FV3-LAM CAM Ensemble Consensus and Machine Learning Products for Predicting Heavy Rain in the Hydrometeorology Testbed Experiments

Keith Brewster
Nate Snook, Chang Jae Lee[1],
Phillip Spencer, Tim Supinie, Ming Xue

Center for Analysis and Prediction of Storms
University of Oklahoma, Norman, Oklahoma
[1]Korea Meteorological Administration, Seoul, Korea

2023 FFaIR Seminar
July 25, 2023

Hydrometeorology Testbed

- **Hydrometeorology Testbed R2O-O2R Experiments**
  - Organized by the NOAA Weather Prediction Center
  - Bring together researchers & operational forecasters
  - FFaIR (June-August)
    - Evaluate new products for flash flood and excessive rainfall forecasts
  - Winter Weather Experiment (November – March)
    - Evaluate new products for snowfall forecasts

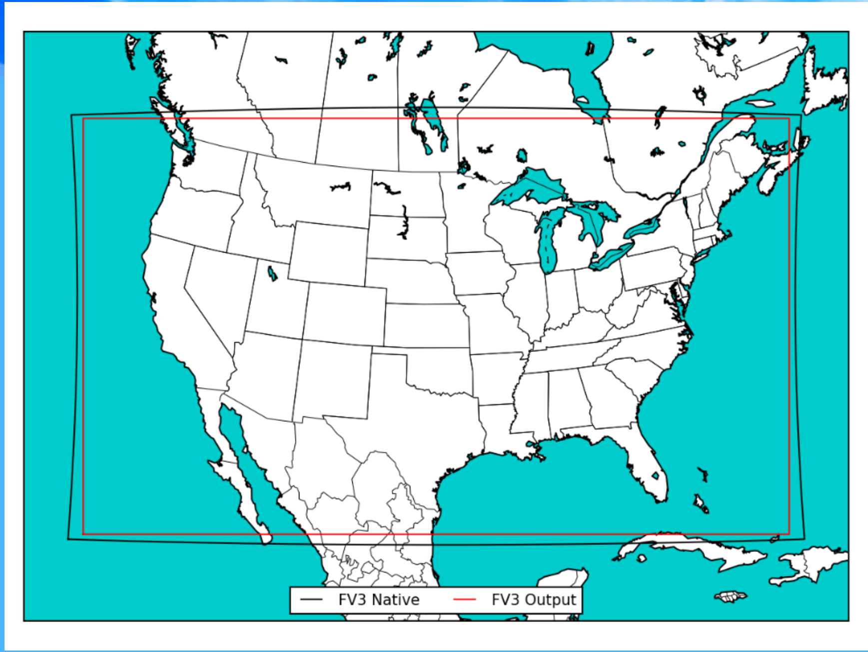
- **CAPS Contribution**
  - HMT participant since 2016 (HWT since 2007)
  - Multi-member 3-km CONUS CAM Ensemble Forecasts
  - Ensemble Consensus Products
  - Participate in Forecasting Exercises (EROs, MRTP, etc)

2022 FFaIR Real-Time Ensembles

Research Goals

- Test various FV3-LAM Physics Combinations
- Contribute to RRFS design & testing (including 2022 HWT SFE)
- Develop and Evaluate Ensemble Consensus Methods
  - LPM Mean
  - Spatial-Aligned Mean
  - Machine Learning Probabilistic Products

2022 HMT FFaIR Configurations



- 16 FV3-LAM members
- 3 km grid spacing CONUS grid
- 84-hr forecasts initialized at 00Z
- Code: Latest UFS FV3-LAM Short Range Weather App 1.0.1 plus NSSL microphysics
  Base code & grid same as EMC FV3
- Run on Frontera at the Texas Advanced Computing Center (TACC)

FRONTERA

TACC ACCESS Advancing Innovation

CAPS

2022 FFaIR 16-Member Ensemble

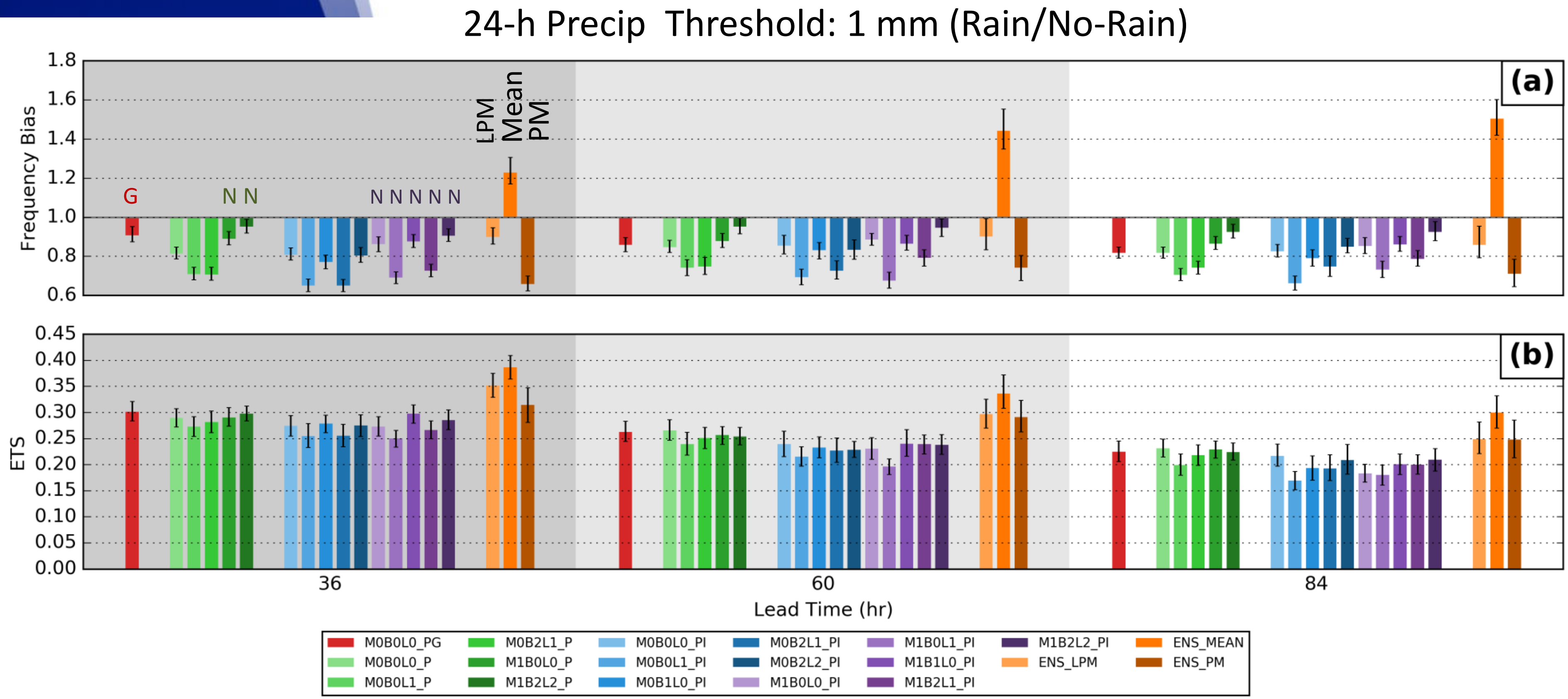
Naming

M: Microphysics
B: Boundary Layer
L: Land Sfc Model

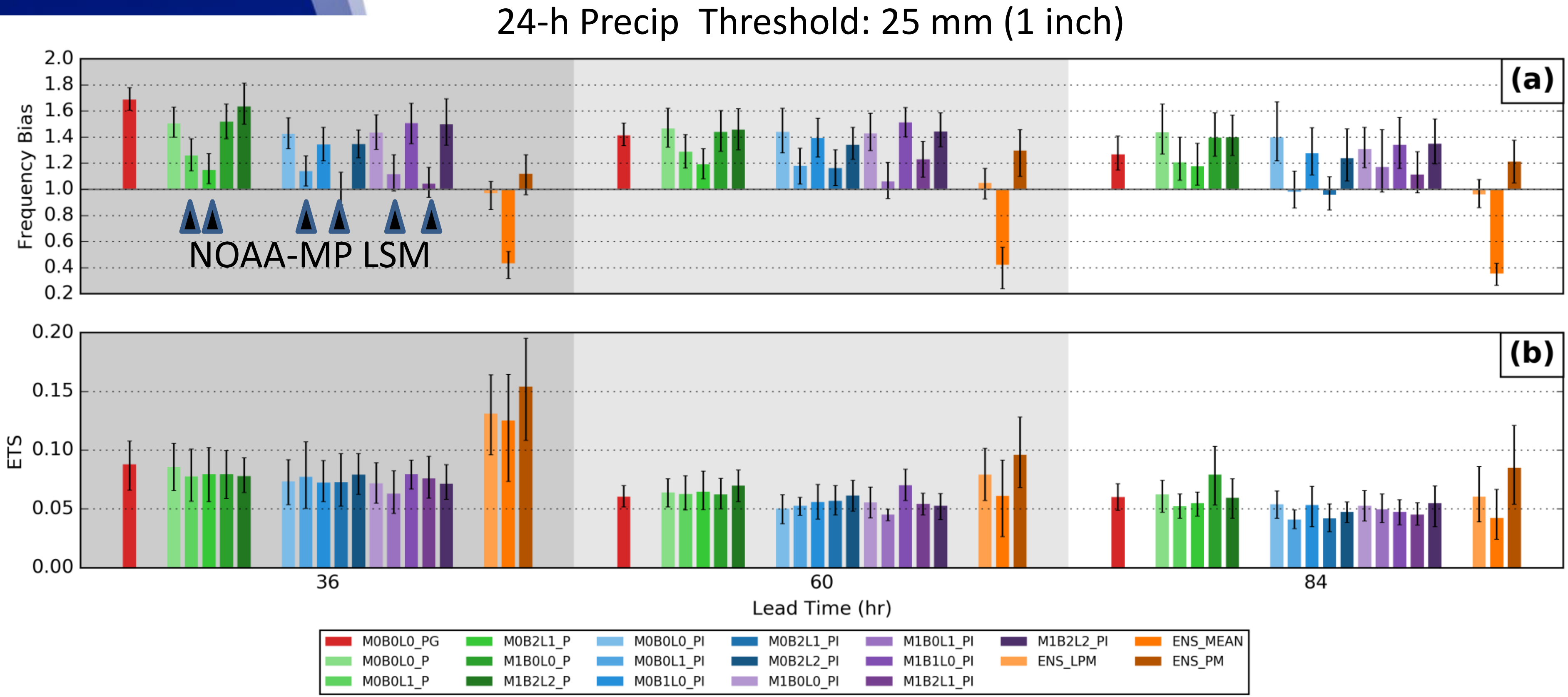
PG: GFS Initial/Bndy Cdx
P: GSL EnKF
PI: Initial perturbations

Experiment	Microphysics	PBL	Surface	LSM	IC/LBC (like system)
GFS IC for Baseline Configuration					
M0B0L0_PG	Thompson	MYNN	MYNN	NOAH	GFS/GFS (RRFSv0)
Multi-Physics Core Configurations, Same IC/LBC					
M0B0L0_P	Thompson	MYNN	MYNN	NOAH	RRFShybrid/GFS (RRFSv0)
M1B0L0_P	NSSL	MYNN	MYNN	NOAH	RRFShybrid/GFS (WoF)
M0B0L1_P	Thompson	MYNN	MYNN	NOAHMP	RRFShybrid/GFS (RRFS)
M1B2L2_P	NSSL	TKE-EDMF	GFS	RUC	RRFShybrid/GFS (Mixed)
M0B2L1_P	Thompson	TKE-EDMF	GFS	NOAHMP	RRFShybrid/GFS (GFS16)
Physics + IC Perturbation Ensemble					
M0B0L0_PI	Thompson	MYNN	MYNN	NOAH	RRFSenkf01/GEFS_m1
M0B1L0_PI	Thompson	Shin-Hong	GFS	NOAH	RRFSenkf02/GEFS_m2
M0B2L1_PI	Thompson	TKE-EDMF	GFS	NOAHMP	RRFSenkf03/GEFS_m3
M0B0L1_PI	Thompson	MYNN	MYNN	NOAHMP	RRFSenkf04/GEFS_m4
M0B2L2_PI	Thompson	TKE-EDMF	GFS	RUC	RRFSenkf05/GEFS_m5
M1B0L0_PI	NSSL	MYNN	MYNN	NOAH	RRFSenkf06/GEFS_m6
M1B1L0_PI	NSSL	Shin-Hong	GFS	NOAH	RRFSenkf07/GEFS_m7
M1B2L1_PI	NSSL	TKE-EDMF	GFS	NOAHMP	RRFSenkf08/GEFS_m8
M1B0L1_PI	NSSL	MYNN	MYNN	NOAHMP	RRFSenkf09/GEFS_m9
M1B2L2_PI	NSSL	TKE-EDMF	GFS	RUC	RRFSenkf10/GEFS_m10

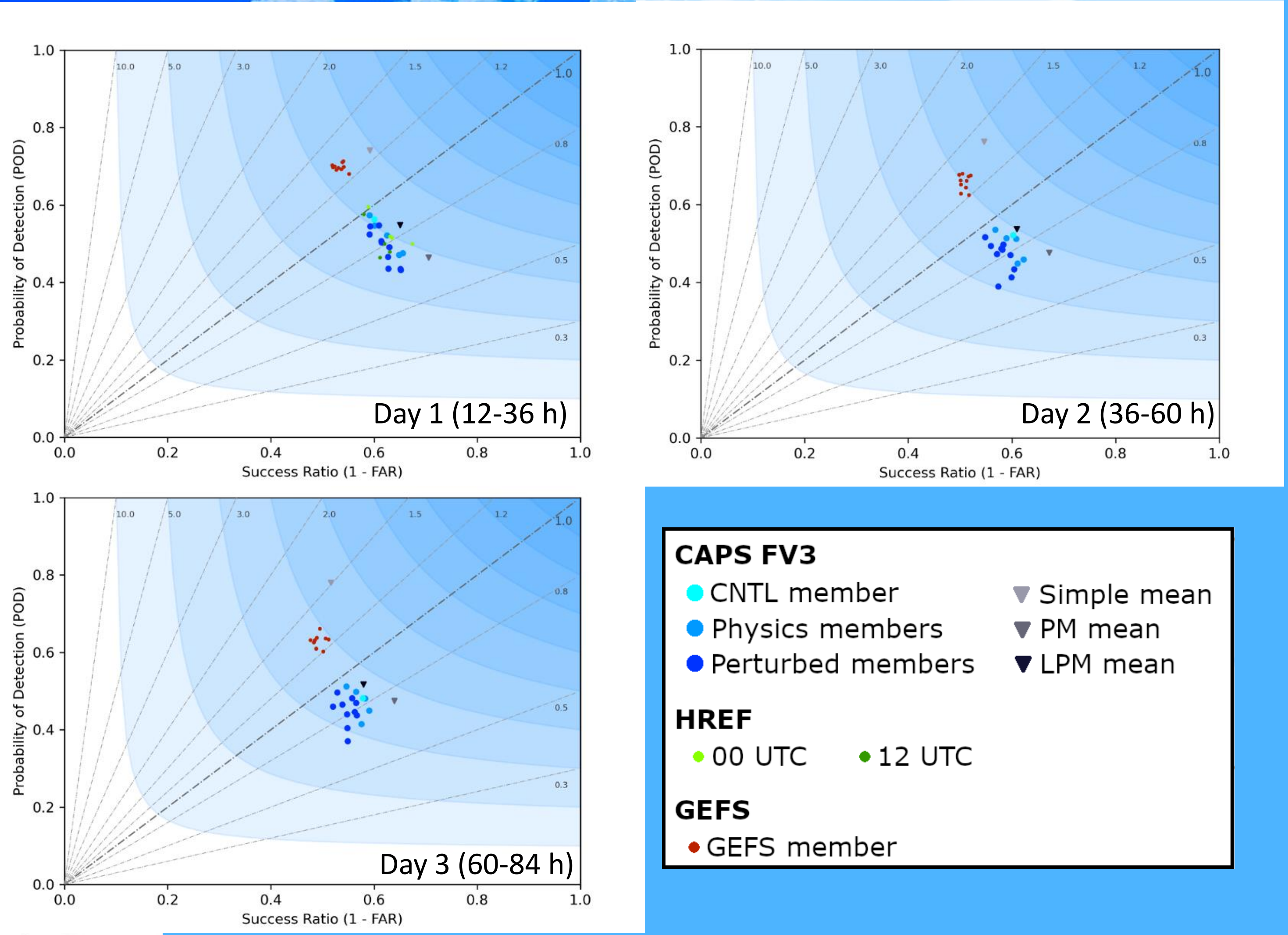
2022 FFaIR Precip Verification 1 mm



2022 FFaIR Precip Verification 25 mm



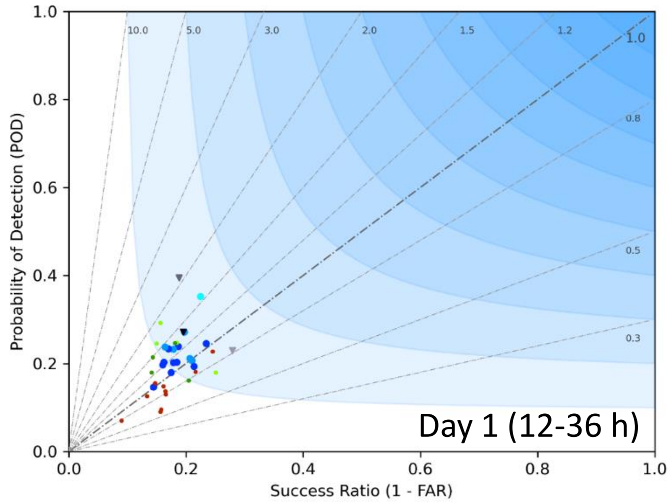
HMT FFaIR 2022
1 mm threshold
rain/no-rain

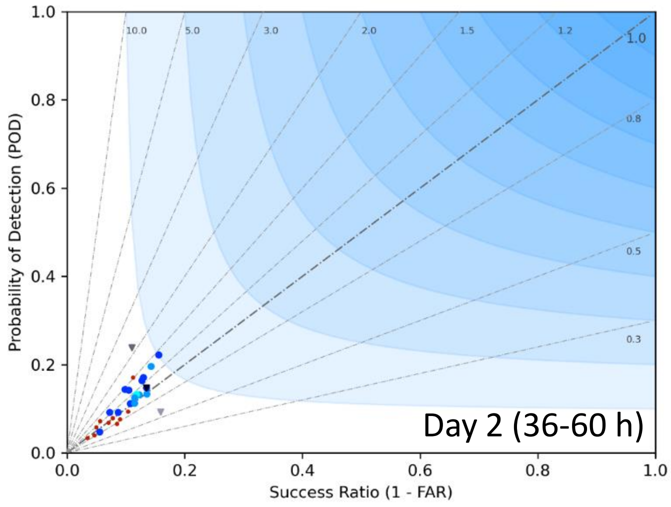


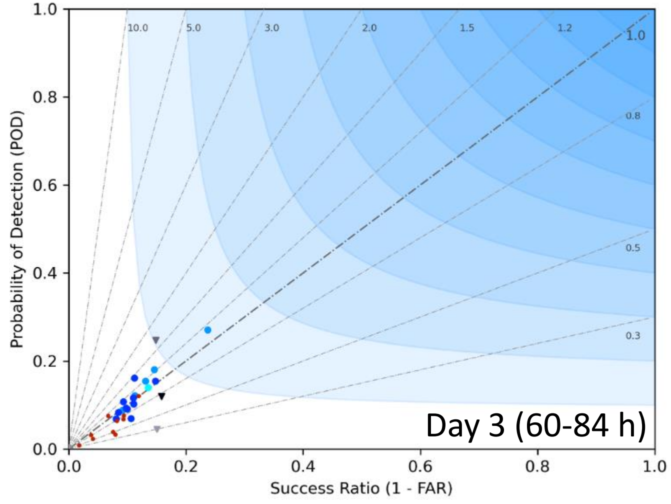
HMT FFaIR 2022

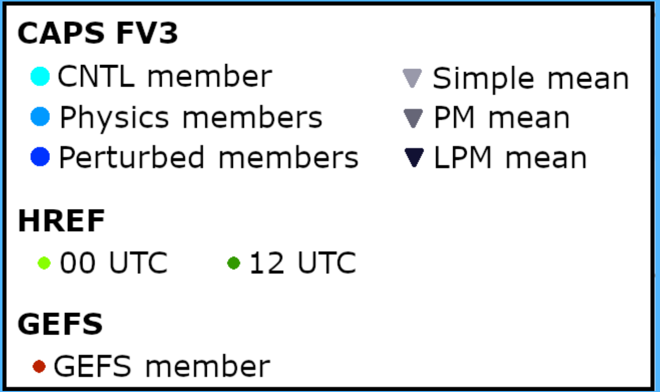
25 mm threshold

30 km neighborhood



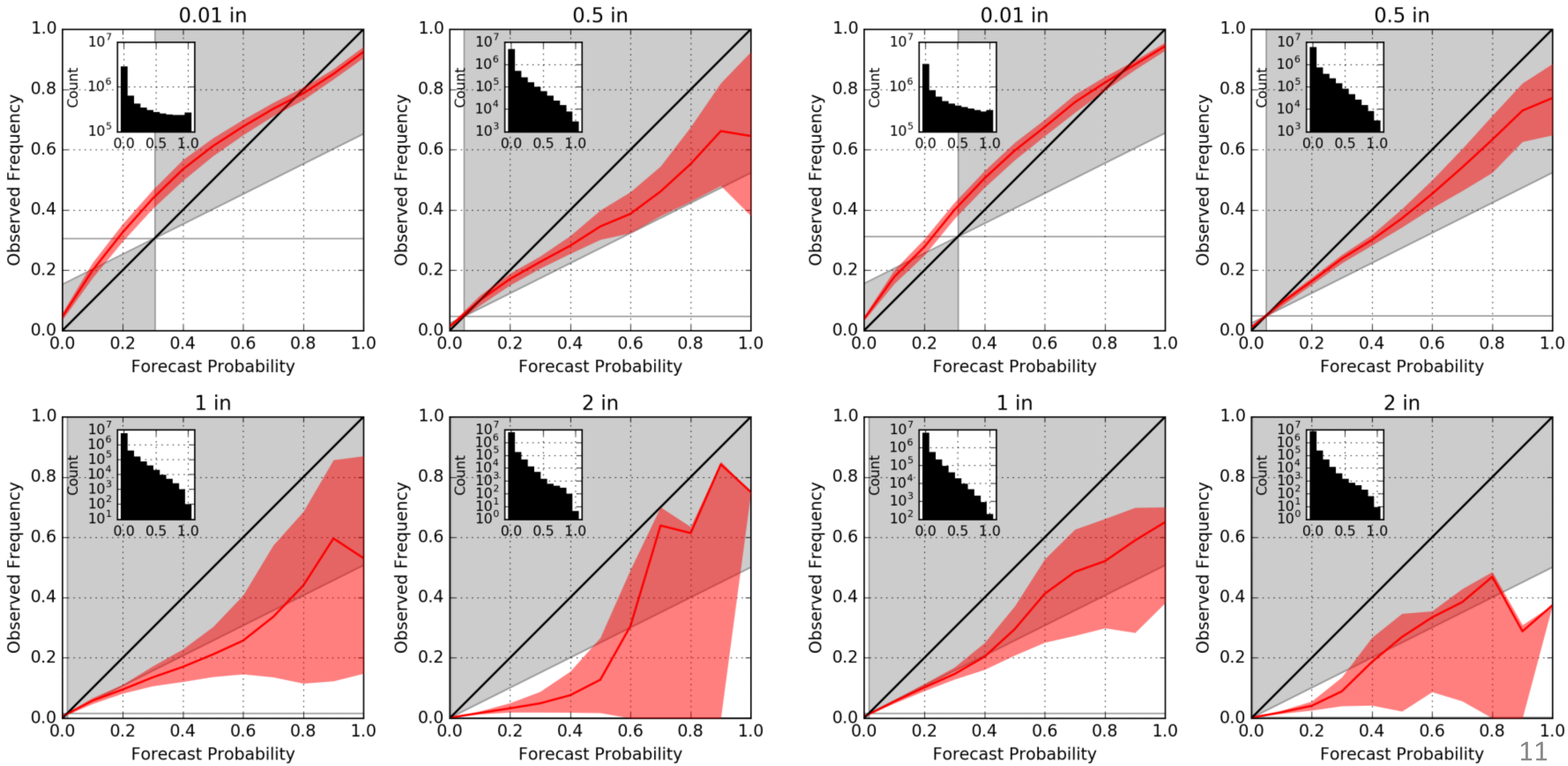






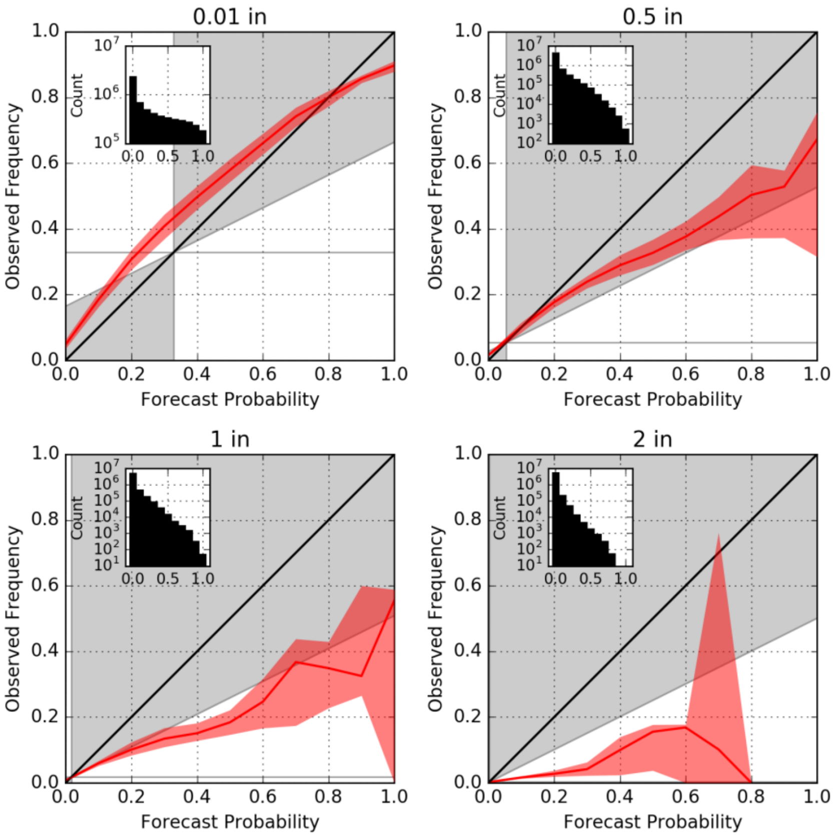
2022 FFaIR 24h Precipitation Reliability

36 h CAPS FV3-LAM **36 h HREF**

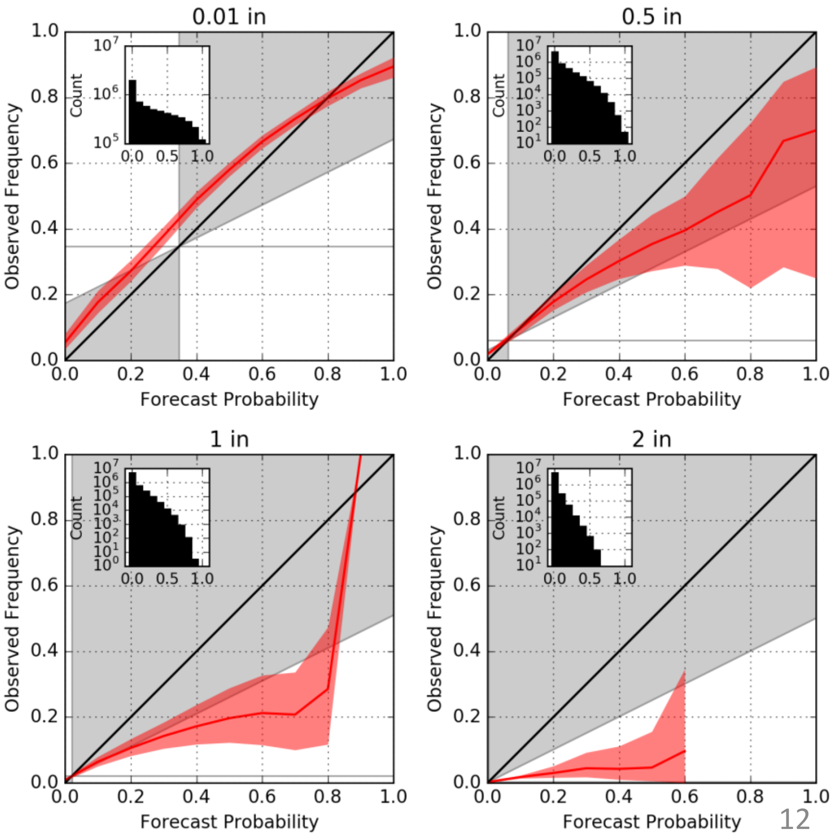


2022 FFaIR 24h Precipitation Reliability

60 h CAPS FV3-LAM



84 h CAPS FV3-LAM

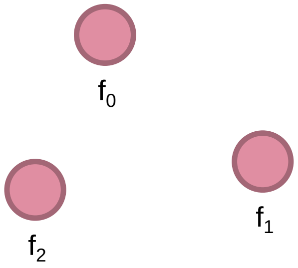


Spatially Aligned Mean

PM and LPM are focused on **the intensity of the fields**

It is common to have Convection Initiation (CI) location and propagation speed differences among models.
To better preserve the spatial structures of the fields: Spatially Aligned Mean

1. Consider three separate forecasts of rain:



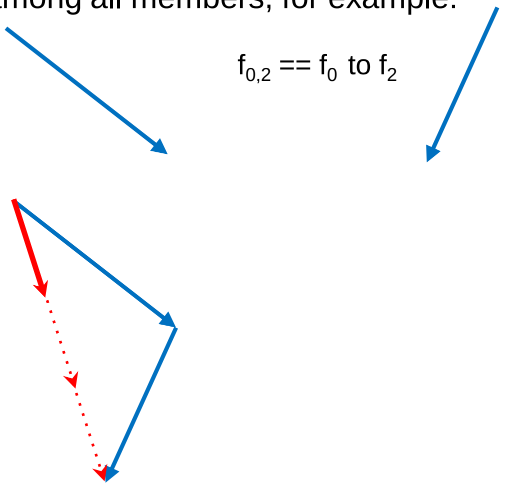
2. Determine individual spatial shifts among all members, for example:

$f_{0,0}$ == f_0 to f_0 = 0

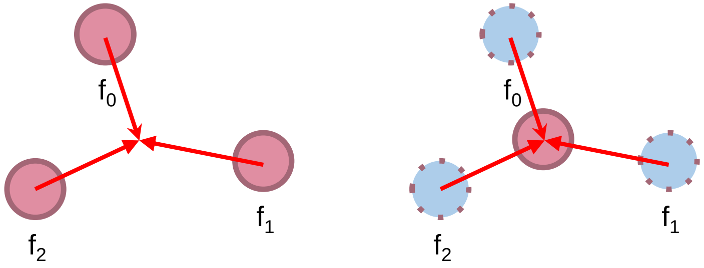
$f_{0,1}$ == f_0 to f_1

$f_{0,2}$ == f_0 to f_2

Correction of f_0 f_{0c}= ($f_{0,0}$ + $f_{0,1}$ + $f_{0,2}$) / 3



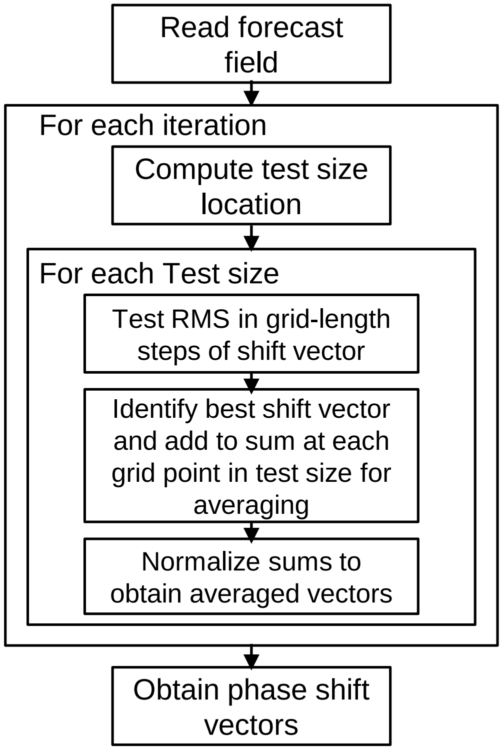
3. Calculate point-wise mean after spatial alignment completed:



Spatially Aligned Mean – the Algorithm

Based on Phase-Correcting Data Assimilation (Brewster, 2003),
a method for spatial alignment of background forecast to observations

Phase Shift Algorithm



1) Domain divided into overlapping patches (test size)

2) For each patch (test size) :

- **Check offsets of +/- 25 grid points in x,y directions**
- **Find the best shift vector which minimize RMS differences between each pair of members including a penalty for larger offset distances**

3) Average the shift vectors among overlapping patches

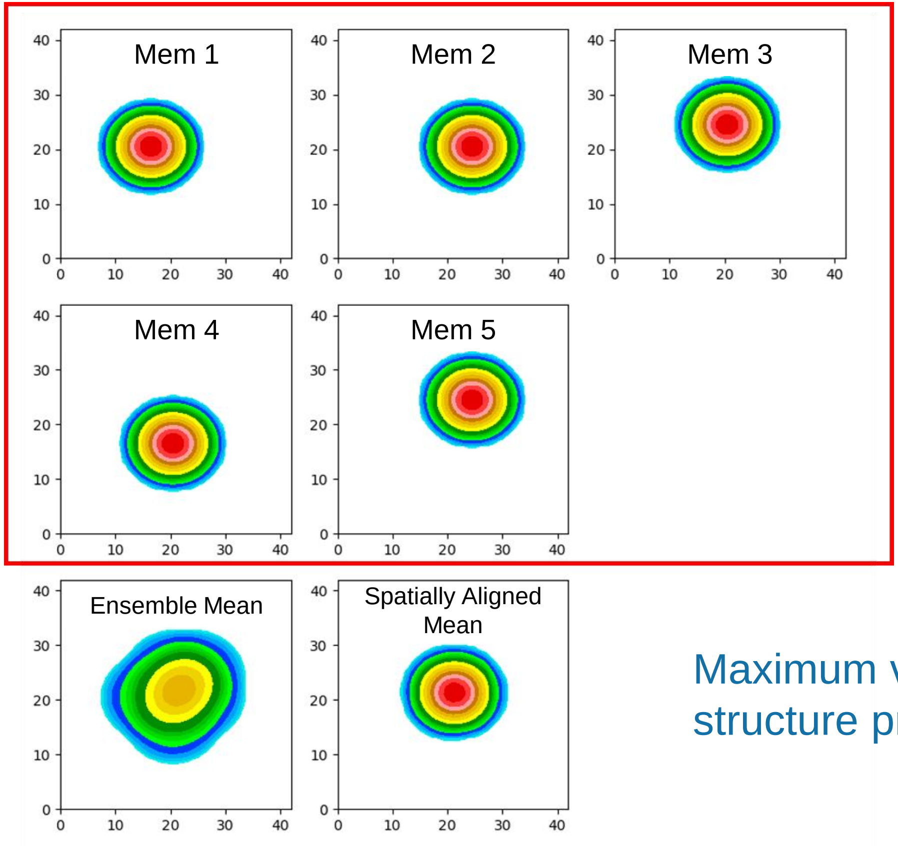
4) Can be applied in multiple steps (iteration) with decreasing patch size (test volume) to correct synoptic scale, mesoscale, storm scale)

- **In this research, 2 steps were applied**
- **1^{st} step's patch size was 600km (synoptic) and 2^{nd} step's patch size was 225km (mesoscale)**

5) Move field using obtained shift vectors and
6) Restore the intensity with the PDF from the original field

Spatially Aligned Mean – The Algorithm

Analytic case example



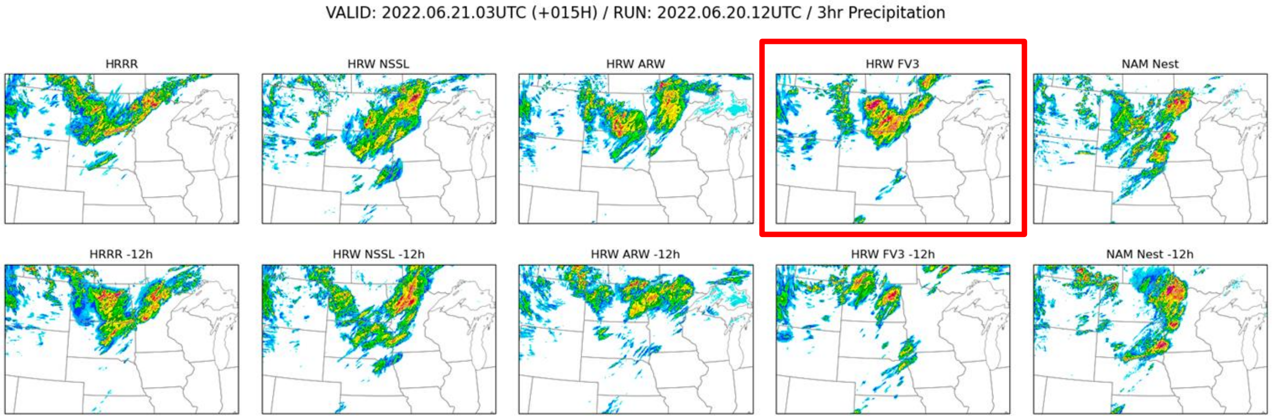
Maximum value *and* structure preserved

Spatially Aligned Mean – the Algorithm

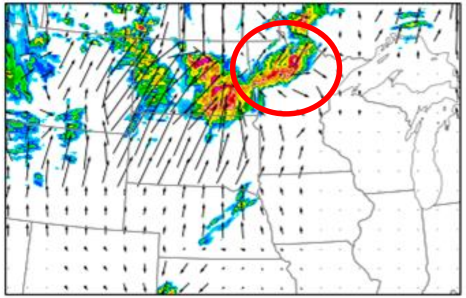
A single vector is found for each patch, but patches overlap and more local variation can be added in following iterations

- Thereby features can be **stretched**, rotated, and contracted

Stretching example:



Individual HREF Members

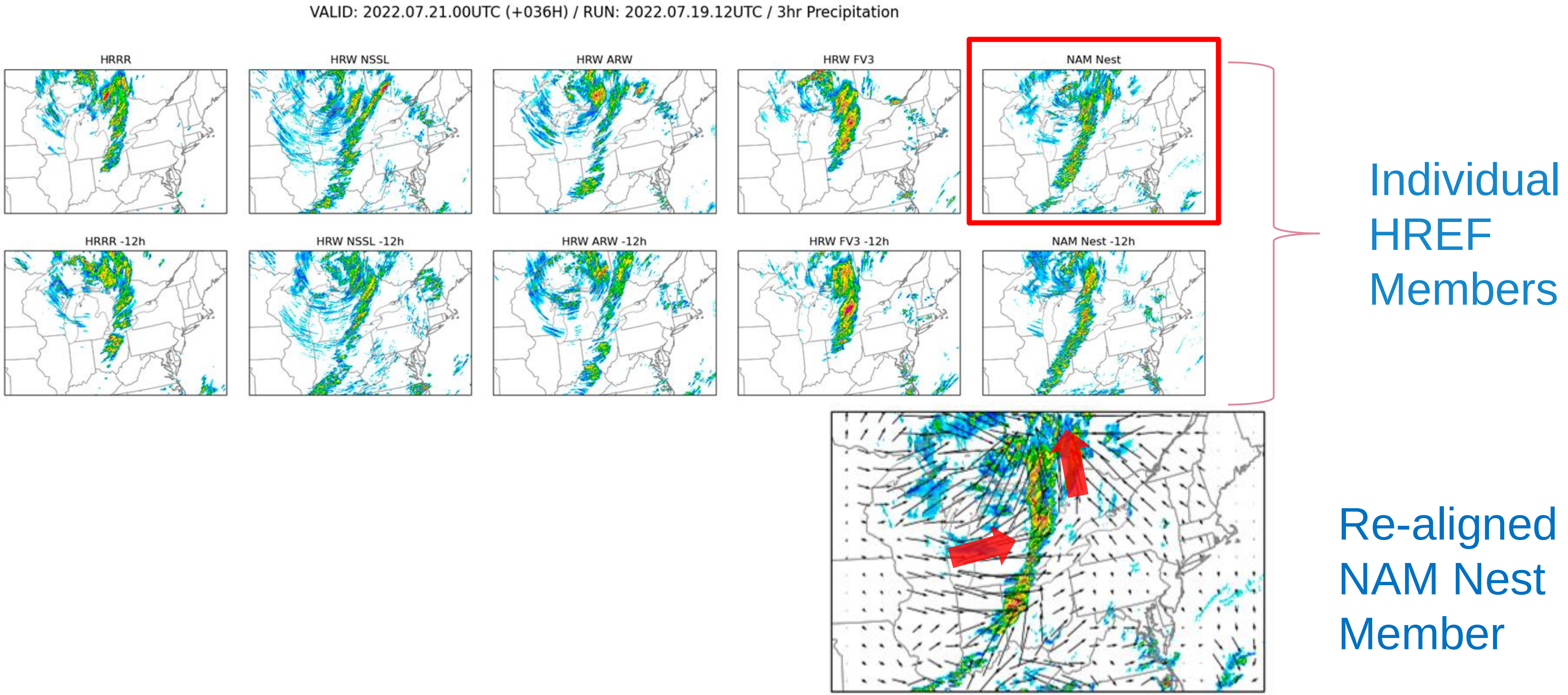


Re-aligned FV3 Member

Spatially Aligned Mean – the Algorithm

Each patch moves one direction, but they overlap with nearby patches and can be applied again with decreasing patch size
- Therefore features can be stretched, **rotated**, and contracted

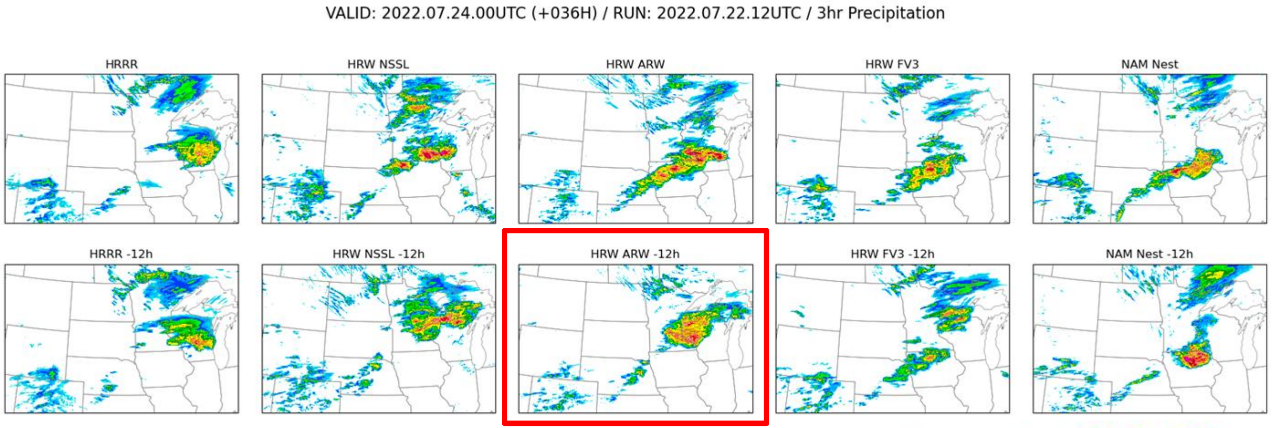
Rotation example:



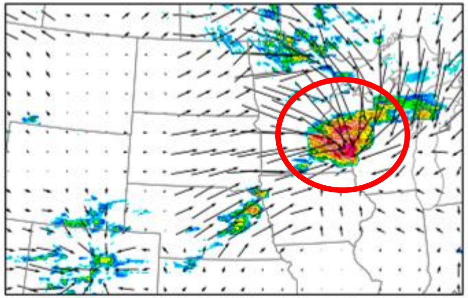
Spatially Aligned Mean – the Algorithm

Each patch moves one direction, but they overlap with nearby patches and can be applied again with decreasing patch size
 - Therefore features can be stretched, rotated, and **contracted**

Contraction Example:



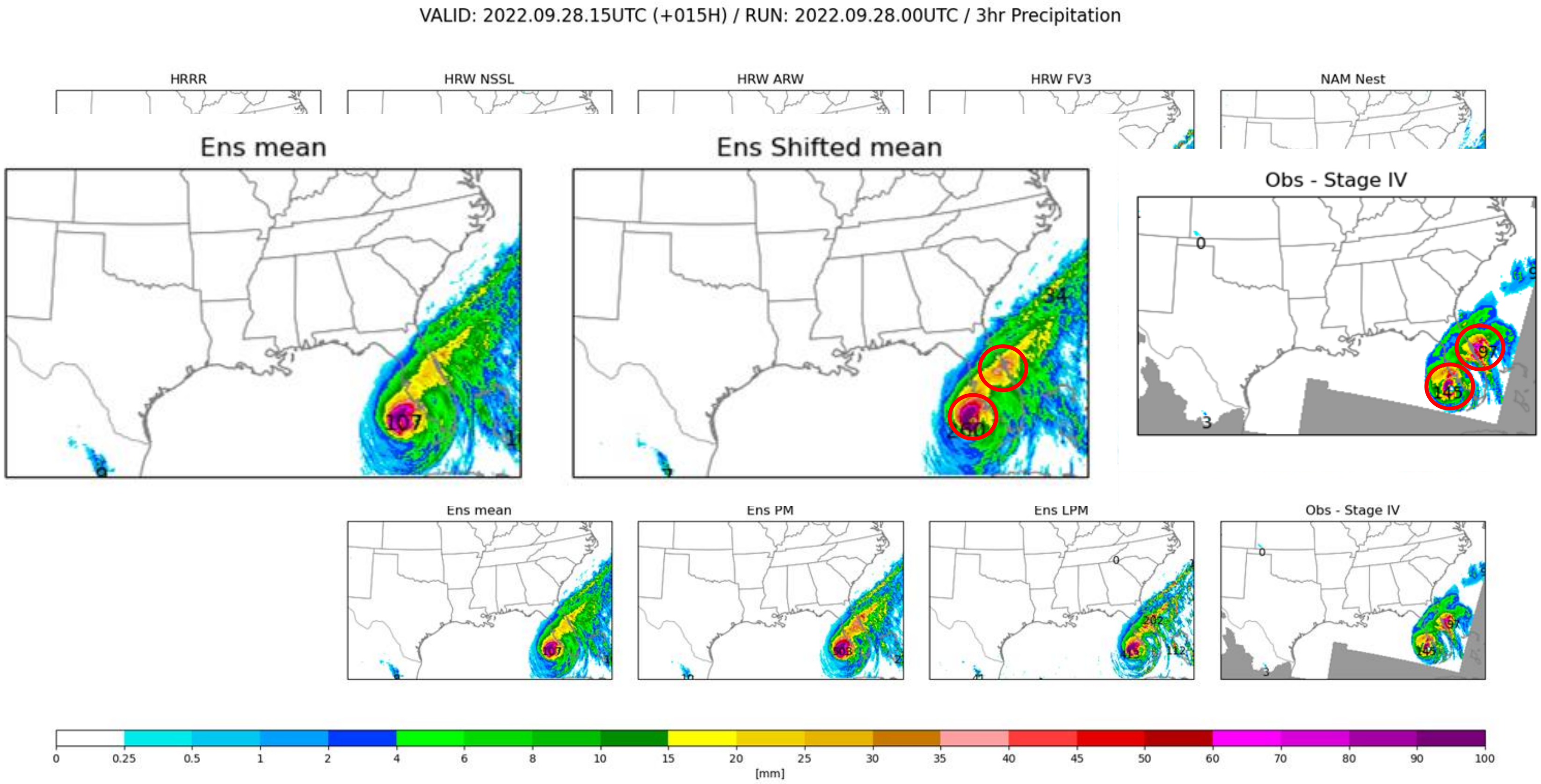
Individual HREF Members



Re-aligned ARW-Lagged Member

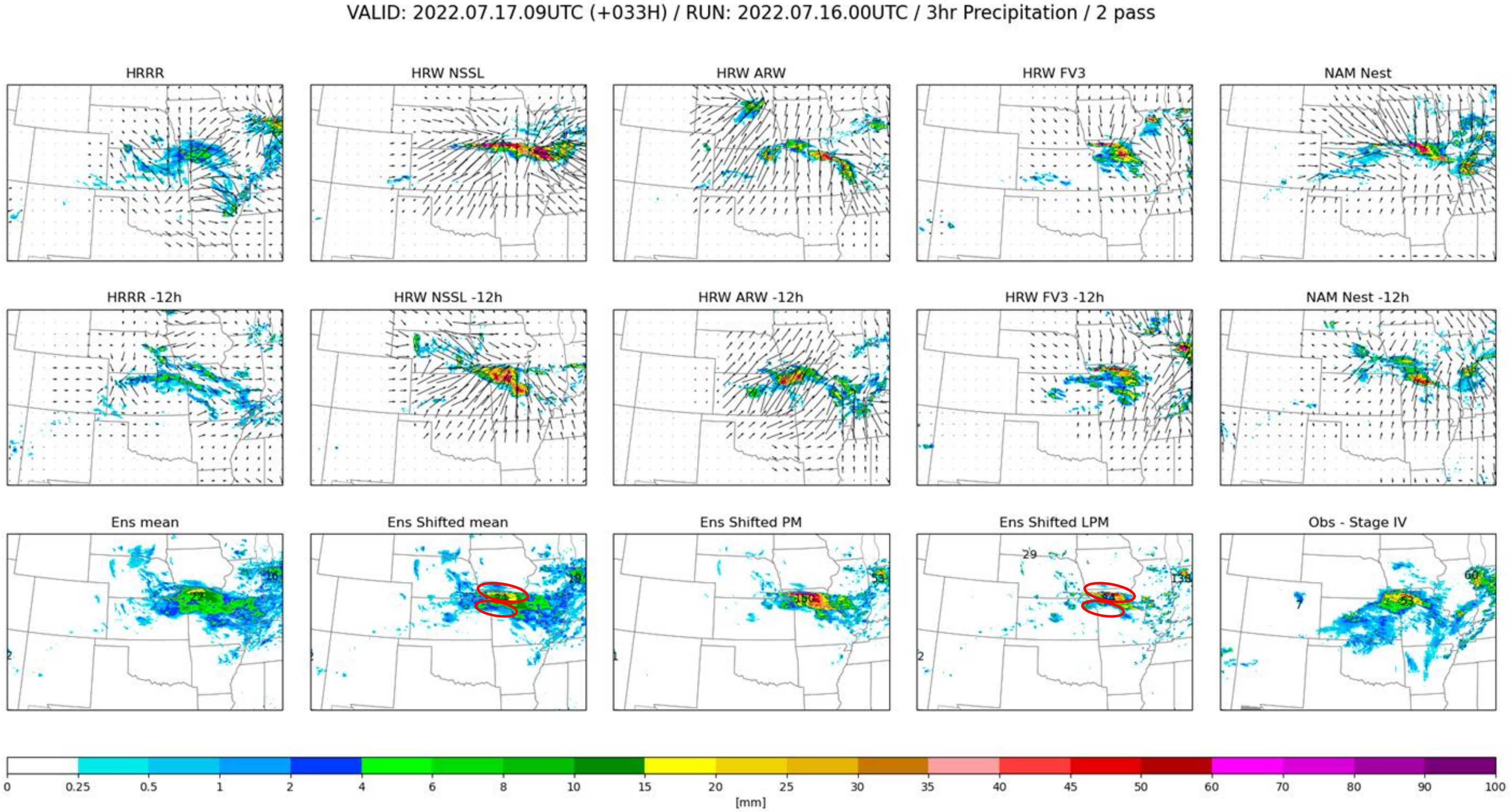
Spatially Aligned Mean – the Algorithm

Example : Hurricane Ian 2022



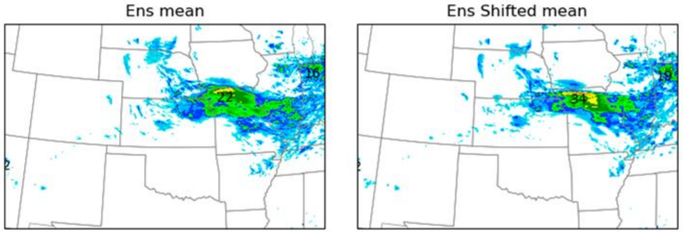
Spatially Aligned Mean Research with HREF in 2022 FFaIR Period

Case: Phase difference in Convective System Propagation

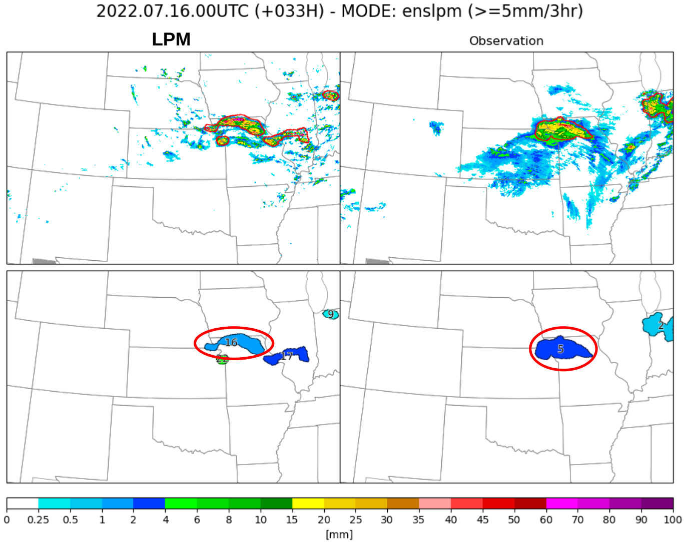


Spatially Aligned Mean Research with HREF in 2022 FFaIR Period

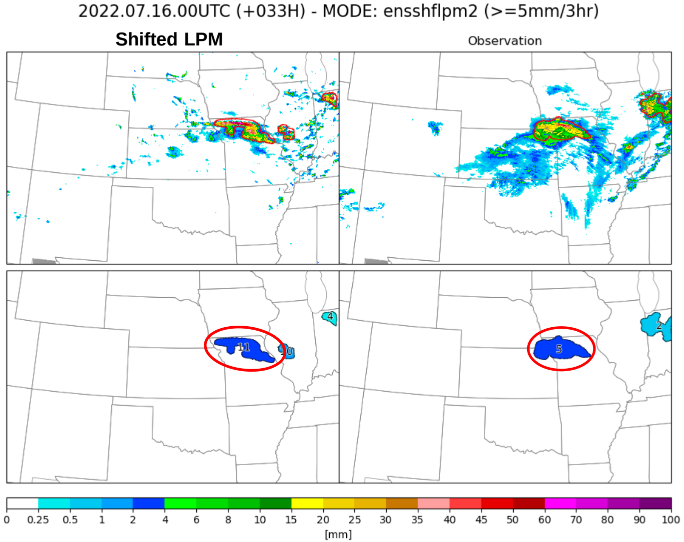
Case: Phase Difference in Convective System Propagation



Apply LPM method to shifted mean
- To preserve the ensemble members’ maxima
- Using same PDF, but with the improved structure



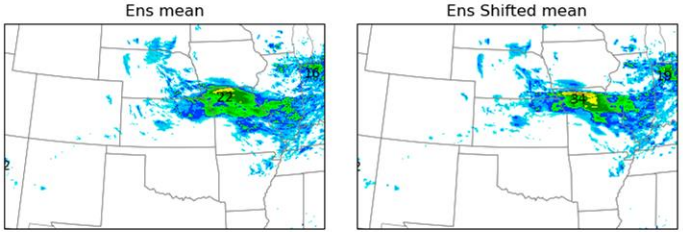
The Interest of MODE : 0.97723
Area ratio : 0.67505
Intersection area(km^2) : 702



The Interest of MODE : 0.98144
Area ratio : 0.80354
Intersection area(km^2) : 877

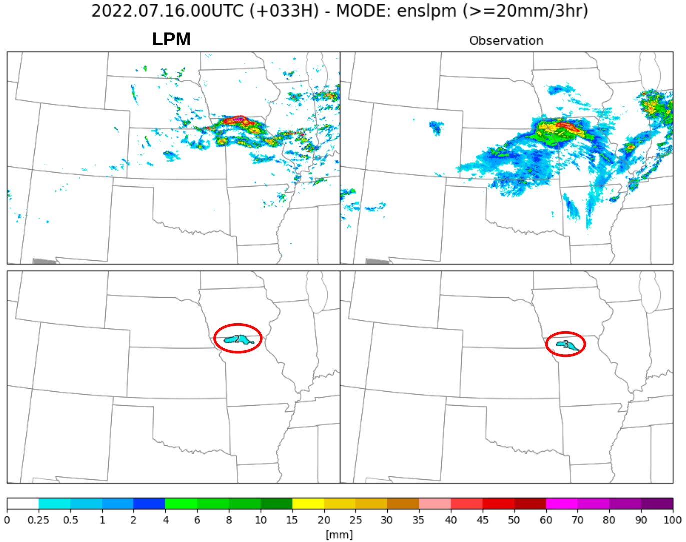
Spatially Aligned Mean Research with HREF in 2022 FFaIR Period

Case: Phase Difference in Convective System Propagation

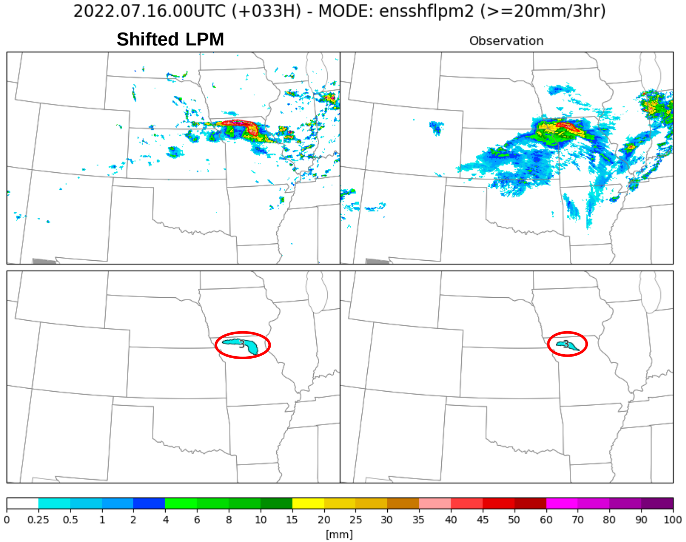


Apply LPM method to shifted mean

- To preserve the ensemble members' maxima
- Using same PDF, but with the improved structure



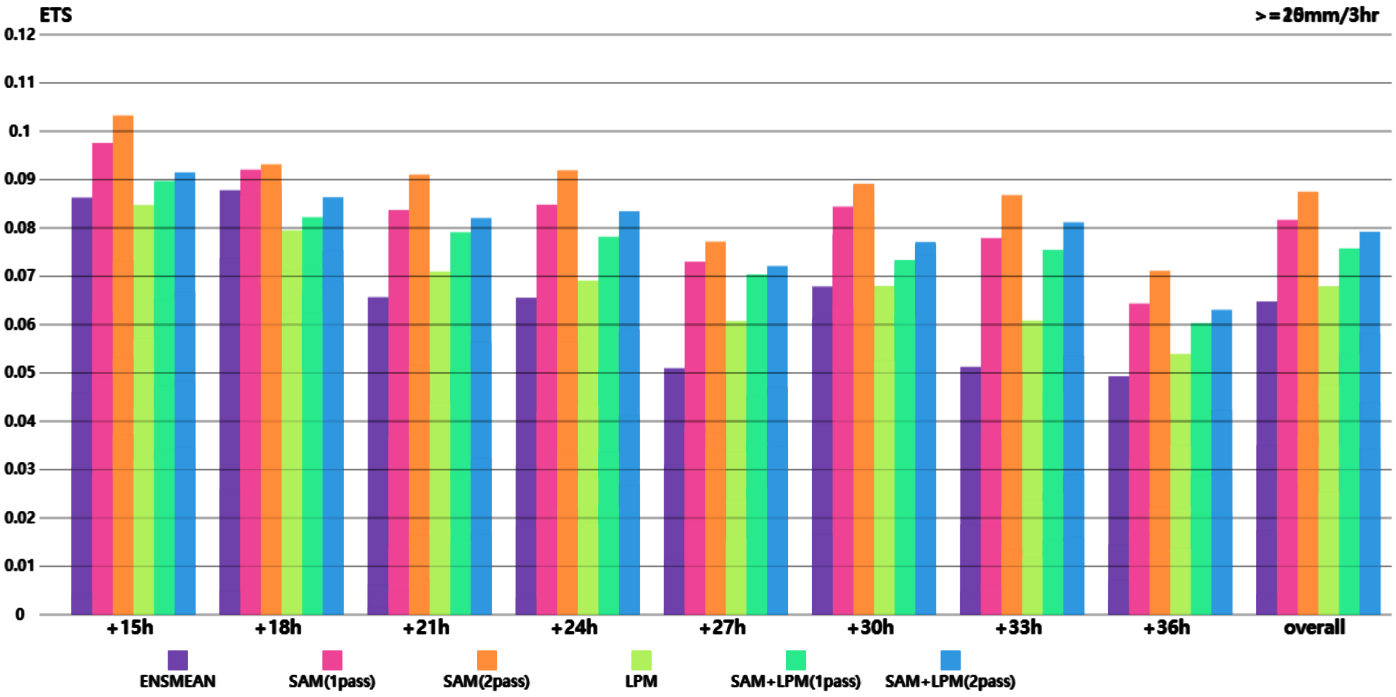
The Interest of MODE : 0.78533
Angle difference(degree) : 13.71772
Intersection area(km^2) : 6



The Interest of MODE : 0.94188
Angle difference(degree) : 0.28638
Intersection area(km^2) : 58

Spatially Aligned Mean Research with HREF in 2022 FFaIR Period

Verification ETS   4 weeks of 2022 FFaIR period

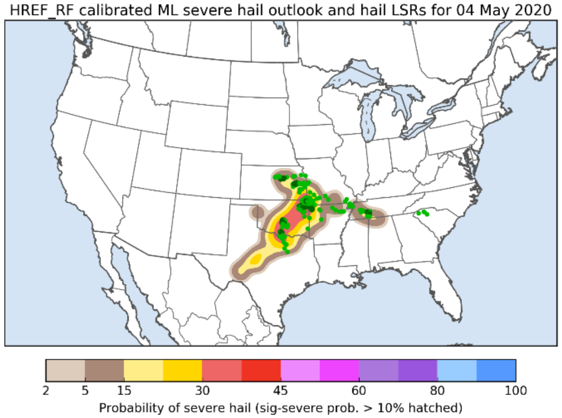


- ETS for shifted mean and shifted LPM increased a lot, compared to regular mean and LPM
- ETS for shifted mean was slightly better than shifted LPM
  (Shifted LPM has a lot better POD, but also has high FAR than shifted mean)

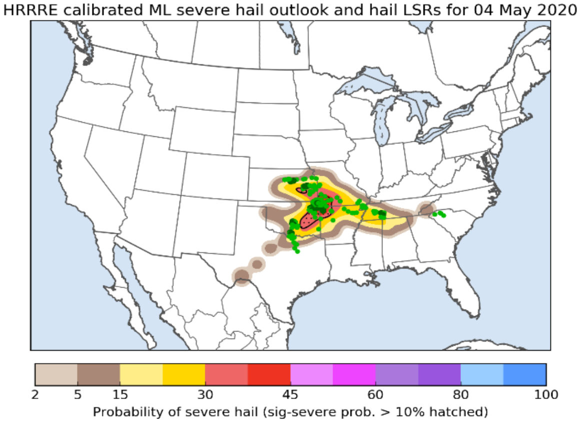
Machine Learning Component

- Collaboration with NSF AI2ES Institute hosted at OU
- U-Net Convolutional Neural Network (Deep Learning)
- Real-time probabilistic rainfall forecasts during 2022 FFaIR
- Builds upon ML hail prediction work in HWT (2017-2021)
- Trained using HREF plus 4 members of 2020-21 CAPS FV3-LAM Ensemble (HREF+)

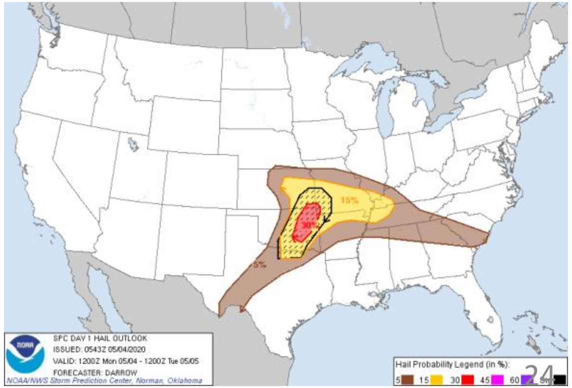
0000 UTC HREFv2



0000 UTC HRRRE

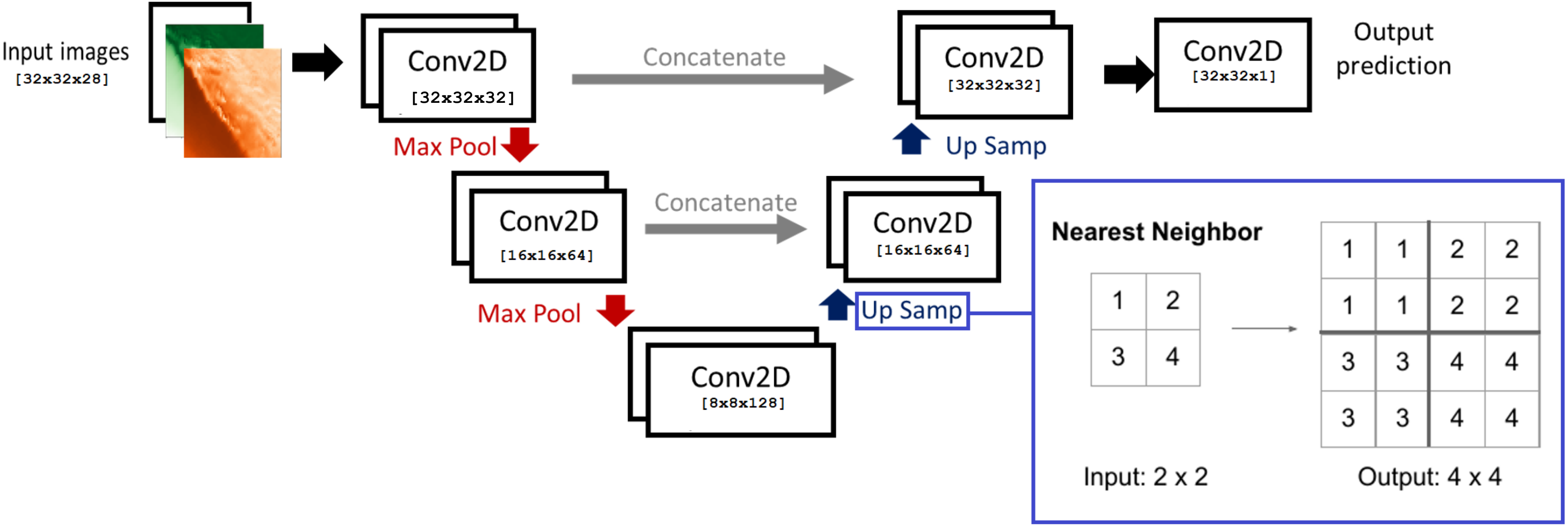


1200 UTC SPC Outlook



U-Net Data/Methods

- Structure for CAPS FV3 Precipitation U-net:
  - Patch size, number of connections, and number of layers are being evaluated as hyper-parameters (architecture shown below may change in later iterations)

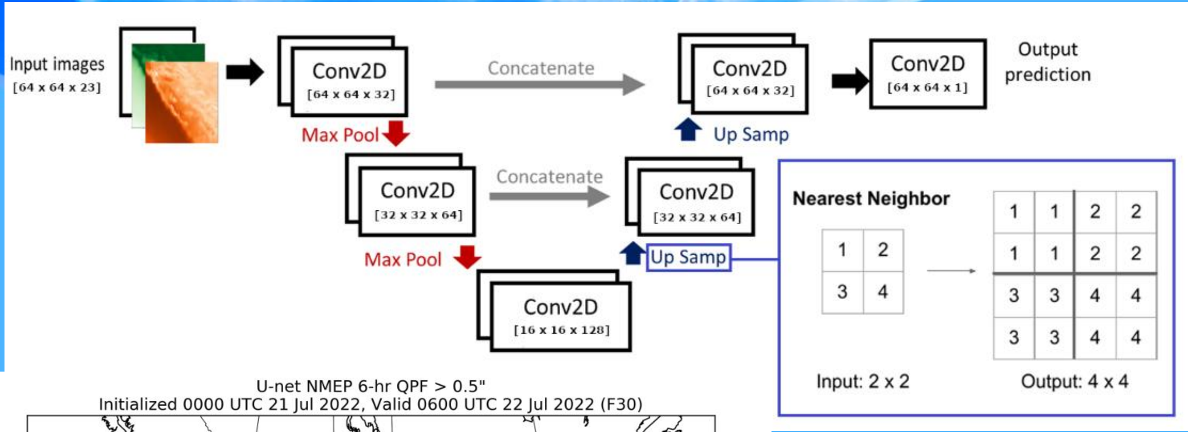


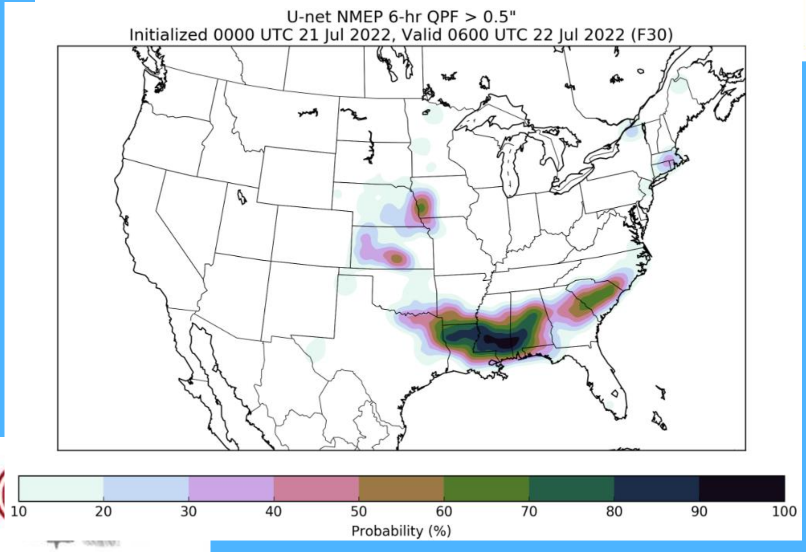
U-Net Data/Methods

CAPS U-Net for Rainfall Prediction uses **23** 2D NWP **forecast variables** relevant to rainfall prediction:

Variable	Level(s) Used
Geopotential height	500 hPa
Temperature	500, 700, 850 hPa; 2 m AGL
Dewpoint	500, 700, 850 hPa; 2 m AGL
u- and v- wind components	500, 850 hPa; 10 m AGL
6-h maximum reflectivity	1 km AGL
Precipitable water	column-integrated
Hourly maximum updraft velocity	column maximum
6-h accumulated precipitation	
Echo-top height	
CAPE	
Mean Sea Level Pressure	
Terrain height	

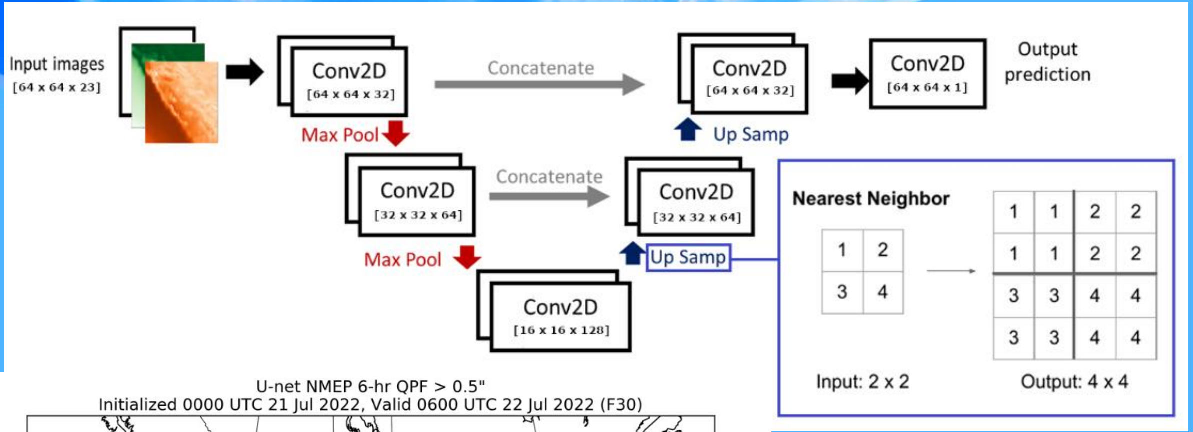
Methods – U-Net for Rainfall Prediction

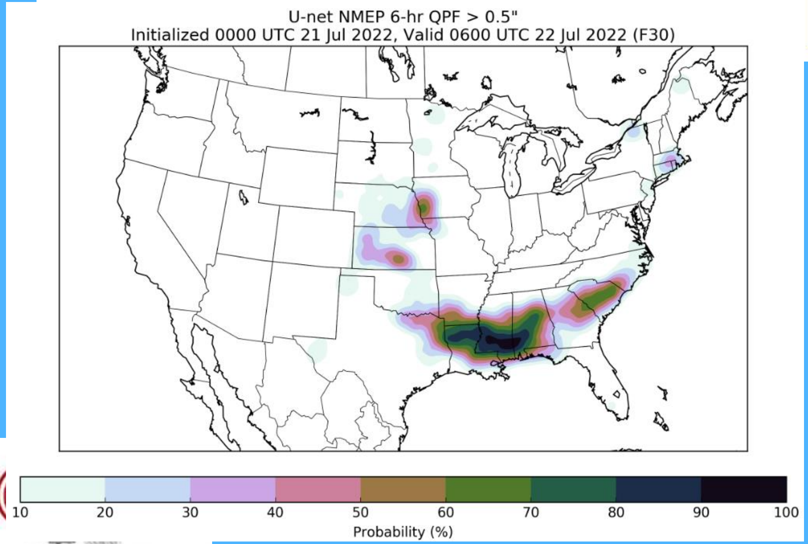




- **2D U-Net** implemented using **Keras**, **Tensorflow**, and the Python "**keras_unet_collection**" library in Python 3.
- The architecture (top left) chosen after preliminary testing with different U-net depths, patch sizes, and training hyperparameters.
- A U-Net using this architecture was trained for each ensemble member, and neighborhood ensemble probability (NEP) and neighborhood maximum ensemble probability (NMEP) were generated from the ensemble of U-Net outputs.
  - Neighborhood radius: 45 km (15 grid points)
  - Gaussian smoother with a standard deviation of 90 km

Methods – U-Net for Rainfall Prediction





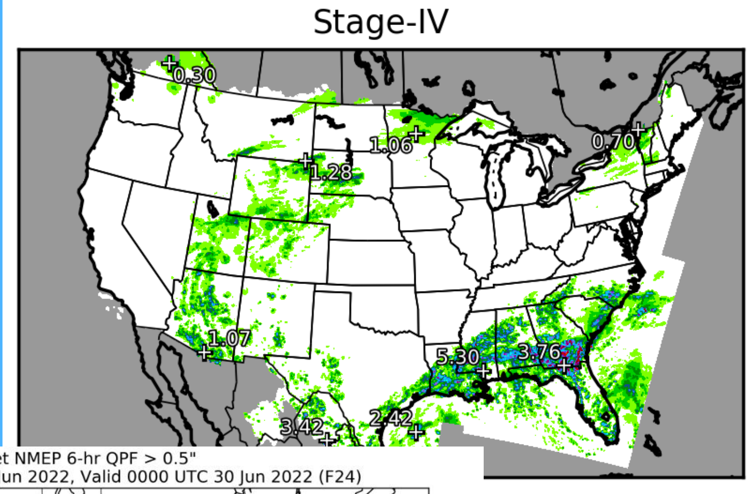
- U-Net output is predicted 6-h accumulated rainfall
  - Predictions are performed on 64 x 64 patches and are stitched together to produce full-conus prediction
  - Patch overlap and light smoothing reduces patch boundary discontinuities.
- Outputs are produced for probability of rainfall/snowfall exceeding given thresholds.
  - 2022: 0.5” in 6 h
  - 2023: 0.5”, 1.0”, and 2.0” in 6 h
- Result: probabilistic forecast product suitable as guidance for areas of moderate- to high-impact rainfall that combines information from the full HREF + 4 CAPS ensemble members (HREF+).

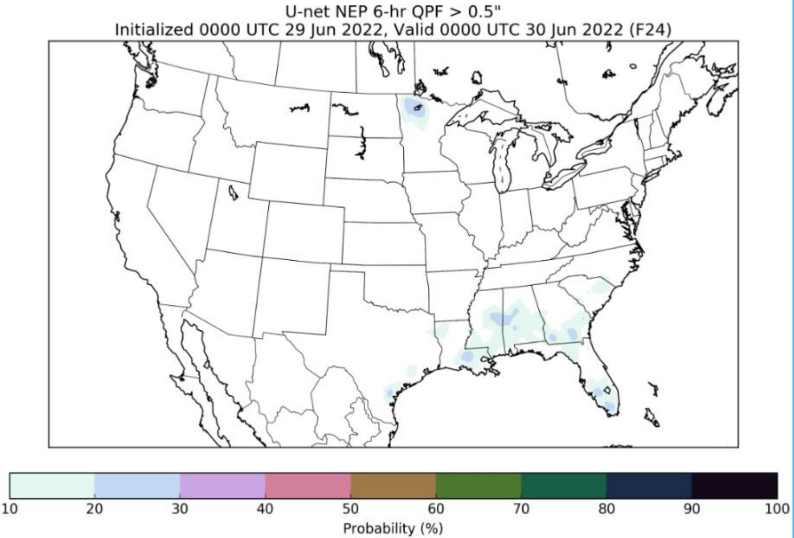
ML Forecasts – 2022 FFaIR

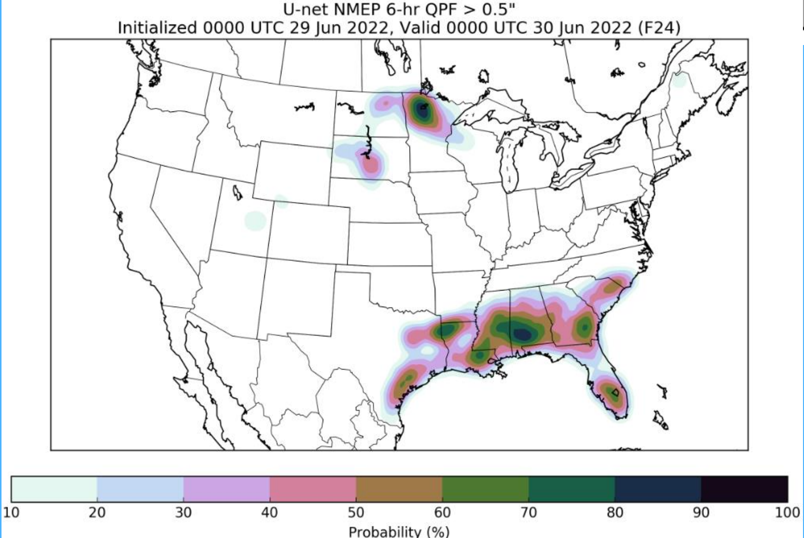
Example: 24 h forecast valid 00 UTC, 30 Jun. 2022

Our initial U-net performs reasonably

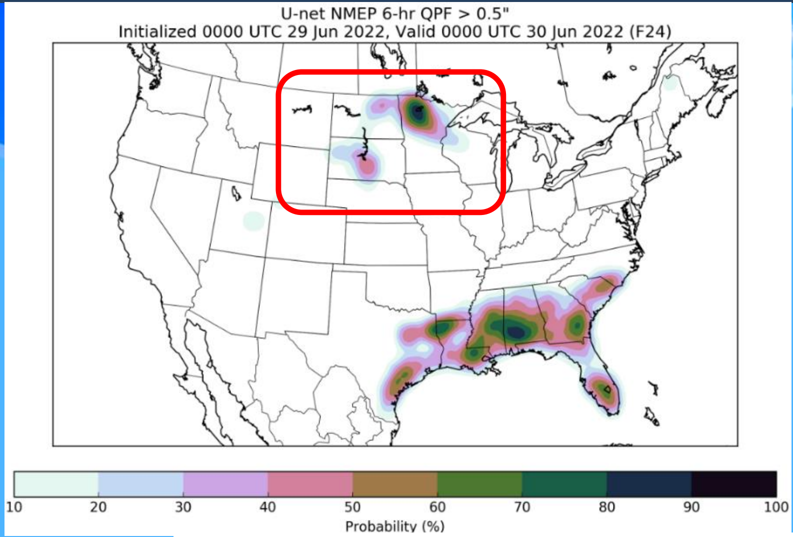
- Successfully identifies heavy rainfall threat over gulf coast states, MN, and SD
- NMEP much better calibrated than NEP
- U-net missed areas of 0.5”+ rainfall over NM – will continue to monitor/investigate regional performance trends.

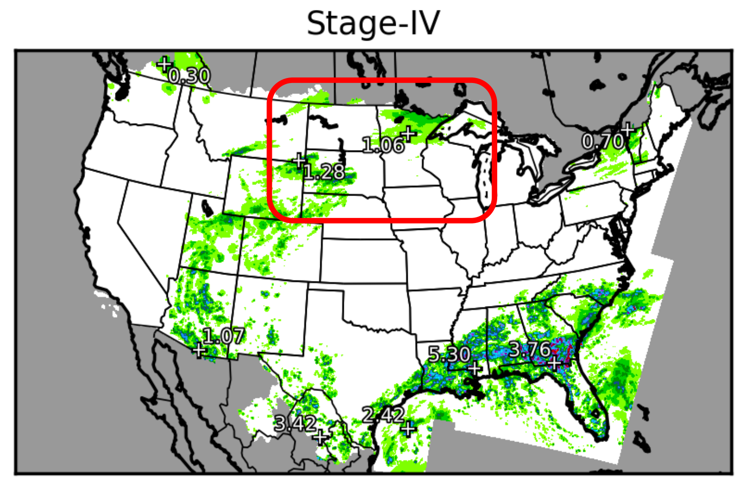


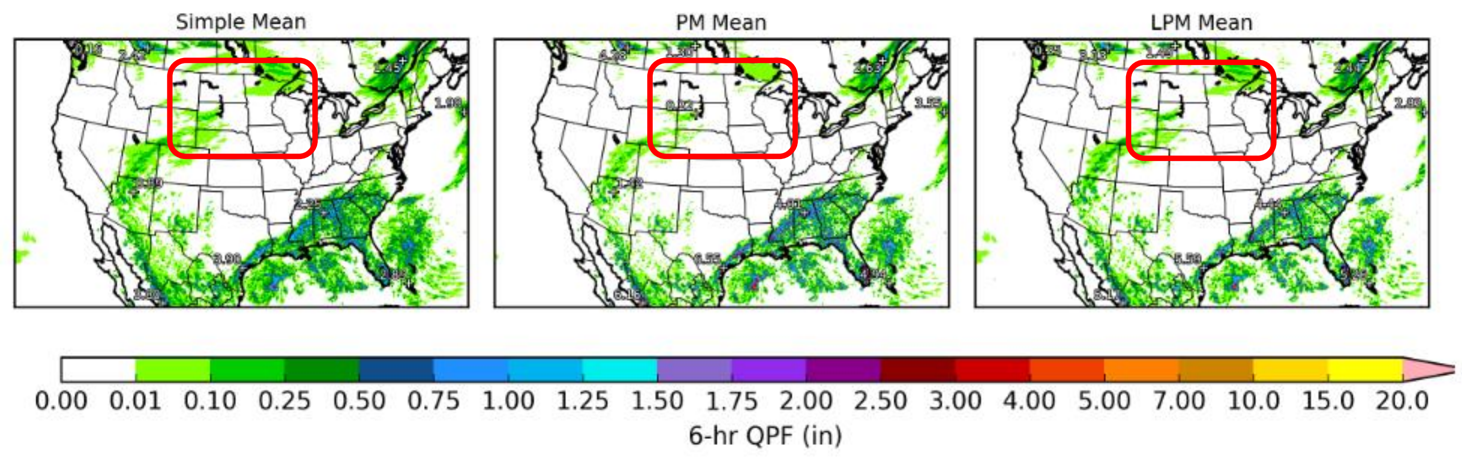




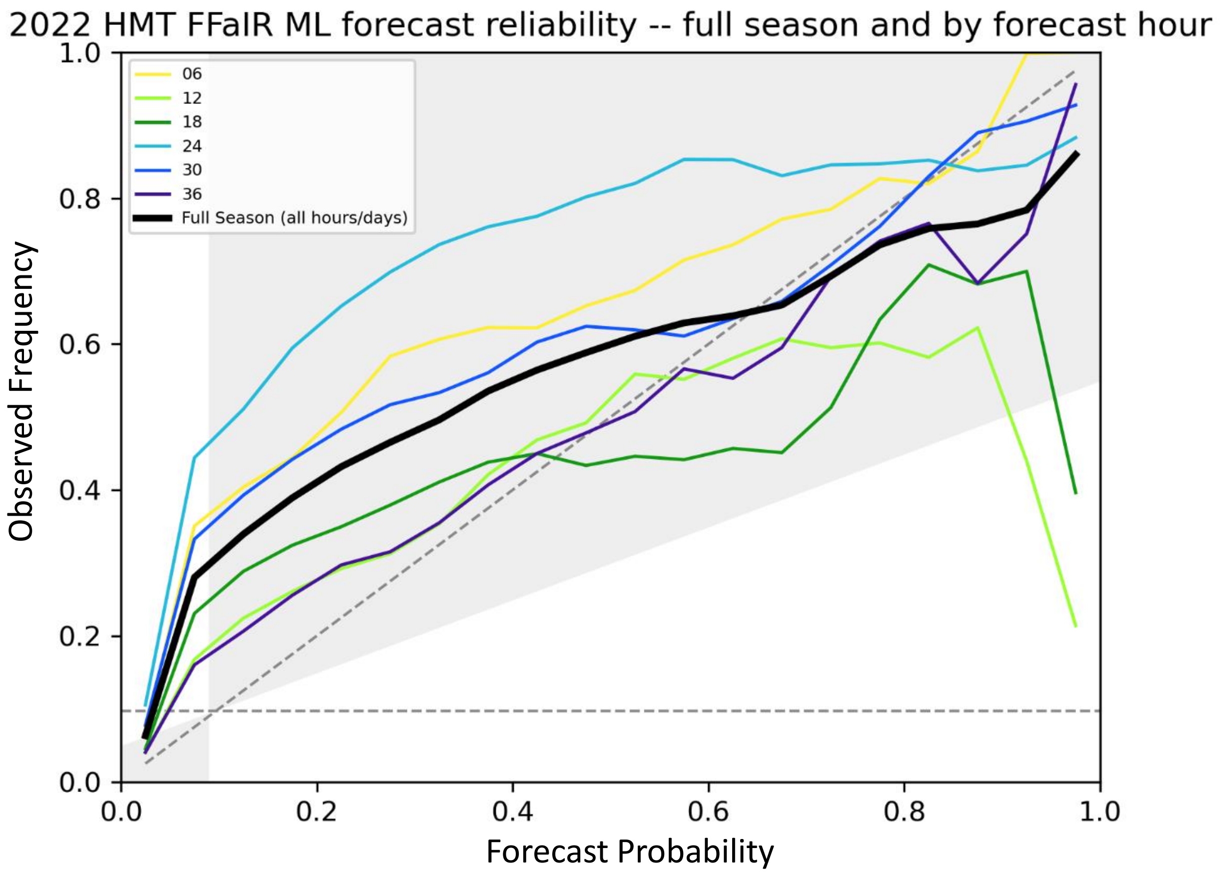
ML Forecasts – 2022 FFaIR







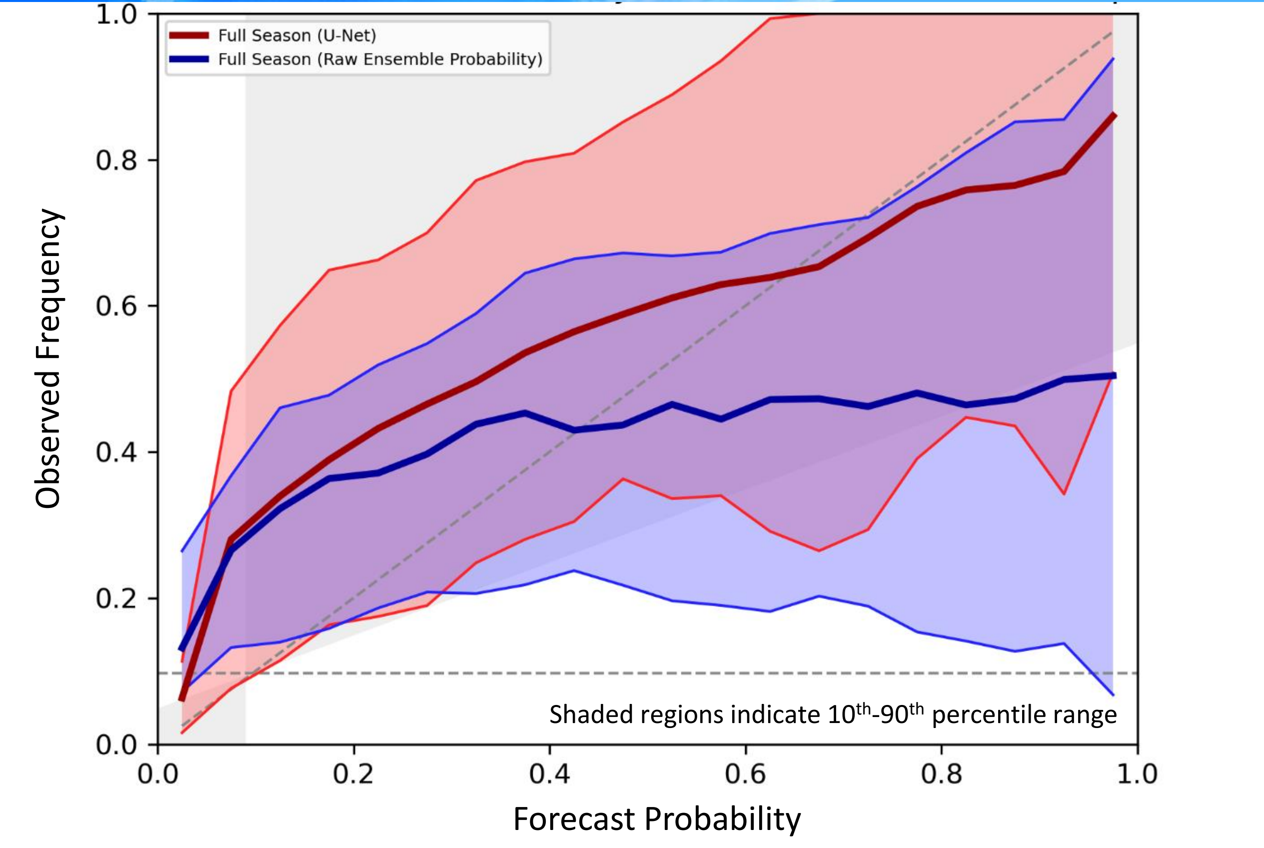
Results – 2022 HMT FFaIR Objective Verification



- U-net ensemble rainfall predictions shows good reliability for NMEP of 6-h accumulated precipitation exceeding 0.5”.
- Substantial diurnal variation
  - Best reliability for 6-, 30-, and 36-hour forecasts (valid at 0600 or 1200 UTC; evening and overnight hours).
  - Worst reliability (substantial under-prediction) for 24-h hour forecasts (valid at 0000 UTC; afternoon hours).
  - General trend toward under-prediction at low probability thresholds.

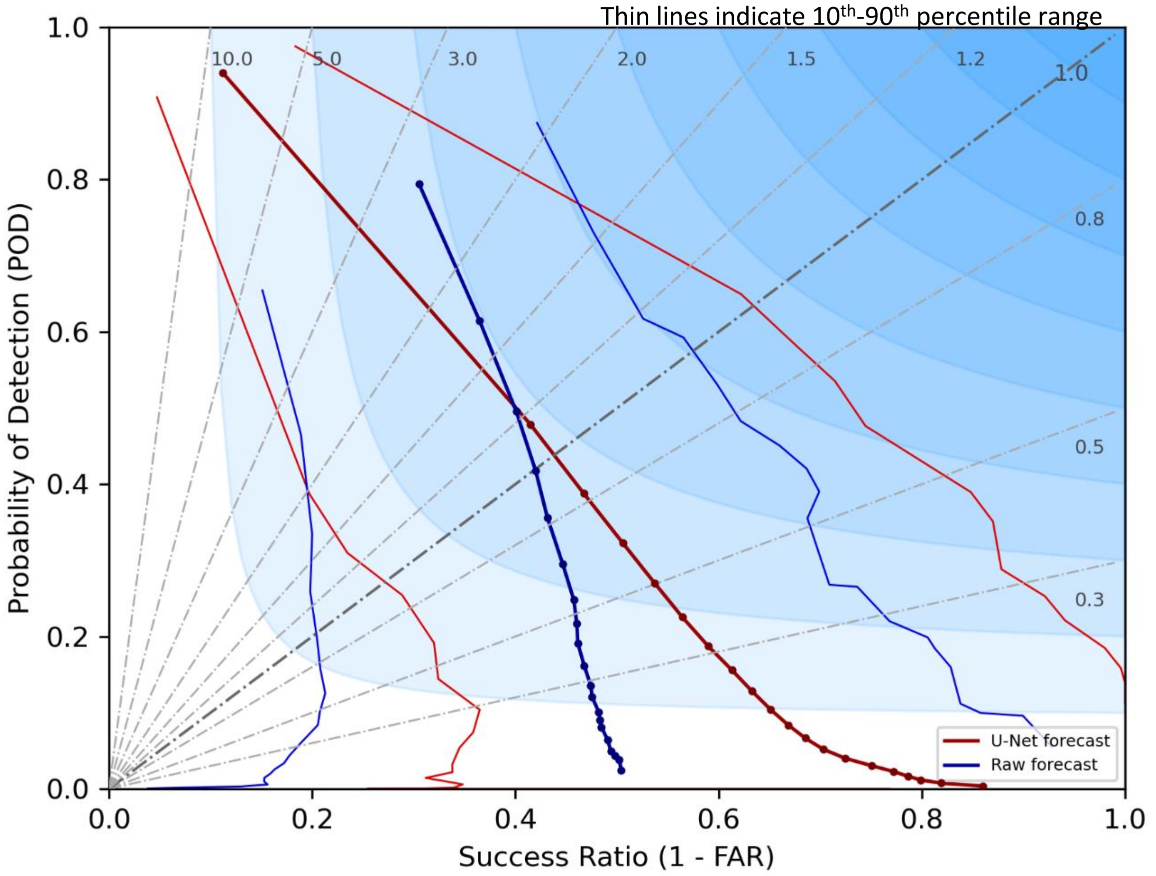
Results – 2022 HMT FFaIR Objective Verification

2022 HMT FFaIR forecast reliability: HREF+ U-Net vs. CAPS FV3 raw output



- During the 2022 FFaIR, the HREF+ U-net ensemble rainfall predictions exhibited desirable reliability compared to raw NWP output from the CAPS FV3 ensemble.

Results – 2022 HMT FFaIR Objective Verification



- U-net ensemble performs comparably to or slightly outperforms raw NWP output from CAPS FV3 ensemble, depending on probability threshold.
- U-net ensemble outperforms CAPS FV3 ensemble at higher probability thresholds (at the expense of greater low bias).

Preliminary ML Conclusions

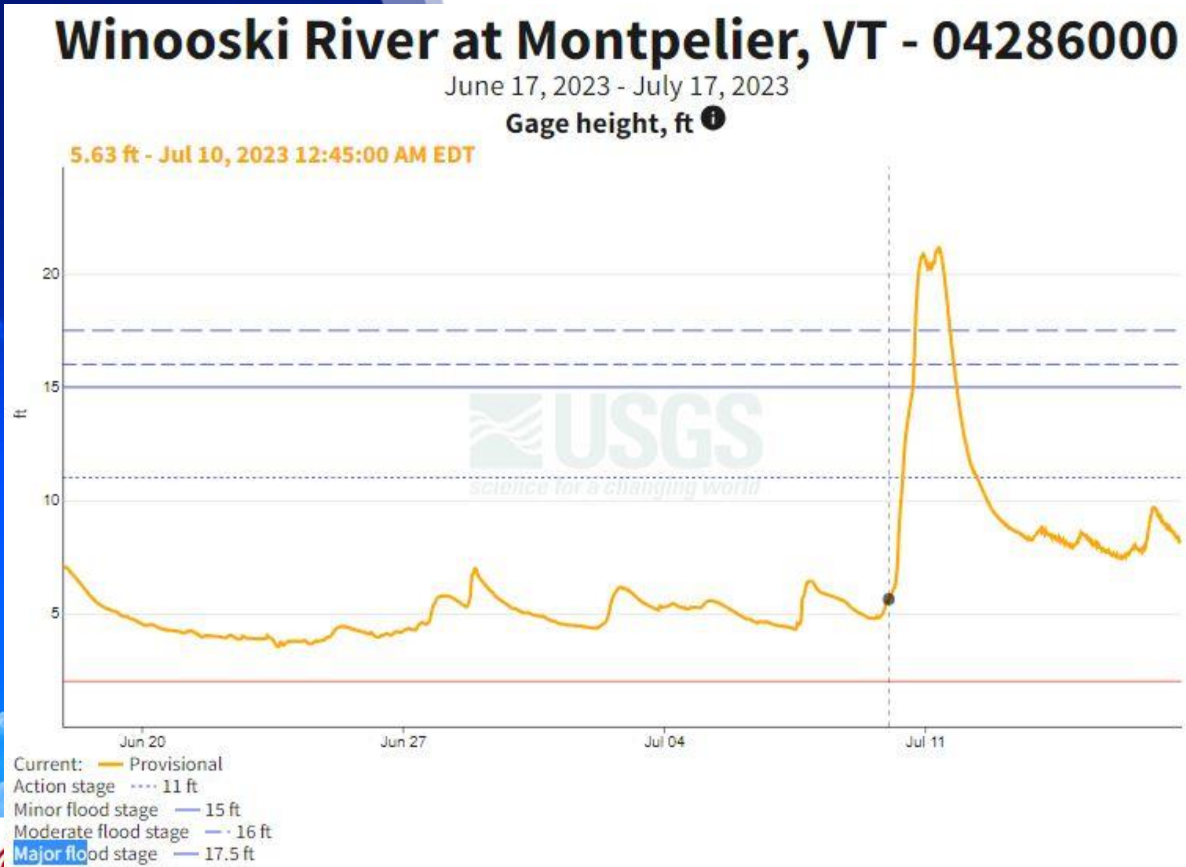
- First iteration CAPS HREF+ U-Net for rainfall prediction performs reasonably, although much room remains for further improvement and refinement.
- The Neighborhood Maximum Ensemble Probability (NMEP) configuration appears to be much better calibrated than the NEP version—NMEP will be used going forward.
- Further improvement and tuning is under way including use of derived fields in addition to the base model output
- Additional rainfall forecast probabilities are planned (e.g., exceedance of return intervals).

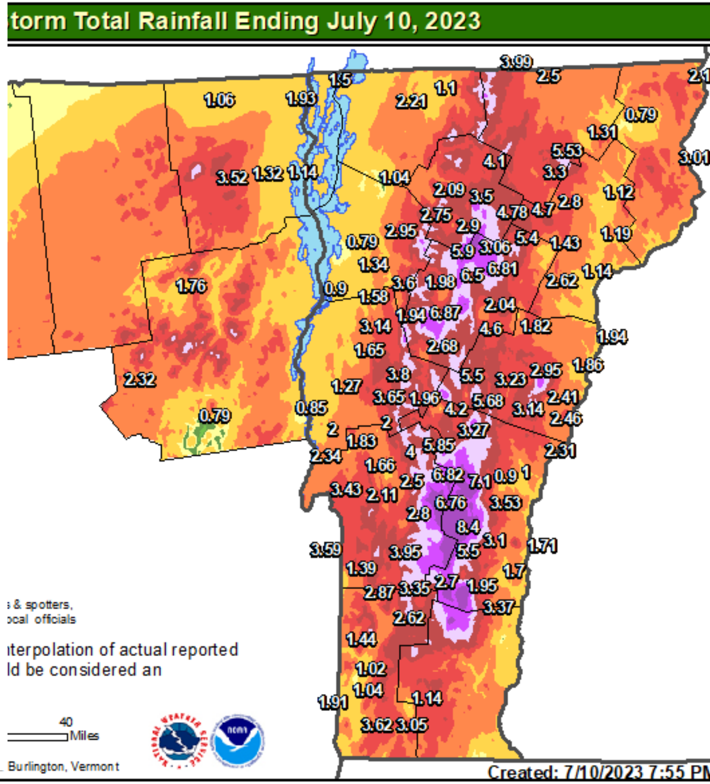
CAPS FV3-LAM 2023 FFaIR 15 Members

Naming
M: Microphysics
B: Boundary Layer
L: Land Sfc Model
P: GFS Initial/Bndy Cdx
PI: Initial perturbations

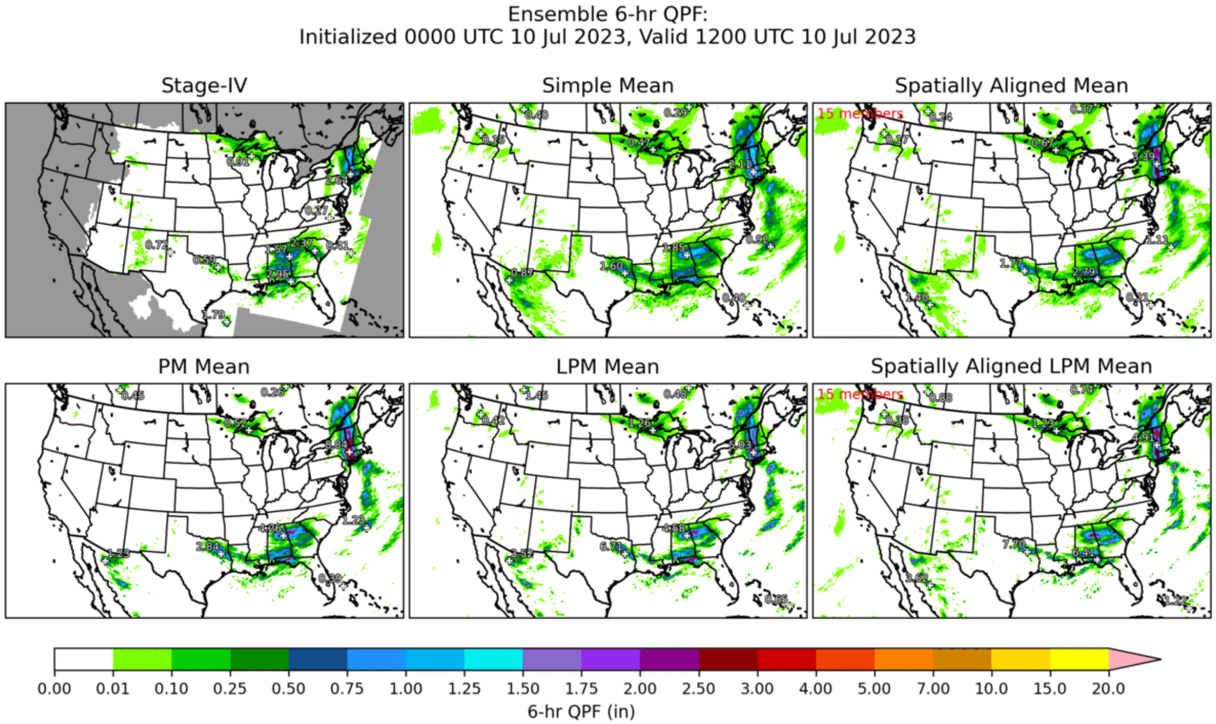
Experiment	Microphysics	PBL	Surface	LSM	IC/LBC (like system)	AI member
GFS IC for Baseline Configuration						
M0B0L0_P	Thompson	MYNN	MYNN	NOAH	GFS /GFS	AI-1
M1B0L0_P	NSSL	MYNN	MYNN	NOAH	GFS/GFS (WoF)	AI-2
M0B0L2_P	Thompson	MYNN	MYNN	RUC	GFS/GFS (RRFSm1)	
M1B2L2_P	NSSL	TKE-EDMF	GFS	RUC	GFS/GFS (RRFSmphys8)	
M0B2L1_P	Thompson	TKE-EDMF	GFS	NOAHMP	GFS/GFS (GFSv16)	AI-3
Physics + IC Perturbation Ensemble						
M0B0L2_PI	Thompson	MYNN	MYNN	RUC	GEFS_m1	
M0B1L0_PI	Thompson	Shin-Hong	GFS	NOAH	GEFS_m2	
M0B2L1_PI	Thompson	TKE-EDMF	GFS	NOAHMP	GEFS_m3	
M0B0L0_PI	Thompson	MYNN	MYNN	NOAH	GEFS_m4	
M0B2L2_PI	Thompson	TKE-EDMF	GFS	RUC	GEFS_m5	AI-4
M1B0L2_PI	NSSL	MYNN	MYNN	RUC	GEFS_m6	
M1B1L0_PI	NSSL	Shin-Hong	GFS	NOAH	GEFS_m7	
M1B2L1_PI	NSSL	TKE-EDMF	GFS	NOAHMP	GEFS_m8	
M1B0L0_PI	NSSL	MYNN	MYNN	NOAH	GEFS_m9	
M1B2L2_PI	NSSL	TKE-EDMF	GFS	RUC	GEFS_m10	

July 10-11, 2023 Vermont Flash Flooding

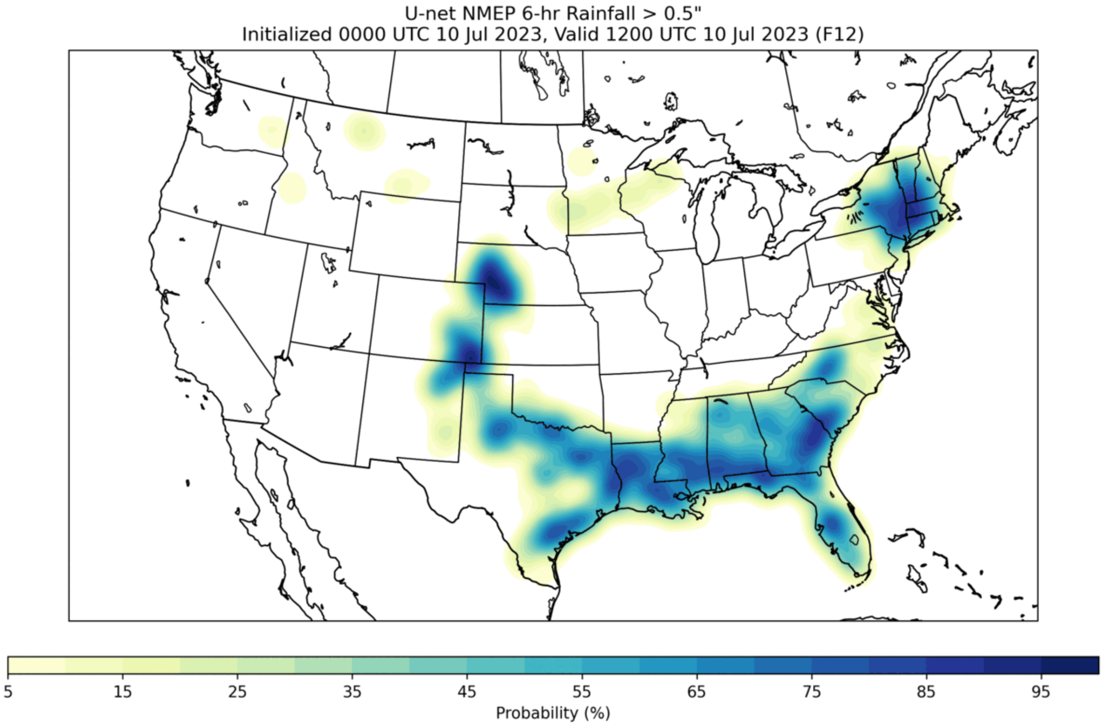




July 10 Forecast
Ensemble Consensus Products



ML NMEP Probabilities



Acknowledgments/Contact

- Computing:
  - NSF ACCESS
    Texas Advanced Supercomputing Center (TACC) Frontera

- Funding: NOAA/OAR/OWAC Testbed Grants:
  NA19OAR4590141 & NA22OAR4590522
  UFS R2O Grant NA16OAR4320115

Realtime FFaIR Forecasts Online

Contact Info:

Keith Brewster: kbrewster@ou.edu

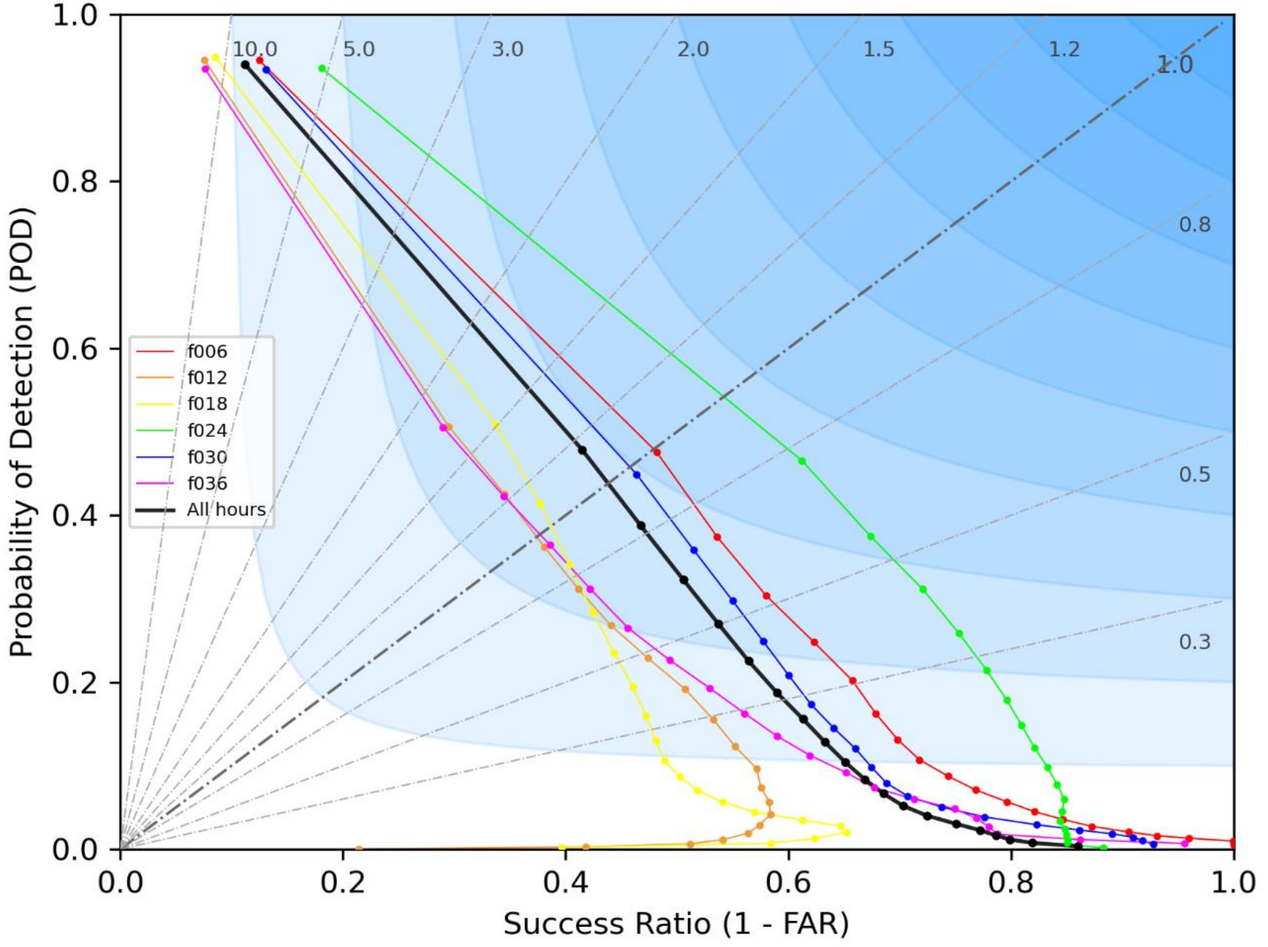
Nate Snook: nsnook@ou.edu

Chang Jae Lee: changjae.lee.3789@gmail.com

https://www.caps.ou.edu/forecast/realtime

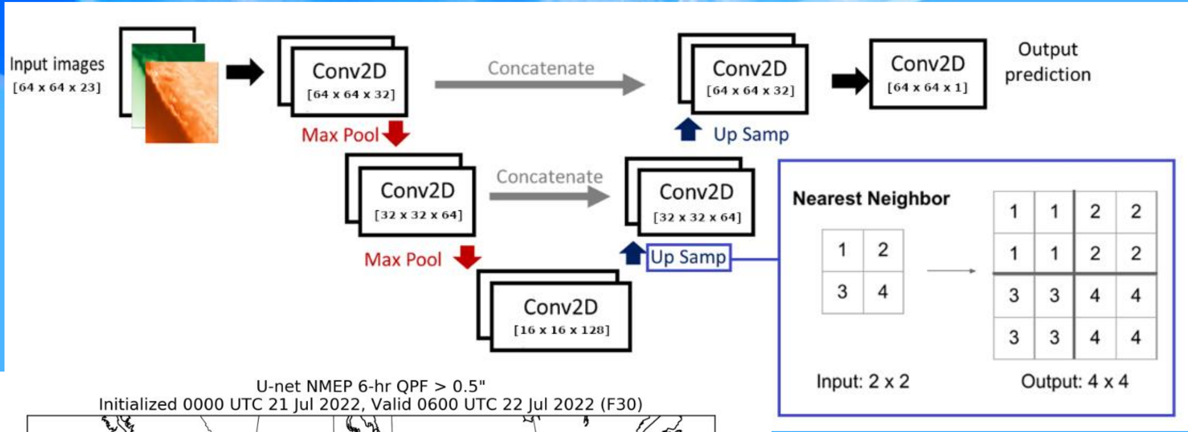
Bonus Slides

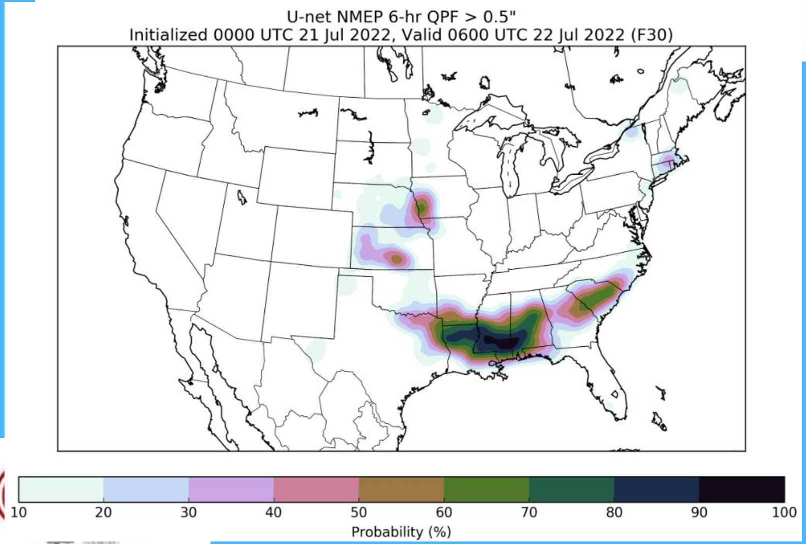
Results – 2022 HMT FFaIR Objective Verification



- U-net ensemble rainfall predictions exhibit best performance for 6-, 24-, and 30-h forecasts (valid at 0000 or 0600 UTC; afternoon and evening hours).
  - 24-h forecasts exhibit high CSI (best at low probability thresholds of ~10%) despite substantial under-prediction.
- Performance is worst for 12- and 36-h forecasts (valid at 1200 UTC; overnight hours).
- Most desirable bias and maximum CSI are generally obtained at low NMEP probability thresholds (10-20%).

Methods – U-Nets for Rainfall/Snowfall Prediction





- During training, for each member, 13,000 patches are randomly selected from the available data meeting criteria for raw NWP 6-h accumulated rainfall at the central pixel of the patch:
  - 9000 “rain” patches with non-zero 6-h accumulated rainfall/snowfall
  - 3000 “heavy rain” patches with 6-h accumulated rainfall of at least 20 mm
  - 1000 “null” patches with 6-h accumulated rainfall of no more than 0 mm
- Validation was performed using 6500 patches (4500 “rain/snow”, 1500 “heavy rain/snow”, and 500 “null”).
- Labels were generated from Stage IV 6-h accumulated rainfall observations.
- Additional U-net configuration details:
  - Loss function: mean squared error
  - Activation: ReLU
  - Batch size of 128 with 20 training epochs