5.5 COMMUNICATING THREAT LEVEL FOR FLASH FLOODS IN THE HAZARDOUS WEATHER TESTBED HYDROLOGY EXPERIMENT 2014

Elizabeth M. Argyle[1,2,3], Jonathan J. Gourley[2], Robert A. Clark, III[1,2], Zachary L. Flamig[1,2], Maria M. Gutierrez[1], Jessica M. Erlingis[2], Steven M. Martinaitis[1,2], Brandon R. Smith[1,2], and Chen Ling[4]

[1] Cooperative Institute for Mesoscale Meteorological Studies, The University of Oklahoma
[2] NOAA/National Severe Storms Laboratory
[3] School of Industrial & Systems Engineering, The University of Oklahoma
[4] Department of Mechanical Engineering, Akron University

1. INTRODUCTION

In July 2014, the inaugural Hazardous Weather Testbed Hydrology experiment (HWT-Hydro) sought to evaluate a suite of hydrologic flash flood forecasting models while gathering knowledge about forecaster decision-making processes. The suite of 30+ products, collectively known as MRMS-FLASH tools (Multi-Radar/Multi-Sensor, and Flooded Locations and Simulated Hydrographs, respectively) were used by forecasters to issue experimental watch and warning polygons throughout each of the four weeks during the experiment. In addition, HWT-Hydro occurred in coordination another experiment hosted by the Weather Prediction Center, the second Flash Flooding and Intense Rainfall Experiment (FFaIR) (Barthold et al., 2015).

The testbed experiment occurred in four cycles, each one week in duration. Forecasters participated for one week only, so each week involved a unique set of participants. Upon arrival, participants received training on the use of the AWIPS-II weather forecasting display platform, the MRMS-FLASH tools, and the expected outcomes from the experiment. The majority of the week was spent in real-time experimental forecasting operations. Each participant worked at an individual workstation, but usually partnered with a participant at a workstation near them in order to forecast over a shared geographic region. Due to the nature of the evaluation, participants were encouraged to rely primarily on the experimental tools, but they were allowed to consult external guidance tools online if the tools were not available on the testbed workstations.

Corresponding author's address: Elizabeth M. Argyle, Cooperative Institute for Mesoscale Meteorological Studies, The University of Oklahoma; e-mail: emargyle@ou.edu

During the experimental operations, forecasters issued experimental watches and warnings across the continental United States. An example of one of the experimentally issued warnings is shown in Figure 1. Participants in the WPC's FFaIR experiment provided guidance to the HWT-Hydro forecasters in the form of a webinar at the beginning of each day.

Evaluation was addressed in a two-fold approach during HWT-Hydro: evaluations focused on (1) tool and forecast performance as well as (2) aspects of the forecaster decision-making process. Tool and forecast performance were evaluated in a subjective manner; each day, participants completed a survey in which they evaluated flash flood events from the prior day's forecasts. The evaluation implemented a survey that assessed how well the experimental tools predicted the actual threat, as represented by flash flood reports and other observations. In addition, the survey also included questions to analyze how well the experimental watches and warnings performed in comparison to operational watches and warnings.

Throughout the week, participants also took part in a human factors-based, mixed methods analysis of warning decision-making behavior. During forecasting operations, participants used desktop recording software to audio- and video-record their forecasting activities; the recordings were used for a time-based analysis of tool usage during the watch/warning issuance timeline. At the end of each week, participants took part in a focus group in which they gave feedback on the tools, discussed challenges in flash flood forecasting, and provided information about how experimental uncertainty attributes allowed them to communicate threat levels in their forecasts.

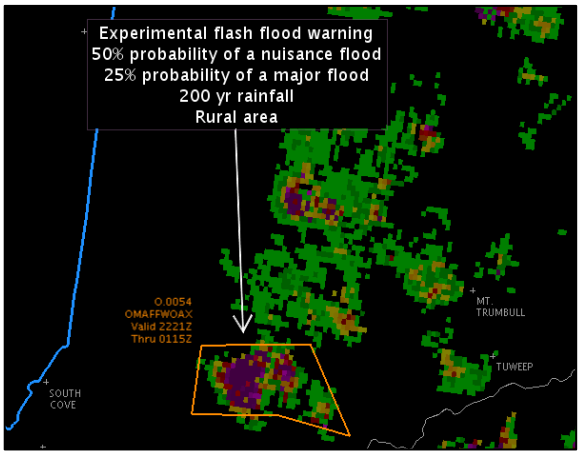

Figure 1. Experimental flash flood warning issued during HWT-Hydro 2014 alongside the CREST maximum return period tool.

2. METHOD

2.1 *Experimental Design Variables*

The HWT-Hydro experiment added a unique aspect to the experimental forecasting operations process; for each experimental watch and warning polygon, forecasters were asked to assign uncertainty attributes. A forecaster was presented with two categories of magnitude (major versus nuisance flooding) and was asked to assign a probability of occurrence of each magnitude level (Figure 2). Probability of each threat could be assigned at one of five thresholds (0%, 25%, 50%, 75%, and 100%). For example, a forecaster could have issued a warning polygon with a 75% probability of nuisance flooding and a 50% chance of major flooding.

2.2 *Participants*

Fifteen participants took part in the focus groups. Participants were all National Weather Service forecasters, with either primary job roles in hydrologic or meteorological forecasting. They primarily worked at Weather Forecast Offices (n = 13), though a few were based out of River Forecast Centers (n = 2). Forecasters were selected from offices and centers around the continental United States, so a wide variety of geographic regions were represented in the sample.

2.3 *Focus Group Design*

The focus group addressed a range of topics related to the general forecasting process as well as the participants' views on uncertainty, probability, and confidence in flash flood forecasting. The questions of interest to the present study were those that sought to elicit feedback on the role of the uncertainty attributes in communicating threat information to end users. During the group discussions, the questions were posed as:

> [1] How did issuing attributes of severity for watches and warnings (nuisance versus major) enable you to communicate threat information? What was helpful? What would you change about the categorization?
>
> [2] How did participating affect how you view probabilities in flash flood forecasting? What factors affected your decisions when assigning probabilistic levels?

Focus groups were audio-recorded and then transcribed. Transcripts were then analyzed using thematic analysis to collect emerging themes from the response sets.

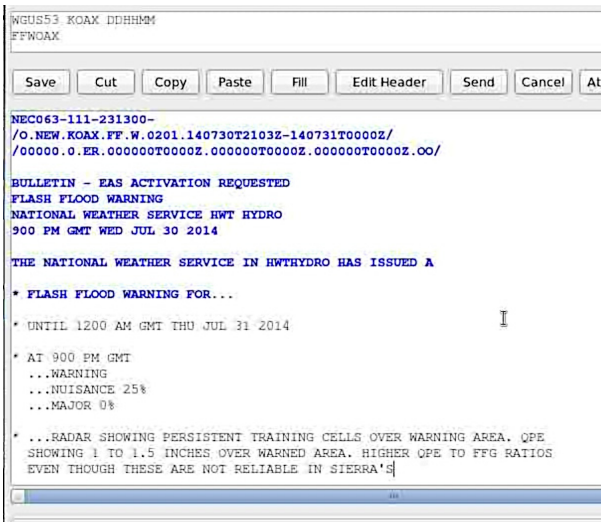

Figure 2. Example of experimental product text with uncertainty attributes.

3. THEMATIC ANALYSIS

Transcripts of the focus groups were analyzed using the thematic analysis framework (Boyatzis, 1998; Braun & Clarke, 2006). Thematic analysis is an iterative analysis method that uses several

cycles of qualitative coding to identify themes that occur within a set of data. In the present study, initial codes were given to responses in the focus group discussions, which were then sorted into broader themes.

3.1 *Findings on Magnitude Attributes*

Several themes emerged from the focus group responses regarding the inclusion of a magnitude estimate. Overall, the experimental requirement to include a magnitude estimate was seen as a positive addition to forecast products. Participants generally expressed a desire to have the ability to issue products with standardized text reflecting threat level in their operational office settings.

Some forecasters discussed their wishes to be able to communicate their mental model to forecast consumers. Including an impact-based uncertainty statement was viewed as a means to such an end. In regard to including the magnitude and uncertainty attributes in the experimental products, one forecaster stated:

> "We kind of do that in our head. I think that's very valuable information for the public, and having this nuisance or major, we're in effect giving them that information that they would have never gotten before."

A theme related to data-driven decision support emerged from some of the discussions; such an addition would allow forecasters to provide data-driven support to end users. Including a magnitude estimate in a watch or warning was seen as a value-adding attribute that would help to provide actionable information that would help consumers like emergency managers to make informed decisions. In the words of one participant:

> "It gives you the ability to quantify the anecdotal information. If we're doing a decision support service brief to emergency managers, you know, on that phone call, we'll say… 'this will be a widespread, minor flood event, or… it's not going to happen everywhere, but if it does, it's going to be really bad.'"

While participants generally adopted a positive affect towards the magnitude uncertainty attributes, they did have some concerns about their design. Themes related to professional interpretation challenges, concern for members of the general public, and training issues emerged from the discussions.

Some participants expressed concern that members of the public would have trouble interpreting both the probabilistic and magnitude components of the threat attributes. Furthermore, participants repeatedly commented that they would expect to see disagreement at a professional level regarding the interpreting of a nuisance versus a major flood. The categorization was seen as subjective; a commonly heard comment was that what may seem like a nuisance flood from a forecasting perspective may feel like a major impact to an individual affected by it. As put by one forecaster,

> "If I get a foot of water in my basement and I'm the only one in 500 miles that did… that's a nuisance, but to me that's major."

Additional concerns tended to revolve around the lack of background experience in issuing magnitude uncertainty attributes. Although some participants stated that they regularly considered threat levels and uncertainty when issuing forecasts, comments from other participants revealed that issuing the experimental attributes created a substantial challenge for some. This may be due to a lack of probabilistic flash flood forecasting in operations and only a short training session on issuing products with the experimental attributes prior to the testbed.

3.2 *Findings on Probabilistic Information*

When asked specifically about the role of probabilities and factors that influence them in flash flood forecasting, positive-affect themes included mental model building, decision support services, and improved forecaster behavior. Almost as a whole, participants commented that they often considered probabilistic information during operational forecasting. While flash flood forecasting is not currently issued probabilistically, some participants suggested that they regularly consider the probability of a threat when before deciding to issue a watch or warning. This is in line with the National Weather Service's Directive 10-922, which creates thresholds for uncertainty that a forecaster must reach before issuing a watch or warning (National Weather Service,

2011). The directive, which requires that there must be a 50-80% chance of flash flooding before issuing a flash flood watch, among other requirements before issuing a flash flood warning, may have led to some bias in the experimental watch and warning products. When asked to give an example of how a forecaster considered probabilistic information in forecasting, one participant responded:

"In issuing a product, [I] will always consider probabilities, because innately in the directive… you must have an eighty percent confidence for something in a warning, or a fifty percent confidence in it happening for a watch. So that's something you're always considering."

Another recurring theme focused on how the experimental threat attributes assisted the participants in making fewer hedged forecasts. Hedging, defined by Murphy (1978) as a forecast in which there is a "difference between a forecaster's judgment and his forecast." Some HWT-Hydro participants felt that by being forced to consider the uncertainty and assign a magnitude uncertainty attribute to each watch and warning, their ability to hedge was reduced; generally, this was a desirable outcome.

In response to the question of how to improve the attribute categorizations, participants felt that the probabilistic intervals of 25% were appropriate for the testbed experiment. However, several participants suggested that smaller probabilistic intervals would allow them to communicate threat information more accurately to forecast consumers in a real-world setting.

4. DISCUSSION

4.1 *Lessons Learned*

From an experimental perspective, the focus groups produced a number of lessons learned regarding the design and evaluation of the experimental threat attributes. When asked whether or not the magnitude and probabilistic categories were appropriate, participants felt that the probabilistic levels were fine for their current forecasting skill level when using the experimental FLASH tools, but it could be useful to have a scale with smaller intervals for operational forecasting. To address this concern, future iterations of the hydrology testbed experiment will allow forecasters to select probabilities at thresholds spaced one percentage point apart.

From an evaluation standpoint, it was very difficult to separate probability from magnitude in the discussion. Both were so closely linked, it was difficult to get a clear picture of how probability and magnitude were chosen separately. In addition, probability thresholds for major and nuisance flooding changed based on environment and socio-geographic constructs. Participants discussed differences in probabilistic thresholds that they needed to reach in order to issue warnings over rural and urban areas, exemplified in the quote by the following participant:

> "As that level of severity increases, especially over an area where you know is, is a wilderness area, you kind of hit that threshold and say, 'boom, I'm [going to] issue at this point.' Whereas… that threshold is [going to] be a lot lower over a metropolitan area."

4.2 *Recommendations*

Based on responses from the focus groups, three recommendations were developed for the future of flash flood forecasting and decision-making research. With regard to the development of impact- and uncertainty-based forecast products, participants expressed the need for consistency, actionable terminology, and a standardized scale for flood threat level.

Participants pointed out that terminology often varies when forecasting for river floods, areal floods, and flash floods. Although the HWT-Hydro focused entirely on flash floods, the participants generally worked in professional roles that required them to issue warnings for other types of flood threats, as well. A unified flood forecasting system requires consistent terminology to facilitate communication between actors in the weather response system.

Testbed participants also indicated that the term "nuisance flooding" was difficult to define from a scientific and a social perspective. There is a great need for future research to address best practices with regard to what type and quantity of information should be shared with different types of forecast consumers. For example, an emergency manager may be able to make a more informed decision after receiving a magnitude uncertainty attribute issued alongside a warning polygon, but this type of information may be

interpreted differently by an individual in a different role.

Although some forecasters stated that they do discuss potential impacts with forecast consumers, there is currently no standardized method of communicating such risks to forecast consumers. Initiatives such as Impact-Based Warnings (IBW) have experimented with the design of text-based forecast products that contain information related to potential impacts. An evaluation of IBWs for tornado threats revealed that up to a certain threshold, including possible impacts in the text product increased the likelihood that an individual would take protective action (Ripberger et al., 2014). Furthermore, following a severe thunderstorm in Abilene, Texas in which an IBW was issued operationally, Guerrero et al. (2015) found that the additional impacts-oriented text gave members of the public actionable information that lessened confusion and clarified the level of risk.

Lastly, future work is needed to develop a scale for flash flood forecasting impacts. Unlike the Enhanced Fujita Scale for tornado threats, there is no scale available for use by National Weather Service forecasters for communicating flash flood threat level. The nuisance and major flood categorizations used in the magnitude attributes in HWT-Hydro attempted to provide a basic structure for flood threat. However, additional research into scientifically and socially appropriate threat levels would be of great benefit to the forecasting community and society at large.

5. CONCLUSION

The 2014 Hazardous Weather Testbed Hydrology experiment allowed forecasters to experiment with new methods for communicating flash flood threat and forecast uncertainty. The threat attributes, represented by a probability of major and nuisance flooding, provided a means to communicate forecasters' mental models to forecast consumers. Future hydrology testbed experiments will build upon the findings from the present study in order to expand the body of knowledge related to impact- and uncertainty-based warning products and the weather forecasting community.

6. ACKNOWLEDGEMENTS

Funding for this research was provided by NOAA/OAR/Office of Weather and Air Quality (OWAQ) under the NOAA cooperative agreement, NA11OAR4320072.

The authors would like to acknowledge the staff of the Weather Decision Training Division and the Hazardous Weather Testbed staff at the National Weather Center for their support in this research.

7. REFERENCES

Barthold, F.E., Workoff, T.E., Cosgrove, B.A., Gourley, J.J., Novak, D.R., Mahoney, K.M., (2015). Improving Flash Flood Forecasts: The HMT-WPC Flash Flood and Intense Rainfall Experiment. Bulletin of the American Meteorological Society.

Boyatzis, R.E. (1998). *Transforming qualitative information: Thematic analysis and code development.* Thousand Oaks, London & New Delhi: SAGE Publications.

Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101.

Guerrero, H., Myers, L., Lyons, S., Dunn, J., & Johnson, M. (2015). *Public Reaction to National Weather Service Impact Based Warnings and The Effectiveness of Decision Support Services Provided During the June 12, 2014, Abilene, Texas Extreme Wind and Hail Event*. Paper presented at the 95th Annual Meeting of the American Meteorological Society, Phoenix, AZ.

Murphy, A. H. (1978). Hedging and the Mode of Expression of Weather Forecasts. *Bulletin of the American Meteorological Society, 59*(4), 371-373.

National Weather Service. (2011). *Weather Forecast Office Hydrologic Products Specification*. (National Weather Service Directive 10-922). Washington, D.C.: Retrieved from http://www.nws.noaa.gov/directives/sym/pd01009022curr.pdf.

Ripberger, J. T., Silva, C. L., Jenkins-Smith, H. C., & James, M. (2014). The Influence of Consequence-Based Messages on Public

Responses to Tornado Warnings. *Bulletin of the American Meteorological Society, 96*(4), 577-590.