

Recommendations for Developing Useful and Usable Convection-Allowing Model Ensemble Information for NWS Forecasters

JULIE L. DEMUTH AND REBECCA E. MORSS

National Center for Atmospheric Research, Boulder, Colorado

ISIDORA JANKOV

Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, and NOAA/ESRL/Global Systems Laboratory, Boulder, Colorado

TREVOR I. ALCOTT, CURTIS R. ALEXANDER, AND DANIEL NIETFELD

NOAA/ESRL/Global Systems Laboratory, Boulder, Colorado

TARA L. JENSEN

National Center for Atmospheric Research, Boulder, Colorado

DAVID R. NOVAK

NOAA/NWS/WPC, College Park, Maryland

STANLEY G. BENJAMIN

NOAA/ESRL/Global Systems Laboratory, Boulder, Colorado

(Manuscript received 24 May 2019, in final form 15 April 2020)

ABSTRACT

U.S. National Weather Service (NWS) forecasters assess and communicate hazardous weather risks, including the likelihood of a threat and its impacts. Convection-allowing model (CAM) ensembles offer potential to aid forecasting by depicting atmospheric outcomes, including associated uncertainties, at the refined space and time scales at which hazardous weather often occurs. Little is known, however, about what CAM ensemble information is needed to inform forecasting decisions. To address this knowledge gap, participant observations and semi-structured interviews were conducted with NWS forecasters from national centers and local weather forecast offices. Data were collected about forecasters' roles and their forecasting processes, uses of model guidance and verification information, interpretations of prototype CAM ensemble products, and needs for information from CAM ensembles. Results revealed forecasters' needs for specific types of CAM ensemble guidance, including a product that combines deterministic and probabilistic output from the ensemble as well as a product that provides map-based guidance about timing of hazardous weather threats. Forecasters also expressed a general need for guidance to help them provide impact-based decision support services. Finally, forecasters conveyed needs for objective model verification information to augment their subjective assessments and for training about using CAM ensemble guidance for operational forecasting. The research was conducted as part of an interdisciplinary research effort that integrated elicitation of forecasters' CAM ensemble needs with model development efforts, with the aim of illustrating a robust approach for creating information for forecasters that is truly useful and usable.

1. Introduction

When there is a risk of hazardous weather, U.S. National Weather Service (NWS) forecasters characterize and

communicate the potential threat and its impacts with the fundamental goal to reduce harm. Forecasters draw on

Publisher's Note: This article was revised on 29 June 2020 to include an additional affiliation for co-author Jankov that was missing when originally published.

Corresponding author: Julie L. Demuth, jdemuth@ucar.edu

DOI: 10.1175/WAF-D-19-0108.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) (www.ametsoc.org/PUBSReuseLicenses).

their meteorological expertise and on-the-job experiences, assess available observations, and interpret deterministic and ensemble numerical weather prediction (NWP) guidance (Murphy and Winkler 1971a,b; Roebber and Bosart 1996a,b; Bosart 2003; Doswell 2004; Roebber et al. 2004; Morss and Ralph 2007; Novak et al. 2008). NWP guidance available to forecasters has evolved tremendously over the last several decades due to advances in computing capabilities, understanding of meteorological processes, observational datasets, data assimilation, model parameterizations, and postprocessing techniques (see Benjamin et al. 2019 for a review). One major development has been convection-allowing models (CAMs), which resolve finescale spatial and temporal processes, including more accurate depictions of convection and its evolution (Weisman et al. 2008) and of orographically influenced processes (Mass et al. 2002; Schwartz 2014; Gowan et al. 2018). The first CAM became operational in the United Kingdom in 2002, and operational implementation expanded in the subsequent years, including to the United States in 2007 (Benjamin et al. 2019, see their Tables 13–7 and 13–8). The years since then have seen development of CAM ensembles, which explicitly characterize uncertainty of weather hazards, and which offer potential to help forecasters assess and communicate weather risks (Roebber et al. 2004; Novak et al. 2008; Kain et al. 2013; Stensrud et al. 2013; Rothfus et al. 2018; Benjamin et al. 2019).

Translating CAM ensemble output into useful and usable information is especially important given new emphasis for NWS forecasters to provide impact-based decision support services (IDSS). In this role, forecasters “connect forecasts and warnings to decisions made,” and they “emphasize expert interpretation, consultation, and communication of forecasts and their impacts” (NWS 2019, p. 7). These forecaster responsibilities focus particularly on supporting NWS “core partners,” which include members of the emergency management communities, water resources communities, other government partners, and electronic media (NWS 2018a). Forecasters’ use of cutting-edge science and technology to help their users make better decisions is not a new concept (Stuart et al. 2006; Novak et al. 2008). However, CAM ensemble development, the capabilities it offers, and NWS’s focus on IDSS creates a new context for operations-to-research and research-to-operations (O2R/R2O) efforts (Jirak et al. 2010; Kain et al. 2013; Evans et al. 2014; Sobash et al. 2016; Gallo et al. 2016; Clark 2017; Greybush et al. 2017; Schwartz et al. 2019; Wilson et al. 2019).

With this new context comes the new—or perhaps more accurately, *renewed*—need to develop and provide model guidance that NWS forecasters can readily use to help them characterize and convey hazardous weather

threats and impacts, including associated uncertainties. As Roebber et al. (2004) explained, “where high resolution model data are available, it is critical that resources be devoted to improving the *use* of the information rather than simply increasing the supply. The output from such models must be tailored to the needs of the forecasters” (p. 941, emphasis in original). These user-oriented ideas reflect the tenets of risk communication research wherein iterative dialogue with users to understand their decision space—including their goals, values, barriers, needs, experiences, and other factors—is essential for developing information that is useful to them (NRC 1989; Fischhoff 1995; Árvai 2014; Árvai and Campbell-Árvai 2014). This approach recognizes that users’ decision-making context is complex and that risk information is *a* factor, not *the* only factor, in managing risk. Risk information that considers this multifaceted decision context can then be developed accordingly.

This risk communication approach underpins the goal of the social science research presented here, which is to understand NWS forecasters’ IDSS-focused decision contexts in order to identify their needs for new and improved CAM ensemble information. Our research builds on a foundation of past work that has investigated public- and private-sector operational meteorologists’ forecast processes, information use and interpretations, and needs (Murphy and Winkler 1971a,b; Stewart et al. 1997; Doswell 2004; Homar et al. 2006; Morss and Ralph 2007; Novak et al. 2008; Demeritt et al. 2010; Daipha 2012, 2015; Evans et al. 2014; Wilson et al. 2019).

For our research, we collected in-depth, qualitative data from NWS forecasters at national forecast centers and local weather forecast offices (WFOs) using two methods: participant observations and semistructured interviews. The data collection focused on forecasters’ processes and decisions, including observations and model guidance used for forecasting and communication with partners. In addition, we developed prototype plots of different CAM ensemble output to represent, hypothetically and conceptually, the kinds of information that could be derived, and we elicited forecasters’ feedback on them. The guiding research questions of the work presented here are as follows:

- What are NWS forecasters’ key forecast challenges and information needs?
- What CAM and CAM ensemble guidance do NWS forecasters interrogate for different hazardous weather types and scenarios? How do they interpret and use the different guidance?
- What CAM ensemble guidance do NWS forecasters want for assessing and communicating different hazardous weather types and scenarios, particularly based on their partners’ needs?

- In what ways do NWS forecasters think about the skill of CAM ensemble guidance?
- How can the knowledge gained by investigating the above research questions inform development of CAM ensemble information?

The path of O2R and R2O in the United States is iterative and involves multiple steps that evolve from initial foundational work to early conceptual prototyping to experimental testing to full deployment and operationalization (NOAA 2017). At the time of the interviews, no CAM ensemble guidance was available through NWS forecasters' Advanced Weather Interactive Processing System (AWIPS) workstations,¹ which is an NWS requirement for it to be deemed operational. Experimental CAM ensemble guidance was available through web-based platforms from different U.S. research laboratories and universities. R2O evaluations of products, which include different CAM ensemble output, often are done with forecasters at the experimental phase through NWS testbeds (Barthold et al. 2015; Gallo et al. 2016, 2017; Clark et al. 2012; Wilson et al. 2019). The research reported here represents an earlier part of the O2R/R2O path, where foundational work with forecasters was conducted iteratively alongside CAM ensemble model development efforts, including early conceptual prototyping of model output.

CAM ensemble development, capabilities, and use are multifaceted topics that involve interconnected issues ranging from ensemble system design, calibration, and limitations (Roebber et al. 2004; Benjamin et al. 2019) to philosophies about the forecaster's role in an increasingly automated environment (Snellman 1977; Bosart 2003; Stuart et al. 2006, 2007; Novak et al. 2014; Henderson 2019). The research conducted here was designed to be agnostic to the CAM ensemble prediction system, so that the results could be applied to the NWS's future operational system. Thus, this paper does not address the design or merits of one CAM ensemble system versus another, nor does it advocate what the role of a NWS forecaster ought to be with respect to model output as the "final forecast" versus "as guidance" (Novak et al. 2008, p. 1079, emphasis in original). Furthermore, this paper does not advocate the use of CAM ensembles in preference to other available information or in a given forecast situation. Rather, the purpose of this research is to recognize that CAM ensemble information is being developed for operational forecast use and to help guide development of information that has the potential to be most useful based on forecasters' perspectives. The results synthesize forecasters' feedback on

and needs for types of CAM ensemble products (section 3), for information relevant to their IDSS roles (section 4), and for model verification and training to make use of the guidance (section 5). The needs that emerge are not all immediately viable for operational implementation, yet establishing them can be useful for guiding future R2O efforts.

2. Methods

a. Research design and data collection

The multimethod social science research approach employed here reflects the iterative nature of the project. In the first year of the project, data with forecasters were collected through a qualitative research method termed "participant observations." The lead author was in the forecast environment to unobtrusively watch the forecast process—including information interrogated, communications, products issued, and so forth—and to occasionally ask follow-up questions about what was observed (Cresswell 2013; Merriam and Tisdell 2016). Seven randomly chosen days of observation were conducted of forecast operations at two NWS national forecast centers. In addition, 10 days of observation of the forecast process were conducted during two NWS Hydrometeorology Testbed experiments during which experimental ensemble guidance from coarse- and convection-allowing models was used and evaluated (NOAA 2020). The testbed observations were conducted across three weeks of the testbed experiments, which are held when the weather phenomena of interest climatologically occur. The participant observations were conducted between October 2015 and July 2016, and more than 20 forecasters were observed. Real-time and reflexive field notes were taken, and some observation periods were audio recorded for later reference.

The participant observations provided knowledge about the existing and experimental model guidance that forecasters examine and use across different forecast scenarios. This knowledge then guided semistructured interviews with WFO forecasters, which were conducted in the second year of the project.

Semistructured interviews include a set of open-ended questions that serve as a guide to elicit information, but they offer the interviewer flexibility to ask follow-up questions to delve deeper into a topic (Cresswell 2013; Merriam and Tisdell 2016). The interview guide, which simply is "a list of questions that you intend to ask in an interview" (Merriam and Tisdell 2016, p. 124), was developed collaboratively among the research team based on what was learned from the participant observations and based on ideas about CAM ensemble information that could be developed.

Forecasters first were asked in the interview to provide background about their job position and core duties. Then,

¹The High-Resolution Ensemble Forecast (HREF) v2 became available in AWIPS in late 2017.

they were asked to select from among severe weather,² winter weather, or heavy rainfall and flash flooding as a hazardous weather focus for the interview, and they were asked to think about forecasting in the short term, from 0 to 24 h out. With these weather and timeframe foci, forecasters were asked questions about 1) their forecast processes, including a synthesis of observational and model data used; 2) details of coarse-scale ensemble, convection-allowing deterministic, and experimental CAM ensemble products accessed and used; 3) their ideas and needs for CAM ensemble guidance, including different parameters, thresholds, verification, and other information about the model or output, for different types of weather scenarios; 4) their interpretations of, feedback on, and potential use of six prototype CAM ensemble products, discussed in the next paragraph; and 5) their processes of and needs for communicating hazardous weather information with their partners. The interview guide is available from the authors upon request.

Prototype products were utilized in the interview in order to test ideas about information that could be derived from a CAM ensemble. Our aim with the prototypes was not that these versions would become operationalized and used by forecasters, but rather that they were initial product ideas to meet forecasters' needs that could be further developed if found to be potentially useful. Because the research presented here was early in the O2R/R2O path (see introduction), the prototypes were model agnostic and were mock creations. Thus, the prototypes were not applicable to forecast operations on the day of the interview.

Six types of products, along with short descriptions of each, were developed by the NOAA/Earth System Research Laboratory (ESRL) research team and provided during the interviews. The products were map-based plots of the following:

- point probabilities of exceeding a threshold,
- neighborhood probabilities of exceeding a threshold in a 40-km radius,
- paintballs (or paint splats) of where ensemble members exceed a threshold,
- a “combination” of the ensemble control run and the 10th, 50th, and 90th percentile neighborhood probability contours,
- mean onset time, showing the ensemble mean hour of the day when a threshold is first exceeded at a point, and
- mean duration time, showing the ensemble mean number of hours that a threshold is exceeded at a point.

The point, neighborhood, and paintball prototypes emulated contemporary ways of portraying probabilistic ensemble information. The combination, onset, and duration prototypes were developed based on forecasters' informational needs that emerged during the participant observations and are further discussed in [section 3](#). The products were created for each of the heavy rainfall, winter, and severe weather scenarios to show rainfall amounts and rates, snowfall amounts and rates, and updraft helicity, respectively. No mean duration plot was generated for the severe weather case because it is uncommon for large magnitudes of updraft helicity to be sustained at a given point for more than one hour. Example sets of the prototype plots are shown in [Figs. 1–3](#) for the heavy rainfall, winter, and severe weather scenarios, respectively. [Figure 1](#) also includes the short descriptions that accompanied the prototype plots when they were shown to the forecaster. The same descriptions were provided for each scenario. To allow the interviewer to show forecasters prototypes that were relevant to their forecast area, the prototype products were generated for four different regions of the country: the Southeastern United States (depicted in [Fig. 1](#)), the Northeastern United States (depicted in [Fig. 2](#)), the Ohio Valley (depicted in [Fig. 3](#)), and the southwestern states of California, Nevada, Arizona and Utah (not shown).

The prototype plots were shown to the forecaster after they had already answered interview questions pertaining to topics 1–3 described above. Paper copies of the prototype plots along with their associated descriptions were shown to the forecaster in the order presented in [Figs. 1–3](#). For each prototype, forecasters were asked to discuss their interpretation of the information, their preferences for additional or different information (e.g., thresholds, fields), and their potential use of such information if it were operationally available.

The semistructured interviews were conducted with 31 forecasters from 12 WFOs across all 4 NWS regions in the continental United States. A total of 27 interviews were conducted in person, and 4 were conducted by phone. The first and second interviews were conducted to pretest the interview guide with a focus on the question ordering, wording, and length. No significant changes to the interview guide were made after these interviews, and thus both interviews were included in the final dataset. All interviews were conducted between February 2017 and December 2017. By the final interview, “saturation” of ideas was reached, meaning no key insights were mentioned that had not been discussed in earlier interviews; this indicates that the sample size is sufficient to generate robust results ([Merriam and](#)

²The U.S. National Weather Service defines “severe weather” only as tornadic storms or thunderstorms that produce hail greater than or equal to 2.54-cm (1-in) and/or winds greater than or equal to 50 kt (1 kt \approx 0.51 m s⁻¹; 58 mph) ([NWS 2018b](#)). Winter weather and heavy rainfall are not considered to be severe weather.

Tisdell 2016). The median interview length was 76 min (mean = 79 min; range: 42–124 min). Data on the forecasters' gender, years of experience working in the NWS, current job position, and type of hazardous weather discussed in the interview are provided in Table 1.

b. Qualitative data analysis and reporting

The aim of qualitative research is to understand “how people make sense of their world and the experiences they have” (Merriam and Tisdell 2016, p. 15). We selected this research approach due to the limited state of knowledge about NWS forecasters' needs for CAM ensemble information. Qualitative research focuses “on meanings rather than on quantifiable phenomena” and with “collection of many data on a few cases rather than a few data on many cases” (Schutt 2012, p. 324). Qualitative data are richly descriptive, with quotes used to depict complex themes that cannot be validly represented through bits of data.

All of the interviews were conducted by the first author, and they were audio recorded and transcribed verbatim. The data were analyzed with a focus on the research goals using a reflexive thematic analysis approach (Braun and Clarke 2006; Braun et al. 2019). Through inductive, iterative analysis, we identified themes related to how forecasters interpret and use model guidance and what are their critical forecast challenges and CAM ensemble needs. The themes identified can capture both “implicit ideas ‘beneath the surface of the data’” and more explicit ideas, which in turn incorporate both the “essence and spread of meaning” (Braun et al. 2019, p. 3), much like ensemble output has an average and a distribution.

Human subjects approval for the observations and interviews was obtained from NCAR's Human Subjects Committee, and all forecasters consented to participating. Per common ethical human subjects research practices and the human subjects approval obtained for this study, we committed to maintaining forecasters' anonymity to the extent possible so that they could freely express their thoughts and opinions. Thus, we do not identify the national forecast centers, weather forecast offices, or individuals who participated in this study. All quotes are anonymized and referenced as interviewee number and, for context, type of hazardous weather discussed. For example, the thirty-first interviewee discussed severe weather and thus is referred to as “No. 31-severe.”

The interview data are rich and nuanced due to forecasters' expertise and roles. Most of the (sometimes lengthy) quotes are presented in tables with alphanumeric references in the manuscript text. This data presentation approach is intended to facilitate manuscript readability while also preserving the forecasters' “voices” and the richness of the information they provided.

Last, the data collected, analyzed, and reported here represent the forecasters' interpretations, perceptions, and

experiences as they *are* from their perspectives—termed an emic focus (Schutt 2012)—not as others might believe they *ought* to be. Although a reader might not understand or agree with a perspective, understanding the state of knowledge, beliefs, and practice can help identify where improvements might be made.

3. Forecasters' needs for specific CAM ensemble guidance

A range of needs emerged from the participant observation and interview data about specific CAM ensemble guidance that forecasters would like to have. These needs emerged directly and indirectly as the forecasters performed their forecast processes and discussed their context, roles, and goals.

From the participant observations, two common themes arose about guidance needs: 1) information to help forecasters transition from utilizing deterministic guidance to probabilistic guidance, and 2) map-based guidance about timing of hazardous weather threats. The former represents a class of needs that different types of products might fulfill, while the latter represents a need for a specific type of product. These two themes from the participant observations informed development of three prototype products (the combination plot and the two timing plots) that were used in the interviews. Sections 3a and 3b discuss these two themes and the associated prototype products in greater depth. Recall that the first part of the interview guide asked about participants' forecast processes, guidance used, and CAM ensemble guidance needs, and then the prototype products were shown and discussed. Thus, the text below includes forecasters' mentions of information needs before they were shown the products as well as their subsequent feedback about the products. Forecasters expressed CAM ensemble guidance needs beyond those discussed in sections 3a and 3b; these needs are summarized in section 3c.

a. Facilitating the deterministic-to-probabilistic transition: The “combination” plot

It is sometimes stated in the weather community—by forecasters about the public and by model developers about forecasters—that people do not want or use probabilistic forecast information because they prefer simpler, single-valued forecasts or cannot understand uncertainty information (e.g., Hirschberg et al. 2011). Such thinking places the burden on the recipient. However, lack of uptake of any forecast product suggests the information is not useful for some reason; the need to understand those reasons shifts the burden back to the information developer.

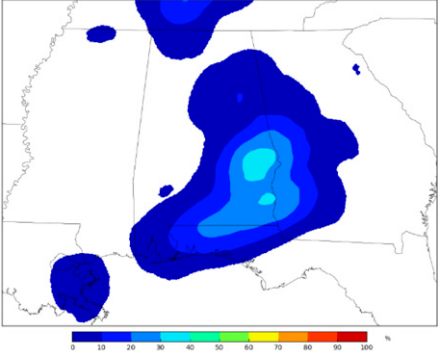
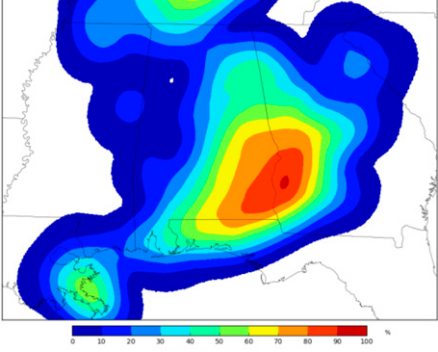
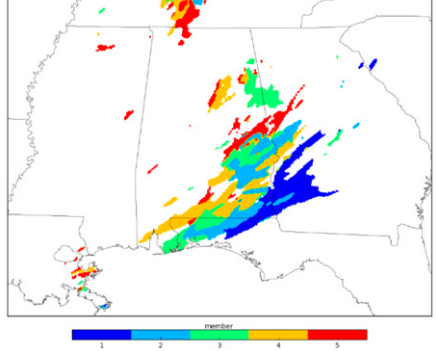
Prototype plot	Accompanying plot description
<p data-bbox="384 258 606 274">Probability of 3-h rainfall > FFG at a point</p>  <p data-bbox="243 647 273 673">(a)</p>	<p data-bbox="768 258 1003 284"><u>Point probability plot</u></p> <p data-bbox="768 290 1248 576">Values indicate the probability of exceeding a threshold at a given point. The probabilities appear smooth due to post-processing techniques that increase spread and improve reliability. Expect a much noisier plot (values mostly 0% or 100/n%, where n is the ensemble size) for any small, under-dispersive ensembles without post-processing.</p>
<p data-bbox="338 701 651 717">Neighborhood probability of 3-h rainfall > FFG within 40 km</p>  <p data-bbox="243 1090 273 1116">(b)</p>	<p data-bbox="768 701 1097 727"><u>Neighborhood probability plot</u></p> <p data-bbox="768 733 1264 939">Values indicate the probability of exceeding a threshold anywhere in an X-km radius around a given point, with post-processing/smoothing applied as for the point probabilities. In almost any real-world situation, these values will be higher than the point probabilities.</p>
<p data-bbox="349 1155 641 1171">Areas where ensemble member 3-h rainfall exceeds FFG</p>  <p data-bbox="243 1544 273 1570">(c)</p>	<p data-bbox="768 1135 913 1161"><u>Paintball plot</u></p> <p data-bbox="768 1167 1264 1514">Points where at least one ensemble member exceeds a threshold are color-coded to indicate which member exceeds the threshold. In this particular version, if multiple members exceed a threshold at the same point, the color goes to the lowest-numbered member. In a non-time-lagged ensemble that color choice is arbitrary. But for a time-lagged ensemble, it might make sense for the most recent member to take precedence where multiple members overlap.</p>

FIG. 1. Example set of heavy rainfall scenario prototype CAM ensemble products that were shown as part of the interview. Plots are (a) point probabilities, (b) 40-km neighborhood probabilities, (c) paintball, (d) “combination” of the ensemble control run and the 10th, 50th, and 90th percentile neighborhood probability contours, (e) mean onset time, and (f) mean duration time. The description that accompanied the plot is shown alongside it. Readers can refer to [Clark et al. \(2014\)](#) to learn more about flash flood guidance (FFG).

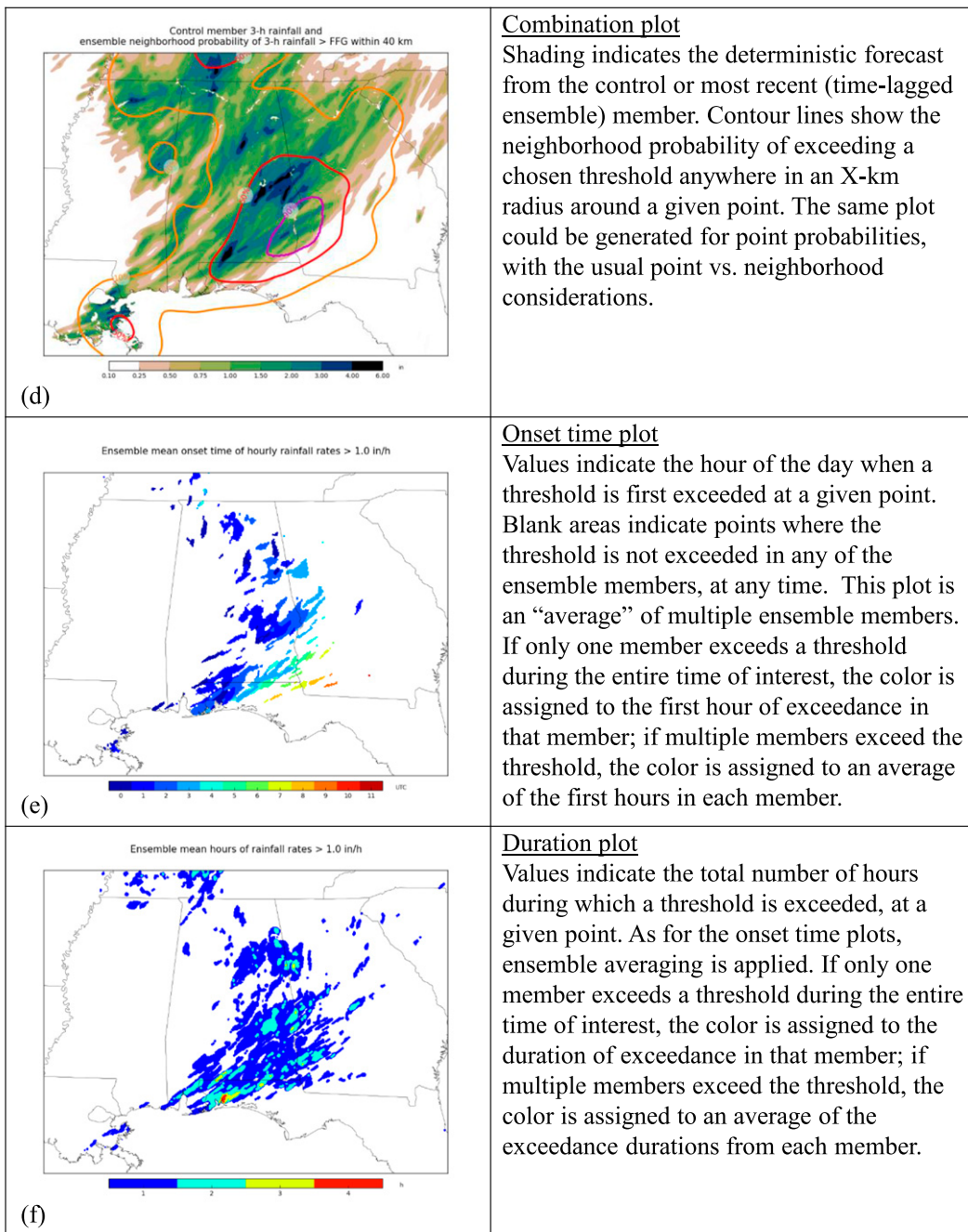


FIG. 1. (Continued)

Most forecasters who were observed and interviewed understand and believe, both theoretically and practically, that ensemble-based guidance confers benefits that deterministic guidance does not. Still, the transition for forecasters and forecast offices from using deterministic guidance to ensemble-based guidance can be challenging, particularly for some types of information.

When postprocessed probabilistic guidance (e.g., neighborhood probability of exceeding some parameter threshold) was shown or discussed during the participant observations, barriers in using the information emerged from many of the forecasters. Forecasters are used to thinking and working spatially and to assessing specific atmospheric features in three and four dimensions. Most current forms of probabilistic guidance do not map, literally

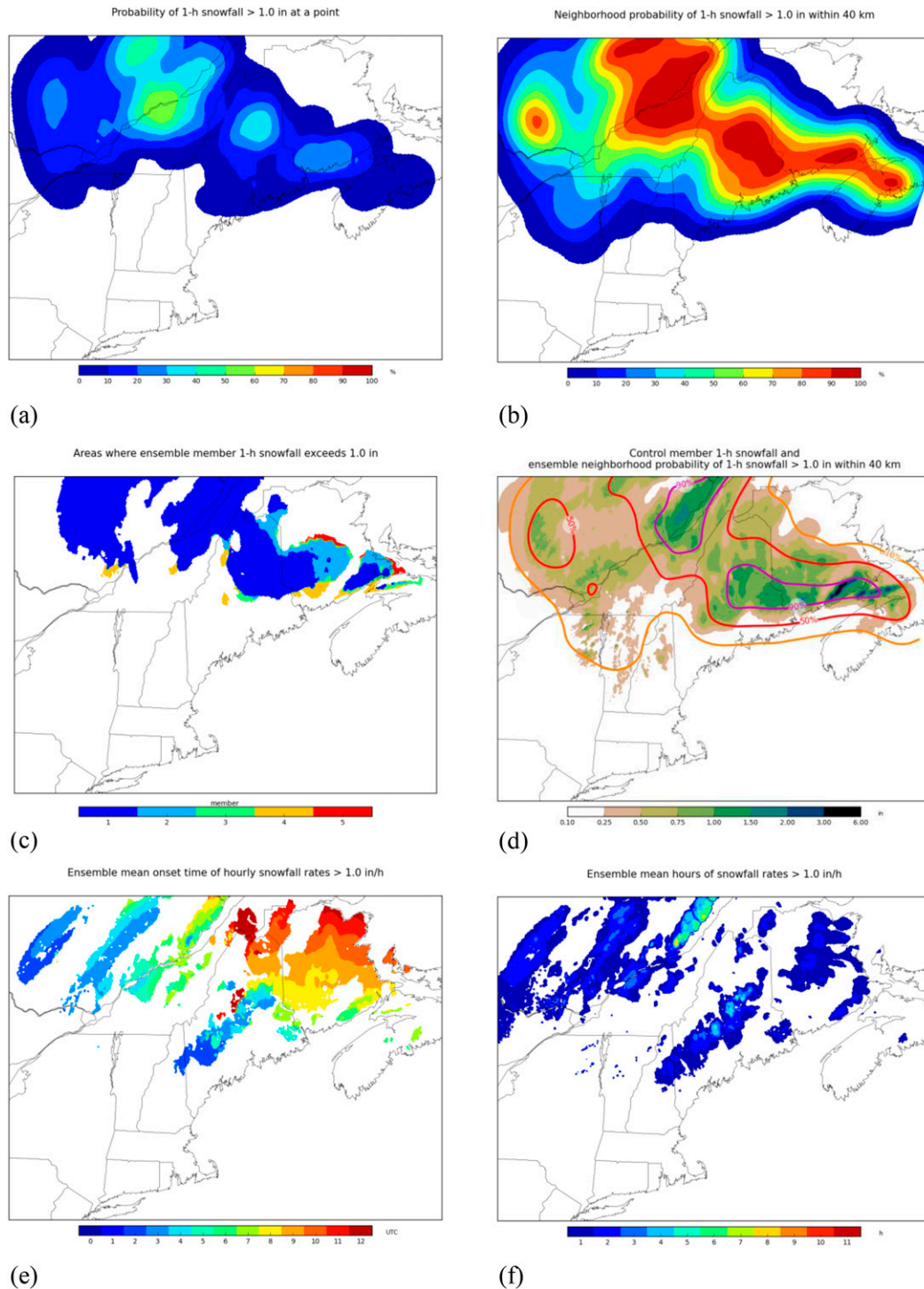


FIG. 2. Example set of winter weather scenario prototype CAM ensemble products that were shown as part of the interview. Plots are (a) point probabilities, (b) 40-km neighborhood probabilities, (c) paintball, (d) “combination” of the ensemble control run and the 10th, 50th, and 90th percentile neighborhood probability contours, (e) mean onset time, and (f) mean duration time.

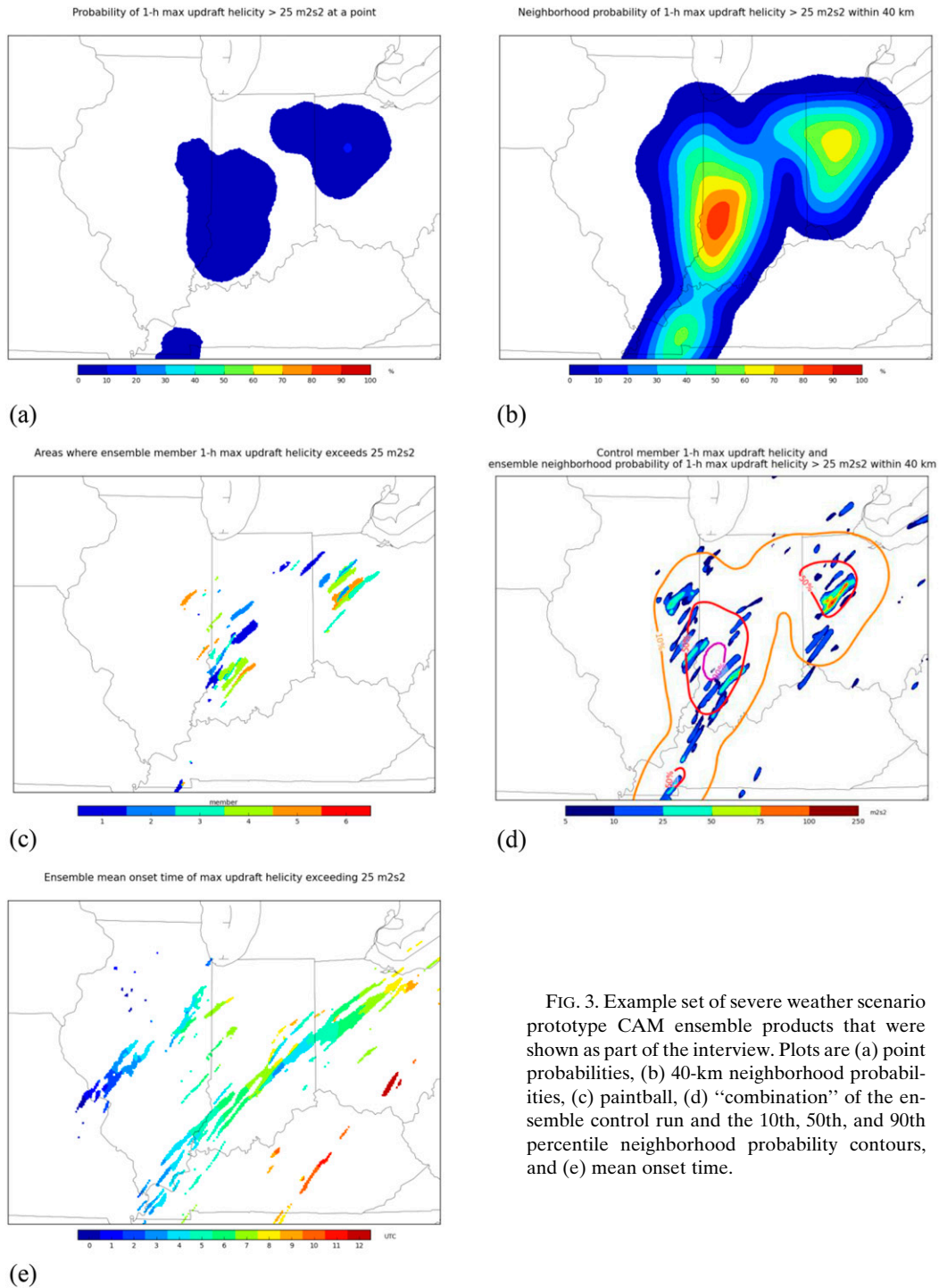


FIG. 3. Example set of severe weather scenario prototype CAM ensemble products that were shown as part of the interview. Plots are (a) point probabilities, (b) 40-km neighborhood probabilities, (c) paintball, (d) “combination” of the ensemble control run and the 10th, 50th, and 90th percentile neighborhood probability contours, and (e) mean onset time.

or figuratively, onto how forecasters view the atmosphere (although there is ongoing work to address this, e.g., Rautenhaus et al. 2018). Forecasters also expressed that, for them, postprocessed probabilistic products are a “black box,” with no easy way to understand what data went into the product or how the resulting information was generated.

Moreover, forecasters know that models have limitations and inaccuracies, and thus part of their forecast process includes assessing critical model errors (see section 4). Doing so is made more difficult with probabilistic guidance, especially guidance that is postprocessed. Because forecasters are scientists, they inherently want to understand how things

TABLE 1. Characteristics of the 31 forecasters interviewed.

Characteristic	Interview sample
Gender	
Female	9
Male	22
NWS experience: median (range)	14 years (4–36 years)
NWS job position	
Intern	2
General forecaster	14
Lead forecaster	14
Service hydrologist	1
Type of hazardous weather discussed	
Nonsevere wind ^a	1
Severe convective weather	7
Heavy rainfall/flash flooding	9
Winter weather	14

^a One forecaster opted to discuss nonsevere wind (winds < 58 mi h⁻¹), which he deemed a common type of high-impact weather for his county warning area.

work. When they cannot easily understand the workings of a probabilistic product or evaluate its accuracy, this reduces their trust in information and their willingness to use it.

To help forecasters overcome the barriers described, the ESRL research team developed a prototype product that we termed the “combination” plot, which includes probability contours derived from postprocessing plotted over output from a single deterministic member. In Fig. 1d, an example combination plot is provided that has the 10th, 50th, and 90th percentile neighborhood probability contours of 3-h rainfall exceeding flash flood guidance (FFG)³ overlaid on deterministic 3-h rainfall amounts from the control member. Similar plots for winter and severe weather are shown in Figs. 2d and 3d. Such plots could be generated in multiple ways, such as with contours of either point or neighborhood probabilities overlaid, or with an ensemble mean, member, or maximum underlain.

When shown the combination plot, most of the forecasters interviewed expressed that they liked it and found it useful (Table 2). Forecasters’ favorable comments indicated that the combination plot helped them better understand and have more confidence in the probabilities (Quotes 2A–2B) and it helped them recognize lower probability risks in some areas (Quotes 2C–2D). Some of the forecasters indicated they would like the ability to scroll among all the different ensemble

members to see each of them as an underlay (Quote 2E), and some forecasters liked being able to compare the probabilities against the ensemble control member, as in the prototype plot developed (Quotes 2E–2G).

However, not all forecasters liked the combination plot. One common reason was that it comprises “way too much information” (No. 21-winter) with “too much going on” (No. 24-winter) and thus would require a lot of time for the forecaster to comprehend. A second reason, expressed by a few forecasters, related to their perceived disconnect between looking at a single piece of guidance and an envelope of guidance from an ensemble (Quotes 2H–2I). For instance, Forecaster No. 11 (-winter) indicated, “I don’t care what members are showing what. If they all have equal probability of occurrence, then it doesn’t matter.” These forecasters do not seem to need a product to transition from using deterministic to probabilistic information because of how they understand and value information from CAM ensembles.

b. Developing new CAM ensemble guidance: Map-based threat timing information

A forecaster’s job is not only to identify *whether* and *where* hazardous weather will occur but also *when* it will occur. The timing of hazardous weather at different locations is inextricably linked to the risk it poses and thus to how a forecaster assesses and communicates the risk to core partners and other users. For instance, forecasters evaluate the risks from precipitation by evaluating amounts over some time frame. This can include assessing multiple waves of heavy precipitation to determine whether there is a threat of flooding or flash flooding (Table 3, Quotes 3A–3B) or assessing whether there is sufficient snowfall to warrant a watch, warning, or advisory. Forecasters also evaluate if hazardous weather might occur at times when people are particularly vulnerable to harm, such as when people are exposed outside with few protective options (Quote 3C), or when weather interacts with transportation systems to amplify negative consequences (Quote 3D).

Because of the importance of the timing of weather at different locations, our data suggest that often forecasters are seeking information about the timing of hazardous weather represented spatially. To ascertain and convey such information, forecasters interrogate observations (e.g., radar reflectivity), deterministic guidance from sequential runs valid at a given time [i.e., $d(\text{prog})/dt$ or time lagged] or from multiple models (i.e., poor man’s ensemble), and ensemble plume diagrams which provide guidance about a particular parameter (e.g., rainfall amount) over time at a geographical point (Quotes 3E–3F). Although these approaches work for forecasters, missing from their toolkit is map-based

³ Flash flood guidance (FFG) is a spatially variable, two-dimensional field. FFG is defined as “a numerical estimate of the average rainfall over a specified area (or pre-defined grid) and time interval required to initiate flooding on small streams” (NWS 2017a, p. 9; see also Clark et al. 2014; Schmidt et al. 2007).

TABLE 2. Forecaster quotes pertaining to feedback on the prototype “combination” plot (see Figs. 1d, 2d, and 3d).

Quote identifier	Forecaster quote
2A	Deterministic is, I think, always easier for people to understand, but I think the probabilities kind of adds to that. [. . .] The combination plot, that’s kind of neat. That overlaying of the probability onto the deterministic. We do not normally see that. Sometimes I’ll have to get probabilities from one model, and I’ll have to get a deterministic from another model. Being able to overlay the probabilities and a deterministic of the same model, I think, gives you a little bit more confidence when you see that. (No. 31-severe)
2B	If it’s 90%, to me that’s pretty highly confident and then I could flip through each of the members and a lot of them would probably show the same thing, I guess. And there might be one or two that show nothing, I guess. But when it’s clustered like that, I mean that’s a pretty good way to visualize it, if I could actually look at each member quickly. That, just seeing 90%, would say, okay, this is probably going to happen but seeing each member on top of that may actually boost confidence a little more. (No. 26-winter)
2C	It kind of tells me what I had seen in [the neighborhood plot] a little bit, is that, you know, there’s fairly heavy amounts quite a bit away from the highest probability. [. . .] You can kind of read between the lines based on how things extend in certain directions, and stuff like that, but seeing one of the members underneath shows actually what some of the values could be. So, there’s value in it. It’s more qualitative for me. [. . .] You know, the fact that this 50% goes quite a ways up the [mountain range], that leads me to believe that there’s probably quite a few that generate heavier precip up that direction. (No. 5-rain)
2D	I like this [combination plot]. It feels better. [. . .] Conceptually, you get your slightly more deterministic look at your one hour that we all use already, but then you’re also met with an ensemble where it’s also saying at any given area, maybe that [updraft helicity] streak, it might be a little further north. (No. 13-severe)
2E	I like [the combination plot] quite a bit. Like I said I’m really big on the neighborhood probabilities and being able to assess each [member] as far as the detail like that is nice. I’d like to be able to see each member if I could. (No. 16-severe)
2F	[The combination plot] has some use, because we operational forecasters do deterministic output, and the control member obviously is sort of the quote-unquote deterministic model forecast, and then you have overlaid the probabilities and you can see that they do match up to a large degree. Of course, the control member has a lot more focal details as far as a lot of the detail where the precipitation will be heavier and not so heavy. [. . .] There is use to this. (No. 1-winter)
2G	I think [the combination plot] is useful [. . .] As I understand it, the control actually behaves differently than the ensemble members. Otherwise, I think the forecaster is probably going to be skewed very hard toward whatever the control member is showing. (No. 17-rain)
2H	The member [underlaid] wouldn’t really help me that much unless I knew something about what the perturbation was and why it might be important to this case. (No. 12-rain)
2I	I think putting any sort of deterministic single member on an ensemble plot kind of takes away from the whole point of it being an ensemble. (No. 25-winter)

ensemble-derived information about threat timing that could more directly meet their needs.

To evaluate the possible benefits of providing this kind of information, the ESRL research team developed two prototype products to illustrate different types of threat timing information that could be derived from an ensemble. One product, termed the “onset” plot, maps the ensemble mean hour of the day when a threshold is first exceeded at a point, with blank areas indicating that no ensemble member exceeds the threshold. The prototype plot shown in Fig. 1e is for the ensemble mean onset time that rainfall rates exceed a threshold. Snowfall rates were shown for the winter weather scenario (Fig. 2e), and updraft helicity values were shown for the severe weather scenario (Fig. 3e). These prototype figures have the data plotted over an 11- or 12-h window, but ensemble mean onset time could be generated over shorter windows to address multiple rounds of hazardous weather. The other product, termed the “duration” plot, maps the ensemble mean number of hours that a

threshold is exceeded at a point. The prototype plot shown in Fig. 1f is for the ensemble mean number of hours that rainfall rates exceed a threshold. Snowfall rates were shown for the winter weather scenario (Fig. 2f), but no duration plot was generated for the severe weather scenario due to the isolated and transient nature of severe convection. These prototype figures have the data plotted over shorter (4 h, Fig. 1f) and longer (11 h, Fig. 2f) windows. The information in Figs. 1f and 2f does not necessarily represent the sequential number of hours that a threshold is exceeded, but such information could be generated. Moreover, the prototype duration plots do not convey the number of ensemble members that contribute to the average values shown, which could be misleading in instances where the information plotted is from only one or few members; thus future prototypes could combine the duration information with probability of occurrence. For both the onset and duration plots, the timing of other parameters or thresholds could be extracted and mapped. Additionally, the time windows over which the output is generated could be a

TABLE 3. Forecaster quotes pertaining to needs for timing information.

Quote identifier	Forecaster quote
3A	[What's important] is timing or being able to identify time-waves of heavier precipitation when they come through. So, a mixture of the models and the satellite imagery, and being able to see the waves in that. Sometimes, if there's a little break between a couple of surges of moisture, that could mean the difference between getting significant flooding and maybe not getting much, because then the rivers have a chance to get rid of a little bit of the water before the next surge comes in. So, being able to time and figure that out. (No. 12-rain)
3B	Timing is really important because, with any big weather event, a lot of our core partners want to know when will it start, when will it be at its worst, and when will it end? [...] Another one of the most important aspects I feel of decision support services, which I think is an avenue we're going down, is that our core partners want to know these waves of precipitation. [...] There could be several low pressure systems to where you have several waves of precipitation moving across the area. [...] We ran into that this past year where we had to describe it in terms of waves, one, two, and three and the timing of all of those. That is essential, not only for core partners, but for us in terms of forecasting because we have to know when are these waves coming, where are they going to be heading, and are they going to be heading the same area which would then be training which would then lead to a higher chance of dealing with flooding. (No. 30-rain)
3C	Let's go back to our DSS ^a support that we provide... I mentioned we've got the barbecue contest going on, but that's one of the events we're keeping a close eye on this weekend. [...] Any probability that we can have like, "well, it could potentially be as early as this if the probability increases by this time", that's really helpful for the DSS services. Especially in the case for most outdoor events—lightning. When is the potential for there to be lightning in the area? Because in many cases if you're dealing with multiple thousands of people outdoors, it's going to take a chunk of time to evacuate people. [...] In some cases our partners request an hour heads up, which can be challenging, so any tools [are needed] that can help us in pinpointing anything like that to assist in the evacuations that might be needed. (No. 18-severe)
3D	What we've come to learn is it's not just the magnitude of the event, but the timing is critically important. If you bring in even just an inch of snow but it's right in the morning rush hour, well, now they cannot plow because there's all the cars in the road, and it's just a mess. And then you wind up with a lot more accidents because, well, they cannot get the plows through because everybody's on the road trying to get to work. So timing is probably one of the questions that we get asked the most: when is it going to start, and when is it going to end? (No. 28-rain)
3E	$d(\text{prog})/dt$ is big in my world. [...] If the models are bouncing all over the place, then I do not even want to touch the forecast because models are struggling from a run-to-run consistency standpoint. [...] How stable is the model solution basically? Stable as in, is there a lot of variability in its output? If there is, then that's just inherently a low confidence forecast. (No. 3-rain)
3F	Some of these [web]sites will show you when the rain will start, so you see the plumes going up, and there's a series of them from all the different runs, all the different members. The SREF ^b will show you in time the increases of rain and then when it will come back down. Or, if you have different waves of precipitation, you're going to see it kind of be volatile. But you'll see how much rain and when, and I like that. I like a plot like that because, if you go into the different points, it will show you kind of in a spatial sense. But in terms of the temporal, you'll see when it thinks the heaviest rain will be coming in. (No. 30-rain)

^a DSS: Decision support services

^b SREF: Short-Range Ensemble Forecast (system)

customizable setting of the interface through which the forecaster analyzes and displays model data.

Most of the forecasters interviewed liked the map-based threat timing information, and they discussed ways the guidance could be useful to them (Table 4). The forecasters mentioned, for example, ways that the mean onset and duration products could help them make a decision about the start time of a warning (Quote 4A) or whether flooding is a risk (Quotes 4B–4C). Many forecasters described how the timing products could help with their messaging to the public and to their core partners (Quotes 4D–4G). It also was mentioned how the timing products, particularly the onset information, could help determine staffing needs for severe weather operations (Quote 4H–4I).

The few forecasters who were less enthusiastic about the map-based timing guidance gave different reasons

for their views. One forecaster described the ensemble mean onset timing product as "a little non-intuitive [...] because you're representing time as space" (No. 15-severe). This same forecaster also questioned whether convection that occurs later in time is being covered up by early convection, a sentiment that also was raised by another forecaster (Quote 4J). A third forecaster expressed confusion about whether long tracks of updraft helicity represent the longevity of a single, strong storm or storm regeneration (Quote 4K).

c. Additional needs for specific CAM ensemble guidance

Beyond the types of information represented in the combination and timing plots, the forecasters interviewed discussed a number of additional types of information

TABLE 4. Forecaster quotes pertaining to feedback on the prototype ensemble mean onset and duration plots (see Figs. 1e,f, 2e,f, and 3e).

Quote identifier	Forecaster quote
4A	This is pretty cool. You know, one of the decisions I was faced with today was whether to maintain the valid start time of a winter storm warning in our mountains, and I ended up pushing it off by a few hours, but it was largely based on just looking at a set of deterministic models and current radar data. [...] You know, when you're changing the valid start time of a warning that is already out there, you want your confidence to be pretty high before delaying the onset time of a warning. [...] In the case like today [...] I probably would have been looking for about a 70 or 80% threshold again, that high confidence that you can change a warning that is already out there. (No. 8-winter)
4B	For our flooding-type events, the mean hours of rainfall greater than a half inch per hour, I think would be very, very useful. Especially concurrent hours because if we have a 12-h forecast and it's hour 1, hour 6, hour 7, hour 9, there might be enough of a break in-between those to let the water flush out without hitting a flood level. (No. 3-rain)
4C	Duration is a big deal. [...] for us with synoptic rains, I mean the duration is what matters. You can have the heavy rain, but if it does not last long enough we do not get enough. The response isn't enough. There's not enough water in the system for us. Has a big influence on the river responses and whether we reach flood stage or not. (No. 5-rain)
4D	The #1 question we get on social media is, "when is it going to hit my house". [...] So for us to be able to pass along that information to our users... I think the onset time is very important. [...] Another example is] If I'm at an event that I'm supporting for an emergency manager and the event is going until 2:00 in the afternoon and so people will be leaving the area between 2:00 and 3:00 p.m. and it looks like there are storms that might be moving in in the afternoon, they want to know is that early afternoon, midafternoon, or is that well after the event is over? And so I think having this ensemble mean onset time for example, of when something is expected to happen I think would be very, very useful for Decision Support Services." (No. 22-rain)
4E	That's actually really cool. [...] That would be helpful as well because again, that goes back to when we're messaging. When's stuff going to develop? How quickly is stuff going to move out of the area? (No. 19-severe)
4F	I like it [the ensemble mean onset plot]. I do. If I've got an EM ^a here, I'd feel really confident that it's not going to snow before 18Z. I feel really confident that it's not going to snow after 0Z. (No. 14-winter)
4G	I like the concept of this [ensemble mean duration plot], of giving an idea of the odds that we're going to have some snowfall rate—especially for DOT ^b crews; it's really important for them—and then how long it could happen for. (No. 6-winter)
4H	I think this could have some utility. I mean, this definitely allows you a quick snapshot look at how the day might unfold vs having to go through the reflectivity swath and seeing, you know, literally, looking at how it goes. [...] I could use it as, okay, there's some kind of hint of rotation through here, and it looks like maybe the beginning stuff could start as early as five. Let's make sure the staffing is here between that and that. (No. 13-severe)
4I	If you have that model consistency [...] of] this looks to really become more organized by maybe this time or we're looking at those discreet super cells really starting to turn up by this time [then the question is] will we have appropriate staffing ready by that time. (No. 18-severe)
4J	This [mean onset plot] can get a little tricky especially if there are [...] multiple rounds of storms that are tracking over the same area at some point. (No. 16-severe)
4K	If [the mean onset plot] is showing that this updraft is persisting that long then [...] that's a long-track super cell. That would be really useful information to have. [Or, is it showing] if it's regenerating? Or if it's a new storm? (No. 23-severe)

^a EM: Emergency manager

^b DOT: Department of Transportation

they would like to have from CAM ensembles, as well as ways that they would like to be able to manipulate or customize that information. Many forecasters volunteered these needs early on in the interview, but viewing the prototype CAM ensemble plots (Figs. 1–3) spurred additional ideas.

We do not discuss these additional suggestions in this article due to space considerations, but a synthesis is provided in Table 5, clustered into two categories. The category “model parameters and outputs” comprises forecasters’ requests for extraction of different characteristics from the ensemble distribution (e.g., earliest onset time of a hazard from the ensemble) and for additional model fields (e.g., ensemble output of precipitable water). The category “statistical and post-processing” comprises

forecasters’ desires to be able to interrogate the ensemble system on-the-fly for criteria that are relevant to their forecast problem of the moment (e.g., query-able exceedance thresholds). The list of needs is not meant to be exhaustive but rather is meant to offer insight into some of the additional types of information and ways of interrogating it that forecasters suggested may be useful to them.

4. Forecasters’ needs for CAM ensemble guidance in support of their shifting role toward IDSS

Coincident with the provision of CAM guidance are ongoing changes in the NWS forecasting environment. Chief among these is a change toward the “partner and customer-centric service delivery model” (NWS 2019)

TABLE 5. Additional forecaster-expressed needs for specific CAM ensemble guidance.

Category	Needs
Model parameters and outputs	For all hazards—Ensemble outputs of end time of hazardous weather, to complement the onset and duration plots (see Figs. 1e,f; 2e,f; and 3e)
	For all hazards—Earliest and latest onset time from the ensemble of a hazard, for a given threshold or for any occurrence (e.g., any snowfall)
	For all hazards—Duration of consecutive hours of threshold exceedance
	For all hazards—Maxima from the ensemble (e.g., maximum precipitation amounts, precipitation rates, spatial extent of precipitation, updraft helicity values)
	For heavy rainfall/flash flooding—Ensemble outputs of precipitable water; runoff
	For severe weather—Ensemble outputs of subsevere hazards (lightning, wind, hail); subhourly output
	For winter weather—Ensemble output of freezing rain, sleet, or other frozen precipitation types; blizzard conditions, rain/snow line
Statistical and postprocessing	Queryable exceedance thresholds
	Selectable exceedance timeframes (e.g., 1-, 3-, 6-h)
	Selectable radii for neighborhood probabilities

of providing impact-based decision support services. IDSS “requires forecasters to ‘go above and beyond the forecast’ to deliver improved service to government agencies” (NWS 2017b) through the “provision of relevant information and interpretive services to enable core partners’ decisions when weather, water, and climate have a direct impact on the protection of lives and livelihoods” (NWS 2018a). The forecasters interviewed regularly referenced IDSS, including how they serve in this role and what challenges they encounter. Because IDSS involves decisions at local levels with NWS core partners and because there is uncertainty in hazardous weather at those spatial and temporal scales, CAM ensemble guidance has the potential to meet forecasters’ IDSS needs.

In discussing the provision of IDSS, some forecasters explicitly emphasized it as a shift from the past in how they do their job, including in how they spend their time and how they think (Table 3: Quotes 3B, 3D, Table 6: Quotes 6A–6B). Forecaster No. 28 (-rain) articulated this shift through an example of a rainfall event that may only produce a moderate amount of rain but that can be high impact in certain circumstances and thus “causes partners all sorts of grief” (Quote 6B). Several other forecasters described their IDSS role through examples of the kinds of information their partners need. Often, these needs pertain to the timing of the risks posed by hazardous weather, as noted in section 3b. Forecaster No. 30 (-rain) succinctly explained that, for significant weather events, “core partners want to know when will it start, when will it be at its worst, and when will it end?” This sentiment was echoed by many of the other forecasters who participated in this study. More specific partner needs for information about threat timing include timing of waves of precipitation in order to gauge flood threats (Quote 3B); timing of convective weather, especially lightning, in order to protect people during

outdoor events (Quote 3C); and whether and when snowfall, especially high precipitation rates, will occur in order for departments of transportation to plan for plowing (Quotes 3D, 4G, 6C). As part of their IDSS role, multiple forecasters also discussed providing to partners the most likely scenario for hazardous weather coupled with goalposts, typically in the form of a best-case and worst-case scenario (Quotes 6D–6E; see also Novak et al. 2008, 2014). Forecaster No. 28 (-rain) explained that when he provides such scenarios to his partners, he draws on ensemble guidance in order to determine how to qualitatively express forecast uncertainty through words and tone, a process he characterized as nuancing the deterministic solution (Quote 6F).

A related theme that commonly emerged as the forecasters discussed providing IDSS is that of conveying forecast confidence (Table 7). Many forecasters discussed how they use (or see the potential to use) CAM guidance to assess and communicate confidence. One way forecasters discussed this topic was using probabilistic output from CAM ensembles to shape their own confidence. For instance, forecasters mentioned that higher probabilities (e.g., of 80% or greater of a parameter that is significant for their forecast process) would increase their confidence, which they may share with users (Quotes 7A–7B). Additionally, Forecaster No. 15 (-severe) discussed how probabilistic guidance can help convey confidence to some partners for low-likelihood, high-consequence risks (Quote 7C). The other way forecasters discussed using CAM guidance to inform their confidence is by looking at model trends for signals of consistency over time (Quote 7D–7E), monotonic changes (Quote 7F), or sharpening of the forecast (Quote 7G). Forecaster No. 30 (-rain) also discussed that she assesses model consistency by examining ensemble member clustering, or lack thereof, as

TABLE 6. Forecaster quotes pertaining to providing IDSS.

Quote identifier	Forecaster quote
6A	In the Weather Service nowadays, there's a lot of different parts to the job besides just sitting in front of a computer and forecasting. There's so many things now. I mean the Weather Service is going the way of decision support so we're taking phone calls from people. We've got [users] that call up, [...] and we have to get ready and do these decisions support slides and everything. (No. 26-winter)
6B	One of the things that we've come to learn—to go down the decision support services [path]—it's not necessarily that 10 inches of rain that gets you that has an impact. It's the 1/2 to 3/4 of an inch that falls at the wrong time. That's one of the things that's taken me a while to learn. You bring in a moderate-to-heavy rainfall during a rush hour say in [city], and it's just a mess. You may not have any flooding at all other than nuisance stuff, but it causes partners all sorts of grief. It would be nice if they could plan for it a little ahead of time. (No. 28-rain)
6C	[State] DOT, ^a when we're talking with them and briefing them, they want to know what kind of snowfall rates are we going to have. Because one inch per hour is bad enough, but at two inches per hour, it's just impossible for them to keep up. And they need to know that. They need to effectively position things the best they can and to start mitigating traffic. (No. 2-winter)
6D	A lot of the time I am working on the IDSS ^b desk, which means talking to a lot of people. So when I'm over there, I want to know the best case scenario and the worst-case scenario, and maybe some sort of most probable situation. (No. 9-rain)
6E	That's why our emergency managers like to come to us. It's like okay, what do I really need to be concerned about and sometimes it comes to point blank asking what do you think is going to happen for my area. That way, we're doing the gathering of the information, and we're deciphering it for them so that they can properly understand the most likely scenario and possible worst-case scenario. Especially our core partners, emergency managers, they appreciate the kind of worst-case scenario for planning purposes. And they seem to have the understanding and appreciation of those challenging obstacles that we have to deal with and things that are variable, so they can take all of that into consideration, but they can kind of plan ahead in that worst-case scenario. (No. 18-severe)
6F	We're doing more decision support services and less actual construction of the gridded dataset. That's the vision of where we're going. We're using ensembles more and more to figure out okay, here is the answer, the deterministic solution, now how do I nuance this? [...] If I'm providing a briefing to state emergency managers, we generally try to [convey], "this is what we think is most likely to happen, this is your plausible worst case scenario, and we're expecting at least this much." [...] The ensembles also give us a shot at when I'm actually briefing an emergency manager or a governor of the state on a winter storm or heavy rainfall event, how do I nuance the words? For example, if I was briefing on [Hurricane] Harvey, am I going to say "it's five days out, it <i>may</i> be scenario where we're going to see a significant amount of rain, or we <i>will</i> see a significant amount of rain"? The ensemble information and having confidence in that is going to color that tone of how the information is presented pretty substantially. (No. 28-rain)

^a DOT: Department of Transportation

^b IDSS: Impact-based decision support services

well as continuity among different models (Quote 7D). Forecaster No. 19 (-severe) extended the idea of determining confidence by evaluating model trends from deterministic output to CAM ensemble output, indicating that he would look for trends in hourly modeled probabilities of updraft helicity to determine if and how they are changing in space and time (Quote 7H).

Although these quotes illustrate that forecasters have developed some understanding about and strategies for providing IDSS, several forecasters discussed challenges for serving in this role (Table 8). One challenge that was raised by a few forecasters is the variability among partners in what constitutes relevant information and interpretive services (Quotes 8A–8B). For some, this challenge is exacerbated by the shift from a long period of the NWS focusing on quantitative, verification-based performance to a more recent focus on impacts, which are more qualitative (Quote 8B). For instance, one forecaster explained that official winter storm warning criteria for his

area of responsibility is 8 in. in 12 h or 12 in. in 24 h, but that "moving forward into more of a DSS world, those thresholds are stretched up and down" (No. 2-winter).

Forecasters also discussed multiple challenges specific to using CAM ensemble and other guidance to provide IDSS. One challenge is how forecasters can interpret and communicate a hazardous weather threat in a way that is meaningful for partners when there is significant uncertainty, such as when there is substantial spread in the ensemble solutions (Quote 8C) or when there is a low probability of a weather event with significant societal impacts (Quote 8D). Related to the difficulty of communicating uncertainty to partners is the challenge of when probabilistic products are not calibrated to be reliable (see section 5 for further discussion), especially when those errors have significant implications, such as money spent on hazard mitigation (Quote 8E) or a visible false alarm to thousands of people (Quote 8F). A final model-based challenge for IDSS is when a marginal

TABLE 7. Forecaster quotes pertaining to determining and communicating confidence when providing IDSS.

Quote identifier	Forecaster quote
7A	If you're expecting supercells, traditionally that's going to give you the strongest max updraft helicity. That's a pretty high probability, 80% to 90% of that type of rotation, that the model is generating some sort of supercell. That builds confidence. (No. 13-severe)
7B	If I knew eight out of ten times I was going to get three inches right there, that's pretty high of a probability. Certainly, our watch and warnings would have more confidence, and I would be telling be able to tell our users with more confidence. (No. 12-rain)
7C	[My primary concern is] when and where we think [hazards are] going to develop and our confidence level. I think getting probabilities may help us communicate our confidence level, even if our confidence level isn't the same as the probability in the model. I think it's still a tool that helps us communicate that. A lot of our higher-end users, EMs and bigger communities, really understand confidence [and] are a little more able to deal with low certainty when you tell them "Hey this is a low probability that happens, but if it does it's going to be a big deal." (No. 15-severe)
7D	Confidence is going to be directly tied to how consistent a model has been. So if a model has been very consistent on showing that the northern portions of [state] are going to be receiving a lot of rain in the next 24 h, then you build confidence with that model. Now, confidence increases even more if you have model continuity – various models all showing a same general idea of what's going to happen, that's also important. So, consistency in time, and continuity between the models. And also especially when you're looking at ensemble members. If you're seeing less of a spread, your confidence is increasing—that model is honing in on a solution that is likely to occur. If there's a lot of variability within the model itself, its different members, then you're confidence decreases. [...] So, you're comparing model to model, comparing its own members, and comparing it in time. (No. 30-rain)
7E	With some of the higher resolution and higher temporal resolution models, if it happens run after run after run, that's an ensemble approach to show this model is capturing this storm in this location. There's a high probability it's going to happen in this location. [...] It gives me confidence that the model's producing something. (No. 23-severe)
7F	If you just see an upward trend in [snowfall probabilities] in the SREF, ^a I know that could be a signal. (No. 29-winter)
7G	[Getting] a track record not only of the deterministic guidance but also how the ensembles have been handling [an event]. Are the ensembles getting more detailed? That would be a sign that we're starting to converge on a solution that is looking to be most likely. So, if we have a wide goalpost and then as we go from [run to run to run] the goalpost keeps getting skinnier and skinnier, the forecast confidence goes up and I gain confidence (No. 28-rain)
7H	The one thing I've learned is not to take any singular forecast verbatim, but to look for trends. Is it a consistent signal? Is it a consistent signal in time and space? If so, what that does is that gives us more confidence that that solution is plausible. If it's changing, how is it changing? Is it the timing of initiation? Is it earlier? Is it later? How's the placement? Is it the same? Is it different? [...] What's the probability of updraft helicity of greater than 25 m ² s ⁻² on an hourly basis [...] What are the trends in this ensemble system? (No. 19-severe)

^a SREF: Short-Range Ensemble Forecast (system)

event occurs that was not captured by the models but that still is high-impact to partners, such as the snow event example discussed by Forecaster No. 2 (-winter) that can lead to flight diversions at the local airport (Quote 8G). Precisely because the models are imperfect and are tools to aid forecasters in better predicting threats, some forecasters emphasized the need to retain their meteorological skills and blend those skills with guidance in order to effectively communicate risks to their users to reduce harm (Quote 8G–8I).

5. Forecasters' needs for model verification and training

In addition to needs for specific CAM ensemble information (section 3) and information relevant to their

IDSS role (section 4), the forecasters interviewed expressed multiple needs for model verification and calibration and for being better trained to use new guidance. As with previous results, these needs emerged both implicitly and explicitly, and in reference to model guidance generally and CAM ensembles specifically.

a. Needs for CAM ensemble verification and calibration

Forecasters know that numerical weather prediction models are imperfect. They discussed many examples of model errors that they have learned about experientially, for example, with precipitation amount, duration, and placement or with certain synoptic situations (Table 9, Quotes 9A–9D). Many forecasters also discussed a broad, general desire to have objective

TABLE 8. Forecaster quotes pertaining to challenges with providing IDSS.

Quote identifier	Forecaster quote
8A	When you're on these webinars, and you say "are there any questions?", and then it's silent, and then you're like, okay, I did a great job in describing what's about to occur. You hang up, and then the phones ring off the hook. They're asking questions that you've already kind of gone through in the webinar. But it makes you realize, okay, I didn't describe this in a way that was most appropriate for this person or this core partner. It depends on the audience. (No. 30-rain)
8B	Our culture has changed quite a bit, too [...] We do not think in impact ways. What impacts one group does not always impact another group. There's so much qualitative stuff in there that, as scientists, we have a hard time. And, the ability to remain consistent becomes difficult. [...] In the early mid-90s, when Jack Kelly came into the Weather Service, everything was numbers-based and verification-based and you would get [admonished] if you did not do this, that, or the other thing. Then, all of a sudden, it went into the impact phase. So after being drilled for so many years using the other way, now we're going into a completely different direction. (No. 11-winter)
8C	[Ensemble model] had a consistent high bias last winter [...] If we'd used it, we basically would have gone out with "expect anywhere from 2 to 48 inches [of snow]" and you cannot do that. [...] I think a lot of customers would say, "well, I could have told you that. That's not why we're coming to you. We're coming to you to help us refine". [...] Everybody can look at models—it's all online. That's why they're coming to us is to provide that skill that maybe they lack, to wade through all of this data and come up with a more realistic answer of what they should be preparing for. (No. 28-winter)
8D	I am worried about probabilities going to partners because they do not mean a lot to them, and I think they are still going to have to have a decision point. To me, giving probabilities to a partner, especially with this high-res data just forces the decision onto them instead of us making that decision. Especially for severe weather information, I do not know what our partners are going to be able to do with percentages that are not 100 or 0. I think it's already confusing, even to some of our really high-end users that know a lot about storms. They get confused with the 10% TOR ^a on SPC, ^b that's really low but that means something bad. So, I think throwing all these numbers at people that do not work in probabilities every day is just going to frustrate them. (No. 15-severe)
8E	I see these 90s [percent probabilities from an ensemble] all the time, and they shift big time or they just completely vanish and they go down [in subsequent model runs]. My biggest problem with these is 90 doesn't always mean 90. [...] What is annoying is when you do find some higher confidence values, and you still have to be a little bit [cautious] because you've been burned before. Especially when it comes to partners. Because when an EM ^c calls, and says at three inches of snow I've got to put trucks out, I've got to pay for overtime and it's going to cost me thousands of dollars, I always hate to say, "yeah, you need to do whatever you got to do because you're going to get your three inches. And then the next day, he's already salted his roads, paid his overtime, and then [the ensemble] says now 60 [percent]. (No. 14-winter)
8F	I do not know if I feel personally [that the models have] been proven [as skillful enough] to let me go out on the limb and [message] that there's a really heightened threat of a supercell and tornado in the next two hours when [no storms have] developed yet [...]. [Model] might show that, and it has 90% [probability of exceeding updraft helicity of some value. . .but] I do not have confidence yet that I can go and message it to several thousands of people [...] Let's be honest, if every time I put [model] solution out when I saw a big supercell and said in the next two hours, "you're going to have a big storm," I'm going to be wrong more than I am right. (No. 13-severe)
8G	Yesterday morning, it snowed before any piece of guidance out there, any short term model, showed any snow. [...] It started early. Yesterday afternoon and evening, it was supposed to snow. But yesterday morning, it was a piece of meteorology that we didn't communicate effectively. The best upper level support was coming across in the morning, which was not in sync with anything else that was happening with the front or anything like that. But it was unstable, we had upper level support, and we had a very shallow upslope component. And probably about 9 times out of 10, thinking back, it usually does something here. It usually does precipitate. And we probably didn't communicate that effectively enough that was a good threat of snow in the morning. And that's probably because maybe we were too reliant on the models. And forgot the little bit about our meteorology backgrounds. [...] It's usually for the marginal events. It's not the classic events. But, at the same time, those events can turn into high-impact events. [...] And they can wreak havoc. Even though they may be short duration, they can end up in 100 flight diversions from [airport]. (No. 2-winter)
8H	I'm definitely more of a not-the-specific numbers person when it comes to probabilities. If [partners] ask for it, yeah, I will give it. But I'm definitely more qualitative, giving them a feel for how I feel about the forecast. But some folks, they want numbers, so, I'll try to convey numbers as best as I can. [...] Again, the high-res guidance will give me confidence, and then I can better formulate a number off of that. Because, a forecast is not just the models. It's your experience. It's climatology. It's all the above. So that's just one piece of the puzzle. (No. 23-severe)
8I	It's just this delicate balance of how I think in an ideal world. If we just do not have time to do proper analysis and spend time doing meteorology, to me it undermines the message and once the message is undermined, it's hard to be good at anything. (No. 19-severe)

^a TOR: Tornado risk^b SPC: Storm Prediction Center^c EM: Emergency manager

TABLE 9. Forecaster quotes pertaining to model verification and calibration needs.

Quote identifier	Forecaster quote
9A	[Model] said we were going to get a bunch of rain across [state], like, 1 and 2 inches. [But] it had been doing this all summer. We've been in a moderate, abnormally dry thing along our coast. [...] The last four or five times we've had a system with this new [model], it said it was going to come in and it was going to dump rain through the area. But the rain has just fallen apart, dried up, and been way less than expected. You see that a bunch of times and you're, like, "No, I'm not kicking that football, Charlie Brown." (No. 25-rain)
9B	In general, [the models] do a pretty good job above 2- or 3000 feet. Where they struggle is what happens in the bottom 2- or 3000 feet. [...] I do not think that there's been an instance this winter—and we've had a lot of events—where the cold air scoured out as fast as what the models showed going into an event. So, we know that the cold air's going to hang on longer [...], but we've had instances where it was [much] longer. We had an ice storm in December where we were initially saying, "the freezing rain would stop [at this time]", but it didn't stop for 36 h after that. (No. 6-winter)
9C	We have the additional issue of, even if [models] are getting the QPF right, sometimes the axis of heaviest snow or freezing rain isn't quite where the model showed it. (No. 7-winter)
9D	I think most people know weakly forced, weakly sheared-type scenarios are very challenging for [models]. But, if we could make a list of what does it struggle with, what does it do really well? Even something as simple as that. But packaged in a way that could concomitantly work with verification scores, that'd be very impactful. (No. 13-severe)
9E	What does a 30% mean for this phenomena, and over what area? (No. 4-winter)
9F	[Probabilistic output would be useful] if they were able to say that the probabilities actually meant what they mean. (No. 11-winter)
9G	I guess without some kind of verification, it would be really hard to know [how to use a probabilistic product. . .] What are the stats on that? Is it 60% of the time that it says there's a 60% chance of an inch of snow, and it snows or does it only verify 40% of the time? [...] I think that [verification] would help give forecasters confidence of, "Okay, I should really believe this value," or "I shouldn't." (No. 6-winter)
9H	In terms of objective verification, I would want to understand what exactly that means. Are they looking at past events to be able to come to this determination? Are they looking at how models have been performing recently? (No. 22-rain)
9I	The big thing for me, especially with anything new, is have they tested it out west? Does it work here? Is it something that's being developed across the Plains? Of course it's going to work great across the Plains. But how does it work out here where you have sparsity of observation data? You do not get to have two or three days of upstream data going into it. I have to gain confidence that it's going to be useful for me here. (No. 3-rain)
9J	The problem I find with objective verification things I see is—and it's no fault, I would do the same thing if I was in [model developers'] boat—but it's always self-serving. Every now and then, they throw in, "we're still having trouble with the cold pools". But they cherry-pick—and maybe it's just perceived that way—three or four excellent, clear-cut cases [...]. I think people would take any kind of objective verification paper or research more seriously if there appeared to be more honesty. [...] It's hard to swallow when you see studies or brown bags like that, and it's like, I hear you telling me this, but I've seen it not work four days in row. (No. 13-severe)
9K	We could break [model verification needs] down into maybe orographics for snow events vs synoptic scale-driven forced snow events. Or, for the West Coast, atmospheric river events vs just your standard weather system moving across. (No. 2-winter)
9L	I would love to have a history of here's how this model has performed. [...] Here is the past 90 days of model performance for grid points exceeding X amount of rain. [...] Or even just how did the 90th percentile amounts pan out? [...] Anything and everything to make it meaningful. You cannot just do it over the entirety of the scenarios because you get too many really light events. I would love to see it focused on high-impact events, whatever you deem those to be. So, here's how it verified where more than an inch was expected or more than two inches [...or] convective vs stratiform events. [...] Or cases where] precipitable water is greater than 1.5 to 2. (No. 17-rain)
9M	But I also want to see, if you look at [model verification] over the whole winter season, this particular model may do the best. But is it doing the best when it really matters? [Like] when there are high winds? Because that's when the impacts are going to be greatest. Maybe a model does best at snowfall amounts over the whole season, but is it catching our higher amounts or does [under forecast]? (No. 24-winter)
9N	What I would probably be doing is calibrating on my own through experience of seeing this is what happens when [model] shows, say, 50%. Or, what is the maximum [model] ever shows in the really big events that should have been obvious—or at least what I think should have been obvious. And what about the run-of-the-mill rain storms we get? What does that look like [from the model]. (No. 12-rain)
9O	It's the same thing with the SPC mesoanalysis pages, where they have all those parameters—they have the supercell parameter, they have the significant tornado parameter. None of these things mean anything. They're calculations based on a forecasted environment through research that's led to us wherever it is, but they're not set values. A day in April with a supercell parameter of 13 might be really good, but in June that might not mean anything. Over time, the forecasters have learned what these values do mean—like, "oh, significant tornado parameter of three today, that's bad". And the numbers start meaning something. I think it's very difficult to make numbers mean anything when you do not have anything you can assign experience to. (No. 13-severe)

TABLE 9. (Continued)

Quote identifier	Forecaster quote
9P	I think that's the biggest issue we're facing is here in the office, or WFOs maybe in general. We're producing these forecasts, and it's on to the next event. You cannot really look back at the last one because you've got to move onto the next one, especially in a busy office like this. And then if you do not have time because you do not have training shifts or extra shifts to do programming, do the verification, then you're just getting through the events and you're producing the forecasts, but you do not really know how you're doing. That kind of bothers me. (No. 24-winter)
9Q	It takes so much to go back and verify. It goes back to the whole time constraint; we're busy doing everything else and we just do not have time to go back and look at a particular model. Also there is accessing the data. We have to try and archive it as fast as we can, and some sites do not have the data going far back. [...] It gets dumped. It only lasts for a week. There is so much data—we're talking hundreds of gigabytes just for an event. You've really got to focus in on what do I really want to verify because we just do not have the space to store all that. (No. 16-severe)
9R	Maybe even showing a sigma [verification statistic] to show how unusual a difference is vs model climatology. I know that there has been a model climatology done at Western region, which I like a lot for the synoptic pattern as far as looking at the sigma level vs model climatology. That's great stuff but we do not have anything for high res guidance in that way. (No. 16-severe)
9S	There has to be some way to collect and display the signal, whether you have to go step by step, hour by hour. [...] Right now, when I want to go back and look at reflectivity and see what are the trends, I have to literally rerun every hour and look online. I look at the 13Z run and see that's what it did. Then, 14Z and 15Z. And I build that model in my head, whereas, if I just had that produced in some fashion, I could quickly look at it. (No. 13-severe)
9T	The question [about developing verification] becomes—especially when you're dealing with convection allowing models—how do you do it? I know there's work out there that's evolving with regard to identifying objects in the model. So you identify objects, and you track those objects, and you can compare those objects to radar data, for instance. [But] what qualifies as a good forecast? If a model develops storms at the right time, and there's a phase error—say it makes the dryline farther east than what happened in reality—you can still consider that a pretty good forecast in all honesty. It was just displaced. And so from an objective standpoint, maybe that model performance would not verify all that well, but I think it adds value to the forecast. (No. 19-severe)

information about model biases and about when a model does and does not perform well (Quote 9D).

In addition to a general need to be informed about model skill, forecasters discussed a number of more specific model verification needs. Some forecasters indicated that in order to use model output, they need to know what it “means.” In the context of probabilistic output, this is one way they expressed wanting to know if the probabilities were calibrated to be reliable (Quotes 9E–9F). This is also implied by Quote 8E from section 4 where Forecaster No. 14 (-winter) indicated that “My biggest problem with [percent probabilities] is 90 doesn't always mean 90.” This quote further reflects that a forecaster can have low confidence in a high-probability forecast. Other forecasters more directly expressed a desire to have reliable probabilities (Quote 9G). These quotes also reflect forecasters' need for training about how to interpret probabilistic guidance, which is discussed further in the next subsection.

Forecasters also expressed wanting to know what the verification “means” with respect to what dataset the metrics were calculated over, including over what timeframes and in what locations (Quotes 9H–9I). Related to this, Forecaster No. 13 (-severe) indicated that he wants model verification to have “more honesty,” meaning done and shown for a collection of

events versus only for a few successful events (Quote 9J). Forecasters also want to be able to stratify the data so that they can assess model performance for the different situations that they forecast for, mapped onto the scenario-based ways that they think about when forecasting. These scenarios include different atmospheric forcing mechanisms (Quotes 9K–9L). They also include weather situations that are high-impact in their forecast area, such as when more than one or two inches of rain is forecast (Quote 9L) or when snow is forecast with high winds, in other words, in the situations that Forecaster No. 24 (-winter) describes as “when it really matters” (Quote 9M). These weather scenarios are high-impact because they can significantly affect society and partners' decisions (Morss et al. 2008), but they are not statistically extreme events, for which model verification would be challenging at best. In the absence of being provided with scenario-based objective verification information, forecasters aim to calibrate guidance for themselves (Quotes 9N–9O). But many forecasters acknowledged how difficult it is to do their own verification due to lack of time and lack of ability to store model data for analysis (Quotes 9P–9Q).

The forecasters discussed a variety of other types of model verification that could help them with utilizing CAM ensembles (and other model guidance) in their

TABLE 10. Forecaster quotes pertaining to model training needs.

Quote identifier	Forecaster quote
10A	I think a well-packaged module that is speaking, very much, the language of a Weather Service employee—and not too on the deep end of a model world and all the parameters that are going into it—would be very useful. The key is [model developers] really have to get real operational people to get feedback [...] because the moment [model developers] go on and show some equations and talk about microphysical schemes and this and that, people are going to tune out. (No. 13-severe)
10B	I know those that are modelers and they want to geek out over the post processing. For me, it's almost cutting it down to how do we communicate with the public? We want to tell them the what, when, where, and how much. For me it's what is this tool, how can I use it, when can I use it, what are the strengths, and what are the limitations of it? I want to be taught how can I use this tool and how is it applicable to my job and how can I use it in operations? [...] When there are so many tools that are out there, it's easy to get lost and then you're like "oh yeah, I forgot about that. I did this one training on it once, but I can't remember it." Actually show and give examples of here's how to use it, here's how it can be helpful to you, or here are the scenarios where you can use this in preparing for your messaging for an event, for staffing, for localized DSS support. (No. 18-severe)
10C	I think ultimately it's helpful to understand the internal dynamics of these models or what not. But, for the most part, people want to know the bottom-line-type stuff. For us in the field, that's the direct application. How can this data source be applied to what we do on a day-to-day basis? So, examples of how you apply the model to certain situations. I think that type of information is beneficial to people more so than going back and looking at the true guts of the model. Because what happens is you go through it and you look at it, but you comprehend it and then it's gone. But if you can actually see how this information can be applied to making a forecast or where it can help benefit the forecast process, I think that that resonates more with an operational person. (No. 19-severe)
10D	I do not look as often at probabilities when it comes to rainfall as I probably should. And I would. So, I guess for me I need to work on that [...] I wish training was maybe more of a priority. And maybe have the regions send out some ways that we can incorporate that into our procedures. (No. 30-rain)
10E	[I want] examples [of probabilities] overlaid this on top of a past event. That way you learn how to interpret the data with an actual weather event going on. So here is an example of where we had these probabilities with these events. And that's where you can also tie in a strength, a limitation, a "hey, be mindful of this, it can overdo this, or don't get too carried away with this." All of that can be tied into an example. (No. 18-severe)
10F	[Regarding a neighborhood probability plot] I do not know what that means. Even the SPC ^a [outlooks] where it says 5% chance of a tornado within 25 miles – there's two qualifiers there. And when you have two qualifiers it's difficult—you've got two moving pieces. [With point probability product], I can at least say one point is fixed, so the location is fixed, and the probabilities are moving. [...] When you have two moving parts, I find it a little bit more difficult to really understand what you're telling me. (No. 14-winter)
10G	We do not really look at neighborhood probabilities much in the field. [...] If we were to get into neighborhood probabilities, we would have to be trained on understanding the difference between that and point probabilities. [...] I've looked at neighborhood probabilities before. I think that they could be pretty useful, as long as we know how to use them right. (No. 1-winter)
10H	[Referring back to the example CAM ^b ensemble products shown (as in Fig. 1) and description of them. . .] This is a good example of showing a plot, of what it looks like, and then a brief application with just a sentence or two. You do not need to go into huge depth for each product. [...] I think just a general web page can be super useful because we can we can go back to that. We do not have to go through a full training module every single time to remember how to use this specific model. That could be really useful. [...] And if you could] click on a question mark and it gives you a brief synopsis of what the model is supposed to be telling you. (No. 23-severe)
10I	There's a lot of sources, locations you can get information [about new guidance]. But if you're not privy to that, if you do not have those right connections, or let's say the SOO ^c doesn't send it out, you'll never know. So you have to be connected to the right outlets to know that they exist. [...] That's why I love conferences because I learn things. I'm like, "wait a second, that exists? I didn't know that." It's a communication issue. Often times, I think things get lost in translation, and it's no one's fault. [...] But we need] training increased to where we notice that [guidance] even exists or what to look at. [...] I'm not sure how much in terms of ensembles for high resolution there are, but I can tell you this right now—it's not as publicized as it's out there. It's not as publicized as I wish it was because then we could tap into that. (No. 30-rain)
10J	When there is a new version of a model or new model that has been developed, usually that kind of information is sent through the layers of the Weather Service, where it originates with headquarters and the Weather Prediction Center, and then it's sent to regions, and then regions send it to local office management, and then local office management sends it to the forecast guy. That can take days or sometimes a couple or a few weeks, but it shouldn't take much more than a few weeks or month at the most to trickle down to a forecaster level, so that we know what the newest datasets and models are available. (No. 1-winter)

TABLE 10. (Continued)

Quote identifier	Forecaster quote
10K	I think there's probably a lot more out there that could use but do not really know about. [We need] a broader awareness of tools that we can use during severe weather, if there are products we should be looking at that we're not. [...] One thing that's been really successful is having a champion in the office—someone that's super excited about it and super involved in it, and when they see something cool they tell everyone. So having an almost a high-res modeling champion in the office that's willing to sell it for you and keep updated on if there is something new coming out [...] not just forwarded e-mails where you have to read down the chain of seven people and then click on a link and then go through a power point. (No. 15-severe)

^a SPC: Storm Prediction Center

^b CAM: Convection-allowing model

^c SOO: Science and Operations Officer (management position in a NWS forecast office)

forecast process. Examples include measures that compare current output to the model's climatology (Quote 9R), measures of model run-to-run consistency and trends (Quote 9S), and measures that quantify spatial or temporal errors (Quote 9T).

b. Needs for CAM ensemble-specific training

The need for training on how to interpret and use model guidance emerged pervasively from the forecasters, revealing the important and multifaceted nature of this issue. Forecasters' opinions were based on their experiences with on-the-job training about forecast tools⁴ generally and on their thoughts on training about CAM ensemble guidance specifically.

A common theme from the forecasters when discussing training about models, including CAM ensembles, was the need for training that is "speaking, very much, the language of a Weather Service employee" (Quote 9A). This means providing less detail (but not zero details) about the "guts of the model" and more content that is clearly and directly relevant to forecast operations, specifically how and when model guidance applies to and benefits their forecast processes (Table 10, Quotes 10A–10C). Forecaster No. 18 (-severe) summarized this as needing to know "what is this tool, how can I use it, when can I use it, what are the strengths, and what are the limitations of it [because] when there are so many tools that are out there, it's easy to get lost" (Quote 10B). Specific to ensemble output, some forecasters specifically mentioned the importance of training to address the challenges they experience with interpreting and using probabilities (Quotes 10D–10G), particularly neighborhood probabilities, which Forecaster No. 14

(-winter) described as having "two qualifiers," the probability and the area, making it difficult to understand (Quote 10F). Correspondingly, Forecaster No. 1 (-winter) observed that forecasters do not look at neighborhood probability products very often and thus would need to be trained on how to understand them if they are to be provided (Quote 10G).

The forecasters had varying opinions about the best mechanisms for training, especially in light of their workloads. Regardless of how a forecaster learns about a new product, they noted that it can be helpful to have readily available and succinct ways of recalling what a given piece of guidance means, especially given the vast number of model products available coupled with the fact that a forecaster may utilize a product only intermittently. An interactive feature, such as a question mark on a web page that a forecaster could click on to obtain a brief description of each product, is one simple refresher mechanism that was suggested (Quote 10H).

In addition to discussing formal training, forecasters described a more general need for improved communication about CAM ensembles. They noted that there are multiple communication sources and channels for learning about new or updated model information, and that this communication is sometimes fragmented, unreliable, slow, or ineffective (Quotes 10I–10K). The ramification of such ineffective communication is that forecasters sometimes are left unaware about existing tools that could aid their forecast process. In other words, even if new CAM ensemble guidance is developed that could be useful to forecasters, the process of operationalizing it breaks down if they are unaware of it.

Understanding forecasters' needs for training and communication about model information, including CAM ensembles, was not a focus of the research conducted here. Nevertheless, we found that addressing these needs is a critical component of enabling more effective use of CAM ensembles by forecasters. Our

⁴ Many forecasters had completed GOES-R training shortly before the interviews were conducted, and thus this experience with learning about the new satellite platform and the observational data provided by it influenced their views.

results also suggest that there is not a one-size-fits-all approach—for either the population of forecasters to be trained or for the content of the training. Thus, more focused work on this topic is needed to advance the development of useful and usable CAM ensemble guidance for forecasters.

6. Summary and discussion

This paper develops knowledge about NWS forecasters' practices, perspectives, and decision contexts in order to identify their needs for CAM ensemble information. The research is based on qualitative data collected from over 50 NWS forecasters through participant observations and semistructured interviews.

We found that forecasters' feedback clusters into three areas, which represent different scales of needs for CAM ensemble information: 1) needs for specific types of products, 2) needs for information that can support forecasters' roles in providing IDSS, and 3) needs for accompanying model verification information and training.

Needs for specific types of CAM ensemble guidance emerged explicitly and implicitly from the forecasters (section 3). Probabilistic guidance can be perceived by forecasters as a black box that is difficult to interrogate to understand the output, including its errors in any given forecast situation. This finding echoes that from Novak et al. (2008) who found that forecasters prefer to interact with ensemble guidance rather than have black-box output. This suggests a need for CAM ensemble information that bridges deterministic and probabilistic forecast representations. The prototype "combination" plot (Figs. 1d, 2d, 3d) developed by the research team is an example of a way to build this bridge. Some forecasters believed this type of information could help them better understand and evaluate ensemble guidance, but it was a too complex for other forecasters. Thus, refining and prototyping additional "bridge" products for forecasters may be useful. A second type of needed CAM ensemble guidance is for information about the potential timing of hazardous weather. Such information can help forecasters better predict and communicate weather risks (e.g., whether precipitation will last long enough to cause flooding or whether snowfall may occur during rush hour), and thus better support partners' decision-making. CAM ensembles are well-suited to providing threat timing information—including forecast uncertainty, such as the earliest possible onset of hazardous weather or exceedance of a threshold—in a map format for efficient use by forecasters. When shown the prototype timing plots (Figs. 1e,f, 2e,f, 3e), most forecasters indicated that these types of information could be useful. This

suggests that more work should focus on deriving, verifying, and providing such output from CAM ensembles. Forecasters articulated additional specific CAM ensemble needs, synthesized in Table 5. These needs include different parameters from those discussed here as well as a desire for a dynamic system that would allow forecasters to define and generate their own type of model-derived information based on the forecast scenario rather than relying on predefined, static outputs.

Needs for information that can help forecasters provide IDSS emerged as a second key area (section 4). The change in the forecasters' role toward providing IDSS is significant and salient. Forecasters are increasingly asked to provide information about the possibility of different high-impact weather threats. Such information requests commonly are for timing information and for specific scenarios, and they pertain to partners' decision-making at refined space and time scales. CAM ensemble output could help forecasters support these IDSS needs. This utility is already being realized to some degree, but there is room for more development of different CAM ensemble parameters and associated forecast uncertainty information to meet these growing needs. Moreover, forecasters discussed that CAM ensemble guidance can help them assess and convey confidence to their partners, a finding that Evans et al. (2014) also reported. However, it is unclear how CAM ensemble guidance shapes their confidence, which suggests that future work should be done to investigate forecasters' thinking in this regard.

Needs for CAM ensemble-specific verification and training were conveyed by many forecasters (section 5). Objective model verification information helps forecasters understand how much to rely on different model guidance, where, and when. Such needs discussed by forecasters included understanding model biases; knowing whether probabilistic guidance is calibrated to be reliable; knowing over which cases, timeframes, and geographies verification was conducted; and having ways to stratify the verification statistics. There has been an increased focus on augmenting traditional verification metrics with metrics that are meaningful to users (e.g., Davis et al. 2006a,b; Gilleland et al. 2010; Sobash and Kain 2017; Wolff et al. 2014). The needs expressed here reinforce and expand on this notion for additional verification metrics, for CAM ensembles and beyond, that are forecaster-oriented such that they are closely aligned with forecast processes and situations.

In order for forecasters to be able to use CAM ensembles, they need to know what information is available, how to access it, and what it means. This need for better and more CAM ensemble-specific training emerged strongly from forecasters. This result is all the

more noteworthy given that no interview questions explicitly asked about training. Forecasters indicated that training content can be disproportionately heavy on how the model was developed (i.e., the research) and light on how to use the model output (i.e., the operations). These forecaster training needs mirror their verification needs in that there is a need for the information to be packaged in a way that is forecaster-oriented. Thus, a shift toward training about the basics of guidance coupled with more information about how output can be used in operations, with forecasting examples, would be beneficial for forecasters. This finding mirrors that from [Novak et al. \(2008\)](#) from over a decade ago, which suggests that this critical need of the forecasters requires more attention.

There are limitations to the research conducted here. The research was simplified in that it did not elicit forecasters' feedback about the prototype products during their actual, real-world forecast process or relative to the existing observational and model data that they have. Although we took care to gather data from NWS forecasters from across the United States about multiple types of hazardous weather, the results cannot be generalized to all forecasters in all areas for all hazard situations. Some CAM ensemble guidance may be of limited use in some situations, such as neighborhood-smoothed probabilities where there are tight gradients (e.g., areas of complex topography, lake-effect snow, bands of heavy snow or rainfall, flash flood guidance values), or when the number of ensemble members is so large that paintball plots become unintelligible. As such, we do not suggest that the CAM ensemble products discussed here—or any product for that matter—serves all forecaster needs all of the time. Furthermore, we recognize that meteorological research will yield more complex CAM ensemble systems in the future, with more members and finer resolution. Also, new approaches to postprocessing are needed because of limitations of neighborhood techniques; related work is ongoing ([Blake et al. 2018](#); [Dey et al. 2014, 2016](#)), and future work could leverage these efforts through additional product development for and evaluation by forecasters. Despite these limitations, we believe that many of the results reported here will apply in the future—particularly the results about forecasters' needs for information about hazardous weather timing; for forecast platforms to dynamically interrogate ensemble guidance; for types of information that help provide IDSS; and for model verification, calibration, and training. The forecaster needs summarized here represent part of the CAM ensemble information development process, and we propose that such research with forecasters should continue in conjunction with future model development efforts.

The research conducted here has important implications for efforts to develop new CAM ensemble information. Such efforts often are motivated by a desire to reduce forecasters' information overload by getting them to adopt new tools that are intended to streamline their forecast process. These motivations are well intentioned. However, the real challenge to address is arguably the continued creation and provision of new products without in-depth understanding of the forecasters' point of view. In other words, we propose that instead of asking how to get forecasters to adopt new tools, the question that ought to be asked is how to effectively create and transition products that forecasters actually want, need, and can use.

Our approach to doing this was to integrate social science research into the traditional R2O process. We employed a risk communication research approach wherein developing new CAM ensemble guidance that is useful and usable to forecasters starts with understanding their job and decision contexts—that is, the roles and functions they perform, the knowledge and information that they do (and do not) possess to carry-out those responsibilities, and their experiences, values, and cultures that comprise the backdrop in which they operate. Utilizing this approach yielded information about forecasters' perspectives and needs beyond their direct feedback about specific products and the concrete informational needs that they can articulate. Robustly incorporating such social science research allowed us to look more broadly and listen more deeply in ways that reveal critical mismatches, gaps, and potential solutions. Some of these pertain to a specific, needed piece of information, such as the combination plot or hazard timing information. Others are broader, such as our findings about the relevance of CAM ensembles in forecasters' complex and evolving role to effectively provide IDSS, and the importance of verification and training for forecasters to effectively utilize the guidance that is available.

The social science research reported here was part of a bigger project and thus was informed by the complementary research into and expertise pertaining to CAM ensemble modeling and postprocessing, verification, and operational forecasting. Such interdisciplinary efforts for improving human weather forecasting, including with probabilistic information, have been advocated for several decades ([Murphy and Winkler 1984](#); [Doswell 2004](#); [Demuth et al. 2007](#); [NAS 2017](#)). Although interdisciplinary research can prove difficult to do meaningfully ([Morss et al. 2018](#)), the benefits are worth the effort.

Acknowledgments. The authors kindly thank all of the NWS forecasters who shared their invaluable time, expertise, and perspectives for this research. We also thank Lance Bosart, reviewer Steve Zubrick, and two

anonymous reviewers for providing helpful comments on the manuscript. This research is supported by National Oceanic and Atmospheric Administration (NOAA), U.S. Weather Research Program (USWRP) Research to Operations (R2O) Award EA-133W-16-CQ-0051. The National Center for Atmospheric Research is sponsored by the National Science Foundation.

REFERENCES

- Árvai, J., 2014: The end of risk communication as we know it. *J. Risk Res.*, **17**, 1245–1249, <https://doi.org/10.1080/13669877.2014.919519>.
- , and V. Campbell-Árvai, 2014: Risk communication: Insights from the decision sciences. *Effective Risk Communication*, J. Árvai and L. Rivers III, Eds., Routledge, 234–257.
- Barthold, F. E., T. E. Workoff, B. A. Cosgrove, J. J. Gourley, D. R. Novak, and K. M. Mahoney, 2015: Improving flash flood forecasts: The HMT-WPC flash flood and intense rainfall experiment. *Bull. Amer. Meteor. Soc.*, **96**, 1859–1866, <https://doi.org/10.1175/BAMS-D-14-00201.1>.
- Benjamin, S. G., J. M. Brown, G. Brunet, P. Lynch, K. Saito, and T. W. Schlatter, 2019: 100 years of progress in forecasting and NWP applications. *A Century of Progress in Atmospheric and Related Sciences: Celebrating the American Meteorological Society Centennial*, *Meteor. Monogr.*, No. 59, Amer. Meteor. Soc., <https://doi.org/10.1175/AMSMONOGRAPHSD-18-0020.1>.
- Blake, B. T., J. R. Carley, T. I. Alcott, I. Jankov, M. E. Pyle, S. E. Perfater, and B. Albright, 2018: An adaptive approach for the calculation of ensemble gridpoint probabilities. *Wea. Forecasting*, **33**, 1063–1080, <https://doi.org/10.1175/WAF-D-18-0035.1>.
- Bosart, L. F., 2003: Whither the weather analysis and forecasting process? *Wea. Forecasting*, **18**, 520–529, [https://doi.org/10.1175/1520-0434\(2003\)18<520:WTWAAF>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)18<520:WTWAAF>2.0.CO;2).
- Braun, V., and V. Clarke, 2006: Using thematic analysis in psychology. *Qual. Res. Psychol.*, **3**, 77–101, <https://doi.org/10.1191/1478088706qp063oa>.
- , —, N. Hayfield, and G. Terry, 2019: Thematic analysis. *Handbook of Research Methods in Health Social Sciences*, P. Liampittong, Ed., Springer, 1–18, https://doi.org/10.1007/978-981-10-2779-6_103.1.
- Clark, A. J., 2017: Generation of ensemble mean precipitation forecasts from convection-allowing ensembles. *Wea. Forecasting*, **32**, 1569–1583, <https://doi.org/10.1175/WAF-D-16-0199.1>.
- , and Coauthors, 2012: An overview of the 2010 Hazardous Weather Testbed experimental forecast program spring experiment. *Bull. Amer. Meteor. Soc.*, **93**, 55–74, <https://doi.org/10.1175/BAMS-D-11-00040.1>.
- Clark, R. A., J. J. Gourley, Z. L. Flamig, Y. Hong, and E. Clark, 2014: CONUS-wide evaluation of National Weather Service flash flood guidance products. *Wea. Forecasting*, **29**, 377–392, <https://doi.org/10.1175/WAF-D-12-00124.1>.
- Cresswell, J. W., 2013: *Qualitative Inquiry and Research Design*. 3rd ed. Sage Publications, 448 pp.
- Daipha, P., 2012: Weathering risk: Uncertainty, weather forecasting, and expertise. *Sociol. Compass*, **6**, 15–25, <https://doi.org/10.1111/j.1751-9020.2011.00437.x>.
- , 2015: From bricolage to collage: The making of decisions at a Weather Forecast Office. *Sociol. Forum*, **30**, 787–808, <https://doi.org/10.1111/sofc.12192>.
- Davis, C. A., B. Brown, and R. Bullock, 2006a: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772–1784, <https://doi.org/10.1175/MWR3145.1>.
- , —, and —, 2006b: Object-based verification of precipitation forecasts. Part II: Application to convective rain systems. *Mon. Wea. Rev.*, **134**, 1785–1795, <https://doi.org/10.1175/MWR3146.1>.
- Demeritt, D., S. Nobert, H. Cloke, and F. Pappenberger, 2010: Challenges in communicating and using ensembles in operational flood forecasting. *Meteor. Appl.*, **17**, 209–222, <https://doi.org/10.1002/met.194>.
- Demuth, J. L., E. Grunfest, R. E. Morss, S. Drobot, and J. K. Lazo, 2007: WAS*IS: Building a community for integrating meteorology and social science. *Bull. Amer. Meteor. Soc.*, **88**, 1729–1738, <https://doi.org/10.1175/BAMS-88-11-1729>.
- Dey, S. R. A., G. Leoncini, N. M. Roberts, R. S. Plant, and S. Migliorini, 2014: A spatial view of ensemble spread in convection permitting ensembles. *Mon. Wea. Rev.*, **142**, 4091–4107, <https://doi.org/10.1175/MWR-D-14-00172.1>.
- , N. M. Roberts, R. S. Plant, and S. Migliorini, 2016: A new method for the characterization and verification of local spatial predictability for convective-scale ensembles. *Quart. J. Roy. Meteor. Soc.*, **142**, 1982–1996, <https://doi.org/10.1002/qj.2792>.
- Doswell, C. A., 2004: Weather forecasting by humans—Heuristics and decision making. *Wea. Forecasting*, **19**, 1115–1126, <https://doi.org/10.1175/WAF-821.1>.
- Evans, C., D. F. Van Dyke, and T. Lericos, 2014: How do forecasters utilize output from a convection-permitting ensemble forecast system? Case study of a high-impact precipitation event. *Wea. Forecasting*, **29**, 466–486, <https://doi.org/10.1175/WAF-D-13-00064.1>.
- Fischhoff, B., 1995: Risk perception and communication unplugged: Twenty years of process. *Risk Anal.*, **15**, 137–145, <https://doi.org/10.1111/j.1539-6924.1995.tb00308.x>.
- Gallo, B. T., A. J. Clark, and S. R. Dembek, 2016: Forecasting tornadoes using convection-permitting ensembles. *Wea. Forecasting*, **31**, 273–295, <https://doi.org/10.1175/WAF-D-15-0134.1>.
- , and Coauthors, 2017: Breaking new ground in severe weather prediction: The 2015 NOAA/Hazardous Weather Testbed spring forecasting experiment. *Wea. Forecasting*, **32**, 1541–1568, <https://doi.org/10.1175/WAF-D-16-0178.1>.
- Gilleland, E., D. A. Ahijevych, B. G. Brown, and E. F. Ebert, 2010: Verifying forecasts spatially. *Bull. Amer. Meteor. Soc.*, **91**, 1365–1376, <https://doi.org/10.1175/2010BAMS2819.1>.
- Gowan, T. M., W. J. Steenburgh, and C. S. Schwartz, 2018: Validation of mountain precipitation forecasts from the convection-permitting NCAR ensemble and operational forecast systems over the western United States. *Wea. Forecasting*, **33**, 739–765, <https://doi.org/10.1175/WAF-D-17-0144.1>.
- Greybush, S. J., S. Saslo, and R. Grumm, 2017: Assessing the ensemble predictability of precipitation forecasts for the January 2015 and 2016 East Coast winter storms. *Wea. Forecasting*, **32**, 1057–1078, <https://doi.org/10.1175/WAF-D-16-0153.1>.
- Henderson, J., 2019: Weather-ready nation or ready-weather agency? Empathic accuracy and sociotechnical resilience in the National Weather Service. *The Sociotechnical Constitution of Resilience*, S. Amir, Ed., Palgrave Macmillan, 93–116, https://doi.org/10.1007/978-981-10-8509-3_5.
- Hirschberg, P. A., and Coauthors, 2011: A weather and climate enterprise strategic implementation plan for generating and communicating forecast uncertainty information. *Bull. Amer. Meteor. Soc.*, **92**, 1651–1666, <https://doi.org/10.1175/BAMS-D-11-00073.1>.
- Homar, V., D. J. Stensrud, J. J. Levit, and D. R. Bright, 2006: Value of human-generated perturbations in short-range ensemble

- forecasts of severe weather. *Wea. Forecasting*, **21**, 347–363, <https://doi.org/10.1175/WAF920.1>.
- Jirak, I. L., and Coauthors, 2010: Evaluation of the performance and distribution of hourly maximum fields from storm-scale ensemble forecasts. *25th Conf. on Severe Local Storms*, Denver, CO, Amer. Meteor. Soc., 13B.3, <https://www.spc.noaa.gov/publications/jirak/ssefmax.pdf>.
- Kain, J. S., and Coauthors, 2013: A feasibility study for probabilistic convection initiation forecasts based on explicit numerical guidance. *Bull. Amer. Meteor. Soc.*, **94**, 1213–1225, <https://doi.org/10.1175/BAMS-D-11-00264.1>.
- Mass, C. F., D. Owens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? *Bull. Amer. Meteor. Soc.*, **83**, 407–430, [https://doi.org/10.1175/1520-0477\(2002\)083<0407:DIHRPM>2.3.CO;2](https://doi.org/10.1175/1520-0477(2002)083<0407:DIHRPM>2.3.CO;2).
- Merriam, S. B., and E. J. Tisdell, 2016: *Qualitative Research: A Guide to Design and Implementation*. 4th ed. Jossey-Bass, 368 pp.
- Morss, R. E., and F. M. Ralph, 2007: Use of information by National Weather Service forecasters and emergency managers during CALJET and PACJET-2001. *Wea. Forecasting*, **22**, 539–555, <https://doi.org/10.1175/WAF1001.1>.
- , J. K. Lazo, B. G. Brown, H. E. Brooks, P. T. Ganderton, and B. N. Mill, 2008: Societal and economic research and applications for weather forecasts: Priorities for the North American THORPEX program. *Bull. Amer. Meteor. Soc.*, **89**, 335–346, <https://doi.org/10.1175/BAMS-89-3-335>.
- , H. Lazrus, and J. L. Demuth, 2018: The “inter” within interdisciplinary research: Strategies for building integration across fields. *Risk Anal.*, <https://doi.org/10.1111/risa.13246>, in press.
- Murphy, A. H., and R. L. Winkler, 1971a: Forecasters and probability forecasts: The responses to a questionnaire. *Bull. Amer. Meteor. Soc.*, **52**, 158–166, [https://doi.org/10.1175/1520-0477\(1971\)052<0158:FAPFTR>2.0.CO;2](https://doi.org/10.1175/1520-0477(1971)052<0158:FAPFTR>2.0.CO;2).
- , and —, 1971b: Forecasters and probability forecasts: Some current problems. *Bull. Amer. Meteor. Soc.*, **52**, 239–248, [https://doi.org/10.1175/1520-0477\(1971\)052<0239:FAPFSC>2.0.CO;2](https://doi.org/10.1175/1520-0477(1971)052<0239:FAPFSC>2.0.CO;2).
- , and —, 1984: Probability forecasting in meteorology. *J. Amer. Stat. Assoc.*, **79**, 489–500.
- NAS, 2017: *Integrating Social and Behavioral Sciences within the Weather Enterprise*. National Academies Press, 198 pp.
- NOAA, 2017: NOAA Administrative Order (NAO) 216-105B: Policy on research and development transitions. NOAA, 24 pp., https://www.corporateservices.noaa.gov/ames/administrative_orders/chapter_216/Handbook_NAO216-105B_03-21-17.pdf.
- , 2020: Hydrometeorology Testbed: Experiments. NOAA, accessed 12 July 2018, <https://hmt.noaa.gov/experiments/>.
- Novak, D. R., D. R. Bright, and M. J. Brennan, 2008: Operational forecaster uncertainty needs and future roles. *Wea. Forecasting*, **23**, 1069–1084, <https://doi.org/10.1175/2008WAF2222142.1>.
- , K. A. Brill, and W. A. Hogsett, 2014: Using percentiles to communicate snowfall uncertainty. *Wea. Forecasting*, **29**, 1259–1265, <https://doi.org/10.1175/WAF-D-14-00019.1>.
- NRC, 1989: *Improving Risk Communication*. National Academies Press, 352 pp.
- NWS, 2017a: National Weather Service Instruction 10-912: River Forecast Center Products Specification. NOAA/NWS, 18 pp., <https://www.nws.noaa.gov/directives/sym/pd01009012curr.pdf>.
- , 2017b: Operations and workforce analysis catalog. NOAA/NWS, 132 pp., https://www.weather.gov/media/nws/OWA_Catalog_09072017.pdf.
- , 2018a: NWS service description document (SDD): Impact-based decision support services for NWS core partners. NOAA/NWS, 24 pp., https://www.nws.noaa.gov/im/IDSS_SDD_V1_0.pdf.
- , 2018b: National Weather Service Instruction 10-511: WFO Severe Weather Products Specification. NOAA/NWS, 64 pp., <https://www.nws.noaa.gov/directives/sym/pd01005011curr.pdf>.
- , 2019: Building a weather-ready nation: 2019-2022 strategic plan. NOAA/NWS, 28 pp., https://www.weather.gov/media/wrn/NWS_Weather-Ready-Nation_Strategic_Plan_2019-2022.pdf.
- Rautenhaus, M., M. Böttinger, S. Siemen, R. Hoffman, R. Kirby, M. Mirzargar, N. Röber, and R. Ruediger, 2018: Visualization in meteorology—A survey of techniques and tools for data analysis tasks. *IEEE Trans. Vis. Comput. Graph.*, **24**, 3268–3296, <https://doi.org/10.1109/TVCG.2017.2779501>.
- Roebber, P. J., and L. F. Bosart, 1996a: The contributions of education and experience to forecast skill. *Wea. Forecasting*, **11**, 21–40, [https://doi.org/10.1175/1520-0434\(1996\)011<0021:TCOEAE>2.0.CO;2](https://doi.org/10.1175/1520-0434(1996)011<0021:TCOEAE>2.0.CO;2).
- , and —, 1996b: The complex relationship between forecast skill and forecast value: A real-world analysis. *Wea. Forecasting*, **11**, 544–559, [https://doi.org/10.1175/1520-0434\(1996\)011<0544:TCRBF5>2.0.CO;2](https://doi.org/10.1175/1520-0434(1996)011<0544:TCRBF5>2.0.CO;2).
- , D. M. Schultz, B. A. Colle, and D. J. Stensrud, 2004: Toward improved prediction: High-resolution and ensemble modeling systems in operations. *Wea. Forecasting*, **19**, 936–949, [https://doi.org/10.1175/1520-0434\(2004\)019<0936:TIPHA5>2.0.CO;2](https://doi.org/10.1175/1520-0434(2004)019<0936:TIPHA5>2.0.CO;2).
- Rothfus, L. P., R. Schneider, D. Novak, K. Klockow-McClain, A. E. Gerard, C. Karstens, G. J. Stumpf, and T. Smith, 2018: FACETS: A proposed next-generation paradigm for high-impact weather forecasting. *Bull. Amer. Meteor. Soc.*, **99**, 2025–2043, <https://doi.org/10.1175/BAMS-D-16-01001.1>.
- Schmidt, J. A., A. J. Anderson, and J. H. Paul, 2007: Spatially-variable, physically-derived flash flood guidance. *21st Conf. on Hydrology*, San Antonio, TX, Amer. Meteor. Soc., 6B.2, https://ams.confex.com/ams/87ANNUAL/techprogram/paper_120022.htm.
- Schutt, R. K., Ed., 2012: Qualitative data analysis. *Investigating the Social World: The Process and Practice of Research*, Sage Publications, 320–357.
- Schwartz, C. S., 2014: Reproducing the September 2013 record-breaking rainfall over the Colorado Front Range with high-resolution WRF forecasts. *Wea. Forecasting*, **29**, 393–402, <https://doi.org/10.1175/WAF-D-13-00136.1>.
- , G. S. Romine, R. A. Sobash, K. R. Fossell, and M. L. Weisman, 2019: NCAR’s real-time convection-allowing ensemble project. *Bull. Amer. Meteor. Soc.*, **100**, 321–343, <https://doi.org/10.1175/BAMS-D-17-0297.1>.
- Snellman, L. W., 1977: Operational forecasting using automated guidance. *Bull. Amer. Meteor. Soc.*, **58**, 1036–1044, [https://doi.org/10.1175/1520-0477\(1977\)058<1036:OFUAG>2.0.CO;2](https://doi.org/10.1175/1520-0477(1977)058<1036:OFUAG>2.0.CO;2).
- Sobash, R. A., and J. S. Kain, 2017: Seasonal variations in severe weather forecast skill in an experimental convection-allowing model. *Wea. Forecasting*, **32**, 1885–1902, <https://doi.org/10.1175/WAF-D-17-0043.1>.
- , G. S. Romine, C. S. Schwartz, D. J. Gagne, and M. L. Weisman, 2016: Explicit forecasts of low-level rotation from convection-allowing models for next-day tornado prediction. *Wea. Forecasting*, **31**, 1591–1614, <https://doi.org/10.1175/WAF-D-16-0073.1>.
- Stensrud, D. J., and Coauthors, 2013: Progress and challenges with Warn-on-Forecast. *Atmos. Res.*, **123**, 2–16, <https://doi.org/10.1016/j.atmosres.2012.04.004>.
- Stewart, T. R., P. J. Roebber, and L. F. Bosart, 1997: The importance of the task in analyzing expert judgment. *Organ. Behav.*

- Hum. Decis. Process.*, **69**, 205–219, <https://doi.org/10.1006/obhd.1997.2682>.
- Stuart, N. A., and Coauthors, 2006: The future of humans in an increasingly automated forecast process. *Bull. Amer. Meteor. Soc.*, **87**, 1497–1502, <https://doi.org/10.1175/BAMS-87-11-1497>.
- , D. M. Schultz, and G. Klein, 2007: Maintaining the role of humans in the forecast process: Analyzing the psyche of expert forecasters. *Bull. Amer. Meteor. Soc.*, **88**, 1893–1898, <https://doi.org/10.1175/BAMS-88-12-1893>.
- Weisman, M. L., C. Davis, W. Wang, K. W. Manning, and J. B. Klemp, 2008: Experiences with 0–36-h explicit convective forecasts with the WRF-ARW Model. *Wea. Forecasting*, **23**, 407–437, <https://doi.org/10.1175/2007WAF2007005.1>.
- Wilson, K. A., P. L. Heinselman, P. S. Skinner, J. J. Chaote, and K. E. Klockow-McClain, 2019: Meteorologists' interpretations of storm-scale ensemble-based forecast guidance. *Wea. Climate Soc.*, **11**, 337–354, <https://doi.org/10.1175/WCAS-D-18-0084.1>.
- Wolff, J. K., M. Harrold, T. Fowler, J. H. Gotway, L. Nance, and B. G. Brown, 2014: Beyond the basics: Evaluating model-based precipitation forecasts using traditional, spatial, and object-based methods. *Wea. Forecasting*, **29**, 1451–1472, <https://doi.org/10.1175/WAF-D-13-00135.1>.