

Improving Flash Flood Forecasts

The HMT-WPC Flash Flood and Intense Rainfall Experiment

BY FAYE E. BARTHOLD, THOMAS E. WORKOFF, BRIAN A. COSGROVE, JONATHAN J. GOURLEY, DAVID R. NOVAK, AND KELLY M. MAHONEY

Despite advancements in numerical modeling and the increasing prevalence of convection-allowing guidance, warm season (June–August) quantitative precipitation forecasts (QPF) remain challenging (Fritsch and Carbone 2004). While overall threat scores have increased steadily over the past 50 years, the majority of that improvement has occurred during the cold season (December–February), with only incremental improvements during the warm season (Fig. 1). The difference in forecast skill and rate of improvement is likely driven by the spatial and temporal scale of precipitation events during each season (Sukovich et al. 2014). While precipitation during the cold season tends to be driven by synoptic-scale events (e.g., extratropical cyclones), precipitation during the warm season is often driven by small-scale convective processes (e.g., thunderstorms) that are more difficult for numerical models to accurately simulate.

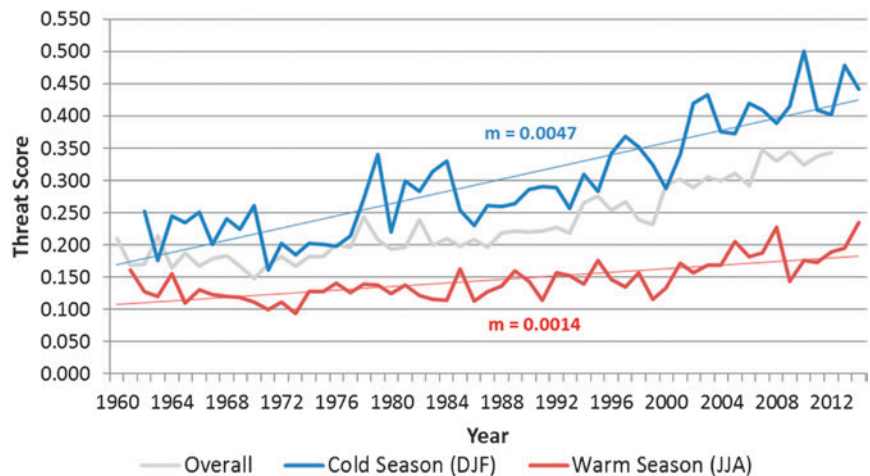


FIG. 1. Comparison of Day 1 WPC 1.0 in QPF threat scores from 1961 to 2014 during the cold season (blue) and warm season (red). The annual average threat score (gray) is provided for reference.

Flash flooding, defined as a rapid and extreme flow of high water into a normally dry area, or a rapid water level rise in a stream or creek above a predetermined flood level that begins within six hours of the causative event (NOAA 2012), introduces another variable into the forecast equation. In addition to accurate QPF information, the hydrologic aspect of a flash flood

AFFILIATIONS: BARTHOLD*—I.M. Systems Group, Inc., Rockville, Maryland, and NOAA/NWS/Weather Prediction Center, College Park, Maryland; WORKOFF*—Systems Research Group, Inc., Colorado Springs, Colorado, and NOAA/NWS/Weather Prediction Center, College Park, Maryland; COSGROVE—NOAA/NWS/National Water Center, Silver Spring, Maryland; GOURLEY—NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma; NOVAK—NOAA/NWS/Weather Prediction Center, College Park, Maryland; MAHONEY—Cooperative Institute for Research in the Environmental Sciences, University of Colorado Boulder, and NOAA/ESRL/Physical Sciences Division, Boulder, Colorado

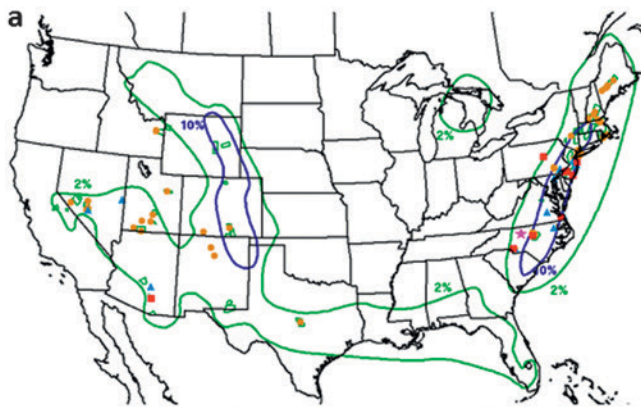
***CURRENT AFFILIATION:** NOAA/NWS/WFO New York, NY, Upton, New York

***CURRENT AFFILIATION:** FirstEnergy, Akron, Ohio

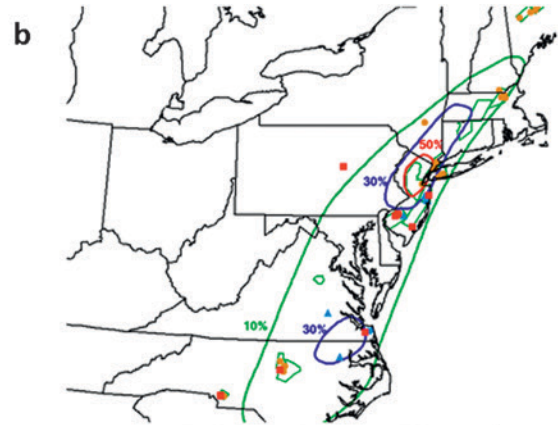
CORRESPONDING AUTHOR: Faye E. Barthold, NOAA/NWS/WFO New York, NY, 175 Brookhaven Avenue, Upton, NY 11973
E-mail: faye.barthold@noaa.gov

DOI:10.1175/BAMS-D-14-00201.1

©2015 American Meteorological Society



FFaIR 18 hr Flash Flood Outlook valid 140716/1200 and flash flood warnings (green), flash flood LSRs (brown), flood LSRs (blue), mPING reports (red), USGS stream gage exceedance (pink)



FFaIR 6 hr Flash Flood Forecast valid 140716/0000 and flash flood warnings (green), flash flood LSRs (brown), flood LSRs (blue), mPING reports (red), USGS stream gage exceedance (pink)

FIG. 2. Example of an experimental (a) 18-h probabilistic flash flood outlook forecast valid 1200 UTC 16 Jul and (b) 6-h probabilistic flash flood forecast valid 0000 UTC 16 Jul from the 2014 Flash Flood and Intense Rainfall Experiment. Both forecasts were defined as the probability of flash flooding within 40 km of a point. The experimental forecast contours are overlaid on the corresponding flash flood warnings (green) and observations including flash flood LSRs (brown), flood LSRs (blue), mPING reports (red), and USGS stream gage exceedance (pink).

forecast represents a challenge on its own as recent rainfall, soil type, slope, land use, basin size, degree of urbanization, etc., all play a role in determining the flash flood threat. This combination of meteorological and hydrologic challenges allows flash flooding to remain one of the deadliest weather phenomena, typically resulting in more fatalities each year than either tornadoes or hurricanes.

Compounding these forecast challenges is the lack of a real-time comprehensive flash flood verification dataset. At present, flash flood observations are available from a variety of different sources (e.g., National Weather Service, U.S. Geological Survey) in a variety of different forms, each with varying levels of detail and accuracy. The development of a real-time dataset that combines the best available observations is vital to assessing the quality of current flash flood forecasts and identifying areas in need of improvement.

As the national center responsible for providing QPF and flash flood forecast guidance, NCEP's Weather Prediction Center (WPC) and the Hydrometeorology Testbed at WPC (HMT-WPC) are uniquely positioned to address the challenges associated with flash flood forecasting. WPC currently issues two products that address the flash flood threat: the Excessive Rainfall Outlook (ERO) and the Mesoscale

Precipitation Discussion (MPD). The ERO is issued at scheduled intervals throughout the day as part of WPC's Day 1–3 QPF product suite and indicates the probability of exceeding flash flood guidance (FFG) at a point across the contiguous United States (CONUS). With the potential of multiday lead times, these products are intended to provide several days of advance notice about the potential for flash flooding. To address the near-term flash flood threat, WPC began issuing MPDs in April 2013. These event-driven forecasts highlight regions where heavy rainfall may lead to flash flooding over the next 1–6 h and are designed to enhance situational awareness among local National Weather Service (NWS) offices, emergency managers, and the media.

THE FLASH FLOOD AND INTENSE RAINFALL (FFaIR) EXPERIMENT. In an effort to improve flash flood forecasts and verification both at WPC and across the NWS, HMT-WPC partnered with the National Severe Storms Laboratory (NSSL) and the Earth System Research Laboratory (ESRL) in 2013 to develop and host the first annual Flash Flood and Intense Rainfall (FFaIR) Experiment. The FFaIR Experiment brings together participants from the operational forecasting, model development, and

research communities during the month of July to explore the challenges associated with flash flood forecasting. In particular, the experiments have focused on

- evaluating the utility of high-resolution convection-allowing models and ensembles for short-term flash flood forecasts,
- exploring new tools and approaches for combining meteorological and hydrologic information, and
- exploring improvements to WPC's operational suite of flash flood forecast guidance.

During the experiment, participants used a combination of operational and experimental model output to issue a series of experimental forecasts (Fig. 2). These forecasts, along with the corresponding model guidance, were then subjectively evaluated to gain insight about the utility of the experimental data under real-time operational conditions. The experimental forecasting environment provides an opportunity to test new forecast tools—such as neighborhood probabilities—and gather feedback about different verification approaches that can help inform future improvements to flash flood forecasts.

NEIGHBORHOOD FFG EXCEEDANCE PROBABILITIES. With the creation of WPC's MPDs, there was a need to develop guidance that combines QPF and hydrologic information to aid forecasters in identifying areas at risk for flash flooding. Using the concept of the “neighborhood maximum” technique (e.g., Schwartz et al. 2009, Schwartz et al. 2010, Ebert 2008), HMT-WPC uses convection-allowing model guidance to develop probabilistic forecast tools highlighting areas where QPF may approach or exceed FFG, which indicates the average amount of rain needed over an area during a specific time period to initiate flooding on small streams (Sweeney 1992). The neighborhood maximum technique is a postprocessing procedure that identifies the maximum value of a parameter within a user-defined search radius of each grid point. Once identified, the original value of the grid point is then replaced by its neighborhood maximum value; the benefit of this technique is to help account for spatial and temporal errors that are inherent in high-resolution forecasts.

HMT-WPC applies this approach to the QPF from members of the ~4-km grid spacing Storm Scale Ensemble of Opportunity (SSEO, Jirak et al.

2012), created by the Storm Prediction Center (SPC). Generated at 0000 and 1200 UTC, the seven-member SSEO combines various operational (NCEP ARW and NMMB high-resolution windows and NAM Nest) and nonoperational (NSSL WRF-ARW and EMC WRF-NMM) deterministic convection-allowing models. For flash flood purposes, the neighborhood maximum QPF (nQPF) fields are compared against CONUS 3-h and 6-h gridded FFG; these CONUS FFG grids are 5-km mosaic grids that combine the gridded FFG (Schmidt et al. 2007) that is generated by each NWS River Forecast Center (RFC). The result is a product indicating the neighborhood probabilities of QPF exceeding FFG (nPQPF > FFG), and highlights areas at risk for flash flooding.

To understand how to combine QPF and FFG into a forecast tool, HMT-WPC has evaluated different versions of the nPQPF > FFG tool in the FFaIR Experiment by comparing experimental forecast guidance to various flash flood verification metrics such as QPE, QPE > FFG, QPE recurrence intervals, flash flood warnings (FFW), and flash flood observations (e.g., NWS local storm reports). These evaluations have included exploring the use of various search radii (point, 20 km, and 40 km) in the neighborhood maximum technique (Fig. 3), as well as using different percentages of FFG (75%, 90%, or 100% of FFG values) as the exceedance threshold (Fig. 4). Having participants test and evaluate various permutations of the nPQPF > FFG products in a real-time environment has been critical in the continued development and improvement of the product.

Results of the FFaIR Experiment have shown that forecasters prefer using a 40-km search radius for nPQPF > FFG because it best highlights the potential for flash flood events (Fig. 3d). Forecasters also found that using 100% of FFG as a threshold value is the most effective way to identify areas at risk of flash flooding (Fig. 4d). They also noted that while the 75% threshold could provide helpful information about the potential for flash flooding in more uncertain situations, it had a tendency to highlight the potential for flash flooding over too large of an area. Beginning 1 June 2014, HMT-WPC began producing a modified version of the SSEO four times a day (0000, 0600, 1200, 1800 UTC) that utilizes the latest convection-allowing models and FFG, and replaces the EMC WRF-NMM (due to its noted high QPF bias) with the High Resolution Rapid Refresh (HRRR) model. As a result of the experiment, all of these products are currently available to WPC forecasters.

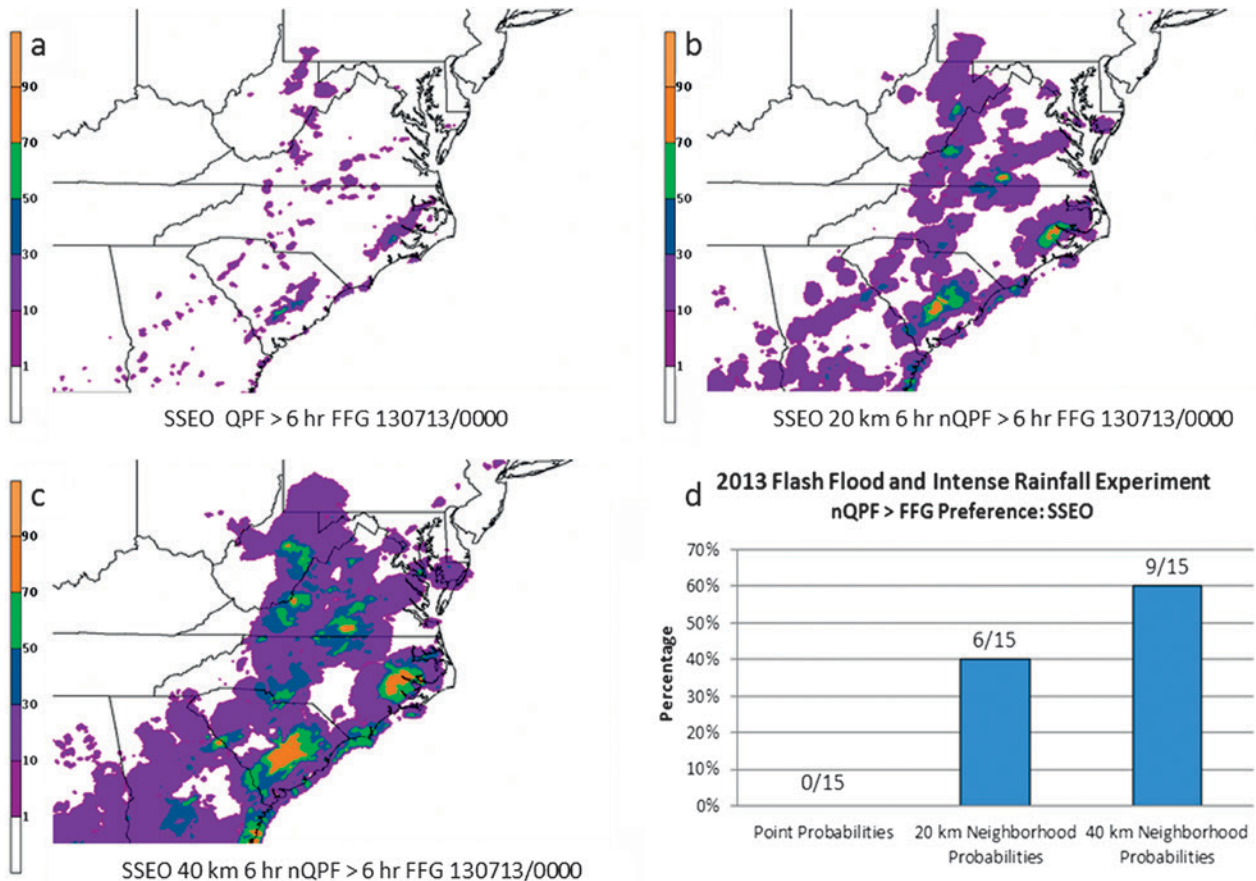


FIG. 3. Ensemble probability of 6-h QPF exceeding 6-h FFG from the SSEO valid at 0000 UTC on 13 Jul 2013 using (a) point probabilities, (b) neighborhood probabilities at a 20-km search radius, and (c) neighborhood probabilities at a 40-km search radius. (d) Shows results of subjective verification conducted in the 2013 FFaIR Experiment in which participants were asked daily which search radius (point, 20 km, or 40 km) provided the best guidance in terms of highlighting areas that received flash flooding.

FLASH FLOOD VERIFICATION. In addition to the development of new forecast tools, obtaining a complete and accurate assessment of when and where flash floods occur is a critical task (Gourley et al. 2013) for calibrating forecasts and improving forecast skill. Unfortunately, a single comprehensive source of real-time flash flood verification data does not currently exist. As such, it was decided to leverage three CONUS-wide hydrologic data sources to create a new merged, real-time verification dataset: NWS flash flood Local Storm Reports (LSRs), NSSL Meteorological Phenomena Identification Near the Ground (mPING, Elmore et al. 2014) reports, and United States Geological Survey (USGS) stream gauge measurements.

LSRs are an official NWS product, although reports can be subjective in nature and are dependent on people witnessing an event; darkness, low population density, etc., can limit reporting. Additionally, event

categorization (flash flood versus flood versus heavy rain) can be inconsistent across the NWS, and location and time stamp errors can also occur. mPING reports, a method of crowdsourcing weather information developed by NSSL, are also dependent on submission by end users. Unlike LSRs, they do not undergo quality control and do not differentiate between floods and flash floods, although NSSL's examination of the reports indicates that they are mainly flash floods.

The third component of the database centers on USGS stream gauge reports. To the best knowledge of the authors, this represents the first CONUS-wide effort to leverage this resource for real-time verification of flash flooding. The dataset is composed of stage and discharge data collected at all-weather automated USGS stream gauges across the CONUS every 5–60 min.

To extract natural flash flood event signals, real-time data from USGS basins smaller than 2,000 km² are

passed through a series of sequential filters. Flooding is judged to occur if 1) discharge exceeds the minor flood stage discharge, 2) the stage exceeds the NWS minor flood stage, or 3) discharge exceeds the two-year recurrence discharge. This latter filter is used only when the flood stage is unavailable, as it is only a rough indicator of bank-full status. Streams that fall into the “flooding” category then pass through a $\geq 3 \text{ ft hr}^{-1}$ rate-of-rise check to differentiate between floods and flash floods, and must be preceded or followed by a $\geq 1 \text{ ft hr}^{-1}$ rate-of-rise to insure that the signal is not a one-time spurious data spike. This series of filters is based on a limited review of approximately 150 flash flood cases and will be refined as the database matures.

These datasets are stored in a searchable Postgres database that is updated throughout the day to capture the latest observations. During the 2014 FFAIR

Experiment, this database was used to plot point observations of flash flooding to aid in the subjective evaluation of the experimental forecasts and model guidance (Fig. 2). This helped promote discussion about the best applications of the various datasets available to evaluate flash flooding. In addition, the development of this database also provided an opportunity to explore a new approach to probabilistic flash flood verification through the SPC-pioneered Practically Perfect analysis technique (Brooks et al. 1998, Hitchens et al. 2013). This experimental technique uses a Gaussian weighted function to convert the location of user-selected flash flood observations into a probabilistic forecast with the goal of producing the forecast a forecaster would have issued had the location of all reported flash flooding been known in advance (Fig. 5).

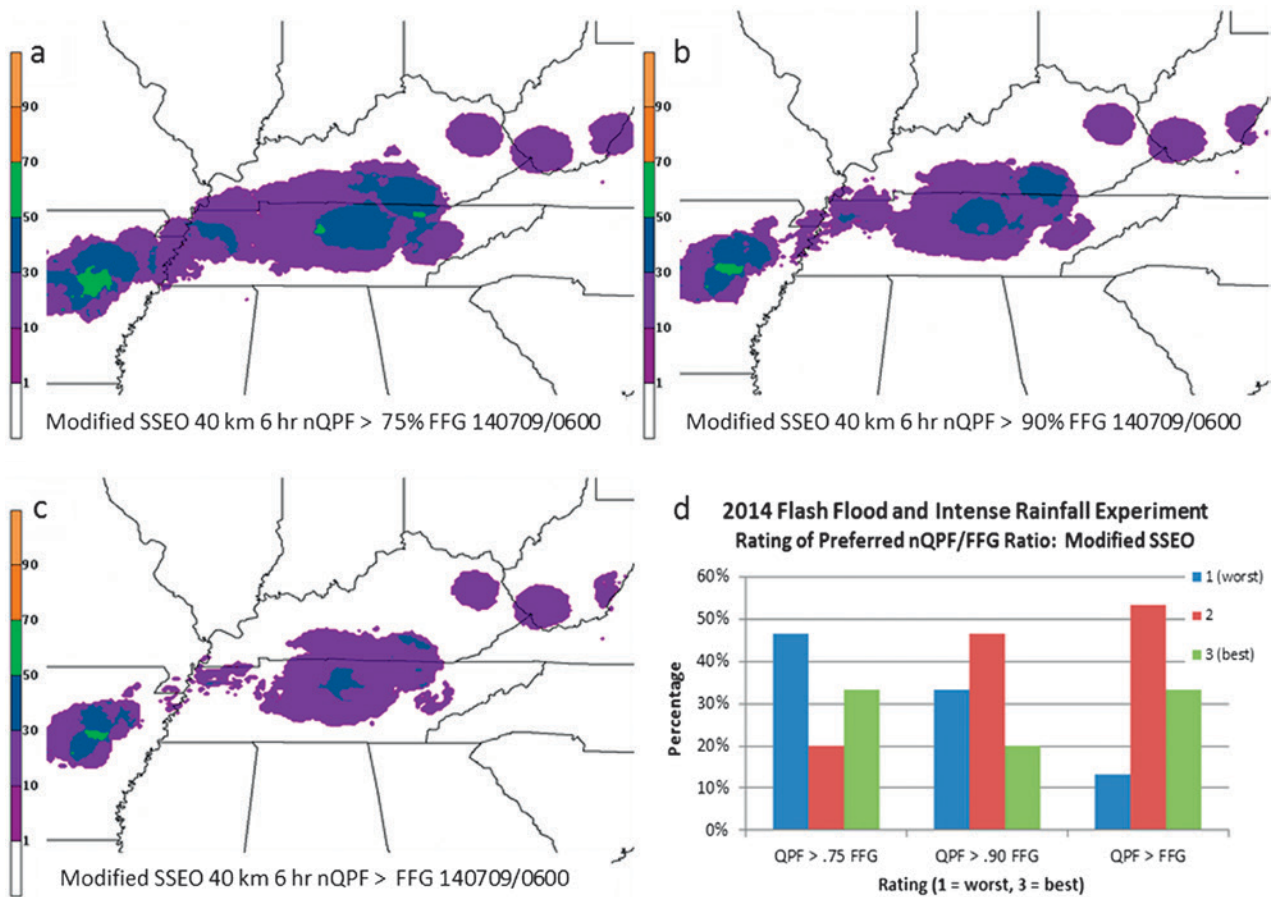


FIG. 4. Ensemble probability of 6-h QPF exceeding 6-h FFG from the modified SSEO valid at 0600 UTC on 9 Jul 2014 using (a) 75% of FFG (b) 90% of FFG, and (c) FFG as the exceedance threshold. (d) Results of subjective verification conducted in the 2014 FFAIR Experiment in which participants were asked daily which FFG threshold (FFG, 75% of FFG, or 90% of FFG) provided the best guidance in terms of highlighting areas that received flash flooding.

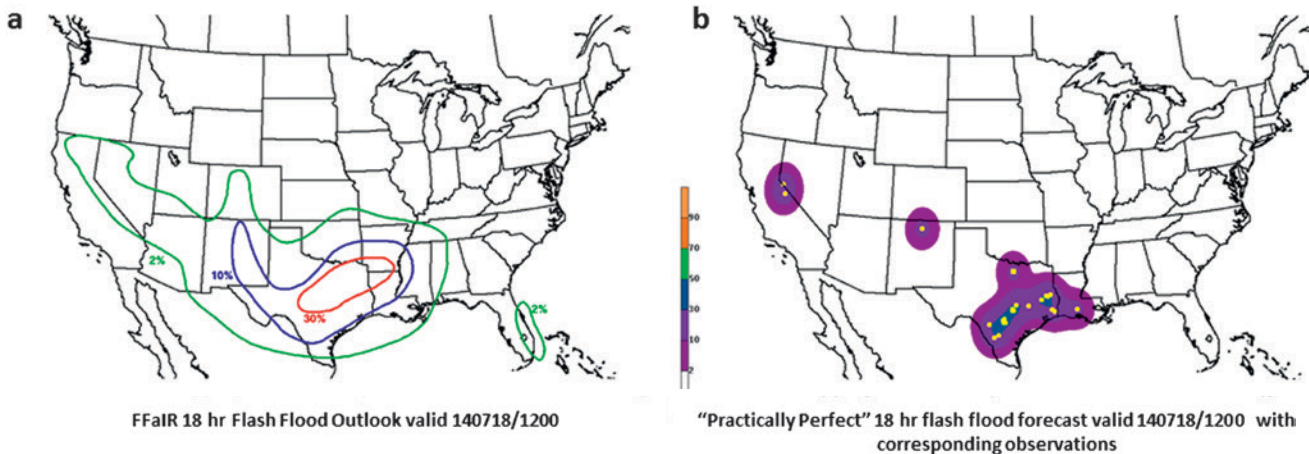


FIG. 5. Example of an (a) experimental 18-h probabilistic flash flood outlook forecast valid 1200 UTC 18 Jul, and (b) the corresponding experimental probabilistic flash flood forecast verification from the Practically Perfect analysis technique. The observations used to derive the Practically Perfect analysis are overlaid on (b).

FINAL THOUGHTS. Despite advancements in convection-allowing numerical modeling and precipitation forecasting in general, flash flooding remains one of the most difficult, yet most important, phenomena to forecast accurately. HMT-WPC initiated the FFaIR Experiment in 2013 to investigate ways to advance flash flood forecasting, better understand ways to communicate the flash flood threat, and develop a method to improve flash flood verification.

In its first two years, the FFaIR Experiment has made steady progress in aiding the development and testing of flash flood forecasting and verification tools. Through its experimental forecasting exercises and verification activities, FFaIR has shown that coupling QPF and hydrologic guidance (e.g., FFG) through the use of the neighborhood probability technique is a useful forecast tool for identifying areas of flash flood risk across the county. As a result, these neighborhood probabilities are now used by WPC forecasters in the process of generating both Mesoscale Precipitation Discussions (MPD) and the Day 1 Excessive Rainfall Outlook (ERO). Additionally, the development of a database which collects various flash flood observations in real time is a valuable resource for improving the awareness and accuracy of identifying where flash flooding occurs. These observations are now being used internally at WPC in the subjective verification of the ERO and MPD products.

While the development of these new forecast tools and the availability of a real-time flash flood verification database represent important steps forward, improvements in the quality of flash flood forecasts will

not be made until the meteorological and hydrologic communities come together to address the remaining forecast challenges:

- *Improved hydrologic guidance*—Although FFG is readily available, it has numerous limitations, and the development of improved hydrologic datasets targeted toward flash flood applications is necessary.
- *Improved warm-season model QPF guidance*—Continued investment in the development of convection-allowing models, including the establishment of an operational storm scale ensemble, is vital to supporting the continued improvement of flash flood forecasts.
- *Improved flash flood forecast tools*—While neighborhood probabilities of QPF > FFG are useful, additional approaches to combining meteorological and hydrologic data need to be explored to increase forecast accuracy and lead time.

Although overcoming these challenges will require community-wide long-term development efforts, work is already under way to address some of these issues. For example, NSSL’s Flooded Locations and Simulated Hydrographs (FLASH) project (Hong and Gourley 2015) and the NOAA Office of Hydrologic Development’s Distributed Hydrologic Model-Threshold Frequency (DHM-TF) project (Reed et al. 2007) are focusing on the direct simulation of routed streamflow. Additionally, the Weather Research and Forecasting Model Hydrological modeling extension package (WRF-Hydro, Gochis et al.

2014) is advancing the use of convection-allowing model and ensemble data to directly force distributed hydrologic models.

Adding to the challenge, the current approach to the flash flood forecast problem is inconsistent across the NWS. These inconsistencies are present beginning with the definition of flash flooding (how should urban and poor-drainage street flooding be handled?) and continue into the forecasting and reporting of these events. For example, when heavy rain occurs in an urban area, there are a plethora of products available to a local NWS forecast office to communicate the threat: Flash Flood Warning, Urban and Small Stream Advisory, and Flood Warning. The best product to use often depends on the impact, which may be unknown. In addition, the product definitions themselves have overlap. Similarly, the classification of reported events may depend on unknown attributes (how fast did the water rise, is this a poor drainage area, etc.). The subjectivity involved in choosing the type of forecast product and classifying the observations results in a confusing picture of flash flooding across the nation and contributes to the challenge of both flash flood forecast verification and forecast improvement.

With the help of both the meteorological and hydrologic communities, HMT-WPC will continue to foster an environment of collaboration between forecasters, researchers, and model developers through the annual FFaIR Experiment. Building off of these initial lessons learned, future FFaIR Experiments will focus on testing the use of additional hydrologic guidance and convection-allowing ensembles, investigating the best ways to convey the flash flood risk through probabilistic forecasts, and exploring enhancements to the verification database to develop a more complete record of flash flooding across the nation.

ACKNOWLEDGMENTS. This work was supported by NOAA's Hydrometeorology Testbed (HMT) program and the U.S. Weather Research Program (USWRP). The authors thank all past FFaIR Experiment participants for their enthusiasm and valuable insight into the flash flood forecast problem. In addition, Michael Smith (National Water Center), Russ Schumacher (Colorado State University), and two anonymous reviewers provided helpful feedback to improve the quality of this paper.

FOR FURTHER READING

Brooks, H. E., M. Kay, and J. A. Hart, 1998: Objective limits on forecasting skill of rare events. *Preprints*,

19th Conf. on Severe Local Storms, Minneapolis, MN, Amer. Meteor. Soc., 552–555.

Ebert, E. E., 2008: Fuzzy verification of high resolution gridded forecasts: A review and proposed framework. *Meteor. Appl.*, **15**, 51–66, doi:10.1002/met.25.

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

Fritsch, J. M., and R. E. Carbone, 2004: Improving quantitative precipitation forecasts in the warm season: A USWRP research and development strategy. *Bull. Amer. Meteor. Soc.*, **85**, 955–964, doi:10.1175/BAMS-85-7-955.

Gochis, D. J., W. Yu, and D. N. Yates, 2014: The WRF-Hydro model technical description and user's guide, version 2.0. NCAR Technical Document, 120 pp.

Gourley, J. J., and Coauthors, 2013: A unified flash flood database across the United States. *Bull. Amer. Meteor. Soc.*, **94**, 799–805, doi:10.1175/BAMS-D-12-00198.1.

Hitchens, N. M., H. E. Brooks, and M. P. Kay, 2013: Objective limits on forecasting skill of rare events. *Wea. Forecasting*, **28**, 525–534, doi:10.1175/WAF-D-12-00113.1.

Hong, Y., and J. J. Gourley, 2015: *Radar Hydrology: Principles, Models, and Applications*. 1st ed. CRC Press, 180 pp.

Jirak, I. L., S. J. Weiss, and C. J. Melick, 2012: The SPC storm-scale ensemble of opportunity: Overview and results from the 2012 Hazardous Weather Testbed Spring Forecasting Experiment. *Preprints, 26th Conf. on Severe Local Storms*, Nashville, TN. Amer. Meteor. Soc. P9.137.

NOAA, 2012: Hydrologic Services Program, Definitions and General Terminology. National Weather Service Manual 10-950, 5 pp, [Available online at www.noaa.gov/directives/sym/pd01009050curr.pdf].

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, doi:10.1016/j.jhydrol.2007.02.015.

Schmidt, J. A., A. J. Anderson, and J. H. Paul, 2007: Spatially variable, physically-derived flash flood guidance. *Preprints, 21st Conf. on Hydrology*, San Antonio, TX, Amer. Meteor. Soc., 6B.2.

Schwartz, C. S., and Coauthors, 2009: Next-day convection-allowing WRF model guidance: A second look at 2-km versus 4-km grid spacing. *Mon. Wea. Rev.*, **137**, 3351–3372, doi:10.1175/2009MWR2924.1.

—, and —, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*, 25, 263–280, doi:10.1175/2009WAF2222267.1.

Sukovich, E. M., F. M. Ralph, F. E. Barthold, D. W. Reynolds, and D. R. Novak, 2014: Extreme quantitative precipitation forecast performance at the Weather Prediction Center from 2001 to 2011. *Wea. Forecasting*, 29, 894–911, doi:10.1175/WAF-D-13-00061.1.

Sweeney, T. L., 1992: Modernized areal flash flood guidance. NOAA Tech. Rep. NWS HYDRO 44, Hydrologic Research Laboratory, National Weather Service, NOAA, 21 pp.



THE UNIVERSAL SOUNDING DECODE & ANALYSIS PROGRAM

RAOB raises the *milliBAR*

Plot any diagram up to the 0.1 mb (hPa) level.

RAOB...for the most versatile and accurate sounding analyses available!
Environmental Research Services, LLC • 570-491-4689 • www.raob.com

 A black and white photograph of a Young HD weather station. The station is a white, cylindrical instrument with a four-bladed anemometer cup on top and a vertical mast. It is positioned in the foreground against a backdrop of rugged, snow-capped mountains under a cloudy sky. The Young logo and 'HD' are visible on the instrument's body.

STRONG. SENSITIVE. SMART.

50 YEARS

YOUNG

SENSORS FOR: WIND, TEMPERATURE, HUMIDITY, PRECIPITATION, SOLAR RADIATION, ATMOSPHERIC PRESSURE

www.youngusa.com