Medium-range Forecasts of Excessive Rainfall with the CSU-MLP

Dr. Aaron J. Hill[1,2] and Dr. Russ Schumacher[3]

[1]*School of Meteorology, University of Oklahoma*

[2]*NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES)*

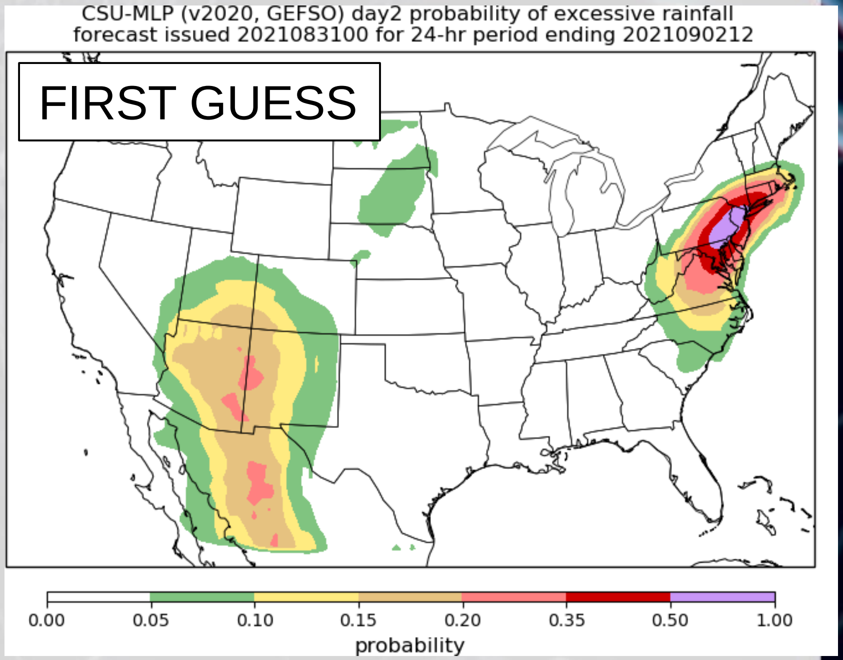
[3]*Department of Atmospheric Science, Colorado State University*

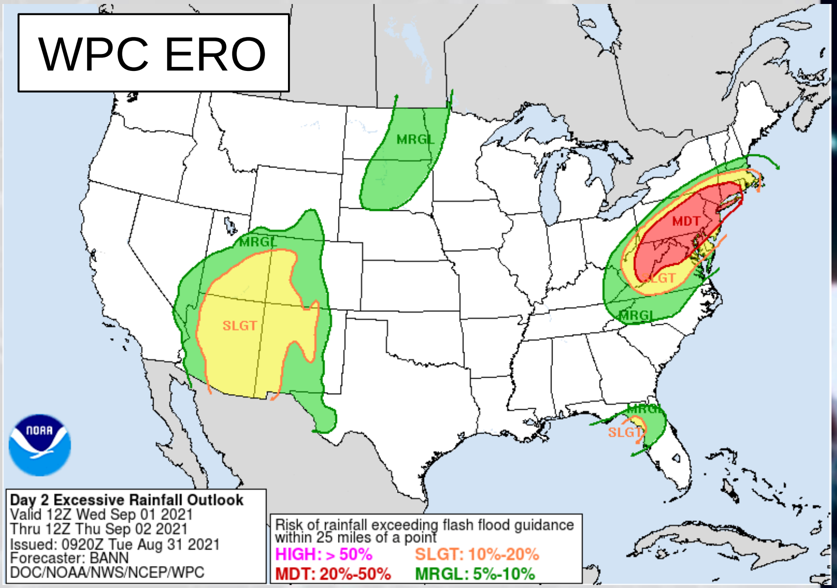
Acknowledgements: Mitchell Green (UNL)
NOAA JTTI Program

HMT Flash Flood and Intense Rainfall Experiment
13 June 2024

Background

- NOAA Weather Prediction Center forecasters routinely issue Excessive Rainfall Outlooks (EROs), indicating regions with the potential for flooding rains across the continental US on days 1-3
- Since 2017, we have developed and tested probabilistic forecasts that apply machine-learning techniques to a reforecast ensemble to help give guidance to WPC forecasters -- a "first guess" when producing these outlooks
- Several versions of the forecast system based on the GEFS are now running operationally at WPC



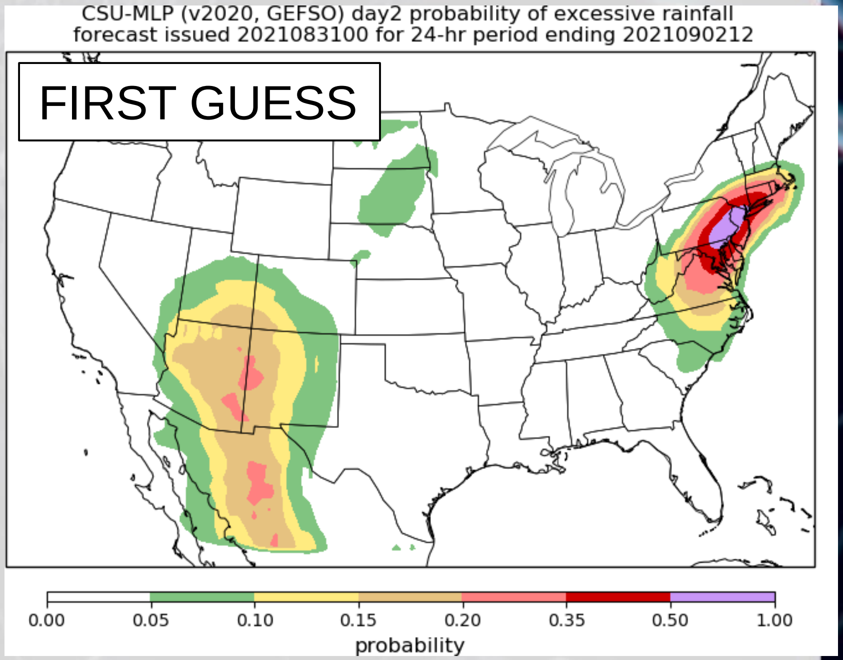


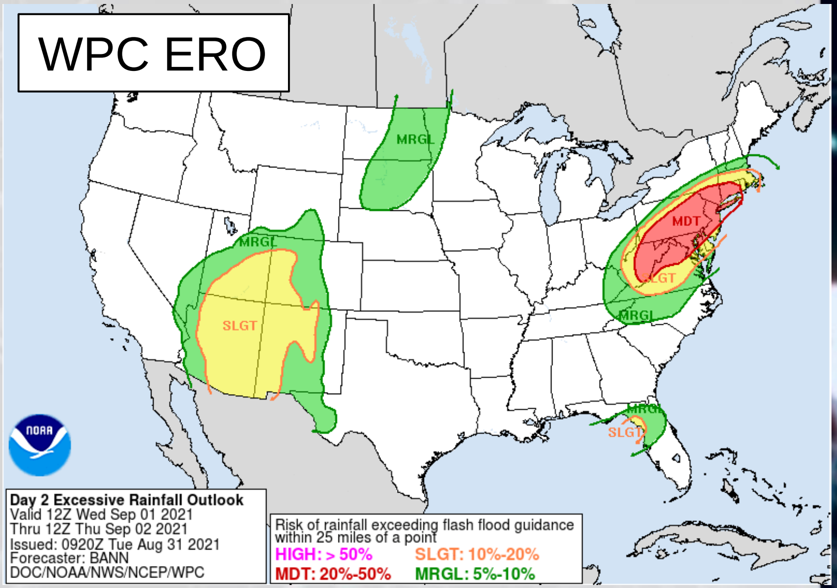
Schumacher et al. (2021, *BAMS*)

Real-time forecast graphics:

Background

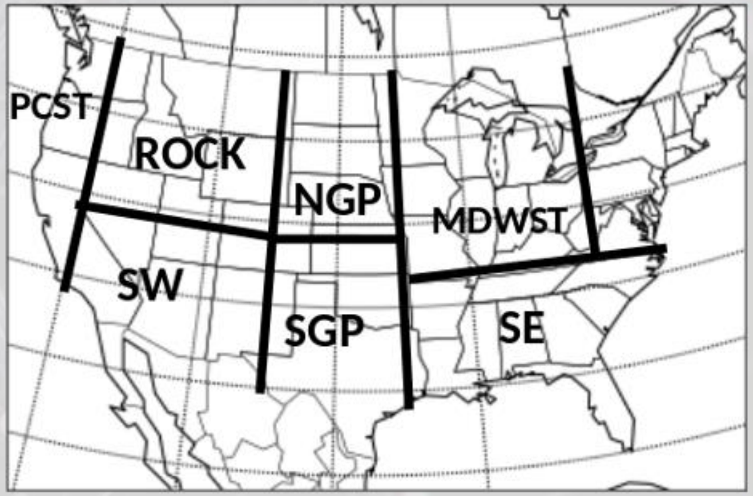
- In 2022, WPC began issuing experimental day 4-5 EROs – these are now operational
- **To support this effort, and to see whether even longer lead times are possible, the CSU-MLP precipitation forecasts have been extended to 8 days, similar to severe weather guidance products (Hill et al. 2023, *WAF*)**
- **Q: But is there actual forecast skill at these lead times?**
- **Q: When and where to forecasts derive skill?**





The Approach

- Data: NOAA’s FV3-GEFS Reforecast Dataset (Hamill et al. 2022): 5 members, matches current GEFSv12
- Use many atmospheric fields as predictors, train random forest models over 8 regions
- We use Jan 2003 – August 2013 as the training period (~10 yrs)
- Probabilistic forecasts mimic the ERO categories/definitions
- Observations to define excessive rainfall...



Symbol	Description
APCP	Precipitation accumulation in past (3) 6 h
CAPE	Surface-based convective available potential energy
CIN	Suface-based convective inhibition
MSLP	Mean sea level pressure
PWAT	Total precipitable water
Q2M	Specific humidity two meters above ground
SHR500	Bulk wind difference magnitude between 10 m and 500 hPa
SHR850	Bulk wind difference magnitude between 10 m and 850 hPa
T2M	Air temperature two meters above ground
U10	Zonal component of 10-m wind
UV10	10-m wind speed
V10	Meridional component of 10-m wind

See Schumacher et al. (2021); also Herman and Schumacher 2018a,b) for more details

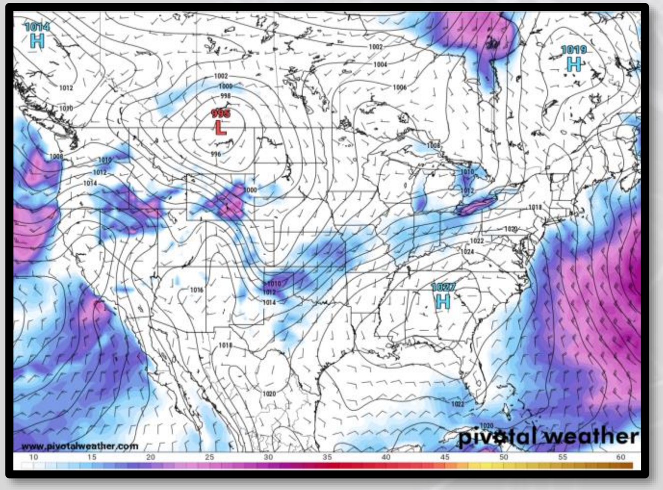
We want to predict excessive rainfall…but what is excessive rainfall?

- A primary motivation for this approach is that forecasters need probabilistic information about the rarity of upcoming rainfall. But...
- We have accepted (if flawed) definitions of tornado, severe hail, severe winds – but nothing analogous for excessive rainfall
- Exceeding flash flood guidance (FFG)?
- Produces a flash flood report?
- More than a certain threshold? (and if so, which one(s)?)
- What quantitative precipitation estimate to use?

Two precipitation/impact datasets for training

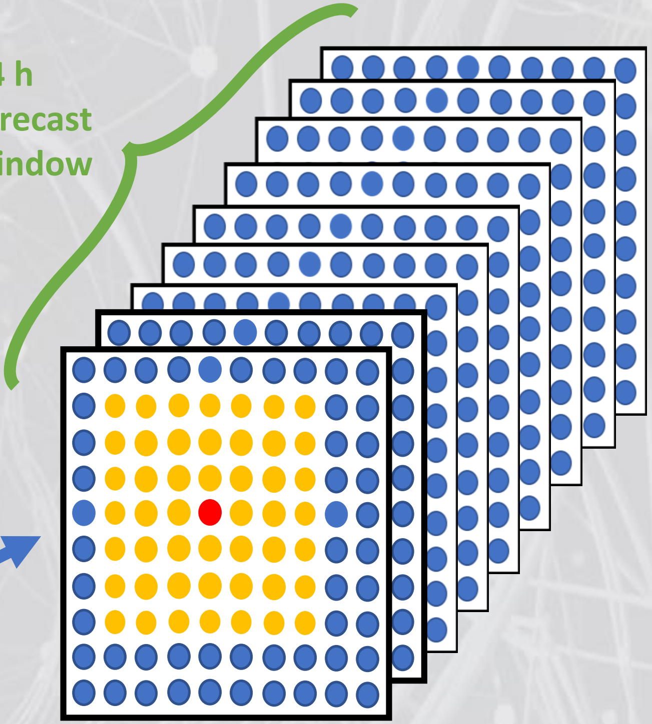
- "Fixed Frequency" – or in other words, we use climatological average recurrence intervals (ARIs) to define a heavy or extreme rain event
    - Better corresponds to actual impacts in a given region than a fixed threshold
    - Doesn't bias the verification statistics toward climatologically wet regions
    - We use the NCEP Climatology-Calibrated Precipitation Analysis (CCPA) to identify historical exceedances of the various average recurrence intervals (1 and 2 yr) for 24-hour rainfall accumulation
- Unified Flood Verification System (Erickson et al. 2019,2021)
    - Local storm reports, exceedances of FFG or the 5-yr ARI for various temporal periods (1-, 3-, 6-, 24-h), and reports of flooding from USGS stream gauges

MLP Prediction System



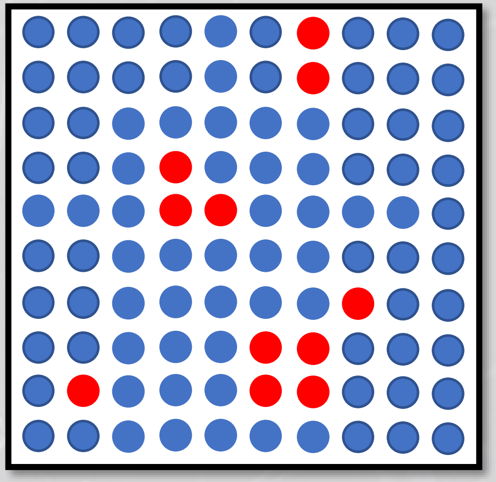
GEFS ensemble median of environmental parameters

24 h forecast window



QPF
CAPE
CIN
MSLP
Shear
etc.

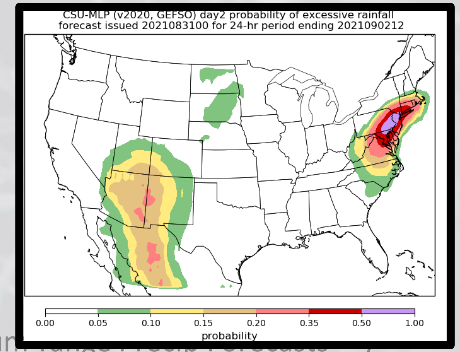
Historical Daily Events Defining Training Sample (10 years)



Real-time GEFS predictors

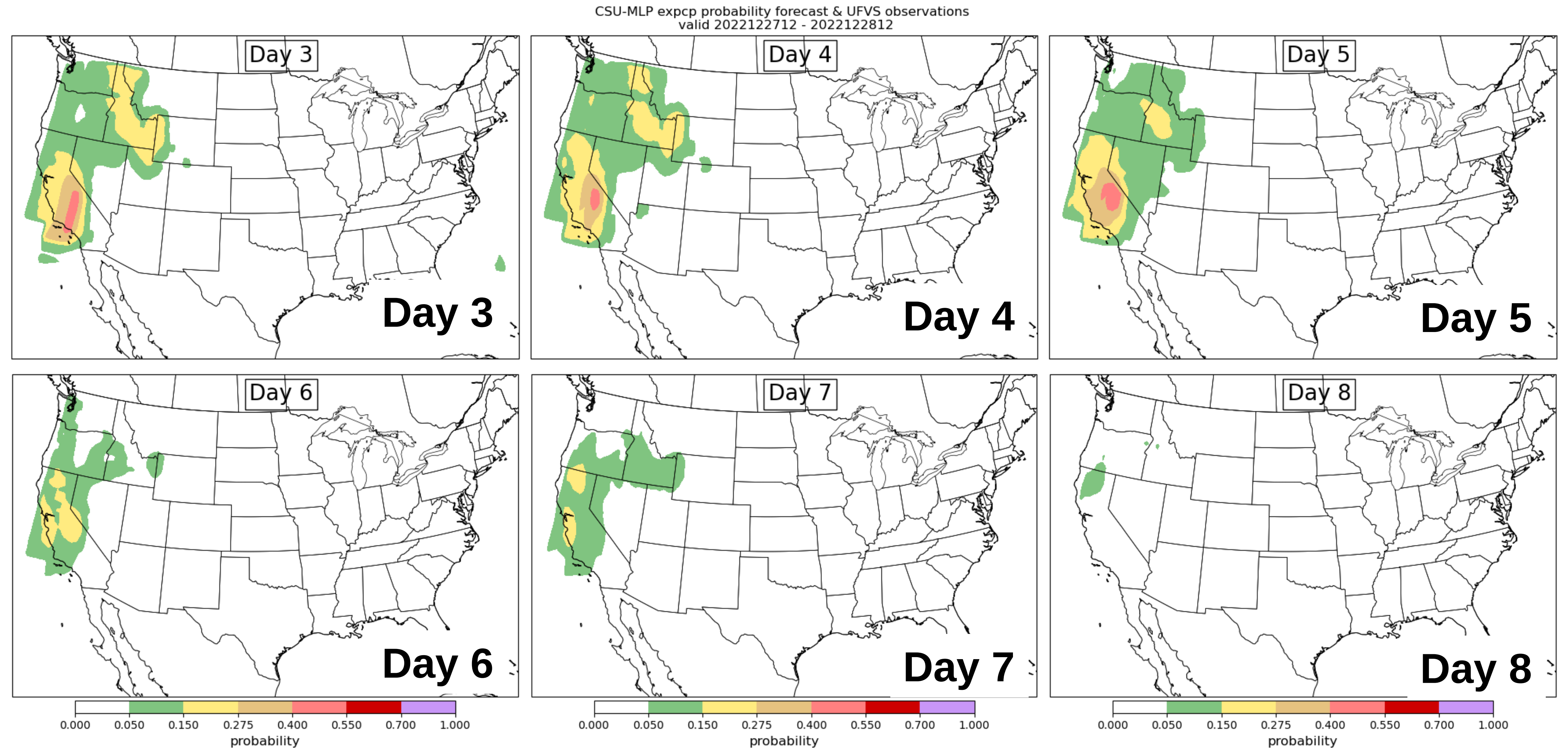
MODEL

Real-time Predictions



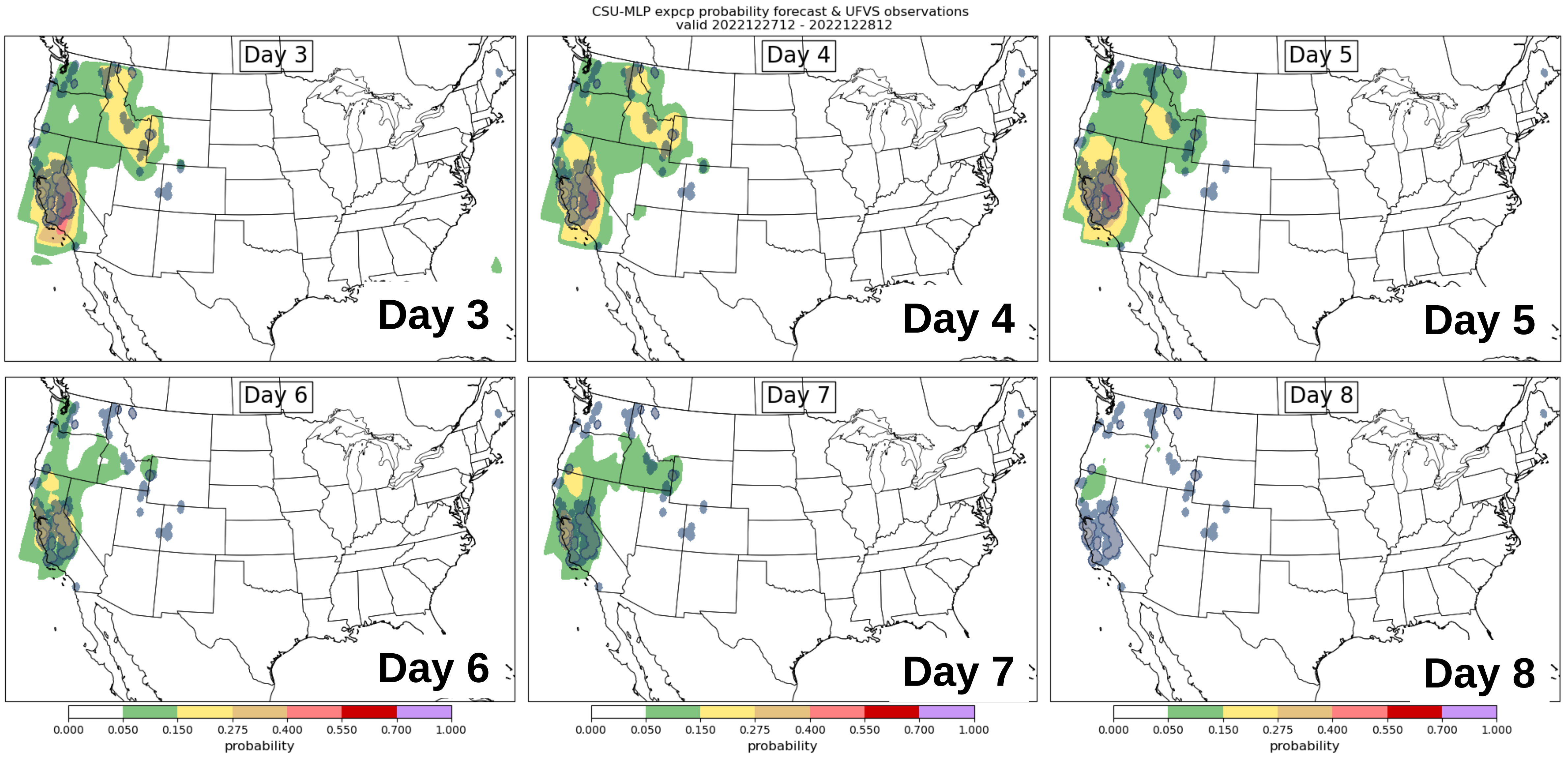
Good forecast example: 27-28 December 2022 (California flooding)

Fixed frequency model



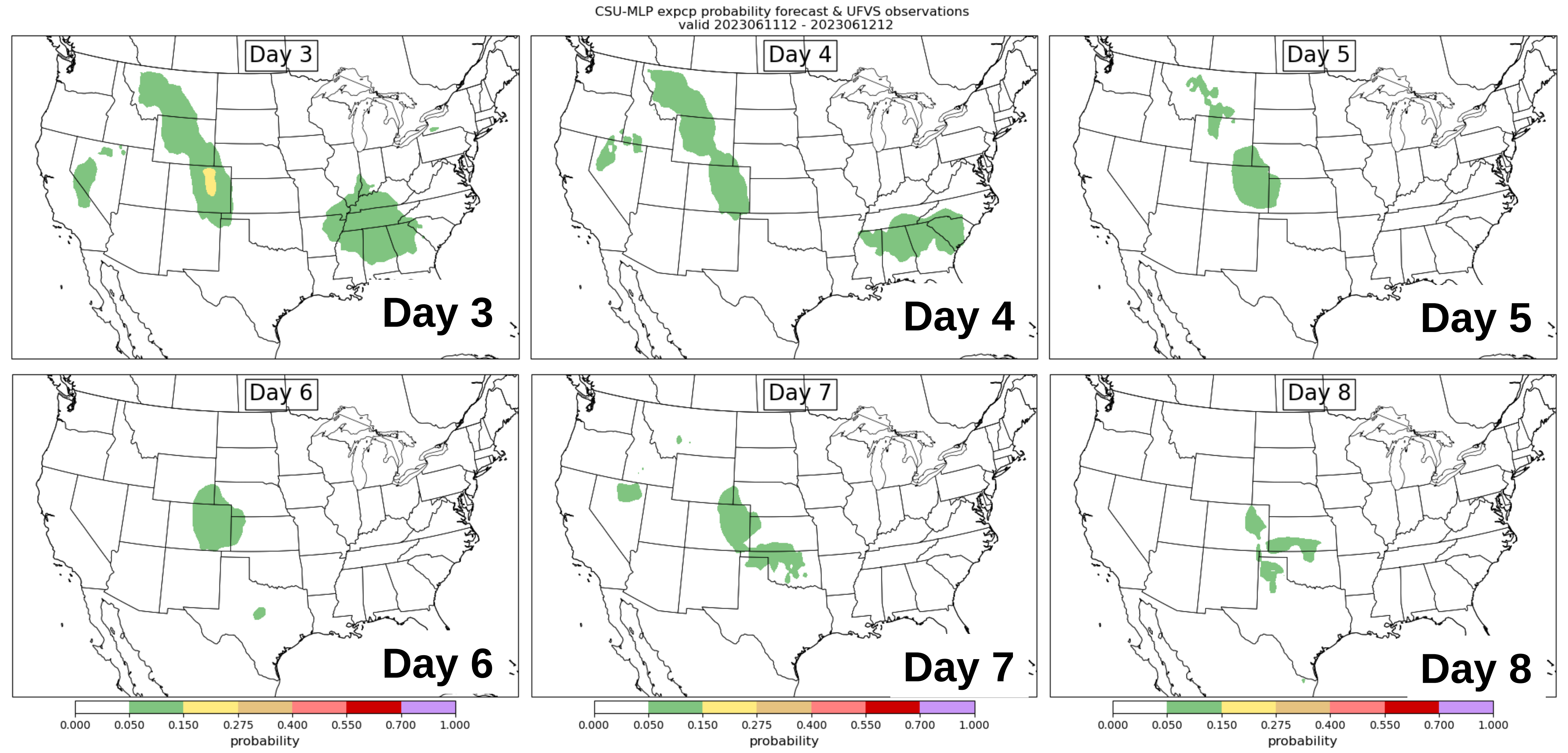
Good forecast example: 27-28 December 2022 (California flooding)

Fixed frequency model



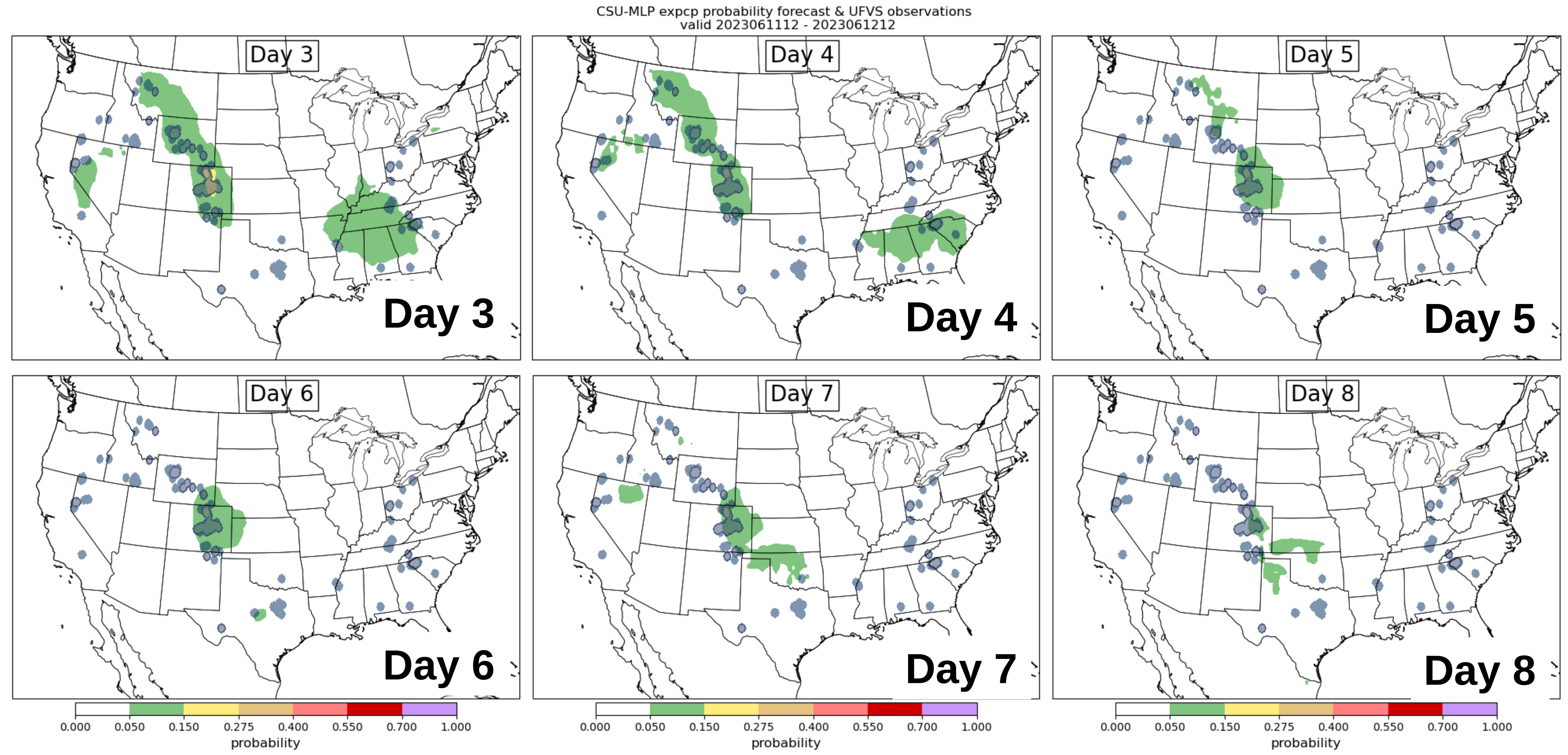
Good forecast example: 11-12 June 2023 (Colorado flash flooding)

Fixed frequency model



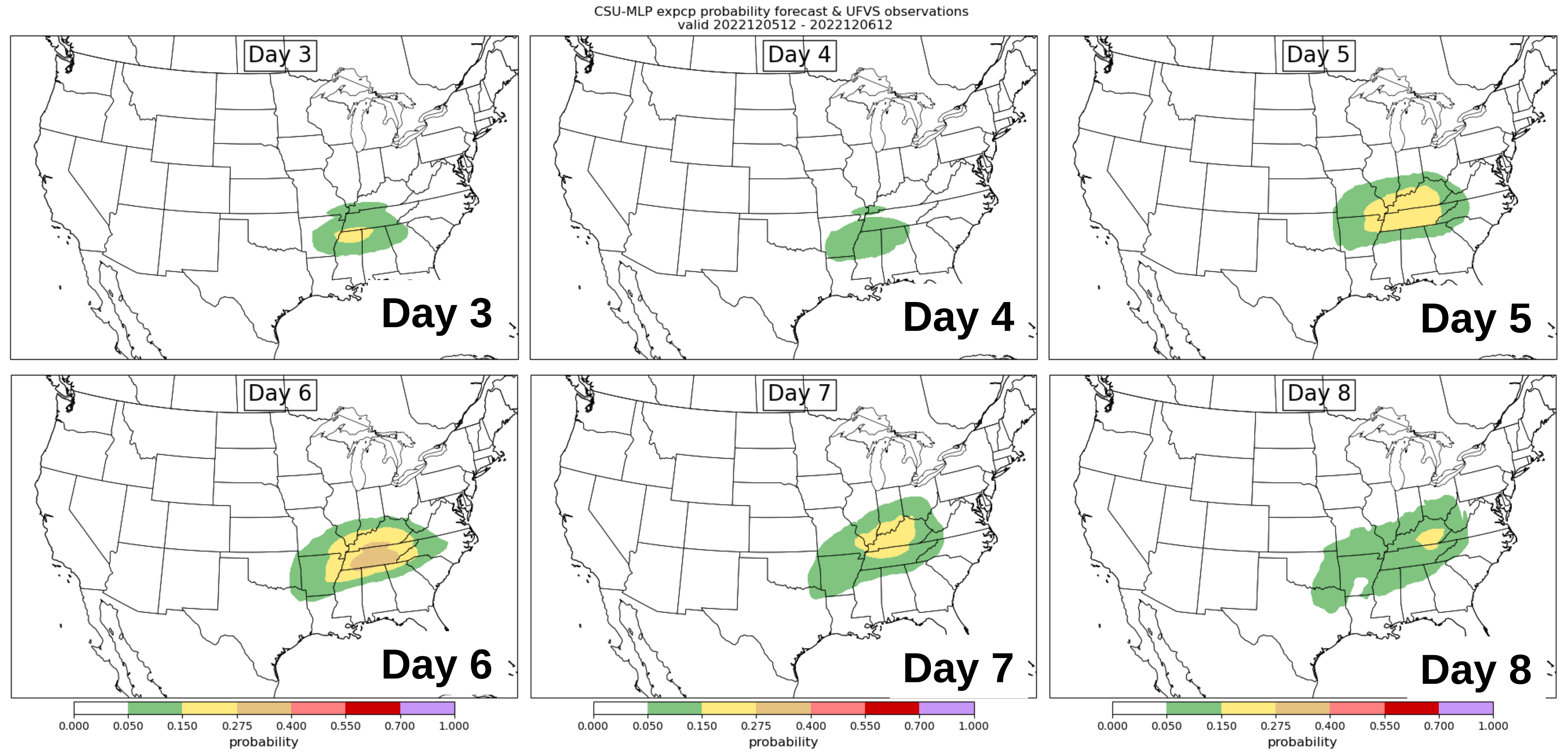
Good forecast example: 11-12 June 2023 (Colorado flash flooding)

Fixed frequency model



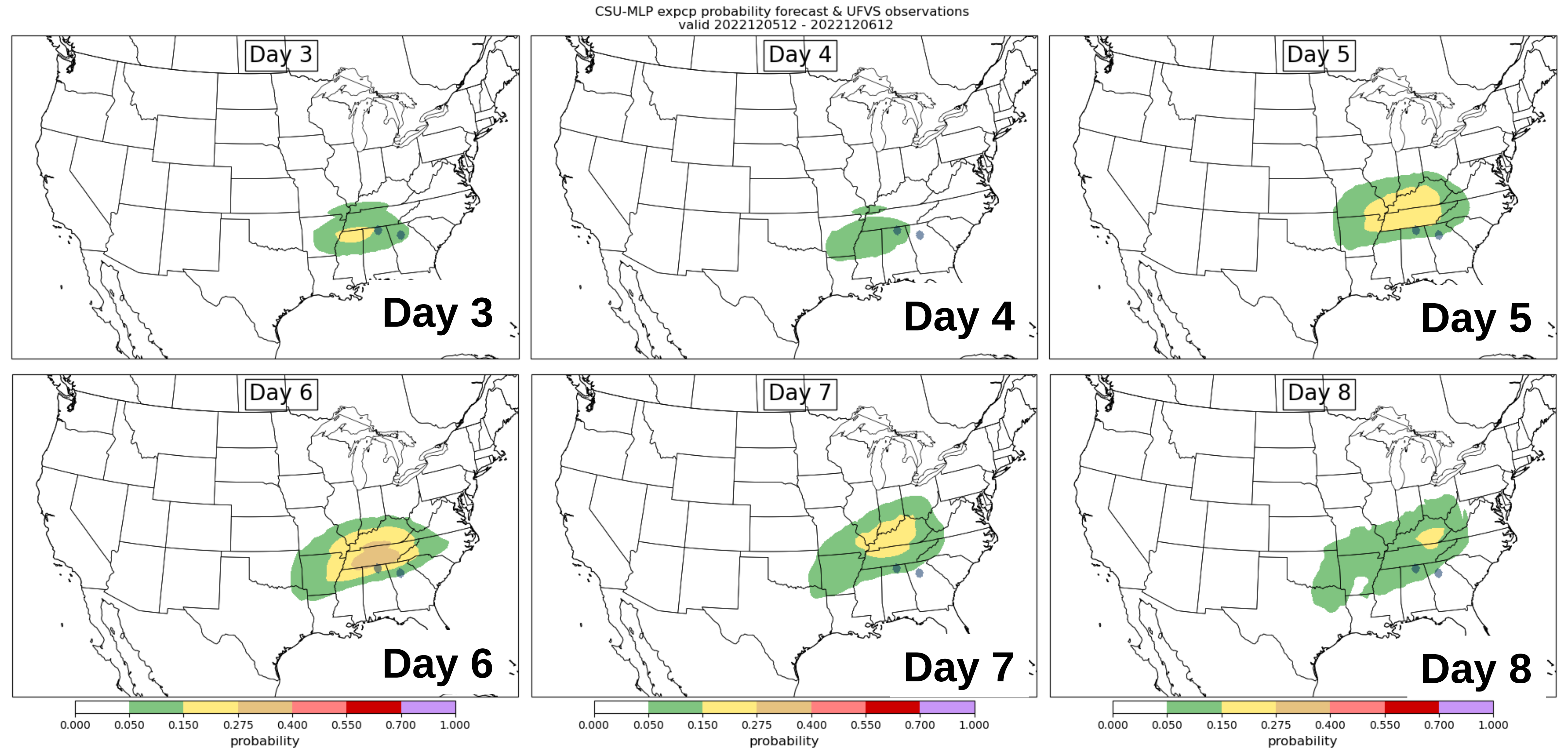
Poor forecast example: 5-6 December 2022

Fixed frequency model

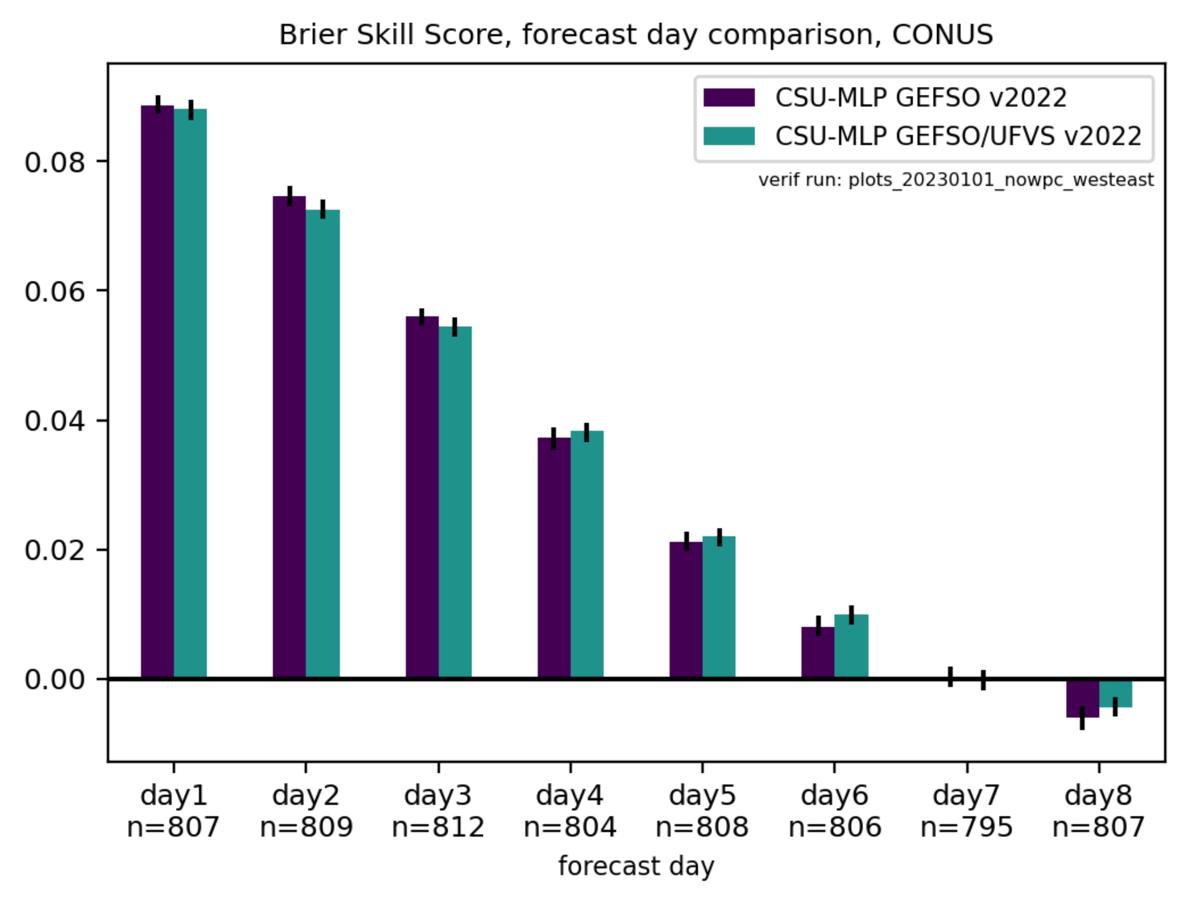


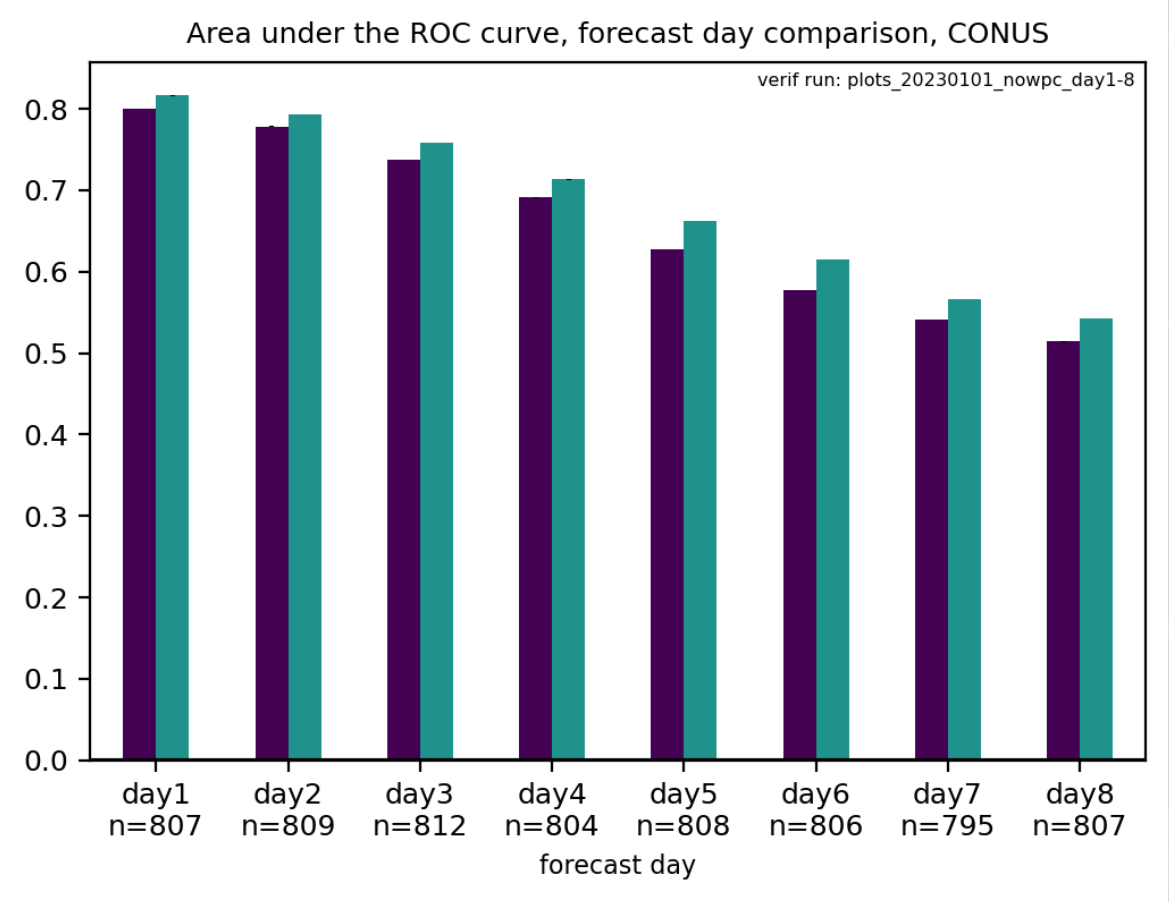
Poor forecast example: 5-6 December 2022

Fixed frequency model



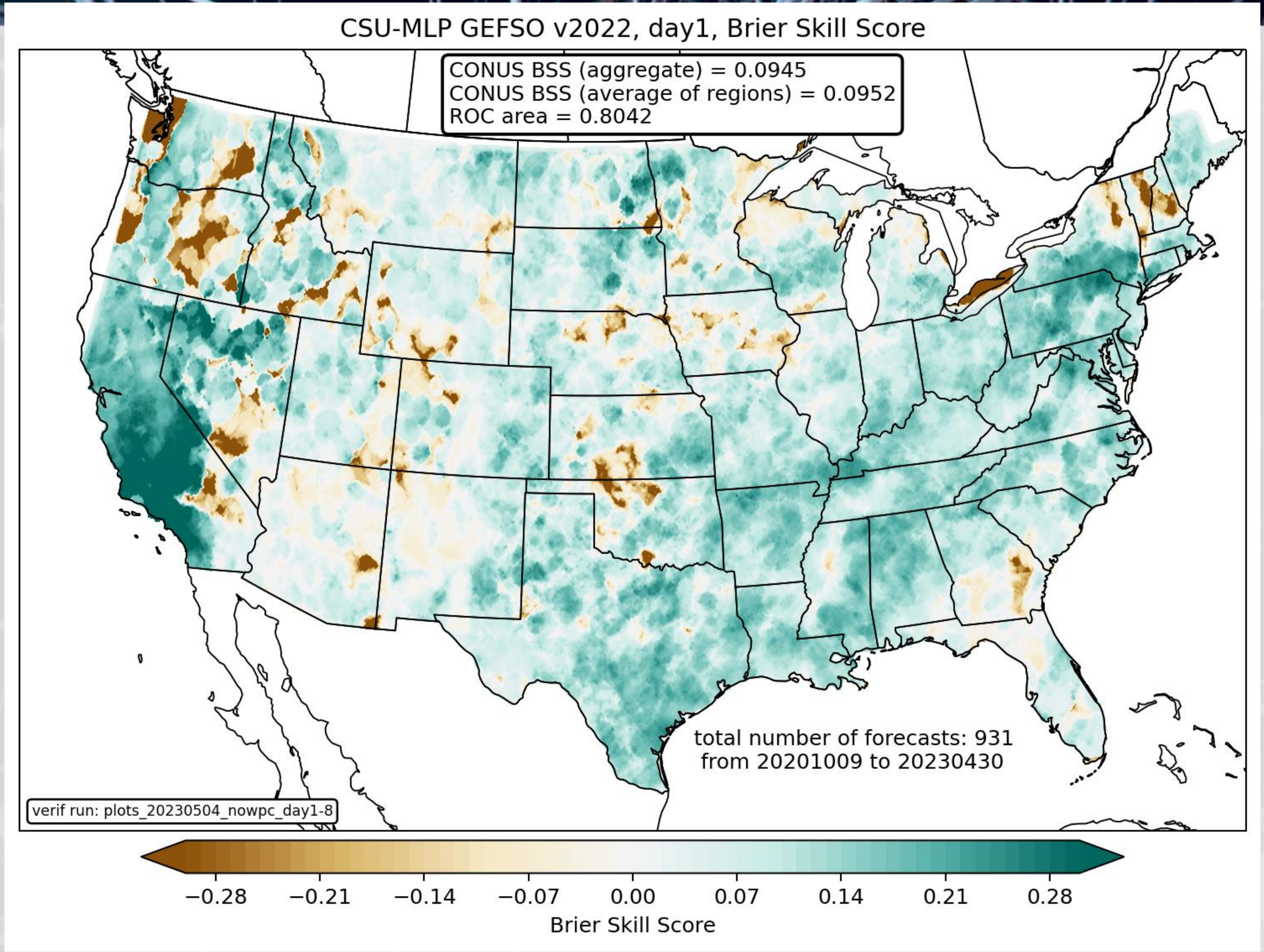
Brier skill score and ROC area, CONUS, by day





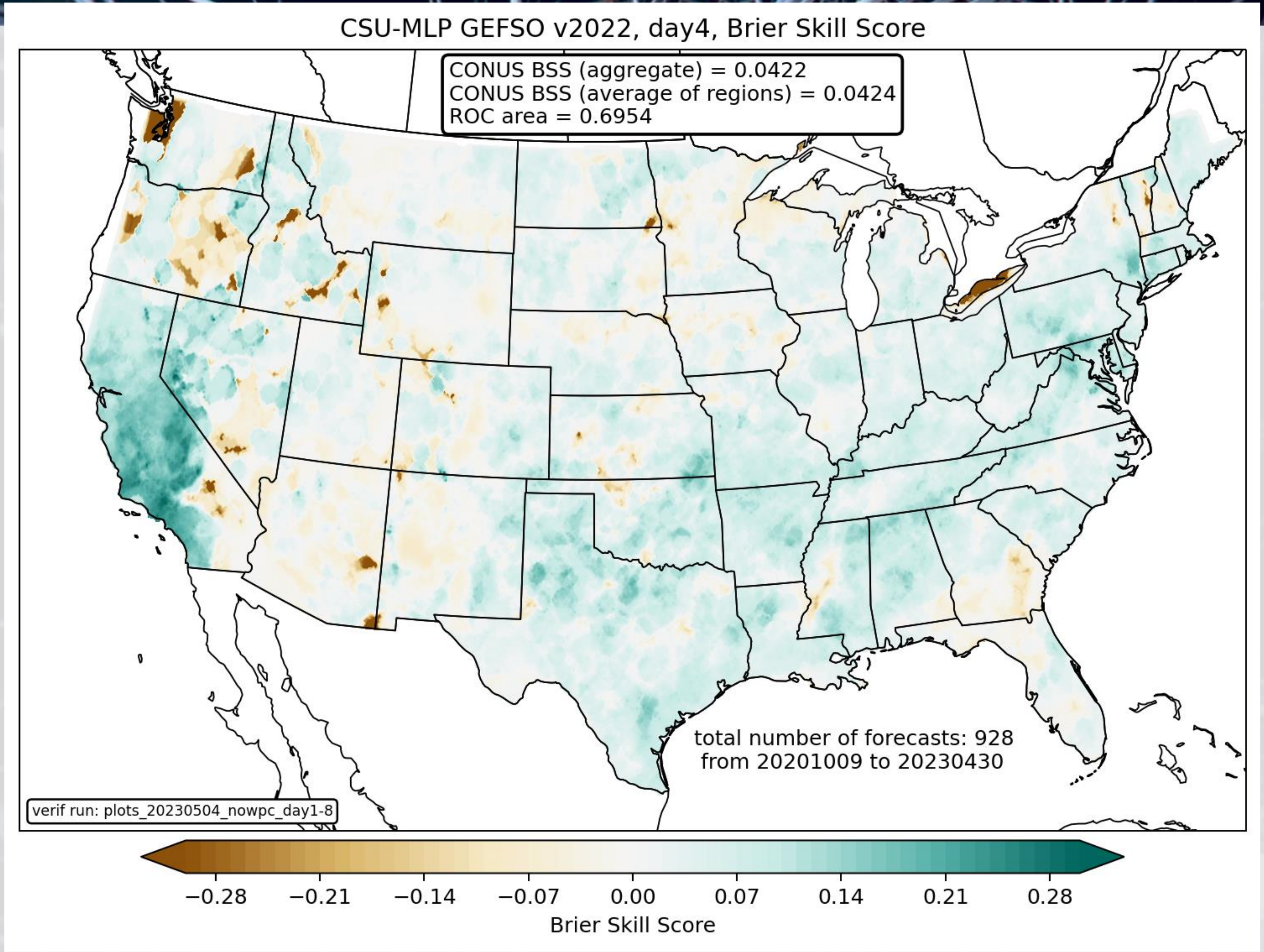
- Fixed frequency model displays slightly better skill while UFVS-trained model has better AUROC

Day 1 BSS



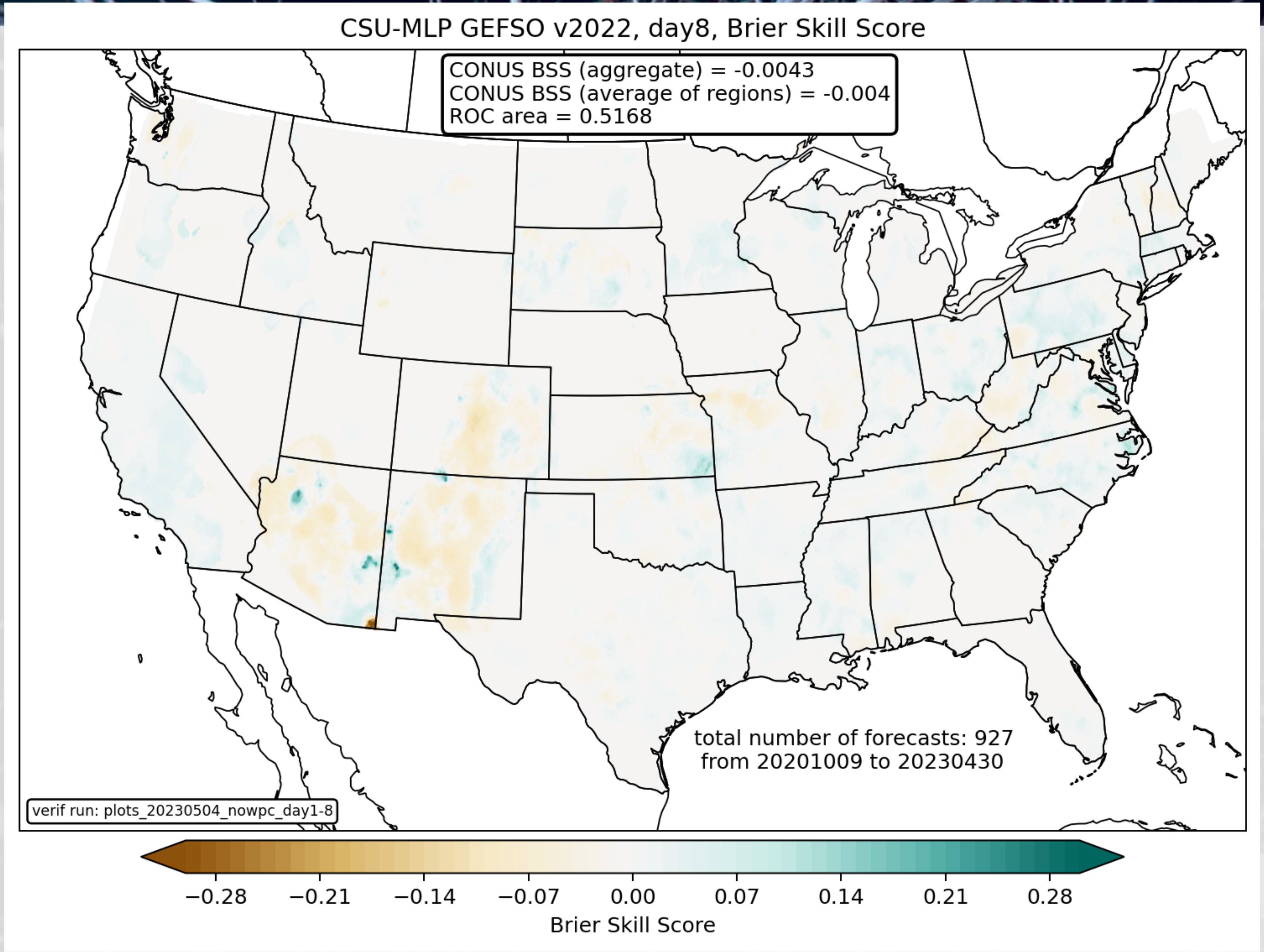
Fixed frequency model

Day 4 BSS



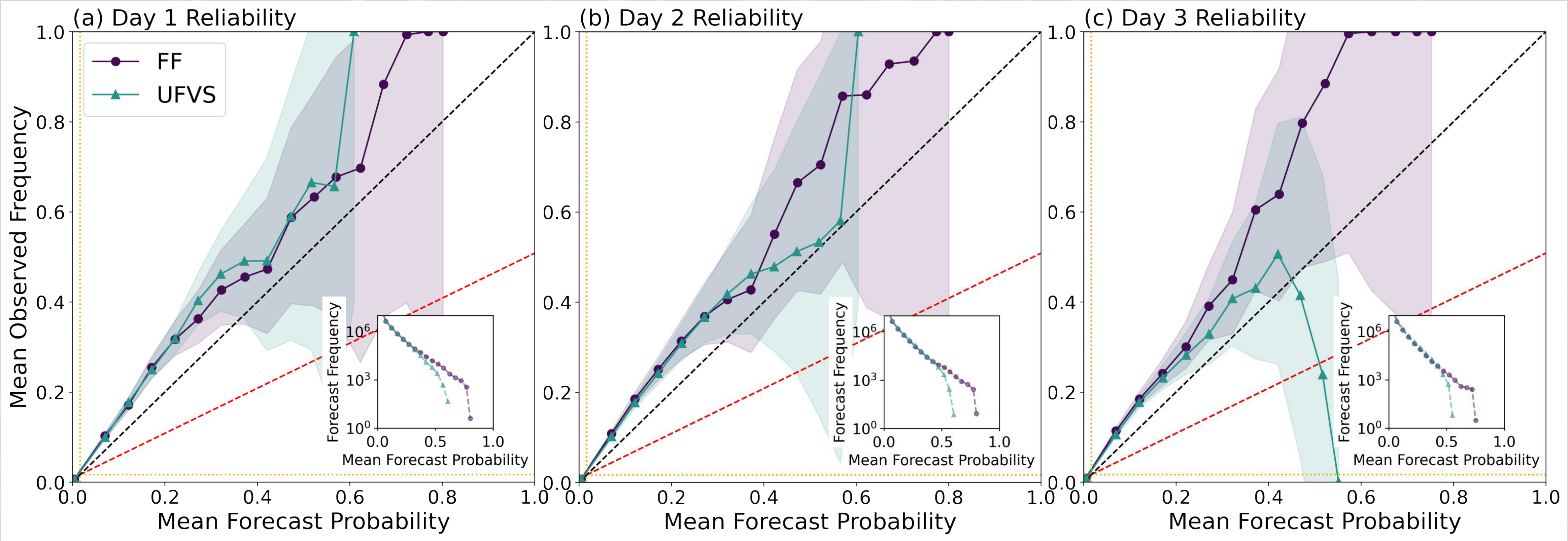
Fixed frequency model

Day 8 BSS



Fixed frequency model

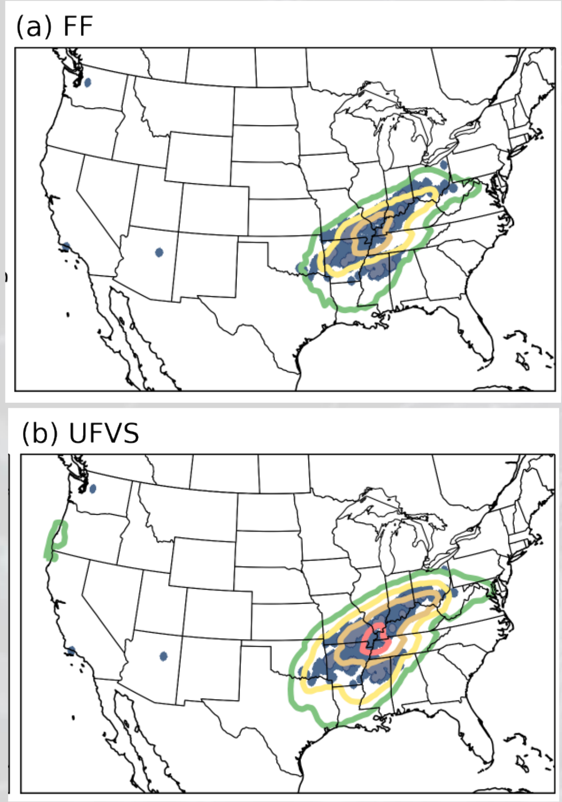
Both systems are underforecasting at nearly all probabilistic levels

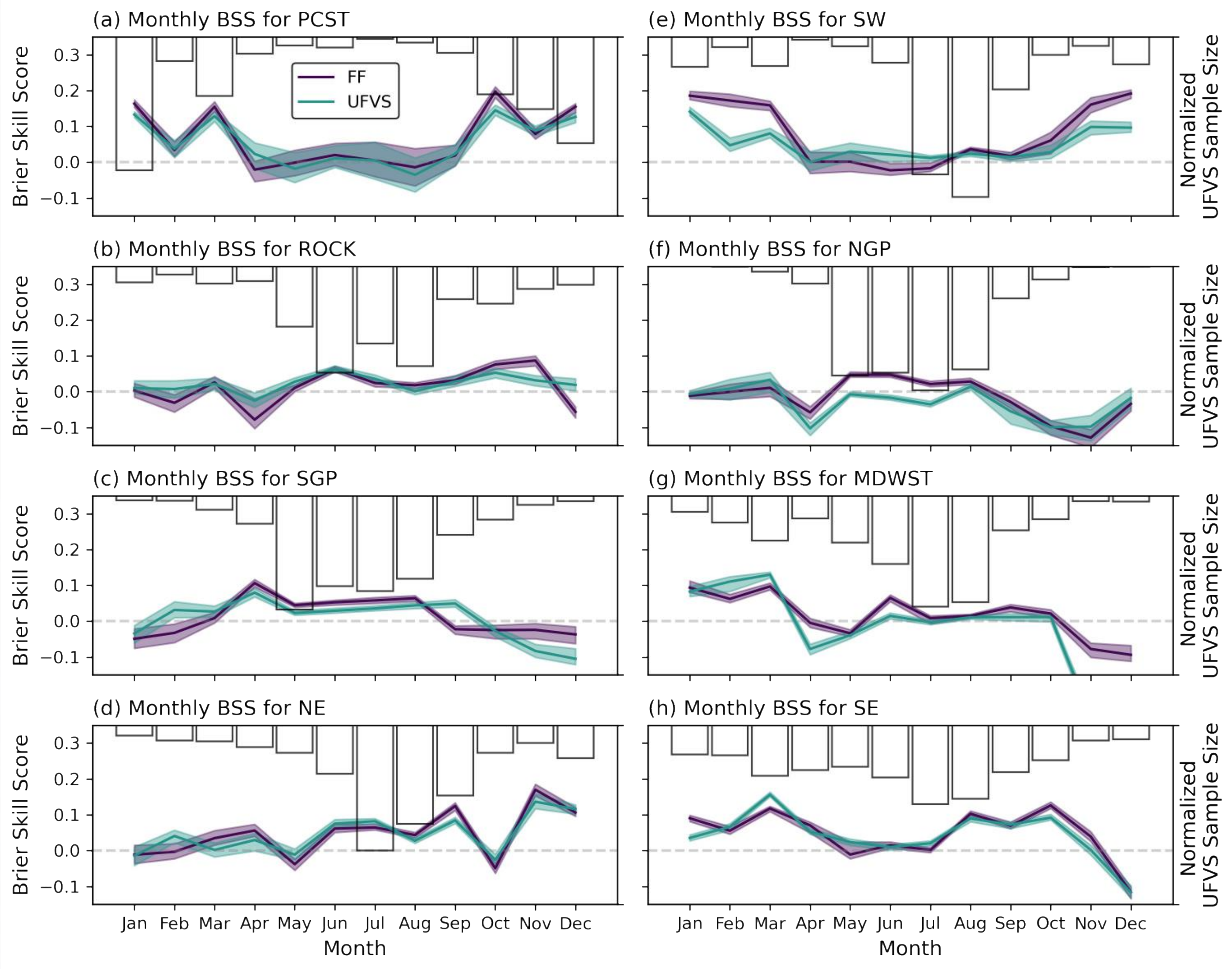


Hill et al. (2024, in review)

Skill between systems varies regionally and seasonally

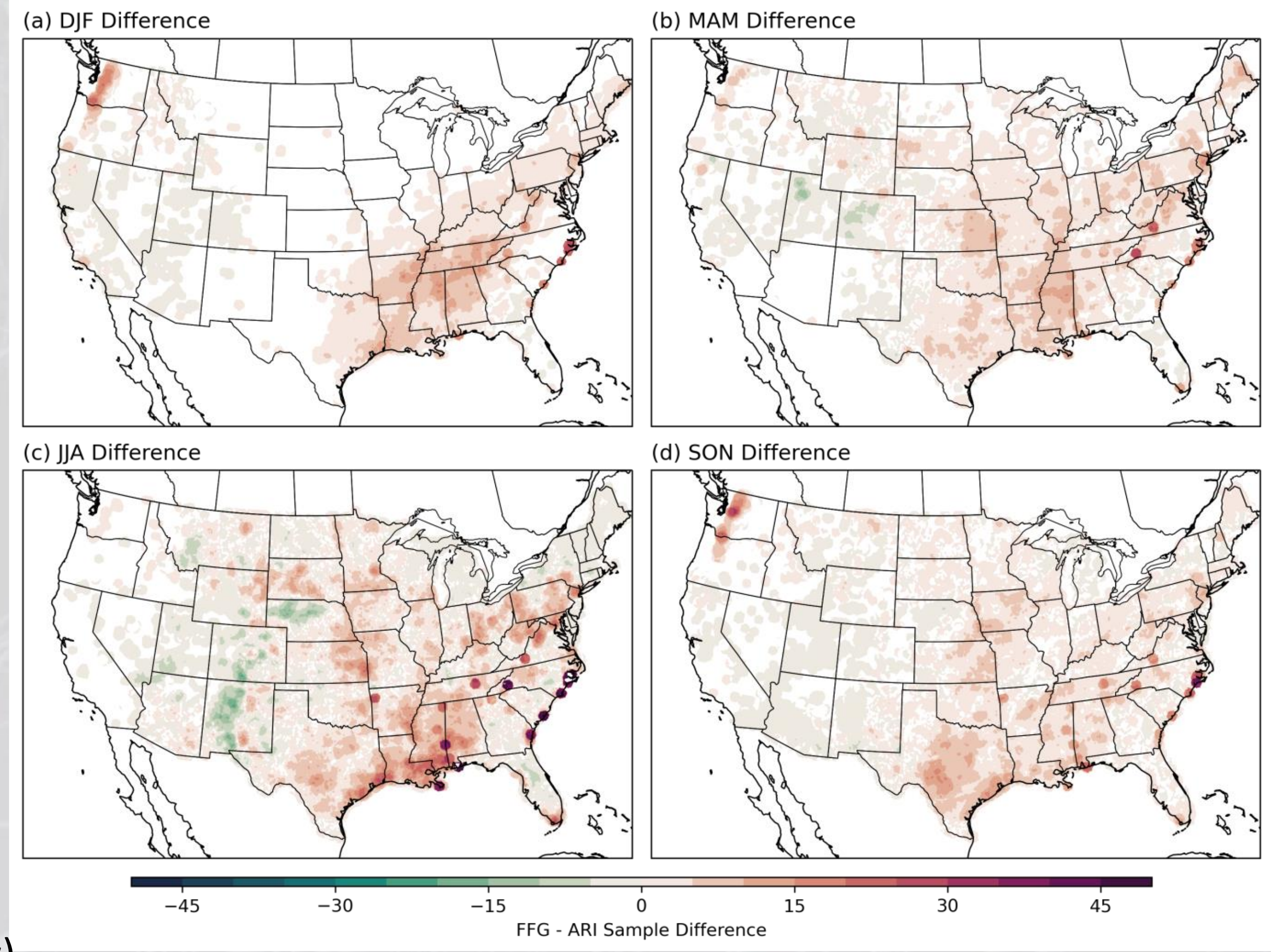
Forecasters have noted that the UFVS model tends to be "hotter" in the SE; FF model tends to be "hotter" in the West





Variability dominated by training sample differences across the CONUS

FFG – ARI samples in UFVS dataset

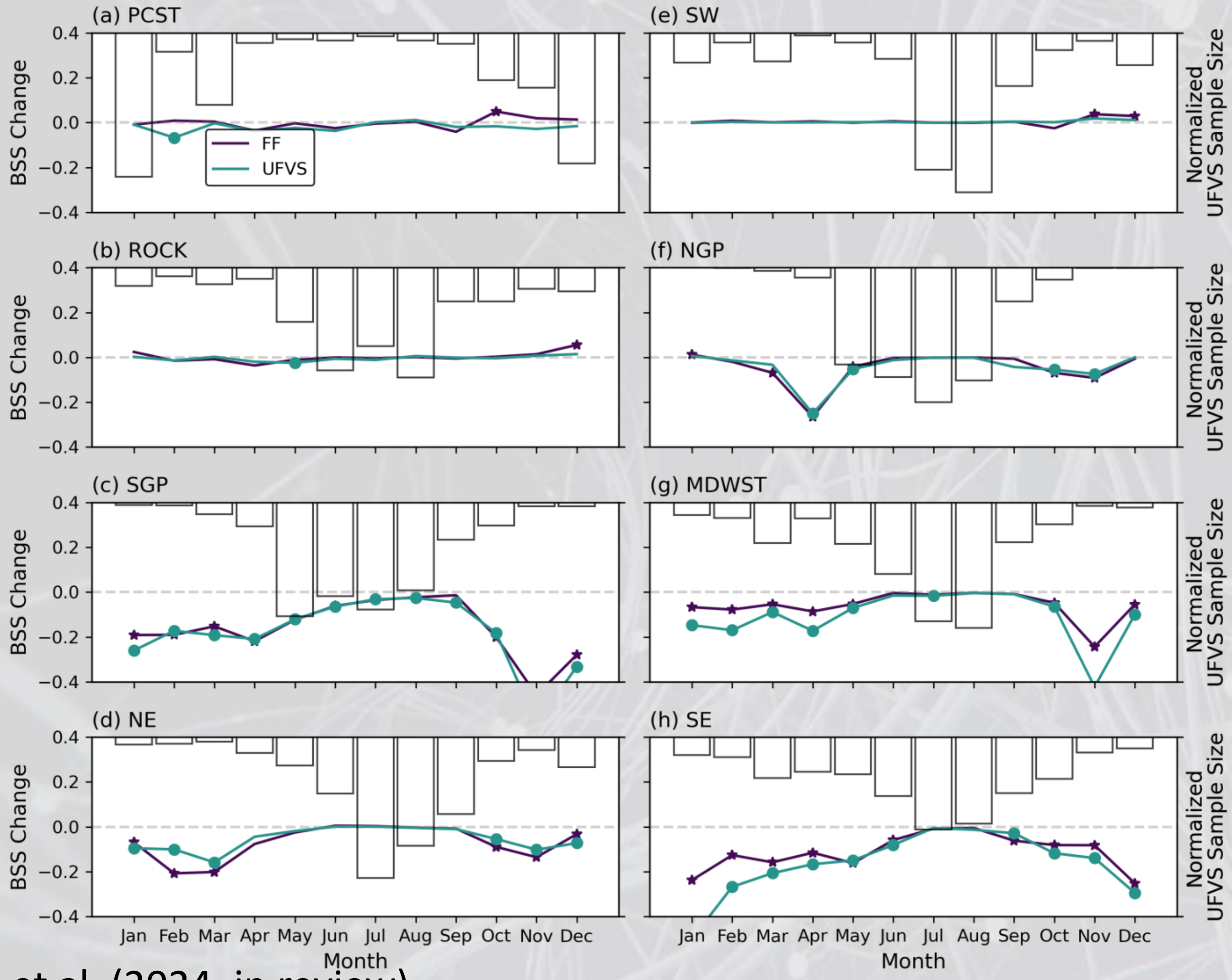


More ARI

More FFG

Hill et al. (2024, in review)

Variability dominated by training sample differences across the CONUS



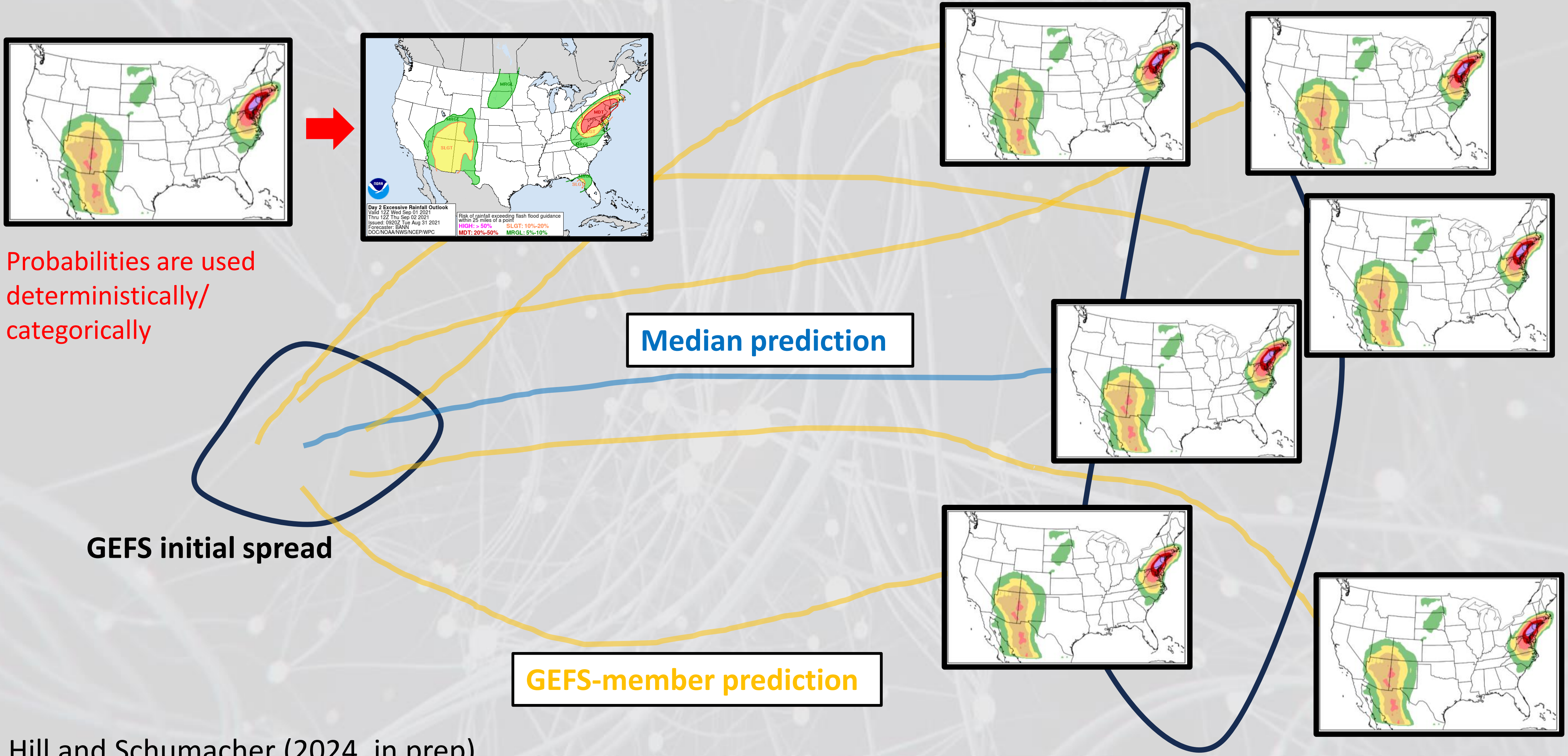
Hill et al. (2024, in review)

Ablation experiments in which FFG is removed from the UFVS dataset for verification only

Skill is reduced in regions where FFG exceedances are more frequent compared to ARI exceedances

Skill in UFVS-trained system degrades more -> deriving forecast skill from learning on FFG exceedance events

Shift from deterministic to ensemble framework

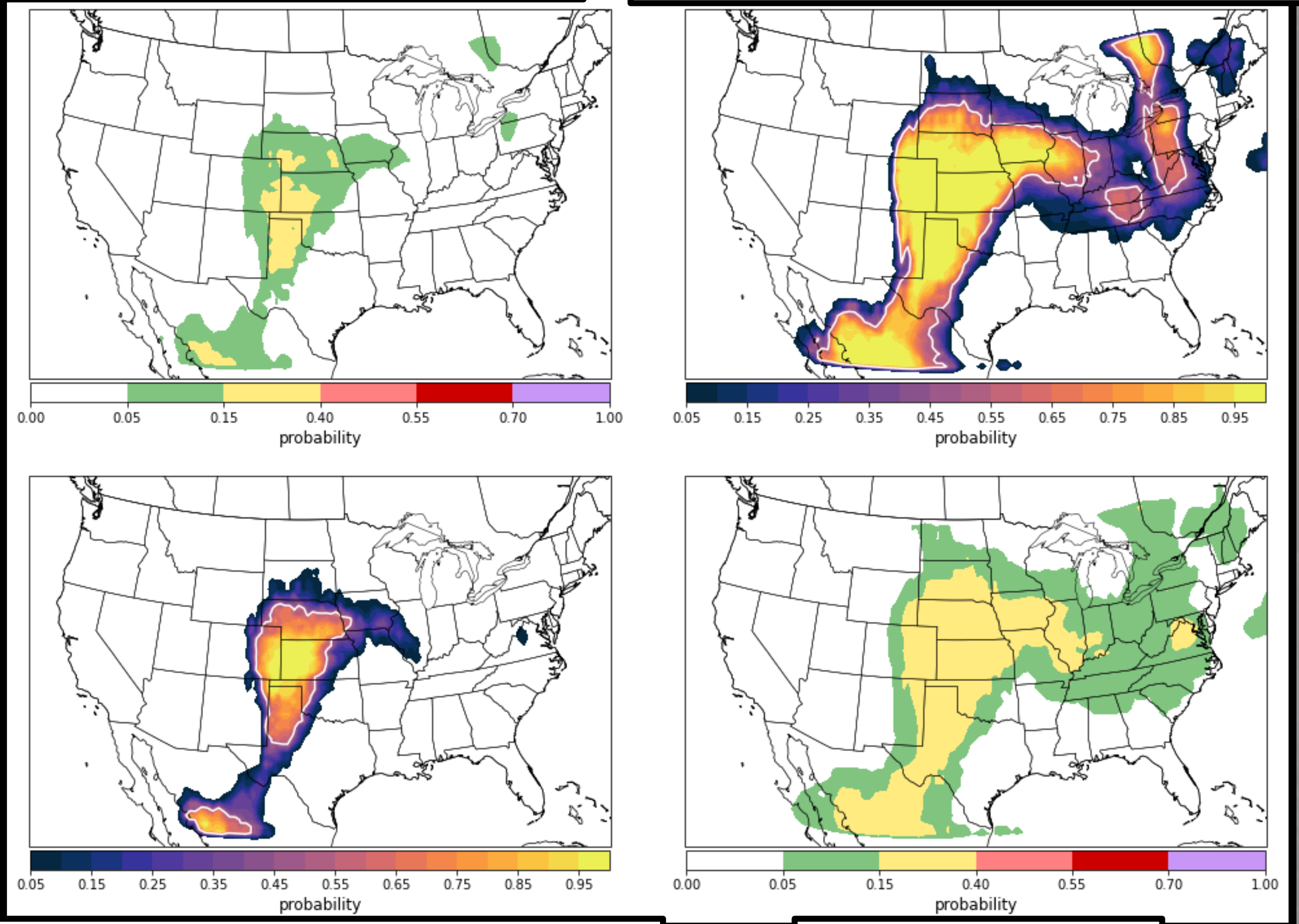


Hill and Schumacher (2024, in prep)

Shift from deterministic to ensemble framework

Day 1 MLP Excessive Rainfall Forecast
Valid 12 UTC 30 June 2023 – 12 UTC 1 July 2023

Probability of at least a “Marginal Risk” (5-15%)
21-member ensemble of MLP forecasts



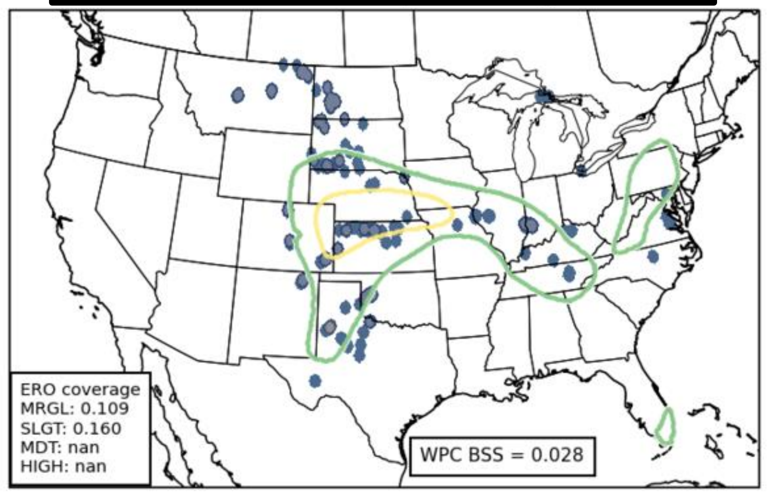
Probability of at least a “Slight Risk” (15-40%)
21-member ensemble of MLP forecasts

Ensemble max probability

Provides probabilistic information to operational forecaster end users who want to know about distribution of ML predictions

Further extends skill of hazard predictions?
TBD

Weather Prediction Center ERO and corresponding observations for valid period



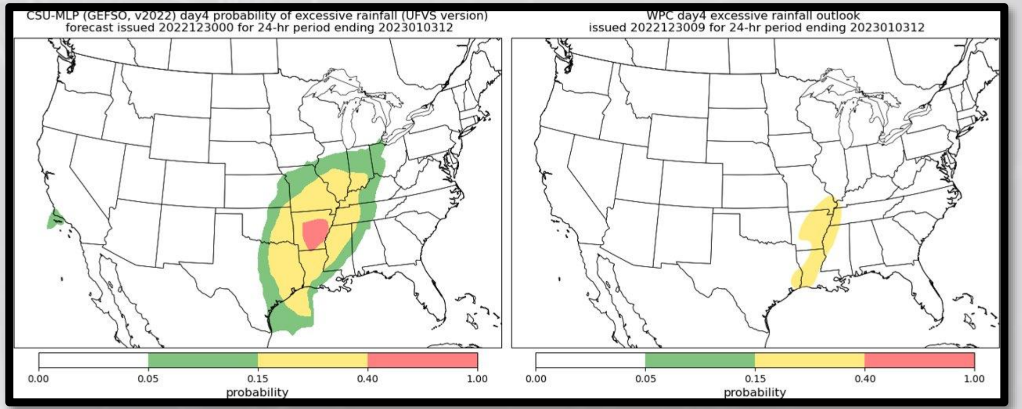
Summary and Discussion

The CSU-MLP is moving towards medium-range precipitation predictions to assist WPC operations – appears there is skill out to 6 days in these products

Comprehensive analysis is still in progress as we try to quantify the strengths and weaknesses of longer-range products, their forecast characteristics, and ultimately their value in the forecast process

The future is in ensemble predictions!

One of the most important aspects of AI-based predictions is building trust with end-users (e.g., forecasters) – models and forecasts should be <u>interpretable</u> and <u>explainable → FFaIR participation is a big component to building trust and getting user feedback to inform future development</u>



MLP real-time forecast graphics:
https://schumacher.atmos.colostate.edu/hilla/csu_mlp

Contact with questions/comments: ahill@ou.edu