Improving the Use of Hydrologic Probabilistic and Deterministic Information in Decision-Making

Rachel Hogan Carr, Kathryn Semmens,
Burrell Montz, and Keri Maxfield

ABSTRACT: Uncertainty is everywhere and understanding how individuals understand and use forecast information to make decisions given varying levels of certainty is crucial for effectively communicating risks and weather hazards. To advance prior research about how various audiences use and understand probabilistic and deterministic hydrologic forecast information, a social science study involving multiple scenario-based focus groups and surveys at four locations (Eureka, California; Gunnison, Colorado; Durango, Colorado; Owego, New York) across the United States was conducted with professionals and residents. Focusing on the Hydrologic Ensemble Forecast System, the Advanced Hydrologic Prediction Service, and briefings, this research investigated how users tolerate divergence in probabilistic and deterministic forecasts and how deterministic and probabilistic river level forecasts can be presented simultaneously without causing confusion. This study found that probabilistic forecasts introduce a tremendous amount of new, yet valuable, information but can quickly overwhelm users based on how they are conveyed and communicated. Some were unaware of resources available, or how to find, sort, and prioritize among all the data and information. Importantly, when presented with a divergence between deterministic and probabilistic forecasts, most sought out more information while some others reported diminished confidence in the products. Users in all regions expressed a desire to "ground truth" the accuracy of probabilistic forecasts, understand the drivers of the forecasts, and become more familiar with them. In addition, a prototype probabilistic product that includes a deterministic forecast was tested, and suggestions for communicating probabilistic information through the use of briefing packages is proposed.

KEYWORDS: Social Science; Probability forecasts/models/distribution; Communications/decision making

AFFILIATIONS: **Hogan Carr, Semmens, and Maxfield**—Nurture Nature Center, Easton, Pennsylvania; **Montz**—East Carolina University, Greenville, North Carolina

The technical capacity for probabilistic forecasting has advanced considerably, and there have been calls for wide dissemination of these forecasts. The U.S. National Research Council (2006, p. 12) asserted that "by providing mainly single valued categorical information, the hydrometeorological prediction community denies its users much of the value of the information it produces—information that could impart economic benefits and lead to greater safety and convenience for the nation." Similarly, Michaels (2015, p. 44) indicated that "the use of probabilistic flood forecasts is in tune with the wider trend in public policy to employ risk-based decision making." While probabilistic forecasts provide information on the range of possible outcomes and thus are explicit about the uncertainty inherent in a given forecast, there remain questions about how to effectively present such information recognizing that there are various public and professional audiences (Budimir et al. 2020). As a result, conveying uncertainty is a significant risk communication challenge (Severtson and Myers 2013). For example, previous research (Hogan Carr et al. 2018) has shown that the hydrograph is a much-preferred product for hydrologic information, and participants have requested that probabilistic information be shown in the context of a deterministic product. But these products have also proven to be difficult for many to understand together. Further, there is evidence that people underestimate the uncertainty in deterministic forecasts and may (mis)interpret a probabilistic forecast as deterministic (Fleischhut et al. 2020).

A number of researchers have recognized that there is still much work to be done to make risk-related, probabilistic information usable by various public and professional audiences (Wood et al. 2012; Spiegelhalter et al. 2011; Ramos et al. 2010), especially relating to "alternative ways to communicate risk and uncertainty for low-probability, high-consequence events" (Bostrom et al. 2008, p. 36). On one hand, some question whether people can successfully make use of uncertainty information given biases and expectations that may influence interpretations of this information (Joslyn and Savelli 2010) while others assert that providing uncertainty information to the public in an accessible format may help people decide how much confidence to place in a given forecast (Morss et al. 2008). Indeed, research has suggested that communicating information about data uncertainty has the potential to increase trust in results and to support decision-making that uses those data, whether it is the public or professional users, but that there is a need to evaluate the techniques used (Kinkeldey et al. 2014; Hullman et al. 2018).

Communicating forecasts effectively requires understanding how intended audiences interpret and use forecast information presented in different ways (Morss et al. 2010). As suggested by Palmer (2002, p. 753), "most of the time, the ordinary person does not have the motivation to digest the extra information that is implicit in a probability weather forecast." This aligns with the findings of Joslyn and Savelli (2010) that many people anticipate some uncertainty in the deterministic forecasts. At the same time, when deterministic and probabilistic forecasts are both available and there are discrepancies between them in the data shown, trust in both declines (Hogan Carr et al. 2018). In contrast, practitioners may make poorer decisions if they do not have the benefit of taking forecast uncertainties and risks into account (Hirschberg et al. 2011). For instance, one study reported that professionals in national hydrologic services in Europe found that threshold forecasts that used both deterministic and probabilistic forecasts were more useful to better evaluate the risk of a potential flood (Ramos et al. 2007). Yet for professionals, probabilistic forecasts can be

challenging in decision-making when binary decisions (issue a warning or not) must be made (Arnal et al. 2020). Combined, these studies illustrate the complications that arise from the fact that there are differences in the understanding of probabilistic forecasts depending on the type of user, but there are also differences within groups (Fundel et al. 2019; Hogan Carr et al. 2018; Kox et al. 2015). Further, the thresholds at which such information will motivate action differ among users (Morss et al. 2010).

In previous National Oceanic and Atmospheric Administration (NOAA)-funded studies, Nurture Nature Center and East Carolina University tested various National Weather Service (NWS) probabilistic flood forecast products among other NWS tools. The first included the significant river flood outlook, watches, and warnings, the Advanced Hydrologic Prediction Service (AHPS) hydrograph, and the Meteorological Model Ensemble Forecast System (MMEFS) used in the eastern region to provide probabilistic hydrologic guidance (Hogan Carr et al. 2016). Among the recommendations from that study were changes to the products so that they would be more easily understood by users and more likely to motivate action.

The other study to address probabilistic products focused on the Hydrologic Ensemble Forecast System (HEFS) and found that the presentation of probabilistic information via HEFS alongside deterministic information (i.e., hydrograph) in scenario-based focus groups created significant barriers to understanding (Hogan Carr et al. 2018). Audiences struggled to understand the deterministic and probabilistic information together and, in one situation, experienced decreased trust in both the hydrograph and the HEFS products as a result. Additionally, the study found that modifications to the display and presentation of the information helped improve user understanding of the forecast.

It is clear from the research that both deterministic and probabilistic information are important to a wide range of users, even though the relative utility of each will vary depending on the users' needs and decisions as well as their understanding of the data presented in a given product. Further, it is not just the availability of ensemble forecasts that is important, but how that information is presented. Indeed, one study reported that among the lessons learned in their research is the need for engagement and collaboration on the design of probabilistic forecasts (Nobert et al. 2010). Thus, in addition to understanding how to present deterministic and probabilistic forecasts simultaneously without diminishing the value of either or both, it is also necessary to consider how uncertainty should be presented to be most effective for various audiences.

This paper shares findings from a study that advances prior research about how various audiences use and understand probabilistic hydrologic forecast information, testing three forecast products, and proposing improvements to the display and communication of uncertainty and probabilistic information in hydrologic forecasts. The three products were studied as they are used in various regions across the country, and include 1) AHPS and regional hydrographs [e.g., hydrographs developed by NWS River Forecast Centers (RFCs) or Weather Forecast Offices (WFOs)]; 2) outputs from the HEFS, including seasonal water supply forecast related products; and 3) briefings for impact-based decision support services (IDSS). Taken together, the findings provide information on the issues associated with presenting probability forecasts and with public understanding. The full study addressed several research questions, including questions related to timing of information and the use of briefings. This study was not designed to deal with or assess the difference between user preference and actual user decision efficacy; rather, the aim was to investigate user understanding of the forecast products, the barriers to understanding, and to recommend product modifications that increase their intention to use forecast products in decision-making. This paper focuses specifically on two of the study's key research questions related to understanding of probabilistic forecast products as addressed above, namely,

- how users tolerate divergence in probabilistic and deterministic forecasts and
- how deterministic and probabilistic river level forecasts can be presented simultaneously without causing confusion for the users.

Methodology

Hydrographs developed from AHPS as well as regional offices, outputs from HEFS, and a variety of emergency briefings (delivered as multipage PDF files containing a range of forecast products and notes about impending events, developed by local WFOs) were tested through three methods (focus groups, in-person surveys, and an online survey), in four different geographic locations, and with two different audiences: residents and professionals (emergency managers, water resources professionals) in Eureka, California; Owego, New York; Gunnison, Colorado; and Durango, Colorado. Two rounds of in-person focus groups were held in each community (round 1 in spring of 2019 and round 2 in fall of 2019), and in each round, two focus groups were held at each of the four locations—one for professionals and one for residents—for a total of 16 focus groups.

Working with RFCs and WFOs in each location, the project team developed four regionally relevant hypothetical weather scenarios (one for each region) that told the story of an impending weather event through forecast products; these scenarios were used as the basis for the two-hour focus groups. The scenarios included a range of NWS forecast products as well as products from other government agencies, including the U.S. Geological Survey (USGS) and the National Resources Conservation Service (NRCS). Scenarios emphasized, through repeated inclusion, the products at the center of this study, namely, HEFS, AHPS and regional hydrographs, and briefings. The scenarios also contained supporting weather information, including precipitation and temperature forecasts, watches and warnings and snowmelt information, as needed to help participants understand the scenario.

The scenarios were structured to start at a given point in time and move forward toward the ultimate weather event, or day T. The duration of the scenarios was established in conjunction with WFO and RFC partners and varied based on the regional patterns for hydrologic weather events; the scenarios ranged from 2 months to 7 days out across the communities, reflecting tendencies for regions to need either seasonal (i.e., long-term drought) or more acute (i.e., imminent flood) weather information. These scenarios were constructed in the ESRI Story Map platform so they could be easily shared and are summarized later in the region-specific tables. A focus group protocol including products and questions shown during the focus groups is provided as an online supplement (https://doi.org/10.1175/BAMS-D-21-0019.2).

The research team has a long history of collaboration with NWS WFOs and RFCs on social science research studies and established this partner arrangement in the conception of this research. NWS partners were guided by the research team on the type and format of the overall scenario during progressive discussions. The NWS partners provided the specific details of the scenario based on previous events and experiences in their region, assisted the research team in finding locations and means of outreach for the focus groups, and provided guidance on the scientific accuracy of the revisions made to the tested products.

The project team worked with NOAA partners to determine easily accessible public meeting places, ranging from a public library in Durango to meeting space at a fairground in Gunnison. Following approval by East Carolina University's University and Medical Center Institutional Review Board, flyers for each focus group were developed and shared through partner's contacts, as well as through social media, local news outlets, and local organizations focused on waterways. Participants were required to register for each session via an online web form. All nongovernmental participants were offered $50 as compensation.

When participants arrived for their respective focus group, they were each given an iPad and asked to complete a presession survey. They kept the iPads to follow along with the scenario.

The scenario was simultaneously projected on a large screen and on each iPad. This allowed participants to zoom in on product details if needed. The facilitator walked the participants through each day in the scenario, asking questions about understanding, motivation to take action, and resource needs. Discussion was recorded and transcribed for analysis with NVivo, a social science software used for qualitative analysis, which involved using the forecast products, and themes related to the research questions, for content research and findings. In that analysis, the products were used as nodes in keeping with the research questions, providing a categorical partitioning of the transcribed discussions (Krippendorf 2018). Such categorization allowed us to synthesize product-specific feedback and from that to identify themes and trends in responses (Hsieh and Shannon 2005).

A postsession survey was completed by all participants (survey instruments are included in supplemental material) during the in-person meeting. The surveys were developed with NOAA partners and designed to collect quantitative information on participants' characteristics and experiences, their reactions to products, and to address the study's research questions. The same process and scenarios were used for both round 1 and round 2, with the exception that round 2 scenarios included the revised HEFS products. Round 2 focus groups were conducted with new participants in order to eliminate the influence of previous exposure to the scenario and products on responses to the revised product and allow for the assessment of differences in understanding of the revised and original products.

The graphical and design revisions to the probabilistic flood forecast products were based on analysis of pre- and postsession surveys as well as focus group notes, transcripts, and content analysis. Survey responses were quantified using Excel to tabulate the data and calculate the percentage of participants answering each choice. These results, combined with the analysis using NVivo, were used to identify trends and themes about individual products from the focus group discussions. Of particular relevance were specific postsession survey questions asking about each design element in the product and whether each was useful or not. Participants could explain which elements were most or least useful.

These survey results combined with focus group discussion analysis, and research team discussions, led to a series of HEFS revisions specific to each site. While specifics of each of these findings are beyond the scope of this paper, revisions to the HEFS products were made that included changes to the color scheme, design, legends, and title, and also included the addition of a forecaster's note, which is a dedicated space near the top of the product designed for forecaster-driven text-based messaging to accompany the graphical product. Region-specific revisions included adding an interactive text box for USGS historical values for Colorado products, and adding a river level exceedance vertical bar as a side box for California and New York, element additions that were supported by findings from the surveys and focus group transcripts.

Following analysis of the second round of focus groups, the HEFS graphics were revised again, and an online survey (round 3) was developed and administered in March 2020 to all participants from both rounds 1 and 2. This survey (included in the online supplemental material) showed the newly revised HEFS products (shown here in Figs. 1–4 next to the original round 1 products), as well as a prototype of a national version of the HEFS. Collective findings from survey and focus group data analysis from rounds 1 and 2 across all regions were used to inform the graphic redesign of the products, which was led by an expert designer using established principles for visual communication, including considerations for accessibility. Adobe software was used to proof images with Color Universal Design (CUD) to help ensure that graphical information was communicated accurately to people with various types of color vision deficiency. Font sizes and weights were considered in every phase of the design process. Responsive screens were used during the focus groups so that participants could
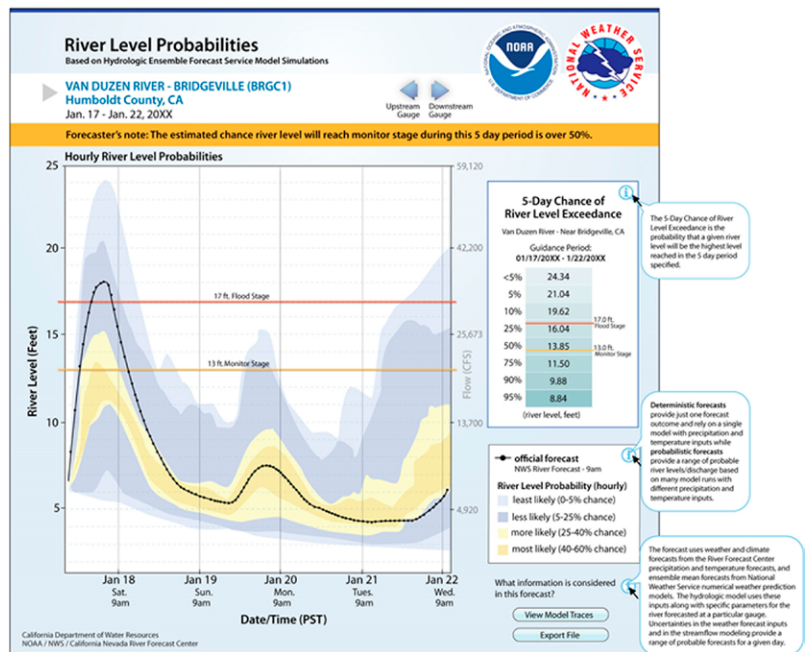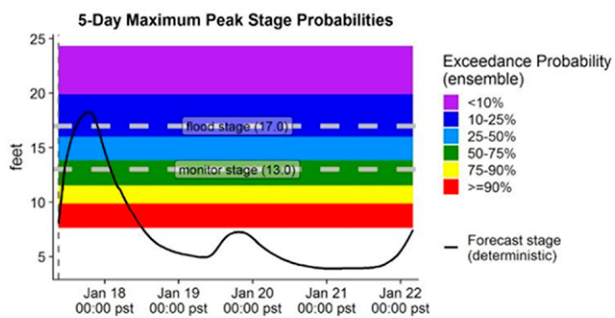
Fig. 1. The probabilistic flood level forecast product (HEFS) shown in (left) round 1 and (right) round 3 after revisions for Eureka.
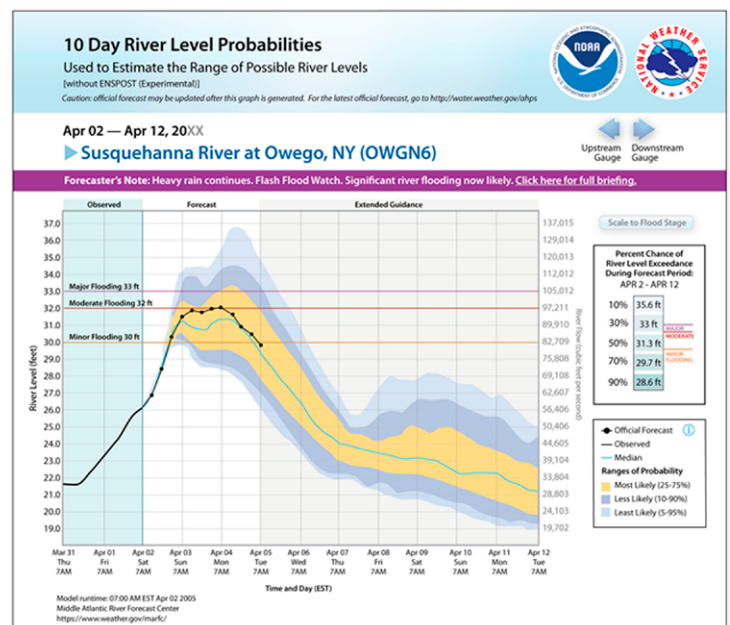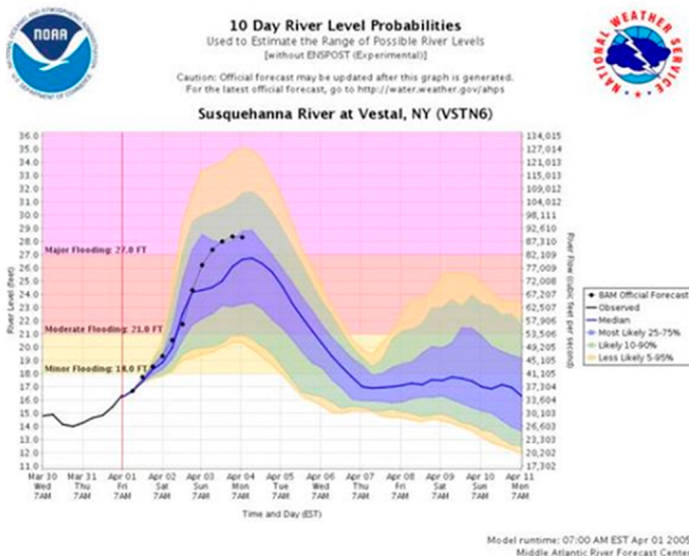

Fig. 2. The probabilistic flood level forecast product (HEFS) shown in (left) round 1 and (right) round 3 after revisions for Owego.

zoom in on a product for better visibility. All revisions were made in consultation with NWS partners to ensure scientific accuracy was retained as changes were made to facilitate user understanding.

While the presession surveys provided an understanding of participants' experiences with extreme events, the postsession surveys sought to quantify the utility and perception of the forecast products shown during the focus groups, as well as how participants report they would respond to a divergence in the probabilistic and deterministic forecasts. In the final online survey, participants were asked about the usefulness of the three main products on which the study focused—the deterministic hydrograph, the HEFS (probability of river level),
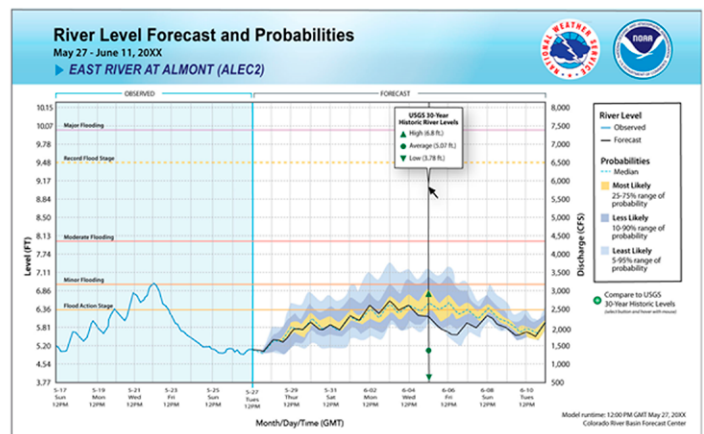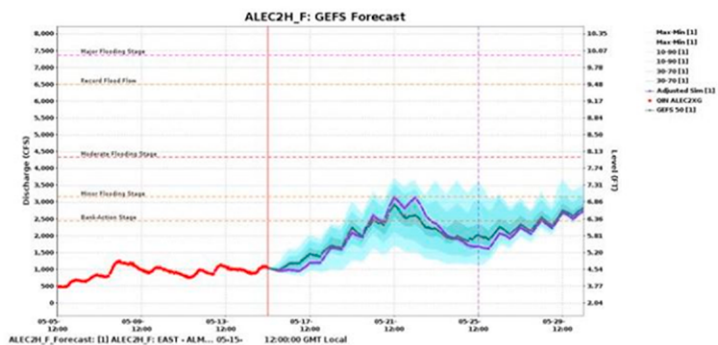
Fig. 3. The probabilistic flood level forecast product (HEFS) shown in (left) round 1 and (right) round 3 after revisions for Gunnison.
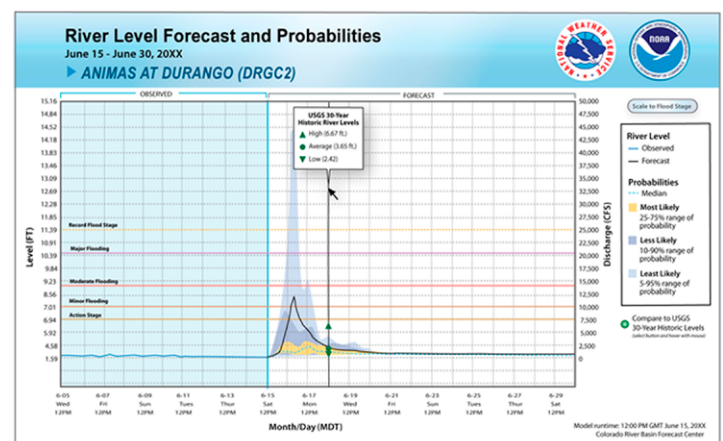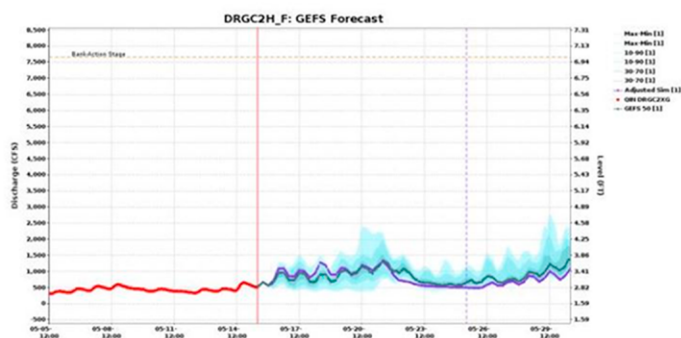

Fig. 4. The probabilistic flood level forecast product (HEFS) shown in (left) round 1 and (right) round 3 after revisions for Durango.

and the briefing package—as well as how they would interpret a divergence in the probabilistic and deterministic forecasts.

Results

Characteristics of participants. The total number of participants by location and focus group for both round 1 and round 2 are shown in Table 1. Because there is a limited number of professionals relevant for the study in each location, and because round 1 participants could not participate in round 2, there were smaller numbers of professionals in round 2 (except in Eureka). For the follow-up online survey, 107 participants (33 professionals and 74 residents) participated, a 75% overall response rate (88% for residents and 56% for professionals). Specific numbers by location for the final survey round include the following: Eureka: 13 professionals and 28 residents; Gunnison: 6 professionals and 6 residents; Durango: 5 professionals and 18 residents; and Owego: 9 professionals and 22 residents.

Focus group participants had varied experience with flooding in all locations and

Table 1. Number of participants in focus groups by location for rounds 1 and 2.

	Professionals		Residents	
	R1	R2	R1	R2
Eureka	11	11	16	12
Gunnison	10	3	6	6
Durango	7	3	11	11
Owego	11	6	12	14

sessions, along with differing perceptions of their flood risk. In Owego, all professionals and residents had flood experience, while in Durango, 57% (R1) and 67% (R2) of professionals had experienced flooding and 55% (R1) and 36% (R2) of residents did. These varying characteristics highlight the diverse factors experienced by different regions as well as the diversity of participants between locations and rounds.

Survey results related to probabilistic product usefulness. Focusing on the usefulness of the probability product by location, Fig. 5 shows the progression of usefulness for the regional HEFS products over all three rounds of testing for each location. For many, but not all locations and user groups, the ratings of very/extremely useful increased from R1 to R2 to R3. Recall that the graphics were revised between each round with the goal of enhancing understandability and usability and R3 was an online survey, different from R1 and R2, which were in-person group settings.

In round 3, the online survey, the participants were also asked about the usefulness of a national HEFS product variation that included the deterministic forecast (Fig. 6); this version was different from the regional product they had seen in the focus groups and surveys, and incorporated the highly rated elements of the regional product designs to inform its development.

A strong majority (over 80%) for each region and user group rated the proposed national HEFS product as very or somewhat useful, with all but Eureka and Colorado residents having over 60% stating it was very useful (Fig. 7a). Most were also very or somewhat likely to use the product (Fig. 7b). Eureka showed the strongest likelihood to use it, while Colorado showed the least, though participants in Colorado tended to report relying mainly on USGS discharge products as we note later.
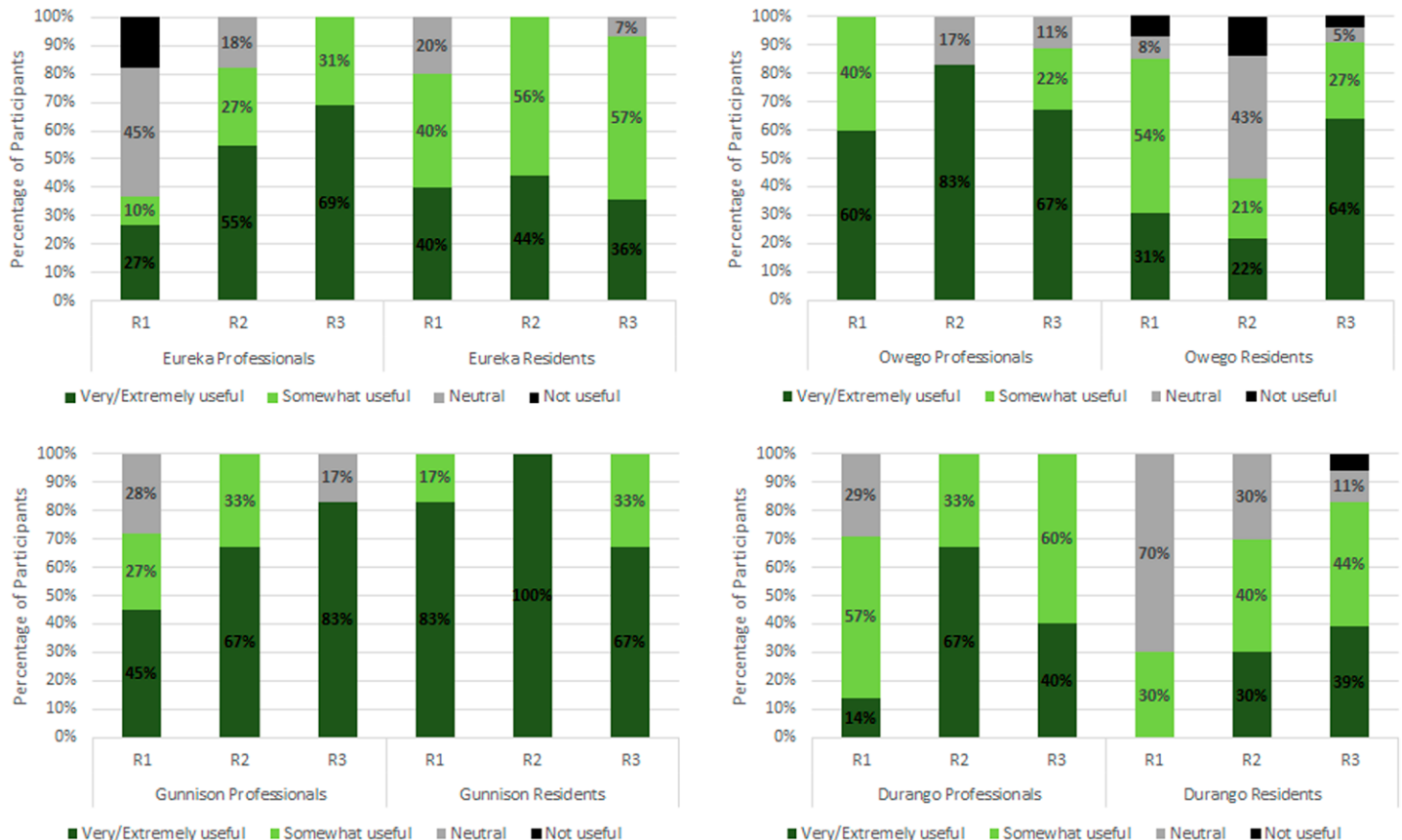

Fig. 5. Percentage of participants reporting usefulness of the probability of river level products in round 1 (R1), round 2 (R2), and round 3 (R3) by focus group location, reflecting changes to the product between rounds.
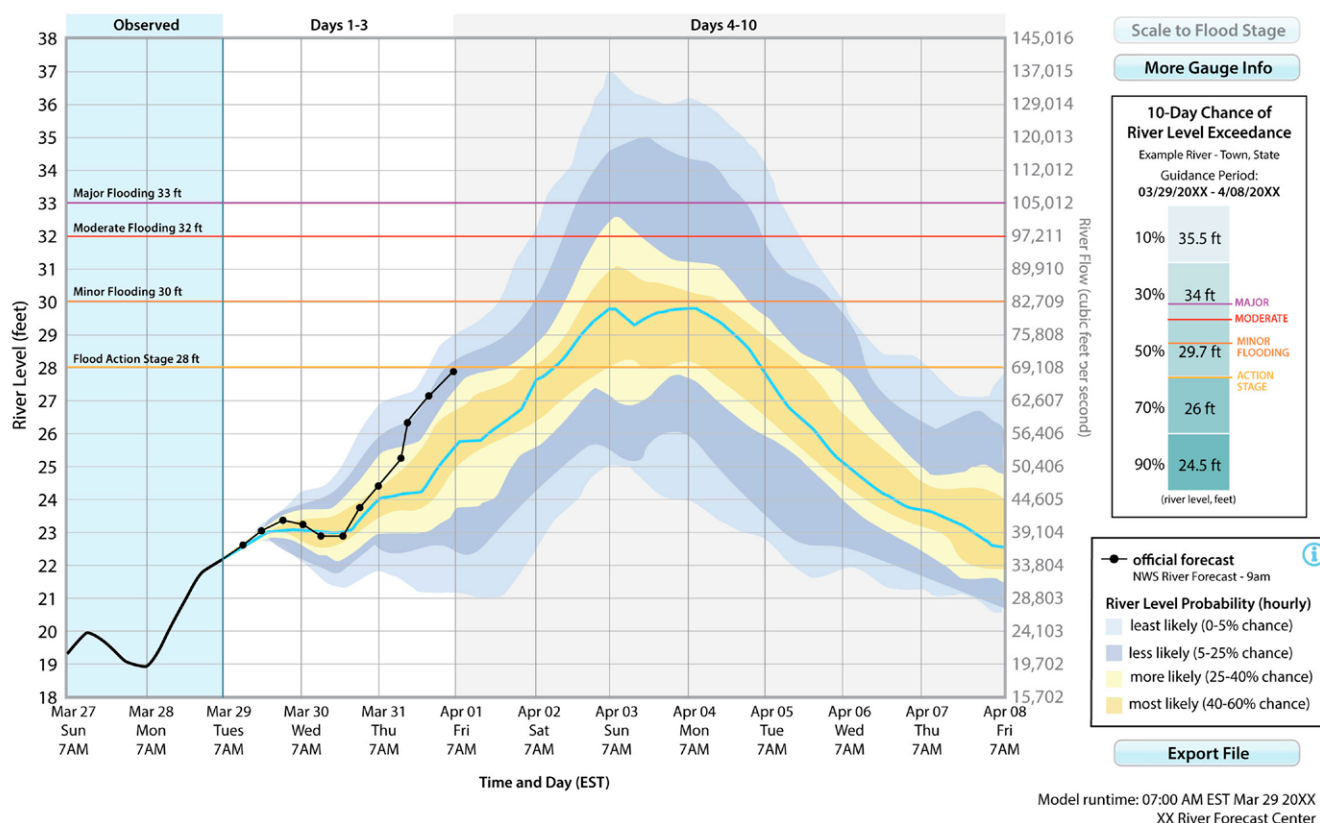
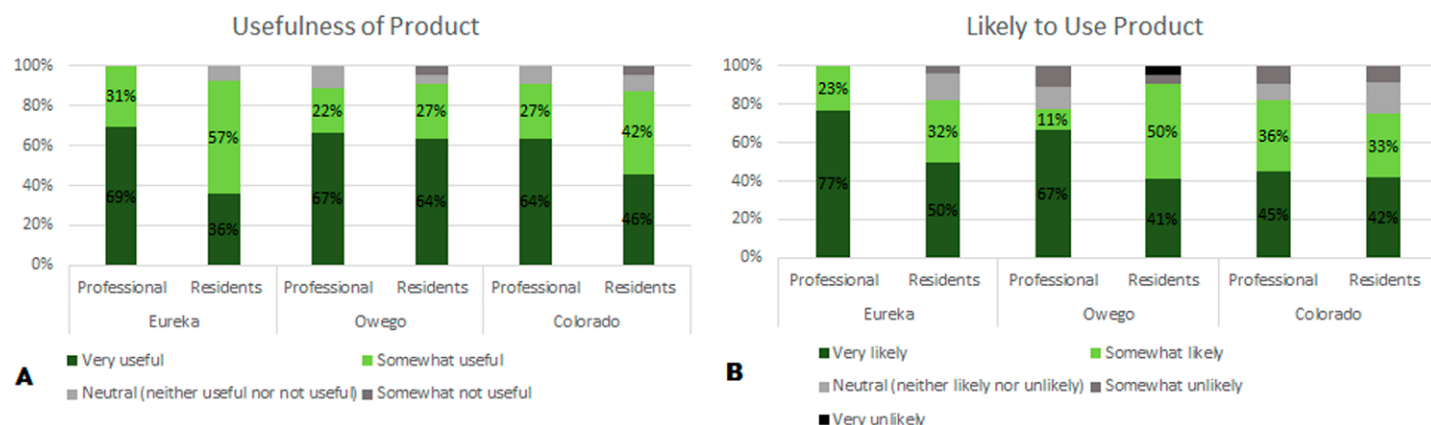Fig. 6. Proposed national HEFS product tested in round 3 online survey.

Fig. 7. Percentage of online survey (round 3) respondents rating the (a) usefulness of the national probabilistic river level product and (b) likelihood to use.

Survey results related to probabilistic and deterministic forecast divergence. A focal research question of this study is the tolerance for divergence between the probabilistic and deterministic river level forecasts. Having gone through a weather scenario and been exposed to both probabilistic and deterministic forecasts, and having seen an example of forecast divergence between these two in each region, participants were asked in the survey how they would react to a divergence, with options including the following: ignoring the forecast, seeking out more information, having less confidence in both or either forecast (probabilistic and deterministic), or asking an expert. Participants could choose more than one response. The highest response among both professionals and residents was to seek more information to better understand the difference between the two forecasts, while asking an expert was another frequent response among professionals (Fig. 8a). Comparing the impact of the divergence from round 1 (different from participants in round 2) to round 3 (included participants from both round 1 and round 2), to see how changes to the product design and elements affected understanding, there were decreases reported across all choices, with the greatest decrease (15%) in professionals having less confidence in the probabilistic product (Fig. 8b).

The revised design of the probabilistic product in round 3 may have increased understanding and confidence in the product resulting in less need to seek out more information or asking experts; yet increased familiarity and learning with the product from participating in the focus groups may also have influenced these results. It is important to note that our sample size is necessarily small (focus groups were capped at 15 participants in order to maximize discussion) and many of the percentage changes are not very high (less than 5% but significant with a p value of 0.05 using a paired t test), and it could be reasoned the revisions did not overwhelmingly influence understanding of the product where there is a divergence between the deterministic and
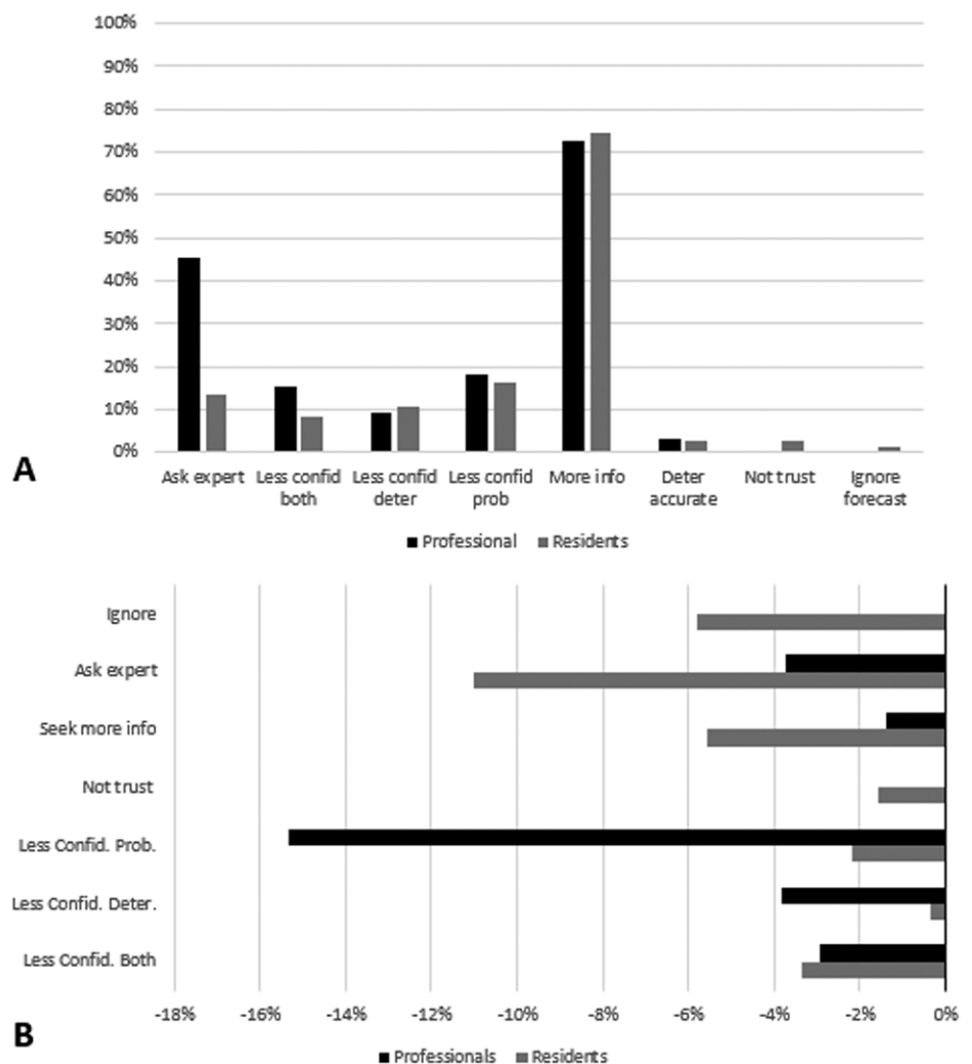

Fig. 8. (a) Round 3 survey responses to actions/reactions when faced with a divergence between a probabilistic and deterministic forecast. **(b)** Difference from round 1 to round 3 in percentage of professionals and residents in response to how a divergence in the probabilistic and deterministic products would impact their action/reactions.

the probabilistic forecasts. More details of the professionals' and residents' perceptions and confidence in the deterministic and probabilistic forecasts when there is a difference are elucidated in the focus group analysis that follows.

Focus group analysis. As described in the "Methodology" section, focus groups were held with professional users and with residential users in each of the four locations. Rather than discuss the results by location, we focus on the overall findings, but Tables 2–5 present location-specific results. Figures 1–4 show the probabilistic hydrologic products shown in round 1 and round 3 in each location. There are some differences in responses to the HEFS among locations, both within and between user groups. For instance, in Durango, professionals expressed great utility in the longer-term product for planning activities while residents needed more information to determine its utility. In Eureka, both professionals and residents appreciated the probabilistic information, though both groups also wanted more information to foster understanding.

Overall focus group results.
Need for validation of the product. Across all regions, both professionals and residents expressed a desire to "ground truth" the accuracy of probabilistic forecasts, and suggested that seeing the past performance of probabilistic forecasts would help ascertain confidence in current probabilistic forecasts. As a Gunnison professional noted, "I would like to see past history of how the probabilistic versus the deterministic forecasts perform, before I would say, I trust one or the other. So really this doesn't, for me doesn't do a whole lot."

When probabilistic and deterministic forecasts diverged, residential and professional participants responded in a variety of ways, based on experience and their level of understanding of product information. Almost always, the HEFS forecast products had to be explained

Table 2. Summary of focus group discussion for Eureka. The first column briefly describes each scenario. The second and third columns show the summaries for rounds 1 and 2.

Eureka	Water resource and emergency management professionals	Residential users
Participants were presented with two separate scenarios, an initial longer scenario which focused on a 5-day lead up to a flooding situation, and a second, briefer 5-day scenario that focused on low flow conditions.	**Round 1** • Expressed that probabilistic information would be important to decision-making • Decreased trust upon finding the difference between probabilistic and deterministic forecasts	**Round 1** • Appreciated receiving probabilistic information • Understanding was generally low
	Round 2 • Response to the revised HEFS was very favorable, more helpful than deterministic forecasts alone • Some needed time to use the product before assessing its usefulness • Asked for a description of the elements that drove the forecast—wondering if the product accounted for dam operations, for instance, or soil saturation • Asked for a spatial map of probabilistic forecasts to monitor a broad region • Value in being able to view two gauge locations at the same time • Important to explain how the deterministic and ensemble forecasts are each calculated, especially if the deterministic and ensemble forecasts diverged	**Round 2** • Interested in using and understanding the HEFS • Not immediately familiar with the data • Required explanation of the deterministic line and its relationship to the ensemble forecast; wanted to understand if historical data and current conditions were used in the models to drive the ensemble forecasts • Big divergence between probabilistic and deterministic would cause less trust in the forecast and require explanation • Requested tutorials, including videos, to describe how the various products should be understood and used

Table 3. Summary of focus group discussion for Owego. The first column briefly describes each scenario. The second and third columns show the summaries for rounds 1 and 2.

Owego	Water resource and emergency management professionals	Residential users
Participants were shown a series of products as part of a 5-day river flood scenario; products included hydrographs, HEFS graphics showing deterministic and probabilistic forecasts on one graph, briefing packages from both the WFO and the Middle Atlantic River Forecast Center (MARFC), quantitative precipitation forecasts, snow products, 5-day river flood outlook, flood watches and warnings, and other products.	**Round 1** • Appreciated the longer time horizon offered by the HEFS • Want to "ground truth" the product during a storm to see how probabilities compared to actual levels or see how probabilities are changing day to day in comparison to river levels	**Round 1** • Mixed in initial response to the HEFS and unfamiliar with probabilistic forecasts • Wanted more information to explain the discrepancy between probabilistic and deterministic forecasts • Some trust in the deterministic because it was perceived to be influenced by forecasters with experience • Some indicated a preference for following the higher forecast when in conflict, to be overprepared rather than underprepared • Wanted HEFS presented as an interactive platform to accommodate less and more experienced users
	Round 2 • Familiar with both the hydrograph and probabilistic forecasts generally • Would use the probabilistic information in planning for emergency response and managing water resources • Needed the probability percentages to be listed alongside any text descriptions like most or least likely • Would be looking for further information when forecasts do not converge, or when facing anomalous information or participate in NWS briefing calls, which allow people to engage in a group conversation and ask questions together	**Round 2** • Welcomed the idea of longer-term probabilistic forecasts • Some questioned the helpfulness of the wide range of probabilities (95%–5%) • Would use the product for situational awareness and would check it daily • Probabilistic information in the graphic made the event feel more intimidating, and that there was a "more imminent danger" than the deterministic alone • Others found the amount of information to be off-putting • Would be helpful to have it paired with a rain forecast to clarify when to prepare • When probabilistic and deterministic forecasts diverged, some indicated they would follow the deterministic forecast • Steadily increasing probability of flooding from previous days would encourage them to be more cautious and to prepare

to participants, with the difference between the black deterministic forecast line and the probabilities described in detail. With repeated use, participants' sense of the utility of the product generally increased.

Professionals expressed an expectation that the deterministic and probabilistic would converge and struggled to understand why the deterministic forecast and the 50% line were not in agreement: "Why on this graph does the deterministic exceed the most likely? Is it because of some setting you just change?" Another said, "I can't quite figure that. It doesn't make sense to me that the black line doesn't jive with, all the, all the fancy colors that are in there." One experienced professional found the forecast very confusing when the deterministic forecast did not align with the median: "Now when I'm looking at this, I'm like, everything I knew was wrong. I don't know what they're doing."

TRUST AND CONFIDENCE IN THE PRODUCT. Some suggested that large divergences can cause them to believe the forecast is incorrect, as said one Eureka resident when probabilities diverged from the forecast: "So like, they say that it [river level] has a low percent chance of being true, why should I believe the official forecast?"

If previous experience suggests that the probabilities are reliable, people will favor those instead of the forecast. When one resident participant noted that the deterministic seemed to

Table 4. Summary of focus group discussion for Gunnison. The first column briefly describes each scenario. The second and third columns show the summaries for rounds 1 and 2.

Gunnison	Water resource and emergency management professionals	Residential users
Participants were presented with a month-long scenario with high snowpack and possible flooding along the East River near Almont.	**Round 1** • Expressed value in longer-term deterministic forecasts, longer than 5 days as shown • Expressed value in receiving probabilistic forecasts as far as a month out—helpful for monitoring and planning water-based activities, such as rafting and festivals • A forecast that diverges from the probabilistic median should be called out on the product and addressed • Needed clarity on the data included in the forecasts, such as historical data	**Round 1** • Valued the HEFS in conjunction with weather forecasts including temperature • Favored showing historical data (such as the 30-yr average), and probabilistic and deterministic forecasts all in one product
	Round 2 • Wanted to know the past performance of the model to assess reliability • Some said a divergence in forecasts caused a lack of trust • Suggested a different sort of risk product that showed a maximum potential along with the deterministic might be less confusing • Would refer to probabilistic to understand the probabilities of various scenarios • Would use the deterministic line as primary go-to for decision-making • Would call the RFC/WFO when the ensemble diverged from deterministic • Favored interactive product showing historical and forecast data together	**Round 2** • Response to the probabilistic forecast was mixed • When the forecasts diverged, participants were confused and sought understanding and more information • Watching the progression of ensemble forecasts over a period of days was helpful in building trust in the product • Participants indicated they would rely more on probabilities

not match the actual impacts occurring in the scenario, he said, "The prediction hasn't been so good. That's what I'm saying. But the probabilities have been really good. They have been very consistent with reality so I'm feeling really good about it."

When given information about the reason for the divergence, users, especially professionals, expressed increased confidence in the products. As noted by one water resource manager, "So you, you made the point that the deterministic is different than the probabilistic, and that the, the black line is not following the average of the probabilistic, so you're actually in two different methods here to calculate those two things. And so it would be useful to just have a button or something that describes how each was calculated. And so that you understand that there actually, the black line is not the average of the probabilistic, it's a different method." Another professional echoed: "So the middle of the yellow zone is fifty percent probability. So according to fifty percent probability, you're not even gonna hit your monitor stage or you're nowhere close, but your deterministic is saying that you're going to be in the flood stage.... Why is that difference there? What is that indicating? So you need to understand the methods behind it."

NEED FOR EXPLANATION. Users were nearly unanimous in suggesting that forecasters should explain clearly the reason for divergence between probabilistic and deterministic forecasts. Residents anticipated problems for public response when the forecasts diverge, with one resident suggesting that people "are going to freak out" if the forecasts diverge, and that the divergence must be explained to avoid panic. One professional suggested that forecasters could

Table 5. Summary of focus group discussion for Durango. The first column briefly describes each scenario. The second and third columns show the summaries for rounds 1 and 2.

Durango	Water resource and emergency management professionals	Residential users
Participants were presented with a 2.5-month low-flow scenario, which included the HEFS and hydrograph, snowpack products from NWS and NRCS, modeled snow conditions, departure from normal temperature, soil moisture, river flow compared to normal, USDA drought monitor and drought products, weather forecast summaries and temperature and precipitation outlooks, hurricane cone, quantitative precipitation forecasts, flood warning and watches, and impact statements.	**Round 1** • Familiar with hydrograph—trust its capacity to handle snowmelt more than rainfall • Utility in the probabilistic forecast, because it provides a fuller range of information than hydrograph • Would call the RFC or WFO if probabilistic and deterministic forecasts diverge **Round 2** • Asked for historical river levels, modeled ensemble forecasts, and deterministic forecast information in one product • Would use probabilistic products for planning, but rely on on-the-ground observations during acute events • Probabilities helpful for focusing them on when to pay attention and as a tool for being vigilant in monitoring weather, but even high probabilities were taken with some doubt	**Round 1** • Almost no experience with probabilistic forecasts • Required more information about the basis of the forecast to determine whether it was useful or not **Round 2** • Found the revised HEFS graphic readable and were mixed on its utility • Lack of experience with the product • Needed to use over time to determine accuracy • Need for an "action stage" for drought • Hydrograph or HEFS more helpful in context with the other products in the scenario instead of a standalone product • Graphics should show progress over time, to let users know how the situation was changing

"flag" forecast information that does not immediately appear sensible, such as a probabilistic forecast that diverges from deterministic data: "I mean if there is something that is suspect on a gauge or in a forecast, then it would be good for it to be flagged as yes, the model is showing this, but for whatever reasons, you know, so that the information is suspect, you know." Another professional said, "I think uncertainty is good to display and communicate … better than not knowing that cause they're being hesitant and pulling back because I don't want to freak people out but I'd prefer to know about the uncertainty in the forecasts." Still another suggested that without explanation of divergence, it would be hard to make decisions: "Yeah, I think the first question would be, why is the black line so far out of all the probabilities? And if that's not explained on the page anywhere … I don't know what to do."

Without explanation for a reason for divergence, participants—as with other decisions about flooding—sometimes relied on personal forecast experience to decide which of the divergent forecasts to follow. One resident, when faced with a divergence, said, "The deterministic doesn't make sense to me knowing our river because that's, I don't think that's when we'd have peak flow so … I would throw that one out and look at the trend on the other line as being more realistic."

SEEKING MORE INFORMATION. Both professional and residential participants across all regions acknowledged overwhelmingly that when faced with a discrepancy, they would seek more information. One resident said, "This would make me come back and check it three days later and see what the trend's doing." Participants suggested adding a phone number for the issuing office directly on the product so users can call the forecasters to ask questions, "'Cause that's what I'm gonna do," said one professional participant. Another professional said he would start looking to other products and forecasts to decide how to proceed: "This would make me want to look for other information. So I look at the um, uh, forecast models for precipitation and really just start following those and seeing are those lining up. And then also really talking with…. Like I would

talk with the forecaster at National Weather Service and just try to get more information. You get an idea of their confidence." Another echoed that without explanation, the divergent forecasts were not viewed as trustworthy: "I'm going to be skeptical. I'm going to look at it ... I'm not going to do much until I talked to somebody who's in this equation. What's this forecast process?"

PREFERENCES FOR DETERMINISTIC OR PROBABILISTIC FORECASTS. Whether people favored the probabilistic or deterministic when they diverged depended on several factors, with some relying on probabilities because they afford more information or because they felt like models with "all of the different data points" would be more trustworthy than the "human aspect that goes into the [deterministic] forecast line," as noted by one resident. Perceiving probabilistic data to be more reliable than deterministic forecast data that are influenced by a forecaster, one resident said, "I would trust that a little bit more than the forecast." Others suggested a growing preference for probabilistic as their familiarity with them grew during the scenario, as expressed by one resident who said, "The probability is always nice now, since like now that they're on there, we've seen them, I don't really trust the black line." A professional with experience with forecasters said they would favor the deterministic, knowing that forecasters are behind the product and "they have picked a black line." Other professionals suggested they would simply refer to the deterministic forecast on a daily basis, to reduce complication in the face of so much data: "I mean, you could run a million scenarios in a model, but at some point that becomes somewhat useless." Some participants expressed a tendency to prepare for the higher forecast if flooding is a risk, to be ready, as noted by one Owego professional: "So I'm not gonna go with the forecast," said one. "I'm gonna go with higher than the forecast."

Some users favored the longer time horizon the probabilistic models provide. One professional noted that the deterministic forecast already contains uncertainty, but that the probabilistic model "makes it more easy to find, I guess." Despite this acknowledgment, users reported that a very large divergence can cause a loss of faith in the models running the forecast, as expressed by one Gunnison professional: "It's almost like why are we using these models to do this forecasting, if then the deterministic forecast is so much different than the models? Because typically you would, you're using the models to help you make the forecast, so your deterministic might be a little bit different than the models. But if your model is that bad, why are you?"

As such, professionals indicated they would be hesitant to share such an example with the public because it would be hard to explain, noting that the "kind of people that call" them to discuss weather would see a divergence and say "well that's why you shouldn't use those damn models." Presenting explanations of the reason for the discrepancy was an important request to counter this concern.

Even as they struggled to understand the occasion of divergence between probabilistic and deterministic forecasts, participants overall welcomed the presence of probabilistic forecasts alongside deterministic and sought information to help make them usable and more understandable. As noted above, users requested clear explanation of any divergence, with one professional summarizing the comments of many across the focus group sessions: "Just the explanation behind why there's that much uncertainty would be really helpful."

Other considerations. Different geographic regions had varying levels of familiarity with probabilistic information and hydrologic forecasts in general. In New York, for example, users were much more familiar with and trusting of the hydrograph's deterministic forecast, which appeared to make interpretation of probabilistic data around the forecast easier for new users. In contrast, participants in Colorado tended to rely on real-time and historical data products rather than forecasts and initially required more time to assess and understand the probabilistic forecasts. Getting users to adopt HEFS products, therefore, may require different

approaches by region. In places where the hydrograph is familiar and trusted by public and professional users, the probabilities could be easily absorbed into the data flows people use for personal and professional use. In places where the hydrograph is less well known and trusted because of geographic factors, the probabilities could be presented to new users as helpful tools to decipher the range of possible outcomes in a region where people already expect uncertainty.

National product. Based on the findings from the round 3 survey, we propose that a prototype (Fig. 6) as tested in that survey will meet the needs of both residential and professional users related to probabilistic and deterministic forecasts. This graphic was developed with the most favored elements of the products tested in each of the three regions, and with strategies designed to improve user understanding. Recognizing that a one-size-fits-all approach is impossible, this product nonetheless aims to provide maximum utility to the broadest set of users—combining, for instance, preferences for discharge and river level in one product as well as numerical representations of probability along with "likely" categories, and adding a vertical side bar graph showing river level exceedances to help interpret the HEFS graph. A forecaster's note, which ranked high in all survey data, is prominent, and formatting is designed to be standard and easy-to-use across all regions.

This product includes a proposed method for sharing historic river level data gathered from the USGS, which serves to situate a given day's forecast river level into context of the low, average, and high data for that gauge compared to the previous 30 years. However, the HEFS platform could also develop a complementary product that shares the probabilistic forecast levels in conjunction with a more complete display of historic river levels for the same period (with the historical levels demarcated by patterns as an overlay to the colorized probabilities).

While a national product will provide a standard for delivery of information across the country, feedback from focus groups in each region nonetheless revealed a need for region-specific information, delivered in formats that are comfortable and familiar to users in the area. During focus group conversations, participants relayed stories revealing that each community has a unique and important culture of cooperation with local NWS offices and established patterns of communication that have developed over many years. As such, regional offices working to communicate probabilistic information may require specific modifications to probabilistic data products to meet regional needs.

Discussion

The findings of this project align with those of others who studied different hazards in different locations (Fundel et al. 2019; Zabini et al. 2015; Ash et al. 2014; Morss et al. 2008, 2010) and who reported that nonexperts anticipate uncertainty in deterministic forecasts. Indeed, users across all focus groups in this study anticipated that deterministic hydrographs contain some inherent uncertainty. Despite this recognition, as noted earlier, participants reported the serious potential for lost faith in forecasts when probabilistic and deterministic forecasts diverge. However, participants also overwhelmingly indicate that they will look for more information to explain the divergence in order to understand and decide how to take action. As such, forecasters need to consider methods for directly identifying and explaining meaningful divergences within product dissemination to avoid confusion.

Recommendations from this study present a range of product design considerations to address this need for explanation, including redesign of legends, inclusion of a forecasters' note for explicit communication about forecast information, and changes to the presentation of data, including color and format. Recommendations also include suggestions to share the

inputs and drivers used to generate forecast data to help explain why deterministic forecasts may not always align with the median of probabilistic forecasts. Recommendations also include a national prototype that can be used in conjunction with regional products that provide context for specific users.

Proposed design changes to the HEFS outputs through this study, including the use of legends that contain categories as well as numeric quantifications of probabilities, reflect findings from a number of studies indicating that including numeric probabilities with forecasts increases both trust in the forecast and the quality of decisions based on it (Grounds and Joslyn 2018; Joslyn and Grounds 2015; Joslyn and LeClerc 2013), though these researchers found that the positive impacts of probabilistic forecasts were not found equally across groups they studied. Among the recommended changes made to the HEFS as a result of the focus groups and surveys is the use of both verbal and numeric descriptions of probability. This is in keeping with the recognition that people have different levels of numeracy (Grounds and Joslyn 2018) and, at the same time, there can be large differences in how people interpret verbal descriptions (Budescu et al. 2012).

While, as we found, one size may not fit all, other research has suggested it may not be necessary to create different products for different user groups. As Grounds and Joslyn (2018, p. 31) note, "A forecast that includes a numeric estimate along with explicit advice may be best for a wide range of users." This finding comports with the recommendation herein to include a forecaster's note as part of the HEFS graphical presentation, which allows for narrative explanation of critical data and impacts that may emerge from the forecast. And, while one-size-fits-all may be difficult to achieve for the HEFS, our research findings suggest that "more is more" when it comes to information describing the use of HEFS data and description of how forecasts are generated.

This study also demonstrated that users need to build experience with ensemble forecasts to determine their utility and to build their confidence in the products. Sharing the past performance of probabilistic forecasts can help new users build confidence, through visual and/or narrative explanations. Uncertainty causes people to seek additional information to confirm a forecast and to consider actions they should take. Having access to additional information—such as easily identified links to active watches and warnings—along with any uncertainty information will help the user find supporting forecast details to inform their decision-making. Further, linking precipitation forecasts to hydrologic forecasts when possible will help users quickly assess the situation and understand their confidence in the forecasts. This can also be done by building interactivity between and across products to direct users to relevant information. New and nonprofessional users of probabilistic forecast information will often be unaware of how to find additional information and this interactivity can strengthen user understanding of probabilistic forecasts.

Finally, this study reinforces that user testing of product design is critical for ensuring that forecast data are correctly interpreted. Users provided discrete and helpful suggestions about the use of color, specific language and word choice in legends, and placement of information to remove barriers to understanding. As forecast data become more complex with joint presentation of probabilistic and deterministic data, product design becomes ever-more important to ensure that the visual presentation does not add needless complication. The prototype products as recommended here contain the benefit of that user testing.

Limitations

This study focused on participants' understanding of forecast products, and perceptions in a real-life context cannot be elucidated. Participants' ratings of product usefulness may be influenced by the exchange and discussion during the focus group and stated usefulness may not reflect preferences in real-life conditions. Repeated exposure to products may also

increase ratings of utility and intent to use, supporting the notion that lack of understanding is a barrier to use of the product.

This study points to the fact that increased understanding shows increased tolerance for divergence. However, this is one study that provides incremental progress on understanding and addressing confusion related to showing these products. Widespread and continued use of the product and more research will provide additional insight into the research questions explored here.

Conclusions and future research

This study found that probabilistic forecasts introduce a tremendous amount of new, and valuable, information into a weather enterprise that already offers much data and many products. Users can be quickly overwhelmed by information and not know how to sort and prioritize. Conversely, lay users may be unaware of valuable resources that are available, and if they are aware, may not know how to find them if located on a website that contains a lot of information. Beyond the issue of divergence in deterministic and probabilistic forecasts, this study also looked at the ability of briefings to convey probabilistic information and recommends them as another option for forecasters looking to explain forecast complications. Future research should explore presentation and dissemination strategies to help NWS design websites, social media, and other mechanisms to 1) alert users to the availability of probabilistic information, 2) help them locate it easily, and 3) direct them to "self-briefing" interactive platforms that would let users set up their own customized data pages.

Respondents also indicated that time spent "ground truthing" the products would be helpful in determining their confidence in the products. There are multiple methods by which a probabilistic forecast could indicate past performance, for instance, showing the previous days of probabilities along with the observed information, to illustrate how closely the probabilities matched actual outcomes, or to show results over a longer period of time, including seasonal results. But such approaches may create new confusion for users and should be studied to identify the best ways to share this information clearly. While this study did test multiple approaches for presenting deterministic and probabilistic river level forecasts together, which was used to develop the final product, there is a need and potential for more research to further explore alternative best practices for presenting deterministic and probabilistic river level forecasts, and when users do or do not tolerate divergences.

Data availability statement. Data (transcripts and raw survey data), documentation, and methods used to support this study are available upon request to rhogan@nurturenature.org at Nurture Nature Center. A report of the full social science research study is available at https://focusonfloods.org/social-science/reports-and-findings/.

References

Arnal, L., L. Anspoks, S. Manson, J. Neumann, T. Norton, E. Stephens, L. Wolfenden, and H. L. Cloke, 2020: "Are we talking just a bit of water out of bank? Or is it Armageddon?" Front line perspectives on transitioning to probabilistic fluvial flood forecasts in England. *Geosci. Commun.*, **3**, 203–232, https://doi.org/10.5194/gc-3-203-2020.

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

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

Budescu, D. V., H. Por, and S. B. Broomell, 2012: Effective communication of uncertainty in the IPCC reports. *Climatic Change*, **113**, 181–200, https://doi.org/10.1007/s10584-011-0330-3.

Budimir, M., and Coauthors, 2020: Communicating complex forecasts: An analysis of the approach in Nepal's flood early warning system. *Geosci. Commun.*, **3**, 49–70, https://doi.org/10.5194/gc-3-49-2020.

Fleischhut, N., S. M. Herzog, and R. Hertwig, 2020: Weather literacy in times of climate change. *Wea. Climate Soc.*, **12**, 435–452, https://doi.org/10.1175/WCAS-D-19-0043.1.

Fundel, V. J., N. Fleischhut, S. M. Herzog, M. Göber, and R. Hagedorn, 2019: Promoting the use of probabilistic weather forecasts through a dialogue between scientists, developers, and end-users. *Quart. J. Roy. Meteor. Soc.*, **145**, 210–231, https://doi.org/10.1002/qj.3482.

Grounds, M. A., and S. L. Joslyn, 2018: Communicating weather forecast uncertainty: Do individual differences matter? *J. Exp. Psychol. Appl.*, **24**, 18–33, https://doi.org/10.1037/xap0000165.

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.

Hogan Carr, R., B. E. Montz, K. Maxfield, S. Hoekstra, K. Semmens, and E. Goldman, 2016: Effectively communicating risk and uncertainty to the public: Assessing the National Weather Service's flood forecast and warning tools. *Bull. Amer. Meteor. Soc.*, **97**, 1649–1665, https://doi.org/10.1175/BAMS-D-14-00248.1.

——, ——, K. Semmens, K. Maxfield, and S. Connolly, 2018: Major risks, uncertain outcomes: Making ensemble forecasts work for multiple audiences. *Wea. Forecasting*, **33**, 1359–1373, https://doi.org/10.1175/WAF-D-18-0018.1.

Hsieh, H.-F., and S. E. Shannon, 2005: Three approaches to qualitative content analysis. *Qual. Health Res.*, **15**, 1277–1288, https://doi.org/10.1177/1049732305276687.

Hullman, J., X. Qiao, M. Correll, A. Kale, and M. Kay, 2018: In pursuit of error: A survey of uncertainty visualization evaluation. *IEEE Trans. Visualization Comput. Graphics*, **25**, 903–913, https://doi.org/10.1109/TVCG.2018.2864889.

Joslyn, S., and S. Savelli, 2010: Communicating forecast uncertainty: Public perception of weather forecast uncertainty. *Meteor. Appl.*, **17**, 180–195, https://doi.org/10.1002/met.190.

——, and J. LeClerc, 2013: Decisions with uncertainty: The glass half full. *Curr. Dir. Psychol. Sci.*, **22**, 308–315, https://doi.org/10.1177/0963721413481473.

——, and M. A. Grounds, 2015: The use of uncertainty forecasts in complex decision tasks and various weather conditions. *J. Exp. Psychol. Appl.*, **21**, 407–417, https://doi.org/10.1037/xap0000064.

Kinkeldey, C., A. M. Maceachren, and J. Schiewe, 2014: How to assess visual communication of uncertainty? A systematic review of geospatial uncertainty visualization user studies. *Cartogr. J.*, **51**, 372–386, https://doi.org/10.1179/1743277414Y.0000000099.

Kox, T., L. Gerhold, and U. Ulbrich, 2015: Perception and use of uncertainty in severe weather warnings by emergency services in Germany. *Atmos. Res.*, **158–159**, 292–301, https://doi.org/10.1016/j.atmosres.2014.02.024.

Krippendorf, K., 2018: *Content Analysis: An Introduction to Its Methodology*. 4th ed. Sage Publications, 472 pp.

Michaels, S., 2015: Probabilistic forecasting and the reshaping of flood risk management. *J. Nat. Resour. Policy Res.*, **7**, 41–51, https://doi.org/10.1080/19390459.2014.970800.

Morss, R. E., J. L. Demuth, and J. K. Lazo, 2008: Communicating uncertainty in weather forecasts: A survey of the U.S. public. *Wea. Forecasting*, **23**, 974–991, https://doi.org/10.1175/2008WAF2007088.1.

——, J. K. Lazo, and J. Demuth, 2010: Examining the use of weather forecasts in decision scenarios: Results from a US survey with implications for uncertainty communication. *Meteor. Appl.*, **17**, 149–162, https://doi.org/10.1002/met.196.

Nobert, S., D. Demeritt, and H. Cloke, 2010: Informing operational flood management with ensemble predictions: Lessons from Sweden. *J. Flood Risk Manage.*, **3**, 72–79, https://doi.org/10.1111/j.1753-318X.2009.01056.x.

Palmer, T. N., 2002: The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. *Quart. J. Roy. Meteor. Soc.*, **128**, 747–774, https://doi.org/10.1256/0035900021643593.

Ramos, M.-H., J. Bartholmes, and J. Thielen, 2007: Development of decision support products based on ensemble forecasts in the European flood alert system. *Atmos. Sci. Lett.*, **8**, 113–119, https://doi.org/10.1002/asl.161.

——, T. Mathevet, J. Thielen, and F. Pappenberger, 2010: Communicating uncertainty in hydrometeorological forecasts: Mission impossible? *Meteor. Appl.*, **17**, 223–235, https://doi.org/10.1002/met.202.

Severtson, D. J., and J. D. Myers, 2013: The influence of uncertain map features on risk beliefs and perceived ambiguity for maps of modeled cancer risk from air pollution. *Risk Anal.*, **33**, 818–837, https://doi.org/10.1111/j.1539-6924.2012.01893.x.

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

U.S. National Research Council, 2006: *Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts*. National Academies Press, 124 pp.

Wood, M. M., D. S. Mileti, M. Kano, M. M. Kelley, R. Regan, and L. B. Bourque, 2012: Communicating actionable risk for terrorism and other hazards. *Risk Anal.*, **32**, 601–615, https://doi.org/10.1111/j.1539-6924.2011.01645.x.

Zabini, F., V. Grasso, R. Magno, F. Meneguzzo, and B. Gozzini, 2015: Communication and interpretation of regional weather forecasts: A survey of the Italian public. *Meteor. Appl.*, **22**, 495–504, https://doi.org/10.1002/met.1480.