2015 Flash Flood and Intense Rainfall Experiment

Final Report

October 1, 2015

INTRODUCTION

During the three weeks from July 6 to July 24, 2015, the Hydrometeorological Testbed at the Weather Prediction Center (WPC-HMT) hosted the third annual Flash Flood and Intense Rainfall (FFaIR) Experiment. In an effort to support improvements to WPC's operational Excessive Rainfall Outlook (ERO) and explore the utility and accuracy of shorter, 6-hourly probability of flash flood forecasts, the FFaIR Experiment brought together 24 participants (Appendix A) from the forecast, research, and modeling communities to investigate methods for improving flash flood forecasting in both the near term (0-6 hour) and short range (Day 1) forecast periods. The WPC-HMT collaborated in tandem with the Norman Hydrometeorology Testbed (HMT-Hydro) with the common goal of improving the forecasting of flash floods.

More specifically, the goals of the 2015 Flash Flood and Intense Rainfall Experiment were to:

- Evaluate ways to maximize the utility of high resolution convection-allowing models (CAMs) and ensembles for short-term flash flood forecasts, not only to benefit WPC forecasters but to identify the best forcing for hydrologic models.
- Identify the most effective forms and proper usage of available hydrologic and climatological guidance for the prediction of flash floods.
- Explore proposed changes to WPC's operational Excessive Rainfall Outlook by evaluating the utility of probabilistic flash flood forecasts for Day 1.
- Explore ways to improve WPC's Mesoscale Precipitation Discussion (MPD) by evaluating the utility of 6-hourly forecasts.
- Enhance cross-testbed collaboration as well as collaboration between the operational forecasting, research, and academic communities on the forecast challenges associated with short-term flash flood forecasting.

This report will provide a summary of the activities, subjective evaluations, and potential enhancements to operations resulting from the experiment.

EXPERIMENT DESCRIPTION

Data

In addition to the full multi-center suite of operational deterministic and ensemble guidance, the 2015 FFaIR Experiment featured several experimental ensemble systems: The Storm Scale Ensemble of Opportunity, or SSEO (Jirak et al., 2012), provided by the Storm Prediction Center (SPC), a modified version of the SSEO provided by WPC, and the High-Resolution Ensemble Forecast (HREF) provided by NCEP's Environmental Modeling Center (EMC). The experiment also featured three experimental high-resolution deterministic models: the High Resolution

Rapid Refresh (HRRR) provided by ESRL, the experimental 3-km North American Model (NAM) provided by EMC, and the ADSTAT model provided by the National Water Center (NWC). Table 1 summarizes the model data that was the focus of the experiment.

***Table 1**. Featured 2015 FFaIR deterministic and ensemble model guidance (Experimental guidance is shaded).*

Model	Provider	Grid Spacing	Forecast Hours	Notes
NAM	EMC	12 km (parent) 4 km (nest)	84 (parent) 60 (nest)	Operational NAM, includes 12 km parent model and 4 km CONUS nest
Flash Flood Guidance	RFCs	5 km	01, 03, 06, 12 and 24 hour values	CONUS mosaic grid created by compiling individual RFC-domain grids
HRRR	EMC	3 km	15	High resolution, hourly updated, convection-allowing run initialized by the Rapid Refresh (RAP) model
NMMB ARW WRFNSSL	EMC/NSSL	4 km	48 36 (WRFNSSL)	High resolution, convection allowing CONUS models
SSEO	SPC/NSSL/EMC	~4 km	36	Multi-physics, convection - allowing ensemble consisting of 7 high-resolution deterministic models
SREF	NCEP	16 km	87	21 Members, Multi-model, Multi-physics, Multi-IC ensemble system
Experimental NAM	EMC	3 km (nest)	60 (nest)	NAMX; features differing data assimilation (ENKF), is higher resolution than the operational NAM nest
HRRRX	ESRL/GSD	3 km	15	Experimental version of the HRRR, hourly updated, convection-allowing
WPC-SSEO (7 members)	SPC/ESRL/WPC	4 km	24	Modification of the original SSEO provided by SPC; removes SPC-WRF and adds HRRR
ADSTAT	NWC	4 km	6	6-hour QPF and PQPF forecasts; extrapolation of radar and satellite rain-rates, and RAP and humidity forecasts, are blended to produce guidance.
Precipitation Recurrence Data (Atlas 14)	NSSL/HDSC/NERFC/CSU	5 km	3- and 6-hour (1, 2, 5, 10, 25 and 100 year intervals)	Precipitation frequency estimates based on historical observations.
HREF (11 members)	EMC	5 km	36	11 member, multi-model, convection-allowing high-resolution ensemble

Deterministic Guidance

The Experimental NAM (**NAMX**) is an evolving experimental version of the operational CONUS NAM. This alternate version features full use of the global Ensemble Kalman Filter (EnKF) members as part of its data assimilation system, as well as increases horizontal grid spacing from 4 km to 3 km. The nest is available at 00, 06, 12 and 18 UTC.

The operational HRRR (http://ruc.noaa.gov/hrrr) has 3 km grid spacing and uses boundary conditions from the hourly updated, radar-assimilated Rapid Refresh (RAP) model. It features a WRF-ARW core, Thompson microphysics, and is fully convection-allowing. The Experimental HRRR (**HRRRX**) in the FFaIR Experiment was initialized with the latest 3-D radar reflectivity using a digital filter initialization (radar-DFI) technique (via the parent 13 km RAP) and provided data hourly. The HRRRX uses grid-point statistical interpolation (GSI) hybrid data assimilation and broadens and weakens the forcing applied from radar reflectivity data assimilation to reduce excessive convective storm development early in the HRRR cycle. Replacing the Rapid Radiative Transfer Model (RRTM)/Goddard radiative schemes with the RRTMG (RRTM with General Circulation Model applications) for both long wave and shortwave, the HRRRX features an enhanced planetary boundary layer (PBL) scheme, increased roughness length values to help reduce a high surface wind speed bias, seasonally-varying vegetation fraction, an update to the land-surface model and updated microphysics.

The Advective-Statistical Forecasts of Rainfall *(***ADSTAT**; Kitzmiller et al. 2011) package produces forecasts of 6-hour total precipitation amounts and 6-hour probabilistic quantitative precipitation forecasts (PQPFs) of 0.25, 2.5, 12.5, 25, 50, and 75 mm thresholds. The grid mesh is the Hydrologic Rainfall Analysis Project (HRAP) grid system polar stereographic projection, with a nominal mesh length of 4 km over the CONUS. Forecasts for FFaIR were provided for 1700-2300 UTC and 1800-0000 UTC periods. Forecasts are based on input from the previous hours Rapid Refresh (RAP) model runs, Multi-Radar/Multi-Sensor (MRMS) radar precipitation rates, and Hydroestimator satellite IR precipitation rates. A combination of extrapolation forecasts of the initial-time radar and satellite rain-rates, and RAP precipitation and humidity forecasts is blended to produce the guidance.

Ensemble Guidance

The **SSEO** is a high-resolution, multi-model, multi-physics, CAM ensemble system produced by SPC. Issued at 00 and 12 UTC, it is composed of seven deterministic high-resolution members (Table 2). At WPC, the ensemble mean is displayed at 4 km, although each member can be viewed independently at its native resolution (Table 2). Two of the members (the operational ARW and NMM hi-res windows) are time-lagged by 12 hours to provide additional initial condition diversity (Jirak et al, 2012). It should be noted that the National Severe Storms Laboratory (NSSL) WRF-ARW and EMC WRF-NMM are non-operational and can be subject to outages, and the four high resolution window members (HRW-ARW and HRW-NMMB) are operational, but can be supplanted with other high resolution runs (e.g. hurricane models) if the need arises (Jirak et al, 2012).

At WPC, a modified version of the SSEO is also employed, which replaces the EMC WRF-NMM (member 6) with the latest cycle of the HRRR. This was done to mitigate the high QPF bias that has been observed with the EMC WRF-NMM member. Additionally, the WPC-SSEO is run at 06 and 18 UTC; the 06 and 18 UTC cycles feature five-time lagged members (members 1, 2, and 4 are time-lagged 6 hours, members 3 and 5 time-lagged 18 hours) along with the 06 UTC cycles of the HRRR and NAM nest.

Table 2. *Membership characteristics of the SSEO and WPC-SSEO. Members denoted by the asterisk (*) are time lagged by 12 hours. For the WPC-SSEO, member six is changed from the EMC WRF-NMM to the HRRR. Adapted from Jirak et al (2012).*

SSEO Member	Model	Provider	Grid Spacing	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 *06*	EMC WRF-NMM *HRRR*	EMC *ESRL*	4 km *3 km*	MYJ *MYNN*	Ferrier *Thompson*
07	NAM-NMMB Nest	EMC	4 km	MYJ	Ferrier

The **HREF** is the latest blend of real-time and time-lagged high resolution models composed of 11 members, of which 8 are time-lagged, from the Hi-Res Window ARW, Hi-Res Window NMMB, and NAM CONUS Nest. Membership varies slightly with forecast range, as the oldest Hi-Res Window runs do not cover the last 12 hours of the 36-hour-long HREF. The native grid spacing of the constituent models are in the below graphic, however, the ensemble itself is output on a 5 km grid.

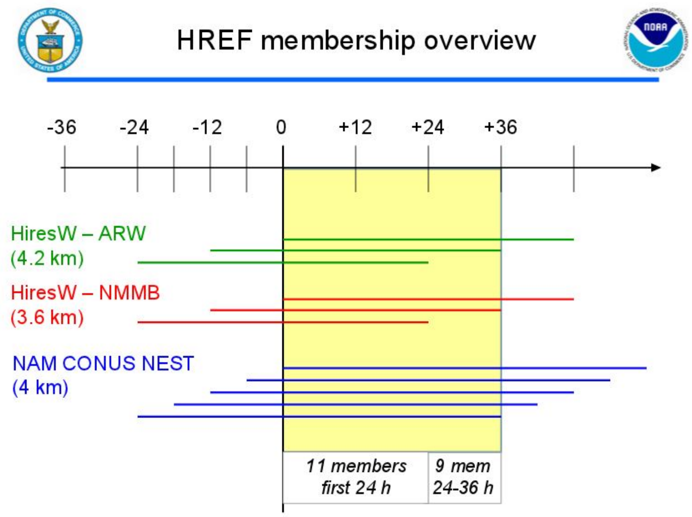


Figure 1. *Membership overview of the HREF Ensemble from EMC.*

Ensemble Forecast Tools

Neighborhood probabilities use the premise of the 'neighborhood maximum value' (Schwartz et al. 2009, Ebert 2008). For each grid point, a search is conducted within a 40 km radius to locate the maximum QPF value. The value of the original grid point is then replaced with this maximum value. This technique, used in alignment with the SPC severe local storms forecasting technique, attempts to offset spatial errors in hi-res model forecasts. Probabilities were available for the WPC-SSEO in hourly intervals out to 24 hours.

- *QPF (3- and 6-hour)*
  The neighborhood probability of the 3-hour and 6-hour QPF exceeding certain notable thresholds (e.g. 3 inches) were available. These probabilities are derived by determining the number of ensemble members which are predicting precipitation to exceed the relevant threshold using the neighborhood maximum QPF value at each individual grid point.

- *QPF > Flash Flood Guidance (FFG) (3- and 6-hour)*
  The neighborhood probability of the 3-hour and 6-hour QPF exceeding FFG values were available. Probabilities are derived by subtracting the FFG value from the neighborhood maximum QPF of each member at each individual grid point, followed by determining how many members predict precipitation to exceed FFG. Neighborhood probabilities of QPF exceeding a certain percentage (e.g. 75% and 90%) of FFG were also provided.

- *QPF > Recurrence Interval (3- and 6-hour)*

The neighborhood probability of the 3-hour and 6-hour QPF exceeding various precipitation recurrence interval values were also available. Probabilities are derived by subtracting the recurrence interval value from the neighborhood maximum QPF of each member at each individual grid point, followed by determining how many members predict precipitation to exceed the recurrence interval. Neighborhood probabilities of QPF exceeding recurrence intervals of 2, 5, 10, and 100 years were provided.

The HREF also provided threshold comparisons. Probabilities are derived from the maximum QPF of each member at each individual grid point. These gridded values included:

- *QPF > 0.25 inches (3-hour)*
  The probability of QPF exceeding 0.25 inch values
- *QPF > 0.5, 1, 2, and 3 inches (3-, 6-, and 24-hour)*
  The probability of QPF exceeding 0.5, 1, 2, and 3 inch values
- *QPF > 5 inches (24-hour)*
  The probability of QPF exceeding 5 inch values.

Additional Guidance

The Cooperative Institute for Meteorological Satellite Studies (CIMSS) **NearCast** model was evaluated, which uses information from the GOES-13 water vapor channels to identify areas most susceptible to convection. Winds from the RUC are used to predict trajectories of precipitable water measured from the GOES-13 sounder. This model can predict areas of destabilization (convective potential) if low-level moisture moves underneath upper-level drying. Fields are presented as moisture change with height. The system is updated hourly, and provides data in half-hour intervals out to nine hours.

The **Atmospheric River Detection Tool (ARDT)**, provided by the Earth Systems Research Laboratory (ESRL), was evaluated online. Output from the ARDT highlights regions forecast to be exposed to elevated atmospheric water vapor transport which can potentially contribute to extreme precipitation. The ARDT operates on forecast or analysis fields of integrated water vapor transport (IVT) and objectively identifies narrow corridors of enhanced transport consistent with the definition of atmospheric rivers (ARs). For the FFaIR experiment, the tool was available out to 48 hours derived from 6-hourly runs of the Global Forecast System (GFS) model. The ARDT was evaluated for its performance over the previous 48 hours during the daily verification session. The direct output includes estimates of the axis location and width of the IVT corridor as well as the IVT values along that axis. Graphical products showed the location of identified atmospheric rivers and indicators of their magnitude expressed as percent anomalies and percentiles of climatological values for AR conditions. Through analysis of multiple forecast cycles, indicators of persistent AR conditions were also displayed.

Daily Activities

Each of the three weeks, participants were grouped with a WPC MetWatch forecaster to form a collaborative forecast team. Each day, the team was tasked with completing several different experimental forecast exercises, which aimed to simulate the timeframe, workload and thought processes associated with creating a Day 1 Excessive Rainfall product and shorter, 6 hourly flash flood forecasts. Unlike their operational counterparts, all experimental forecasts in FFaIR employed the neighborhood approach and were defined as the probability of flash flooding occurring with 40 km of a point.

To start each day, the team generated a CONUS-wide 21-hour (15 – 12 UTC) experimental neighborhood probabilistic 'excessive rainfall outlook' using a similar process used to create WPC's operational ERO. In the afternoon, the team was tasked with creating two smaller-scale 6-hourly probabilistic flash flood forecasts and a Domain Limiting Discussion (DLD) similar to WPC's MPD. For these forecasts, the team was asked to identify a multi-state region where the risk of flash flooding was assessed to be the highest, then create a probabilistic flash flood forecast for the given 6-hour period. A detailed version of the daily schedule can be found in Appendix B.

21-hour (15 – 12 UTC) excessive rainfall outlook (ERO), due at 1400 UTC. Participants were asked to draw contours of 2%, 5%, 10% and 30% probability of flash flooding within 40 km of a point, when applicable, over the entire CONUS. This forecast was similar in scope to WPC's current operational ERO, but examined the applicability of the neighborhood approach, used different probability thresholds for each risk category, and is not based on an exceedance of FFG.

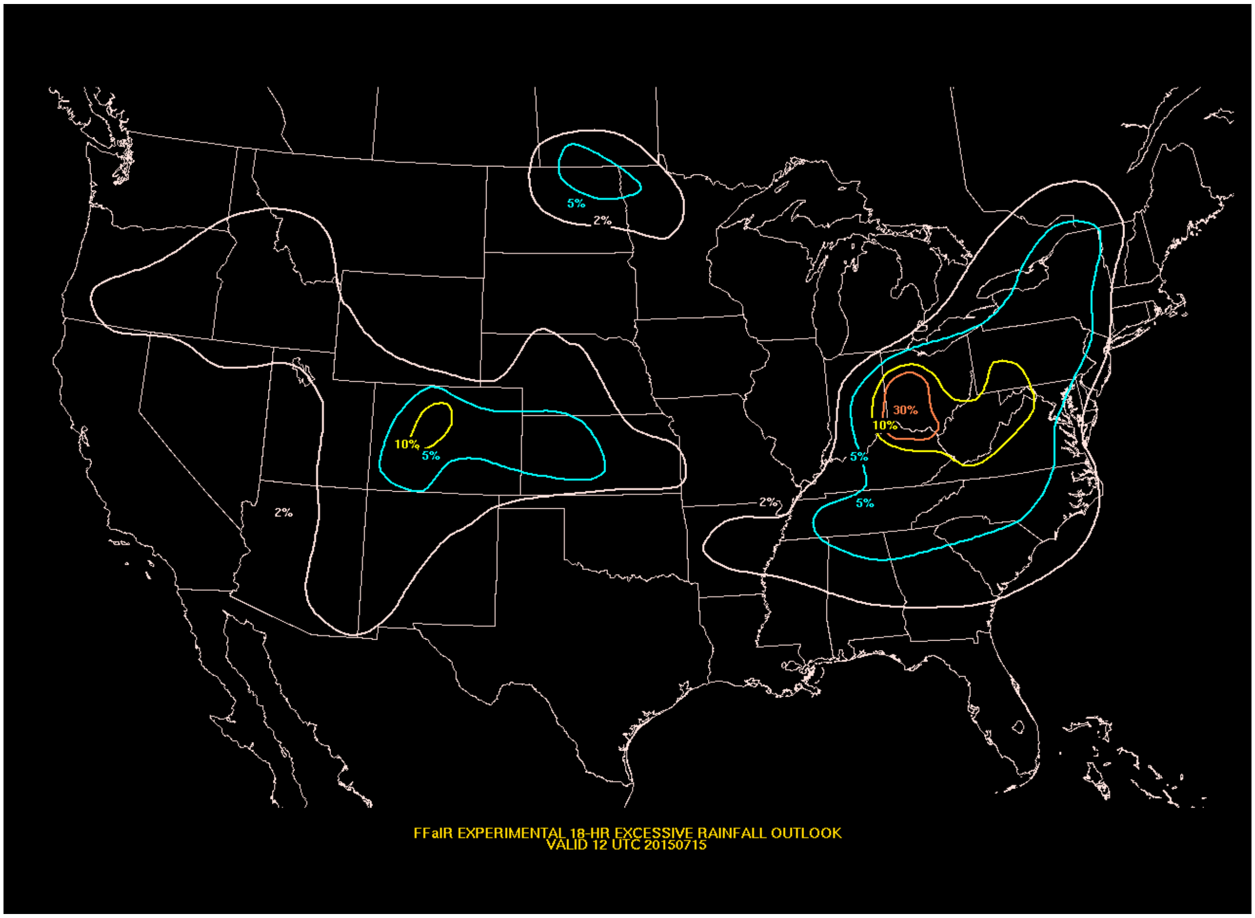


Figure 2. *Displaying the 21-hour excessive rainfall outlook forecast from July 15, 2015. 2% probabilistic contours are white, 5% cyan, 10% yellow, and 30% orange.*

6-hour (18 – 00 UTC) probability of flash flooding (PFF1), due at 1800 UTC. This forecast mirrored elements of WPC's MPD, except participants were instructed to draw contours of a 10%, 30% and 50% probability of flash flooding occurring within 40 km of a point, when applicable, over their chosen area of interest. This required the forecast team to consider both hydrologic and meteorological information to assess the flash flood threat and issue a forecast for the likelihood of flash flooding. An example of this forecast is shown in Figure 3.

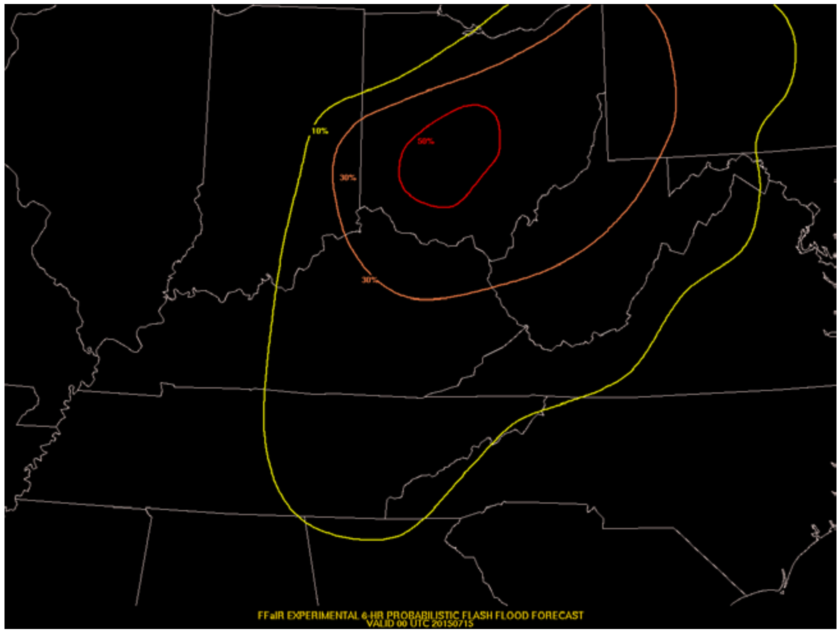
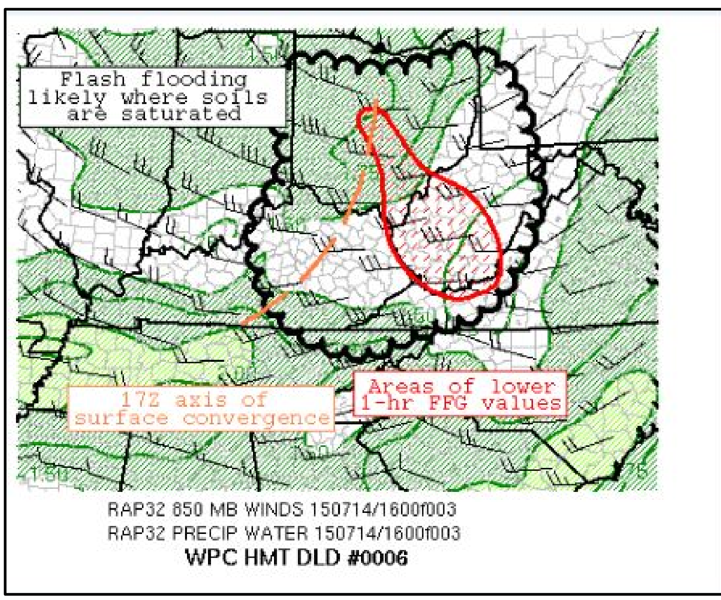
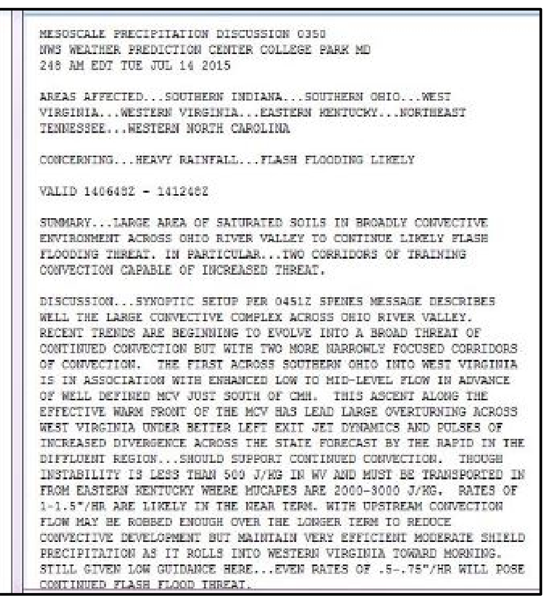


Figure 3. *Experimental probability of flash flooding forecast graphic valid from 18Z on July 14, 2015 to 00Z on July 15, 2015. 10% probabilistic contours are in yellow, 30% in orange, and 50% in red.*

6-hour (18-00 UTC) Domain Limiting Discussion, due at 1800 UTC. This product was a non-operational replica of the WPC MPD and an example of the one issued on July 14, 2015 can be seen in Figure 4. Created using the same software package and process that occurs on the MetWatch Desk, forecasters chose the area of highest risk of flash flooding within a 6 hour period based on the results of the PFF1 and accompanied the graphic with an in-depth text discussion.




MESOSCALE PRECIPITATION DISCUSSION 0350
NWS WEATHER PREDICTION CENTER COLLEGE PARK MD
248 AM EDT TUE JUL 14 2015

AREAS AFFECTED...SOUTHERN INDIANA...SOUTHERN OHIO...WEST VIRGINIA...WESTERN VIRGINIA...EASTERN KENTUCKY...NORTHEAST TENNESSEE...WESTERN NORTH CAROLINA

CONCERNING...HEAVY RAINFALL...FLASH FLOODING LIKELY

VALID 140645Z - 141245Z

SUMMARY...LARGE AREA OF SATURATED SOILS IN BROADLY CONVECTIVE ENVIRONMENT ACROSS OHIO RIVER VALLEY TO CONTINUE LIKELY FLASH FLOODING THREAT. IN PARTICULAR...TWO CORRIDORS OF TRAINING CONVECTION CAPABLE OF INCREASED THREAT.

DISCUSSION...SYNOPTIC SETUP PER 0451Z SPENES MESSAGE DESCRIBES WELL THE LARGE CONVECTIVE COMPLEX ACROSS OHIO RIVER VALLEY. RECENT TRENDS ARE BEGINNING TO EVOLVE INTO A BROAD THREAT OF CONTINUED CONVECTION BUT WITH TWO MORE NARROWLY FOCUSED CORRIDORS OF CONVECTION. THE FIRST ACROSS SOUTHERN OHIO INTO WEST VIRGINIA IS IN ASSOCIATION WITH ENHANCED LOW TO MID-LEVEL FLOW IN ADVANCE OF WELL DEFINED MCV JUST SOUTH OF CMH. THIS ASCENT ALONG THE EFFECTIVE WARM FRONT OF THE MCV HAS LEAD LARGE OVERTURNING ACROSS WEST VIRGINIA UNDER BETTER LEFT EXIT JET DYNAMICS AND PULSES OF INCREASED DIVERGENCE ACROSS THE STATE FORECAST BY THE RAPID IN THE DIFFLUENT REGION...SHOULD SUPPORT CONTINUED CONVECTION. THOUGH INSTABILITY IS LESS THAN 500 J/KG IN WV AND MUST BE TRANSPORTED IN FROM EASTERN KENTUCKY WHERE MUCAPES ARE 2000-3000 J/KG. RATES OF 1-1.5"/HR ARE LIKELY IN THE NEAR TERM. WITH UPSTREAM CONVECTION FLOW MAY BE ROBBED ENOUGH OVER THE LONGER TERM TO REDUCE CONVECTIVE DEVELOPMENT BUT MAINTAIN VERY EFFICIENT MODERATE SHIELD PRECIPITATION AS IT ROLLS INTO WESTERN VIRGINIA TOWARD MORNING. STILL GIVEN LOW GUIDANCE HERE...EVEN RATES OF .5-.75"/HR WILL POSE CONTINUED FLASH FLOOD THREAT.

Figure 4. *The 18-00 UTC Domain Limiting Discussion valid July 14, 2015.*

6-hour (00 – 06 UTC) probability of flash flooding (PFF2), due at 2000 UTC. This forecast required the forecast team to consider an altered collection of guidance, including available 1200 UTC model guidance, and submit their forecast with a longer lead time (4 hours). Participants were again instructed to draw contours of a 10%, 30% and 50% probability of flash flooding within 40 km of a point, when applicable, over their chosen area of interest. An example is shown in Figure 5 valid from 0000-0600 UTC on July 15, 2015.

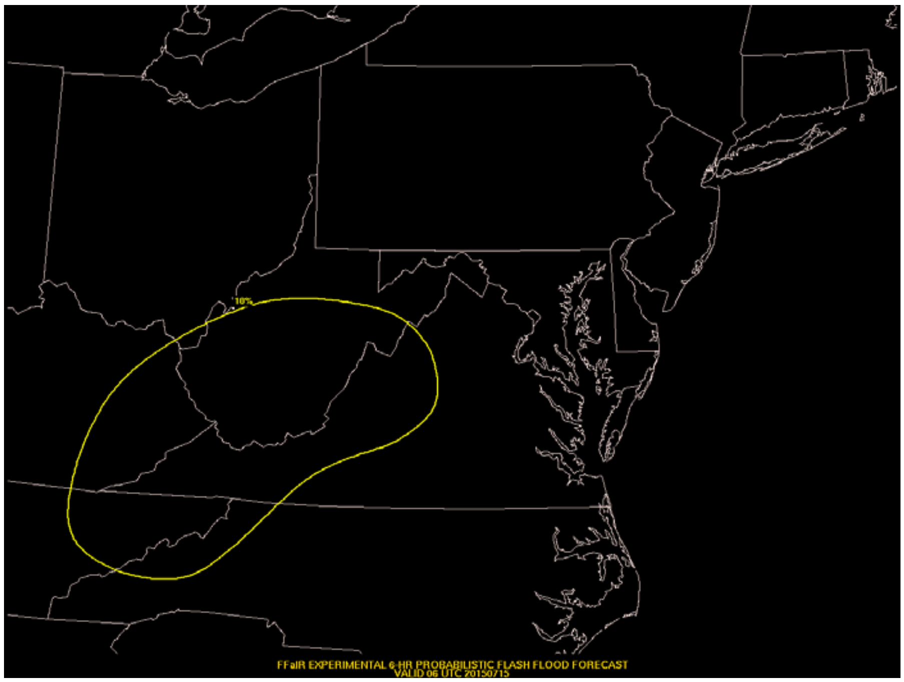


Figure 5. *Experimental probability of flash flooding forecast graphic valid 00- 06 UTC on July 15, 2015*

During the development of these experimental forecasts, participants were asked to prepare a PowerPoint briefing comprised of a collection of real-time situational awareness graphics such as current radar, visible satellite, and water vapor imagery, guidance data, as well as other information explaining their forecast rationale and highlighting the major areas of concern across the country. This discussion was then used to provide a daily forecast briefing to participants in the Hydrometeorological Testbed Hydro (HMT-Hydro) program in Norman, Oklahoma, which was conducting a concurrent flash flood watch and warning exercise.

Lastly, participants were also asked to visually and subjectively evaluate their experimental forecasts and the corresponding experimental model guidance and forecast tools. The subjective evaluations graded the relative accuracy and effectiveness of the experimental forecasts and model guidance against a combination of flash flood indicators, including radar-estimated QPE from the MRMS system, flash flood warnings (FFWs), areas of QPE-to-FFG exceedance, local storm reports (LSRs) of flooding and/or flash flooding, stream flow data from the U.S. Geological Survey (USGS), and NSSL Meteorological Phenomena Identification Near the Ground (mPING) reports.

SYNOPTIC OVERVIEW AND DAILY IMPACTS DURING THE EXPERIMENT

Throughout the three weeks of the experiment, dynamic weather throughout the U.S. allowed the forecasters to issue their experimental products over multiple geographic regions, from the West Coast, through the Great Plains, and the East Coast and Southeast Region. The mean 500 hPa heights, shown in Figure 6A, were characterized by a pronounced trough just off the coast of California, a broad ridge through the middle part of the country, and a weak trough just off of the East coast. The mean 500 hPa height anomalies in Figure 6B show the dramatic trough over the West coast that was atypical for this time of year. Mean precipitable water anomalies over the course of the three weeks are displayed in Figure 7. Anomalously high precipitable water was present stretching from the Southwest Region in Arizona and New Mexico northeast into the Central Plains and east to the Ohio River Valley. These areas were indeed some of the most active zones over the three week period. Anomalously high precipitable water also was present over northern California and into the Northwest Region.

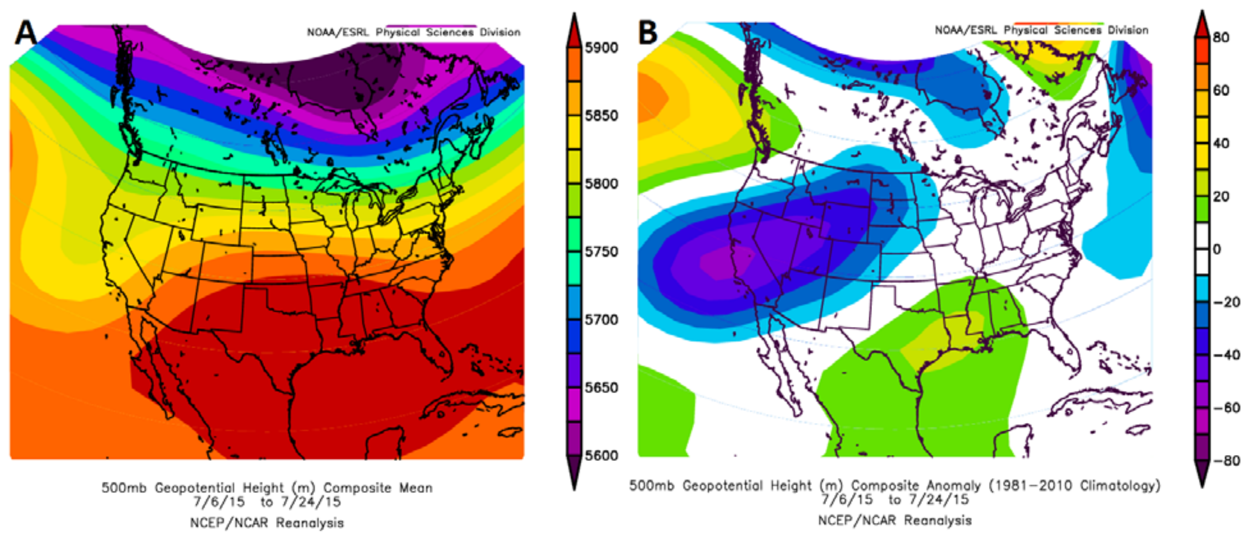


Figure 6. *A) Composite mean 500 hPa geopotential heights over the three week experiment period from July 6-July 24, 2015. B) Mean 500 hPa geopotential height anomalies over the same time period. Images generated from the NCEP/NCAR Reanalysis provided by NOAA/ESRL/Physical Sciences Division (http://www.esrl.noaa.gov/psd/data/composites/day/).*

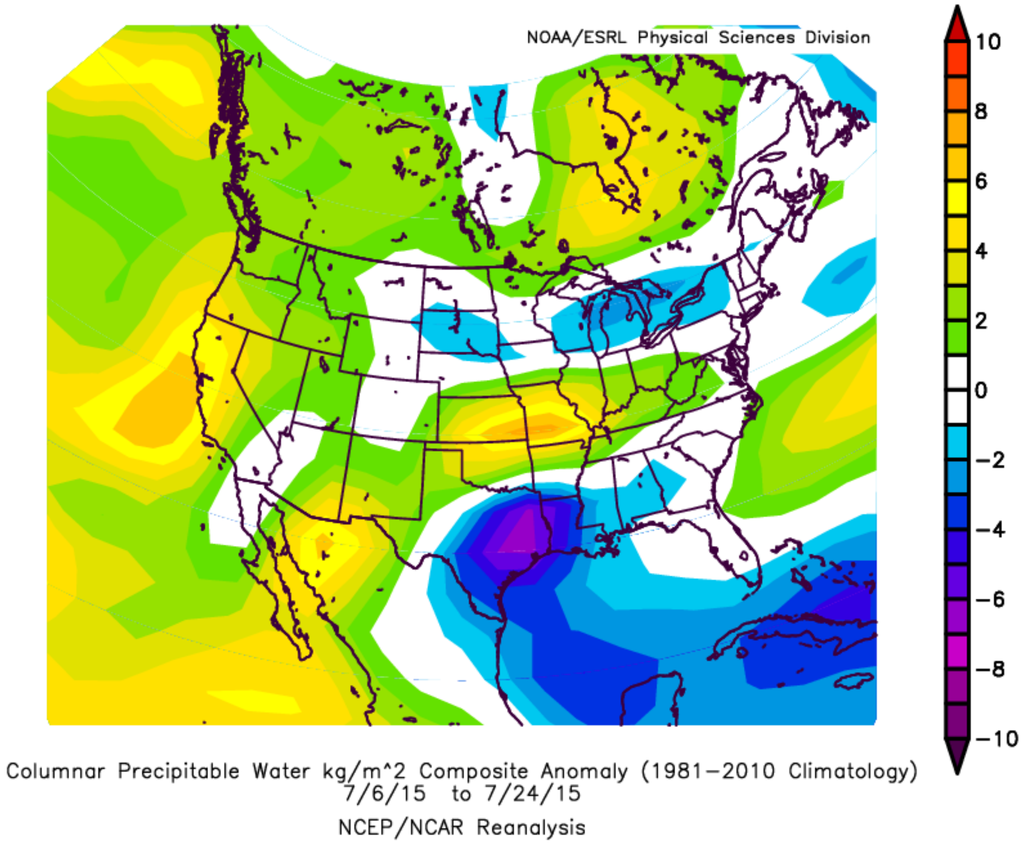


Figure 7. *Precipitable water anomalies over the U.S. during the weeks of the experiment, July 6-July 24, 2015. Images generated from the NCEP/NCAR Reanalysis provided by NOAA/ESRL/Physical Sciences Division (http://www.esrl.noaa.gov/psd/data/composites/day/).*

Some of the more notable events that occurred over the course of the three weeks were a flash flooding event in the Las Vegas, NV area on July 8 in which a number of water rescues were performed, severe flash flooding in Kentucky on July 14 where three people lost their lives, and a flash flooding event in Brooklyn, NY on July 16 where a number of vehicles became stranded in the high water. A complete list of the where the PFF1 and PFF2 forecasts were focused and any notable impacts during this year's experiment can be found in Table 3.

Table 3. *Experimental forecasts issued during the 2015 Flash Flood and Intense Rainfall Experiment. Note that on July 11, July 18, and July 25, no 0600 UTC forecast was issued due to changes in the experiment schedule.*

Forecast Valid Date	Valid Time	Forecast Area	Notes
7 July 2015	00 UTC	Southern Plains	
	06 UTC	Texas Panhandle	Abilene, Texas, set a new all-time record for the highest daily rainfall total - 8.26 inches
8 July 2015	00 UTC	Northern Texas, Southern Plains	Over a foot of rain fell in some locations - more than 10 water rescues in the Las Vegas area
	06UTC	Northern California, Nevada	
9 July 2015	00 UTC	Southern Plains/Midwest, including	Flooding reported in Muskogee,

		Oklahoma, Missouri, Illinois.	OK.
	06 UTC	Texas Panhandle, SE Colorado	
10 July 2015	00 UTC	Northern California, Oregon	Flash Flood in Ashland, Oregon recorded on video
	06 UTC	Texas Panhandle	
11 July 2015	00 UTC	Ohio River Valley	Flooding reported in Wallback, WV.
	06 UTC	N/A	
14 July 2015	00 UTC	Ohio River Valley and Midwest.	KY: **3 DEAD**, 5 missing; flooding damaged or destroyed 150 homes; impacted a total of 500 – state of emergency declared
	06 UTC	Illinois, Indiana, Kentucky	
15 July 2015	00 UTC	Ohio River Valley, eastern Tennessee	OH: Businesses closed/basements flooded in Chagrin Falls; multiple water rescues in Novelty
	06 UTC	West Virginia, extreme western Virginia	
16 July 2015	00 UTC	Mid Atlantic to Southern New England	Rain rates up to 1 inch in 30 minutes swamped Brooklyn, NY; stranded vehicles in NJ; flooded basements in PA
	06 UTC	Iowa, Missouri	
17 July 2015	00 UTC	Iowa, Illinois	Flash flooding reported in Monmouth, IL
	06 UTC	Illinois – Quad Cities	
18 July 2015	00 UTC	Arizona, New Mexico	AZ: over 3.50 inches of rain over two hours; motor homes swept away; 7 structures completely flooded; roads closed
	06 UTC	N/A	
21 July 2015	00 UTC	Colorado, New Mexico, Texas Panhandle	Amarillo, TX: multiple water rescues; roads closed, barricades set up
	06 UTC	Ohio Valley	
22 July 2015	00 UTC	Southern Plains, Northern Arkansas	Significant flash flooding in Wichita, KS.
	06 UTC	Oklahoma	Disaster Recovery Center set up to help people in McAlester affected by flooding
23 July 2015	00 UTC	Idaho, Montana	
	06 UTC	Arkansas, Tennessee	
24 July 2015	00 UTC	North Carolina, eastern Tennessee	NC: people stranded in cars from floods in Morehead City; impassible roads in Greenville, structural flooding in Beaufort, Newport
	06 UTC	Northern Plains	
25 July 2015	00 UTC	Florida panhandle, NE Arkansas	
	06 UTC	N/A	

DETERMINISTIC HIGH RESOLUTION MODEL PERFORMANCE

As part of the subjective evaluation process, participants were asked to rate the QPF guidance provided by each of the deterministic CAMs on a scale of 1 (very poor) to 5 (very good) based on the observed precipitation during the 00-06 UTC period. Each model solution was evaluated independently; the models were not ranked from best to worst. The results, shown in Figure 8 below, are based primarily on these subjective responses.

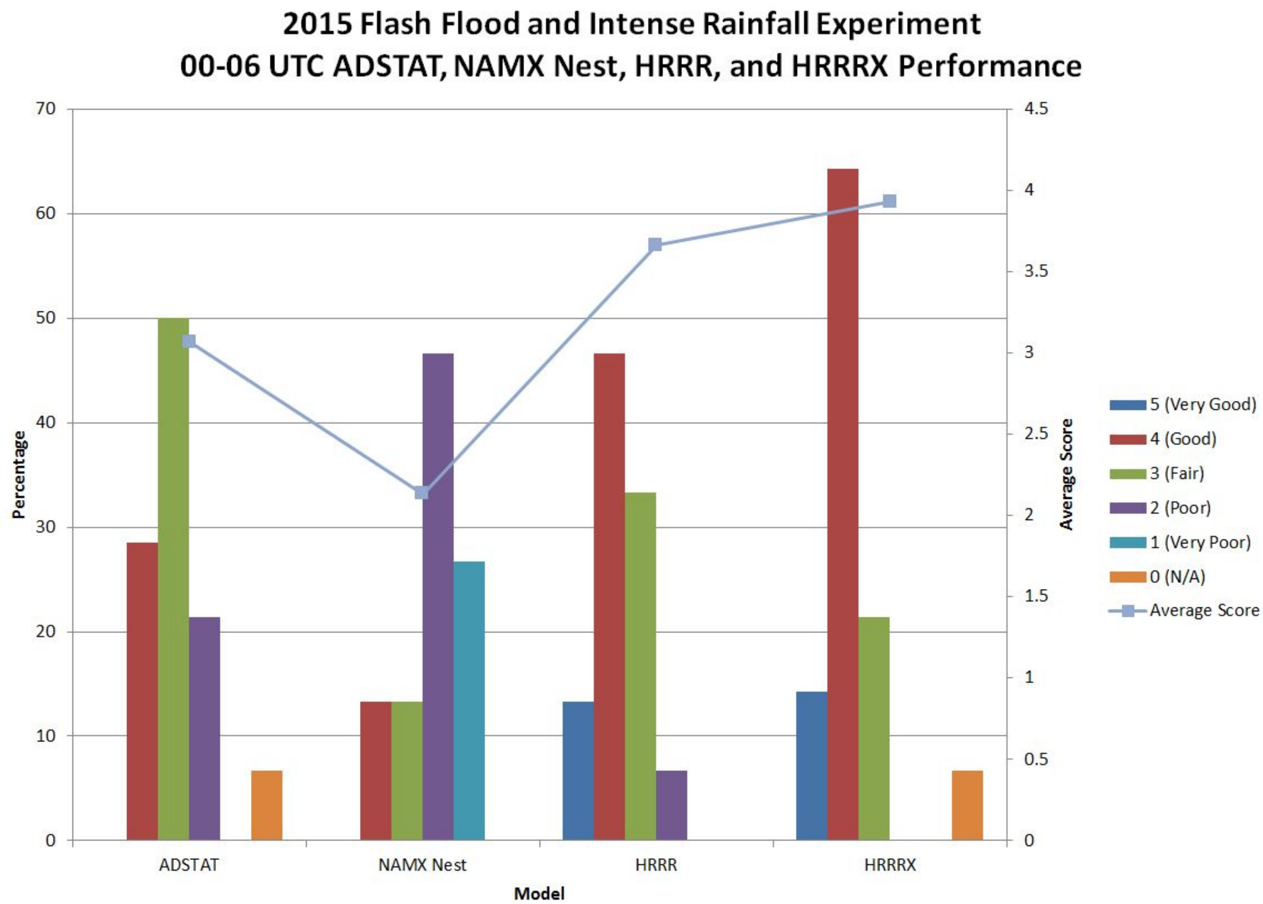


***Figure 8**. Experimental deterministic model performance based on feedback from subjective model evaluations conducted during the 2015 Flash Flood and Intense Rainfall Experiment. Participants were asked to rate the performance of each model on a scale of 1 (very poor) to 5 (very good). The Average Score is a numerical average of all scores calculated over the experiment, not an average of the percentages shown.*

Each of the high resolution CAM brought value to the forecast process in location and intensity of QPF. Figure 9 shows all of the models together with the NSSL QPE and practically perfect forecast[1] (Hitchens et al, 2013) as an example of how the models were subjectively evaluated

[1] *The "practically perfect" technique converts point observations into contoured areas using a Gaussian weighted function (Hitchens et al. 2013) with the goal of producing the probabilistic forecast a forecaster would have issued had the location of all reported flash flooding been known in advance. At WPC this is done by placing flash flood LSRs, mPING reports, and USGS stream gauge exceedance locations onto a 5 km grid. Once on the grid, any grid cell within 40 km of an observation (consistent with the definition of our experimental forecasts) is assigned a 100% probability of flash flooding. These values are then smoothed over 80 km to yield an approximation of the probability of flash flooding within 40 km of a point.*

by the participants. It was determined that although the Experimental HRRR (HRRRX) did not vary greatly from the operational, there was an improvement in location and intensity of precipitation maxima. The HRRRX model statistically performed the best scoring a 3.93 subjective verification score. The operational HRRR had a score of 3.67 and was the second most favorably ranked model. The HRRRX and the HRRR were the only high-resolution models to receive a perfect score on two different days of the 15-day experiment.

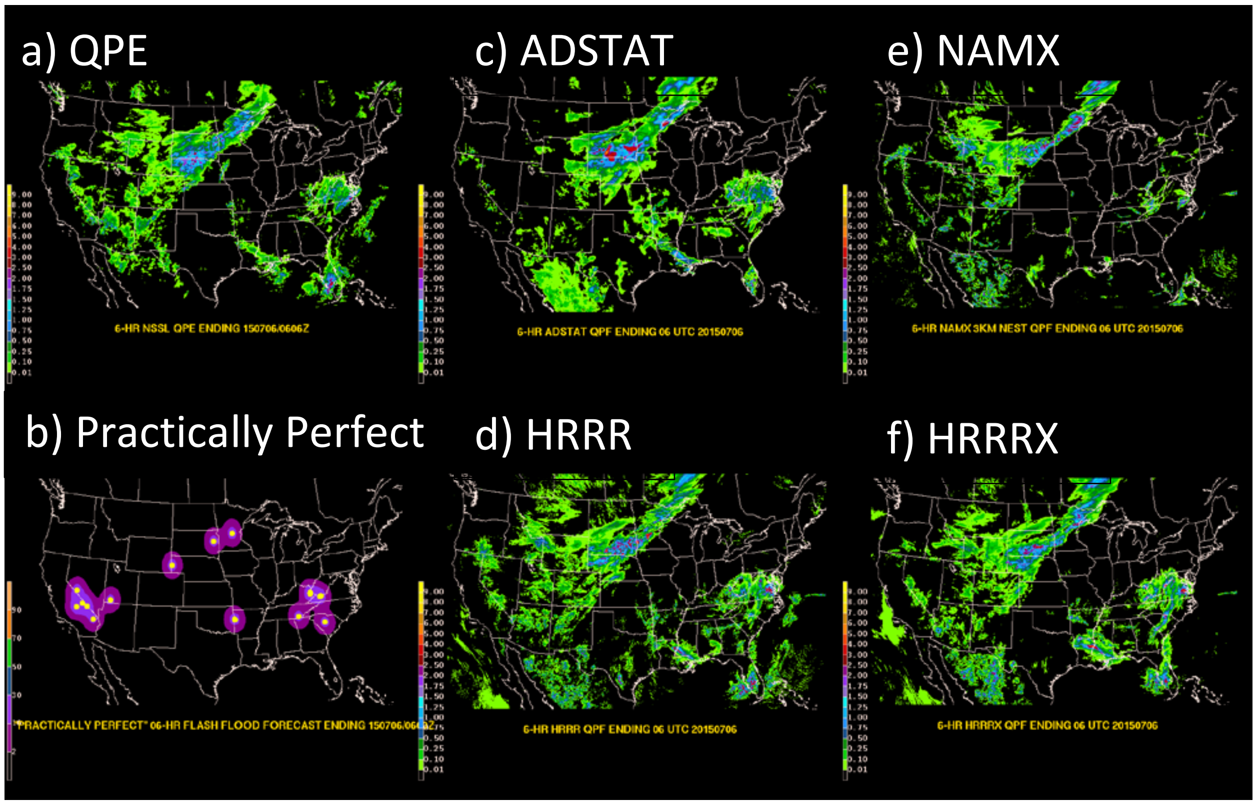


Figure 9*. (a) The corresponding radar-estimated 6 hour NSSL QPE and (b) Practically Perfect Forecast for the 6 hour precipitation forecasts ending 00 UTC July 7, 2015 from the (c) ADSTAT, (d) HRRR, (e) NAMX Nest, (f) HRRRX.*

The Experimental 3 km NAM (NAMX) received an average subjective rating of 2.13. It was noted by participants that the QPF maxima tend to both under- and over-perform on different occasions, and the areas of lighter precipitation were scoured out. This displacement of both location and magnitude of precipitation consistently misled forecasters when determining areas of most risk. EMC participants commented that the spin-up time of the 00 UTC NAMX was too long to capture convective precipitation for this short, 00-06 UTC time frame and contributed to the lack of accuracy when relying on the 00 UTC cycle run for the forecast.

The ADSTAT Model struggled to instill forecaster confidence due to its high bias of maximum QPF amounts in the east and conversely its low bias in the west where there is more complex terrain. Although generally the location of the axis of precipitation and areal coverage was adequate, the magnitude of the QPF did not typically verify. The ADSTAT model scored an average of 3.07 as it consistently generated a too little precipitation in the west and too much elsewhere in the CONUS.

As shown in Figure 9 above, in which the high-resolution models are compared against QPE (9a), the ADSTAT (9b) has reduced QPF magnitude in the west, the NAMX (9c) is too light over most of the CONUS, and the HRRR (9e) and HRRRX (9f) are not only similar to each other but closest the to QPE.

LOW-RESOLUTION ENSEMBLE PERFORMANCE

Low-resolution ensemble guidance was subjectively evaluated during experiment operations. Recognizing that some forecasters in the field do not have access to the high resolution models and that EMC was seeking feedback on the performance of the Parallel SREF (SREF-P), the lower-resolution ensembles were included in the FFaIR Experiment for evaluation.

The low-resolution ensemble guidance evaluated during the experiment included the SREF, the SREF-P and the WPC ENSBC model. The performance of each system's mean 6-hour QPF during the 18 – 00 UTC forecast period were subjectively rated as *very poor* (1), *poor* (2), *fair* (3), *good* (4) and *very good* (5). Participants were asked to score each ensemble mean QPF on the quality of guidance they gave the forecaster, independent of the other three models (e.g. not ranking the models from best to worst). The results are shown in Figure 10.

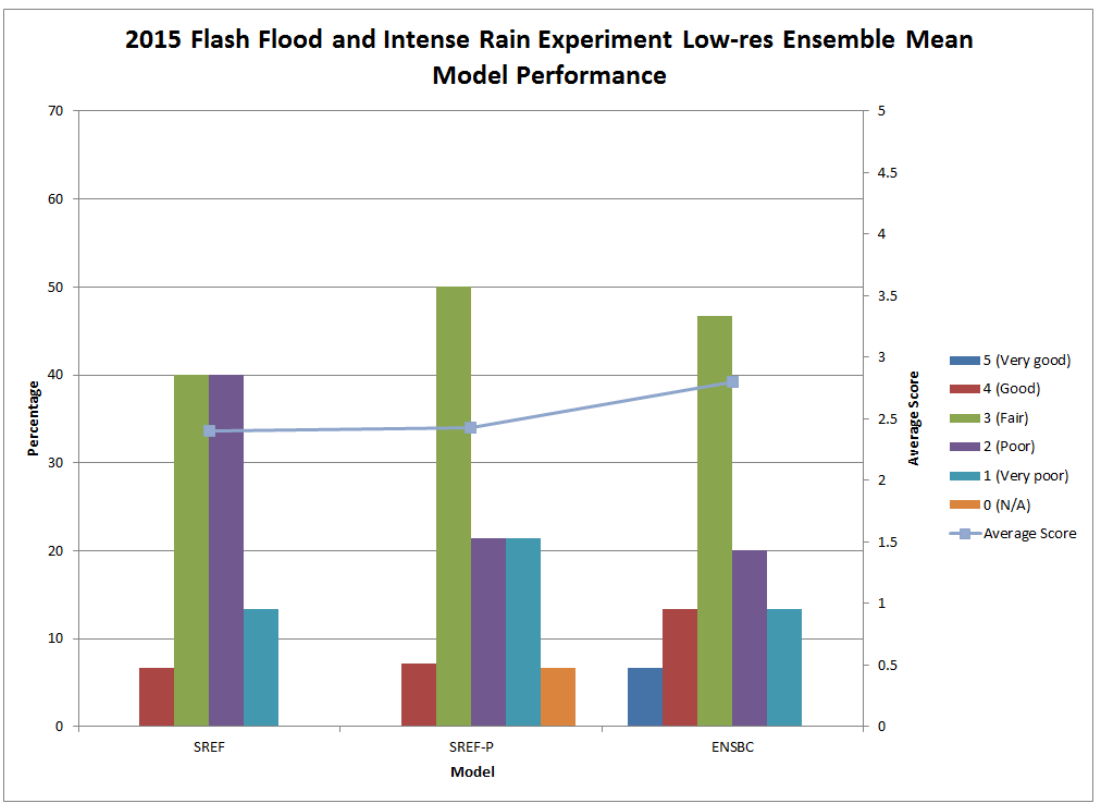


Figure 10*. Operational and experimental low-resolution ensemble model performance based on feedback from subjective model evaluations conducted during the 2015 Flash Flood and Intense Rainfall Experiment. Participants were asked to rate the performance of each model on a scale of 1 (very poor) to 5 (very good). The Average Score is a numerical average of all scores calculated over the experiment, not an average of the percentages shown.*

The average rating for the SREF over the experiment was 2.40 and the SREF-P was 2.43 out of 5 indicating that the participants found little difference in the performance of the two ensembles. Comments captured by the participants supported the rating as it was frequently noted that the SREF-P did slightly better with magnitude and location of the mean QPF.

The ENSBC (a bias-corrected ensemble which computes the final QPF as a weighted average of the ensemble mean and the mean of the deterministic runs of the SREFHR, NAM Nest, GFS, ECMWF, UKMET, and CMC models and run operationally at WPC) was rated 2.8 out of 5.

EXPERIMENTAL HIGH-RESOLUTION ENSEMBLE PERFORMANCE

The high-resolution ensemble guidance systems and corresponding experimental ensemble probabilistic forecast tools were also subjectively evaluated during experiment operations. When assessing the performance of each system's mean 6-hour QPF during the 18 – 00 UTC forecast period, the SPC-SSEO and the HREF were subjectively rated from 1 (*very poor*) to 5 (*very good*) with the results shown in Figure 11. Participants were asked to score each ensemble mean QPF on the quality of guidance they gave the forecaster and not rank the models against each other.

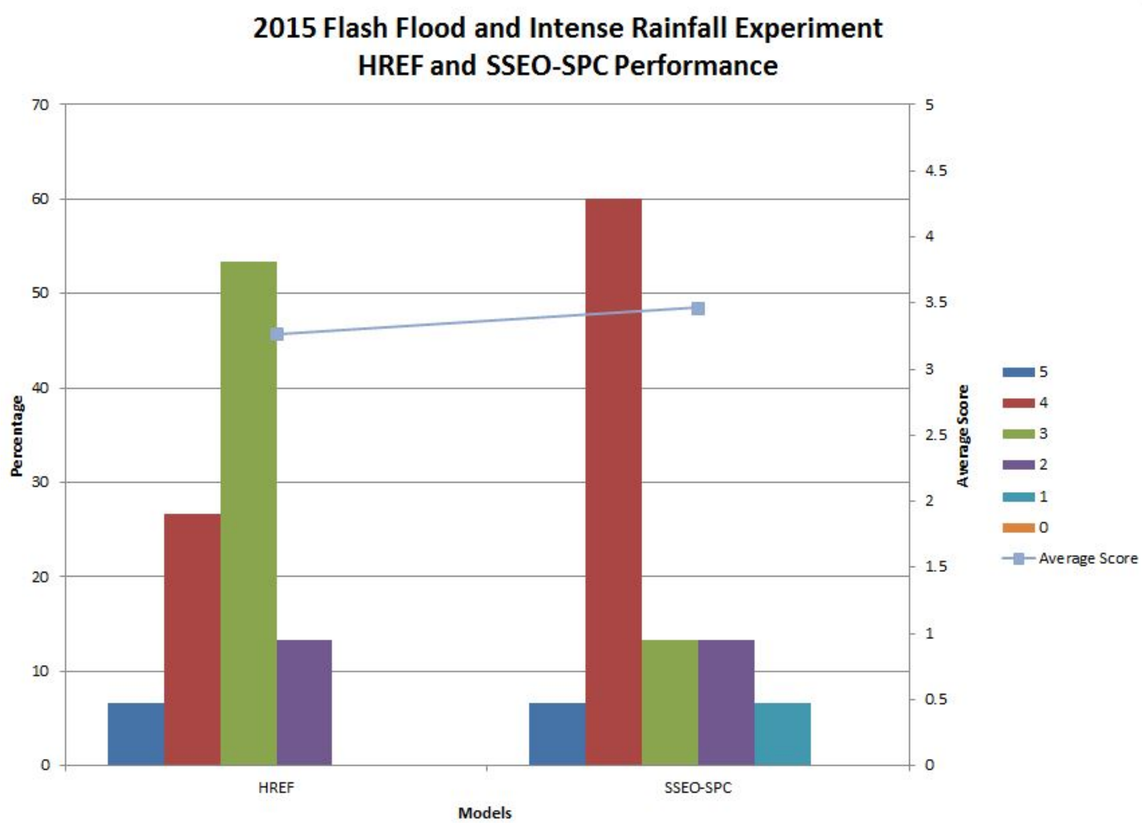


Figure 11 *Experimental ensemble model performance based on feedback from subjective model evaluations conducted during the 2015 Flash Flood and Intense Rainfall Experiment. Participants were asked to rate the performance of each model on a scale of 1 (very poor) to 5 (very good). The Average Score is a numerical average of all scores calculated over the experiment, not an average of the percentages shown.*

QPF

As shown in Figure 11, the SPC SSEO was evaluated as the superior high-resolution ensemble for predicting heavy rain events, much like it was during the 2014 FFaIR Experiment. With an overall rating of 3.47 out of 5, the SSEO again consistently provided the best mean QPF guidance to forecasters with both magnitude and placement of precipitation. It was noted by participants that at times the SSEO exhibited a wet bias, but overall performed very well.

The HREF, provided by EMC, was not far behind the SSEO with an average utility rating of 3.27 out of 5. Deterministic membership of both ensembles is quite similar (see Table 2 and Figure 1) and should be noted when evaluating its performance. Despite having slightly courser resolution (5 km as compared to the 4 km SSEO), the HREF performed well throughout the experiment and increased forecaster confidence in both magnitude and placement of QPF events. Participants commented that although the HREF did consistently well with location, magnitude of the precipitation was not as proficient as the SSEO.

Probabilistic Tools

In addition to the mean QPF, the forecasters and participants were provided with probability threshold tools derived from the SSEO and HREF ensembles. Throughout the experiment, the thresholds offered by the SSEO neighborhood probability of the mean QPF exceeding Flash Flood Guidance (FFG) over a 3-hour period received high praise during the 2015 FFaIR Experiment for its utility in helping forecasters identify risk areas for flash flooding. Figure 12 shows the SSEO forecast for 3-hour mean QPF on July 15, 2015. Figure 13 shows the neighborhood probability of the 3-hour mean QPF from the SSEO model exceeding 3-hour FFG.

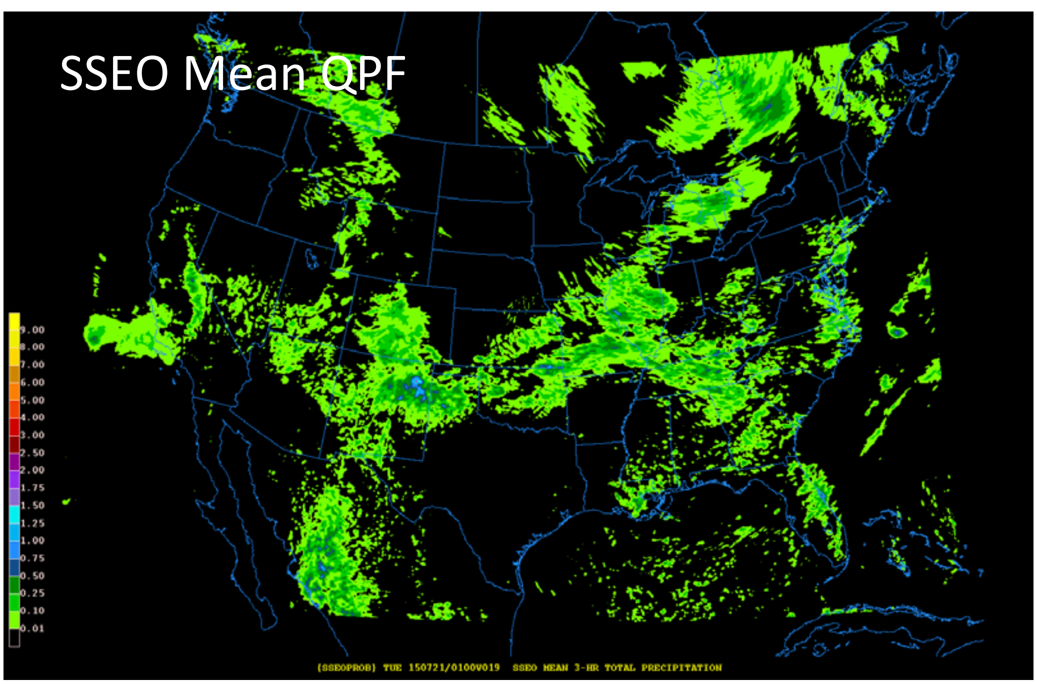


Figure 12. *The SSEO forecast for 3-hour mean QPF on July 15, 2015.*

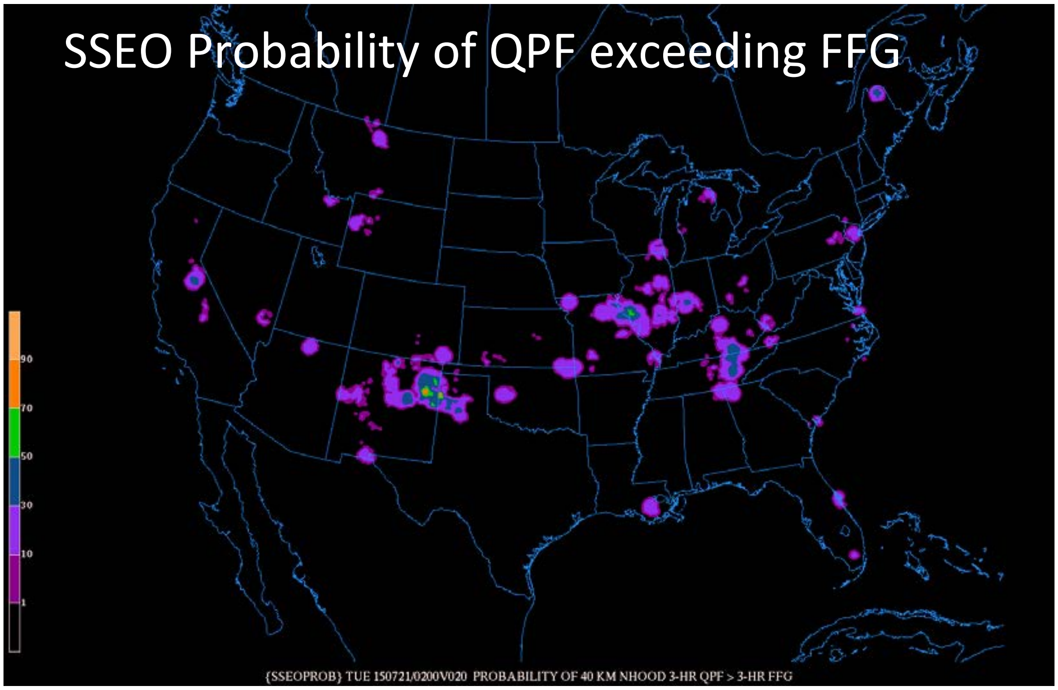


Figure 13. *The SSEO neighborhood probability of the same 3-hour mean QPF exceeding the 3-hour Flash Flood Guidance values on July 15, 2015.*

Jointly with the SSEO neighborhood probability tools for this example, the HREF offered probabilities at a grid point as opposed to a neighborhood approach to forecasters of the 3-hour mean QPF being greater than certain rainfall thresholds and is shown in Figure 14.

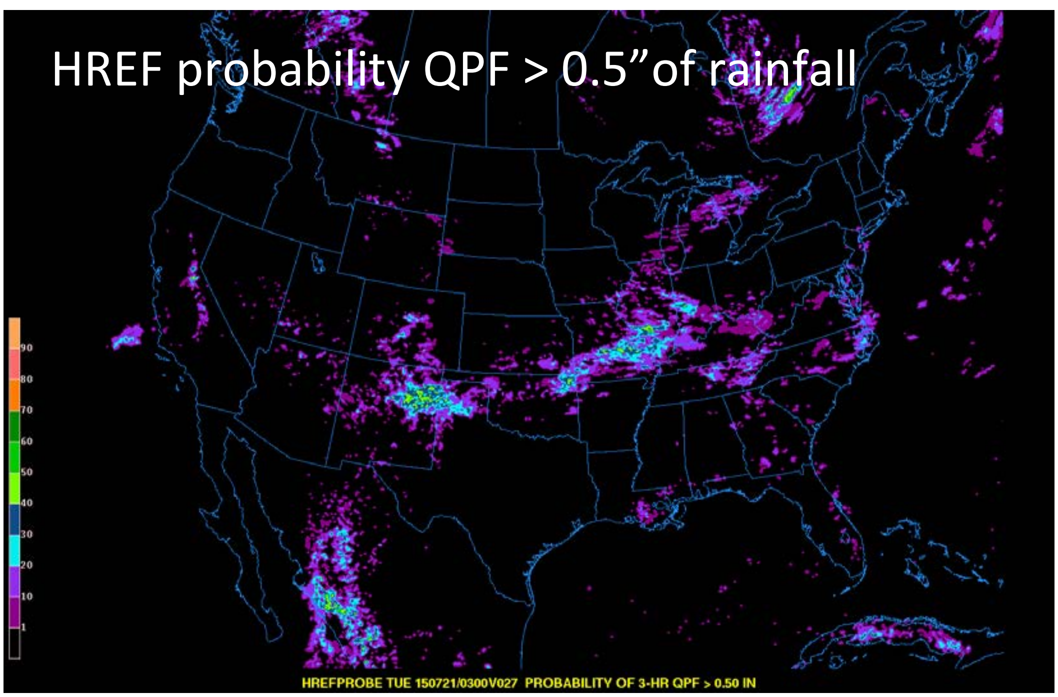


Figure 14. *The HREF probability of the same 3-hour mean QPF being greater than 0.5 inches of rainfall on July 15, 2015.*

The Flash Flood Guidance produced by our nation's River Forecast Centers is generally the accepted form of current ground water conditions used as an indicator of how much rainfall would be required to trigger flash flood conditions. In recent years, forecasters have also found utility in the Recurrence Intervals (RIs) defined by NOAA's Atlas-14[2] climatological database of USGS rain gage information conveying the frequency of rare precipitation events that may lead to flash flooding given certain conditions. When used in conjunction with the QPF produced by high-resolution, CAMs (and ensembles comprised of these models), climatological RIs can help forecasters identify emerging signals of likely flash flooding. For the 2015 FFaIR Experiment, the forecasters and participants were asked to evaluate the utility of a new guidance tool which compares SSEO neighborhood ensemble mean QPF and Recurrence Intervals of 2, 5, 10 and 100 years over 3- and 6-hourly time periods (see above section *Ensemble Forecast Tools*). Figure 15 shows an example of the display the forecaster's used to evaluate the 6-hour RIs.

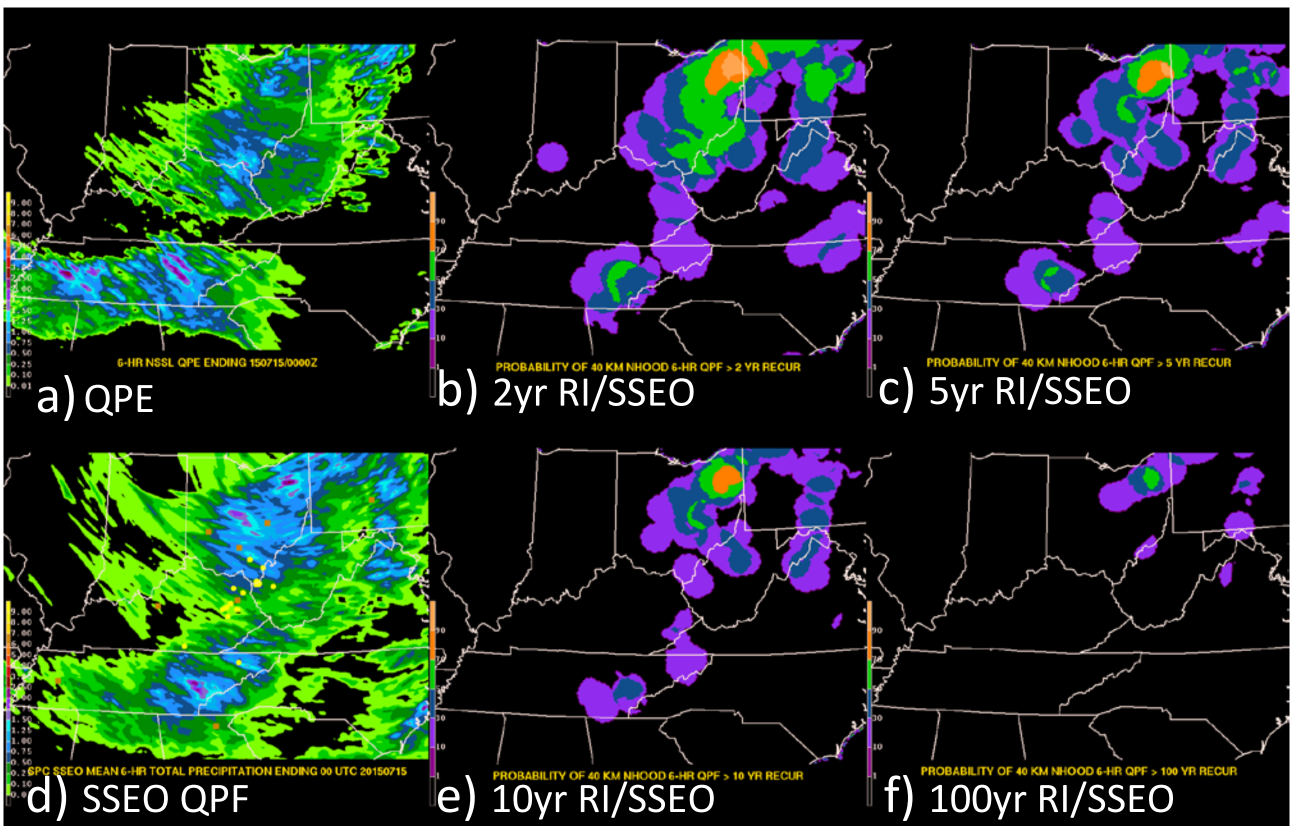


Figure 15. *FFaIR Participants used this display to subjectively rank the value of the Recurrence Intervals. (a) NSSL 6-hr QPE, (b) 2-year Recurrence Intervals as derived by the SSEO 6-hr mean QPF, (c) 5-year Recurrence Interval, (d) SSEO 6-hr mean QPF and flash flooding reports, (e) 10-year Recurrence Interval, (f) 100-year Recurrence Interval.*

It was determined in the evaluation discussions that the value and utility of the RIs were highly event-driven, and that it was difficult to pick a best interval overall. Statistically, the 100-year interval provided the least amount of useful guidance and received a subjective score of 2.47 out of 5. However, if there was a signal in the 100-year interval, it drew the attention of the forecaster at the possibility a very rare event occurring. Participants found the 2-year RI to be

[2] *For more information on NOAA's Atlas-14 see http://www.nws.noaa.gov/oh/hdsc/index.html*

most useful with applications and gave it a subjective score of 3.00 out of 5. The 5- and 10-year intervals, with scores of 2.87 and 2.60 out of 5 respectively, seemed to give the same basic guidance for each event and did not vary enough to differentiate value or valuable signals for flash flooding in slightly more than half of the cases.

Overwhelmingly, forecasters agreed that the RI climatology was valuable guidance in the determination of a potential highly-impactful flooding event. Additionally, the process of drawing neighborhood probabilities from high-resolution ensembles and RIs added value to the flash flood forecasting process and many participants expressed interest in acquiring these tools for scientific and operational applications in the future.

EXPLORING BEST GUIDANCE FOR A DAY 2 EXCESSIVE RAINFALL FORECAST

In addition to creating a Day 1 Excessive Rainfall Outlook using existing and experimental guidance, participants and forecasters were asked to evaluate longer-range guidance to create a Day 2 Excessive Rainfall Outlook. Participants were provided with the QPF from the operational NAM Nest, experimental NAM Nest (NAMX), SREF Mean, GFS CONUS and the GFS Hi-Res. The GFS CONUS has grid spacing of 1-degree and the GFS Hi-res grid spacing is 0.25 degrees.

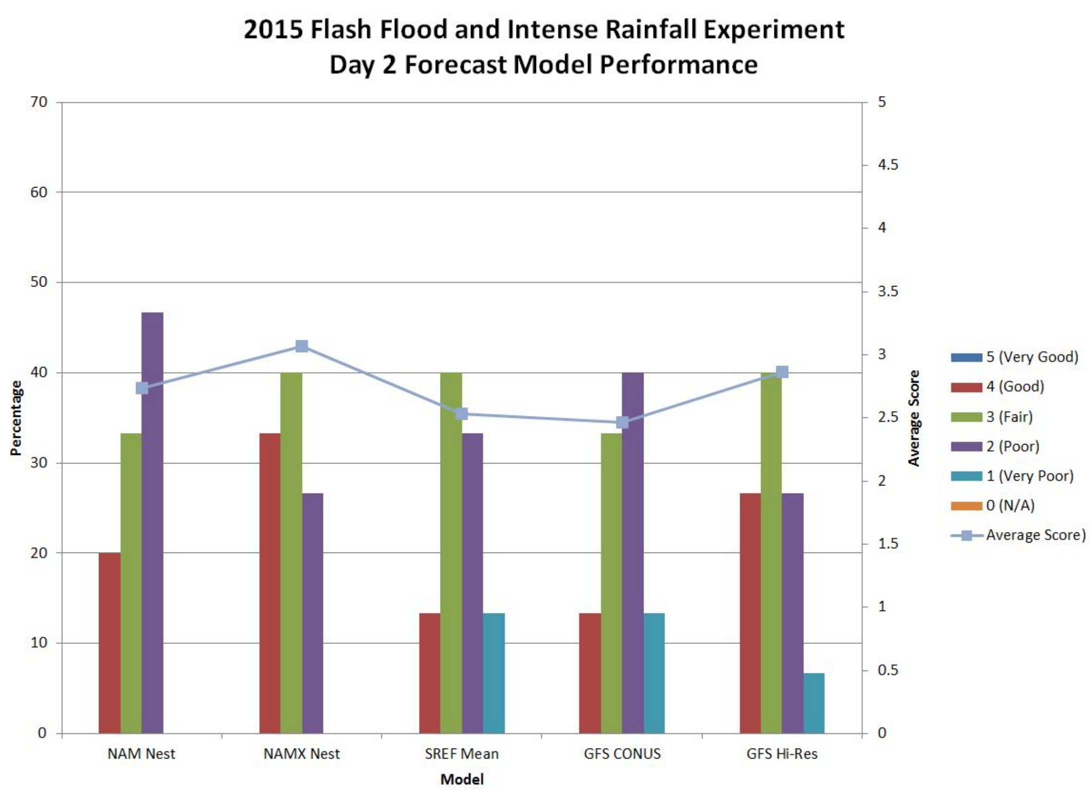


Figure 16. *Model evaluation for creating a Day 2 Excessive Rainfall forecast based on feedback from subjective feedback collected during the 2015 Flash Flood and Intense Rainfall Experiment. Participants were asked to rate the performance of each model on a scale of 1 (very poor) to 5 (very good). The Average Score is a numerical average of all scores calculated over the experiment, not an average of the percentages shown.*

As expected, Figure 16 shows the courser SREF and 1-degree GFS CONUS models received lower rankings on the prediction of QPF magnitude and location in the Day 2 time period mostly due to the lack of detail. The only difference between the GFS CONUS and GFS Hi-Res was resolution, as mentioned earlier. Therefore, the score improvement for the GFS Hi-Res was based entirely on the improved resolution. Recognizing that some forecasters in the field do not have access to the high resolution models, we included them in the FFaIR Experiment for evaluation.

Conversely to the shorter range evaluation (00-06 UTC time period), the experimental 3 km NAMX Nest received the highest ranking with an average of 3.07 out of 5 edging out its 4 km operational counter-part, the NAM Nest, which received an average ranking of 2.73 out of 5.

The high-resolution GFS model ranked second with a score of 2.87, out-performing the courser version, the 1-degree GFS CONUS, which received the lowest subjective ranking of 2.47 for a Day 2 flood risk outlook. And the SREF, also of a courser resolution, did slightly better than the GFS CONUS with a ranking of 2.53 out of 5.

TIME-LAGGED ENSEMBLE MEMBER PERFORMANCE

As part of this year's FFaIR Experiment, time-lagged ensemble members were subjectively evaluated against MRMS QPE and their latest-run counterparts and the results are shown in Figure 17. Participants compared the hi-res windows of the 12Z parallel NMMB and parallel ARW runs against the run 12 hours previous (00Z, -12 hour) and the time-lagged 06Z NAM Nest (-6 hour). The purpose of this exercise was to foster discussion and visual comparisons of the value of time-lagged deterministic members that often contribute to ensembles.

As shown in Figure 17, the most recent runs had higher average ratings than the time-lagged members, although the preferred runs varied with each event. On three of the 15 days, the time-lagged members received a higher subjective score than the more current runs. They were rated of equal score on a few days as well. But overall, the newer runs demonstrated more skill with magnitude and location of QPF than the older.

The 12Z parallel ARW was rated highest of the deterministic members, averaging a score of 3.00 out of 5 over the 2015 FFaIR experiment. It's time-lagged equivalent was rated 2.73. The 12Z parallel NMMB was rated a 2.80 with its time-lagged equivalent receiving a score of 2.60.

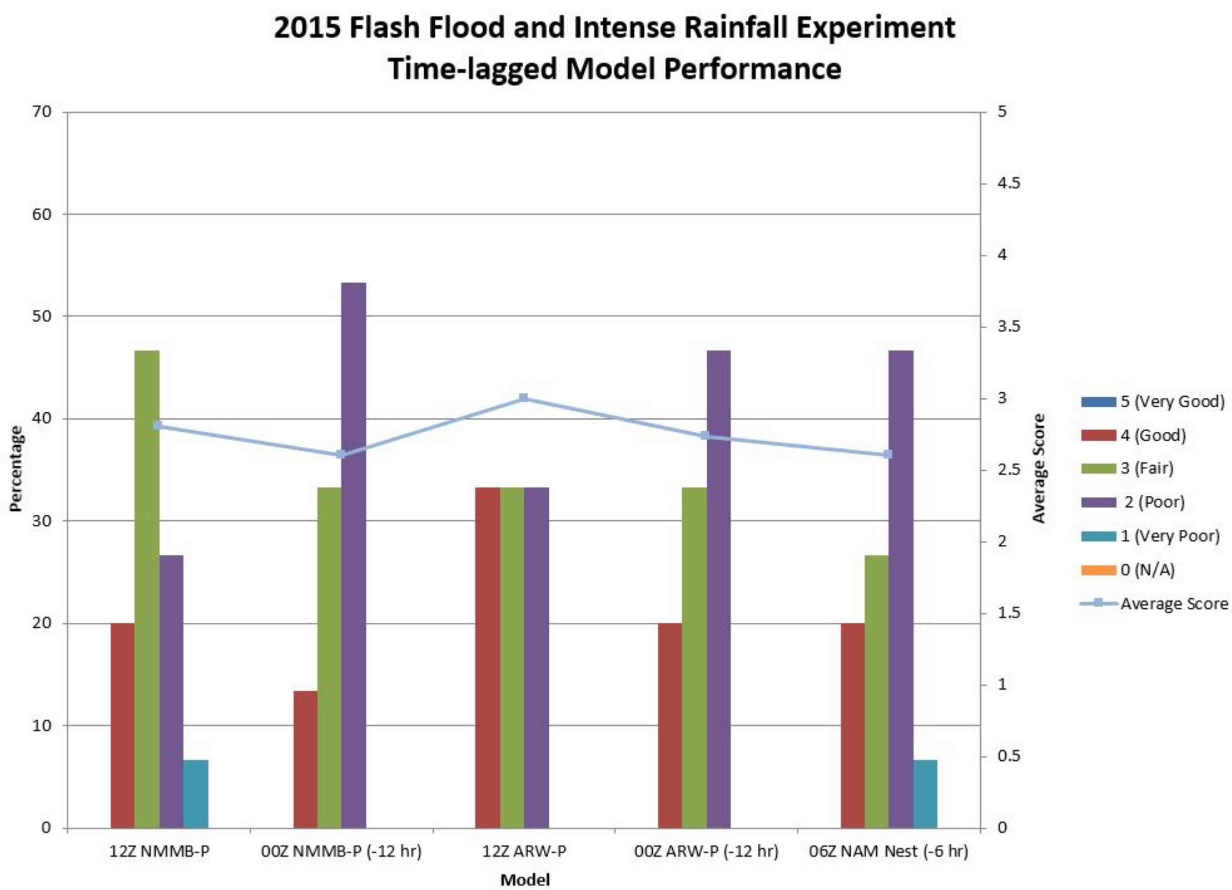


Figure 17*. Time-lagged member performance based on feedback from subjective model evaluations conducted during the 2015 Flash Flood and Intense Rainfall Experiment. Participants were asked to rate the performance of each model on a scale of 1 (very poor) to 5 (very good). The Average Score is a numerical average of all scores calculated over the experiment, not an average of the percentages shown.*

The 6-hour time-lagged 4 km operational NAM CONUS Nest received an average rating of 2.60 out of 5 during subjective evaluation. Participants often noted that the NAM CONUS Nest presented a high bias in areas of more significant QPF maxima, while exhibiting a low bias for lighter precipitation amounts. Most often the NAM CONUS Nest had a better handle on the location of the axes of precipitation but amounts were either over- or under-forecast which misled forecasters and created distrust in creating a forecast. Some of these differences are noted in an example in Figure 18.

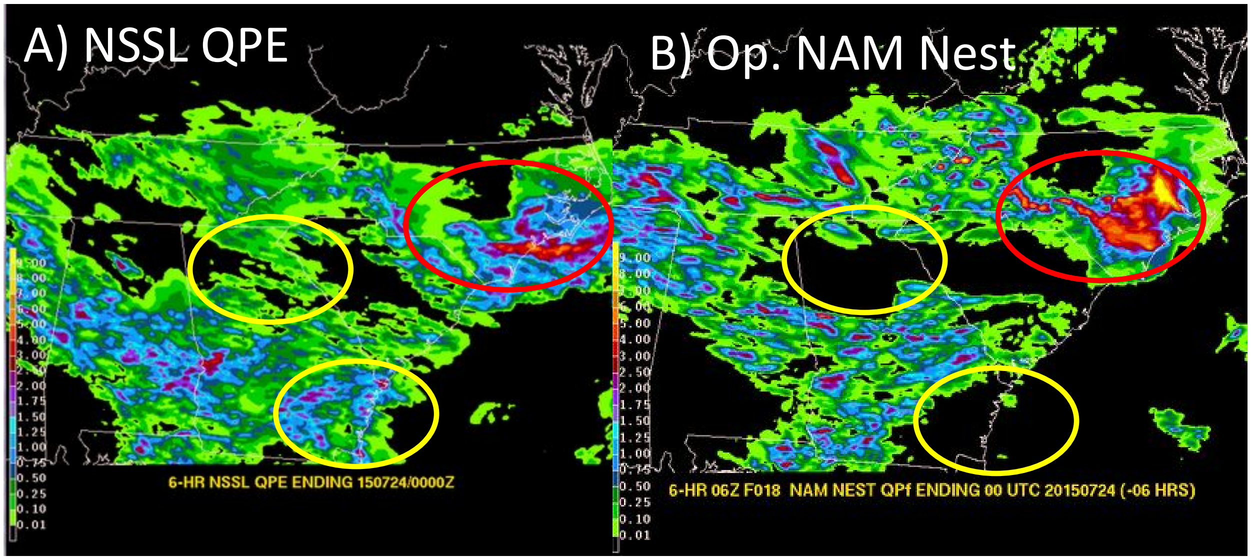


Figure 18. *A) An example of the NSSL QPE the B) time-lagged 4 km operational NAM Nest and during the 2015 FFaIR Experiment. The red oval indicates an area in which the NAM Nest the maximum QPF was too intense. The yellow ovals indicate lighter precipitation areas in which the NAM Nest was too light or non-existent.*

PERFORMANCE OF THE ATMOSPHERIC RIVER DETECTION TOOL

As additional guidance, the 2015 FFaIR Experiment featured a web-based tool provided by ESRL designed to highlight locations at which Integrated Vapor Transport (IVT) detected by the fields applied to the GFS model met the threshold of an atmospheric river[3]. The ARDT was offered each day to forecasters both as guidance, and to be subjectively evaluated as *very good (5), good (4)*, *fair (3)*, or *poor (2), or very poor (1)* for its value as a Day 1 and a Day 2 excessive rainfall forecasting tool.

As a Day 1 forecasting tool, the ARDT received a score of 2.47 out of 5. Participants commented that the view and resolution was not ideal for identifying small areas of high flooding risk. However, it was noted that having the deviation from climatology and the persistence of the forecasts from several runs helped increase forecaster confidence that moisture amounts were significant. The presentation of the data as a web tool was difficult for the participants to grasp (Figure 19) and had to be explained several times before it was understood.

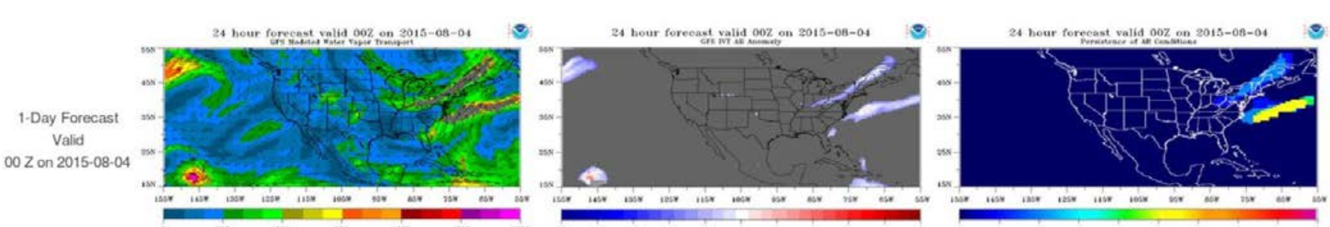


Figure 19*. A sample of the ARDT web tool provided by ESRL for the 2015 FFaIR Experiment, offering GFS IVT (left), climatological anomaly (center), and forecast persistence over three consecutive runs (right).*

[3] *ARDT AR Detection Criteria: IVT > 500 kg/m/s, Width < 1000 km, Length > 1500 km*

The participants preferred to use the ARDT as a Day 2 excessive rainfall forecasting tool not based on accuracy, but based on lead time. The concept of the wide geographic view and the resolution of the ARDT provided the forecasters with an indication for areas to watch in Day 2. However, when evaluating skill of the location and magnitude of moisture, the ARDT received an average score of 2.27 out of 5 which is lower than the Day 1 score.

Overall, the participants struggled to find value in utilizing the ARTD as a warm-weather short term excessive rainfall forecasting tool. Many indicated that it would be more valuable in the winter, or as medium-range guidance.

EXPERIMENTAL PROBABILISTIC FORECAST PERFORMANCE

Figure 20 shows the results of the subjective verification for the three forecasts made each day: the 15-12 UTC excessive rainfall outlook (ERO), the 18-00 UTC probabilistic flash flood forecast (PFF1), and the 00-06 UTC probabilistic flash flood forecast (PFF2). Each of the forecasts were visually compared to NSSL's MRMS QPE, various flash flood observations (FFWs, areas of QPE-to-FFG exceedance, LSRs, USGS stream flow data, mPING reports, etc.), as well as plots of the 'practically perfect' forecast, and then subjectively rated as *very good (5), good (4)*, *fair (3)*, or *poor (2), or very poor (1)*.

The participants in the 2015 FFaIR Experiment were very satisfied with the resulting probabilistic Day 1 forecasts as 83% were rated as *fair* or *good*. This increased slightly from the 80% rating over the 2014 FFaIR Experiment (note that the 2014 ranking was *poor (1), fair (2) and good (3*)). Of the 12 ERO forecasts evaluated, one did receive a *very good* rating (the highest option on the scale). The evaluators noted on several occasions that the observational database of local storm reports and flash flood warnings may not be reliably populated for various reasons, but considered forecast areas drawn that included 75% exceedance of Flash Flood Guidance as well-captured.

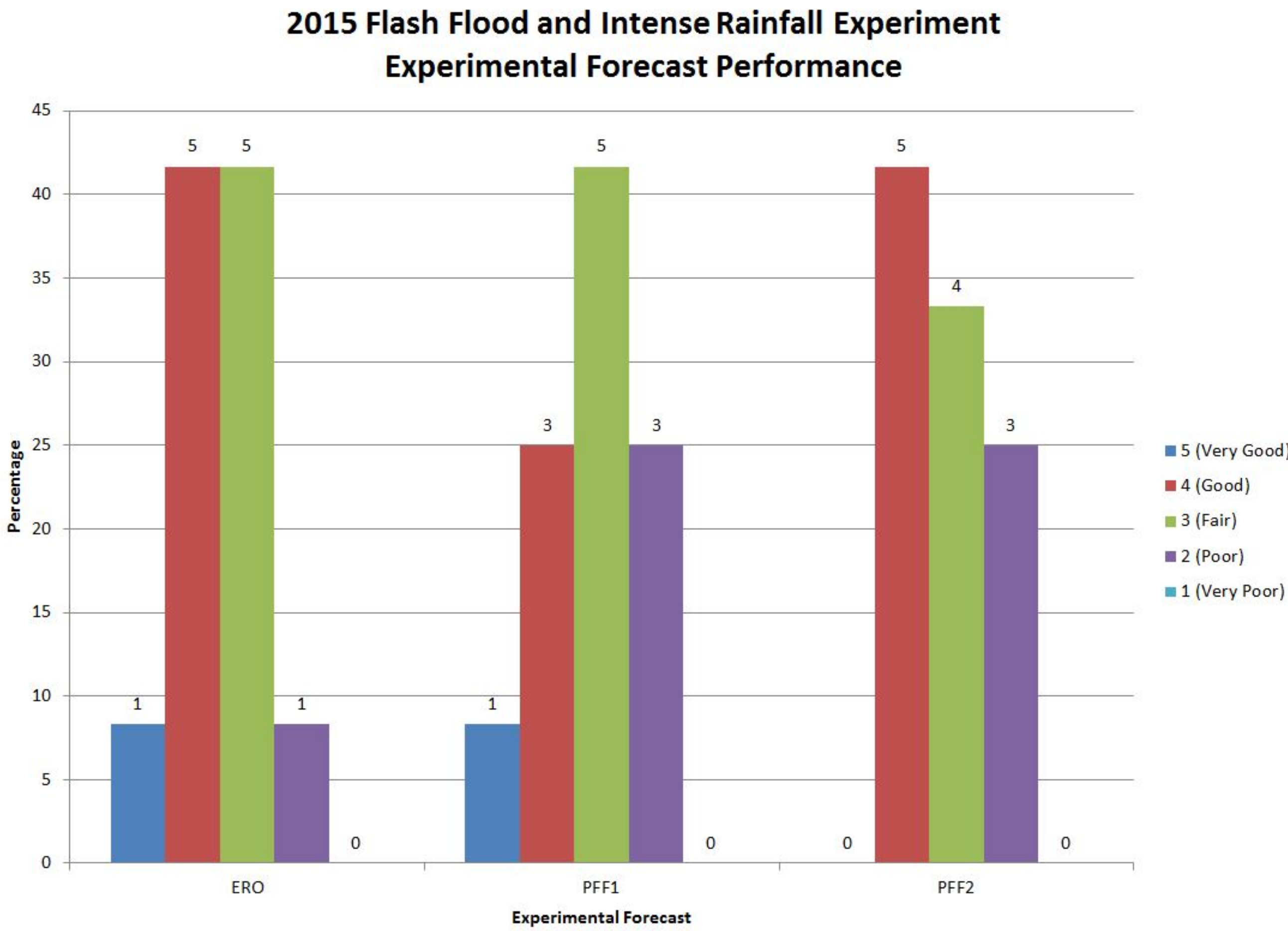


***Figure 20**. The rating for each of the three experimental forecasts completed in FFaIR: the 15-12 UTC probabilistic excessive rainfall outlook (ERO), the 18-00 UTC probabilistic flash flood forecast (PFF1) and the 00-06 UTC probabilistic flash flood outlook (PFF2).*

The average rating of the PFF1 and PFF2 were equal at 2.53 out of 5. However, upon further examination of the frequency of "*Good*" ratings, the 18-00 UTC 6-hour probabilistic forecasts (PFF1) received a lower performance rating than the longer-range 00-06 UTC forecasts (PFF2). The PFF1 was rated 4 out of 5 in just 25% of cases, but the PFF2 was rated 4 out of 5 in 41.7% of cases. The PFF1 received one "*Very Good*" rating (5 out of 5) whereas the PFF2 did not. The participants noted that the 0600 UTC and 1200 UTC guidance available when making the PFF2 (0000-0600 UTC) forecast demonstrated more skill in capturing QPF at this time range than the models initialized at 0000 UTC that were predominately used for the PFF1 forecast (1800-0000 UTC).

EVALUATING THE STATISTICAL SKILL OF THE FORECAST CONTOURS

In the experimental ERO in FFaIR, the forecast is defined as the probability of flash flooding occurring within 40 km of a point. In an effort to calibrate the experimental contours and evaluate their skill, the average percent of the area of the contours within 40 km of a flood observation was calculated for each contour. To calculate statistics on the forecasts created during the 2014 and 2015 FFaIR Experiments, the forecast vector graphics files were converted in to a gridded format using the General Meteorology Package (GEMPAK) program. GEMPAK and Perl scripts were then utilized to calculate the skill of each contour in predicting where flash flooding would occur based on its capture of a flood observation within 40 km. The averages from both years' experiments are shown in Figure 21.

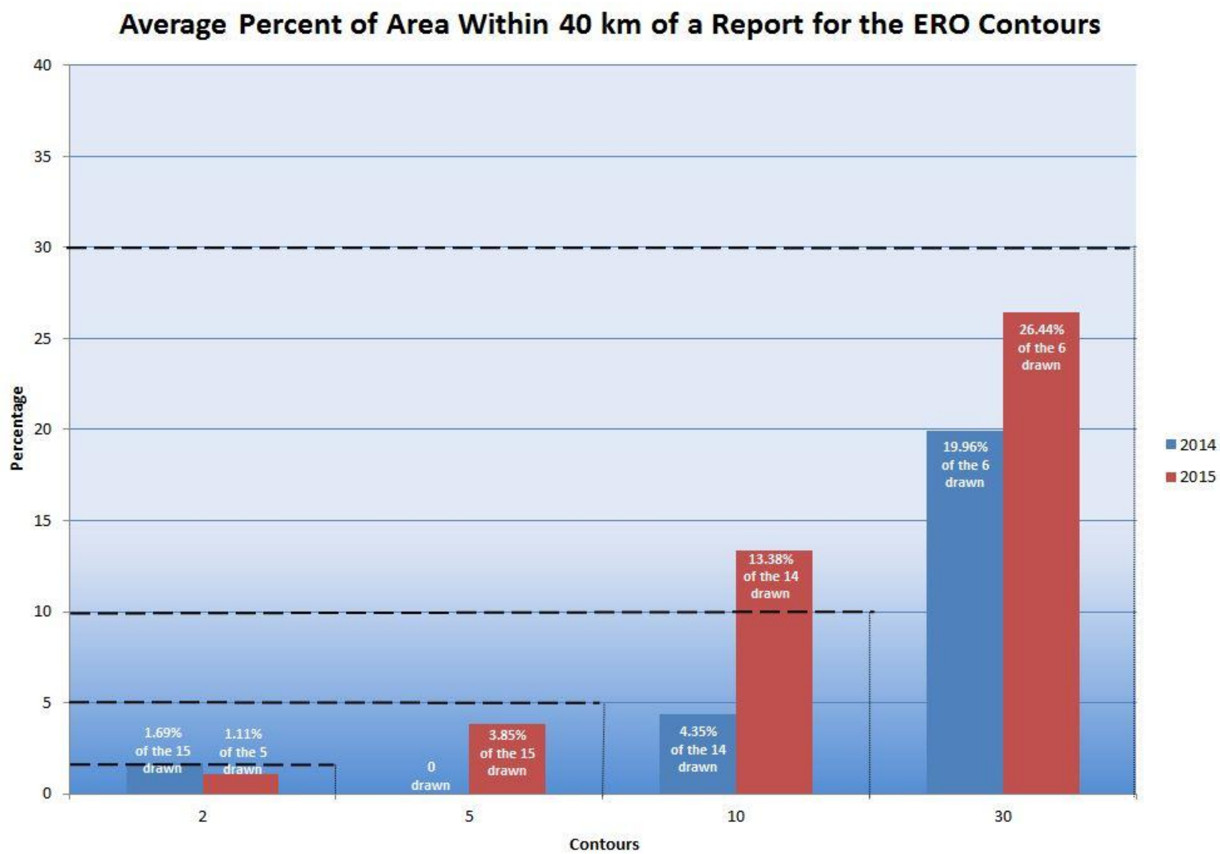


Figure 21. *A comparison of the average percent of area of the ERO contours drawn from the 2014 (blue) and 2015 (red) experiments falling within 40 km of a reported observation of flooding or flash flooding. The horizontal dashed lines indicate the percent of each contour and the value the forecasts should equal if they showed perfect skill/calibration. Note that the 5% contour was not available during the 2014 FFaIR Experiment.*

As stated earlier in this document, the experimental ERO features low probability thresholds less than 5%, whereas the operational product currently provides "SEE TEXT" on the map for areas below a 5% probability with a subsequent text discussion for those areas. For the experimental ERO, the 2014 experiment featured 2%, 10%, and 30% forecast probabilities for flooding with the 2015 experiment introducing an additional 5% contour. Ideally with perfect skill and calibration, the average percent of the contour areas within 40 km of the report would be equal to the value of the probability contour.

In 2014 (no 5% contour), 1.69% of the drawn **2% contour** areas (15 forecasts) were within 40 km of a report. In 2015, 1.11% of the drawn 2% contour areas (15 forecasts) were within 40 km of a report. Some of the decrease from 2014 to 2015 could possibly be attributed to the addition of the 5% contour noting that reports that fell within a 5% contour would not be verified for the 2015 2% contour. On average, the 2015 2% contour performed with .58% lesser skill, but was still within .89% of 2%. In the 2015 FFaIR experiment (15 forecasts), 3.85% of the drawn **5% contour** area was within 40 km of a report on average, which is 1.15% below perfect calibration level.

The 2 and 5% contours were explored as an alternative to the current "<5% - SEE TEXT" label in the current operational WPC Excessive Rainfall Outlook. In particular, the 2% threshold received overwhelming support from participants with 67% in favor of its addition when polled. Although experimental testing and verification accumulated over an entire warm season is still needed and a now proposed pilot project for the 2016 warm season is planned, the 2013, 2014 and 2015 FFaIR Experiments have shown increased skill and value of the experimental neighborhood probabilistic ERO with lower-threshold contours (2% and 5%) to convey flood risk. Participants agreed alignment among the NWS National Centers was of highest importance when conveying impacts of significant weather events and that the 2% excessive rainfall contour is equivalent to the SPC "General Thunder."

In both 2014 and 2015, 14 out of the 15 experimental ERO forecasts utilized the **10% contour**. In the 2014 experiment, the average percent of the drawn 10% contour area within 40 km of a report was 4.35%, which is significantly below the 10% level. However, in 2015, the average percent within 40 km of a report was actually higher than 10%, at 13.38%. Despite being 3.38% over the 10% calibration in 2015, overall forecaster's showed a 2.3% improvement in skill when considering in 2014 that the 10% contour was 5.65% off of the calibration. Comments collected from participants and surveys indicate that the addition of the 5% contour for lower confidence areas may have led to more skillful 10% forecasts as it allowed forecasters to use the 10% contour in areas they had higher confidence. Improved convection allowing models and better understanding on how to utilize them also contributed to the skill increase.

The highest probabilistic forecast contour forecasters could utilize in the creation of the ERO was the **30% contour**. In both 2014 and 2015, 6 out of the 15 forecasts utilized the 30% contour. In 2014, the average percent of the 30% area within 40 km of a report was 19.96%. In 2015, the average percent of the 30% area within 40 km of a report was 26.44%, 6.48% higher than the previous year and 3.56% below the perfect calibration level of 30%. With all aspects very similar between the two years, the skill of the 30% contour increased the most (6.48%) over the two experiments.

SUMMARY AND OPERATIONAL IMPACTS

The third annual Flash Flood and Intense Rainfall Experiment was conducted from July 6 – 24, 2015 at the NOAA Center for Weather and Climate Prediction in College Park, MD. Over the course of the three week experiment, 24 forecasters, researchers, and model developers used a variety of innovative high resolution model and ensemble output to issue a series of experimental probabilistic flash flood forecasts. This year's experiment expanded upon the utility of CAM guidance in the flash flood forecast process and ensemble tools. A number of the experiment's findings are relevant to operational forecasters focused on the flash flood threat:

- In order to improve flash flood forecasts, **the NWS must work toward an agency-wide approach to the flash flood forecast problem, ranging from definition, to warning practices, to reporting requirements.** Subjective opinion and geographic inconsistencies with regard to the definition of a flash flood, combined with inconsistent reporting practices from WFO to WFO, leads to a severe lack of database verification when it comes to identifying where flash flooding has occurred.
- **Convection-allowing ensembles such as the SSEO and HREF are providing high value to the flash flood forecast process, and even more so when combined with hydrologic data.** The neighborhood probability of QPF > FFG again proved to be a useful forecast tool, and this year the neighborhood probability of QPF > Recurrence Intervals also gained high praise for its contribution to the flash flood forecast process. The probabilistic fields of QPF greater than pre-determined rainfall thresholds from the HREF guided forecasters to more confidence when building a probabilistic forecast. More high resolution ensembles and tools are desired.
- While high resolution models have proven to be a useful forecast tool east of the Rockies, **flash flood forecasting in the inter-mountain west represents a more significant challenge.** Improving forecast skill in this region will require both improved model guidance and a focused effort to build forecaster understanding of the factors that govern flash flooding in the complex topography in the western U.S.
- **WPC will now pursue changes to the current Excessive Rainfall Outlook product** to highlight broader areas at risk of flash flooding. During the experiment, the forecast teams were consistently able to successfully distinguish between areas with a risk for flash flooding and those without. Tested for 3 years now, the neighborhood probabilistic ERO has proven skill and will be tested experimentally in a parallel operational environment over the next warm season.
- The daily forecasts briefings provided useful information to the HMT-Hydro experimental flash flood watch and warning activities. Export and exchange of GIS files better enabled collaboration among the groups despite the use of different software platforms (N-AWIPS versus AWIPS II). However, **continued effort needs to be made to overcome platform inconsistencies and maximize cross-testbed interactions.**

ACKNOWLEDGEMENTS

The Flash Flood and Intense Rainfall Experiment would not have been possible without the dedication of many people including Sarah Perfater (HMT-WPC), Ben Albright (HMT-WPC), Mike Bodner (WPC), Mark Klein (WPC), Scott Jacobs (WPC/NCO), Brian Cosgrove (NWC), and Dave Novak (WPC). Thank you to Matt Pyle (EMC) for proving the experimental HREF and Eric Rogers (EMC) providing the Experimental NAM 3 km (NAMX). Special thanks to JJ Gourley, Steve Martinaitas and Zac Flemming of HMT-Hydro in Norman, OK for collaboration on our flash flood workshops.

REFERENCES

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.

Hitchens, N.M., H.E. Brooks, and M.P. Kay, 2013: Objective limits on forecasting skill of rare events. *Wea. Forecasting,* **28**, 525-534.

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.

Kitzmiller, D., W. Wu, S. Wu, and D. Miller, 2011: Development of a short-range probabilistic precipitation forecast algorithm based on radar and numerical prediction model input. *Preprints, 35th Conference on Radar Meteorology and Hydrology*, **145**.

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., **13**, 3351–3372, doi:10.1175/2009MWR2924.1.

APPENDIX A

Participants

Week	WPC Forecaster	WFO/RFC	Research/Academia	EMC
July 6 – 10	Rich Otto	Stanley Czyzyk (VEF) Maureen Hastings (CAR) Jon Zeitler (EWX)	Curtis Alexander (ESRL) Yu Zhang (NWC)	Matt Pyle
July 13–17	Brendon Rubin-Oster	Brian Boyd (IKN) Link Crawford (OHRFC) Steve Willington (UKMET) Bryan Smith (SPC)	Kelly Mahoney (ESRL) Charlie Pilling (UKMET)	Jacob Carley, Eric Aligo
July 20-24	Patrick Burke	Troy Lindquist (BOI) Cathy Zapotocny (OAX) Jeff Colton (GJT)	Keith Brewster (OU) Isadora Jankov (ESRL)	Brad Ferrier Jeff McQueen

APPENDIX B

Daily Schedule

8:00 am – 10:00 am	Excessive rainfall outlook, valid 15 – 12 UTC
10:00 am -10:15 am	Prepare discussion/PPT
10:15 am – 10:30 am	Break
10:30 am – 11:45am	Subjective model evaluation
11:45 am – 12:45pm	Lunch
12:45 pm – 1:45 pm	Probabilistic flash flood forecast, valid 18 – 00 UTC; DL discussion
1:45 pm - 2:00 pm	Finish discussion/PPT
2:00 pm – 2:45 pm	HMT-Norman forecast briefing
2:45 pm - 3:00 pm	Break
3:00 pm – 3:45 pm	Probabilistic flash flood forecast, valid 00 – 06 UTC
3:45 pm - 4:00 pm	Update discussion