The 2016 Multi-Radar/Multi-Sensor (MRMS) HMT-Hydro Testbed Experiment

Final Report

28 September 2017

Prepared By

Jonathan J. Gourley[1] and Steven Martinaitis[2]

[1]NOAA/OAR/National Severe Storms Laboratory, Norman, OK
[2]Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, OK

Table of Contents

List of Figures

Introduction

The Hydrometeorology Testbed Multi-Radar Multi-Sensor (MRMS) Hydro Experiment (hereinafter denoted as the HMT-Hydro Experiment) was a part of the 2016 United States Weather Research Program (USWRP) Hydrometeorology Testbed (HMT). The HMT-Hydro Experiment was conducted in the Hazardous Weather Testbed (HWT) at the National Weather Center (NWC) in Norman, OK. It was conducted in conjunction with the Flash Flood and Intense Rainfall (FFaIR) Experiment at the Weather Prediction Center (WPC) in College Park, MD.

The HMT-Hydro Experiment operated for three weeks during the period from 20 June to 15 July 2016 with a 1-week break during the 4th of July holiday. Forecasters from National Weather Service (NWS) Weather Forecast Offices (WFOs) and River Forecast Centers (RFCs) worked with research scientists to assess emerging hydrometeorological technologies and products to improve the prediction, detection, and warning of flash flooding. The primary focus of the experiment in 2016 was the forecaster evaluation of short-term predictive tools derived from the MRMS radar-only quantitative precipitation estimates (QPE) and the Flooded Locations and Simulated Hydrographs (FLASH) hydrologic modeling framework. The decision-making process for each experimental flash flood watch and warning that was issued was also evaluated through the Hazard Services platform. The HMT-Hydro Experiment also explored the utility of experimental flash flood watches and warnings conveying uncertainty and magnitude. Lastly, a statistical approach using a random forest based on GFS model products to forecast flash flooding out to several hours was evaluated. Results from the HMT-Hydro Experiment will help in determining operationally relevant best practices.

Activities included training on the suite of MRMS-FLASH tools, forecast shifts to issue experimental flash flood watches and warnings, daily sessions to evaluate experimental forecasts and the tools used to generate them, and "Tales from the Testbed" webinars to spread initial findings and recommendations to NWS local and regional offices. Researchers from the National Severe Storms Laboratory (NSSL) and the University of Oklahoma (OU) administered the project and the Hazardous Weather Testbed (HWT) provided physical space and computing resources.

The 3rd annual HMT-Hydro experiment had four specific goals: 1) evaluate the existing and new national tools for flash flood forecasting, 2) rate the experimentally issued flash flood watches and warnings, 3) identify what forecasters used in their decision-making process for watches and warnings, and 4) evaluate and rate the experimental Day 1 and 2 products issued from the FFaIR Experiment. Every effort was undertaken to mimic the general operational organization of flash flood forecasting within the NWS. This report discusses the activities of the HMT-Hydro Experiment and presents findings from it with a specific emphasis on operational impacts and recommendations for future investigation.

Experimental Activities and Schedule

Each of the three weeks of the experiment followed a similar schedule; participants arrived at the testbed Monday morning and departed Norman early on Friday afternoon. Table 1 is a general outline of the experimental schedule. Forecasters spent a total of thirty-eight hours per week in the testbed. Of that time, 15 hours were spent in experimental forecasting shifts, ten were spent collecting data via three survey instruments, and the rest in other activities.

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 scientific research received preference. Prior to their arrival in Norman, participants were given general information about the principal scientific goals of the experiment, but were not officially exposed to any experimental products or tools until the Monday afternoon training session. In this session, four separate presentations were given: a reiteration of the scientific goals of the project; detailed descriptions and usage examples of all constituents of the suite of FLASH tools; AWIPS II training that focused on the differences between it and AWIPS I; and an explanation of the survey and audio/visual recording data to be collected throughout the experiment.

Table 1. Weekly experimental schedule of HWT-Hydro. Gray shading corresponds to non-working hours.

	Monday	Tuesday	Wednesday	Thursday	Friday
8 AM					Evaluation
9 AM					Evaluation
10 AM					Webinar Prep.
11 AM	Training				FFaIR Evaluation
Noon	Training	FFaIR Evaluation	FFaIR Evaluation	FFaIR Evaluation	Webinar
1 PM	Wx Briefing	Wx Briefing	Wx Briefing	Wx Briefing	Feedback Survey & Group Photo
2 PM	Forecasting	Evaluation	Evaluation	Evaluation	
3 PM	Forecasting	Forecasting	Forecasting	Forecasting	
4 PM	Forecasting	Forecasting	Forecasting	Forecasting	
5 PM	Forecasting	Forecasting	Forecasting	Forecasting	
6 PM	Forecasting	Forecasting	Forecasting	Forecasting	
7 PM	Forecasting	Forecasting	Forecasting	Forecasting	

Weather Briefings with WPC

One benefit of conducting HMT-Hydro during the summer is the timing overlap with the FFaIR experiment at WPC (Barthold and Workoff 2014). HMT-Hydro coordinated with FFaIR in an attempt to mimic the operational cascade of responsibilities from heavy rainfall guidance from WPC down to the issuance of flash flood watches and warnings from local WFOs. In the HMT-Hydro framework, FFaIR provided daily guidance on synoptic-scale heavy rainfall potential, numerical weather prediction diagnostics, and probabilistic forecasts of various heavy rainfall and flash flooding parameters. HMT-Hydro participants took on the role of a floating, national WFO, using FFaIR's guidance as a starting point. In general, FFaIR was responsible for forecasting heavy rainfall and flash flooding potential for timescales greater than six hours with HMT-Hydro taking on the responsibility for forecasting less than six hours prior to an event. The main conduit for interaction between the two experiments was a daily videoconference weather briefing from 1–2 PM CDT, Mondays through Thursdays. HMT-Hydro participants had the chance to ask questions of the FFaIR participants at each briefing and frequently took the opportunity to do so.

Experimental Forecast Shift

Experimental forecast shifts lie at the core of the HMT-Hydro Experiment. These sessions were nominally slated to begin at 3 PM CDT but were flexible based on the present weather scenario. At the latest, forecast shifts ended at 8 PM CDT, although weekly experiment coordinators had wide latitude to dismiss participants early if weather conditions were not conducive to flash flooding. Within forecast shifts, participants were expected to issue experimental flash flood watches and warnings, as necessary, for any portion of the CONUS they believed flash flooding was imminent. Specific characteristics of these experimental watches and warnings are in the subsequent Experimental Datasets section of this report.

Evaluation Session

Forecast evaluation sessions generally took place before the real-time forecasting from Tuesday to Friday of the experiment. In this session, weekly experiment coordinators walked the participants through an online survey with questions about the relative ability of the forecast tools and the observations to properly diagnose the spatial extent and magnitude of the flooding that occurred. They also evaluated the spatial extent and magnitude of the forecast flooding from a new, GFS-based flash flood prediction tool. The FLASH products were also fed with quantitative precipitation forecasts from the HRRR-X model, in attempt to increase lead time. These products were evaluated and compared to those that were forced by MRMS QPE.

All forecast products were displayed in AWIPS II and the forecasters used the Hazard Services software for the issuance of flash flood watches and warnings. The evaluations of the prior days' products, forecaster-issued watches and warnings, and FFaIR guidance products were accomplished through a specifically designed web interface that permitted the overlay of the forecast products on flash flood point observations from NWS local storm reports, USGS stream gauge data, and citizen scientist reports from the mPING app. Each participant had an equal vote in the evaluation of prior days' products and forecasts. TurningPoint™ software and individual clickers were used to collect, display, and archive forecaster ratings for each question.

'Tales from the Testbed'

In association with the NWS WDTD, participants were asked each week to prepare a short presentation on what they learned during their time in the testbed. The WDTD invited all NWS forecast offices, River Forecast Centers (RFCs), and regional centers to join these webinars, which took place Friday afternoons at 12 PM CDT during the experiment. Participants used their webinar time to share tips for how to use various components of the FLASH tool suite in operations, to describe interesting flash flooding cases they encountered during their experimental shifts, and to answer questions about future development work on the FLASH suite and its constituents. Participants were instructed during the Monday training sessions to collect screenshots of interesting or important FLASH tools from AWIPS II as desired throughout the week. Experiment coordinators assisted participants in preparing webinar segments Friday mornings after the final evaluation survey.

Feedback Survey

The final activity of each week of the experiment was a short online feedback survey administered via the Qualtrics system. This feedback survey gave participants a chance to expound on experimental activities including the amount of time assigned to each endeavor, the level of mental stress experienced during various activities, the physical setting and technical set-up of the testbed, and suggestions for improvement in future experiments. The feedback survey consisted of seven questions and one comment box. Participants were asked to rate the Monday introductory activities, the time allotted for various activities, and if they had the appropriate tools to issue experimental products and if discussion and evaluation helped to improve their forecasts. Forecasters ranked their workload during various activities and were asked if they would want to participate in the future or if they would recommend participation to their colleagues.

Experimental Datasets

Forecast Tools

Within AWIPS II, experiment participants had access to a range of operational NWS forecast guidance, including the regular runs of the GFS (Global Forecast System), NAM (North American Mesoscale), and High Resolution Rapid Refresh (HRRR) models. Forecasters also had the ability to view observed soundings from the NWS upper-air network of rawinsondes, surface observations from the national network of ASOS (Automated Surface Observing System) stations, and data from the GOES (Geostationary Operational Environmental Satellite) program. Forecasters had access to the European Centre's global forecast system but were limited to freely available, unencrypted model outputs. Local radar data from the Terminal Doppler Weather Radar (TDWR) or WSR-88D (Weather Surveillance Radar-1988 Doppler) networks was generally not available in AWIPS II due to bandwidth limitations. Outside of AWIPS II, forecasters could access additional tools via web browser or personal device. Forecasters were not instructed not to view any operational flash flood watches and warnings to prevent any bias in their operational decision making.

Several experimental forecast products were made available to the forecasters. They fall into the following two categories: 1) flash flood monitoring and prediction (primarily for the issuance of flash flood warnings) and 2) short-term forecasting (primarily for the issuance of flash flood watches). In category one, forecasters were provided QPE accumulations from the radar-only products in MRMS ranging from 2-min up to 6 hr. These QPEs are compared to the NOAA Atlas 14 precipitation frequencies to compute precipitation average recurrence interval products (ARIs) from 30-min up to 12 hours in duration. The MRMS QPE products were are compared to flash flood guidance (FFG) values produced at NWS RFCs via a mosaic created at WPC. The ratios of QPE/FFG are generated for 1-, 3-, and 6-hr accumulations and are updated every 2 min. Forecasters were also provided access to the CREST maximum unit streamflow products that are valid at each grid cell between thirty minutes prior to the forecast valid time and 12 hours into the future. Note that this CREST product was forced by MRMS radar-only QPEs and QPFs from the HRRR-X model. Category two consists of four tools: two surface – 300 hPa precipitable water analyses, one from rawinsonde observations and one from the RAP weather model. These analyses were compared to a gridded monthly precipitable water climatology developed by M. Bunkers (2014).

QPE and QPF were provided from the MRMS suite of tools (Zhang et al. 2016) and the experimental HRRR-X suite of tools, respectively. Flash flood guidance is provided in gridded format by RFCs across the U.S. (Clark et al. 2014). This mosaic is then compared to MRMS QPE or HRRR QPF and to produce grids of QPE-to-FFG ratio (also referred to as FFG ratio). Precipitation average recurrence intervals (ARIs) consist of MRMS QPE grids compared to ARI grids from NOAA Atlas 14 (Perica et al. 2013). NOAA Atlas 14 analyses are not yet available for states in the Pacific Northwest, northern Intermountain West, 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. The comparisons of QPE to FFG and ARI along with the products related to the hydrologic models (CREST, SAC-SMA, and hydrophobic) are provided by the FLASH system (Gourley et al. 2017).

Finally, two MRMS radar reflectivity factor mosaics were provided to participants. The first, “MRMS Quality-Controlled Composite Reflectivity”, consists of the maximum reflectivity factor value, regardless of vertical level, at each grid point. The second, “MRMS Seamless Hybrid-Scan Reflectivity”, consists of the reflectivity factor at the lowest unblocked vertical level at each grid point. Appendix B contains names of and basic information about each of these tools. The full MRMS product suite was also available, but was confined to a floating domain that could be directed to areas of interest via FFaIR briefings and forecaster requests.

Observations

During the experiment, three separate sources of flash flood observations were available to participants and staff: automated streamgage measurements collected by the United States Geological Survey (USGS), Local Storm Reports (LSRs) collected by NWS WFOs, and unsolicited public geolocated smartphone or mobile phone reports from the mPING (Meteorological Phenomena Identification Near the Ground) project run by NSSL and OU (Elmore et al., 2014).

USGS streamgages are located on catchments of various sizes across the U.S. In order to qualify for inclusion in this observation database, a flash flood event recorded at a streamgage 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 ($0.9\ m \bullet hr^{-1}$) of the stage (B. Cosgrove 2014, personal communication). Only streamgages with contributing drainage areas of less than 2,000 km^2 are considered.

NWS LSRs are issued during or immediately after a given hazardous weather event (Horvitz 2012). They 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. Flash flooding LSRs will typically include a short description of the exact impact of the reported event in plain English.

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

mPING uses the recent proliferation of GPS-enabled smart phones and other mobile devices to crowd-source surface weather conditions. Users can identify the relative severity of the observed flood using a 1-4 integer scale, where “1” corresponds to the least risk to human life and limb and “4” corresponds to the greatest risk to life and limb. For example, a “1” flood corresponds to a river or creek out of its banks, or flooding in a yard, basement, or over cropland. A “4” flood requires homes, buildings, or cars to be swept away by floodwaters.

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, four types of products are considered: operational flash flood warnings, operational flash flood watches, experimental flash flood warnings, and experimental flash flood watches.

Operational flash flood warnings 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, but in the HMT-Hydro context those caused by heavy rainfall are of chief interest. These products 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 (or AWIPS II, depending on the office), called WarnGen. In the case of HMT-Hydro, Hazard Services software was used for issuing experimental products. The forecaster draws a polygon with as many vertices as needed to accurately encompass the threat. Based on this polygon, WarnGen determines which counties and locations should be in the warning text, produces the appropriate text, and then disseminates the warning.

Operational flash flood watches are used to alert the public that flooding is possible 6–48 hours before an event (Clark 2011). They are also issued at the WFO level and do not cross the boundaries of WFOs. These watches are not polygons in the same sense as *Storm Data* reports or operational flash flood warnings. Instead, watches cover a set of counties or parishes, or in areas with large counties, forecast zones defined at a sub-county level. They are generated operationally in the Graphical Hazards Generator (GHG) software program. Unlike warnings, watches can be issued before they officially enter into effect. Watches contain a generalized non-technical synopsis of the anticipated event. Operational flash flood watches are supposed to be issued when the forecaster's confidence in flooding occurring within two days is between 50–80%. Operational watches were processed to include only those *valid* (not just issued) during some portion of an HMT-Hydro Experimental forecast shift.

Experimental flash flood warnings work similarly to their operational counterpart but with some important differences. In the testbed, WFO boundaries are unimportant. Participants were told to act as a national forecast office; in other words, they were responsible for forecasting and monitoring conditions for flash flooding across the entire CONUS, and experimental warnings could cross WFO boundaries. The investigators used a modified version of Hazard Services that required forecasters to quantify their uncertainty about the magnitude of flooding expected in each polygon. The probability of minor flooding (corresponding to mPING impact classes "1" and "2") ranged from 10–100% and the probability of major flooding (corresponding to mPING impact classes "3" and "4") could be 0–100%, with values 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 30 min to 6 hrs), the default warning length of 3 hrs was set in Hazard Services. In all, 78 experimental warnings were issued during the 3-week forecast period.

Experimental flash flood watches contain elements of their operational equivalents as well as of experimental flash flood warnings. Like operational flash flood watches, experimental watches are larger in area and longer in time than warnings. However, experimental watches are not county-based but are drawn with the same Hazard Services polygon methodology used for warnings. Participants could also draw watch polygons that spanned WFO boundaries, unlike in the operational realm. A final important difference concerns lead time: official NWS watches are valid somewhere 6–48 hrs prior to an event. In the testbed, watches were generally valid for 6–12 hours starting immediately from the time of issuance. Forecasters could issue watches at the beginning of an experimental shift, to catch flooding during that entire shift, or they could issue watches later in the shift to catch flooding they forecast to occur overnight after their shift had ended. There were 14 experimental watches issued during the 2016 HMT-Hydro Experiment.

All of the flash flood observations, tools, and experimentally issued watches and warnings were evaluated by the participants. The evaluation survey consists of these questions: four relating to the forecast tools, two relating to the degree of accuracy and lead time offered by HRRR-X forcings, and six relating to the experimental watches and warnings. A set of questions regarding the spatial coverage and magnitude is asked of the forecast tools (MRMS QPE, QPE ARI, QPE-to-FFG ratio, CREST Max Return Period, GFS Probability Tool). The watch and warning section of the survey consists of five Likert scale questions where experimental watches and warnings are compared to their operational counterparts in the realms of spatial accuracy, uncertainty estimates, and magnitude assignment. A dialog box was customized within Hazard Services so that the participants were required to enter the influence of the experimental products in the issuance of flash flood watches and warnings. Lastly, the spatial accuracy and probability values assigned to the FFaIR-issued excessive rainfall outlooks (EROs) and probability of flash flooding forecasts (PFFFs) were evaluated by the HMT-Hydro participants.

Results

Product Evaluation

Participants were asked to evaluate the spatial accuracy and magnitude of the four MRMS-forced products as compared to reports of flash flooding from the aforementioned observations. Each forecaster supplied a ranking value for the products using the TurningPoint™ software and clickers. Figure 1 shows a summary of the responses for all events evaluated throughout the experiment. All products were ranked similarly with average rankings near 75. The lowest ranking was given to the QPE ARI product with an average ranking of 72. However, each of the products yielded value in the flash flood warning and decision making process.

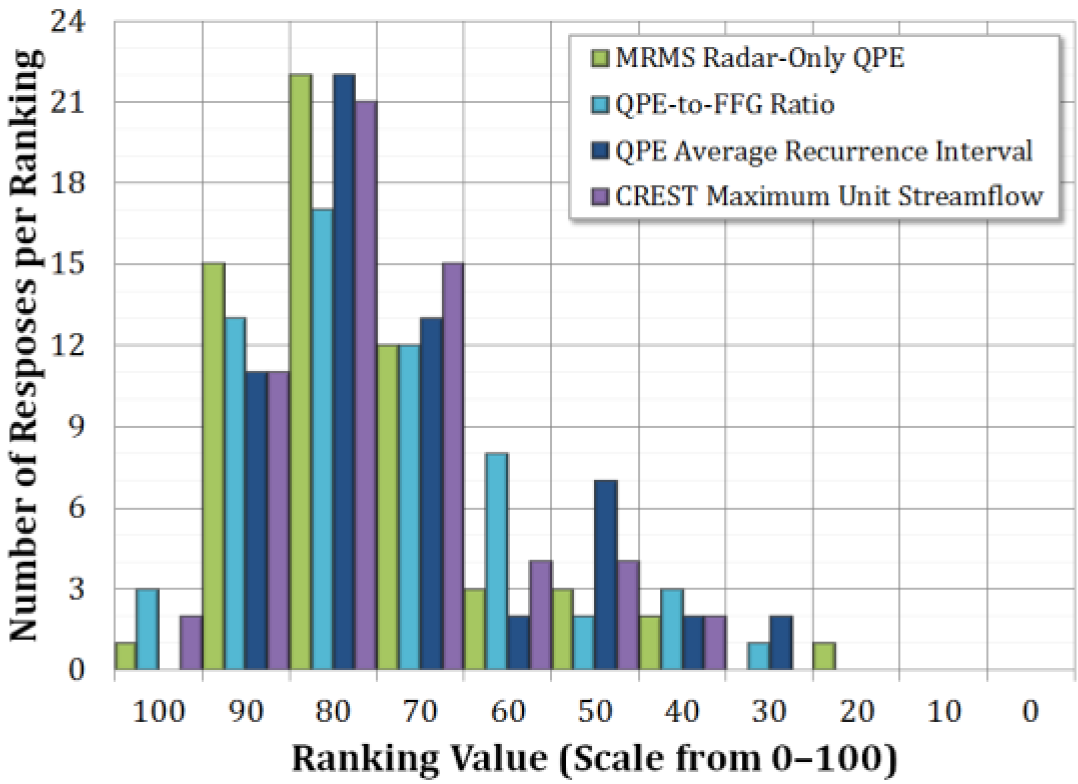

Figure 1. Forecaster rankings of the spatial coverage of the flash flood impacts for the FLASH product.

Figure 2 shows the rankings for the evaluated products as they revealed the magnitude of flash flooding impacts. In this case, there was more disparity in the rankings of the products. The MRMS Radar-Only QPE was ranked the highest with an average of 73 while the QPE ARI was ranked the lowest with an average of 59. The CREST unit streamflow product had an average ranking of 67 and the QPE-to-FFG ratio product was ranked at 61. Nonetheless, each of the products showed some capability to provide information in the identification of locations and severity of flash floods. One recommendation is to continue providing support for each of the products.

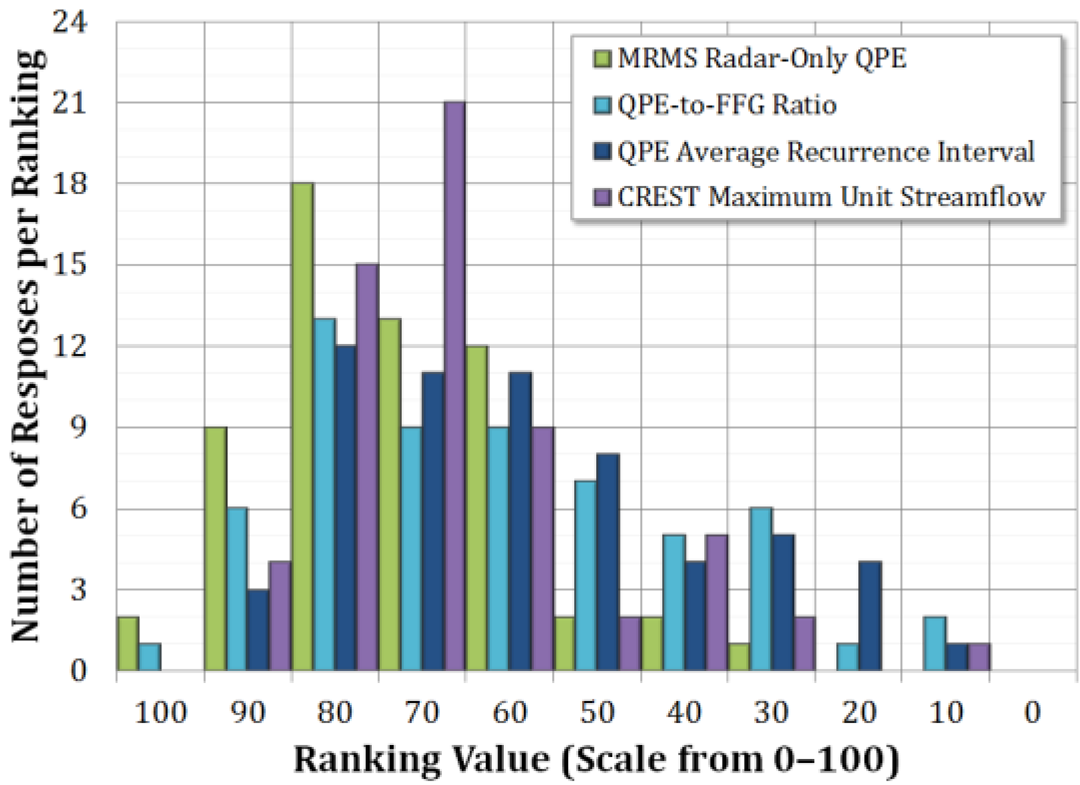

Figure 2. Forecaster rankings of the magnitude of the flash flood impacts for the FLASH product.

A new, experimental product that was developed specifically for the HMT-Hydro 2016 experiment was the GFS Prediction Probability Tool. This is a machine-learning product trained on GFS variables and observed flash flood LSRs. It is available globally but the training dataset is specific to the U.S. It also differs from the other tools in that it provides several hours of forecast lead time. Figure 3 shows the participants' responses to the following statement: "The spatial accuracy of the GFS prediction probability forecast for the previous day was skillful." The responses indicate that most of the forecasters disagreed with the statement and saw little value in the GFS-based flash flood forecasts. This is not too surprising given the lack of hydrology in the machine-learning approach. Furthermore, participants noted some skill in the probabilities when the flash flooding events had synoptic scale forcing that was well represented and forecast by the GFS. The tool was much less skillful for the smaller scale events, which were more numerous.

Figure 4 shows the participants' responses to the following statement: "The probability values of the GFS prediction probability forecast for the previous day were accurate." The rankings were better with the magnitude assessment of the tool as compared to the spatial accuracy. However, the distribution is approximately normal with a mean response of neutral. As with the spatial accuracy rankings, forecasters noted some skill with the tool for the events that were strongly forced at synoptic scale, but not with the smaller scale events. The assessments indicate that additional research needs to be conducted with the machine-learning approach that presently relies on GFS variables alone. Future approaches could be developed with the HRRR-X model and/or hydrologic model outputs.

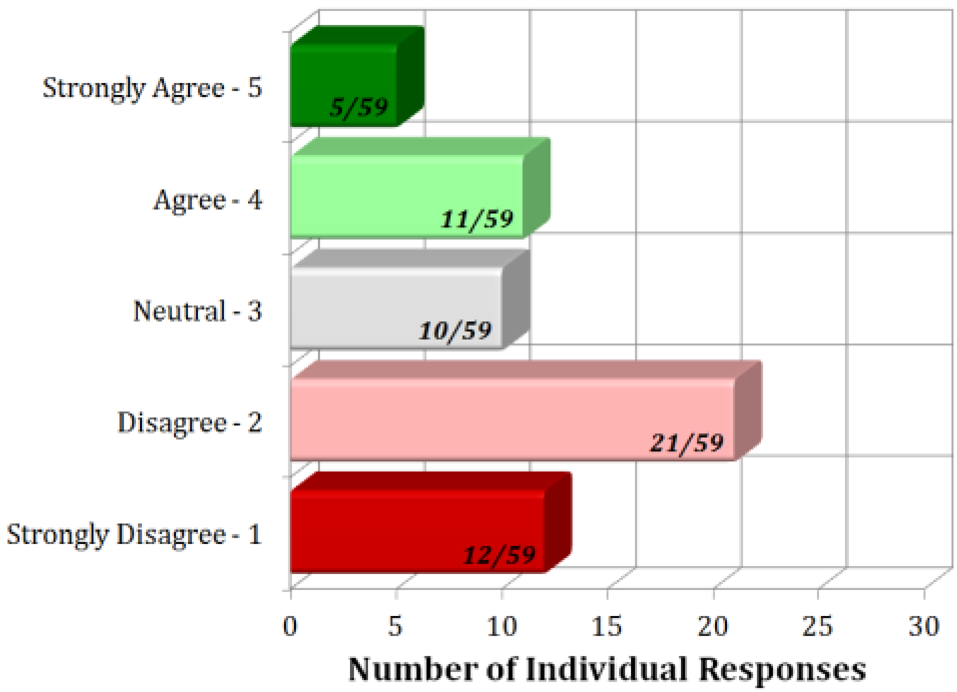

Figure 3. Forecaster rankings of the spatial accuracy of the GFS prediction probability tool.

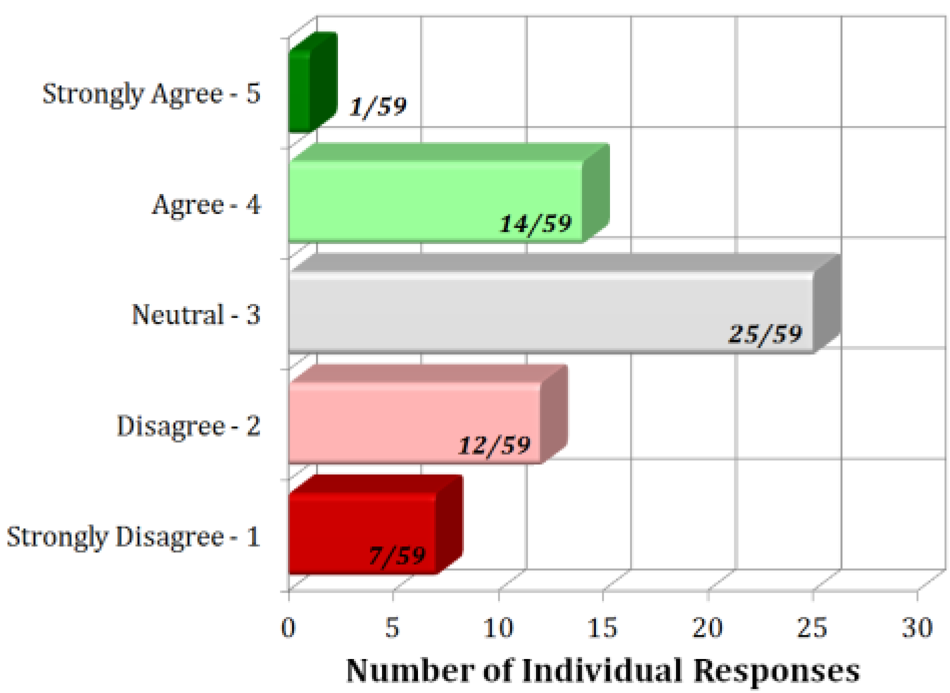

Figure 4. Forecaster rankings of the magnitude of the GFS prediction probability tool.

QPF forcings from the HRRR-X model were input to the CREST model during forecast periods. The MRMS rainfall estimates were used during prior times up to the analysis period. Forecasters considered all aspects of the CREST unit streamflow product including detection, false alarming, spatial accuracy, and magnitude. Figure 5 shows how the forecasters ranked the QPF-forced product relative to the QPE-forced CREST unit streamflow. In general, forecasters rated the QPF-forced product either slightly better or about the same as the QPE-forced one. More positive results were noted in larger, or synoptic, scale events, while the HRRR-X was rated less favorably with mesoscale events or isolated convection, largely due to placement of related to model initiation of convection. In the events where the forecasters noted that there was some skill in the QPF-forced hydrologic products, they were asked to assess how much lead time was provided. Figure 6 shows that there was very little lead time offered by the HRRR-forced product; however, 27 out of 59 cases studied yielded some improvements in lead time up to at least 30 min. HMT-Hydro experiments in prior years have also evaluated the utility of QPF forcings for flash flood warning purposes. This is the first year at which the results indicated there was some utility in identifying flash flood cases.

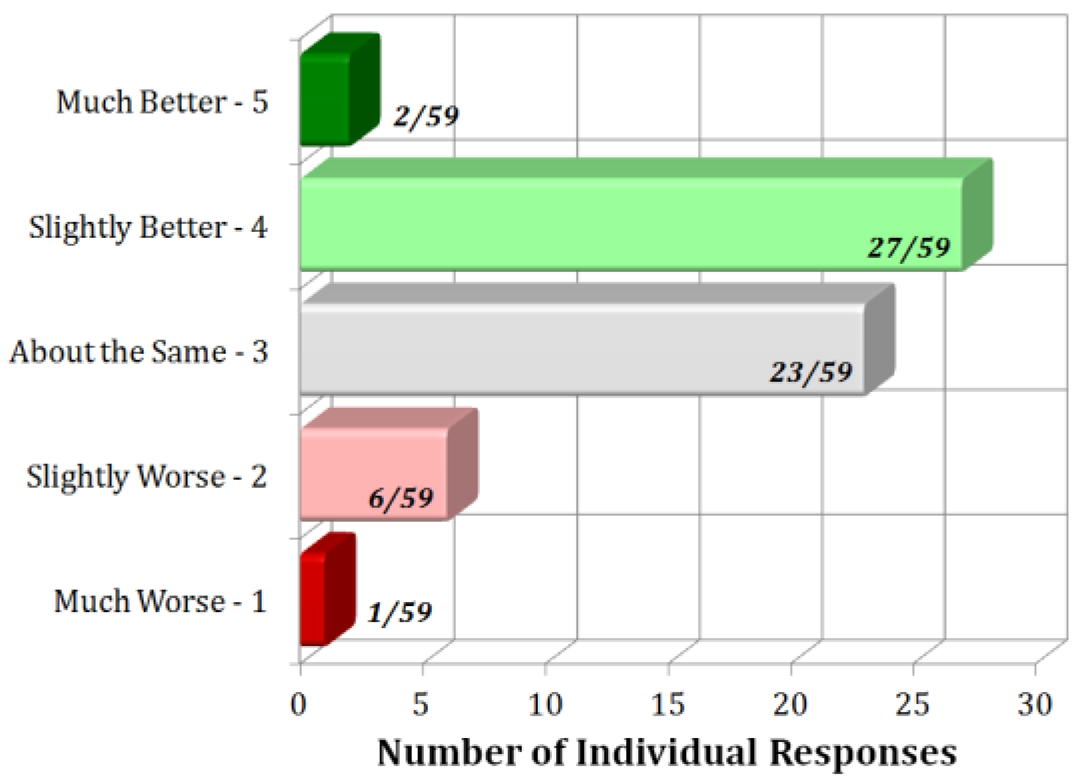

Figure 5. Forecaster rankings of HRRR-X-forced CREST unit streamflow product relative to the one forced by MRMS alone.

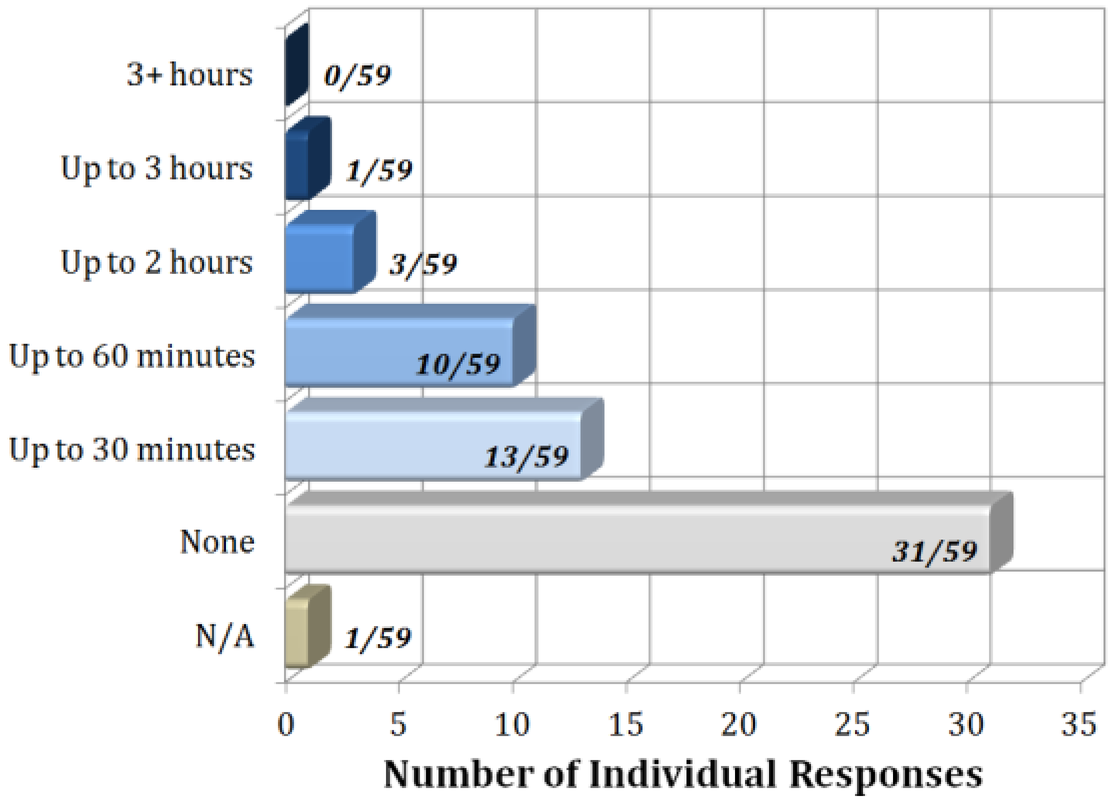

Figure 6. Forecaster assessments of lead time offered by the HRRR-X-forced CREST unit streamflow product relative to the one forced by MRMS alone.

Evaluation of Experimental Watches and Warnings

Forecasters assessed the spatial coverage of experimentally issued flash flood watches and warnings as compared to those that were issued operationally by local forecast offices. During the experiment, participants did not have access to the operationally issued flash flood watches or warnings. Otherwise, it would have been much easier to improve the spatial coverage of a polygon that had already been designated. There are some important differences between the experimental and operational flash flood watches and warnings. First, experiment participants had unique access to the FLASH products and were encouraged and trained to use those during their decision-making process. Hazard Services enables the issuance of watches and warning without regard to county warning areas. Operational flash flood watches are generally issued several hours or even days prior to an event. In the case of HMT-Hydro, forecasters were encouraged to issue experimental flash flood watches in a similar manner as severe thunderstorm and tornado watches are issued, on the order of 6 hours prior to the anticipated event. Participants were operating in regions of the U.S. that were often unfamiliar to them and thus did not have local knowledge about streams that are known to be particularly susceptible to flash flooding. Lastly, participants were asked to evaluate their own products relative to those that were issued operationally, so the evaluation was not completely independent.

Figure 7 shows the rankings of the spatial accuracy of the experimentally issued flash flood watches relative to the operational ones. As with prior HMT-Hydro experiments, the flash flood watches were ranked significantly higher than the operational ones. As many as 17 watches out of 42 were ranked a 5 (Much Better). In general, operational flash flood watches were issued prior to the experimental ones, which contributes some to the improved spatial coverage. Figure 8 reveals that the experimentally issued flash flood warnings were generally not as accurate in terms of spatial coverage as the operational ones. This finding is consistent with prior years' findings. Some of differences are attributable to the advantage of having local knowledge contribute to the decision-making process. However, it is noted that the forecasters who issued the operational flash flood warnings were often involved in the process of collecting local storm reports to validate them. Thus, there is some dependence between the forecast product and observations used for validation.

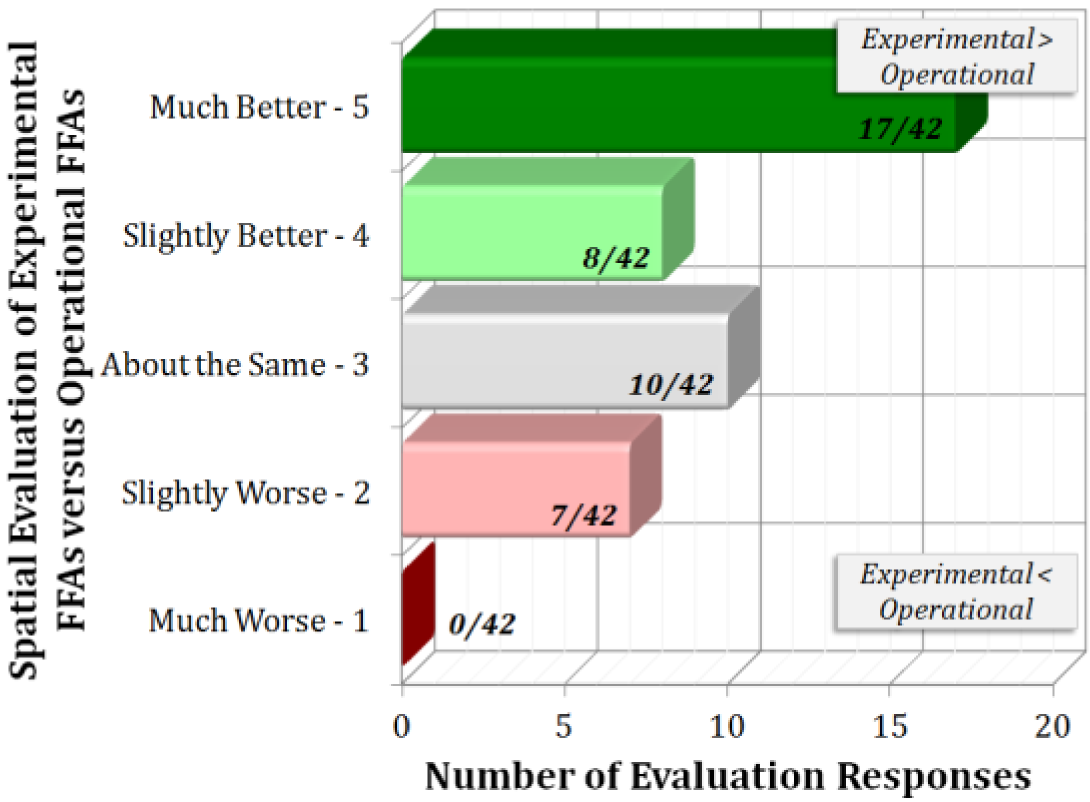

Figure 7. Forecaster rankings of experimental flash flood watches relative to those that were issued operationally.

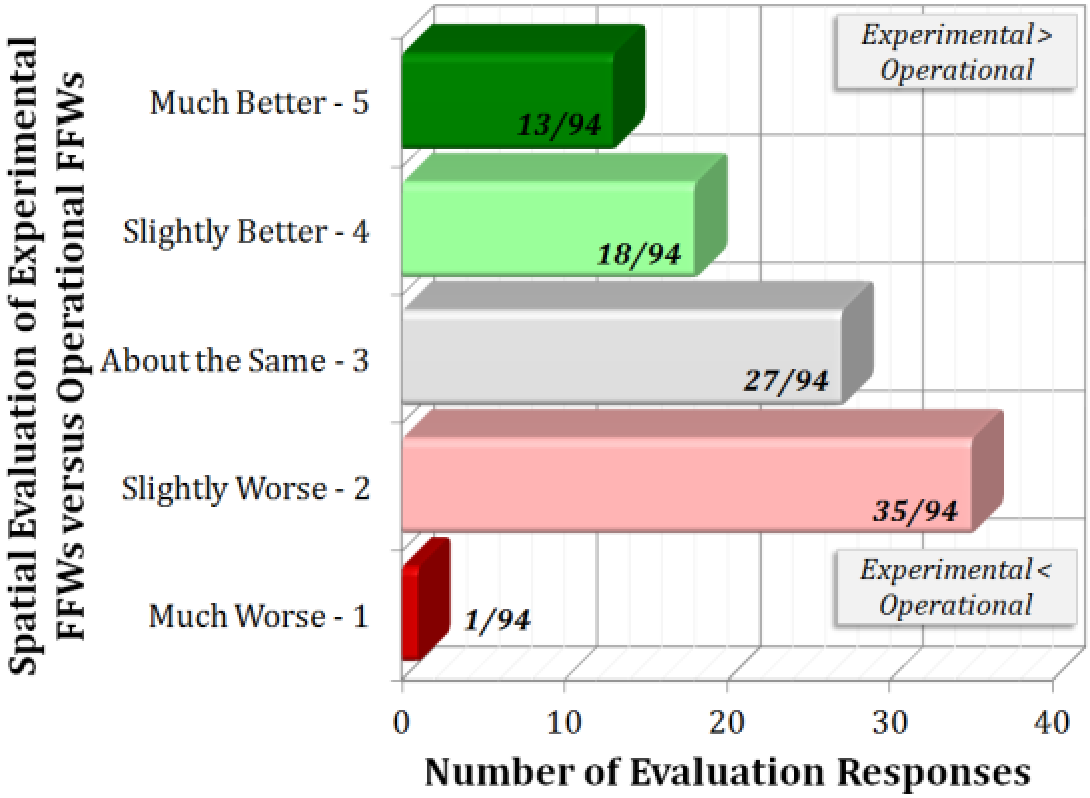

Figure 8. Forecaster rankings of experimental flash flood warnings relative to those that were issued operationally.

Assigned Magnitudes and Probabilities to Experimental Watches and Warnings

A unique aspect of the HMT-Hydro experiment is the requirement that forecasters must assign a probability to the experimental flash flood watches and warnings of being associated to both minor and major impacts. The details contained within the LSRs were used to subjectively assign the impact severity. Figure 9 shows a reliability diagram of the probability assignments for flash flood watches. The points and lines in gray shade show results from 2014 and 2015. In general, there is reasonable reliability with the probabilities but with a slight tendency to assign too low probabilities to the major events. Figure 10 reveals that the probability assignments to the flash flood warnings were quite reliable. Furthermore, the tendency to assign too high probabilities to the minor events has largely been mitigated compared to results from prior years; however, some of the overestimation of probabilities for experimental warnings could be attributed to the dependence on the NWS LSR verifications (i.e., verification is usually not available for experimental warnings that occurred where an operational warning did not exist).

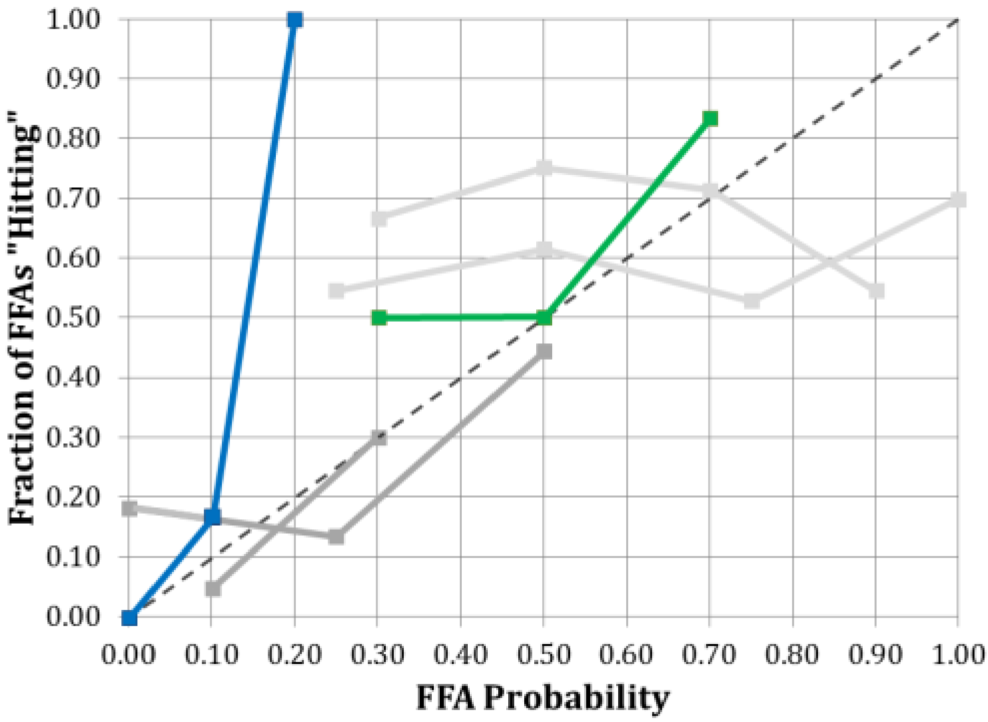

Figure 9. Objective assessment of the reliability of experimentally issued flash flood watches for major (blue) and minor (green) flash flood events.

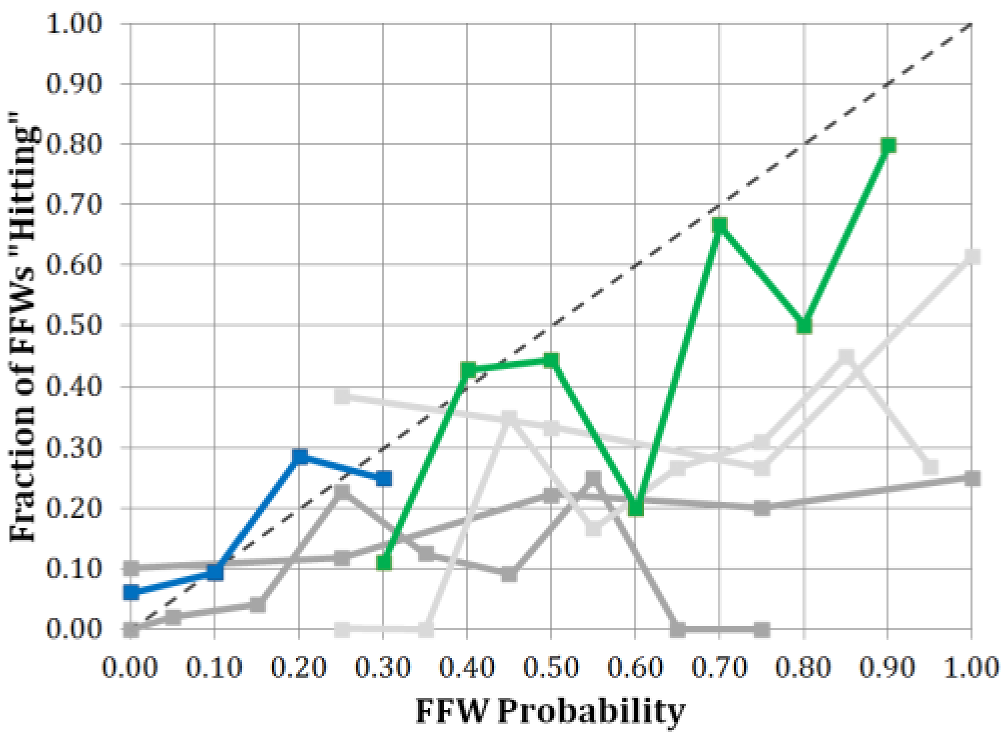

Figure 10. Objective assessment of the reliability of experimentally issued flash flood watches for major (blue) and minor (green) flash flood events.

Consideration of products used in decision-making process

The Hazard Services software was modified by developing a GUI template that prompted the participants to record the products they used in their decision-making process for issuing flash flood watches and warnings. Figure 11 shows the responses for products that contributed to the issuance of experimental flash flood watches. The top three products that were used most frequently were the FFaIR-issued excessive rainfall outlook, meteorological ingredients, and precipitable water values. None of the QPE-forced FLASH products were considered by forecasters for issuing flash flood watches, since they were not intended for the watch phase. Figure 12 reveals how forecasters considered all four MRMS and FLASH tools when issuing flash flood warnings. The greatest consideration was given to the two products that were most familiar to the forecasts: MRMS QPE and the QPE-to-FFG ratio product. The least considered products was the rainfall ARI product, owing to less confidence in the product values; however, the forecasters noted how important it is in a situational awareness sense. These results are consistent with the subjective evaluation of the products provided Figs. 1 and 2. The CREST maximum unit streamflow was shown to be influential in the decision making process, especially in urban areas where a signal in QPE-to-FFG ratio or QPE ARI products would be much less than expected.

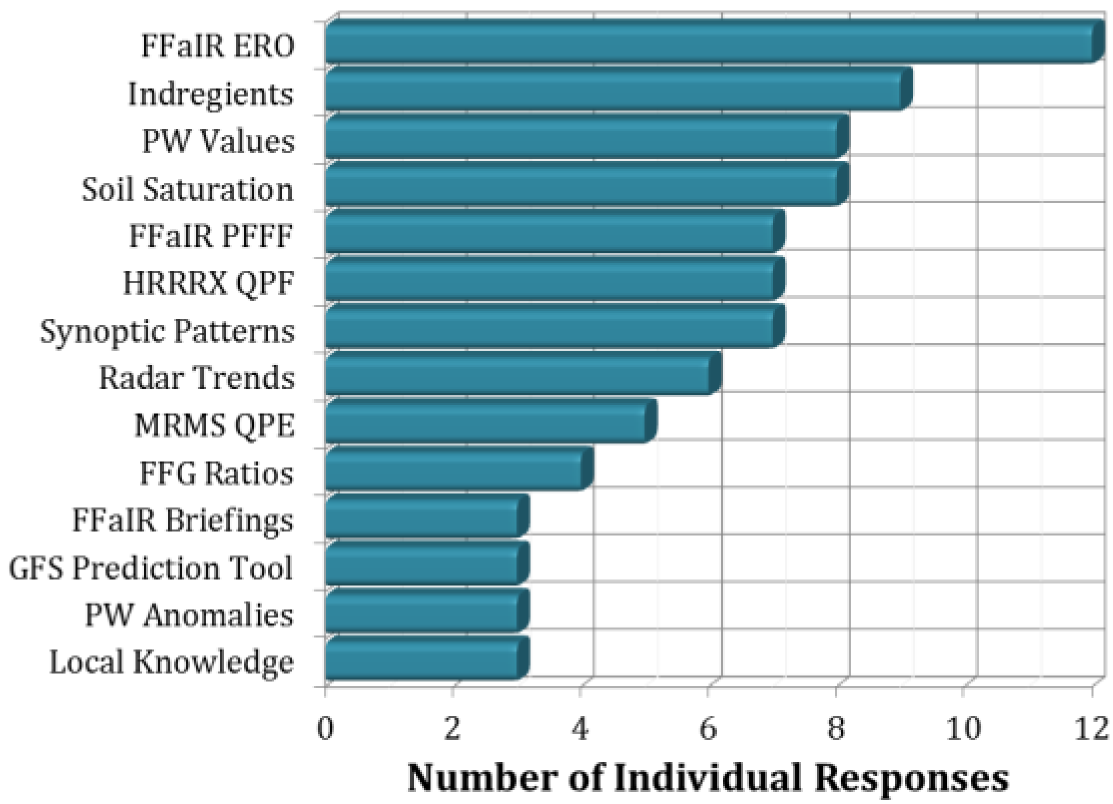

Figure 11. Products that were used by participants to issue flash flood watches.

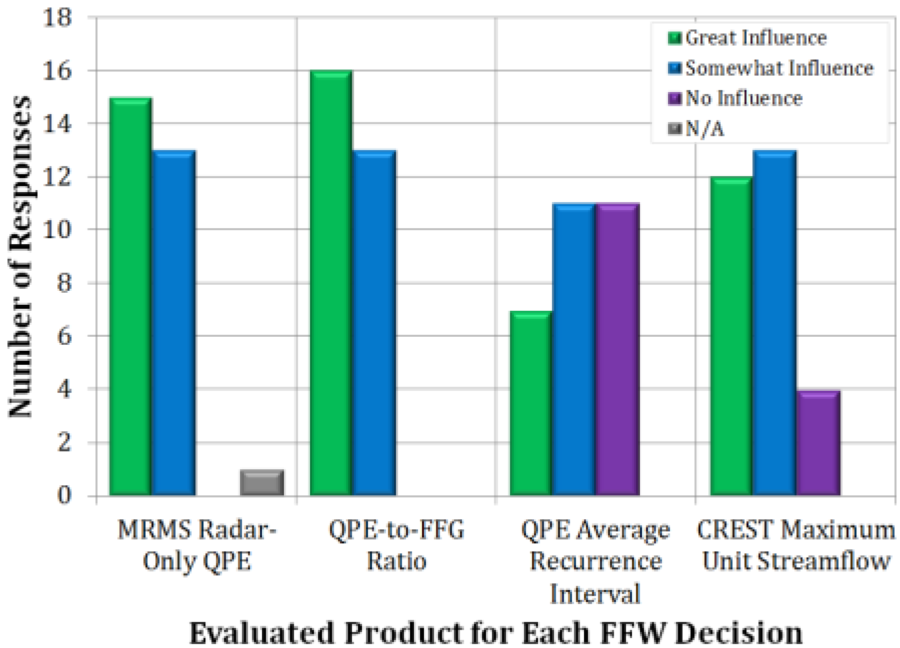

Figure 12. The influence of the FLASH products on issuing experimental flash flood warnings.

Evaluation of warning lead time and coverage area

The lead times and warning area associated with the experimental flash flood warnings were also assessed. Of the 25 isolated flash flood warning events that were studied, 14 of them has positive lead time increases. The average warning lead time increase for all events was six minutes, yet there were five instances where the lead time compared to operation flash flood warnings were at least 40 minutes longer. Analysis of the polygon warning area was conducted for isolated events (i.e., a threat are was contained by a single polygon and not a series of polygons). From a total of 12 events, the experimental warnings had an area that was 705 km^2 larger than the collocated operational warnings; however, five of the 12 warnings had a smaller warning area (i.e., smaller false alarm area) than the collocated operational warnings. Three warnings had a smaller polygon by 1000–3000 km^2. These numbers would have to take into consideration that the participating forecasters were unfamiliar with the areas they were working in (i.e., lacked local knowledge of area and flashiness of some basins) and were dependent of verification from local NWS WFOs.

Evaluation of FFaIR-issued guidance products

The HMT-Hydro participants utilized the FFaIR excessive rainfall outlooks and probability of flash flooding products in their decision-making process. They also evaluated the products in terms of their spatial accuracy, probability assignment, and their overall utility in benefitting situation awareness and decision-making. The forecasters were asked to rate the following statement: “The spatial accuracy of the Day 1 FFaIR Excessive Rainfall Outlook for the previous day was skillful”. Figure 13 indicates that the HMT-Hydro participants considered the ERO to be quite skillful. This result is consistent with

Figure 11 in that the forecasters placed high confidence with the ERO in guiding the placement of flash flood watches. Figure 14 shows an evaluation of the probability assignments to the ERO product. Forecasters generally agreed that the probabilities assigned to the ERO product were accurate. One facet of the bridging between the HMT-Hydro and FFaIR experiments is a daily weather briefing. The weather briefing typically begins by showing products and tools that are primarily based on observational data. The intent is to improve situation awareness amongst the HMT-Hydro participants. The participants were later asked to rank the FFaIR weather briefings and products as they pertained to improving situation awareness. Similar to the findings with the ERO product, HMT-Hydro participants generally agreed that the FFaIR weather briefing and products increased their situation awareness as they began their forecasting shifts. The responses were more neutral, however, when asking how the FFaIR products influenced the issuance of flash flood watches (Fig. 16).

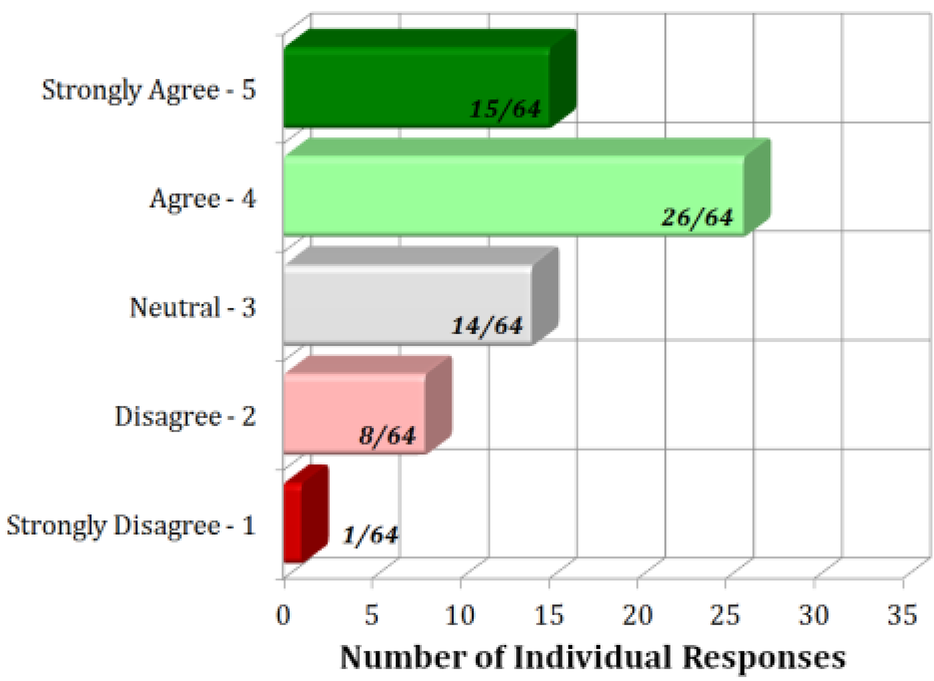

Figure 13. Evaluation of the spatial accuracy with the FFaIR-issued Excessive Rainfall Outlook for flash flood forecasting.

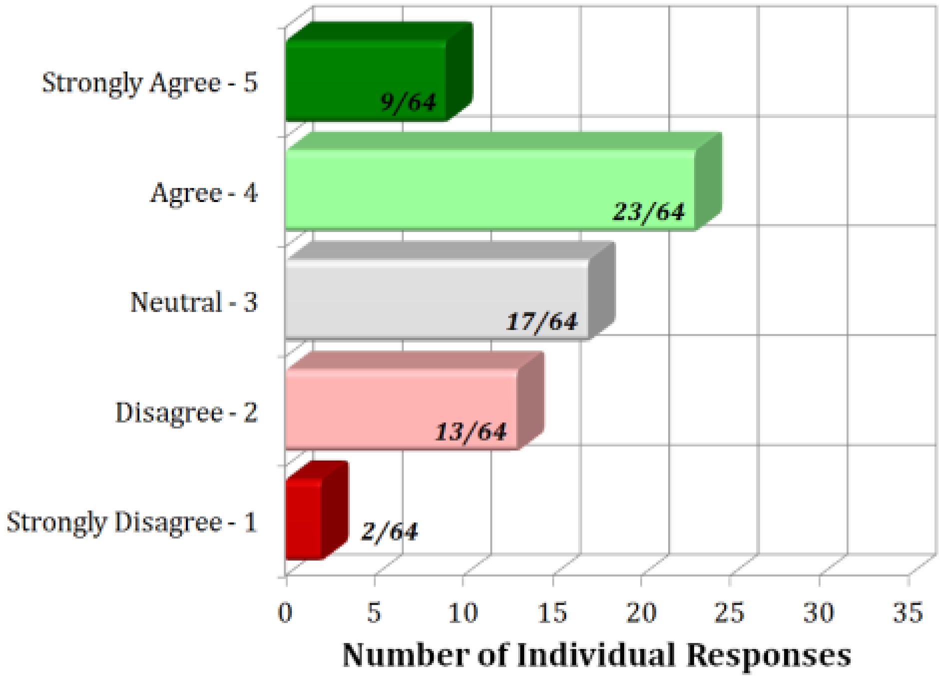

Figure 14. Evaluation of the probability assignments with the FFaIR-issued Excessive Rainfall Outlook for flash flood forecasting.

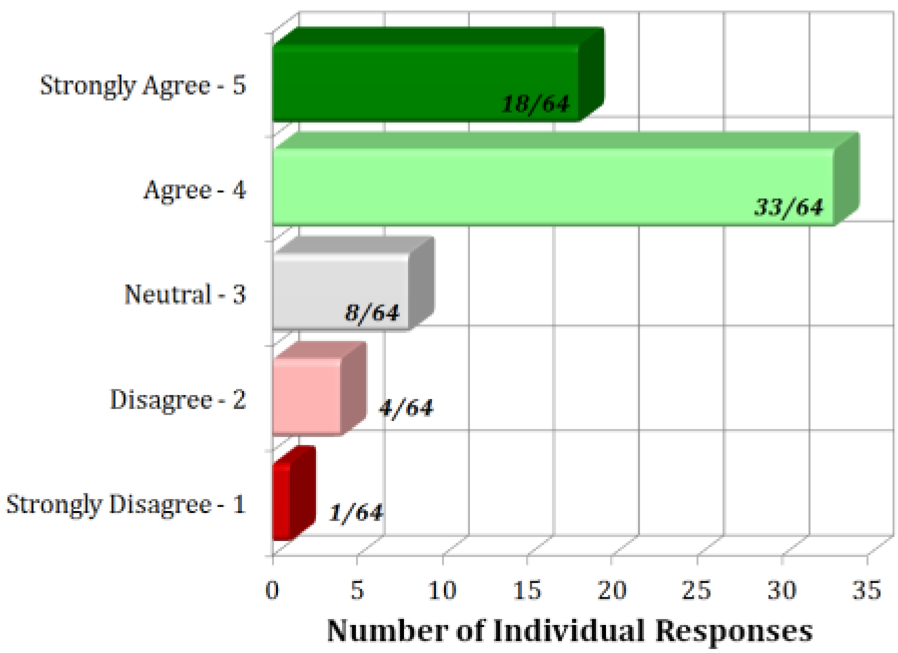

Figure 15. Evaluation of the weather briefings and products provided by FFaIR as they pertained to improving situation awareness.

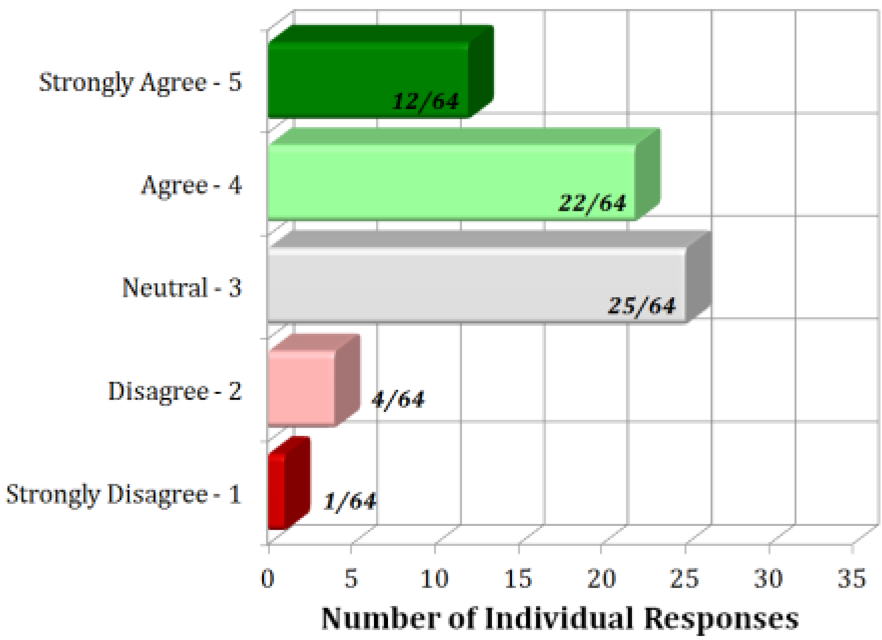

Figure 16. Evaluation of the weather briefings and products provided by FFaIR as they pertained to issuing flash flood watches.

Results of the Feedback Survey

HMT-Hydro participants were asked to fill out a feedback survey on the overall functioning of the experiment. The detailed results are provided in Appendix D. The responses indicate favorable evaluations on the training materials, tools provided, workload, and time allocated to the various tasks. In the written section, forecasters noted that there were some technical limitations and constraints related to AWIPS II, Hazard Services, and the data feeds. Being in a testbed environment, there were instances in which the data feeds were down. They also asked for more concise FFaIR weather briefings, more in-depth information about the FLASH products, and adaptive forecast shifts to better capture the entire events.

Analysis and Recommendations

For Operations

Participants were required to assign probability values for minor and major flash flooding impacts for watches and warnings. The assignment of these probabilities has shown improvement over the last three years. As such, a recommendation is to consider issuing these probabilities in operations to provide more information to end-users. These forecast products were issued using Hazard Services software. This was the second summer for HMT-Hydro to use Hazard Services. The functionality had improved and will become a necessary tool for issuing contemporary products, such as probability assignments for watches and warnings.

Participants issued flash flood watches and warnings across the CONUS. While the sample is rather small, the spatial accuracy of experimentally issued flash flood watches was better than those issued on an operational basis. The experimental flash flood warnings were not as accurate though, presumably due to specific knowledge by local forecasters; however, there were some improvements in lead time and warning area with the experimental flash flood warnings.

The FFaIR weather briefings and issued products were well received by the HMT-Hydro participants. The spatial coverage of the excessive rainfall outlooks were rated well, slightly higher than the probability values that were assigned to them. Participants noted that the FFaIR weather briefings and products improved their situation awareness, but did not necessarily guide their issuance of flash flood watches.

A major limitation of the HMT-Hydro experiment has always been the dearth of observations to completely describe the spatial coverage and specific impacts of flash flooding. Prior years had used the SHAVE experiment to collect additional, independent reports on warned events, but this experiment has come to an end. Promotion of the mPING project via NWS text products, NWS social media, and the NWS website should be undertaken, as this app allows the wisdom of the crowds to be leveraged into appropriately identifying and classifying flash flood events largely independent of watches and warnings, unlike the *Storm Data* publication or LSRs.

For Tool Development

All MRMS and FLASH products provided utility in identifying the spatial coverage and magnitude assigned to flash flooding events. They should all be supported. There was a slight preference toward the use of the MRMS QPE and CREST unit streamflow products for spatial accuracy and magnitude assessments. The CREST unit streamflow product has been rated increasingly higher with each year as the product improved. Furthermore, prior experiments had established thresholds that have now been incorporated in training materials. The GFS Probability Prediction tool was still in an early development stage. It general, it was not rated very highly, but did provide useful information for the synoptically forced events. Future research should focus on developing machine-learning approaches on the HRRR forecast variables and consider guiding the forecasts further with hydrologic model outputs.

The HRRR-X QPFs were used as inputs to the CREST model as they had been used in previous years. There was a noted increase in the utility of these products in terms of providing some forecast lead time. Future experiments should consider using forcings from an ensemble of QPFs that are produced at flash flood scale.

For Future Iterations of HMT-Hydro

The inaugural HMT-Hydro Experiment was held in the month of July and extended into August; June or July is recommended for future experiments. The summer allows for the inclusion of monsoon-driven events in the Desert Southwest (over three-quarters of the experimental shifts in HMT-Hydro had some sort of activity in this area). The summer also allows for close coordination with the FFaIR experiment and avoids interfering with springtime severe convection studied by other experiments under the HWT umbrella.

Despite a number of flash flooding events during the 2016 HMT-Hydro Experiment, some days were notably slow. This is inevitable in this sort of research, and the experiment administrators should develop at least one – and probably two – displaced real time AWIPS II flash flood simulations for this eventuality. These simulations should showcase positive and negative aspects of the experimental tools and should require the length of an experimental shift to complete.

Acknowledgements

HMT-Hydro was funded by NOAA/OAR/Office of Weather and Air Quality (OWAQ) under the NOAA cooperative agreement, NA11OAR4320072. Regional NWS headquarters also provided some funding for certain participants. Tiffany Meyer (OU/CIMMS) provided assistance with establishing data feeds and displays in AWIPS II and Hazard Services. This experiment would not have been possible without the enthusiastic whole-hearted participation of the NWS forecasters from around the country.

References

Bangor, A., P. Kortum, and J. Miller, 2009: Determine what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, **4**, 114–123.

Barthold, F., and T. Workoff, 2014: The 2014 Flash Flood and Intense Rainfall Experiment. Hydrometeorological Testbed at the Weather Prediction Center. [Available online at http://www.wpc.ncep.noaa.gov/hmt/FFaIR_2014_final_report.pdf.]

Brooke, J., 1996: SUS - A ‘quick and dirty’ usability scale. *Usability Evaluation in Industry*, P. Jordan, B. Thomas, B. Weerdmeester, and I. McClelland, Ed., Taylor & Francis, 189–194.

Bunkers, M., 2014, cited 2014: Upper-Air Climatology Plots. US Department of Commerce, National Weather Service Rapid City, SD Weather Forecast Office. [Available online at www.crh.noaa.gov/unr/?n=pw.]

Burnash, R., R. Ferral, and R. McGuide, 1973: A generalized streamflow simulation system – conceptual modeling for digital computers. US Department of Commerce, National Weather Service and State of California, Department of Water Resources.

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.

DeGaetano, A. and D. Zarrow, 2010, cited 2014: Extreme Precipitation in New York & New England: Technical documentation & user manual. Northeast Regional Climate Center, Cornell University. [Available online at http://precip.eas.cornell.edu/docs/xprecip_techdoc.pdf.]

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., J. M. Erlingis, T. M. Smith, K. L. Ortega, and Y. Hong, 2010: Remote collection and analysis of witness reports on flash floods. *J. Hydrol.*, **394,** 53–62, doi: 10.1016/j.jhydrol.2010.05.

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.]

Likert, R., 1932: A technique for the measurement of attitudes. *Archives of Psychology*, **140**, 1–55.

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.]

Ortega, K.L., T.M. Smith, K.L. Manross, K.A. Scharfenberg, A. Witt, A.G. Kolodziej, and J.J. Gourley, 2009: The severe hazards analysis and verification experiment. *Bull. Amer. Meteor. Soc.*, **90,** 1519–1530.

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.]

Sauro, J., 2011, cited 2014: Measuring usability with the System Usability Scale (SUS), Measuring Usability, LLC. [Available online at http://www.measuringu.com/sus.php.]

Reed, S., J. Schaake, and Z. Zhang, 2007: A distributed hydrologic model and threshold frequency-based method for flash flood forecasting at ungauged locations. *J. Hydrol.*, **337**, 402–420.

Wang, J., Y. Hong, L. Li, J. J. Gourley, S. Khan, K. Yilmaz, R. Adler, F. Policelli, S. Habib, D. Irwn, A. Limaye, T. Korme, and L. Okello, 2011: The coupled routing and excess storage (CREST) distributed hydrologic model. *Hydrolog. Sci. J.*, **56**, 84–98.

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

Week	Name	Affiliation
1	John Goff	NWS Burlington, VT
1	Nick Greenawalt	NWS Syracuse, NY
1	Mike Hardiman	NWS El Paso, TX
1	Greg Heavener	NWS Corpus Christi, TX
1	Lara Pagano	NWS Morehead City, NC
2	Glenn Carrin	NWS Morristown, TN
2	Derek Giardino	NWS West Gulf RFC
2	Eric Seymour	NWS Wakefield, VA
2	Tina Stall	NWS San Diego, CA
2	Aaron Treadway	NWS New Braunfels, TX
3	Jessica Brooks	NWS Davenport, IA
3	Joseph Dandrea	NWS San Diego, CA
3	Stephen Hrebenach	NWS Wilmington, NC
3	Adrienne Leptich	NWS Upton, NY
3	Patrick Sneeringer	NWS West Gulf RFC
3	Robert Stonefield	NWS Blacksburg, VA

Name	Title	Affiliation	Week
J. J. Gourley	Principal Investigator	NOAA/OAR/NSSL	All
Steven Martinaitis	Experiment Coordinator	OU/CIMMS	All
Race Clark	Logistics Coordinator	OU/CIMMS	All
Zac Flamig	Technical Coordinator	OU/CIMMS	All
Gabe Garfield	HWT Coordinator	OU/CIMMS	All
Tiffany Meyer	HWT Technical Support	OU/CIMMS	All
Michael Bowlan	WDTD Seminar Coordinator	OU/CIMMS	All
Ami Arthur	Weekly Coordinator	OU/CIMMS	2
Manab Saharia	Weekly Coordinator	OU	1
Daniela Spade	Weekly Coordinator	OU	3

Appendix B: Experimental Tools Used in HMT-Hydro

Tool Name (as shown in AWIPS II)	Tool Category	Additional Versions Available	Units
HRRR-Forced CREST Unit Streamflow	Hydrologic Model		$ft^3 \bullet s^{-1} \bullet mi^{-2}$
GFS Probability of flooding tool	Machine-learning algorithm		%
CREST Soil Moisture	Hydrologic Model		%
CREST Streamflow	Hydrologic Model		$ft^3 \bullet s^{-1}$
CREST Unit Streamflow	Hydrologic Model		$ft^3 \bullet s^{-1} \bullet mi^{-2}$
SAC-SMA Soil Moisture	Hydrologic Model		%
SAC-SMA Streamflow	Hydrologic Model		$ft^3 \bullet s^{-1}$
SAC-SMA Unit Streamflow	Hydrologic Model		$ft^3 \bullet s^{-1} \bullet mi^{-2}$
MRMS Radar-Only QPE	QPE/QPF	4: Instantaneous rate, 1-, 3-, and 6-h	in or $in \bullet hr^{-1}$
HRRR QPF	QPE/QPF	3: 1-, 3-, and 6-h	in
MRMS Radar-Only QPE to FFG Ratio	FFG	4: 1-, 3-, 6-h, and maximum of any	%
Precipitation Return Period	Precipitation Return Period	6: 1-, 3-, 6-, 12-, 24-h, and maximum of any	Year
Precipitable Water Analysis	Precipitable Water	2: RAOBs or RAP	in
Precipitable Water Standard Anomalies	Precipitable Water	2: RAOBs or RAP	Unitless
MRMS Quality-Controlled Composite Reflectivity	Radar		dBZ
MRMS Seamless Hybrid-Scan Reflectivity	Radar		dBZ

Appendix C: Tips for Displaying Tools in AWIPS-II

A substantial portion of the planning for the HWT-Hydro Experiment was devoted to developing the capability to display the FLASH suite in AWIPS II. AWIPS II consists of two main components: a display interface – CAVE (Common AWIPS Visualization Environment) –and the data server component – EDEX (Environmental Data Exchange System). Development of FLASH display capabilities was undertaken on a standalone (i.e., EDEX and CAVE running on the same computer) installation of AWIPS II version 14.1.1. Attempts were made at using version 14.2.X for HWT-Hydro but later versions of the code proved buggy and, due to time constraints, HWT-Hydro proceeded with the older version.

In the testbed, there were four CAVE workstations running Red Hat Enterprise Linux version 6 and a fifth computer acting primarily as an EDEX server but with CAVE enabled so that experiment staff could diagnose any problems reported by participants. These computers were equipped with 48 Gb of random access memory (RAM) and 16 Intel Xeon processing cores. One of the four CAVE workstations had only 16 Gb of RAM and was noticeably slower than the other three, according to participants. Some even referred to it as nearly unusable. The EDEX server was equipped with a conventional 1 Tb hard drive. Based on monitoring of the EDEX processor and memory usage, the hard drive acted as a speed bottleneck for most of the experiment, and solid-state or hybrid hard drives are recommended for EDEX servers whenever economically possible. (Additionally, forecasters suggested that at least two 22" diagonal monitors per workstation are desirable.) Several modifications to EDEX are possible to increase speed and reduce chances of extreme latency. FLASH tools were brought into AWIPS II as GRIB2 files during the experiment. AWIPS II plugin `com.raytheon.edex.plugin.grib.properties` was modified to read:

```
grib-decode.count.threads=10
```

The `com.raytheon.uf.edex.datadelivery.bandwidth.properties` plugin was modified as follows:

```
bandwidth.dataSetMetaDataPoolSize=6
bandwidth.retrievalPoolSize=12
bandwidth.subscriptionPoolSize=12
```

In `ingestGrib.sh`, the following modifications were made:

```
export INIT_MEM=1024 # in Meg
export MAX_MEM=8196 # in Meg

export JMS_POOL_MIN=4
export JMS_POOL_MAX=24
export METADATA_POOL_MIN=4
export METADATA_POOL_MAX=16
export EDEX_DEBUG_PORT=5007
export EDEX_JMX_PORT=1618
export MGMT_PORT=9603
```

In `request.sh`, the following modifications were made:

```
export INIT_MEM=128 # in Meg
if [ "$EDEX_ARCH" == "64-bit" ]; then
```

```
    export MAX_MEM=4096 # in Meg
else
    export MAX_MEM=1280 # in Meg
fi
export SERIALIZE_POOL_MAX_SIZE=24
export SERIALIZE_STREAM_INIT_SIZE_MB=2
export SERIALIZE_STREAM_MAX_SIZE_MB=8

export JMS_POOL_MIN=16
export JMS_POOL_MAX=32
export EDEX_DEBUG_PORT=5005
export EDEX_JMX_PORT=1616
export MGMT_PORT=9601
```

Finally, `default.sh` read as follows:

```
export INIT_MEM=512 # in Meg
export MAX_MEM=4096 # in Meg
export MAX_PERM_SIZE=128m
export EDEX_JMX_PORT=1616
export EDEX_DEBUG_PORT=5005
export JMS_POOL_MIN=64
export JMS_POOL_MAX=128
export METADATA_POOL_MIN=5
export METADATA_POOL_MAX=50
export DEBUG_PARAM_1=""
export DEBUG_PARAM_2=""
export DEBUG_PARAM_3=""
export DEBUG_PARAM_4=""
export PROFILER_PARAM_1=""
export PROFILER_PARAM_2=""
export PYPIES_MAX_CONN=50

export SERIALIZE_POOL_MAX_SIZE=16
export SERIALIZE_STREAM_INIT_SIZE_MB=2
export SERIALIZE_STREAM_MAX_SIZE_MB=6

export LOG4J_CONF=log4j.xml
export MGMT_PORT=9600
```

All CAVE workstations had their `cave.ini` files modified to avoid memory issues after several out-of-memory crashes in the first day of the experiment. In that file, the following lines were changed:

```
-Dthrift.stream.maxsize=200
-XX:MaxPermSize=256m
-Xmx4096m
```

These modifications permit AWIPS II to display high-resolution (1-km grid cell) grids that extend across the entire Lower 48. Although AWIPS II may still run more slowly than desired, these modifications are believed necessary to successfully load national FLASH and MRMS-Hydro grids.

In general, FLASH data can be stored in GeoTIFF format. Experiment staff wrote a utility to automatically convert these GeoTIFF files into GRIB2 format, with headers corresponding to the properties described hereafter. The GRIB2 table properties file resides in the following location:

```
awips2/edex/data/utility/edex_static/base/grib/tables/
161/1/4.2.0.16.table
```

Inside this table should be a list of tools, line-by-line, using the following format: grib_varID:grib_varID:product_menu_name:units:parameterID, where grib_varID is an integer between 192 and 254. The product_menu_name should be identical to the string displayed for the tool's menu entry in CAVE. Units can take many formats, including "year", "%", "in", and "m^3*s^-1".

Next, /awips2/edex/data/utility/edex_static/base/grib/models/ was modified to add gribModels_FLASH.xml which contains the following XML code:

```
<gribModelSet>
    <model>
        <title>FLASH</title>
        <name>FLASH</name>
        <center>161</center>
        <subcenter>1</subcenter>
        <process>
            <id>100</id>
        </process>
    </model>
</gribModelSet>
```

Then in /awips2/edex/data/utility/edex_static/base/distribution/, grib.xml was edited to add <regex>FLASH</regex> after the pre-existing entries. Modifications to Warngen templates (see Appendix D) were desired, and the files to do so exist in this location:

```
awips2/edex/data/utility/common_static/site/[WFO_code]/
warngen/
```

Finally, inside /awips2/edex/data/utility/cave_static/user/[username]/, the styleRules directory contains a file named gridImageryStyleRules.xml, which should be modified to add the parameters from the GRIB2 table described above. This file also contains information about the units to be displayed in CAVE (which should correspond to those in the GRIB2 table), the colormap to be used in CAVE, the type of scale used in the display (e.g., linear or logarithmic), the minimum and maximum values corresponding to the beginning and end of the colormap, the values to be explicitly displayed on the colormap legend in CAVE, and whether smoothing will be turned on by default for the tool.

CAVE menus are controlled by the menus directory. For HWT-Hydro, a subdirectory FLASH was created and inside it, index.xml, which contained the following text:

```
<menuContributionFile>
    <include
    installTo="menu:org.eclipse.ui.main.menu?after=n
```

```
        cephydro” filename=”menus/FLASH/flash.xml”/>
    </menuContributionFile>
```

Then that `flash.xml` file in the `menus` directory contains the actual organization and text that will appear within CAVE. Colormaps are defined inside the `colormaps` directory. Each colormap is stored a .cmap file with the following format:

```
    <color r=”____” g=”____” b=”____” a=”____” />
```

where each line is a color in the RGB system, such that “r” is between zero and one and represents a normalized value of the red component of a color in the RGB system (and “g” and “b” correspond to green and blue, respectively). “a” stands for alpha and represents, on a zero to one scale, the degree of transparency to be used in CAVE for that color, where zero corresponds to full transparency and one corresponds to no transparency. CAVE automatically interpolates between lines in the .cmap file, so the more lines, the smoother the color transitions. CAVE limits .cmap files to 8,192 lines. Experiment staff created a Python script to automate colormap creation.

Finally, in the `bundles` directory an XML file must be created that corresponds to the bundle name used in the XML file establishing the CAVE menus. This bundle file sets the default CAVE values for density, magnification, brightness, contrast, transparency (unless otherwise specified in the colormap), and blinking.

Data enters EDEX from an LDM server. Depending on the tools desired, the LDM configuration files should be modified according to that program’s instructions, at all times remembering that filenames and other properties must remain consistent with the properties fed to CAVE and EDEX.

Appendix D: Responses from Feedback Survey

The Experiment introduction on Monday afternoon:						
Field	**Strongly Disagree**	**Disagree**	**Neither**	**Agree**	**Strongly Agree**	**Average Value**
The introduction helped me to understand experimental flash flood products sufficiently.	0	1	2	10	3	3.94
I understood the anticipated outcomes and methodology after the presentations.	0	0	2	10	4	4.13
The introduction was effective in giving me more familiarity with the AWIPS capabilities during the experiment.	0	1	0	12	3	4.06

The Experiment introduction on Monday afternoon:						
Field	**Strongly Disagree**	**Disagree**	**Neither**	**Agree**	**Strongly Agree**	**Average Value**
The introduction helped me to understand experimental flash flood products sufficiently.	0	1	2	10	3	3.94
I understood the anticipated outcomes and methodology after the presentations.	0	0	2	10	4	4.13
The introduction was effective in giving me more familiarity with the AWIPS capabilities during the experiment.	0	1	0	12	3	4.06

Please indicate your level of agreement or disagreement with the following statements:						
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 watches and warnings.	0	0	1	11	4	4.19
The evaluation and discussion sessions helped me to improve my forecasts as the week progressed.	0	0	1	11	4	4.19
The FFaIR briefings gave me sufficient situation awareness to start the day.	0	1	1	8	6	4.19
The FFaIR briefings gave me all the information I needed to identify areas at risk.	0	3	2	8	3	3.69

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 issuance of flash flood watches/warnings (forecasting sessions)	1	3	8	4	0	2.94
Tools/watch/warning evaluation and discussion sessions	0	0	14	2	0	3.13
FFaIR evaluations and FFaIR briefings	0	1	14	1	0	3.00
Webinar preparation and broadcast sessions	1	5	10	0	0	2.56

Was the material provided before the experiment helpful in understanding and preparing for the experiment?						
Field	Not at All Helpful	Somewhat Helpful	Neutral	Somewhat Helpful	Very Helpful	Average Value
Forecaster Response	0	2	4	6	4	3.75

Would you consider participating in this experiment again in the future?			
Field	Yes	No	Undecided
Forecaster Response	15	0	1

Would you recommend participating in this experiment to colleagues?			
Field	Yes	No	Undecided
Forecaster Response	15	0	1