; Demuth et al. 2009; Doherty and Barnhurst 2009; Keul and Holzer 2013). However, broadcast meteorologists remain the primary source of NWS weather warnings for most of the general population (Silva et al. 2017, 2018, 2019; Krocak et al. 2020), despite the rise in smartphone notifications. Viewers cited trust of broadcaster advice as a main influence in deciding to seek shelter (Hammer and Schmidlin 2002; Sherman-Morris 2005). Therefore, the research and development process of future NWS severe weather products must be user centric, considering the needs of NWS forecasters *and* core partners, including broadcasters. The job duties of these users have differing responsibilities and inherent pressures; however, user-centered research methodologies increase the likelihood that new products will be successful upon transition for all groups. An example of this type of success was the WSR-88D upgrade. Researchers engaged the television broadcast industry *before* the NEXRAD installation through nationwide surveys (Robertson and Droegemeier 1990). Results showed broadcasters were willing to use the technology and wanted to attend regional educational workshops. Today, NEXRAD is perhaps the most critical component of television weather coverage. The work in the paper seeks to expand end-user engagement beyond surveys to include full-fledged usability testing. To accomplish this, broadcast meteorologists (2016 to present) were incorporated into the probabilistic hazard information

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/WCAS-D-21-0171.s1.

Corresponding author: Holly Obermeier, holly.obermeier@ou.edu

DOI: 10.1175/WCAS-D-21-0171.1

(PHI) experiments (Calhoun et al. 2014; Karstens et al. 2015, 2018) in NOAA's Hazardous Weather Testbed (Calhoun et al. 2021). In these Hazardous Weather Testbed experiments, initial object-based probabilistic forecasts for severe wind/hail, tornadoes, and lightning (known as ProbSevere for wind and hail; Cintineo et al. 2014; Calhoun et al. 2018) were created and/or modified by NWS forecasters to produce grid-based probabilistic forecasts (i.e., PHI plumes) using the NSSL Prototype PHI Tool (Karstens et al. 2015, 2018) (Fig. 1). The PHI plumes moved with the identified hazards, updating through machine and/or forecaster-controlled automation every 2 min. In these experiments, NWS forecasters created storm-based PHI, which was then displayed in an experimental version of the NWS Enhanced Data Display (Wolfe 2014) (Fig. 1). Broadcast meteorologists used the information displayed in the experimental Enhanced Data Display to make decisions in a simulated work environment. The NWS forecasters, broadcast meteorologists, and emergency managers functioned as an integrated warning team, using NWSChat (https://nwschat.weather.gov) to interact across rooms.

The PHI experiments are part of the broader Forecasting a Continuum of Environmental Threats initiative (FACETs; Rothfusz et al. 2018). FACETs is a framework to modernize the NWS forecast and warning paradigm from a reliance on deterministic products (e.g., watches and warnings) to one that adds gap-filling, continuously updating PHI spanning from days to within minutes of an event for all environmental hazards.

The initial vision for FACETs specified that social and behavioral sciences research would identify appropriate probabilistic thresholds for automated warnings for tornadoes, severe wind/hail, and cloud-to-ground lightning. As a result, multiple methods for automated severe warnings were tested in the Hazardous Weather Testbed with broadcast meteorologists and emergency managers. In the 2015 experiment, emergency managers received PHI only (no warnings). Emergency managers stated that NWS warnings served as triggers in their emergency operations plans. Without them, participants individually attempted to determine a probabilistic threshold they felt warranted a warning (Karstens et al. 2018). During the 2016 experiment, forecasters chose their own probabilistic threshold to define the warning polygon boundaries. Forecasters were inconsistent on what probability threshold constituted a warning (Karstens et al. 2018). In addition, the higher the threshold a forecaster specified for the warning, the smaller the resulting warning polygon. Broadcast meteorologists found a warning product was needed to justify coverage decisions to station management, beyond relying on probability alone. In 2017, the forecaster participants used 0% as the warning threshold. While this approach created sufficiently large warning polygons (Karstens et al. 2018), broadcast meteorologists found it extremely difficult to communicate the need to seek shelter at such low probabilities of tornadoes. During the 2018 experiment, broadcast meteorologists and emergency managers received PHI and the actual warning polygons issued by the NWS during each event (Fig. 1). This approach would ideally allow the broadcasters and emergency managers to make decisions at their own discretion using both products.

This paper documents the development and evolution of a recurring experiment in the Hazardous Weather Testbed in which broadcast meteorologists tested and evaluated the usability and communication of warning-scale PHI on air in a simulated television studio environment. Researchers summarize the evolution of a research television studio and discoveries from four years of testing how broadcasters could present PHI on air.

2. Method

a. *Evolution of a research television studio*

Researchers developed K-Probabilistic Hazard Information–Television (KPHI-TV) over a period of three years as a research television studio environment. In the framework of the Hazardous Weather Testbed, broadcasters performed work routines typical of severe weather events while using experimental products issued by NWS forecasters. Researchers used the Oklahoma Weather Laboratory to operate KPHI-TV. The Oklahoma Weather Laboratory is a student-run forecasting organization within the University of Oklahoma School of Meteorology that produces short, routine weather broadcasts made by students for the University of Oklahoma Nightly television program. With its modern broadcasting equipment and setup, the Oklahoma Weather Laboratory proved to be a flexible and modifiable space to accommodate both student and researcher needs.

Beginning in 2016, a single broadcast meteorologist participated as a member of an integrated warning team in each of the three weeks of the PHI Prototype experiment, under the NWS Hazard Simplification project (Eastern Research Group 2016; Nemunaitis-Berry and Obermeier 2017; Obermeier et al. 2017; LaDue et al. 2016). Broadcasters tested both PHI and hazard simplification messaging while using the Enhanced Data Display to visualize and communicate PHI. The participant stood in front of two wall monitors to perform improvised on-air coverage while a web camera streamed the video and audio to NWS forecasters and emergency managers in separate rooms. Researchers sat nearby, observed, and took notes. The studio concept was effective for basic experimental needs but lacked several more complex factors. The design proved unrealistic for test cases that featured high-impact severe weather threats, which is often the circumstance in Hazardous Weather Testbed experiments. In a real-world setting, a team of broadcast meteorologists, rather than a single person, would handle high-impact severe weather coverage. Additional support from the newsroom (e.g., reporters, anchors, producers, and directors) would be common. Many work duties and pressures that would be expected in a television studio environment were not represented. Realizing these shortcomings, researchers sought to incorporate additional elements of realism over subsequent years of the experiment.

In the 2017 PHI Prototype experiment, one broadcaster participated each week and researchers retained the double monitor studio format. To add more depth and realism for the participants, researchers added both an assigned local designated market area (DMA) and a television programming

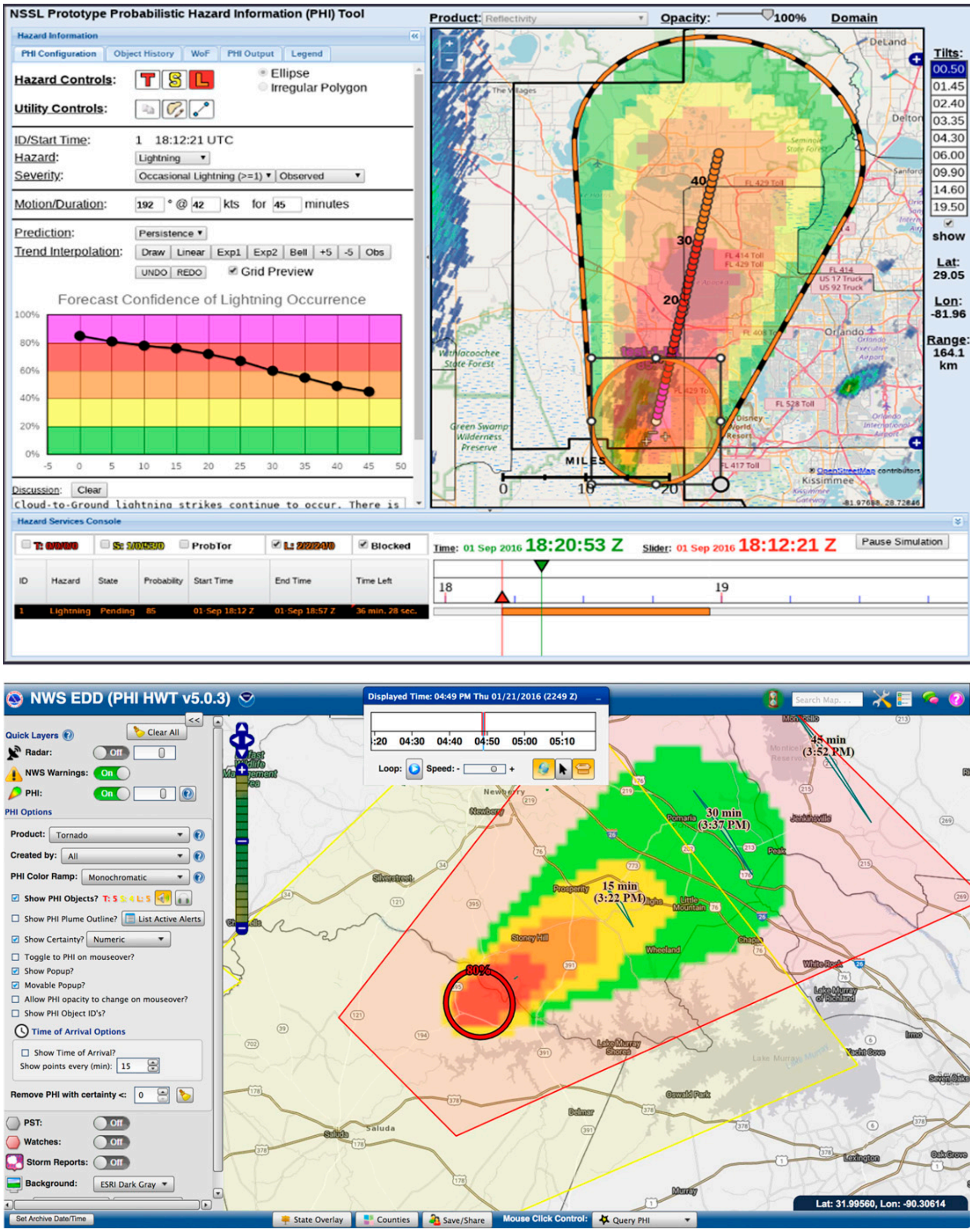

FIG. 1. (top) The graphical user interface of the NSSL Prototype PHI Tool, displaying the creation of a PHI object and probability plume by an NWS forecaster. Within the user interface, an NWS forecaster could manipulate the storm vector and a trend line of probability over time. The forecaster could also change the shape of the probability plume. (bottom) The graphical user interface of the Enhanced Data Display, displaying a PHI object and plume along with warning polygons. Broadcast meteorologists could toggle the PHI and warning layers on or off. In both images, "warmer" tones such as red and purple indicated higher probabilities and "cooler" colors such as yellow and green corresponded to lower probabilities.

FIG. 2. The 2018 KPHI-TV studio, with one participant using the green screen (far middle) and another participant at the social media desk (lower right). A student director is seated (lower left). Researchers and observers fill the remaining space.

schedule. Researchers assigned the DMA based on case location. Schedules included programming from all four major networks (ABC, CBS, NBC, and Fox). Depending on the time and date of the case, the rundown could include both local newscasts and programming content. In addition, participants were informed that ratings periods, or "sweeps," were under way if the case occurred during February, May, July, or November. During sweeps, television stations calculate their audience size, a critical measurement that ultimately determines advertising revenue (Buzzard 2015). Adding the combination of geographic location and television programming required a realistic analysis of the severe weather threat before a coverage decision was made by the participant.

Other technological additions included text-crawling software that allowed the participants to construct and run their own custom messages along the bottom of the television screen. To expand visualization capabilities, PHI placefiles were incorporated into GR2Analyst (http://www.grlevelx.com). GR2Analyst allowed participants to show PHI plumes along with any radar variable, including dual-polarimetric variables. While a webcam captured video and audio in 2017, the content was confined to a private channel and not broadcast to other participants. This arrangement allowed the broadcaster to retain the effect of feeling "on air" but reduced overall performance burdens.

Additional funding in 2018 allowed researchers to fully expand KPHI-TV resources. Upgrades included a green wall, high-definition camera, source switchboard, lapel microphones for audio, studio lighting, and a social media desk complete with protected Facebook and Twitter accounts (Fig. 2). The 2018 experiment included two broadcasters in each of the three weeks working as a team, for a total of six participants. One broadcaster was stationed on-camera against the green screen while the other was stationed at the social media desk. Both had active, or "hot," microphones and could banter back and forth during on-air coverage. An undergraduate student acted as a director and handled the switchboard. This student changed sources between the Enhanced Data Display or GR2Analyst at the request of the on-camera broadcaster. Depending on the case location and population size of the DMA, participants were given the option of asking for sources such as mock helicopter coverage, live shots, or tower cameras. They could also "toss" coverage to a fictitious news anchor or field reporter, which allowed the participants to take a brief break or drink of water. The culmination of the design features implemented from 2016 to 2018 led to enhanced testing and technological stability with each successive iteration of the experiment.

b. *Participant selection*

Using snowball sampling (Atkinson and Flint 2001), researchers sought participants by distributing a nationwide recruitment letter during early 2018. Distribution sources included the American Meteorological Society, the National Weather Association, and NWS media and emergency management email lists through local weather forecast offices. Selection criteria of most importance included availability, geographic location of the applicant's home DMA, typical shift (e.g., mornings vs evenings; weekdays vs weekends) and number of years of experience. Broadcast meteorologists on different shifts often have varying responsibilities, especially with regard to how newscasts are structured. In addition, storm mode can vary with region or time of day (Smith et al. 2012). Researchers hoped to capture applicants with a range of backgrounds concerning these criteria. Researchers also recognized the need for diversity in gender, race, ethnicity,

TABLE 1. Broadcast meteorologist participant employment attributes for the 2018 Hazardous Weather Testbed PHI experiment.

Participant	Years of experience	DMA size	DMA location	Shift time
A	12	Large	Tennessee	Evening
B	13	Large	Connecticut	Evening
C	14	Small	Mississippi	Evening
D	18	Large	Tennessee	Morning
E	8	Large	Michigan	Morning/evening
F	8	Small	Virginia	Evening

age, and types of educational training, and strove to select a mix of participants with differing kinds of backgrounds (Table 1). The six selected broadcasters in 2018 had television experience ranging from 8 to 18 years. DMA representation included markets from all sizes (small to major), and covered Midwest, northeast, and southern states. Most of the participants had previously worked in other DMAs across the country and brought that experience into the context of the experiment. To maintain anonymity in accordance with institutional review board requirements, each of the six participants are represented by an alphabetical pseudonym (A–F).

c. *Experiment design and data collection*

The 2018 experiment took place over three weeks in June. Prior to each week, the participants received an emailed form to complete and return. The form was a modified cognitive task analysis (CTA; Hoffman 2005), in which participants were asked to list and define their work responsibilities according to the Storm Prediction Center's (SPC) day-1 convective outlook. Researchers asked the participants to document their routines and workload according to days with no risk, slight risk, and moderate/high risk for severe weather. Since little formal information exists about the day-to-day procedures of a broadcast meteorologist, researchers hoped to gain more thorough insight using the CTA. Following completion of this step, each pair of selected broadcast participants traveled to Norman, Oklahoma, for their experiment week. Each week followed the same schedule (Table 2). The day-1 schedule started with introductions, followed by a detailed dialogue with each participant reviewing their completed CTA. Researchers walked each participant through every detail of the broadcasters' duties, which typically began well before their formal shift started and sometimes ended hours later than their shift ended, depending on the severity of the hazardous weather event. The CTA interviews allowed researchers to establish a critical baseline for broadcast meteorologists' duties and decisions on nonsevere and severe weather days (Ernst 2020).

Next, researchers conducted a semistructured interview with both participants (see also all semistructured interview guides in the online supplemental material). The day-1 interview allowed researchers to gather participant feedback about the current watch, warning, and advisory paradigm. Specifically, researchers were interested in learning about participants' views concerning any spatiotemporal gaps in information and perception of probabilistic information, before being exposed to the suite of experimental PHI products. This early week feedback provided researchers with the broadcasters' initial and unbiased perceptions. After a lunch break, a training case was simulated in the afternoon. During this case, participants became oriented with the concept of the PHI plume, PHI objects, ProbSevere, the Enhanced Data Display, GR2Analyst, and the studio layout. Researchers collected no data during this training activity.

Days 2–4 featured a consistent schedule in which participants simulated a case in the morning and afternoon, for a weekly total of six cases with differing hazards, SPC convective outlooks, and watch types (Table 3). These archived cases represented a variety of DMAs across the nation and severe convective hazards, including tornado, wind, and hail (Fig. 3). One case focused solely on experimental cloud-to-ground lightning PHI (Calhoun et al. 2018; Meyer et al. 2019). Prior to each case, participants watched a prerecorded weather briefing that included both synoptic and mesoscale observations, SPC day-1 convective outlooks, SPC mesoscale discussions, and watches for the case area. Participants then worked through the case, which was typically less than 2 h long. During cases, participants focused on executing their work duties (such as radar interrogation, on-air coverage, social media updates, and checking NWSChat), but were encouraged to think aloud or ask questions. Researchers administered postcase semistructured interviews after each case. The interview questions included participants' thoughts on multiple aspects of PHI: usefulness for their decision-making, characteristics that were or were not helpful, and perceived challenges. On day 5, participants completed a final semistructured interview covering their impressions from the entire week. Questions covered such topics as the participants' overall thoughts about PHI and whether they would or would not use PHI operationally.

In addition to the CTAs and semistructured interviews, researchers collected many other data and media before, during, and after simulated cases. Because the broadcast experiment recreates a television studio environment, these data included the following:

- television video and audio;
- television coverage decisions, including probabilistic thresholds, hazard type, programming and coverage type, and length of time on air;
- social media posts, including Twitter and Facebook (Kolakoski et al. 2019; Trujillo et al. 2020);
- NWSChat logs;
- modified NASA task load index surveys (Hart and Staveland 1988; Hart 2006); and
- researcher case observation notes.

TABLE 2. Schedule for each week of the 2018 Hazardous Weather Testbed PHI experiment.

	Day 1 (Mon)				Days 2–4 (Tue–Thurs)				Day 5 (Fri)
Before week	Morning			Afternoon	Morning		Afternoon		Morning
Preweek CTA form	Introductions	In-depth CTA review	Day-1 semistructured interview	Training case (no data collected)	Weather briefings and cases 1, 3, and 5	Postcase semistructured interviews	Weather briefings and cases 2, 4, and 6	Postcase semistructured interviews	End of week semistructured interview

Researchers documented the probability of the PHI objects each time a coverage decision was made, as well as the storm identifier, type of hazard, type of coverage, and length of time on air (if applicable). Tracking these elements allowed researchers to assess probabilistic thresholds for decision-making by each participant, hazard type, and coverage type.

As part of the Enhanced Data Display functionality, broadcasters could change the update frequency of the PHI. Options included 20-, 10-, 5- or 2-min updates. The update speed each participant chose at the beginning of every case was documented to determine an ideal PHI update frequency to meet the needs of the broadcast media. Researchers also investigated whether the participants preferred different color schemes of PHI based on previous research with the public (Ash et al. 2014; Miran et al. 2017; Klockow-McClain et al. 2020). To explore broadcasters' color preferences for communication to their viewers, researchers provided PHI in both rainbow and monotonic color schemes (Fig. 4). Researchers acknowledged that color schemes could be confusing for those experiencing color vision deficiency. To date, no participants have expressed any specific problems with the presented color schemes.

d. *Thematic analysis*

A single coder conducted a qualitative thematic analysis (Braun and Clarke 2006) on the semistructured interviews from the day-1, postcase, end-of-week, and researcher case observation notes to identify common themes and feedback between participants. To corroborate these documents, the audio recordings from the experiment were transcribed and content was compared to verify the completeness. Patterns were drawn from the codes to devise and verify the themes. After verifying that the themes were representative of the data and research questions, themes were named and defined. Concrete examples from the transcripts were connected to themes.

3. Results

a. *Coverage decisions in the current warning paradigm*

Five types of coverage decisions emerged during the 2018 experiment that included: on-air messages known as crawls; short on-air "cut-ins" (often 5 min or less); weather segments during local newscasts (including a primary weathercast and any additional segments); continuous on-air coverage ("wall to wall"); and social media posts (Facebook and Twitter). The day-1 interviews revealed that participants typically balance several factors when making coverage decisions concerning severe weather. These factors include the hazard type, presence (or nonpresence) of a watch/warning, SPC day-1 convective outlook, television programming, ratings periods, management/newsroom policies, and the population/geography of their DMA. Factors were weighted with importance depending on each unique scenario. Ultimately, participants expressed that all coverage decisions were situationally dependent to a degree, and no firm rules exist across the industry. Some hazard types took precedence in coverage over others, and this precedence was often dictated by the presence of a watch and/or warning.

TABLE 3. Attributes for the six archived displaced real-time cases.

Case date	Primary hazard	Secondary hazard	State	Watch status	SPC convective outlook
21 Jan 2016	Tornado	Straight-line winds	MS	Tornado watch	Enhanced
25 May 2016	Tornado	Straight-line winds	KS	Tornado watch	Slight
25 May 2017	Tornado	Hail	KS	Severe thunderstorm Watch	Enhanced
24 May 2017	Straight-line winds	Hail	SC	Tornado watch	Enhanced
1 Sep 2016	Lightning	Tornado	FL	Tornado watch	Slight
22 July 2016	Lightning	None	CO	None	General Thunderstorm

1) TORNADO WARNINGS

According to all of the 2018 participants, tornado warnings were an immediate cue to begin continuous weather coverage, no matter the current programming or if sweeps were under way (Table 4). Most often, coverage would continue until the warning expired. If coverage ended before the tornado warning expiration, it would often resume with subsequent updates to the warning. Also, the participants' own analysis of radar signatures and viewer reports influenced their coverage decisions.

2) SEVERE THUNDERSTORM WARNINGS

Unlike tornado warnings, participants typically used short cut-ins as coverage for severe thunderstorm warnings, primarily during commercial breaks (Table 4). Most voiced that television programming played an important role in this decision. Some programs have higher viewership than others, which is often related to the time of day. Many participants listed shows such as *Judge Judy*, *Wheel of Fortune*, *The Voice*, *The Bachelor/Bachelorette*, soap operas, and sporting events as programming they prefer not to interrupt with a severe thunderstorm warning. The decision to interrupt programming was made more difficult if the case occurred during sweeps, especially if the severe threat was marginal. Participants even expressed dread about program interruption because of past experiences with negative viewer feedback. Most of the participants felt that cut-ins for severe hail and wind were best reserved for national commercial breaks, unless significant hail [>2-in. (5.1 cm) diameter] or wind [>70 mi h^{-1} (113 km h^{-1})] were expected.

3) GENERAL COVERAGE POLICIES/PRACTICES

Participants preferred cutting in over national commercial breaks, since local markets focus on the advertising revenue from local commercials. Other times, participants chose to cut in over station identification. Station identifications typically occur at the end of a commercial break and can last approximately 15 s. For network programming, station management policies could be strict about cut-in decisions. Participants from network owned-and-operated stations said that the choice to interrupt primetime network programming with weather coverage could be a very complicated decision, which was weighted heavily toward input from station management. However, in the majority of the participants' DMAs, these restrictions were less prevalent and broadcast meteorologists were the primary decision-makers concerning weather coverage. With regard to

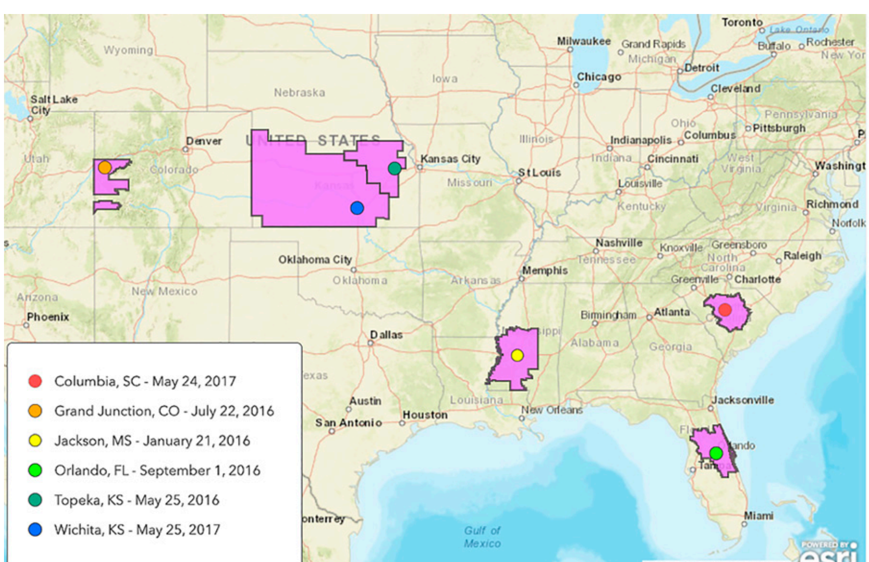

FIG. 3. Primary market locations and dates of the six archived cases simulated each week of the 2018 Hazardous Weather Testbed experiment. DMA coverage regions for the case locations are shaded in purple (Sood 2016).

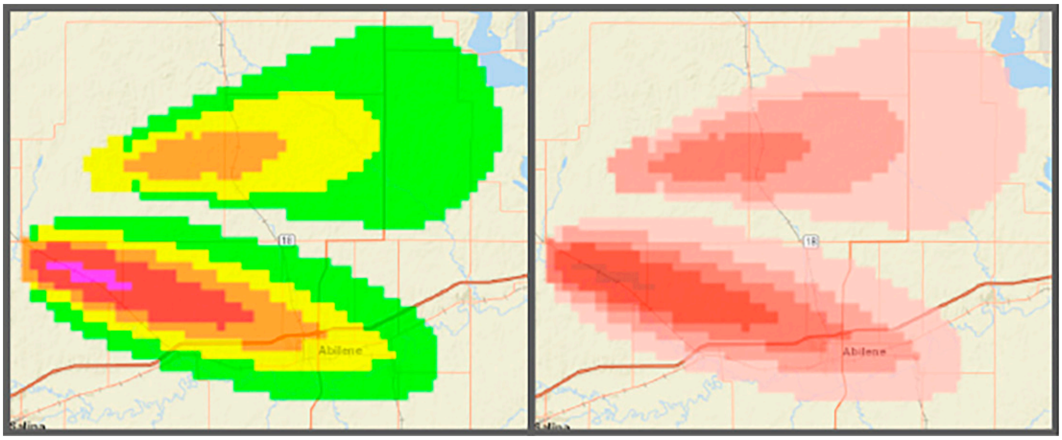

FIG. 4. Comparison of rainbow tornado plumes and red monotonic tornado plumes.

social media, broadcasters expressed a great deal of freedom in their ability to share severe weather information, both in frequency and amount of content. The interviews revealed that the participants shared numerous posts concerning weather information daily, and that social media was a useful tool for severe weather coverage.

b. Coverage decisions in the PHI paradigm

Results from the 2018 experiment indicate that the addition of PHI into the warning paradigm added more complexity into participants' coverage decisions and communication. Because most participants had an established mental checklist for making coverage decisions, it was not immediately clear how PHI should fit into their workflow. To understand their thoughts on probability, both interview questions and a statistical analysis of their decisions were implemented. As part of the day-1 semistructured interview, participants were asked their thoughts about what percent probability of severe weather they *perceived* should warrant television coverage: "Thinking about likelihood in terms of numeric probabilities, at what probability level do you feel immediate coverage is warranted from the threat of tornado, severe, and lightning?" (Table 5). The same question was asked in the end-of-week interview on day 5. For the duration of the experiment week, the participants' perceptions of tornado probabilities changed very little. For severe, half of the participants' perceived thoughts on probabilistic thresholds increased. For the other half, the answer shifted from a numeric probability to a qualitative response. For example, participant F changed their answer from "20%–30%" on day 1 to "intensity dependent" on day 5, perhaps suggesting that after using PHI, the participant considered factors beyond hazard probability. This was not surprising, since most participants described so many factors they consider when making a coverage decision.

Over the course of the 3-week experiment, participants made a total of 377 coverage decisions using all five modes of coverage types (crawl, social media, weathercast, short cut-in, continuous). Of these coverage decisions, 308 were made using PHI guidance. To investigate if any common probabilistic threshold patterns occurred, decisions were analyzed by the PHI value of the hazard when the decision was made. When examined by hazard type alone, PHI thresholds for tornado coverage decisions were lowest (median = 65%), followed by severe (median = 80%) and lightning (median = 92%). Variability existed, likely due to the other influencing factors as described in the answers to the CTAs and day-1 interviews. To investigate if thresholds varied depending on coverage type, decisions were examined under the lens of both hazard and coverage type (crawl, social media, weathercast, short cut-in, continuous; Fig. 5). For tornado PHI (Fig. 5, top left), participants performed short cut-ins at a comparably low threshold (median = 25%) with much less variability (standard deviation = 8%). The median thresholds were much higher for the other types of coverage. For severe thunderstorm PHI (Fig. 5, top right), the threshold was also lower for short cut-ins relative to the other coverage types (median = 42%; standard deviation = 25%). More variability existed for social media posts, although the

TABLE 4. Answers from each participant to the day-1 interview question: "What is your policy on breaking into programming/commercials during severe weather?" Answers are in reference to current television station policy.

Participant	Severe thunderstorm warning	Tornado warning
A	No cut-in [unless highly populated area affected, or threat of significant hail/wind]	Continuous coverage
B	Cut-in during national commercial breaks	Continuous coverage
C	Typically, no cut-in	Continuous coverage
D	Cut-in during national breaks; situational	Continuous coverage
E	Situational	Continuous coverage
F	Cut-in during commercial breaks	Continuous coverage

TABLE 5. Answers from each participant to the day-1 and end-of-week question: "Thinking about likelihood in terms of numeric probabilities, at what probability level do you feel immediate coverage is warranted from the threat of tornado, severe, and lightning?"

	Tornado		Severe		Lightning	
Participant	Day 1	Day 5	Day 1	Day 5	Day 1	Day 5
A	25%	30%	50%	75%	"No action"	"No action"
B	30%	30%	60%	75%	"No action"	"No action"
C	10%	"Depends"	70%–90%	"Higher in general"	"100% during outdoor events"	"Cover during newscasts"
D	30%	20%	50%	"Higher but conditional"	"No action"	"Location dependent"
E	20%	30%	30%	65%	60%	100%
F	20%	20%	20%–30%	"Intensity dependent"	50%	"Timing dependent"

median threshold was 90%. Lightning was solely covered by social media posts, except for two short cut-ins (Fig. 5, bottom). Figure 6 shows coverage decisions according to each participant by hazard. The median threshold for a coverage decision for tornado threats was lower than for severe or lightning threats. Variability in thresholds for each participant was common.

Figure 7 illustrates the percentage of decisions allocated for each hazard type and coverage type, along with the duration of on-air coverage. Despite social media coverage representing the greatest percentage of decision points, three other coverage types (continuous, short cut-in, or weathercast) resulted in nearly 17 h of combined on-air coverage between all six participants. Of these 17 h, approximately 15 h were spent covering primarily tornado PHI and approximately 2 h were spent covering severe PHI. Only about 2 min were spent covering lightning PHI on air. Having an additional participant in the 2018 Hazardous Weather Testbed experiment focused on social media allowed for a much higher volume of posts than in 2016 or 2017, especially to Twitter. Posting severe weather information to social media did not require interrupting the television station's main source of revenue, translating to less financial risk and potential viewership frustration. Twitter or Facebook posts were a lower-risk method for the participants to spread PHI quickly as updates arrived from the NWS forecasters. Also, some participants were much more prolific posters to social media than others. One participant posted 108 times throughout the week, which was almost 2 times the total number of posts of the next highest participant. A comparison of social media posts in the Hazardous Weather Testbed with those from the actual events revealed that posts created by participants using PHI included more specific temporal and spatial information (Kolakoski et al. 2019; Trujillo et al. 2020).

c. *Thematic analysis*

Two primary themes emerged in the analysis of the semi-structured interviews:

1) the complex relationship between PHI and the warning polygon, including
   - preference for PHI,
   - need for a warning threshold or trigger,
   - inconsistencies between the two products, and
   - importance of trends in probability with each update (on the order of minutes), and
2) communication challenges in visually and verbally telling the "weather story" to the audience, including
   - messy, cluttered displays,
   - the usefulness of ProbSevere/PHI object,
   - PHI color schemes,
   - words for PHI, and
   - update frequency.

Throughout the experiment, the complex relationship between the PHI plume and the warning polygon was apparent. Participants freely experimented with combinations of both products for all hazards, but consistently voiced their preference for PHI and its usefulness. The addition of probabilistic information added more context than participants felt they received from the warning polygon alone:

> The plume allows you to focus on the highest probabilities; it gives you a graded product instead of a "yes or no." (Participant D)

> I liked the plumes better because of the shades of certainty that they communicated. They're more focused than the legacy warnings. Did layer both in some tweets. (Participant D)

> To be honest, if this were me, I would base all of my decisions on the PHI and not the polygon to stay on or off air. (Participant F)

> It [PHI] is a product that works best by itself. Having that and radar is a lot going on. If you turn everything else off, it's [PHI] pretty self-explanatory. (Participant D)

> There's this relief knowing you're not missing something when you have that PHI layer. You can take a part of your brain from radar analysis and focus on the communication of the hazard. (Participant E)

Despite the preference for PHI, participants still voiced the need for a warning threshold or trigger of some kind. Current decision practices are based on the issuance of warnings, and participants still wanted a warning product:

> There has to be some threshold to trigger a warning. (Participant B)

> The boxes [warning polygons] were extra assurance, comfort level that we've had for years. (Participant E)

> . . . there should be a tornado warning threshold . . . I can see a news manager saying "you cannot go on the air unless the tornado probability is over 100%. It can add a lot of ambiguity to coverage decisions. (Participant D)

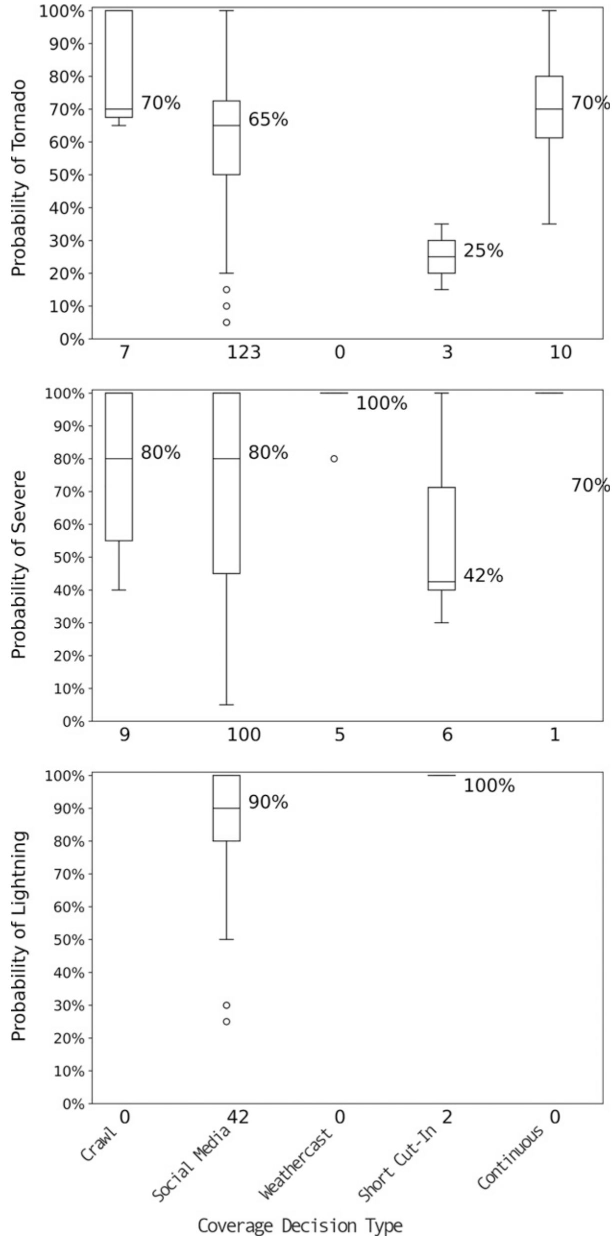

FIG. 5. Boxplots of hazard probability values for each type of coverage decisions for (top) tornado, (middle) severe, and (bottom) lightning. Each of the five boxes represents a different type of coverage, along the *x* axis from left to right (crawl, social media, weathercast, short cut-in, and continuous). The number of decisions for each coverage type is also listed along the *x* axis. The median value for each type of coverage is shown to the right of each box.

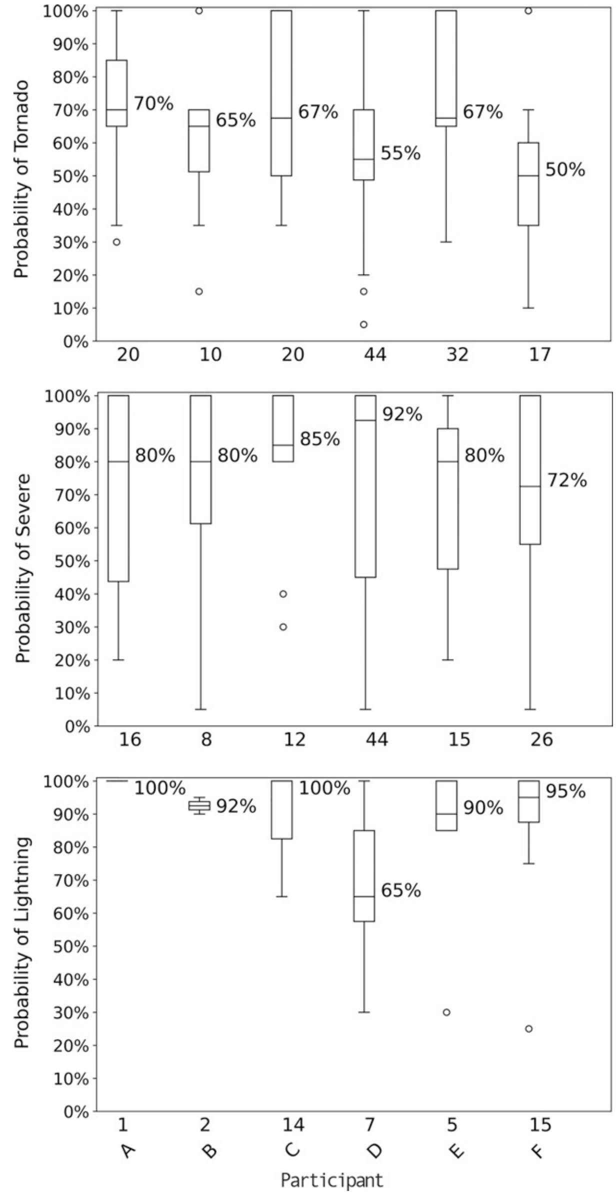

FIG. 6. Boxplots of hazard probability values for coverage decisions by each participant for (top) tornado, (middle) severe, and (bottom) lightning. Each of the six participants is represented by an alphabetical letter along the *x* axis, from A to F. The number of decisions made by each participant is also listed along the *x* axis. The median value of decision thresholds by each participant is shown to the right of each box.

However, the visual presentation of PHI and the warning polygon together presented some issues. These concerns were most prevalent when the two products were inconsistent and did not spatially align with one another. Feedback during and after cases indicated that discrepancies between the two products were problematic and challenging to communicate:

> Polygons may be too broad, but plumes are too specific. (Participant A)

> That becomes difficult to communicate [the two products]. (Participant A)

> I was turning off the warnings a lot . . . most of the time actually. Because explaining in 280 characters [Twitter] with both [warning and PHI] is difficult. (Participant D)

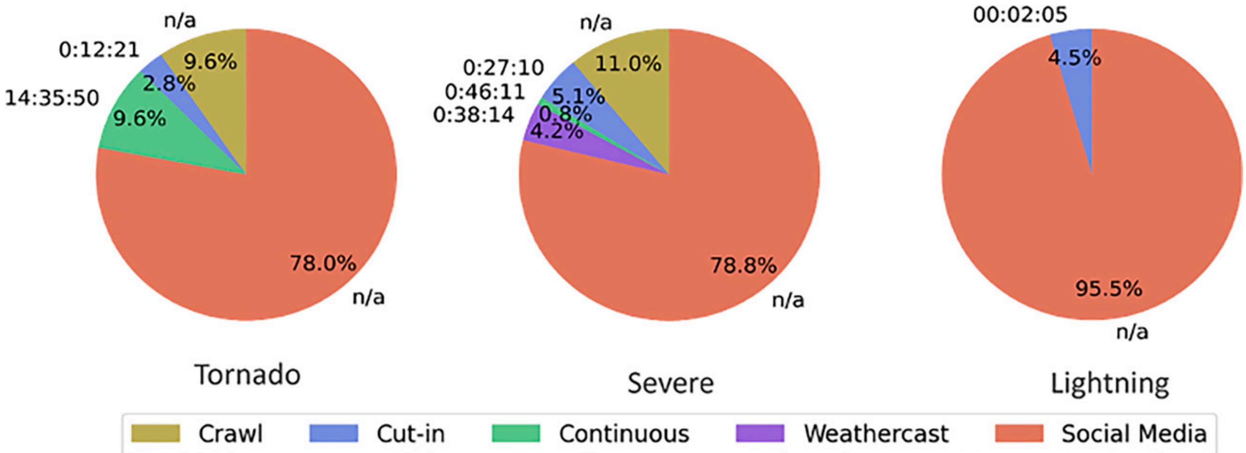

FIG. 7. Percentage of total coverage decisions according to coverage type (crawl, short cut-in, continuous, weathercast, and social media) and hazard (tornado, severe, and lightning). If applicable, total duration of on-air coverage is listed outside the pie chart (given as hours:minutes:seconds).

> From a messaging standpoint that [PHI overlaid with the warning polygon] becomes very challenging. It's two different messages. (Participant B)

> The warning and PHI have to make sense together. You can't have [a storm with] no warning with a higher probability than a storm with a warning [with a lower probability]. (Participant B)

Numerous instances occurred in which high-probability PHI existed with no corresponding warning. The opposite situation occurred as well, in which a warning was valid without the presence of a concurrent PHI plume. In addition, situations arose in which a PHI plume for a specific hazard type was collocated with a warning for a different hazard type. Feedback concerning these situations was also mixed:

> . . . so many warnings without PHI corresponding . . . I do not know what I would have done in a real-life situation. (Participant F)

> Nightmare communicating an 80% tornado plume and a severe thunderstorm polygon. (Participant B)

> [Plume and polygon] not a problem if I do my job correctly and communicate it. (Participant C)

Despite the consistency issues, participants did find some effective combinations of the two products. For example, a decreasing trend in PHI allowed a participant to end coverage and return to scheduled programming before a warning expired. Participants often noted that *trends* in probability over time were key information and would often be their trigger to investigate a PHI plume more closely:

> I enjoyed seeing the ramp up in percentage, which was another tool to get the end user [viewer] to react like I would like. I'm not just looking at graphics making up numbers. (Participant C)

> The trend up and the trend down on the tornado threat were helpful. (Participant C)

> Instead of dropping it [the tornado plume] all at once, it was sort of a gradual decrease. (Participant B)

> I liked the ones that showed up before the legacy warning came out because in some instances, I was mentally checking the time and we gave them an extra 20 minutes' heads up. That's really important. (Participant A)

Increasing probability trends in the plume, especially at lower percentages, gave participants more lead time on a strengthening hazard. Participants felt they could communicate this lead time to their viewers, hoping to allow them more time to take protective action. The potential enhancement in lead time was particularly helpful when a plume appeared before a warning polygon was issued. However, one participant mentioned that decreasing trends could have the opposite effect:

> I might see some managers trying to get people off the air at lower percentages. (Participant D)

In other words, a broadcaster may want to remain on air for a warning, but new probabilistic-oriented policies could pressure the broadcaster to end coverage before they are comfortable doing so. Overall, the *change* in the probability values over short time periods was important and sometimes provided more value than the single probability value in a given moment.

As mentioned, over the 3-week experiment, participants spent nearly 17 h on air, in addition to creating a steady stream of social media content. Such long-duration coverage requires the broadcast meteorologist to be in a constant state of information-sharing through the telling of the "weather story" of the ongoing situation. With so much information available through PHI, participants became concerned about the appearance of their on-air displays. Many struggled with how to layer the data without diminishing the message that they wished to communicate. Mentions of "messy" displays were frequent each week, especially in reference to polygons and time of arrival. Participants often spent time trying to "clean up" their displays while on air:

> It is a little busy and hard to get used to. My question is "what you are going to put on the screen to the viewer?" (Participant F)

> I filtered out the lower PHI values a few times to cut down cluttering. It wasn't due to being overwhelmed. Just because it was cluttered. (Participant D)

> Too much stuff on a graphic at one time. That is up to the meteorologist to clean it up. I can see my consultant telling me that finally we have something that can replace the radar that no one understands anyway. (Participant C)

Participants also relied heavily on the ProbSevere or PHI object guidance. This guidance was consistently layered with the plume, for all three hazards. Participants voiced that the tornado guidance was most useful and appeared nicely on air, essentially providing a visual anchor for the hazard. Exceptions occurred when the tornado guidance and PHI plumes were not well correlated with radar signatures on velocity and correlation coefficient modes. Severe and lightning objects were less useful for the participants.

> I think when I toggled over to the severe plume, it was a big area covered by the highest category. That might lead people to disregard it. (Participant D)

> All objects look the same, but no indication of severity. Tagged with sigsevere [significant severe]. SPC outlooks throw in sigsevere [significant severe] and tags for higher stuff. We need a way to distinguish between severe and sigsevere [significant severe]. (Participant B)

> I didn't really use the severe objects, but I think the tornado objects were really great for severe weather days, multiple tornadoes, especially if you were by yourself. (Participant A)

> The tornado object is very useful because it is intuitive to people at home. (Participant D)

> Wow . . . those are big blobs [lightning PHI] . . . Not really telling me much, they are so noisy. Not bringing me value, even behind the scenes. Wouldn't show them on the air, just too many of them. (Participant B)

Participants also struggled with the nomenclature of PHI. For researchers, "PHI" and "plume" are the most used jargon to describe this probabilistic information. However, participants did not always choose to use this terminology to communicate with their viewers. Instead, broadcasters spontaneously experimented with other words to communicate the plume. A range of words emerged, such as "cone" or "zone":

> I'm not a fan of the word plume, like polygon. I'm still thinking of the right word. It will come to me. (Participant C)

> Plume, bullseye (for non-moving objects), layer cake, baubles, bubbles, I think I used "danger zone" at one point, but that made me think of Kenny Loggins. (Participant D)

> I would not call it a PHI plume, if the tornado warning was based on it, we could call it a tornado warning, "something we are watching," "low risk." (Participant B)

Plume color tables were another important part of the visual "weather story" the participants tried to communicate. All six participants were immediately drawn to the rainbow color scheme at the beginning of the experiment weeks. The intensity of the reds and purples of the highest probabilistic threat regions translated well on screen. Participants felt viewers would be drawn to pay attention to the bright colors. However, concern existed about whether viewers would confuse the rainbow plumes with radar reflectivity. Also, when multiple hazard types were all displayed in rainbow colors, it became confusing to discern between each hazard. In these situations, participants experimented with the monotonic color scheme. Once the monotonic color scheme was chosen, the participant typically did not revert to rainbow. By the end of the experiment weeks, all but one participant used the monotonic color scheme exclusively.

During the 2018 PHI experiment, the default update frequency for both automatic and forecaster-manipulated PHI was every 2 min (Karstens et al. 2018). The swiftly updating speed of PHI was often overwhelming for a broadcaster working a case alone during 2016 and 2017. Concerned about this potential workload burden, researchers gave the participants the option of 20-, 10-, 5-, or 2-min update speeds and allowed them to choose which was most comfortable. The addition of a second participant in the 2018 experiment allowed the team to more easily handle the demands of 2-min updates, and all participants wanted to receive new information as quickly as possible, especially during the less severe cases. Despite some degree of comfort with the 2-min option, concerns existed.

> For me it felt like with it [PHI] moving and wobbling so much, it may turn this way and that way, it turns into the cat with a laser pointer. I feel more comfortable stepping back when the motion is uncertain. (Participant A)

> . . . during social I was overwhelmed. By the time I composed a tweet and got a screen grab and looked at towns, it was old and I needed to grab a new shot. (Participant E)

> Two minutes would have been fine if the forecaster was not so specific with the plume. (Participant B)

Participants frequently mentioned looking for new information in the forecaster discussion box within the Enhanced Data Display with each update. The updates that provided the most value were those that included a fresh note from the NWS forecaster. These notes helped the participants to confirm their own observations and increased their confidence in their decisions. Questions still existed concerning the ability of current television station technology to adapt to the dissemination of rapidly flowing information in a real-life setting, such as the crawl and social media.

4. Discussion and conclusions

Perhaps the clearest result of the work in KPHI-TV was that all participants chose to use PHI in their decision-making, and preferred PHI over the warning polygon alone. While PHI added another layer of complexity to their decision-making process, it also allowed the broadcasters to make more informed decisions. On average, participants chose to make coverage decisions about tornado PHI at lesser probabilities than severe or lightning PHI, although variability in their thresholds was common. Participants spent most of their on-air time covering tornado PHI, at a much higher rate than severe or lightning PHI. However, social media coverage accounted for most individual decision points, meaning the participant at the social media desk was very busy creating content with the swift updates to the PHI. At times, participants felt that their tweets were outpaced by the PHI updates and the tweet information was outdated once posted.

According to the participants, PHI allowed them to justify remaining on air longer and to return to scheduled programming sooner as threats diminished. The precise coverage timing resulting from PHI benefitted both the broadcaster and, at least from a programmatic perspective, the viewer. Participants felt that the PHI plume could allow them to communicate additional lead time and arrival times, which may allow the viewer more time to take protective action. However, several issues still require additional research and testing, namely the relationship between the warning polygon and PHI. When the two products are spatially or temporally mismatched, numerous communication issues arise for the broadcaster that may lead to confusion for the viewer. In addition, no clear probabilistic decision threshold emerged during the testing period. Because of the myriad of factors considered when making coverage decisions, great variation in probabilistic thresholds exists. With this variation in mind, simply curating a warning polygon derived from PHI at a probabilistic threshold is not an adequate solution. Other Hazardous Weather Testbed research with NWS forecasters aligns with this finding (Karstens et al. 2018); perhaps the answer lies instead with a time and probability fused warning that works in tandem with PHI. Under this design, both time and probability could be considered. A proposed idea that could capture this concept is known as "Threats-in-Motion" (Wolf et al. 2013; Bates et al. 2020; Stumpf and Gerard 2021). Threats-in-Motion is a rapidly updating polygon that requires both a mix of automation and forecaster intervention. While the Threats-in-Motion prototype is currently time based, the ability to fuse probabilistic information that is closely aligned with the warning is imperative. Any eventual operational warning product must work in alignment with PHI, lest serious communication issues emerge for end users and potentially the public. Achieving and refining this alignment should be of great concern for researchers who continue to press forward with this potential new warning framework.

Other outstanding issues exist, such as the nomenclature for PHI. While researchers did not necessarily intend participants use the term "PHI," participants voluntarily made it clear that "PHI" and the word "plume" did not adequately serve their needs and may not resonate with their audiences or the public. Participants often experimented with their own words during their coverage. However, participants did not appear to settle on any specific words that worked in a more generalizable manner, suggesting that future research should investigate this issue more closely. Researchers sought guidance from the National Hurricane Center on this topic, given that the current depiction of a PHI plume evokes similarities to the National Hurricane Center Tropical Cyclone Track and Watch/Warning Graphic, unofficially known as the "cone of uncertainty." Numerous publications refer to the product as the "cone of uncertainty" as far back as 2007, but no research specifically addressed the nomenclature of the product (Eastern Research Group 2018). Researchers could choose to let a term grow organically from end users while making efforts to ensure that problematic jargon does not take hold.

Given the importance of update frequency and probability trends, the quick changes observed over the course of the life of a PHI plume must be considered carefully. Working as a team, the broadcast participants felt confident in their ability to interpret the increased information provided by 2-min update intervals. This was especially true when the NWS forecaster included update information in the forecaster discussion box. However, concern existed when rapid spatial fluctuations occurred with the plume, affecting their ability to communicate. Participants also struggled with keeping the crawl and social media updated when information arrived rapidly, especially during more severe cases. Ultimately, while participants wished to have information as quickly as possible, the rapid flow was sometimes a disadvantage.

A limitation of this study is the limited sample size of broadcast meteorologists due to logistical challenges and funding broadcasters for weeklong experiments. Each broadcaster participated in several archived scenarios, surveys, and focus groups. The resulting data gathered were comprehensive and provided numerous data points per participant. However, the data do not represent a statistically significant sample size. We are not presenting the data and conclusions as generalizable. Rather, this experiment is a foundation for future research to increase our sample size and evaluate a wider array of opinions and workflows. Although the experimental design sought to simulate a realistic environment within a television station, it is not possible to completely simulate the studio environment in an experimental setting.

The landscape of television technology is changing. Traditional severe weather coverage always consisted of a broadcaster interrupting programming, assuming the viewer was sitting at home or their workplace. However, the future of television will be on mobile devices also, meaning the viewer could be located anywhere. Advanced Television Systems Committee, version 3.0 (ATSC 3.0 or NEXTGEN TV), is a new digital television transmission standard that leverages internet capabilities and allows for advanced emergency alerting, separate from cellular networks [Advanced Warning and Response Network (AWARN) 2022]. Emergency alerts are geolocated to the user, and will allow for multiple streams of information simultaneously, such as evacuation routes and hazard mapping (AWARN 2022). Ultimately, ATSC 3.0 could shift some of the pressures of decision-making about severe weather coverage away from the broadcaster to the viewer. The implications of this shift and how ATSC 3.0 will change the current paradigm are not fully known. ATSC 3.0 is already being implemented in many domestic television markets. As researchers continue to develop and prototype products such as PHI, it is essential to keep in mind the capabilities of incoming broadcast technology.

The creation of KPHI-TV and the inclusion of broadcast meteorologists in the Hazardous Weather Testbed will continue to provide a framework for the coproduction and prototyping of research concepts and products. The number of data collected through KPHI-TV experiments have provided investigators a wealth of research and analysis potential, of which only a portion has been covered in this paper. Student opportunities have also been abundant under the umbrella of KPHI-TV, providing both research and networking prospects. Through the voices of broadcast meteorologists and others, researchers hope to continue to pave a path forward for FACETs, including valuable updates to the forecast and warning paradigm that will ideally benefit the greater public. By utilizing this user-centric, coproduction approach, the likelihood that future operational PHI will be adopted and effectively utilized by end users in an effective manner is increased.

Acknowledgments. Funding was provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreement NA16OAR4320115 and NOAA–University of Colorado Cooperative Agreement NA17OAR4320101, U.S. Department of Commerce. Special thanks are given to Tiffany Meyer, Jonathan Wolfe, James Murnan, Shawn Riley, and all of our participants for making this research possible.

Data availability statement. Because of privacy and ethical concerns outlined in human-subjects research guidelines and adherence to confidentiality agreements with the participants of this study, neither the data nor additional information about the data or the study participants can be made available.

REFERENCES

Ash, K. D., R. L. Schumann III, and G. C. Bowser, 2014: Tornado warning trade-offs: Evaluating choices for visually communicating risk. *Wea. Climate Soc.*, **6**, 104–118, https://doi.org/10.1175/WCAS-D-13-00021.1.

Atkinson, R., and J. Flint, 2001: Accessing hidden and hard-to-reach populations: Snowball research strategies. *Soc. Res. Update*, **33**, 1–4.

AWARN, 2022: Advanced Warning and Response Network. Accessed 27 April 2022, https://awarn.org.

Bates, A. V., and Coauthors, 2020: Toward a new warning method of threats in motion: Improving warning lead and departure times with innovative hazard communication and dissemination techniques. *10th Conf. on Transition of Research to Operations*, Boston, MA, Amer. Meteor. Soc., 7.3A, https://ams.confex.com/ams/2020Annual/meetingapp.cgi/Paper/365303.

Berry, K., H. Obermeier, K. Klockow, S. Jasko, and D. LaDue, 2017: Broadcast meteorologist decision making in the 2016-2017 hazardous weather testbed probabilistic hazard information project. *NWA 42nd Annual Meeting Broadcaster Workshop*, Garden Grove, CA, National Weather Association.

Braun, V., and V. Clarke, 2006: Using thematic analysis in psychology. *Qual. Res. Psychol.*, **3**, 77–101, https://doi.org/10.1191/1478088706qp063oa.

Buzzard, K. S. F., 2015: The rise of market information regimes and the historical development of audience ratings. *Hist. J. Film Radio Telev.*, **35**, 511–517, https://doi.org/10.1080/01439685.2015.1052219.

Calhoun, K. M., T. M. Smith, D. M. Kingfield, J. Gao, and D. J. Stensrud, 2014: Forecaster use and evaluation of real-time 3DVAR analyses during severe thunderstorm and tornado warning operations in the Hazardous Weather Testbed. *Wea. Forecasting*, **29**, 601–613, https://doi.org/10.1175/WAF-D-13-00107.1.

——, and Coauthors, 2018: Cloud-to-ground lightning probabilities and warnings within an integrated warning team. *Special Symp. on Impact-Based Decision Support Services*, Austin, TX, Amer. Meteor. Soc., 4.4, https://ams.confex.com/ams/98Annual/webprogram/Paper329888.html.

——, K. L. Berry, D. M. Kingfield, T. Meyer, M. J. Krocak, T. M. Smith, G. Stumpf, and A. Gerard, 2021: The experimental warning program of NOAA's Hazardous Weather Testbed. *Bull. Amer. Meteor. Soc.*, **102**, E2229–E2246, https://doi.org/10.1175/BAMS-D-21-0017.1.

Cintineo, J. L., M. J. Pavolonis, J. M. Sieglaff, and D. T. Lindsey, 2014: An empirical model for assessing the severe weather potential of developing convection. *Wea. Forecasting*, **29**, 639–653, https://doi.org/10.1175/WAF-D-13-00113.1.

Demuth, J. L., B. Hearn Morrow, and J. K. Lazo, 2009: Weather forecast uncertainty information: An exploratory study with broadcast meteorologists. *Bull. Amer. Meteor. Soc.*, **90**, 1614–1618, https://doi.org/10.1175/2009BAMS2787.1.

Doherty, R., and K. G. Barnhurst, 2009: Controlling nature: Weathercasts on local television news. *J. Broadcast. Electron. Media*, **53**, 211–226, https://doi.org/10.1080/08838150902907710.

Eastern Research Group, 2016: Hazard simplification project: Findings from the October Kansas City Workshop. NOAA Office for Coastal Management Rep., 57 pp.

——, 2018: Cone of uncertainty: Social and behavioral science research. NOAA Office for Coastal Management Rep., 45 pp.

Ernst, S. R., 2020: Colorful language: Investigating the interpretation of the Storm Prediction Center's convective outlook by broadcast meteorologists and the U.S. public. M.S. thesis, School of Meteorology, University of Oklahoma, 124 pp, https://shareok.org/handle/11244/325336.

Hammer, B. O., and T. W. Schmidlin, 2002: Response to warning during the 3 May 1999 Oklahoma City tornado: Reasons and relative injury rates. *Wea. Forecasting*, **17**, 577–581, https://doi.org/10.1175/1520-0434(2002)017<0577:RTWDTM>2.0.CO;2.

Hart, S. G., 2006: NASA-Task Load Index (NASA-TLX); 20 years later. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, **50**, 904–908, https://doi.org/10.1177/154193120605000909.

——, and L. E. Staveland, 1988: Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Adv. Psychol.*, **52**, 139–183, https://doi.org/10.1016/S0166-4115(08)62386-9.

Hoffman, R. R., 2005: Protocols for cognitive task analysis. Advanced Decision Architectures Collaborative Technology Alliance Doc., 108 pp., https://apps.dtic.mil/sti/pdfs/ADA475456.pdf.

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, https://doi.org/10.1175/WAF-D-14-00163.1.

——, and Coauthors, 2018: Development of a human–machine mix for forecasting severe convective events. *Wea. Forecasting*, **33**, 715–737, https://doi.org/10.1175/WAF-D-17-0188.1.

Keul, A. G., and A. M. Holzer, 2013: The relevance of legibility of radio/TV weather reports to the Austrian public. *Atmos. Res.*, **122**, 32–42, https://doi.org/10.1016/j.atmosres.2012.10.023.

Klockow-McClain, R. A. McPherson, and R. P. Thomas, 2020: Cartographic design for improved decision-making: Trade-offs in uncertainty visualization for tornado threats. *Ann. Amer.*

Assoc. Geogr., **110**, 314–333, https://doi.org/10.1080/24694452.2019.1602467.

Kolakoski, C., K. Berry, A. Gerard, H. Obermeier, J. T. Ripberger, P. A. Campbell, T. C. Meyer, and K. E. Klockow-McClain, 2019: Broadcast meteorologist use of social media in the 2018 Hazardous Weather Testbed probabilistic hazard information project. *Ninth Conf. on Transition of Research to Operations*, Phoenix, AZ, Amer. Meteor. Soc., 8B.5, https://ams.confex.com/ams/2019Annual/meetingapp.cgi/Paper/351124.

Krocak, M. J., J. T. Ripberger, C. L. Silva, H. C. Jenkins-Smith, S. Ernst, A. Bell, and J. Allan, 2020: Measuring change: Public reception, understanding, and responses to severe weather forecasts and warnings in the contiguous United States, version 2. University of Oklahoma Center for Risk and Crisis Management, accessed 5 May 2021, https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/EWOCUA.

LaDue, D. S., and Coauthors, 2016: Identifying key characteristics of severe weather communication strategies for optimal emergency managers' and broadcast meteorologists' decision making. *41st NWA Annual Meeting*, Norfolk, VA, National Weather Association, 445, https://nwas.org/meetings/nwas16/abstracts_html/Other/abstract_0003.html.

Meyer, T. C., K. Berry, K. M. Calhoun, H. Obermeier, A. Campbell, J. Wolfe, and K. E. Klockow, 2019: Visualization and communication of probabilistic lightning hazard information for a broad spectrum of end users. *Ninth Conf. on the Meteorological Application of Lightning Data*, Boston, MA, Amer. Meteor. Soc., 6.3, https://ams.confex.com/ams/2019Annual/meetingapp.cgi/Paper/349844.

Miran, S. M., C. Ling, J. James, A. E. Gerard, and L. P. Rothfusz, 2017: User perception and interpretation of tornado probabilistic hazard information: Comparison of four graphical designs. *Appl. Ergon.*, **65**, 277–285, https://doi.org/10.1016/j.apergo.2017.06.016.

Nemunaitis-Berry, K. L., and H. M. Obermeier, 2017: Conference Notebook, broadcast meteorologist decision making in the 2016 Hazardous Weather Test probabilistic hazard information project. *Bull. Amer. Meteor. Soc.*, **98**, 2045–2047, https://doi.org/10.1175/BAMS_9810_2039-2052_Nowcast.

Obermeier, H., K. L. Nemunaitis-Berry, S. A. Jasko, D. LaDue, C. Karstens, G. M. Eosco, A. Gerard, and L. Rothfusz, 2017: Broadcast meteorologist decision making in the 2016 Hazardous Weather Test Probabilistic Hazard Information Project. *12th Symp. on Societal Applications: Policy, Research and Practice*, Seattle, WA, Amer. Meteor. Soc., 4.3, https://ams.confex.com/ams/97Annual/webprogram/Paper314239.html.

Robertson, M. M., and K. K. Droegemeier, 1990: NEXRAD and the broadcast weather industry: Preparing to share the technology. *Bull. Amer. Meteor. Soc.*, **71**, 14–18, https://doi.org/10.1175/1520-0477(1990)071<0014:NATBWI>2.0.CO;2.

Rothfusz, L., R. Schneider, D. Novak, K. Klockow, A. Gerard, C. Karstens, G. Stumpf, and T. Smith, 2018: FACETs: A proposed next-generation paradigm for high-impact weather forecasting. *Bull. Amer. Meteor. Soc.*, **99**, 2025–2043, https://doi.org/10.1175/BAMS-D-16-0100.1.

Sherman-Morris, K., 2005: Tornadoes, television, and trust: A closer look at the influence of the local weathercaster during severe weather. *Environ. Hazards*, **6**, 201–210, https://doi.org/10.1016/j.hazards.2006.10.002.

Silva, C., J. Ripberger, H. Jenkins-Smith, and M. Krocak, 2017: Establishing the baseline: Public reception, understanding, and responses to severe weather forecasts and warnings in the contiguous United States. Center for Risk and Crisis Management Reference Rep., 31 pp., http://risk.ou.edu/downloads/news/WX17-Reference-Report.pdf.

——, ——, ——, ——, and W. Wehde, 2018: Refining the baseline: Public reception, understanding, and responses to severe weather forecasts and warnings in the contiguous United States. Center for Risk and Crisis Management Reference Rep., 29 pp., http://risk.ou.edu/downloads/news/WX18-Reference-Report.pdf.

——, ——, ——, ——, S. Ernst, and A. Bell, 2019: Continuing the baseline: Public reception, understanding, and responses to severe weather forecasts and warnings in the contiguous United States. Center for Risk and Crisis Management Reference Rep., 33 pp., http://risk.ou.edu/downloads/news/WX19-Reference-Report.pdf.

Smith, B. T., R. L. Thompson, J. S. Grams, C. Broyles, and H. E. Brooks, 2012: Convective modes for significant severe thunderstorms in the contiguous United States. Part I: Storm classification and climatology. *Wea. Forecasting*, **27**, 1114–1135, https://doi.org/10.1175/WAF-D-11-00115.1.

Sood, G., 2016: Geographic information on designated media markets, version 9. Harvard Dataverse, accessed 7 April 2021, https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/IVXEHT.

Stumpf, G. J., and A. E. Gerard, 2021: National Weather Service severe weather warnings as Threats-in-Motion. *Wea. Forecasting*, **36**, 627–643, https://doi.org/10.1175/WAF-D-20-0159.1.

Trujillo, J. E., K. Nemunaitis-Berry, H. Obermeier, A. Gerard, K. E. Klockow-McClain, P. A. Campbell, T. C. Meyer, and J. T. Ripberger, 2020: Plumes, probabilities, and posts: How social media coverage evolved in the 2019 Hazardous Weather Testbed Probabilistic Hazard Information Experiment. *15th Symp. on Societal Applications: Policy, Research and Practice*, Boston, MA, Amer. Meteor. Soc., 12A.5, https://ams.confex.com/ams/2020Annual/meetingapp.cgi/Paper/364474.

Wilson, K., 2008: Television weathercasters as potentially prominent science communicators. *Public Understanding Sci.*, **17**, 73–87, https://doi.org/10.1177/0963662506065557.

Wolf, P., and Coauthors, 2013: A next generation warning concept of Threats in Motion (TIM): An operational warning decision making collaborative exercise. *41st Conf. on Broadcast Meteorology/Second Conf. on* Weather Warnings and Communication, Nashville, TN, Amer. Meteor. Soc., 3.5, https://ams.confex.com/ams/41BC2WxWarn/webprogram/Paper225997.html.

Wolfe, J. P., 2014: An open source approach to communicating weather risks. *10th Free and Open Source (FOSS4G) Conf.*, Portland, OR, Open Source Geospatial Foundation, 20, https://av.tib.eu/media/31617.