Advancing probabilistic prediction of high-impact weather using ensemble reforecasts and machine learning

Russ S. Schumacher and Aaron J. Hill

Department of Atmospheric Science, Colorado State University

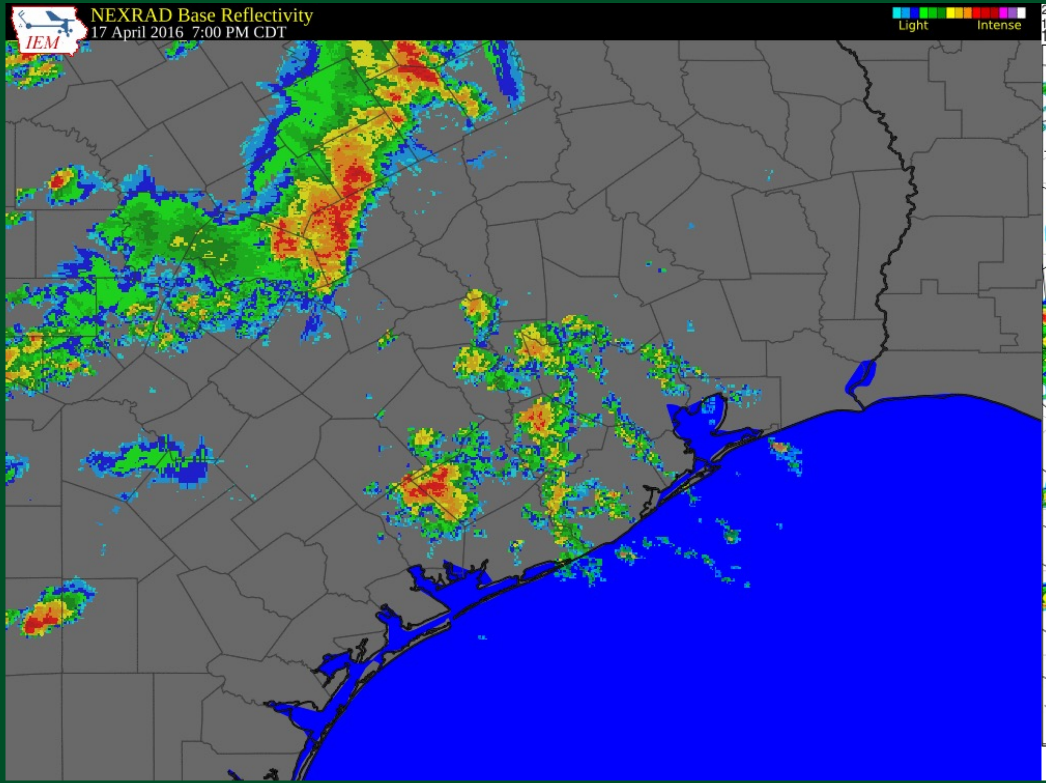
Along with: Greg Herman, Eric James, Jacob Escobedo, Mark Klein, Jim Nelson, Mike Erickson, Sarah Trojniak

ATMOSPHERIC SCIENCE
COLORADO STATE UNIVERSITY

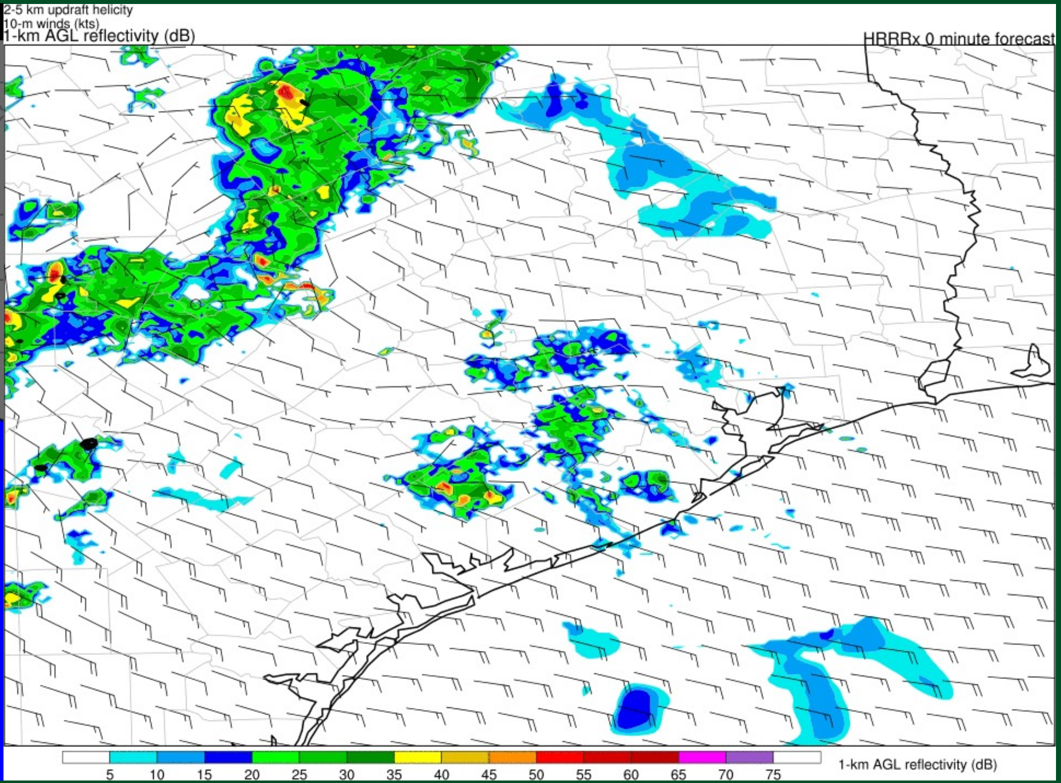
Research supported by NOAA JTTI grants NA16OAR4590238, NA18OAR4590378 and NA20OAR4590350

UFS Webinar Series
June 2021

Deterministic forecasts of convective storms are nearly impossible...this is one of the best forecasts of an extreme event I’ve seen, yet it still has notable errors in details



Radar mosaic

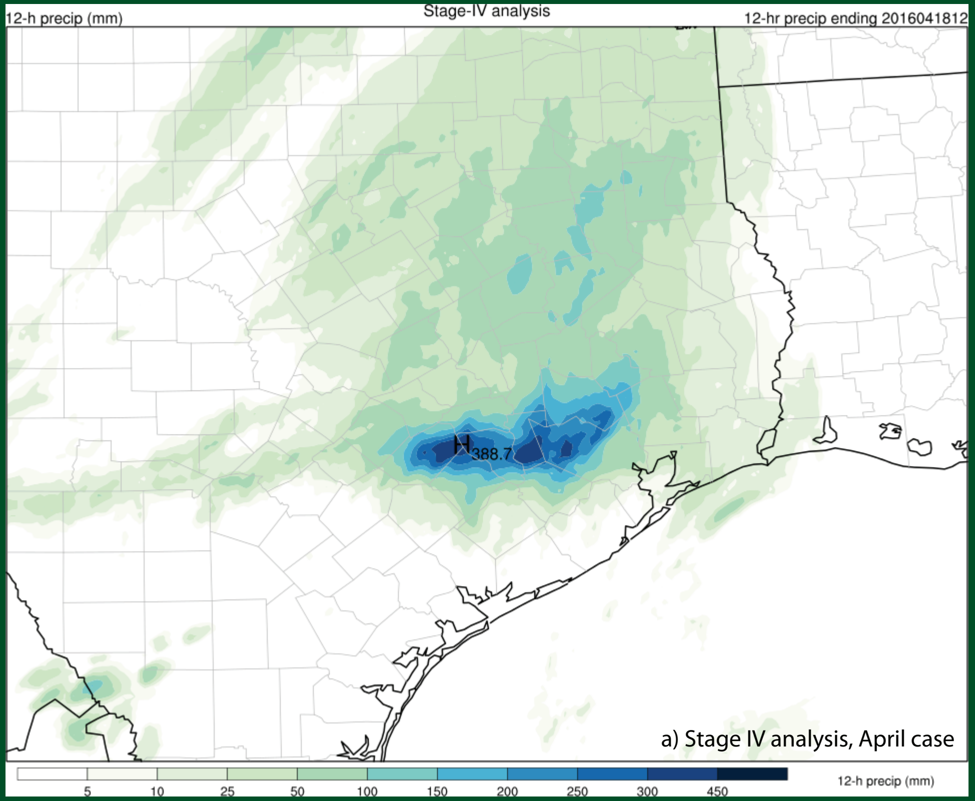


00Z HRRRx model forecast
(1-km AGL reflectivity and updraft helicity)

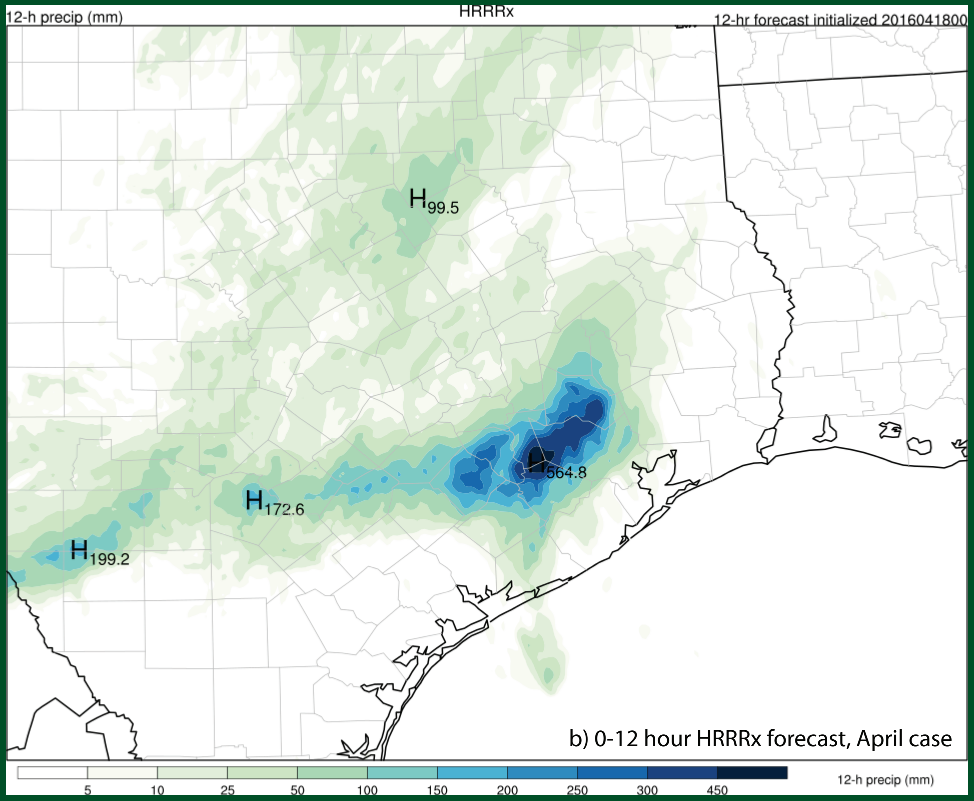
(see also Nielsen and Schumacher 2020, *Monthly Weather Review*)

Houston, Texas, 18 April 2016

But deterministic forecasts of convective storms are nearly impossible...this is one of the best forecasts of an extreme event I've seen, yet it still has notable errors in details



Stage-IV precipitation analysis
12-hour precip



00Z HRRRx model forecast
12-hour precip

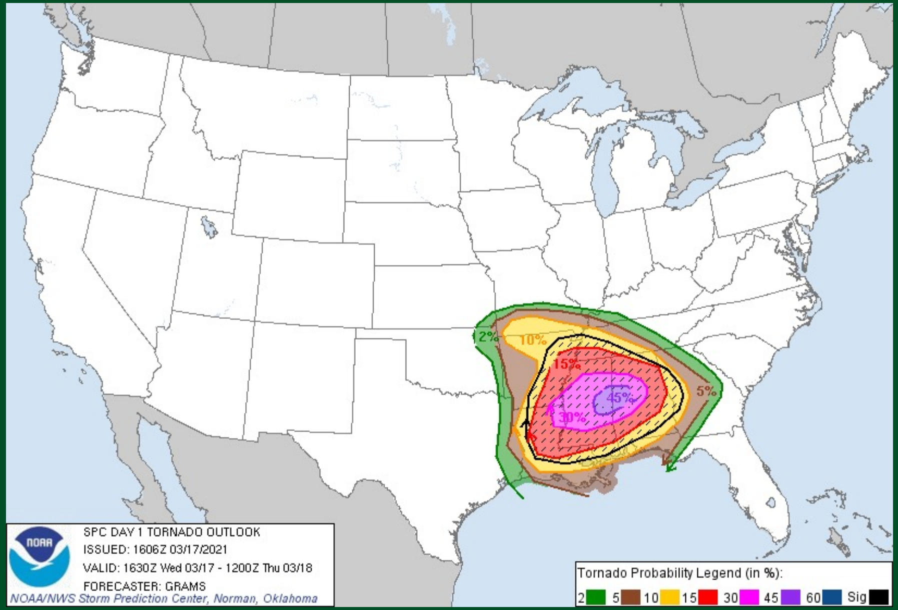
(see also Nielsen and Schumacher 2020, *Monthly Weather Review*)

Houston, Texas, 18 April 2016

We need to think and forecast probabilistically for convective storms and precipitation!

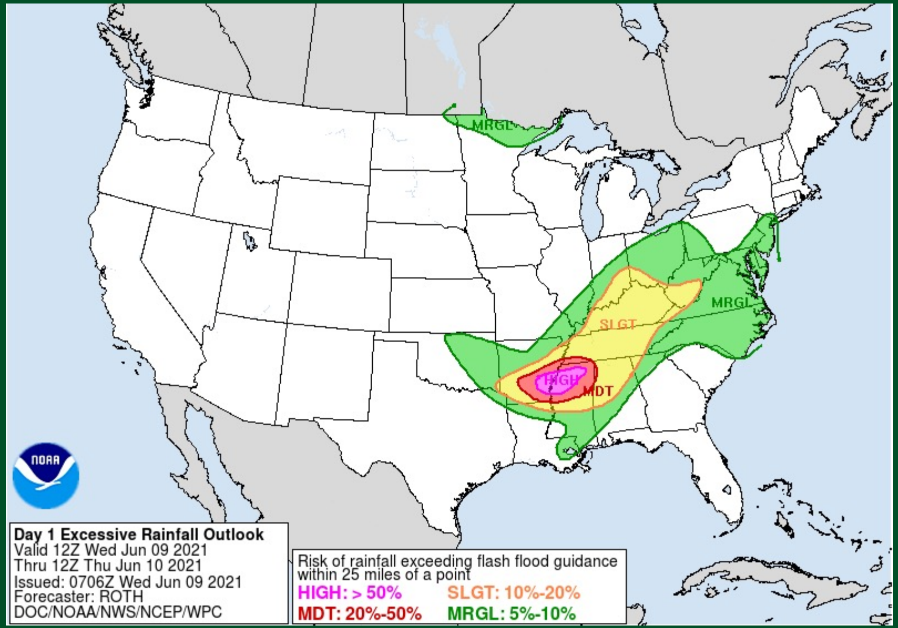
- Unfortunately, this can be difficult for a number of reasons:
  - Probabilistic information can be challenging to understand, even for experts!
  - Ensembles are expensive to run (at least good ones are)
  - Ensembles can also be hard to interpret
  - And our ensembles aren’t yet good enough to rely on – at least not at convection-allowing scales (But a lot of progress is being made!)

- People and forecasters are accustomed to “probability of precipitation” and probabilistic outlooks, even if they don’t know their formal definitions



SPC DAY 1 TORNADO OUTLOOK
ISSUED: 1606Z 03/17/2021
VALID: 1630Z Wed 03/17 - 1200Z Thu 03/18
FORECASTER: GRAMS
NOAA/NWS Storm Prediction Center, Norman, Oklahoma

Tornado Probability Legend (in %):
2 🟩 5 🟫 10 🟨 15 🟥 30 🟪 45 🟪 60 🟦 Sig ⬛



Day 1 Excessive Rainfall Outlook
Valid 12Z Wed Jun 09 2021
Thru 12Z Thu Jun 10 2021
Issued: 0706Z Wed Jun 09 2021
Forecaster: ROTH
DOC/NOAA/NWS/NCEP/WPC

Risk of rainfall exceeding flash flood guidance within 25 miles of a point
HIGH: > 50% SLGT: 10%-20%
MDT: 20%-50% MRGL: 5%-10%

Why Machine Learning (ML)?

Do it better

- e.g. Models do not specifically forecast severe hazards, use ML to make explicit forecasts

Do it faster

- e.g. Radiation code in models is very slow (but we know the right answer) - use ML methods to speed things up

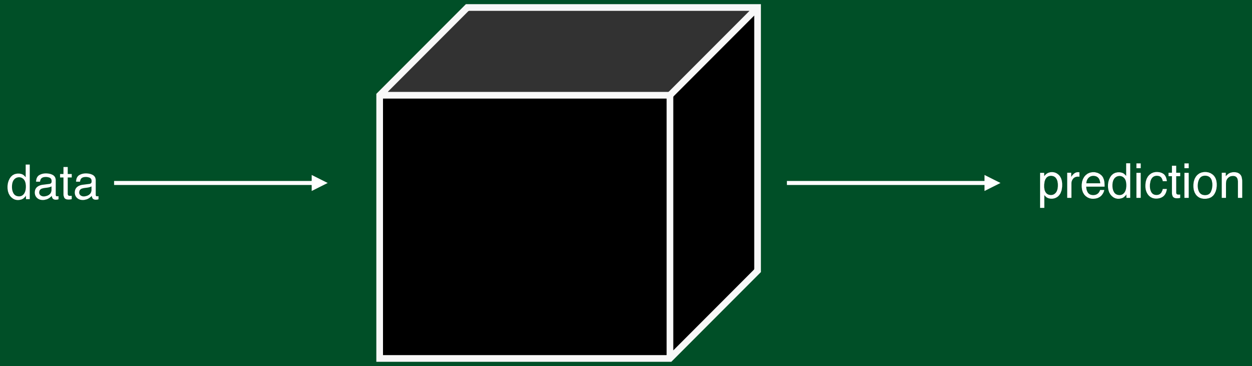
Do something new

- e.g. Go looking for non-linear relationships you didn’t know were there

Credit: CSU machine Learning Workshop

The ML “Black Box”

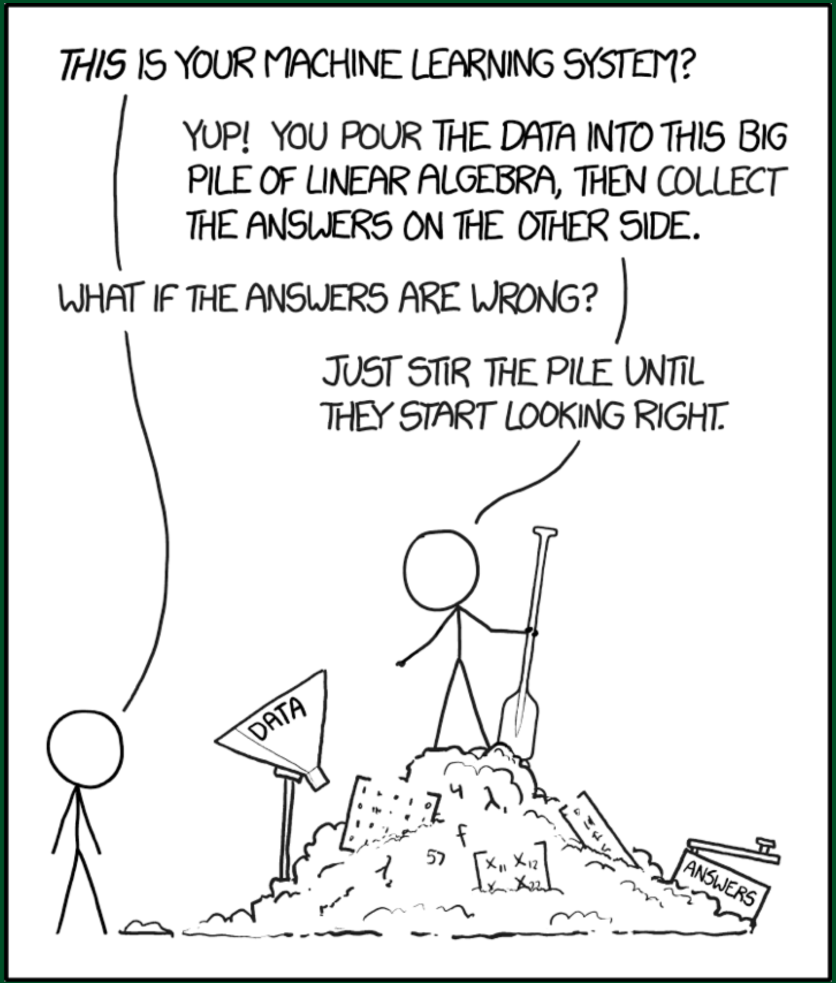
ML is not a futuristic, black-box idea anymore, specifically within geophysical applications…



data

prediction

The box is configurable and ***interpretable***


THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS

xkcd.com

MOS is a ”primitive” example: Predict Y from records of X(s)

Recommended reading: McGovern et al. (2019), “Making the Black Box More Transparent: Understanding the Physical Implications of Machine Learning”, BAMS

(slide from Aaron Hill)

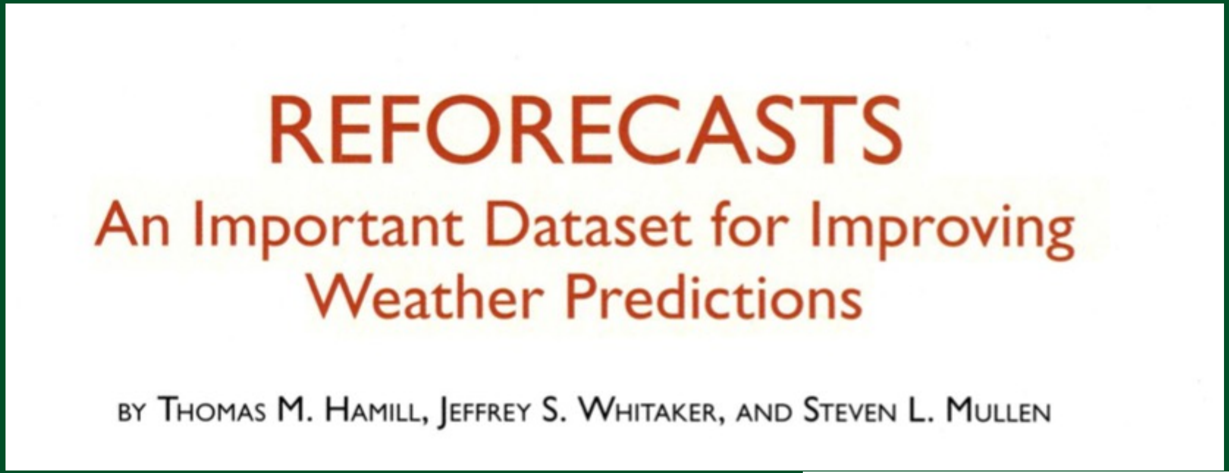
Application to convection hazards

- Our focus to this point has been on excessive rainfall and severe weather (tornado, large hail, damaging winds)

- NWP models don’t explicitly predict any of these
  - For excessive rainfall, we have QPF, but not much about whether it’s “excessive” or likely to cause flash flooding
  - For severe hazards, need to use proxies like updraft helicity (e.g., Sobash et al. 2016)

- But we do now have the observations we need (mostly) and a long record of “reforecasts” that we can use

- Prime opportunity to ***do something better***



REFORECASTS

An Important Dataset for Improving Weather Predictions

BY THOMAS M. HAMILL, JEFFREY S. WHITAKER, AND STEVEN L. MULLEN

- To successfully post-process NWP forecasts, it helps to have a long record of previous forecasts from the NWP model or ensemble

- Enter “reforecasts”: retrospective forecasts with a consistent model version over a long time period (years or decades)

- This allows for identification of systemic errors and biases in the NWP output, which can be accounted for in post-processed forecasts

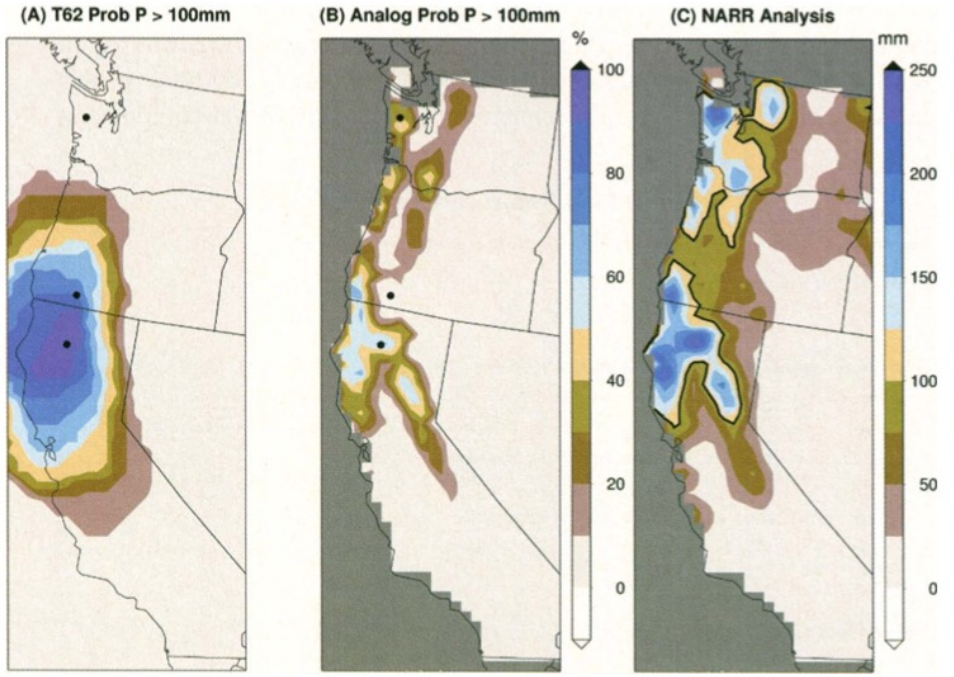


FIG. 3. (a) Raw ensemble-based probability of greater than 100 mm of precipitation accumulated during days 4–6 for a forecast initialized at 0000 UTC 26 Dec 1996 (from 0000 UTC 29 Dec 1996 to 0000 UTC 1 Jan 1997). Dots indicate locations used in Fig. 4. (b) As in (a), but where probabilities have been estimated from relative frequency of historical NARR analogs. (c) Observed precipitation from NARR (mm). The 100-mm threshold is highlighted.

From Hamill et al. (2006)

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

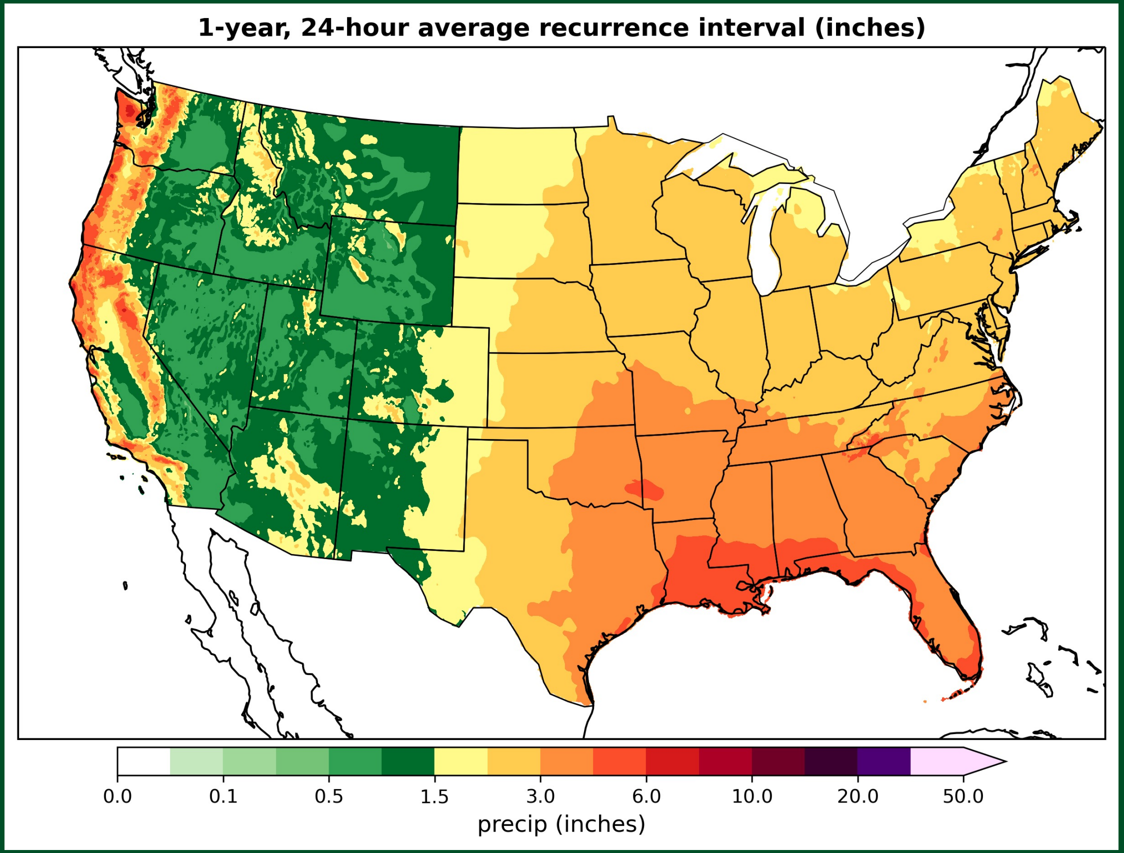
- A primary motivation for this approach is that forecasters need probabilistic information about the rarity of upcoming rainfall. But...
- We have accepted (if flawed) definitions of tornado, severe hail, severe winds – but nothing analogous for excessive rainfall
- Exceeding flash flood guidance?
- 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?

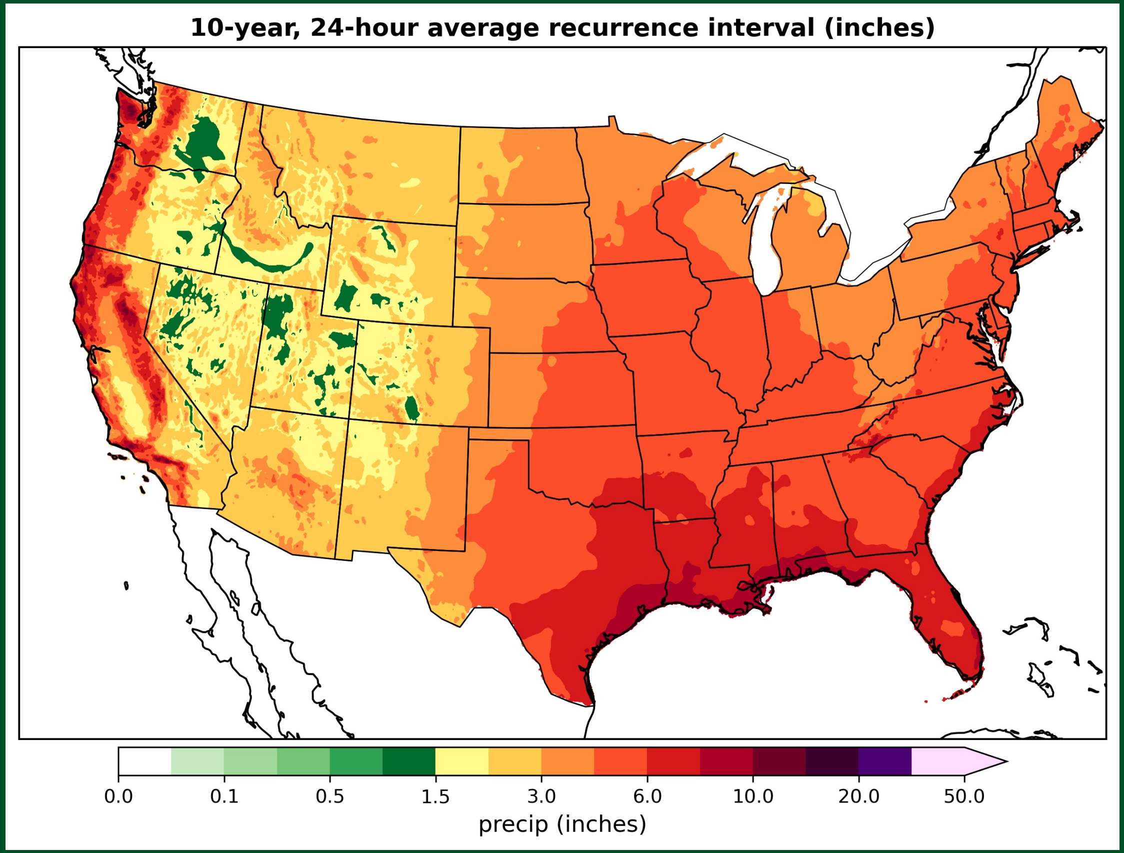
- We have chosen to use a “fixed frequency” framework – or in other words, we use climatological average recurrence intervals to define a heavy or extreme rain event
  - Better corresponds to actual impacts in a given region than a fixed threshold
  - Doesn’t bias the verification statistics toward climatologically wet regions

Average recurrence intervals over the US

1-yea, 24-hour ARI



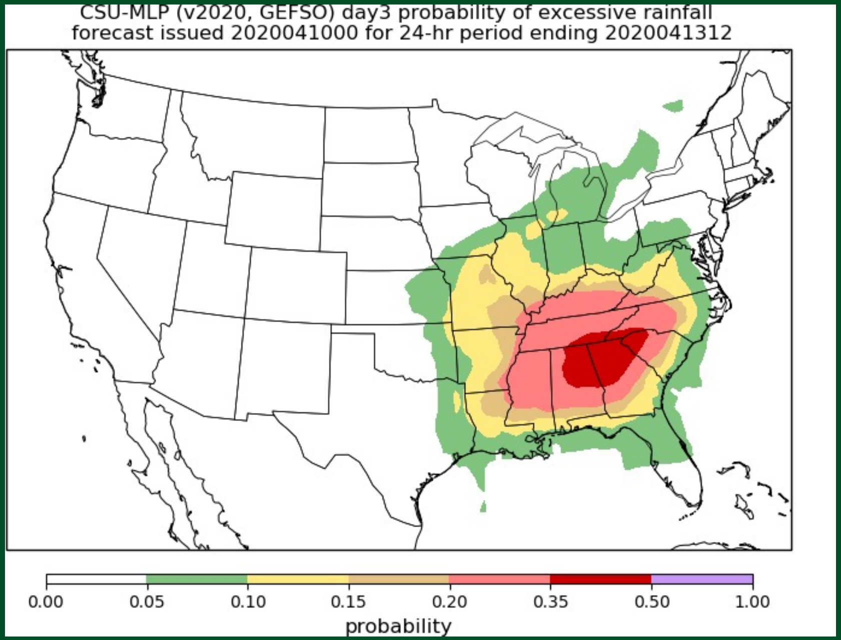
10-year, 24-hour ARI (10% probability of exceedance in any given year)

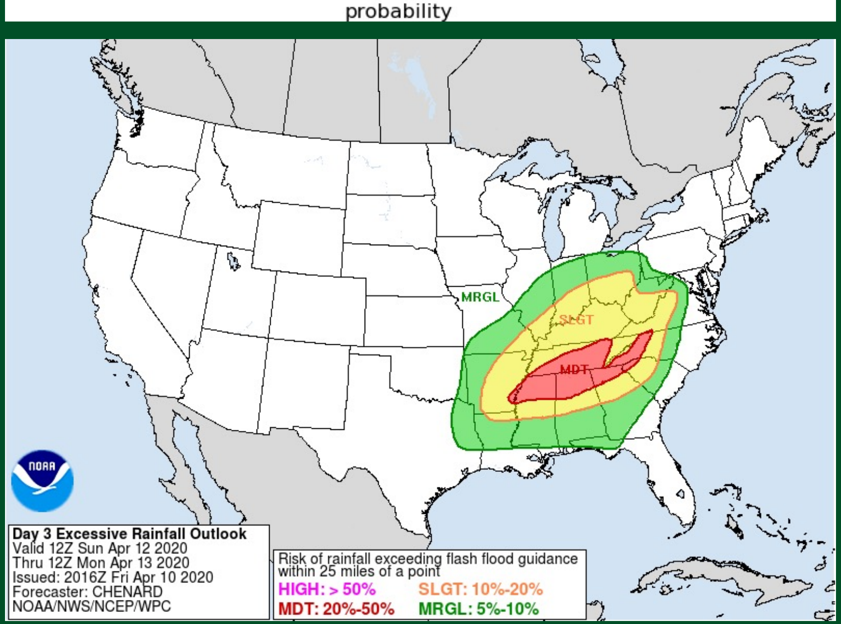


(From NOAA Atlas 14 where available, otherwise Atlas 2)

Our current approach

- 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 are now running operationally at WPC


CSU-MLP (v2020, GEFSO) day3 probability of excessive rainfall
forecast issued 2020041000 for 24-hr period ending 2020041312
0.00
0.05
0.10
0.15
0.20
0.35
0.50
1.00
probability



Day 3 Excessive Rainfall Outlook
Valid 12Z Sun Apr 12 2020
Thru 12Z Mon Apr 13 2020
Issued: 2016Z Fri Apr 10 2020
Forecaster: CHENARD
NOAA/NWS/NCEP/WPC

Risk of rainfall exceeding flash flood guidance within 25 miles of a point
HIGH: > 50% SLGT: 10%-20%
MDT: 20%-50% MRGL: 5%-10%

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

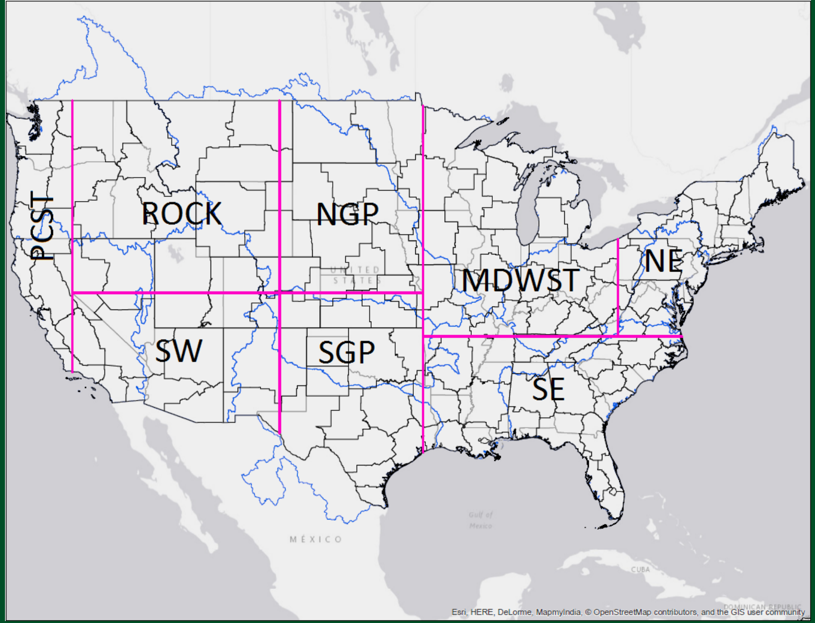
Schumacher et al. (2021, *BAMS,* in press)

Background: Predictor data

- NOAA’s Second-Generation Global Medium-Range Ensemble Reforecast Dataset (GEFS/R; Hamill et al. 2013)
  - 00Z initialization; forecasts out to 384h
  - 11 ensemble members
  - Operational version of GEFS on 2/14/2012
  - T254L42 resolution (~40 km)
  - Data from Dec. 1984-Present
  - Same model configuration throughout
- Use the NCEP Climatology-Calibrated Precipitation Analysis (CCPA) to identify historical exceedances of the various average recurrence intervals for 24-hour rainfall accumulation

The approach

- We break the CONUS into 8 distinct regions which are reasonably hydrometeorologically homogenous
- Use many atmospheric fields as candidate predictors: model QPF; CAPE; CIN; PWAT; MSLP; 2-m T and mixing ratio; 10-m U and V
- We use January 2003 – August 2013 as the training period (almost 11 yrs)


PCST
ROCK
NGP
MDWST
NE
SW
SGP
SE
MÉXICO
Esri, HERE, DeLorme, MapmyIndia, © OpenStreetMap contributors, and the GIS user community

Consider:

All days per year

Forecast values at all 9 3-hour times during meteorological day (1200-1200 UTC)

Forecast information up to 4 grid points (~2°) displaced relative to forecast location

Threshold criteria for return period exceedances and forecast point information

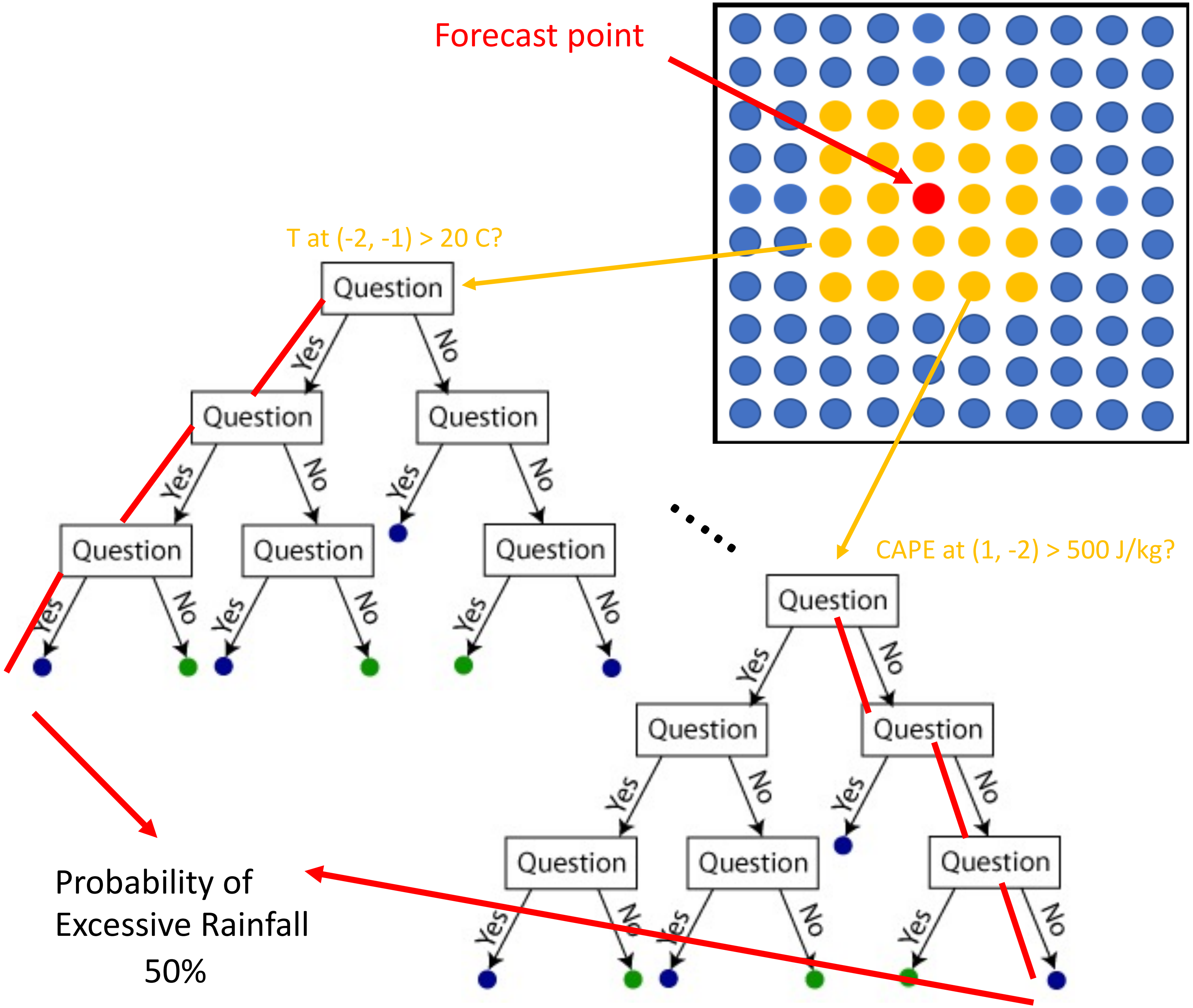
Total: ~16.0M forecasts, 6575 predictors

Two papers in MWR with all the details!

- Herman, G.R. and R.S. Schumacher, 2018: Money Doesn't Grow on Trees, but Forecasts Do: Forecasting Extreme Precipitation with Random Forests. *Mon. Wea. Rev.*, **146**, 1571–1600, https://doi.org/10.1175/MWR-D-17-0250.1
- Herman, G.R. and R.S. Schumacher, 2018: “Dendrology” in Numerical Weather Prediction: What Random Forests and Logistic Regression Tell Us about Forecasting Extreme Precipitation. *Mon. Wea. Rev.*, **146**, 1785–1812, https://doi.org/10.1175/MWR-D-17-0307.1

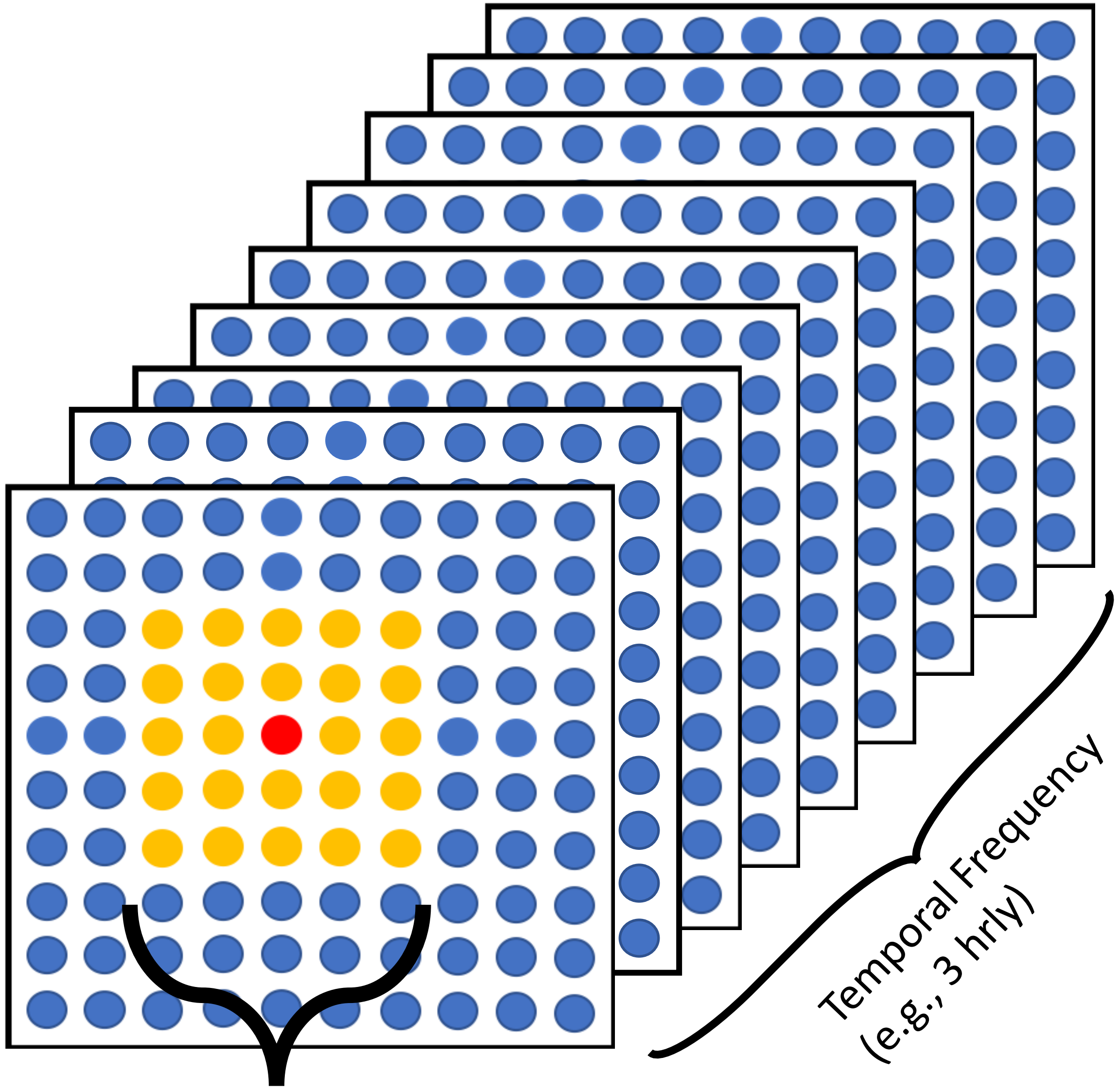
Random Forests

- 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



Probability of Excessive Rainfall 50%

(slide from Aaron Hill)

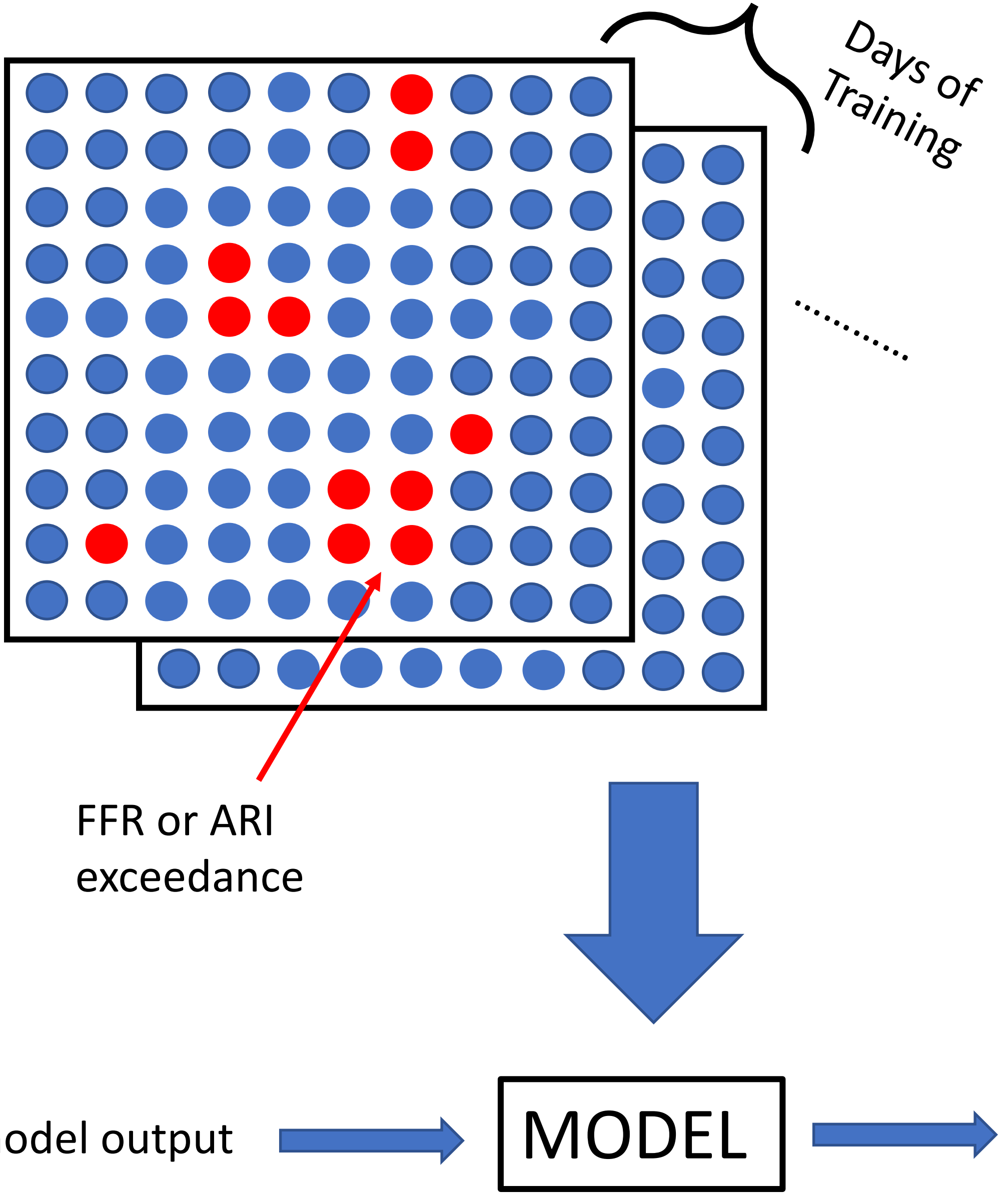


Temporal Frequency (e.g., 3 hrly)

Spatial frequency

Predictor
APCP
CAPE
CIN
MSLP
PWAT
Q2M
T2M
UPHL
U10
U6000
V10
V6000
W3000
Z500

Variable Type



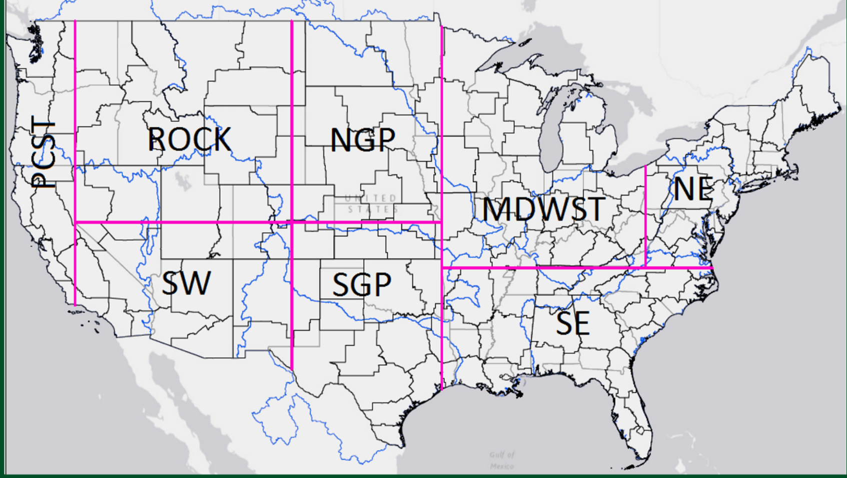
Days of Training

FFR or ARI exceedance

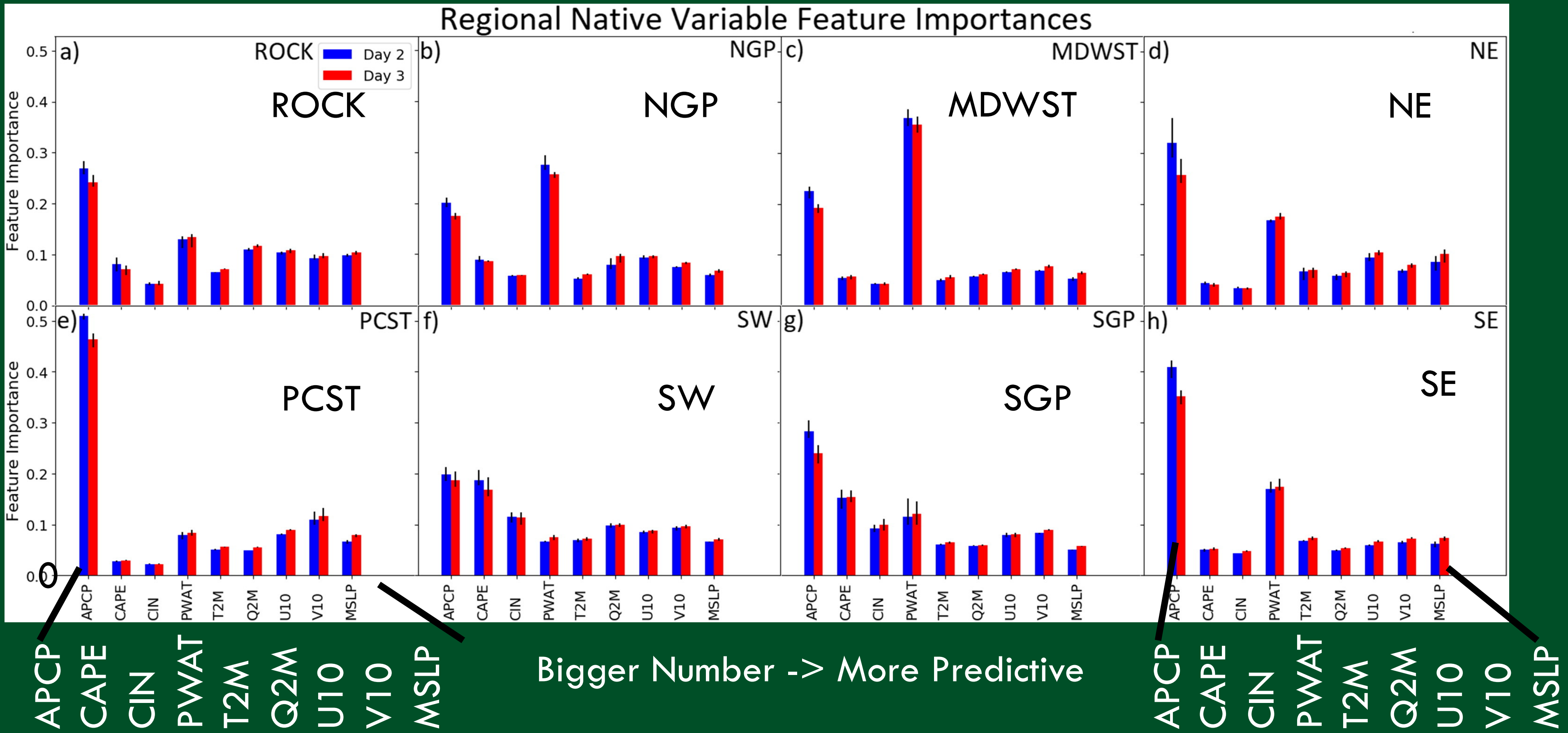
Real-time model output

MODEL

Variable Importances

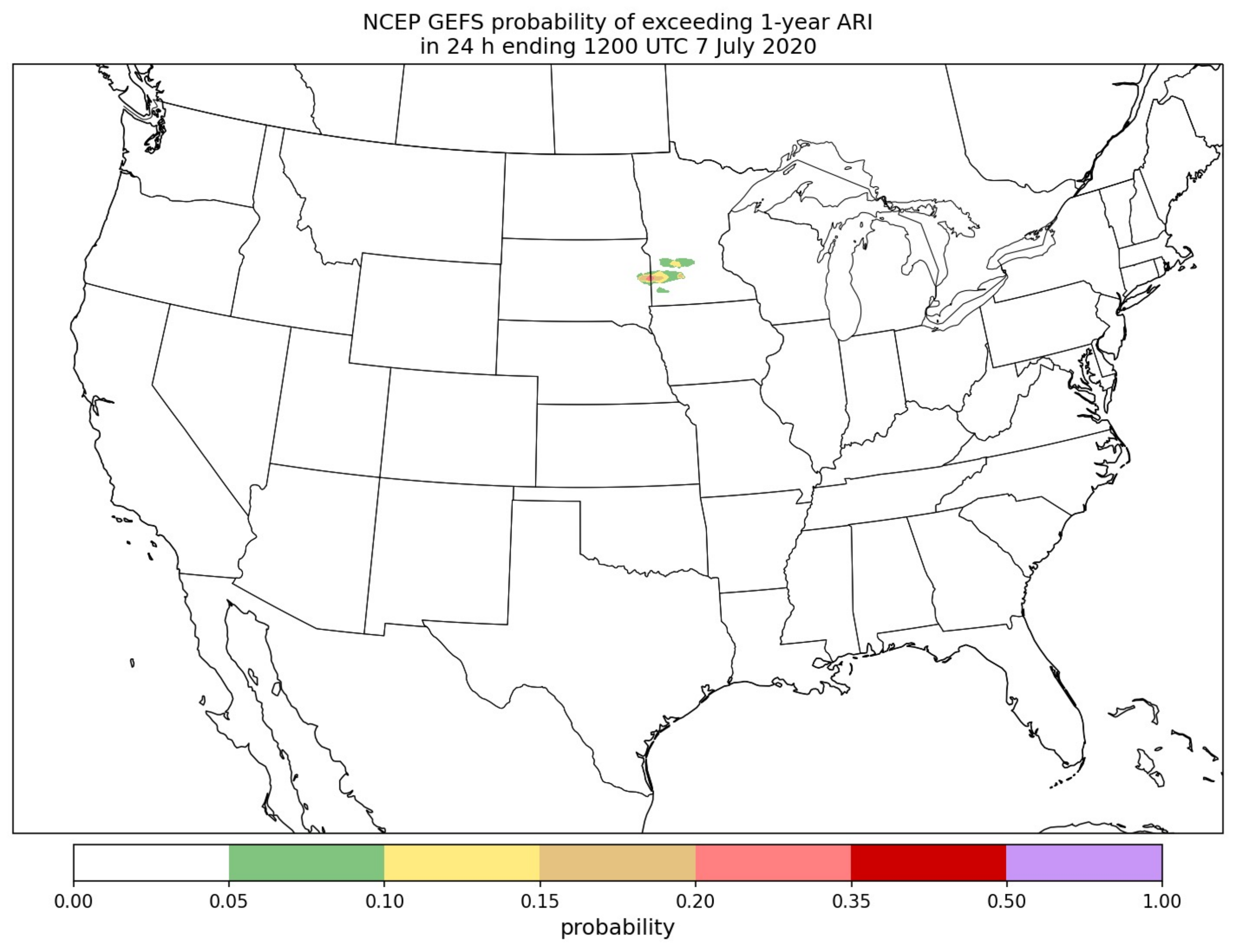


- Model QPF (APCP) most predictive region of ARI exceedances in most CONUS regions
- In regions where extreme precipitation driven by large-scale processes, APCP identified as even more predictive
- In highly convectively active regions (e.g. NGP, MDWST), model PWAT identified more predictive than APCP
- CAPE very predictive in SW region; not as much elsewhere


Regional Native Variable Feature Importances

Bigger Number -> More Predictive

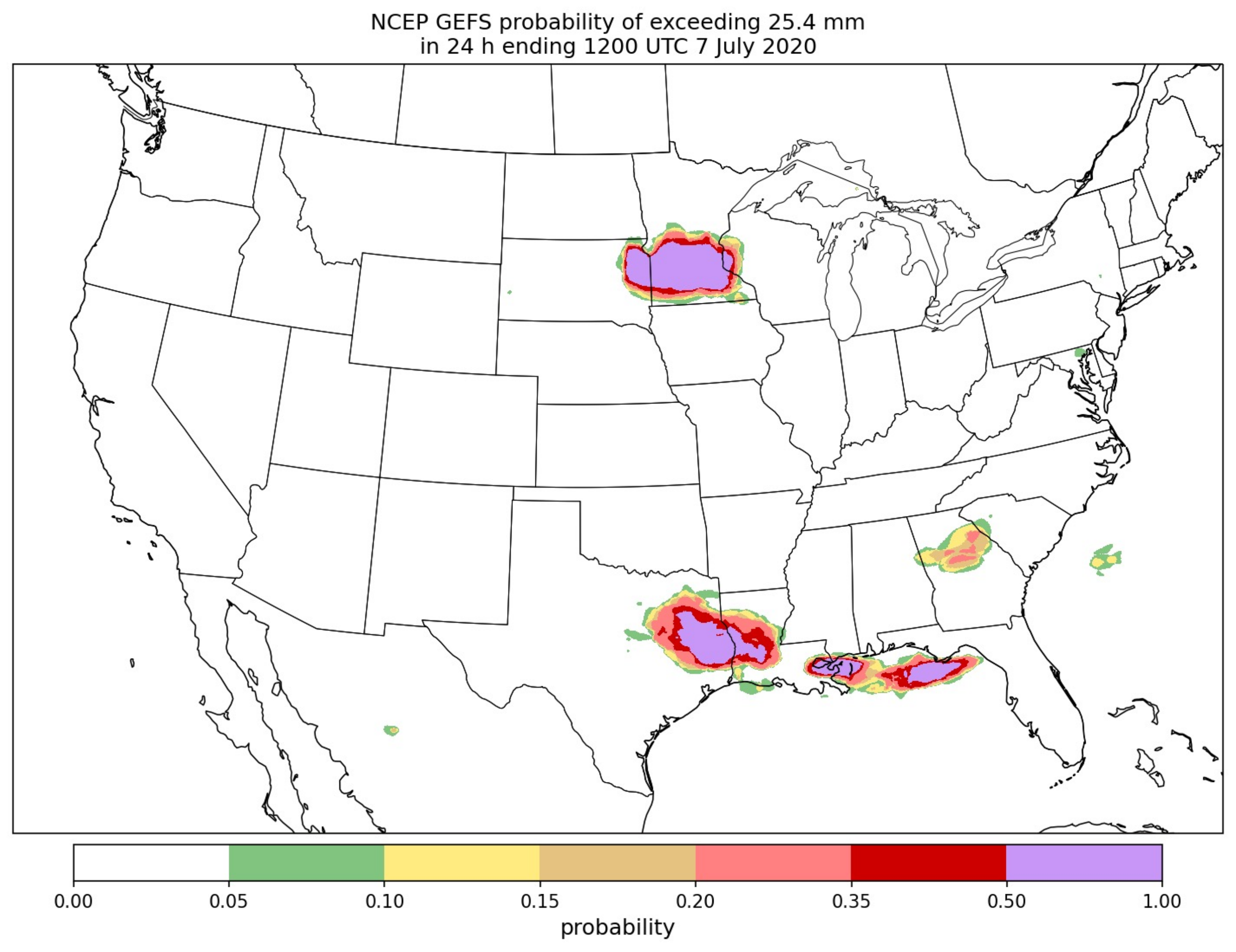
Example forecast from July 2020


NCEP GEFS probability of exceeding 1-year ARI
in 24 h ending 1200 UTC 7 July 2020
0.00
0.05
0.10
0.15
0.20
0.35
0.50
1.00
probability

Raw GEFS probability of exceeding 1-year ARI

0000 UTC 6 July 2020, 12—36-h forecast valid 1200 UTC 8 July 2020

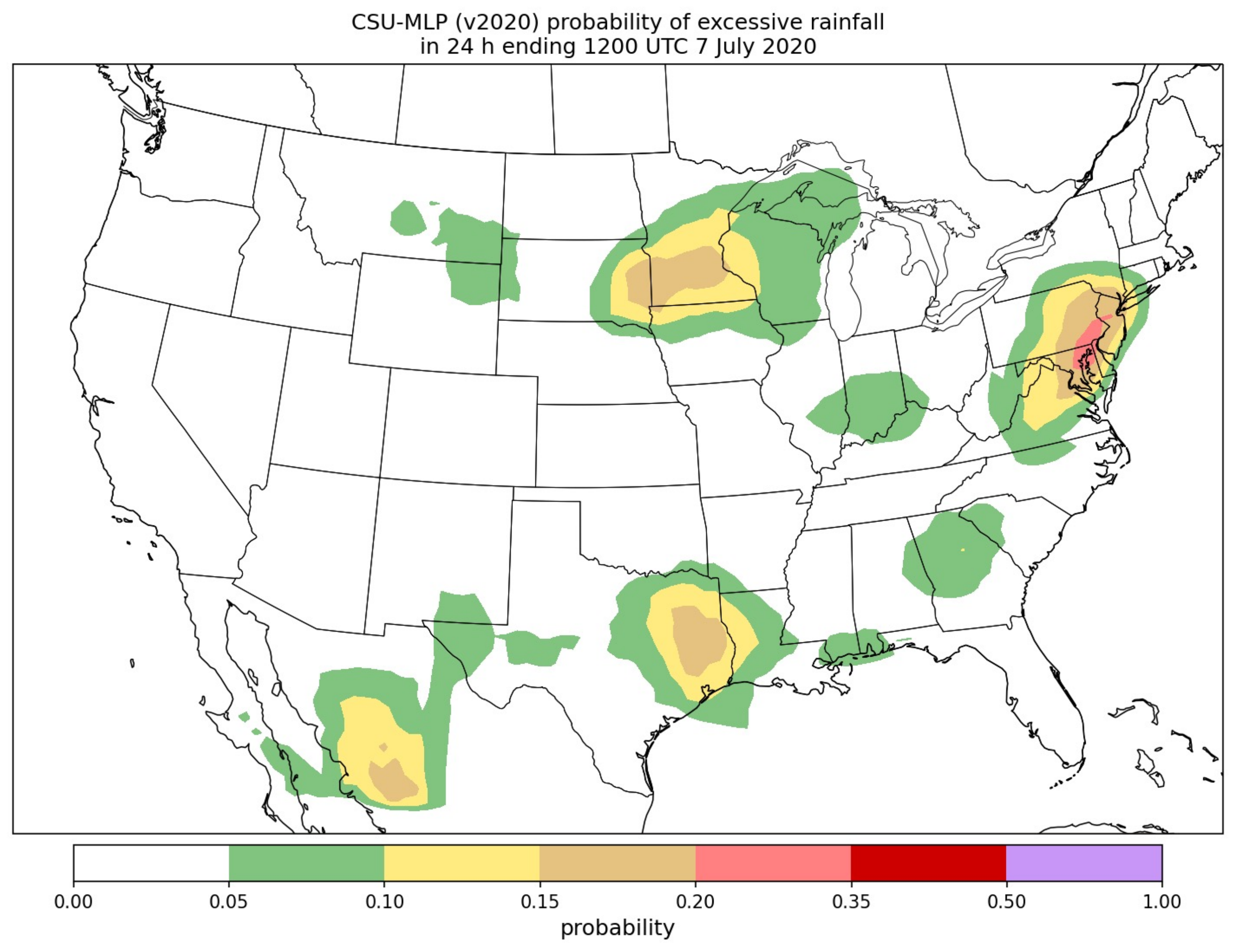
Example forecast from July 2020


NCEP GEFS probability of exceeding 25.4 mm
in 24 h ending 1200 UTC 7 July 2020
0.00
0.05
0.10
0.15
0.20
0.35
0.50
1.00
probability

Raw GEFS probability of 1” in 24 h

0000 UTC 6 July 2020, 12—36-h forecast valid 1200 UTC 7 July 2020

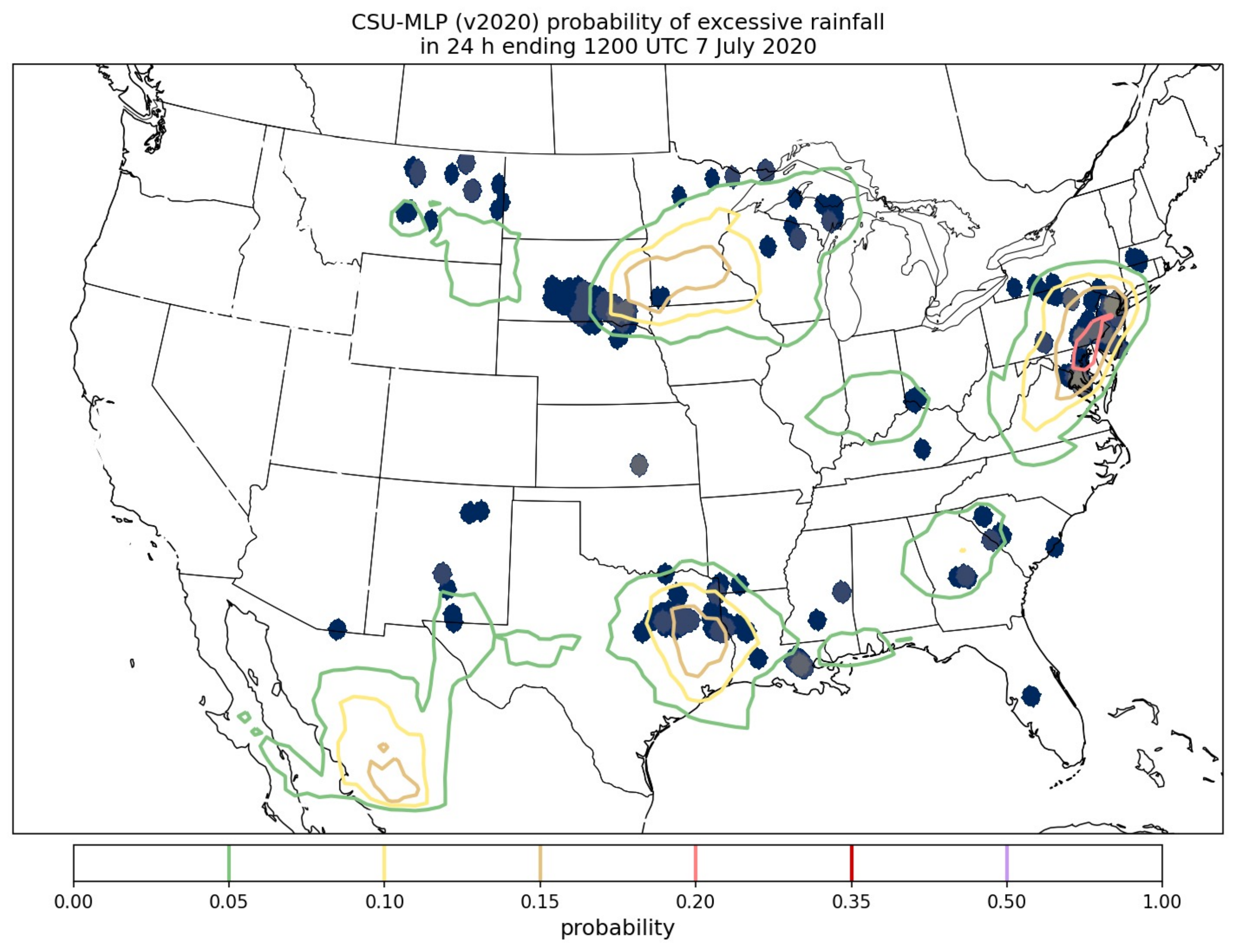
Example forecast from July 2020


CSU-MLP (v2020) probability of excessive rainfall
in 24 h ending 1200 UTC 7 July 2020
0.00
0.05
0.10
0.15
0.20
0.35
0.50
1.00
probability

CSU-MLP day-1 forecast, probability of excessive rainfall

0000 UTC 6 July 2020, 12—36-h forecast valid 1200 UTC 7 July 2020

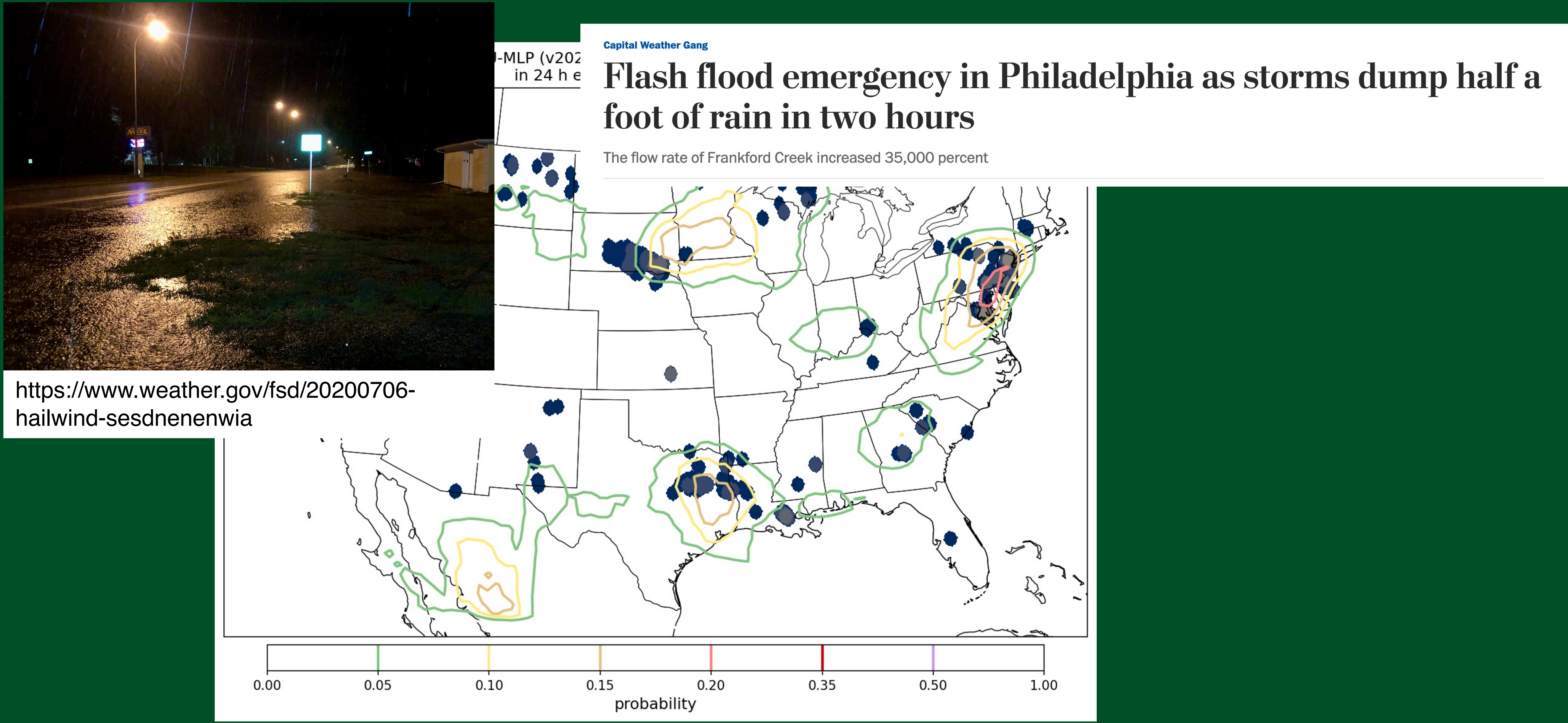
Example forecast from July 2020


CSU-MLP (v2020) probability of excessive rainfall
in 24 h ending 1200 UTC 7 July 2020
0.00
0.05
0.10
0.15
0.20
0.35
0.50
1.00
probability

CSU-MLP day-1 forecast, probability of excessive rainfall

0000 UTC 6 July 2020, 12—36-h forecast valid 1200 UTC 7 July 2020

Dots: observations of excessive rainfall from Unified Flood Verification System



https://www.weather.gov/fsd/20200706-hailwind-sesdnenenwia

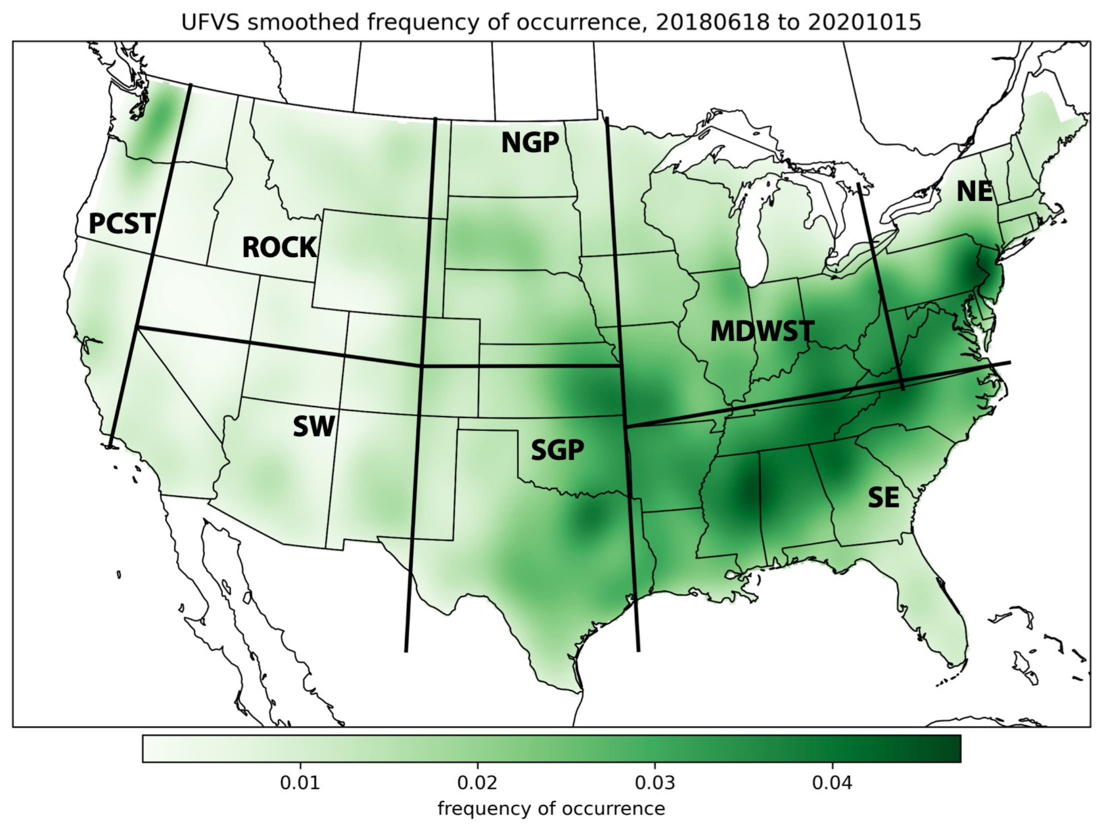
Capital Weather Gang

Flash flood emergency in Philadelphia as storms dump half a foot of rain in two hours

The flow rate of Frankford Creek increased 35,000 percent

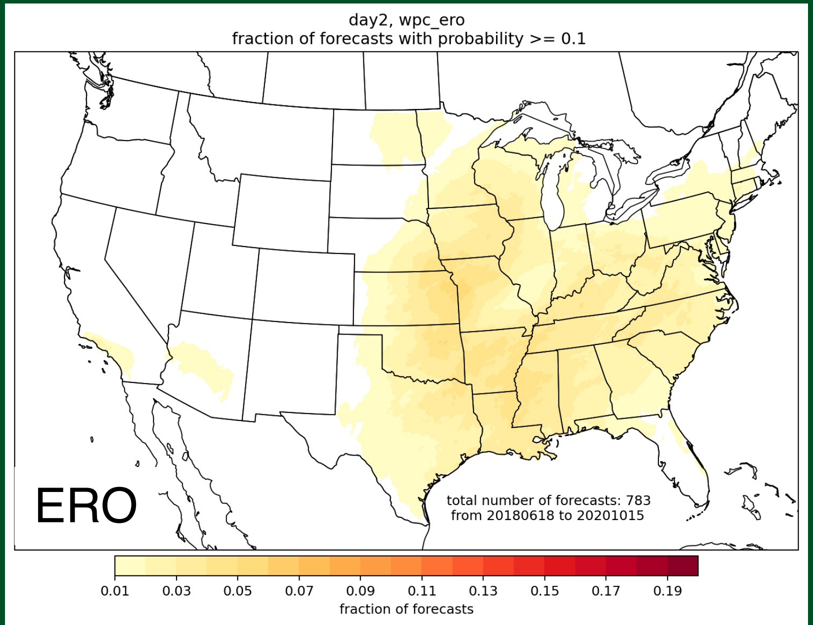
Verification of CSU-MLP forecasts

- Methods similar to those used in WPC's in-house verification
- Includes percent of area covered by observations within a contour, Brier Skill Score, area under the ROC curve
- Observation dataset is WPC's Unified database, includes flash flood guidance exceedances, 5-yr ARI exceedances, flash flood LSRs, USGS and MPING flood reports
- Verification from 18 June 2018 through 15 October 2020 for days 2-3
  - 3 March 2019 to 15 October 2020 for day 1
- Verification is done CONUS-wide as well as for each CSU-MLP region
- Comparison is to 09Z WPC operational EROs

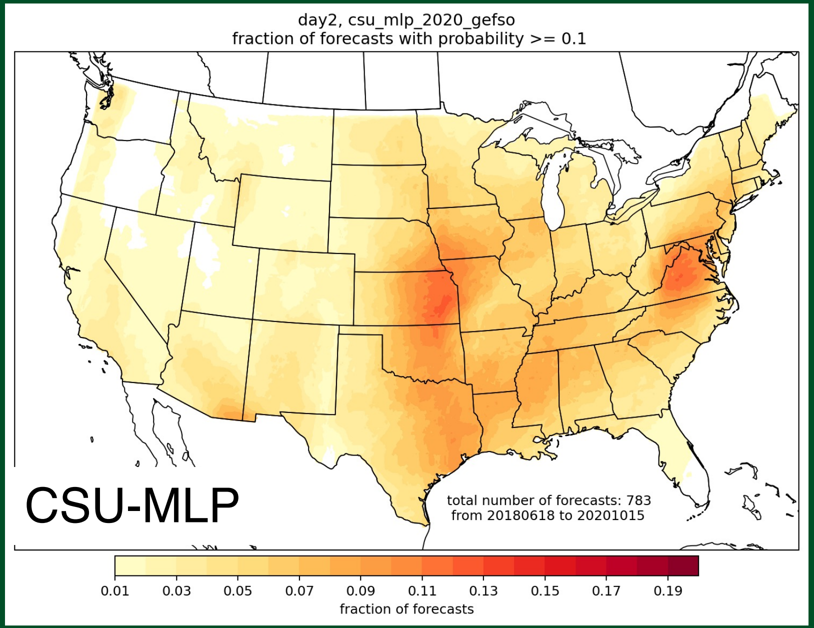


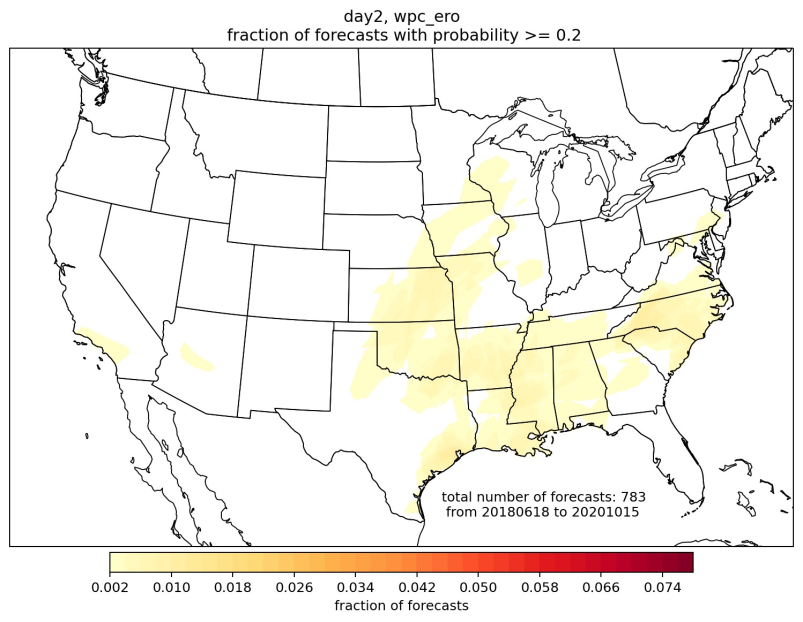
More details in Schumacher et al. (2021, *BAMS,* in press)

Frequency of forecast probabilities

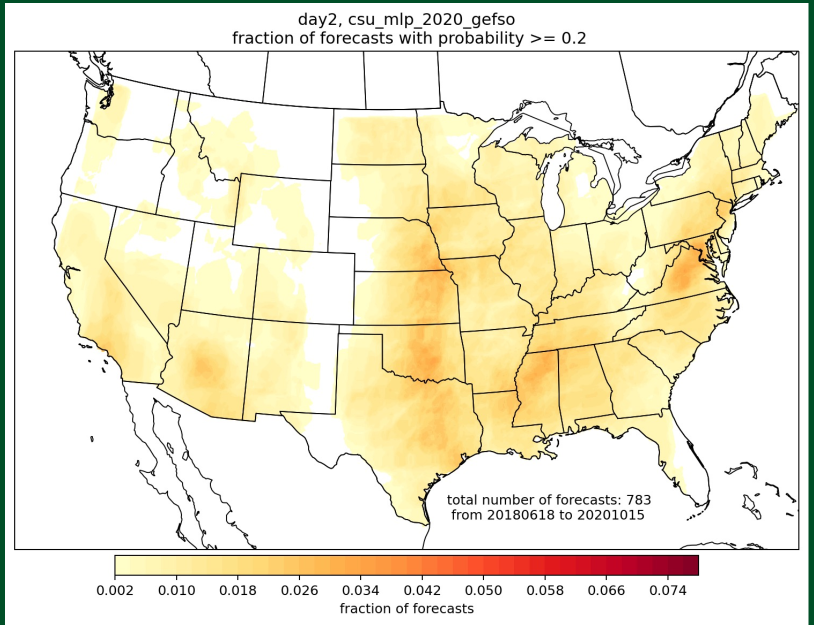

day2, wpc_ero
fraction of forecasts with probability >= 0.1

SLGT

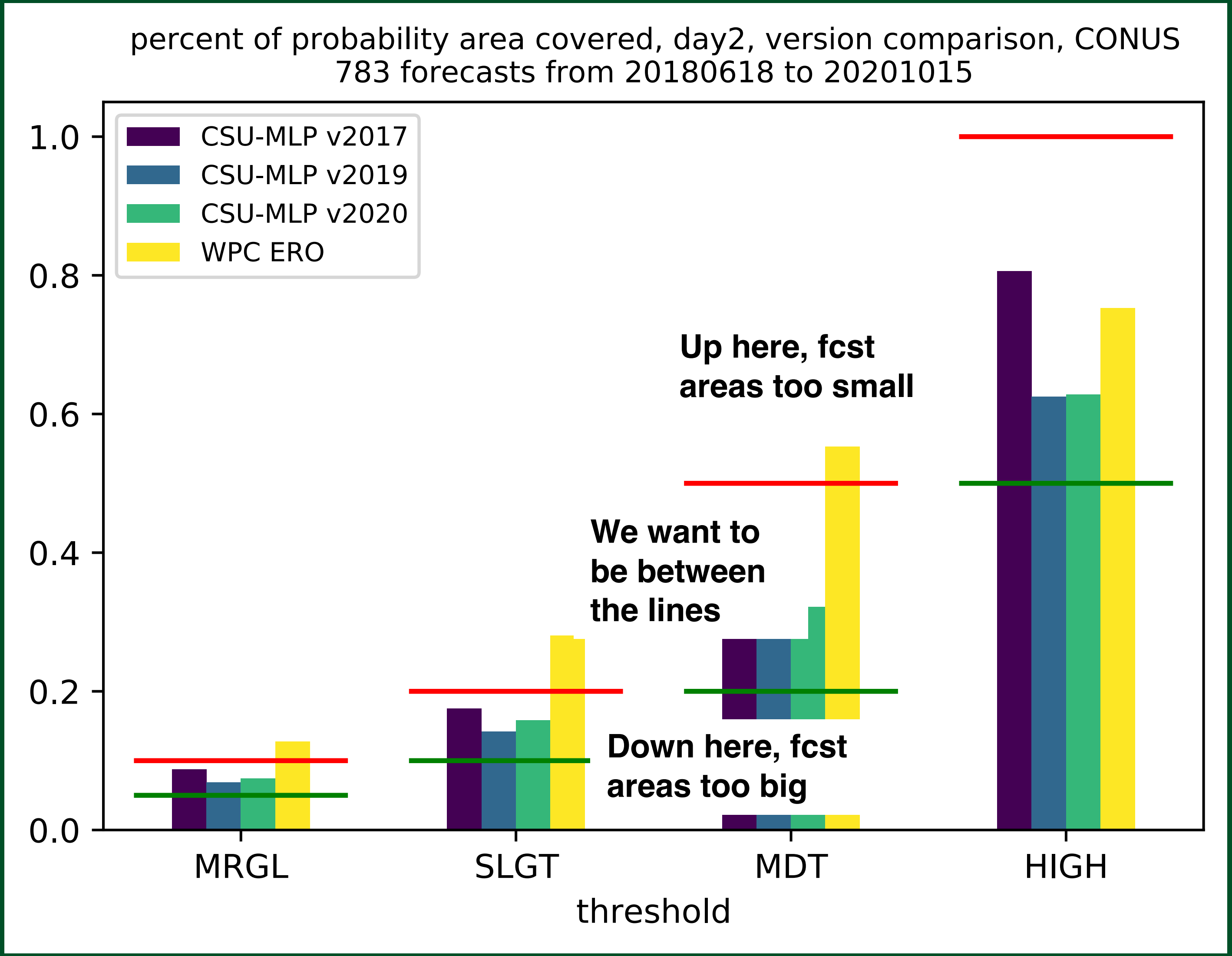

day2, csu_mlp_2020_gefso
fraction of forecasts with probability >= 0.1


day2, wpc_ero
fraction of forecasts with probability >= 0.2

MDT


day2, csu_mlp_2020_gefso
fraction of forecasts with probability >= 0.2

Day-2 outlooks: full CONUS



Up here, fcst areas too small

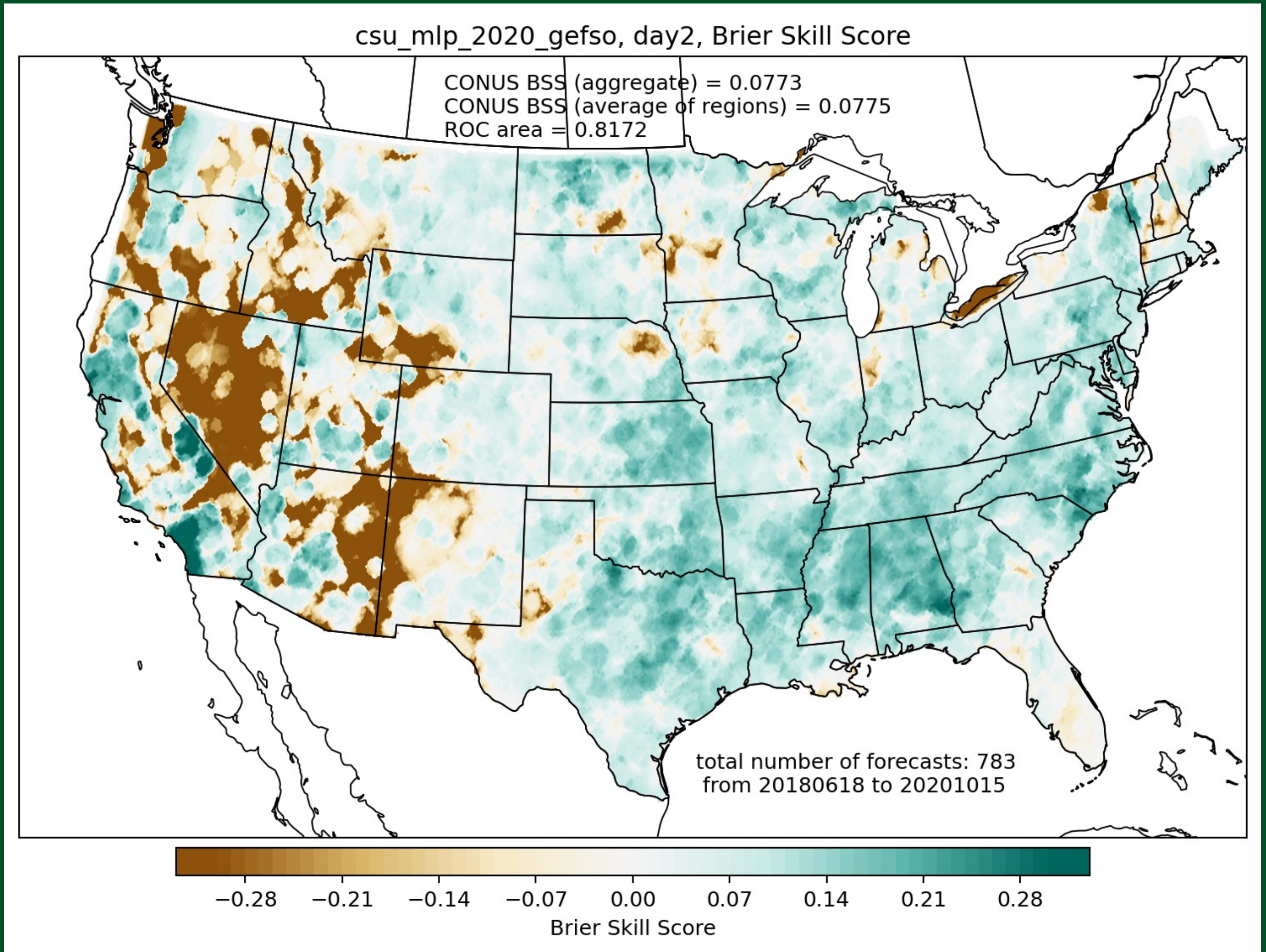
We want to be between the lines

Down here, fcst areas too big

- At least with this “truth”, WPC day-2 ERO contours (except for high risk) are too small/infrequent in coverage
- CSU-MLP versions are reasonably well calibrated for MRGL through MDT categories (sample size is very small for HIGH)
- The 2020 version improves slightly on the 2019 version, especially in problematic regions

Brier skill score comparison: day 2

CSU-MLP (v2020)

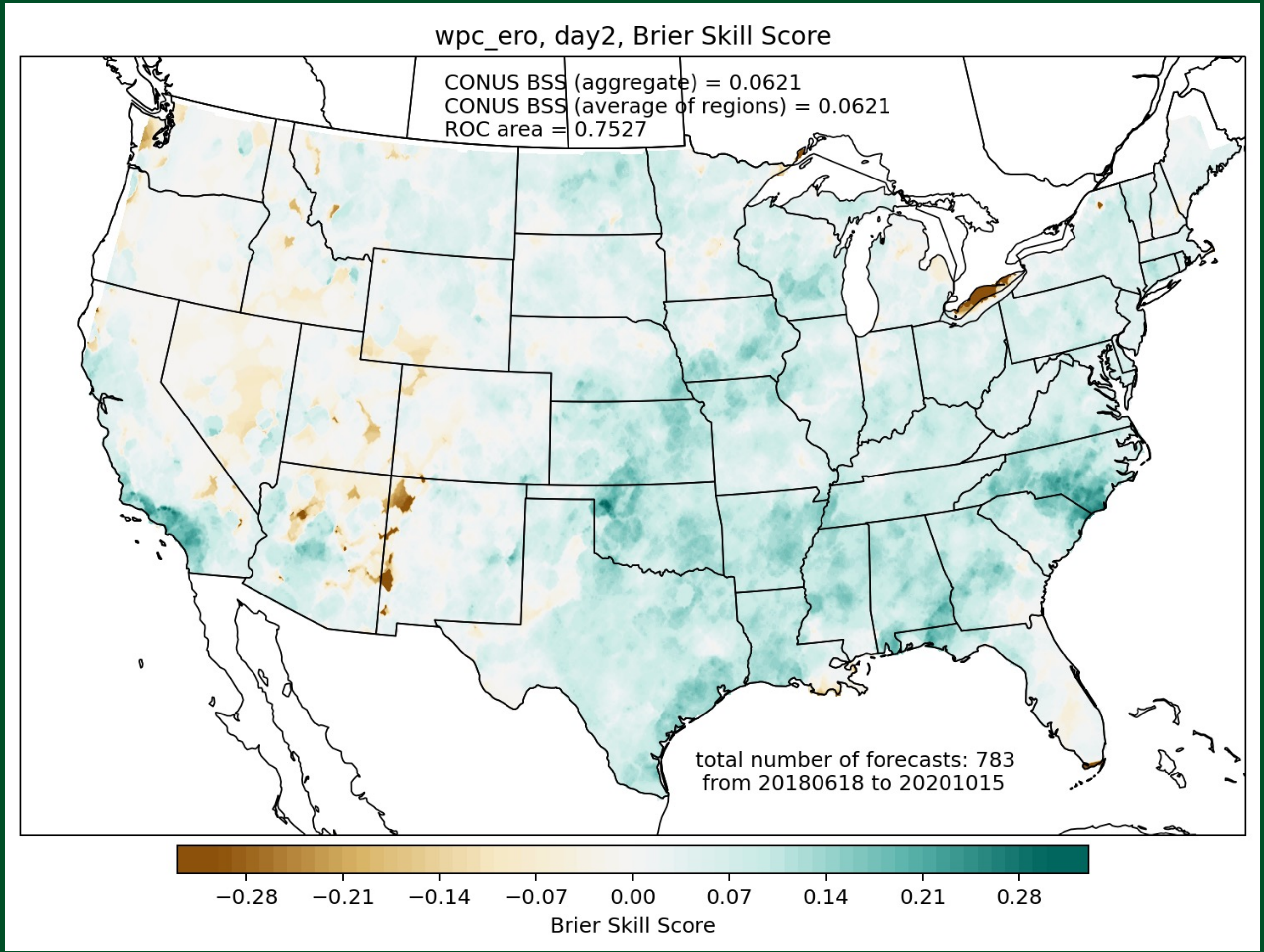


csu_mlp_2020_gefso, day2, Brier Skill Score

CONUS BSS (aggregate) = 0.0773
CONUS BSS (average of regions) = 0.0775
ROC area = 0.8172

total number of forecasts: 783
from 20180618 to 20201015

WPC ERO



wpc_ero, day2, Brier Skill Score

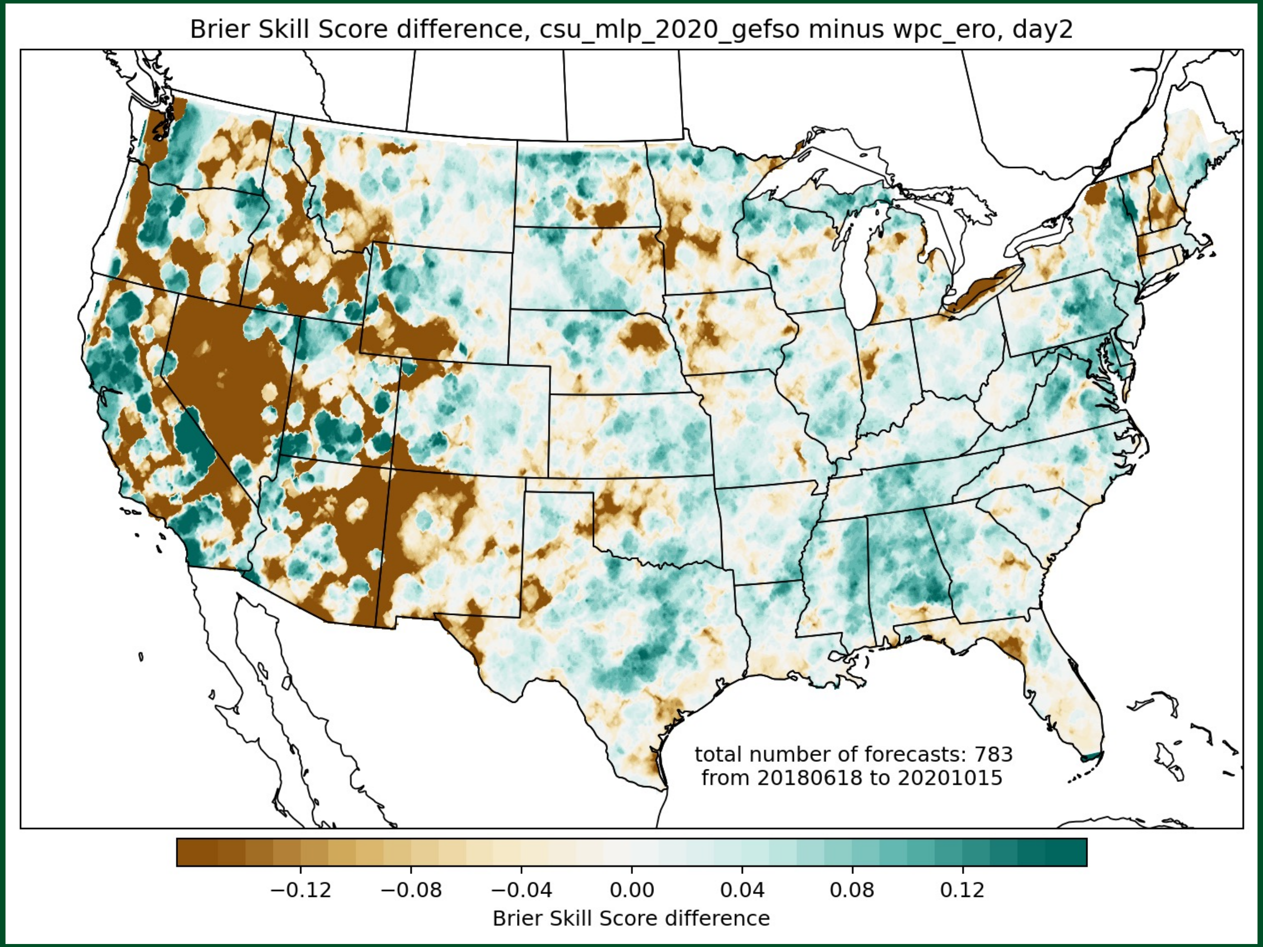
CONUS BSS (aggregate) = 0.0621
CONUS BSS (average of regions) = 0.0621
ROC area = 0.7527

total number of forecasts: 783
from 20180618 to 20201015

- Spatial patterns very similar: BSS is highest east of the Rockies, low in the west (where observed events are rare)

Brier skill score comparison: day 2

CSU-MLP minus WPC ERO

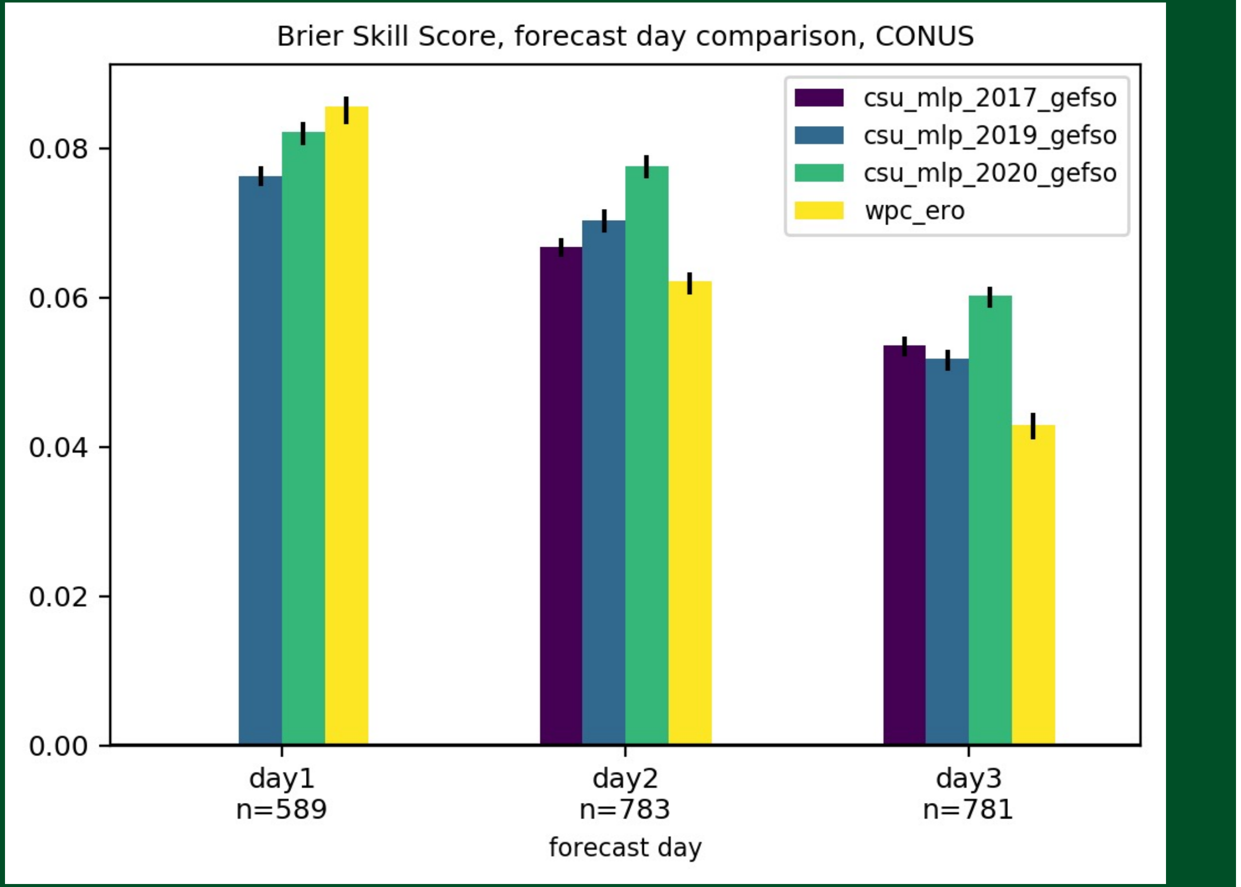

Brier Skill Score difference, csu_mlp_2020_gefso minus wpc_ero, day2
total number of forecasts: 783
from 20180618 to 20201015
−0.12
−0.08
−0.04
0.00
0.04
0.08
0.12
Brier Skill Score difference

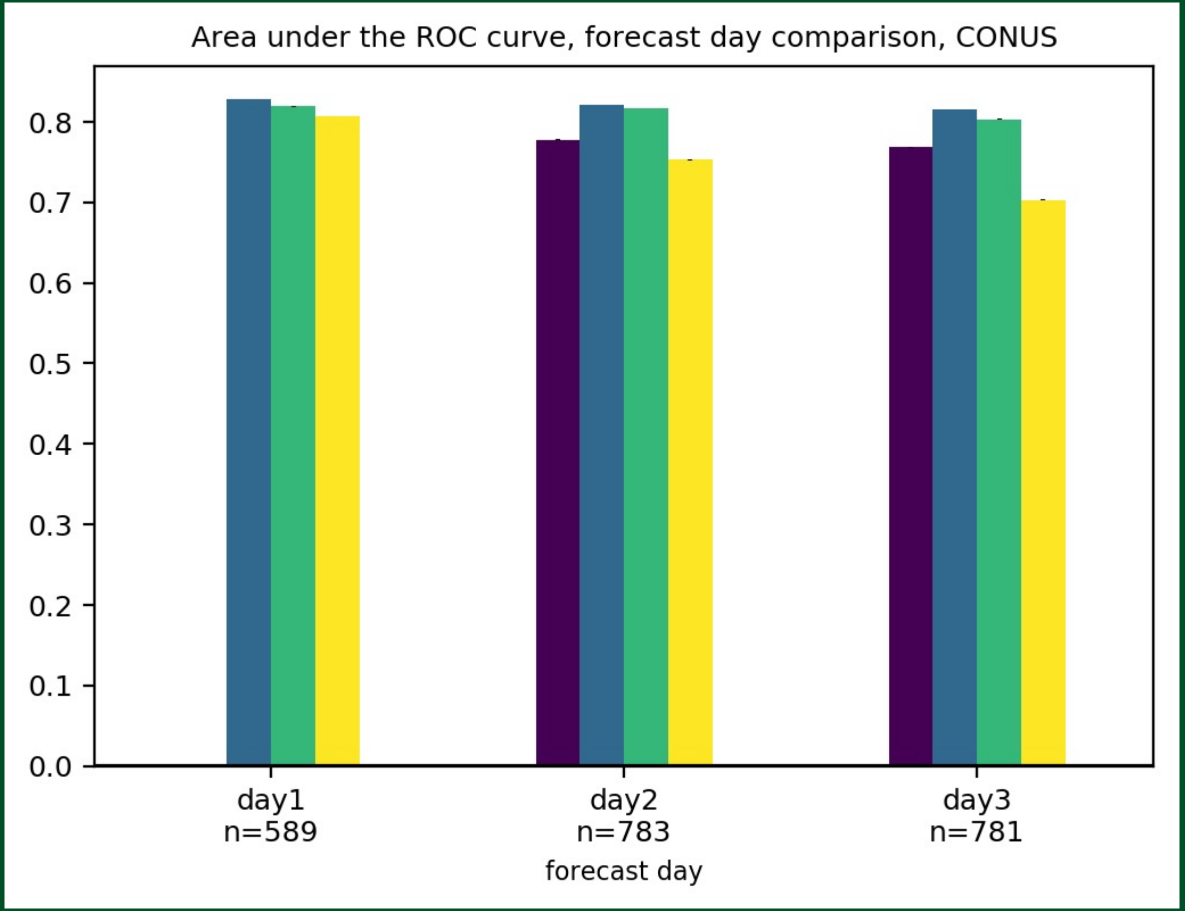
Greens: CSU-MLP has more skill
Browns: WPC ERO has more skill

CSU-MLP has:

- more skill in much of the south, southeast, and northeast
- much less skill in the interior west, but the sample size is extremely small here
- mixed performance in the rainy areas of the west coast

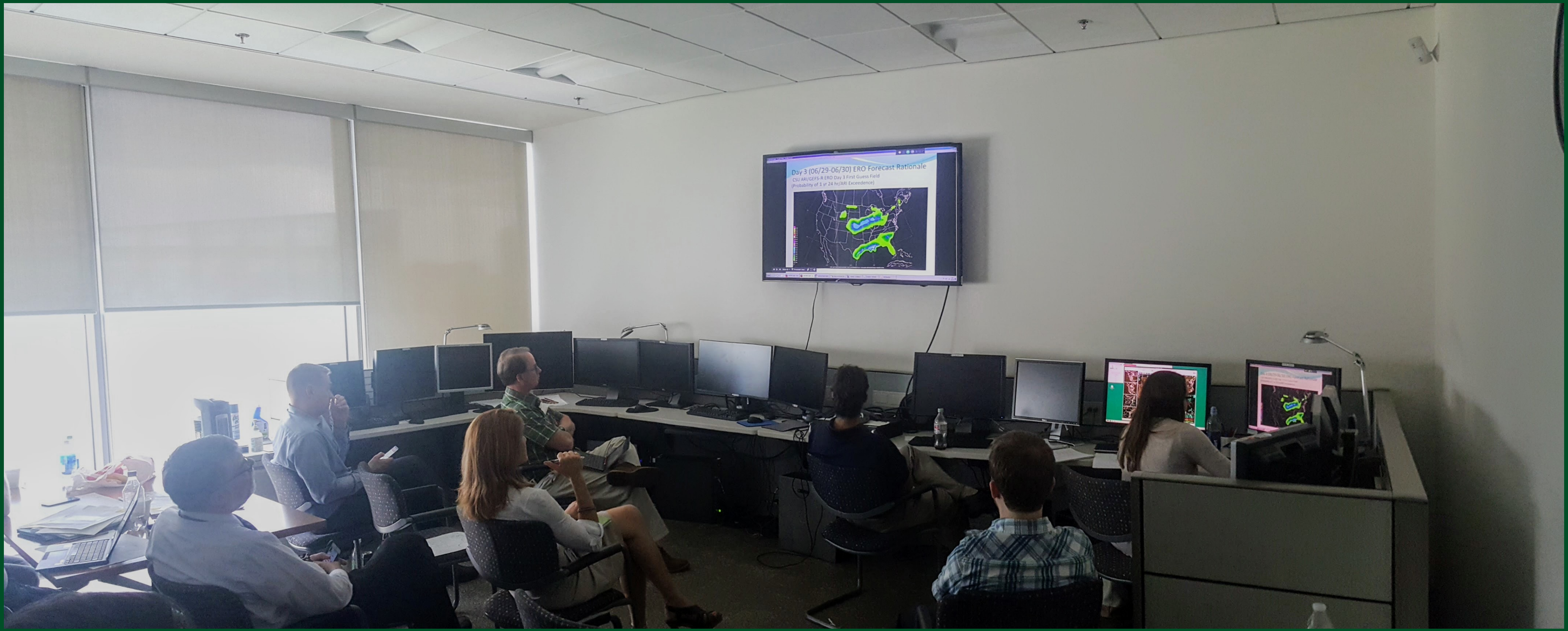
Brier skill score and ROC area, CONUS, by day





- For BSS, the WPC ERO is best on day 1, but 2020 version of CSU-MLP is best on days 2 and 3
- Big improvements from 2019 to 2020 versions on all days
- CSU-MLP generally has higher ROC area, mainly because fewer events fall outside the 5% contour (again, this is partly a function of the definitions used for ‘excessive rainfall’ here)

Collaborative improvements to the system through the Flash Flood and Intense Rainfall (FFaIR) testbed experiments, 2017-2020


Day 3 (06/29-06/30) ERO Forecast Rationale
CSU ARI/GEFS-R ERO Day 3 First Guess Field
(Probability of 1 yr 24 hr/ARI Exceedence)

See also Erickson et al. (2019, JAMC)

2017: “participants overwhelmingly agreed that the CSU-MLP First Guess Field is an excellent step in providing an initial starting point for WPC
ERO forecasts, which has been a long-requested tool.”

2018: “The WPC-HMT recommends the Day 2 and Day 3 ERO CSU-MLP First Guess Field for operations as it showed great potential and was scored well by participants. It is recommended that the CSU developers work to refine some of the high probabilities in the High Plains and low probabilities in the Southeast. . .”

2019: day 2-3 models transitioned to operations
2020: day 1 model transitioned to operations

Use in WPC operations

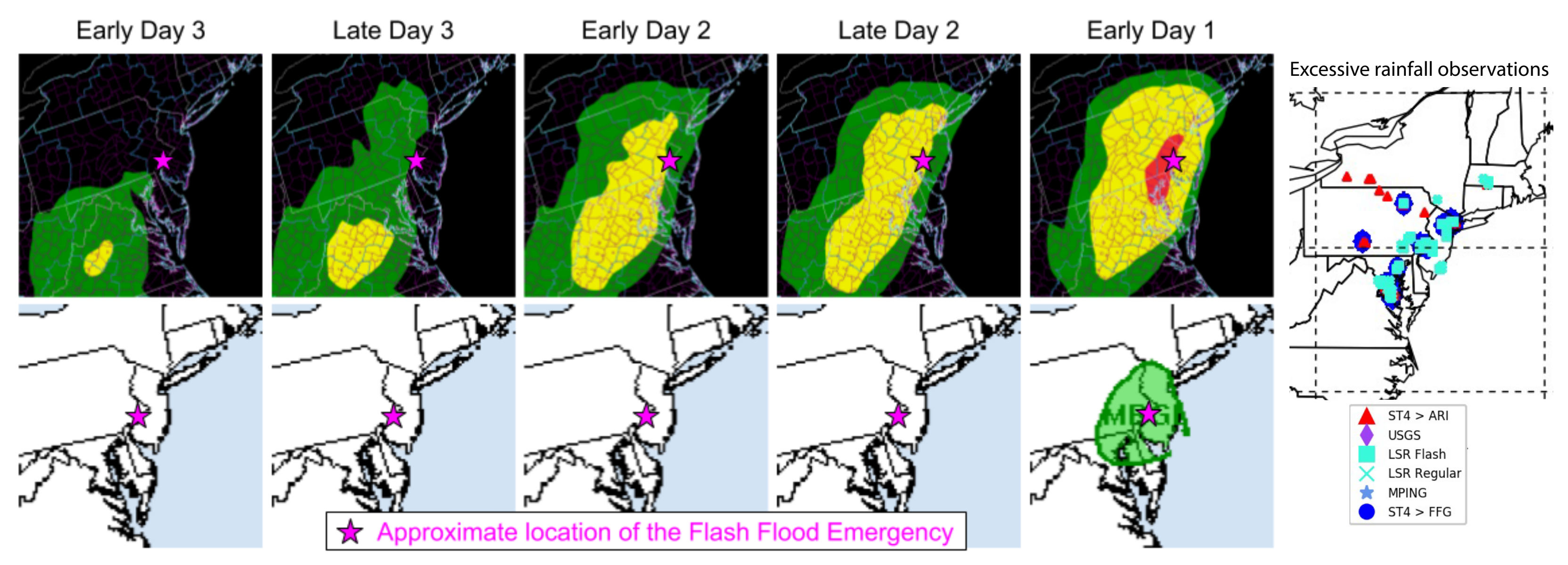
Prior to the CSU-MLP’s availability, WPC forecasters examined an abundance of numerical model and observational data to create the product, which was often challenging given tight product deadlines. Consequently, the CSU-MLP has resulted in notable time savings to forecasters.

Forecaster feedback has been positive, and while there is still a perception of a high bias, objective evidence that the CSU-MLP is well-calibrated with respect to fractional coverage has resulted in a general trend of forecasters to increase the areal extent of their risk areas. In addition, there have been several cases where the CSU-MLP has correctly highlighted a high impact event where WPC had relatively low risk potential.

CSU-MLP

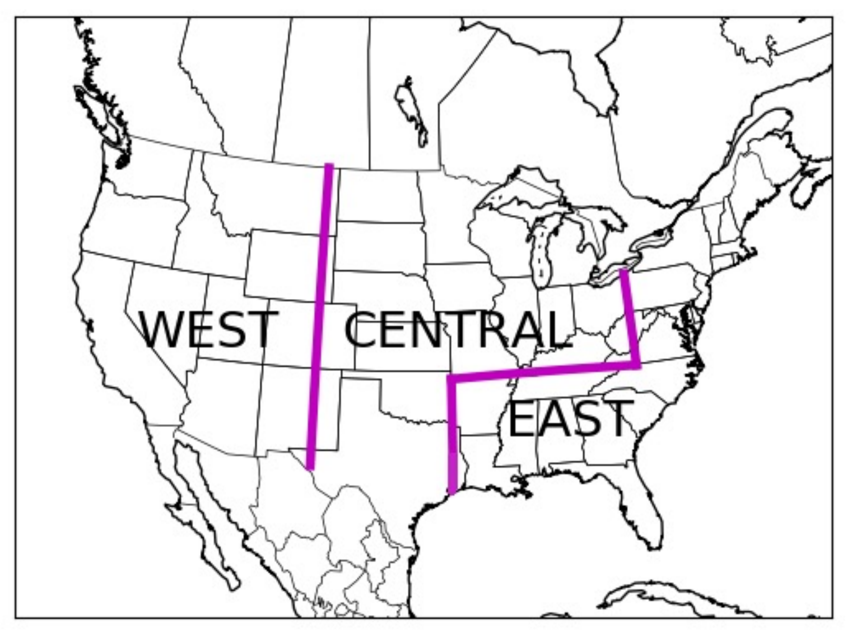
6-7 July 2020 event

WPC ERO



Severe weather CSU-MLP guidance

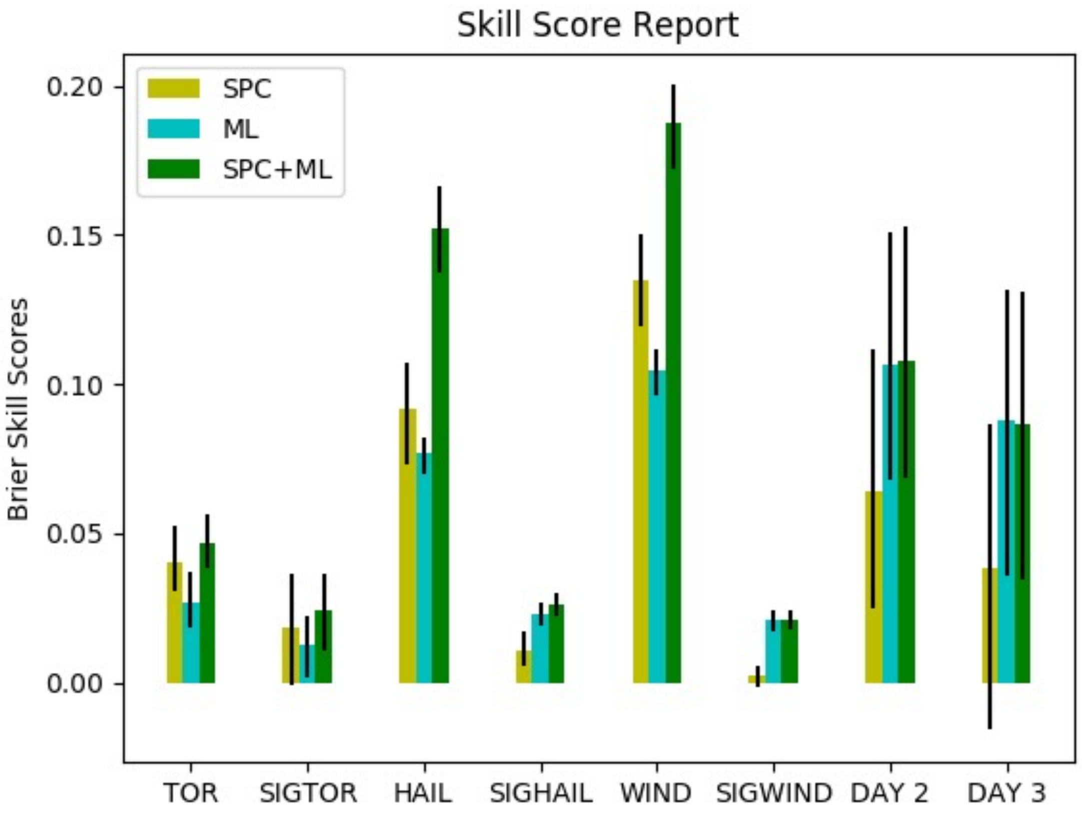
- Same general approach as the excessive rainfall forecasts, except:
  - Only 3 regions
  - Slightly different set of environmental variables (here we also include SRH and LCL height)
- Tornado, hail, and wind reports are labels/predictands
  - Predict individual hazards on Days 1 and 2, aggregate hazards on Day 3; sig severe too at Days 1 and 2
  - Forecasts analogous to SPC outlooks
- Compare to best operational forecasts: SPC outlooks
- Where does the statistical guidance benefit forecasters?
  - Blend RF forecasts with SPC outlooks based on relative skill



Hill, Herman, and Schumacher, 2020: Forecasting severe weather with random forests. *Monthly Weather Review*, 148, 2135--2161

Aggregated Forecast Skill

- Compared to RFs, SPC generally cannot be beat at Day 1
- RFs generate more skillful Day 2 and 3 forecasts
- Blended model produces best forecasts across all predictands, but is generally equal in skill to RF forecasts at Days 2 and 3

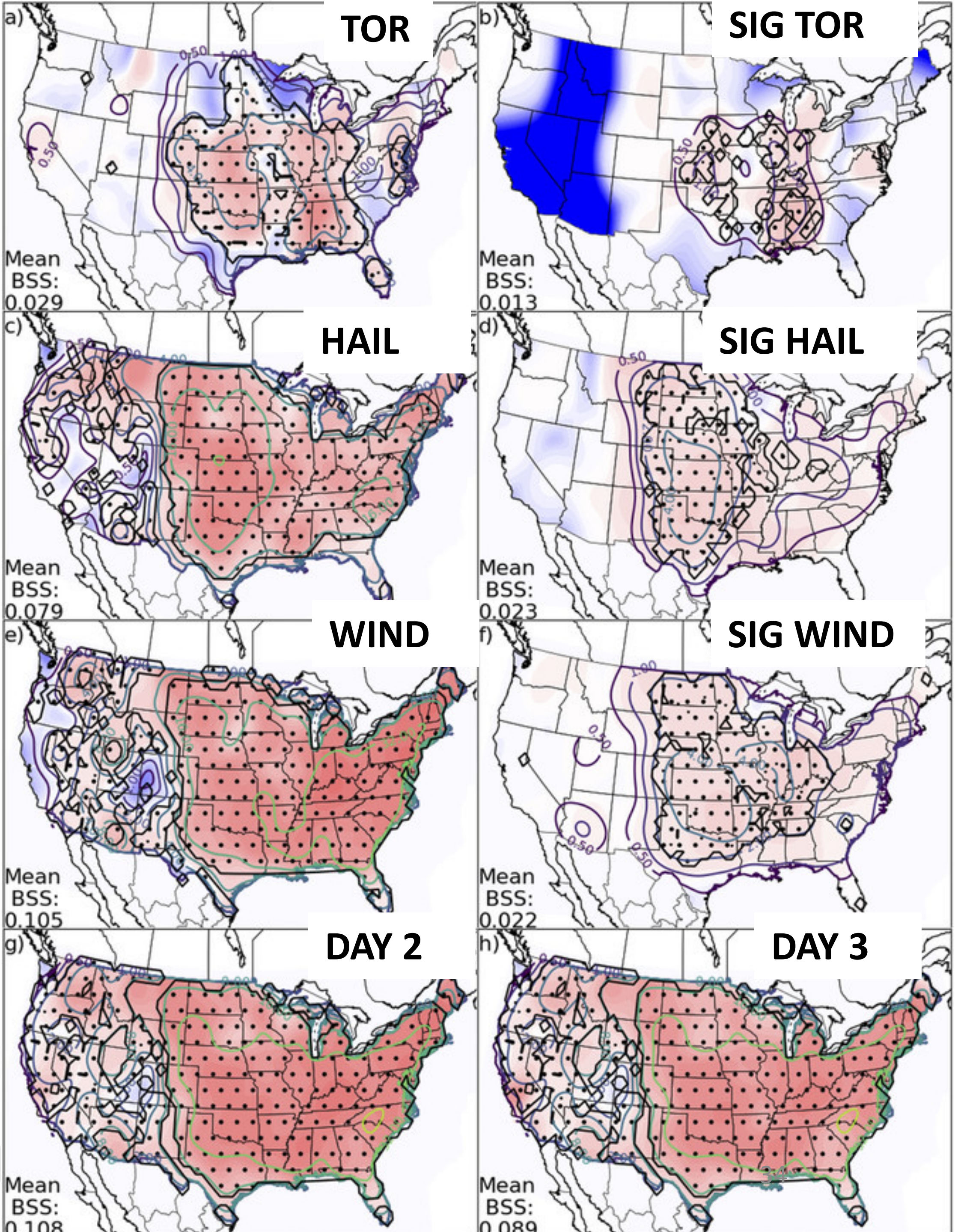


Blend Weights:

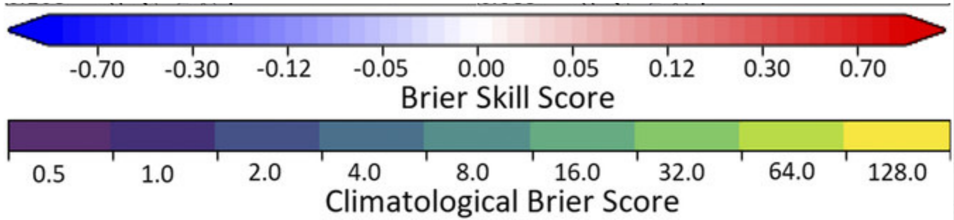
$$W_{\mathrm{SPC}} = \frac{\dfrac{1}{1-\mathrm{BSS}_{\mathrm{SPC}}}}{\dfrac{1}{1-\mathrm{BSS}_{\mathrm{SPC}}} + \dfrac{1}{1-\mathrm{BSS}_{\mathrm{RF}}}}; \quad W_{\mathrm{RF}} = 1 - W_{\mathrm{SPC}}. \tag{1}$$

Are the forecasts skillful?

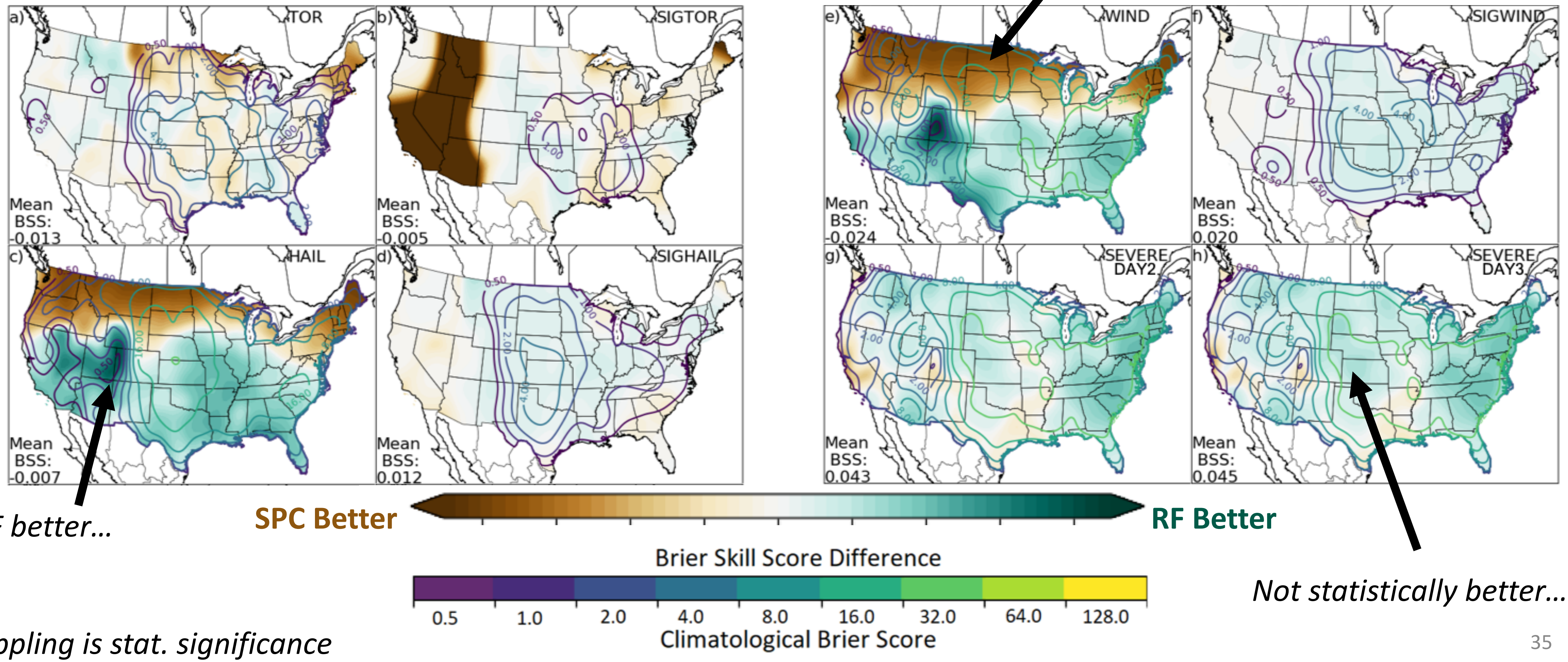
- For almost all hazards and lead times, yes!
- The significant severe forecasts are less skillful than their ”regular severe” counterparts



Red: positive skill
Blue: negative skill



CSU-MLP BSS minus SPC BSS



SPC much better…

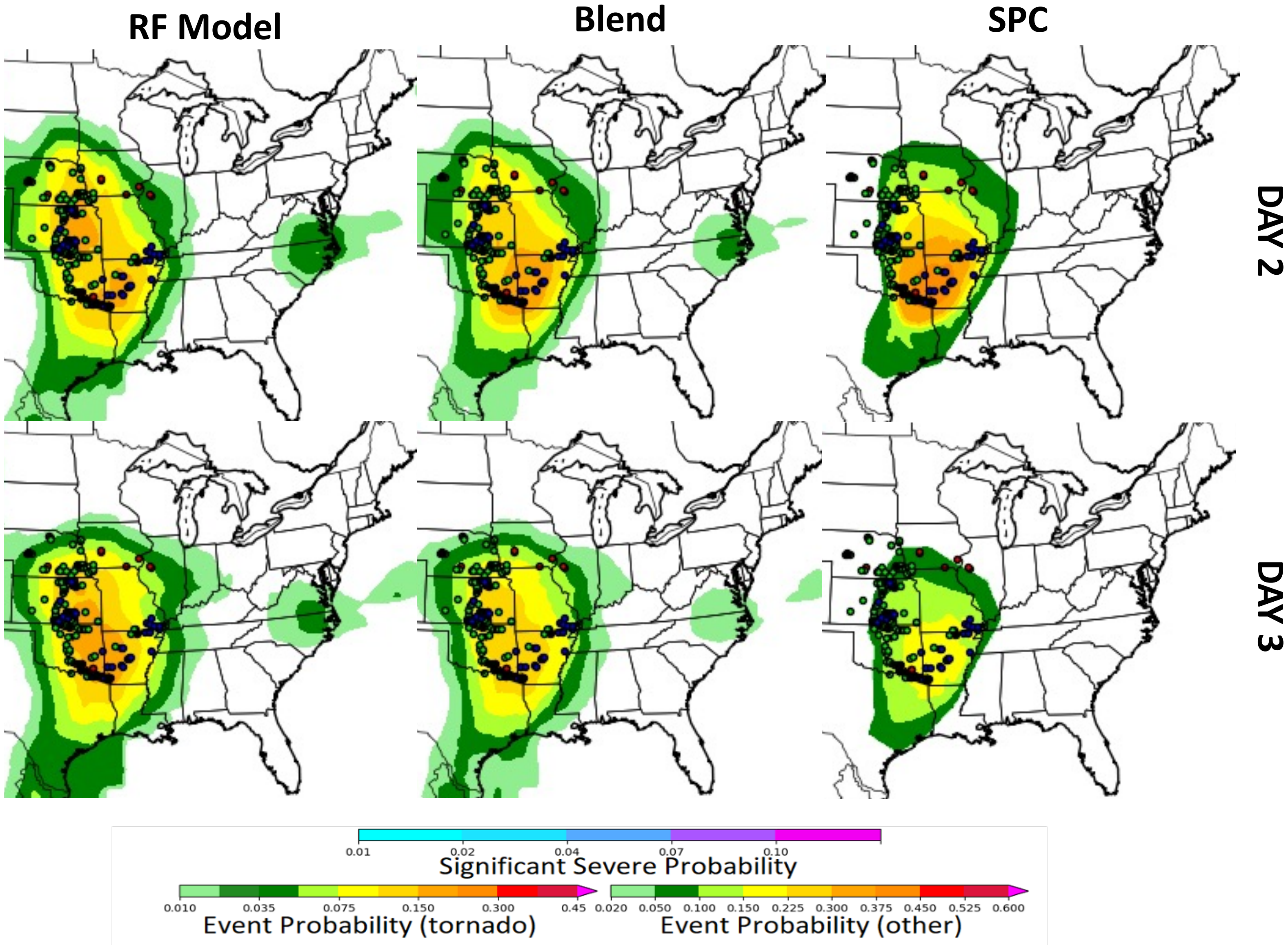
RF better…

Not statistically better…

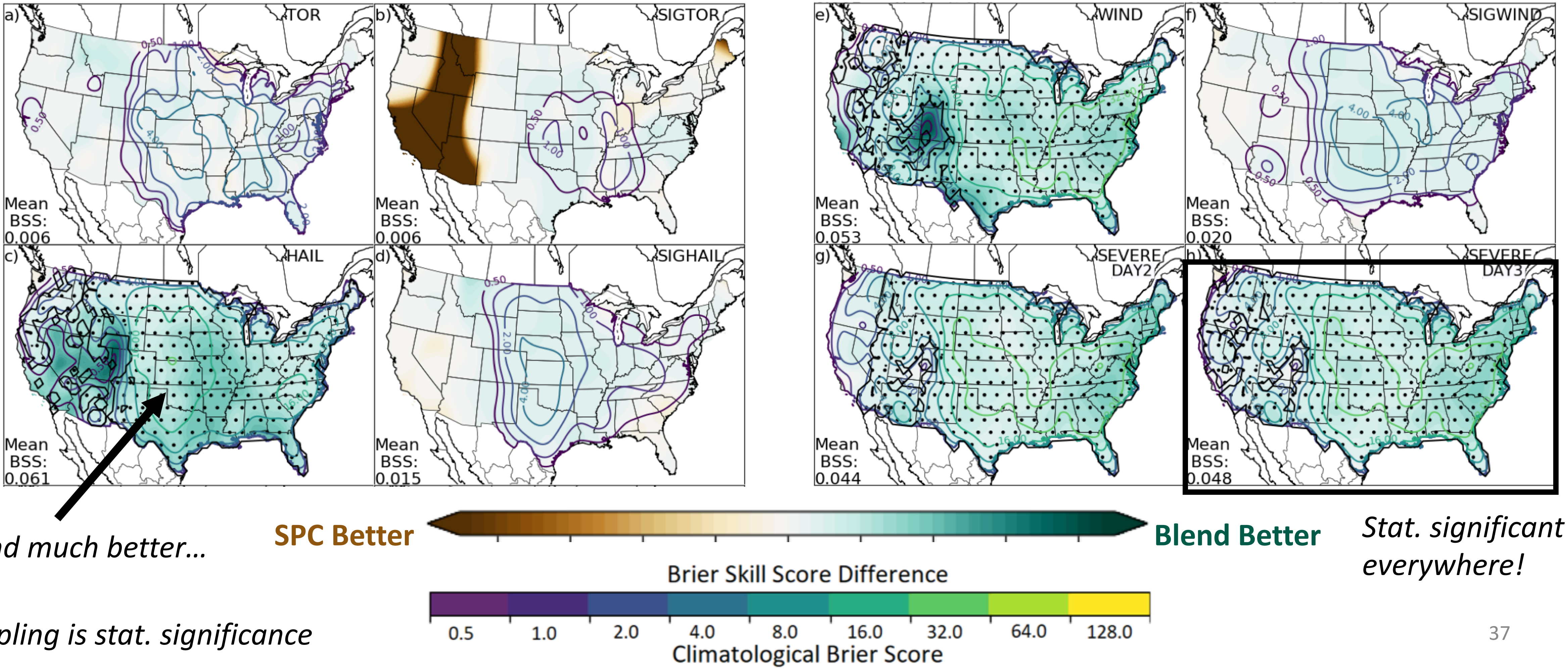
Stippling is stat. significance

Example Day 2 and 3 Forecasts

24-hr period ending
1200 UTC 10 May 2016



Blended Model BSS Differences

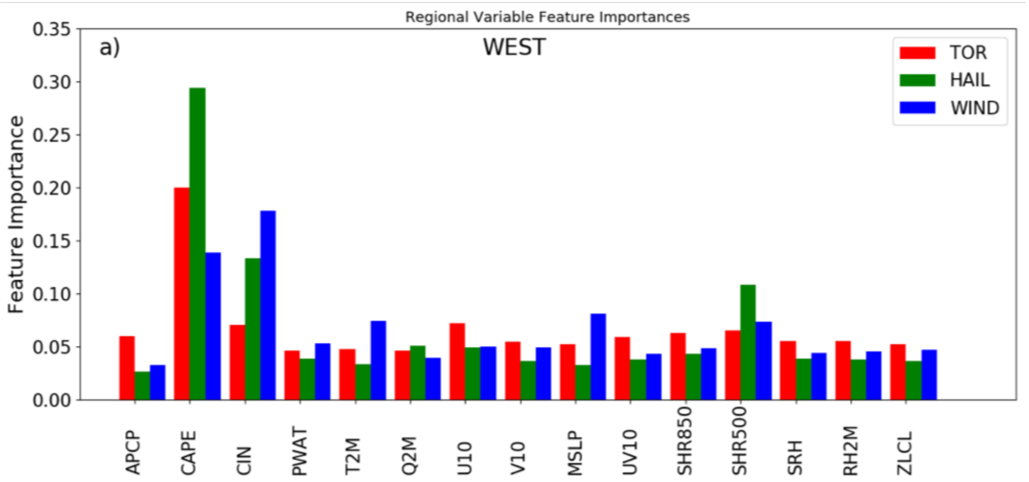


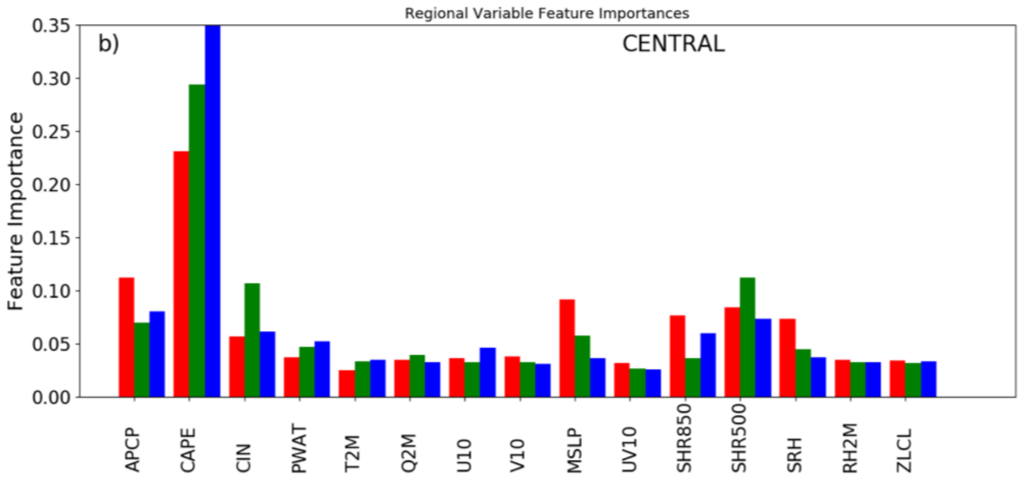
Blend much better…

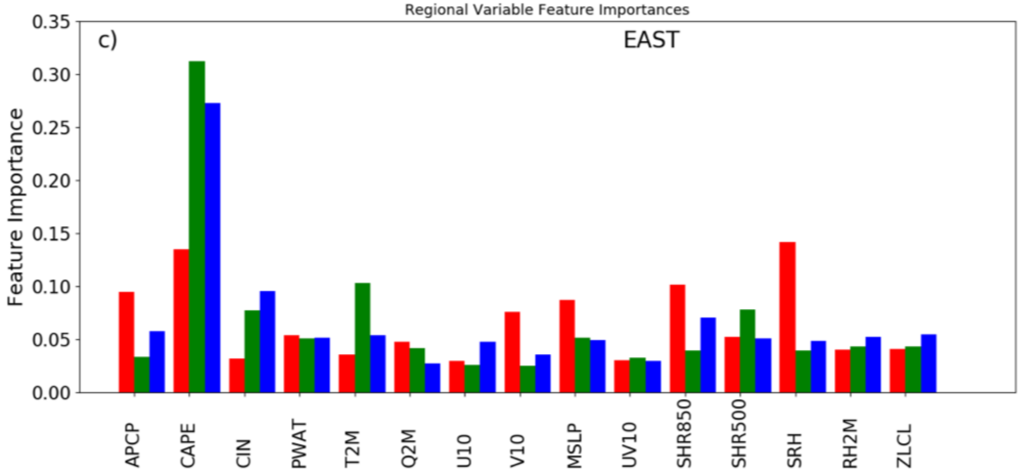
Stat. significant everywhere!

Stippling is stat. significance

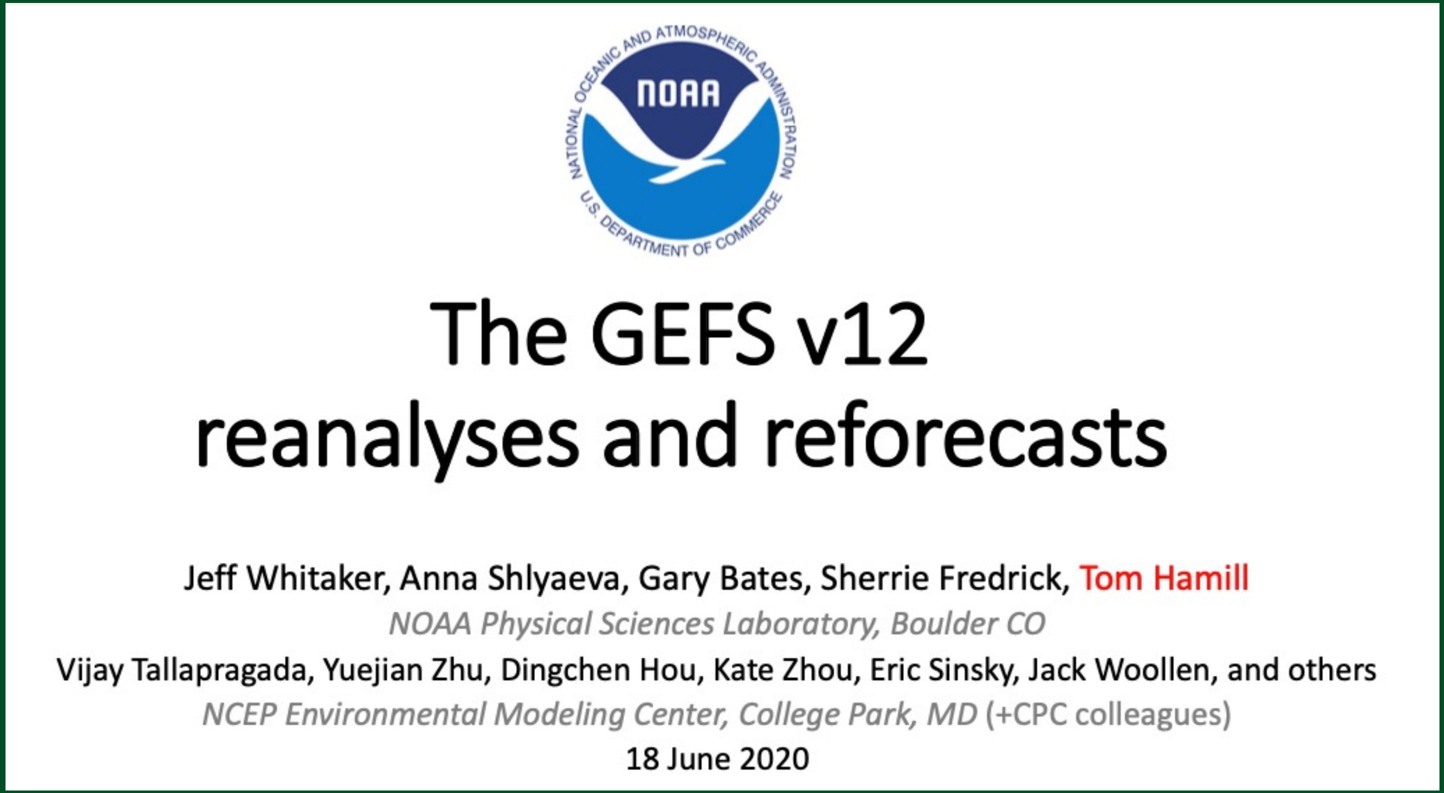
Interpretability







- CAPE is the most important predictor for all hazards, and particularly for hail
- SRH eclipses CAPE for tornado prediction in the east region
- CAPE and CIN are most important for wind prediction
- Shear, CIN, MSLP, and accumulated precipitation (APCP) tend to “pop out” as important



The GEFS v12 reanalyses and reforecasts

Jeff Whitaker, Anna Shlyaeva, Gary Bates, Sherrie Fredrick, Tom Hamill
NOAA Physical Sciences Laboratory, Boulder CO
Vijay Tallapragada, Yuejian Zhu, Dingchen Hou, Kate Zhou, Eric Sinsky, Jack Woollen, and others
NCEP Environmental Modeling Center, College Park, MD (+CPC colleagues)
18 June 2020

- The upgraded GEFS has a reforecast dataset that’s consistent with the current operational ensemble – **hugely important for post-processing/machine learning applications!**

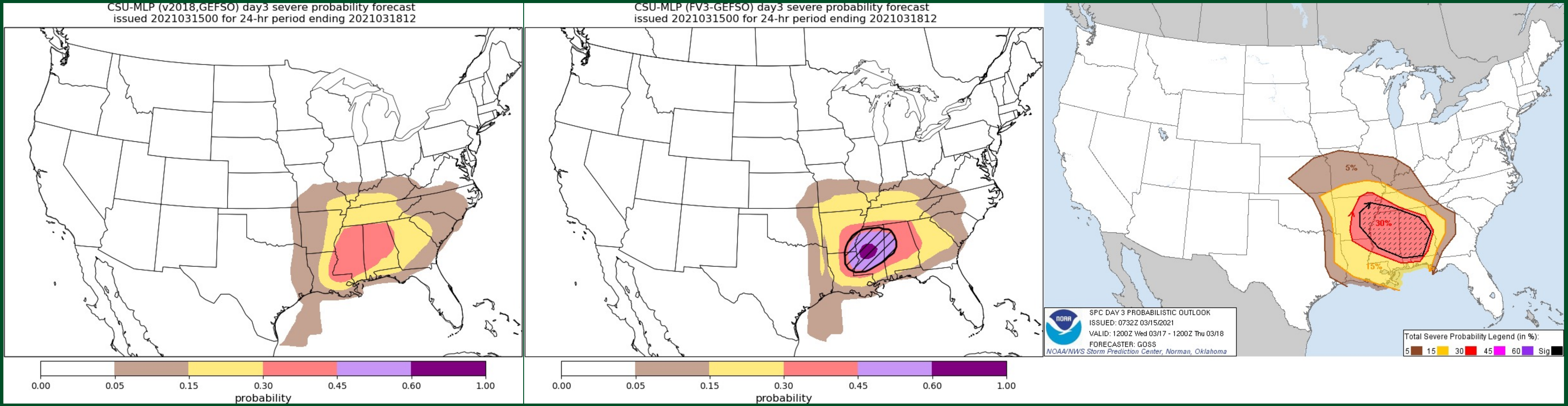
- Reforecast has 5 ensemble members daily from 2000-2019 (11 members once per week)

Our first project: re-train the severe weather models using the GEFS v12 reforecast dataset

“old” day-3 forecast: trained on reforecast v2, driven by GEFS v12

new day-3 forecast: trained on & driven by GEFS v12

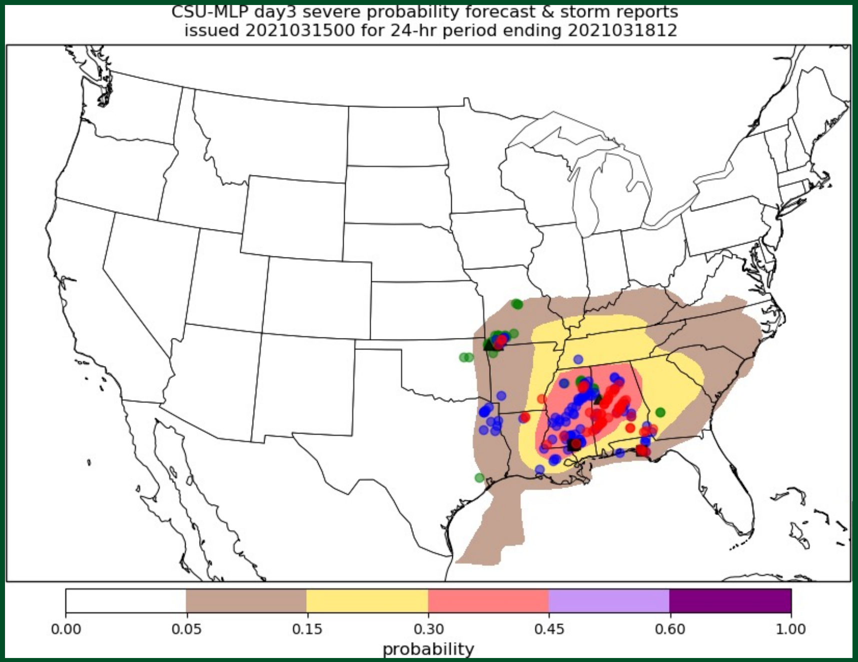
SPC day-3 outlook


CSU-MLP (v2018,GEFSO) day3 severe probability forecast
issued 2021031500 for 24-hr period ending 2021031812
0.00
0.05
0.15
0.30
0.45
0.60
1.00
probability
CSU-MLP (FV3-GEFSO) day3 severe probability forecast
issued 2021031500 for 24-hr period ending 2021031812
0.00
0.05
0.15
0.30
0.45
0.60
1.00
probability
5%
30%
15%
SPC DAY 3 PROBABILISTIC OUTLOOK
ISSUED: 0732Z 03/15/2021
VALID: 1200Z Wed 03/17 - 1200Z Thu 03/18
FORECASTER: GOSS
NOAA/NWS Storm Prediction Center, Norman, Oklahoma
Total Severe Probability Legend (in %):
5
15
30
45
60
Sig

Day-3 forecasts made on 15 March 2021, for 24-h period ending 18 March 2021

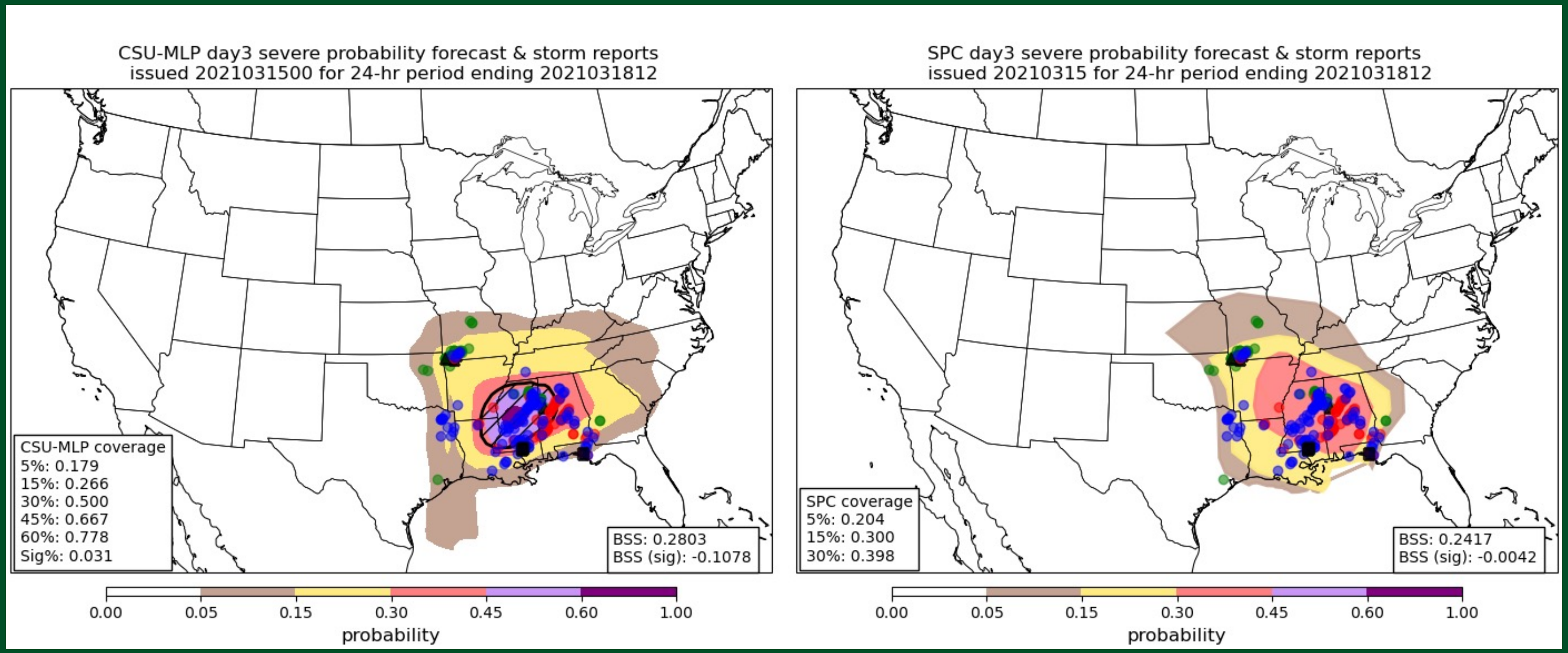
Our first project: re-train the severe weather models using the GEFS v12 reforecast dataset

“old” day-3 forecast: trained on reforecast v2, driven by GEFS v12


CSU-MLP day3 severe probability forecast & storm reports
issued 2021031500 for 24-hr period ending 2021031812
0.00
0.05
0.15
0.30
0.45
0.60
1.00
probability

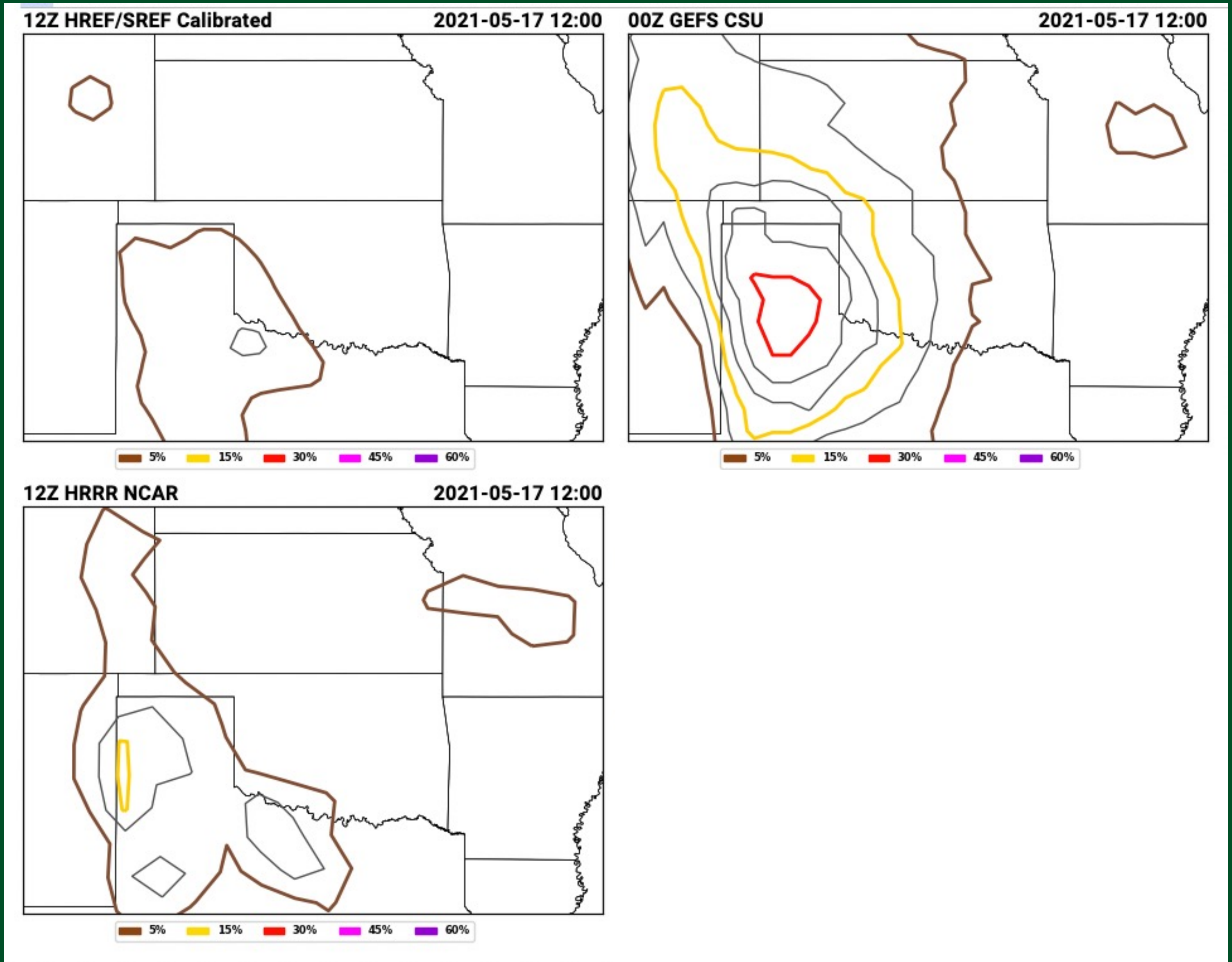
new day-3 forecast: trained on & driven by GEFS v12

SPC day-3 outlook



Day-3 forecasts made on 15 March 2021, for 24-h period ending 18 March 2021

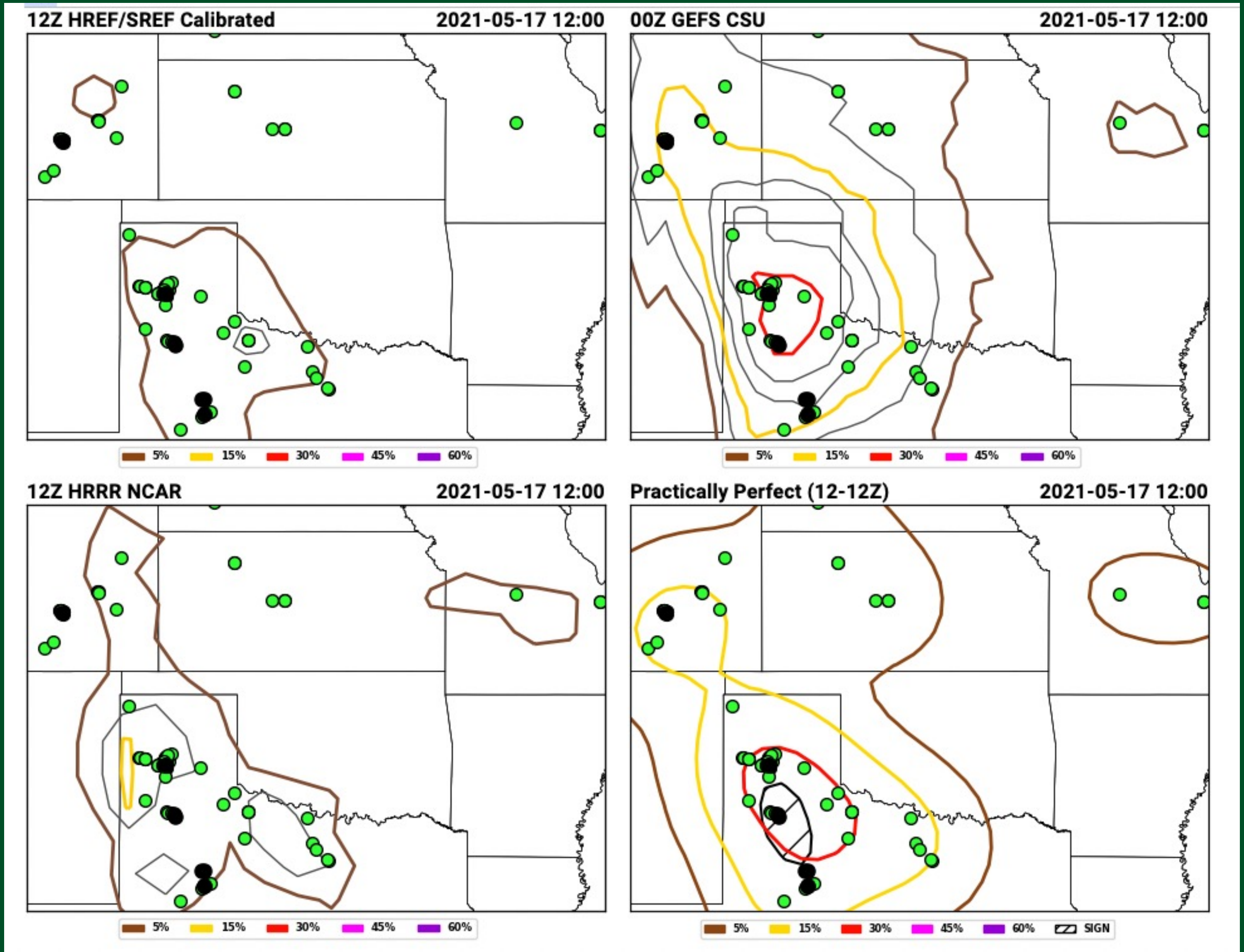
These forecasts were demonstrated and evaluated during the Hazardous Weather Testbed Spring Forecasting Experiment



Day-2 hail forecasts issued 15 May 2021, valid 16-17 May 2021

https://hwt.nssl.noaa.gov/sfe_viewer/2021/outlook_verification/

These forecasts were demonstrated and evaluated during the Hazardous Weather Testbed Spring Forecasting Experiment



Day-2 hail forecasts issued 15 May 2021, valid 16-17 May 2021

https://hwt.nssl.noaa.gov/sfe_viewer/2021/outlook_verification/

Can we push skillful guidance beyond day 3?

- We have a current project with the Storm Prediction Center, and a new project starting soon with the Weather Prediction Center (thanks to the JTTI program for continued support!)

- Objectives of both projects include developing seamless CSU-MLP guidance for days 1-8 for both severe weather and excessive rainfall

- How do machine learning forecasts perform at these longer lead times? Do we need to configure the models in a different way? (For example, to include more information about the evolution of the NWP output?)

- How could this medium-range guidance be most helpfully incorporated into SPC/WPC operations?

Other ongoing work

- Does incorporating convection-allowing models into a similar ML system offer benefits for shorter term forecasts?
  - A big limitation is that we don't have a 'reforecast' dataset for any CAMs, and they change configurations routinely
  - Results have been mixed thus far, but with some promising avenues for further exploration
  - At the Flash Flood and Intense Rainfall (FFaIR) experiment this summer, we will evaluate a blend of CSU-MLP models that include output from both the GEFS and CAMs
- Understanding synoptic patterns associated with successful and failed forecasts (graduate student Jacob Escobedo)
- Exploring other "flavors" of machine learning (convolutional neural networks, etc.) for these types of applications

Summary

- Machine learning techniques can help in post-processing NWP output to yield useful “first guess” guidance for operations
- ML models for severe weather and excessive rainfall are skillful, and competitive with operational outlooks – especially beyond day 1
- 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 for the opportunity to be here!
russ.schumacher@colostate.edu

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

Contact us if interested in gridded output!

Backup slides