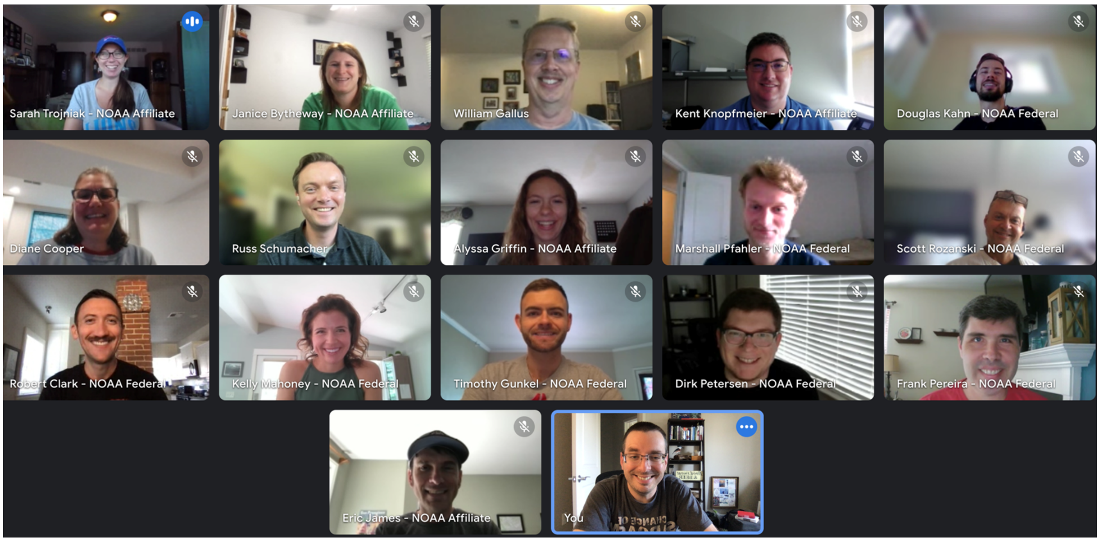

2023 Flash Flood and Intense Rainfall (FFaIR) Experiment

Operations Plan

June 5 - August 11, 2023
Weather Prediction Center
Hydrometeorology Testbed

Sarah Trojniak - CIRES CU Boulder, NOAA/NWS/WPC/HMT
James Correia Jr. - CIRES CU Boulder, NOAA/NWS/WPC/HMT

Contents

1 Introduction

The Flash Flood and Intense Rainfall (FFaIR) Experiment is part of the Hydrometeorology Testbed (HMT) at the Weather Prediction Center (WPC). FFaIR is centered around the challenge of forecasting for warm season precipitation and is one avenue through which research and development in this area is evaluated for transition into National Weather Service (NWS) operations. It brings together people from across the weather enterprise, from forecasters to developers to hydrologist, for a week at a time. During this time, the participants use experimental data in a pseudo-operational setting, allowing them to become immersed in the guidance and tools as they use them to create various forecasts. This results in hands-on evaluation of the guidance and products, which provides valuable information to developers that normal verification metrics might miss.

The importance of tackling the challenge of forecasting heavy rainfall and flash flood events is becoming increasingly important. Although large scale heavy rainfall events were lack-luster during the FFaIR 2022 season (Trojniak and Correia, Jr., 2023), research on the impacts of the warming climate suggest that heavy and extreme precipitation will increase (ex. Martinez-Villalobos and Neelin (2018) and Visser et al. (2022)). Therefore, working to better understand the performance of deterministic and ensemble models at predicting heavy rainfall and development of new forecasting tools for such events is crucial. FFaIR strives to be a reliable platform to work towards decreasing the challenge in forecasting for heavy/extreme rainfall and the flash flooding that can be associated with it.

2 Experiment Operations

For the first time since the start of the COVID pandemic, FFaIR will host participants onsite at the National Center for Weather and Climate Prediction (NCWCP). Additionally, differing from the previous years of FFaIR, the experiment will be expanded from four to six weeks. The experiment will have participants for a week at a time[1], on and off starting June 5 and ending August 11 and will consist of four virtual weeks and two hybrid weeks, with both in-person and virtual participants during the hybrid weeks. The weeks that FFaIR will be hosting participants are:

Week 1: June 5 - 9 (virtual)
Week 2: June 12 - 16 (virtual)
Week 3: June 26 - 30 (hybrid)
Week 4: July 10 - 14 (virtual)
Week 5: July 31 - Aug 4 (hybrid)
Week 6: Aug 7 - 11 (virtual)

In addition to the daily activities of the experiment, FFaIR will also host a science seminar series. These will take place at 2:00pm EDT (18 UTC) every Tuesday and Thursday that the experiment is operating. The seminars, like in the past, are open to everyone within the NWS and our partners. Presenters will range from academics to model developers to operational forecasters. A list of the seminars can be seen in Table 1.

2.1 Daily Activities

The experiment will run from 930am-5pm EDT (1330-21 UTC) with planned time for breaks and lunch. Due to the strain caused from actively looking at a screen for a prolonged period of time, even outside of the planned breaks the FFaIR team will be encouraging participants to step away from the screen whenever they feel the need to. The virtual platform that will be used during the experiment will once again be Google Meet. The general activities will consist of a morning forecasting activity, verification of products and tools, and an afternoon forecasting activity. On Mondays, time will be spent completing an ice-breaker to help the participants meet one another and reviewing FFaIR operations and the guidance being evaluated. The complete daily schedule can be seen in Table 2

[1]In some cases participants will not be able to attend the entire time of scheduled FFaIR week.

Table 1: The 2023 FFaIR seminar schedule. Seminars will take place at 2:00pm EDT (18 UTC). The google link can be found here. The seminars highlighted in red will be given outside of the 6 weeks FFaIR is in session.

Seminar Date	Name(s)	Topic/Title	Affiliation
Tues. May 30	Sarah Trojniak and Jimmy Correia	How to be FFaIR	CIRES/CIESRDS@ WPC-HMT
Thurs. June 1	Peggy Lee	An overview of the NWC's experimental products: the FHO, AHD, and NHD	NWC
Tues. June 6	Andrew Osborne	MRMS Machine Learning QPE	OU-CIWRO @ NOAA/OAR NSSL
Thurs. June 8	Jane Marie Wix	"A Recap of the July 2022 Eastern Kentucky Flooding"	WFO Jackson, KY
Tues. June 13	Erik Nielsen and Jen Henderson	"Current Knowledge about TORFFs in both the social and physical science realms"	TTU
Thurs. June 15	Jacob Carley	"The Status of the First Version of the Rapid Refresh Forecast System"	EMC
Tues. June 27	Aaron Hill and Russ Schumacher	"Progress towards medium range excessive rainfall forecasts with the CSU-MLP"	CSU
Thurs. June 29	Kristie Franz	"QPF driven ensemble streamflow predictions using three different hydrologic models"	ISU
Tues. July 11	Marc Chenard	"WPC Excessive Rainfall Outlook: Overview, recent verification, and a look ahead"	WPC
Thurs. July 13	Janice Bytheway and Diana Stovern	"Characterization of extreme precipitation in the HREF"	PSL
Tues. July 25	Keith Brewster and Nate Snook	"FV3-LAM & HREF CAM Ensemble Consensus and Machine Learning Products for FFaIR"	OU CAPS
Tues. August 1	JJ Gourley	Flash Flood Flashiness	NSSL
Thurs. August 3	Brenda Philips	Flash Flood Response	UMass
Tues. August 8	Mark Glaudemans	"Water Model Geospatial tools and Inundation Maps"	NWS
Thurs. August 10	Steve Martinaitis	"Initial Work on Precipitation Nowcasting within MRMS"	OU-CIWRO @ NOAA/OAR NSSL

The flow of FFaIR this year will be similar to last year. Typically the day will begin by briefly discussing what happened weather-wise over the past 24 hours. This will be followed by a weather briefing given by a WPC forecaster. The participants will then be broken into two groups for the Day 1 forecasting activity. One group will work on creating an Excessive Rainfall Outlook (ERO), mimicking what is done in WPC operations. The second group will work on creating an AERO, based on Average Recurrence Interval (ARI) exceedances. Both of these will be discussed further in Section 2.3. The groups will change throughout the week so that the participants have a chance to interact with different people. Participants will be able to create their own ERO or AERO each day and work together to create a collaborative ERO or AERO.

Table 2: Daily Schedule for the 2023 FFaIR Experiment.

Monday	Tuesday/Thursday	Wednesday/Friday
1330 - 1500 Ice Breaker and Orientation **1500 - 1600** Day 1 ERO and AERO Forecasting Activity **1600 - 1700** Verification (det./ens.) **1700 - 1800** Lunch **1800 – 1900** Verification (MLP/forecasts) **1900 - 2000** Day 1 MRTP Forecast **2000 - 2100** Day 2 MRTP Forecast	**1330 - 1600** Morning review of yesterday's weather and Day 1 ERO and AERO Forecasting Activity **1600 - 1700** Verification (det./ens.) **1700 - 1800** Lunch **1800 – 1830** DIFFERENT GOOGLE LINK - Science Seminar **1830 - 1930** Verification (MLP/forecasts) **1930 - 2100** Day 1 MRTP Forecast (poss. D2)	**1330 - 1600** Morning review of yesterday's weather and Day 1 ERO and AERO Forecasting Activity **1600 - 1700** Verification (det./ens.) **1700 - 1800** Lunch **1800 – 1900** Verification (MLP/forecasts) **1900 - 2000** Day 1 MRTP Forecast **2000 - 2100** (Wed) Day 2 MRTP Forecast (Fri) End of week debrief

Completion of the Day 1 forecasting activity will be followed by verification activities. Verification will be broken up by lunch. Much of this year's verification will revolve around the Rapid Refresh Forecast System (RRFS). The RRFS is planned for implementation in fall of 2024 and is expected to replace all of the NWS's convection allowing models (CAM), including the ensembles. The design of the RRFS will be discussed further in Section 3. Verification will also focus on evaluation of machine learning products and evaluation of participants' own forecasts. As mentioned previously, on Tuesdays and Thursdays participants will attend the FFaIR seminar series as well. The last activity of the day will be the Maximum Rainfall and Timing Product (MRTP), which will include another weather briefing from a WPC forecaster. When possible, participants will complete two different MRTPs.

2.1.1 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.

2.2 Overview of Science Questions and Goals

The questions asked during the verification sessions will address the science questions and goals of the 2023 FFaIR Experiment. The science questions and goals listed below are both subjective (qualitative) and objective (quantitative) and are not meant to be overly specific. As stated above, verification will focus on the RRFS, both on its deterministic and ensemble configurations. Additionally this year, the FFaIR team is collaborating with the Storm Prediction Center's (SPC) Hazardous Weather Testbed (HWT). This took two forms, one was a developmental product in the form of a webiste to view model soundings. The other was to have some RRFS verification questions asked at both the Spring Forecasting Experiment (SFE) and FFaIR. This will allow for a more in-depth subjective analysis of various model parameters, ensemble products, and the impact of data assimilation (DA). We hope that this collaboration will help better inform the RRFS developers since the data will be collected in nearly 4 months of results across spring and summer.

- Evaluate the performance, focusing on Quantitative Precipitation Forecast (QPF) and precipitation rate, of the RRFS_a (referred to as the RRFSp1 in the FFaIR experiment) compared to the HRRR and NAMnest.
- Evaluate the performance of different configurations of the RRFS other than the planned operational version of the RRFS deterministic.
- Analyze the impact of the RRFS DA and compare it to the DA done in the HRRR.

- Identify the pros and cons of a multi-physics ensemble compared to a single physics ensemble with stochastic perturbations. Compare their performances to the HREF.
- Evaluate the performance of ensembles with time-lagged members.
- Provide a large, time-lagged ensemble created by the FFaIR team to determine how the impact using all available cycles of the RRFSp1 and its stochastic members has on the predictability of extreme precipitation.
- In addition to evaluating "classic" ensemble probabilities, FFaIR will be evaluating a machine learning product (MLP) for the probability of exceedance from the CAPS group.
- 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.
- Analysis of the Colorado State University (CSU) ERO MLPs. This will include an updated version to the HRRR-based ERO MLP evaluated last year and a comparison of GEFS-trained ERO MLPss trained on observational datasets.
- Explore the addition of an ERO risk category between a Slight and Moderate risk.
- Explore including an intensity contour on the ERO, defined by exceeding some ARI threshold.
- Continue to analyze the the utility of using 6-h ARI QPF exceedances as a proxy to identify rainfall intensity via the AERO and work to develop a verification methodology for the product.
- Evaluate the performance (CSI, max QPF) of the various models for specific 6-h precipitation extreme events via the Maximum Rainfall and Timing Product.

2.3 Forecasting Activities

The forecasting activities will closely follow last year's. As stated, the forecasting activities will be the ERO, AERO, and MRTP. The ERO and the AERO will both be a Day 1 product, issued by 16 UTC each day and valid from 16 UTC to 12 UTC the following day. Participants will create their own ERO or AERO each day depending on which group they are assigned and work together to create a collaborative ERO or AERO as well.

The definition of the ERO is the probability of exceeding flash flood guidance (FFG) within 25 miles of a point. As the risk categories increase, one can expect to see more instances of flash flooding; See Fig 1. The ERO will consist of the four operationally defined risks: Marginal (5%-15%), Slight (15%-40%), Moderate (40%-70%), and High (>70%). Additionally, the FFaIR ERO will include an Enhanced Slight (hereafter called Enhanced), denoted as 25% chance of exceeding FFG. This threshold was chosen because WPC is internally testing an additional contour at this threshold. Also added to the ERO forecasting activity is a contour for identifying where the rainfall intensity could be large; this contour will be referred to as "Hatched". The addition of an intensity contour comes from feedback from participants during FFaIR last year (Trojniak and Correia, Jr., 2023), suggesting using the AERO and ERO in tandem. The expected exceedance of the 6-h 10-y ARI will be used as guidance for the Hatched contour. This is because over the last two years of creating an AERO this threshold was not commonly drawn. However when the 6-h 10-y ARI was included in the AERO product, it was often associated with higher end events not always associated with higher end ERO risk categories. An example of what an ERO would look like if only the operational risks were drawn and what one with the additional risks would look like can be seen in Fig. 2.

The methodology for the AERO this year will be similar to last year. Once again the participants will have the option to draw contours for the 6-h 2-y, 5-y, 10-y, 25-y, and 50-y ARIs, with no set probability of exceedance for the drawing the thresholds. Therefore the AERO identifies the 6-h ARI that is most likely to be exceeded within 25 miles of a point, for any six-hour time period within the

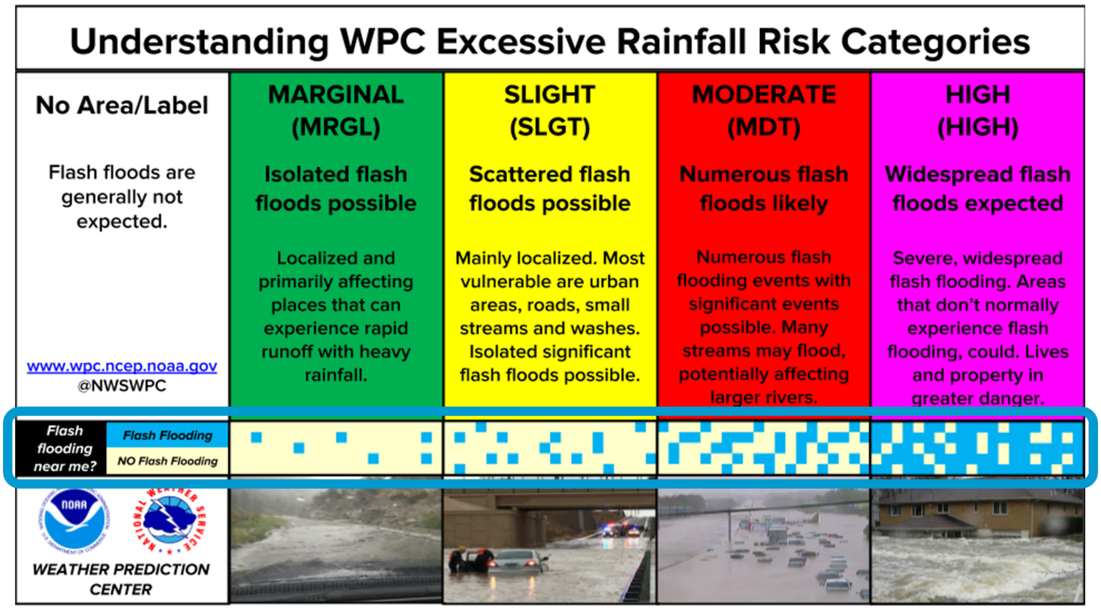

Figure 1: WPC graphic depicting what impacts can be expected for a given ERO category. Circled in blue is the expected coverage of flash flooding with each category.

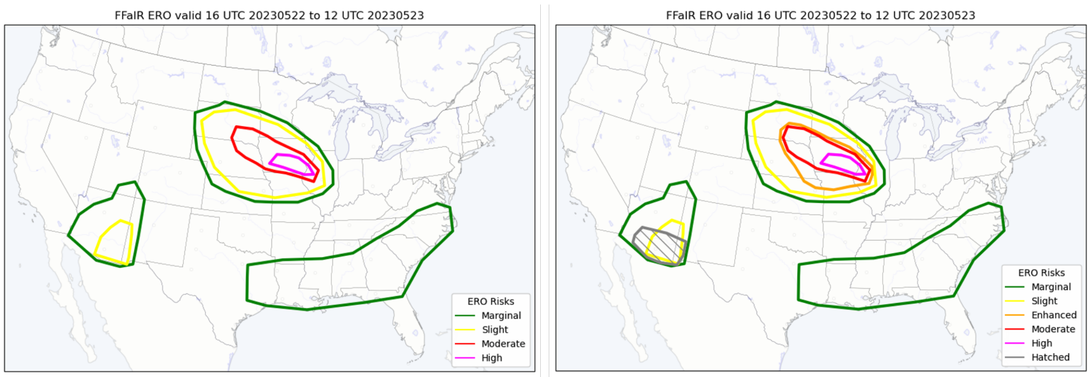

Figure 2: Comparison of how different an ERO might look with (right) and without (left) the inclusion of an Enhanced risk and the hatched area for intensity. This is an idealized case. The Enhanced is contoured in orange, the "hatched" is contoured grey.

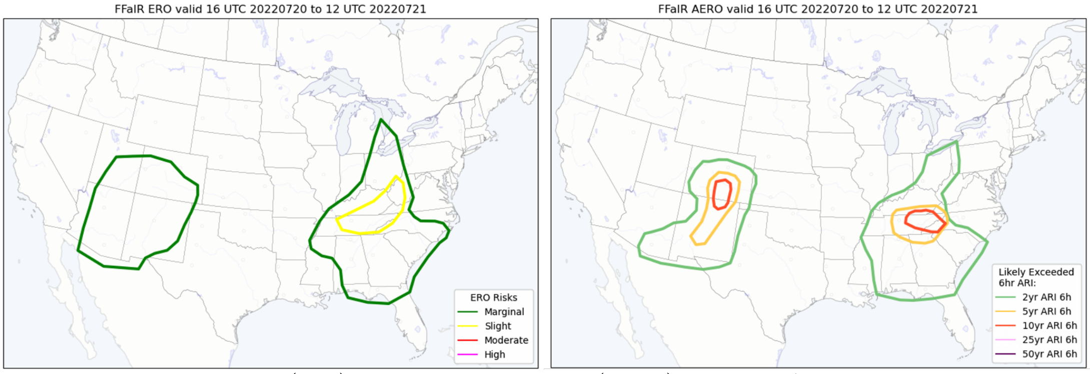

Figure 3: The Day 1 (left) FFaIR ERO and (right) FFaIR AERO valid 16 UTC 20 July to 12 UTC 21 July 2022. This case was analyzed in the 2023 FFaIR Final Report (Trojniak and Correia, Jr., 2023).

valid time of the product (16 UTC to 12 UTC). It was discussed whether the 2-y ARI should be dropped from the AERO but it was decided to keep the contour since it works as a good baseline on challenging what the participants want the product to accomplish. For instance, during the summer it is likely that any given thunderstorm could exceed the 6-h 2-y ARI so often the conversion amongst the AERO group last year focused on whether or not this should warrant drawing the contour across the general thunderstorm area. An example of an ERO and AERO from FFaIR last year can be seen in Fig 3.

The MRTP is an individual forecasting activity, requiring the participants to draw a 6-h QPF forecast, along with identifying the location where they think the maximum rainfall will occur inside the MRTP domain. The group will work together to determine the region and 6-h window where the most rainfall or largest areal coverage will occur. The 6-h window for the MRTP must be between 21 UTC and 12 UTC the following day. Once the domain and time period are chosen, the participants will work on their own to create their MRTP. The participants have the option to draw for 0.5, 1, 2, 3, 4 and 5 inches. New this year, they can also depict where they think flooding will occur.

In addition to drawing areas for the aforementioned thresholds, participants will also be randomly assigned a model or ensemble to evaluate during their forecasting process. They are not required to base their forecast off the model or ensemble but they are expected to provide feedback about it. The participants will also be required to forecast the following:

- 6-h maximum rainfall.
- Maximum 6-h ARI to be exceeded.
- 1-h maximum rainfall.
- Probability of flash flooding.
- Probability of the flash flooding leading to damage.
- The probability that the 6-h maximum rainfall will exceed a pre-determined value; i.e. the participant's confidence an extreme event of the chosen threshold will occur.
- Across the possible 6-h time periods that could have been chosen as a valid MRTP time, what is the probability that the maximum 6-h rainfall will occur in that time window (discussed further below).

The last of the bullet points is a way to analyze the timing aspect of the MRTP. There are 10 possible six hour windows between 21 UTC and 12 UTC, with each of these time windows ending at: 03, 04, 05, 06, 07, 08, 09, 10, 11, and 12 UTC. The participants will be asked to provide what they think the probability of the maximum 6-h rainfall will be for each of these windows, adding up to 100%. For instance, if the MRTP is valid from 23-05 UTC, then a participant might think there is a 60% chance the maximum will be in the 05 UTC 6-h window. That means they have 40% left to split between the other 9 time windows. They might think there is an equal chance of the maximum occurring in the 6-h window before and after the valid MRTP time, so they would input 20% for the 04 UTC time window and 20% for the 06 UTC one. Then they would input 0% for the reminder of the 6-h time windows. The final input would look like: 0,20,60,20,0,0,0,0,0,0.

Depending on factors such as time, screen fatigue, and the weather, a day 2 MRTP will be completed as well. Like the day 1 MRTP, the product can be valid starting at 21 UTC until 12 UTC but starting the following day and ending the day after that. The ability to complete a second MRTP for the following day will

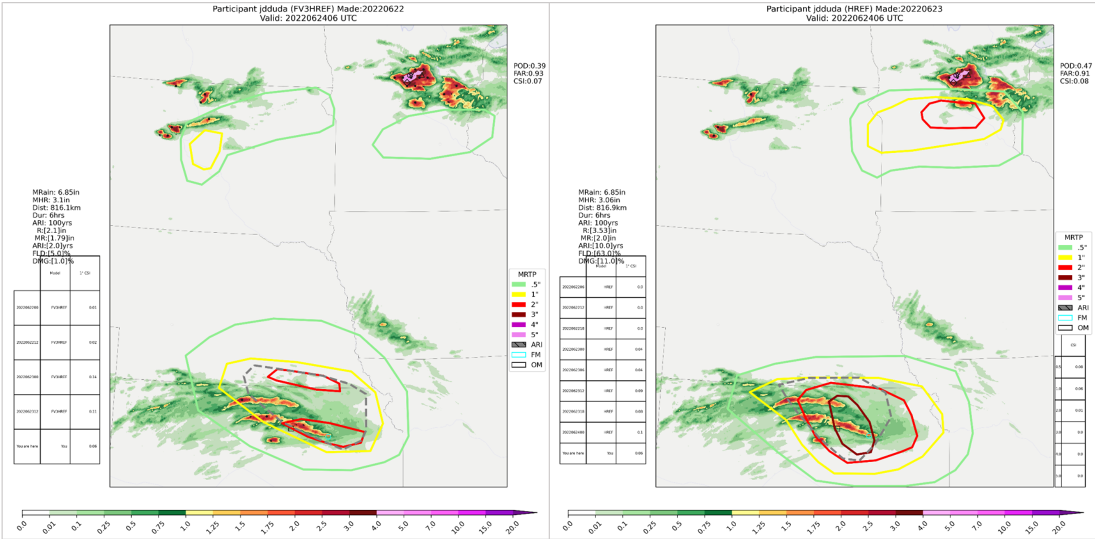

Figure 4: An example of the verification graphics for a Day 2 (left) and Day 1 (right) MRTP issued by a participant in last year's experiment, valid 06 UTC 24 June 2022.

allow for analysis of the utility of the CAMs to forecast heavy rainfall events at longer lead times. This is a particularly useful analysis since the RRFS is expected to extend CAM forecasts out to 60-h. If a day 2 MRTP was issued, the next day's Day 1 MRTP will be valid over the same time and region. An example of a day 1 and day 2 MRTP from last year can be seen in Fig. 4.

3 Guidance, Products, and Data to be Evaluated

The following subsections provide an overview of the guidance and tools that will be evaluated in FFaIR this year. Most of the data will be evaluated daily via the verification sessions however some of it will only be evaluated on a weekly basis. These products require general feedback utility rather than verification against observations. For instance, satellite products that will be provided by CSU Cooperative Institute for the Atmosphere (CIRA) will be used in forecasting activities and at the end of the week participants will be asked to provide feedback on things they noticed such as if high layer vapor transport values were associated with heavy rainfall.

3.1 RRFS

The RRFS uses the Finite Volume Cubed-Sphere (FV3) core. Multiple configurations of deterministic FV3 CAMs will be evaluated this year. These will follow the naming convection RRFSp#[2]. The primary objective, however, is to identify strengths and weakness of the RRFS_a, which is planned for implementation in Fall 2024. Although the formal name of the configuration is RRFS_a, it will be called RRFSp1 during the FFaIR experiment to stay consistent with the terminology used in the FFaIR 2022 Experiment. Since the model is still in active development, there is a possibility that updates to the RRFSp1 will occur during FFaIR. If science changes are made during FFaIR that are expected to impact results, the team will take note of the change and include it in the final report.

As stated in Section 2.2, the RRFS refers to the entirety of the new NWS CAM system. It is a rapidly-updating, convection-allowing (3 km) ensemble forecast system. Four configurations of the RRFS will be provided by the RRFS development team for testing this year with: single physics membership, time-lagged single physics membership, multi-physics membership, and time-lagged multi-physics membership. The 00z, 06z, 12z and 18z cycles of the models will be evaluated. The ensembles are initialized using 3-km ensemble perturbations drawn directly from the RRFS Data Assimilation System's (RDAS) ensemble Kalman filter analysis members. The control member of the ensembles will be the same and is the NWS's planned deterministic FV3 CAM (what FFaIR call's the RRFSp1). The control member forecast is initialized from the hybrid 3DEnVar analysis. The RDAS uses a wide variety of conventional observations along with radar reflectivity. It also includes a nonvariational cloud analysis. For gravity wave drag, the small scale and turbulence orographic form drag options are used in all members. The configuration for the RRFSp1 as of the end of May can be found in Table 3.

For the single physics (hereafter RRFSe1) and time-lagged single physics (hereafter RRFSe1tl) ensembles, stochastically perturbed parameterization tendencies (SPPT) are applied to all perturbed members (i.e., RRFS01-09). Stochas-

[2]There will be no deterministic FV3 CAM called RRFSp2. This is to avoid confusion since last year this name referred to the DA version of RRFSp1.

Table 3: The deterministic model configurations that will be evaluated in FFaIR. For the models provided by the OU CAPS team, if the model is part of their machine learning product, the member number is superscripted as AI-#.

Members (data provider)	ICs	LBCs	Microphysics	PBL	Surface	LSM
RRFSp1 (EMC)	RRFS hybrid 3DEnVar	GFS	Thompson	MYNN	MYNN	RUC
RRFSp3[AI-2] (OU CAPS)	GFS	GFS	NSSL	MYNN	MYNN	NOAH
RRFSp4[AI-1] (OU CAPS)	GFS	GFS	Thompson	MYNN	MYNN	NOAH
RRFSp5[AI-3] (OU CAPS)	GFS	GFS	Thompson	MYNN	MYNN	RUC
RRFSp6 (OU CAPS)	GFS	GFS	NSSL	TKE-EDMF	GFS	RUC
RRFSp7[AI-4] (OU CAPS)	GFS	GFS	Thompson	TKE-EDMF	GFS	NOAHMP

tic parameter perturbations (SPP) are applied to the microphysics, PBL, surface layer, radiation, and gravity wave drag parameterizations in the perturbed members. The configuration for the membership of RRFSe1 and RRFSe1tl can be found in Table 4, where the members with $*$ next to them are used in the RRFSe1tl. For the multi-physics (hereafter RRFSe2) and time-lagged multi-physics (hereafter RRFSe2tl) ensembles SPPTs are applied to all perturbed members. SPP is applied to Thompson in members RRFSphys02-05; MYNN PBL and surface layer physics in members RRFSphys04, 06, and 09; and LSM, radiation, and gravity wave drag parameterizations in all perturbed members. For the NSSL microphysics (Mansell 2010) members, SPP is applied using a parameter perturbation following a Latin hypercube sampling with multidimensional uniformity technique. The configuration for the membership of RRFSe2 and RRFSe2tl can be found in Table 5, where the members with $*$ next to them are used in the RRFSe2tl. The time-lagged members for both the RRFSe1tl and RRFSe2tl are lagged by 12-h.

A planned update of the RRFS is set to occur prior to the start of FFaIR. This will likely result in some performance differences in the RRFSp1 during the SFE compared to FFaIR. The update will back out a bug found in the PBL scheme in the version running during SFE. It will also change the timestep of the model from

Table 4: The member configuration for the RRFSe1 and RRFSe1tl (i.e. the single physics ensembles) that will be evaluated in the 2023 FFaIR. The members that will provide the 12-h time-lagged members for the RRFSe1tl have a $*$ next to their name.

Members	ICs	LBCs	Microphysics	PBL/SFC	LSM	Radiation	Shallow Cumulus
*RRFSp1 (cnt)	RRFS hybrid 3DEnVar	GFS	Thompson	MYNN/MYNN	RUC	RRTMG	n/a
*RRFS01	enkf1	GEFS m1	Thompson	MYNN/MYNN	RUC	RRTMG	n/a
*RRFS02	enkf2	GEFS m2	Thompson	MYNN/MYNN	RUC	RRTMG	n/a
*RRFS03	enkf3	GEFS m3	Thompson	MYNN/MYNN	RUC	RRTMG	n/a
*RRFS04	enkf4	GEFS m4	Thompson	MYNN/MYNN	RUC	RRTMG	n/a
*RRFS05	enkf5	GEFS m5	Thompson	MYNN/MYNN	RUC	RRTMG	n/a
RRFS06	enkf6	GEFS m6	Thompson	MYNN/MYNN	RUC	RRTMG	n/a
RRFS07	enkf7	GEFS m7	Thompson	MYNN/MYNN	RUC	RRTMG	n/a
RRFS08	enkf8	GEFS m8	Thompson	MYNN/MYNN	RUC	RRTMG	n/a
RRFS09	enkf9	GEFS m9	Thompson	MYNN/MYNN	RUC	RRTMG	n/a

Table 5: The member configuration for the RRFSe2 and RRFSe2tl (i.e. the multi-physics ensembles) that will be evaluated in the 2023 FFaIR. The members that will provide the 12-h time-lagged members for the RRFSe2tl have a $*$ next to their name.

Members	ICs	LBCs	Microphysics	PBL/SFC	LSM	Radiation	Shallow Cumulus
*RRFSp1 (cnt)	RRFS hybrid 3DEnVar	GFS	Thompson	MYNN/MYNN	RUC	RRTMG	n/a
*RRFSphys01	enkf1	GEFS m1	Thompson	H-EDMF/GFS	RUC	RRTMG	saSAS Shal
*RRFSphys02	enkf2	GEFS m2	Thompson	TKE-EDMF/GFS	RUC	RRTMG	saSAS Shal
*RRFSphys03	enkf3	GEFS m3	Thompson	MYNN/MYNN	RUC	RRTMG	saSAS Shal
*RRFSphys04	enkf4	GEFS m4	Thompson	TKE-EDMF/GFS	RUC	RRTMG	saSAS Shal
*RRFSphys05	enkf5	GEFS m5	NSSL	MYNN/MYNN	RUC	RRTMG	saSAS Shal
RRFSphys06	enkf6	GEFS m6	NSSL	H-EDMF/GFS	RUC	RRTMG	saSAS Shal
RRFSphys07	enkf7	GEFS m7	NSSL	TKE-EDMF/GFS	RUC	RRTMG	saSAS Shal
RRFSphys08	enkf8	GEFS m8	NSSL	MYNN/MYNN	RUC	RRTMG	saSAS Shal
RRFSphys09	enkf9	GEFS m9	NSSL	TKE-EDMF/GFS	RUC	RRTMG	saSAS Shal

60-s to 36-s. The timestep shift also applies to when the microphysics is called, changing from every 20-s to being called every 36-s, matching the model timestep. This update will also add SPPT as a source of dispersion to the ensemble members.

In addition to the RRFSp1, 5 FV3 CAM configurations will be provided for evaluation by the Center for Analysis and Prediction of Storms (CAPS) at the University of Oklahoma (OU). These will be referred to as RRFSp3-RRFSp7. Their configurations can be seen in Table 3. These configurations a nearly the same as the CAPS members provided last year for FFaIR but the initial conditions are from the GFS rather than from various members of the RDAS; refer to Table 1 in the 2022 FFaIR Final Report (Trojniak and Correia, Jr., 2023). Another difference is that the CAPS team have updated their version of the RRFS to the latest version release as of early May. The CAPS configuration referred to RRFSp4 last year, initialized from the GFS, will once again be called the RRFSp4 since this configuration will only differ from last year due to the model version upgrade.

CAPS will also be providing a 10 member, mixed physics ensemble. The configuration for the ensemble can be seen in Table 6 and like last year the ensemble will be referred to as CAPS_RRFSe. The Initial Conditions (ICs) and Lateral Boundary Conditions (LBC) will be forced by different GEFS[3] members. All CAPS data will only be provided for the 00z cycle.

3.2 Spatially-Aligned Mean

OU-CAPS will also be providing a mean product that has been in development and testing for a few years, called the Spatially-Aligned Mean (SAM). The SAM is calculated using a spatial alignment algorithm based on the Phase-Correcting Data Assimilation method used in Brewster (2003) for the spatial alignment of the background forecast to observations. The method was adapted to be used to align fields from ensemble members to one another. To do so, the domain is divided into overlapping patches and then the algorithm checks shifts of +/- 25 grid points in 2-dimensions for a minimum squared difference in fields (i.e. precipitation fields) including a penalty for larger offsets. These spatial alignment offset vectors are

[3]The NWS's Global Ensemble Forecast System.

Table 6: The member configuration for the CAPS_RRFSe that will be evaluated in the 2023 FFaIR.

Members	IC/LBCs	Microphysics	PBL	Surface	LSM
M0B0L2_PI	GEFS_m1	Thompson	MYNN	MYNN	RUC
M0B1L0_PI	GEFS_m2	Thompson	Shin-Hong	GFS	NOAH
M0B2L1_PI	GEFS_m3	Thompson	TKE-EDMF	GFS	NOAHMP
M0B0L0_PI	GEFS_m4	Thompson	MYNN	MYNN	NOAH
M0B2L2_PI	GEFS_m5	Thompson	TKE-EDMF	GFS	RUC
M1B0L2_PI	GEFS_m6	NSSL	MYNN	MYNN	RUC
M1B1L0_PI	GEFS_m7	NSSL	Shin-Hong	GFS	NOAH
M1B2L1_PI	GEFS_m8	NSSL	TKE-EDMF	GFS	NOAHMP
M1B0L0_PI	GEFS_m9	NSSL	MYNN	MYNN	NOAH
M1B2L2_PI	GEFS_m10	NSSL	TKE-EDMF	GFS	RUC

averaged after comparing all permutations of pairs in the ensemble to bring all the members to a common point. A visualization of the spatial alignment concept can be seen in Fig. 5 while a comparison of how the SAM and simple mean reflectivity could differ given the 5 ensemble forecasts can be seen in Fig. 6. The simple mean has broader footprint and reduced maximum than the individual fields, while the SAM was able to recover the maxima and better retain the shape.

Recently, extensive testing was done to explore the parameter space (size of patches, length scales applied, maximum offset allowed, etc) and to improve algorithm efficiency in massively parallel computing environment. The offsets were also tested at telescoping scales, applying large scale (meso-alpha scale, fronts and waves) corrections first, then applying smaller scale adjustments (meso-beta scale, closer to individual cells) using smaller patches. It was found that a two-step correction was most effective, with only very small changes with a third iteration. Thus the two-step correction (or pass) method will be applied to the SAM for FFaIR

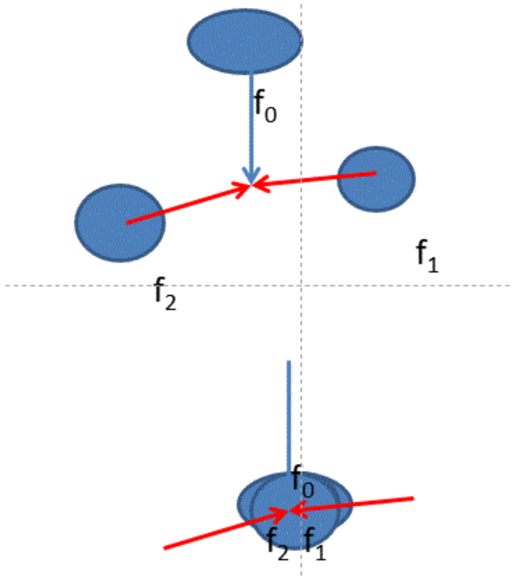

Figure 5: Conceptual diagram of spatial alignment of storm cell among 3 ensemble members. Shift vectors are applied to move the fields to a common central point.

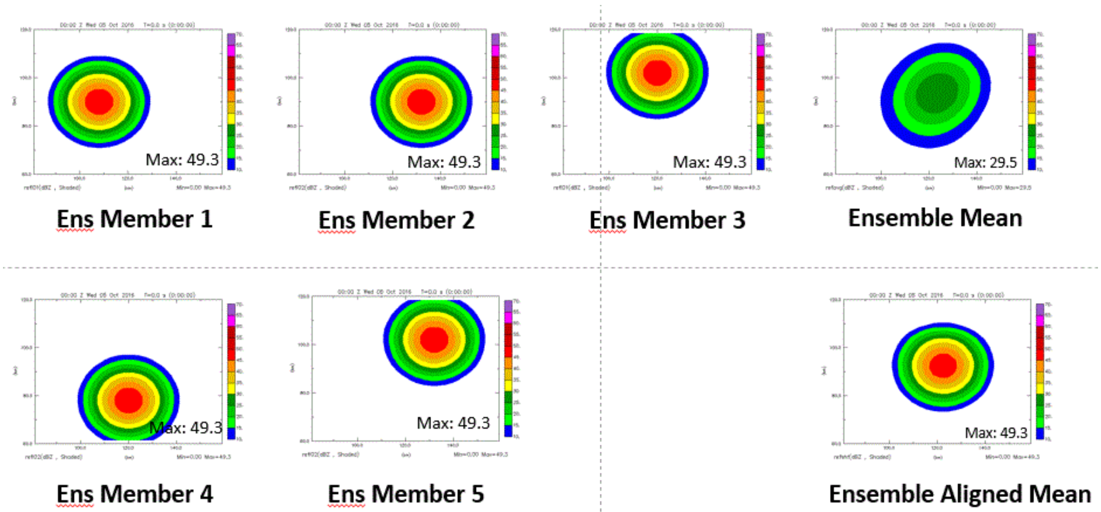

Figure 6: Demonstration using analytic cell with 5 different offsets as ensemble members (labeled as members 1-5). Right side of image shows the resulting simple mean field (top right) and SAM field (bottom right) given these members.

In addition to the ordinary SAM, CAPS will also provide the SAM-LPM, which is a product combining the SAM and LPM. For this product the ensemble members are first spatially aligned to one another then the LPM is computed from the shifted members. A comparison of the simple mean, SAM, and SAM-LAM for a 3-h precipitation forecast for Hurricane Ian can be seen in Fig. 7. Note the SAM is better able to capture the structure of the storm, including the open eye. Then applying the LPM technique using the precipitation distribution of the original members rescales values and refines the features.

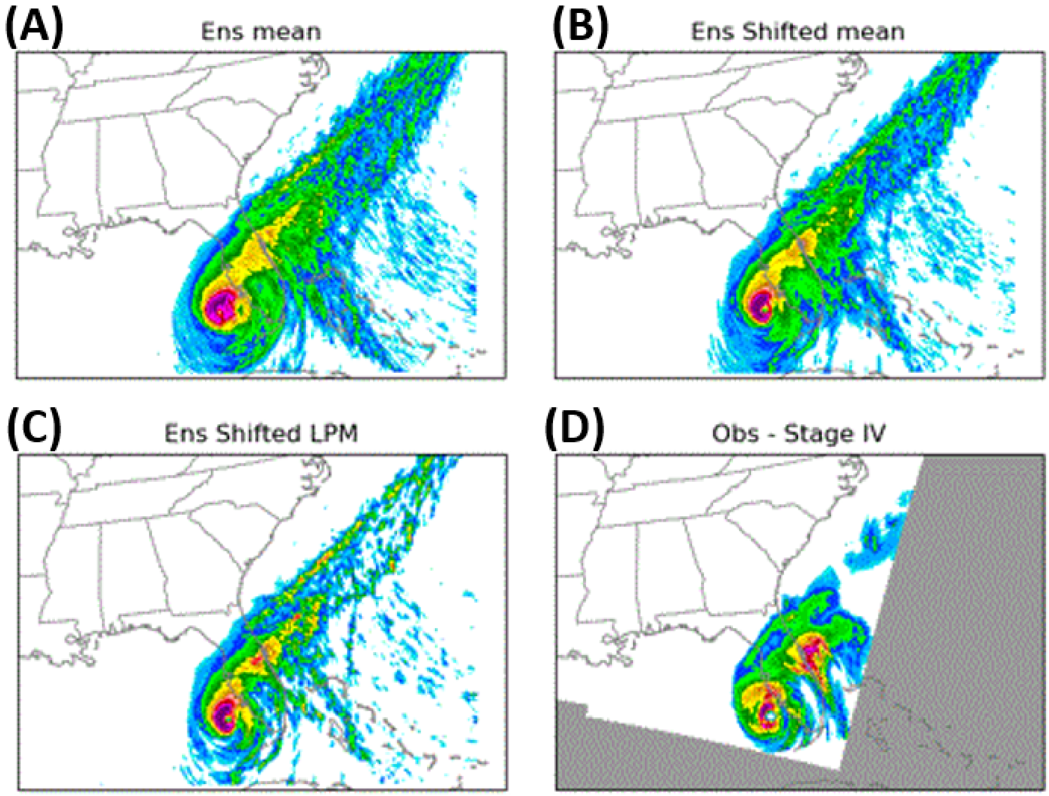

Figure 7: HREF 3-h precipitation forecasts for 15-h (27-h time-lagged members) forecasts of Hurricane Ian valid 15 UTC 28 September 2022 for the (A) ensemble mean, (B) SAM, and (C) SAM-LPM. The Stage-IV QPE verification is (D).

3.3 Machine Learning

Similar to last year, the CAPS team will be providing ensemble-based artificial intelligence/machine learning (AI/ML) products for rainfall prediction. Also like last year, the ML team from CSU will be providing multiple versions of their ERO First-Guess MLPs, including an updated version of the HRRR-based ERO.

3.3.1 CAPS MLP

The ensemble-based MLPs that will be provided by CAPS uses a 12-member combined ensemble consisting of 4 real-time FV3-CAM forecasts produced by the CAPS_RRFSe and 8 members of the HREF ensemble (which includes 4 HREF members and their corresponding 12-hour time-lag counterparts). This is referred to as the HREF+. The members and their time lagged counterparts from the HREF are the HRRR, NAMnest, hiresw_arw, and hiresw_nssl. Because HREF data are only available for 48 hours of forecast time, and due to the need to include 12-hour time-lagged HREF members, these AI rainfall forecast products are being produced for six-hour intervals from 0-36 hours of forecast time. This year, the daily HREF+ forecasts will be generated for rainfall exceeding thresholds of 0.5, 1, and 2 inches during each 6-h period from 0 to 36 hours of forecast time.

To produce these probabilistic ML forecasts, individual ensemble member ML forecasts are generated using a U-Net (a deep learning approach which uses convolutional neural networks and is designed to identify spatial patterns in images). These individual member ML forecasts use a set of 23 input variables from their corresponding member of the HREF+ combined ensemble; variables used include wind, temperature, and moisture information at different vertical levels, as well as predictions of reflectivity, QPF, and precipitable water. To produce a gridded forecast covering the full CONUS, the U-Net considers data over 64x64 patches, generating predictions on these patches which are stitched together (with slight overlap to mitigate the impacts of patch-boundary discontinuities) for each member to produce a full CONUS forecast. A neighborhood maximum ensemble probability (NMEP) is applied to this ensemble of U-Net predictions, using the three thresholds of interest (0.5, 1, and 2 inches) to produce the final ML forecast products.

3.3.2 CSU ERO MLPs

A focus of evaluation for the CSU Day 1 ERO MLPs this year will be on comparison of the GEFS-based MLPs trained on ARI exceedances and the one trained on the Unified Flooding Verification System (UFVS). The UFVS is an observational dataset used by both WPC and CSU for ERO verification (Erickson et al., 2019). Both systems train RF models regionally, and make predictions in those regions which are then stitched together to make a full-CONUS prediction. The 2022 ARI-based version, hereafter FV3GEFSR, was recommended for transition to WPC and represents a baseline system for CSU ERO evaluation. Continued evaluation of the UFVS-based version, hereafter UFVSGEFSR, is necessary before transitioning the system as there are discrepancies between qualitative and quantitative evaluations performed within and outside of FFaIR. Examples of the FV3GEFSR and UFVSGEFSR can be seen in Figs. 8C-D along with the current operational GEFS-based MLP ERO (Fig. 8B) and the WPC Day 1 ERO (Fig. 8A).

An additional focus will be evaluated the updated version of the Day 1 HRRR-based ERO MLP. This updated versions has several aspects of improvement relative to that evaluated at FFaIR 2022. In particular, (1) the system uses a three-year training period that is exclusively based on HRRRv4 (instead of a mix of HRRRv3 and HRRRv4). (2) The system uses hourly predictors instead of 3-h predictors. (3) The HRRR predictors are aggregated spatially, using a spatial max for accumulated precip / 1-h max updraft helicity / 700 hPa updraft speed, a spatial min for 1-h min updraft helicity, and a spatial mean for all the remaining environmental predictors. An comparison of the 2022 version and 2023 version can be seen in Fig. 9.

The improvements to the HRRR-based system are based on sensitivity experiments carried out during the 2022-2023 period. These experiments showed a slight degradation in forecast skill coming from a mismatch in HRRR versions used in training vs. that used in the daily forecasts, compared to using a consistent

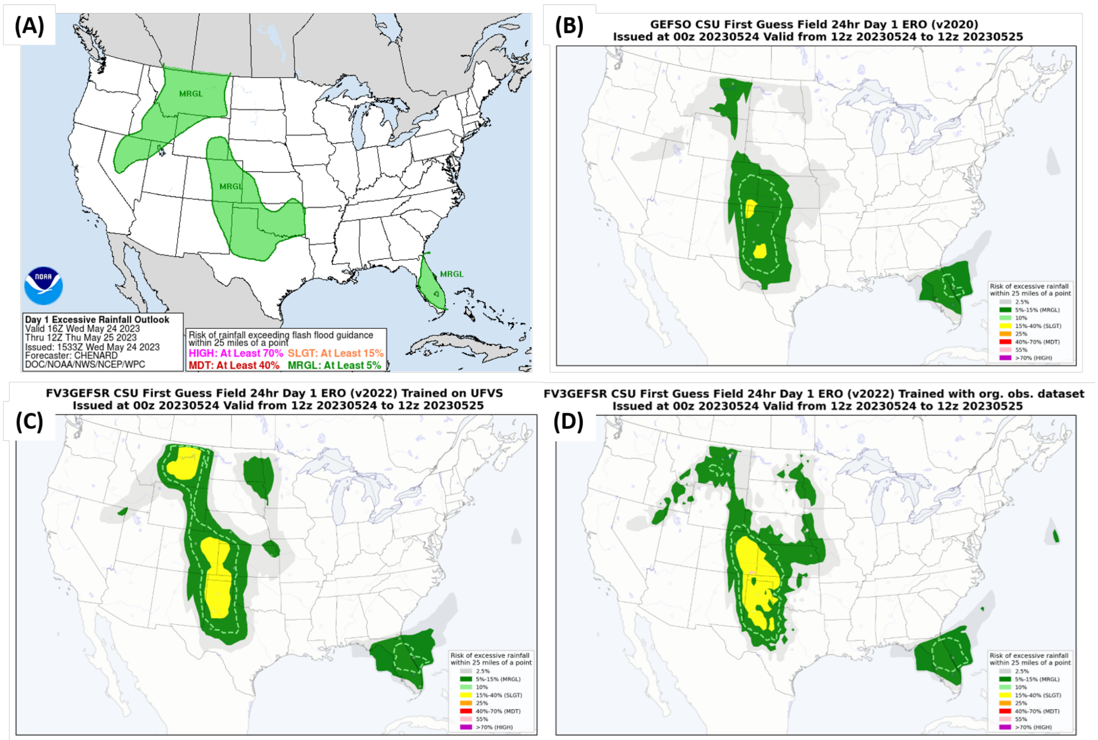

Figure 8: (A) WPC Day 1 ERO valid 16 UTC 24 May to 12 UTC 25 May 2023. (B) GEFSO, (C) 00z FV3GEFSR, and (D) 12z FV3GEFSR valid 12 UTC 24 May to 12 UTC 25 May 2023. The WPC ERO risk probabilities contoured Marginal: 5% green, Slight: 15% yellow, Moderate: 40% red and High: 70% purple/pink. On the CSU MLP EROs, addition contours are: 2.5% gray, 10% light green, 25% orange, and 50% pink.

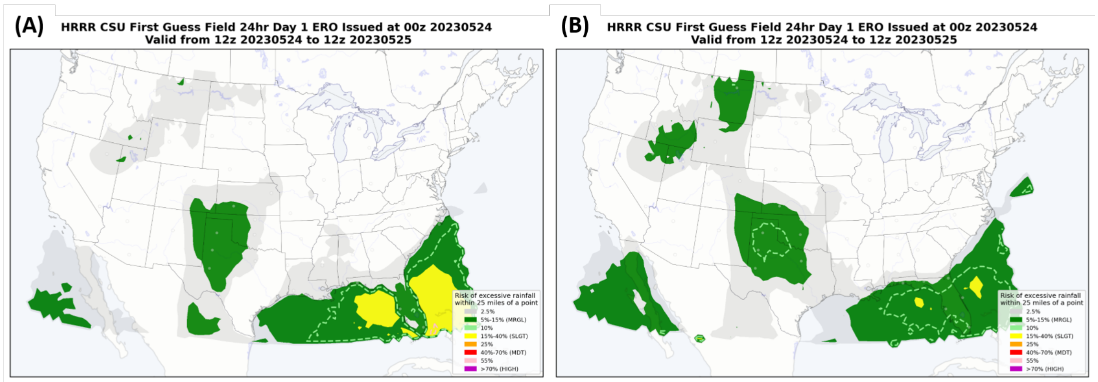

Figure 9: The HRRR-based Day 1 ERO version (A) 2022 and (B) 2023 valid 12 UTC 24 May to 12 UTC 25 May 2023. The WPC ERO risk probabilities contoured Marginal: 5% green, Slight: 15% yellow, Moderate: 40% red and High: 70% purple/pink; addition contours are: 2.5% gray, 10% light green, 25% orange, and 50% pink.

HRRR version. Other experiments also revealed benefit from using hourly HRRR output instead of 3-h output, particularly in the central and interior western US, and general benefit in most regions from using spatial aggregation of predictors.

3.4 Satellite Products

FFaIR will once again be partnering with the CIRA Satellite team at CSU led by John Forsythe. This group focuses on providing perceptible water (PWAT or PW) data to forecasters via their derived satellite products. Multiple products developed by them have been transitioned or are in the process of being transitioned into operations at the recommendation of FFaIR and WPC. This includes the Advected Layer Precipitable Water (ALPW) product that is being transitioned to operations in Fall 2023. For evaluation, they will be providing two new derived satellite products: (1) an hourly percentile ranking of ALPW by layer and (2) a Layered water Vapor Transport (LVT). Both of these products are created hourly and are available by :40 past the hour.

3.4.1 Percentile Ranking

ALPW is derived from passive microwave soundings which are moved to a common time using GFS winds. More details are in (Gitro et al., 2018). An example of an atmospheric river (AR) approaching the California coast can be seen in Fig. 10. Missing areas (black) are due to gaps in polar orbiter swaths, precipitation, or high terrain impacting the layer. The new ALPW percentile ranking product to be evaluated in FFaIR uses monthly LPW fields dating back to 2013 are used to rank the current ALPW field in terms of percentiles, with monthly background fields are used as the background. Using the same event in Fig. 10, Fig. 11 provides an example of the ranked percentile ALPW product. The 95^{th}, 99^{th} and maximum values are shown by red shading, as well as the $<5^{th}$ percentile for very dry values.

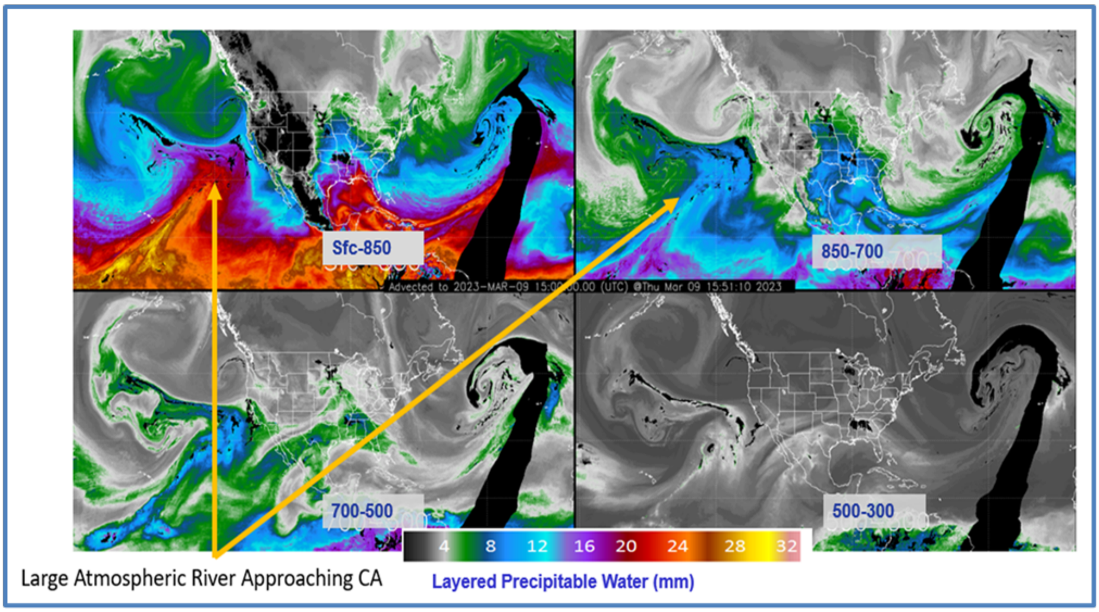

Figure 10: ALPW product for 15 UTC 9 March 2023, with a large atmospheric river is approaching California.

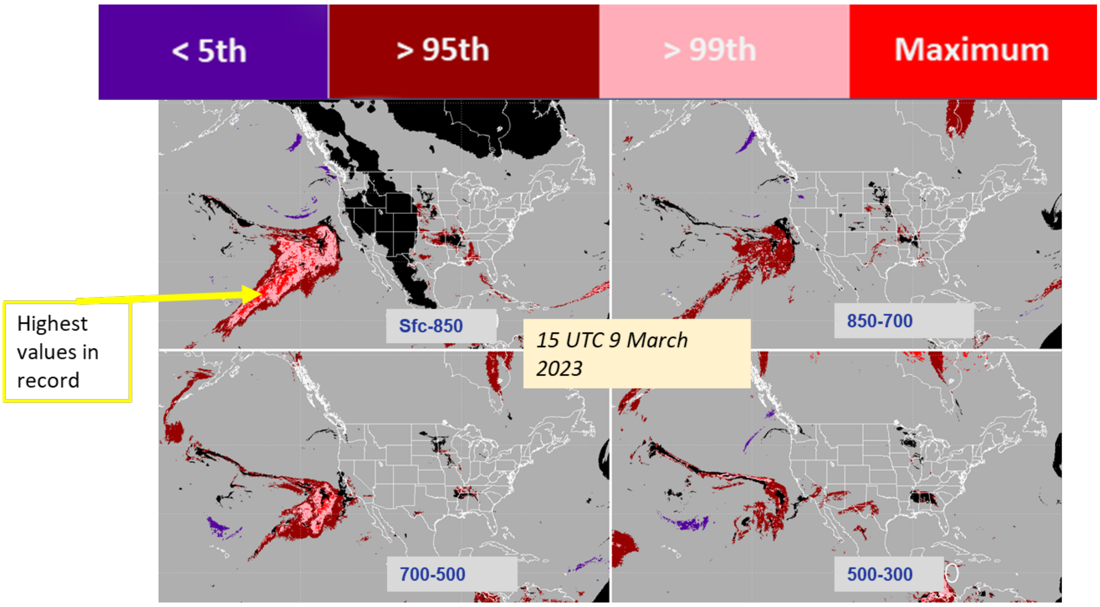

Figure 11: Percentile rankings for ALPW shown in Fig. 10. The atmospheric river is reflected at all four layers by $<95^{th}$ percentile values, including some 10-y maxima in the surface-850 mb layer.

3.4.2 Layered Vapor Transport

Integrated Vapor Transport (IVT) is commonly used to rank the intensity of atmospheric rivers. IVT between 1000 and 300 mb is defined as:

$$IVT = \sqrt{(\frac{1}{g}\int_{1000}^{300} qu\,dp)^2 + (\frac{1}{g}\int_{1000}^{300} qv\,dp)^2}$$

where q is the layer-averaged specific humidity, u and v are the layer-averaged easterly and northerly wind components, g is the acceleration, and dp is the pressure difference between two adjacent pressure levels. The units of IVT are $kg/m/s$. It is expected that during FFaIR forecasters will use IVT from various model sources.

ALPW, with some manipulations to input the correct units, can be used to derived layered water vapor transport (LVT). GFS u-v winds at 900, 800, 600, and 400 mb are used to multiply the ALPW values to derive LVT, which has the same units as IVT. An example LVT field for the atmospheric river shown in Fig. 10 can be seen in Fig. 12. LVT would typically be at a maximum in the surface-850 layer due to higher water vapor amounts, but due to higher-level moisture or higher wind speeds it could be higher at higher layers. It is hypothesized that LVT might help forecasters to interpret IVT and be especially helpful in mountainous terrain, where water vapor can be lifted to fuel precipitation.

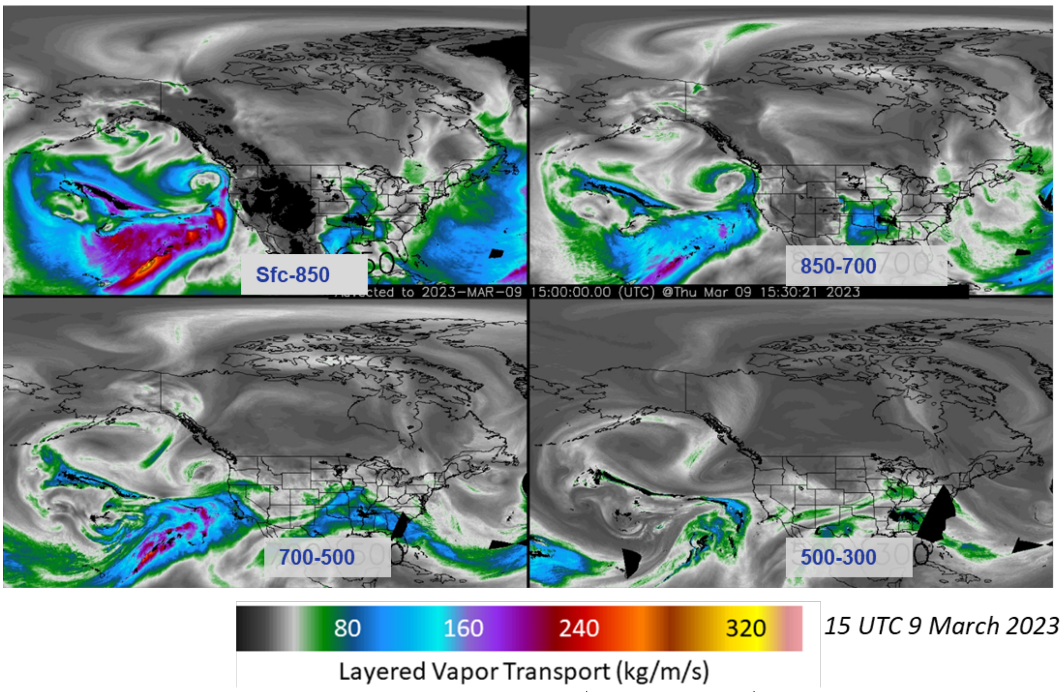

Figure 12: Four-layer water vapor transport (aka the LVT) for same data and time in Fig. 10 and Fig. 11. The black areas missing due to terrain or precipitation.

References

Brewster, K. A., 2003: Phase-correcting data assimilation and application to storm-scale numerical weather prediction. part i: Method description and simulation testing. *Mon. Wea. Rev.*, **3**, 480–492, https://doi.org/https://doi.org/10.1175/1520-0493(2003)131\%3C0480:PCDAAA\%3E2.0.CO;2.

Erickson, M. J., J. S. Kastman, B. Albright, S. Perfater, J. A. Nelson, R. S. Schumacher, and G. R. Herman, 2019: Verification results from the 2017 hmt–wpc flash flood and intense rainfall experiment. *J. Appl. Meteor. Climatol.*, **58 (12)**, 2591 – 2604, https://doi.org/10.1175/JAMC-D-19-0097.1, URL https://journals.ametsoc.org/view/journals/apme/58/12/jamc-d-19-0097.1.xml.

Gitro, C. M., and Coauthors, 2018: Using the multisensor advected layered precipitable water product in the operational forecast environment. *J. Operational Meteor.*, **6**, 59–73.

Martinez-Villalobos, C., and J. D. Neelin, 2018: Shifts in precipitation accumulation extremes during the warm season over the united states. *Geophys. Res. Lett.*, **45 (16)**, 8586–8595, https://doi.org/https://doi.org/10.1029/2018GL078465.

Trojniak, S., and J. Correia, Jr., 2023: 2022 flash flood and intense rainfall experiment: Findings and results. Tech. rep., NCEP WPC-HMT. URL https://www.wpc.ncep.noaa.gov/hmt/2022_FFaIR_Final_Report.pdf.

Visser, J. B., S. Kim, C. Wasko, R. Nathan, and A. Sharma, 2022: The impact of climate change on operational probable maximum precipitation estimates. *Water Resources Research*, **58 (11)**, https://doi.org/https://doi.org/10.1029/2022WR032247.