THE EXPERIMENTAL WARNING PROGRAM

2019 EXPERIMENT SUMMARY

NOAA Hazardous Weather Testbed, Norman, OK

30 January 2020

Kodi Berry[1] and Tiffany Meyer[2,1]

M. Bowlan[2,3], B. Smith[2,1], M. Krocak[2,1], S. Martinaitis[2,1], J.J. Gourley[1], G. Stumpf[5,9], K. Manross[5,6], A. Bates[2,8], C. Ling[7], H. Obermeier[4,6]

[1] NOAA/OAR/National Severe Storms Laboratory
[2] OU/Cooperative Institute for Mesoscale Meteorological Studies
[3] NOAA/NWS/NCEP Storm Prediction Center
[4] CIRES/Global Systems Laboratory
[5] CIRA/Global Systems Laboratory
[6] NOAA/Global Systems Laboratory
[7] University of Akron
[8] NOAA/Warning Decision Training Division
[9] NOAA/NWS/Meteorological Development Laboratory

TABLE OF CONTENTS

1. INTRODUCTION

The NOAA Hazardous Weather Testbed (HWT) is a joint project of the National Weather Service (NWS) and the National Severe Storms Laboratory (NSSL). The HWT provides a conceptual framework and a physical space to foster collaboration between research and operations to test and evaluate emerging technologies and science for NWS operations. The HWT emerged from the "Spring Program" which, for more than a decade, tested and evaluated new forecast models, techniques, and products to support NWS Storm Prediction Center (SPC) forecast operations. Now, the HWT consists of two primary programs: the original Spring Program, which is part of the Experimental Forecast Program (EFP), and the Experimental Warning Program (EWP).

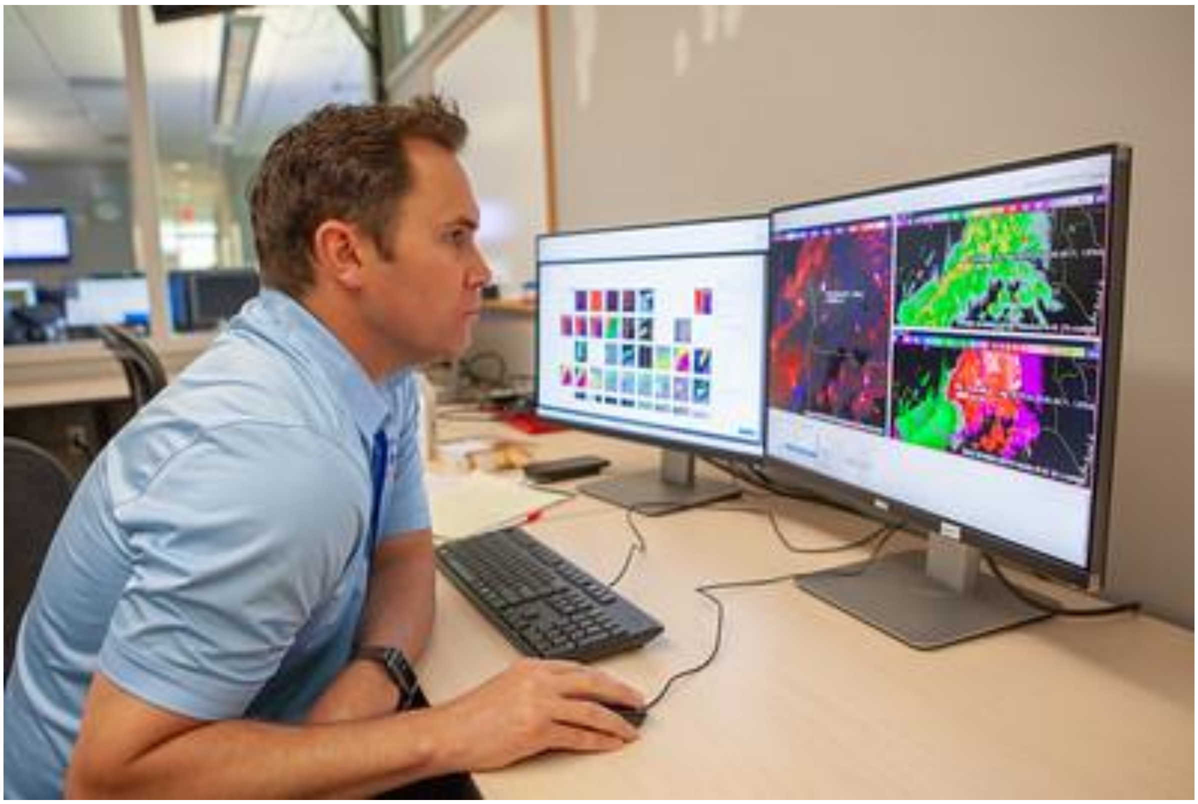

Figure 1. A participant examines radar products for an archived case.

The EWP tests and evaluates new applications, techniques, and products to support Weather Forecast Office (WFO) severe convective weather warning operations. This was the thirteenth year for warning activities in the HWT. Feedback was gathered from NWS operational meteorologists, broadcast meteorologists, and emergency managers. The experiment participants issued experimental warnings, published live blogs, engaged in focus groups, and completed surveys. User comments were also collected during shifts, which have been used to inform product development. This input is vital to improving the NWS warning process, which ultimately leads to saved lives and property.

2. OVERVIEW

The HWT Experimental Warning Program (EWP) at the National Weather Center (NWC) in Norman, Oklahoma hosted five primary experiments over 14 calendar weeks to improve National Weather Service severe weather warnings.

EWP Project	Dates	Length	Number of Participants
Satellite & Radar Convective Applications Experiment	Apr 22 - Jun 7	6 weeks	30 forecasters 3 broadcasters 2 military
Probabilistic Hazard Information (PHI) Emergency Manager Experiment	May 13 - 24	2 weeks	8 emergency managers
HMT Hydro Experiment	Jun 24 - Jul 19	3 weeks	10 forecasters
Hazard Services - Probabilistic Hazard Information (HS-PHI) Experiment	Oct 7 - Nov 1	3 weeks	6 forecasters
Hazard Services - Probabilistic Hazard Information (HS-PHI) End User Experiment	Oct 7 - Nov 1	3 weeks	6 broadcasters 11 emergency managers

Table 1. Details for the 2019 Experimental Warning Program.

3. PROJECT DETAILS AND RESULTS

Radar & Satellite Convective Applications Experiment

Summary by Michael Bowlan and Brandon Smith

Overview

The HWT provides an opportunity to conduct demonstrations of baseline, future capabilities, and experimental products associated with the next generation GOES-R geostationary and JPSS polar satellite systems, as well as many radar related products. Many of these products have the potential to improve hazardous weather nowcasting and short-range forecasting. Feedback from forecasters in the HWT has led to the continued modification and development of these algorithms for possible future operational implementation.

Experiment Design

During the HWT 2019 Satellite & Radar Spring Experiment, products were demonstrated within the real-time, simulated warning operations environment of the EWP using AWIPS-II (see figure below). Some radar products were also demonstrated in the form of case studies using the Warning Decision Support System-Integrated Information (WDSS-II) software suite. This experiment was conducted Monday-Friday during the weeks of April 22, April 29, May 6, May 13, May 20, and June 3. Participants included a new group of visiting forecasters each week. In total, 35 forecasters (30 NWS, 3 broadcast, 2 U.S. Air Force) participated in the experiment in 2019. Product developers from various collaborating institutions observed the activities and interacted with the forecasters. Monday through Thursday included eight-hour forecast/warning shifts and case study evaluation. Friday was a half-day dedicated to final feedback collection. During the simulated forecast shifts, forecasters utilized the baseline GOES-16 and WSR-88D data along with other experimental satellite products, in conjunction with operationally-available meteorological data, to issue non-operational short-term mesoscale forecast updates and severe thunderstorm and tornado warnings. Forecaster feedback was collected through the completion of daily and weekly surveys, daily and weekly debriefs, and blog posts.

GOES-16/17 Products

GOES-R series algorithms demonstrated during the experiment included: GOES-16/17 baseline ABI cloud and moisture imagery and GOES-16/17 baseline derived products, such as GOES derived stability indices, clear-sky TPW, derived motion winds, GOES-16/17 multi-spectral RGB composites, GOES-16/17 channel differences, UW/CIMSS ProbSevere Version 2, the GOES-16/17 Geostationary Lightning Mapper (GLM), the All-Sky LAP suite of products (Layered PW, TPW, MLCAPE, LI, K-index, and Total Totals), and the CIRA Advanced Blended Total Precipitable

Water product. This was the first year for data from GOES-17 ABI and GLM to be evaluated within the testbed environment. The experiment proved invaluable in testing this data and on the fly modifications were made, especially with experimental GLM products, per forecaster request. Ultimately the GLM products viewed most positively in the 2019 experiment will look to be released to the field in the future.

JPSS Products

From the JPSS program, the NOAA Unique Combined Atmospheric Processing System (NUCAPS) temperature and moisture profiles were demonstrated in the AWIPS-II NSHARP sounding analysis program. An experimental version of the NUCAPS profiles, which applied a boundary layer correction, was demonstrated again this year with further updates from the previous year. The NUCAPS soundings were available from the newly launched NOAA-20 polar orbiting satellite and evaluated in the testbed. NOAA-20 greatly reduced the latency of the soundings, which was well received by forecasters in the experiment. Gridded NUCAPS plan views of many parameters, such as lapse rates, were evaluated and subsequently released to the field in recent AWIPS build updates. A NUCAPs forecast product was evaluated for the first time in the HWT in 2019. The NUCAPS forecast product displayed plan views of instability parameters derived from NUCAPS and advanced the soundings forward in time using model data to arrive at a forecast of these profiles out to six hours. The idea of this product was applauded by many forecasters, but the algorithm still has a way to go to be operationally ready.

Single-Radar Products

The NWS Radar Operations Center (ROC) demonstrated two single-radar products that are proposed for the WSR-88D: 1) the New Mesocyclone Detection Algorithm (NMDA) and 2) velocity-derived azimuthal shear (AzShear). The evaluation provided an avenue for product developers and the ROC to obtain NWS and DoD forecaster feedback that helps determine their usefulness if deployed on WSR-88D network. In the case of the NMDA, feedback from participants helped to drive targeted improvements to the NMDA that are in-turn slated to be examined during a re-evaluation of the algorithm within the 2020 HWT Radar Convective Applications Experiment. These improvements include enhancements to the quality control of the mesocyclone detections, fine-tuning of the NMDA's ability to detect circulations located within quasi-linear convective systems (QLCS), as well as improving the overall visualization package within AWIPS-II. Perhaps the most important result related to the single-radar products evaluated in the 2019 HWT pertained to the participant feedback received for the AzShear product. An overwhelming number of participants noted that single-radar AzShear could positively impact NWS warning operations given its ability to aid in highlighting key features in velocity data that are precursors to tornadic circulations, especially those that are associated with challenging QLCS events.

Multi-Radar Products

Funded by NOAA's Office of Weather and Air Quality (OWAQ), the Cooperative Institute for Mesoscale Meteorological Studies (CIMMS) at the University of Oklahoma demonstrated a multi-radar product developed for the Multi-Radar/Multi-Sensor (MRMS) system called the Conditional Probability of Tornado Intensity (CPTI). Working in conjunction with and utilizing published tornado intensity research developed by the NWS SPC, CPTI provides a real-time probability of a tornado's intensity, conditional on the presence of a tornado. This product turns the manual approach developed by SPC forecasters into a fully automated, CONUS-wide, gridded probability product that can be used in real-time as guidance for the intensity of an ongoing tornado. Results from the evaluation of CPTI showed some potential in the utility of the product to aid in the adjustment of the impact wording within impact-based warnings (IBW) through the quantification that CPTI provides via its probabilistic intensity information. Feedback from the participants also highlighted the fact that providing context of the probabilities through training as-well-as allowing time for the building of personal calibration to the product, are both necessary for proper utilization of a probabilistic product by the end-user.

Figure 2. Forecasters evaluate Satellite and Radar products during the 2019 Spring Experiment in the Hazardous Weather Testbed.

Plans for Future Experiments

For the 2020 Spring Experiment in the HWT, the Satellite and Radar Convective Applications Experiment will be partially split into separate experiments, but will overlap for two weeks. The Radar experiment will run for four weeks starting in April and ending the first week of June. The Satellite Experiment will run for three weeks starting in mid-May through mid-June. Both experiments will evaluate similar products to the 2019 experiment, with a few new products. The structure for both experiments will be very similar with the satellite experiment evaluating products in real-time within AWIPS-II and the radar experiment using a mix of real-time and case study evaluations of products. Feedback will continue to be collected from the evaluation and reported out to the various funding programs for future analysis.

Satellite & Radar Convective Applications Experiment

Web Presence

GOES-R HWT Blog	http://goesrhwt.blogspot.com/
EWP Blog	http://blog.nssl.noaa.gov/ewp/
Forecaster Training	http://hwt.nssl.noaa.gov/ewp

Project Contacts

Kristin Calhoun	NSSL	Lightning
Michael Bowlan	CIMMS/SPC	Satellite
Tiffany Meyer	CIMMS/NSSL	Lightning/AWIPS-II Support
Brandon Smith	CIMMS/NSSL	Radar
Justin Monroe	CIMMS/NSSL	AWIPS-II Support

Probabilistic Hazard Information Emergency Manager Experiment

Summary by Makenzie Krocak

Overview

Three emergency managers were included in the PHI Emergency Manager experiment each week with the main objective of learning how the continuous flow of probabilistic information from days before the event through the warning time scale may impact them and their decision making. Each morning, experimental products on the 1- to 4-day time scales were viewed alongside SPC's operational convective outlooks and watches. Each afternoon, experimental probabilistic products and warnings were viewed in the NWS Enhanced Data Display and GR2Analyst. The probabilistic products were graphically rendered to quickly allow for quick identification of the hazard type, severity, direction of motion, level of automation, and time of arrival.

Emergency management participants simulated decisions for towns or areas that matched the scale of their jurisdiction (e.g., university, city, county, state). Researchers investigated how the experimental products changed the decisions they were making or the timeframes during which they were making decisions during hazardous weather events.

Figure 3. EM participants use the NWS Enhanced Data Display to investigate prototype probabilistic and warning information to make decisions during a simulated severe weather event.

Experiment Design

Standard operating plans for emergency managers have elements (e.g., sounding outdoor warning sirens) based upon forecasts, products, or warnings from the NWS. The addition of prototype products on longer time scales allowed participants to make decisions earlier or be more confident in the decisions they were making. For example, if a prototype product showed the event beginning around shift change at a hospital, that EM may make the decision to have the second shift come in early. This decision would have been made hours before the event started, allowing staff the time to get to work before the weather impacted travel. On the warning scale, traditional warning information in the PHI system helped forecasters and users connect with necessary and effective elements of the current warning system. In addition, preliminary results show that EM participants used both severe and sub-severe information in their decision making. EMs carefully watched the trends in probabilities, and depending upon circumstance, they made decisions based first on time of arrival, second on severity. For example, if a dorm at a university requires 18 minutes to get students to safe areas on the lowest floors, that EM might make decisions ahead of a warning because more time is required than a typical warning lead time.

Plans for Future Experiments

The 2020 PHI Emergency Manager Experiment will take place in the spring during which three EMs will participate each week. Research interests include understanding if and how longer lead time (12 hours to 2 days) prototype products impact event preparation decisions. In addition, researchers are interested in understanding how probabilities on different time and space scales can create a cohesive forecast story.

Prototype Probabilistic Hazard Information Emergency Manager Experiment

Web Presence

Emergency Manager Project Blog https://blog.nssl.noaa.gov/phi-em/
FACETs Program https://nssl.noaa.gov/projects/facets/

Project Contacts

Kim Klockow	CIMMS/NSSL	Lead Social Scientist
Katie Wilson	CIMMS/NSSL	Warn-on-Forecast Scientist
Kenzie Krocak	CIMMS/NSSL	Graduate Student Researcher
Zach Stanford	OEM	EM Facilitator
Cassandra Shivers-Williams	CIMMS	Postdoctoral Researcher
Adrian Campbell	CIMMS/NSSL	Computer Scientist

Multi-Radar Multi-Sensor Hydrometeorological Testbed – Hydrology Experiment

Summary by Steven Martinaitis and J.J. Gourley

Overview

The Multi-Radar Multi-Sensor (MRMS) Hydrometeorological Testbed - Hydrology Experiment (hereinafter denoted as "HMT-Hydro Experiment") focused on the issuance of experimental flash flood warnings (FFWs) for extreme hydrologic flash flood events during a select period of the warm season. The experiment was conducted in close coordination with the seventh annual Flash Flood and Intense Rainfall (FFaIR) Experiment at the NOAA/NWS Weather Prediction Center (WPC) located in College Park, MD and ran for three weeks from 24 June to 19 July 2019. Forecasters from the NWS Weather Forecast Offices (WFOs) and River Forecast Centers (RFCs), along with participation from other NWS entities and research institutions, assessed emerging hydrometeorological technologies and products to improve the prediction, detection, and warning of flash floods.

The next evolution of flash flood prediction will integrate probabilistic and uncertainty information into the warning decision-making process. The 2019 HMT-Hydro Experiment included real-time experimental warning operations and archived case studies using prototype products and techniques. The three topics of interest for the 2019 HMT-Hydro Experiment were: 1) the use of probabilistic information to convey uncertainty of the flash flood threat, 2) the use of Warn-on-Forecast System (WoFS) quantitative precipitation forecasts (QPFs) for short-term prediction of potential flash flooding, and 3) the initial evaluation of probabilistic quantitative precipitation estimation (PQPE) products. Activities were followed by structured discussions and subjective evaluations to understand the perception of these products during flash flooding (Fig. 4). The primary scientific goals of the 2019 HMT-Hydro Experiment were:

- Evaluate the relative skill of experimental probabilistic flash flood monitoring and short-term predictive tools.
- Determine the potential benefits/limitations of utilizing precipitation forecasts for flash flood prediction and warning decision making.
- Assess the utility and skill of experimental FFWs that communicate the uncertainty and magnitude of the flash flood threat.
- Evaluate the skill of PQPE precipitation accumulations and the associated probabilistic and uncertainty products.
- Enhance cross-testbed collaboration and coordination, as well as collaboration between the operational forecasting, research, and academic communities, on the forecast challenges associated with short-term flash flood forecasting.

- Identify how the use of probabilistic information can advance the science and communication of impacts of flash flooding within the Forecasting a Continuum of Environmental Threats (FACETs) paradigm.

Figure 4. Follow-up discussions during the subjective evaluation of experimental probabilistic flash flood products from the Minneapolis, MN flash flood event of 16 July 2019.

Results

Real-time experimental warning operations focused on the use of four probabilistic gridded flash flood products forced by the MRMS experimental dual-pol synthetic radar QPE with evaporation correction (e.g., Fig. 5):

- Probability of Receiving a Flash Flood local storm report (LSR)
- Probability of Minor Flash Flooding
- Probability of Moderate Flash Flooding
- Probability of Major Flash Flooding

Participants utilized various probabilistic data to issue experimental FFWs using a modified WarnGen application. Each FFW issued had user-assigned probabilities for minor and major flash flood potential. Participants filled out survey questions on how the products influenced their warning decision. Subjective evaluations and feedback from the participants showed perceived biases in the different probabilistic fields but were improved from the 2018 experiment. Evaluated experimental warnings with user-assigned minor and major probabilities were favorably rated with respect to timing, warning area, and assigned probabilities. Reliability diagrams for all experimental warnings showed a tendency for overestimation of the assigned probabilities, yet the overestimation bias was improved upon from the 2018 experiment.

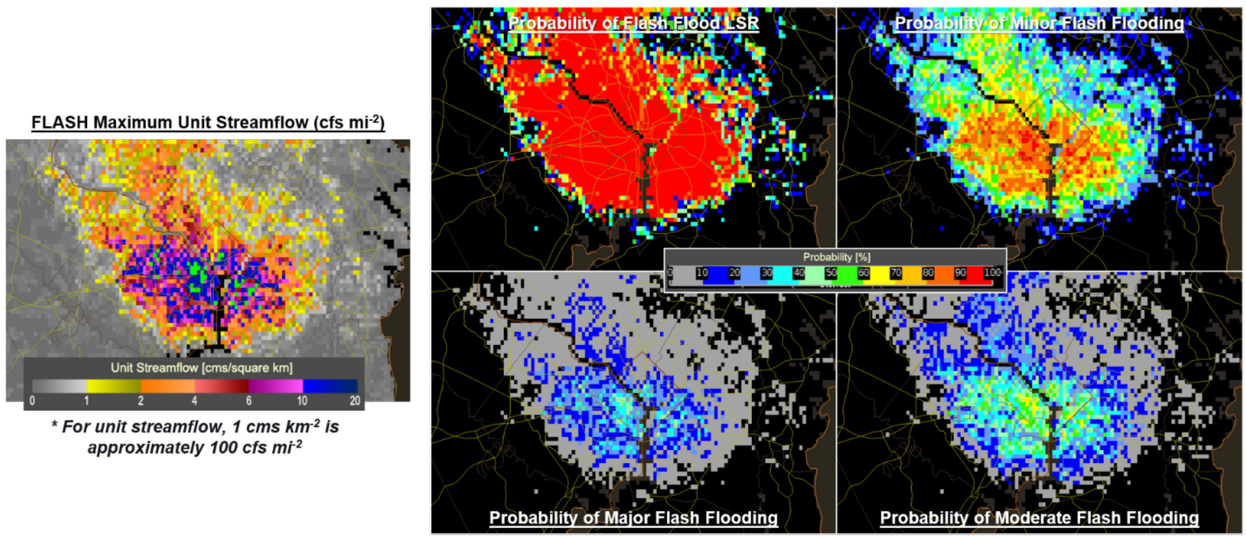

Figure 5. The Washington, D.C. flash flood event as seen from the deterministic (left) and probabilistic (right) data at 1400 UTC 8 July 2019.

Changes in warning lead time and warning area were also assessed with the verified experimental FFWs. Twelve of the experimental FFWs had an increase in warning lead time compared to the operational FFWs. Three warnings had an increase of lead time > 30 minutes, with a warning issued on 11 July having an 87-minute increase in lead time. Eight isolated flash flood events had smaller warned areas from the experimental FFWs. Five of the warning areas were over 50% smaller than the operational warning area. In all events except one, the reduction in the warned area was hundreds of square kilometers. The possibility is there to reduce false alarm area in FFWs, and the topic should consider future research.

Archived case studies focused on the evaluation of the potential impacts of ingesting the WoFS QPFs into the flash flood prediction process. Forecasters were presented with three cases of varying degrees of flash flood potential and severity:

- Ellicott City, MD Event (1900 UTC 27 May 2018 to 2110 UTC 27 May 2018)

- Central Iowa Event (2200 UTC 30 June 2018 to 0140 UTC 1 July 2018)
- Sioux Falls Event (2130 UTC 12 July 2018 to 0140 UTC 13 July 2018)

The cases focused on the coupling of the probabilistic FLASH data with the WoFS QPFs in a displaced real-time environment. Participants were given a new 0-3 h forecast every 30 minutes with a survey every half-hour regarding the uncertainty and confidence of the evolving flash flood threat. Participants were able to issue FFWs using the experimental probabilistic data that included the WoFS QPFs as a forcing.

Seven of the eight events had a positive increase in warning lead time for warnings that were issued prior to the first flash flood LSR. Four of the events saw an increase in average warning lead time by over 30 minutes using the experimental probabilistic FLASH data coupled with the WoFS QPFs, including an average 30.4-minute warning lead time increase for the historic Ellicott City, MD flash flood of 2018 (Fig. 6). The collocated flash flooding in Baltimore, MD had the average warning lead time increased by over an hour to an average lead time of 71.6 minutes using the experimental data (Fig. 7). There were also multiple instances of the participants issuing flash flood statements (FFSs) to update their assigned minor and major probabilities based on the probabilistic FLASH data with WoFS QPFs.

The 2019 HMT-Hydro Experiment was the first time that participants were introduced to the PQPE product suite. A one-month data set for June 2017 was provided in a web-based format with six specific cases identified for initial analysis. Forecasters evaluated a variety of products related to the uncertainty information portrayed within the QPE data and the probabilities of exceeding various rainfall rate values. Analysis of instantaneous exceedance probabilities of rainfall rates of 25 mm h^{-1} and 50 mm h^{-1} along with temporal periods of 1 and 3 hours were found to be most relevant in understanding the magnitude of the rainfall event and the likelihood of flash flooding. Certain products were shown as being more favorable; moreover, feedback was given on how best to display the PQPE product suite. The initial set of results for PQPE will be used to refine the overall PQPE product suite.

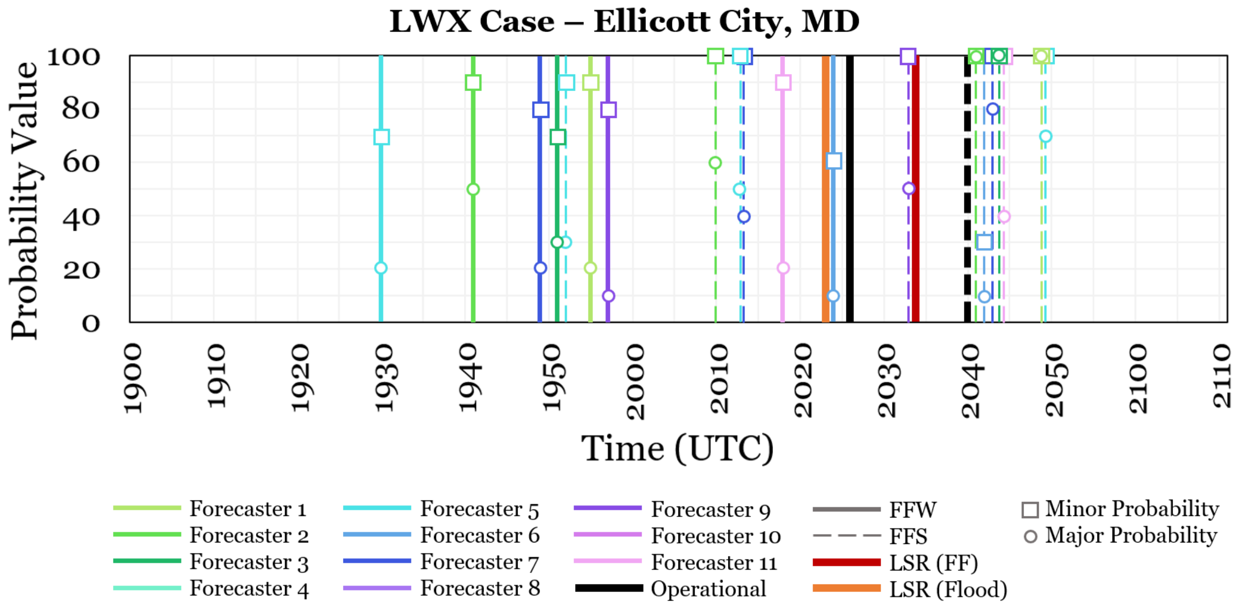

Figure 6. Timeline of all products issued along with LSRs for the Ellicott City, MD event in the Baltimore/Washington (LWX) case. Forecaster experimental FFW (FFS) issuance are denoted by solid (dashed) lines in cool colors. The assigned minor (major) probabilities for each experimental FFW and FFS are denoted by a square (circle) at each time. The operational FFW (FFS) is denoted by a thick black solid (dashed) line. Flood LSRs are denoted by a thick orange line. Flash Flood LSRs are denoted by a thick red line.

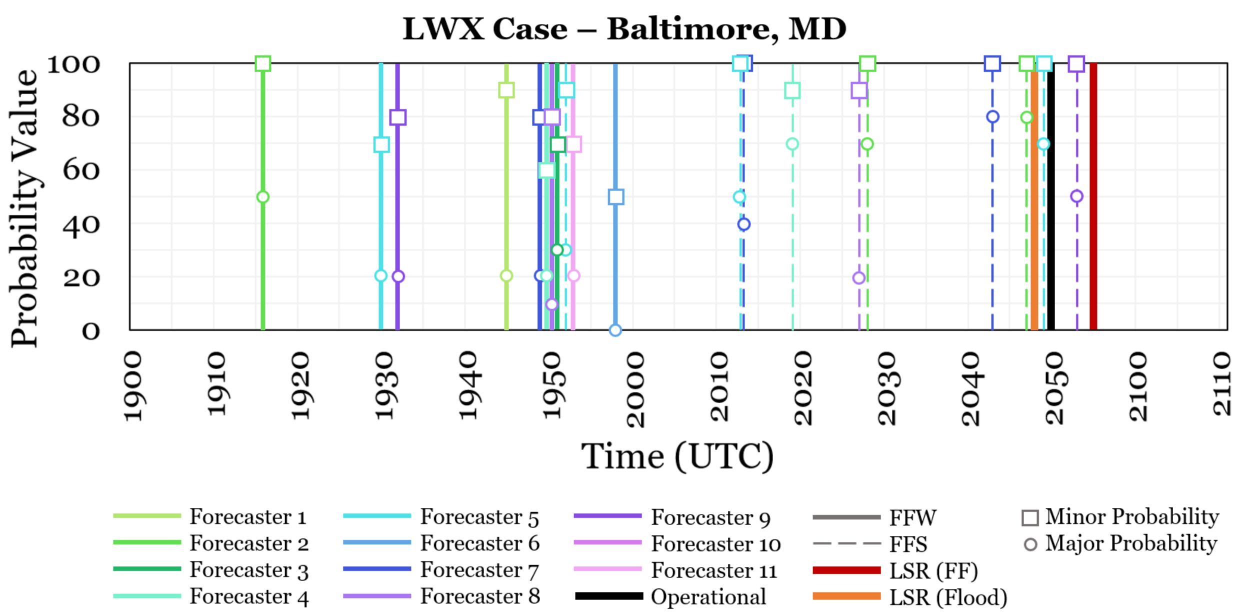

Figure 7. Same as Figure 6 except for Baltimore, MD in the LWX case.

Future Planning

The information and feedback gained during the 2019 HMT-Hydro Experiment will further advance the science and application of probabilistic products and WoFS QPFs for flash flood prediction and warning issuance. Work will continue to refine the probabilistic information portrayed in the experimental FLASH products and investigate how to the apply it to the new NWS impact-based warning design for FFWs. Future work will look at implementing the probabilistic data into the Hazard Services software platform.

The core of the WoFS will shift to FV3; thus, studies will continue with the coupling of the short-term ensemble WoFS with FLASH to improve the prediction and warning of flash flooding. Long-term planning will investigate the real-time running of FLASH with WoFS; moreover, the lessons learned from the HMT-Hydro Experiment will help in the development of the FACETs framework for a next-generation severe weather watch/warning framework.

Hydrometeorology Testbed Multi-Radar Multi-Sensor Hydro Experiment

Web Presence

HMT-Hydro Page	https://blog.nssl.noaa.gov/flash/
FLASH System Data Page	http://flash.ou.edu/
FACETs Program	https://nssl.noaa.gov/projects/facets/

Project Contacts

Jonathan J. Gourley	NOAA/OAR/NSSL	Principal Investigator
Steven Martinaitis	OU-CIMMS/NSSL	Principal Investigator
Katie Wilson	OU-CIMMS/NSSL	Warn-on-Forecast Coordinator
Nusrat Yussouf	OU-CIMMS/NSSL	Warn-on-Forecast Coordinator
Humberto Vergara-Arrieta	OU-CIMMS/NSSL	Real-Time Operations Coordinator
Andres Vergara-Arrieta	OU-CIMMS/NSSL	Real-Time Operations Coordinator
Pierre-Emmanual Kirstetter	University of Oklahoma	PQPE Coordinator
Micheal Simpson	OU-CIMMS/NSSL	PQPE Coordinator
Nathaniel Indik	OU-CIMMS/NSSL	PQPE Coordinator
Tiffany Meyer	OU-CIMMS/NSSL	HWT IT Coordinator
Justin Monroe	OU-CIMMS/NSSL	HWT IT Coordinator
Kodi Berry	NOAA/OAR/NSSL	Executive Officer – HWT
Ami Arthur	OU-CIMMS/NSSL	Daily Coordinator

Hazard Services – Probabilistic Hazards Information Experiment

Summary by Greg Stumpf, Kevin Manross, Alyssa Bates, and Chen Ling

Overview

NSSL has been developing a prototype tool for testing the early concepts of FACETs known as Probabilistic Hazard Information (PHI). This PHI Prototype Tool was evaluated by NWS forecasters and human factors experts in the HWT from 2014 to 2017. A NOAA Joint Technology Transfer Initiative (JTTI) grant is funding an effort to transfer the capabilities of the prototype into AWIPS-II Hazard Services (HS). The project has just concluded its fourth year of experimentation. The fourth HS-PHI experiment was conducted during October-November 2019. As with previous experiment years, this evaluation included NWS forecasters and human factor experts. New for this year, we operated in tandem with emergency managers and broadcast meteorologists (Hazard Services - Probabilistic Hazard Information Experiment with End Users). We evaluated the software design and the concept of PHI as it relates to hazardous weather warning operations using archived data.

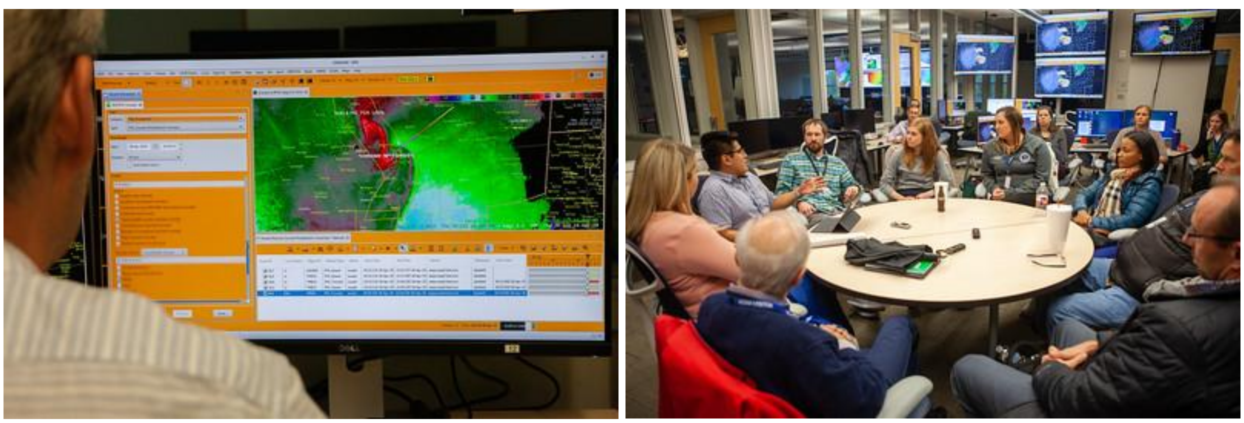

Figure 8. Images from the HS-PHI experiment: Forecaster using HS-PHI on an AWIPS-II workstation (left); post-scenario group discussion (right).

2019 Accomplishments

The following is a list of accomplishments from the 2019 HS-PHI Experiment:

- Completed the development of HS-PHI, with the goal to match the functionality of the 2015 version of the PHI Prototype prior to this experiment. In addition to these capabilities, these new components were available for 2019:
    - Deterministic Threats-In-Motion (TIM) polygons,
    - Warning product output (with VTEC, etc.), and

    - Lightning PHI.
- Developed six new archived cases from 2018-2019 to test the software on a variety of severe weather conditions. All of these new cases used Supplemental Adaptive Intra-volume Low-level Scan (SAILS) tilts, which comprise rapid updates to the low-elevation radar scans. Cases included:
    - Single isolated violent long-tracked tornado event in Alabama and Georgia,
    - Isolated supercells in Iowa,
    - Tropical outer convective rainband event for testing Lightning PHI,
    - Quasi-Linear Convective System (QLCS) squall line and long-lived tornado event in Mississippi,
    - Complex severe storm system with numerous storm mergers and splits in the High Plains, and
    - Major tornado outbreak in Ohio.
- Tested HS-PHI in the HWT for three alternating weeks from October 7 through November 1 with 6 NWS forecaster participants. The outcomes of this test included:
    - Gathering feedback on software performance and design, with bug-fixes and improvements developed and tested during the off-weeks of the test;
    - Collecting forecaster workload data in collaboration with human factors scientists;
    - Capturing discussions on the FACETs and PHI concepts in NWS severe weather warning operations, including how adjacent forecast offices would collaborate and share severe storm objects to provide seamless service across forecast area boundaries;
    - Integrated warning team (IWT) collaboration with emergency managers and broadcast meteorologists for the afternoon cases, coordinated by a scientist performing the duties of a station Warning Coordination Meteorologist (WCM). These key partners provided valuable feedback on the products and data generated by HS-PHI, specifically the continuously-updating warnings and probability plumes, and how they would use them in their own decision making.

Mental Workload (NASA TLX) Survey

The NASA-TLX survey is a questionnaire-based mental workload rating tool. The tool encompasses six sub-dimensions of mental workload: mental demand, physical demand, temporal demand, performance, effort, and frustration. The analysis of mental workload includes a weighting dimension used to calculate an overall workload score. The raw scores of the mental workload ranges from 1 to 100, where 1 represents extremely low workload and 100 represents extremely high workload. The ratings were averaged from all of the cases for each of the six sub-dimensions of workload, and the importance factors were calculated for each aspect.

Figure 9 shows the average mental workload for all forecasters from all archived hazardous weather events. Each bar in the figure represents the average mental workload for each of the six sub-dimensions of workload. The width of each bar represents the importance of each sub-dimension. The red line represents the average workload. The average workload for 2019 HS-PHI was 54.4 (out of 100, standard deviation 18.8), which is an increase from 2018 workload average of 47.7. All workload sub-dimensions increased from 2018.

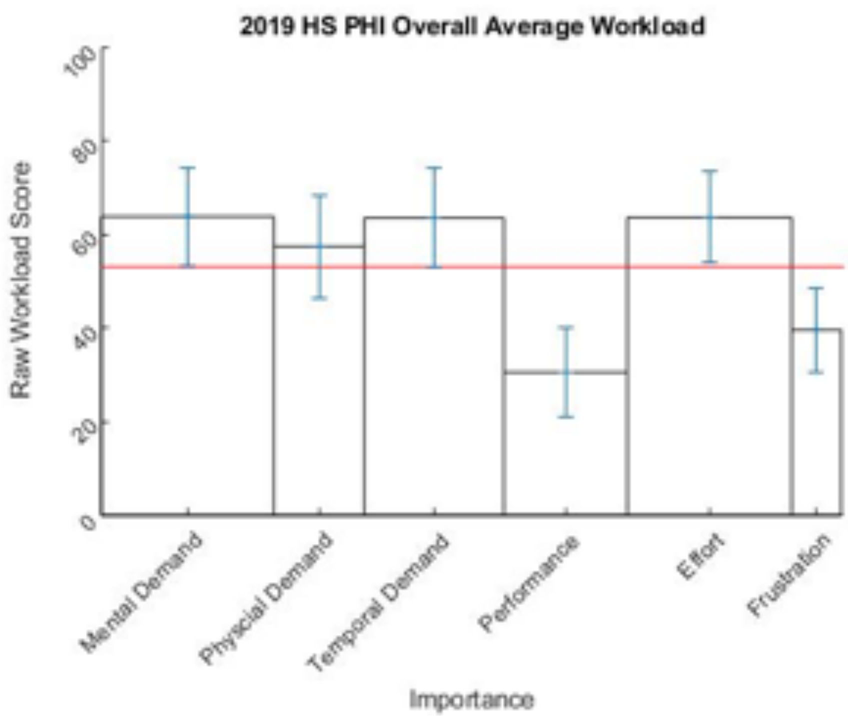

Figure 9. NASA-TLX workload sub-dimensions with the average mental workload for the 2019 HS-PHI experiment represented by the red line.

Figure 10 shows the average mental workload for the past four years of the HS-PHI experiment. The 4-year average of HS-PHI workload is 52.6. Figure 11 shows the mental workload for cases when forecasters created tornado and severe thunderstorm PHI (average of 54.4), lightning PHI (average of 48.2), and Threats-in-motion (average of 52.9).

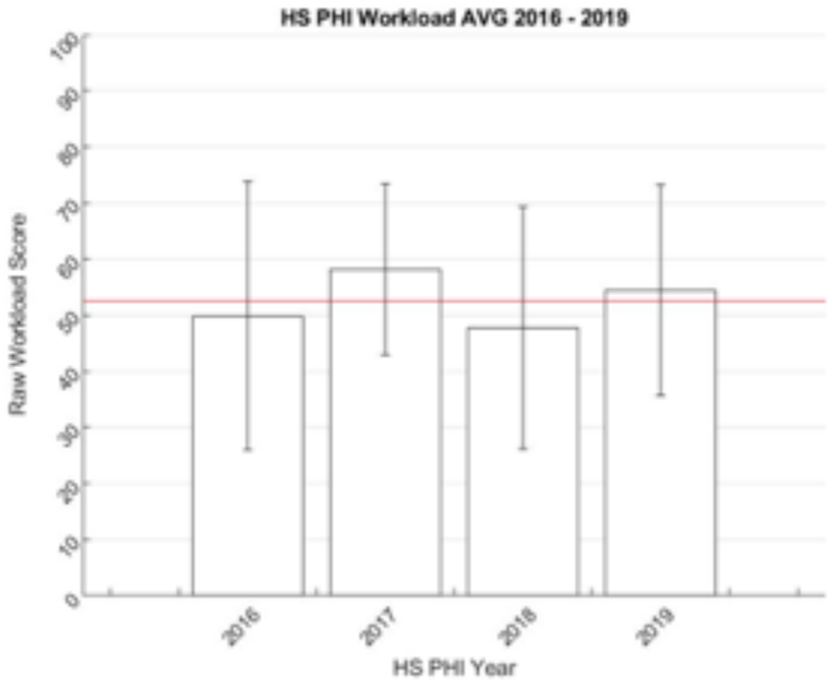

Figure 10. NASA-TLX mental workload with the average mental workload for all four years of the HS-PHI experiment represented by the red line.

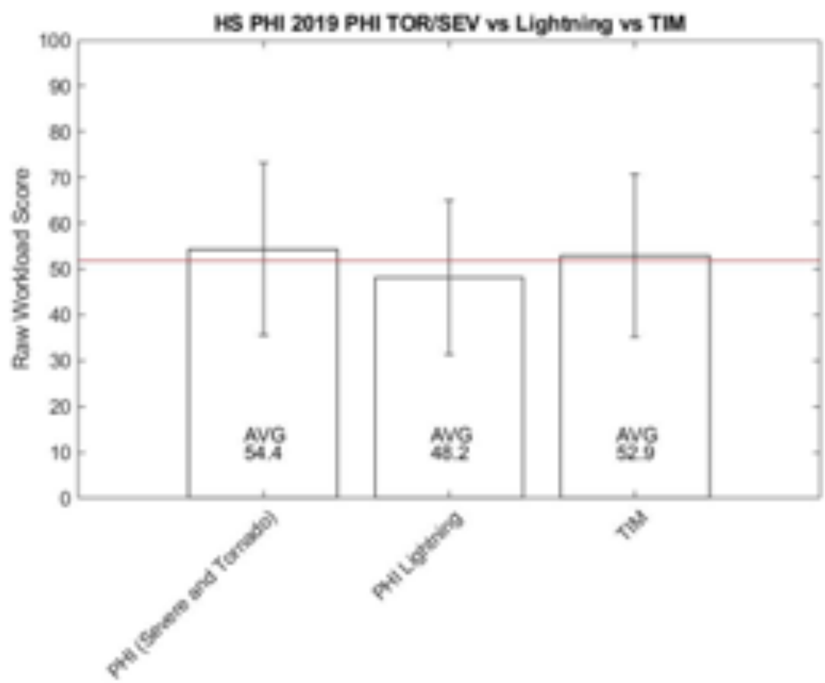

Figure 11. NASA-TLX mental workload comparing PHI (severe and tornado) to PHI (lightning) and TIM for the 2019 HS-PHI experiment.

Confidence Survey

Participants were asked to complete a confidence survey using a 7-point Likert scale, ranging from not confident at all (rating of 1) to very confident (rating of 7). The neutral response had a rating of 4. They were asked to rate their confidence in their understanding of the weather, in the automated guidance, and in producing PHI while working archived hazardous weather scenarios. They were given an optional text box to provide more details about what situations or events contributed to their confidence level designation. Analysis of the text was performed to determine the top contributing factors to forecaster confidence.

Figure 12 shows the average confidence score of 5.5 (out of 7, standard deviation 0.9). Forecasters stated their confidence was positively influenced by the pre-event weather briefing, well-organized storm development, and gaining experience with PHI and HS throughout the week. Some negative contributions to confidence were due to software issues, unfamiliarity with geography and storm type, and lagging automated guidance. Overall the confidence level for 2019 was higher than those reported in 2018, with an average of 5.

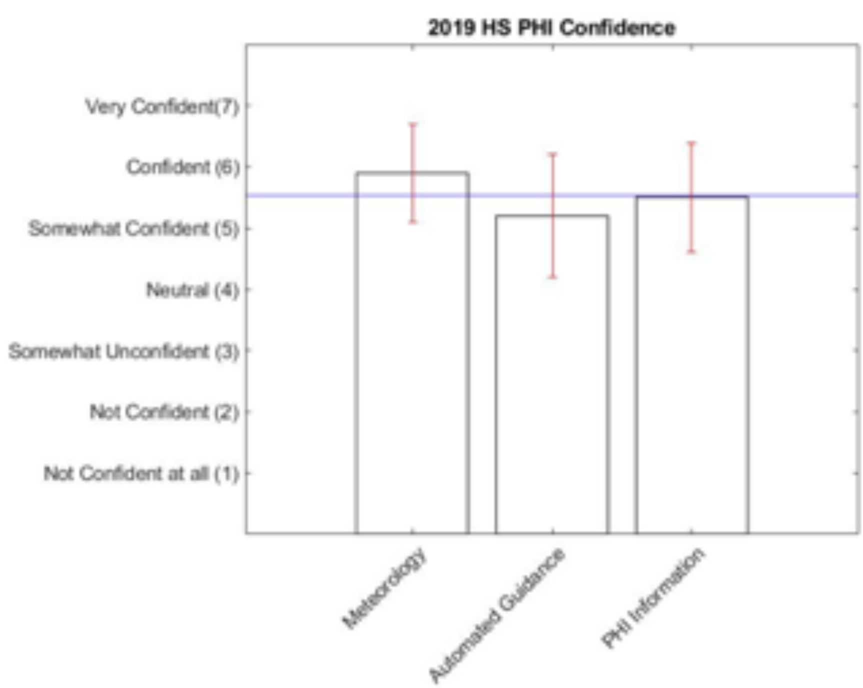

Figure 12. Forecaster confidence for meteorology, automated guidance, and probabilistic hazard information for the 2019 HS-PHI experiment.

PSSUQ Usability Questionnaire

The Post Study System Usability Questionnaire (PSSUQ) is a survey tool designed to evaluate the usability of a computer system. The tool is designed with 19 usability questions to ass 4 different areas of system usefulness, information quality, interface quality, and overall usability. The rating ranges from 1 to 7, where 1 corresponds to low level of usability, 7 corresponds to a high level of usability, and 4 corresponds to neutral level of usability.

The PSSUQ questionnaire was completed by the participants on the final day of the experiment. Table 2 shows the average responses for each of the four categories.

	2016	**2017**	**2018**	**2019**
Overall Usability	4.62	5.39	4.95	4.92
System Usability	4.96	5.56	5.21	4.88
Information Quality	4.37	5.00	4.37	4.81
Interface Quality	4.72	5.72	5.50	5.18

Table 2. Usability results based on the PSSUQ for 2016 to 2019 HWT (7-point scale)

Plans for Future Experiments

The following is a list of goals for the next iterations of the HS-PHI experiments:

- Complete development of Year 4 version of HS-PHI by January 24, 2020 for another HWT experiment in February 2020 that focuses on inter-forecast office collaboration issues. The development will include:
    - Addressing performance issues and software bugs
    - Expansion of collaboration capabilities to include “domain permission keys” so that a neighboring WFO cannot initiate new storm objects over another WFO without that WFO giving permission.
- Continue software refinement and development of new functionality, including:
    - Storm and environmental trends,
    - Storm Longevity (if available), and
    - ProbTornado recommender (if available).
- Integration of HS-PHI into the operational version of Hazard Services to take advantage of new features available in AWIPS-II versions 19x and 20x.
- Another HWT experiment in the June-July timeframe.

Hazard Services – Probabilistic Hazards Information Experiment

Web Presence

PHI – Hazard Services https://vlab.ncep.noaa.gov/group/facets/hwt-2019-hs-phi-resources

Project Contacts

Greg Stumpf	CIRA/NWS-MDL	POC and Co-PI
Alyssa Bates	CIMMS/NWS-WDTD	Co-PI
Chris Golden	CIRES/GSD	Developer
Yujun Guo	CIRA/GSD	Developer
Joe James	U. Akron	Project Scientist
James LaDue	NWS-WDTD	Project Scientist
Chen Ling	U. Akron	Co-PI
Kevin Manross	CIRA/GSD	Co-PI and Developer
Tiffany Meyer	CIMMS/NSSL	AWIPS-II Support
Justin Monroe	CIMMS/NSSL	AWIPS-II Support

Hazard Services - Probabilistic Hazard Information Experiment with End Users

Summary by Kodi Berry, Makenzie Krocak, and Holly Obermeier

Overview

The 2019 Hazard Services - Probabilistic Hazard Information (HS-PHI) Experiment with End Users was conducted during the weeks of October 7-11, October 21-25, and October 28-November 1. During this experiment, two broadcast meteorologists and four emergency managers worked in an integrated warning team with the HS-PHI forecasters each week. The emergency managers and broadcasters used the warning and probabilistic information to make simulated decisions in mock work environments. After each event, researchers brought the three groups together for discussions focused on the forecast information relevant to each forecast hazard type (tornadoes, severe thunderstorms, and lightning) and how each element could be improved.

The main objective of the HS-PHI End User Experiment was to learn how the continuous flow of probabilistic and warning information may impact them and their decision making. Probabilistic forecasts and warnings generated by the forecasters were viewed in an experimental version of the NWS Enhanced Data Display and GR2Analyst. The PHI objects were graphically rendered to quickly allow for quick identification of the hazard type, severity, direction of motion, level of automation, and time of arrival.

Experiment Details and Results

Emergency management participants simulated decisions for towns or areas that matched the scale of their jurisdiction (e.g., university, city, county, state). Researchers investigated how the experimental products changed the decisions they were making or the timeframes during which they were making decisions during hazardous weather events.

Standard operating plans for emergency managers have elements (e.g., sounding outdoor warning sirens) based upon forecasts, products, or warnings from the NWS. The addition of prototype products on longer time scales allowed participants to make decisions earlier or be more confident in the decisions they were making. For example, if a prototype product showed the event beginning around shift change at a hospital, that EM may make the decision to have the second shift come in early. This decision would have been made hours before the event started, allowing staff the time to get to work before the weather impacted travel. On the warning scale, traditional warning information in the PHI system helped forecasters and users connect with necessary and effective elements of the current warning system. In addition, preliminary results show that EM participants used both severe and sub-severe information in their decision making. EMs carefully watched the trends in probabilities, and depending upon circumstance, they made decisions based first on time of arrival, second on severity. For example, if a dorm at a university requires 18 minutes to get students to safe areas on the lowest floors, that EM might make decisions ahead of a warning because more time is required than a typical warning lead time.

Broadcast participants performed typical job functions under a simulated television studio environment as they received experimental probabilistic information from forecasters. Research protocols were used to systematically study how broadcast meteorologists interpreted, used, and communicated probabilistic information and rapidly-updating warnings. Decision points of interest included when to run “crawls,” post to social media, interrupt commercials, and interrupt programming. A protected Twitter account allowed researchers to study how participants communicate probabilistic information in a social media environment, which can differ from on-air methods. In addition, probability thresholds for coverage decisions were tracked. During one of the three experiment weeks, a bilingual broadcast meteorologist presented on air and posted on social media in both English and Spanish. This bilingual component of the project allows researchers to begin to study the challenges of communicating probabilistic information in the second-most used language in the United States.

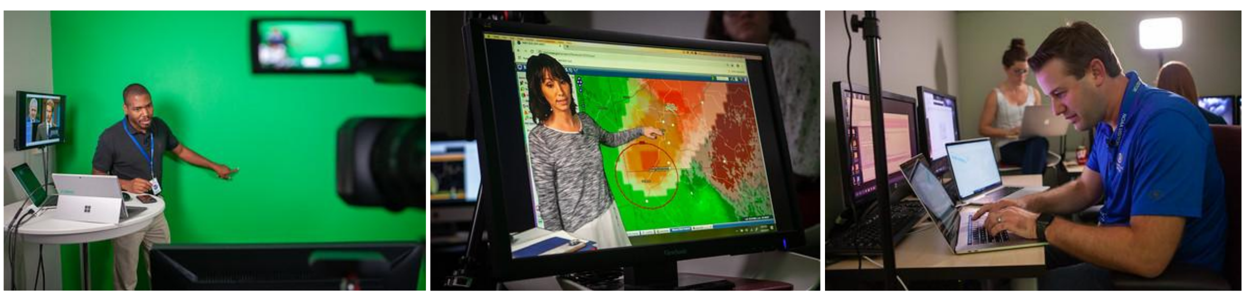

Figure 13. Broadcast meteorologists assess hazard probabilities and communicate with “viewers” during the HS-PHI End User Experiment.

Preliminary results reveal that broadcast participants often preferred to display and communicate the rapidly-updating warning polygons and probabilistic plumes separately. Inconsistencies between the warning polygon and the plume regarding spatial continuity made the two products difficult to communicate when layered together. The bilingual broadcast participant did not explicitly mention probabilities due to concern that Spanish-speakers would not be able to understand what they meant. In addition, the participant struggled to translate some of the weather jargon used by forecaster participants into Spanish.

Overall, research with end-users continues to refine contemporary ideas about how continuous probabilistic information may be useful, usable, and used.

Plans for Future Experiments

The 2020 HS-PHI Experiment with End Users will take place in the summer during which three emergency managers and two broadcast meteorologists will participate each week. Research interests common to both end user groups include continuing to improve the way multiple probabilistic products and warnings at different time and space scales are best communicated as a collection.

Hazard Services - Probabilistic Hazard Information Experiment with End Users

Web Presence

FACETs Program https://www.nssl.noaa.gov/projects/facets/
Broadcaster Project Blog https://blog.nssl.noaa.gov/kphi/
Emergency Manager Project Blog https://blog.nssl.noaa.gov/phi-em/

Project Contacts

Kim Klockow	CIMMS/NSSL	Emergency Manager Lead
Holly Obermeier	CIRES/GSD	Broadcast Co-Lead
Kodi Berry	NSSL	Broadcast Co-Lead
Pam Heinselman	NSSL	EM Researcher
Kenzie Krocak	CIMMS/NSSL	EM Graduate Researcher
Adrian Campbell	CIMMS/NSSL	Computer Scientist
James Murnan	AceInfo	A/V Specialist
Skylar Williams	CIMMS/NSSL	Broadcast Researcher
Zach Stanford	OEM	EM Facilitator
Cassandra Shivers-Williams	CIMMS	Postdoctoral Researcher
Joseph Trujillo	OU COMM & CIMMS/NSSL	Broadcast Graduate Researcher
Sean Ernst	OU School of Meteorology	Broadcast Graduate Researcher
Austin MacDonald	OU COMM & CIMMS/NSSL	EM Graduate Researcher
Emma Landeros	OU School of Meteorology	Broadcast Undergraduate Researcher

4. PERSONNEL

Alan Gerard
Acting HWT Co-Executive Officer alan.e.gerard@noaa.gov

Kristin Calhoun
Acting HWT Co-Executive Officer kristin.calhoun@noaa.gov

Tiffany Meyer
EWP Coordinator & Technical Lead tiffany.meyer@noaa.gov

Justin Monroe
EWP Technical Advisor justin.monroe@noaa.gov

5. ACKNOWLEDGMENTS

We would like to acknowledge the contributions of the following individuals: Tanya Riley, Tracy Reinke, Keli Pirtle, Emily Summars, Debra Molenar, Jack Dostalek, Emily Berndt, Nadia Smith, Rebekah Esmali, Kris White, Jonathan Case, Brian Kahn, John Forsythe, Dale Morris, Eric Bruning, Chris Schultz, Lee Cronce, John Cintineo, Justin Sieglaff, Karen Cooper, Carrie Langston, Vicki Farmer, Alan Gerard, Shadya Sanders, and Kenya Goods.

HWT Infrastructure Support was provided, in part, by the U.S. Weather Research Program and Joint Technology Transfer Initiative. Research was funded by the U.S. Weather Research Program, Joint Technology Transfer Initiative, National Severe Storms Laboratory, GOES-R Program Office, JPSS Program Office, NWS Radar Operations Center, and NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA16OAR4320115, U.S. Department of Commerce.