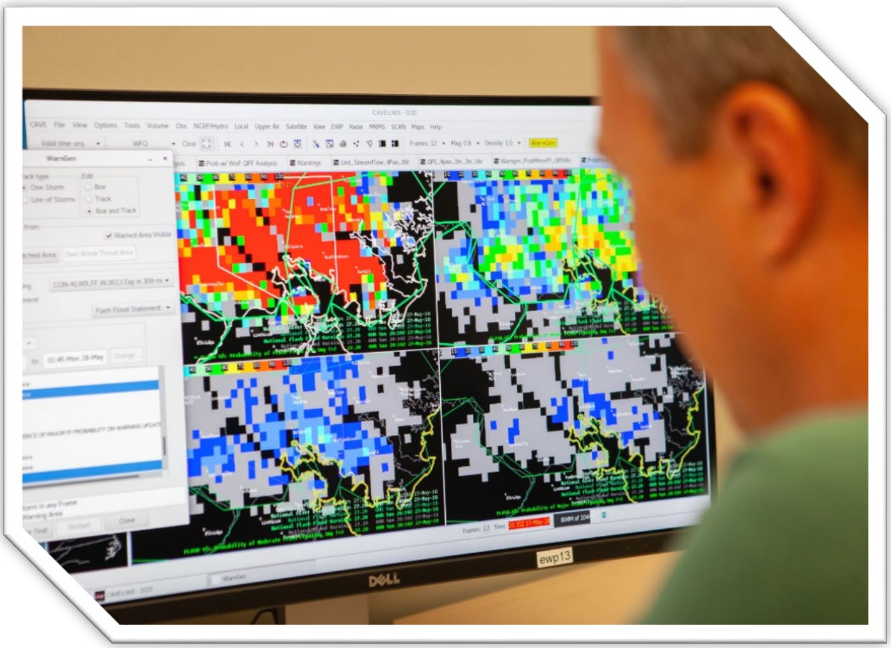
NOAA U.S. Weather Research Program (USWRP)

Hydrometeorology Testbed (HMT)
Multi-Radar Multi-Sensor (MRMS) Hydro Experiment

In Coordination with the HMT Flash Flood and Intense Rainfall (FFaIR) Experiment



-- 2019 HMT-Hydro Experiment Final Report --

June 24 – July 19, 2019

Hazardous Weather Testbed Facilities
National Weather Center
Norman, OK

Updated September 30, 2019
Version 1.0

TABLE OF CONTENTS

LIST OF TABLES 3

LIST OF FIGURES 4

I. Introduction 7

II. Objectives 8

III. Experiment Design and Activities 9

IV. Experiment Datasets 17

V. Results: Real-Time Operations 23

VI. Results: Archived Case Evaluations 34

VII. Results: Probabilistic QPE Analysis 44

VIII. Analysis, Other Findings, and Recommendations 50

Acknowledgements 53

References 54

APPENDIX A: HMT-Hydro Participants and Staff 55

APPENDIX B: Hours of Operations for HMT-Hydro Experiment 56

APPENDIX C: Products Used in HMT-Hydro Experiment 57

APPENDIX D: Responses from Feedback Survey 59

LIST OF TABLES

Table 1. List of the 12 experimental FFWs that had an increase in lead time over the collocated operational FFWs. Listed are the location (CWA) of each warning, the time of the first flash flood LSR, the time of the experimental and operational warnings, and the difference in lead time.

Table 2. List of events where the experimental warning area was smaller than that of the operational warning area. Listed are the events by date and CWA, the number of warnings involved in each calculation, the experimental and operational warning area, and the percent reduction of warning area.

Table 3. List of flash flood events analyzed for experimental warning lead times using WoFS QPFs in the probabilistic FLASH products.

Table 4. List of the operational FFW lead time, the average experimental FFW lead time, and the difference between the operational and average experimental FFW lead time.

Table 5. Operational warning metrics for all three archived cases.

Table 6. Experimental warning metrics for all three archived cases.

Table 7. Results of ranking the uncertainty information from the PQPE product suite.

Table 8. Results for expression of uncertainty value of PQPE.

Table 9. Results for expression of quantile value of PQPE.

Table 10. Results for rating the instantaneous exceedance probabilities for different rate values with respect to the understanding the rain rate magnitude.

Table 11. Results for rating the instantaneous exceedance probabilities for different rate values with respect to the understanding the flash flood threat.

Table 12. Results for expression of exceedance probability values with PQPE.

Table 13. Results for understanding the flood risk using exceedance probabilities over a given time period.

Table 14. Results for rating the various rain rate thresholds over a given time period.

Table 15. Results for rating the different time period accumulations for exceedance probability counts and percentages.

Table 16. Results for ranking the overall utility of each PQPE product.

LIST OF FIGURES

Figure 1. FFaIR ERO (left) and PFFF (right) for daily briefing given on 16 July 2019. The ERO was valid from 1500 UTC 16 July to 1200 UTC 17 July. The PFFF was valid from 1800 UTC 16 July to 0000 UTC 17 July.

Figure 2. Discussion of the experimental probabilistic flash flood products from the Minneapolis, MN flash flood event of 16 July 2019.

Figure 3. Percent of time over a 1-h period that the instantaneous rate exceeded 25 mm h^{-1} over northern Missouri ending 0300 UTC 29 June 2017 as seen from the web viewer utilized by HMT-Hydro participants.

Figure 4. WarnGen GUI used to create experimental FFWs. WarnGen was modified to survey participants about their warning decision making process using the experimental probabilistic flash flood products.

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. The images were taken from the flash.ou.edu web page, which was used during the evaluations. Note that the units for the FLASH CREST Maximum Unit Streamflow product is in metric units (m^3 s^{-1} km^{-2}).

Figure 6. Subjective ranking of the spatial coverage of the FLASH CREST Maximum Unit Streamflow product and the four experimental probabilistic products when compared to verified flash flood events using a box-and-whisker plot. The top (bottom) of each box represents the 75th (25th) percentile with the line in the middle of each box representing the median subjective ranking value. The top (bottom) whisker represents the maximum (minimum) ranking. The black dot represents the mean subjective ranking. The mean (μ) and standard deviation (σ) values for each product are shown below each box-and-whisker plot.

Figure 7. Subjective evaluation of the magnitude of values for the FLASH CREST Maximum Unit Streamflow product.

Figure 8. Subjective evaluation of the magnitude of values for the a) Probability of Receiving a Flash Flood LSR product, b) Probability of Minor Flash Flooding, c) Probability of Major Flash Flooding, and d) Probability of Major Flash Flooding.

Figure 9. Subjective evaluation of the spatial extent of experimental FFWs when compared to collocated operational FFWs for areas with verified flash flooding.

Figure 10. Same as Figure 9 except for the timing of FFWs.

Figure 11. Subjective evaluation of the assigned minor (left) and major (right) flash flood probabilities in the experimental FFWs containing verified flash flooding.

Figure 12. Objective assessment of the reliability of experimentally-issued flash flood warnings for major (blue) and minor (red) flash flood events for the 2019 experiment (dark shade) compared to the 2018 experiment (light shade).

Figure 13. Percent of responses regarding the influence of each probabilistic FLASH product on the warning decision for all warnings.

Figure 14. Percent of responses regarding the influence of each probabilistic FLASH product on the warning decision for (left) verified warnings and (right) unverified warnings.

Figure 15. Verified isolated flash flood event over the RLX CWA on 11 July 2019 where the area calculations utilized two experimental FFWs (orange polygons) compared to the single operational FFW (white polygon).

Figure 16. Verified isolated flash flood event over the PHI CWA on 17 July 2019 where the area calculations utilized one experimental FFW (orange polygon) compared to the three operational FFWs (white polygons).

Figure 17. Timeline of all of the products issued along with LSRs for the Ellicott City, MD event in the 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) are 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 18. Same as Figure 17 except for Baltimore, MD in the LWX case.

Figure 19. Same as Figure 17 except for southwest Minnesota in the FSD case.

Figure 20. Same as Figure 17 except for Gregory Co., SD in the FSD case.

Figure 21. Same as Figure 17 except for Douglas Co., SD in the FSD case.

Figure 22. Same as Figure 17 except for Story Co., IA in the DMX case.

Figure 23. Same as Figure 17 except for Dallas Co., IA in the DMX case.

Figure 24. Same as Figure 17 except for Polk Co., IA in the DMX case.

Figure 25. Same as Figure 17 except for Blackhawk Co., IA in the DMX case.

Figure 26. The domain areas for the DMX (top left), LWX (top right), and FSD (bottom) county warning areas. The DMX and LWX areas were restricted to the regions highlighted in yellow. Participants could forecast for the entire FSD domain but were directed to have increased attention for the two areas highlighted in blue.

I. Introduction

The National Oceanic and Atmospheric Administration (NOAA) Hydrometeorology Testbed Program (HMT) is administered by the Office of Water and Air Quality (OWAQ). The HMT promotes hydrometeorological research that will have quick and direct impact on operations within the National Weather Service (NWS), especially in regards to flash flood forecasting. The HMT provides a conceptual framework to foster collaboration between researchers and operational forecasters to test and evaluate emerging technologies and science for NWS operations. The project described herein is unique in that it addresses objectives of the HMT program while leveraging the physical facilities of the Hazardous Weather Testbed (HWT) at the National Weather Center (NWC) located in Norman, OK.

The fifth edition of the Multi-Radar Multi-Sensor (MRMS) Hydro Experiment (hereinafter denoted as "HMT-Hydro Experiment”) focused on the issuance of experimental flash flood warnings for the hydrologic extreme of flash flooding during a select period of the warm season. The 2019 HMT-Hydro Experiment contained a blend of experiments with real-time data and archived case playback using prototype products and techniques. 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.

The 2019 HMT-Hydro Experiment ran for three weeks during a period from 24 June to 19 July 2019 with a one-week break during the Fourth of July holiday. Forecasters from the NWS Weather Forecast Offices (WFOs) and River Forecast Centers (RFCs) along with participation from other NWS entities and research institutions worked with researchers to assess emerging hydrometeorological technologies and products to improve the prediction, detection, and warning of flash flooding. There were three primary topics of interest with the 2019 HMT-Hydro Experiment: 1) the use of probabilistic information to convey uncertainty of the flash flood threat, 2) the use of Warn-on-Forecast quantitative precipitation forecasts (QPFs) for short-term prediction of potential flash flooding, and 3) the initial evaluation of probabilistic quantitative precipitation estimation (PQPE) products. Various objectives of the experiment were primarily conducted through real-time experimental warning operations and archived case studies. Each week finished up with a group discussion focusing on the probabilistic products and use of QPFs in the warning decision making process.

Researchers from the National Severe Storms Laboratory (NSSL) and the University of Oklahoma (OU) Cooperative Institute for Mesoscale Meteorological Studies (CIMMS) administered the project and the HWT provided physical space and computing resources. This report discusses the activities of the 2019 HMT-Hydro Experiment and presents findings from it with a specific emphasis on operational impacts and recommendations for future investigations.

II. Objectives

HMT-Hydro Experiments prior to 2018 focused on deterministic products, including high-resolution distributed hydrologic model forecasts that operated on the flash flood time scale (Martinaitis et al. 2017). The hydrologic models examined in the past experiments were forced by MRMS radar-only quantitative precipitation estimates (QPEs). Different deterministic quantitative precipitation forecasts (QPFs) in combination with the QPE forcing for one of the hydrologic models were also evaluated.

The next evolution of hydrologic modeling and flash flood prediction will integrate probabilistic information and uncertainty into the warning decision making process. Activities within the 2019 HMT-Hydro Experiment were primarily split between real-time operations and archived case playback. For real-time operations, forecasters focused on the decision to issue experimental flash flood warnings using probabilistic gridded information. Archived case playback for a variety of flash flood events analyzed flash flood prediction tools from a hydrologic model that was forced by QPE combined with ensemble QPFs from the Warn-on-Forecast System (WoFS). There was time during the experiment that allowed forecasters to also evaluate a series of products from a probabilistic quantitative precipitation estimation (PQPE) product suite.

The HMT-Hydro Experiment was conducted in collaboration with the FFaIR Experiment (Barthold et al. 2015) to simulate the real-time workflow from forecast and guidance products in the 6–24 h timeframe from WPC to experimental flash flood warnings issued in the 0–6 h timeframe. For the days utilizing real-time experimental warning operations, the HMT-Hydro Experiment team acted as a "virtual, floating forecast office" to shift its area of responsibility to where heavy precipitation events and subsequent flash flooding was anticipated to occur. The participating forecasters had the ability to issue products for any county warning area (CWA) in the CONUS.

The primary scientific goals of the 2019 HMT-Hydro Experiment were as follows:

- 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 perceived skill of experimental flash flood warnings 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 the 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 societal impacts of conveying the threat of flash flooding within the Forecasting a Continuum of Environmental Threats (FACETs) paradigm.

III. Experiment Design and Activities

The HMT-Hydro Experiment ran Monday through Friday for three weeks from 24 June to 19 July 2019 with a break taken during the week of 1 July. The physical location of the experiment was in the Hazardous Weather Testbed (HWT) on the second floor of the National Weather Center (NWC) in Norman, OK. A total of 13 participants from a variety of NWS field offices as well as other participants came from NWS headquarters, training centers, and research facilities. Each participant contributed to the operational and evaluation activities of the HMT-Hydro Experiment. See Appendix A for the list of participants and the officers that conducted the experiment.

The mix of real-time experiment warning operations and archived case studies varied the day-to-day running of the HMT-Hydro Experiment. The period from Monday through Thursday contained two days of real-time operations and two days of case studies. It should be noted that analysis of the PQPE products were conducted during slow real-time operations, between archived cases, and/or on Friday morning. The weather and daily briefings from the FFaIR Experiment dictated what activities would be conducted on each day, and more importantly, which days would have the focus on real-time experimental warning operations. Basic information on the daily schedules for each week are provided in Appendix B.

Experiment Introduction and Training

Participants underwent an application and selection process under the aegis of the HWT in the months prior to the commencement of the experiment. NWS service hydrologists and forecasters expressing interest in storm-scale hydrology and in related scientific research received preference. Other participants outside of NWS field offices were brought in using testbed funding. Prior to their arrival in Norman, participants were given general information about the principal scientific goals of the experiment, including a copy of the operations plan and links to associated training material. Participants were not officially exposed to any experimental products or tools until the Monday afternoon session.

The introduction session on Monday morning featured a series of presentations given by the HMT-Hydro Experiments principal investigators (PIs) and officers. The presentations focused on the following topics:

- A reiteration of the scientific goals along with a history and past results from the HMT-Hydro Experiment
- Detailed descriptions and usage examples of all products available for use in both the real-time operations and archived case studies
- Description of the WoFS, the experimental design of the coupled WoFS and hydrologic modeling system and expectations on the use of Warn-on-Forecast with the archived case studies
- Description of the PQPE product suite
- Use of AWIPS and Hazard Services in the HMT-Hydro Experiment

The introduction session was modified based on the what the afternoon activities would be. Warn-on-Forecast archived case study presentation would occur just prior to the first case being used instead of it being given only on Monday morning. This was a result of the potential of the archived cases not being used until Wednesday; thus, this allowed the training on the cases to be fresh for the forecasters prior to starting the first case. A similar approach was given to the PQPE presentation.

FFaIR Daily Briefings

The daily weather briefing given at 12:30PM CDT was directed by the FFaIR Experiment held at the NWS Weather Prediction Center (WPC) in College Park, MD. HMT-Hydro Experiment participants and officers joined the briefing in the HWT using screen-sharing software, which were projected on to the large situational displays in the HWT. The primary goals of these briefings were to:

- Conduct a post-mortem on experimental products issued the prior day,
- Provide present synopsis of atmospheric conditions, rainfall, flooding for situational awareness, and
- Summarize experimental model-based forecasts of heavy rainfall and guidance for probabilistic flash flooding for the day and for the first probabilistic flash flood forecaster (which occurred during HMT-Hydro operations).

The briefing provided by the FFaIR Experiment along with the follow-up question-and-answer session usually lasted approximately 30 minutes. Each briefing stepped through various experimental products focused on the 6–24 hour forecasting of flash flooding. The final products derived by the FFaIR Experiment were the Excessive Rainfall Outlook (ERO) valid through 1200 UTC the following day and the first six-hour probabilistic flash flood forecast (PFFF; Figure 1). The first PFFF issued by the FFaIR Experiment was valid from 1800–0000 UTC, which encompassed the general real-time operations period of the HMT-Hydro Experiment. The details and products from these briefings, along with a beginning-of-the-week forecast, helped determine what days featured real-time experimental warning operations and what days focused on archived case studies.

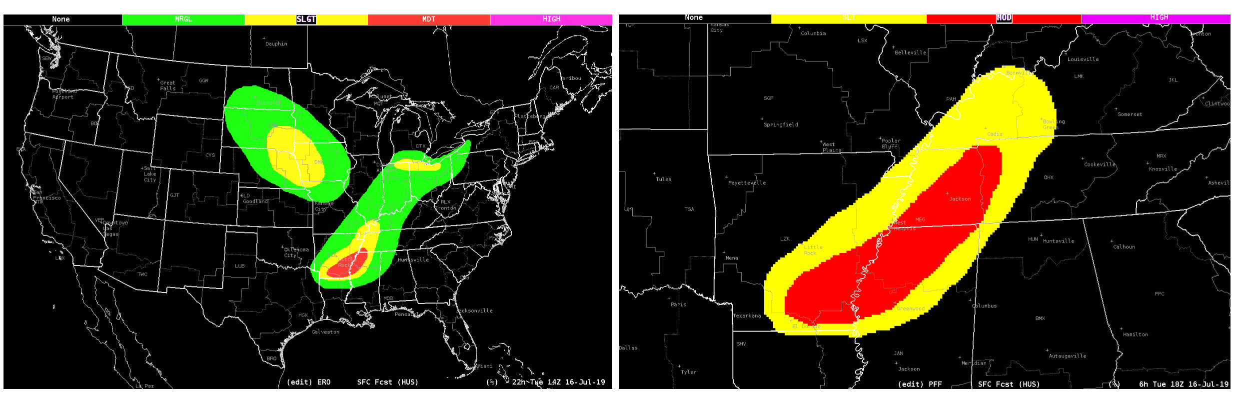


Figure 1. FFaIR ERO (left) and PFFF (right) for daily briefing given on 16 July 2019. The ERO was valid from 1500 UTC 16 July to 1200 UTC 17 July. The PFFF was valid from 1800 UTC 16 July to 0000 UTC 17 July.

Experimental Warning Operations

Experimental warning operations depended on the daily schedule and the weather observations and forecasts. The focus region(s) initially corresponded to the FFaIR Experiment forecast guidance and current observations. Participating forecasters primarily used products with the Flooded Locations and Simulated Hydrographs (FLASH) system with radar reflectivity and QPEs provided by MRMS.

The experimental warning operations design was intended to mimic the responsibilities of a local forecast office, but with the ability to change to any county warning area in the testbed. In the event of multiple flash flooding events occurring in separate regions of the CONUS, the HMT-Hydro Experiment officers prioritized the operations to the domain(s) with the anticipated biggest impacts and perhaps population density (in order to obtain a greater density of local storm reports). Participating forecasters were allowed to work the separate regions or collaborate in a specific region.

The experimental warnings differed from those issued in operations in that they included two probability of occurrence corresponding to flash flooding magnitudes. The WarnGen application was used to issue experimental products. The WarnGen GUI was tailored to solicit information from the forecaster regarding the decision-making process and the products used for the issuance of each experimental warning. The participant responses were then analyzed to see what products and probabilities influenced the warning decision. More information regarding the use of the WarnGen software in the HMT-Hydro Experiment are detailed in the Experiment Datasets section.

Subjective Evaluation Sessions

Experimental product and warning evaluation sessions followed up real-time experimental warning operations. This usually occurred on the following day; however, deviations in the operational schedule due to current weather conditions meant that evaluations were rescheduled to more appropriate periods. Multiple evaluation sessions sometimes occurred following a single real-time operational session. One of the HMT-Hydro Experiment officers guided the participants through a series of questions related to the experimental products and the associated experimental FFWs issued for verified flash flood events. Each participant had an equal vote in the evaluation process through the use of TurningPoint™ software and individual clickers used to collect, display, and archive forecaster responses for each question. A discussion followed each question with comments captured by HMT-Hydro Experiment officers (e.g., Figure 2).

The subjective evaluation sessions were broken into two sections: Experimental products and experimental FFWs that were primarily based on the products. The first session on experimental products focused on a single flash flood event that occurred during real-time warning operations. If no flash flood event occurred or if a more significant flash flood event occurred outside of the experimental operational hours and within 24-hours of the evaluation period, it could be utilized during this session. The experimental product evaluation focused on the following products:

Figure 2. Discussion of the experimental probabilistic flash flood products from the Minneapolis, MN flash flood event of 16 July 2019.

- FLASH CREST Maximum Unit Streamflow
- Probability of Receiving a Flash Flood LSR
- Probability of Minor Flash Flooding
- Probability of Moderate Flash Flooding
- Probability of Major Flash Flooding

For each of the five evaluated products, the participants were given the following two statements to subjectively rank the product performance: 1) Using all available flash flood observations, supply a score quantifying how the [product] depicted the spatial extent of flash flooding, and 2) Using all available flash flood observations, rate how the [product] depicted the magnitude of flash flooding. The subjective analysis of the spatial coverage of each product compared to verified flash flooding was conducted on a scale from 0–100 in ten-point intervals. The subjective analysis of each product depicted the magnitude of the verified flash flooding was on a seven-point Likert scale from “Much Too Low” to “Much Too High.” The deterministic FLASH CREST Maximum Unit Streamflow product is operational within the NWS and was used as the control product for probabilistic product comparisons.

The second session focused on the experimental FFWs issued by the participants. Only experimental FFWs where flash flooding occurred both within the polygon and within the valid time were used in this analysis. The participants were given the following three statements to subjectively evaluate the experimental FFW performance: 1) Using all available flash flood observations, rate the spatial accuracy of the experimental FFW vs. FFW(s) issued operationally, 2) Using the timing of the warnings, rate the issuance of the experimental FFW vs. FFW(s) issued operationally, 3) Using all available flash flood observations and tools, rate the probability of minor flash flooding that was given in the experimental flash flood warning, and 4) Using all available flash flood observations and tools, rate the probability of major flash flooding that was given in the experimental flash flood warning. Up to four experimental FFWs could be evaluated in a single evaluation session.

Archived Case Studies

Two days during the testbed week were dedicated to the evaluation of past flash flood events and the potential implications of including WoFS ensemble QPFs into the flash flood prediction process. Three cases of different threat varieties were utilized throughout the three weeks of the experiment:

- 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 CWA Event (2130 UTC 12 July 2018 to 0140 UTC 13 July 2018)

The timing and location of the WoFS events were selected based on the availability of WoFS 2018 real-time runs. The participant workload was also taken into account. The Ellicott City, MD and Central Iowa events were geographically constrained for focused analysis. The entire Sioux Falls CWA was available for analysis, yet the forecasters were directed to keep their central focus on two regions: southwest Minnesota and southeast South Dakota. The Sioux Falls CWA event originally began at 2030 UTC 12 July 2018; however, participant feedback after the first week led to the reduction of the case by one hour. This was to reduce forecaster fatigue for being on the warning desk for a prolonged period of time. The period removed (2030-2130 UTC) also had no significant data to collect per the forecaster feedback.

The three cases utilized the Weather Event Simulator (WES) to play back the data in a displaced real-time (DRT) format. This allows forecasters to view the data as it would come in real-time during an event. Each case was loaded and was paused at the start time. This allowed for the participant to watch a pre-brief presentation to increase their environmental awareness of the event. Information provided to the participants included a synoptic overview of the atmospheric conditions, sounding and precipitable water analysis, and the latest relevant products (WPC precipitation outlooks, mesoscale precipitation discussions, etc.). Once the participant completed the presentation, the simulated event can be played.

The archived cases utilized in the 2018 HMT-Hydro Experiment focused on three different conditions of data: 1) deterministic FLASH and MRMS output, 2) probabilistic FLASH forced with MRMS QPE, and 3) probabilistic FLASH forced with MRMS QPE plus WoFS QPFs. Participants analyzed all three conditions in order at the top of each hour. For the 2019 HMT-Hydro Experiment, only the third condition (probabilistic FLASH coupled with WoFS) was considered. Participants were directed to use the available deterministic products to understand the current conditions and then utilized the probabilistic FLASH data with WoFS QPFs to help drive their warning decision-making process.

At the onset of each case, various datasets from one hour prior to the start time were available to the participant forecasters. The first available run of the deterministic FLASH data set came in eight minutes past the start time. The first probabilistic FLASH data utilizing the WoFS QPFs as forcing were made available at 13 minutes past the valid (start) time. Even though it is unknown how the coupling of WoFS and FLASH would work in a real-time environment, the update of the FLASH data forced with WoFS ran on the update cadence of WoFS. A new forecast that went ten minutes farther into the future was presented every minute until a full 3-h (180 min) forecast was available at 30 minutes past valid time.

At nn:01 and nn:31 times for each hour, the WES Scripting Language (WESSL) application would prompt questions to the participants. These questions focused on the expectations and confidence of flash flooding from the perspective of the forecaster. The five questions that were generated every half-hour were as follows:

- What is your current expectation for a flash flooding threat?
- In what area is this threat likely to happen and over what time frame?
- How confident are you in this expectation?
- Please bullet point the product(s) that you are using to form this expectation for flash flooding and the corresponding product values that you consider important for this expectation. Please provide additional information on if any of the product(s) were weighted more than others.
- Please bullet point any actions you would take based on this expectation (e.g., contacting partners, communicating to the public, office planning logistics, etc.).

All questions had written responses except for the first and third questions. The first question (current expectations of flash flooding threat) had a four-point scale ranging from no flash flooding to major flash flooding. The third question (confidence of threat) had a five-point scale ranging from not at all confident to very confident. The forecaster would have to hit the “Okay” button after every response to register it to an archived text file.

Another function of the archived cases using WES is the ability for the forecaster to issue a FFW. The warning application (WarnGen) and the questions provided were very similar to that utilized during experimental real-time warning operations. The forecasters also had the ability to issue a follow-up statement, and were heavily encouraged to do so to modify the minor and major flash flood probabilities based on the evolving threat and the updating WoFS QPFs. The Ellicott City event had one warning provided to the participants at the start of the simulation in northeastern Baltimore City and east central Baltimore County, MD. This warning was the operational one issued by NWS LWX and ensured that the forecasters did not have to worry about the ongoing threat at the start of that simulation.

At the end of each simulation, the WESSL script launched a post-brief presentation for the participants. The post-brief presentation included information about the event, including the operational warning decisions, a description of the flash flood events, and the MRMS QPE totals.

Group Discussion

A group discussion was held on Friday morning to garner feedback on the activities that occurred during the week, with an emphasis on the WoFS archived cases. HMT-Hydro Experiment officers and PIs presented focused questions to guide the discussion to gain insight and feedback on the probabilistic grids and the use of short-term QPFs in the flash flood prediction and warning process. The questions asked were as follows:

- You worked three archived cases in total. Reflecting back on these three cases, what are some of your general thoughts on the three events?
- What about the overall procedure (i.e., working independently, using the WESSL interface, types of questions asked)?
- What are your thoughts on the pre- and post-briefing videos?
- What are your thoughts on the information provided in the:
    - 3-hr forecasts of probability of receiving an LSR?
    - 3-hr forecasts of probability of minor/moderate/major flash flooding?
    - 3-hr WoF QPF summary images?
- What do you see as some of the benefits and challenges associated with using these three products in real time?
- How do you compare the importance of these products relative to your use of the deterministic flash flood products?
- How do you envision the WoF-driven guidance being integrated into operational decisions?
- Any other comments?

PQPE Analysis

The 2019 HMT-Hydro Experiment was the first opportunity for participants to view and evaluate the experimental Probabilistic Quantitative Precipitation Estimation (PQPE) product suite (Kirstetter et al. 2015). A one-month data set for June 2017 was provided in a web-based format, yet six specific cases were identified within this period for initial analysis. More details about the PQPE product suite can be found in the next section (Experiment Datasets).

Accompanying this one-month PQPE data set was a survey. The survey provided in the Qualtrics web system first asked which one of the six they were analyzing. Prompts were provided if the participants completed the six cases and identified another event he/she would like to analyze. From there, the participants addressed the following questions and statements:

- How important do you think the uncertainty information is for PQPE with respect to this case?
- What way would you like to see uncertainty information expressed for precipitation rates that would be most beneficial for this case?
- What way would you like to see quantile information expressed for precipitation rates that would be most beneficial for this case?

- Rate the instantaneous exceedance probabilities for understanding the instantaneous rain rate magnitude for the event.
- Rate how the instantaneous exceedance probabilities help your understanding of the likelihood for flooding for this event.
- What way would you like to see instantaneous exceedance information expressed for precipitation rates that would be most beneficial for this case?
- Rate how the 1) number of times and 2) the percent of times that the probability of exceedance > 50% for a given threshold over a given time period helps your understanding of flood risk.
- Rate the importance of the various instantaneous rain rate thresholds that were used to create the number or proportion (percent) of times that a probability of exceedance > 50% for the various rain rates.
- Rate the importance of the various temporal accumulation periods that were used to create the number or proportion (percent) of times that a probability of exceedance > 50% for the various rain rates.
- Please rank the overall utility of the basic product concept for analyzing the precipitation for this event.

The time that was used to analyze the PQPE product suite varied from day-to-day. Participants rotated through the six designated cases during slow real-time operational periods. Leftover time after archived case studies allowed for opportunities to complete case analyses. A two-hour period was also designated on Friday for PQPE analysis; however, the time could also be utilized for other testbed activities, usually subjective real-time evaluations if Thursday consisted of busy real-time operations.

Feedback Survey

At the end of the week, participants filled out an online feedback survey via the Qualtrics web system. The feedback received within this survey was useful in determining what was working during the experiment and what changes could be made either between weeks or for future HMT-Hydro experiments. Results from the feedback survey can be found in Appendix D.

IV. Experiment Datasets

This section details the various gridded fields and observations that were utilized throughout each week of the 2019 HMT-Hydro Experiment and the associated products issued by the forecasters during each week.

Forecast Tools

HMT-Hydro Experiment participants had access to a range of NWS operational products and experimental products during real-time experimental warning operations. The suite of NWS operational products contains products that were available via satellite broadcast network (SBN) from the National Centers for Environmental Prediction (NCEP). This included surface observations and precipitation gauges, observed soundings from the NWS upper-air network of rawinsondes, full Multi-Radar Multi-Sensor (MRMS) product suite, forecast guidance from various numerical and convective-allowing models, and satellite data. The satellite data included base and derived products from the recently implemented GOES-16 (Geostationary Operational Environmental Satellite) satellite. Up to ten individual WSR-88D radars were made available during the operational period. The radars were chosen in regions defined by current or anticipated heavy rainfall and flash flooding.

In contrast, because of the limited availability of WSR-88Ds that could be turned on and the ability to work anywhere in the CONUS, the Flash Flood Monitoring and Prediction (FFMP) and Four-dimensional Stormcell Investigator (FSI) software were not available. Local numerical and convective-allowing models were also not available. Operational products, including FFWs were not viewable; moreover, the forecasters were asked not to look at NWS web pages and other applications that would display operational NWS products.

A separate menu was dedicated to the HMT-Hydro Experiment that included select operational MRMS (Zhang et al. 2016) and Flooded Locations and Simulated Hydrographs (FLASH; Gourley et al. 2017) products for use during operations. A list of available products used in both the real-time operations and archived case studies is provided in Appendix C. The deterministic products focused on the MRMS precipitation products, more specifically the radar-only QPE, the seamless hybrid scan reflectivity, Radar Quality Index (RQI; Zhang et al. 2012), and surface precipitation type. This allowed for participants to have high spatio-temporal resolution radar reflectivities and QPEs across the CONUS to compensate for the limited number of WSR-88D data feeds into the HWT.

There were two sections of the deterministic FLASH product suite provided to participants: Hydrologic model output and QPE comparison products. The featured hydrologic model in FLASH is the Coupled Routing and Excess Storage (CREST) model. It is a fully distributed hydrologic model that provides maximum streamflow, maximum unit streamflow, and soil moisture products to forecasters. Similar data were also provided for the Sacramento Soil Moisture Accounting (SAC-SMA) model and a hydrophobic model. The QPE comparison products has a suite of grids creating a ratio between various MRMS radar-only QPE accumulations and flash flood guidance produced at NWS RFCs (Clark et al. 2014) via a mosaic created at WPC. Another suite of products in this section compares the MRMS radar-only QPEs to precipitation frequencies from NOAA Atlas 14 to compute average recurrence intervals (Perica et al. 2013). NOAA Atlas 14 analyses are not yet available for states in the Pacific Northwest or Texas. Rainfall frequencies were modeled in these regions using a multivariate regression approach, thus enabling the computation of rainfall ARI products in these states.

There were a total of four probabilistic gridded products generated from the FLASH system for the HMT-Hydro Experiment:

- Probability of Receiving a Flash Flood LSR
- Probability of Minor Flash Flooding
- Probability of Moderate Flash Flooding
- Probability of Major Flash Flooding

The Probability of Receiving a Flash Flood LSR product corresponds to a range of unit streamflow values that have been associated with flash flood LSRs. The probability values depicted in this product indicates how frequent the corresponding values of simulated unit streamflow have been associated with flash flood LSRs. A value of 100% means that the corresponding simulated unit streamflow value has always been associated with flash flood LSRs, while a value of 0% means that the corresponding simulated unit streamflow value has never been associated with flash flood LSRs. These probabilities are generated in a post-processing algorithm using probabilistic models that have been trained on historical data. This means that the hydrologic model is integrated in a deterministic way and the resulting output (unit streamflow) is used as input to the probabilistic model. For this product, the flash flood LSR probabilistic model is based on reports from 2005–2011, and uses a logistic regression algorithm.

The Probability of Minor/Moderate/Major Flash Flooding products use three different thresholds of CREST maximum unit streamflow values to represent a different potential magnitude of the flash flood hazard. The determination of these values for the different magnitudes was based on past studies and observations from previous events and HMT-Hydro Experiments. The probabilities are generated in a post-processing algorithm using probabilistic models that have been trained on historical data. This means that the hydrologic model is integrated in a deterministic way and the resulting output (unit streamflow) is used as input to the probabilistic model. The probabilistic model is based on probabilities of observed USGS unit streamflow values conditional to values of simulated unit streamflow with a bias correction based on the USGS observations. This probabilistic model was trained on archived data from 2002–2011.

All work was conducted within the AWIPS-II (Advanced Weather Interactive Processing System) software in the Display 2-Dimension (D2D) perspective. No other perspectives were utilized in the testbed environment.

Probabilistic QPE Products and Tools

The PQPE product suite was broken down into the following categories:

- Expected Precipitation
    - Instantaneous rate with 1, 3, and 6-h accumulations
- Instantaneous Uncertainty of QPE
- Instantaneous Quantiles of QPE
- Instantaneous Probability of Exceeding Rate Values
    - 10, 25, 50, and 100 mm h^{-1}
- Count and Percentage of Times Instantaneous Probability of Exceeding Rate Values
    - For 1, 3, and 6-h periods based on 10, 25, 50, and 100 mm h^{-1} rates

The expected precipitation is what a deterministic QPE product would look like except it is based on a moment derived from the distribution of precipitation. The uncertainty in quantifying precipitation rates is associated with the spread of the distribution of possible precipitation rates given the radar observations. It strongly depends on the type and the magnitude of the measured reflectivity, and can be attributed to error factors in the precipitation estimation process. The probability of exceeding the predefined precipitation rate thresholds compares the distribution of possible precipitation rate values to the given threshold. The count and percentage of times the probability of exceeding these rate values over the various time periods (1, 3, and 6 hours) are based on the number of times that each probability exceeded 50% for each of the two-minute products over said time periods.

The PQPE data was provided via a web display similar to that of the MRMS QPE web interfaces (Figure 3). The participants were able to select the experimental PQPE product plus a select set of operational MRMS products for comparison purposes. The associated survey was conducted in a web browser using the Qualtrics system (see previous section).

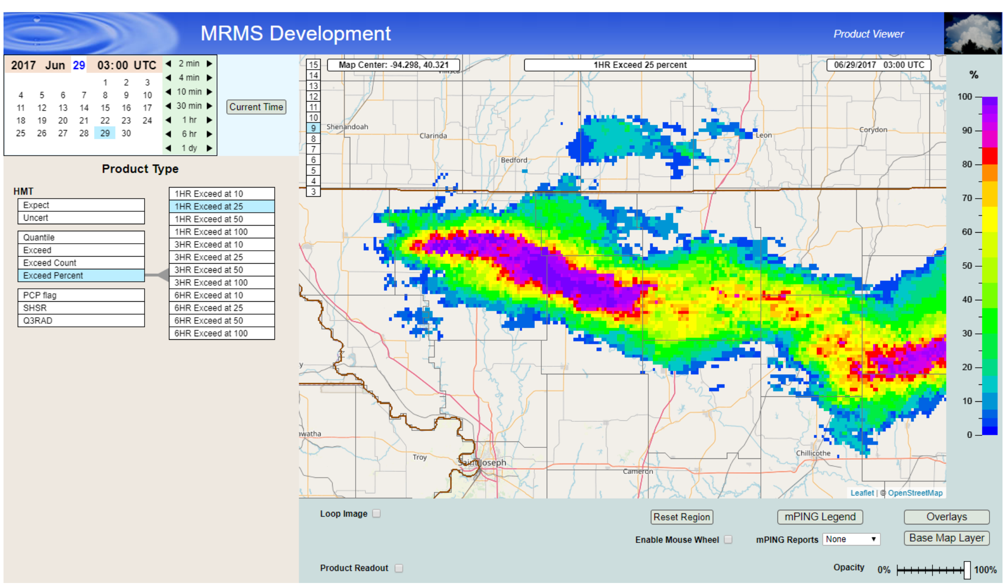


Figure 3. Percent of time over a 1-h period that the instantaneous rate exceeded 25 mm h^{-1} over northern Missouri ending 0300 UTC 29 June 2017 as seen from the web viewer utilized by HMT-Hydro participants.

Observations

Three separate sources of flash flood and flood observations were available to participants and officers during the experiment:

- Local Storms Reports (LSRs) gathered by local NWS Weather Forecast Offices (WFOs)
- Automated stream gauge measurements collected by the United States Geological Survey (USGS)
- Unsolicited public geolocated smartphone or mobile devices reports from the mPING (Meteorological Phenomena Identification Near the Ground) product run by NSSL and OU (Elmore et al. 2014)

NWS LSRs are issued during or immediately after a given hazardous weather event (Horvitz 2012). These reports include the date and time of the event, the city and county of the event, the type of event, the source of the report, and the location in decimal degrees. Flood and flash flood LSRs typically include a short description of the exact impact of the reported event in plain English.

Closely related to NWS LSRs are reports in the NWS publication *Storm Data* (MacAloney 2007). In contrast to LSRs, they can contain a range of times and a spatial range (a series of latitude/longitude points versus a single point in a LSR). *Storm Data* reports are generally correlated with LSRs, but there are situations when a flash flood event only comes to light days after an event, and thus, is absent from the LSR database but present in the *Storm Data* database.

USGS stream gauges are located on catchments of various sizes across the United States. In order to quantify for inclusion in this observation database, a flash flood event recorded at a stream gauge must exceed the NWS-defined minor flood stage for the gauged location or the USGS-defined two-year return period for the gauged location and satisfy a requirement for a quick time-of-rise (0.90 m h^{-1}) of the stage (Cosgrove 2014, personal communication). Only stream gauges with contributing drainage areas of less than 2000 km^2 were considered.

The mPING project uses the recent proliferation of GPS-enabled smart phones and other mobile devices to crowd-source surface weather conditions (Elmore et al. 2014). Users can identify the relative severity of an observed flood or flash flood using a 1-4 integer scale defined by the following:

1) River/creek overflowing or cropland/yard/basement flooding
2) Street/road flooding or closure; Vehicles stranded
3) Homes/buildings with water in them
4) Homes/buildings/vehicles swept away

NWS local storm reports were rated in a similar fashion by HMT-Hydro Experiment officers. Any report will be used to validate a minor flood, while a report rated as a 3 or 4 is required for a major flood. The major flood category also includes personal impacts such as rescues, evacuations, injuries, and fatalities. If a flood is captured by a USGS stream gauge, then the reported flood stage can be used to validate the magnitude associated with the warning. The experimental coordinators will also examine social media and local news stations for reports that are informative to the validation process.

Issued Products

In common NWS parlance, "product" refers to a text message disseminated by an operational unit of the agency. Common products include watches, warnings, and advisories. In this report, two types of products are considered: operational flash flood warnings and experimental flash flood warnings.

Operational flash flood warnings (FFWs) are issued for "storm-term events which require immediate action to protect life and property" (Clark 2011). Warnings are polygons that can be drawn independent of county or other political boundaries. They can be issued for multiple causative factors; however, in the context of the HMT-Hydro Experiment, flash flood events caused by heavy rainfall are of chief interest. FFWs are issued by local WFOs and therefore cannot cross County Warning Area (CWA) boundaries. They are created in these WFOs by an add-on application to AWIPS-II called WarnGen. The forecaster draws a polygon with as many vertices as needed to accurately encompass the threat. Based on this polygon, the WarnGen application determines which counties and locations should be in the warning text, produces the appropriate text, and then disseminates the warning.

Experimental FFWs work similarly to their operational counterpart but with some important differences. In the testbed environment, participants were told to act as a national forecast office (i.e., the participants were responsible for monitoring conditions for and predicting potential flash flooding across the entire CONUS). The participant would have to localize their AWIPS to the CWA that requires the warning issuance. If the participant changes geographic locations and then decides to issue a warning, the participant would have to close their AWIPS and re-open it to re-localize to the corresponding CWA.

The investigators used a modified version of WarnGen that required forecasters to quantify their uncertainty about the magnitude of flash flooding expected within the warning polygon. The probability of minor flash flooding (corresponding to mPING impact classes "1" and "2") ranged from 10–100%, and the probability of major flash flooding (corresponding to mPING impact classes "3" and "4") ranged from 0–100%. The flash flood probability values were available in ten-point increments for both minor and major probabilities. Although forecasters could identify a variety of valid lengths for their experimental warnings (ranging from one to six hours), the default warning length of three hours was set in WarnGen. A total of 83 experimental FFWs were issued during the three weeks of the 2019 HMT-Hydro Experiment.

WarnGen was further modified to survey the participant about their warning decision making process (Figure 4). The following questions were presented to the participants while they are issuing an experimental FFW:

- Rate the influence of each probabilistic product on your warning decision:
    - Significant Influence
    - Somewhat Influence
    - Little Influence
    - No Influence
- Discuss if your flash flood probability value(s) deviated from the gridded products. Please provide some reasoning why:
    - [Text response]
- Additional notes about the FFW issuance:
    - [Text response]

The participants had the option to issue follow-up statements (FFS) to their experimental FFWs. The WarnGen application had the same questions and prompts for FFSs as those presented when issuing the initial FFW. Only ten FFS were issued during the three-week period of the 2019 HMT-Hydro Experiment. All FFWs and follow-up statements were archived as text products for evaluation by HMT-Hydro Experiment officers.

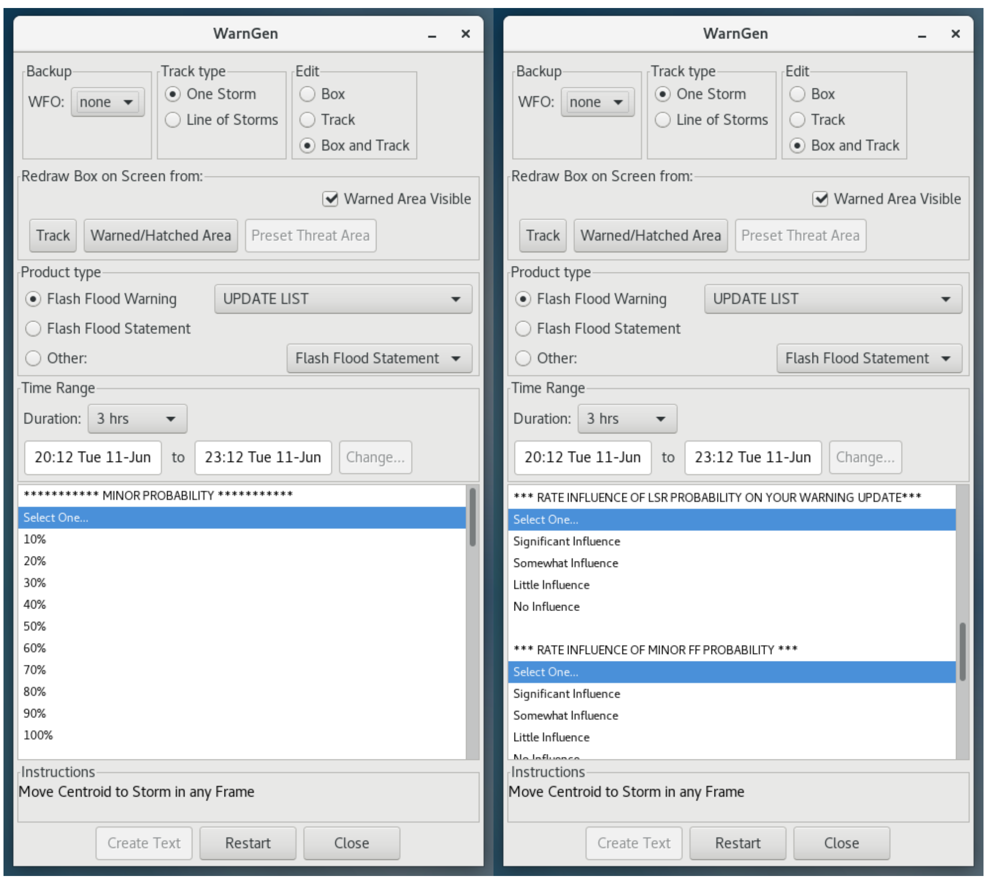


Figure 4. WarnGen GUI used to create experimental FFWs. WarnGen was modified to survey participants about their warning decision making process using the experimental probabilistic flash flood products.

V. Results: Real-Time Operations

As described in Section III, the subjective evaluations of the real-time experimental warning operations were broken down into two sections: 1) the experimental products and 2) the subsequently issued experimental FFWs. This section of the final report will look at the subjective evaluation results with some objective results and warning statistics as well.

Subjective Product Evaluations

A verified flash flood event was chosen after each real-time experimental warning operations. Each event was chosen based on its significance and the availability of LSRs. The event did not have to occur during the real-time experimental warning operations, but events that did occur during operations were favored. A total of eight events were analyzed during the three week period. There were additional evaluations that occurred after two real-time operation session given the evolution of the weather event, which allowed the time for the extra evaluations. The events that were analyzed varied in severity, ranging from more nuisance-type events to significant urban events, like Washington D.C. flash flood event that occurred before daily HMT-Hydro operations on 8 July 2019 (Figure 5). Using the available observations, it was determined by HMT-Hydro Experiment officers and participatns if the flash flood event was a minor or major flash flood event. This was based on the criteria outlined in Section IV.

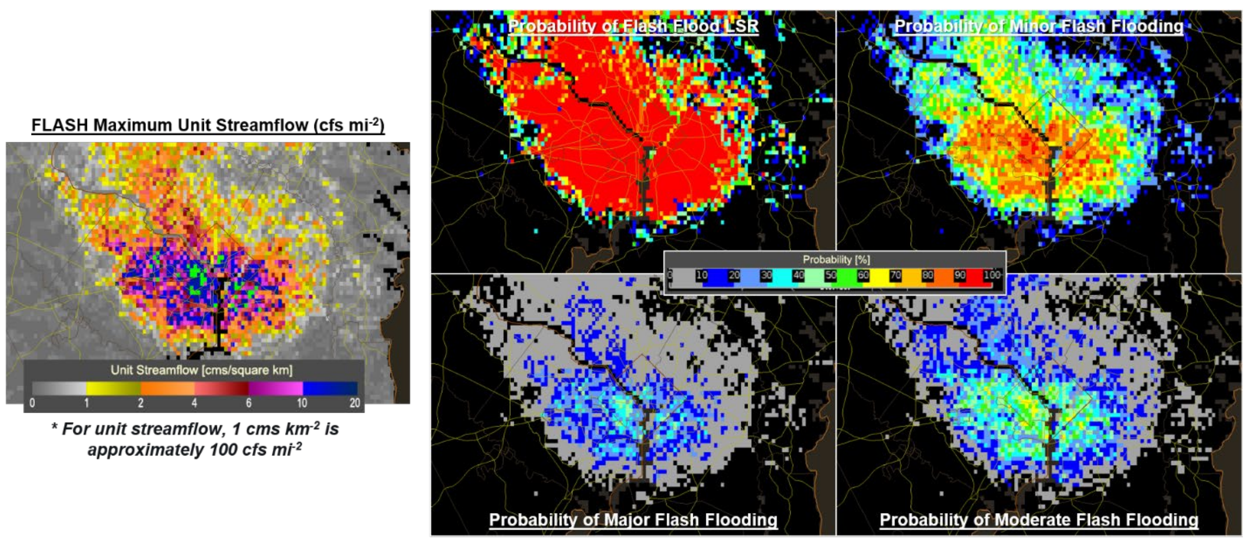


Figure 5. The Washington D.C. flash flood event as seen from the deterministic (left) and probabilistic (right) data at 1400 UTC 8 July 2019. The images were taken from the flash.ou.edu web page, which was used during the evaluations. Note that the units for the FLASH CREST Maximum Unit Streamflow product is in metric units ($m^3\ s^{-1}\ km^{-2}$).

Participants first focused on the spatial coverage of the FLASH CREST Maximum Unit Streamflow product and the four experimental probabilistic products (Figure 6). Using the FLASH CREST Maximum Unit Streamflow product as the control data, most probabilistic products were similarly ranked and had similar trends compared to the 2018 analysis; however, one deviation from this was the lower ranking of the Probability of Major Flash

Flooding. This was on average ranked ten points or more lower than the other probabilistic products with an average score of 67.1. It also had the greatest standard deviation value (18.2), which was more than five points higher than the next highest value. Participants noted that there were times it was hard to analyze given events where significant flash flooding did not occur. One forecaster noted that it was "hard to quantify the moderate and major products spatially because it did not have the values" to analyze. There were other instances of the spatial distribution of this and all probabilistic products that influenced the product rankings.

Of all of the probabilistic products, the Probability of Minor Flash Flooding was scored the highest with a nearly equal average ranking compared to the control FLASH CREST Maximum Unit Streamflow product; moreover, it even had the lowest standard deviation value compared to all products.

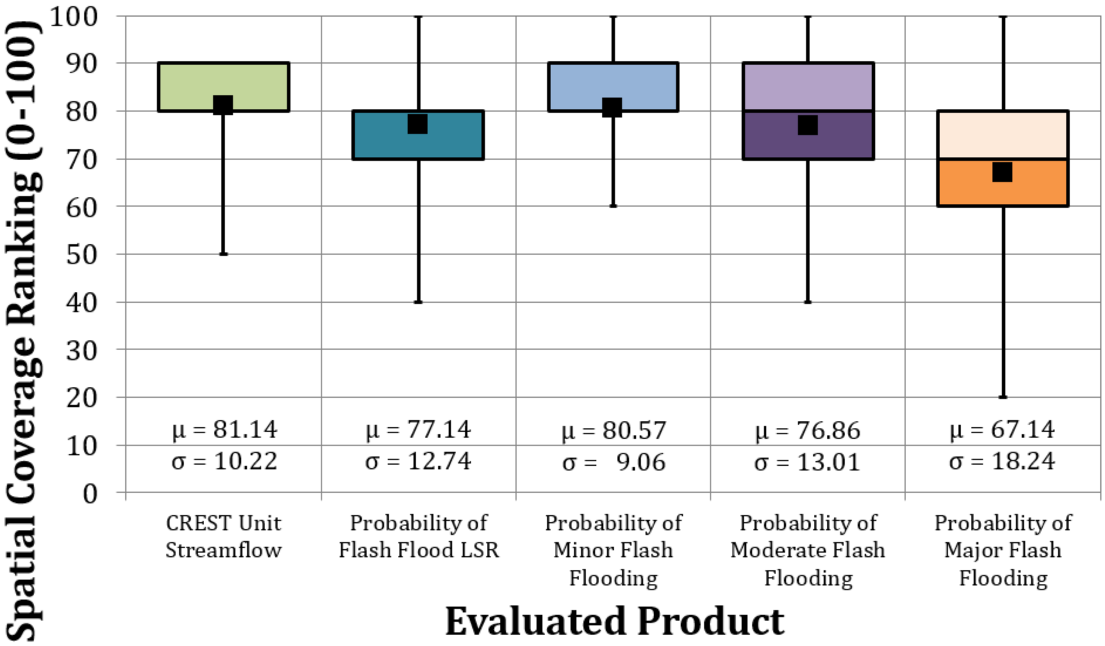


Figure 6. Subjective ranking of the spatial coverage of the FLASH CREST Maximum Unit Streamflow product and the four experimental probabilistic products when compared to verified flash flood events using a box-and-whisker plot. The top (bottom) of each box represents the 75th (25th) percentile with the line in the middle of each box representing the median subjective ranking value. The top (bottom) whisker represents the maximum (minimum) ranking. The black dot represents the mean subjective ranking. The mean (μ) and standard deviation (σ) values for each product are shown below each box-and-whisker plot.

The subjective evaluations of the probability magnitudes for the four probabilistic flash flood products had varying results. The FLASH CREST Maximum Unit Streamflow product was evaluated again as the control dataset. Approximately 66% of the evaluation responses ranked this product as "About Right" with a nearly equal distribution on either side of that response (Figure 7).

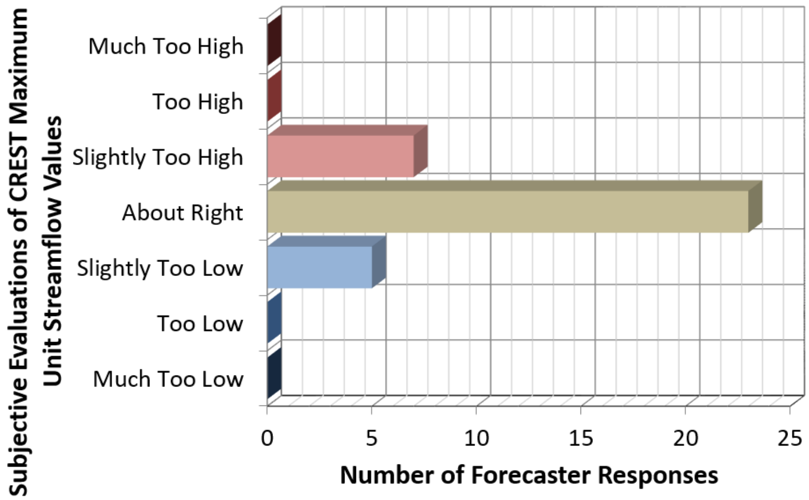


Figure 7. Subjective evaluation of the magnitude of values for the FLASH CREST Maximum Unit Streamflow product.

The evaluation of the four probabilistic products yielded different results for each product (Figure 8). The magnitude of the Probability of Receiving a Flash Flood LSR was generally perceived as higher than expected, while the Probability of Minor Flash Flooding was generally perceived as lower than expected. The perceptions of the Probability of Moderate and Major Flash Flood products were more centric but had significant spread across the categories, notably with the Major flash flood product.

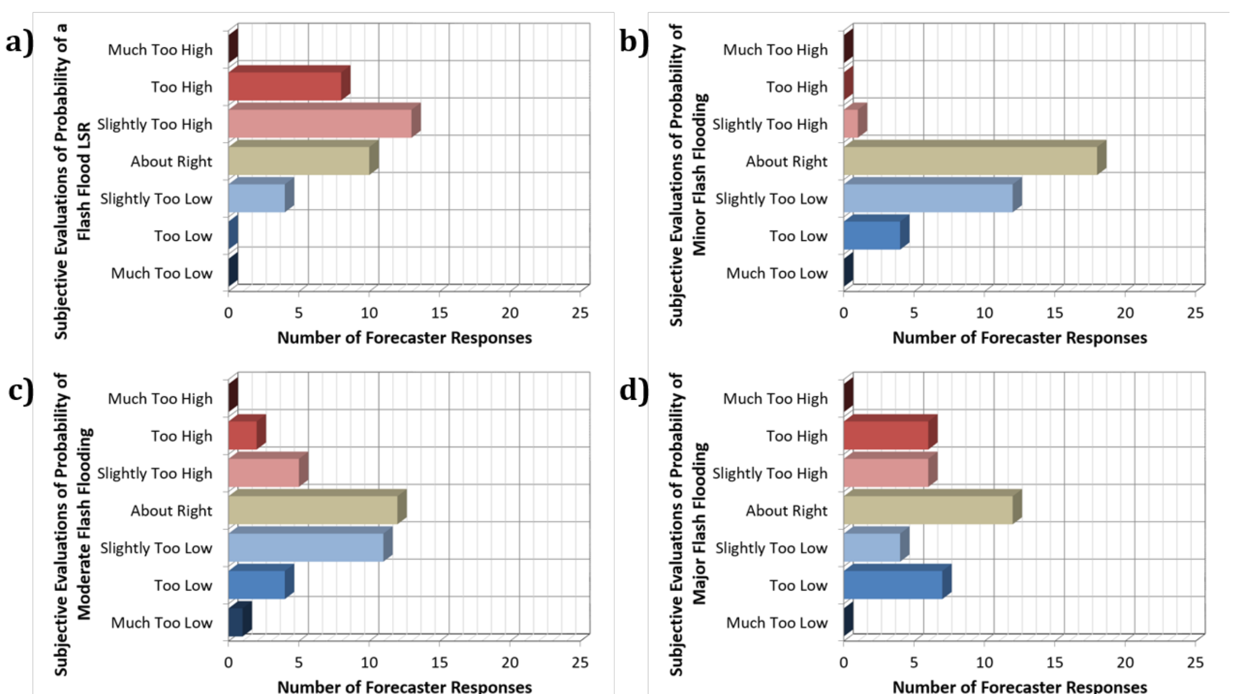


Figure 8. Subjective evaluation of the magnitude of values for the a) Probability of Receiving a Flash Flood LSR product, b) Probability of Minor Flash Flooding, c) Probability of Major Flash Flooding, and d) Probability of Major Flash Flooding.

Similar to the 2018 HMT-Hydro Experiment, the most striking difference amongst the probabilistic products were the perception of the biases in the Probability of Receiving a Flash Flood LSR and the Probability of Minor Flash Flooding products. Both of these products were generally considered as proxies for determining the predictability of flash flooding. The probabilities for the Minor/Moderate/Major flash flood product suite used a bias correction technique during post-processing, which determined that the deterministic unit streamflow values were too high when compare to observed unit streamflow values derived by USGS observations. The 2019 experiment adjusted the product that is now the Probability of Minor Flash Flooding and saw a considerable change in the bias perception of the product. While still biased towards slightly too low, the majority of responses felt that the product was perceived as about right.

There were also considerable differences between the Probability of Moderate/Major flash flooding results when assessed against the 2018 counterparts; however, this can be attributed to the lack of major flash flood events during the 2018 experiment versus a number of major events being analyzed during the 2019 experiment. The large spread of responses is likely based on the various events that were analyzed in the subjective evaluations. Further research would need to be conducted to analyze how the products were received during different levels of flash flood severity.

The follow-up discussions to these spatial and product magnitude evaluations yielded that the participants were using all available data and not just one preferred product to make their warning decisions. And while the attention was more focused on the Probability of Receiving a Flash Flood LSR and the Probability of Minor Flash Flooding products, some forecasters noted that they were using the Probability of Moderate/Major Flash Flooding to determine the flash flood potential. This was similar to the discussion had in the 2018 experiment, yet the emphasis on the Moderate/Major products in the 2019 experiment was not as prominent.

Subjective FFW Evaluations

A total of 18 experimental FFWs were subjectively analyzed during the group evaluation sessions. These experimental FFWs were collocated with verified flash flooding. Most flash floods were verified with NWS local storm reports (LSRs); however, there were a few instances where USGS streamflow gauges verified flash flooding. Two of the 18 evaluated experimental FFWs did not have an associated operational FFW.

The evaluation of the experimental FFWs compared to the operational FFWs were split into two categories for the 2019 HMT-Hydro Experiment: 1) spatial coverage and 2) timing. Starting with the spatial coverage, there were no conclusive trends in the subjective comparisons, though more responses had a more favorable view of the experimental FFWs than the operational FFWs (Figure 9). Some comparisons considered the spatial extent of both FFWs, while acknowledging that some operational FFWs were more constrained in area due to local knowledge of the terrain and flashiness that the participants did not have. Other noted how the location of the flash flood threat within the FFW polygons (more centered within the polygon area versus near the polygon border).

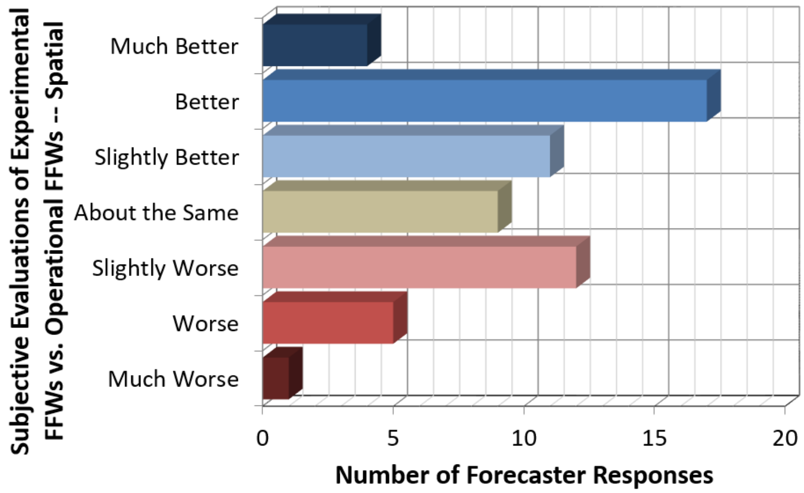


Figure 9. Subjective evaluation of the spatial extent of experimental FFWs when compared to collocated operational FFWs for areas with verified flash flooding.

In regards to the timing of the warnings, there were also variabilities amongst the responses (Figure 10). Only 25% of the responses felt that the timing of the warnings were similar. For the instances when the timing of the experimental FFWs were after the operational FFWs, they were perceived as "Slightly Worse." When looking at experimental FFWs that were issued ahead of the operational FFWs, they were generally perceived as "Better" or "Much Better." These results are also dependent on the events that were evaluated.

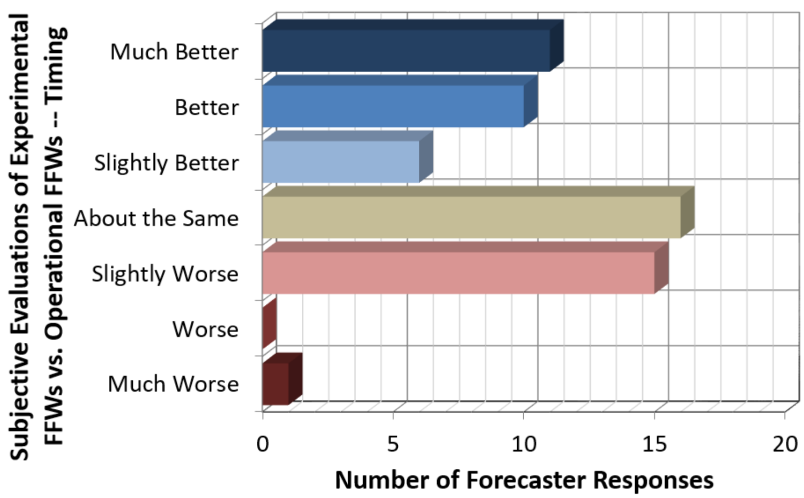


Figure 10. Same as Figure 9 except for the timing of FFWs.

The participants analyzed the probability values assigned within the experimental FFWs. The average minor flash flood probability value in the evaluated experimental FFWs was 73.3% and the average major flash flood probability was 16.7%. The participants evaluated the minor flash flood probabilities as being generally "About Right" (Figure 11). The major flash flood probabilities had some bias towards being "Slightly Too High."

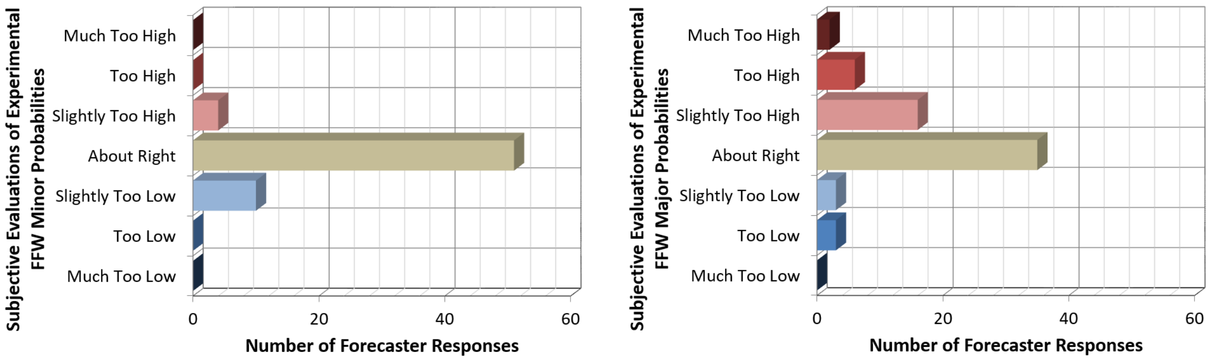


Figure 11. Subjective evaluation of the assigned minor (left) and major (right) flash flood probabilities in the experimental FFWs containing verified flash flooding.

A total of 83 experimental FFWs were issued throughout the 2019 HMT-Hydro Experiment. A total of 31 of the experimental FFWs were verified for flash flooding while 52 experimental FFWs went unverified. The average minor (major) flash flood probabilities for the verified warnings were 74.5% (15.8%). The average minor (major) flash flood probabilities for unverified warnings were 67.1% (12.9%). The 7.4% spread in the minor flash flood probabilities for verified and unverified warnings was a noticeable change from the results of the 2018 experiment, where the difference in minor probabilities between the verified and unverified warnings was only 1%. It is hypothesized that the change in the Probability of Minor Flash Flooding product influenced the probabilities assigned. The spread in the major probabilities from verified to unverified warnings was similar to the 2018 experiment.

A reliability diagram of all experimental FFWs showed that the participants generally over-predicted flash flooding (Figure 12). Probabilities that had a minimum sample size of five were plotted for both the minor and major values assigned to the experimental FFWs. None of the five warnings that were assigned a minor probability of 50% were verified. The percent of verified events for the greater minor probability values increased with increasing probabilities, with the best reliability seen at the 80% minor probability value (62.5% verification rate). Moreover, there is a significant year-to-year improvement in the verification of warnings that had minor probability values of 80% and 90%. The major probability values also saw a marked improvement over 2018, partially from the fact that major flash flood events occurred in 2019 (versus none in 2018). While there was still an over-prediction of major flash flooding, the major probability values at 10% and 20% were reasonable and near the ideal value. The best result came from the major probability value of 20%, which had an observed verification rate of 12.5%.

It should be noted that the verification was conducted based on reports garnered by the local NWS offices, meaning that they would look for verification within their warning areas and not those considered by the HMT-Hydro Experiment. This would impact the verification statistics and the subsequent reliability plot.

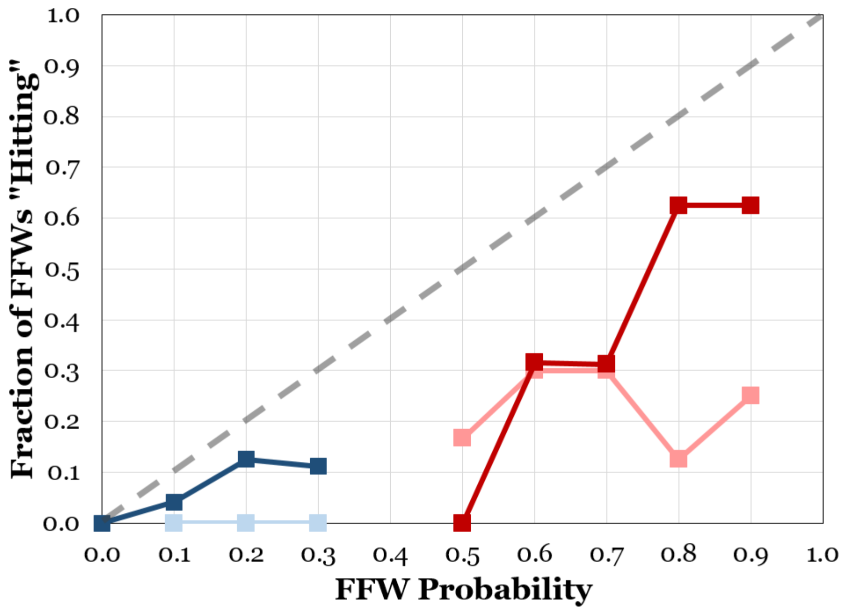


Figure 12. Objective assessment of the reliability of experimentally-issued flash flood warnings for major (blue) and minor (red) flash flood events for the 2019 experiment (dark shade) compared to the 2018 experiment (light shade).

Participants provided data regarding their warning decision making process in their issuance of their experimental FFWs through the modified WarnGen GUI (see Section IV). When asked about ranking the influence of the various probabilistic products in their warning decision making process, the product with the greatest influence was the Probability of Minor Flash Flooding (Figure 13). 73.5% of the experimental warnings were issued with that product having a "Significant Influence" on the decision to warn. The Probability of Receiving a Flash Flood LSR had the next highest percentage of "Significant Influence" responses at 43.4%. The products assessing the potential for moderate and major flash flooding had a lesser influence in the process, with "Somewhat Influence" being the dominant response for the Probability of Moderate Flash Flooding and "Little Influence" being the majority response for the Probability of Major Flash Flooding.

When separating the experimental FFWs based on verification (Figure 14), the different probabilistic products had somewhat similar rankings of influence with the Probability of Receiving a Flash Flood LSR and the Probability of Minor Flash Flooding products. It should be noted that the Probability of Minor Flash Flooding product did not have a single influence selection less that "Somewhat Influence." A greater percent of responses for the Probability of Moderate/Major Flash Flooding products went to the "Little Influence" category. The analysis of the unverified warnings showed that a little bit more "Significant Influence" was recorded to the Probability of Receiving a Flash Flood LSR product compared to the verified FFWs (48.1% to 35.5%). The Probability of Moderate/Major Flash Flooding products did have a greater percentage of "No Influence" in the unverified warnings but also had a greater percent of "Significant Influence" as well.

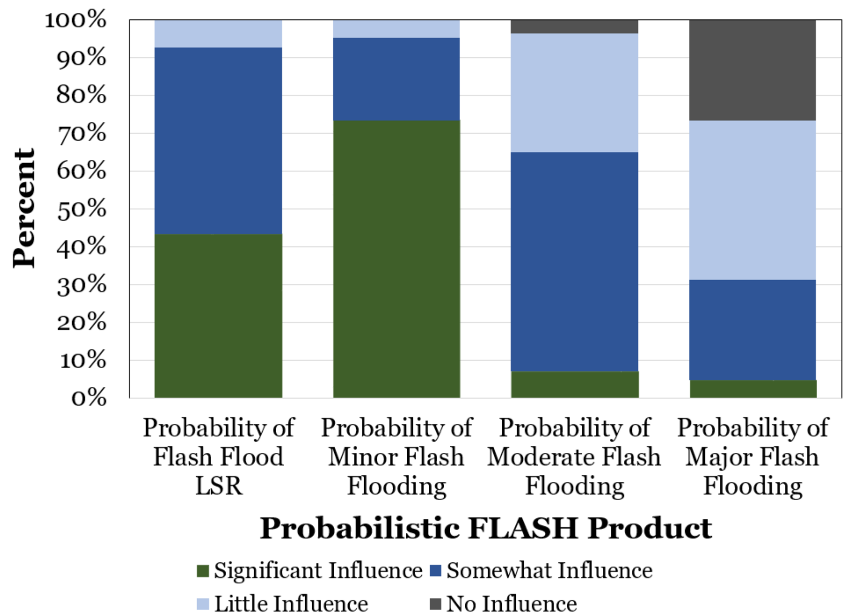


Figure 13. Percent of responses regarding the influence of each probabilistic FLASH product on the warning decision for all warnings.

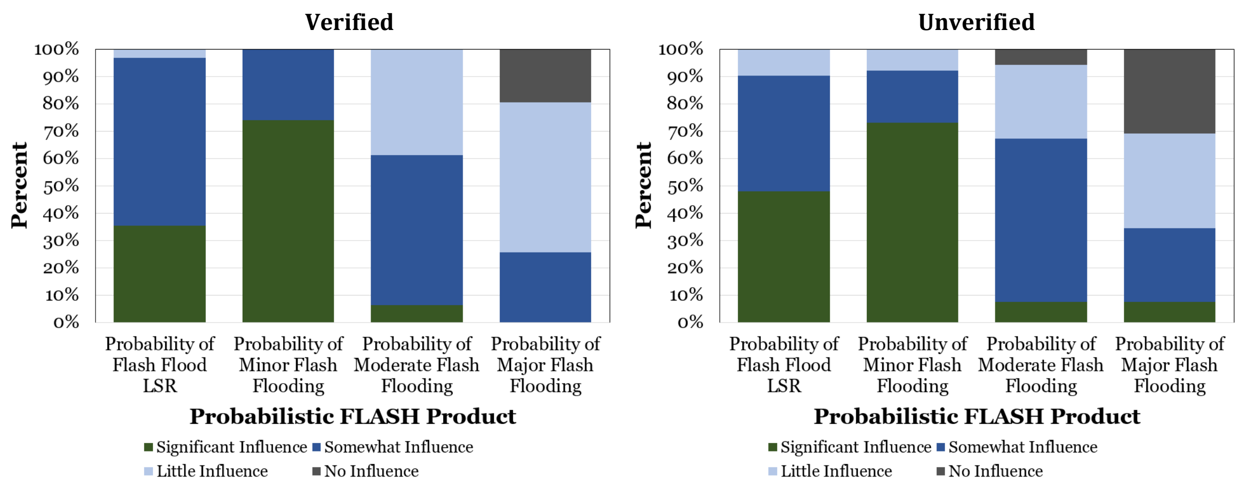


Figure 14. Percent of responses regarding the influence of each probabilistic FLASH product on the warning decision for (left) verified warnings and (right) unverified warnings.

Other Warning Statistics

Two other aspects of the experimental FFWs were analyzed: 1) changes in warning lead time and 2) changes in warning area. Twelve of the experimental FFWs had an increase in warning lead time compared to the operational FFWs. One warning had two different LSRs in different operational FFWs and had two different lead time values for this analysis. The 12 experimental FFWs that had a positive increase in lead time had an average lead time of 65.8 minutes, while the corresponding operational FFWs had an average lead time of 39.3 minutes (Table 1). Three warnings had an increase of lead time > 30 minutes, with the OKX warning issued on 11 July having an 87-minute increase in lead time.

Table 1. List of the 12 experimental FFWs that had an increase in lead time over the collocated operational FFWs. Listed are the location (CWA) of each warning, the time of the first flash flood LSR, the time of the experimental and operational warnings, and the difference in lead time.

FFW ID	CWA	Date	LSR Time (UTC)	Experimental Time (UTC)	Experimental Lead Time (min)	Operational Time (UTC)	Operational Lead Time (min)	Lead Time Diff (min)
1	EWX	24 June	2105	2014	51	2052	13	38
18	LZK	8 July	2326	2316	10	2331	-5	15
23	RLX	11 July	1800	1603	117	1613	107	10
26	PBZ	11 July	1710	1635	35	1653	17	18
			1800	1635	85	1722	38	47
27	PBZ	11 July	1800	1654	66	1722	38	28
32	CTP	11 July	2045	1917	88	1918	87	1
34	PHI	11 July	2025	1917	68	1945	40	28
42	LWX	11 July	2121	2021	60	2040	41	19
51	OKX	11 July	2245	2116	89	2243	2	87
72	ILN	15 July	2230	2107	83	2131	59	24
77	MPX	16 July	2152	2126	26	2129	23	3
81	LSX	17 July	0100	2348	77	0014	51	26
					65.8		***39.3***	***26.5***

The analysis of the warning areas had three constraints. First, the warnings analyzed had to occur over a verified flash flood event. Second, the events had to generally be isolated. Complex, multi-warning events with widespread flash flooding were not considered (example, central Nebraska on 8 July 2019) due to challenges in calculating the warning areas and/or the event continued beyond the operational period. And finally, with the isolated events, multiple operational and/or experimental warnings can be used in the calculation. This is because the multiple operational FFWs were also isolated to this event and were issued within minutes of each other and were thus influenced by similar observations. One example of multiple experimental FFWs is the event over the RLX CWA on 11 July 2019 (Figure 15) where the two experimental FFWs were issued 14 minutes apart. One example of multiple operational FFWs is the event over the PHI CWA on 17 July 2019 (Figure 16). This situation had all operational FFWs issued between 2306 UTC and 2336 UTC with the experimental FFW issued at 2321 UTC.

Given the aforementioned constraints, eight experimental FFW events shown to have a smaller warning area than that of the operational FFW events (Table 2). 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.

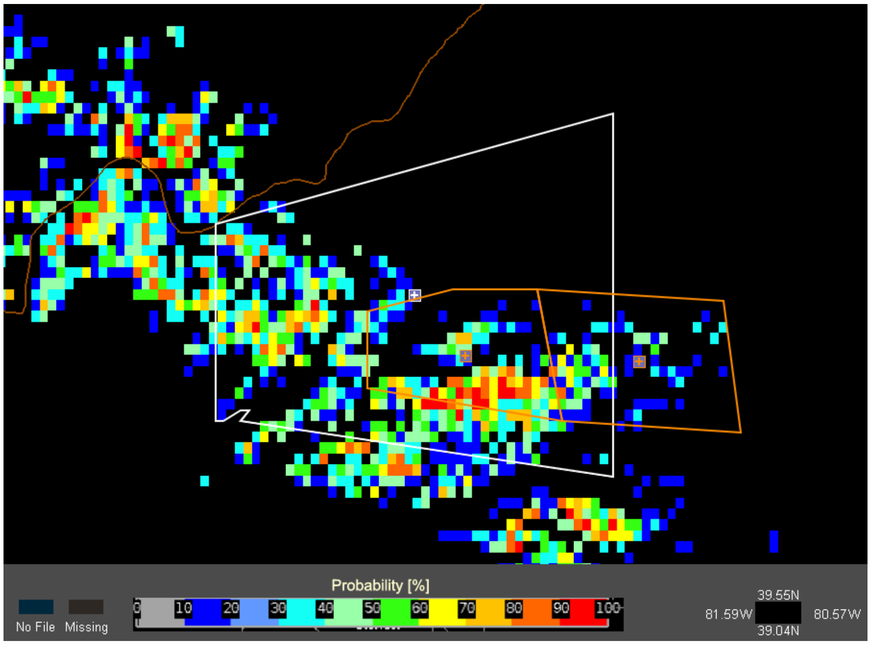


Figure 15. Verified isolated flash flood event over the RLX CWA on 11 July 2019 where the area calculations utilized two experimental FFWs (orange polygons) compared to the single operational FFW (white polygon).

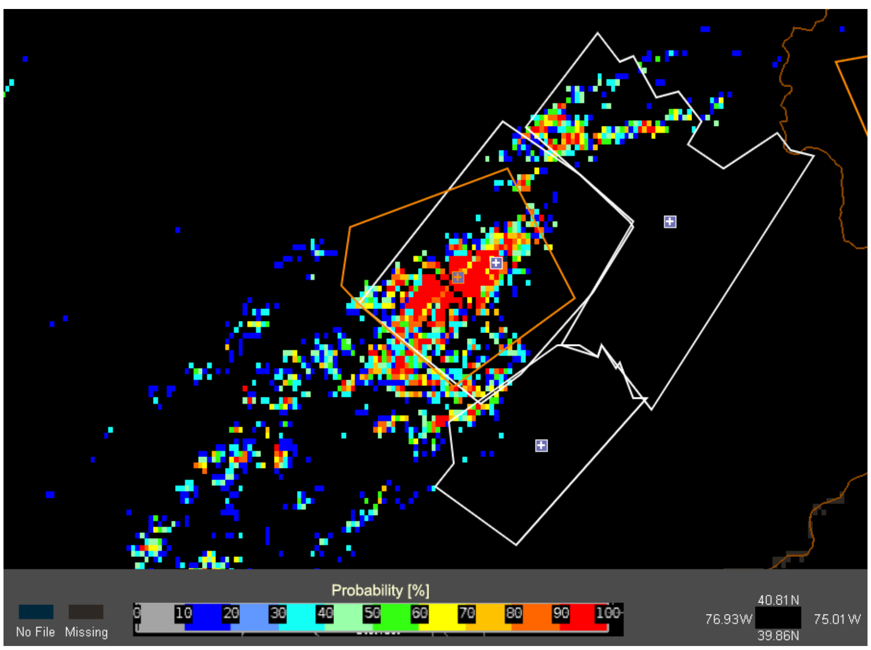


Figure 16. Verified isolated flash flood event over the PHI CWA on 17 July 2019 where the area calculations utilized one experimental FFW (orange polygon) compared to the three operational FFWs (white polygons).

Table 2. List of events where the experimental warning area was smaller than that of the operational warning area. Listed are the events by date and CWA, the number of warnings involved in each calculation, the experimental and operational warning area, and the percent reduction of warning area.

			Warning Count		Warning Area (km^2)		
ID	**CWA**	**Date**	**Experimental**	**Operational**	**Experimental**	**Operational**	**Percent Reduction**
18	LZK	8 July	1	1	165.6	689.9	76.0
23	RLX	11 July	2	1	449.5	1139.8	60.6
32	CTP	11 July	1	1	263.2	728.7	63.9
42	LWX	11 July	1	3	1602.3	2780.9	42.3
61	LCH	15 July	1	1	5728.5	12215.5	53.1
72	ILN	15 July	2	1	961.1	1027.8	6.5
79	PHI	17 July	1	3	1077.7	3887.6	72.3
82	PHI	18 July	1	1	1312.5	2170.5	39.5

VI. Results: Archived Case Evaluations

A number of trends and decision-making characteristics were found throughout the evaluation of the three displaced real-time archive case studies using WoFS. This section here will focus on the warning issuance during the three archived cases and trends with the assigned minor and major probabilities throughout the event. A more detailed set of results along with analysis of the evaluation and decision-making questions presented every half-hour during the cases will be discussed in the upcoming Warn-on-Forecast report.

Warning Lead Times

There were a total of eight locations between the three cases that were evaluated with respect to warning lead time (Table 3). Multiple local storm reports (LSRs) were received at a number of these locations; thus, the focus will be on the first flash flood LSR that was received for each local area of interest. There were instances when a flood LSR was the first LSR recorded; however, those were not used in the statistical analysis.

Table 3. List of flash flood events analyzed for experimental warning lead times using WoFS QPFs in the probabilistic FLASH products.

CWA	Location	Date	Flash Flood LSR Time
LWX	Ellicott City, MD	27 May 2018	2034 UTC
LWX	Baltimore, MD	27 May 2018	2055 UTC
DMX	Story Co., IA	1 July 2018	0045 UTC
DMX	Dallas Co., IA	1 July 2018	0100 UTC
DMX	Polk Co., IA	1 July 2018	0111 UTC
DMX	Blackhawk Co., IA	1 July 2018	0120 UTC
FSD	Gregory Co., SD	13 July 2018	0000 UTC
FSD	Douglas Co., SD	13 July 2018	0124 UTC

Each event had a sample size of 11 participants that conducted the case evaluations; however, a technical error reduced the recorded sample size for the LWX case down to ten. Using the first available flash flood LSR for each identified location, all but one event (Dallas Co., IA) had a positive operational FFW lead time.

With the introduction of the probabilistic FLASH data coupled with the WoFS QPFs, seven of the eight events had a positive increase in warning lead time (Table 4) for warnings that were issued prior to the first flash flood LSR. The exception to this was the Douglas Co., SD event for the FSD case, where the average experimental FFW lead time was similar to the operational lead time. 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, and the Baltimore, MD event had the average warning lead time increased by over an hour to an average lead time of 71.6 minutes. 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.

Table 4. List of the operational FFW lead time, the average experimental FFW lead time, and the difference between the operational and average experimental FFW lead time.

CWA	Location	Ops FFW Date/Time	Operational Lead Time (min)	Experimental Average Lead Time (min)	Difference (min)
LWX	Ellicott City, MD	2026 UTC 27 May	8	38.38	+30.38
	Baltimore, MD	2050 UTC 27 May	5	71.60	+66.60
DMX	Story Co., IA	0033 UTC 1 July	12	22.22	+10.22
	Dallas Co., IA	0116 UTC 1 July	−16	32.70	+48.70
	Polk Co., IA	0102 UTC 1 July	9	42.00	+33.00
	Blackhawk Co., IA	0042 UTC 1 July	38	53.67	+15.67
FSD	Gregory Co., SD	2331 UTC 12 July	29	37.80	+8.40
	Douglas Co., SD	0023 UTC 12 July	61	60.45	−0.55

Event Timelines

A timeline for each area of interest was created to map the various warning decisions. Each timeline encompasses the entire case simulation period. All FFW and FFS issuances are plotted along with the minor and major probabilities assigned to each product. The operational FFW and FFS (if available) issuances were also plotted. Both flood and flash flood LSRs were plotted on the timeline, since all were available to the forecasters during the simulation. While the flood LSRs were not used for the statistical verifications, their existence could have driven both the experimental and operational warning decisions.

The LWX event was arguably the most successful case with respect to the use of WoFS QPFs and warning lead time improvements. Starting with the Ellicott City, MD event, the operational FFW was issued at 2026 UTC immediately following a flood LSR at 2023 UTC (Figure 17). Eight participants issued an experimental FFW before the operational FFW, with the time difference ranging from 2 minutes to 56 minutes. Most experimental FFWs were issued between 1940–2000 UTC. The two participants that missed the event did have nearby warnings out; however, those warnings did not extend far enough west. The edge of both warnings were within five kilometers of Ellicott City. Four participants were able to issue a FFS prior to the flash flood LSR at 2034 UTC. Most minor probabilities prior the flash flood LSR ranged from 70-90% with major probabilities between 10–60%. These values increased with the FFSs in response to the LSR in Ellicott City, MD.

The Baltimore, MD event was the most successful event amongst all three of the archived cases. All ten recorded FFW responses from participants occurred before the flash flood LSR and the operational FFW (Figure 18). Six of the ten warning decisions came in an eight-minute window between 1945–1953 UTC. The earliest warning decision came from Forecaster #2 at 1916 UTC, nearly two hours before the flash flood LSR. The last warning decision came from Forecaster #6 at 1958 UTC, which is still 52 minutes before the operational FFW and 57 minutes before the flash flood LSR. The minor and major probabilities were similar to the Ellicott City event at initial warning issuance, but forecasters were confident to issue probabilities of 100% minor flash flooding at 50–80% major flash flooding prior to the first report.

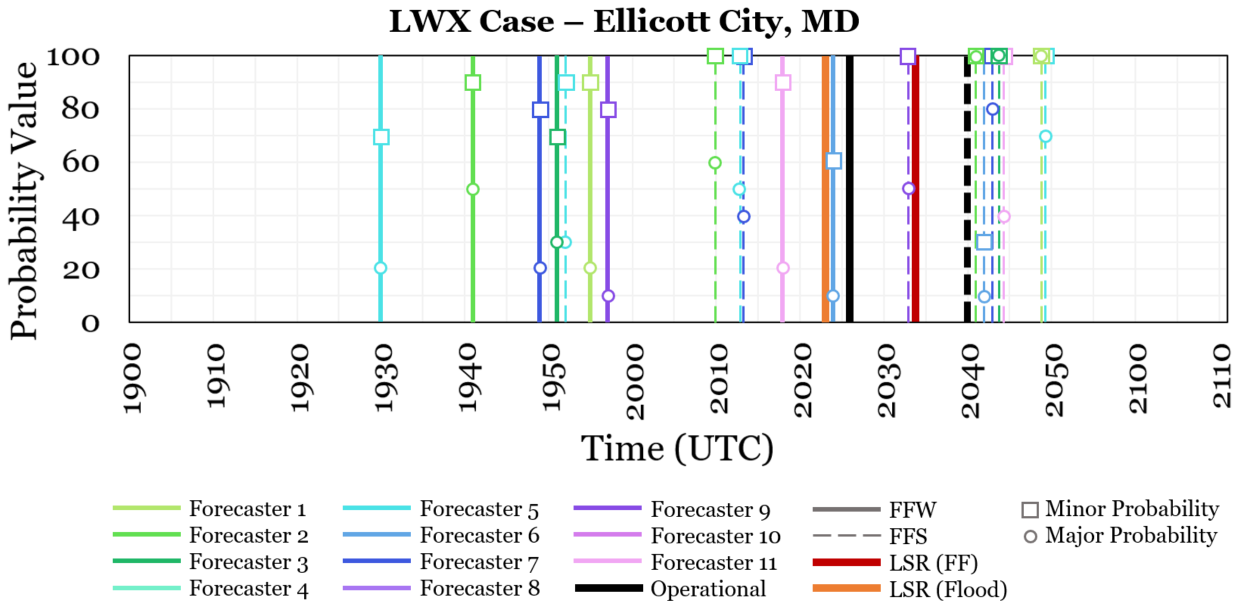


Figure 17. Timeline of all of the products issued along with LSRs for the Ellicott City, MD event in the LWX case. Forecaster experimental FFW (FFS) issuance are denoted by solid (dashed) lines in cool colors. The assigned minor (major) probabilities for each experimental FFW and FFS are denoted by a square (circle) at each time. The operational FFW (FFS) are 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.

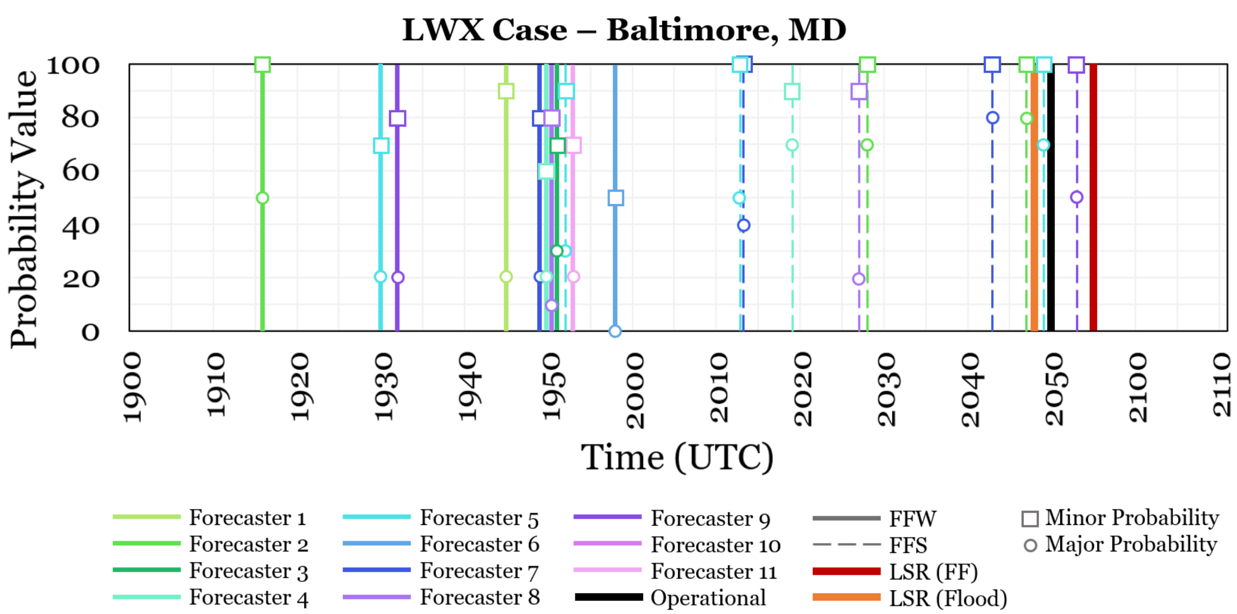


Figure 18. Same as Figure 17 except for Baltimore, MD in the LWX case.

The FSD event had three areas of interest for the participants: the two areas with verified flash flooding (Gregory and Douglas Counties in SD) and a null event in southwest Minnesota. Starting with the null event in southwest Minnesota, eight of the 11 participants placed a warning in this region (Figure 19). All of the experimental warning decisions from the eight participants who issued a FFW occurred in a 27-minute period from 2226–2253 UTC 12 July

2018. Six of the eight experimental FFWs were issued after the operational FFW (2235 UTC). Most assigned minor probabilities were from 50-70% while most of the assigned major probabilities were 10% or 0%. The written responses in the warnings and the data collection platforms noted that the observational data was driving the warning decision rather than the WoFS QPF-driven data. It is hypothesized that the WoFS was not portraying as significant of a threat in that region.

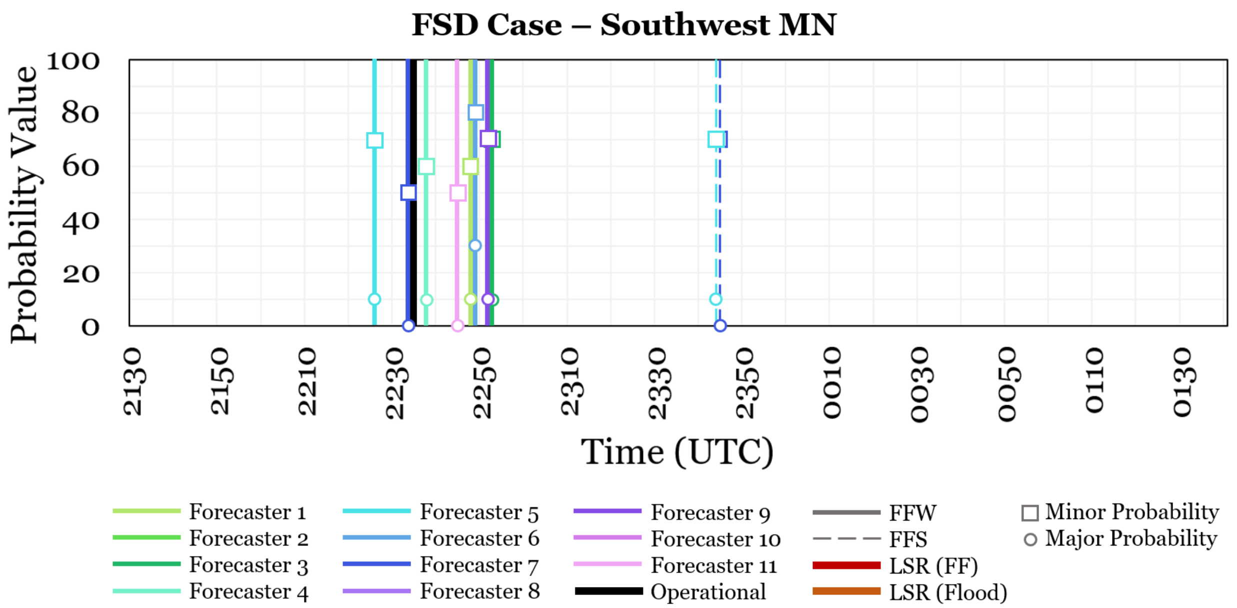


Figure 19. Same as Figure 17 except for southwest Minnesota in the FSD case.

The Gregory Co., SD flash flood event saw a mixed result of experimental warning decisions (Figure 20). Ten of the 11 forecasters issued a warning for this region, but the warning decisions were spread out over a 90 minute period. Five of the experimental FFWs occurred before the LSR and were used to calculate the average lead time of 37.4 minutes. Three of those five warnings occurred before the operational FFW at 2331 UTC, with the first experimental FFW issued at 2257 UTC. The minor probabilities were around 60–70% with major probabilities in the 0–20% range. While some forecasters were confident enough with the WoFS forcing, others were observationally driven.

The Douglas Co., SD event was another event that had some influence with WoFS but was primarily driven by observational data and trends. All 11 participants issued for this area before the LSR with seven participants issuing FFWs before the operational FFW (Figure 21). Six of those warnings occurred between 0016–0019 UTC. Most responses on the influence of the WoFS QPFs and QPF-driven products in the warning decision was “Somewhat” or “Little” influence. The minor and major flash flood probabilities assigned to the FFWs and FFSs did increase in time up to the flash flood LSR.

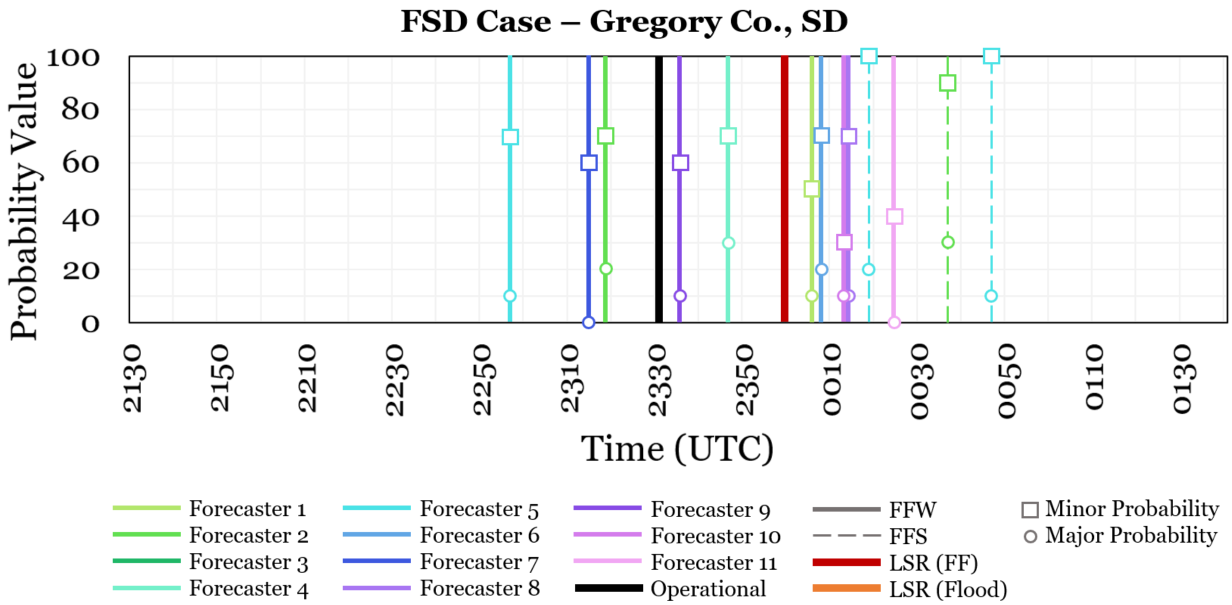


Figure 20. Same as Figure 17 except for Gregory Co., SD in the FSD case.

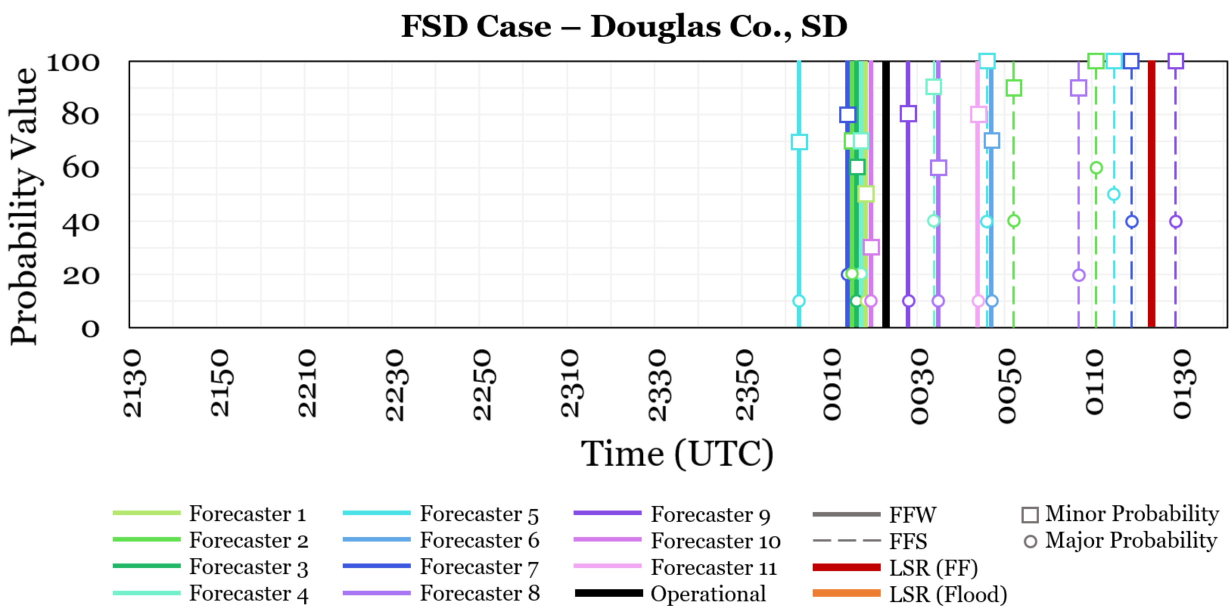


Figure 21. Same as Figure 17 except for Douglas Co., SD in the FSD case.

The experimental warning performance varied for the four areas of interest in the DMX case. The Story Co., IA event was warned by all participants with nine of the FFWs issued before the first flash flood LSR at 0045 UTC 1 July (Figure 22). Seven of the experimental FFWs were issued before the operational FFW. The earliest waring decision was from Forecaster #1 at 2359 UTC 30 June and all of the warning decision were spread out over a 55 minute period. There were a number of instances when the WoFS QPF-driven data had "Significant" influence over the decision, yet there were a few cases that observations drove the warning issuance. While the assigned major flash flood probabilities were pretty steady throughout the timeline, the minor flash flood probabilities did increase in time.

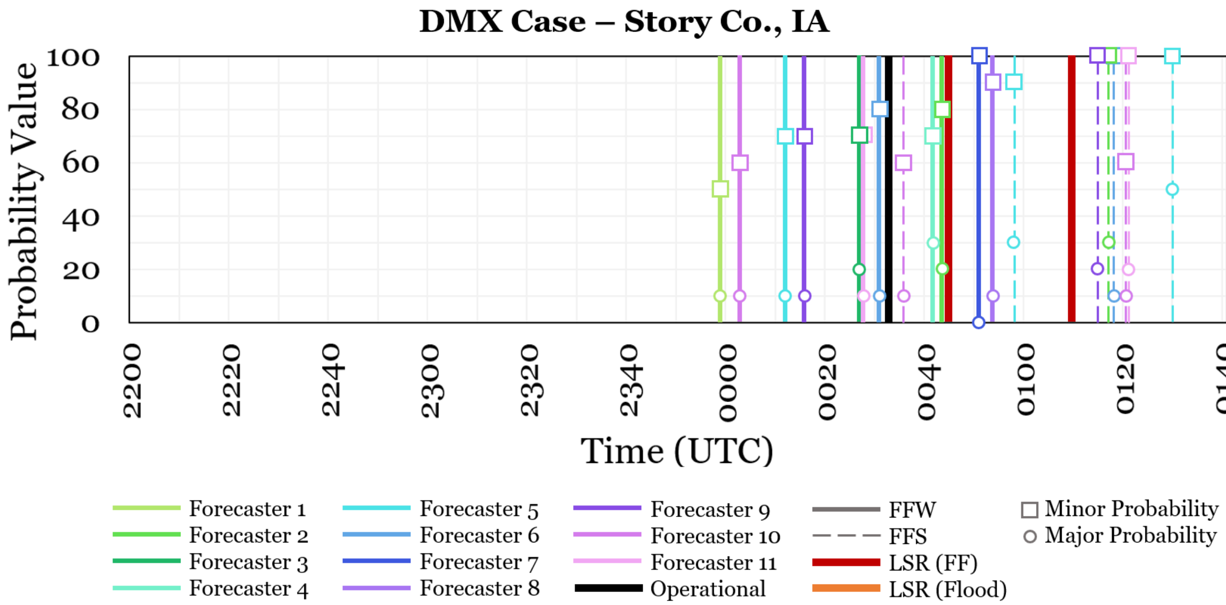


Figure 22. Same as Figure 17 except for Story Co., IA in the DMX case.

Both the Polk County and Dallas County, IA event had similar experimental warning performances. The Dallas Co. event had ten of the 11 participants warn for this area; moreover, all ten experimental warnings were issued before the flash flood LSR and the operational FFW (Figure 23), resulting in an average lead time of 32.7 minutes. Seven of the experimental FFWs were clustered in a 15 minute period around 0030 UTC with the first warning issued by Forecaster #10 at 0003 UTC. The probabilities for minor flash flooding before the LSRs were around 60–80% with most assigned major flash flood probabilities at 10%.

The Polk Co. event saw all 11 participants issue warnings for this region, nine of which were issued before the first LSR and operational FFW (Figure 24). The average lead time of those nine warnings was 42 minutes, which was 33 minutes more than the operational FFW. The majority of the participants drew an experimental warning polygon that encompassed both this event and the Polk Co. event; thus, the warning issuance times are similar between the two events.

The Blackhawk Co., IA event was shown to be the most difficult event to warn for amongst all the cases and event. Five of the 11 participants did not put out a warning for where the LSR was located. It should be noted that some forecasters did issue warnings for nearby areas. Of the six issued experimental FFWs, all six occurred before the flash flood LSR and four occurred before the operational FFW at 0042 UTC 1 July (Figure 25). The six experimental warnings did provide an additional 15.67 minutes of lead time. The six warning decisions were spread out over a large period of time and were dependent upon the influence of WoFS on their decisions. The first experimental FFW was issued by Forecaster #5 at 2331 UTC 30 June while the last one issued was at 0058 UTC 1 July. Assigned minor flash flood probabilities ranged from 50% to 80% while major flash flood probabilities were 20% or less.

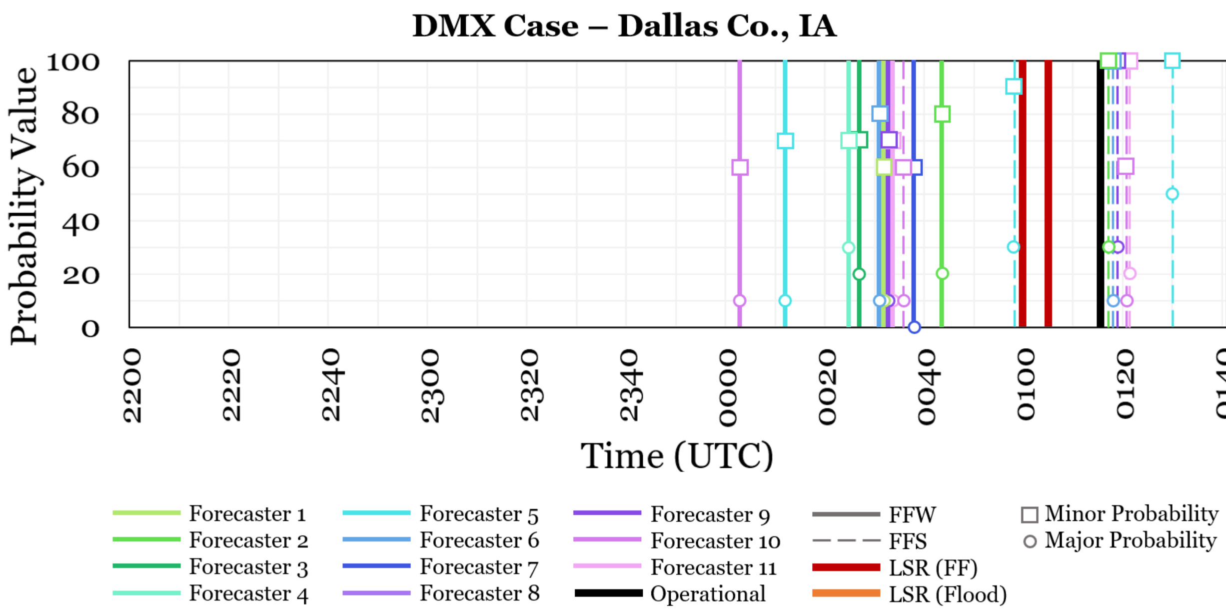


Figure 23. Same as Figure 17 except for Dallas Co., IA in the DMX case.

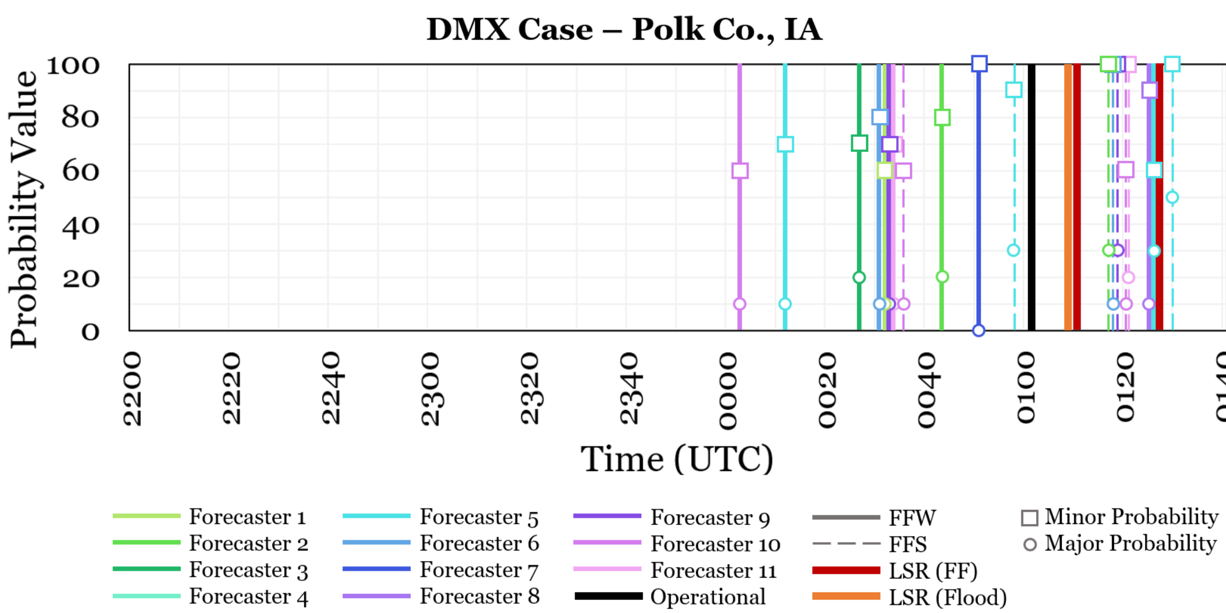


Figure 24. Same as Figure 17 except for Polk Co., IA in the DMX case.

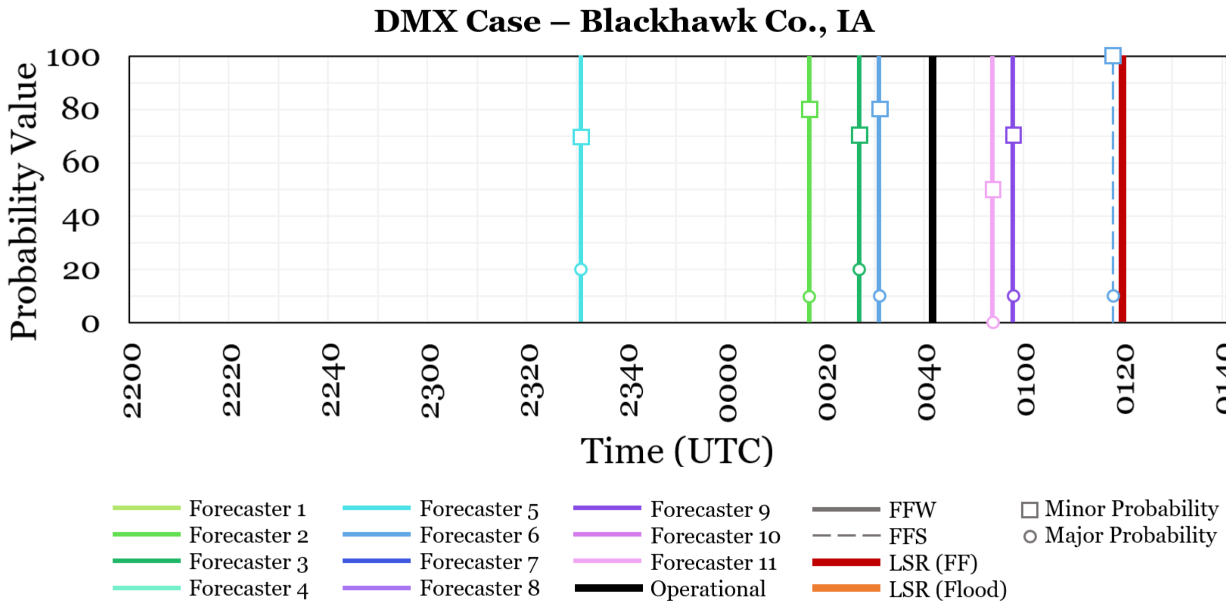


Figure 25. Same as Figure 17 except for Blackhawk Co., IA in the DMX case.

Overall Warning Metrics

Another perspective from the warning issuance that was assessed was the warning performance metrics for each case. To conduct this, the following guidelines were implemented:

- Flash flood LSRs were the verification source (i.e., flood LSRs were not used to verify the warnings or set their lead times).
- FFWs that had positive lead times were marked as "Hit Events" and had their lead times counted in the analysis.
- FFWs that had negative lead times were set as "Missed Events" and did not have their lead times incorporated into the analysis.
- Only FFWs that were inside the designated areas of interest were considered (Figure 26).
    - Participants were constrained in area for both the DMX and LWX cases
    - Participants were able to forecast over the entire FSD CWA; yet, it should be noted that in the case pre-brief presentation and evaluation questions, two areas were highlighted in particular for flash flood threat assessments

A total of nine operational warnings were issued across the three cases during the time periods of each case. Using the aforementioned guidelines, seven of the nine warnings were verified with positive lead time and considered a "hit" event (Table 5). One FFW during the DMX case was considered a "missed" event given the negative lead time, and there was one unverified warning from the FSD case. The overall probability of detection (POD) was 0.88 with a false alarm rate (FAR) of 0.13 and a critical success index (CSI) of 0.78. The average FFW lead time across all of the cases was 23.14 minutes.

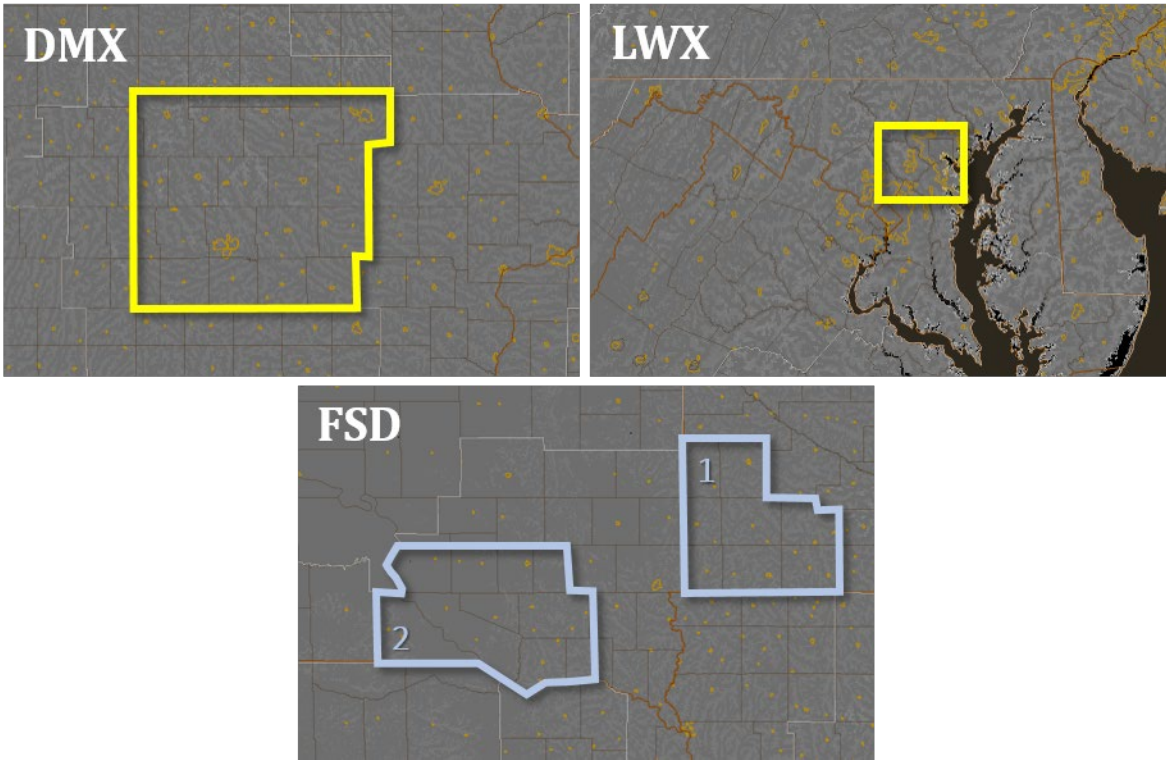


Figure 26. The domain areas for the DMX (top left), LWX (top right), and FSD (bottom) county warning areas. The DMX and LWX areas were restricted to the regions highlighted in yellow. Participants could forecast for the entire FSD domain but were directed to have increased attention for the two areas highlighted in blue.

Table 5. Operational warning metrics for all three archived cases.

	Hit Events	**Missed Events**	**Unverified FFWs**	**POD**	**FAR**	**CSI**	**Lead Time (min)**
FSD Case	2	0	1	1.00	0.33	0.67	45.00
DMX Case	3	1	0	0.75	0.00	0.75	19.67
LWX Case	2	0	0	1.00	0.00	1.00	6.50
Overall	***7***	***1***	***1***	***0.88***	***0.13***	***0.78***	***23.14***

A total of 103 experimental FFWs were issued across all three cases by the 11 participants who conducted the archived case analyses. Half of the experimental warnings (52) were verified by flash flood LSRs, while there were 18 missed events and 33 unverified warnings (Table 6). The most successful event was the LWX case that focused on areas near Baltimore and westward. The POD was 0.88 with a CSI of 0.65. The average lead time for this event was 54 minutes. The case also had the second lowest FAR rate of 0.29. The two missed events were the two aforementioned warnings that were within 5 km of Ellicott City but did not encompass the Ellicott City flash flood LSR. The greatest FAR rate and lowest CSI was with the FSD case. This was primarily from the null event in southwest Minnesota; however, there were additional warnings stretching northeast from the Gregory and Douglas Co. events that were also unverified. There were a number of unverified events with the DMX case that focused on a strong flash flood signal using WoFS QPFs that did not verify over Marshall County (east of Story County).

Table 6. Experimental warning metrics for all three archived cases.

	Hit Events	Missed Events	Unverified FFWs	POD	FAR	CSI	Lead Time (min)
FSD Case	16	6	20	0.73	0.56	0.38	52.56
DMX Case	21	10	7	0.68	0.25	0.55	29.81
LWX Case	15	2	6	0.88	0.29	0.65	53.87
Overall	***52***	***18***	***33***	***0.74***	***0.39***	***0.50***	***43.75***

The overall POD amongst all three cases and all participants was 0.74 (Table 6), which is a slight decrease from the POD of 0.88 with the operational warnings. The FAR tripled to 0.39 with the experimental FFWs issued, while the CSI for the experimental FFWs was 0.50. The average lead time increased for all three cases, and the overall FFW lead time was 38.48 minutes, an approximately 12 minute increase.

The incorporation of the WoFS ensemble QPFs into the probabilistic FLASH data likely contributed to three notable warning characteristics. First is the overall increase in lead time, especially with the LWX case where the QPFs from the WoFS was perceived as a well-skilled forecast. The average increase in lead time across all cases was 20 minutes. The increase in lead time also came with an increase in the FAR. All cases had an increase in FAR between 0.20–0.30, which could be attributed to areas being forecasted to experience potential flash flooding but did not transpire. There were also some instances of missed events, and combined with the FAR, decreased the overall CSI. The skill of the WoFS varied between the three events and are being investigated by researchers.

Other Observations

The 2018-edition of the HMT-Hydro Experiment noted two different trains of thought with applying WoFS QPFs to the warning decision making process. Some forecasters utilized it to create earlier warning decisions while others were hesitant to make a warning decision in areas that had not experienced much or any rainfall yet. This trend did continue this year with the cases provided. Similar trends with lower probabilities towards the end of the 3-h forecast period were also noted, which is likely attributable to the generation of probabilities with ensembles.

Approximately two hours were dedicated on each Friday to a group discussion on the probabilistic products and the short-term QPFs from the Warn-on-Forecast system. Eight questions were used to drive the conversation during these discussions. Analysis was not completed in time for this version of the report; however, a detailed set of results presented in the upcoming Warn-on-Forecast report and other forums.

VII. Results: Probabilistic QPE Analysis

A total of six cases were identified by the experiment officers for analysis using the prototype PQPE product suite. Each case was centered on a six-hour period and are as follows:

- Event #1: 1800 UTC June 2 to 0000 UTC 3 June – Central North Dakota
- Event #2: 0600-1200 UTC 6 June – Southern Florida
- Event #3: 1800 UTC 11 June to 0000 UTC 12 June – Northern Wisconsin
- Event #4: 1800 UTC 21 June to 0000 UTC 22 June – Southwest Louisiana
- Event #5: 0000-0600 UTC 27 June – Northeast New Mexico
- Event #6: 0000-0600 UTC 29 June – Northern Missouri

Each of the six cases were analyzed by the participants. Only one additional event was analyzed outside of the six pre-selected cases and the combined results are presented here.

<u>*Uncertainty Information*</u>

The forecasters were asked to rank how important the two types of uncertainty information in the context of flash flood forecasting were for the cases presented (Table 7). The results for the uncertainty and quantile values of PQPE were similar and tended to cluster around the "Slightly Important" to "Moderately Important" categories, though some rankings were in the "Very Important" category. It was potentially perceived as less informative because it showed less spatial and temporal dynamics. Middle-of-the-road average scores of 2.94 and 3.02 for uncertainty and quantile values, respectively, show that there were little overall trends in either direction.

Table 7. Results of ranking the uncertainty information from the PQPE product suite.

How important do you think the uncertainty information is for PQPE with respect to this case?						
Field	**Not At All Important**	**Slightly Important**	**Moderately Important**	**Very Important**	**Extremely Important**	**Average Value**
Uncertainty Values	1	24	22	16	3	2.94
Quantile Values	1	22	22	15	5	3.02

On a ranking from 1 to 5 where 1 is "Not At All Important" and 5 is "Extremely Important"… 5.00 average is ideal.

When asked how best to display uncertainty information, there were no consensus response (Table 8). A total of 28 of the 66 responses went with the option for a margin of error value with a plus/minus value in inches. The other responses were spread nearly evenly among the original display of data in percent uncertainty and with a range of values in inches. The fact that 70% of the responses were not the percent of uncertainty that was presented was a deviation from the a priori assumption. The results do indicate that bracketing the average (expected) with some value would be more meaningful. When asked about the display of quantile information, 68% of the responses went with -50 to +50 range that was utilized in the prototype data set (Table 9). Responses that fell into the "Other" category included the differences in values and the percent difference in values. This was based on the different perception of ranges and results for lighter rate values versus that of more significant rate values.

Table 8. Results for expression of uncertainty value of PQPE.

What way would you like to see uncertainty information expressed for precipitation rates that would be most beneficial for this case?		
Field	**Count**	**Percent**
Percentage of uncertainty (50%)	20	30.30%
Range of values (1.00 to 1.50 inches)	18	27.27%
Margin of error value (+/- 0.25 inches)	28	42.42%
Other	0	0.00%

Table 9. Results for expression of quantile value of PQPE.

What way would you like to see quantile information expressed for precipitation rates that would be most beneficial for this case?		
Field	**Count**	**Percent**
-50 to +50 (value of 0 when PQPE == MRMS Q3RAD)	45	68.18%
0 to 100 (value of 50 when PQPE == MRMS Q3RAD)	8	12.12%
Other	13	19.70%

If other, describe how you would like to see quantile information displayed?

Exceedance Probabilities

Two questions were posed regarding the understanding of the threat of the event via exceedance probabilities: the magnitude of the instantaneous rain rate and the likelihood of flash flooding. Similar trends were shown with both sets of results (Tables 10 and 11). The most frequent response for exceedance probabilities of 10 mm h^{-1} were in the "Slightly Important" category (likely due to being less relevant to defining a flash flood threat, while "Very Important" was the most frequent response for rates of 25 and 50 mm h^{-1}. It should be noted that there was significant spread in responses for all exceedance probability rate categories. The responses for exceedance probabilities of 100 mm h^{-1} were nearly even distributed between "Slightly Important" to "Extremely Important"; however, this could potentially be attributed to the different magnitude of cases presented to the participants and the rarity of these events.

Regardless of the large spread in responses, there is a significant disconnect between the importance of exceedance probability values for 10 mm h^{-1} versus the other, greater rate values. It should be noted that these lesser rates could be important for urban events and burn scars. When asked about how best to express exceedance rate probabilities, over 95% of the responses were to keep it in the same format as presented: percentages (Table 12). Two of the 66 responses were for standard deviations and one response discussed the use of a worst-case precipitation values as well as 90% and 10% exceedance probabilities.

Table 10. Results for rating the instantaneous exceedance probabilities for different rate values with respect to the understanding the rain rate magnitude.

Rate the instantaneous exceedance probabilities for understanding the instantaneous rain rate magnitude for the event.						
Field	**Not At All Important**	**Slightly Important**	**Moderately Important**	**Very Important**	**Extremely Important**	**Average Value**
Exceeding 10 mm h^{-1}	7	32	19	7	1	2.44
Exceeding 25 mm h^{-1}	0	9	16	27	14	3.70
Exceeding 50 mm h^{-1}	0	10	15	24	17	3.73
Exceeding 100 mm h^{-1}	2	17	17	14	16	3.38

On a ranking from 1 to 5 where 1 is "Not At All Important" and 5 is "Extremely Important"… 5.00 average is ideal.

Table 11. Results for rating the instantaneous exceedance probabilities for different rate values with respect to the understanding the flash flood threat.

Rate how the instantaneous exceedance probabilities help your understanding of the likelihood for flooding for this event.						
Field	**Not At All Important**	**Slightly Important**	**Moderately Important**	**Very Important**	**Extremely Important**	**Average Value**
Exceeding 10 mm h^{-1}	18	24	18	6	0	2.18
Exceeding 25 mm h^{-1}	3	16	17	21	9	3.26
Exceeding 50 mm h^{-1}	1	11	17	27	10	3.52
Exceeding 100 mm h^{-1}	5	15	15	17	14	3.30

On a ranking from 1 to 5 where 1 is "Not At All Important" and 5 is "Extremely Important"… 5.00 average is ideal.

Table 12. Results for expression of exceedance probability values with PQPE.

What way would you like to see instantaneous exceedance information expressed for precipitation rates that would be most beneficial for this case?		
Field	**Count**	**Percent**
Standard deviations about a certain value	2	3.03%
Percentage	63	95.45%
Other	1	1.52%

If other, describe how you would like to see exceedance information provided?

Count and Percent of Exceedance Probabilities

The number of times that the instantaneous rain rate probabilities exceeded 50% for the different rate thresholds were counted for three different temporal periods: 1, 3, and 6 hours. The percent of time that the 50% exceedance threshold was also calculated (i.e., given the 2-min update of MRMS QPE, if a rate exceeded 50% a total of 12 times over a one-hour period, then the percent of time for rate exceedance is 40% of that period). When asked about which method better helps the understanding of the flood risk, the percent value had a greater average response by 0.76 over the counts (Table 13). Discussions with participants noted that having the percentages would mean they would not have to perform the calculations of the count versus the number of product instances per time period.

Table 13. Results for understanding the flood risk using exceedance probabilities over a given time period.

Rate how the 1) number of times and 2) the percent of times that the probability of exceedance > 50% for a given threshold over a given time period helps your understanding of flood risk.						
Field	**Not At All Important**	**Slightly Important**	**Moderately Important**	**Very Important**	**Extremely Important**	**Average Value**
Exceedance Count	1	23	29	10	1	2.80
Exceedance Percentage	0	5	23	31	5	3.56

On a ranking from 1 to 5 where 1 is "Not At All Important" and 5 is "Extremely Important"… 5.00 average is ideal.

Participants were asked to rate the importance of the various rain rate thresholds using the counts and percentages values (Table 14) and the temporal periods (Table 15). The importance of the rate values were very similar to that of the initial analysis of the exceedance probabilities in Tables 10 and 11. When asked about the various temporal accumulation periods of the exceedance probabilities, the 1-h and 3-h periods were generally ranked as "Very Important," while the 6-h accumulations had a greater spread in values and more centered around an average value closer to "Moderately Important."

Overall Product Utility

The overall utility rankings of the various PQPE products can be binned into two categories (Table 16). The expected PQPE values, the probability of exceeding rain rate thresholds, and the percent of probability exceedance had higher overall rankings. The mean values from the participants had them between 69 and 78 with 100 being an ideal value. The uncertainty, quantile, and count of probability exceedance values had mean values less than 51. The average standard deviation value for the higher rated products was 3.25 points lower than that of the three lower rated products (21.78 vs. 25.03).

Table 14. Results for rating the various rain rate thresholds over a given time period.

Rate the importance of the various instantaneous rain rate thresholds that were used to create the number or proportion (percent) of times that a probability of exceedance > 50% for the various rain rates.						
Field	**Not At All Important**	**Slightly Important**	**Moderately Important**	**Very Important**	**Extremely Important**	**Average Value**
10 mm h^{-1}	10	28	20	7	0	2.37
25 mm h^{-1}	2	11	19	25	8	3.40
50 mm h^{-1}	0	8	15	30	12	3.71
100 mm h^{-1}	2	13	23	14	13	3.35

On a ranking from 1 to 5 where 1 is "Not At All Important" and 5 is "Extremely Important"… 5.00 average is ideal.

Table 15. Results for rating the different time period accumulations for exceedance probability counts and percentages.

Rate the importance of the various temporal accumulation periods that were used to create the number or proportion (percent) of times that a probability of exceedance > 50% for the various rain rates.						
Field	**Not At All Important**	**Slightly Important**	**Moderately Important**	**Very Important**	**Extremely Important**	**Average Value**
1 hour	0	3	17	26	19	3.94
3 hours	0	6	15	29	14	3.80
6 hours	4	16	18	23	4	3.11

On a ranking from 1 to 5 where 1 is "Not At All Important" and 5 is "Extremely Important"… 5.00 average is ideal.

Table 16. Results for ranking the overall utility of each PQPE product.

Please rank the overall utility of the basic product concept for analyzing the precipitation for this event.		
Field	**Mean**	**Standard Deviation**
Probabilistic QPE	77.78	24.13
Uncertainty of QPE	50.94	27.48
Quantiles of QPE	48.31	23.70
Probability of Exceeding Values	74.15	20.60
Count of Probability Exceedance	45.85	23.92
Percent of Probability Exceedance	68.92	20.62

Conducted on a scale from 10-100 where a 100.00 average is ideal.

VIII. Analysis, Other Findings, and Recommendations

A summary of each section of the HMT-Hydro Experiment is presented here, which includes some recommended actions to improve the design of future experiments. Some of the findings here stemmed from the end-of-week feedback survey. The questions and their rankings are provided in Appendix D. A summary of the written feedback from the end-of-week survey can be made available upon request.

For Probabilistic Tool Development

The suite of probabilistic products made available within the FLASH system provided utility in identifying the spatial coverage and the potential magnitude of the flash flood threat. Refinements were made to the product now designated as the Probability of Minor Flash Flooding, and the perceived bias in the product improved from the 2018 analysis; however, perceived biases still exist in all probabilistic products. This is still most noticeable in the Probability of Receiving a Flash Flood LSR product that has been described as "too hot."

Another notable observation about the probabilities were how the probability curves were used to generate the probabilistic guidance for minor, moderate, and major flash flooding. For example, the Probability of Minor Flash Flooding never reached 90%, even when the FLASH CREST Unit Streamflow far exceeded 10 $m^3 s^{-1} km^{-2}$. The Probability of Major Flash Flooding rarely exceeded 40%, even for major flash flood events. The values generated should still be investigated and adjusted as necessary.

Computational resources limited the probabilistic values generated in AWIPS and the FLASH web site. Probability values for areas where the FLASH CREST Unit Streamflow is less than 0.3 $m^3 s^{-1} km^{-2}$ were not displayed. Thus, probability values around 15% or less for the Probability of Receiving a Flash Flood LSR product are not seen. As computation resources improve, the ability to show more of the low-end probability values should be considered.

For Real-Time Operations

Times were more flexible for real-time operations, including the call to bring participants in early (9:00AM CDT) on Monday 15 July 2019 to capture the end of Hurricane Barry. Moreover, there was time at the end of other sessions (i.e., archived cases) that led to additional warning operations that led to additional verified experimental FFWs. A more flexible schedule for real-time operations should continue to be utilized.

The Hazard Services software was not available for this experiment, but WarnGen was available and easily modified to survey participants about their warning decision. The survey-style approach allowed for valuable feedback on the warning decision making process and how the experimental products have been utilized. This should continue with additional and/or modified questions depending on the testbed objectives.

Once the Hazard Services software is re-introduced, the use of flash flood recommenders should be re-evaluated. The 2015 HMT-Hydro Experiment explored the use of flash flood recommenders for an automated drawing of a first-guess warning polygon (Martinaitis et al. 2017). The experiment used a single product with a user-defined threshold to draw these recommended polygons. Given the use of probabilistic information, it is recommended that the reintroduction of the flash flood recommender should be considered with the complexity of its calculations to be determined. Similar analysis to the Martinaitis et al. (2017) should be conducted in regards to the use of fully automated versus manually drawn or edited polygons.

Since the inaugural HMT-Hydro Experiment in 2014, forecasters have issued experimental FFWs with user-assigned probabilities for the potential of minor and major flash flooding. This action continued in the 2019 HMT-Hydro Experiment. The combination of forecasters assigning minor/major flash flood probabilities with the multiple gridded probabilistic products developed within the FLASH system provided a platform to not only give the uncertainty of a potential hazard existing but the uncertainty of the severity of the potential hazard. This approach should continue to be explored with flash flooding and other storm-scale hazards and within the concepts of the FACETs paradigm.

For Archived Cases

Multiple improvements were made to the archived cases compared to the 2018 HMT-Hydro Experiment. The cases went from a paused analysis to a displaced real-time mode. The long data collection forms were replaced with simple survey questions that were prompted every 30 minutes. The participants had the ability to issue warnings. All of this was well received by the participants this year.

The archived cases should continue to use the 0–3 h WoFS QPFs. The timing/data latency of this product with the probabilistic FLASH system was based solely on the availability of WoFS 2018 real-time data. It is unknown what the timing of FLASH would be with WoFS, and this should be investigated. Moreover, the WoFS QPFs themselves were not available in AWIPS but as separate images via WESSL. Efforts are underway to display the various products from WoFS QPFs for the entire 0–3 h forecast period in AWIPS.

The case lengths were relatively short; however, the FSD case was reduced in length by one hour due to participant fatigue. Operational forecasters are not on the warning "hot seat" for periods of over four hours, so cases should continue to be constrained in both space and time to prevent fatigue.

For Probabilistic QPE

This was the first year for the use of PQPE, and the use of these products are being evaluated. The participants were given one-month of data and six pre-defined cases to analyze within this period. However, the survey tool made it very difficult to separate the results out for each case. It is recommended to have a different survey for each case.

The survey results did show that some products were not as favored as others; moreover, feedback was given on how best to display the PQPE product suite. The expected QPE value from PQPE and the exceedance values were perceived as more relevant, while uncertainty information and comparison with MRMS were used to explain the shift from deterministic to probabilistic information, even though not as well received. It is recommended that the PQPE product suite be trimmed to the products, rate, and time periods that the participants found to be most useful.

The PQPE data was only available in an archived format. In future testbed experiments, the PQPE product suite should be considered for real-time viewing in the AWIPS system for further analysis.

For Future HMT-Hydro Experiments

The time of the year of the HMT-Hydro Experiment allowed for close coordination with the FFaIR Experiment and avoids interfering with springtime severe convection studies by other experiments under the HWT umbrella. Running the HMT-Hydro Experiment in the summer also allows for the inclusion of monsoon-driven events in the southwestern CONUS and the potential for impacts from tropical cyclones; however, some operational days were notably slow despite the fact that only two days per week were dedicated to real-time operations.

The use of archived cases was a benefit to the schedule, allowing for activities to occur on slower weather days. The use of archived cases should be used in future experiments to test other objectives related to WoF and the FACETs paradigm.

Subjective evaluations conducted during the HMT-Hydro Experiment utilized a laptop containing the TurningPoint™ software and individual clickers used to collect, display, and archive forecaster responses. This laptop was borrowed from the NWS Warning Decision Training Division (WDTD). The use of the TurningPoint™ software allowed for participants to provide independent, anonymous feedback to evaluation questions and statements without other participants influencing responses (via open discussions during the scoring portion of the evaluation). The purchasing of a laptop with the TurningPoint™ software should be considered by the HWT for use in the HMT-Hydro Experiment and potentially for other experiments conducted within the HWT for evaluation purposes.

Acknowledgements

First and foremost, the officers of the HMT-Hydro Experiment would like to thank all of the participants for their hard work and contribution during the three weeks of this experiment. The insight gained throughout this process is invaluable to furthering the science and application of future products to improve flash flood prediction.

The principal investigators would like to thank everyone who helped make the 2019-edition of the HMT-Hydro Experiment a success. From the various data sets to the technical and managerial aspects of running the HWT, this would not be possible without your time and dedication to the project.

The HMT-Hydro Experiment is funded through the Hydrometeorology Testbed by the Office of Weather and Air Quality under NOAA Award Number NA17OAR4590281 for the project "Probabilistic Warn-on-Forecast System for Heavy Rainfall and Flash Flooding." Regional NWS headquarters/offices also provided some funding for certain participants.

References

Barthold, F., T. Workoff, B. Cosgrove, J. Gourley, D. Novak, and K. Mahoney, 2015: Improving flash flood forecasts: The HMT-WPC Flash Flood and Intense Rainfall Experiment. *Bull. Amer. Meteor. Soc.*, **96**, 1859–1866.

Clark, E., 2011: Weather Forecast Office hydrologic products specification. US Department of Commerce, National Weather Service, Instruction 10-922. [Available online at http://www.nws.noaa.gov/directives/sym/pd01009022curr.pdf.]

Clark, R. III, J. J. Gourley, Z. Flamig, Y. Hong, and E. Clark, 2014: CONUS-wide evaluation of National Weather Service flash flood guidance products. *Wea. Forecasting*, **29**, 377–392.

Elmore, K. L., Z. L. Flamig, V. Lakshmanan, B. T. Kaney, V. Farmer, H. D. Reeves, L. P. Rothfusz, 2014: MPING Crowd-Sourcing Weather Reports for Research. *Bulletin of the American Meteorological Society*, **95,** 1335–1342.

Gourley, J. J., and Coauthors, 2017: The FLASH project: Improving the tools for flash flood monitoring and prediction across the United States. *Bull. Amer. Meteor. Soc.*, **98**, 361–372.

Horvitz, A., 2012: Multi-purpose weather products specification. US Department of Commerce, National Weather Service, Instruction 10-517. [Available online at http://www.nws.noaa.gov/directives/sym/pd01005017curr.pdf.]

Kirstetter, P. E., J. J. Gourley, Y. Hong, J. Zhang, S. Moazamigoodarzi, C. Langston, and A. Arthur, 2015: Probabilistic precipitation rate estimates with ground-based radar networks. *Wat. Resour. Res.,* **51,** 1422–1442.

MacAloney, B., 2007: Storm Data preparation. . US Department of Commerce, National Weather Service, Instruction 10-1605. [Available online at http://www.ncdc.noaa.gov/stormevents/pd01016005curr.pdf.]

Martinaitis, S. M., and Coauthors, 2017: The HMT Multi-Radar Multi-Sensor Hydro Experiment. *Bull. Amer. Meteor. Soc.*, **98**, 347–359.

Perica, S., D. Martin, S. Pavlovic, I. Roy, M. St. Laurent, C. Trypaluk, D. Unruh, M. Yekta, and G. Bonnin, 2013, cited 2014: NOAA Atlas 14: Precipitation-frequency atlas of the United States, Volume 9, Version 2.0. US Department of Commerce, National Weather Service, Hydrologic Design Studies Center. [Available online at http://www.nws.noaa.gov/oh/hdsc/PF_documents/Atlas14_Volume9.pdf.]

Zhang, J., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 621–637.

APPENDIX A: HMT-Hydro Participants and Staff

A total of 13 participants were a part of the 2019 HMT-Hydro Experiment. Nine participants were from local NWS Weather Forecast Offices (WFOs) or River Forecast Centers (RFCs), while other participants came from NWS headquarters, training centers, and research facilities. Eleven of the participants were able to evaluate the archived cases due to their NWS and warning expertise.

Name	Affiliation	Week #
Mike Dutter	NWS WFO Wakefield, VA	1
Amanda Schroeder	NWS West Gulf RFC (Fort Worth, TX)	1
Jane Marie Wix	NWS WFO Jackson, KY	1
Linda Cheng	NWS WFO Salt Lake City, UT	2
Justin Gibbs	NWS WDTD (Norman, OK)	2
Matthew Kelsch	COMET Program (Boulder, CO)	2
Brian Schoettmer	NWS WFO Louisville, KY	2
Jimmy Taeger	NWS WFO San Diego, CA	2
Kate Abshire	NWS WRSB (Silver Springs, MD)	3
Emilie Nipper	NWS Lower Mississippi RFC (Slidell, LA)	3
Nina Oakley	Desert Research Institute (Reno, NV)	3
Jeremy Wesely	NWS WFO Hastings, NE	3
Alexander Zwink	NWS WFO Norman, OK	3

The officers of the 2019 HMT-Hydro Experiment were responsible for the facilitating of all operational activities during each week. At least one or two HMT-Hydro Experiment officers were in attendance through all daily activities, while other officers focused on the technical aspects, logistics, or evaluations of specific products during the experiment.

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

APPENDIX B: Hours of Operations for HMT-Hydro Experiment

This appendix details the times that were kept for each of the three active weeks of the HMT-Hydro Experiment. Hours varied each day based on the current weather during the experiment. The Mondays through Thursdays of each week featured the combination of real-time warning operations and archived cased studies featuring Warn-on-Forecast data. The Fridays revolved around discussions, final evaluations, and surveys. While the weekly schedule stayed within the framework defined by the operations plan, there were a few modifications to the times and activities. The biggest deviation from the schedule occurred on 15 July when forecasters started 2.5 hours earlier than scheduled to capture the end of the flash flooding component of Hurricane Barry from that weekend. The tables below define the time of operations for each week and the primary activity of the day (real-time operations, cases, and/or discussions).

Week One – 24 June to 28 June 2019				
	Date of Week	Start Time (CDT)	End Time (CDT)	Activities
Monday	24 June	11:30 AM	07:30 PM	RT
Tuesday	25 June	10:00 AM	06:00 PM	AC
Wednesday	26 June	10:30 AM	06:30 PM	AC
Thursday	27 June	10:30 AM	06:30 PM	RT
Friday	28 June	09:00 AM	02:00 PM	DS

Week Two – 8 July to 12 July 2019				
	Date of Week	Start Time (CDT)	End Time (CDT)	Activities
Monday	8 July	11:30 AM	07:30 PM	RT
Tuesday	9 July	11:00 AM	07:00 PM	AC
Wednesday	10 July	10:00 AM	06:00 PM	AC
Thursday	11 July	10:00 AM	06:00 PM	RT
Friday	12 July	09:00 AM	02:00 PM	DS

Week Three – 15 July to 19 July 2019				
	Date of Week	Start Time (CDT)	End Time (CDT)	Activities
Monday	15 July	09:00 AM	05:00 PM	RT
Tuesday	16 July	11:00 AM	07:00 PM	RT
Wednesday	17 July	11:00 AM	07:00 PM	AC, RT
Thursday	18 July	10:00 AM	06:00 PM	AC, RT
Friday	19 July	09:00 AM	02:00 PM	DS

Primary Activities: Real-Time Operations – RT
Archived Case Studies – AC
Discussions/Surveys – DS

APPENDIX C: Products Used in HMT-Hydro Experiment

Subjective evaluations of the real-time operations focused on the gridded flash flood probabilities and the experimental warning issued. Archived cases utilized the same probabilistic products with additional focus on the Warn-on-Forecast QPFs. The archived cases played in a displaced real-time environment that allowed for warning issuance and the ability to received reports like a NWS office would in real-time.

The table below summarizes the primary products and observations that were available for both the archived cases (AC) and the real-time operations (RT) in the 2019 HMT-Hydro Experiment. It should be noted that other data sets were also available during both the real-time and archived case sessions (e.g., WSR-88D radar data, numerical weather prediction models, surface observations, etc.), and they are not outlined here.

Product	Provider	Description	AC	RT
Flash Flood Observations				
Local Storm Reports	NWS	Operational reports of flash flooding used to validate warnings	X	X
mPING	NSSL	Citizen-scientist reports of flash flooding defined by four levels of severity		X
Streamflow	USGS/NWS/NSSL	Measurement of streamflow that have exceeded flood stage or a nominal return period flow (e.g., 5-yr return) in small, gauged basins		X
Quantitative Precipitation Estimations and QPE Comparison Products				
MRMS QPE (Dual-Pol Synthetic)	NSSL	Precipitation estimates from radar-only algorithm using various dual-polarization variables; Derives instantaneous rates and multiple accumulation periods	X	X
QPE-to-FFG Ratio	RFCs/WPC/NSSL	Compares a 1, 3, and 6-h rolling sum of MRMS QPE to most recently issued 1, 3, and 6 h FFG*	X	X
QPE Average Recurrence Interval	NWS/NSSL	Compares various MRMS QPE accumulations from 30-min to 24-h to precipitation frequencies from NOAA Atlas 14**	X	X
Quantitative Precipitation Forecasts				
Warn-on-Forecast System	NSSL	Ensemble QPFs provided to FLASH on a 3-km resolution, 900x900 km domain every hour for a lead time of 0–3 h or 0–6 h (case depending)	X	
Hydrologic Modeling Products				
Max Streamflow	NSSL	Maximum streamflow forecast during an interval spanning up to 12 hours after valid time	X	X
Max Unit Streamflow	NSSL	Maximum unit streamflow forecast during an interval spanning up to 12 hours after valid time	X	X
Soil Moisture	NSSL	Analysis of soil saturation	X	X
Probability of LSR	NSSL	Gridded probabilities of receiving a local storm report (on a scale of 0.00 to 1.00)	X	X
Probability of Minor/Moderate/ Major Flash Flooding	NSSL	Gridded probabilities of exceeding defined maximum unit streamflow values to determine hazard magnitude (on a scale of 0.00 to 1.00)	X	X

* RFCs typically update FFG at synoptic (0000, 1200 UTC) and sub-synoptic (0600, 1800 UTC) times, but the FLASH server queries all RFCs once an hour for FFG updates. During heavy rainfall events, some RFCs produce intermediate FFG products and hourly queries ensure that FLASH catches these intermediate FFG issuances. The FFG product displayed in FLASH is a national mosaic. There are different methodologies used to produce FFG across the country (including gridded and lumped FFG as well as the flash flood potential index), so discontinuities in FFG values across RFC boundaries may exist. Since the FFG values are being obtained from the RFCs, no locally forced FFG values from NWS forecast offices are included in this national FFG mosaic.

** NOAA Atlas 14 does not yet include precipitation frequency estimates for the Northwestern United States. Precipitation frequency values were derived by NSSL for use in this product until the official grids are published.

APPENDIX D: Responses from Feedback Survey

This section provides the results from the end-of-week survey that was given to each participant on their respective Friday afternoon. Answers with greater responses have a darker shade of red for a background color.

With regard to the HMT-Hydro Experiment introductions on Monday:

Field	Strongly Disagree	Disagree	Neither	Agree	Strongly Agree	Average Value
The introduction helped me to understand experimental flash flood products sufficiently.	0	0	1	2	9	4.67
The introduction helped me to understand the experimental WoF quantitative precipitation forecasts sufficiently.	0	0	1	7	4	4.25
The introduction helped me to understand the probabilistic QPEs sufficiently for evaluation.	0	0	2	8	2	4.00
The introduction was effective in giving me more familiarity with the AWIPS capabilities during the experiment.	0	0	3	6	3	4.00
I understood the anticipated outcomes and methodology after the presentations.	0	0	1	5	6	4.42

On a ranking from 1 to 5 where 1 is "Strongly Disagree" and 5 is "Strongly Agree"… 5.00 average is ideal.

With regard to the time allotted for each activity:

Field	Far Too Little	Too Little	About Right	Too Much	Far Too Much	Average Value
Introduction session	0	1	11	0	0	2.92
Real-time experimental flash flood warning operations	0	2	10	0	0	2.83
Case evaluations with the Warn-on-Forecast QPFs	0	0	9	2	0	3.18
Evaluation and discussion of the products/warnings from the prior day	0	0	12	0	0	3.00
FFaIR daily briefings	0	0	10	2	0	3.17
Assessment of PQPE product suite	0	5	6	1	0	2.67
End-of-week discussion on WoF QPFs and probabilistic grids for flash flood prediction	0	1	11	0	0	2.92

On a ranking from 1 to 5 where 1 is "Far Too Little" and 5 is "Far Too Much"… 3.00 average is ideal

Please indicate your level of agreement or disagreement with the following statements regarding the real-time experimental flash flood warning operations:						
Field	**Strongly Disagree**	**Disagree**	**Neither**	**Agree**	**Strongly Agree**	**Average Value**
In the forecasting sessions, I was given the tools that I needed to issue flash flood warnings.	0	0	0	3	9	4.75
The evaluation and discussion sessions helped me to improve my forecasts as the week progressed.	0	0	1	7	4	4.25
The FFaIR briefings gave me sufficient situation awareness to start the day.	0	0	4	7	1	3.75
The FFaIR briefings gave me all the information I needed to identify areas at risk.	0	2	5	5	0	3.25

On a ranking from 1 to 5 where 1 is "Strongly Disagree" and 5 is "Strongly Agree"… 5.00 average is ideal.

Please indicate your level of agreement or disagreement with the following statements regarding the archived case evaluations:						
Field	**Strongly Disagree**	**Disagree**	**Neither**	**Agree**	**Strongly Agree**	**Average Value**
I was given the tools that I needed to evaluate the flash flood threat during the archived cases that focused on WoF QPFs.	0	0	0	5	6	4.55
The AWIPS procedures were beneficial to the evaluation process.	0	0	0	1	10	4.91
The data collection form allowed me to provide detailed information on the experimental products.	0	0	0	6	5	4.45

On a ranking from 1 to 5 where 1 is "Strongly Disagree" and 5 is "Strongly Agree"… 5.00 average is ideal.

In terms of workload, please indicate the levels you felt across the whole week during each of the primary sessions:						
Field	**Much Lower than Average**	**Somewhat Lower than Average**	**About Average**	**Somewhat Higher than Average**	**Much Higher than Average**	**Average Value**
Experimental real-time flash flood warning operations	0	2	9	1	0	2.92
Tools/warning evaluation and discussion sessions	0	0	12	0	0	3.00
Archived case evaluations	0	0	6	5	0	3.45
PQPE analysis	0	3	7	2	0	2.92
FFaIR daily briefings	0	4	8	0	0	2.67

On a ranking from 1 to 5 where 1 is "Much Lower than Average" and 5 is "Much Higher than Average"… 3.00 average is ideal

Was the material provided before the experiment helpful in understanding and preparing for the experiment?						
Field	**Not at All Helpful**	**Somewhat Not Helpful**	**Neutral**	**Somewhat Helpful**	**Very Helpful**	**Average Value**
Forecaster Response	0	0	1	2	8	4.64

On a ranking from 1 to 5 where 1 is "Not at All Helpful" and 5 is "Very Helpful"… 5.00 average is ideal.

Would you consider participating in this experiment again in the future?			
Field	**Yes**	**No**	**Undecided**
Forecaster Response	11	0	1

Would you recommend participating in this experiment to colleagues?			
Field	**Yes**	**No**	**Undecided**
Forecaster Response	11	0	1