2024 Flash Flood and Intense Rainfall (FFaIR) *Operations Plan*

June 10 - August 2, 2024
Weather Prediction Center (WPC)
Hydrometeorology Testbed (HMT)

Sarah Trojniak[1], **James Correia Jr.**[1], **W. Massey Bartolini**[1], and **Tomer Burg**[1]

[1]CIRES-CIESDRS CU Boulder, NOAA/NWS/WPC/HMT

Contents

1 Introduction

The Hydrometeorology Testbed (HMT) at the Weather Prediction Center (WPC) will once again be hosting the Flash Flood and Intense Rainfall (FFaIR) Experiment this summer. It is a NWS Testbed Experiment that focuses on the challenges associated with forecasting flash flooding and heavy rainfall. Like other testbeds, it beings together a variety of expertise across the weather enterprise, including forecasters, model developers, hydrologists, and academia. Running in a pseudo-operational setting and using HMT-developed web tools, participants are challenged to create heavy rainfall forecasts using experimental data and tools. These are also evaluated subjectively by participants.

This year, FFaIR will comprise of 5 week-long sessions spanning June to the beginning of August. These will be a mix of virtual and hybrid, with the in-person attendees coming to the National Center for Weather and Climate Prediction (NCWCP) in College Park, MD. Virtual attendance will expand across the country including Alaska and Hawaii and include some territories like Guam and American Samoa. The dates FFaIR is in session are as follows:

Week 1: June 10 - 14 (virtual)
Week 2: June 24 - 28 (virtual)
Week 3: July 8 - 12 (hybrid)
Week 4: July 22 - 26 (virtual)
Week 5: July 29 - Aug 2 (hybrid)

This year's experiment is designed to evaluate the Rapid Refresh Forecast System (RRFS) as it pertains to forecasting warm season precipitation. To do this, the experiment will run in a data denial format by splitting the participants into two Groups, one that only has access to the current operational models and the other only having access to the RRFS suite. This is being done to simulate the planned retirement of nearly all operational convection allowing models (CAMs), which are slated to be replaced by the RRFS.

2 Experiment Operations

Typically the experiment will run from 930am-5pm EDT (1330-21 UTC) every day. The exception is on Mondays, when it will start at 915am to accommodate some introductions and general FFaIR information. The anticipated daily schedule can be seen in Table 1. Although not listed, we tend to take breaks after each activity and sometimes during the activities as we feel necessary. That said, we encourage our participants, especially the ones attending virtually, to take screen breaks when they feel they need a break. As part of the FFaIR Experiment, FFaIR also hosts a seminar series during the week prior to the start of FFaIR and the weeks FFaIR is in session. These are from 2-230 pm EDT and are centered around current and ongoing research and development associated with heavy rainfall. The seminar schedule can be seen in Table 2. These are open to NWS employees and our partners. The seminars are not recorded so that honest and open discussions can happen.

Table 1: Daily Schedule for the 2024 FFaIR Experiment.

Monday	Tuesday/Thursday	Wednesday/Friday
1315 - 1500 Ice Breaker and Orientation **1500 - 1600** Day 1 ERO Forecasting Activity **1600 - 1700** Verification **1700 - 1800** Lunch **1800 – 1930** Verification **1930 - 2100** Day 1 MRTP Forecast	**1330 - 1600** Morning review of yesterday's weather and Day 1 ERO Forecasting Activity **1600 - 1700** Verification **1700 - 1800** Lunch **1800 – 1830** **DIFFERENT GOOGLE LINK - Science Seminar** **1830 - 2000** Verification **2000 - 2100** Day 1 MRTP Forecast	**1330 - 1600** Morning review of yesterday's weather and Day 1 ERO Forecasting Activity **1600 - 1700** Verification **1700 - 1800** Lunch **1800 – 1930** Verification **1930 - 2100** Day 1 MRTP Forecast (Wed) possible Day 2 MRTP Forecast (Fri) End of week debrief

As shown in Table 1, the mornings will consist of creating a Day 1 forecast via the Excessive Rainfall Outlook (ERO). This activity is designed to mimic WPC's operational ERO. Participants will create an individual ERO then work together

Table 2: The 2024 FFaIR seminar schedule. Seminars will take place at 2:00pm EDT (18 UTC). The Google Meet link can be found here.

Dates of seminars – all seminars are 2-230pm EDT	Presenter(s)	Title/Theme of Seminar	Affiliation
Tues - June 4	Sarah Trojniak and Jimmy Corriea	How to FFaIR	CIRES-CIESRDS @WPC-HMT
Thurs - June 6	Erica Bower	Objective Verification of the Weather Prediction Center's Mesoscale Precipitation Discussions	CIRES-CIESRDS@WPC
Tues - June 11	Trevor Alcott	MPAS Ensemble Forecasts of Heavy Rainfall: Does adding members add value?	GSL
Thurs - June 13	Aaron Hill	Medium-range Forecasts of Excessive Rainfall with the CSU-MLP	University of Oklahoma
Tues - June 25	Bill Gallus	A Machine Learning Postprocessor to Mitigate QPF Errors for Improved Hydrometeorological Forecasting	Iowa State University
Thurs - June 27	Keith Brewster	FV3-LAM CAM Ensemble Consensus and Machine Learning Products for Predicting Heavy Rain for the FFaIR Experiment	CAPS @ University of Oklahoma
Tues - July 9	Matt Pyle	Current Status of RRFS and REFS, with an emphasis on QPF	EMC
Thurs - July 11	Eric James	Evaluating HREF probabilistic forecasts of excessive rainfall	GSL
Tues - July 23	Austin Coleman	Advancing Situational Awareness with Ensemble Clustering and Sensitivity Analysis Tools	CIRES-CIESRDS@WPC
Thurs - July 25	Mike Seaman	Leveraging Machine Learning and Probabilistic Guidance to Improve Flash Flood Forecasting Across Southern Utah	WFO- SLC
Tues - July 30	Brenda Philips	Societal Responses to Flash Floods	University of Massachusetts
Thurs - Aug 1	Ben Moore and Leif Swenson	Advances and Challenges In Atmospheric River Forecasting	PSL and CIRES-CIESRDS@PSL
Follow the WPC HMT Calendar for the seminar schedule: https://calendar.google.com/calendar/u/0?cid=Y19mNGNuZHZlcXJidWU3NjNqb2ZoN2h1cmQxNEBncm91cC5jYWxlbmRhci5nb29nbGUuY29t Google Meet Link: https://meet.google.com/fhb-spep-zui			

to create a collaborative ERO. After this, verification activities will occur. These will be broken up by lunch. Once the verification activities are completed, the day will wrap up with the Maximum Rainfall and Timing Product (MRTP). This is a short-term product that attempts to forecast which 6-h window from 21 UTC to 12 UTC will see the highest rainfall accumulation. Both of these products have been used to evaluate the performance of model guidance and tools in previous FFaIR experiments.

2.0.1 Hybrid Information

During the two weeks that the experiment is held in hybrid form, the FFaIR team will be utilizing technology that has been installed in some conference rooms at NCWCP. The technology allows for online participants to hear the conversation going on in the room and for them to actively participate in the conversation. In-person participants will be required to bring a laptop to the experiment. This will be used to both log onto the Google Meet so they can participate in any conversation occurring on the chat feature and so they can use the tools developed by the FFaIR team for the forecasting activities. During breakout Groups, the room will be separated into two spaces and the Groups will be a mix of in-person and virtual attendees.

2.1 Overview of Science Questions and Goals

As noted in the introduction, the main objective of FFaIR this year is the evaluation of the RRFS suite, hereafter referred to as the REFS. The REFS, its control (deterministic) member the RRFSp1 (also referred to as RRFS and RRFS_a across the NWS), and the ensemble's membership are all included in this evaluation. These models are all slated to replace nearly every operational CAM and the ensemble they comprise, the High Resolution Ensemble Forecast (HREF). One of the requirements from the NWS for this plan to be implemented is that the deterministic RRFSp1/REFS have comparable or better performance to the HRRR/REFS. To test how heavy rainfall forecasting will be impacted by the switch, the HREF and REFS systems will be pitted against each other. Therefore, many of our science questions and goals will be centered around scrutinizing differences in performance.

That said, we will also be evaluating additional guidance and tools. The Center for Analysis and Prediction of Storms (CAPS) at the University of Oklahoma (OU) is providing an ensemble system and a Machine Learning Product (MLP). A bias-corrected HREF forecast for Mesoscale Convective Systems (MCSs) ML forecast will be provided from Bill Gallus's and Kristie Franz's team at Iowa State University (ISU). The Hazardous Weather Testbed (HWT) and the NSSL are providing an MPAS model configuration (NSSL-MPAS) while Austin Coleman, part of WPC and Cooperative Institute for Research in Environmental Sciences (CIRES), will be providing REFS QPF clusters.

- Evaluate the performance, focusing on Quantitative Precipitation Forecast (QPF) of the RRFSp1 compared to the HRRR, NAMnest and NSSL-MPAS.
- Although not the main focus, other aspects of the RRFSp1 compared to the aforementioned models will also be noted, such as reflectivity, with performance differences conveyed to developers.
- Evaluate the performance of the REFS through means and probability of exceedances.
- Perform data denial experiments during forecasting activities to simulate what it will be like to forecast only having access to the REFS.
- In addition to evaluating "classic" ensemble probabilities, FFaIR will be evaluating a MLP for QPF probability of exceedance from the CAPS group.
- Continued evaluation of a OU-CAPS Spatially-Aligned Mean (SAM) and a SAM with local probability matched mean (LPM) applied with the SAM methodology, called the SAM-LPM.
- Evaluations of a bias-corrected HREF forecast over domains of intense precipitation.
- Use and evaluate cluster analysis based on QPF from the REFS using the previous four cycles of the membership.
- Continue to explore the addition of an ERO risk category between a Slight and Moderate risk, called the Enhanced Risk.
- Explore including an intensity contour on the ERO, defined by exceeding the 10-y or 25-y 1-h ARI threshold.
- Evaluate the performance (CSI, max QPF) of models and participants for specific 6-h precipitation extreme events via the MRTP.
- Explore participants' perception of probabilities through the MRTP.

Table 3: The models, ensembles and tools that each Group will have access to for the 2024 FFaIR Experiment. The Group names are the names of the operational and experimental ensembles; HREF and REFS respectively. The current, operational models/ensembles are in black, experimental models/ensembles/tools are in red. Note that the HRRR and GFS will be used in both Groups.

HREF Group	REFS Group
GFS HRRR NAMnest ARW-HREF ARW HREF2 FV3-HREF HREF NSSL-MPAS ISU HREF MLP	GFS HRRR RRFSp1 RRFSm2-6 REFS CAPS Det. CAPS Ensemble REFS Clusters

3 Description of Activities

Each day, participants will be broken into two Groups, either the HREF or REFS Group. The facilitators will work to mix up the Groups so that participants are not with the same people all week. Table 3 lists the guidance and tools that each Group will have access to. When creating their ERO and MRTP forecasts, they are expected to not look at any of the guidance listed in the other Group. One exception will be the HRRR, which is part of both ensemble suites.

The ERO activity will mimic what is done in WPC operations, with a Day 1 product valid from 16 UTC to 12 UTC each morning. Like in operations, the goal is to complete the activity by 16 UTC. However it has been the experience of the HMT team that on some days, especially Mondays, aspects of the experiment such as orientation and learning how to use our tools, leads to a delay in the completion of the ERO by 16 UTC. If this occurs, the participants will be instructed that they can no longer look at current observations or new data that comes out. The ERO, defined as the probability of exceeding flash flood guidance (FFG) within 25 miles of a point, can be thought of as a product that highlights the risk of flood/flash

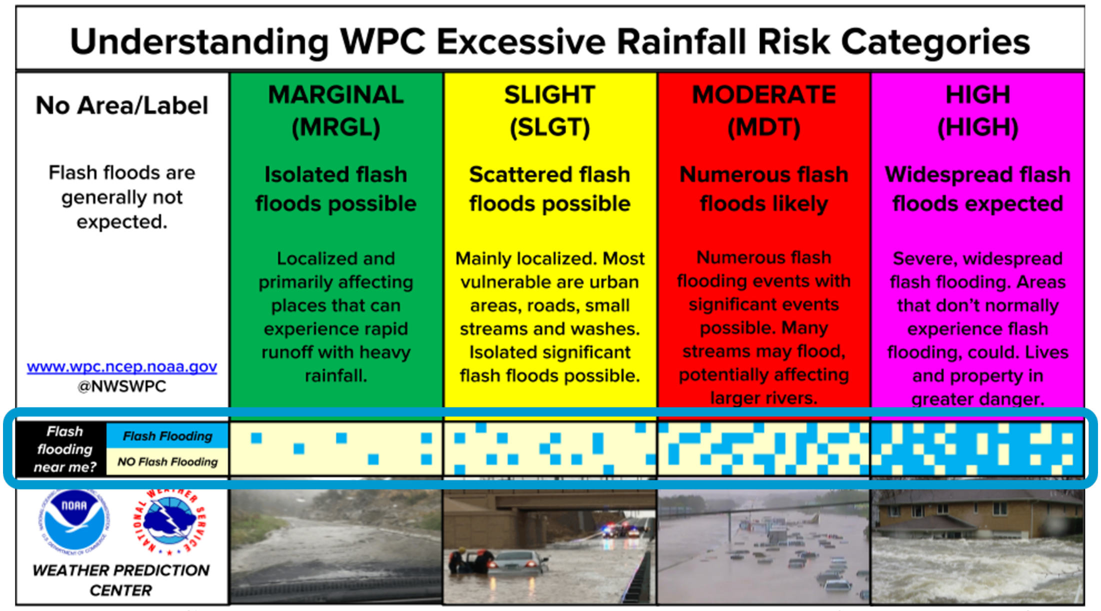


Figure 1: WPC graphic depicting what impacts can be expected for a given ERO category. Circled in blue is the expected coverage of flash flooding with each category.

flood coverage. This is shown in the WPC's summary of what impacts can be expected given each risk category in Fig. 1. Like last year, the ERO will have two additional contours that will be drawn that differ from the operational ERO. The first is the Enhanced Risk, which is set between the current Slight and Moderate Risk. This risk category is also being tested internally at WPC. The ERO risk categories are as follows: Marginal (5%-15%), Slight (15%-25%), Enhanced (25%-40%), Moderate (40%-70%), and High (>70%).

The other contour, referred to as Intensity (it was called Hatched in the 2023 FFaIR Experiment), will attempt to highlight where the high intensity rainfall rates will occur. The goal is to have this independent of flash flooding risk. Meaning that unlike the classic ERO Risk definition, focusing on coverage of impacts, the Intensity contour is attempting to highlight the meteorological aspect of heavy rainfall rather than its impacts. Last year we explored multiple ways to define intensity, including 6-h 10-y ARI exceedances. This year we will be tweaking the definition, with the Intensity definition being "An area where multiple hours of exceeding with 10-y 1-h ARI will occur withing 25 miles of a point." By stating that the high rainfall rate must occur for multiple hours, we hope to eliminate drawing an Intensity contour for areas of general thunderstorms.

Finally, part of the ERO activity will include the participants drawing both their own ERO and a collaborative ERO, as well as provide “Key Messages” for the ERO.These Key Messages are expected to include important meteorological and hydrologic information. For example, looking at the ERO in Fig. 2A, a Key Message that would be acceptable could say something like “The area within the Moderate Risk is expected to have training storms, with rates exceeding 2 inches per hour. Where the training line sets up expect widespread flooding.” We would like to avoid Key Messages like “Saint Louis is under a Moderate Risk” or “There is a Marginal Risk for flash flooding across the southwest today.” Although participants were given the option to draw their own EROs last year, these were not verified. This year they will be able to subjectively verify both there forecasts and Key Messages. An example of how the collaborative and individual EROs might vary can be seen in Fig. 2.

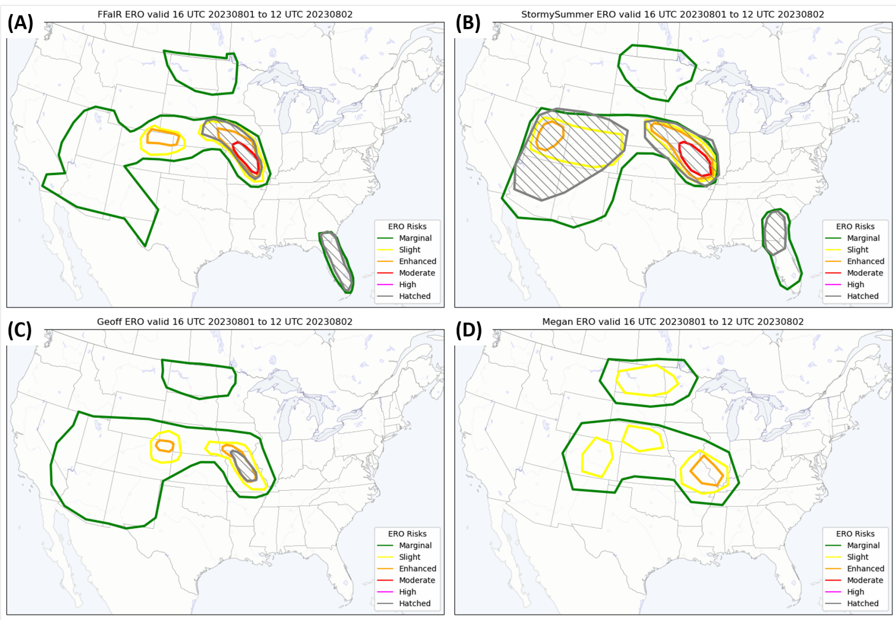


Figure 2: The Day 1 ERO valid 16 UTC 01 Aug. to 12 UTC 02 Aug. 2023 for the (A) FFaIR ERO and (B)-(D) three of the participants’ individual EROs that helped lead to the collaborated FFaIR ERO in (A). The ERO Risk contours are - Marginal: 5%-15% (green), Slight: 15%-25% (yellow), Enhanced: 25%-40% (orange), Moderate: 40%-70% (red) and High: >70% (purple/pink). The Intensity contour is grey with hatching.

The MRTP is a mix of identifying where and when the greatest 6-h rainfall accumulation will occur, thinking in probabilistic space, and scrutinizing guidance. Participants will be tasked with drawing 0.5”, 1”, 2”, 3”, 4” and 5” contours, a flood contour, and maximum rainfall points. Associated with these, they will provide the probability that they think each will be exceeded. For the flood contour, they will provide a Key Message about what could lead to the flooding. Like with the ERO, the Key Message should not be general verbiage, but should include information about rates or exceedances. For example, the full Key Message in Fig 3 is “the Dallas-Fort Worth region has been impacted by numerous rain events resulting in saturated grounds. Another 2” to isolated higher amounts, could lead to flooding in the urban and low-laying areas.” The same Group participants were assigned for the ERO will also be their Group for the MRTP activity. In the past, the participants have worked together to determine a domain and time for the MRTP to be valid. However, because of the data denial aspect of this year’s forecast the facilitators will be determining the domain and time. The MRTP can be valid for any 6-h window with a valid end time from 03 UTC to 12 UTC; i.e. 21 UTC to 03 UTC all the way to 06 UTC to 12 UTC. This allows for 10 possible time windows that can be selected. As part of the activity, participants will indicate the probability they think that each time window has on receiving the greatest 6-h rainfall in that domain. Participants will also complete a survey that asks questions about the guidance in the Group they were in. They will be able to subjectively verify their forecast. An example of what a completed activity might look like can be seen in Fig 3 while an example of a verification graphic can be seen in Fig. 4. An in-depth tutorial on the MRTP activity can be found here. If time permits, a Day 2 MRTP will be drawn as well.

4 Description of the Guidance and Tools

4.1 HREF

The HREF is the operational convection allowing ensemble for the NWS. The current operational version is version 3. The ensemble and the models that are included in it have been frozen since HREF version 3’s implementation in January

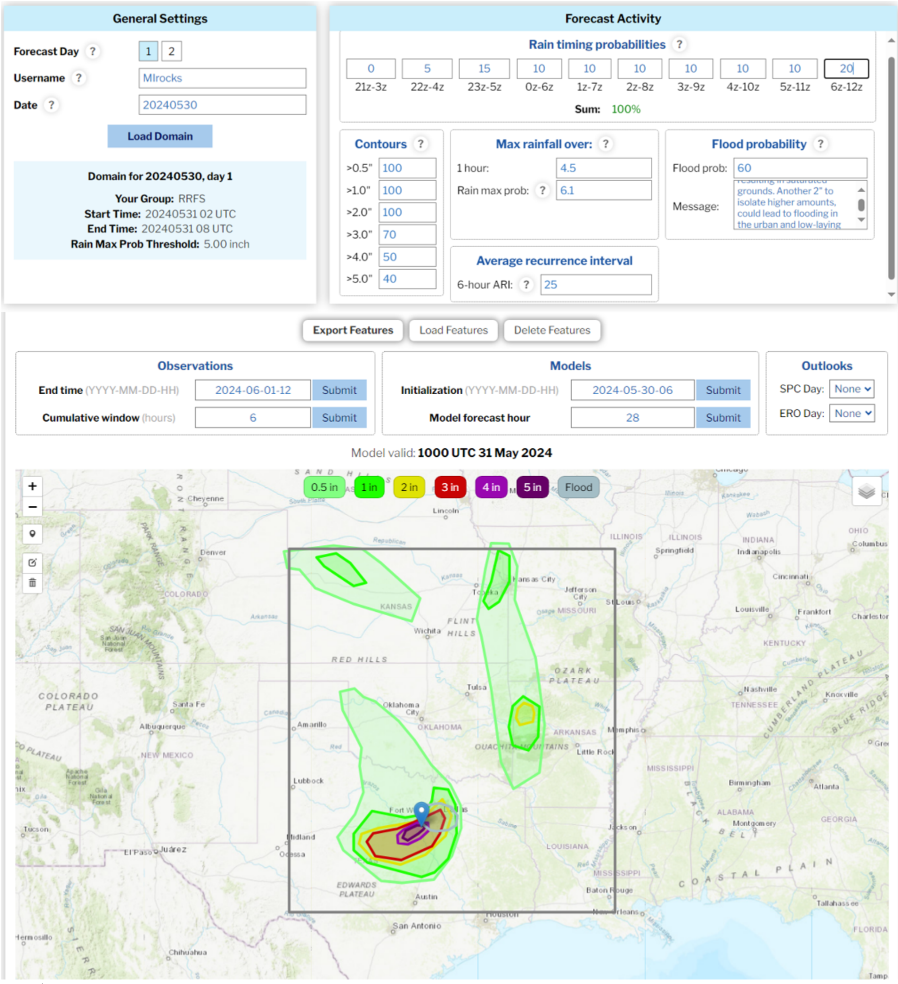


Figure 3: An example of a completed MRTP forecast on the MRTP Drawing Website. The domain is highlighted by the gray box. The forecast is valid from 04 UTC to 10 UTC 31 May 2024. This includes the contours for the the 6-h rainfall accumulations: 0.5” (green), 1” (yellow), 2” (red), 3” (dark red), 4” (purple), and 5” (pink) and an area of flood concern (gray). Along the top is the probabilistic information, information about maximum 6-h and hourly accumulation and the Key Message about flooding.

2021. It is comprised of the HRRR, NAMnest, HRW-ARW (ARW-HREF), HRW-NSSL (ARW-HREF2), and the FV3-HREF along with their 6-h time-lagged runs. The probabilities and means shown from the HREF are the operationally derived ones from EMC. When/if the REFS is approved for implementation, the HREF and all its members except for the HRRR will be retired.

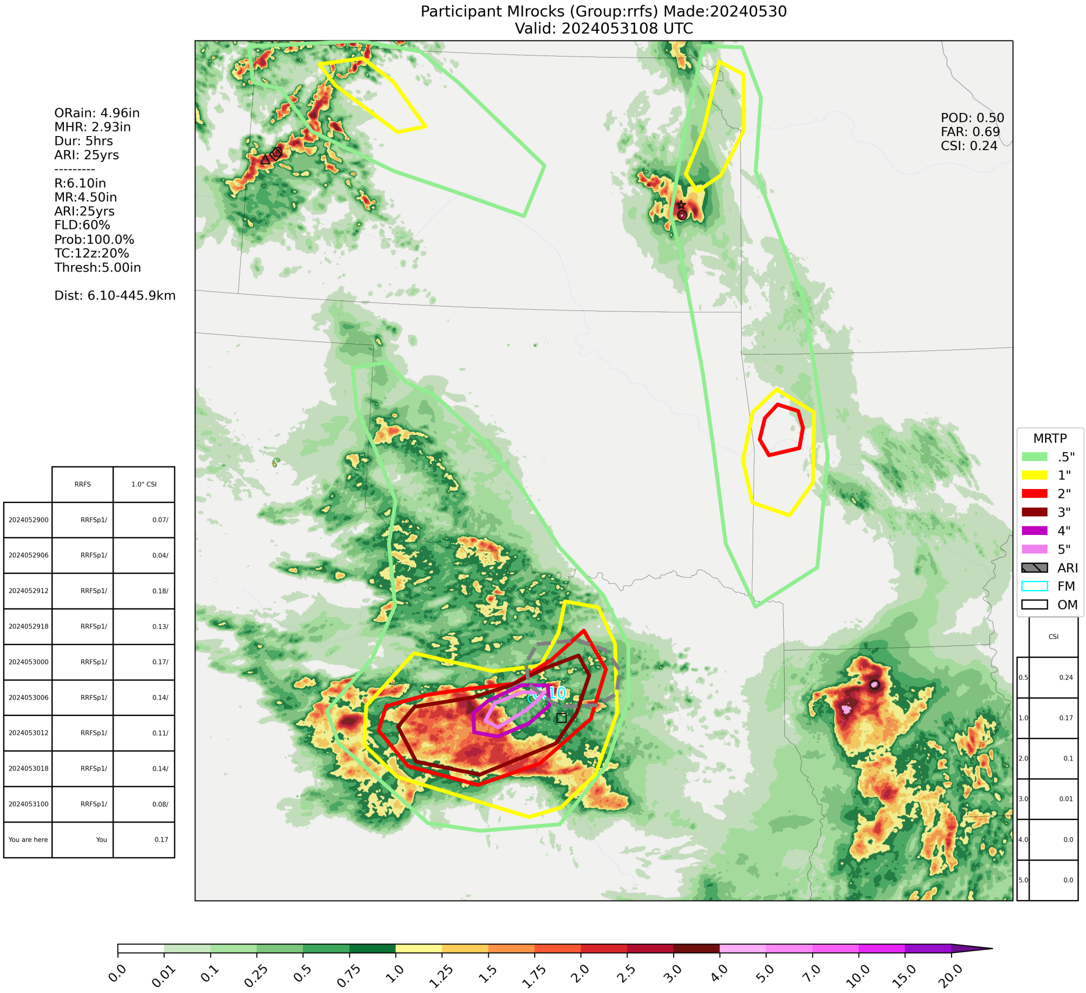


Figure 4: Example verification image for the MRTP experiment using the MRTP forecast from Fig 3. The graphic includes both the drawn MRTP (contoured) and the MRMS 6-h (filled), performance metrics like CSI and FAR, information about their maximum rainfall point compared to the MRMS maximum, and how the model's in the Group they were assigned performed.

4.2 REFS and RRFSp1

The REFS and its membership has been under development since roughly 2018. The original plan was for the REFS to be a single core system, using the Finite Volume Cubed Sphere (FV3). It is a high-resolution (3-km), hourly updated ensemble prediction system. It domain is larger than the HREF and its membership, covering North America. Its forecast ensemble includes multiple physics schemes, stochastic parameter perturbations, time-lagging, initial/lateral and boundary condition diversity. There are 6 FV3 members along with 6-h time-lagged members. The HRRR and its 6-h time-lagged member is also included. The membership and their configurations can be see in Table 4. Note that although this is a 3-km ensemble, in an attempt to address the high bias in convective development and QPF, cumulus (CU) schemes were added to the REFS and the control member in Fall of 2023. Differences in the CU schemes also add diversity to the ensemble.

The control member is the deterministic member (referred to as the RRFSp1 by the HMT team) and is initialized from a hybrid 3DEnVar analysis. The ensemble component of the 3DEnVar uses the RRFS Data Assimilation System (RDAS) ensemble Kalman filter. The rest of the REFS membership, excluding the HRRR, uses initial conditions from the corresponding member in the RDAS. The RRFSp1 is recentered around the RDAS at 07 and 19 UTC. Deterministic and ensemble forecasts are integrated out to 60 hours four times per day at 00, 06, 12, and 18 UTC. The ensemble size decreases as a function of forecast lead time due to the time-lagging nature and the fact that the HRRR is only run out 48 hours. Like with the HREF, the probabilities and means provided by the HMT are the ones provided by EMC.

4.3 CAPS Ensemble and MLPs

The CAPS team will be providing an ensemble, spatially-aligned means, and 6-h QPF exceedance MLPs. The ensemble employs the FV3 core and is comprised of 16 members. The members will have approximately 3-km grid spacing covering the contiguous United States (CONUS) and will be initialized at 00 UTC daily. A Local Ensemble Transform Kalman Filter (LETKF) will be run with 30 members during a one-hour spin-up (23-00 UTC) to assimilate standard observations and radar reflectivity data. This will be used to initialize 12 of the forecast members. The LETKF is part of the JEDI package in contrast to the GSI-sourced DA being used in the the REFS. Ensemble diversity is provided by physics variations combined with stochastic physics perturbations as well as lateral boundary condition

Table 4: The REFS membership configurations. The RRFSp1 (highlighted red) is the control member for the REFS. Sources of spread in the ensemble are: EnKF ICs, GEFS LBCs, time-lagging, multi-physics, stochastic parameter perturbations($*$), and fixed parameter perturbations (#).

	Micro-physics	PBL Scheme	Surface Scheme	LSM	Cumulus Scheme	IC/LBC
m1 (ctrl) aka RRFSp1	Thompson	MYNN	MYNN	RUC	G-F deep	RRFS hybrid/GFS
m2	Thompson*	TKE-EDMF	GFS	RUC*	G-F dp*+sh	RRFS enkf1/GEFS m1
m3	Thompson*	MYNN*	MYNN*	RUC*	saSAS deep	RRFS enkf2/GEFS m2
m4	NSSL#	MYNN*	MYNN*	RUC*	G-F deep*	RRFS enkf3/GEFS m3
m5	NSSL#	TKE-EDMF	GFS	RUC*	G-F dp*+sh	RRFS enkf4/GEFS m4
m6	NSSL#	MYNN*	MYNN*	RUC*	saSAS deep	RRFS enkf5/GEFS m5
m7 (m1-6h)						
m8 (m2-6h)						
m9 (m3-6h)						
m10 (m4-6h)						
m11 (m5-6h)						
m12 (m6-6h)						
m13 (HRRR)	Thompson	MYNN	MYNN	RUC	None	HRRRDAS / RAP
m14 (m13-6h)						

variation (refer to Table 5). Lateral boundary conditions will come from the GFS and various members of the operational GEFS.

Table 5: The CAPS ensemble membership configurations. In the Notes column it is indicated if the member has DA. If there is an AI-# then it indicates that it is a member used in the CAPS MLP.

Name	MP	PBL	SFC	LSM	Cu	IC/LBC	Notes
M0B0L2C0_Z	Thompson	MYNN	MYNN	RUC	G-F deep	ensmean/GFS	ZDA CNTL
M0B2L2C0	Thompson*	TKE-EDMF	GFS	RUC*	G-F dp*+sh	m01/GEFS m01	ZDA m001
M0B0L2C1	Thompson*	MYNN*	MYNN*	RUC*	saSAS deep	m02/GEFS m02	ZDA m002
M0B0L1C0	Thompson*	MYNN*	MYNN*	Noah MP	G-F deep*	m03/GEFS m03	ZDA m003
M0B2L1C0	Thompson*	TKE-EDMF	GFS	Noah MP	G-F dp*+sh	m04/GEFS m04	ZDA m004
M0B0L1C1	Thompson*	MYNN*	MYNN*	Noah MP	saSAS deep	m05/GEFS m05	ZDA m005
M1B2L2C0	NSSL#	TKE-EDMF	GFS	RUC*	G-F dp*+sh	m06/GEFS m06	ZDA m006
M1B0L2C1	NSSL#	MYNN*	MYNN*	RUC*	saSAS deep	m07/GEFS m07	ZDA m007
M1B0L1C0	NSSL#	MYNN*	MYNN*	Noah MP	G-F deep*	m08/GEFS m08	ZDA m008
M1B2L1C0	NSSL#	TKE-EDMF	GFS	Noah MP	G-F dp*+sh	m09/GEFS m09	ZDA m009
M1B0L1C1	NSSL#	MYNN*	MYNN*	Noah MP	saSAS deep	m10/GEFS m10	ZDA m010
M0B0L0	Thompson	MYNN	MYNN	Noah	-	GFS/GFS	AI-1
M1B0L0	NSSL	MYNN	MYNN	Noah	-	GFS/GFS	AI-2
M0B2L1	Thompson	TKE-EDMF	GFS	Noah MP	-	GFS/GFS	AI-3
M0B2L2	Thompson	TKE-EDMF	GFS	RUC	-	GFS/GFS	AI-4
M17 (MPAS)	Thompson	MYNN	MYNN	Noah	-	GFS/GFS	MPAS member
M0B0L2C0_L	Thompson	MYNN	MYNN	RUC	G-F deep	ensmean/GFS	LDA CNTL
M0B0L2C0_N	Thompson	MYNN	MYNN	RUC	G-F deep	ensmean/GFS	NoDA CNTL

Four of the real-time members will be run with 00 UTC GFS initial conditions to maintain a stable set of members needed to preserve the sub-set of members used to train the Machine Learning algorithm using forecasts from prior FFaIR experiments. Separately, in delayed real-time mode, the CAPS team will run some members utilizing GOES lightning data in place of radar data during DA and a control with no data assimilation for statistical comparisons.

4.3.1 Means

In additional to the traditional means, the arithmetic mean, the probability matched mean (PM mean or PMM) and the local probability matched mean (LPM mean or LPMM), the CAPS team will also be providing Spatially-Aligned Means (SAMs) for their ensemble system. The SAM, as implemented for 2024 FFaIR, aligns fields from ensemble members to a common position, that of the PMM. To do so, the domain is divided into overlapping patches and then the algorithm checks grid shifts of $+/-$ 25 grid points in 2-dimensions to find a minimum squared difference in the precipitation fields, including a penalty for larger offsets. After doing this, to align at synoptic scale, the process is repeated using smaller patches and scale length of penalty function to account for mesoscale displacements. Each member is then repositioned using the combined field of shift vectors and a mean is calculated. The CAP's team has found that at this step the LPM mean is the most accurate mean for this stage, hence SAM-LPM was implemented.

The process is illustrated for a sample case from the 2023 FFaIR experiment in Figs. 5-7. Figure 5 shows the 6-h QPF for all 15 members of the CAPS ensemble over the Midwest and Central Plains regions for the forecast hour 84 valid 12 UTC on 3 Aug 2023. Nearly all members had a mesoscale convective complex moving through the center of this region, but there are differences in the location of the feature and intensity. Figure 6 shows the shift vectors that were found using the 2-pass method aligning to PM mean, and the resultant shifted precipitation for each member. Looking at Fig. 7 it can be seen that the standard ensemble mean (upper left) is quite washed-out due to the spatial position differences while the LPMM (lower left) restores the intensity of the maxima but the character is very spotty. The SAM-LPM (lower middle panel), in contrast, has a well-defined swath of precipitation similar to many of the individual members and similar to the Stage-IV precipitation verification (upper right). Although the SAM has been found to help retain the intensity and characteristic of the individual members it does to solve the problem of a missed forecast by the membership, which can be see by the area of precipitation in eastern Tennessee, that was not well forecasted by any of the members. Further details on SAM and verification results for FFaIR 2023 are available in a (Lee et al., 2024).

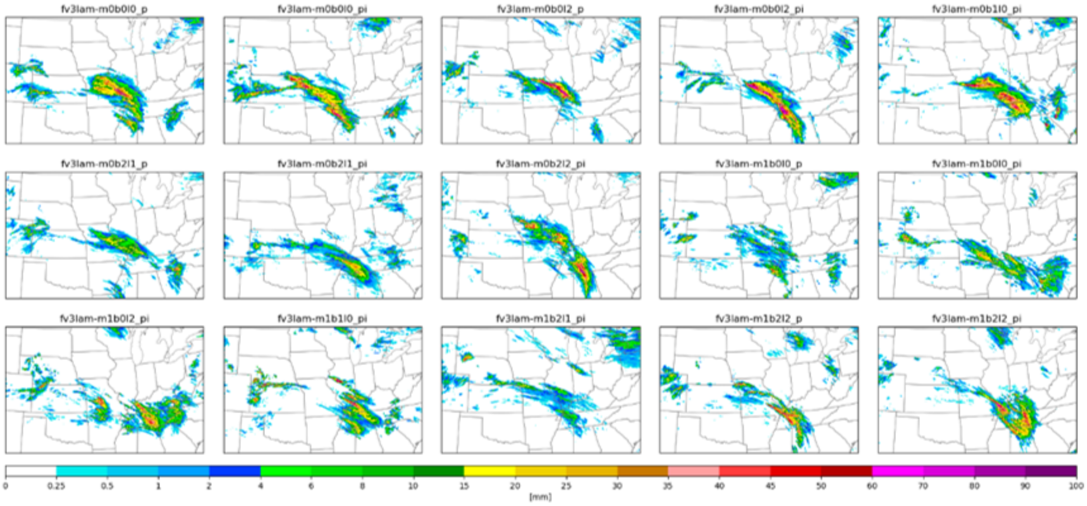


Figure 5: 6-h precipitation forecasts from CAPS ensemble for 2023 mid-South rainfall event for forecast hour 84 valid at 12 UTC 3 Aug 2023). NOTE: QPF is in mm.

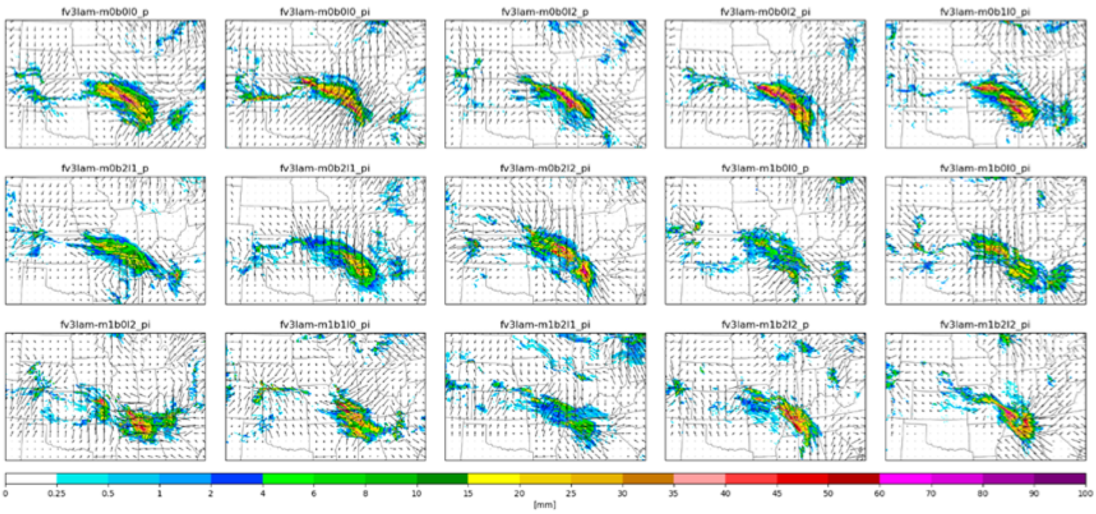


Figure 6: Like Fig. 5 but for the phase-shifted fields of CAPS ensemble members with shift vectors.

4.3.2 MLP

CAPS has developed and tested ensemble-based AI/machine learning forecast products for rainfall and snowfall prediction using a 12-member super-ensemble consisting of 4 FV3-LAM forecasts run by CAPS and the 8 members of the operational HREF (consisting of all members and their 12-hour time-lag counterparts expect the FV3-HREF). This super ensemble is referred to as the HREF+. Due to

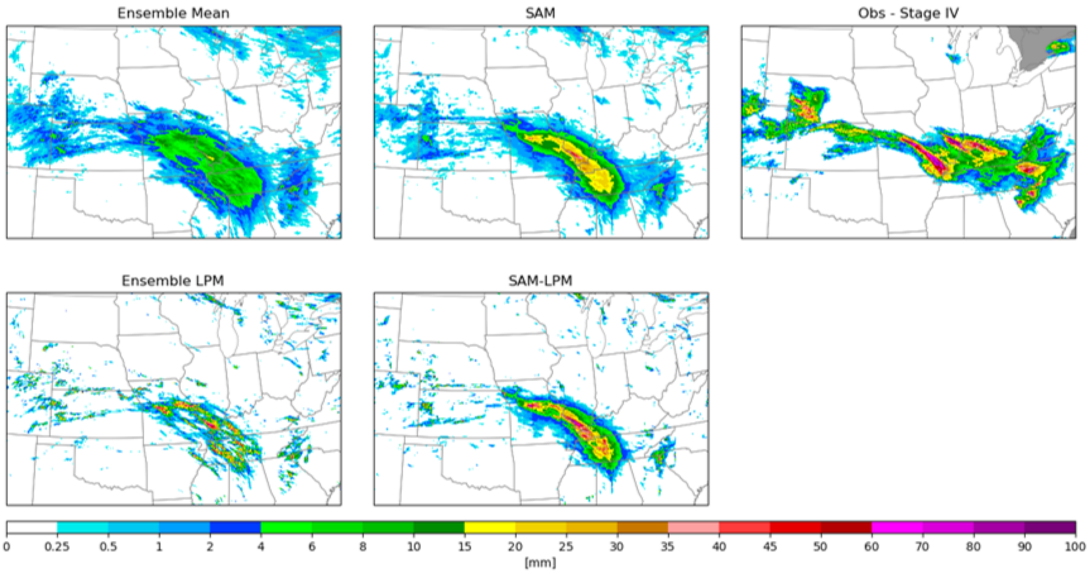


Figure 7: 6-h rainfall (mm) of the simple mean (top left), SAM (top middle), LPMM (bottom left), and SAM-LPM (bottom middle) fields for CAPS ensemble and observation Stage IV QPE (top right) valid the same time as Fig. 5.

constraints in the forecast hours covered by HREF and the use of a time-lagged member, the AI rainfall forecast products are only produced out to 48-h.

Forecasts are produced daily for ensemble-based probabilities of rainfall exceeding 0.5 and 1.0 inch during that period as well as ensemble mean ML 6-h precipitation. AI forecasts are generated for each member using U-Net, which is a deep learning approach that uses convolutional neural networks and is designed to identify spatial patterns in images; the term image here referred to one 2D field of forecast output. A set of 30 2D fields are used from each forecast member, including predictions of wind, temperature, and moisture information at different vertical levels, as well as predictions of reflectivity, QPF, and precipitable water. New for the 2024 FFaIR Experiment are the consideration of some static fields such as terrain height, terrain standard deviation and slopes. Derived fields such as moisture transport and low-level moisture convergence have also been added as inputs. The U-Net considers input data over 64x64 patches, producing forecasts for each patch which are stitched together for each member to produce a CONUS forecast. A neighborhood maximum ensemble probability (NMEP) is applied to the collection of individual U-Net predictions for each member to produce the final product for probabilistic forecast products.

4.4 ISU Bias Corrected HREF

The product designed by the ISU team attempts to improve the location errors of HREF forecasts of mesoscale convective systems (MCS). It consists of two ML models trained on 24-h QPF from MCSs identified from 2018-2023, along with Storm Prediction Center (SPC) Mesoanalysis data. All of the membership but the FV3-HREF is used. This member was excluded because insufficient data were available for ML training. The real-time steps are as follows:

1. Data Collection
   (a) On the day of an event, real-time HREF 24-h QPF is gathered from NOMADS for the 8 members of the HREF aside from the FV3-HREF.
2. Object Identification:
   (a) The QPF data is processed through MODE to identify objects.
   (b) MODE output images are manually analyzed to identify objects of interest for each member.
   (c) The centroids for these objects are automatically obtained from the MODE output files.
3. First Centroid Prediction:
   (a) The forecasted centroids of each member are fed into a ML model that has been pre-trained to calculate a weighted ensemble average centroid.
   (b) The output from this ML model is in latitude and longitude coordinates; Fig. 8.
   (c) Using the coordinates of the ensemble average centroid, 21 severe weather parameters are collected from the SPC Mesoanalysis for a $5x5$ grid ($200x200$ km area) around the weighted centroid.

(d) The averages of the northwest and southeast quadrants of the $5x5$ grid for each variable are calculated for each parameter.

4. Final Centroid Prediction:

   (a) The averaged Mesoanalysis data and weighted centroid are fed into a pre-trained ML model built using the Lasso method and a final centroid prediction is output (Fig. 9).

   (b) Within a $700x700$ km domain centered around the system of interest, the QPF information for each member is shifted so that the location of each QPF centroid matches the ML-predicted centroid.

5. QPF Adjustment and Visualization:

   (a) This adjusted data of each member is then plotted with 5 km grid spacing, alongside the ensemble mean precipitation and PMM.

   (b) The probabilities of exceedance for 1 inch, 2 inches, 3 inches, 5 inches, and 8 inches of precipitation will also be calculated using the ML-adjusted locations of precipitation.

   (c) Image generation is only for the immediate region around the MCS domain. It will not stitched back to the full CONUS.

   (d) If there are multiple MCSs identified there will be multiple domains, each as their own image.

4.5 QPF Clustering

The WPC DTB team will be providing an ensemble clustering product based on a time-lagged version of the REFS, including the HRRR member. Clusters of 6-h QPF will be generated daily for both the ERO and MRTP activities for the 10th, 25th, 50th, 75th, 90th, and 95th QPF percentiles. The ensemble clustering approach uses the forecasts from the most recent run and the previous three cycles, resulting in 28 members for the cluster analysis. Clusters will be generated for both the ERO and MRTP forecast activities. The ERO activity will be launched daily at 1300 UTC, using the 06 z cycles as its “current” forecast since the 12z

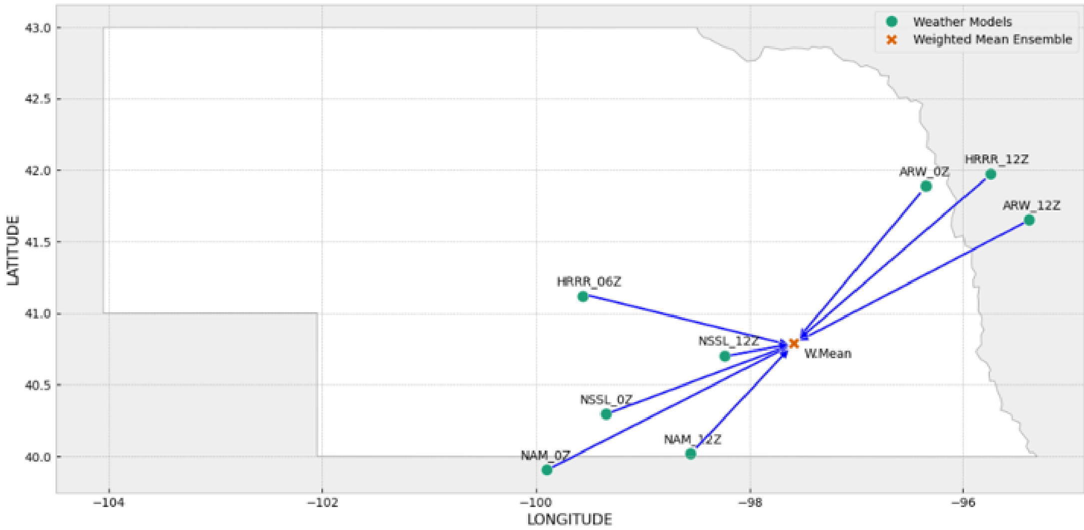


Figure 8: Example of weighted average centroid from the QPFs from the 8 HREF members used for the ISU MLP.

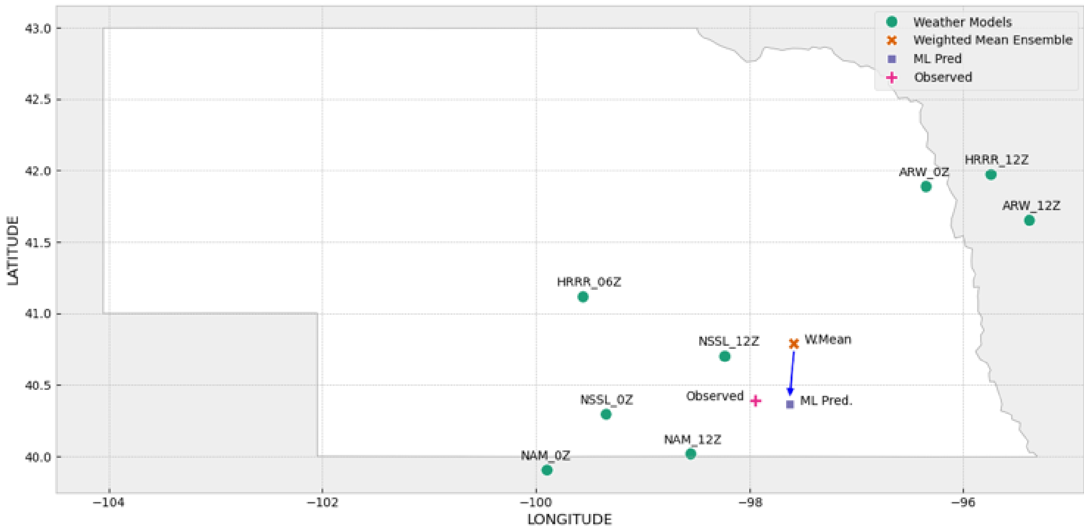


Figure 9: Example of the final ML predicted centroid using inputs of the ensemble average centroid and SPC Mesoanalysis data centered on the ensemble average, along with the observed centroid.

data will not be available by 13 UTC. This means the 06z and 00z cycles from the current day and the 18z and 12z from the previous day will be used for the cluster analysis avaiable for the ERO activity. Three preset clustering windows, 18–00 UTC, 00–06 UTC, and 06–12 UTC and domains chosen by the HMT team the evening prior will be used for the clustering.

Shortly after the ERO activity, a domain for the MRTP activity will be chosen. This domain and its corresponding time window will be fed to the clustering algorithm to produce one or two 6-h QPF clusters within the MRTP timeframe. The hope is that the REFS membership for the 12 z cycle will be available between 18 and 19 UTC so that the its “current” forecast cycle will be 12 z and the time-lagged cycles will be the 06z and 00z cycles from the current day and the 18z from the previous day. An example of clustering around a MRTP domain can be seen in Fig. 10. Verification will be conducted on these clusters and their full ensemble counterparts. Participants will be able to compare the cluster scenarios for both the ERO and MRTP activities to MRMS QPE contours to decide which forecasts fared best. The dominant patterns in the ensemble forecast are derived by calculating the first and second Empirical Orthogonal Functions (EOFs)

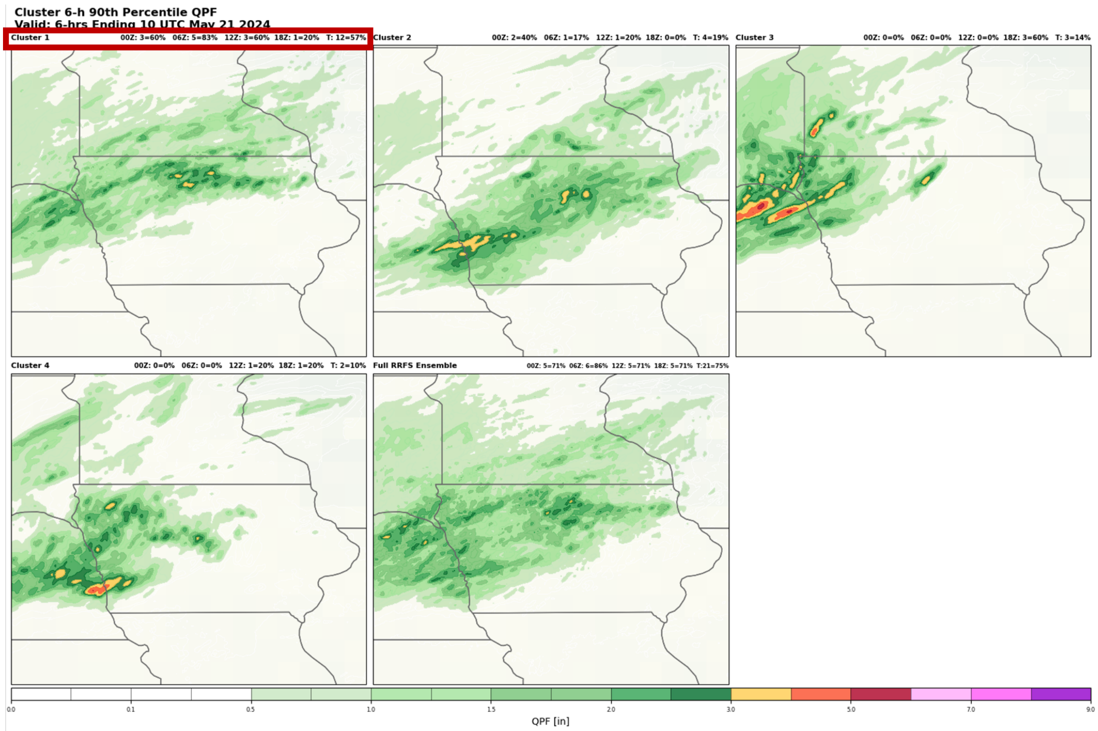


Figure 10: Example of the 6-h 90^{th} percentile from the 4 clusters identified by the EOF along with the 90^{th} percentile for the full membership. Above each panel, the number of members from each of the 4 cycles used in the clustering analysis are listed. For example, Cluster 1 is comprised the following number of members for each cycle (boxed in red): 00z - 3, 06z - 5, 12z - 3, and 18z - 1.

References

Lee, K. B., C-J., P. Spencer, and J. Park, 2024: Spatial aligned mean: A method to improve consensus forecasts of precipitation from convection allowing model ensembles. *Wea. Forecasting*, conditionally accepted as of June 2 2024.

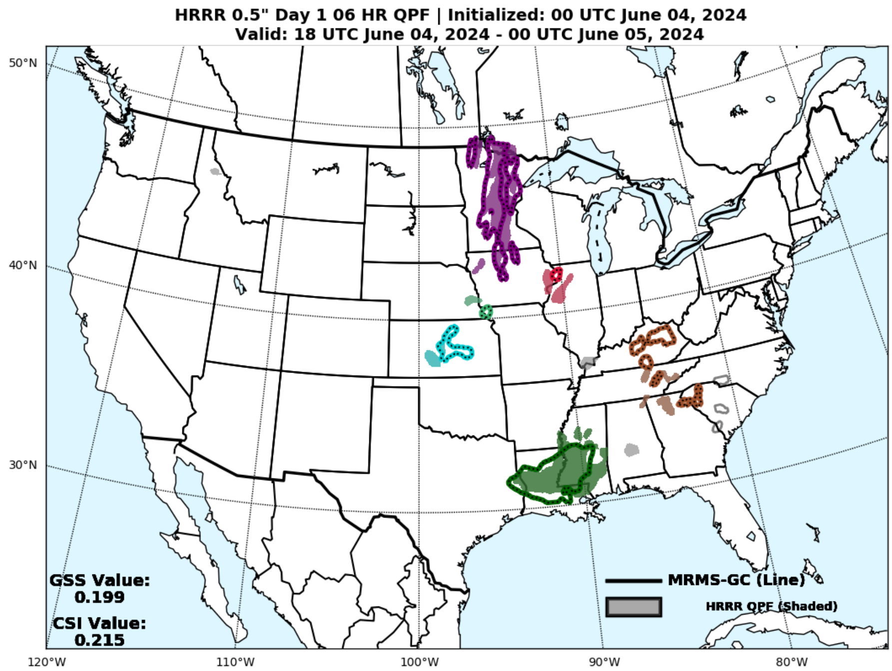
Appendices

A MET-MODE Settings

MODE will be run in realtime for 6-h verification and retroactively for 24-h verification. For the 6-h verification it will be run for the four cycles that would cover the Day 1 time period (i.e 12 UTC to 12 UTC). Meaning the "oddest" cycle would be 18z and the "newest" would be 12z when evaluating the graphics. The 6-h windows that will be evaluated are 12-18 UTC, 18-00 UTC, 00-06 UTC, and 06-12 UTC. An example of a MODE graphic that will be generated for FFaIR can be seen in Fig. 11. Below is some additional information about MODE and its setting for FFaIR verification:

1. For QPF and QPE a CONUS mask applied to common grid.
2. Grid stats harvested from MODE CTS
3. Circular convolution radius of 5 grid squares used
4. Double thresholding technique applied
5. Area threshold of 50 grid squares to keep an object
6. Total interest threshold for determining matches = 0.6

The HMT team would like to thank Ben Albright for is help with running MODE verification for FFaIR.



HRRR

Cluster ID	Interest	Angle	Cent Dist	Fcst Area	Obs Area	Intrsct Area
CF001_CO001	0.99924	1.81841	15.5323	3856	3500	2609
CF002_CO002	0.73035	20.73732	22.67134	140	107	0
CF003_CO003	0.91702	55.7505	25.17181	320	627	74
CF004_CO004	0.6678	56.95401	12.70673	619	103	4
CF005_CO005	0.72258	8.68838	35.01837	563	1004	8
CF006_CO006	1	7.92195	2.98728	4063	3954	2320

Figure 11: Example of the 6-h MODE graphic for the half inch threshold for the 04 June 2024 00z HRRR forecast, valid 18 UTC 04 June to 00 UTC 05 June 2025. On the graphic the HRRR half inch objects are shaded and the MRMS-QPE are contoured. Forecast and observed objects that are match by MODE are the same color and are called clusters. Each MODE cluster's attributes are listed at the bottom of the graphic.