BAMS Article

A Path Toward Short-Term Probabilistic Flash Flood Prediction

Steven M. Martinaitis, Katie A. Wilson, Nusrat Yussouf, Jonathan J. Gourley, Humberto Vergara, Tiffany C. Meyer, Pamela L. Heinselman, Alan Gerard, Kodi L. Berry, Andres Vergara, and Justin Monroe

ABSTRACT: There are ongoing efforts to move beyond the current paradigm of using deterministic products driven by observation-only data to make binary warning decisions. Recent works have focused on severe thunderstorm hazards, such as hail, lightning, and tornadoes. This study discusses one of the first steps toward having probabilistic information combined with convective-scale short-term precipitation forecasts available for the prediction and warning of flash flooding. Participants in the Hydrometeorology Testbed–MRMS Hydrology (HMT-Hydro) experiment evaluated several probabilistic-based hydrologic model output from the probabilistic Flooded Locations and Simulated Hydrographs (PRO-FLASH) system during experimental real-time warning operations. Evaluation of flash flood warning performance combined with product surveys highlighted how forecasters perceived biases within the probabilistic information and how the different probabilistic approaches influenced warnings that were verified versus those that were unverified. The incorporation of the Warn-on-Forecast System (WoFS) ensemble precipitation forecasts into the PRO-FLASH product generation provided an opportunity to evaluate the first coupling of subhourly convective-scale ensemble precipitation forecasts with probabilistic hydrologic modeling at the flash flood warning time scale through archived case simulations. The addition of WoFS precipitation forecasts resulted in an increase in warning lead time, including four events with ≥29 min of additional lead time but with increased probabilities of false alarms. Additional feedback from participants provided insights into the application of WoFS forecasts into warning decisions, including how flash flood expectations and confidence evolved for verified flash flood events and how forecast probabilistic products can positively influence the communications of the potential for flash flooding.

KEYWORDS: Hydrometeorology; Probability forecasts/models/distribution; Decision making; Flood events

https://doi.org/10.1175/BAMS-D-22-0026.1
Corresponding author: Steven M. Martinaitis, steven.martinaitis@noaa.gov
In final form 12 July 2022

AFFILIATIONS: Martinaitis, Wilson, H. Vergara, and Monroe—Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma, and NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma; **Yussouf**—Cooperative Institute for Severe and High-Impact Weather Research and Operations, and School of Meteorology, University of Oklahoma, and NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma; **Gourley and Heinselman**—NOAA/OAR/National Severe Storms Laboratory, and School of Meteorology, University of Oklahoma, Norman, Oklahoma; **Meyer**—University Corporation for Atmospheric Research, Boulder, Colorado; **Gerard and Berry**—NOAA/OAR National Severe Storms Laboratory, Norman, Oklahoma; **A. Vergara**—Gallogly College of Engineering, University of Oklahoma, Norman, Oklahoma

A flash flood is defined as a rapid rise of water into a normally dry area within minutes to hours of the causative event (NOAA 2019a). Most flash floods are a result of excessive rainfall, yet flash floods can occur from other phenomena, such as rapid snowmelt, ice jams, or a levee or dam failure. The primary means of alerting the public to a flash flood is through the dissemination of a Flash Flood Warning (FFW) generated by forecasters from local National Weather Service (NWS) weather forecast offices. A FFW is issued "when flooding is imminent or likely" for "short-term events which require immediate action to protect life and property" (NOAA 2019b). The modern advent of FFWs can be traced to a pilot project started at the Oklahoma City and Des Moines United States Weather Bureau offices in 1956, which coincided with research using WSR-3 radars for the detection of excessive precipitation that could induce flash flooding (Kutschenreuter 1958). Multiple flash flood events during April–June 1957 validated the utilization of radars to supplement rain gauge observations for detecting precipitation and the ability to issue timely warnings for a flash flood threat.

Reliance upon gauge observations and radar-derived quantitative precipitation estimates (QPEs) along with radar trends became the primary means for predicting flash floods; however, the 4–5 July 1969 derecho and flash flood event in northeast Ohio, which claimed 25–30 lives from flooding impacts, resulted in the emergence of the flash flood guidance (FFG) product that was implemented by August 1970 (Zevin 1994). Initial FFG criteria were based upon rainfall intensities and corresponding runoff (Zevin 1994). FFG is currently defined as the average rainfall needed to initiate flooding on small waterways over a specified area or predefined grid over a given time (NOAA 2017). NWS river forecast centers over the decades employed various methodologies driven by the current FFG principle to calculate FFG values (Sweeney 1992; Clark et al. 2014). The corresponding creation of the NWS Flash Flood Program in 1970 served to address the watch and warning needs for flash flooding, the implementation of the FFG technique, and the nationwide Excessive Rainfall Outlook based on the prediction of rainfall that could exceed FFG (Sweeney 1992; NOAA 2020).

More recent advancements introduced real-time hydrologic modeling to NWS warning operations, notably the implementation of the Flooded Locations and Simulated Hydrographs (FLASH) system in 2018 (Gourley et al. 2017). The FLASH system is the first to generate hydrologic modeling products at the flash flood space and time scales in real time for the entire CONUS and outer territories. The FLASH product suite contains forecast discharge outputs and unit streamflow values, which are defined as discharge normalized by the upstream basin area, from different distributed hydrologic models. The FLASH system also includes comparisons of QPEs to FFG and average recurrence intervals for rainfall. Driving the hydrologic models and QPE comparison products in the FLASH system are high spatiotemporal resolution radar-derived QPEs from the Multi-Radar Multi-Sensor (MRMS) system (Zhang et al. 2016). Testbed evaluations demonstrated how the FLASH system can help identify areas of increased flash flood potential (Martinaitis et al. 2017, 2020).

Table 1. A 2 × 2 contingency table as defined by Barnes et al. (2009), where *X* is the number of events with correct forecasts, *Y* is the number of events that occurred but were not forecast, *Z* is the number of forecasts that did not have a corresponding verified event, and *W* is the number of correct null forecasts.

		Event Observed	
		Yes	No
Event Forecast	Yes	*X*	*Z*
	No	*Y*	*W*

Real-time precipitation estimates, derived guidance products, hydrologic model output, and software programs (e.g., Flash Flood Monitoring and Prediction program; Smith et al. 2000; Arthur et al. 2005) are all available for forecasters to use to determine if a FFW requires issuance. Advancements designed for flash flood prediction (e.g., FFG and FLASH) along with other significant technological and scientific upgrades (e.g., the implementation of dual-polarization technology to the WSR-88D network; Kumjian 2013a,b,c) have enhanced the observationally based tools available for the warning decision process, yet FFW performance improvements have slowed. NWS forecasters issued 4,287 FFWs per year from 1986 to 2021 according to the NWS Performance Management System (https://verification.nws.noaa.gov/). Annual FFW probability of detection (POD; Schaefer 1990), probability of false alarm (POFA; Barnes et al. 2009), and critical success index (CSI; Schaefer 1990) values were defined by the following equations using the 2 × 2 contingency table provided in Table 1: POD = $X/(X + Y)$, POFA = $Z/(X + Z)$, and CSI = $X/(X + Y + Z)$. Annual POD values improved through 1997 but remained steady afterward (Fig. 1a). The 5-yr POD average has remained between 0.85 and 0.89 since 2001. Annual POFA values have continuously decreased across the evaluated period (Fig. 1a), yet the slowed improvement in POFA values resulted in the 5-yr POFA average remaining between 0.33 and 0.39 since 2011. A corresponding increase in CSI values was shown across the 1986–2021 period,

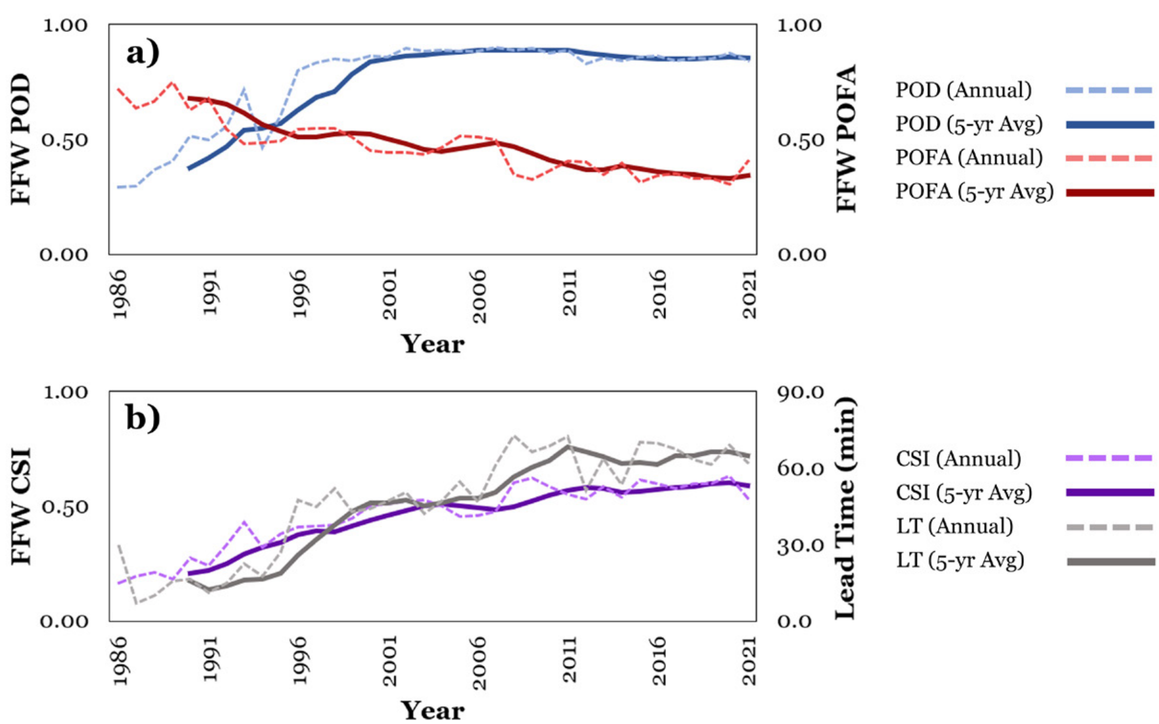

Fig. 1. Statistical analysis of the annual (dashed line) and 5-yr average (solid line) of the (a) probability of detection (POD; blue) and probability of false alarm (POFA; red) as well as the (b) critical success index (CSI; purple) and lead time (LT; gray) for FFWs from 1986 to 2021.

yet it has also remained stagnant with a 5-yr CSI average between 0.56 and 0.60 since 2011 (Fig. 1b). FFW lead times (LTs) increased from a 5-yr average of 15.7 min in 1990 to a 5-yr average of 64.7 min by 2021 (Fig. 1b); however, the 5-yr average LT has remained between 61 and 68 min since 2010. These statistical trends are similar to results found by Karstens et al. (2015) for NWS tornado and severe thunderstorm warnings.

The warning decision and application of an FFW is also binary. Persons within a warning polygon are notified of an imminent or ongoing flash flood hazard. Persons outside a warning polygon receive no notification (though signal bleedover from weather radio and wireless emergency alerts can occur). Warnings are currently issued using a storm-based warning polygon methodology, where the warned area is drawn to the geographically specific region of the hazard (Sutter and Erickson 2010; Harrison and Karstens 2017); however, substantial false alarm areas exist with FFW polygons. The average storm-based FFW polygon from 2008 to 2021 covered 2,068 km^2, while the average flash flood event area was only 107 km^2 (https://verification.nws.noaa.gov/). Karstens et al. (2015) detailed various limitations of the storm-based polygon approach, including the use of static warning areas for dynamic hazards and the generation of large warning areas based upon uncertainty of receiving hazard verification.

The authors concur with Rothfusz et al. (2018) that the current observation-oriented approach resulting in deterministic hazard products has limitations and can benefit from the exploration of incorporating and disseminating probabilistic information with storm-scale forecast components. An NWS-commissioned National Research Council report recommended ways to improve the estimation and communication of risk and uncertainty through probabilistic products and services (NRC 2006). A future direction of the NWS warning paradigm was outlined by Rothfusz et al. (2018) to meet the report recommendations with a concept called Forecasting a Continuum of Environmental Threats (FACETs). The FACETs paradigm proposes the change from the deterministic watch and warning construct to a high-resolution gridded set of probabilistic hazard information (PHI) that encompass the period from days to within minutes of the event (Karstens et al. 2015, 2018; James et al. 2020). PHI relating to the warning time scales are achieved through the fusion of current observations and short-term model forecasts. Probabilistic guidance for short-fused convective severe weather hazards have demonstrated the potential of using PHI grids in the warning time frame (e.g., Karstens et al. 2015). The activities and results presented in this study focus on flash flood PHI development in accordance with the first three components of the FACETs framework: 1) method and manner, 2) observations and guidance, and 3) the forecaster.

Probabilistic FLASH system

A probabilistic FLASH (hereinafter denoted as PRO-FLASH) system was developed at the National Severe Storms Laboratory to generate PHI grids based on hydrologic model output. PRO-FLASH products were derived from statistical analyses of unit streamflow values simulated with the Coupled Routing and Excess Storage Model (CREST; Wang et al. 2011) and the kinematic wave approximation to the Saint-Venant equations of open channel flow. The application of the CREST model and associated physically based parameters were conducted within the Ensemble Framework For Flash Flood Forecasting (EF5; Flamig et al. 2020), the hydrologic modeling core in FLASH. An MRMS reanalysis dataset with rainfall rates fields every 5 min during the 2002–11 period was used as forcing inputs for the simulation, which produced CONUS-wide fields of daily maximum unit streamflow and 5-min unit streamflow values at the 1,643 locations with stream gauges from the United States Geological Survey (USGS) used in the Gourley et al. (2017) study. The probabilistic products generated within the PRO-FLASH system were derived with two postprocessing algorithms developed from two supervised machine learning techniques trained on these deterministic outputs (Table 2).

Table 2. List of PRO-FLASH products delineated by the two different probabilistic approaches. Included are the reference levels for the probabilistic values of each product and the spatiotemporal resolution.

Probabilistic approach	PRO-FLASH product	Reference level	Spatiotemporal resolution
Flash flood storm reports	Prob-LSR	NWS local storm reports	0.01° × 0.01°; 10 min
Unit streamflow values	Prob-Minor	CREST maximum unit streamflow: 2.0 $m^3 s^{-1} km^{-2}$ (2018) 1.0 $m^3 s^{-1} km^{-2}$ (2019)	0.01° × 0.01°; 10 min
	Prob-Moderate	CREST maximum unit streamflow: 5.0 $m^3 s^{-1} km^{-2}$	0.01° × 0.01°; 10 min
	Prob-Major	CREST maximum unit streamflow: 10.0 $m^3 s^{-1} km^{-2}$	0.01° × 0.01°; 10 min

Two different PHI-type model output approaches were developed within the PRO-FLASH system. The first approach focused on the probability of receiving a flash flood local storm report (LSR). The Probability of Receiving a Flash Flood LSR (Prob-LSR) product was derived from training a binary classifier to a range of simulated unit streamflow values related to flash flood LSRs regardless of the flash flood severity. Flash flood LSRs recorded in NWS *Storm Data* from 2005 to 2011 were matched in space and time with the daily maximum unit streamflow values from the deterministic FLASH model to establish the LSR-positive class (labeled as "1"). Nonzero values of the same daily maximum unit streamflow were randomly sampled across the CONUS from locations without LSRs but with coincident dates to establish the LSR-negative class (labeled as "0"). A logistic function was then trained with the labeled maximum unit streamflow dataset, which was then used to produce probability estimates of receiving a flash flood LSR.

The second approach focused on the potential flash flood severity using unit streamflow thresholds. The Probability of Minor, Moderate, and Major Flash Flooding (denoted as Prob-Minor, Prob-Moderate, and Prob-Major) products were the first known effort to develop PHI-type grids with respect to hazard severity levels at the warning time scale. These products were derived from the application of a nonlinear regression model of conditional distributions of measured unit streamflow values from USGS stream gauges, where the conditioning or explanatory variable was simulated unit streamflow values from FLASH. The Generalized Additive Models for Location, Scale, and Shape (GAMLSS; Rigby and Stasinopoulos 2005) framework was utilized for the regression exercise. The probability of a particular severity level was calculated by applying this model of conditional distributions as a postprocessor of deterministic FLASH unit streamflow values and computing the cumulative probabilities at each severity level. The reference unit streamflow levels for the severity-based PRO-FLASH products were determined from past research (e.g., Martinaitis et al. 2017) and feedback from NWS forecasters. The Prob-Minor product was initially developed around the 2.0 $m^3 s^{-1} km^{-2}$ value in 2018 and was later adjusted to 1.0 $m^3 s^{-1} km^{-2}$ in 2019. The Prob-Moderate and Prob-Major products were generated using a value of 5.0 and 10.0 $m^3 s^{-1} km^{-2}$, respectively. The nomenclature of the products follows that of NWS river forecasts and observations on USGS-gauged streams.

Real-time probabilistic product perceptions

The Hydrometeorology Testbed (HMT) MRMS Hydrology (hereinafter denoted as HMT-Hydro) experiment hosted at the National Weather Center in Norman, Oklahoma, facilitated the evaluation of emerging technologies and applications that support flash flood prediction and warning decision-making (Martinaitis et al. 2017). The studies conducted during the 2018–19 HMT-Hydro experiment summers were separated into two primary components: evaluating probabilistic hydrologic model output and incorporating short-term model precipitation forecasts into the warning process. The first phase examined the different approaches for the PHI-based flash flood gridded datasets from the PRO-FLASH system. Real-time warning operations focused on PRO-FLASH products forced by MRMS radar-derived QPEs using an experimental dual-polarization synthetic algorithm (Wang et al. 2019; Cocks et al. 2019; Zhang et al. 2020) with an evaporation correction component (Martinaitis et al. 2018). Various flash flood events throughout the 2018–19 HMT-Hydro experiments provided ample opportunities to utilize PRO-FLASH probabilistic information in a warning environment.

Retrospective subjective evaluation surveys were conducted on 16 select flash flood events that were verified by LSRs with 62 total participant responses per evaluation statement. Each selected event was chosen based on its significance along with the quantity and coverage of flash flood LSRs. Flash flood events that occurred during real-time experimental warning operations were favored. A mix of flash flood impact severity was chosen and delineated into minor and major events. The definitions of minor and major flash flooding corresponded to Meteorological Phenomena Identification Near the Ground (mPING) impact classifications (Elmore et al. 2014). Flash flood reports related to classes 1 (overflowing creek; cropland/yard flooding) and 2 (road flooding or closure; stranded vehicles) were defined as minor, and classes 3 (water in structures) and 4 (structures or vehicles swept away) were defined as major. Water rescues were also classified as a major impact. The Prob-Minor and Prob-Major products were designed to align with the major and minor classification assignments of LSRs that continue from past HMT-Hydro experiments (Martinaitis et al. 2017). The Prob-Moderate product was designed to see if any potential application can be gleaned from having an intermediate severity-level gridded information.

Minor flash flood events were commonplace during the experiment and were characterized by Prob-LSR values > 90% and Prob-Minor values ranging from 40% to 60% (e.g., Fig. 2). The Prob-Moderate and Prob-Major products were typically characterized by probability values < 20%. Major flash floods contributed to greater probability values across the product suite (e.g., Fig. 3). The Prob-LSR product would tend to reach its maximum value, while the Prob-Minor product would generally achieve values of 70%–80%. The Prob-Moderate and Prob-Major values would typically reach probability values of 50%–70% and 30%–40%, respectively.

Subjective product evaluations highlighted how forecasters perceived any biases with the PRO-FLASH product magnitudes (Fig. 4). The Prob-LSR product had 58.1% of responses stating that its probability values were at least slightly too high. The Prob-Minor product had contrasting results compared to the Prob-LSR product with 77.8% of responses stating

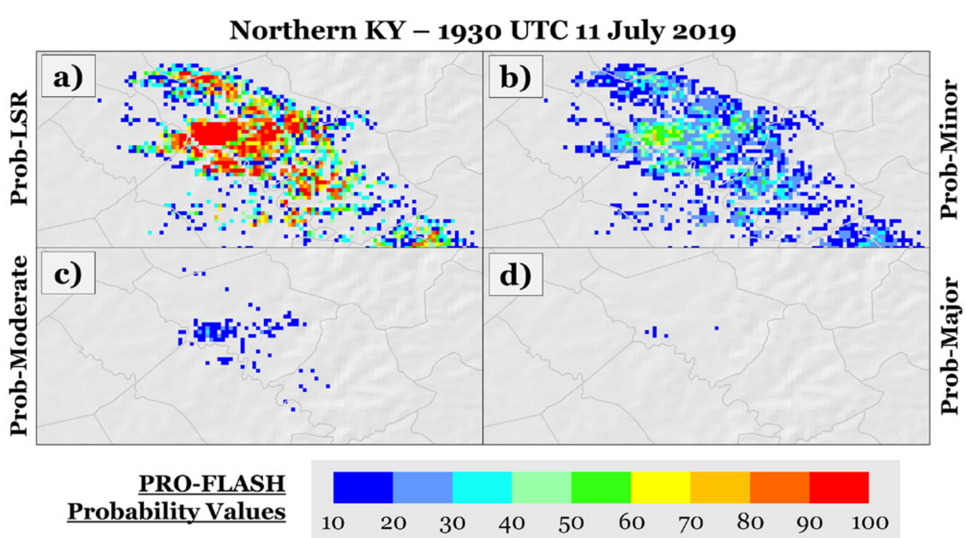

Fig. 2. Representation of a minor flash flood event using the PRO-FLASH system. The following PRO-FLASH products were depicted at 1930 UTC 11 Jul 2019 over northern Kentucky: (a) Prob-LSR, (b) Prob-Minor, (c) Prob-Moderate, and (d) Prob-Major.

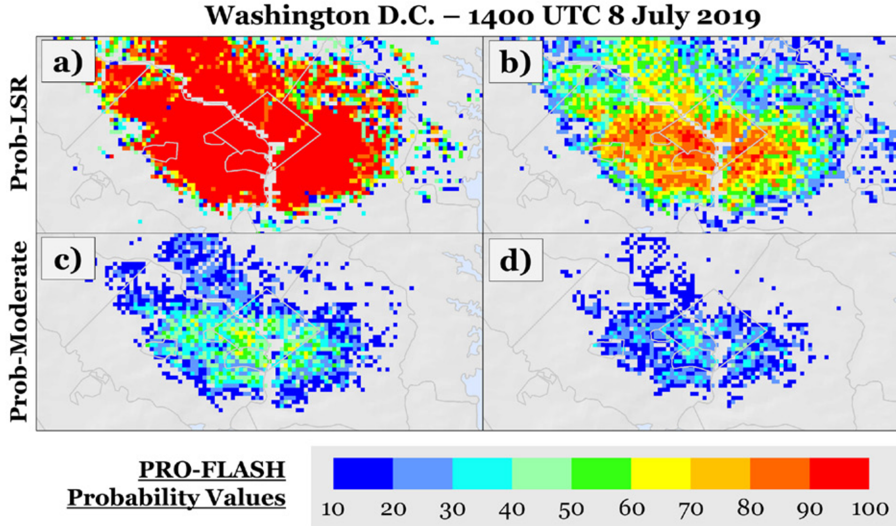

Fig. 3. As in Fig. 2, but for a major flash flood event as depicted at 1400 UTC 8 Jul 2019 over the Washington, D.C., area.

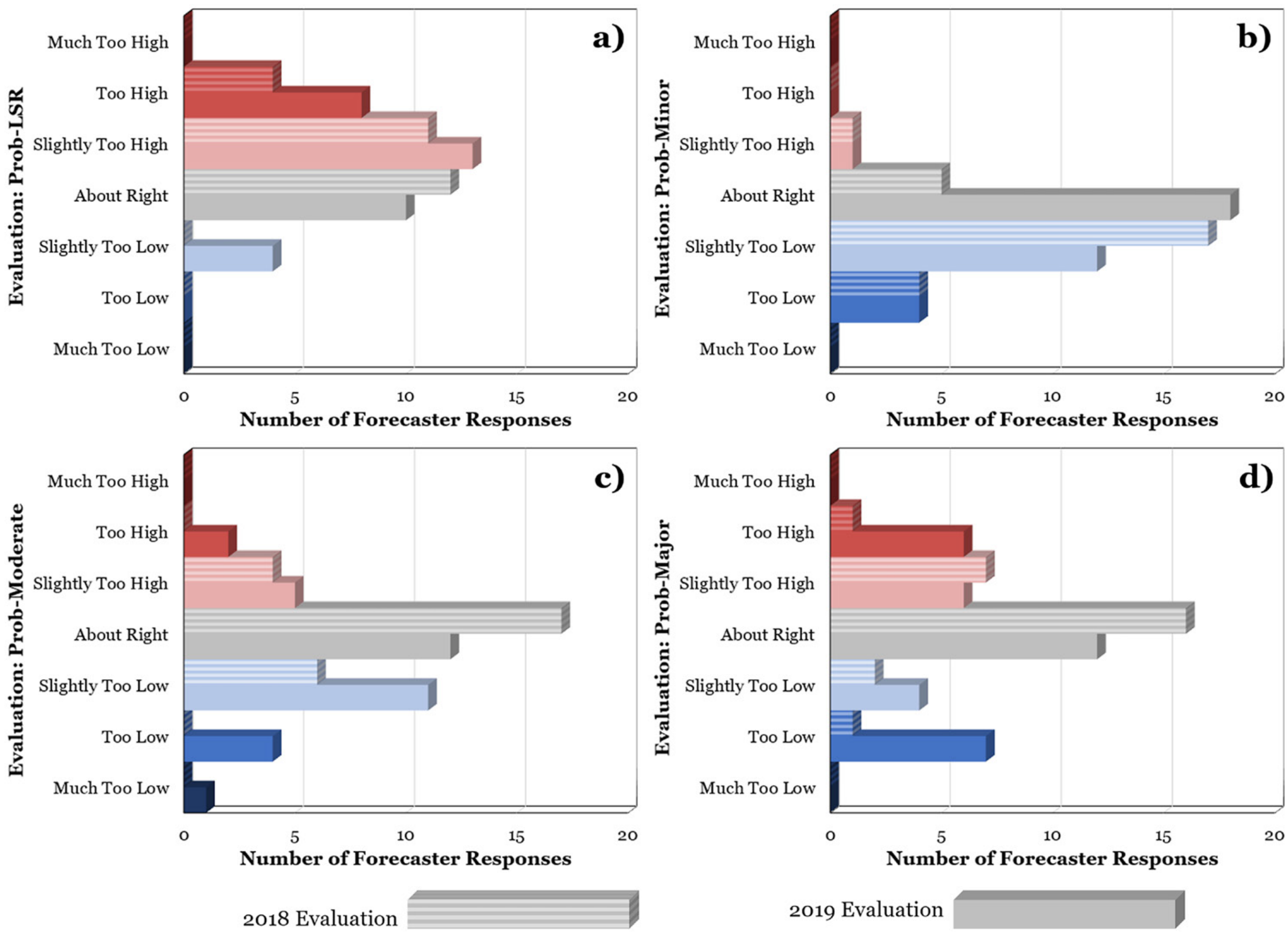

Fig. 4. Perceived bias evaluation of the PRO-FLASH (a) Prob-LSR, (b) Prob-Minor, (c) Prob-Moderate, and (d) Prob-Major products during the 2018 (striped bars) and 2019 (solid bars) HMT-Hydro experiment based on the following statement: "Using all available flash flood observations, rate the gridded probability values from the [PRO-FLASH product]."

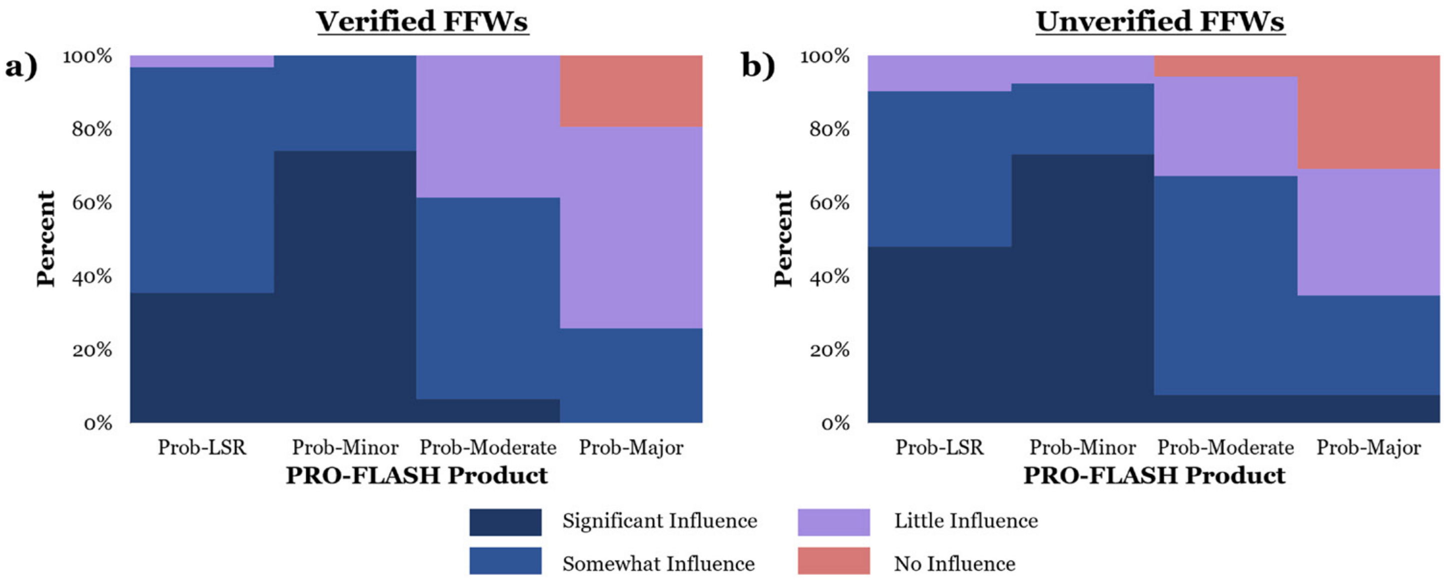

Fig. 5. Percent of participant responses denoting the influence of each PRO-FLASH product in the warning decision for (a) verified and (b) unverified experimental FFWs.

that the Prob-Minor values were at least slightly too low. The perception of the Prob-Minor product changed in 2019 when the product guidance value was adjusted to 1.0 $m^3 s^{-1} km^{-2}$. The percentage of surveyed responses that stated that the Prob-Minor values were at least slightly too low dropped to 45.7% (16 of 35 responses). Most responses for the Prob-Moderate and Prob-Major products were surveyed as "about right" in their product values; however, participants during the 2019 HMT-Hydro experiment perceived some tendencies for both the Prob-Moderate and Prob-Major as biasing low during major flash flood events and biasing high during minor events.

Application with experimental probabilistic warnings

Participants utilized the PRO-FLASH products for the issuance of experimental FFWs during real-time HMT-Hydro experiment operations. Storm-based warning polygons generated by participants were similar to those created during NWS operations; however, the experimental FFWs included participant-defined probability values for minor and major flooding based on their use of the PRO-FLASH products for flash flood prediction. Participants were also tasked with rating how each PRO-FLASH product influenced each warning decision. A total of 140 experimental FFWs were issued during real-time operations over the two test bed summers. Flash flood LSRs collected by local NWS offices were used for verification; thus, it is likely that attempts to obtain verification were not conducted in areas where experimental FFWs were not collocated with operational FFWs. This would impact any warning performance statistics, yet the authors can gain some insight into how the PRO-FLASH products can influence the decision-making process and warning polygons.

Experimental FFWs issued during the 2019 experiment allowed participants to assign a level of influence to every PRO-FLASH product during the FFW issuance using a four-point scale ranging from the product having no influence on the warning decision to having a significant influence on the warning decision. The two products that can be used as a baseline to help predict the potential for any flash flooding (Prob-LSR and Prob-Minor) had the greatest influence on the warning decision, yet the Prob-Minor product was shown to be more effective with verified FFWs. The Prob-Minor product was listed most often as having a significant influence on the warning decision (74.2%) followed by Prob-LSR (35.5%) for verified FFWs (Fig. 5a). The significant influence of the Prob-LSR product increased to 48.1% for unverified FFWs, which

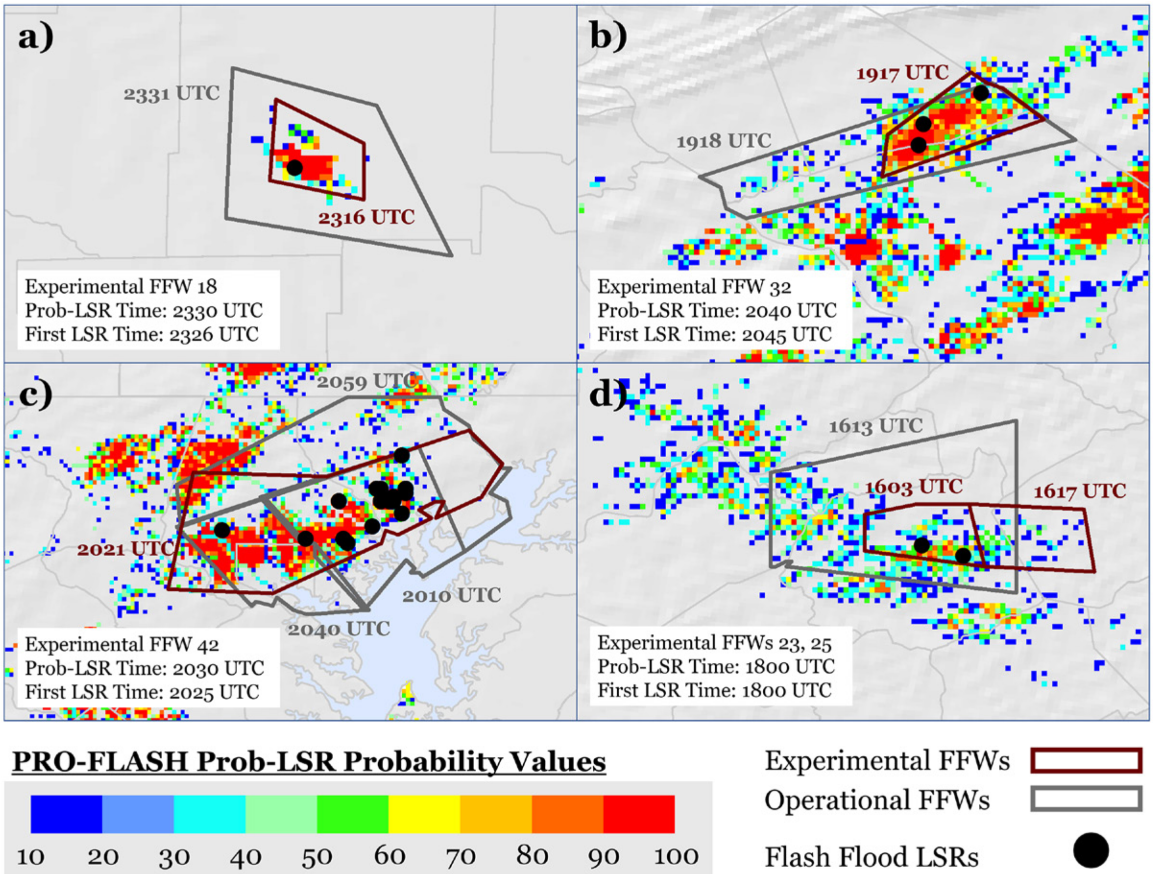

Fig. 6. Example flash flood events that occurred during the real-time experimental warning operations in 2019 that were characterized by reduced warning areas with the experimental FFWs (dark red polygons) compared to the operational FFWs (gray polygons). The issuance time of each operational and experimental FFW is labeled next to each polygon. Plotted in the background are the flash flood LSRs (black dots) and the Prob-LSR product around the time of the first flash flood LSR.

can be attributed to the higher probability values generated by the product. Both the Prob-Minor and Prob-LSR products had at least somewhat influenced 96.8% of verified warnings issued, yet they also at least somewhat influenced 90.4% of unverified FFWs (Fig. 5b). The Prob-Moderate and Prob-Major products were less influential in the warning decision-making process, yet multiple participants stated how the Prob-Moderate and Prob-Major product values provided some increased confidence in determining whether to issue a warning.

The experimental FFWs based on PRO-FLASH showed instances of highlighting the localized flash flood threat and reducing false alarm area. When considering only localized flash floods (i.e., where FFWs were isolated to a singular LSR or cluster of LSRs that were captured within a single area of increased probability values), there were eight events where the warned polygonal area was reduced (Fig. 6). Five events had the warned area reduced by over 50%. Instances of reduced warned areas included events that had multiple operational FFWs (Fig. 6c) or multiple experimental FFWs (Fig. 6d) valid for the same causative event. Participants commented on the concise regions of PRO-FLASH product signals and how the spatial coverage of product values encompassed the flash flood LSRs.

The overall results presented how the products highlighted localized flash flood threat areas along with documented the understanding gained on how the products influenced warning decisions and the associated participant perceptions and biases of the probabilistic products. This allows for research initiatives to modify and align PRO-FLASH products to current NWS operations and warning applications (see the "Advancing current operational impact-based warnings" sidebar).

Advancing Current Operational Impact-Based Warnings

The NWS Weather-Ready Nation initiative (Uccellini and Ten Hoeve 2019) evolved how forecast products and communications of hazardous weather are disseminated to the public and end-users. NWS impact-based warnings (IBWs) communicate the potential effects of a hazard through predefined tags for various hazard severity (e.g., Hudson et al. 2013; Ripberger et al. 2015). Flash flood IBWs are formatted to include a hazard statement, defined source of information, an impact statement within the warning text, and IBW tags at the bottom of the warning (Fig. SB1). The impact information provided is separated into three damage threat levels: Base, Considerable, and Catastrophic (NOAA 2019b). A Considerable tag is reserved for events "of unusual severity of impact where urgent action is needed to protect lives and property," while a Catastrophic tag is applied for "exceedingly rare, violent flash floods which threaten lives and cause disastrous damage."

Contextualizing FFWs to the potential or ongoing severity of the event is critical for impact-based decision support services. There are various local qualitative assessments for categorizing flash flood reports to IBW damage threat tags. One example assessment is the work by Gaviria Pabon et al. (2021) that categorized flash flood events within the NWS Norman, Oklahoma, county warning area by focusing on keywords and phrases within flash flood LSRs. While reports were segregated into the three damage threat categories, Gaviria Pabon et al. (2021) observed that this work contained a substantial level of subjectivity and that some keywords "were ambiguous and vague, which created a level of uncertainty at the time of the classification"; moreover, the application of an IBW tag could be reactive to received reports (i.e., employing the Catastrophic tag after receiving reports of multiple swift water rescues).

The IBW tag application lends itself to two intertwined efforts to improve impact-based decision support services for flash floods: 1) a more objective classification of flash flood reports that can be aligned with IBW tag levels and 2) a predictive component to assess the potential flash flood severity for damage threat level tag application. The Flash Flood Severity Index (FFSI) is an emerging effort to objectively categorize the magnitude of a flash flood using a five-tiered classification index (Schroeder et al. 2016). Initial evaluations of operationally based applications of FFSI demonstrated how postevent evaluations allow for the assignment of an FFSI category (Schroeder et al. 2020). Objective flash flood categorizations via FFSI can result in direct relationships to IBW damage threat tags and hydrologic model outputs from the PRO-FLASH system.

There are two ongoing efforts investigating the coupling of FFSI and PRO-FLASH with IBW damage threat tags. The first is defining direct relationships between the five-tier FFSI categorization and the three IBW damage threat tag levels. The second is refining the three severity-based PRO-FLASH probabilistic outputs to work with the three defined IBW damage threat tags using flash flood LSRs binned into IBW categories, unit streamflow values, and other inputs in a machine learning algorithm. Ultimately, aligning all three components (PRO-FLASH, FFSI, and IBW damage threat tags) can improve the characterization and referencing of flash floods for future risk communications. These efforts plus the lessons learned from the HMT-Hydro experiment can also advance FACETs-based development with flash flooding for current NWS needs, while also bridging the gap between the integration of PHI-based fields and the final FACETs design (Rothfusz et al. 2018) for hydrometeorological phenomena.

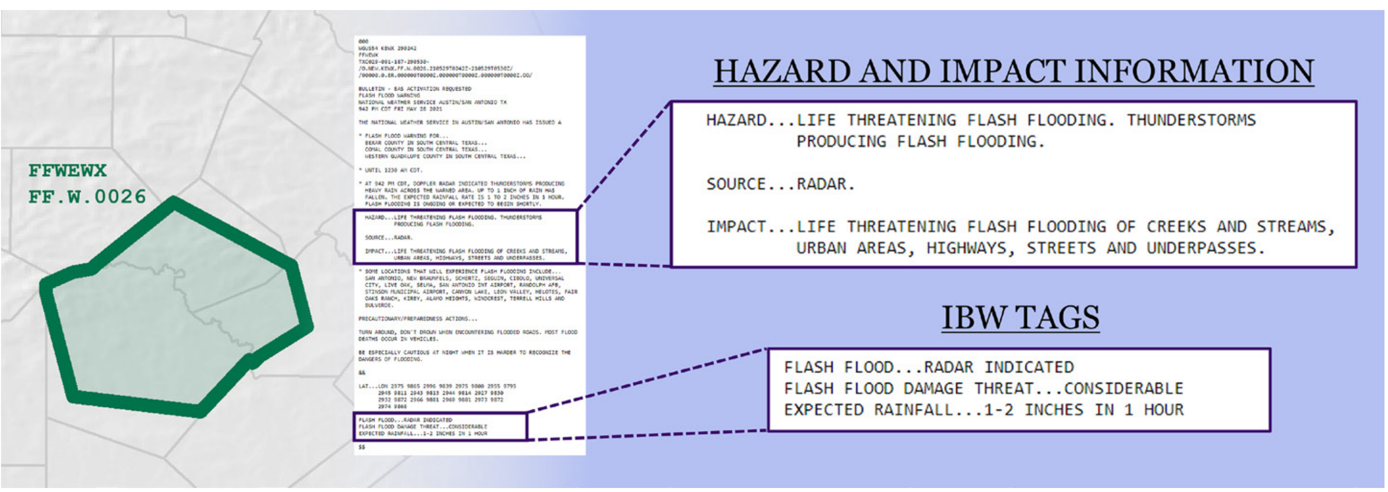

Fig. SB1. Example FFW that highlights the hazard and impact information statements along with the IBW tags for an impact-based FFW.

Incorporating short-term forecast modeling

Previous research demonstrated the value of utilizing high-resolution short-term numerical weather prediction model forecasts to improve the forecast and warning of severe weather hazards (e.g., Yussouf et al. 2015; Wilson et al. 2021). The 2014–2016 HMT-Hydro experiment provided the first opportunity to integrate hourly convective-scale quantitative precipitation forecasts (QPFs) into a real-time storm-scale flash flood prediction system (Martinaitis et al. 2017). Challenges related to the placement and coverage of QPFs along with run-to-run forecast inconsistencies limited the use of QPFs for warning decisions. The need for more accurate numerical weather prediction with higher forecast product cadence on the warning time scale was fulfilled with the NOAA/NSSL Warn-on-Forecast System (WoFS; e.g., Stensrud et al. 2009, 2013; Wilson et al. 2021). The current WoFS configuration is a convective-scale, frequently cycled ensemble that produces 36-member analyses and 18-member probabilistic forecasts of individual thunderstorm hazards for the next 0–6 h. The WoFS guidance is produced over a 900 km × 900 km domain with a 3-km horizontal grid spacing. WoFS provides forecasts on time and space scales that other convection-allowing models are not designed for, thus enabling a more continuous flow of probabilistic weather information between the typical watch and warning spatiotemporal scales. Combining the PRO-FLASH system with WoFS represented the first coupling of subhourly convective-scale ensemble QPFs with probabilistic hydrologic modeling at the FFW time scale.

The first integration of the WoFS QPFs into the PRO-FLASH system (hereinafter denoted as WoFS-FLASH) during the 2018 HMT-Hydro experiment revealed earlier threat assessment and communication opportunities to core partners when WoFS QPFs were introduced, along with numerous indications for potential increased FFW LTs. Please refer to Yussouf et al. (2020) for full details and results of the 2018 case evaluations. The anticipated benefits of the coupled WoFS-FLASH probabilistic products motivated the development of archived case simulations using real-time data playback for the 2019 HMT-Hydro experiment. The simulations utilized WoFS ensemble 90th-percentile QPFs with new model runs initialized every 30 min with forecasts at 10-min intervals up to a 180-min forecast. The PRO-FLASH system ran for each forecast interval to create the WoFS-FLASH forecast output. Participants independently evaluated three cases (Table 3) that included typical operational data, the WoFS-FLASH products, and WoFS QPF graphical information that included the ensemble 90th percentile of forecast precipitation accumulation along with probabilities of exceeding 25.4 mm (1.00 in.), 50.8 mm (2.00 in.), and 76.2 mm (3.00 in.) of precipitation. Each participant independently examined the evolution of WoFS-FLASH product outputs during the playback of the three simulations (e.g., Fig. 7) and were tasked with warning issuance responsibilities.

The benchmark operational FFWs across the three simulated cases had eight verified warnings to one unverified warning without a missed event (Table 4). This resulted in a POD of 1.00 and a CSI of 0.91. The 11 participants during the 2019 HMT-Hydro experiment issued a total of 96 experimental FFWs across the three cases. A high POD was achieved (0.94), while the CSI was reduced to 0.72. The lowered CSI resulted from an increased POFA from 0.10 to 0.25 combined with seven occurrences of missed flash flood events. The increase in unverified FFWs and missed flash flood events can be attributed to uncertainty in QPF coverage and magnitude, which was then reflected in the probabilistic hydrologic output. Five of the seven missed events were associated with the Blackhawk County flash flood LSR in the Des Moines, Iowa, simulation where the WoFS-FLASH products underrepresented the potential threat.

Table 3. The archived case events that were utilized during the 2019 HMT-Hydro experiment for the evaluation of the WoFS-FLASH system. The start and end times, the location of the event based on the NWS county warning area (CWA), and the description of each event is provided. An asterisk for the LWX and DMX cases denotes that a restricted region was defined for the participants to focus experimental warning efforts.

Case event characteristics			
Start date and time	End date and time	NWS CWA	Case event description
1900 UTC 27 May 2018	2110 UTC 27 May 2018	LWX (Sterling, VA)*	Ellicott City and Baltimore, MD, flash flood
2200 UTC 30 Jun 2018	0140 UTC 1 Jul 2018	DMX (Des Moines, IA)*	Multi-area event in central IA evolving into significant flash floods
2130 UTC 12 Jul 2018	0140 UTC 13 Jul 2018	FSD (Sioux Falls, SD)	Minor flash floods in SD, null event in MN

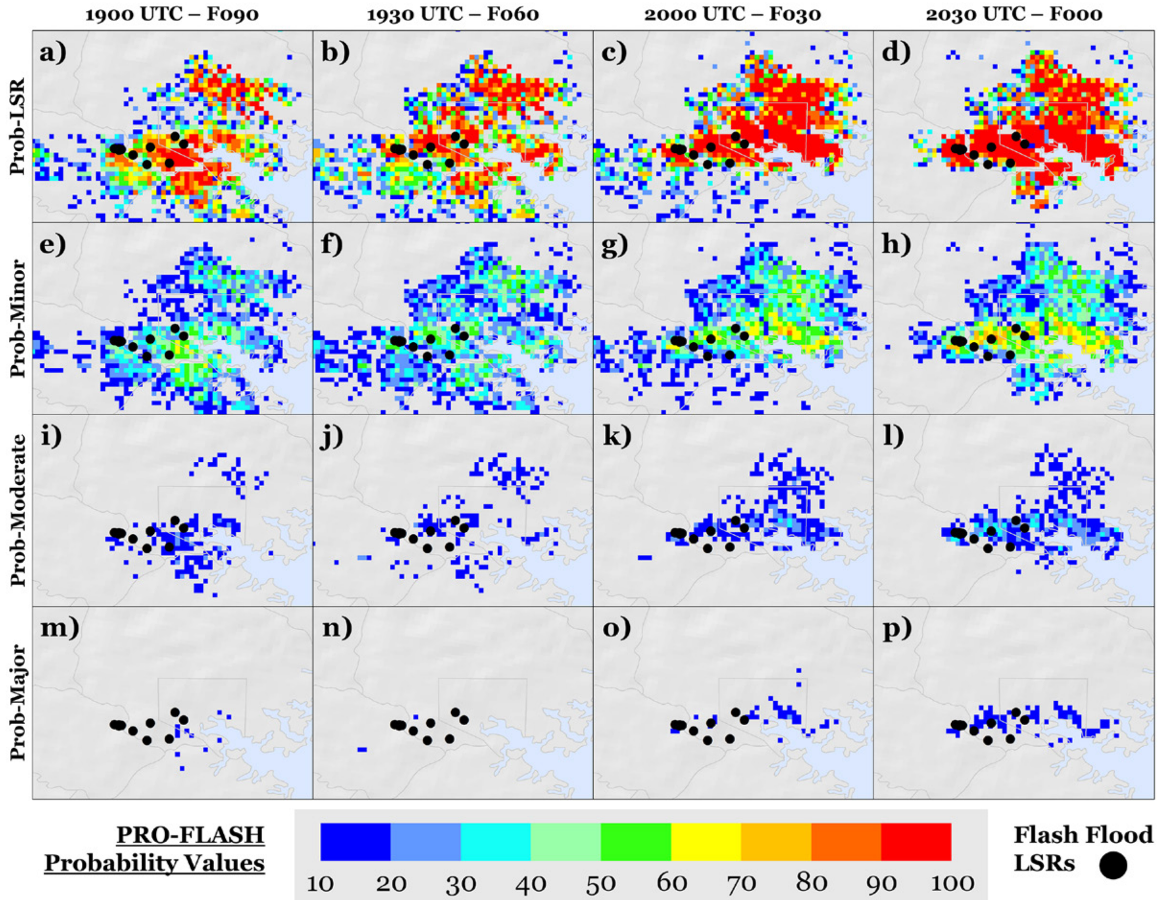

Fig. 7. Evolution of the PRO-FLASH products coupled with the WoFS QPFs during the Sterling, Virginia, displaced real-time case that encompassed the Ellicott City and Baltimore, Maryland, flash floods of 27 May 2018. The PRO-FLASH probabilistic products shown are the (a)–(d) Prob-LSR, (e)–(h) Prob-Minor, (i)–(l) Prob-Moderate, and (m)–(p) Prob-Major products. These products are depicted (from left to right) at 1900 UTC using a 90-min forecast from the WoFS, 1930 UTC using a 60-min forecast from the WoFS, 2000 UTC using a 30-min forecast from the WoFS, and 2030 UTC using QPE-only forcing, which corresponds to the time of the first flash flood LSR received at 2026 UTC. Each plot shows all the flash flood LSRs (black dots) from the event as a reference for where the flash floods occurred with respect to the probabilistic data using the WoFS precipitation forecasts.

Table 4. Comparison of the operational FFW statistics vs the experimental FFW statistics from the three archived case events from 2019. Included in the table are the number of verified and unverified FFWs, the number of missed flash flood events, the probability of detection (POD), the probability of false alarm (POFA), the critical success index (CSI), and the sample size. Operational FFW statistics were obtained from the NWS Performance Management System (https://verification.nws.noaa.gov/).

Case	Verified FFWs	Missed events	Unverified FFWs	POD	POFA	CSI	Sample size
			Operational FFW Statistics				
LWX	2	0	0	1.00	0.00	1.00	1
DMX	4	0	0	1.00	0.00	1.00	1
FSD	2	0	1	1.00	0.33	0.67	1
Overall	8	0	1	1.00	0.10	0.91	3
			Experimental FFW Statistics				
LWX	17	0	6	1.00	0.23	0.77	10
DMX	25	6	7	0.91	0.12	0.82	11
FSD	21	1	20	0.95	0.49	0.50	11
Overall	63	7	33	0.94	0.25	0.72	32

Table 5. Analysis of FFW LTs between the operational FFWs and the experimental FFWs issued for each specific flash flood event and the combined overall average from the three archived case simulations from 2019. Included in this table are the medial experimental FFW LT, the initial operational FFW LT, the difference between the two LT values, and the number of instances of experimental FFWs having positive and zero initial LT.

Case	Flash flood location	Median initial experimental FFW LT (min)	Initial operational FFW LT (min)	Median initial LT difference (min)	Experimental FFWs with positive/zero initial LT
LWX	Ellicott City, MD	29.0	0.0	29.0	9/1
	Baltimore, MD	57.5	0.0	57.5	10/0
DMX	Story County, IA	17.0	12.0	5.0	9/2
	Dallas County, IA	29.0	0.0	29.0	10/1
	Polk County, IA	39.0	9.0	30.0	9/2
	Blackhawk County, IA	22.0	38.0	–16.0	6/5
FSD	Gregory County, SD	0.0	29.0	–29.0	5/6
	Douglas County, SD	65.0	61.0	4.0	11/0
Combined overall average		32.3	18.6	13.7	69/17

The overall average median LT for the experimental FFWs using WoFS-FLASH outputs increased by 13.7 min over the LTs for the operational FFWs (Table 5); moreover, the percent of warnings with positive LTs increased from 62.5% for operational FFWs to 80.2% with experimental FFWs. Four of the flash flood events had a median LT increase ≥29 min, yet two specific flash flood events did not see overall improvements in LTs. The Blackhawk County, Iowa, and Gregory County, South Dakota, events had five and six instances, respectively, where no experimental warnings were issued prior to the flash flood report (i.e., initial LT was declared as 0 min), thus reducing the median initial LT for each event. The 11 instances of zero initial LT between the Blackhawk County and Gregory County events accounted for 64.7% of all zero initial LT occurrences.

The simulation focused on the high-impact Ellicott City and Baltimore, Maryland, flash floods of 27 May 2018 encapsulated the various benefits and trade-offs with incorporating short-term storm-scale ensemble precipitation forecasting into warning operations: increased warning LTs and increased potential to forecast the flash flood severity but increased POFA. The median initial LTs for the first flash flood report for the Ellicott City and Baltimore areas were increased by 29.0 and 57.5 min, respectively (Table 5). And while participant-issued experimental FFWs captured all flash flood LSRs, there were also six unverified warnings across all simulation playbacks. The chronology of all participant FFWs and follow-up statements (i.e., updated warnings) issued during the 27 May 2018 simulation highlighted both the issuance time variability and the evolution of participant-assigned major and minor flash flood probabilities (Fig. 8). Initial warning LTs ranged from 0 to 49 min for the Ellicott City event and from 50 to 91 min for the Baltimore event. FFWs issued well in advance of the first LSR generally had assigned minor flash flood probabilities of 60%–80%, which then steadily increased closer to the event occurrence. The more important trend was the assigned major flash flood probabilities, which also increased from 10% to 20% early in the event simulation to some participants assigning major flash flood probability values of 50%–70%. This signifies that the WoFS-FLASH products can portray the possibility of a major flash flood event before it occurs.

Participant perspectives of probabilistic hydrologic forecasts

Participants were surveyed every 30 min during the archived case simulations to assess their perceived expectations for flash flood severity, forecast confidence, experimental product usage, and operational decisions. The prompted feedback combined with FFW statistical evaluations offered insights into various characteristics of the human-element component

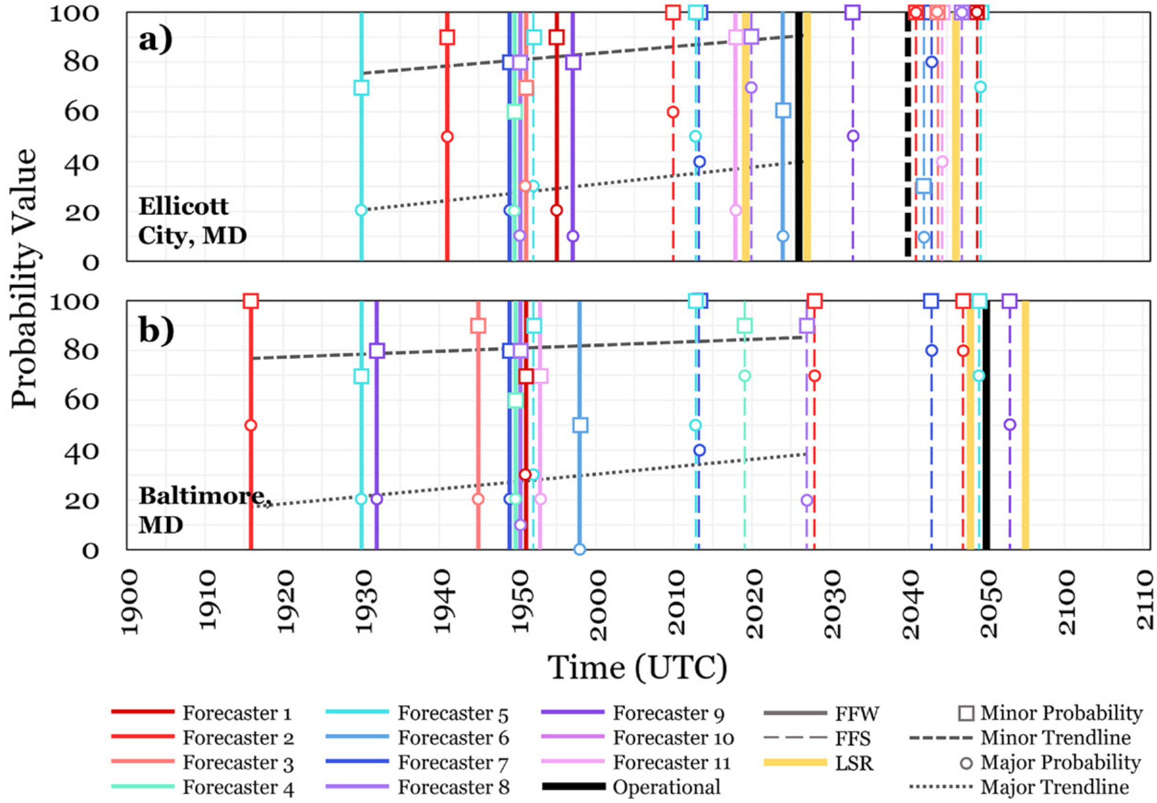

Fig. 8. Timeline of all products issued along with flash flood LSRs for (a) the Ellicott City, Maryland, event and (b) the Baltimore, Maryland, event during the Sterling, Virginia (LWX), simulation. Each timeline denotes the issuance of each experimental flash flood warning (FFWs; solid thin line) and FFW follow-up statements (FFS; dashed thin line) in red–blue–purple colors, the operational FFWs and FFS (thick black lines), and the flash flood LSRs (thick yellow lines). The assigned minor probabilities for each experimental FFW and FFS are denoted by a square. The assigned major probabilities for each experimental FFW and FFS are denoted by a circle. The dashed gray trendlines denote the increase in assigned minor and major probability values within experimental FFWs and FFSs up to the point of the first flash flood LSR at 2026 UTC.

of these simulations, including 1) differences in the decision-making thought process, 2) how the products and thought processes influenced warning performance, and 3) the evolution of flash flood expectations and confidence.

Participants providing longer warning LTs consistently incorporated experimental, probabilistic guidance into their decision-making, made sense of it alongside deterministic data, and often perceived the probability values to be sufficient for warning issuance (Fig. 9). Participant 5 (denoted as P5) was either the first or second participant to issue a FFW for each of the eight flash flood events across the three archived simulations. P5 discussed the incorporation of both deterministic and probabilistic data into their assessment strategy: "[There is a] natural progression of looking at a lot of WoFS data early, then gradually switching to deterministic data as the event ramps up."

Participants that provided shorter LTs or missed flash flood events relied more heavily on the deterministic data to make decisions. Participant 4 (P4) noted that the experimental products had "no influence" in their warning decisions during the simulation within the Sioux Falls, South Dakota, county warning area. Participant 8 (P8) noted hesitancy to trust the experimental guidance enough to issue a warning, especially when rainfall was not yet observed; moreover, P8 also perceived the forecasted probability values to be too low for flash flooding to occur. This resulted in both P4 and P8 having initial warning LTs under 20 min with instances of missed events.

Another trend identified within the participant surveys was the increase in flash flood forecast confidence and expectations as the simulation progressed. The progression of the

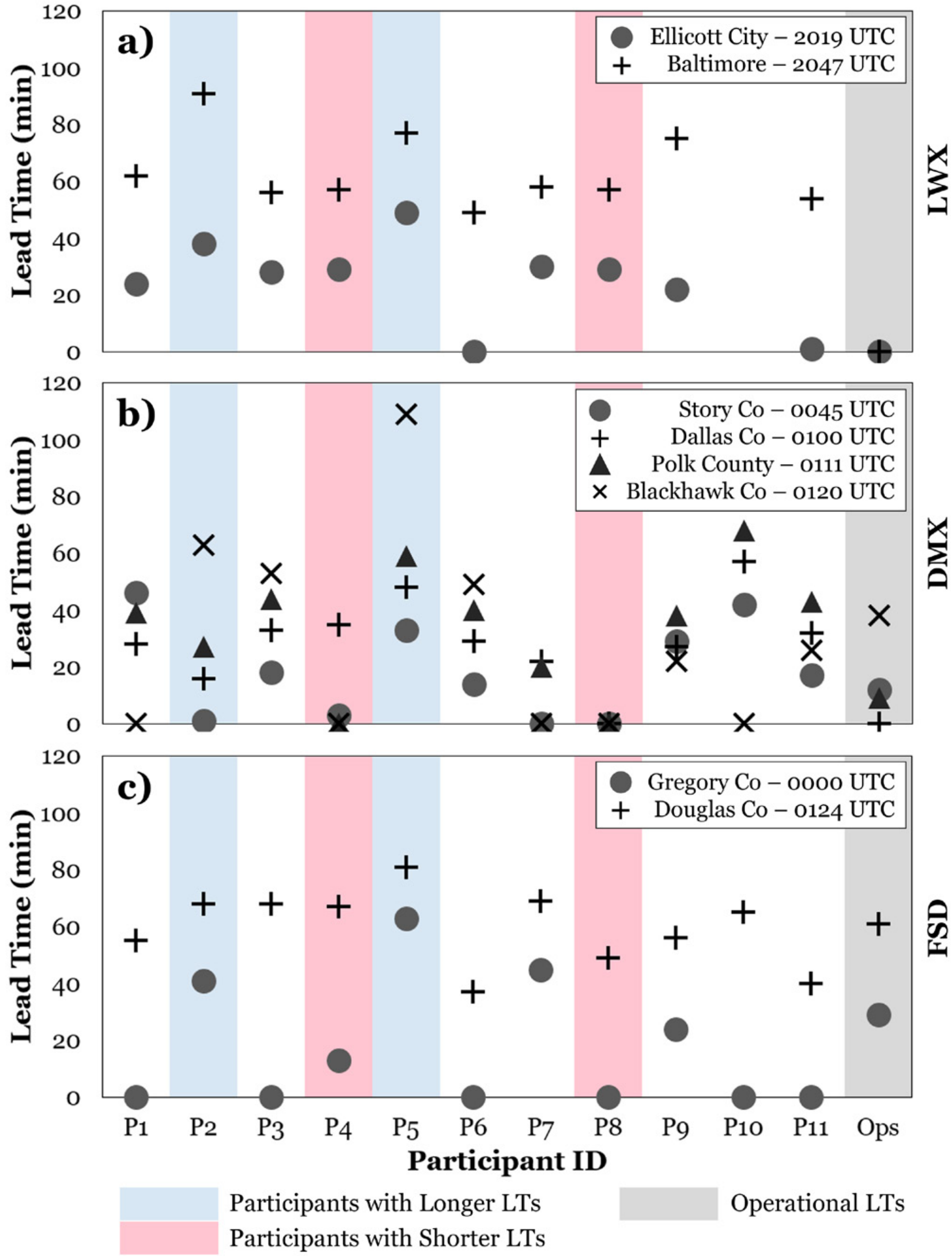

Fig. 9. **Analysis of the experimental FFW lead time (LT) for each participant (P1–P11) compared to the operational FFW LT (Ops) for the (a) Sterling, Virginia (LWX); (b) Des Moines, Iowa (DMX); and (c) Sioux Falls, South Dakota (FSD) simulations. The different markers represent the LTs for each flash flood event within the respective simulation. Two participants noted for longer LTs are highlighted in blue. Two participants noted for shorter LTs are highlighted in red.**

WoFS-FLASH data as it approached the time of the flash flood events provided increased forecast confidence levels (Fig. 10). The notable exception to this was the null event in Minnesota, when the majority confidence level of "confident" was achieved as observed rainfall and peak hydrologic modeling response occurred from 2300 to 0000 UTC and then declined after 0000 UTC (Fig. 10c). A similar pattern was observed in the flash flood expectations throughout each simulation (Fig. 11). The majority expectation level for flash flooding increased at least one level across the simulation period. The notable exception was again the null event in Minnesota (Fig. 11c), which had a rise-and-fall pattern similar to the participant confidence level.

Additional feedback through a structured end-of-week discussion provided further insight into some of the human factors that arose from testing WoFS-FLASH. The propensity to

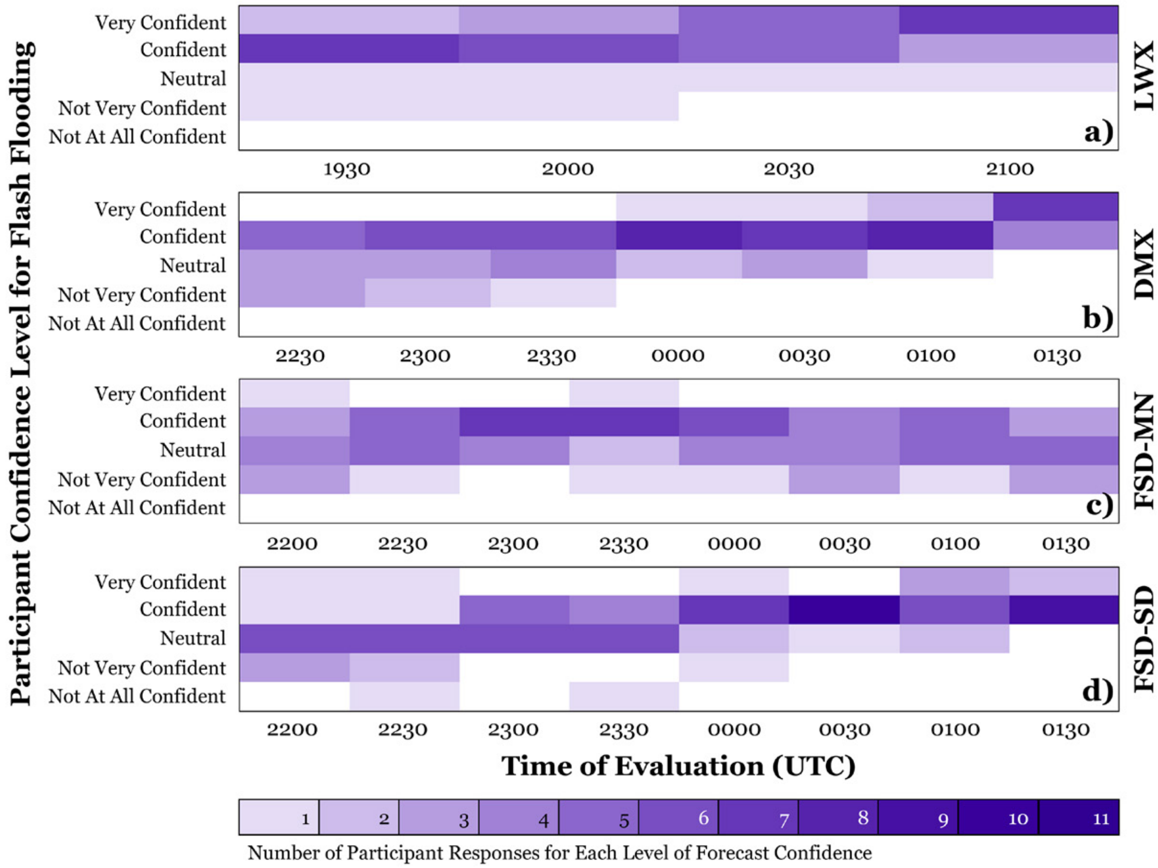

Fig. 10. **Distribution of participant responses on the confidence level of flash flooding every 30 min during each of the simulations: (a) Sterling, Virginia (LWX); (b) Des Moines, Iowa (DMX); and (c),(d) Sioux Falls, South Dakota (FSD). The FSD analysis is broken into two separate analyses: (c) the null event in southwest Minnesota and (d) the verified flash floods in South Dakota.**

trust the experimental products with QPF forcings was not uniform across the forecasters. Participants anticipated challenges with incorporating the WoFS-FLASH analysis into operations due to needing time to calibrate decision-making to model probabilities and to establish trust with the output, especially in cases where rain has not fallen yet in areas where forecast probabilities were generated. Hesitancy to trust the experimental flash flood forecasts by some forecasters was also documented; however, a more uniform consensus focused on incorporating WoFS guidance into the ability to positively impact communications prior to warnings. One 2018 participant noted how using WoFS-FLASH guidance can have its biggest advantage in the "period between watch and warning… [to] provide spatial focus, where threat is increasing… and use it to communicate to special users." A 2019 participant commented that even though QPF might not be spatially accurate, "I can go tell an emergency manager that the southern half of that county has a good chance of seeing flash flooding over the next three hours… a huge decision support services tool." Understanding the differences in forecasters' propensity to trust WoFS-driven experimental products will be important for determining the various means in which forecasters are comfortable using and applying new information to their existing forecast and decision-making processes.

Moving toward FACETs flash flood applications

The creation of flash flood PHI grids is inherently interdisciplinary to account for the identification and tracking of the causative storm(s), the accumulation of rainfall, and the forecast routing of pluvial and fluvial floodwaters. Utilizing probabilistic hydrologic model output within the PRO-FLASH framework and its coupling with WoFS QPFs showed potential for

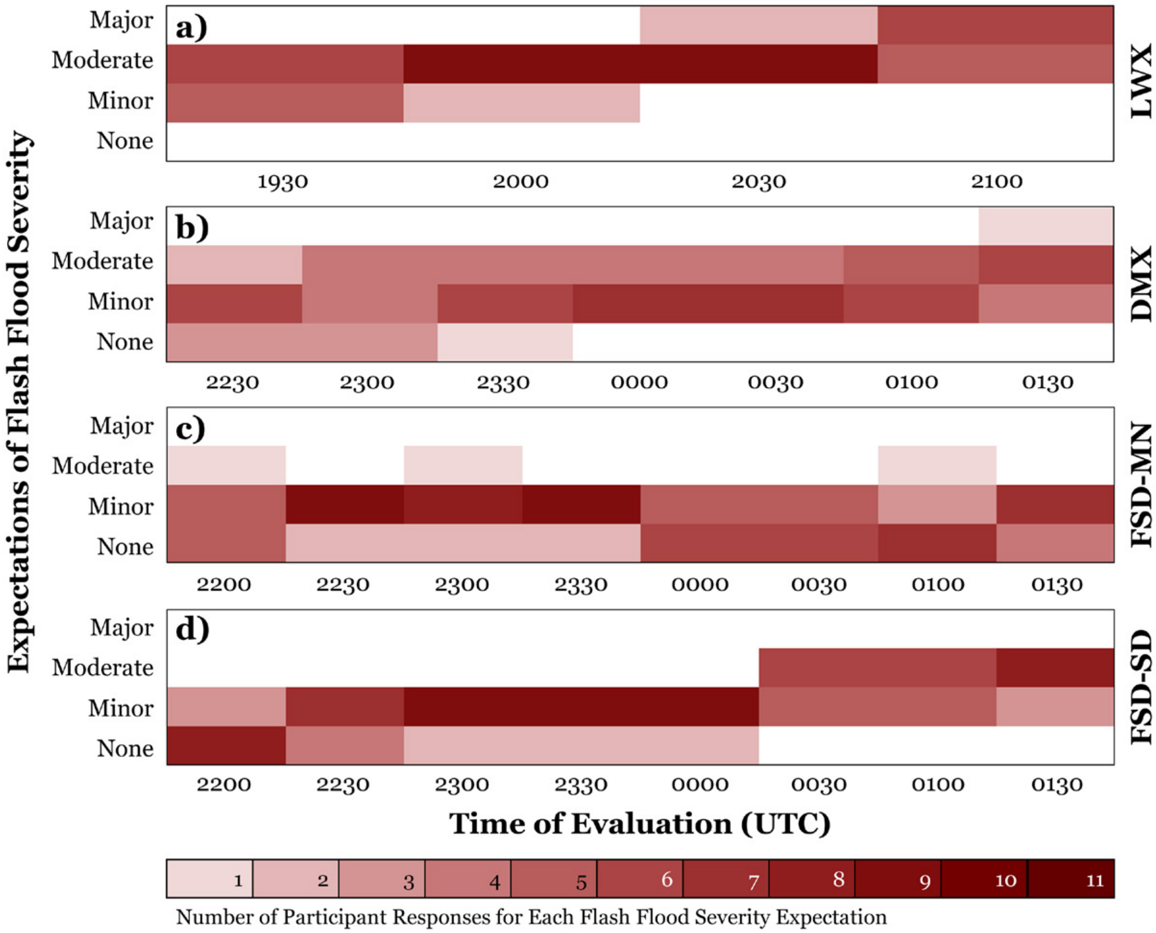

Fig. 11. As in Fig. 10, but for the participant expectations of flash flooding severity.

improving flash flood prediction and warnings LTs. The results also demonstrated how utilizing ensemble short-term model precipitation forecasts as a forcing can improve situational awareness and communication of the flash flood potential. The evolution of flash flood PHI grids will inherently focus on overland and channel routing from both observed precipitation and the forecast precipitation from the WoFS, yet the authors recognize that hydrologic modeling output is only one set of guidance that is incorporated into an operational warning decision. Continued development of a multivariable flash flood PHI grid(s) can leverage both hydrologic model outputs along with precipitation-based products and its uncertainty, such as probabilistic QPE (Kirstetter et al. 2015).

Advancing the FACETs concept for flash flooding beyond PHI grid development shall build upon the current research for the tracking and forecasting of a severe weather hazard. Previous efforts demonstrated how algorithms can identify a PHI object, a 2D geographic area for a defined severe-weather hazard, with an associated prediction of storm path and probabilistic trend (Karstens et al. 2015, 2018; James et al. 2020). The identification of a PHI object and its predictive attributes is imperative to warning applications within the FACETs concept. A PHI object can be generated using a recommender algorithm within the Hazard Services software (e.g., Nietfeld et al. 2018); moreover, a human-machine balance allows for forecaster modifications to PHI object characteristics (Karstens et al. 2018).

Attaching a warning to a PHI object is being tested using a threats-in-motion (TIM) warning approach, which advances the warning polygon with the movement of the PHI object (Stumpf and Gerard 2021). Flash flooding brings forth the additional complexity of forecasting and tracking the associated overland runoff. While it is important to identify and forecast the movement of the causative storm(s), impacts from excessive rainfall can occur after the

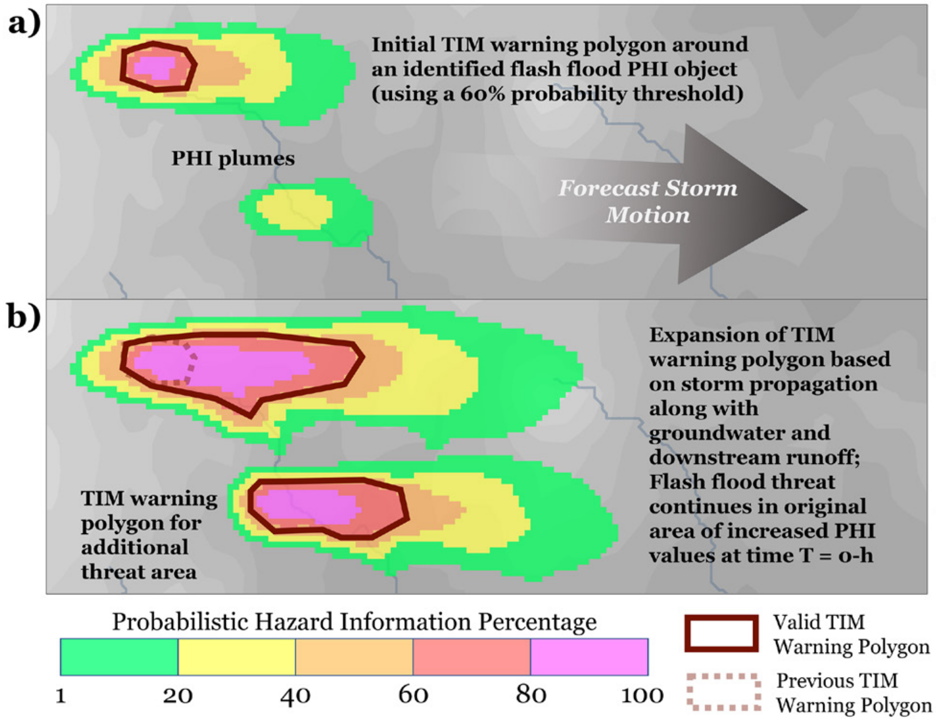

Fig. 12. **Conceptual model of flash flood PHI grid with a TIM warning polygon attached to a PHI object at (a) an initial time when the flash flood PHI grid would indicate the need for a TIM warning polygon and (b) a time in the future depicting the evolution of the flash flood probabilities and the expansion of the TIM warning polygon to encompass the expanding flash flood threat.**

causative precipitation has passed and beyond the area receiving precipitation due to flood wave propagation down channels. A future conceptual design of FACETs for flash flooding would encompass the creation of a PHI object that can expand based on the accumulation of excessive precipitation and the routing of floodwater (Fig. 12). The application and movement associated with a TIM warning polygon for a flash flood PHI object would differ from that of other convective severe weather hazards (e.g., tornadoes and hail). While causative precipitation features would travel, the lag time between the precipitation and the hydrologic response combined with the residence time of flood waters would mean that the TIM warning polygon could remain over an area while expanding for both the movement of the causative precipitation feature (assuming the feature is projected to continue generating excessive rainfall that resulted in high probabilistic values for flash flooding) and the predicted overland and channel runoff that could result in flash floods.

The lessons learned from both the HMT-Hydro experiment along with the ongoing studies and experiments conducted with convective severe weather hazards will further the maturation of the FACETs concepts for flash flooding; furthermore, developing more focused probabilistic threat areas and associated warnings would likely decrease the false alarm area and reduce the indirect costs for responding to warnings (Howard et al. 2021). This continued scientific research combined with associated societal studies regarding the understanding and application of probabilistic information are critical to maximizing the potential for informing end users and eliciting the appropriate actions to protect life and property.

Acknowledgments. The authors first want to thank the participants from the 2018–2019 HMT-Hydro experiment. The insights and information gathered from their involvement were valuable to these efforts to advance hydrometeorological products and techniques. The authors also thank the anonymous reviewers for their feedback and recommendations, as well as Patrick Burke, who provided a thorough review of the work presented here. The authors also want to thank Kate Abshire for her assistance with the NWS directives and the history of FFWs within the NWS. Funding was provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreements NA16OAR4320072 and NA21OAR4320204, U.S. Department of Commerce, as well as by the Hydrometeorology Testbed Program under Grant NA17OAR4590281.

Data availability statement. The PRO-FLASH system and its probabilistic products were developed using the Ensemble Framework For Flash Flood Forecasting (EF5; http://ef5.ou.edu/), streamflow values observed by the United State Geological Survey (USGS) stream gauges (https://nwis.waterdata.usgs.gov/nwis), and the CREST hydrologic model (http://hydro.ou.edu/research/CREST/). The PRO-FLASH system was forced by experimental QPEs from the MRMS system and the ensemble QPFs from the WoFS. The MRMS and WoFS forcings as well as PRO-FLASH system probabilistic output were archived internally at the National Severe Storms Laboratory and may be shared upon request pending consent of the original dataset creators. Flash flood reports used in experimental real-time operations and archived case playback were obtained through NWS *Storm Data* via the NWS Performance Management web portal (https://verification.nws.noaa.gov) and the NWS Local Storm Report application (https://nwschat.weather.gov/lsr/). Operational FFW polygons and statistics were also obtained through the NWS Performance Management web portal.

References

Arthur, A. T., G. M. Cox, N. R. Kuhnert, D. L. Slayter, and K. W. Howard, 2005: The National Basin Delineation Project. *Bull. Amer. Meteor. Soc.*, **86**, 1443–1452, https://doi.org/10.1175/BAMS-86-10-1443.

Barnes, L. R., D. M. Schultz, E. C. Gruntfest, M. H. Hayden, and C. C. Benight, 2009: Corrigendum: False alarm rate or false alarm ratio? *Wea. Forecasting*, **24**, 1452–1454, https://doi.org/10.1175/2009WAF2222300.1.

Clark, R. A., J. J. Gourley, Z. L. Flamig, Y. Hong, and E. Clark, 2014: CONUS-wide evaluation of National Weather Service flash flood guidance products. *Wea. Forecasting*, **29**, 377–392, https://doi.org/10.1175/WAF-D-12-00124.1.

Cocks, S. B., and Coauthors, 2019: A prototype quantitative precipitation estimation algorithm for operational S-band polarimetric radar utilizing specific attenuation and specific differential phase. Part II: Performance verification and case study analysis. *J. Hydrometeor.*, **20**, 999–1014, https://doi.org/10.1175/JHM-D-18-0070.1.

Elmore, K. L., Z. L. Flamig, V. Lakshmanan, B. T. Kaney, V. Farmer, H. D. Reeves, and L. P. Rothfusz, 2014: MPING: Crowd-sourcing weather reports for research. *Bull. Amer. Meteor. Soc.*, **95**, 1335–1342, https://doi.org/10.1175/BAMS-D-13-00014.1.

Flamig, Z. L., H. Vergara, and J. J. Gourley, 2020: The Ensemble Framework for Flash Flood Forecasting (EF5) v1.2: Description and case study. *Geosci. Model Dev.*, **13**, 4943–4958, https://doi.org/10.5194/gmd-13-4943-2020.

Gaviria Pabon, A. R., J. Hardy, J. J. Gourley, and T. Lindley, 2021: Understanding the correlation between NWS impact-based flash flood warning categories and various NWS flash flood tools. *20th Annual Student Conf.*, Online, Amer. Meteor. Soc., 98, https://ams.confex.com/ams/101ANNUAL/meetingapp.cgi/Paper/385131.

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, https://doi.org/10.1175/BAMS-D-15-00247.1.

Harrison, D. R., and C. D. Karstens, 2017: A climatology of operational storm-based warnings: A geospatial analysis. *Wea. Forecasting*, **32**, 47–60, https://doi.org/10.1175/WAF-D-15-0146.1.

Howard, S. P., K. E. Klockow-McClain, A. P. Boehmer, and K. M. Simmons, 2021: Firm behavior in the face of severe weather: Economic analysis between probabilistic and deterministic warnings. *Wea. Forecasting*, **36**, 757–767, https://doi.org/10.1175/WAF-D-20-0107.1.

Hudson, M. J., P. Browning, K. Runk, K. Harding, K. Galluppi, J. Losego, and B. Montz, 2013: Central Region impact-based warnings demonstration: Helping to build a weather-ready nation. *Eighth Symp. on Policy and Socio-Economic Research*, Austin, TX, Amer. Meteor. Soc., JPD1.1, https://ams.confex.com/ams/93Annual/webprogram/Paper222556.html.

James, J. J., C. Ling, C. D. Karstens, J. Correia Jr., K. Calhoun, T. Meyer, and D. LaDue, 2020: Forecasters' cognitive task analysis and mental workload analysis of issuing probabilistic hazard information (PHI) during FACETs PHI Prototype experiment. *Wea. Forecasting*, **35**, 1505–1521, https://doi.org/10.1175/WAF-D-19-0194.1.

Karstens, C. D., and Coauthors, 2015: Evaluation of a probabilistic forecasting methodology for severe convective weather in the 2014 Hazardous Weather Testbed. *Wea. Forecasting*, **30**, 1551–1570, https://doi.org/10.1175/WAF-D-14-00163.1.

——, and Coauthors, 2018: Development of a human-machine mix for forecasting severe convective events. *Wea. Forecasting*, **33**, 715–737, https://doi.org/10.1175/WAF-D-17-0188.1.

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. *Water Resour. Res.*, **51**, 1422–1442, https://doi.org/10.1002/2014WR015672.

Kumjian, M. R., 2013a: Principles and applications of dual-polarization weather radar. Part I: Description of the polarimetric radar variables. *J. Oper. Meteor.*, **1**, 226–242, https://doi.org/10.15191/nwajom.2013.0119.

——, 2013b: Principles and applications of dual-polarization weather radar. Part II: Warm and cold season applications. *J. Oper. Meteor.*, **1**, 243–264, https://doi.org/10.15191/nwajom.2013.0120.

——, 2013c: Principles and applications of dual-polarization weather radar. Part III: Artifacts. *J. Oper. Meteor.*, **1**, 265–274, https://doi.org/10.15191/nwajom.2013.0121.

Kutschenreuter, P. H., 1958: Rainfall and floods of April, May and June 1957 in the South-Central States. Weather Bureau, Department of Commerce Tech. Paper 33, 357 pp., https://hdsc.nws.noaa.gov/hdsc/files25/TP33.pdf.

Martinaitis, S. M., and Coauthors, 2017: The HMT Multi-Radar Multi-Sensor Hydro experiment. *Bull. Amer. Meteor. Soc.*, **98**, 347–359, https://doi.org/10.1175/BAMS-D-15-00283.1.

——, H. M. Grams, C. Langston, J. Zhang, and K. Howard, 2018: A real-time evaporation correction scheme for radar-derived mosaicked precipitation estimations. *J. Hydrometeor.*, **19**, 87–111, https://doi.org/10.1175/JHM-D-17-0093.1.

——, and Coauthors, 2020: The 23 June 2016 West Virginia flash flood event as observed through two hydrometeorology testbed experiments. *Wea. Forecasting*, **35**, 2099–2126, https://doi.org/10.1175/WAF-D-20-0016.1.

Nietfeld, D., K. L. Manross, D. M. Kingfield, N. Hardin, T. L. Hansen, Y. Guo, C. Golden, and H. Obermeier, 2018: The use of hazard services recommenders for convective weather products: Facilitating severe weather research to operations. *29th Conf. on Severe Local Storms*, Stowe, VT, Amer. Meteor. Soc., 118, https://ams.confex.com/ams/29SLS/webprogram/Paper348775.html.

NOAA, 2017: Operations and Services Water Resources Services Program, NWSPD 10-9, River Forecast Center Products Specification. National Weather Service Instruction 10-912, 21 pp., www.nws.noaa.gov/directives/sym/pd01009012curr.pdf.

——, 2019a: Operations and Services Water Resources Services Program, NWSPD 10-9, Definitions and General Terminology. National Weather Service Manual 10-950, 5 pp., www.nws.noaa.gov/directives/sym/pd01009050curr.pdf.

——, 2019b: Operations and Services Water Resources Services Program, NWSPD 10-9, Weather Forecast Office Water Resources Products Specification. National Weather Service Instruction 10-922, 92 pp., www.nws.noaa.gov/directives/sym/pd01009022curr.pdf.

——, 2020: Operations and Services Water Resources Services Program, NWSPD 10-9, Water Resources Services Program Management. National Weather Service Instruction 10-901, 28 pp., www.nws.noaa.gov/directives/sym/pd01009001curr.pdf.

NRC, 2006: *Completing the Forecast: Characterizing and Communicating Uncertainty for Better Decisions Using Weather and Climate Forecasts.* National Academies Press, 122 pp., https://doi.org/10.17226/11699.

Rigby, R. A., and D. M. Stasinopoulos, 2005: Generalized additive models for location, scale and shape. *J. Roy. Stat. Soc.*, **54**, 507–554, https://doi.org/10.1111/j.1467-9876.2005.00510.x.

Ripberger, J. T., C. L. Silva, H. C. Jenkins-Smith, and M. James, 2015: The influence of consequence-based messages on public responses to tornado warnings. *Bull. Amer. Meteor. Soc.*, **96**, 577–590, https://doi.org/10.1175/BAMS-D-13-00213.1.

Rothfusz, L. P., R. Schneider, D. Novak, K. Klockow-McClain, A. E. Gerard, C. Karstens, G. J. Stumpf, and T. M. Smith, 2018: FACETs: A proposed next-generation paradigm for high-impact weather forecasting. *Bull. Amer. Meteor. Soc.*, **99**, 2025–2043, https://doi.org/10.1175/BAMS-D-16-0100.1.

Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting*, **5**, 570–575, https://doi.org/10.1175/1520-0434(1990)005<0570:TCSIAA>2.0.CO;2.

Schroeder, A. J., and Coauthors, 2016: The development of a flash flood severity index. *J. Hydrol.*, **541**, 523–532, https://doi.org/10.1016/j.jhydrol.2016.04.005.

——, R. Smith, T. Lindley, G. Kendrick, A. Perroux, A. Treadway, M. C. Oaks, and P. Yura, 2020: Flash Flood Severity Index (FFSI): Operational application in the field. *34th Conf. on Hydrology*, Boston, MA, Amer. Meteor. Soc., 569, https://ams.confex.com/ams/2020Annual/meetingapp.cgi/Paper/368073.

Smith, S. B., M. T. Filiaggi, M. Churma, J. Roe, M. Glaudemans, R. Erb, and L. Xin, 2000: Flash flood monitoring and prediction in AWIPS Build 5 and beyond. *15th Conf. on Hydrology*, Long Beach, CA, Amer. Meteor. Soc., 3.7, https://ams.confex.com/ams/annual2000/techprogram/paper_18.htm.

Stensrud, D. J., and Coauthors, 2009: Convective-scale warn-on-forecast system: A vision for 2020. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1499, https://doi.org/10.1175/2009BAMS2795.1.

——, and Coauthors, 2013: Progress and challenges with Warn-on-Forecast. *Atmos. Res.*, **123**, 2–16, https://doi.org/10.1016/j.atmosres.2012.04.004.

Stumpf, G. J., and A. E. Gerard, 2021: National Weather Service severe weather warnings as Threats-in-Motion. *Wea. Forecasting*, **36**, 627–643, https://doi.org/10.1175/WAF-D-20-0159.1.

Sutter, D., and S. Erickson, 2010: The time cost of tornado warnings and the savings with storm-based warnings. *Wea. Climate Soc.*, **2**, 103–112, https://doi.org/10.1175/2009WCAS1011.1.

Sweeney, T. L., 1992: Modernized areal flash flood guidance. NOAA Tech. Rep. NWS HYDRO 44, 37 pp., https://repository.library.noaa.gov/view/noaa/13498.

Uccellini, L. W., and J. E. Ten Hoeve, 2019: Evolving the National Weather Service to build a Weather Ready Nation: Connecting observations, forecasts, and warnings to decision-makers through impact-based decision support services. *Bull. Amer. Meteor. Soc.*, **100**, 1923–1942, https://doi.org/10.1175/BAMS-D-18-0159.1.

Wang, J., and Coauthors, 2011: The Coupled Routing and Excess STorage (CREST) distributed hydrological model. *Hydrol. Sci. J.*, **56**, 84–98, https://doi.org/10.1080/02626667.2010.543087.

Wang, Y., S. B. Cocks, L. Tang, A. Ryzhkov, P. Zhang, J. Zhang, and K. Howard, 2019: A prototype quantitative precipitation estimation algorithm for operational S-band polarimetric radar utilizing specific attenuation and specific differential phase. Part I: Algorithm description. *J. Hydrometeor.*, **20**, 985–997, https://doi.org/10.1175/JHM-D-18-0071.1.

Wilson, K. A., B. T. Gallo, P. Skinner, A. Clark, P. Heinselman, and J. J. Choate, 2021: Analysis of end user access of Warn-on-Forecast guidance products during an experimental forecasting task. *Wea. Climate Soc.*, **13**, 859–874, https://doi.org/10.1175/WCAS-D-20-0175.1.

Yussouf, N., D. C. Dowell, and L. J. Wicker, 2015: Storm-scale data assimilation and ensemble forecasts for the 27 April 2011 severe weather outbreak in Alabama. *Mon. Wea. Rev.*, **143**, 3044–3066, https://doi.org/10.1175/MWR-D-14-00268.1.

——, K. A. Wilson, S. M. Martinaitis, H. Vergara, P. L. Heinselman, and J. J. Gourley, 2020: The coupling of NSSL Warn-on-Forecast and FLASH systems for probabilistic flash flood prediction. *J. Hydrometeor.*, **21**, 123–141, https://doi.org/10.1175/JHM-D-19-0131.1.

Zevin, S. F., 1994: Steps toward an integrated approach to hydrometeorological forecasting services. *Bull. Amer. Meteor. Soc.*, **75**, 1267–1276, https://doi.org/10.1175/1520-0477-75.7.1267.

Zhang, J., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 621–638, https://doi.org/10.1175/BAMS-D-14-00174.1.

——, L. Tang, S. Cocks, P. Zhang, A. Ryzhkov, K. Howard, C. Langston, and B. Kaney, 2020: A dual-polarization radar synthetic QPE for operations. *J. Hydrometeor.*, **21**, 2507–2521, https://doi.org/10.1175/JHM-D-19-0194.1.