BEYOND THE TYPICAL WEATHER FORECAST: HARNESSING PROBABILISTIC GRAPHICS FOR EFFECTIVE EMERGENCY MANAGEMENT

Gretchen R. Gillenwater[1,2], Dr. Daphne LaDue[3], and Alex Marmo[3]

[1]National Weather Center Research Experiences for Undergraduates Program
Norman, Oklahoma
[2]University of North Carolina at Pembroke
Pembroke, North Carolina
[3]Center for Analysis and Prediction of Storms, University of Oklahoma
Norman, Oklahoma

ABSTRACT

Severe weather is challenging to predict but even more so to prepare for. County Emergency Managers (EMs) and other local officials are tasked with the dilemma regularly of being fiscally prudent yet always prepared. Tabletop exercises were conducted in northern California to better understand whether two experimental graphics could better support the difficult decisions being made. First, background interviews were conducted with EMs in two counties before the exercises to better understand local concerns and their decision-making process. All interactions with participants were recorded and transcribed by a professional transcription company. Special attention was given during the analysis regarding the two probabilistic graphics: a box-and-whisker graph and a stacked bar graph. They contained similar information but displayed it in different ways. Because the tabletop exercises had many participants, an anonymous online survey about the two graphics helped ensure everyone could provide their input. Surveys and tabletop discussions were analyzed to better understand how EMs and other local officials use weather forecasts. In particular, how understandable the probabilistic graphics were and whether they would be used prior to a potential flooding event. The results show that both graphs are helpful in the decision-making process. Still, timing and adequate meteorologists' input would need to be provided to EMs and other local officials to optimize their usability.

1. INTRODUCTION

Severe weather is challenging to predict but even more so to prepare for. Conditions for severe weather are forecasted to be present more often than they occur. When severe weather does happen in a local National Weather Service (NWS) Warning Forecast Office's (WFO) area, it does not occur in every county. Emergency Managers (EMs) are tasked with these sometimes high-risk decisions surrounding severe weather. It can be exceptionally challenging for EMs and other local officials to decide whether to prepare for severe weather and flooding events because these decisions result in disruption and/or cost, such as cancellation of afterschool activities or staffing response agencies after regular working hours. EMs do their best to manage the risk for their communities.

EMs are trained by the Federal Emergency Manager Agency (FEMA) to think differently (FEMA 2013). They are constantly thinking about what could or will happen next. EMs are taught that their community is in one of the following stages at all times: Mitigation, Preparedness, Response, or Recovery (FEMA 2013). Mitigation refers to any actions taken to prevent or reduce the effects of disasters. Preparedness involves preparing the community for any disasters that cannot be mitigated. This includes disaster preparedness plans in the event of a hazard. These plans contain what to do, where to go, and who to call in an emergency (FEMA 2013). The response stage occurs immediately after a disaster (FEMA 2013). It involves enacting disaster response plans or starting search and rescue missions. After the response stage, recovery begins and consists of

[1] *Corresponding author address:* Gretchen Gillenwater, University of North Carolina at Pembroke, gretchg03@gmail.com.

restoring the community after a disaster. This could range from rebuilding houses to reducing stress in community members. EMs constantly think about what could happen next within their community.

The NWS is crucial to those in emergency management. Hence, the NWS provides impact-based decision support services (IDSS) to help EMs make weather-based decisions (NWS 2024); IDSS can come in many forms: specialized weather briefings and a 24/7 helpline for weather-related questions. The main focus of this study is to test and evaluate how well probabilistic graphics in weather briefings assist EMs in making decisions within a weather briefing.

2. Literature Review

In the days and hours preceding severe weather, EMs and other local officials make decisions so that they can respond quickly and minimize impacts. Research suggests that EMs need understandable and concise forecast information (Morss and Ralph 2007). This is critical since EMs face 78% of actual weather-related events, with 63% being expected events, while only 31% of training is weather-related (Weaver et al. 2014). The lack of knowledge can prove to be a hindrance to decision-making without adequate explanation on the meteorologist's end. Because of this knowledge barrier, EMs highly value their relationships with meteorologists (Cross and LaDue 2021).

EMs practice proactive thinking to keep their community aware. EMs care about how their decisions affect their community and what they can do to protect them (Olson et al. 2023). The NWS informs EMs and other core partners of any weather-related hazard to their communities (NWS 2024). Core partners are government and non-government officials who make weather-related decisions (NWS 2024). Considering this, EMs want clear and concise forecast information, which they often pass to their communities (Cross and LaDue 2021). Not only do EMs receive little weather training, but the information is passed along to local officials who vary in weather knowledge. This is a significant knowledge barrier for EMs and meteorologists to consider.

A way to bridge that gap in knowledge could be probabilistic graphics. Other sources agree that when it comes to probabilistic information, simple terms are needed for a greater understanding (e.g., Savelli and Joslyn 2013). However, probabilistic information can be challenging to interpret if not explained adequately. For instance, Gonzalez and Wu (1999) describe how decision-makers do not treat probability linearly. For example, one EM interprets the graphic differently than another EM. This could be an issue in terms of analyzing if the graphics are or could be more understandable and comprehensive. Savelli and Joslyn (2013) state that graphics could encourage misunderstanding, as a low-probability event seems more likely when described in larger terms and less likely in smaller terms. For example, the odds of 1 in 10 can be perceived differently than 10 in 100 despite the same odds. This concept of numeric probability is difficult for those with low numeracy, which is also shown by Spiegelhalter et al. (2011). Grounds and Joslyn (2018) corroborate the possible misunderstanding of numeric uncertainty estimates because of cognitive differences or a lack of background knowledge. As previously stated, most EMs lack a weather background, so the likelihood of a misunderstanding could be high. For these reasons, meteorologists who use probabilistic graphics must explain them efficiently, or the meaning could be lost.

Since probabilistic information could be misinterpreted, the relationship between an EM and the meteorologist is unique. According to Morss and Ralph (2007), EMs trust their NWS weather forecast office (WFO) and the personnel they interact with. EMs often want to talk to a meteorologist to learn more about the forecast and get their personal opinion (Cross and LaDue 2021). Explaining the probabilistic graphics further could lessen the misunderstanding of EMs and other local officials caused by them.

While using probabilistic information could be complicated and easily misinterpreted, this study aims to use probabilistic graphics to assist EMs and other local officials with their decision-making needs. Probabilistic graphics can help summarize data, lead to a higher perception of risk, and assist those with low numeracy (Spiegelhalter et al. 2011).

This study aims to assess whether probabilistic graphics assist EMs and other local officials with their decision-making needs.

3. Data and Methods

Part of the research team met with a meteorologist at the NWS San Francisco / Monterey Bay (NWS MTR) WFO four times between December 2023 and May 2024 to plan an exercise to test whether two probabilistic graphics would be helpful for core partner's decision-making.

Background interviews were completed with EMs in two counties where the exercises would be held. Questions included personal experience with weather, current role, how long they had been in the role, what makes their jurisdiction unique, how they use weather forecast information before a flooding event, and what parts of a weather forecast they find the most useful. These questions were asked to understand further what types of decisions are made and how they are made. These helped the research team understand why certain decisions were made. It was beneficial to the exercise to establish a connection with the participants prior to the tabletop exercise. This interaction needed to be done before the exercise as time during the exercise did not allow for this type of questioning.

Two tabletop exercises, hereafter referred to simply as exercises, were held on 16 May 2024 in two county Emergency Operations Centers (EOC). The tabletop exercises were designed to mimic how EMs usually engage with other local decision-makers in their jurisdiction. The exercise simulated a briefing three days prior to potential extreme precipitation and flooding scenario. At the start of the exercise, the NWS delivered their weather briefing containing two probabilistic graphics. The NWS MTR WFO chose these graphs for analysis to assess their usefulness to their core partners. The first graph (Fig. 1) was presented as the 24hr Rainfall Spread and is a box-and-whisker graph depicting the greatest amount of rainfall being 5.31in and the lowest amount being 2.55in. The second graph shown (Fig. 2) is the 24hr Rainfall Probability and is a stacked bar graph representing the likelihood of rainfall amounts through percentages; for instance, there is a 2% chance of 6in of rain or more. Participants called these graphics different things, but very few referred to them by their official titles. Participants were then asked about decisions being made and their opinions on the two graphics.

The background interviews and tabletop exercises were recorded. Researchers took notes during all interactions to record any head nods or general sentiments that would otherwise not be captured in the recording. Because the exercises had over 20 participants each, researchers noted the starting sentences of statements so they could be correctly assigned in the transcripts.

During the tabletop exercise, surveys were given out to participants after the weather briefing to gather data from more participants. Of the 55 participants, 31 filled out a survey. Four surveys were blank. One survey was completed but the answer to the final question about which graphic they preferred, simply said, "the prob graph." Because both graphics contained probability, this answer was not assigned to one or the other graphic.

The audio of each exercise was sent to a professional transcription company. The transcripts were then corrected, and annotations were added to include non-verbal communication, such as hesitation, laughter, certainty/uncertainty, agreement, disagreement, etc. This study was conducted under the purview of the University of Oklahoma's Institutional Review Board.

Structural coding (Saldaña 2021) was used to identify larger portions of text on the following topics, related to the research question: Probability, Weather, Decisions, and Timing. Participants' responses to the weather briefing and questions like, what did you think about these graphics, were analyzed using structural coding. For example, a participant would comment on a graphic assigned to probability. After structural coding was complete, inductive, thematic coding was used to create subcodes (Boyatzis 1998). For instance, the box-and-whisker and stacked bar graphs were my subcodes for probability.

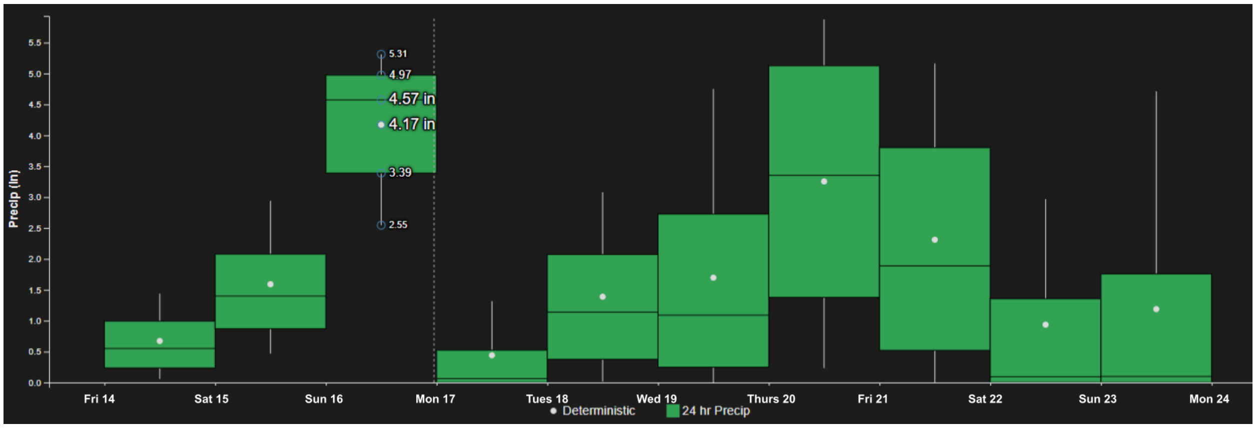

Figure 1: This box-and-whisker plot shows how potential rainfall amounts were distributed across the runs of the model. Precipitation totals are on the y-axis, and days are on the x-axis. The ends of the whiskers are the 10th (2.55in) and 90th percentile (5.31in). The white dot in the middle of the box is the deterministic forecast from a forecast model.

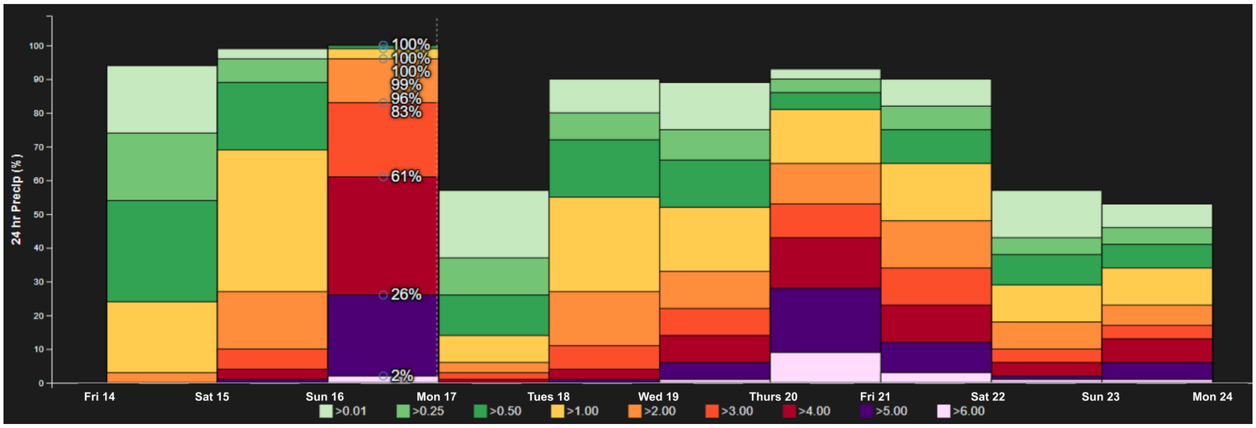

Figure 2: This graph shows the probability of exceeding 9 different threshold amounts ranging from 0.01 to 6.00in. Percentage is on the y-axis, and days are on the x-axis. The colors represent different exceedance thresholds. For example, on Sunday the 16th, there is a 26% chance of at least 5in of rain.

4. Results

The analysis of the tabletop exercises and surveys aimed to determine which graphic better-suited decision-making needs and, if needed, what could be changed or added to them. The exercises were separated by counties for analysis; survey responses were analyzed together.

4.1 *Exercises*

After the weather briefing, participants were first asked about the decisions they would make, then about their interpretation and use of the graphics. Participants who spoke up during the exercise generally favored the stacked bar graph over the box-and-whisker. The key themes (subcodes) regarding the box-and-whisker graph were: information that the graphic does not show, its usefulness, the need for interpretation, and what event is coming after it. Each of these themes is described in Section 4.3. As for the stacked bar graph, key themes were helpfulness, general dislikes, more detailed timing, participant interpretation, difficulty in comprehension, and impacts of color. These are described in Section 4.4. Some participants stated during the exercise that they favored hearing the meteorologists' opinions and experiences over either graphic.

4.2 *Survey*

The survey was an important way to capture information from a greater number of participants than the limited time of the exercise could allow for. The open-ended survey responses were analyzed with the rest of the data, but graphic preference was observed separately. A specific question was asked on the surveys: if you could only have one of the graphics, which would it be? The participants favored the box-and-whisker graph (n=16 of 27 respondents, 59%) over the stacked bar graph (n=10, 37%); one respondent provided input on each graphic but not a clear preference (n=1, 4%).

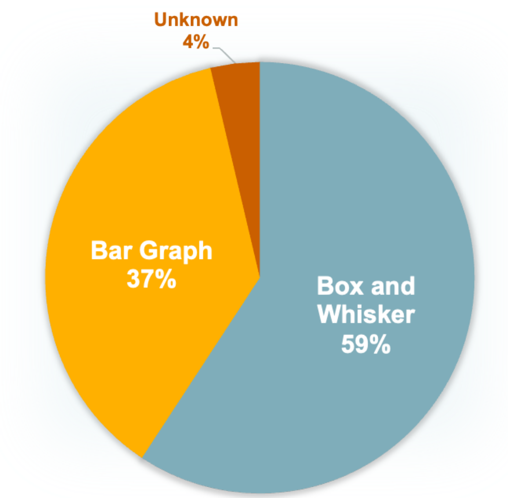

Figure 3: The Pie Chart shows how many participants preferred which graphic based on survey responses.

4.3 *Box-and-Whisker Plot*

The most common theme was what the box-and-whisker graphic does *not* show. These are pieces of information local officials need to make decisions that were not included in the graphic. For instance, multiple participants pointed out that the graphic lists only the 90th percentile, and as one participant stated, they had experienced "many instances where the 90th percentile was far exceeded." One cannot "hang your hat on that" that "this is the worst case" because "what breaks the system is even higher. Or sometimes even lower." Someone else shared similar feelings, saying, "While it's more simplified and doesn't have all that additional detail," so when they see these graphs, they "don't trust that that's the full extent of possibilities." Another participant mentioned, "We should be aware that anything could actually happen." This furthers the need for additional detail. On the survey, a participant pointed out that the forecast was for only one point and added, "It does not give any information on the likely response in the corresponding creeks and rivers and whether they will reach flood stages."

The second theme was the usefulness of the graphic. This theme was mainly found in the survey responses and captured participants' comprehension of the graphic and what they looked at in the graphic. For example, one participant "look[ed] at the depth of the boxes for spread of best and worst case scenarios," as well as "the distance from peak to peak for the compounding effect." The participants were mainly concerned about, for example, "bouts of heavy rainfall on our more vulnerable neighborhoods and facilities." It was unclear to several participants whether the graphic applied to one point or was meant to represent the forecast for the entire county.

The third theme was the need for expert interpretation. A group of participants would regularly refer to the meteorologist's opinion or a specific meteorologist for their opinions. For instance, multiple participants said, as one stated, "I'm not gonna worry about cat whiskers. But I'm gonna lean heavily on [another participant] and his team of experts to guide us." Issues of interpretation were also brought up in the survey, with many expressing they were unfamiliar with statistical plots and that they would, for example, "wait for NWS to describe what I'm seeing."

The fourth theme was needing to know about the next rainfall event. A participant noted, "What's after this storm is what's concerning me." Looking at the graphic, they inferred that the first upcoming rainfall event would not be as impactful as the one following four days later. Their focus was looking ahead and considering whether to prepare for that rather than the simulated event.

4.4 *Stacked Bar Graph*

The first theme was helpfulness, meaning how helpful the graphic was to participants. One participant said they loved “knowing, I love adding the probabilities” and that they “can visually scan it … then I can scan the next one and compare them and go, okay, here's where we are in this color spectrum.” This same participant was also “pretty darn confident that we're going to be in the severe range” and that the graphic gives her “a snapshot of confidence in a way that I find useful.” Another participant said they “get more information out of this than the green one,” which was echoed by another who said, “That graphic is more helpful.” These participants liked the graph and understood what was trying to be communicated.

The next theme was general dislikes, which mainly consisted of statements from a few participants who did not understand the graphic. One said, “I don’t understand this graph,” and another said, “The second graph drives me nuts. I don’t like it.”

The third theme was more detailed timing. Several participants asked for more specific time intervals. During the discussion in one county, someone asked, “Could you take that column of data and divide it into four six-hour periods and apply the percentages per those periods?” The meteorologist responded, “Yes,” after which another participant expressed, “It would be a lot more useful if there were time intervals,” agreeing with the previous participant. A third participant then added, “Even if that was [done] only 24hrs in advance,” it was “still helpful.”

The next theme was participant interpretation. This is what the participants understood from the graphic and if other input was needed. On the survey, one participant said they “Would wait for NWS to describe what I’m seeing.” This comment was representative of another response as well. During the exercise, it was observed by a participant that “even though we don't think the 2 percent is likely … it's not until 2 hours before…or 2 hours after that it hits.” This is when they know what is going to happen. It is believed that this comment shifted other people's opinions on this graphic as another participant “changed my mind on it…I picked this one originally,” and what the previous participant said made them realize that they “got a probably very big false sense of security on the 2 percent.”

The fifth theme was difficulty in comprehension. These participants said that it was hard to understand but still understandable. One participant observed that it was “very busy, but understandable.” Another agreed with them and stated, “I feel like once you get used to it, it's super helpful.”

The final theme was the impact of color. Three participants agreed “that the color coordination for severity really, really, really translates well.” Another participant described the colors starting conversations, “and then when you are talking to those who don't speak NWS or don't speak EM.” The colors help them continue conversations with their organization or constituents.

5. Discussion

We did not find a consensus on which graphic is better. However, participants did provide ways in which each graphic could be improved. This study showed that probabilistic information is difficult to interpret for decision-makers (Gonzalez and Wu 1999, Grounds and Joslyn 2018, Savelli and Joslyn 2013). With both graphics, end-users had difficulty comprehending the information provided. However, those who did understand the graphics were able to be aware of the severity shown in each. This relates to Spiegelhalter et al. (2011), as some participants knew the risk associated with the simulated event.

These results suggest that we continue using both graphics, though with some adjustments. The box-and-whisker, 24-hour Rainfall Spread was liked for its simplicity; however, some were intimidated by the presentation of statistical information. Those confused stated they were unfamiliar with statistics and would want training or more explanation to understand it fully. This demonstrates how EMs and local officials need clear and concise forecast information (Morss and Ralph 2007). In this case, the box-and-whisker graphic was too unfamiliar. The survey results favored the Box and Whisker graphic more than the stacked bar graph.

Other participants liked the stacked bar graph more than the box-and-whiskers, particularly in the exercises. This was possibly due to a recent severe flooding event. The box-and-whisker graph does not illustrate the probability of a 2% chance of more than 6 inches of precipitation, which was pointed out numerous times during the exercise. After this discussion, a participant said, “I kind of changed my mind on it… what [he] said made me realize, I've got a probably very big false sense of security on the 2 percent, I'm going to ignore.” It is believed that the results could possibly be more consistent if the survey were shown after some discussion occurred instead of directly after the briefing.

Three participants were adamant about wanting more detailed timing information. One participant described that even if detailed timing were only given 24hrs in advance, it would still be more helpful than the current 24-hour averages. Knowing a breakdown on timing is crucial to the decision-making process as it helps EMs decide if and when to activate the emergency operations center (EOC) or take other actions.

This study helps articulate the importance of the meteorologist-to-EM relationships; the EM trusts the meteorologist to provide the most accurate insight so they can make the best decisions possible (Cross and LaDue 2021). Some participants strongly preferred the meteorologist’s interpretation. The meteorologist’s expertise in the area’s weather gives the EMs and other end users confidence in the forecast. One participant in public works, for example, stated that they have no expertise to decipher either graph by themselves, so they make decisions solely on what the meteorologist says will happen. Adding more explanation to the graphic, such as timing, would be helpful to the interpreter.

6. Conclusions

In conclusion, opinions on the 24-hour Rainfall Probability stacked bar graph and the 24-hour Rainfall Spread box-and-whisker graph are mixed. Therefore, both probabilistic graphs may prove helpful to EMs with meteorologist explanation or interpretation and timing intervals. These additions are necessary for EMs and their partners to better understand the graphics. After any changes are made, a study would be merited to ensure changes result in increased understanding and utility. EMs and their partners require more detailed timing and explanation to adequately prepare and plan for a severe weather or flooding event.

6.1. *Assumptions and Limitations*

We assume that participants gave their honest opinions and that those opinions represent those who did not actively participate in the discussion or the survey. It is also assumed that those present for the exercise are typically present for a regular weather briefing.

The way the survey was administered limited the quality of the data. The survey was given at the start of the discussion, which did not allow participants to learn how to interpret and use them. The contradiction we saw may be an artifact of the method as evidenced by a participant stating they had changed their mind about graphic preference during the discussion.

8. ACKNOWLEDGMENTS

The lead author thanks her fellow interns for support and sharing efforts in our literature research, especially Gregory Allen and Ivy Jeffries. This work was prepared by the authors with funding provided by National Science Foundation Grant No. AGS-2050267 and the NOAA Weather Program Office, U.S. Department of Commerce under award NA21OAR4590210. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of the University of Oklahoma, the University of North Carolina-Pembroke, the National Science Foundation, NOAA, or the U.S. Department of Commerce.

9. REFERENCES

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

Cross, R.N., and D. S. LaDue, 2021: When uncertainty is certain: a nuanced trust between emergency managers and forecast information in the Southeast U.S. *Wea. Climate Soc.* **13 (1)**, 137–146, https://doi.org/10.1175/WCAS-D-20-0017.1.

FEMA. 2013: Emergency management in the United States. Livestock in disasters Unit 4, 16 pp, https://training.fema.gov/emiweb/downloads/is111_unit%204.pdf

Gonzalez, R., and G. Wu, 1999: On the Shape of the Probability Weighting Function. *Cognitive Psychology*, **38**, 129–166, https://doi.org/10.1006/cogp.1998.0710.

Grounds, M.A., and S.L. Joslyn, 2018: Communicating weather forecast uncertainty: Do individual differences matter? Journal of Experimental Psychology: Applied, **24**(1), 18–33. https://doi.org/10.1037/xap0000165

Morss, R. E., and F. M. Ralph, 2007: Use of Information by National Weather Service Meteorologists and Emergency Managers during CALJET and PACJET-2001. *Wea. Forecasting*, **22**, 539–555, https://doi.org/10.1175/WAF1001.1.

NOAA. 2024: Impact-based decision support services. National Weather Service Directive 10-2403, 8 pp, https://www.weather.gov/media/directives/010_pdfs/pd01024curr.pdf

Olson, J. R., C. M. Doyle, D. S. LaDue, A. N. Marmo, 2023: End-user threat perception: Building confidence to make decisions ahead of severe weather. *J. Operational Meteor.*, 95–109, https://doi.org/10.15191/nwajom.2023.1108.

Saldaña, J., 2021: *The Coding Manual for Qualitative Researchers.* 4th ed. Sage, 414 pp.

Savelli, S., and S. Joslyn, 2013: The Advantages of Predictive Interval Forecasts for Non-Expert Users and the Impact of Visualizations. *Applied Cognitive Psychology*, **27**, 527–541, https://doi.org/10.1002/acp.2932.

Spiegelhalter, D., M. Pearson, and I. Short, 2011: Visualizing Uncertainty About the Future. *Science*, **333**, 1393–1400, https://doi.org/10.1126/science.1191181.

Weaver, J., L. C. Harkabus, J. Braun, S. Miller, R. Cox, J. Griffith, and R. J. Mazur, 2014: An Overview of a Demographic Study of United States Emergency Managers. *Bull. Amer. Meteor. Soc.,* **95**, 199–203, https://doi.org/10.1175/BAMS-D-12-00183.1.