Progress towards medium range excessive rainfall forecasts with the CSU-MLP

Russ S. Schumacher and Aaron J. Hill
Department of Atmospheric Science, Colorado State University

ATMOSPHERIC SCIENCE
COLORADO STATE UNIVERSITY

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NOAA
U.S. DEPARTMENT OF COMMERCE

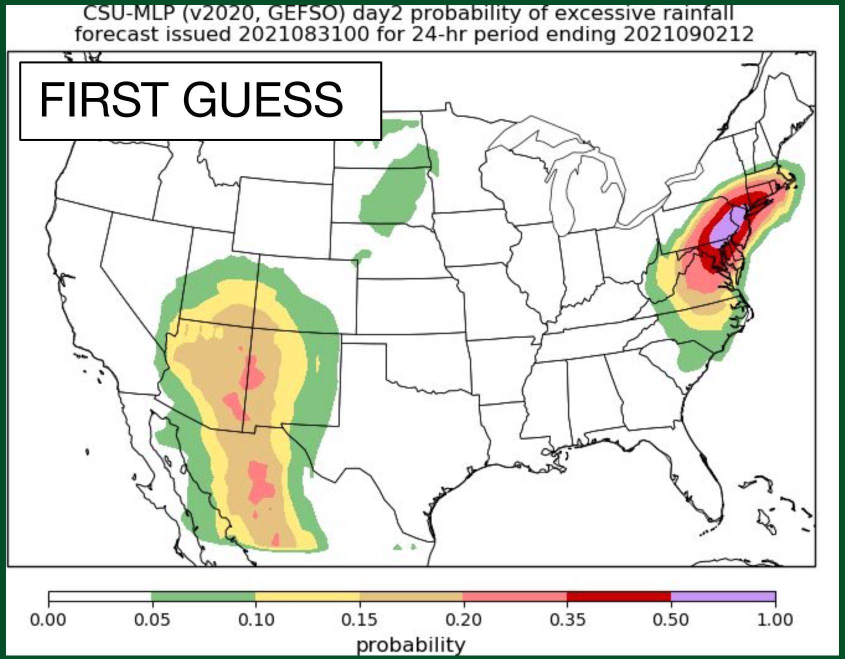
Along with NOAA partners: Mark Klein and Jim Nelson (WPC)
And contributions from Allie Mazurek (CSU) and Hanna McDaniel (FSU)

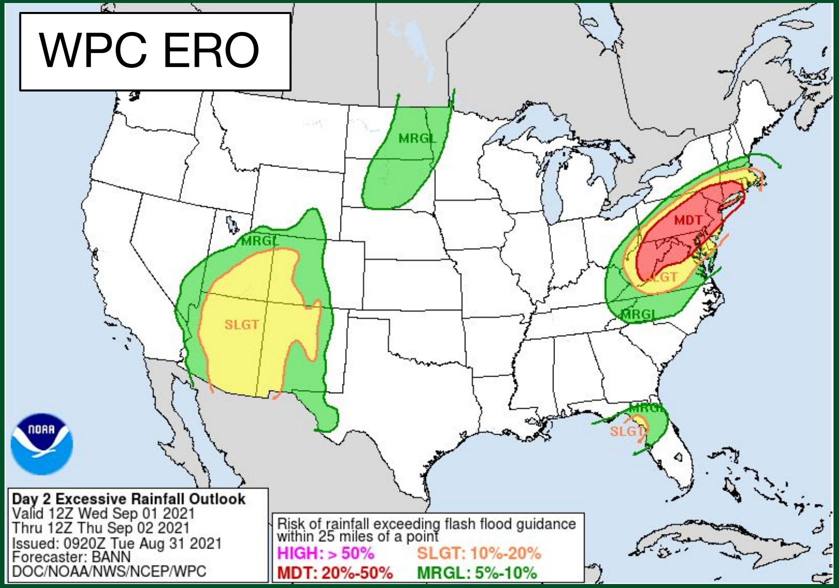
Research supported by NOAA JTTI grant NA21OAR4590187

Flash Flood and Intensive Rainfall Experiment Seminar Series
June 2023

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



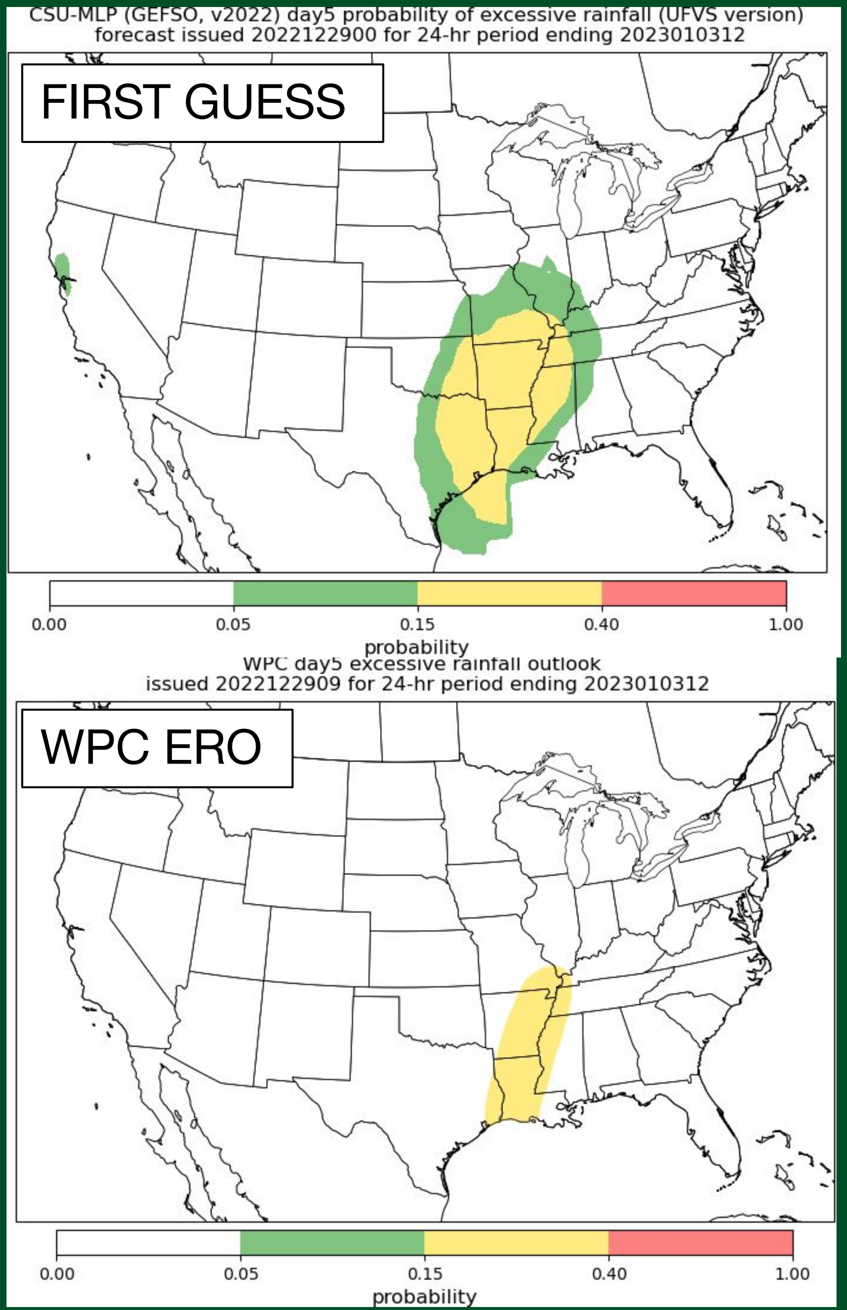


Real-time forecast graphics:
http://schumacher.atmos.colostate.edu/hilla/csu_mlp/

Schumacher et al. (2021, *BAMS*)

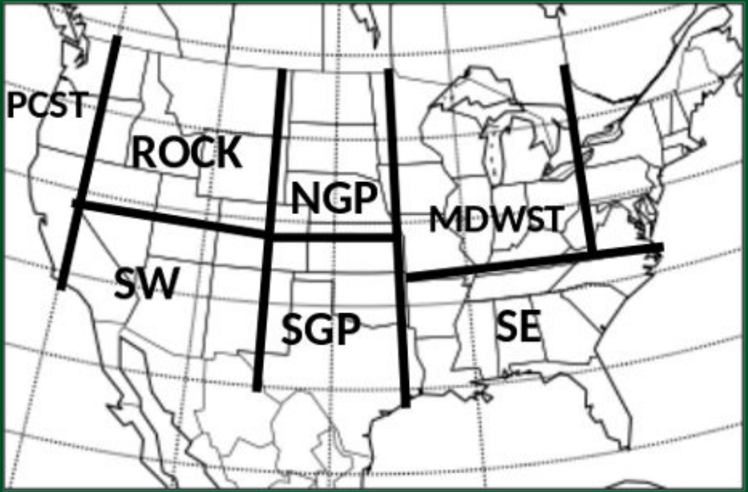
Background

- In 2022, WPC began issuing experimental day 4-5 EROs
- **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*)**
- **But is there actual forecast skill at these lead times?**



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?
- Produces a flash flood report?
- More than a certain threshold? (and if so, which one(s)?)
- What quantitative precipitation estimate to use?

What are we trying to predict?

We have chosen to use two frameworks/datasets:

- "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
  - 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)
  - Flash flood reports, exceedances of FFG or the 5-yr ARI, and reports of flooding from USGS stream gauges, MPING

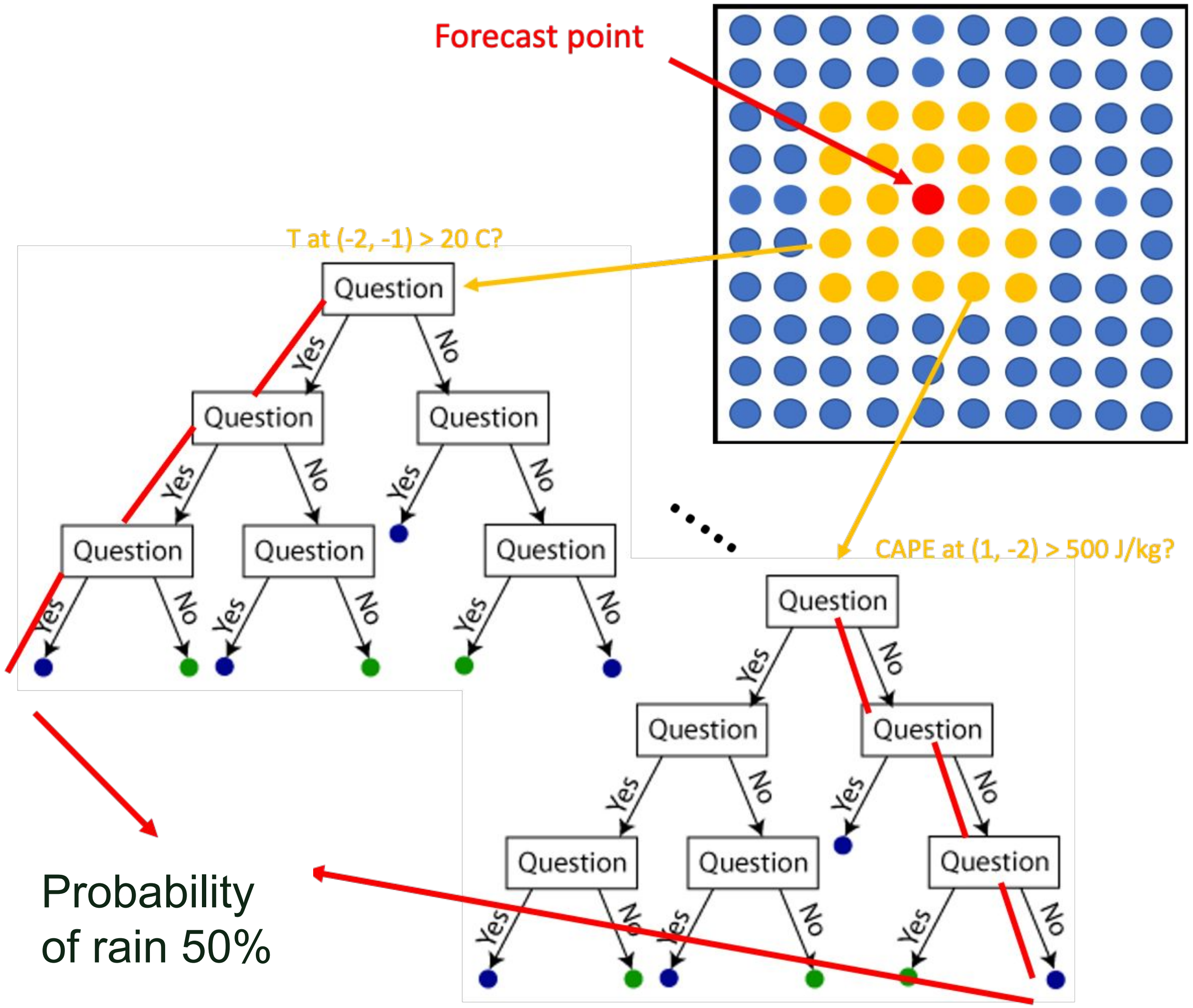
Random Forest Primer

A set of decision trees that contain a series of yes/no questions (branches) based on input predictors that allow traversal of the tree

Corresponding events of excessive rainfall are assigned to the “leaf” nodes

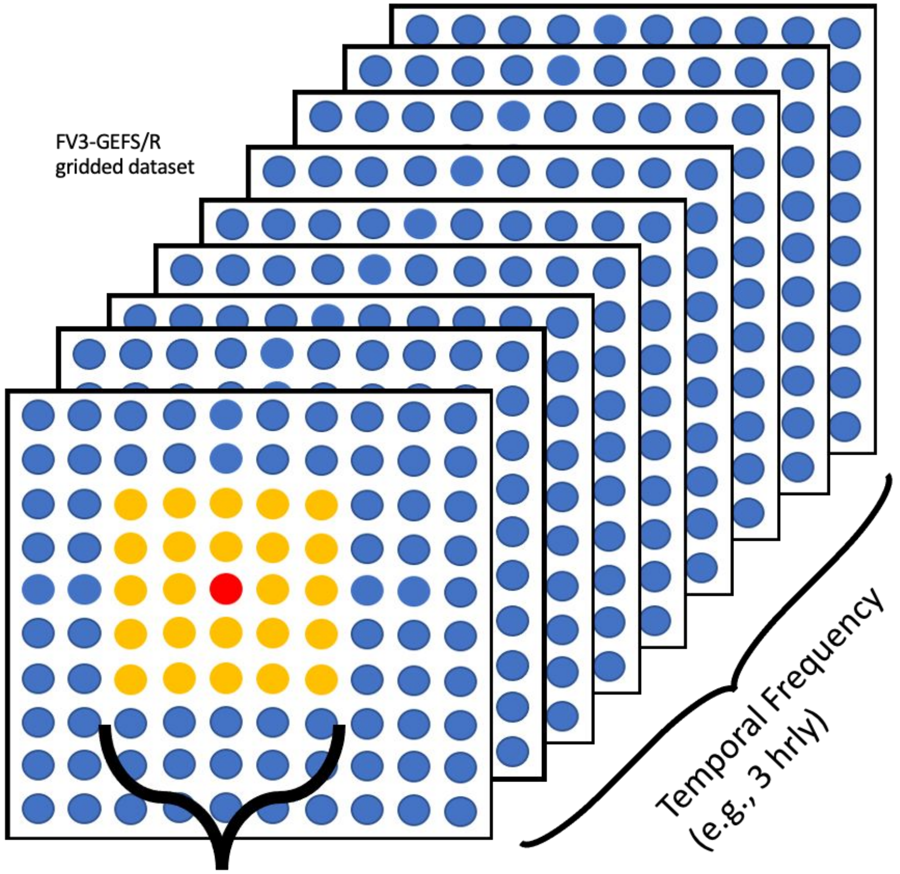
Relative frequency of events in the forest is the forecast probability

Python package "determines" best predictors that discriminate events



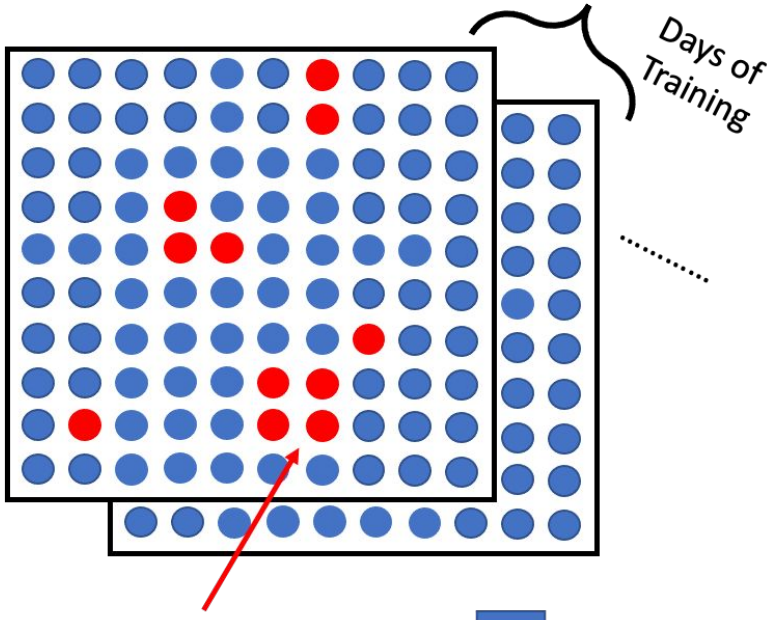
Random Forest Configuration

FV3-GEFS/R
gridded dataset

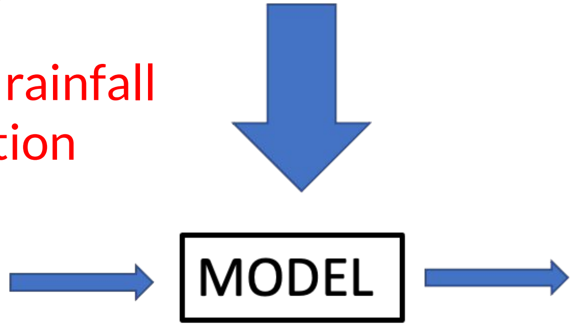


Spatial frequency (5 grid points)

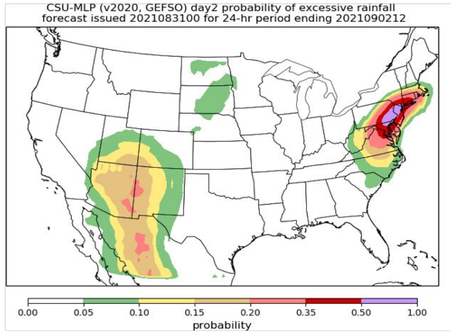
Variables
APCP
CAPE
CIN
MSLP
PWAT
Q2
T2M
0-500 Shear
0-850 Shear
U10
V10
UV10



Gridded rainfall
observation

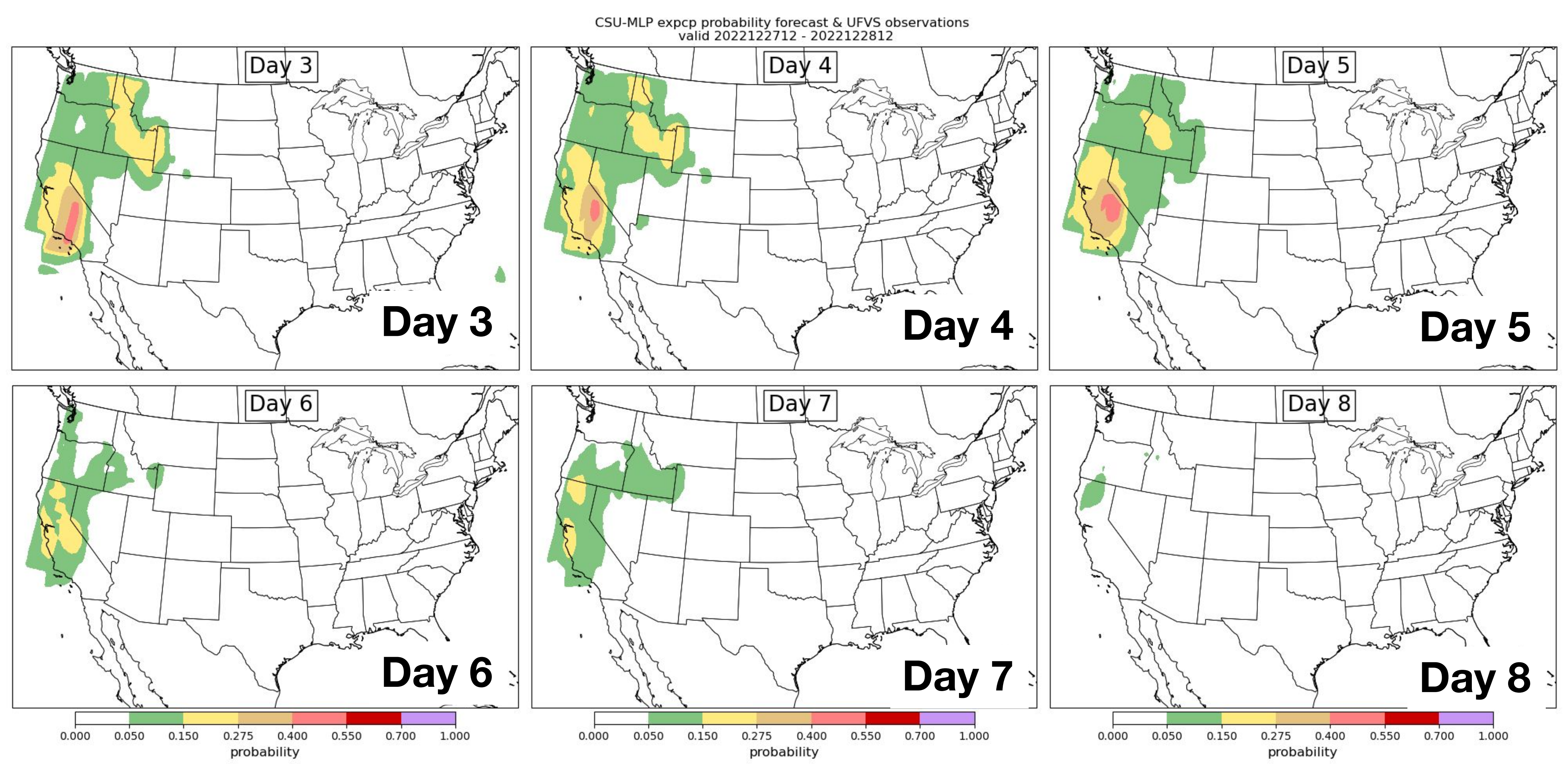


12

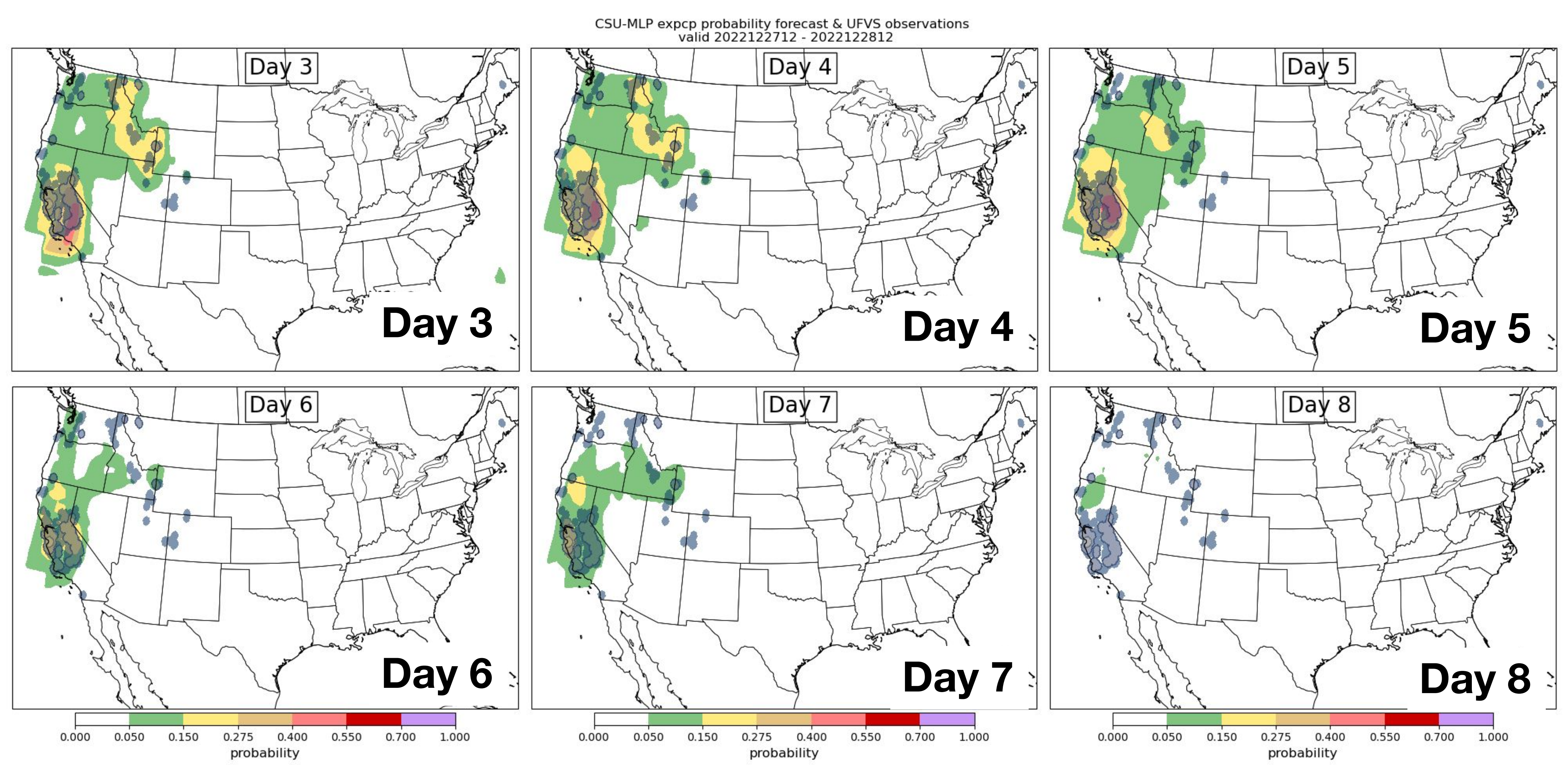


Good forecast example: 27-28 December 2022 (California flooding)

Fixed frequency model

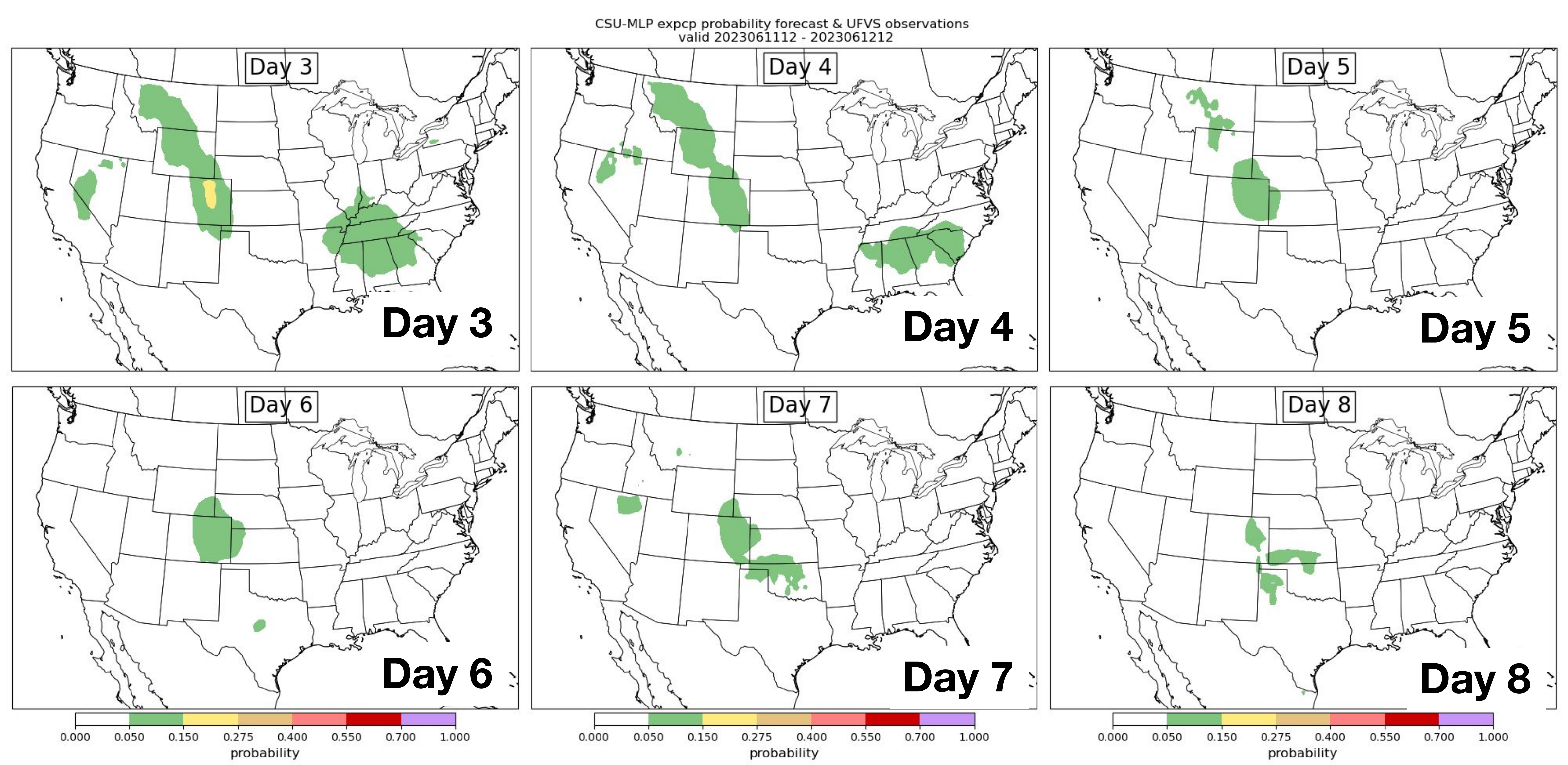


Good forecast example: 27-28 December 2022 (California flooding)

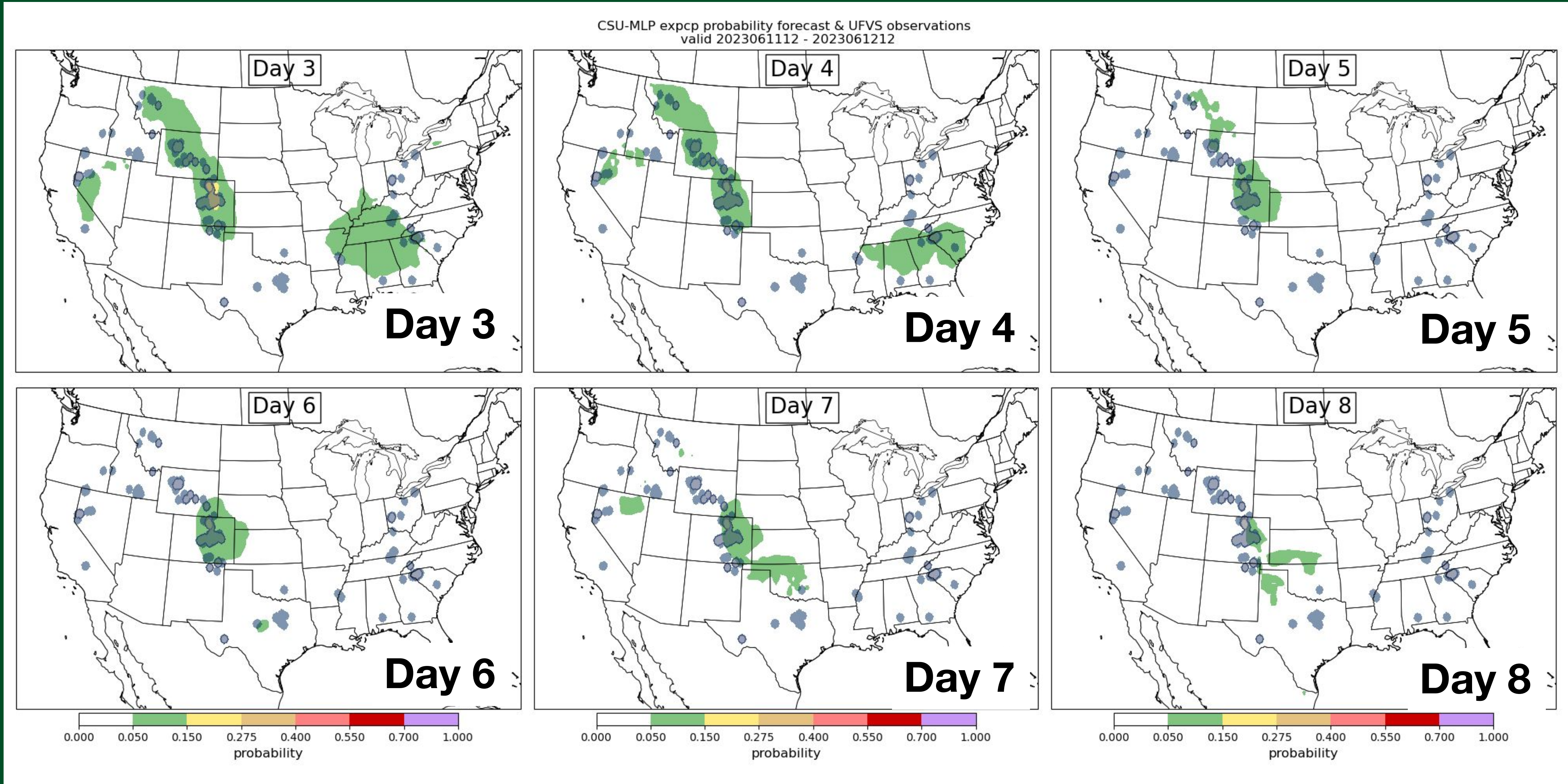


Good forecast example: 11-12 June 2023 (Colorado flash flooding)

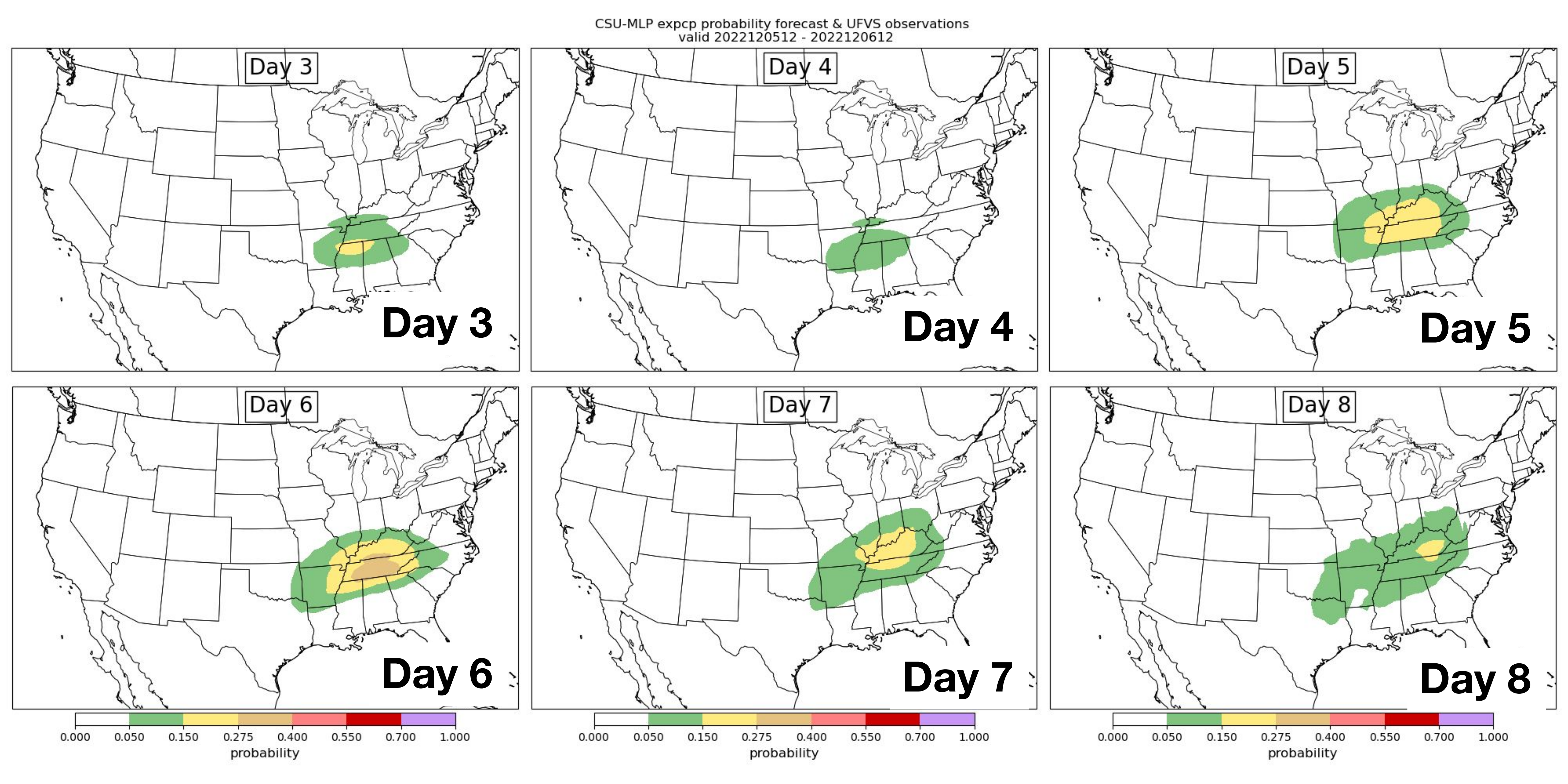
Fixed frequency model



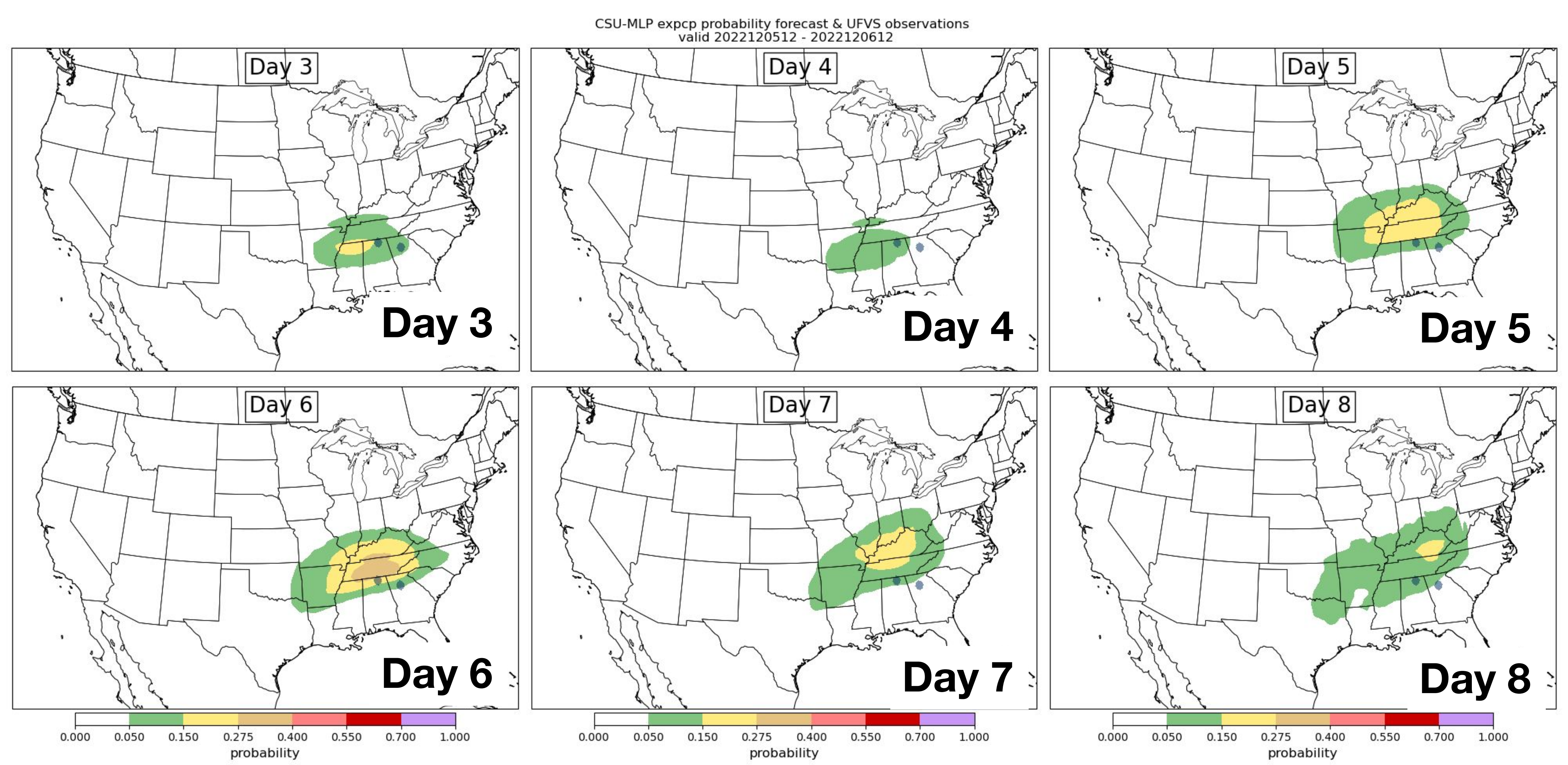
Good forecast example: 11-12 June 2023 (Colorado flash flooding)



Poor forecast example: 5-6 December 2022



Poor forecast example: 5-6 December 2022

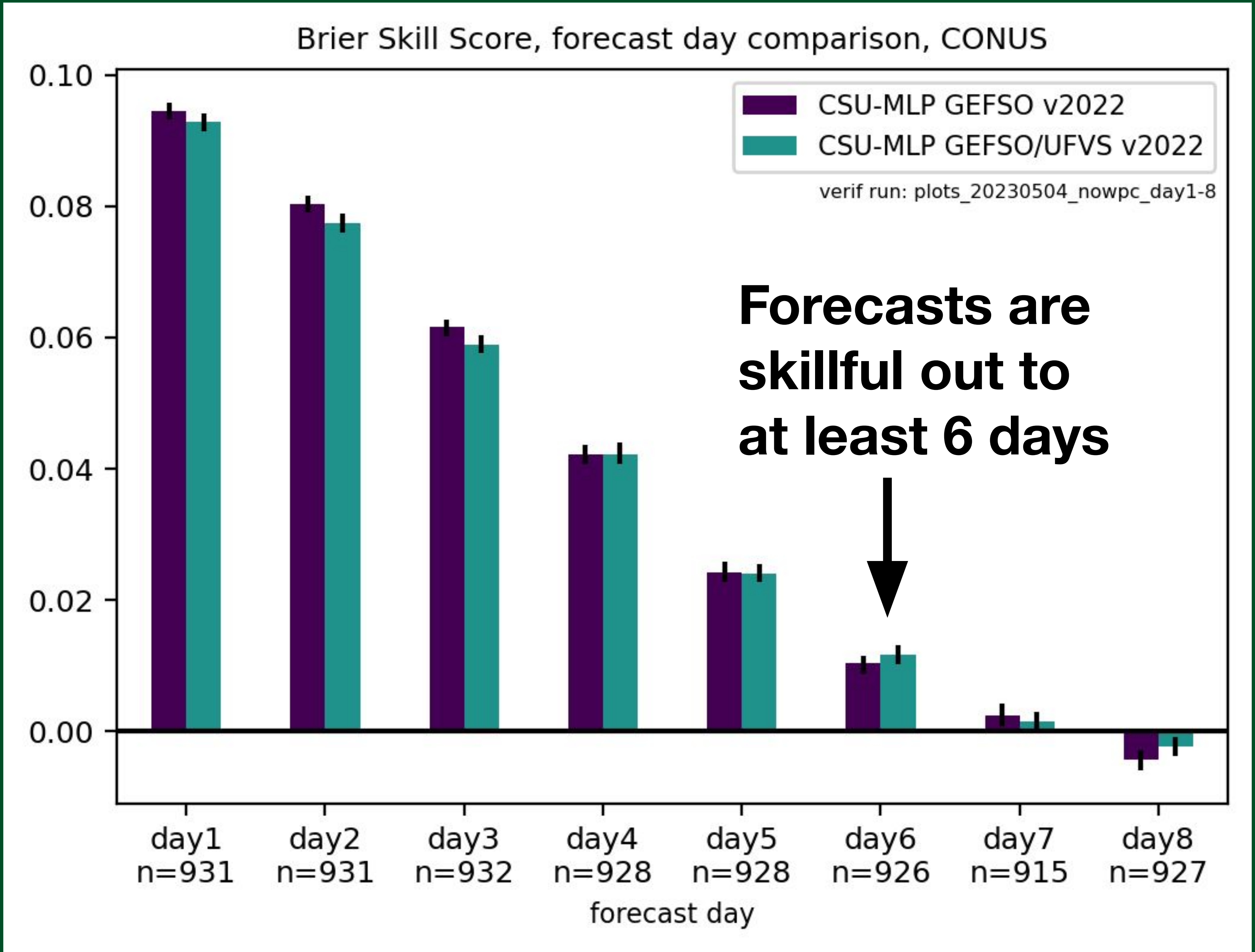


Verification methods

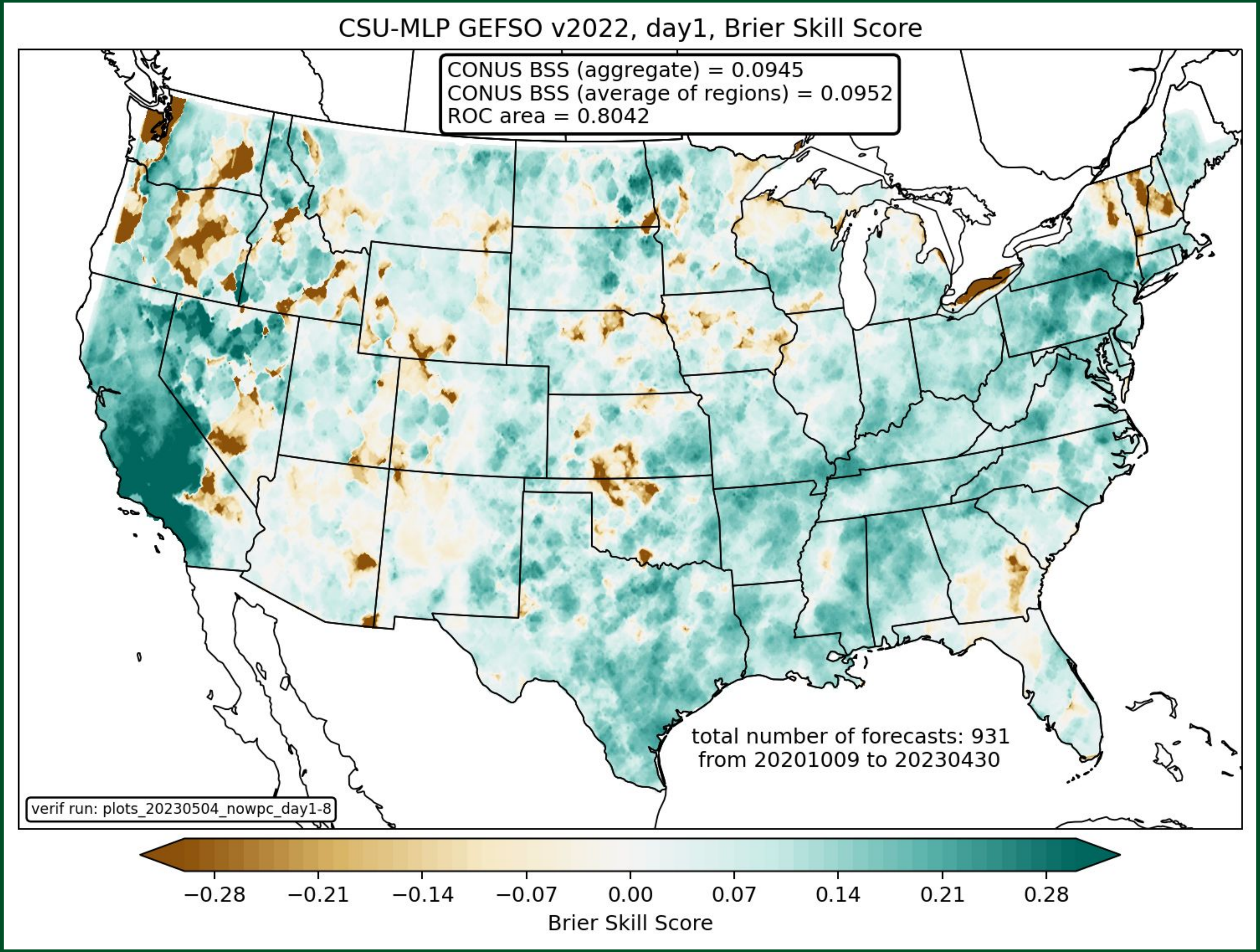
- Two versions of CSU-MLP forecasts were run retrospectively from October 2020 through April 2023
- Forecasts evaluated against WPC's Unified Flood Verification System (Erickson et al. 2019, 2021), includes flash flood guidance exceedances, 5-yr ARI exceedances, flash flood LSRs, USGS and MPING flood reports
- **For all forecasts, we use the "new" definitions for ERO probabilities: 5, 15, 40, 70%**

Brier skill score with respect to smoothed daily climatology, higher is better

~2.5 years of forecasts (Oct 2020-Apr 2023)

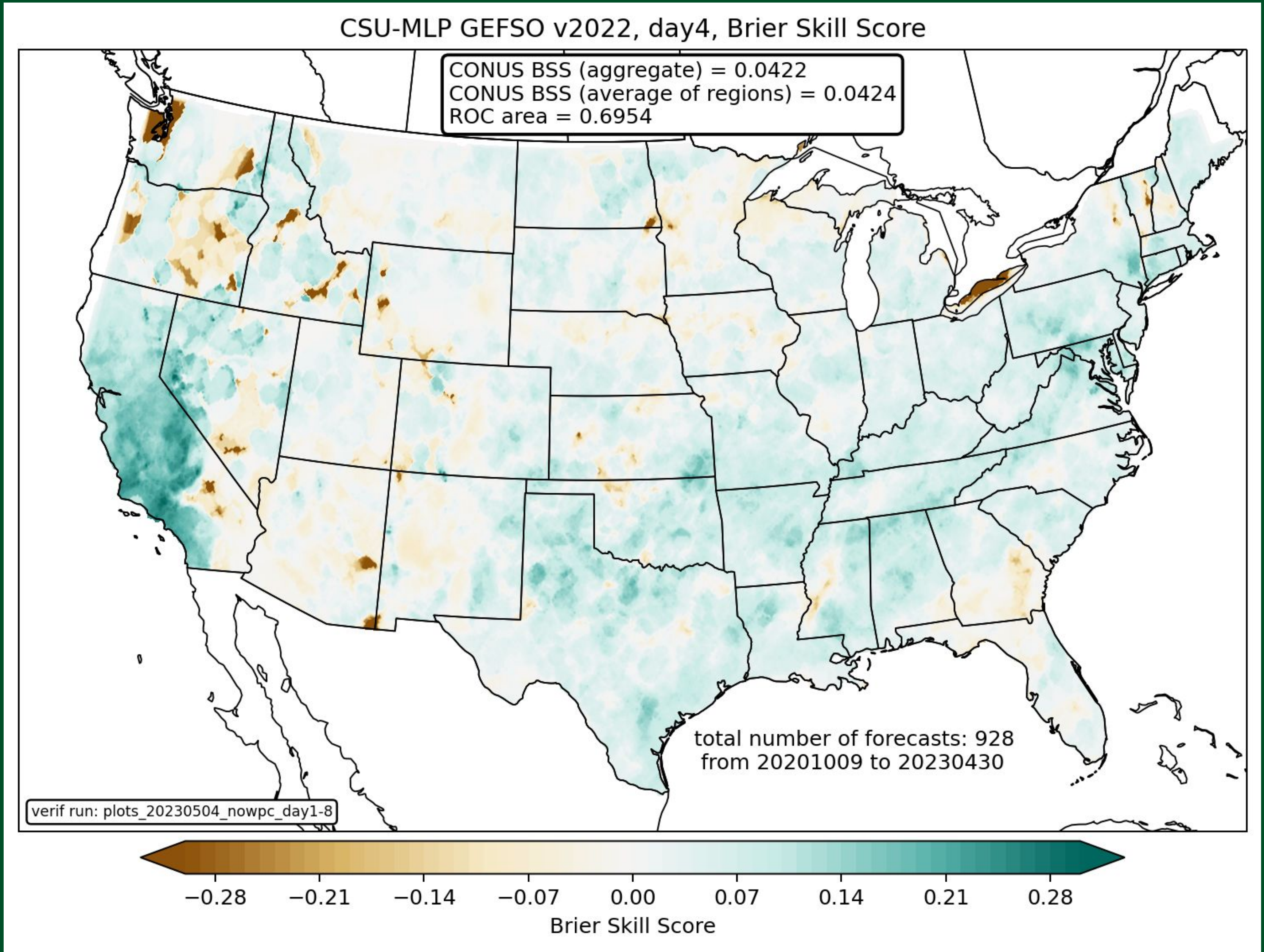


Day 1



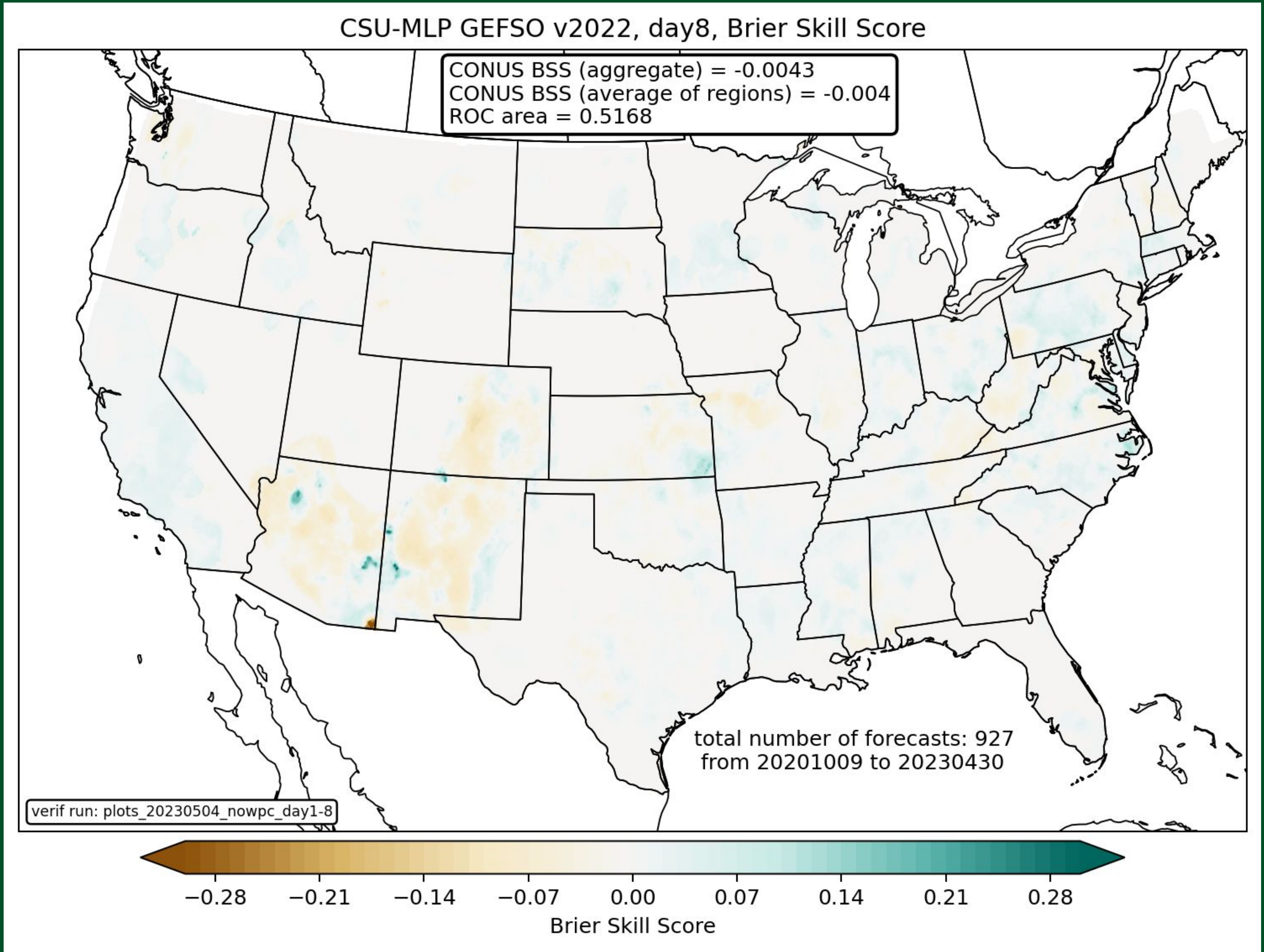
Fixed frequency model

Day 4



Fixed frequency model

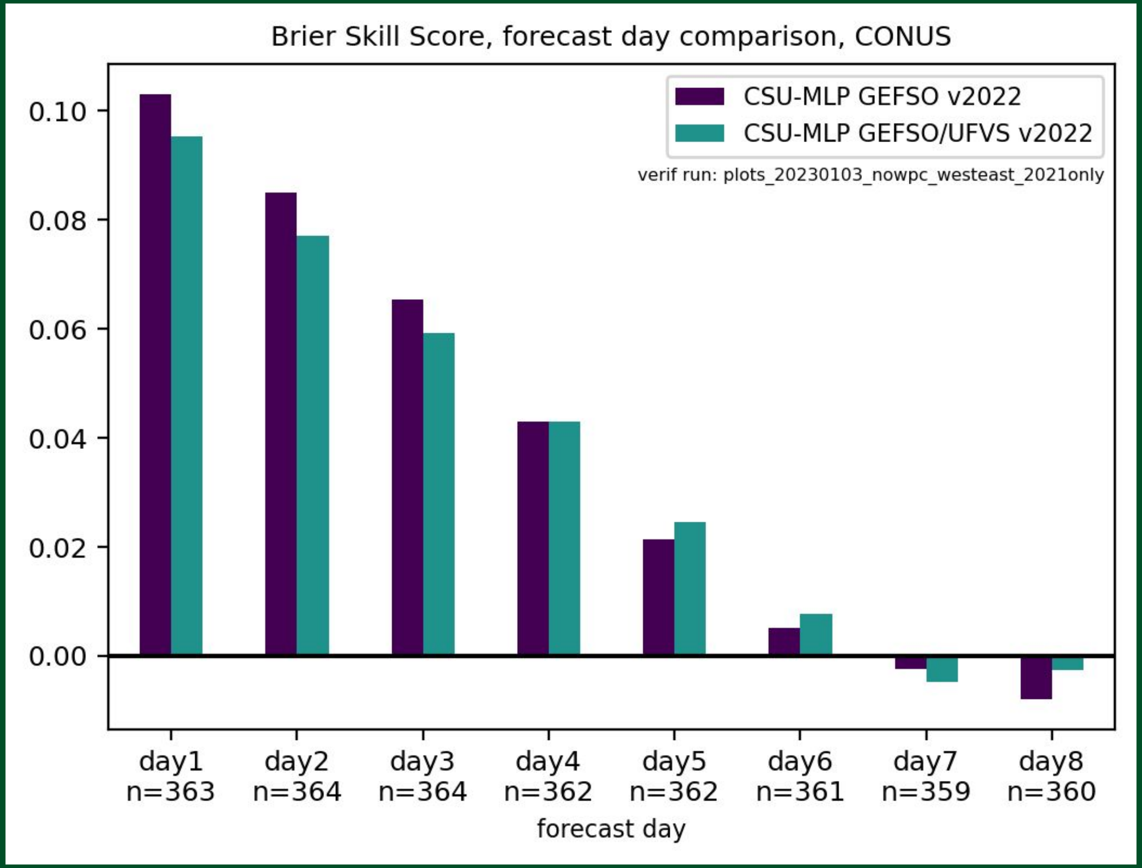
Day 8



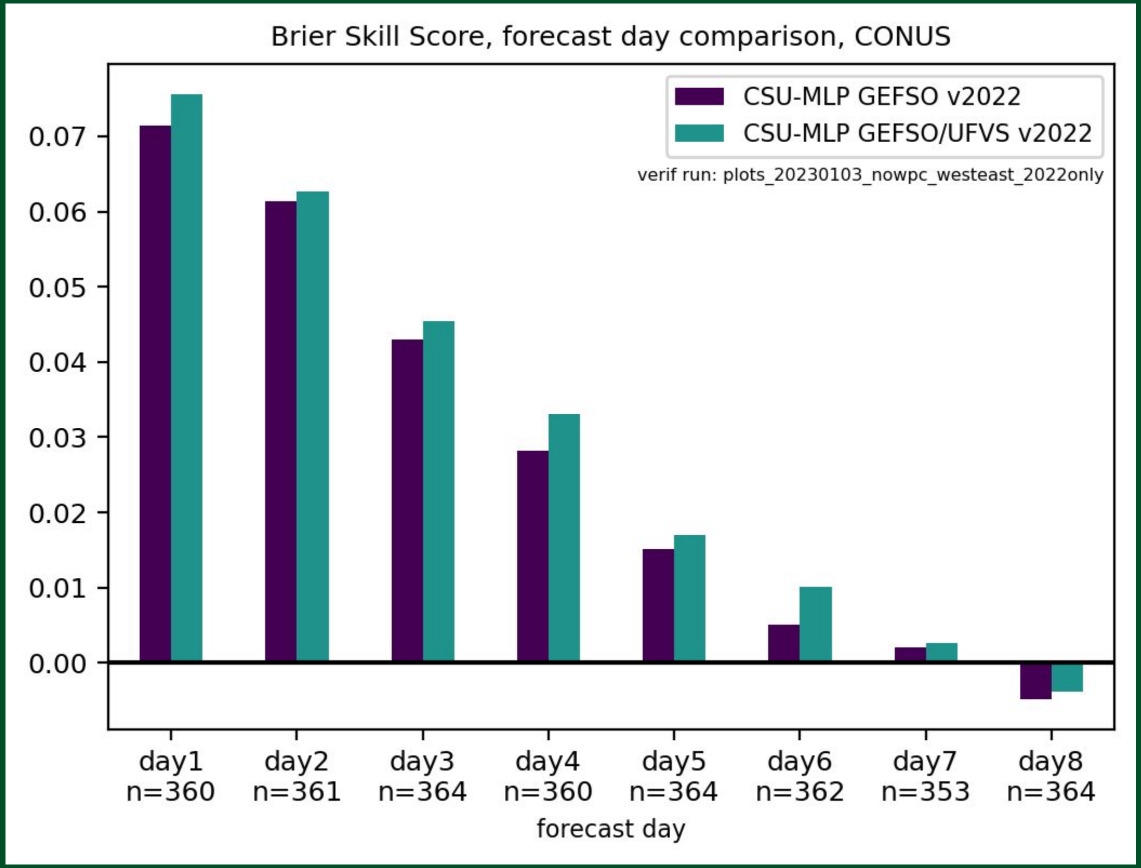
Fixed frequency model

Brier skill score and ROC area, CONUS, by day and year

2021



2022

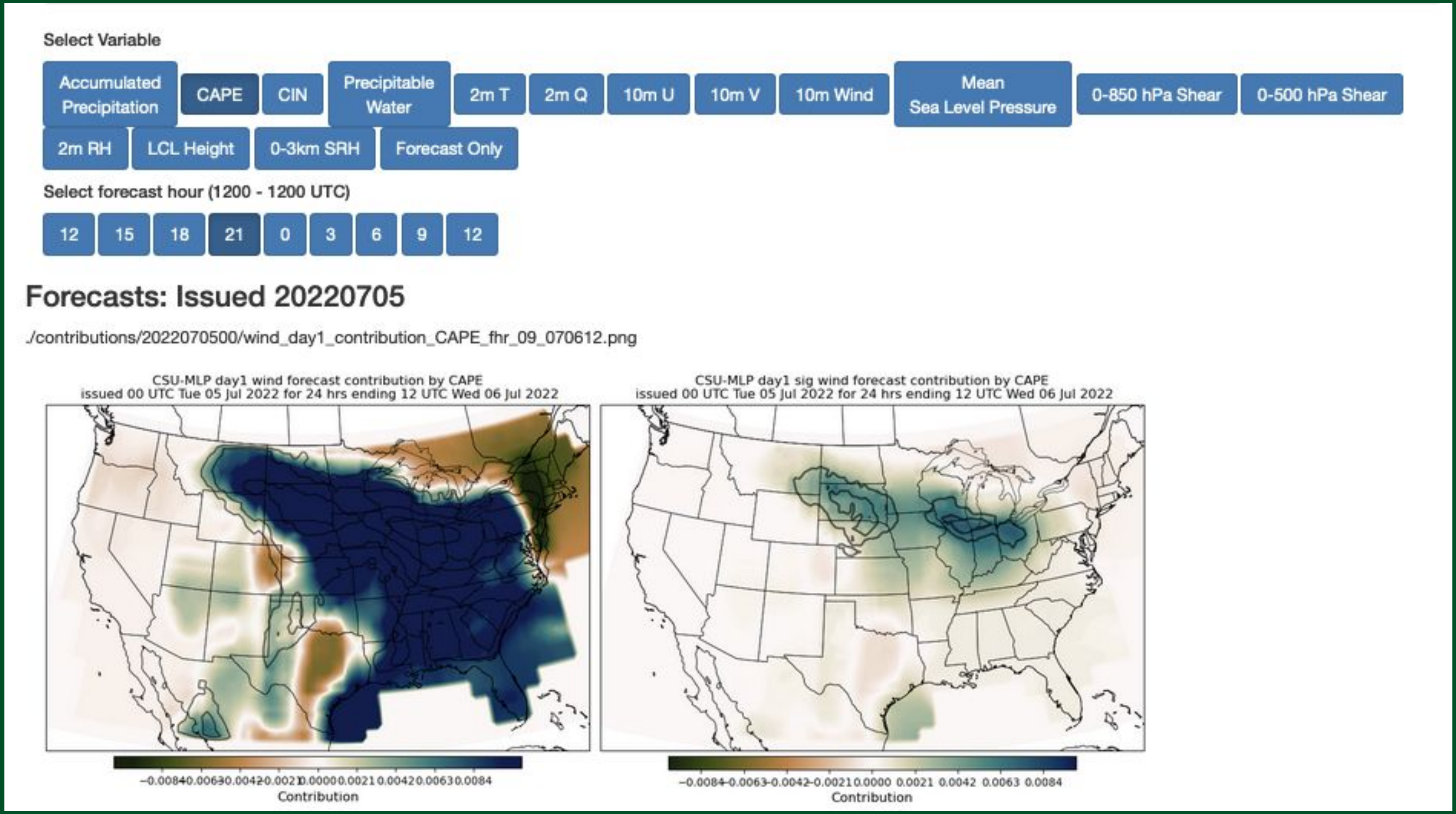


- UFVS models more skillful days 1—4 in 2022 compared to 2021
    - Consistent with impressions from participants in the 2022 HMT FFaIR experiment
    - *More active monsoon?*
- *There may be regimes, seasons, and regions in which one model is more skillful*

The future: partly cloudy?

- ”Fixed frequency” models will be transitioned to WPC operations soon, after recommendation from FFaIR last year
- We anticipate the UFVS-trained models will also be transitioned to operations
- Under current situation, no further support for CSU-MLP beyond December 2023
- So future development is unclear – but we still have plenty of ideas!

Ongoing work: diagnosis of how environmental parameters are influencing the forecast

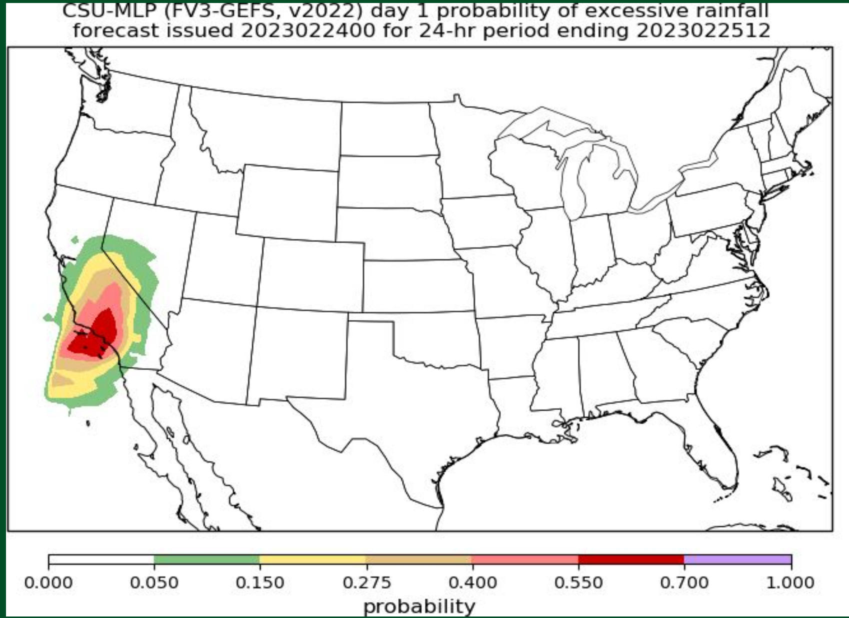


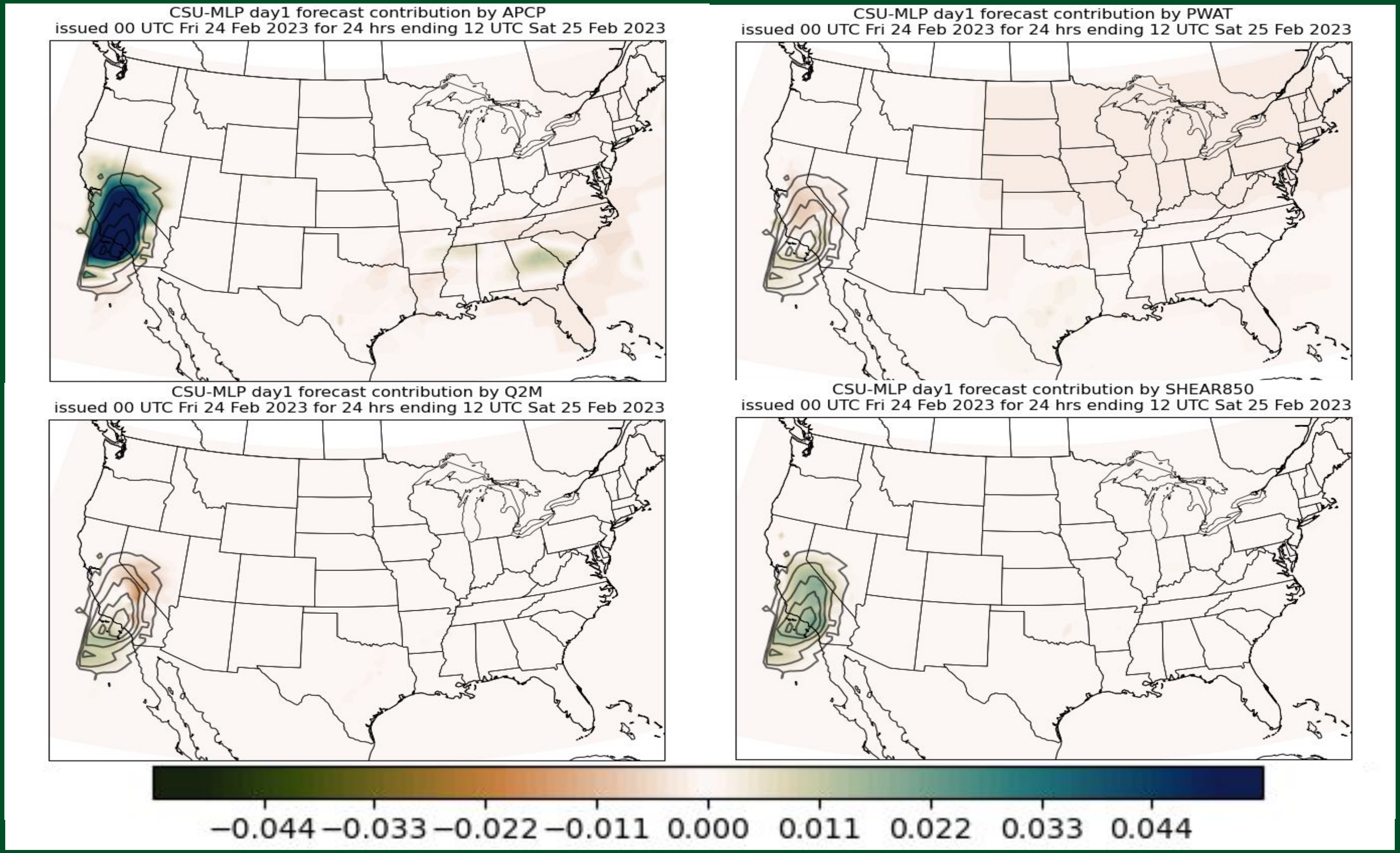
Work by grad student Allie Mazurek and
REU student Hanna McDaniel, using Tree Interpreter package

Ongoing work: diagnosis of how environmental parameters are influencing the forecast

Example: Day-1 forecast contributions for 24 February 2023 at forecast hour 0000 UTC

Can spatially interpret which fields are supplying key information to the forecast probabilities

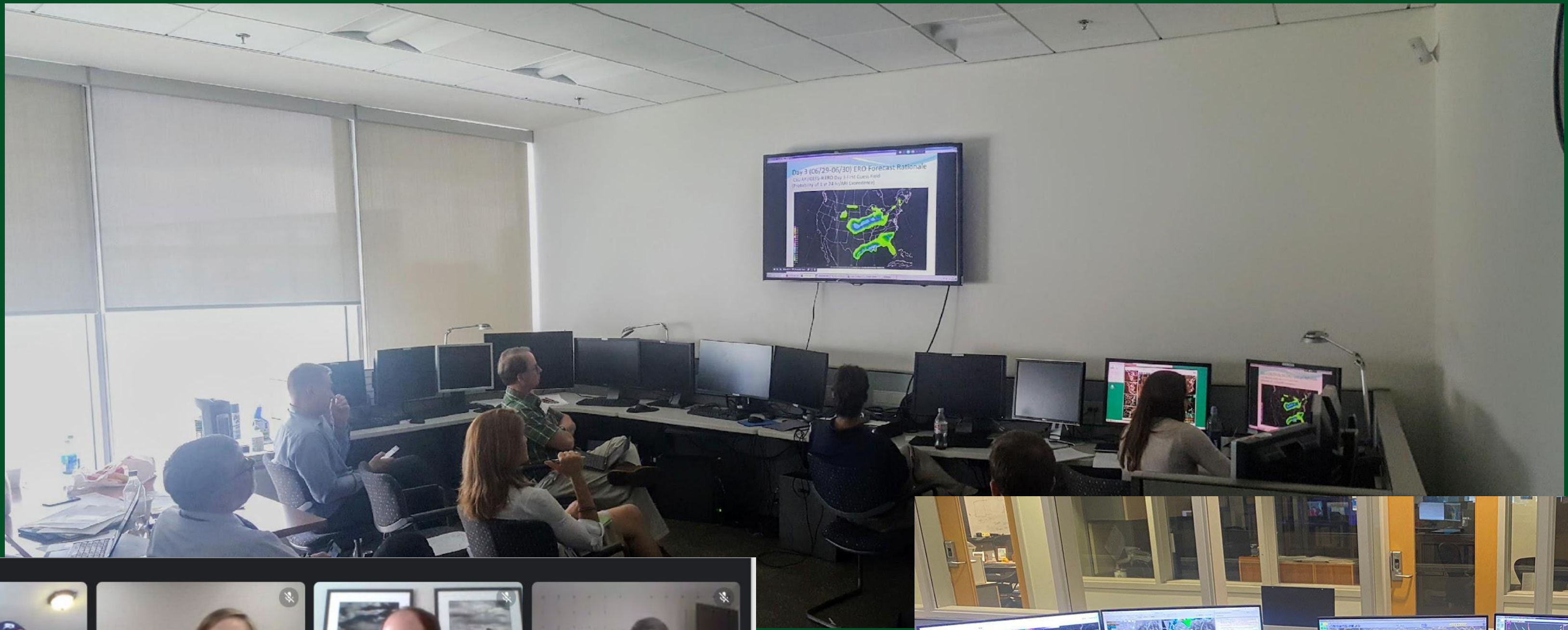


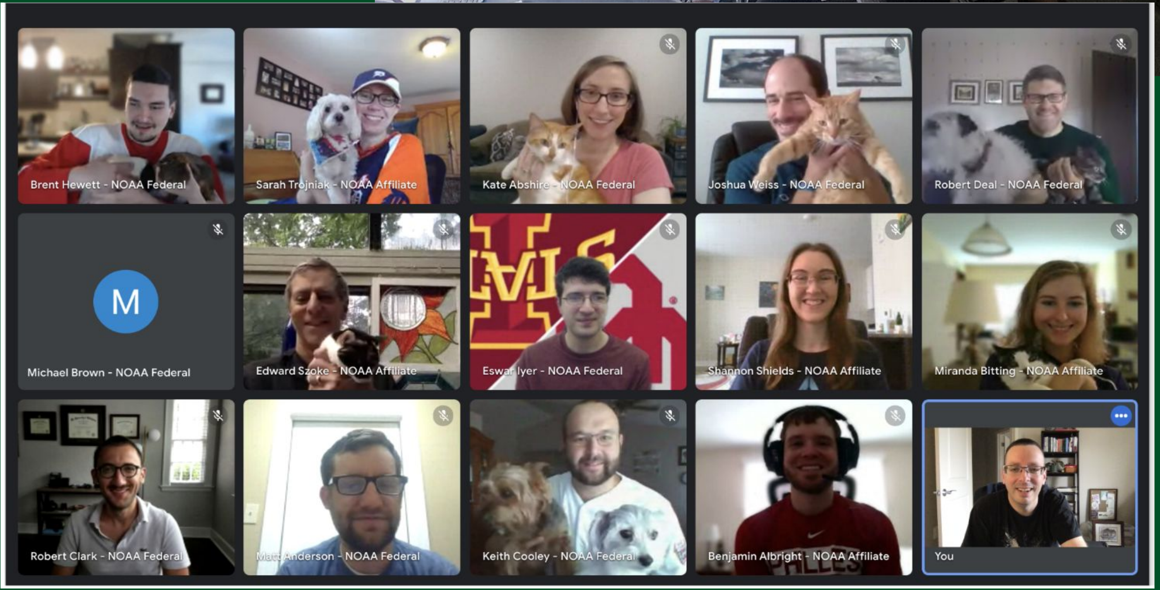


Negative Contribution

Positive Contribution

Collaboration with forecasters has been key to the improvement of these systems!





Summary

- Machine learning techniques can help in post-processing NWP output to yield useful “first guess” guidance for operations
- ML models for excessive rainfall are skillful beyond day 1, but current approaches reach a limit by day 6
- Plenty of opportunities for further advances, both in the forecasts themselves, and how they can be applied: what’s the best way to make them useful and trustworthy for forecasters?

Thank you!

russ.schumacher@colostate.edu
aaron.hill@colostate.edu

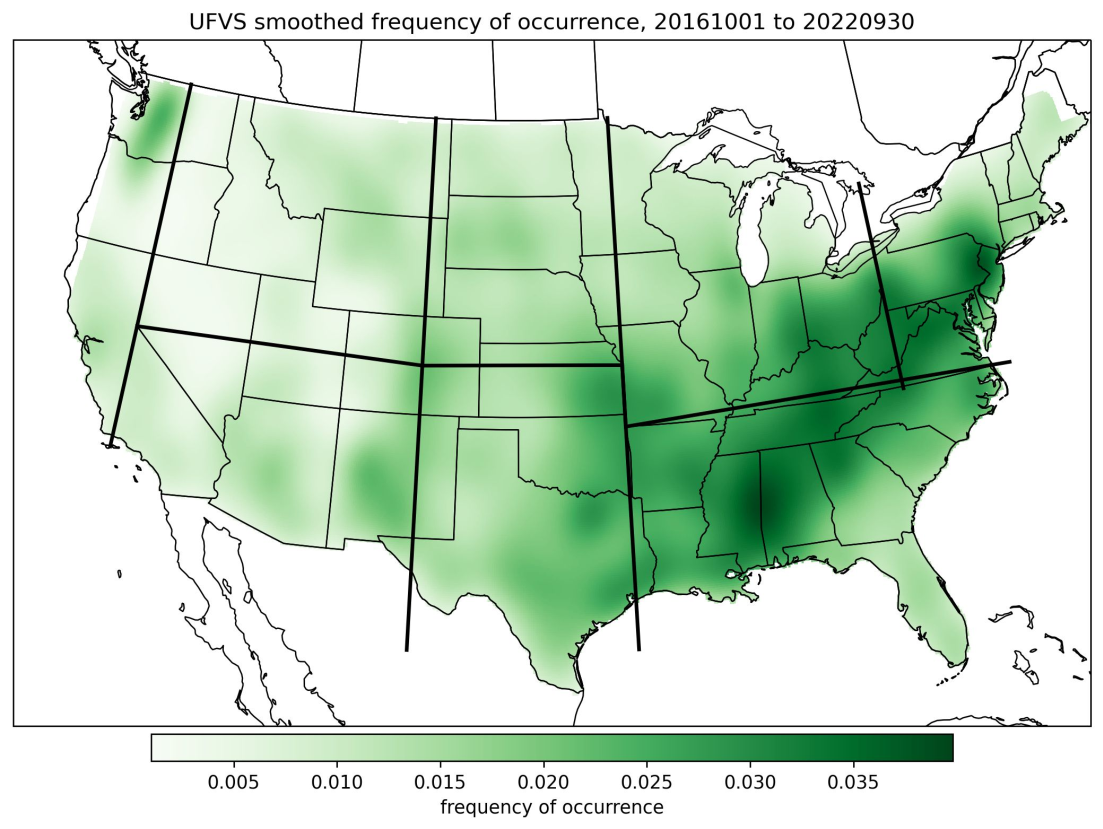
Real-time forecast graphics:
http://schumacher.atmos.colostate.edu/hilla/csu_mlp/

Contact us if interested in gridded output!

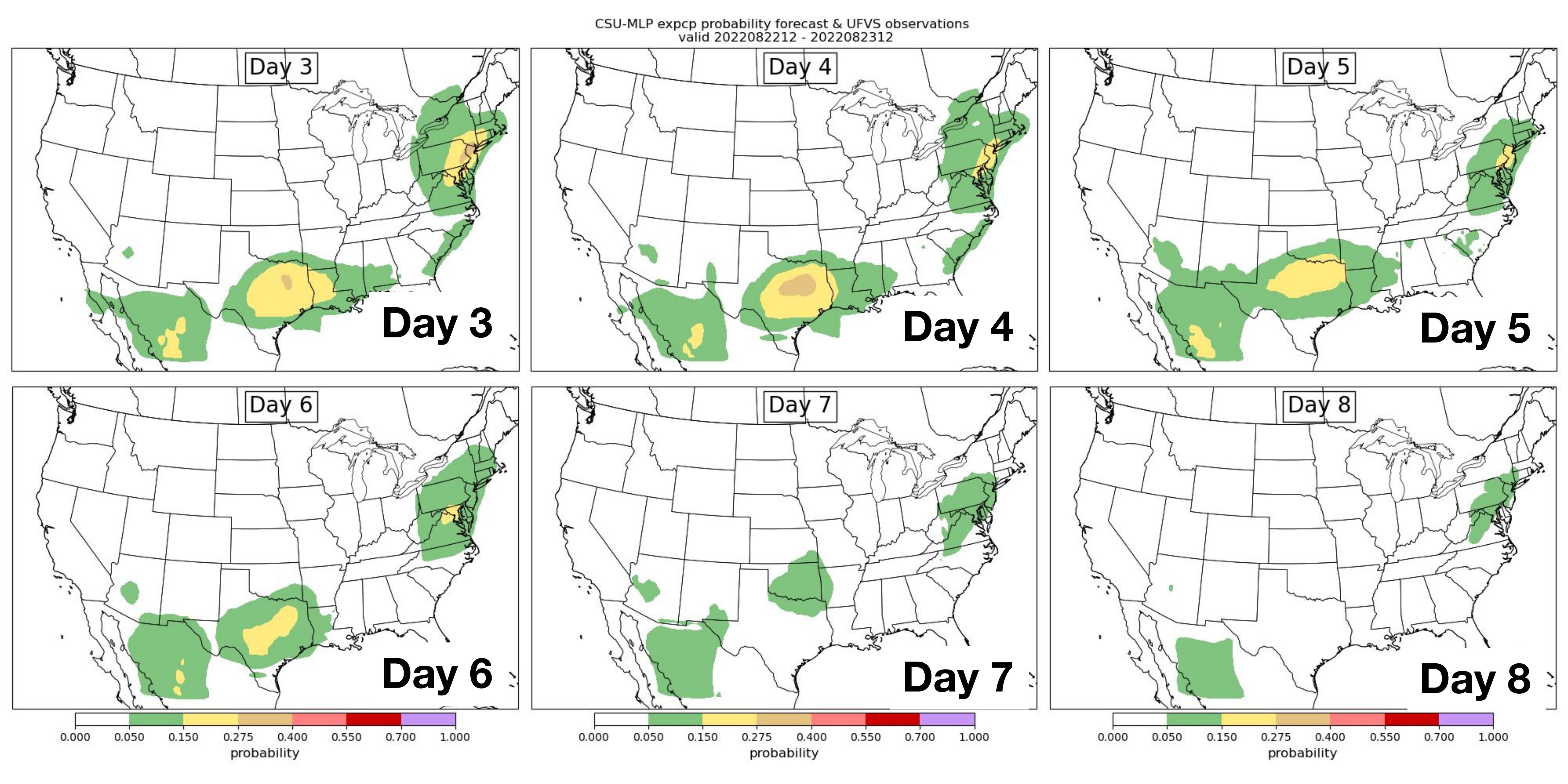
Backup slides

Verification of CSU-MLP excessive rainfall forecasts

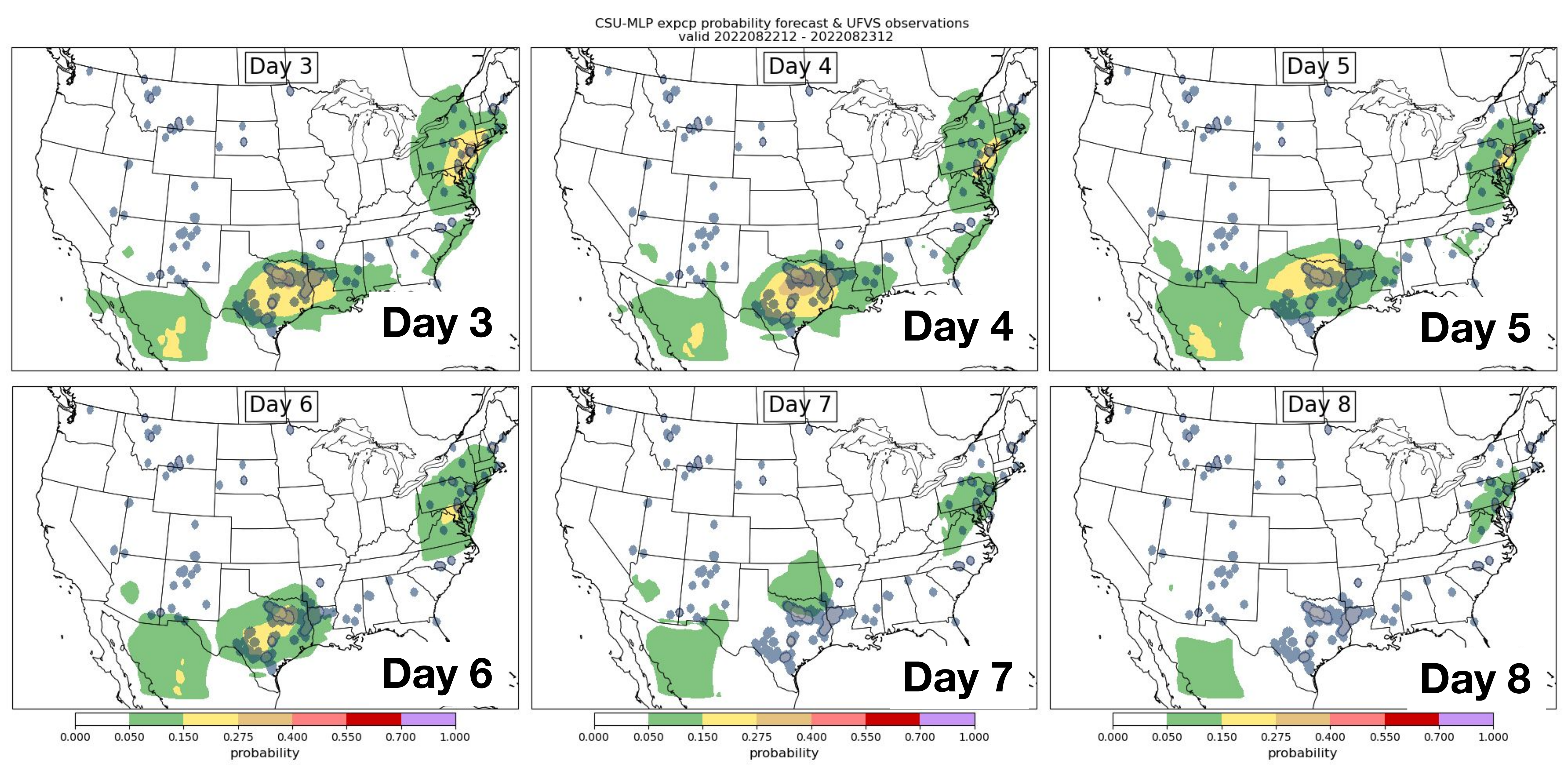
- Observation dataset is WPC's UFVS, includes flash flood guidance exceedances, 5-yr ARI exceedances, flash flood LSRs, USGS and MPING flood reports
- Retrospective forecasts run back to 2 October 2020 (when GEFSv12 became operational) through 31 May 2022.
- Verification is done CONUS-wide and for the western/eastern US
- Comparison is to 09Z WPC operational EROs
- **All evaluation uses the new definitions for ERO categories: 5, 15, 40, 70%**



Good forecast example: 22-23 August 2022 (Texas & northeast flooding)

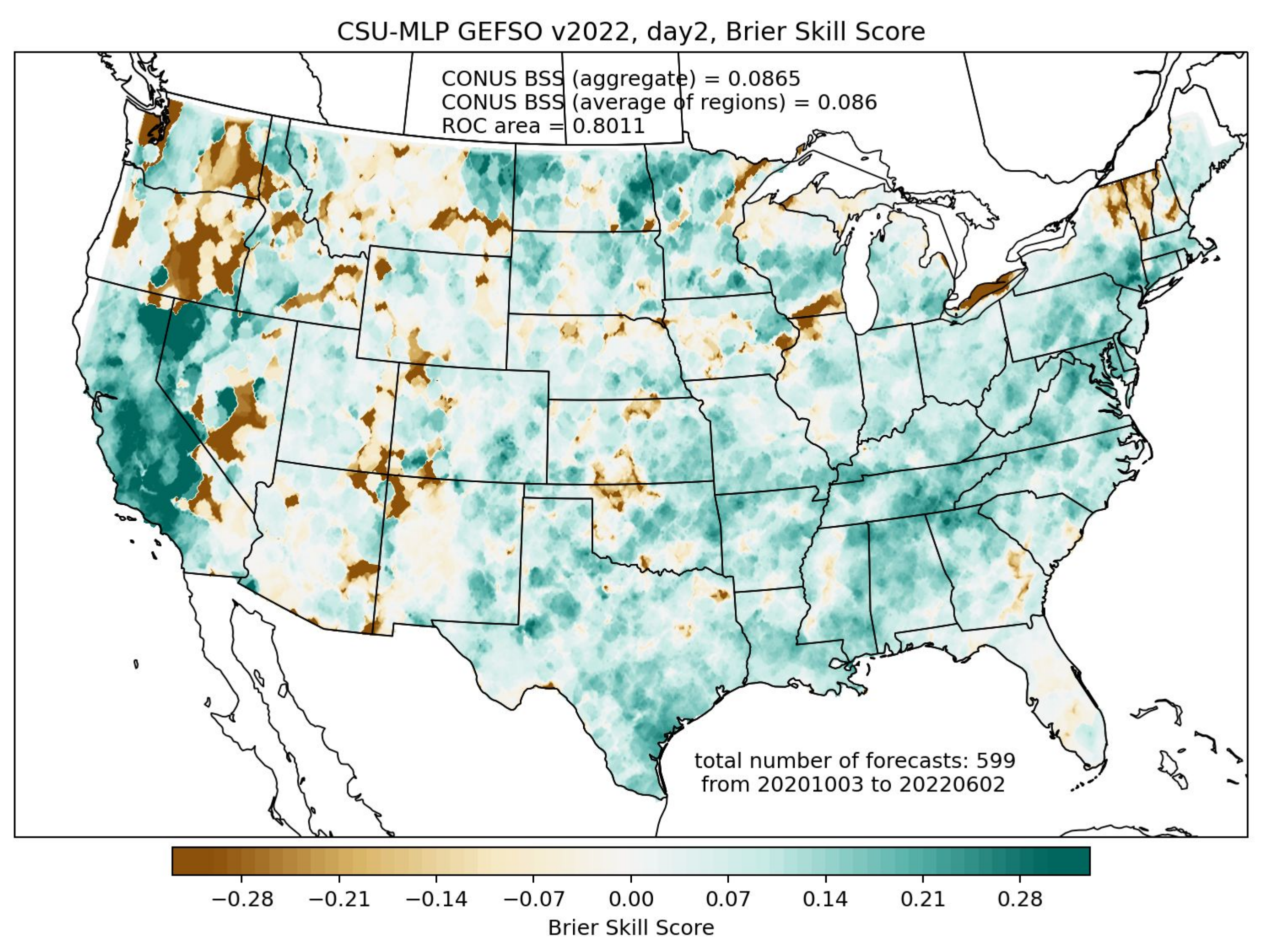


Good forecast example: 22-23 August 2022 (Texas & northeast flooding)



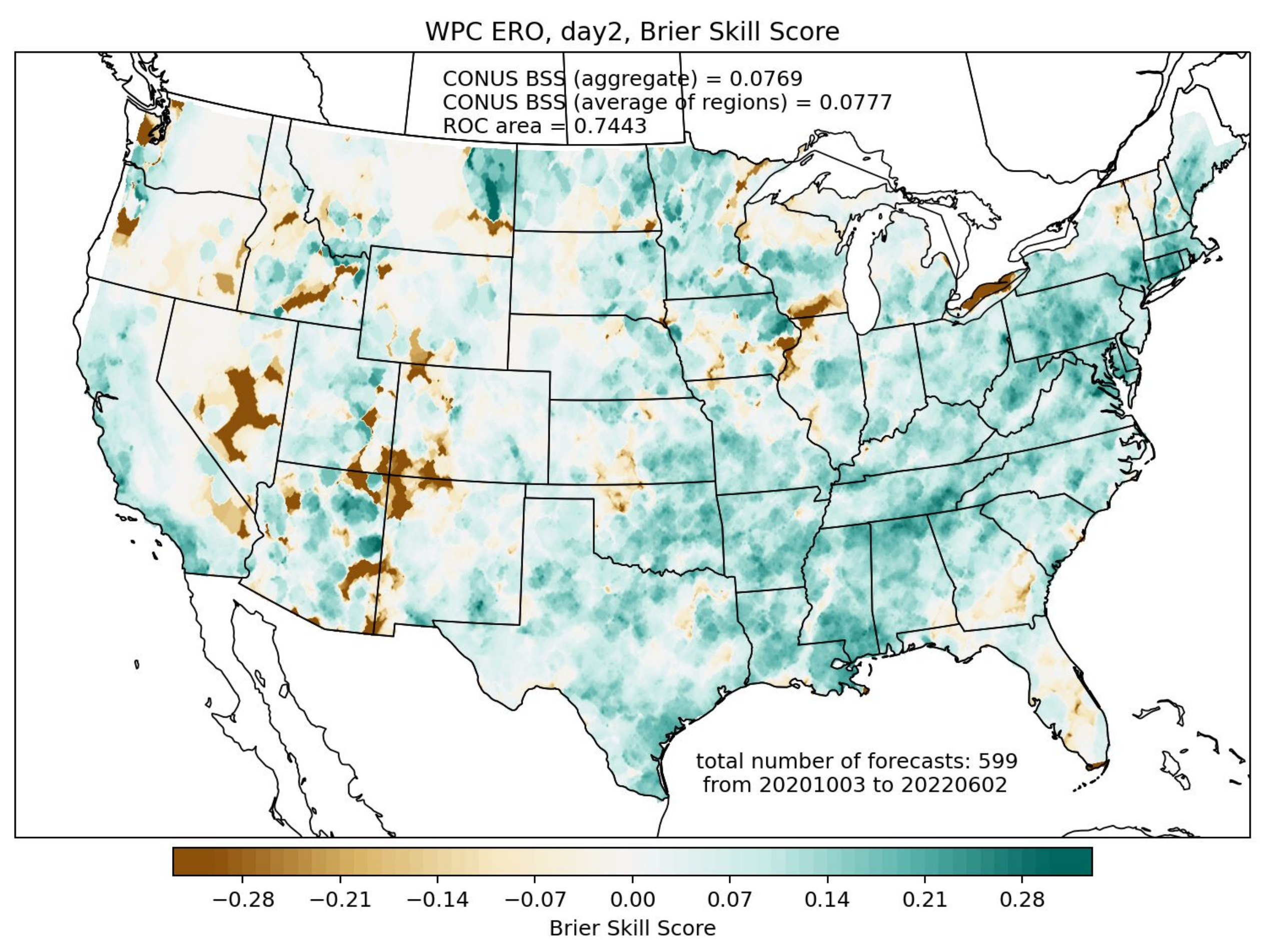
Brier skill score comparison: day 2

CSU-MLP (v2022)



CONUS Brier Skill Score: 0.0865

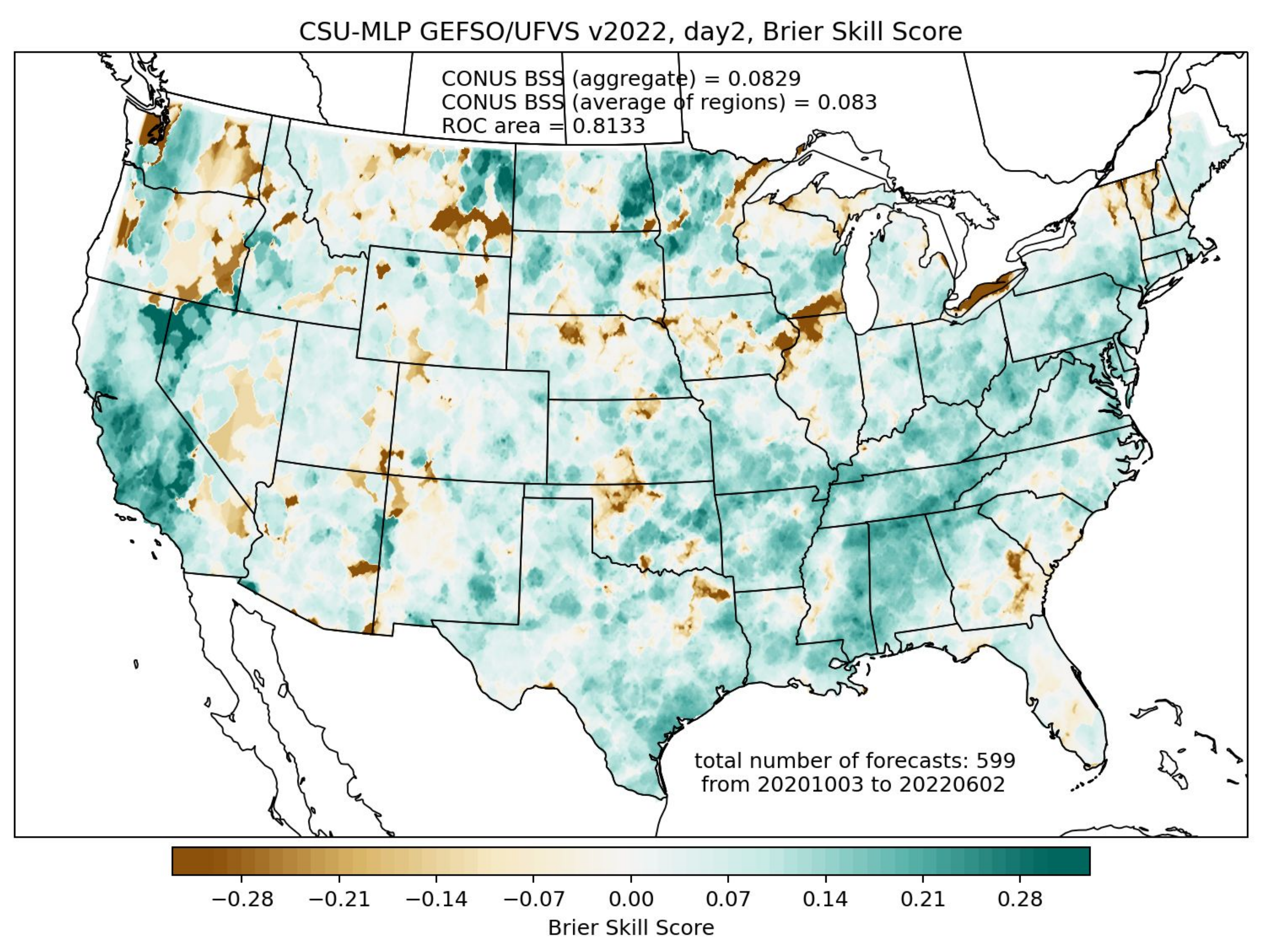
WPC ERO



CONUS Brier Skill Score: 0.0769

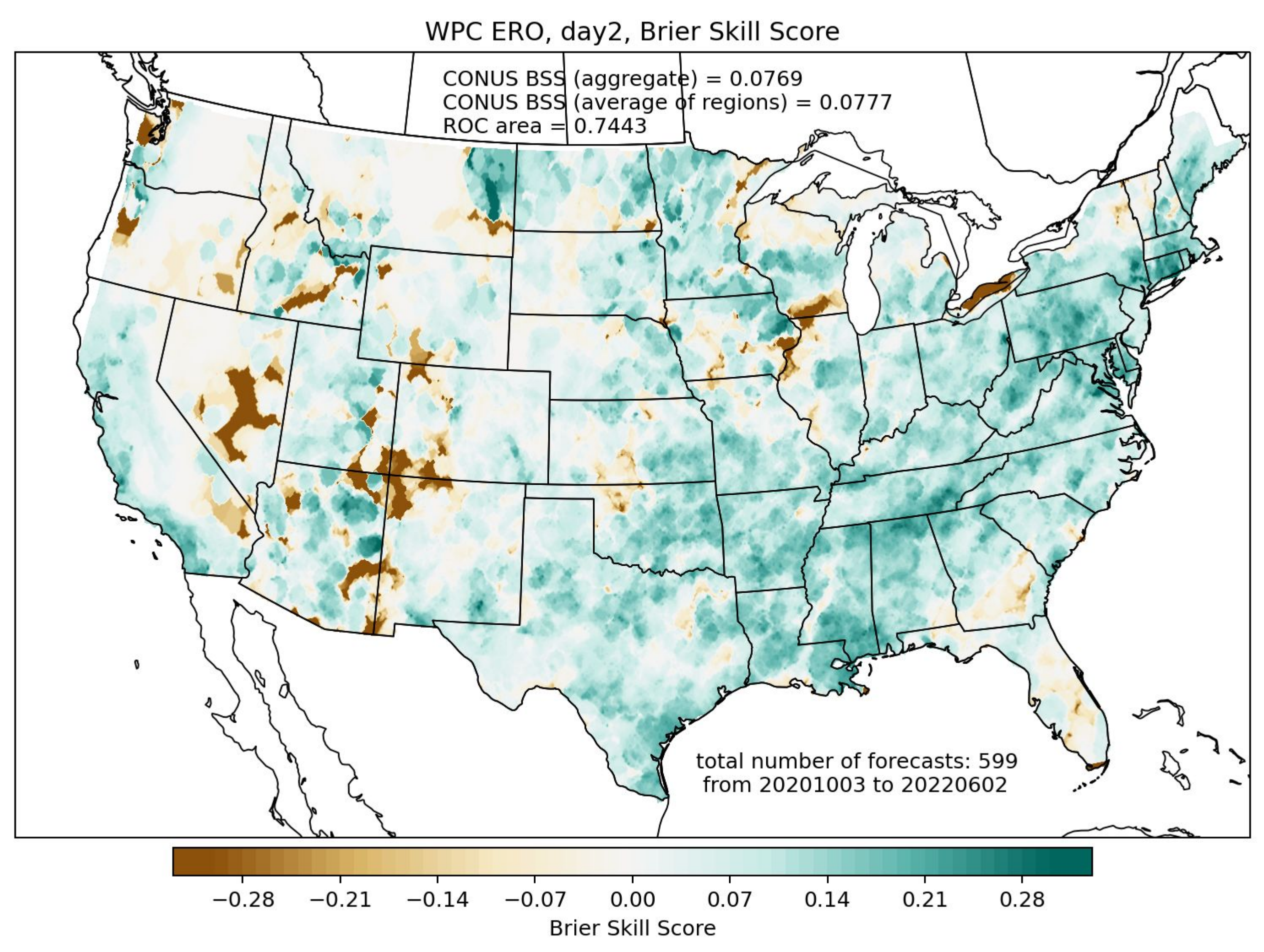
Brier skill score comparison: day 2

CSU-MLP (v2022, UFVS-trained)



CONUS Brier Skill Score: 0.0829

WPC ERO



CONUS Brier Skill Score: 0.0769

Calculation of forecast skill

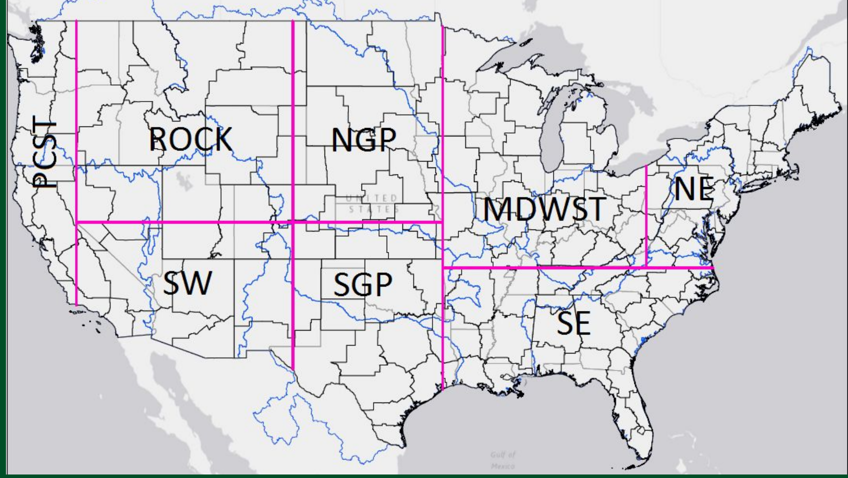
- The Brier Skill Score is used to assess forecast skill:

$$\mathrm{BSS} = 1.0 - \frac{\mathrm{BS}}{\mathrm{BS}_{\mathrm{clim}}} = 1.0 - \frac{\sum_c (p_c - o_c)^2}{\sum_c (p_{\mathrm{clim}_c} - o_c)^2},$$

- Here, we use a smoothed, temporally varying climatology as the reference forecast. So skill on a given day can come from:
  - Correctly predicting high probabilities when/where an event occurs
  - Correctly predicting low probabilities when climatological frequency is high
- Likewise, comparing CSU-MLP to SPC forecast skill, large differences arise when:
  - One forecast is very skillful (“nails it”) and the other is not
  - One forecast has near-zero skill, and another has negative skill that is large in magnitude (“busts”)
  - One forecast nails it and the other busts (this is quite rare)

Variable Importances

- Model QPF (APCP), CAPE, and PWAT most predictive of UFVS-like events in most CONUS regions
- In regions where extreme precipitation driven by large-scale processes, APCP identified as even more predictive
- In highly convectively active regions (e.g. NGP, MDWST), PWAT identified more predictive than APCP (or Q2M as in exceedance models)



UFVS models