WPC-HMT 2019 Flash Flood and Intense Rainfall (FFaIR) Experiment Executive Summary

Sarah Trojniak, Ben Albright, Michael Erickson, Jim Nelson, Mark Klein *October 2019*

Introduction

The 2019 FFaIR experiment focused on forecasting flooding and intense rainfall across the continental United States (CONUS) in the Day 1 time period. The experiment was run from June 17, 2019 to July 19, 2019 by the Hydrometeorology Testbed at WPC (HMT-WPC). The HMT team partnered with National Weather Service (NWS) meteorologists, hydrologists, and the development and research communities to help identify the usefulness of various experimental model guidance and tools in the Day 1 time period. The design of the experiment centered around mimicking operations at WPC on the Day 1 Quantitative Precipitation Forecast (QPF) and MetWatch Desks, which issue the Excessive Rainfall Outlooks (ERO) and Mesoscale Precipitation Discussions (MPD). Participants were tasked with producing four experimental forecasts each day comprising of one ERO and three MPD-like products. These were created and issued almost exclusively off of the analysis of the experimental data that was evaluated during FFaIR, which included high resolution guidance from deterministic and ensemble models, hydrological guidance and remote sensing tools. During this process, the participants were encouraged to have lively discussions about not only where the threat of intense rainfall and flooding would occur but also about the experimental tools they were utilizing. The goals of the 2019 FFaIR experiment were to:

- Evaluate the usefulness of operational and experimental products from high resolution convective-allowing deterministic and ensemble models for forecasting near-term flash flood events.
- Evaluate, subjectively and objectively, the utility of the FV3-Stand-Alone Regional (SAR) in comparison to the nested version of the FV3-GFS (FV3-Nest), specifically at later forecast times, to determine if the FV3-SAR is a viable alternative to the FV3-Nest.
- Access the forecasters' understanding of ensemble tools such as probability matched mean (PMM) and local probability matched mean (LPMM) and identify their usefulness in the forecast process.
- Identify ways to incorporate hydrological model guidance into the decision making process for flash flooding guidance issuance.
- Evaluate, subjectively and objectively, the ability of the CSU-MLP "First Guess Field" to predict the Marginal, Slight, Moderate, and High Risks for the Day 1 ERO.
- Identify ways to incorporate advanced remote-sensing and difference fields into the flash flood forecasting process.

Activities

The forecast process was guided by the expertise of a WPC forecaster, with collaboration among the participants and analysis of experimental guidance that drove the decision making. The experimental forecasts issued were all probabilistic products assessing the potential for flash flooding within 40 km of a point and varied on the timescale in which they were valid. The first product issued was the Day 1 Experimental ERO valid from 1500 UTC to 1200 UTC and was valid over the CONUS, with probabilistic contours of 5% (Marginal), 10% (Slight), 20% (Moderate), and 50% (High). Two 6 h forecasts were also issued, referred to as the Probability of Flash Flooding (PFF#) forecasts. The PFF1 and PFF2 were valid for six hours from 1800 UTC to 0000 UTC and 0000 UTC to 0600 UTC, respectively. A PFF3 was also issued that was valid for 3 h, from 2100 UTC to 0000 UTC. All three of these products were regional products and had probabilistic contours similar to the ERO, though a Marginal Risk was not included in the forecast. In addition to issuing experimental forecasts the participants also subjectively evaluated the experimental guidance provided by our partners as well as their own forecasts.

Summary and Research-to-Operations Recommendations

Table 1: *Research to Operations Transition Metrics for the 2019 FFaIR Experiment.*

Major Models/ Products Evaluated	Recommended for transition to operations	Recommended for further development and testing	Rejected for further testing	Funding Source
FV3-Nest		x		EMC
FV3-SAR		x		EMC
FV3-GSD-SAR		x		ESRL/GSD
HRRRv4	x*			ESRL/GSD
HREFv3		x		EMC
HRRRE		x		ESRL/GSD (OWAQ)
SSEF		x		OU/CAPS (OWAQ)
CSU-Machine Learned Day 1 ERO First-Guess Fields		x		CSU (JTTI)
CIRA-CSU Merged TPW v1.0		x		CSU/CIRA (OWAQ)
CIRA-CSU HRRR ALPW Difference Fields		x		CSU/CIRA (OWAQ)
Total	**1**	**9**	**0**	

- Only the **HRRRv4** is being recommended for transitions to operation, albeit conditionally. Before it can be implemented, retrospective runs and testing of the model must be done to ensure the bug fix in mid-July did not negatively impact model performance.
- Both the **FV3-Nest** and the **FV3-SAR** from EMC performed well based on the subjective scores given by the participants. The FV3-Nest narrowly outperformed the FV3-SAR in both the 24 h and 6 h QPF subjective verification. Despite the high scores from the subjective portion of the experiment, when evaluated using objective metrics, the

FV3-Nest and FV3-SAR performance was not as impressive. Most notable was the high wet bias seen in both models, which at the 1 inch threshold was higher than the wet bias seen in the NAM-Nest. Therefore, until the wet bias is addressed the models should not be transitioned into operations.

- The **FV3-SAR-GSD** from ESRL/GSD differed from the EMC FV3-SAR in the physics suite that was used. Although the model was only available at a limited capacity during the FFaIR experiment, objective and subjective results indicate that FV3-SAR-GSD underperforms in comparison to the FV3-SAR from EMC.

- The **HREFv3** was the best performing ensemble both objectively and subjectively and would have been recommended for transition into operations. However, the models that the HREF are comprised of are not all frozen (e.g. the model dynamics are still being tested and altered). The still fluid nature of the models that form the HREFv3 could impact model performance in the future. Therefore, until the models are frozen, the HREFv3 must continue to be evaluated.

- Per recommendation from the 2018 FFaIR Experiment, partners providing ensemble data to be evaluated in 2019 FFaIR were asked to produce a **LPMM** product. All three of the ensembles assessed, the **HREFv3**, **HRRRE**, and the **SSEF,** supplied this product. Like last year, overall the product was preferred by the participants over the PMM. However objective results suggest that LPMM performance is somewhat dependent on how the LPMM is calculated. Analysis of ensemble biases showed that the LPMM products from the SSEF and HREFv3 had a lower bias their respective PMM products while the HRRRE LPMM had a higher bias compared to its PMM. Therefore the FFaIR team suggests that the method used by the SSEF and HREF be the method used for LPMM calculations and be the product be produced by all operational ensembles.

- The **CSU-MLP First Guess Day 1 ERO** products were well received by the participants and participants felt that the guidance provided a great “starting spot” for creating the experimental FFaIR ERO. However, there are a few regions across the CONUS that the products do not appear to be well-calibrated for, such as the Northern Rockies and Northern Plains in the NSSL ERO product. Further refinement of the how the flooding risk is determined should be done in these regions. These refinement will likely help with the calibration of the marginal risk, which was generally too large spatially for both the NSSL and GEFS products.

- The **CSU-CIRA Merged TPW v1.0** appears to have improved upon the operational BTPW but needs to be further verified against observations to determine if the new method of deriving the TPW from multiple satellites and combining them is accurate.

- The utility of **CSU-CIRA HRRR ALPW Difference** product was not fully understood by the FFaIR participants and they felt the product would not aid them in their forecasting process. They did however feel that the product would be highly beneficial to model developers.