The 2013 Flash Flood and Intense Rainfall Experiment

Final Report

September 16, 2013

1. INTRODUCTION

In collaboration with the National Severe Storms Laboratory (NSSL) and Earth System Research Laboratory (ESRL), the Hydrometeorological Testbed at the Weather Prediction Center (HMT-WPC) hosted the first annual Flash Flood and Intense Rainfall Experiment (FFaIR) from 8 – 26 July, 2013. FFaIR was developed in support of WPC's new MetWatch Desk, which was established in April 2013 and is responsible for issuing short-term (1-6 hr) event-driven Mesoscale Precipitation Discussions (MPDs) that highlight regions of heavy rainfall that may lead to flash flooding. The experiment brought together 26 forecasters, researchers, and model developers (Appendix A), including 8 participating remotely, to explore the challenges faced by MetWatch Desk forecasters related to short-term flash flood and quantitative precipitation forecasting (QPF) during the warm season. In particular, the goals of the experiment were to:

- Evaluate the utility of high resolution convection-allowing models and ensembles for short-term QPF and flash flood forecasts.
- Explore the use of high resolution rapidly updating hydrologic models for identifying areas vulnerable to flash flooding.
- Explore new tools and approaches to rapidly incorporate new data into the forecast process.
- Enhance collaboration among the operational forecasting, research, and academic communities on forecast challenges associated with warm season QPF and flash flood forecasting.

This report summarizes the activities, findings, and operational impacts of the experiment.

2. EXPERIMENT DESCRIPTION

Data

In addition to the full multi-center suite of operational deterministic and ensemble guidance available to WPC forecasters, the 2013 FFaIR experiment featured two experimental ensemble systems: the Storm-Scale Ensemble of Opportunity (SSEO; Jirak et al. 2012) provided by the Storm Prediction Center (SPC) and the Experimental Regional Ensemble Forecasting System (ExREF) provided by ESRL's Global Science Division (ESRL/GSD). In addition, a parallel version (NAMX) of NCEP's North American Model (NAM) was available, as well as the High Resolution Rapid Refresh (HRRR). Table 1 summarizes the model data that was the focus of the experiment.

The SSEO is a high-resolution, convection-allowing, multi-model, multi-physics ensemble system. Issued at 00 and 12 UTC, it is composed of seven deterministic high-resolution members[1] (Table 2). At WPC, the ensemble mean is displayed at 4 km, although each member

[1]While membership is consistent, the members of the SSEO used in FFaIR may be numbered differently in order to assure proper domain display when utilizing GEMPAK's ensemble probability functions.

Table 1. Featured 2013 FFaIR deterministic and ensemble model guidance. Experimental guidance is shaded.

Provider	Model	Resolution	Forecast Hours	Notes
EMC	SREF (21 members)	16 km 32 km (displayed)	87	Operational SREF
EMC	NAM	12 km (parent) 4 km (nest)	84 (parent) 60 (nest)	Operational NAM, includes 12 km parent model and 4 km CONUS nest
RFCs	Flash Flood Guidance (FFG)	5 km	01, 03, 06, 12 and 24 hour values	CONUS mosaic grid created by compiling individual RFC-domain grids
SPC	SSEO (7 members)	4 km	36	Multi-physics, convection allowing ensemble consisting of 7 high-resolution deterministic models
ESRL/GSD	ExREF (8 members)	9 km	84	Multi-physics, multi-initial condition, convection allowing ensemble
EMC	NAM Parallel (NAMX)	12 km (parent) 4 km (nest)	84 (parent) 60 (nest)	Features differing analysis (ENKF) and convective schemes from operational NAM
ESRL	HRRR	3 km	15	High-resolution, hourly updated, convection allowing nest of the radar-assimilating Rapid Refresh (RAP) model

can be viewed independently at its native resolution (Table 2). Two of the members (the operational ARW and NMM high-resolution windows) are time-lagged by 12 hours to provide additional initial condition diversity (Jirak et al. 2012). It should be noted that availability of all 7 members is not guaranteed, and they were not always present during FFaIR. The NSSL WRF-ARW and EMC WRF-NMM are non-operational and are subject to outages; the four high-resolution window members (HRW-ARW and HRW-NMM) are operational, but can be supplanted with other high-resolution runs (e.g. hurricane models) if the need arises (Jirak et al. 2012).

The ExREF is a multi-physics, multi-initial condition, multiple-boundary-condition ensemble system (Table 3). 7 of its 8 members feature use of the Local Analysis and Prediction System (LAPS; laps.noaa.gov) for their initial conditions, with the first member using the GFS analysis. The system (both the individual members and the mean) uses the Kain-Fritsch convective scheme and is run at 9 km resolution.

The SSEO and ExREF were used to create two sets of experimental probabilistic forecast tools: point and neighborhood probabilities. Point probabilities were derived by determining how many ensemble members predicted precipitation to exceed a relevant threshold at each individual grid point. Additionally, neighborhood probabilities (e.g., Schwartz et al. 2009, Schwartz et al. 2010, Ebert 2008) were generated for the two systems based on the

Table 2. Membership characteristics of the SSEO. Members denoted by the asterisk (*) are time lagged by 12 hours. Adapted from Jirak et al. (2012).

SSEO Member	Model	Provider	Resolution	PBL	Microphysics
01	WRF-ARW	NSSL	4 km	MYJ	WSM6
02	HRW-ARW	EMC	5.15 km	YSU	WSM6
03	HRW-ARW*	EMC	5.15 km	YSU	WSM6
04	HRW-NMM	EMC	4 km	MYJ	Ferrier
05	HRW-NMM*	EMC	4 km	MYJ	Ferrier
06	WRF-NMM	EMC	4 km	MYJ	Ferrier
07	NAM-NMMB Nest	EMC	4 km	MYJ	Ferrier

Table 3. Membership characteristics of the ExREF. Member denoted by asterisk (*) denotes use of the 'variational' version of the LAPS analysis; all others use the 'traditional' version.

Member	Initial Conditions	Boundary Conditions	Microphysics
m00	GFS	GFS	Thompson
m01	LAPS	GFS	Thompson
m02	LAPS	GEFS 01	Ferrier
m03	LAPS	GEFS 02	WSM6
m04	LAPS	GEFS 03	Thompson
m05	LAPS	GEFS 04	Ferrier
m06	LAPS	GEFS 05	WSM6
m07	LAPS*	GFS	Thompson

'neighborhood maximum value'. This technique accounts for spatial uncertainty in high-resolution model forecasts by conducting a search within a certain radius (e.g. 40 km) of each grid point to locate the maximum value of a parameter (e.g. precipitation) within that radius. The *value of the original grid point is then replaced with this maximum value,* and probabilities of exceedance are calculated.

The point and neighborhood probabilities were created for two threshold concepts:

- QPF exceeding a certain amount (e.g. 1 inch; QPF>1")
- QPF exceeding flash flood guidance (QPF>FFG)

The QPF probabilities were created for 3, 6 and 12 hour time periods, while the QPF>FFG probabilities were created at 3 and 6 hour time periods using the corresponding flash flood guidance values. Flash flood guidance is produced by the individual NWS River Forecast Centers (RFCs, Fig. 1), and WPC compiles the guidance to create a 5 km CONUS mosaic FFG grid. Since RFCs can update FFG at their discretion, WPC checks hourly for any new guidance and recompiles the mosaic. There are several methods currently employed to create FFG; therefore, the method of producing FFG is inconsistent across RFCs.

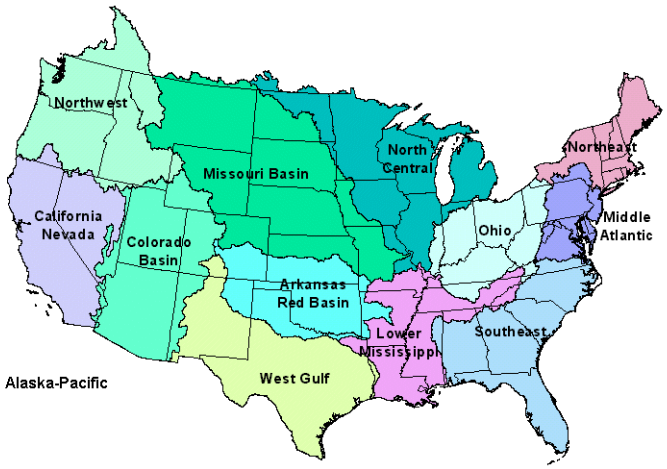

Figure 1. NWS River Forecast Centers. Image courtesy of NOAA/NWS (water.weather.gov).

In addition to the SSEO and ExREF, several experimental deterministic models were also featured in FFaIR. The NAMX, provided by EMC, is the parallel version of the NAM and features both a 12 km parent and a 4 km CONUS nest. This alternate version features uses the global Ensemble Kalman Filter (EnKF) members as part of its regional GSI data assimilation system, and employs the Rapid Radiative Transfer Model (RRTM) radiation scheme (as opposed to the GFDL scheme in the operational NAM). Additionally, the parent 12 km features the Betts-Miller-Janic (BMJ) convective scheme, and its 4 km nest is fully convection allowing; this differs from the operational NAM in which the parent 12 km uses the base BMJ parameterization and the nest uses a modified version of the BMJ scheme to initially trigger convection.

The High Resolution Rapid Refresh (HRRR, http://ruc.noaa.gov/hrrr) was also featured during the experiment. The HRRR is the 3 km nest of the hourly updated, radar-assimilated Rapid Refresh (RAP) model. It features a WRF-ARW core, Thompson microphysics, and is fully convection allowing. The HRRR is initialized with the latest 3-D radar reflectivity using radar-DFI (digital filter initialization) technique (via the parent 13 km RAP) and provides output hourly. WPC currently ingests the HRRR in 3-hour intervals (00, 03, 06, 09, 12, 15, 18, and 21 UTC), but hourly HRRR data was available to participants through the HRRR website.

Lastly, in addition to the high-resolution atmospheric models, the experiment also featured Flooded Locations and Simulated Hydrographs (FLASH, flash.ou.edu), a high resolution (1 km) rapidly updating (5 min) distributed hydrologic model developed by NSSL. FLASH uses radar-derived quantitative precipitation estimates (QPE) from the Multi-Radar Multi-Sensor (MRMS, http://nmq.ou.edu/) system to simulate surface water flows six hours into the future. These forecasts are in the form of return periods, which were developed based on a 10-year retrospective model simulation using Stage IV precipitation estimates (Lin and Mitchell 2005). During the experiment, data limitations in the western United States resulted in areas of constant high return periods that were not representative of the observed rainfall in that region.

In addition to the output from the hydrologic model itself, the FLASH website also featured other potential flash flood indicators including flash flood warnings (FFWs), local storm reports (LSRs), mPING reports (http://www.nssl.noaa.gov/projects/ping/), QPE recurrence intervals, QPE to FFG ratios, observed precipitable water (PW) values, PW anomalies, and satellite-derived inundation maps from the Moderate Resolution Imaging Spectroradiometer (MODIS). QPE recurrence intervals were calculated by comparing the real time MRMS radar-estimated precipitation to climatological precipitation frequency estimates from NOAA Atlas 14 (Vols. 1-9; available at http://hdsc.nws.noaa.gov/hdsc/pfds/index.html). Like the output from the hydrologic model, the recurrence intervals were provided in the form of return periods, indicating how frequently the observed precipitation amount would be expected to occur at a given location. QPE-to-FFG ratios were calculated by comparing the radar-estimated precipitation to the most recent FFG issued by the RFCs. The MODIS inundation maps provided a satellite interpretation of areas of dry land, known surface water (lakes, rivers, etc.), clouds, and flooded land (Smith 1997).

Daily Activities

Each week, participants were paired with a different WPC MetWatch forecaster to form a collaborative forecast team. Each morning the team chose a multi-state forecast 'area of interest' where they anticipated a threat of heavy rainfall that might lead to flash flooding during the 12 – 00 UTC time period. The team used a combination of operational and experimental guidance to create four experimental probabilistic QPF and flash flood forecasts valid at various times during the following 24 hour period (12 – 12 UTC). These forecasts simulated the timeframe, workload and thought processes associated with creating WPC's MPD and Day 1 Excessive Rainfall products.

12-hour (12 – 00 UTC) probability of QPF (PQPF) of greater than 1", due at 15 UTC. Participants were asked to draw contours of 10%, 30% and 50% probability of exceedance of 1", when applicable, over their chosen area of interest. This forecast gave participants the opportunity to evaluate the heavy rainfall threat through an initial investigation of the available observational and numerical model guidance (Fig. 2a).

6-hour (18 – 00 UTC) probability of flash flooding, due at 1730 UTC. Participants were instructed to draw contours of a 10%, 30% and 50% probability of flash flooding, when applicable, over their chosen area of interest (Fig. 2b). This forecast required the forecast team to consider both hydrologic and meteorological information to assess the flash flood threat to issue a forecast for the likelihood of flash flooding.

Update to the 6-hour (18 – 00 UTC) probability of flash flooding, due at 19 UTC. Participants were asked to update their contours of a 10%, 30% and 50% probability of flash flooding, when applicable, over their chosen area of interest (Fig. 2c) by incorporating updated model guidance and real-time observations (e.g. radar, satellite).

12-hour (00 – 12 UTC) probability of flash flooding, due at 20 UTC. Participants were asked to draw contours of a 10%, 30% and 50% probability of flash flooding, when applicable, over the entire CONUS area for the overnight period (Fig. 2d).

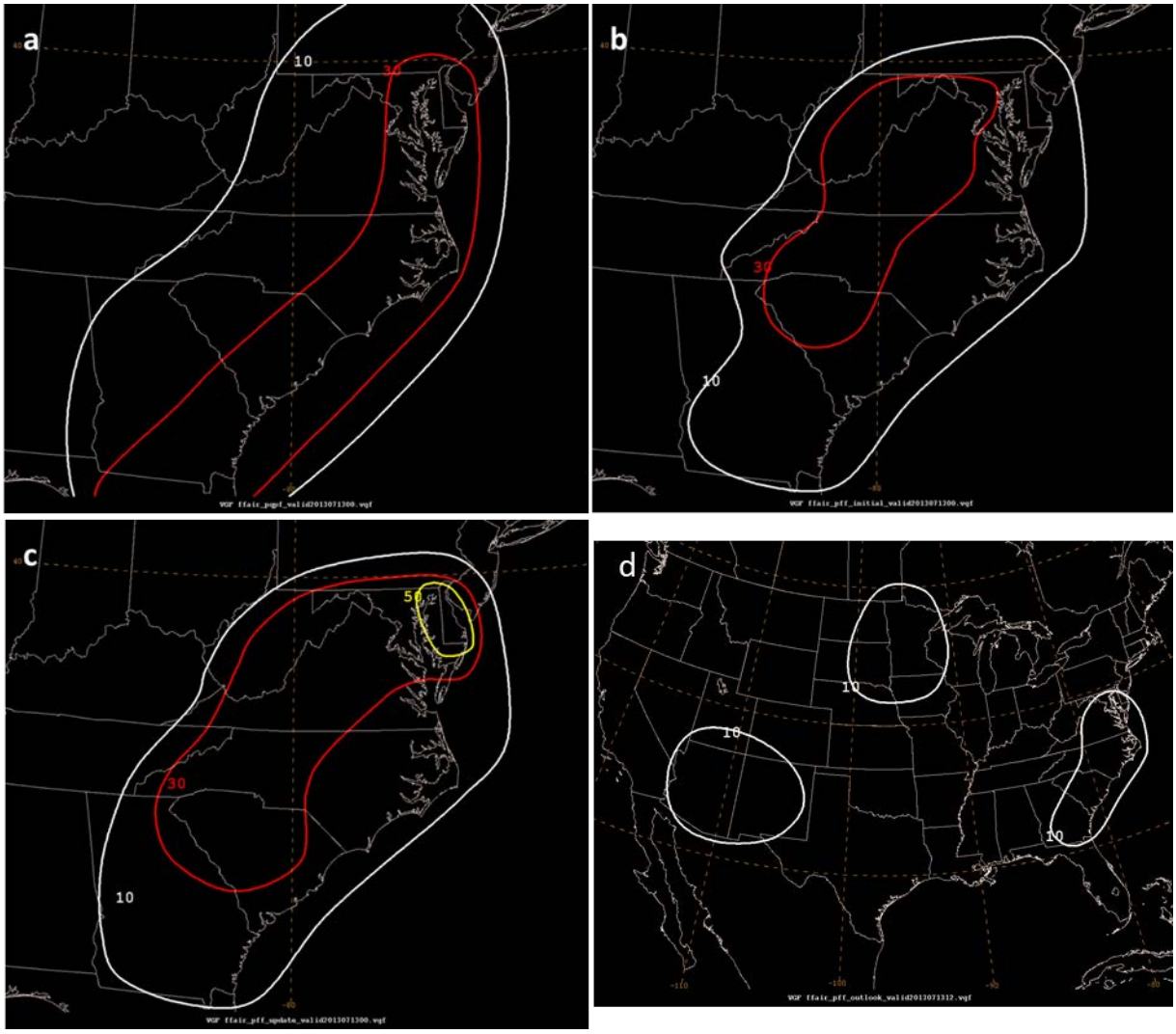

Figure 2. Displaying the (a) 12-hour probability of QPF>1" forecast, (b) the 6-hour probability of flash flooding forecast, (c) the updated 6-hour probability of flash flooding and (d) overnight, 12-hour probability of flash flooding for the exercises from July 12, 2013.

Additionally, the team completed various exercises throughout the day to subjectively evaluate the performance of the experimental model guidance, forecast tools and flash flood diagnostics. The subjective model evaluations compared the relative performance of the experimental guidance to the operational NAM nest (deterministic models) and the operational Short Range Ensemble Forecast System (SREF, ensemble systems). Evaluation of the numerical model guidance and experimental forecasts was conducted using a combination of radar-estimated QPE from the MRMS system as well as FFWs, QPE recurrence intervals, QPE-to-FFG ratios, LSRs and mPING reports.

3. CASES

While the summer of 2013 was characterized by a mean trough in the eastern United Sates, a ridge over the Rockies, and trough just off the west coast (not shown), the experiment period was characterized by a moderation in the trough over the eastern U.S. and a more pronounced ridge over the much of the western and central part of the country (Fig. 3a). Within this regime, the experiment period was characterized by anomalously wet conditions from the Mid-Atlantic and Ohio Valley into the southern plains (Fig. 3b). Although precipitation rate anomalies were near normal across much of the southwest, precipitable water anomalies show a pronounced maximum over the desert southwest (Fig. 3c) during this period.

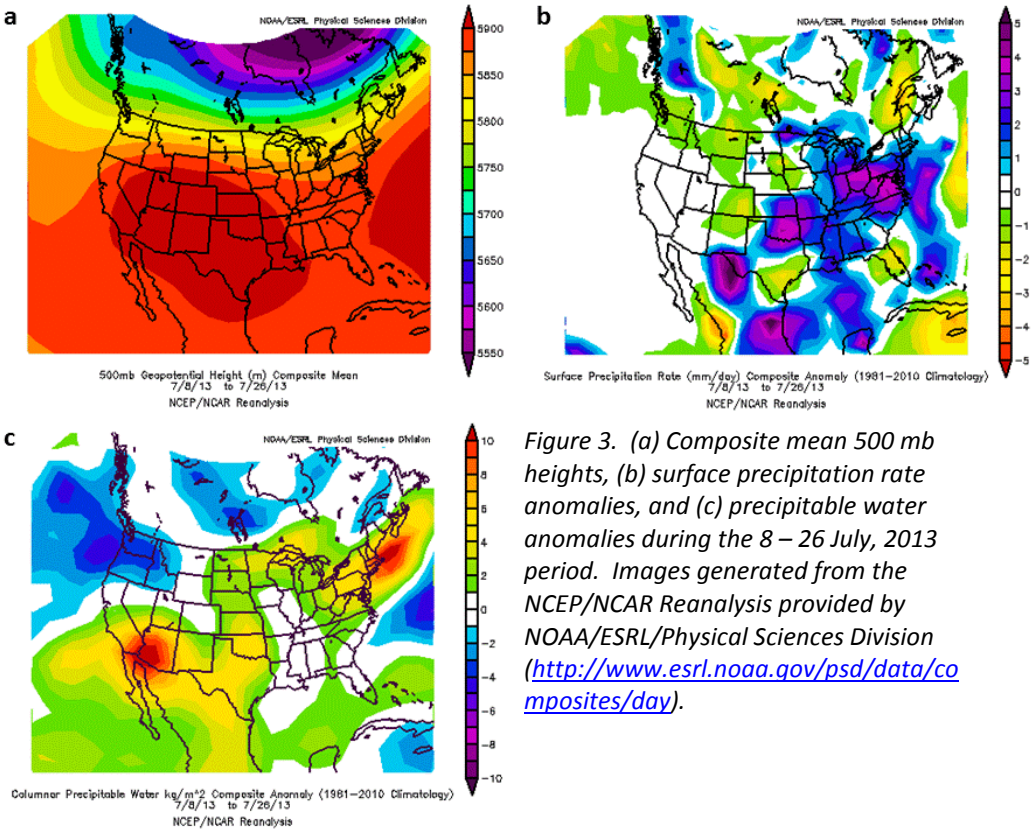

Figure 3. (a) Composite mean 500 mb heights, (b) surface precipitation rate anomalies, and (c) precipitable water anomalies during the 8 – 26 July, 2013 period. Images generated from the NCEP/NCAR Reanalysis provided by NOAA/ESRL/Physical Sciences Division (http://www.esrl.noaa.gov/psd/data/composites/day).

The anomalous moisture across a good portion of the country during the experiment provided an opportunity to investigate the flash flood threat in a variety of different events. The first week of the experiment was highlighted by widespread flash flooding after several days of rain across Ohio and western Pennsylvania. In the second week, the threat shifted to the desert southwest where a retrograding low and monsoonal moisture combined to create several active days with widespread thunderstorms in the terrain. The last week of the experiment featured a series of mesoscale convective systems (MCSs) across the plains. Finally, a significant flash flood event occurred in western North Carolina at the end of the experiment that resulted in two fatalities and damage to numerous homes. A complete list of the events that were investigated during this year's experiment can be found in Table 4.

Table 4. Experimental forecasts issued during the 2013 Flash Flood and Intense Rainfall Experiment. In addition to the valid times and forecast areas listed here, corresponding CONUS-scale outlook forecasts valid at 12 UTC the same day were also issued.

Forecast Valid Time	Forecast Area	Notes
00 UTC 9 July 2013	Northern Plains	
00 UTC 10 July 2013	Ohio Valley to Lower Great Lakes	
00 UTC 11 July 2013	Ohio Valley to Lower Great Lakes	Widespread flash flooding in OH and western PA
00 UTC 12 July 2013	Lower Mississippi Valley to Northeast	
00 UTC 13 July 2013	Southeast to Mid Atlantic	
00 UTC 16 July 2013	Southwest to Southern Rockies	
00 UTC 17 July 2013	Intermountain West to Southwest	
00 UTC 18 July 2013	Southern Plains	
00 UTC 19 July 2013	Upper Mississippi Valley to Ohio and Tennessee Valley	
00 UTC 20 July 2013	Southwest to Central Rockies	
00 UTC 23 July 2013	Lower Great Lakes to Mid Atlantic and Ohio and Tennessee Valleys	Numerous flash flood reports across NV and southern CA during outlook period
00 UTC 24 July 2013	Central Plains to Southeast	
00 UTC 25 July 2013	Southwest and Southern Plains	
00 UTC 26 July 2013	Upper Mississippi Valley to Southern Plains	
00 UTC 27 July 2013	Middle Mississippi Valley to Southern Plains	Significant flash flood event in western NC at end of outlook period (~12 UTC 27 July)

4. DETERMINISTIC HIGH RESOLUTION MODEL PERFORMANCE

As part of the subjective evaluation process, participants were asked to rate the high-resolution deterministic QPF guidance as *much better, better, about the same, worse*, or *much worse* than the operational 4 km NAM nest based on the observed precipitation during the 12 – 00 UTC period. The results below are based primarily on these subjective responses. Although objective verification was considered, ultimately no objective verification statistics were calculated because of the small sample size (15 cases).

Overall, all of the deterministic high-resolution models provided useful guidance for short-term precipitation forecasts. In general, the high-resolution models were able to identify the location of the heaviest rainfall, and unlike previous evaluations conducted during the QPF component of the Hazardous Weather Testbed's (HWT) Spring Experiment, the pronounced high bias in precipitation amounts often associated with these models was not observed during this experiment. In particular, the HRRR consistently provided better forecast guidance than the operational NAM nest and was the best performing deterministic model for 12-hr QPF (Fig. 4). Compared to the operational NAM nest, the HRRR was generally better able to locate the areas of heaviest rainfall across a variety of meteorological phenomena from MCSs to scattered convection in the desert southwest (Fig. 5). Participants liked that the HRRR included radar data assimilation and often used the HRRR website to access more recent model runs not

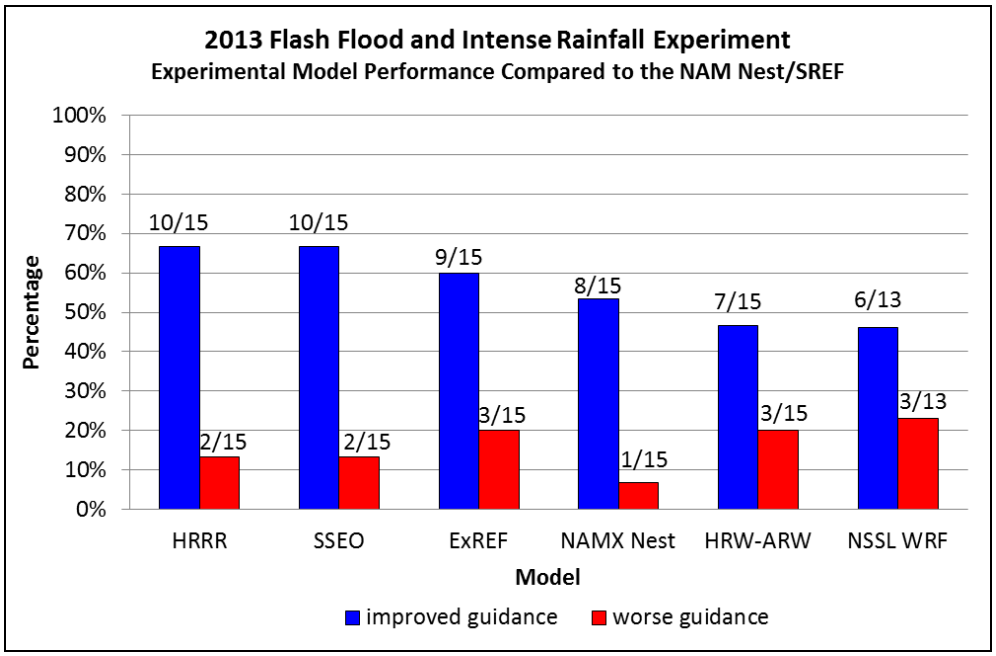

Figure 4. Experimental model performance based on participant feedback from subjective model evaluations conducted during the 2013 Flash Flood and Intense Rainfall Experiment. Participants were asked to determine whether the precipitation forecasts from the 00 UTC experimental guidance (09 UTC HRRR) were much better, better, about the same, worse, or much worse than the corresponding operational guidance from the 00 UTC NAM nest.

available in NAWIPS. Data from the HRRR often played a significant role in the forecast process, particularly in situations with otherwise limited model agreement.

Similarly, the NAMX nest, NSSL WRF, and HRW-ARW also provided better forecast guidance than the operational NAM nest. In comparison to the NAMX nest and the HRRR, however, the NSSL WRF and the HRW-ARW were more likely to have larger errors in the location and coverage of the heaviest precipitation. For example, Figure 6 shows an event from the desert southwest. In this case, the NAMX nest provided much better forecast guidance than the operational NAM nest, highlighting the threat for heavy rainfall over much of central New Mexico, with scattered heavier amounts extending into Arizona, Utah, and Colorado. Like the operational NAM nest, both the NSSL WRF and the HRW-ARW incorrectly focus the heaviest precipitation along the Arizona-New Mexico border. Both models also predicted convection in western Texas where little to no precipitation was observed.

It is important to note that the HRRR forecasts included in the subjective model evaluations are from a more recent model run (09 UTC) than the other deterministic model guidance (00 UTC). The 09 UTC HRRR is both the first run of the HRRR that covers the entire 12 – 00 UTC period of the experimental PQPFs and the most recent run available to participants in WPC's NAWIPS system at the time the experimental forecasts were started (~1330 UTC). While the difference in initialization time may have given the HRRR an advantage in the evaluation process, it also demonstrates the value of a rapidly updating model that includes radar assimilation compared to those run on the standard synoptic schedule.

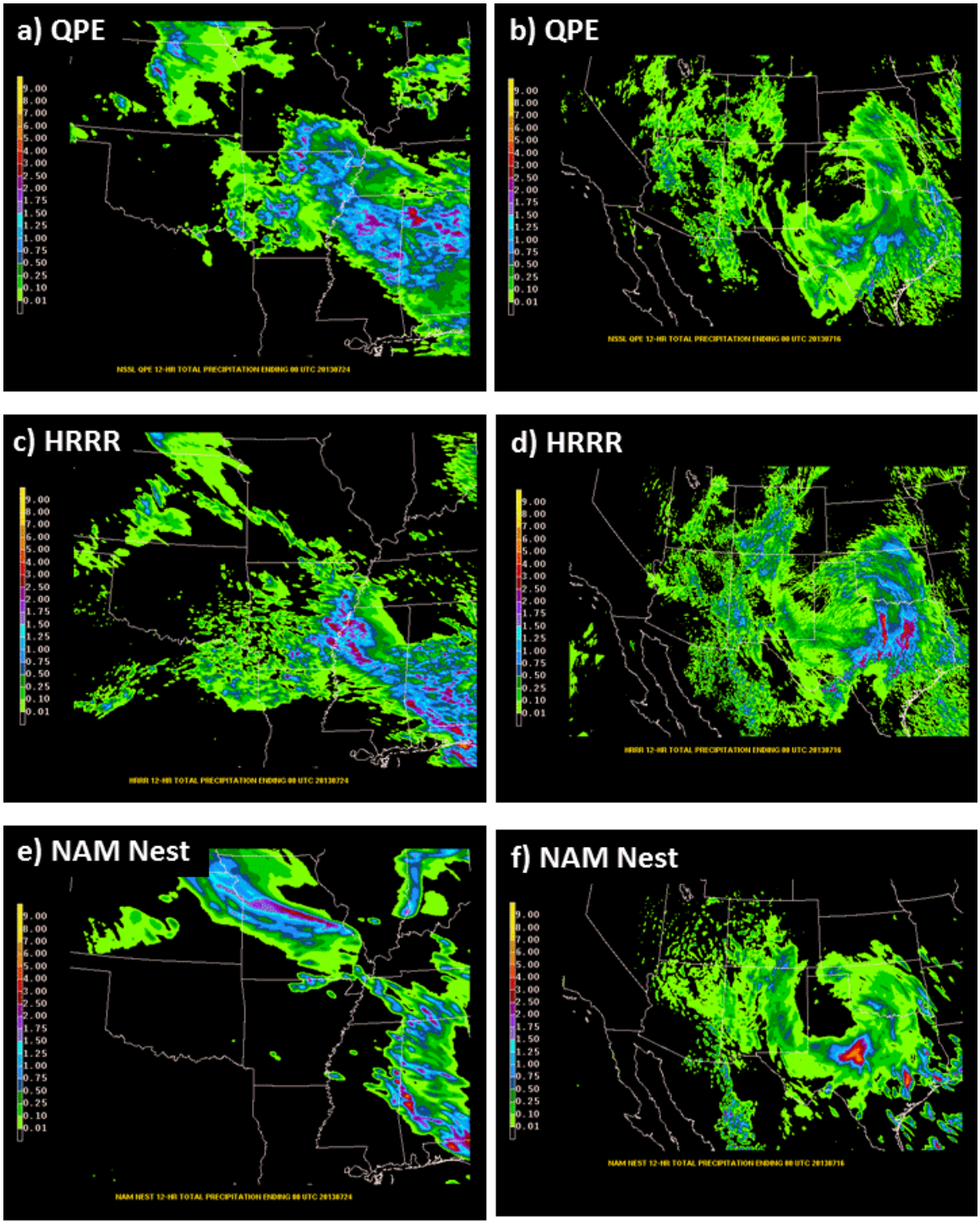

Figure 5. Observed 12-hr precipitation ending (a) 00 UTC 24 July 2013 and (b) 00 UTC 16 July 2013 from MRMS QPE and the corresponding (c,d) 15-hr forecast from the 09 UTC HRRR and (e,f) 24-hr forecast from the 00 UTC operational NAM nest.

Finally, although model forecasts during the 00 – 12 UTC outlook period were not specifically evaluated, there were several cases in which the high-resolution deterministic models (and ensembles) contained signals for heavy precipitation and flash flood events that occurred

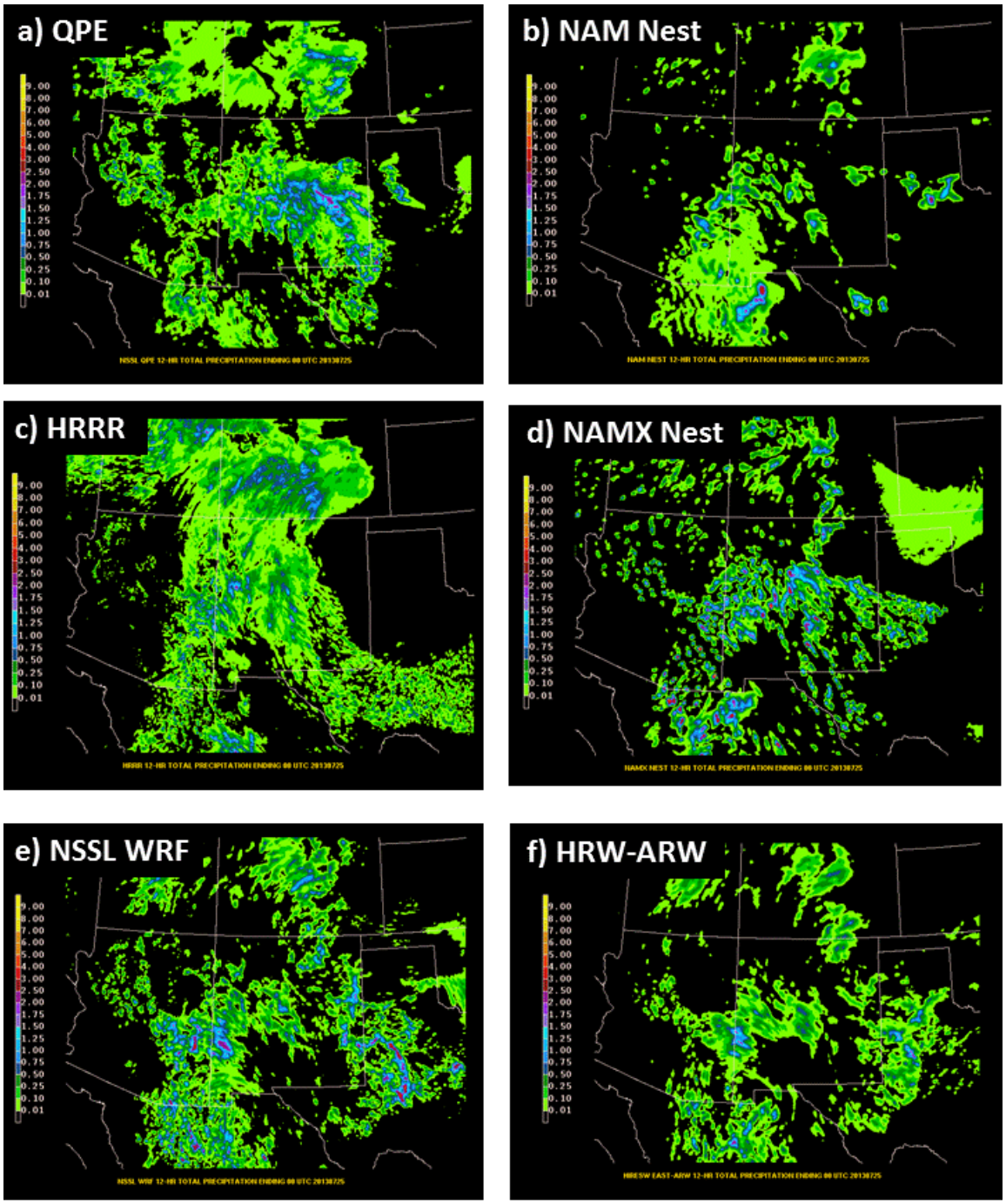

Figure 6. *(a) Observed-12 hr precipitation ending 00 UTC 25 July 2013 from MRMS QPE and the corresponding 24-hr forecasts from the (b) operational NAM Nest, (c) HRRR (15-hr forecast), (d) NAMX Nest, (e) NSSL WRF, and (f) HRW-ARW.*

during the overnight period. For example, the majority of the 12 UTC 11 July 2013 high-resolution runs showed the potential for extremely heavy rainfall in central South Carolina between 00 – 12 UTC 12 July (Fig. 7). MRMS QPE data indicates widespread observed precipitation amounts of 2 in, with isolated 5 in maxima, and the FLASH website reveals numerous reports of flash flooding over the same area (Fig. 7g). Combined with the subjective

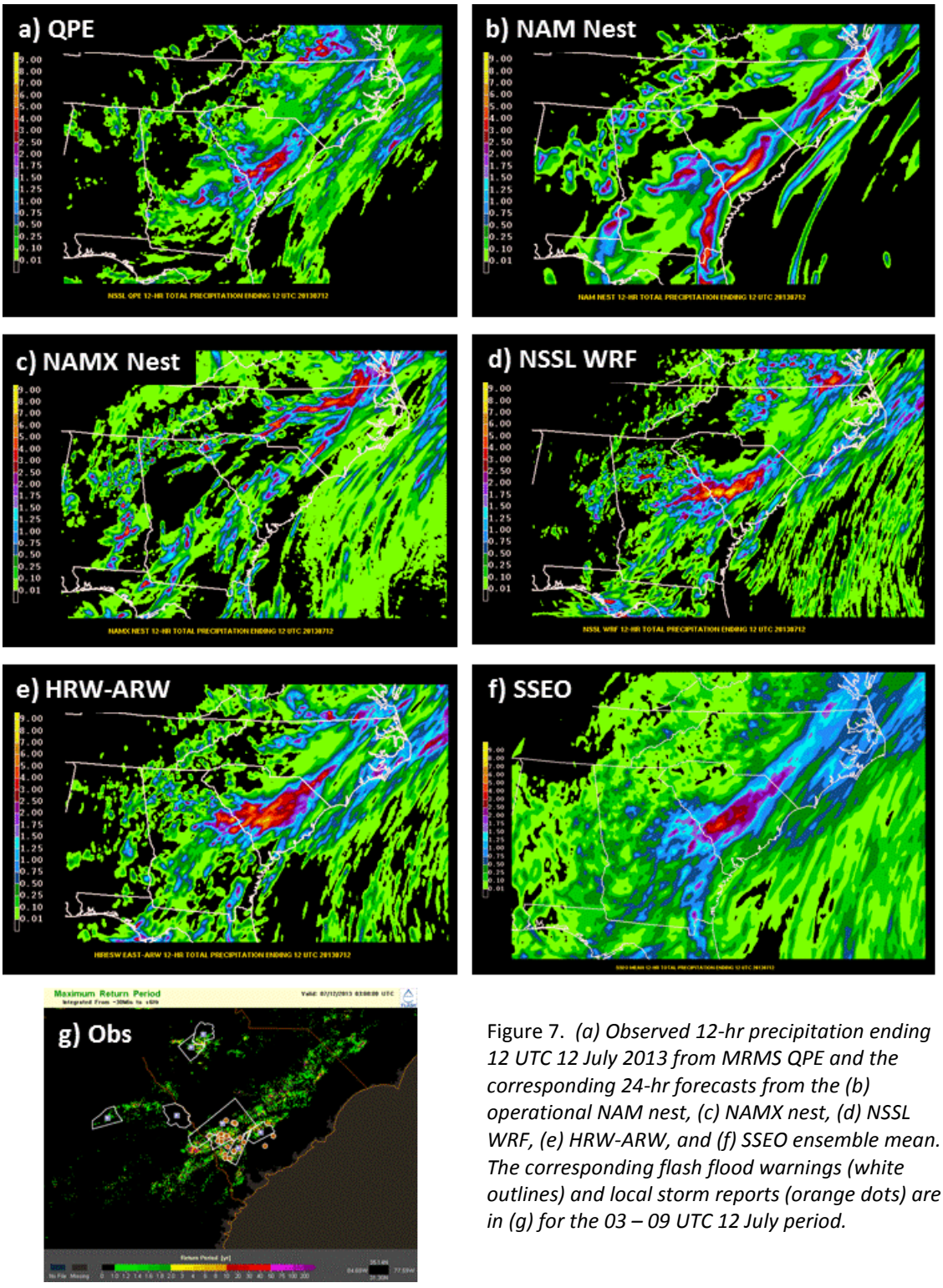

Figure 7. *(a) Observed 12-hr precipitation ending 12 UTC 12 July 2013 from MRMS QPE and the corresponding 24-hr forecasts from the (b) operational NAM nest, (c) NAMX nest, (d) NSSL WRF, (e) HRW-ARW, and (f) SSEO ensemble mean. The corresponding flash flood warnings (white outlines) and local storm reports (orange dots) are in (g) for the 03 – 09 UTC 12 July period.*

evaluation results for the 12 – 00 UTC period, cases like this one demonstrate that high resolution guidance can be a valuable tool for identifying regions of heavy rainfall that may lead to flash flooding in the next 12 – 24 hours. Moreover, although high resolution guidance is not perfect, model agreement can significantly increase confidence in the occurrence of an event.

5. EXPERIMENTAL ENSEMBLE PERFORMANCE

In addition to the deterministic high-resolution numerical model guidance, the ensemble guidance and corresponding experimental probabilistic forecast tools were also subjectively evaluated. When assessing their performance during the 12 – 00 UTC forecast period, the SSEO and ExREF ensemble mean QPF were subjectively rated as *much worse*, *worse*, *about the same*, *better*, or *much better* than the operational SREF mean, using the MRMS precipitation observations. As seen in Figure 4, both the SSEO and ExREF provided forecasts of additional value when compared to the SREF, as 67% of the SSEO mean and 60% of the ExREF mean forecasts were rated as either better or much better than the SREF mean.

Knowing that these were ensemble mean forecasts, forecasters did not try to compare the guidance directly to observations, and evaluated the guidance based on the overall trends and signals for the potential for heavy rain. One of the main benefits that the participants found was the increased resolution of the ExREF (9 km) and SSEO (4 km) compared to the SREF (16 km, but displayed on 32 km). An example of this can be seen in Figure 8, which shows the SSEO, SREF, ExREF and MRMS QPE observations for the 12-hour QPF forecast valid at 00 UTC July 12. In this particular case, participants liked the additional detail and heavier rain amounts shown by the SSEO (Fig. 8b) and ExREF (Fig. 8d) in Maryland and Delaware, as well as along the

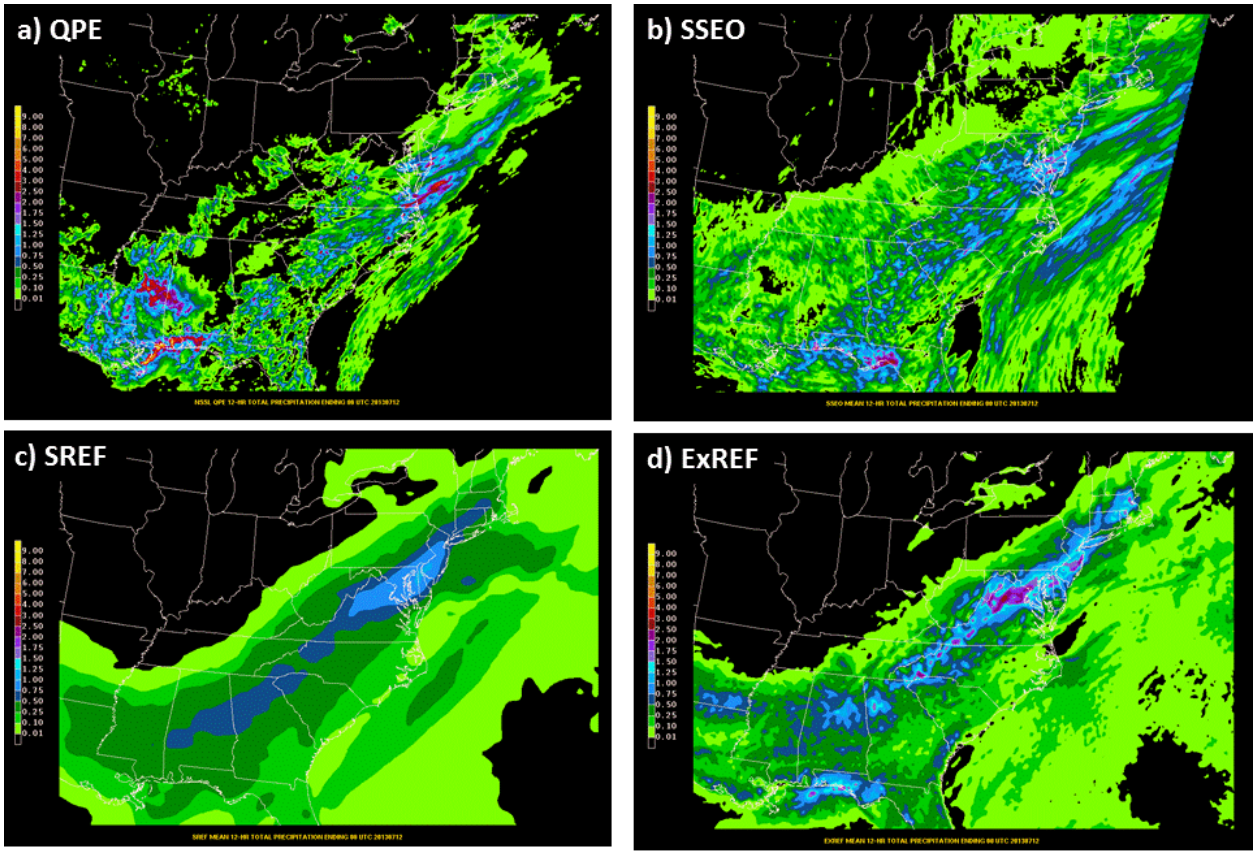

Figure 8. (*a) Observed 12-hr precipitation ending 00 UTC 12 July 2013 from MRMS QPE and the corresponding 12-hr QPF from the (b) SSEO mean, (c) SREF mean, and (d) ExREF mean valid at 00 UTC 12 July.*

Florida and Alabama coasts. While the location and amounts of the ExREF and SSEO may not have collocated with the observations, participants liked that there was a more defined signal for the potential of heavy rain in those general areas than the SREF (Fig. 8c), which, due to its coarser display resolution, showed a more general solution with lower QPF amounts.

With regards to ensemble performance of the point probability of QPF >1" over the same 12 – 00 UTC period, participants were asked if both the ExREF and SSEO *captured*, *nearly captured*, or *did not capture* the entire area that was observed to receive >1" of rain within the base 1% probability contour. Neither model was able to completely capture the >1" areas in any instance during the experiment, but the SSEO nearly captured all the >1" areas 40% of the time, while the ExREF nearly captured all the areas 20% of the time (not shown). An example for the July 11 case is shown in Figure 9a,d, where the model point probabilities of >1% (shaded) are overlaid with all areas that received >1" of precipitation (white contours with dashed filling).

Participants were also asked to evaluate the 20 km and 40 km neighborhood probabilities of QPF>1" (Fig. 9) and QPF>FFG (for the 18 – 00 UTC period, Fig. 10) from both ensemble systems. Figure 11 shows that neighborhood probabilities were deemed to give more useful guidance than the point probabilities, but the most effective radius differed depending on the probability tool and ensemble used. For the SSEO, the 20 km neighborhood probabilities provided the best guidance in a majority (80%) of the cases for the QPF>1" probabilities, but the 40 km neighborhood provided more effective guidance for the QPF>FFG probabilities (60% of cases). Like the SSEO, the 40 km neighborhood probabilities from the ExREF provided the best guidance for the QPF>FFG probabilities (80% of events), but there was a split between the 20 km and 40 km neighborhoods for QPF>1" (each 40% of events).

The subjective evaluation also provided additional information regarding the use and efficiency of each ensemble and probability tool. When asked to directly compare the 40 km neighborhood probabilities of QPF>FFG for the ExREF and SSEO, the SSEO provided more useful guidance in the majority (80%) of cases. The reasons for this were threefold: 1) the lack of dispersion in the ExREF in the 0-24 hour forecast period often failed to provide the proper spread needed for it to capture the observed range of flash flood reports, 2) the 9 km resolution of the ExREF limited its ability to resolve small-scale convection, making it more difficult to generate QPFs that exceed flash flood guidance, and 3) the ExREF appeared to have a tendency to miss/under-develop convection on the southern edge of systems.

An example of this discrepancy can be seen in Figure 10c,f, where the SSEO places probabilities of QPF>FFG throughout Georgia and South Carolina, where flash flooding was observed, but the ExREF does not identify a threat. Meteorologically, the ExREF seemed to be unable to adequately resolve the late afternoon and evening surface-instability-based convection that often develops in the southeast and failed to identify any flash flood threat as a result. In addition, the ExREF's coarser resolution likely limited its ability to produce the small-scale convection often seen in these cases, and the short forecast lead times that were the focus of the experiment correspond with the ExREF forecasts that lack sufficient dispersion. It is

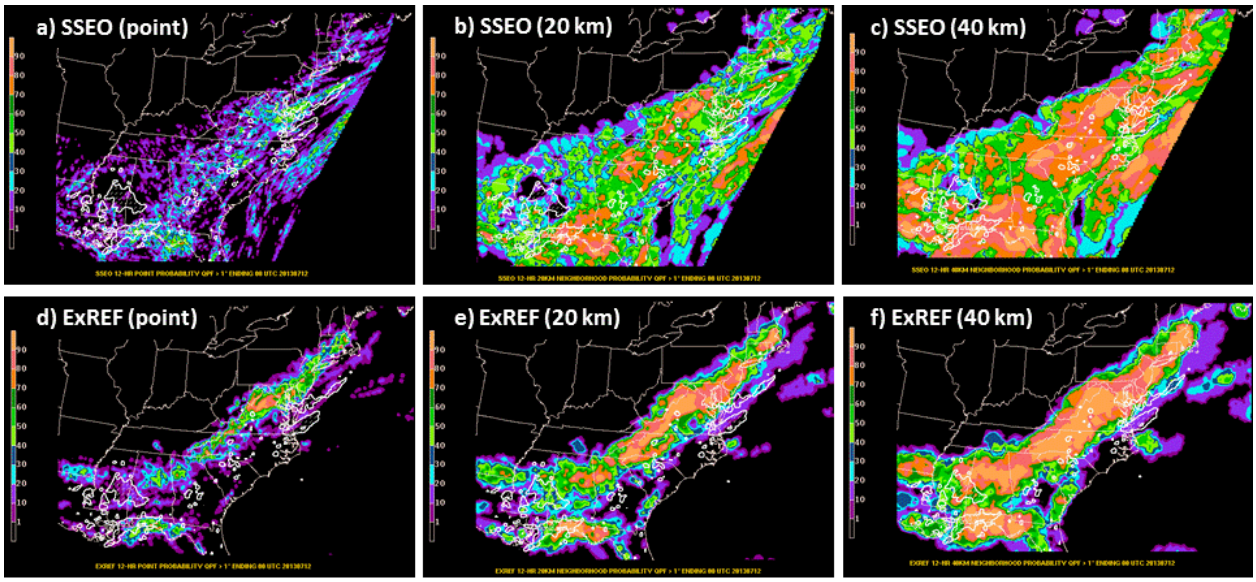

Figure 9. Showing the (a,d) point, (b,e) 20 km and (c,f) 40 km probabilities for 12-hr QPF>1" for the SSEO (top row) and the ExREF (bottom row) valid at 00 UTC 12 July. Probabilities are shaded, and areas >1" are designated by white hatched contoured areas.

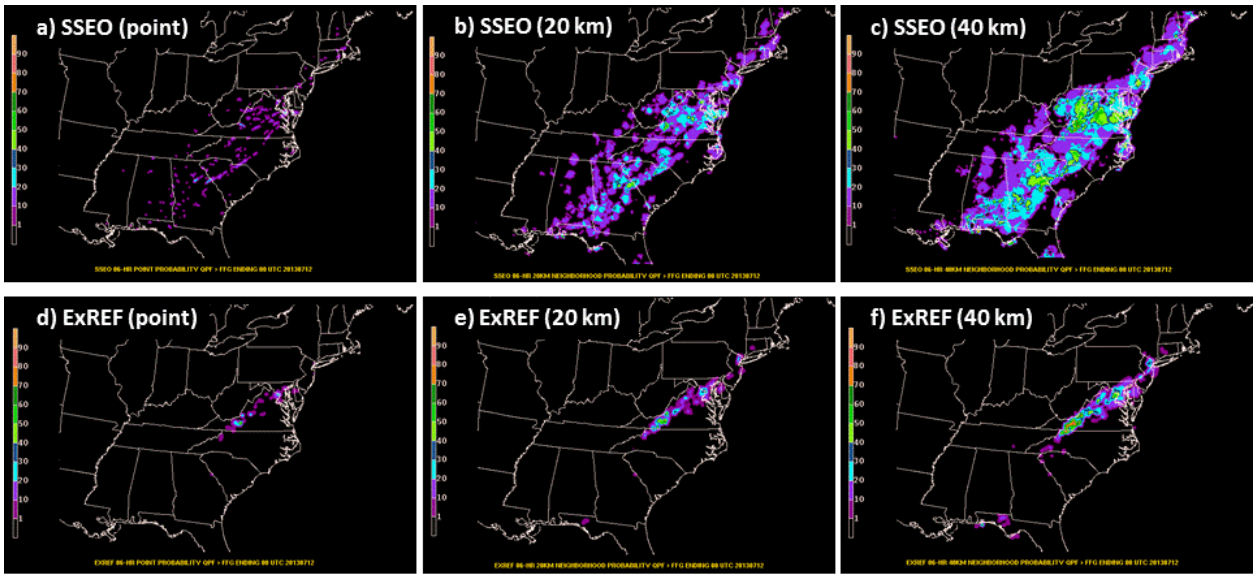

Figure 10. Showing the (a,d) point, (b,e) 20 km and (c,f) 40 km probabilities for 6-hr QPF>FFG for the SSEO (top row) and the ExREF (bottom row) valid at 00 UTC 12 July. Probabilities are shaded.

important to note that evaluating the ExREF at longer forecast lead times may have mitigated any impacts related to ensemble dispersion.

Given these results, it is reasonable to conclude that the preferred probabilistic guidance tool varies depending on the ensemble. As mentioned above (and seen in Fig. 11) the probability of QPF>FFG (40 km) product provided better guidance from the SSEO, while the probability of QPF>1" (40 km) product provided better guidance from the ExREF. The ExREF's coarser resolution and corresponding difficulty developing enough precipitation from small-scale

convection limited the applicability of its QPF>FFG probabilities, while at the same time the short-term focus of the experimental forecasts likely magnified the impact of the initial under-dispersion in the ensemble members. Correspondingly, the known high-bias associated with some of the high-resolution members (ex: EMC WRF-NMM) of the SSEO often created a large area of high probabilities that was unrealistic (Fig. 9c), making its QPF>1" product relatively less useful. It was also noted that throughout the experiment participants felt that the QPF>FFG product had less value in the west, where topography, rain rates, precipitable water and 1 or 3-hour QPF correlated better with flash flooding.

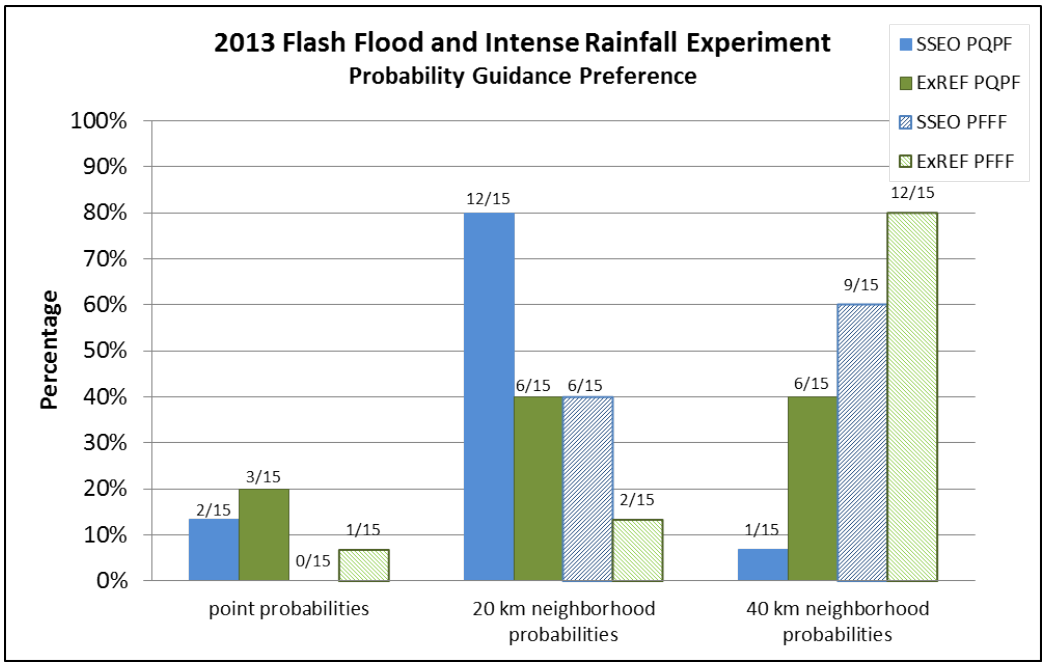

Figure 11. Experimental probability guidance performance based on participant feedback from subjective model evaluations. Participants were asked to determine whether the point, 20, or 40 km probabilities provided better guidance from the 00 UTC SSEO and ExREF for both the QPF>1" (PQPF) and QPF>FFG (PFFF) tools.

6. FLASH FLOOD DIAGNOSTICS

A recurring challenge throughout the experiment was determining when and where flash flooding had occurred since there is no single observational dataset that accurately depicts all flash flood events (Gourley et al. 2013). For example, while FFWs can be a useful diagnostic for identifying areas where flash flooding is either imminent or occurring, warnings are not issued for every event, and an event does not occur every time a warning is issued. In addition, philosophies can vary between forecast offices as to whether issuing a FFW or an urban and small stream advisory is more appropriate. On the other hand, while LSRs provide ground truth, they are population-dependent; i.e., a flash flood can't be reported if there is no one there to observe it. Even the definition of flash flooding itself can be problematic. While two inches of water running over a road in a poor drainage area isn't technically a flash flood, it is likely to be reported as one.

Clearly, accurate identification and reporting of flash flood events is a complex challenge. To start to address this issue, the experiment featured a number of different tools designed to indicate when and where flash flooding is occurring. Participants were asked to rank the utility of four different observational datasets (FFWs, LSRs, mPING reports, and MODIS inundation maps) and three diagnostic/forecast tools (FLASH return periods, QPE recurrence interval, and QPE-to-FFG ratio) based on their ability to provide useful information about the extent, location, and impact of flash flooding.

Of the observational datasets evaluated, LSRs were consistently considered to be the most useful dataset for identifying flash flood events (Fig. 12). Participants considered LSRs to be particularly trustworthy since they are vetted by the local National Weather Service (NWS) offices before being released. FFWs were also considered to be a useful dataset for identifying flash flood events, although their utility varied across the country. Warnings were often considered more useful in the western United States where lower population density may limit flash flood reports, but over the eastern two thirds of the country there was concern about both over-warning and missed events resulting from "warn-on-report" practices. Compared to both LSRs and FFWs, participants considered mPING reports to be considerably less useful because of concerns about the limited number of reports and the quality of reports obtained directly from the public. Finally, participants found that the MODIS inundation maps were not useful for identifying areas of flash flooding since the area of interest was often still cloud-covered at the time of the satellite pass. Given the short time frames associated with flash flooding and the limitation of having only one satellite pass over a point each day, the consensus was that this dataset would likely be more useful for assessing longer term flood events.

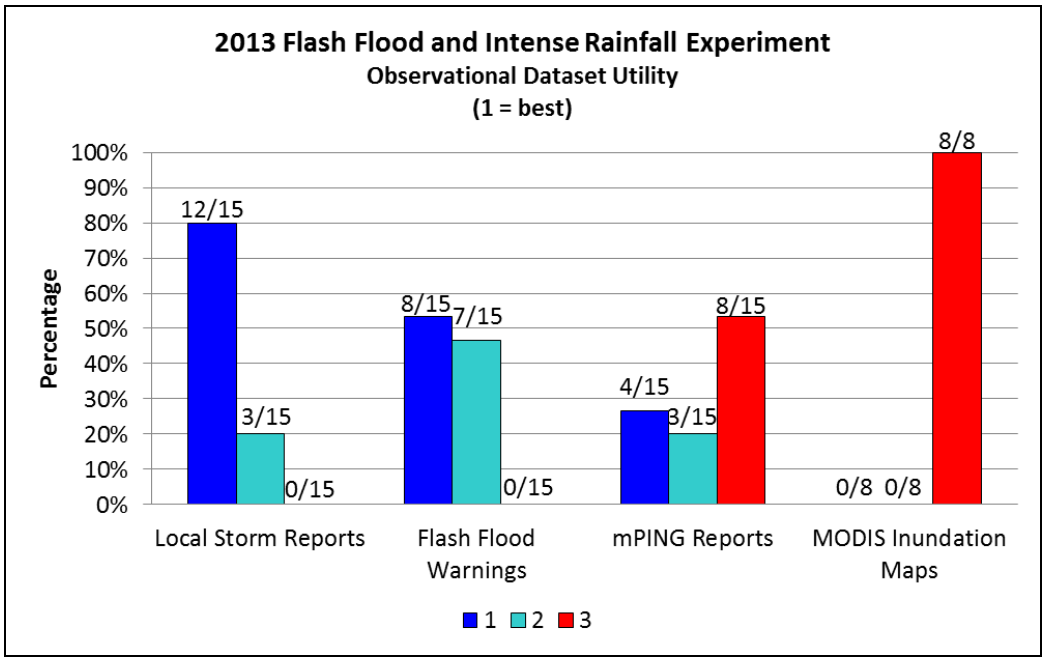

Figure 12. Relative utility of observational datasets for providing useful information about the extent, location, and impact of flash flooding. Datasets were ranked from 1 to 3 with 1 indicating the best dataset. Two datasets could be assigned the same ranking if they provided similar value.

Similarly, participants ranked the three flash flood diagnostic/forecast tools based on their ability to identify both the broad regions that were impacted by flash flooding (loosely defined as the correct County Warning Area) as well as the specific locations (cities, neighborhoods, etc.) that were impacted. FLASH return periods were considered to provide the best indication of flash flood events in both situations, but they were particularly useful for identifying the broad areas impacted (Fig. 13a). Although FLASH return periods were typically able to identify the broad areas of interest well, participants expressed concern about the tendency of the values to appear to be unrealistically high (not shown). In addition, when examining specific locations more closely, participants often mentioned that the FLASH return period output

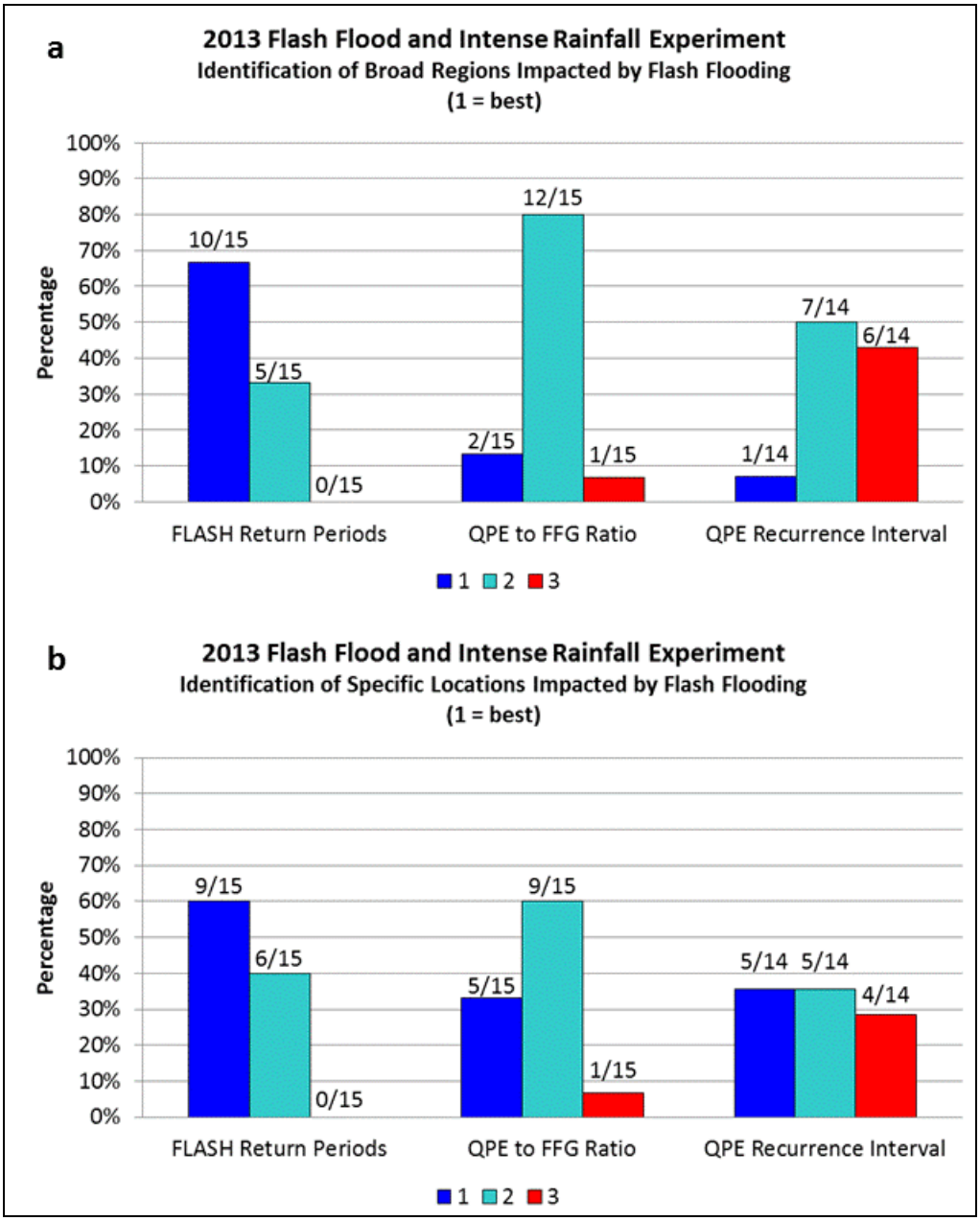

Figure 13. *Relative utility of flash flood diagnostic/forecast tools for identifying (a) the broad regions impacted by flash flooding and (b) specific locations impacted by flash flooding.*

appeared noisy, especially relative to both the QPE recurrence interval and the QPE-to-FFG ratio (Fig. 13b). Because of their more compact nature, participants generally considered these products to be more useful for identifying specific locations of flash flooding than for identifying broad areas of impact.

Some of the perceived differences in utility between these three tools may be the result of the different nature of the tools themselves. For example, since FLASH is high-resolution (1 km) and uses a hydrologic model to route water downstream, its tendency to appear noisy or spread out may be the result of its ability to move water away from the region that received the heaviest rainfall (Fig. 14a). Conversely, both the QPF-to-FFG ratio (Fig. 14b) and the QPE recurrence interval (Fig. 14c) rely on a strict comparison of QPE to either flash flood guidance or climatological precipitation values at a specific point. While QPE-to-FFG ratio incorporates some information about the expected hydrologic response due to its use of flash flood guidance, QPE recurrence intervals do not incorporate any hydrologic information, and neither tool dynamically moves water downstream. While both QPE-to-FFG ratio and QPE recurrence intervals provide useful information about locations that have received heavy rainfall, heavy rainfall alone is not always sufficient to produce flash flooding, and areas that exceed flash flood guidance do not always correspond to flash flood observations. In addition, the method of displaying the data may have also influenced the evaluation results. Additional investigation of this and other issues related to the display of this data on the FLASH website is being conducted at the University of Oklahoma in partnership with NSSL.

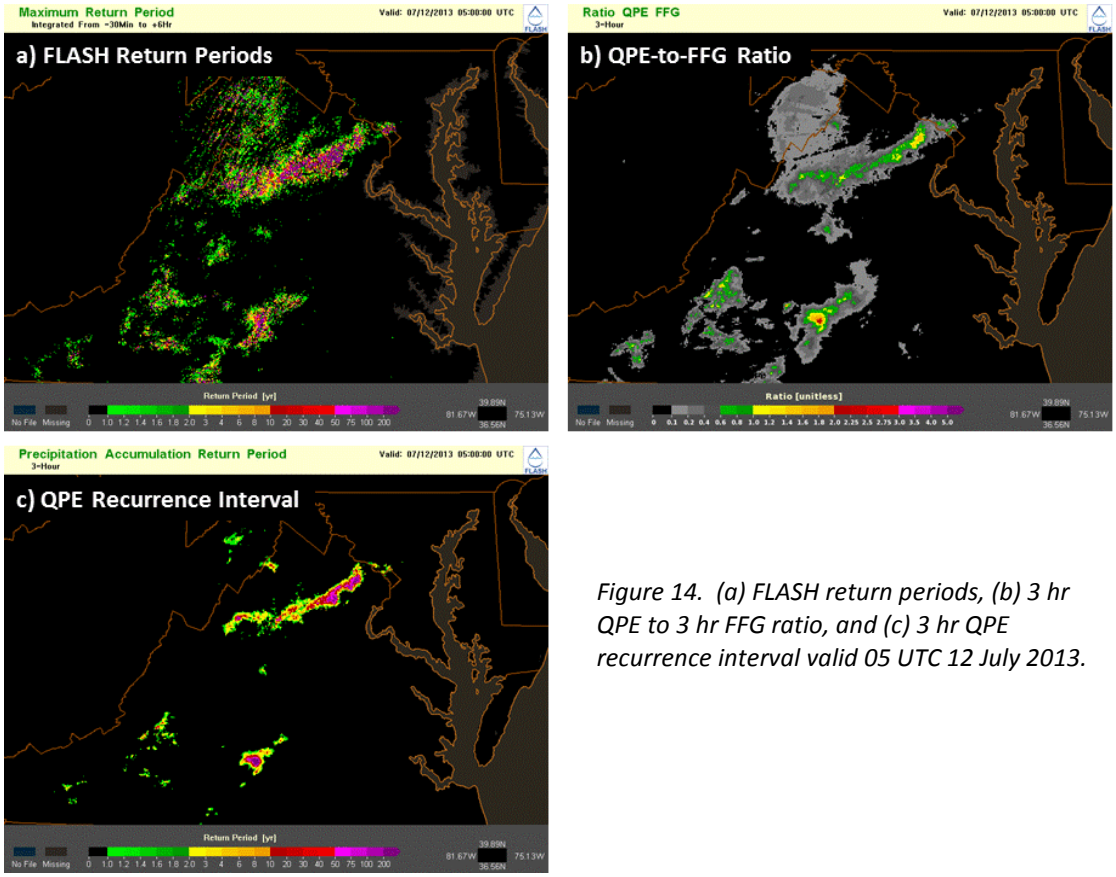

Figure 14. (a) FLASH return periods, (b) 3 hr QPE to 3 hr FFG ratio, and (c) 3 hr QPE recurrence interval valid 05 UTC 12 July 2013.

7. EXPERIMENTAL FORECAST PERFORMANCE

Figure 15 shows the results of the subjective verification for the four forecasts made each day in FFaIR. The PQPF (leftmost column in each cluster, in blue), initial probabilistic flash flood (middle, in shades of red) and flash flood outlook (right, in green) were rated as *good*, *fair*, or *poor*. The updated probabilistic flash flood forecast (different shading of the middle columns) was rated as *better* (pink), *about the same* (red), or *worse* (maroon*)*, depending on how the group felt it compared to the initial flash flood forecast. To evaluate the forecasts, the 12-hour PQPF forecast was compared against MRMS radar-estimated QPE data from the corresponding 12-hour period; the three probabilistic flash flood forecasts were compared to MRMS QPE and various flash flood diagnostics provided on the FLASH website (e.g. FFW, LSRs, FLASH recurrence intervals, QPE>FFG, precipitation recurrence intervals).

Overall, participants thought the 12-hour QPF forecasts performed well (Fig. 15), with 80% of forecasts receiving a *good* evaluation and 20% receiving a *fair*. None of the 12-hr QPF forecasts in the experiment received a *poor* rating. This differed from the initial 6-hour and outlook flash flood forecasts, which saw a more diverse distribution (47% of *good* and *fair*, 7% *poor)*.

When considering the updated 6-hour flash flood forecast, a majority (8, 53%) were rated as *better* (pink) than the original forecast while 5 (33%) were *about the same* (red) and 2 (13%) were *worse* (maroon). The forecast team noted that having the additional observational data was a significant help in updating the forecast. However, the fact that the update took place from roughly 1830-19 UTC, which was already an hour into the forecast period (valid from 18-00 UTC), occasionally allowed forecasters to adjust for rain and flash flooding that had already occurred. This tended to produce a better forecast, as the team would get credit for

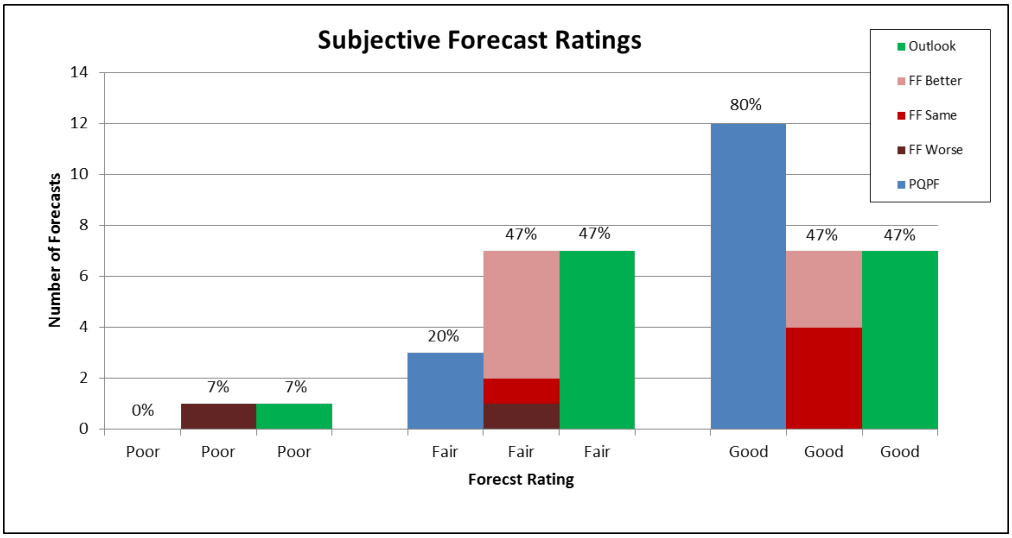

Figure 15. Experimental forecast ratings based on subjective model evaluations. Participants were asked to rate the 12-hr PQPF (blue, left columns), 6-hr probability of flash flooding (reds, center columns) and 12-hr probability of flash flooding (green, right columns) forecast as either good, fair, or poor. Additionally, participants rated the updated probability of flash flooding forecast as better (pink), about the same (red), or worse (maroon) than the original probability of flash flood forecast.

'forecasting' flash flooding that had already or currently was occurring. There was also the issue of verifying the flash flood forecasts, as the most trusted flash flood diagnostic data (e.g. LSRs and FFW) differ in their dependability and usefulness depending on location and time of day. Participants took this into account when issuing their rankings.

One of the main challenges of subjectively evaluating the forecasts, aside from issues associated with the lack of consistent flash flood observations, was how to interpret the probability contours. Throughout the three weeks of the experiment, participants noted that probability can mean different things meteorologically: a percentage of areal coverage, percent chance occurring within a radius of a certain point, number of times out of 100 forecasts that an event will be observed at one location, etc. This was one of the main difficulties identified by the forecast team: the message they were trying to convey with their forecast was often not the message received by the remote participants who called in to help with the forecast evaluation process, but were not involved in the creation of the forecast itself.

Despite these challenges, an analysis of the higher probabilities (30% and 50%) indicated in the three flash flood forecasts (initial, update, and outlook) was conducted in order to gain insight into the ability to correctly highlight areas of elevated flash flood risk (Table 5). For this analysis, FFWs and LSRs were used as 'truth', and only one warning or report was needed in order to be considered a forecast hit. While this is an overly simplistic technique that has a number of limitations, it represents an initial attempt to quantify the value of these higher probability forecasts.

The 30% contour performed well in all three forecasts (Table 5a), with 12 out of 19 (63%) of all 30% areas receiving at least one FFW or LSR for the initial forecast, 78% for the update forecast, and 63% for the outlook forecast. More importantly, the 30% probability areas that were considered forecast hits averaged at least 10 warnings/reports. The same trend can be seen for the 50% contour (Table 5b)—2/2 (100%) for the initial forecast, 80% for the update, and 0% for the outlook. Despite the small sample size, these results are encouraging because they suggest that forecasters are generally able to correctly identify areas with a higher risk for flash flooding. The strong performance of 30% contours drawn in the outlook forecast is of particular interest since when paired with the overall subjective forecast rating (Fig. 15, 14 out of 15 rated as *fair* or *good)*, suggests that there is value in flash flood forecasting with a 6-12 hour lead time, and that there is skill in identifying and designating broad areas of higher flash flood risk.

Table 5. Showing the evaluation of the (a) 30% and (b) 50% probability contours issued in FFaIR, using FFWs and LSRs as verification data. Also included are the number of 30% and 50% contours, and the corresponding verification, that were included in forecasts that were rated as good, fair or poor during the subjective forecast evaluation process. Values designated with an asterisk (*) are values determined from only one forecast.

a)	Initial 6-hour	Updated 6-hour	Outlook
Days with 30% forecast area	11/15	12/15	11/15
Total # of 30% areas issued	19	18	19
Forecast Hits	12	14	12
Forecast Misses	7	4	7
Avg # of reports/warnings (all areas)	7.32	9.33	6.84
Avg # of reports/warnings (forecast hits)	11.25	12	10.83
Max # of reports/warnings	84	88	44
Fcst rated good (better)	5	6	5
Fcst rated fair (same)	5	5	4
Fcst rate poor (worse)	1	1	1
Avg report/warn fcst 'good' (better)	13.11	4.38	11.10
Avg report /warn fcst 'fair' (same)	2.67	21.40	1.00
Avg report/warn fcst 'poor' (worse)	1.00*	2.00*	10.00*

b)	Initial 6-hour	Updated 6-hour	Outlook
Days with 50% forecast area	2/15	5/15	1/15
Total # of 30% areas issued	2	5	1
Forecast Hits	2	4	0
Forecast Misses	0	1	1
Avg # of reports/warnings (all areas)	30	14.8	0
Avg # of reports/warnings (forecast hits)	30	18.5	0
Max # of reports/warnings	59	59	0
Fcst rated good (better)	1	2	0
Fcst rated fair (same)	1	2	1
Fcst rate poor (worse)	0	1	0
Avg report/warn fcst 'good' (better)	59.00*	7.00	-
Avg report/warn fcst 'fair' (same)	1.00*	30.00	0*
Avg report/warn fcst 'poor' (worse)	-	0*	-

8. SUMMARY AND OPERATIONAL IMPACTS

The inaugural Flash Flood and Intense Rainfall Experiment was conducted from 8 – 26 July 2013 at the NOAA Center for Weather and Climate Prediction in College Park, MD. Over the course of the three week experiment, 18 forecasters, researchers, and model developers used a variety of high-resolution model guidance to issue a series of short-term QPF and flash flood forecasts. For the first time, an additional 8 forecasters participated remotely in the experimental forecast evaluation process. The remote participation component was an overwhelming success, and demonstrated that remote participation can provide a valuable, if limited, experiment experience.

Although the high-resolution models provided valuable forecast guidance in many cases, the experiment highlighted a significant gap in understanding between the meteorological and hydrologic aspects of flash flood forecasting. A number of the experiment findings are directly relevant to forecasters tasked with monitoring and forecasting the flash flood threat:

- **High-resolution model guidance is a vital component to a full evaluation of the flash flood threat** over the next 24 hours. While models will not correctly forecast every event every time, they can provide valuable information that can highlight the potential for an event *before* precipitation develops on radar, particularly when there is model consensus (e.g., Fig. 7).
- **The HRRR consistently provided valuable forecast guidance, and was the best performing deterministic high-resolution model during the experiment.** Since the HRRR is currently only available every three hours in WPC's NAWIPS system, participants frequently turned to the web-based graphics to get the latest model updates. Based on these results, HMT-WPC is exploring the feasibility of ingesting these data hourly.
- While not specifically evaluated during the experiment, **flash flood guidance appears to have a number of significant limitations** and typically does not provide a complete assessment of the flash flood threat. Participants noted that flash flood guidance seemed to be particularly problematic in regions of complex terrain and along RFC boundaries, and that data latency can be a significant issue in regions receiving multiple rounds of precipitation.
- Although no single forecast tool can accurately predict every flash flood event and despite the known limitations of flash flood guidance, **probabilities of QPF>FFG provided valuable forecast guidance**. QPF>FFG probabilities were particularly useful over the eastern two thirds of the country where they helped highlight areas with a significant flash flood threat within larger areas of heavy rainfall. However, the utility of these probabilities was limited in the West, where QPF amounts, precipitable water values and rain rates proved to be better flash flood indicators.
- Given the limitations of both warm season precipitation forecasts and the flash flood guidance, **neighborhood probabilities can be a particularly useful forecast tool**. By accounting for some of the spatial uncertainty that is inherent in both the model guidance and the hydrologic response, neighborhood probabilities can provide a more realistic depiction of the threat areas.
- While participants were consistently able to correctly identify areas with a higher risk of flash flooding, **drawing one broad probability contour was often a more effective means of conveying an increased flash flood threat than drawing multiple smaller contours.** The presence of multiple contours with the same probability in close proximity to one another tended to imply higher overall forecast confidence; while a forecaster may have high meteorological confidence in the situation, the non-linear nature of convective precipitation and the corresponding hydrologic response inherently creates uncertainty that must be accounted for.

The Flash Flood and Intense Rainfall Experiment provided a unique opportunity to bring the meteorological and hydrologic communities together to explore the challenges of both short-

term QPF and flash flood forecasting. The experiment provided valuable feedback that will help guide the development of new forecast tools to support WPC's MetWatch Desk and raised awareness about the limitations of the currently available forecast guidance. In addition, HMT-WPC and NSSL were able to receive valuable feedback regarding the applicability and optimization of the experimental datasets, which will go a long way toward continuing the development of these flash flood forecasting tools. In the coming months, HMT-WPC will work to implement the lessons learned, which will include making more intelligent use of the data and optimizing data display to forecasters. HMT-WPC will also discuss some of the alternate definitions of probabilistic forecasts that were proposed during the experiment to determine whether changing the meaning of these probabilistic forecasts would better serve WPC's mission.

ACKNOWLEDGEMENTS

The Flash Flood and Intense Rainfall Experiment would not have been possible without the dedication of a host of individuals including Faye Barthold (HMT-WPC), Tom Workoff (HMT-WPC), Wallace Hogsett (WPC), JJ Gourley (NSSL), Kelly Mahoney (ESRL), and Dave Novak (WPC). Ligia Bernardet (ESRL) was instrumental in providing access to the ExREF during the experiment. Eric Rogers (EMC) provided the data from the NAMX. WPC forecasters Brendon Rubin-Oster, Patrick Burke, and Andrew Orrison helped lead participants through the experimental forecast process.

REFERENCES

Ebert, E.E., 2008: Fuzzy verification of high resolution gridded forecasts: A review and proposed framework. *Meteor. Appl.*, **15**, 53-66.

Gourley, J.J., and Coauthors, 2013: A unified flash flood database across the United States. *Bull. Amer. Meteor. Soc.*, **94**, 799-805

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. Sever Local Storms, Nashville, TN. Amer. Meteor. Soc. P9.137.

Lin, Y. and K. Mitchell, 2005: The NCEP Stage II/IV hourly precipitation analyses: development and applications. Preprints.*19th Conf. on Hydrology*, San Diego, CA., 1.2.

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.

Schwartz, C.S., and Coauthors, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea.* Forecasting, 25, 263-280.

Smith, L. C., 2007: Satellite Remote Sensing of River Inundation Area, Stage, and Discharge: A Review. *Hydro. Processes,* **11**, 1427-1439.

APPENDIX A
Participants

Week	WPC Forecaster	WFO/RFC*	Research/Academia	EMC
July 8 – 12	Brendon Rubin-Oster		Ed Clark (HSD) Roham Abtahi (HSD) Bill Gallus (ISU) JJ Gourley (NSSL) Zac Flamig (OU) Jessica Erlingis (OU) Elizabeth Mintmire (OU)	Jacob Carley
July 15 – 19	Patrick Burke	Ron Horwood (NERFC) Chad Kahler (SLC) Jon Zeitler (EWX)	Ligia Bernardet (ESRL) Kelly Mahoney (ESRL) Jessica Erlingis (OU) Race Clark (OU) Elizabeth Mintmire (OU)	Geoff Manikin
July 22 – 26	Andrew Orrison	Sarah Jamison (CLE) Dean Hazen (PIH) Bill Martin (GGW) Greg Forrester (GGW) Josh Palmer (SERFC)	Mark Antolik (MDL) Dave Kitzmiller (OHD) Race Clark (OU) Jill Hardy (OU) Elizabeth Mintmire (OU)	Matt Pyle

*remote participants

APPENDIX B
Daily Schedule

8:00am – 9:30am Subjective evaluation of FLASH/flash flood diagnostics and the previous day's experimental forecasts

9:30am – 11:00am Determine forecast area of interest. Use observations and 00 UTC guidance to issue 12-hour (12 – 00 UTC) probability of quantitative precipitation forecast (PQPF) of greater than 1"

11:00am – 12:15pm WPC-CPC map discussion and lunch

12:15pm – 1:30pm Use observations and 00 UTC guidance to issue 6-hour (18 – 00 UTC) probability of flash flooding forecast

1:30pm – 2:30pm Subjective evaluation of the numerical model guidance and forecast tools

2:30pm – 3:00pm As 12 UTC model guidance becomes available, update the 18 – 00 UTC probabilistic flash flood forecast

3:00pm – 4:00pm Use all available guidance to issue the 12-hour (00 – 12 UTC) probability of flash flooding forecast for the overnight period

4:00pm – 4:30pm Group discussion