Investigating How NWS Meteorologists, Emergency Managers, and the Public Interpret Conditional Intensity Forecasts for Severe Weather

SEAN ERNST,[a,b,d] BENJAMIN J. FELLMAN,[a,b,d,e] ZOEY ROSEN,[a] MAKENZIE J. KROCAK,[c,a,e] ISRAEL L. JIRAK,[d] JOSEPH RIPBERGER,[a] AND HANK JENKINS-SMITH[a]

[a] *Institute for Public Policy Research and Analysis, University of Oklahoma, Norman, Oklahoma*
[b] *OU Cooperative Institute for Severe and High-Impact Weather Research and Operations, Norman, Oklahoma*
[c] *NOAA/National Severe Storms Laboratory, Norman, Oklahoma*
[d] *NOAA/Storm Prediction Center, Norman, Oklahoma*
[e] *School of Meteorology, University of Oklahoma, Norman, Oklahoma*

(Manuscript received 19 July 2024, in final form 3 January 2025, accepted 16 January 2025)

ABSTRACT: Continued research of severe convective storms has enhanced the forecast capabilities of products like the Storm Prediction Center's (SPC) convective outlook. Since 2003, the outlook has presented information about the likelihood of convective hazards within 25 mi of a point, as well as a "hatched" area where a 10% or greater chance of hail larger than 2 in. in diameter, thunderstorm winds greater than 75 mph, or tornadoes of EF2 strength or greater exists. The SPC has begun testing more detailed forecasts of potential storm intensity and is now seeking to design a product that can effectively communicate this new information. To aid in the development of effective intensity forecast information for the SPC outlook, this study conducted surveys and focus groups with members of the public, National Weather Service meteorologists, and emergency managers, recording their feedback on how they thought this information would change their perceived concern and intended behavior on severe weather event days. We also investigated how different presentations of intensity information impact risk perceptions and understanding of the weather event. The inclusion of intensity information increased the perceived concern of members of the public and emergency managers. Changes to the way that intensity forecast visuals were presented also impacted perceived concern and likelihood of response, suggesting that caution must be taken in deciding what the operational version of the convective outlook should look like.

SIGNIFICANCE STATEMENT: As severe weather science advances, the Storm Prediction Center (SPC) has begun to develop the capacity to forecast where significant tornadoes (EF2 or greater), winds (75 mph or greater), and hail (2 in. in diameter or greater) will occur. However, this new forecast information needs to be packaged into a visual format that can effectively communicate severe weather intensity to users ranging from emergency managers to members of the public. Through a series of surveys and focus groups, this study investigates how different user groups interpret several conditional intensity forecast prototypes. Findings suggest that simplified, separated forecasts graphics are preferred to graphics that layer information in one image, and that, although more development is needed, users are able to incorporate conditional intensity information into their decisions around protecting themselves from severe thunderstorms.

KEYWORDS: Social Science; Convective storms; Communications/decision making; Decision support; Societal impacts

1. Introduction

Severe thunderstorm (TSTM) hazards, including hail, thunderstorm winds, and tornadoes, are some of the most widespread and impactful weather events forecast by the National Weather Service (NWS). Thunderstorm hazards claimed over 130 lives in 2023 alone (NWS 2024; calculated using lightning, tornado, and thunderstorm wind fatalities) and caused an estimated 55.5 billion dollars in damage across 19 separate events across the United States that year (NCEI 2024). However, not all thunderstorm events are created equal. Although the NWS has defined any thunderstorm capable of producing straight-line thunderstorm wind gusts in excess of 58 mph, hail with a diameter in excess of 1 in., and any strength tornado as a "severe" thunderstorm (Schoor 2021), the most severe storms are capable of producing winds in excess of hurricane strength, giant hail larger than baseballs, and tornadoes with wind speeds greater than 100 mph. Over 95% of tornado fatalities occur in tornadoes of F2/EF2 strength or greater (111 mph or greater winds; Simmons and Sutter 2011), even though the majority of tornadoes are only rated F/EF0. Significant wind and hail events also rank among the most costly and harmful in U.S. history, as the costliest thunderstorm wind event in U.S. history resulted from straight-line winds in excess of 100 mph that caused 13.3 billion dollars in damage on 10 August 2020 (NCEI 2024), while in June 2023, nearly 100 people were injured at a concert at Colorado's Red Rocks Amphitheater by storms producing 2-in. diameter hail (Jarpe 2023).

To communicate the risks posed by intense severe thunderstorm events like these, the Storm Prediction Center (SPC) produces a suite of risk communication graphics that forecast the likelihood of severe weather (Figs. 1a,b). The foundation of the convective outlook is the probabilistic outlook, which presents a percent probability of severe hail (>1 in. in diameter),

Corresponding author: Sean Ernst, sean.ernst@ou.edu

DOI: 10.1175/WAF-D-24-0109.1

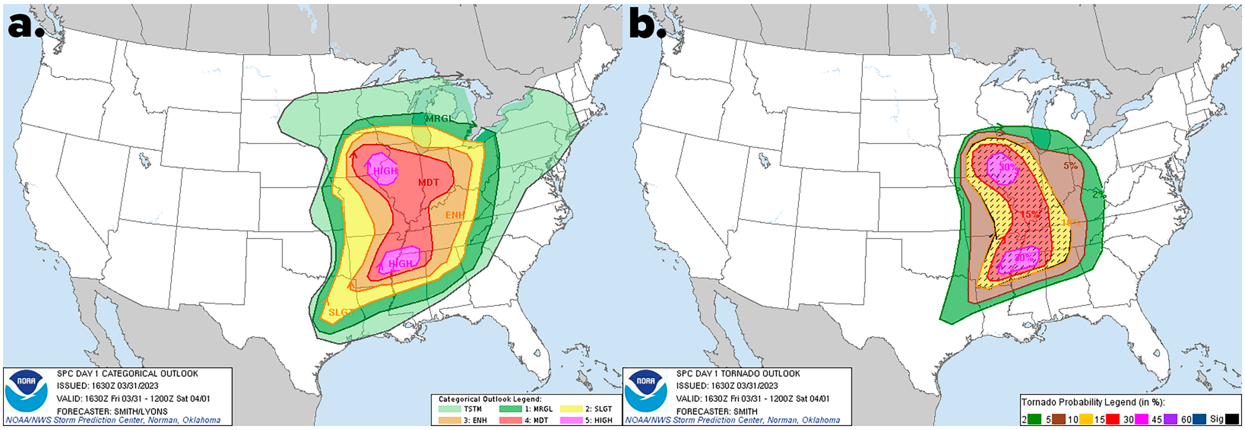

FIG. 1. (a),(b) Example of the day 1 SPC convective outlook product, with the categorical outlook (a) above the corresponding probabilistic outlook for tornadoes (b). Note that the abbreviations in the categorical outlook represent areas expecting general TSTM as well as MRGL, SLGT, ENH, MDT, and HIGH risks for severe weather, while the tornado probabilities are displayed in percent likelihood of a tornado within 25 mi of a point. The probabilistic outlook also includes a black "SIG" item in the legend, which represents the hatching that can be found over the Mississippi River valley and highlights areas with a greater than 10% chance of an EF2 or greater magnitude tornado occurring within 25 mi of a point.

damaging convective straight-line winds (in excess of 58 mph), and any tornadoes within 25 mi of a point (Fig. 1b). However, these severe hazard probabilities are complimented by an intensity forecast, which predicts locations where a 10% or greater chance exists for significant severe weather, defined as hail of 2 in. in diameter or larger, winds equal to or in excess of 75 mph, or EF2 or greater tornadoes, within 25 mi of a point (displayed using black hatching, as can be seen over the magenta, red, and yellow polygons in Fig. 1b). The thresholds for defining significant hazards were set in 1987, to help aid in verifying NWS warning performance for the events most likely to impact life and property (Hales 1987). Around 12% of annual severe weather reports constituted "significant" reports based on the three thresholds in the 1980s, and that proportion has remained relatively unchanged over the decades that the SPC has officially used them (Hales 1987; Grams and Bunting 2020; SPC 2024).

The forecasts for the different probabilities of wind, hail, and tornado hazards, combined with the 10% or greater probability of those hazards surpassing their significant severe thresholds, are then used to define a "categorical outlook" (Fig. 1a), which depicts "up to five risk (categories) based on the coverage and intensity of organized severe weather" (Grams and Bunting 2020). In this use case, risk is defined as a function of the likelihood of a threat occurring and the potential consequences that threat would have (Bostrom et al. 2008). Given this definition, the outlook attempts to forecast risk by fusing the severe hazard probabilities (threat likelihood) with the significant severe forecast hatching (potential consequences, through the lens of potential storm intensity; see Fig. 2). Indeed, the presence of a hatched significant severe contour can increase the level of the categorical outlook, such as for the 15% likelihood contour for tornado forecasts (Fig. 2). However, the current significant severe forecast is effectively deterministic, as it is restricted to a single probability threshold of 10% or greater (Fig. 1b).

The limitations of the current significant severe weather forecast "hatched" area in the SPC outlook are apparent in events like 19 April 2023, where a significant local tornado outbreak occurred around Norman, Oklahoma. The 1630 UTC SPC outlook for the day forecasted a slight (SLGT) risk of severe weather for parts of central Oklahoma, with a 5% chance of tornadoes, a 15% chance of hail, and a greater than 10%

Day 1 Outlook Probability	TORN	WIND	HAIL
2%	MRGL	Not Used	Not Used
5%	SLGT	MRGL	MRGL
10%	ENH	Not Used	Not Used
10% with Significant Severe	ENH	Not Used	Not Used
15%	ENH	SLGT	SLGT
15% with Significant Severe	MDT	SLGT	SLGT
30%	MDT	ENH	ENH
30% with Significant Severe	HIGH	ENH	ENH
45%	HIGH	ENH	ENH
45% with Significant Severe	HIGH	MDT	MDT
60%	HIGH	MDT	MDT
60% with Significant Severe	HIGH	HIGH	MDT

FIG. 2. Key table that translates the day 1 SPC probabilistic outlook forecasts for tornadoes, winds, and hail into the categorical outlook (from Grams and Bunting 2020). Note that the abbreviations for the words in the categorical outlook are the same as in Fig. 1.

chance of significant hail forecast in the SPC probabilistic outlook (SPC 2023). In the outlook forecast discussion, SPC forecasters highlighted that while "storm development is uncertain" for the Oklahoma threat area, "if storms form, the environment favors a threat of isolated very large hail with any sustained supercell" (SPC 2023). Forecasters were less certain that storms would become established enough that they could produce tornadoes but mentioned that "confidence in occurrence/coverage is low, but intensity is conditionally significant" (SPC 2023). Any forecast users that did not read this discussion would have missed the potential for significant tornado activity in the forecast graphics because the forecast tornado risk contour value of 5% was not high enough to allow for the addition of a significant tornado contour (i.e., 10%). Unfortunately, the conditional threat was realized on the evening of 19 April, when a cluster of supercells produced significant hail and over a dozen tornadoes across central Oklahoma, two of which resulted in EF3 damage (NWS Norman 2023). Events like the 19 April 2023 outbreak highlight how important conditional severity information is and warrant exploration into how SPC could graphically communicate these challenging forecasts.

2. Background

In an effort to help the current convective outlook graphical forecast better characterize and communicate low likelihood/high-impact events like 19 April 2023, the SPC developed a product they call the conditional intensity forecast. Past research has identified that analyses of relevant severe weather variables can be used to discriminate between more or less favorable atmospheric environments for significant tornadoes, hail, and straight-line wind (Thompson et al. 2003, 2012; Smith et al. 2012; Wade et al. 2023). Using knowledge of these variables, the SPC began to test conditional intensity forecasts in the Hazardous Weather Testbed Spring Forecasting Experiment (HWT SFE; Clark et al. 2019) in 2019. Conditional intensity forecasts were intended to graphically highlight areas where the likelihood of EF2 and greater tornadoes, 2-in. diameter and greater hail, or 75 mph and greater straight-line winds is highest, assuming a severe storm occurs. Thus, the forecast is conditional on severe weather actually occurring, unlike the current significant severe weather forecast that highlights the unconditional 10% or greater probability of tornadoes of EF2 strength or greater, hail larger than 2 in. in diameter, or greater than 75 mph wind gusts.

Initially, SPC forecasters were asked to consider three possible levels of conditional intensity, which were graphically presented as an overlay to the probabilistic outlook: "nonhatched" (i.e., no conditional intensity contours drawn), suggesting that significant severe was not expected to occur; hatched (i.e., represented by a polygon filled with dashed diagonal lines), which suggested that significant severe was possible; and "double hatched" (i.e., represented by a polygon with cross hatching), which suggested that high-impact significant severe weather was expected (Clark et al. 2019; Figs. 3 and 4). Note that "high-impact significant severe," as described in the terminology row for the double-hatched condition in Fig. 4, does not have a strict definition but was intended to better describe significant severe

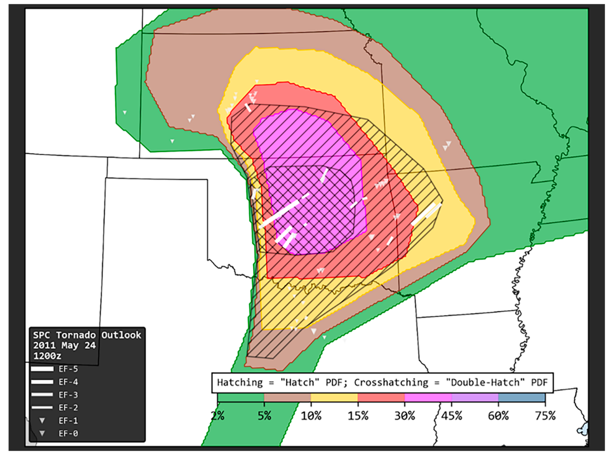

FIG. 3. Example of conditional intensity prototype for 24 May 2011, used as a training example during the 2019 HWT SFE. In this graphic, hatching refers to the black diagonal lines overlaid on the colors that display tornado probability, with hatching or single hatching referring to areas covered by a single diagonal black line, and cross hatching or double hatching referring to areas covered by two perpendicular diagonal lines. The white lines in the image highlight EF2 and greater tornado tracks on this day, with line thickness corresponding to tornado strength, while gray and white triangles indicate EF0 and EF1 tornadoes, respectively.

weather that would be deserving of a moderate (MDT)- or high (HIGH)-risk categorical outlook forecast in this specific HWT experiment. Across the 5 weeks of the HWT in 2019, 70% of HWT SFE participants that created conditional intensity forecasts found it "neither difficult nor easy," "easy," or "very easy" to create these experimental outlooks for the first time, indicating that most forecasters were not overly burdened by this task despite there being some room for improvement of the forecast process (Clark et al. 2019). Participants also reviewed conditional intensity and coverage forecasts issued by an expert "lead forecaster" during the experiment and gave overall positive reviews for the day 1 and day 2 conditional intensity outlook (Clark et al. 2019). The subjective ratings for the day 1 initial coverage and day 1 initial conditional intensity forecasts were very similar, with a median rating of 7 on a 1-to-10 scale recorded for all three convective hazards across the two outlooks (Clark et al. 2019).

Although the initial 2019 test of conditional intensity forecasts was more subjective in nature, future studies sought to verify conditional intensity forecasts using severe weather reports. Vancil et al. (2022) used the National Centers for Environmental Information's Storm Data to verify experimental, internal-to-SPC conditional intensity forecasts issued by the SPC between November 2021 and May 2022. During that time, the in-house hatched and double-hatched areas in the convective outlook contained 70% of all significant tornadoes, 90% of significant severe hail reports, and 90% of significant severe wind reports (Vancil et al. 2022). Further, the ratio of significant to nonsignificant severe weather reports generally increased from unhatched regions to single-hatched regions to double-hatched regions, suggesting that forecasters were

	None	Non-Hatched	Hatched	Double-Hatched
Terminology	Significant severe unlikely	Significant severe not expected	Significant severe possible	High-impact significant severe is expected
Environment	Non-supportive environment	Standard CAPE/shear space for severe events	High-end CAPE/shear space	Extreme CAPE/shear space
Mode	None or disorganized	Disorganized/multi-cell/messy	Tornadoes and hail: Supercells Wind: Supercells, organized clusters, or squall line with bowing segments	Tornadoes and hail: Discrete supercells Wind: Well-organized MCS
Recurrence interval (rough estimate, from past tornado outlooks)	160 days per year	180 days per year	20 days per year	5 days per year
Sub-grid scale impacts from significant severe	None	None or isolated	Sporadic or sparse	Dense

FIG. 4. Conditional intensity prototype scale, used during the 2019 HWT SFE to guide forecaster efforts in creating conditional intensity forecasts during the experiment. This also defined the initial version of the conditional intensity forecast presented to participants in our first study. Note that high-impact significant severe, as described in the terminology row for the double-hatched condition, does not have a strict definition but was intended to better describe significant severe weather that would be deserving of a MDT or HIGH risk categorical outlook forecast in this specific HWT experiment.

capable of discriminating between significant and nonsignificant severe weather events in an operational setting (Vancil et al. 2022).

While these initial feasibility results for conditional intensity forecast areas are promising, forecast skill is only one requirement of an effective risk communication tool. Recipients must also receive, interpret, understand, and be able to act upon the risk information included in the conditional intensity product; in other words, they must be capable of using the forecast to generate value (Murphy 1993; Lindell and Perry 2012). The process through which recipients generate value from forecast information can be described using risk communication models, like the Protective Action Decision Model (PADM; defined in Lindell and Perry 2012). Protective actions, as modeled by PADM, begin with receiving cues and warnings, such as seeing a tornado watch on television or getting a text about a tornado threat from a friend (Lindell and Perry 2012). Individuals then decide how to respond to a potential threat, either through the collection of more information, the dismissal of the threat, or by taking direct protective actions. Through this sequence of events, PADM suggests that risk communication message frequency, consistency, and certainty can affect individuals' response rates (Lindell and Perry 2012). Earlier studies have supported this theory, highlighting the importance of well-constructed messaging in risk communication (Sorensen 2000).

Other research efforts have focused on identifying the types of information that individuals seek out to make protective action decisions in advance of severe thunderstorms and have found that there is indeed a desire for more storm intensity forecast details among forecast users. Both members of the public (Krocak et al. 2023) and emergency managers (Wanless et al. 2023) were asked to highlight whether they prioritized forecast information about storm location, timing, chance, severity (defined in the study as "how intense is the storm going to be," and thus synonymous with intensity), impacts, or recommended protective actions during several different time periods before an event. Both user groups ranked chance, or storm likelihood, information as the most important information type 3 days or more in advance of storms, which the current SPC convective outlook currently provides (Krocak et al. 2023; Wanless et al. 2023). However, as a potential storm event approaches in time, users' information needs shifted to prioritize location and timing information when storms were a day out and then severity information when storms were 4 h from impacting them (Krocak et al. 2023; Wanless et al. 2023). Severity (or intensity) information steadily increased in importance to respondents across the 3-day to 4-h time frame, particularly for the emergency managers surveyed by Wanless et al. (2023). This happens to be the time frame where the convective outlook is intended to be the primary severe weather risk communication tool, suggesting that there is a user desire for information that a conditional intensity product could provide.

Although there does appear to be public desire for intensity information during the convective outlook forecast period, prior work has shown that small changes in graphical presentations of intensity forecast information can have direct impacts on the actions that recipients take. Potter et al. (2018) found that when individuals were provided with impact-based warning information, they displayed a better understanding of

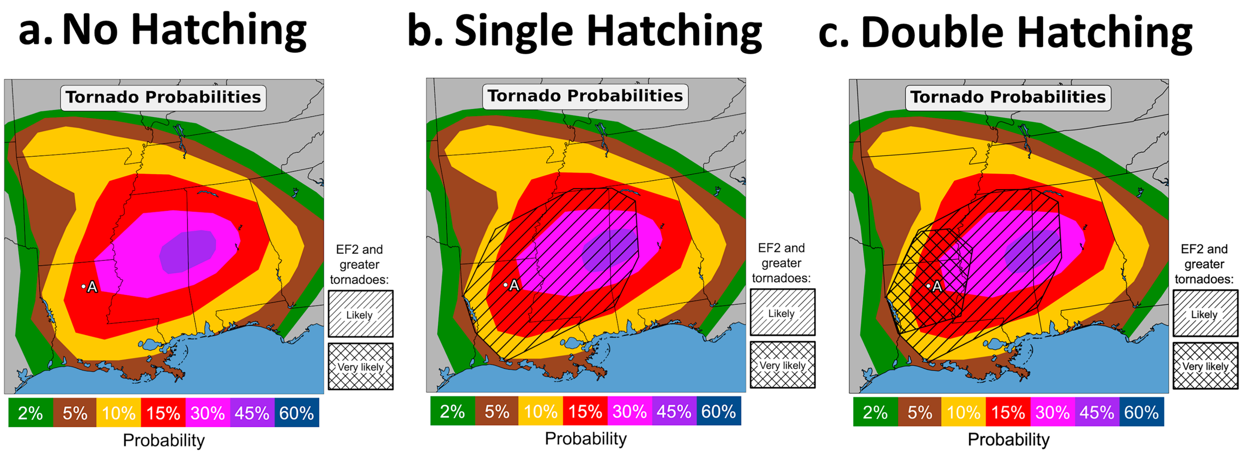

FIG. 5. (a)–(c) Participants in the WX22 and WxEM surveys were shown one of these three graphics, which present increasing levels of conditional intensity in the form of higher levels of hatching on the tornado likelihood graphic.

the forecast than when shown only phenomenon-based warnings, which the researchers theorized could lead to an increased likelihood of protective action for message recipients. Similarly, Ripberger et al. (2015) identified that harsher language in impact forecasts for tornadoes increased recipients' perceived likelihood of protective action, suggesting that the stronger language was leading individuals to process the threat as more severe and thus act according to the processes described in PADM. However, participants who showed increased forecast likelihoods of EF2 or greater tornadoes more often suggested that they would attempt to evacuate their homes rather than shelter in place, a behavior change that has the potential to put individuals at greater risk from the tornado hazard (Ripberger et al. 2015).

Due to the impacts that language has even in graphical forecasts like conditional intensity, consideration needs to be given to the language used to present probability to forecast recipients. We communicate uncertainty verbally using words of estimative probability (WEPs), which include words like "chance," "risk," or "expected" (Kent 1964; Lenhardt et al. 2020). WEPs can be further described as qualified or unqualified, where qualified WEPs include qualifier words like "slight" or "high end" before a WEP like chance, while unqualified WEPs include no qualifier or a modifying word that only makes sense within a context that message recipients may not be aware of (e.g., "increasing potential," "lesser chance"; Lenhardt et al. 2020). Researchers in the past have not defined the SPC categorical outlook risk levels [marginal (MRGL), slight, enhanced (ENH), moderate, and high risk] as WEPs, as they are linked to specific numerical values and thresholds (Fig. 2) rather than to a broader sense of uncertainty that forecasters are communicating (Lenhardt et al. 2020). However, the descriptors "unlikely," "not expected," "possible," and expected used to describe the likelihood of significant severe in areas with no probabilistic forecast, nonhatched areas, single-hatched areas, and double-hatched areas in the 2019 HWT version of conditional intensity are all unqualified WEPs (Fig. 4). Lenhardt et al. (2020) identified that unqualified WEPs like these can lead to a broad range of numerical risk estimates when presented to members of the public and suggest that in cases where precise numerical probabilities are not available to be provided, qualified WEPs should instead be used.

Given the SPC's success thus far at producing skillful forecasts of significant severe weather (e.g., Clark et al. 2019; Vancil et al. 2022; Wade et al. 2023) and previous research that has shown that the design of impact or intensity forecast messages can meaningfully influence forecast recipients' protective action decisions (e.g., Ripberger et al. 2015; Lejano et al. 2016; Potter et al. 2018; Otto et al. 2018; Lenhardt et al. 2020), we seek to identify how forecast users from multiple backgrounds respond to the SPC's prototype conditional intensity forecast product. Similar to recent studies that have developed improved forecast visualizations for broad user groups (e.g., Radford et al. 2023; Semmens et al. 2023), we used an iterative design process to improve upon the original prototype conditional intensity forecast product using feedback from each successive experiment to refine and then retest the prototype in the next experiment. Thus, we present our efforts as a series of three studies that included surveys and focus groups, all of which occurred between 2022 and 2023. The overarching goal of the three studies was to answer the question: How do disparate user groups interpret and use conditional intensity forecasts, and how can these forecasts be improved to better suit user needs?

3. Study 1—Surveying members of the public

a. Methods

During the first study of how users interact with conditional intensity visualizations, we used the initial SPC prototype design of the product (Figs. 3 and 4) to develop a test version of a conditional intensity forecast visualization (Figs. 5b,c). This initial version of the conditional intensity visualization hewed close to the SPC's internal prototype design and highlighted areas with a higher likelihood of significant severe weather (in this case tornadoes of EF2 rating or greater) using "no hatching," "single hatching," and "double hatching" layers over the preexisting probabilistic outlook (again for tornadoes). The graphics used the unqualified WEPs "not likely," "likely," and "very likely" that in the internal SPC prototype described the likelihood of EF2 or greater tornadoes for

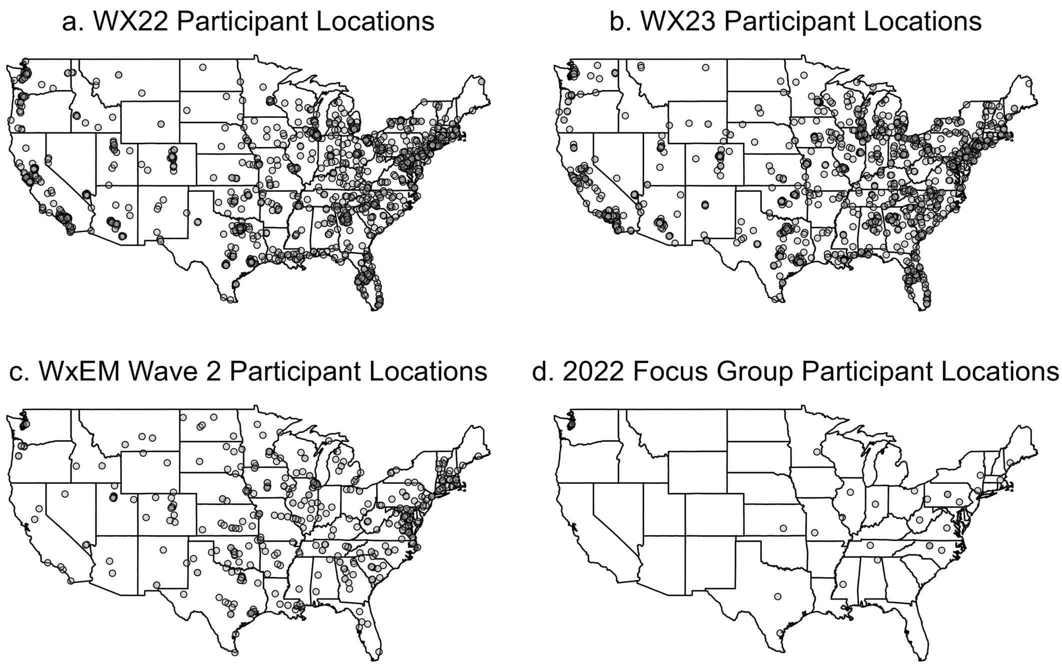

FIG. 6. Maps that display the locations of survey participants in (a) the WX22 survey, (b) the WX23 survey, (c) the WxEM wave 2 survey, and (d) the 2022 HWT focus groups. Each dot on the map represents the location of a single participant. Further demographic breakdowns for the surveys can be found in Bitterman et al. (2022, 2023) and Wanless et al. (2023). Note that seven participants in the WxEM survey in (c) were located in Alaska and Hawaii and thus do not appear in this map projection.

unhatched, single-hatched, and double-hatched regions, but we chose not to include the not likely level in the legend due to concerns about how the language might undersell the risk of severe weather in those regions (Figs. 5a–c).

We randomly presented each participant in the University of Oklahoma's Institute for Public Policy Research and Analysis (OU IPPRA) 2022 Severe Weather and Society Survey (WX22) and 2022 Extreme Weather and Emergency Management Survey (WxEM) with one of the three forecasts shown in Figs. 5a–c. The WX22 survey was administered through Qualtrics, which uses a dynamic sampling process to invite participants from their managed panel of U.S. residents that matches the demographic makeup of the U.S. population (Bitterman et al. 2022, 2023). A mailing list of over 700 emergency managers maintained by OU IPPRA was used to solicit responses for the WxEM survey (Wanless et al. 2023). Overall, 444 emergency managers participated in the WxEM survey, while 1409 U.S. residents completed the WX22 survey (Fig. 6a for WX22 and Fig. 6c for WxEM).

Each of the surveys included a short series of questions about participants' perceptions after being shown one of the three outlook graphics. Members of the public participating in WX22 were then asked how much risk the forecast suggested for location A (marked on the forecast maps in northern Louisiana, Figs. 5a–c) and how likely they would be to change their plans for the day if they were at location A (Table 1). Responses were collected in the form of a 5-point Likert scale, where as an example for the perceived risk question,

TABLE 1. List of the questions posed to respondents in the two surveys that comprised study 1.

WX22 questions	Finally, please look at a different forecast from a few months ago: (random assignment to one of the 3 options in Figs. 5a–c) • How much risk does this forecast suggest for people in location A? (choice of: 1. no risk, 2. low risk, 3. medium risk, 4. high risk, and 5. extreme risk; recorded as 1–5 values) • How likely is it that you would change your plans for the day if you were in location A? (choice of: 1. very unlikely, 2. somewhat unlikely, 3. about as likely as not, 4. somewhat likely, 5. very likely; recorded as 1–5 values)
WxEM questions	Finally, please look at a different forecast from a few months ago: (random assignment to one of the three options in Figs. 5a–c) • How much risk does this forecast suggest for people in location A? (choice of: 1. no risk, 2. low risk, 3. medium risk, 4. high risk, and 5. extreme risk; recorded as 1–5 values) • How likely is it that you would recommend people to change their plans for the day if they were in location A? (choice of: 1. very unlikely, 2. somewhat unlikely, 3. about as likely as not, 4. somewhat likely, 5. very likely; recorded as 1–5 values)

TABLE 2. Mean responses from participants in the WX22 and WxEM surveys, with mean values listed in bold next to standard error, presented in italics within parentheses.

	Outlook graphic viewed	Public respondents (WX22)	Emergency manager respondents (WxEM)
Mean risk perception	No hatching	**2.96** (*0.04*)	**2.86** (*0.05*)
	Single hatching	**3.07** (*0.04*)	**3.04** (*0.05*)
	Double hatching	**3.36** (*0.04*)	**3.66** (*0.08*)
Mean response likelihood	No hatching	**3.58** (*0.05*)	**3.50** (*0.08*)
	Single hatching	**3.55** (*0.05*)	**3.73** (*0.08*)
	Double hatching	**3.73** (*0.05*)	**4.16** (*0.08*)

participants were asked to choose whether they felt that a location in the forecast had "no risk," "low risk," "medium risk," "high risk," or "extreme risk." These responses were then translated into numbers, ranging from 1 to 5 in order, so that the statistical analysis of responses could be performed on the data. Emergency managers that participated in the WxEM survey were asked a similar pair of 5-level Likert-scale questions, although they were asked whether they would recommend people at location A change their plans based on the forecast instead of what they would do at that location (see Table 1).

b. Results

Responses to the WX22 survey reveal that members of the public, on average, become increasingly concerned about potential severe weather risk when shown increasing levels of hatching on the probabilistic outlook (Table 2; Fig. 7a). Participants' mean risk perception increased significantly from the single-hatched to the double-hatched outlook condition ($M_{\mathrm{hatching}} = 3.07$ vs $M_{\mathrm{Dblhatching}} = 3.36$, $p = <0.0001$; see Table 3 for p values), as well as from the no-hatching to single-hatching condition ($M_{\mathrm{no\_hatching}} = 2.96$ vs $M_{\mathrm{hatching}} = 3.07$, $p = 0.023$). Likelihood of response for the public participants was higher than participants' mean risk perception at each level of hatching (see Table 2, Fig. 7b), but likelihood of response decreased very slightly, although not significantly, between the no-hatching and single-hatching conditions ($M_{\mathrm{no\_hatching}} = 3.58$ vs $M_{\mathrm{hatching}} = 3.55$, $p = 0.667$). Likelihood of response significantly increased between the single- and double-hatching conditions ($M_{\mathrm{hatching}} = 3.55$ vs $M_{\mathrm{Dblhatching}} = 3.73$, $p = <0.0001$).

Emergency managers that participated in the WxEM survey also perceived increasing risk with increasing levels of hatching in the forecast product (Fig. 7a), with a mean of 2.86 for the no-hatching condition, 3.04 for the single-hatching condition, and 3.66 for the double-hatching condition. Unlike members of the public, emergency managers' likelihood of recommending that people at location A change their plans increased across all three conditions as well (Fig. 7b), rising from a mean of 3.50 to 3.73 and then 4.16. Emergency managers were significantly more likely (Table 4, $p < 0.0001$) to suggest that people take protective actions for the double-hatched condition than members of the public reported being likely to act, suggesting a difference in interpretation of the risk communicated by the double-hatched condition between the two groups. This difference was also present for risk perception, which was significantly higher ($p = 0.002$) for emergency managers than for members of the public. Thus, although increasing levels of hatching appear to increase perceived risk and likelihood of response, emergency managers interpret double hatching to suggest more significant risk than members of the public.

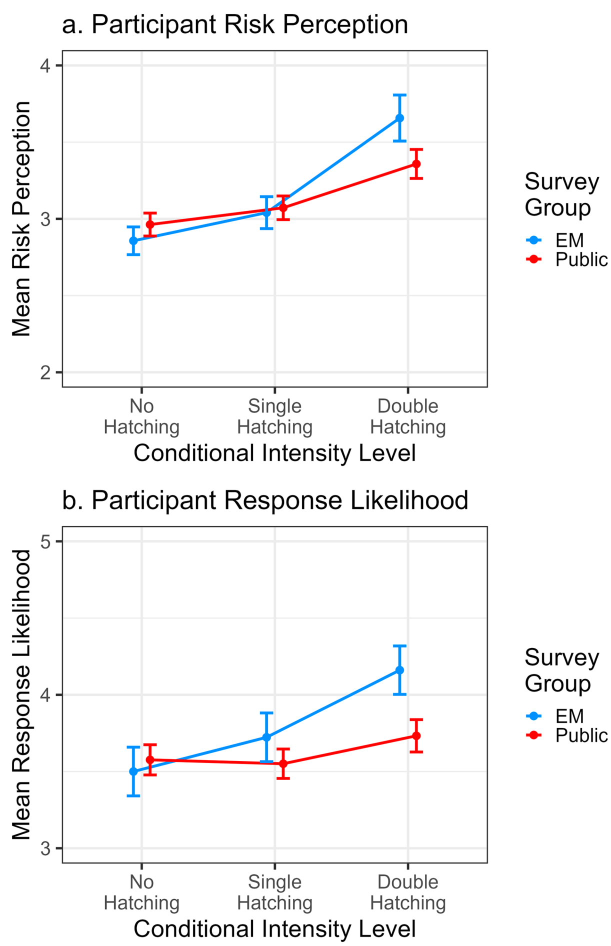

FIG. 7. (a)Mean risk perception and (b) response likelihood selected using five-level Likert scales by participants in the WX22 (public) and WxEM (EM) surveys in 2022. Error bars represent the 95% confidence interval around each mean value. Mean values were obtained by replacing participant Likert scale selections with numbers from 1 to 5 and then calculating a mean.

TABLE 3. Mann–Whitney–Wilcoxson difference of means test p values across the three experimental groups in the WX22 and WxEM surveys. Asterisks highlight the significance of the reported p values (* = significance at the $p = 0.05$ level, ** = significance at the $p = 0.01$ level, and *** = significance at the $p = 0.001$ level).

	Outlook graphic groups compared	WX22 respondent p value	WxEM respondent p value
Mean risk perception	No hatching; single hatching	0.023*	0.015*
	Single hatching; double hatching	<0.0001***	<0.0001***
	No hatching; double hatching	<0.0001***	<0.0001***
Mean response likelihood	No hatching; single hatching	0.667	0.075
	Single hatching; double hatching	0.004**	<0.0001***
	No hatching; double hatching	0.019*	<0.0001***

4. Study 2—User focus groups

a. Methods

After analyzing the data from our experiments with the original prototype of the conditional intensity visualization, we tested the product with NWS forecasters, emergency managers, and members of the public in a series of focus groups. Although surveys of users can collect data on product interpretation from a broad sample, focus groups can highlight the nuance that emerges when individuals are tasked with utilizing experimental SPC forecasts like those for conditional intensity. The focus groups were performed as part of the 2022 HWT experiments and were hosted using the video call application Google Meet across 2 weeks in October of that year. We recruited our participants through email, with NWS forecasters and emergency managers selected from a list of those interested in HWT participation, while public participants were recruited through a convenience sample. Overall, 18 NWS forecasters, 7 emergency managers, and 7 members of the public participated in our focus groups (see Fig. 6d for NWS forecaster and emergency manager locations). Participants signed consent forms before being permitted to participate, and the interview guide developed to organize each focus group was also approved by the Institutional Review Board (protocol No. 13320).

Participants in the focus groups followed a 3-day schedule. On the first day, NWS forecasters were introduced to conditional intensity products and led through a series of discussion prompts (appendix A). Based on the results from our survey testing of the original prototype of the conditional intensity visualization, we developed a new design for the conditional intensity product for our participants to use (in addition to the original design in Fig. 5c), which separated the hatched areas from the probabilistic outlook graphic and presented conditional intensity forecast information alone on a map (as in Fig. 8). Hereafter, we refer to this as the "separated" conditional intensity forecast because it separates conditional intensity from the underlying probabilistic outlook for tornadoes. The separated forecast graphic displays the three levels of conditional intensity with filled regions colored yellow for areas where EF2 and greater tornadoes were not likely, orange for where EF2 and greater tornadoes were likely, and red for where EF2 and greater tornadoes were very likely. This version used the same unqualified WEPs as the WX22 survey design, with not likely for areas with a probabilistic outlook for tornadoes but no "hatching," likely for areas that would have been "single hatched" in the original design, and very likely for areas that would be double hatched. Based on the early participant feedback, the IPPRA research team theorized that reducing the amount of information presented in the conditional intensity product in this way might make the conditional intensity forecast more intuitive for users without prior understanding of what the product was trying to communicate. The discussion prompts on the first day of focus

TABLE 4. Mann–Whitney–Wilcoxson difference of means test p values for independent sample tests across the WX22 and WxEM surveys for each of the three experimental groups. Asterisks highlight the significance of the reported p values (** = significance at the $p = 0.01$ level, and *** = significance at the $p = 0.001$ level).

	Outlook graphic group compared	WX22 vs WxEM p value
Mean risk perception	No hatching	0.433
	Single hatching	0.911
	Double hatching	0.002**
Mean response likelihood	No hatching	0.425
	Single hatching	0.102
	No hatching	<0.0001***

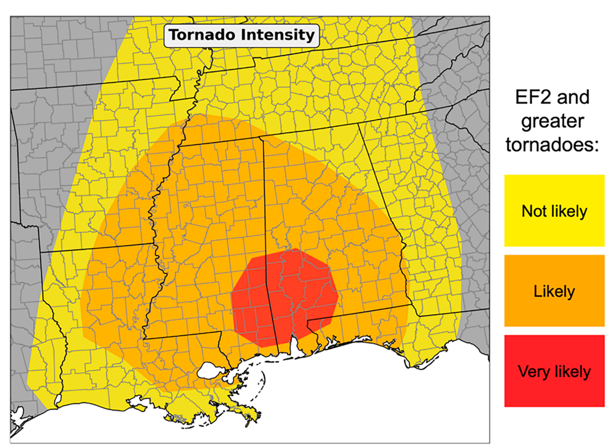

FIG. 8. Participants in the 2022 Fall HWT focus groups were shown several conditional intensity graphics that all followed the same design as this one. Note that this forecast was for a different severe weather event than the forecast presented to 2022 HWT SFE participants as shown in Fig. 5.

TABLE 5. List of thematic codes applied to notes of the focus groups performed at the 2022 Fall HWT experiment.

Code name	Description
Verbal	References to the phrasing and language used to communicate risk in the conditional intensity product (i.e., "I don't think likely is a good word to use").
Visual	References to the visuals used to communicate risk in the conditional intensity product, particularly expressed preferences between the separated and hatched versions of the product.
Behavioral intention	Mentions of actions that would potentially be taken or behaviors that would change after seeing a conditional intensity graphic, particularly mentions of how "I might do things differently from normal" after discussing the graphic.
Comprehension	Discussion of how meaning is extracted from the product, and what final message or understanding participants gain from the product.
Threat perception	Mentions of the threat as comprehended by participants after viewing the product, particularly any mentions of perceived impacts or perceived severity of the threat forecast by the conditional intensity product.

groups centered on asking forecasters how they might use this separated version of the conditional intensity outlook for wind, hail, and tornado hazards and whether forecasters preferred the separated version (Fig. 8) or the hatched version (Fig. 5c) of the visualization for their purposes.

On the second day of focus groups, forecasters were tasked with building forecast briefings and social media posts using conditional intensity forecasts and meteorological information from one of two past severe weather events. Briefings and social media graphics were chosen as these two formats are the most common ones through which NWS forecasters will reach emergency managers and members of the public. NWS social media graphics are frequently posted on their office homepages, and their forecasts are often otherwise repackaged for television by on-air broadcasters. The forecasters were then asked to present their intensity forecasts to our emergency manager participants in two simultaneous but separate discussions, and then the entire group discussed their experience working with the forecasts for the two events (appendix B). Finally, on the third day, members of the public were presented with the social media posts developed by the forecasters, and the NWS forecasters and public participants discussed the conditional intensity products and their use in the social media posts (appendix C). We will note here that because the public participants were recruited through preexisting relationships with the researchers, we have focused our analysis in this paper on the feedback collected from the NWS forecasters as they interacted with the public participants.

While some researchers acted as moderators for the focus groups, others took field notes on the group discussion, recording key ideas about the interpretation and use of conditional intensity forecasts that the participants discussed in their groups. These notes were then analyzed through a thematic coding process. To begin the analysis, two researchers developed a deductive coding scheme based on the research goals of the project (Table 5). The two researchers then analyzed the notes following the coding scheme using the thematic coding analysis software NVivo. Identified codes were compared through negotiations, developing a final set of themes. Negotiations suggested that the codes were reasonably reliable and consistent across the two coders.

b. Results

1) CONDITIONAL INTENSITY USE

When first shown the conditional intensity product in its separated format, forecasters from local NWS offices in our focus groups were generally split between thinking that the product was best for their use internally with other forecasters and core partners and those that felt that they would share conditional intensity information with members of the public. Nine of the 18 forecasters mentioned their preference for internal use of the product, generally due to the fear that the public would misinterpret the graphic. One participant feared that "if we put this up to the public. . . (they) might go 'OMG EF2 tornadoes likely today!'" (FCST17), while another suggested that they were "not sure I would post those on social media, but internally, with training or general use forecasters could (use conditional intensity)" (FCST3). Despite these fears about how members of the public would interpret the product, six of the 18 forecasters believed they would share conditional intensity with the public. These forecasters felt that any information they had available about dangerous weather should also be shared with the public, as "if my mom is in that bullseye, I want her to know that she's within the most impactful storms. I want to share that with the public" (FCST11). Additionally, when presenting their forecasts to the members of the public that were convenience sampled to participate in the focus groups, the forecasters opened up more to the idea of members of the public accessing conditional intensity information and at times tried to present ideas on how to present the forecast in graphics for the public participants to give feedback on. The public participants also mentioned that they might use conditional intensity graphics to make protective action decisions, which the forecasters often asked further questions about.

Emergency managers were much more unified in their reaction to conditional intensity information, as six out of the seven participants in our focus groups stated that they would share the product with both partners and members of the public. Participants quickly described uses for the conditional intensity graphic, ranging from adding the graphic to emails and social media posts (EM4), helping make school dismissal decisions (EM5), and preparing staffing and arranging any external aid needs (EM6). Other emergency managers mentioned that they "would share the conditional intensity graphic with the public, just to add more information" (EM2) and that "knowing the impacts. . . really helps us know which direction to take with our public messaging" (EM7). The conditional intensity product was seen overall as useful for understanding the risk posed by severe weather and that it could be shared to inform other response personnel and the public about those risks.

2) Serious language for serious forecasts

The different groups of participants also expressed that the unqualified WEPs not likely, likely, and very likely used in the prototype conditional intensity graphic (Fig. 8) needed to capture the seriousness of what they were forecasting, and some participants felt that those words did not always match the severity of the potential for storms. Eight different NWS forecasters noted that "we have to convey that (the conditional intensity forecast) is if a storm happens, and it's not like a typical convective outlook," (FCST14) and that "we'll need to ensure on our end that this graphic shows high end tornadoes" (FCST11). Across the eight forecaster groups that were tasked with developing graphics for weather briefings and social media, three used alternative scales for the briefings while four did for the social media graphics, with one group showing no legend at all on the conditional intensity graphic. Notably, the alternative scales were also generally unqualified WEPs (chance, likely, very likely or possibly, likely, expected) although one group labeled the different colors in the graphic in Fig. 8 as highlighting a "very low, low, medium, and high" risk of strong tornadoes, a set of qualified WEPs. Forecasters also discussed their ideas for how to change the descriptors with the members of the public invited to the focus groups, with forecasters in the first week settling on "low," "medium," and "high" risks for strong tornadoes, a series of qualified WEPs, while the second week of discussions highlighted issues both groups had with the use of not likely and how those words decreased concern about significant severe weather. Six emergency managers across the focus groups also noted that "we don't want to overalert or overalarm. . . we need (people) to think 'I have to open this up and pay attention to it'" (EM7). The emergency managers were worried that "the public might see (conditional intensity graphics) and say, 10%. . . and blow it off" (EM6).

3) Specific descriptions of hazards

Some emergency managers also mentioned that they wanted more specific language in the conditional intensity outlook that described exactly what hazards to expect for different levels of conditional intensity. One emergency manager noted that "when we hear wind speeds of 75, it's different from gusts to 75" (EM4). More discussion about specific hazard information occurred between the forecasters and the convenience-sampled members of the public, as forecasters asked whether specific numbers for hail size (FCST1, FCST8) or expected impacts from winds of a given speed (FCST17, FCST18) would change how forecast recipients in the public considered significant severe hazards. The public participants generally liked the greater specifics, although the forecasters highlighted that precise estimates would be difficult for them to confidently issue. Overall, the desire for specific descriptions or details about potential hazard strength was well summarized by the unanimous positive response from the public participants when one forecaster asked, "Instead of the wording "strong tornadoes," would it help if we put a picture of weak versus strong tornado damage, like we say tornadoes on this day have a potential of causing this kind of damage?" (FCST12).

4) Preferences between the separate and hatched versions

Participants were also asked to indicate whether they preferred the separated or hatched overlay versions of the conditional intensity product (Figs. 5 and 8), with most suggesting they preferred separating conditional intensity from the current probabilistic outlook. Forecasters preferred the separated version of the graphic, with 14 out of 18 agreeing with FCST10's sentiment that "there's already enough to explain and interpret. Aside from us, nobody would be able to interpret (hatched conditional intensity)" (FCST10). Others found the hatching to be "a little confusing. . . (but the separated) graphic I could put on a slide and (people would) quickly understand what's going on" (FCST11). Emergency manager participants generally agreed with forecasters' thoughts but felt less strongly that the separated version was superior to the hatched graphic. In the first group, EM2 voiced that they "like it separately, but I guess hatching might be ok. . . Easier to read when they are separate and explain how probability is different than severity to officials," while the three participants in the second group all agreed that the "side by side is very helpful" (EM5, EM6, EM7).

5) Need for greater interactive capabilities

One additional piece of feedback that multiple forecast users shared was a desire for more interactivity options for viewing the convective outlook and its associated products. Emergency managers expressed a desire to be able to zoom in on forecast maps to better view their local area, and more features on maps like roads or political boundaries to help them localize the forecast to their personal contexts. Three emergency managers (EM2, EM6, EM7) mentioned that they wanted the ability to zoom in on the outlook and to toggle features like population centers and roads on the map, as "major cities. . . would be a good reference" (EM7). The ability to display more details and choose map zoom was also discussed four different times between forecasters and the convenience-sampled public participants. Although not directly related to the conditional intensity graphics, forecast users' need for the ability to interact with forecast graphics for more effective contextualization of risk should be considered in future development of conditional intensity graphics.

5. Study 3—Surveying public perceptions of separated visualizations

a. Methods

For the third experiment, we sought to compare our focus group results testing the separated conditional intensity

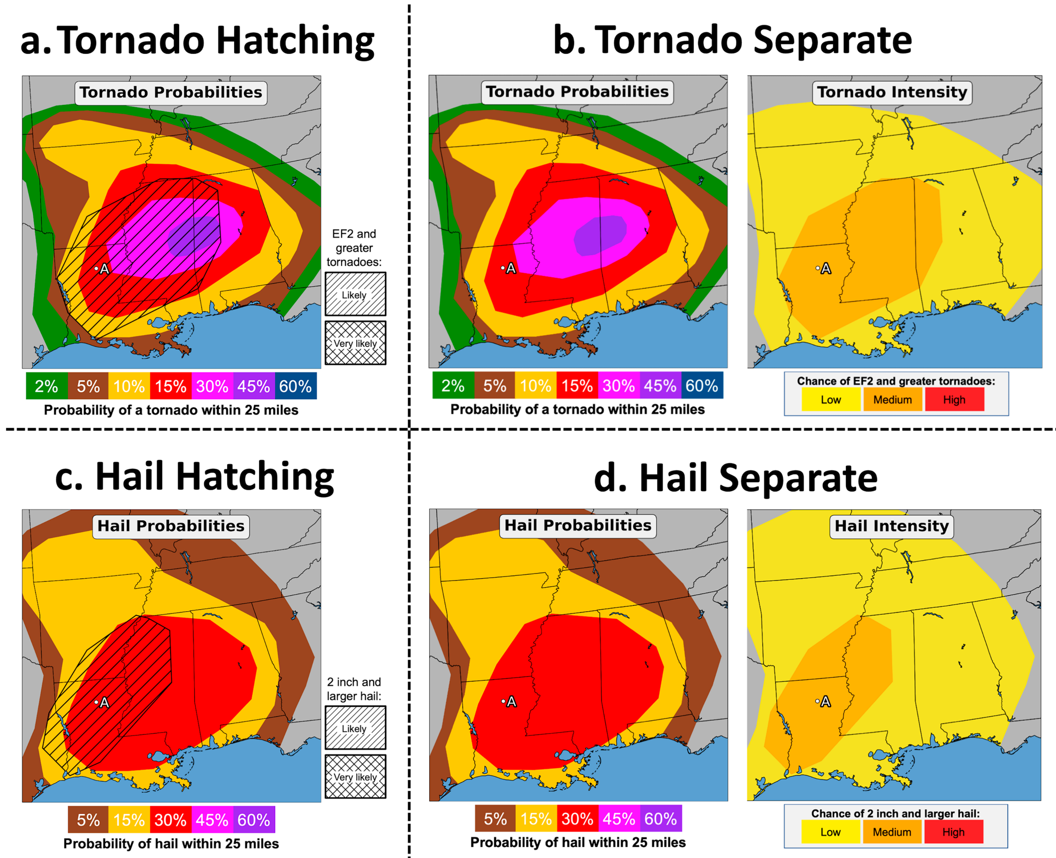

FIG. 9. Participants in the WX23 survey were shown one of these four graphics, which displayed (a),(b) tornado and (c),(d) hail likelihood with conditional intensity in either the (a),(c) hatched or (b),(d) separated format.

visualization with a broader sample of members of the public, to ensure that both versions of the product were tested with a larger and more random sample of the user population. To that end, we developed a survey experiment for the 2023 Severe Weather and Society Survey [WX23; see Bitterman et al. (2023) for demographic breakdowns of participants] that would compare participant reactions to hatched and separated conditional intensity visualizations. Additionally, in response to the focus group discussions about the unqualified WEPs used for the three levels in the conditional intensity graphics, we replaced not likely, likely, and very likely in the graphic with the qualified WEPs low, medium, and high chance. Finally, we also tested graphics displaying forecast hail intensity to capture public reactions to nontornado significant severe weather hazards.

The 1513 public participants (see Fig. 6b) in the WX23 survey were each shown one of the four conditional intensity visuals: a tornado forecast with intensity presented with a hatching overlay (Fig. 9a), a tornado forecast with intensity presented in a separate visual next to the probabilistic outlook (Fig. 9b), a hail forecast with intensity presented with a hatching overlay (Fig. 9c), and a hail forecast with intensity presented in a separate visual next to the probabilistic outlook (Fig. 9d). After being shown one of the four forecasts, participants were asked to report their perceived risk and likelihood of response as if they were at location A, which we assigned to a point in northern Louisiana, as well as whether they agreed that they were confident in their ability to interpret the conditional intensity graphic they were shown (Table 6). As in the first study, responses were collected in the form of Likert scale responses, which were converted from text form to numbers ranging from 1 to 5 to allow statistical analysis of the data.

b. Results

Survey results showed that when members of the public were shown a separated version of the conditional intensity product (Figs. 9b,d), they on average perceived their risk as higher than those shown in the hatched version (Figs. 9a,c; $M_{\text{tor\_hatched\_concern}} = 3.06$ vs $M_{\text{tor\_separated\_concern}} = 3.26$, $p = 0.002$; see Table 7; see Fig. 10b). Interestingly, the difference in mean risk perception between the two tornado forecast visualizations was twice as large as the difference between the two hail forecasts, although reported risk perceptions for the hail forecast were higher on average than risk perception for the tornado forecast for both visualizations ($M_{\text{hail\_hatched\_concern}} = 3.36$ and $M_{\text{hail\_separated\_concern}} = 3.45$, $p = 0.089$; see Fig. 10a). Mean responses for the likelihood of response question were higher, but not significantly so, for participants shown the separated tornado forecast visualization ($M_{\text{tor\_hatched\_response}} = 3.62$ vs $M_{\text{tor\_separated\_response}} = 3.77$, $p = 0.130$; see Fig. 10d), and there was no significant change in mean response likelihood across the two hail forecasts

TABLE 6. Questions presented to participants in the WX23 survey.

WX23 questions	To begin, imagine that it is a Saturday morning at 8:00am and you see this forecast (random assignment to one of the four options in Figs. 9a–d) • How much risk does this forecast suggest for people in location A? (choice of: 1. no risk, 2. low risk, 3. medium risk, 4. high risk, and 5. extreme risk; recorded as 1–5 values) • How likely is it that you would change your plans for the day if you were in location A? (choice of: 1. very unlikely, 2. somewhat unlikely, 3. about as likely as not, 4. somewhat likely, 5. very likely; recorded as 1–5 values) • How much do you agree or disagree with the following statement? I am confident I understand the risk for severe weather for people in location A next Saturday. (choice of: 1. strongly disagree, 2. disagree, 3. neither disagree nor agree, 4. agree, 5. strongly agree; recorded as 1–5 values)

($M_{\text{hail\_hatched\_response}} = 3.92$ vs $M_{\text{hail\_separated\_response}} = 3.89$, $p = 0.512$; see Fig. 10c). Finally, the WX23 participants self-reported that they understood both versions of the forecast graphics about equally, with no significant differences across visualizations or hazard types ($M_{\text{tor\_hatched\_response}} = 3.81$ vs $M_{\text{tor\_separated\_response}} = 3.85$, $p = 0.484$; and $M_{\text{hail\_hatched\_response}} = 3.91$ vs $M_{\text{hail\_separated\_response}} = 3.90$, $p = 0.223$; see Figs. 10e,f).

6. Discussion

Combined, the results of our three iterative studies of conditional intensity forecast visualization suggest that intensity forecast visuals can help a variety of user groups form a more complete understanding of forecast severe weather risk, but that further development is needed to present conditional intensity intuitively and effectively. In our first experiment, after testing of the SPC's first prototype of the conditional intensity visualizations with emergency managers and members of the public, we found that individuals perceived higher risk and were more likely to make protective action decisions when presented with forecasts displaying increased likelihoods of significant tornadoes using a hatching overlay (Figs. 5b,c). Increases in perceived risk for recipients of severity and impact forecasts comparable to conditional intensity have also been observed in previous studies (e.g., Ripberger et al. 2015; Lejano et al. 2016; Potter et al. 2018; Semmens et al. 2023), although the majority of prior studies did not observe participant responses to forecast graphics.

Next, our focus groups with NWS forecasters, emergency managers, and convenience-sampled members of the public revealed more nuanced impressions of the experimental conditional intensity graphics, as well as participant reactions to different visual designs for the product. While the NWS forecasters appreciated conditional intensity forecast information, they were hesitant to share the forecasts in their original form with their partners and especially with the public due to their perceptions of the complexity of the graphic. Emergency managers were more ready to share conditional intensity forecasts, as they mentioned that they would share the visuals as is with their partners and the public, but they did share some concerns about potentially overwarning members of the public. This is not the first HWT experiment where experts and experienced decision-makers shown new forecast products have described being hesitant to share those new products with partners or message recipients, as some forecast communication researchers have begun to describe this behavior as information hoarding (K. L. Berry 2024, personal communication). Communicators engaging in information hoarding believe that new forecast information is valuable to them and can act as a decision aid but feel that the new information will be too complicated, too misleading, or too frightening for the decision-makers they communicate with to use effectively. Previous studies have similarly identified that meteorologists and emergency managers are frequently concerned about whether their messaging will "overwarn" the public (League et al. 2010), as both groups of practitioners are often concerned that

TABLE 7. Mean responses from participants in the WX23 survey, with mean values listed in bold next to standard error values in italics within parentheses. The p values were derived using Mann–Whitney–Wilcoxson difference of means tests that compared the average reported values from participants shown the hatched and separated conditional intensity graphics across the three WX23 experimental questions. Asterisks highlight the significance of the reported p values (** = significance at the $p = 0.01$ level).

	Graphic version	Hail forecast	Tornado forecast
Mean risk perception	Hatched	**3.36** (*0.05*)	**3.06** (*0.05*)
	Separated	**3.45** (*0.05*)	**3.26** (*0.05*)
	p value	$p = 0.089$	$p = 0.002$**
Mean response likelihood	Hatched	**3.92** (*0.06*)	**3.63** (*0.06*)
	Separated	**3.89** (*0.06*)	**3.77** (*0.06*)
	p value	$p = 0.512$	$p = 0.130$
Mean confidence in understanding	Hatched	**3.91** (*0.05*)	**3.81** (*0.04*)
	Separated	**3.90** (*0.04*)	**3.85** (*0.04*)
	p value	$p = 0.223$	$p = 0.484$

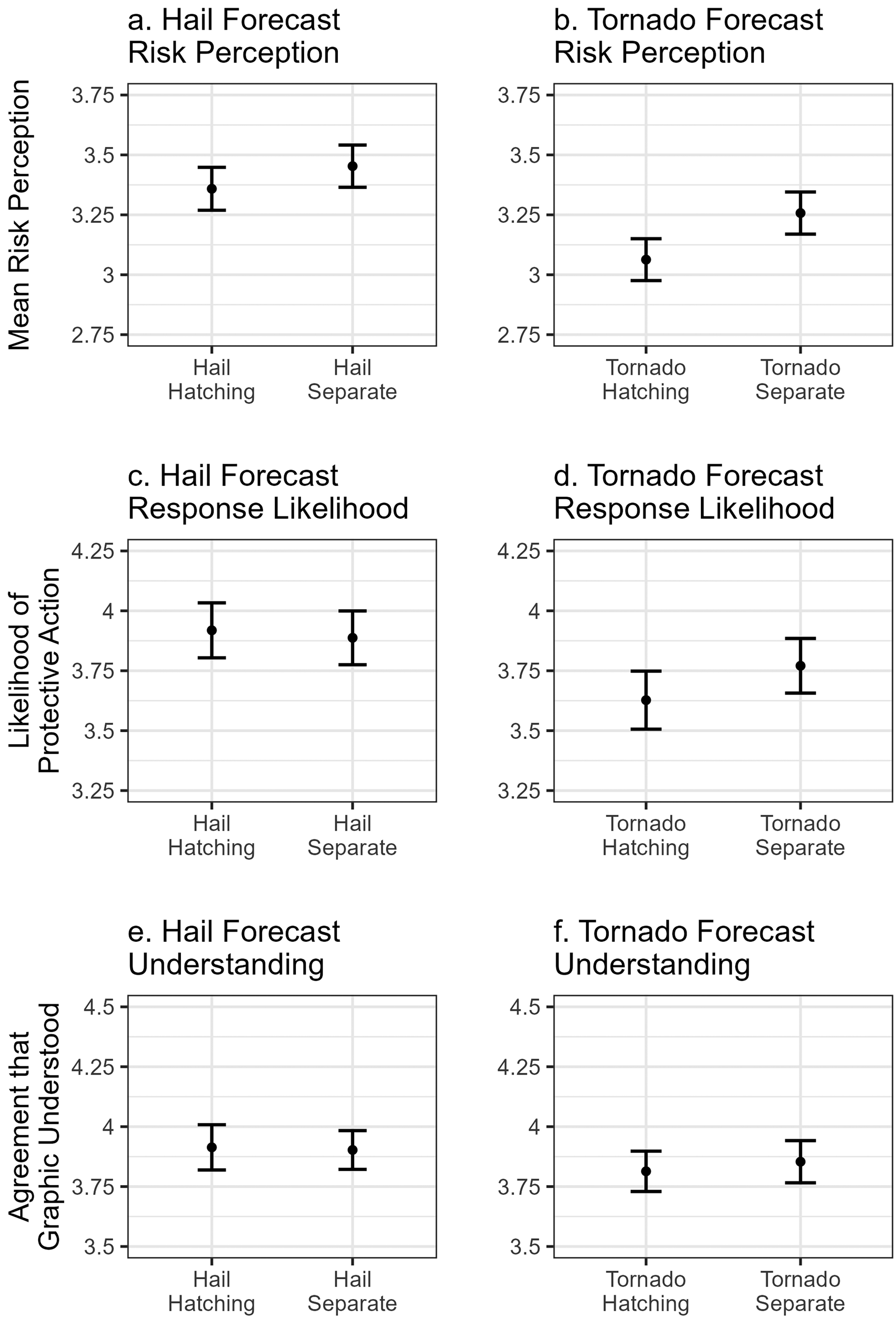

FIG. 10. (a)–(f) Mean responses to the WX23 conditional intensity questions measuring participants' perceived risk, likelihood of response, and subjective understanding of the prototype graphic shown (see Fig. 9). Responses to questions were made using a five-level Likert scale for each question, which can be seen in Table 1. Participants were grouped by whether they were shown forecasts for hail or tornadoes, and whether participants were shown the hatched or separated version of the conditional intensity forecast. Error bars represent the 95% confidence interval around each mean value.

members of the public are uninformed about and apathetic toward weather threats (Littlefield et al. 2010; Stewart et al. 2016).

Discussions between the NWS forecasters and the members of the public that the researchers invited to participate in the focus groups also revealed potential strengths and weaknesses of the NWS forecasters' reinterpretations of conditional intensity products, although the convenience sampling strategy used to recruit public participants means their direct feedback may not be applicable to the public at large. While considering the limitations posed by this study method, it is notable that members of the public discussed actions they might take after receiving conditional intensity graphics with the NWS forecasters. Additionally, their feedback appeared to confirm the NWS forecasters' concerns about the unqualified WEPs used in the prototype version of the conditional intensity graphics, with not likely particularly derided by both parties in the focus group discussions. Prior studies have shown that meteorologists and emergency managers often rely on similar decision biases and interpretations of forecast information to those used by members of the public (Stewart et al. 2016; Roberts and Wernstedt 2019; Lenhardt et al. 2020), which further highlights the value of collecting feedback on new products from all user groups before finalizing any new risk communication graphics design and language.

Participants in our focus groups also shared their impressions of the separated version of the conditional intensity forecast that the IPPRA team had designed for testing after processing the results of the WX22 and WxEM surveys. Prior work has identified that message design can have important impacts on risk perception and protective action decisions (Lindell and Perry 2012; Potter et al. 2018), and the survey results that showed a difference between emergency manager and public risk perception encouraged us to test alternative designs that might be more intuitive to nonexperts. The NWS forecasters in our focus groups felt strongly that the separated version of the graphic was more effective than the version where layers of hatching were applied over the probabilistic outlook, as the separated graphic was less complicated and easier to quickly read than the busier hatched version. Simplified graphics, like the separated version of the conditional intensity product, have previously been found to be more easily understood by message recipients (Peters et al. 2007; Hildon et al. 2012; Clive et al. 2023). In contrast, emergency managers felt less strongly that the separated graphic was best but preferred it to the hatched overlay version. Finally, several of the forecast user participants suggested that map zoom tools and overlays with road and population center names would allow them to contextualize the forecast more easily than they could with the static images provided here. It is worth noting that the SPC provides access to forecast viewers with geographic information system overlays, including on their Convective Outlook viewer (https://www.spc.noaa.gov/products/outlook/day1otlk.html), the High Resolution Ensemble Forecast page (https://www.spc.noaa.gov/exper/href/), and some zoomed map options on the SPC partners page (https://www.spc.noaa.gov/partners/outlooks/), but more intuitive solutions that offer intuitive user experiences with greater control over data viewing should also be explored as potential communication aids.

Finally, our 2023 survey experiment testing the separated and hail conditional intensity forecasts found that members of the public perceived greater risk when shown the separated conditional intensity visualization for both hail and tornado forecasts (Figs. 10a,b). It is possible that, as suggested by Potter et al. (2018) and Clive et al. (2023), the more simplified presentation in the separated graphic improved message recipients' ability to understand their risk from severe weather. However, no major differences in perceived ability to interpret the graphics were measured (Figs. 10e,f), contrasting with both the theories that understanding the graphic increased risk perception and the assumption that NWS forecasters in the focus groups made that users would find the separated graphic easier to understand. The relatively high mean understanding scores across all four versions of the conditional intensity graphics (near a value of four for all, suggesting participants agreed with the statement that they could understand the graphic shown them) also suggest that the qualified WEPs that were used in this version of conditional intensity effectively communicated the threat to participants, although the lack of survey data on understanding from the original unqualified WEPs does not allow us to compare results between the two WEP choices. Further, the WX23 participants were not any more or less likely to consider changing their behaviors after being shown separated or hatched versions of the hail forecast, and there was only a small increase in likelihood of changing plans when shown a separated tornado forecast (Figs. 10c,d). Overall, there did not appear to be any meaningful differences between the hatched and separated visualizations, at least in terms of perceived risk, likelihood of changing plans, or perceived understanding of the graphics.

Beyond our interest in the conditional intensity visualizations in this study, we also found that our participants wanted a more interactive viewer for conditional intensity forecasts. This desire for tools to better contextualize risk highlights the need for further development of visualizations for severe weather forecasts that can present many types of localizable details for different users. Studies of the winter storm severity index (WSSI; Nelson and Perfater 2022), a product developed at the Weather Prediction Center and intended to forecast potential winter weather hazards across the United States, have also provided further evidence for the need for forecasts of intensity or impact that can be more easily accessed and compared in an interactive web-based format (Semmens et al. 2023). As diverse types of information are increasingly requested by users of weather products, including forecasters and emergency managers (Demuth et al. 2012), an online, interactive interface for the SPC convective outlook and its components could act as a valuable "one-stop-shop" for users looking to access detailed information about a potential severe weather event.

7. Conclusions

The conclusions of our study should be considered with regard to the limitations of our research design. First, we relied on a convenience sample for the collection of focus group participants, which limits the generalizability of our findings. This does mean that data from those public participants are not generalizable or unbiased, but we are confident that some of the things NWS forecasters said in discussions with those members of the public relied on their presence in the conversation. Future studies should seek to use alternative recruitment strategies to include members of the public in focus groups about experimental forecast products, as the discussions involving these convenience-sampled participants were rich with ideas and engaging for all those present. Further, both the focus group experiment and surveys only attempted to replicate the messaging that might be received during the lead-up to a severe weather event, and the interpretations and decisions made in real-world situations may differ from those in such a simulated experience. The focus on social media and weather briefing graphics from NWS forecasters also fails to capture the broad spectrum of weather information sources available to the public, such as phone applications, television broadcasts, and word of mouth, and future studies of conditional intensity forecasts should also seek to include these communication formats in their user testing.

Despite these limitations, our results suggest that the conditional intensity forecasts currently under development at the SPC can help improve recipients' protective action decisions by providing valuable information about the potential impactfulness of forecast severe weather. The survey findings that show members of the public considering changing their behaviors at high rates after being shown these graphics is a particularly exciting finding, as it suggests that they are using conditional intensity information to make protective action decisions (Lindell and Perry 2012). However, it is important to keep in mind that the results of our surveys and focus groups suggest that further iterative development is needed before the words of estimative probability used to describe different levels of conditional intensity are finalized. Survey results also revealed that the only difference in people's response to the "hatched" version versus the "separated" version was a slight increase in perceived risk, suggesting that future work should explore in detail how members of the public interpret these different graphical formats in a broad range of severe weather forecast scenarios before any final graphical designs are operationalized.

Further meteorological development of conditional intensity has also continued at SPC, as forecasters there continue to generate internal real-time conditional intensity outlooks in addition to the currently available product suite. These internal, experimental forecasts help refine the forecasting approach for conditional intensity, as well as the development of a statistical foundation for the forecast. In the past year, the primary focus of conditional intensity development has been on improving the meteorological and statistical understanding of what are being called conditional intensity groups or CIGs. Each CIG represents a statistical distribution of severe-hazard intensity (i.e., tornadic wind speeds, hail size, and straight-line wind gusts), with the central tendency of each distribution shifting to a greater likelihood of more intense severe weather from the lowest category (CIG0, formerly "no hatching," where the severe weather is possible but the intensity distribution is expected to be similar to baseline severe climatology) to the higher categories (CIG1, formerly hatched; and CIG2, formerly "double hatched"). The CIG2 (and in the future potentially CIG3) areas will be reserved for forecasts of high intensity and high-impact severe weather events.

The SPC has also begun to intentionally move away from the "hatching" nomenclature, given that hatching is only one graphical approach to displaying layered information and culturally has become associated with the current 10% or greater unconditional probability of significant severe weather. Some options that the SPC is considering for a graphical replacement to hatching include combining the likelihood probabilities and conditional intensity forecast into unconditional probabilities at higher-end thresholds (e.g., probability of EF3 + tornadoes or probability of 3 in. in diameter hail within 25 mi) as well as potentially merging the likelihood and intensity information into a "continuous" categorical risk (e.g., a scale from 1 to 100). Continued codevelopment of conditional intensity products will ensure these products meet both forecast capabilities and user needs as the SPC continues to reinvent its product suite.

Acknowledgments. This research was supported by funding from NOAA Grant NA21OAR4590173. The authors would also like to recognize Dr. Kodi Berry and Dr. Matthew Flournoy for reviewing the paper internally before publication. We would also like to recognize Rich Thompson at the SPC for helping create conditional intensity forecasts for the studies in this manuscript. Finally, the authors thank the forecasters, emergency managers, and members of the public who shared their feedback on the conditional intensity products in the Convective Innovations Hazardous Weather Testbed experiment focus groups, as this work would not have been possible without their participation!

Data availability statement. Survey data analyzed in this study can be accessed online through the Harvard Dataverse at https://dataverse.harvard.edu/dataverse/wxsurvey. However, due to a desire to maintain participant privacy and follow Institutional Review Board policies, data from focus group notes are not publicly available. Inquiries about data access should be forwarded to the lead author.

APPENDIX A

Forecaster Conditional Intensity Discussion Prompts

- Notes:
  - You DO NOT have to ask all of these questions, OR in this order.
  - Let the conversation flow naturally. These are topical prompts to get people talking.
  - Some groups will not need prompts, and others will be more naturally quiet.
  - <u>**BOLDED**</u> questions should be prioritized if you have a chatty group.

This focus group is going to discuss one of the innovations that the SPC is working on for their convective outlook product—conditional intensity information.

- As a refresher, the goal of this new product is to separate severe hazard coverage probabilities from intensity forecasts.

- Currently, intensity information is tied to coverage of severe weather (i.e., forecasters can draw 10% coverage of significant reports).
- However, we know that SPC forecasters have the skill to differentiate more typical severe weather environments from those that produce more significant severe weather [like EF2 + tornadoes, hail larger than 2 in. in diameter, and winds greater than 65kt (1 kt ≈ 0.51 m s^{-1})].
- These conditional intensity (CI) forecasts would divorce coverage from intensity and allow forecasters to draw areas that delineate between a lower and higher end distribution of reports.
- Here is an example of an experimental CI forecast drawn yesterday in the SFE.

We want to know your thoughts on this product from an operational and communication standpoint. We are particularly interested in if/how you might communicate this information to partners and members of the public.

- **What are your first impressions? Let's first talk about the benefits of such a product.**
  - **What do you think works well?**
- **What challenges or issues do you see?**
  - **What do you wish was different?**
- Would this type of product be useful to your operations?
  - How so?
  - What products or communication tools do you see this information contributing to?
  - When communicating severe weather information with community members, do you get specific questions about coverage or intensity?
  - Do you think this separation would make these conversations easier?
- **How would you prioritize coverage versus intensity information from the SPC?**
  - **Do you usually get questions about specifically coverage versus intensity?**
  - **From whom?**
- Should coverage and intensity be displayed on the same product/map?
- How do you imagine translating conditional intensity information to community members? Partners?
  - What about coverage? Both? Together?
  - Or would you want to give more of an overall threat?
- **Speaking to the no/single/double hatch, the difference between the levels is descriptive, qualitative, or more nebulous**
  - **Do you like that the SPC has the flexibility of having more nebulous levels, or would you like a more specific threshold tied to the levels.**

APPENDIX B

Emergency Manager Conditional Intensity Briefing Discussion Prompts

Now that you have seen the briefings from the forecasters, we are interested in your thoughts

- Was there anything that stood out as particularly helpful to you?
  - How would it help you in your workflow?
- Did you notice any big differences in these briefings versus information you are usually using during severe weather events?
- Did any of the new information seem particularly useful?
  - Was there anything that was confusing or not very useful?
- Did you notice any more explicit information about probability or intensity?
- Do you feel like you **need** more explicit probability or intensity information?
- How would this change how you communicate threats to your stakeholders?

APPENDIX C

General Public Conditional Intensity Briefing Discussion Prompts

Now that you have seen the social media posts about a potential severe weather event, we are interested in your thoughts about what you took from them, how concerned you might be, and what you might do in response, etc.

- How concerned would you be if you saw one of these posts?
- Would you do anything after seeing it?
  - This includes ANYTHING. Like talking to a coworker or to sheltering. Anything.
- Do you think you understood everything in the posts?
  - Was there something that you want to revisit?
  - Something that you thought was very helpful?

REFERENCES

Bitterman, A., M. J. Krocak, J. T. Ripberger, C. L. Silva, H. C. Jenkins-Smith, S. Ernst, and S. Stormer, 2022: Maintaining the series: Public reception, understanding, and responses to severe weather forecasts and warnings in the contiguous United States. Harvard Dataverse, accessed 14 August 2023, https://doi.org/10.7910/DVN/TD5DGD.

——, ——, ——, ——, ——, Z. Rosen, and V. Dunham, 2023: Tracking change: Public reception, understanding, and responses to severe weather forecasts and warnings in the contiguous United States. Harvard Dataverse, accessed 14 August 2023, https://doi.org/10.7910/DVN/PPXI8H.

Bostrom, A., L. Anselin, and J. Farris, 2008: Visualizing seismic risk and uncertainty: A review of related research. *Ann. N. Y. Acad. Sci.*, **1128**, 29–40, https://doi.org/10.1196/annals.1399.005.

Clark, A., and Coauthors, 2019: Spring Forecasting Experiment 2019 conducted by the Experimental Forecast Program of the NOAA Hazardous Weather Testbed: Preliminary results and findings. NOAA, accessed 4 September 2024, https://hwt.nssl.noaa.gov/sfe/2019/docs/HWT_SFE_2019_Prelim_Findings_FINAL.pdf.

Clive, M. A. T., E. E. H. Doyle, S. H. Potter, C. Noble, and D. M. Johnston, 2023: How visual design of severe weather

outlooks can affect communication and decision-making. *Wea. Climate Soc.*, **15**, 979–997, https://doi.org/10.1175/WCAS-D-23-0010.1.
Demuth, J. L., R. E. Morss, B. H. Morrow, and J. K. Lazo, 2012: Creation and communication of hurricane risk information. *Bull. Amer. Meteor. Soc.*, **93**, 1133–1145, https://doi.org/10.1175/BAMS-D-11-00150.1.
Grams, J., and B. Bunting, 2020: SPC convective outlooks. Storm Prediction Center, accessed 2 April 2024, https://www.spc.noaa.gov/misc/SPC_probotlk_info.html.
Hales, J. E., Jr., 1987: An examination of the National Weather Service severe local storm warning program and proposed improvements. NOAA Tech. Memo. NWS NSSFC-15, 36 pp., https://repository.library.noaa.gov/view/noaa/14123/noaa_14123_DS1.pdf.
Hildon, Z., D. Allwood, and N. Black, 2012: Impact of format and content of visual display of data on comprehension, choice and preference: A systematic review. *Int. J. Qual. Health Care*, **24**, 55–64, https://doi.org/10.1093/intqhc/mzr072.
Jarpe, S., 2023: "One hail of a year": Data shows more than 150 damaging hail storms this year. Accessed 10 September 2024, https://kdvr.com/news/local/one-hail-of-a-year-data-shows-more-than-150-damaging-hail-storms-this-year/.
Kent, S., 1964: Words of estimative probability. *Stud. Intell.*, **8**, 49–65.
Krocak, M. J., J. T. Ripberger, S. Ernst, C. Silva, H. Jenkins-Smith, and A. Bitterman, 2023: Public information priorities across weather hazards and time scales. *Bull. Amer. Meteor. Soc.*, **104**, E768–E780, https://doi.org/10.1175/BAMS-D-22-0190.1.
League, C. E., W. Díaz, B. Philips, E. J. Bass, K. Kloesel, E. Gruntfest, and A. Gessner, 2010: Emergency manager decision-making and tornado warning communication. *Meteor. Appl.*, **17**, 163–172, https://doi.org/10.1002/met.201.
Lejano, R. P., J. M. Tan, and A. M. W. Wilson, 2016: A textual processing model of risk communication: Lessons from typhoon Haiyan. *Wea. Climate Soc.*, **8**, 447–463, https://doi.org/10.1175/WCAS-D-16-0023.1.
Lenhardt, E. D., R. N. Cross, M. J. Krocak, J. T. Ripberger, S. R. Ernst, C. L. Silva, and H. C. Jenkins, 2020: How likely is that chance of thunderstorms? A study of how National Weather Service forecast offices use words of estimative probability and what they mean to the public. *J. Oper. Meteor.*, **8**, 64–78, https://doi.org/10.15191/nwajom.2020.0805.
Lindell, M. K., and R. W. Perry, 2012: The Protective Action Decision Model: Theoretical modifications and additional evidence. *Risk Anal.*, **32**, 616–632, https://doi.org/10.1111/j.1539-6924.2011.01647.x.
Littlefield, R., and Coauthors, 2010: "We tell people. It's up to them to be prepared." Public Relations Practices of Local Emergency Managers. *The Handbook of Crisis Communication*, W. T. Coombs and S. J. Holladay, Eds., Wiley, 245–260.
Murphy, A. H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281–293, https://doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2.
NCEI, 2024: Time series: Billion-dollar weather and climate disasters. Accessed 10 September 2024, https://www.ncei.noaa.gov/access/billions/time-series/US/cost.
Nelson, J., and S. Perfater, 2022: Winter Storm Severity Index (WSSI). NWS, 7 pp., https://www.wpc.ncep.noaa.gov/wwd/wssi/WSSI_PDD_2022-23.pdf.
NWS, 2024: Weather related fatality and injury statistics. NWS, accessed 11 September 2024, https://www.weather.gov/hazstat/.
NWS Norman, 2023: The severe weather and tornado outbreak of April 19, 2023. NWS, accessed 2 April 2024, https://www.weather.gov/oun/events-20230419.
Otto, P., A. Mehta, and B. Liu, 2018: Mind the gap: Towards and beyond impact messaging to enhance tropical cyclone risk communication. *Trop. Cyclone Res. Rev.*, **7**, 140–151, https://doi.org/10.6057/2018TCRR02.05.
Peters, E., N. Dieckmann, A. Dixon, J. H. Hibbard, and C. K. Mertz, 2007: Less is more in presenting quality information to consumers. *Med. Care Res. Rev.*, **64**, 169–190, https://doi.org/10.1177/10775587070640020301.
Potter, S. H., P. V. Kreft, P. Milojev, C. Noble, B. Montz, A. Dhellemmes, R. J. Woods, and S. Gauden-Ing, 2018: The influence of impact-based severe weather warnings on risk perceptions and intended protective actions. *Int. J. Disaster Risk Reduct.*, **30**, 34–43, https://doi.org/10.1016/j.ijdrr.2018.03.031.
Radford, J. T., G. M. Lackmann, J. Goodwin, J. Correia Jr., and K. Harnos, 2023: An iterative approach toward development of ensemble visualization techniques for high-impact winter weather hazards: Part I: Product development. *Bull. Amer. Meteor. Soc.*, **104**, E1 630–E1 648, https://doi.org/10.1175/BAMS-D-22-0192.1.
Ripberger, J. T., C. L. Silva, H. C. Jenkins-Smith, and M. James, 2015: The influence of consequence-based messages on public responses to tornado warnings. *Bull. Amer. Meteor. Soc.*, **96**, 577–590, https://doi.org/10.1175/BAMS-D-13-00213.1.
Roberts, P. S., and K. Wernstedt, 2019: Decision biases and heuristics among emergency managers: Just like the public they manage for? *Amer. Rev. Public Adm.*, **49**, 292–308, https://doi.org/10.1177/0275074018799490.
Schoor, G., 2021: National Severe Weather products specification. NWSI 10-512 Tech. Doc., 73 pp., https://www.weather.gov/media/directives/010_pdfs/pd01005012curr.pdf.
Semmens, K., R. Hogan Carr, B. Montz, and K. Maxfield, 2023: What impact? Communicating severity forecast information through the winter storm severity index. *Wea. Climate Soc.*, **15**, 747–758, https://doi.org/10.1175/WCAS-D-23-0023.1.
Simmons, K. M., and D. Sutter, 2011: *Economic and Societal Impacts of Tornadoes*. Amer. Meteor. Soc., 282 pp.
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.
Sorensen, J. H., 2000: Hazard warning systems: Review of 20 years of progress. *Nat. Hazards Rev.*, **1**, 119–125, https://doi.org/10.1061/(ASCE)1527-6988(2000)1:2(119).
SPC, 2023: SPC severe weather event review for Wednesday April 19, 2023. NOAA/NWS Storm Prediction Center, accessed 2 April 2024, https://www.spc.noaa.gov/exper/archive/event.php?date=20230419.
——, 2024: Severe weather report database. NOAA/NWS Storm Prediction Center, accessed 11 September 2024, https://www.spc.noaa.gov/climo/summary/.
Stewart, A. E., C. A. Williams, M. D. Phan, A. L. Horst, E. D. Knox, and J. A. Knox, 2016: Through the eyes of the experts: Meteorologists' perceptions of the probability of precipitation. *Wea. Forecasting*, **31**, 5–17, https://doi.org/10.1175/WAF-D-15-0058.1.
Thompson, R. L., R. Edwards, J. A. Hart, K. L. Elmore, and P. Markowski, 2003: Close proximity soundings within supercell environments obtained from the rapid update cycle. *Wea.*

Forecasting, **18**, 1243–1261, https://doi.org/10.1175/1520-0434(2003)018<1243:CPSWSE>2.0.CO;2.

——, B. T. Smith, J. S. Grams, A. R. Dean, and C. Broyles, 2012: Convective modes for significant severe thunderstorms in the contiguous United States. Part II: Supercell and QLCS tornado environments. *Wea. Forecasting*, **27**, 1136–1154, https://doi.org/10.1175/WAF-D-11-00116.1.

Vancil, J. T., I. L. Jirak, and C. D. Karstens, 2022: Conditional intensity forecast verification using significant severe local storm reports. NOAA Tech Doc. 214, 7 pp., https://www.spc.noaa.gov/publications/vancil/condints.pdf.

Wade, A. R., I. L. Jirak, V. A. Gensini, and J. Vancil, 2023: Using Practically perfect intensity hindcasts to identify the environments of significant severe weather outbreaks. NOAA Tech Doc. 180, 9 pp., https://www.spc.noaa.gov/publications/wade/ppsigsvr.pdf.

Wanless, A., S. Stormer, A. Fox, J. T. Ripberger, M. J. Krocak, A. Bitterman, C. L. Silva, and H. C. Jenkins-Smith, 2023: The extreme weather and emergency management survey, Wave 2: Emergency manager use of severe weather forecast information. Harvard Dataverse, accessed 14 August 2023, https://doi.org/10.7910/DVN/RUUQRP.