Design Strategies for Skillful and Reliable Regional UFS Ensemble Forecasts

Glen Romine
NCAR

Contributions from Craig Schwartz, Ryan Sobash, and May Wong
Collaborators include Tara Jensen, David Dowell, Jacob Carley, and Curtis Alexander

UFS Webinar Series

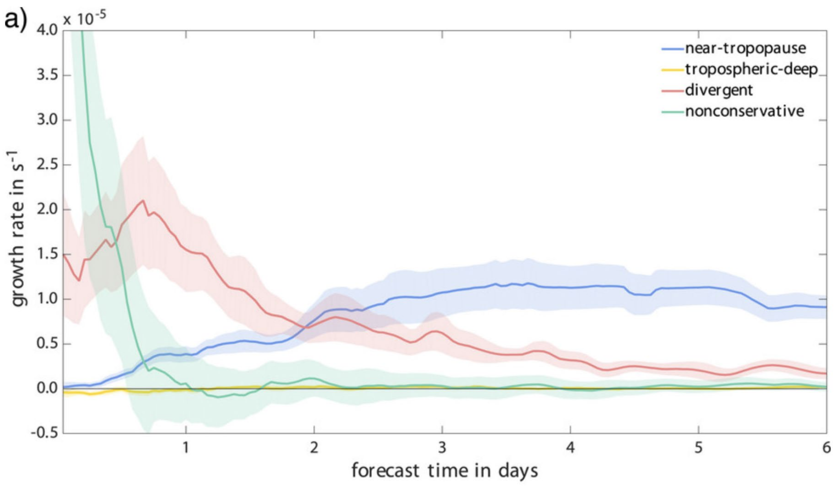
May 6, 2021

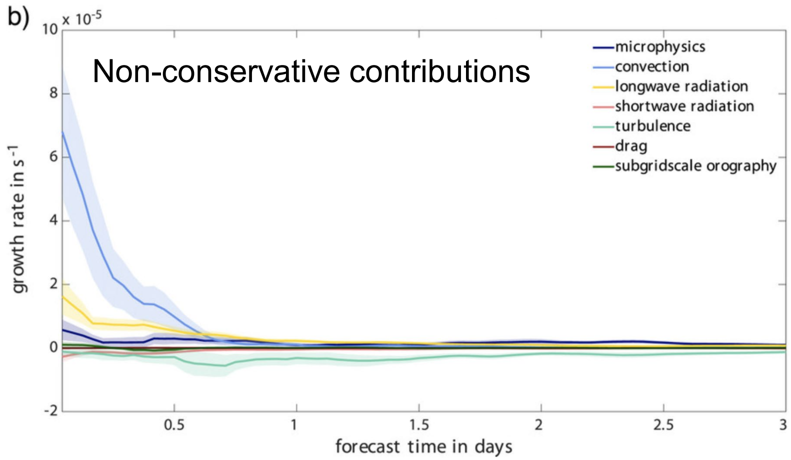
This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977.

Predictability (and prediction) of high-impact weather

Intrinsic predictability – predictability under optimal conditions

- Assumes perfect model and tiny initial condition errors
- Lorenz (1969), Lilly (1990): error growth owing non-linearity limit predictability horizon (upscale growth)
- Judt (2019): error growth (loss of predictability) is flow dependent
- **Moist physics** is the leading process family in upscale error growth (Hohenegger and Schar 2007; Baumgart et al. 2019), with limited dependence on microphysical complexity (Wang et al. 2012)





Predictability (and prediction) of high-impact weather

Practical predictability – the best we can do with current capability

- Still limited by intrinsic predictability, but also limited by an imperfect model, modern observing capability, and data assimilation methods
- Melhauser and Zhang (2012): more accurate initial conditions can lead to further improvement in prediction skill, warrants further progress in observing capability and data assimilation

“Good” Forecast member

“Bad” Forecast member

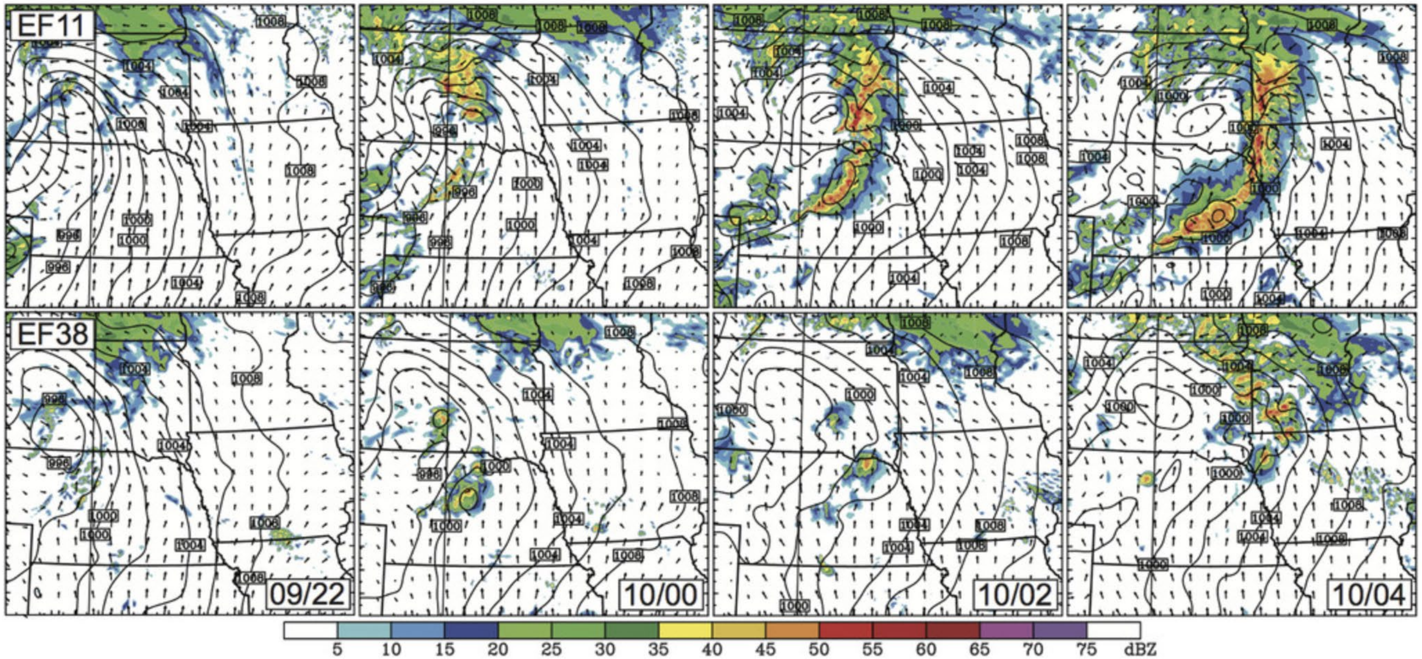


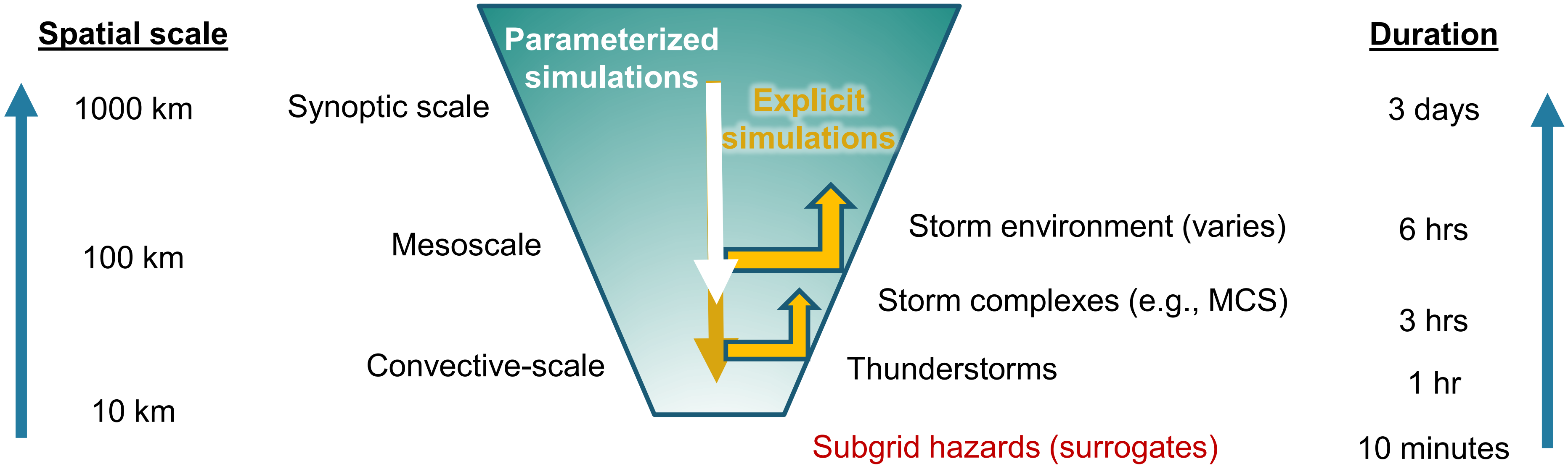
Fig. 6 from Melhauser and Zhang (2012)

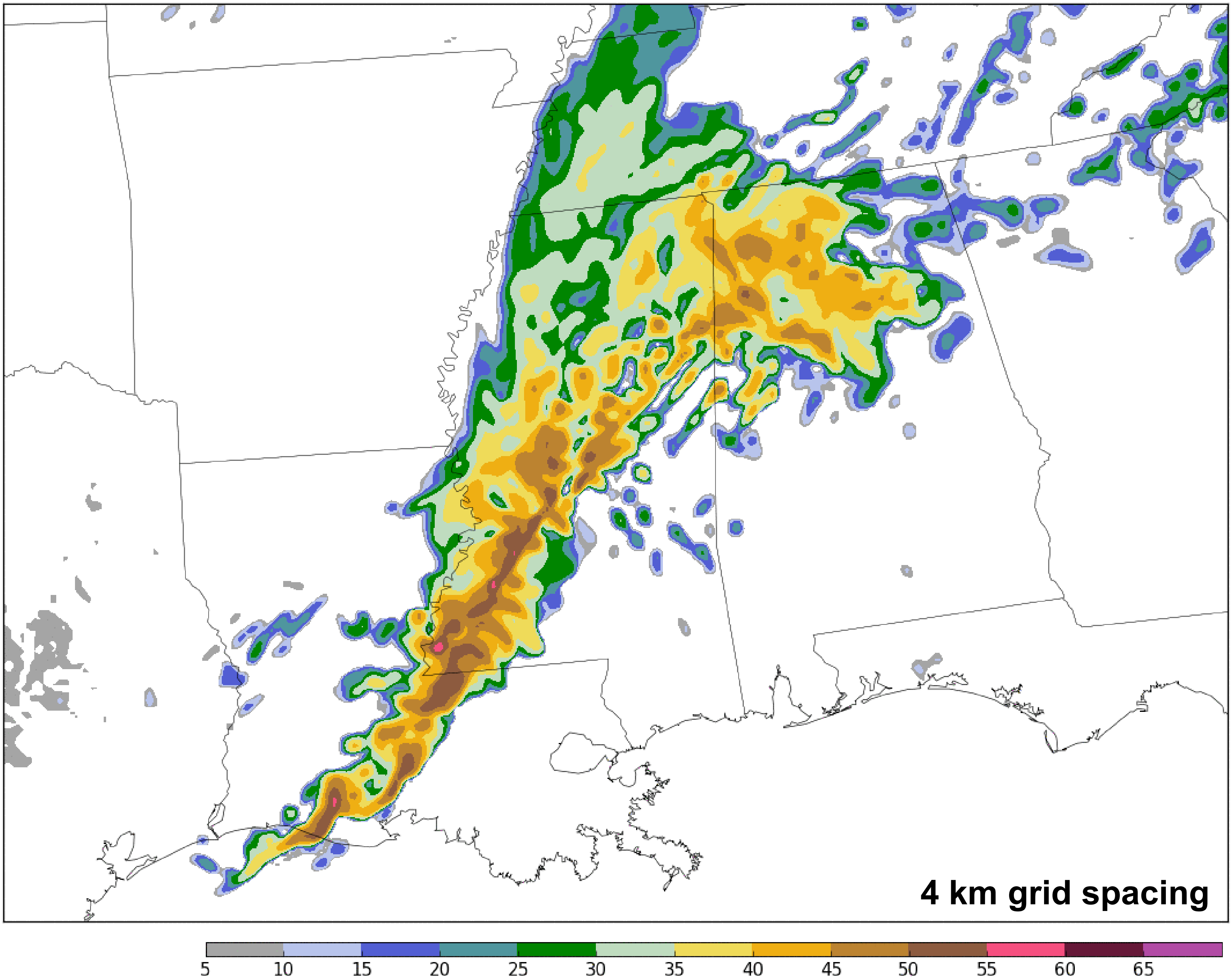
Limitations bridging practical to intrinsic predictability

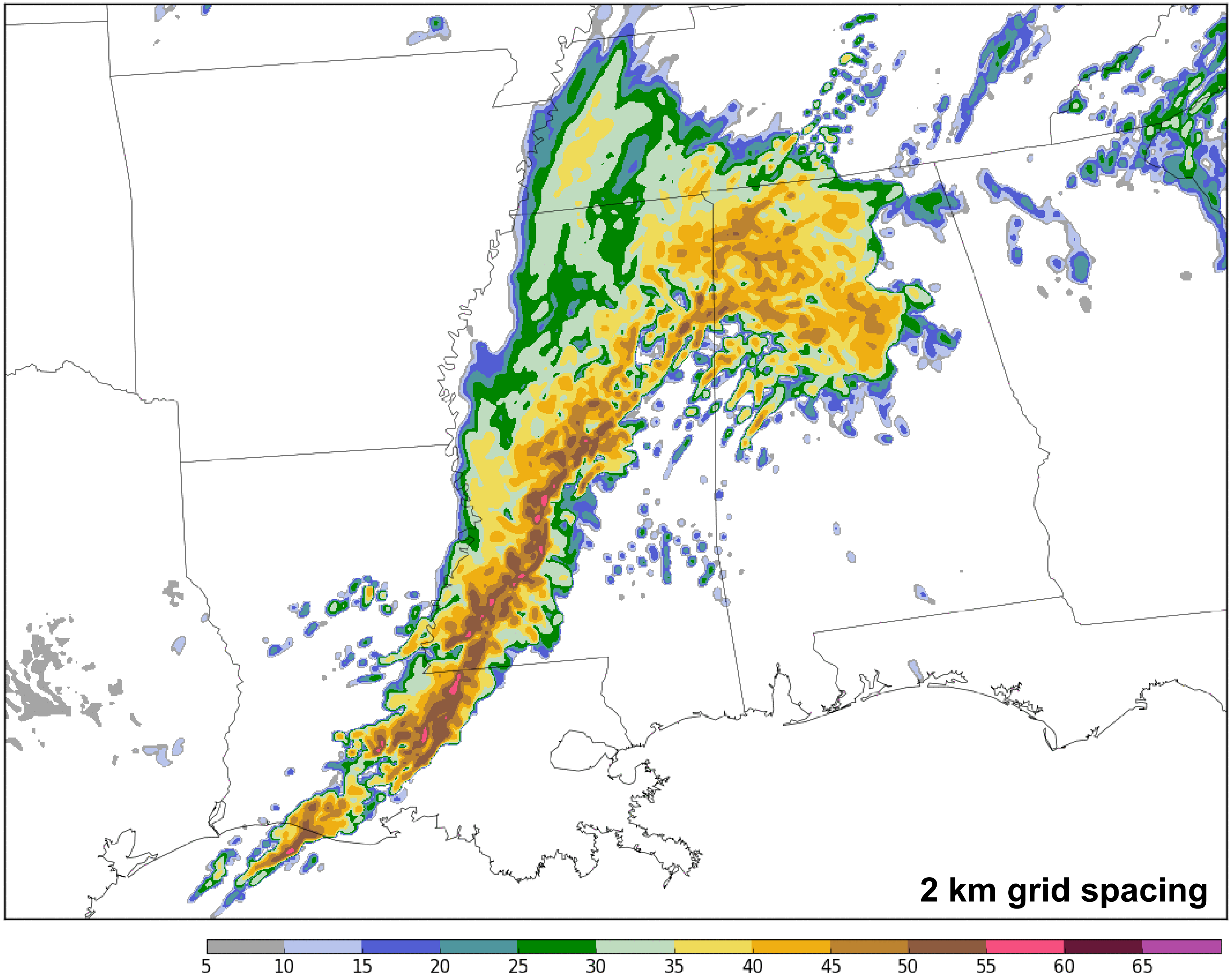
- **Imperfect observations**
  - Errors in measurements, spatial gaps in observing key features, limited temporal sampling, measurements are often not of model state variables
- **Imperfect model**
  - Simplified representation of key processes, unresolved scales, must balance model complexity with available computational resources
- **Imperfect analysis capabilities**
  - Model errors conflate in data assimilation system, simplified and imperfect assimilation methods; leads to initial condition errors, including errors in the estimate of analysis certainty

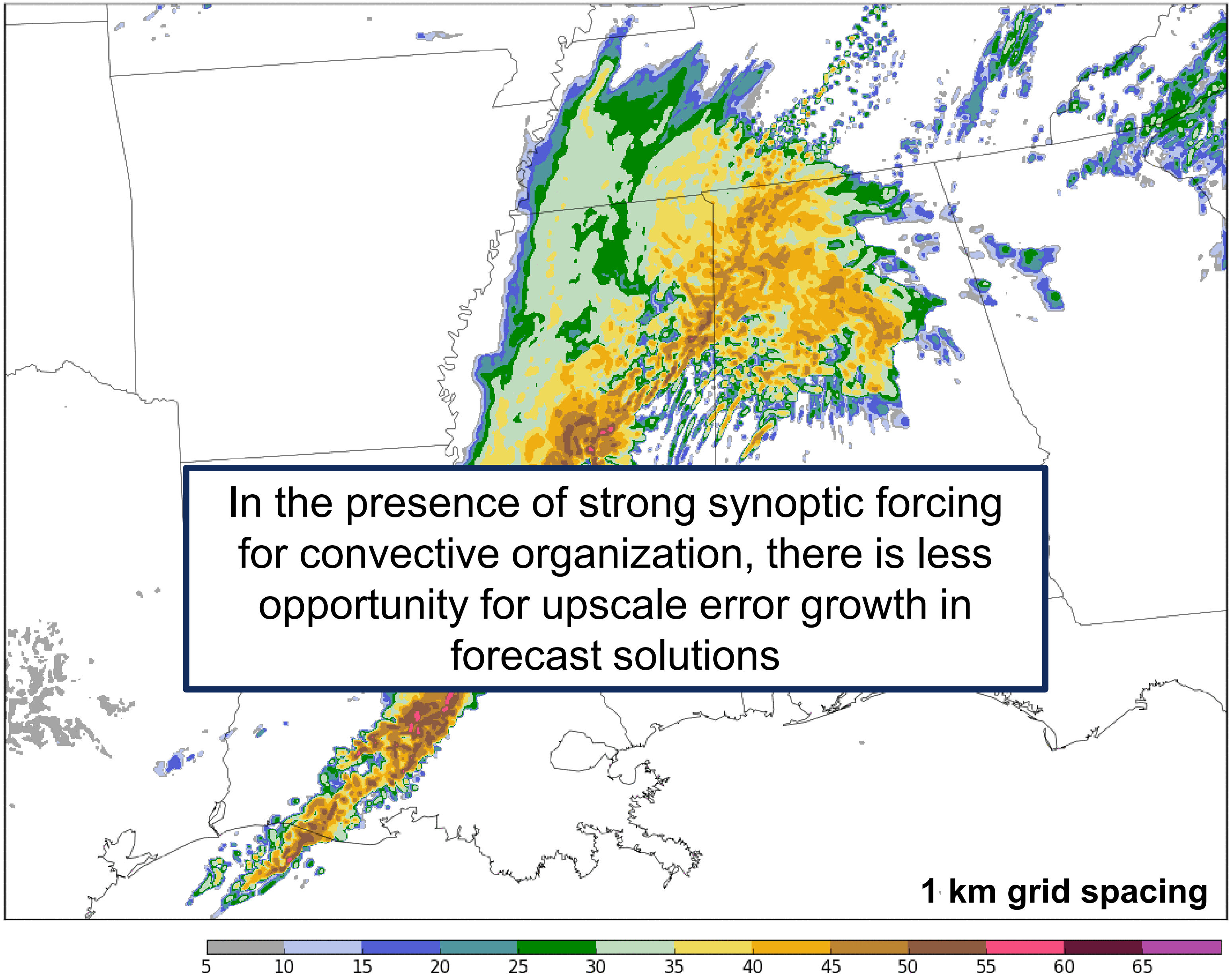
Prediction of high-impact weather

- Predictive skill for convection: storm environment and triggers
  - Even coarse models pretty good at forecasting mesoscale storm environments (well-resolved scale ~ 80 km)
  - Best practice among many operational forecasters – 'forecast funnel'
  - Explicit simulations better at representing upscale feedbacks (large errors)

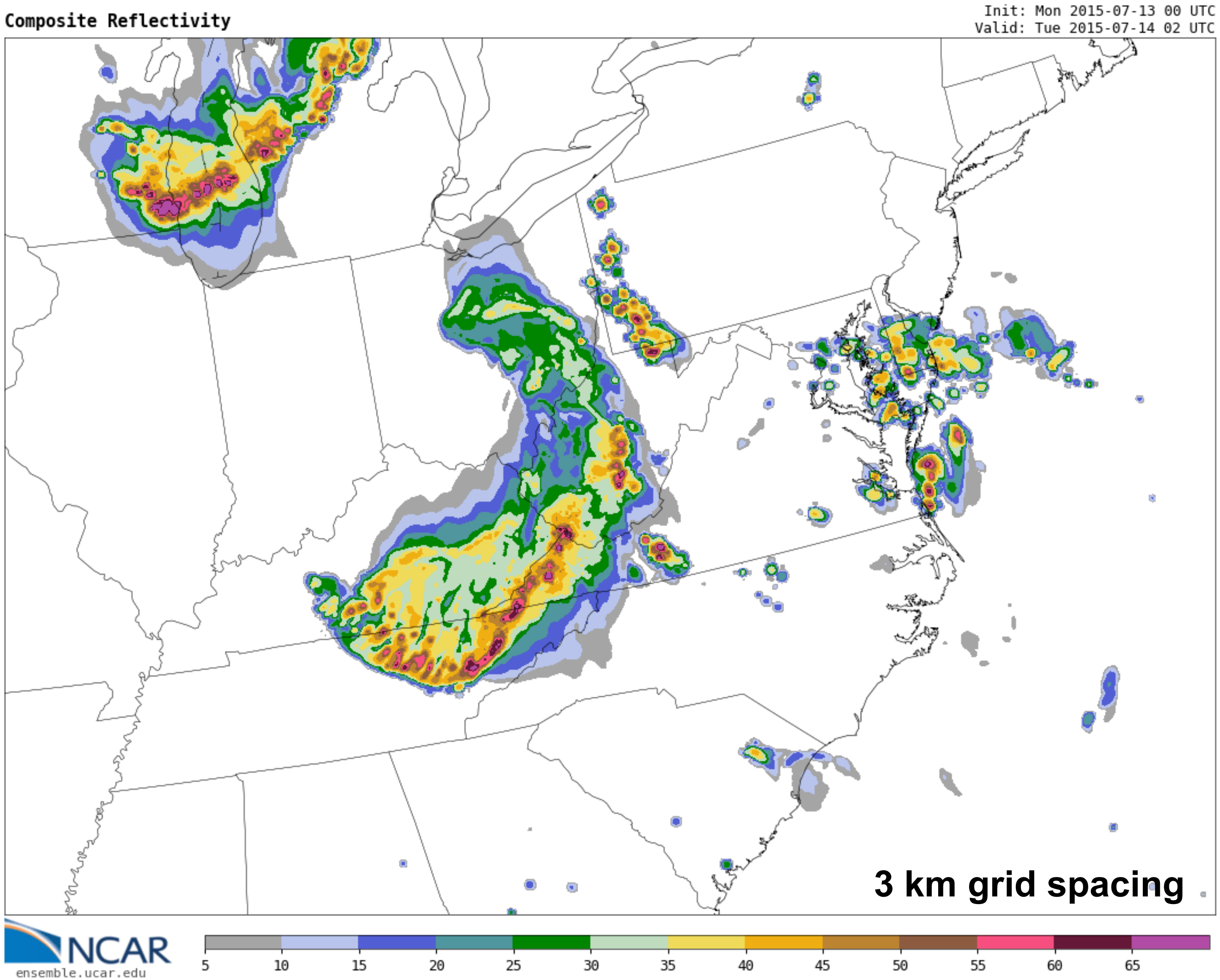


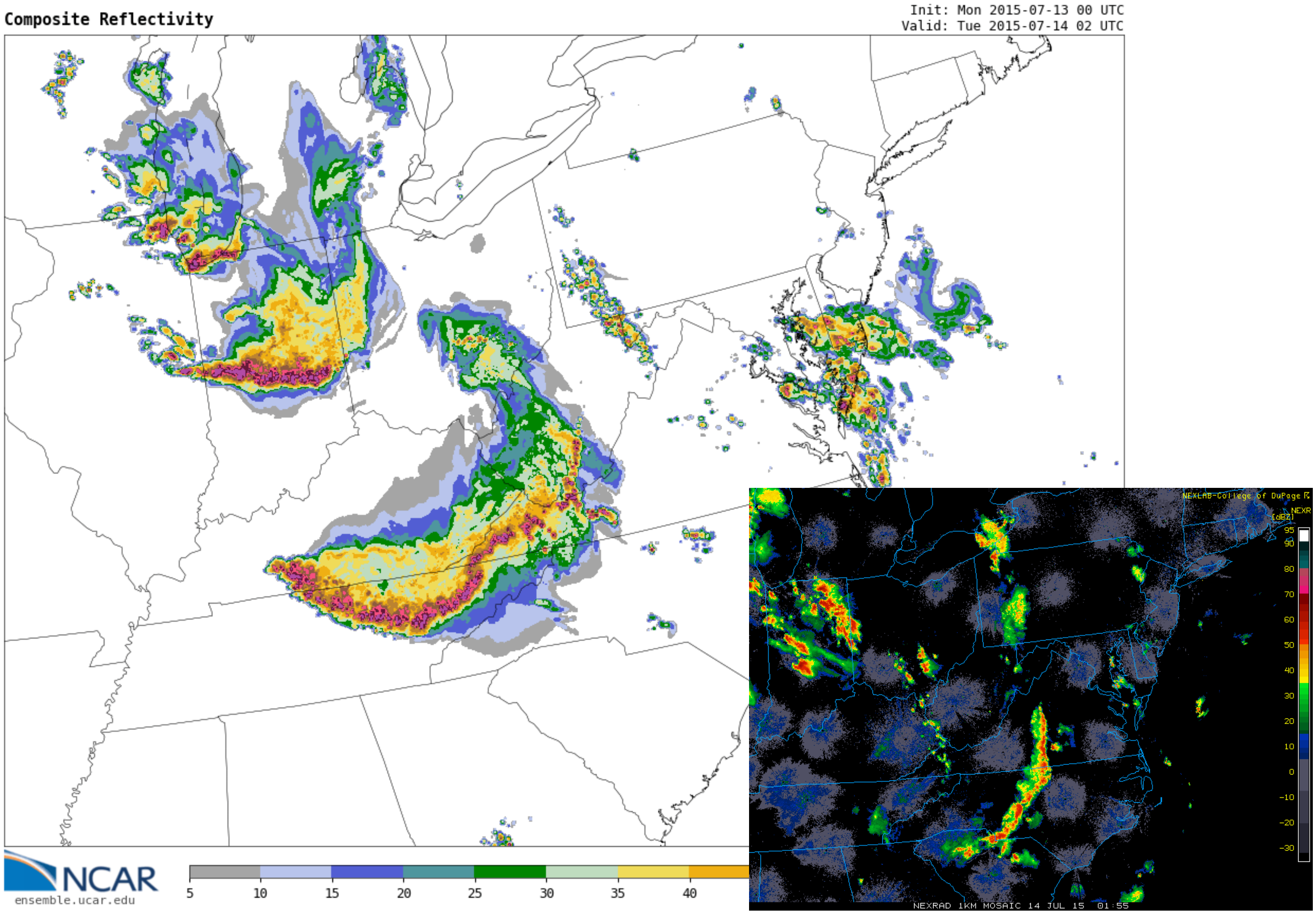

4 km grid spacing
5
10
15
20
25
30
35
40
45
50
55
60
65


2 km grid spacing
5
10
15
20
25
30
35
40
45
50
55
60
65

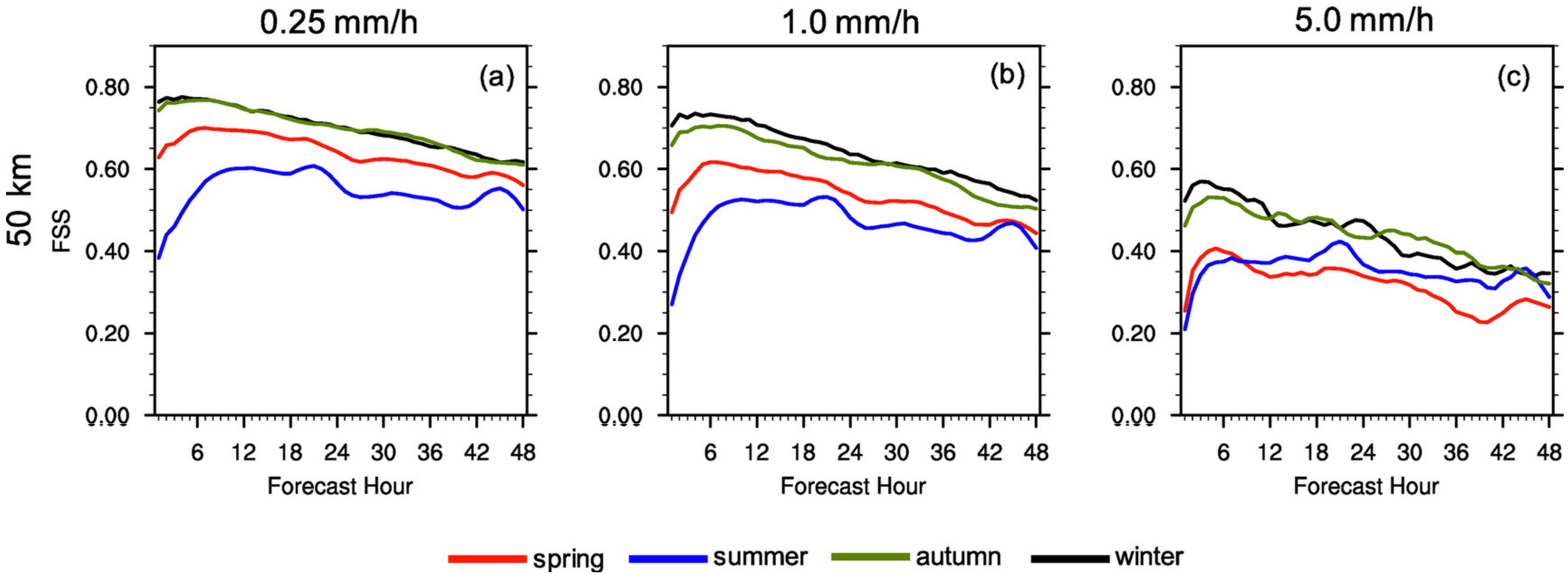


In the presence of strong synoptic forcing for convective organization, there is less opportunity for upscale error growth in forecast solutions


Composite Reflectivity
Init: Mon 2015-07-13 00 UTC
Valid: Tue 2015-07-14 02 UTC
3 km grid spacing
NCAR
ensemble.ucar.edu
5
10
15
20
25
30
35
40
45
50
55
60
65


Composite Reflectivity
Init: Mon 2015-07-13 00 UTC
Valid: Tue 2015-07-14 02 UTC
NCAR
ensemble.ucar.edu
5
10
15
20
25
30
35
40
NEXLAB-College of DuPage
95
90
80
70
60
50
40
30
20
10
0
-10
-20
-30
NEXRAD 1KM MOSAIC 14 JUL 15 01:55

Seasonality of skillful CAM predictions



- Ensemble explicit forecasts are more skillful during seasons with strong synoptic forcing, as predictable features on the mesoscale drive initiation

- Forecast skill degrades with increasing rainfall intensity and during summer where foci for convective development are less skillfully predicted

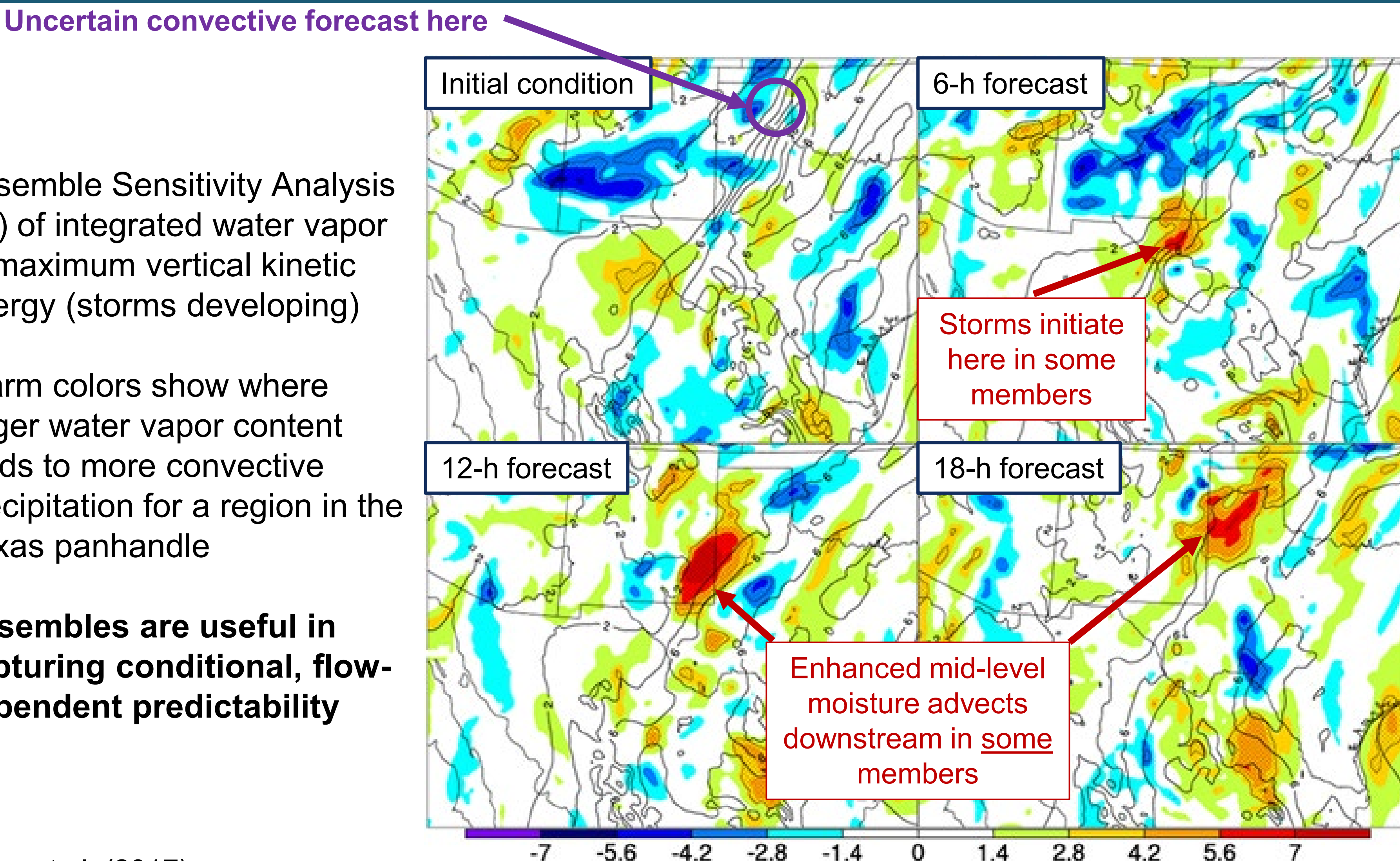
Schwartz et al. (2019)

Why ensembles?

Ensemble Sensitivity Analysis (fill) of integrated water vapor to maximum vertical kinetic energy (storms developing)

Warm colors show where larger water vapor content leads to more convective precipitation for a region in the Texas panhandle

Ensembles are useful in capturing conditional, flow-dependent predictability



Torn et al. (2017)

Incentives toward CAM ensemble R&D

NOAA, through the NGGPS, is moving toward a unified forecast system (UFS) to simplify the production suite

Opportunities:

- Concentrate efforts in common shared model environment
- Share physics between global and regional configurations, e.g. CCPP
- Eventually, a coupled model framework (e.g., CIME)

Challenges:

- Forklift change of several core forecast system components underway: dynamic model, (some) physics, and DA system (JEDI)
- Future Rapid Refresh Forecast System (RRFS) is envisioned to include a convection-permitting (CP) ensemble analysis and forecast system with a <u>single dynamic core (FV3) and common physics suite</u>
- Current systems based on WRF with GSI EnKF, or ad hoc conglomerates of deterministic CP forecasts (HREF)
- Best practice in CAM ensemble design is not yet well defined

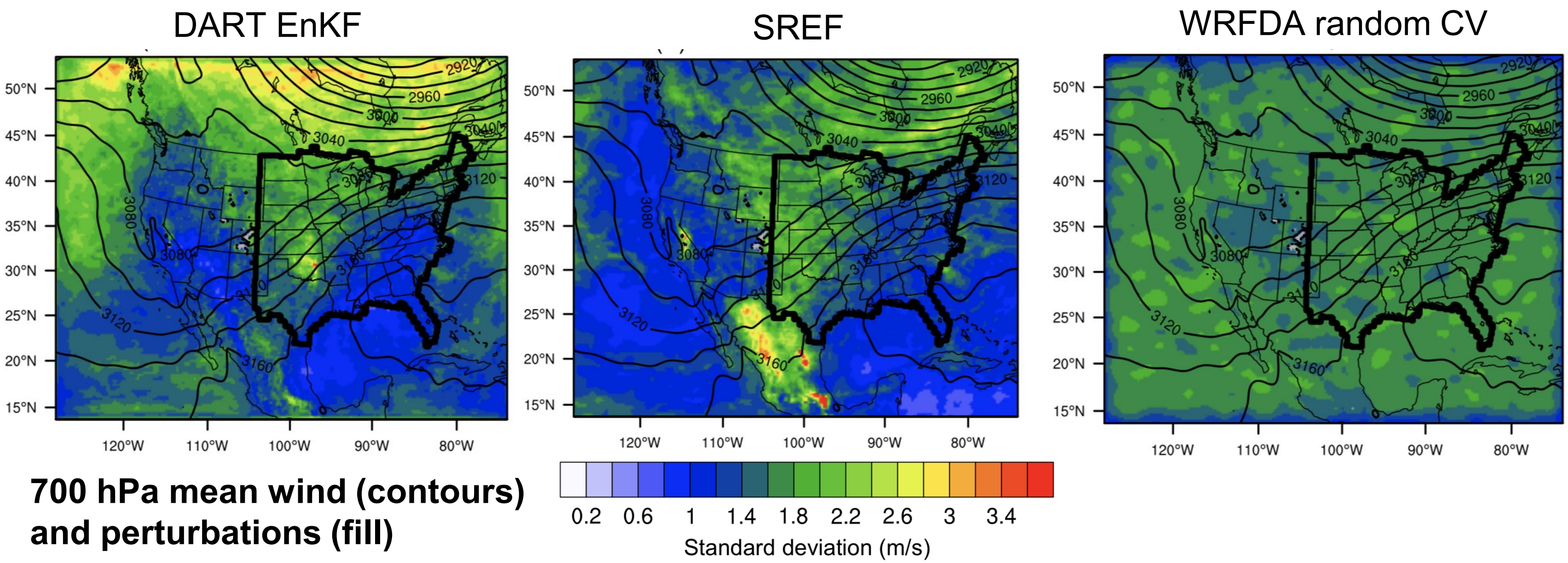
Select NCAR contributions to build skillful CAM ensembles

- NCAR ensemble forecast demonstration system (2015-2017)
- Participation in NOAA testbeds with experimental forecast systems and products (2015-2021)
- Horizontal grid spacing dependence for analyses and forecasts
- **EnKF based initial perturbation ensembles with single dynamic core and physics**
- **Novel methods for tracing spread-error consistency**
- Reducing systematic model errors in continuously cycled regional DA, including:
  - time-averaged initial tendency method to trace model error
  - **High vs. low resolution ensemble analysis**
- **Global analysis blending to enable continuous cycling with simpler workflow**
- **Post-processing to increase the usability of CAM ensemble forecasts**

NOAA sponsored research key catalyst for much of this work!

Initial perturbation ensembles - climatology

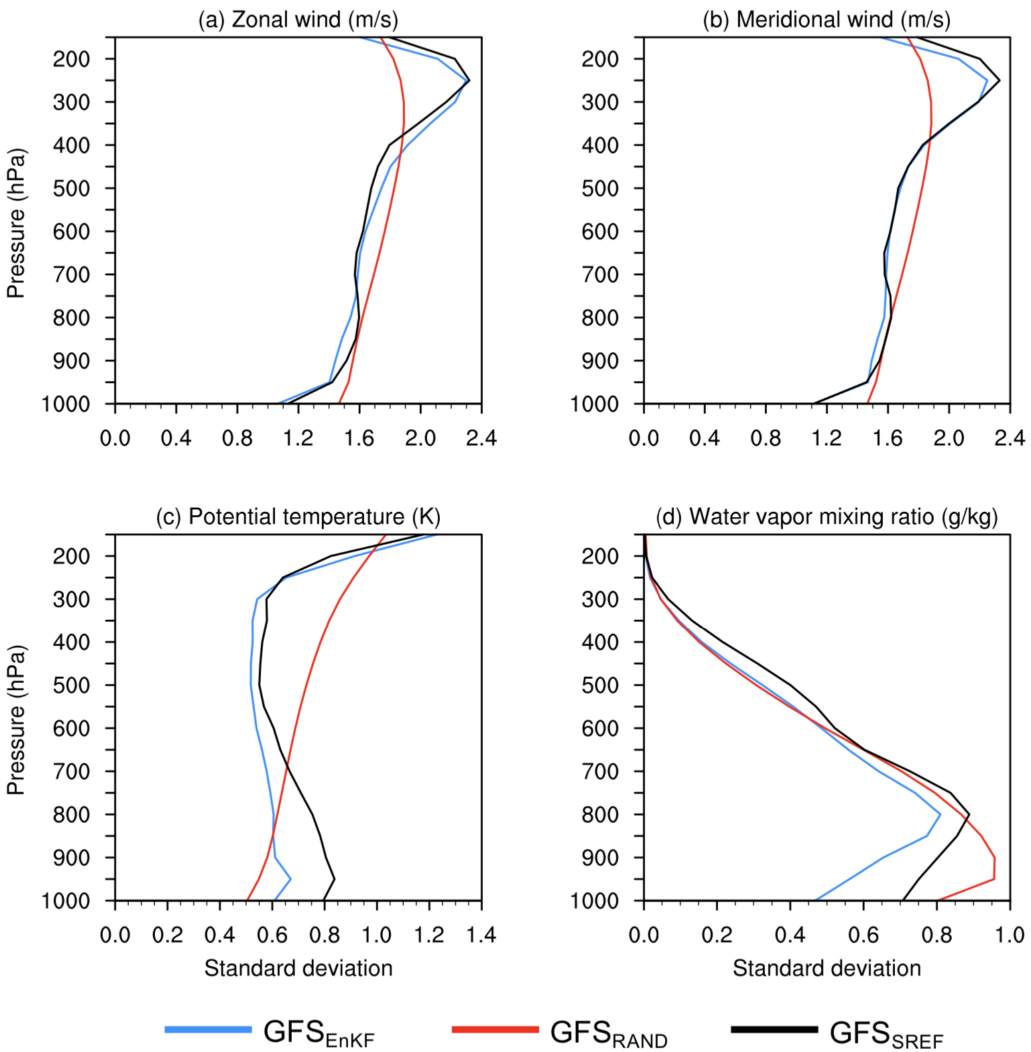
- Consider impact on ensemble dispersion from different initial perturbation sources



700 hPa mean wind (contours) and perturbations (fill)

- DART EnKF analysis and 3-h lead SPC's Short-Range Ensemble Forecast (SREF) give pseudo-flow-dependent perturbations
- Random correlated errors are drawn from WRFDA
- See Schwartz et al. (2020)

Initial perturbation ensembles – vertical structure



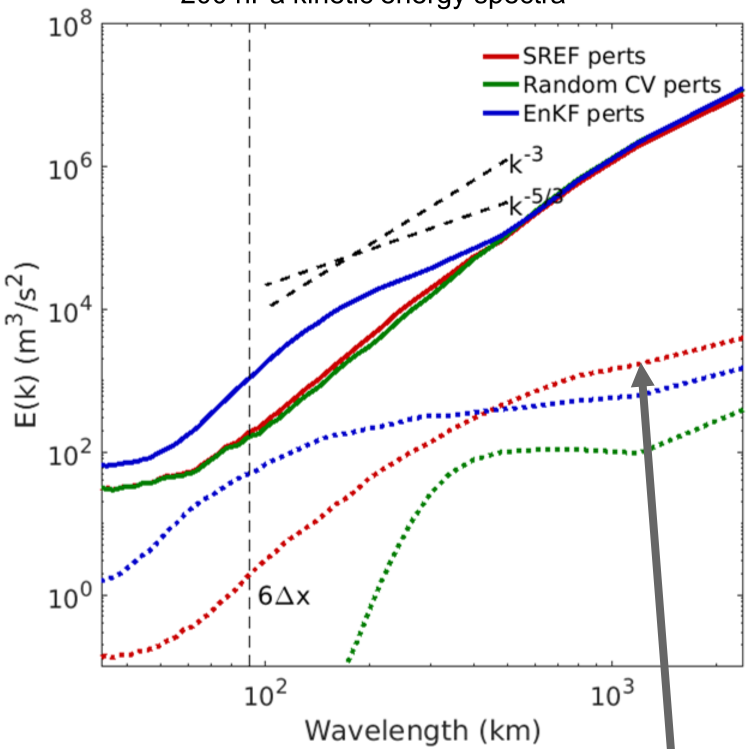
Random error amplitudes were most unique in structure: larger low-level moisture perturbations, smallest wind perturbations at jet level

EnKF perturbations were smaller amplitude for temperature and moisture

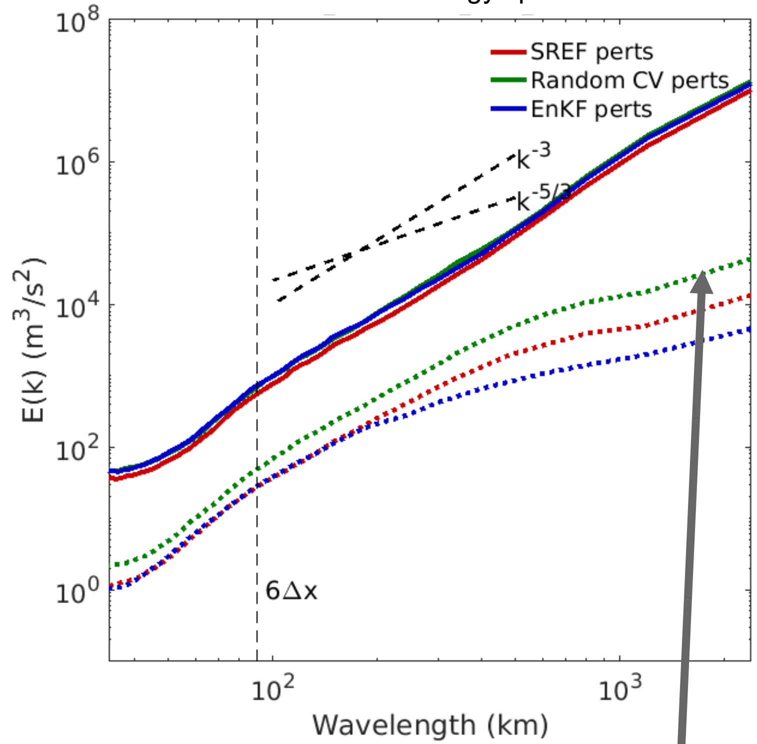
Initial perturbation ensembles – kinetic energy spectra

Analysis state

Solid - total energy
Dashed – perturbation energy



12 hr forecast state

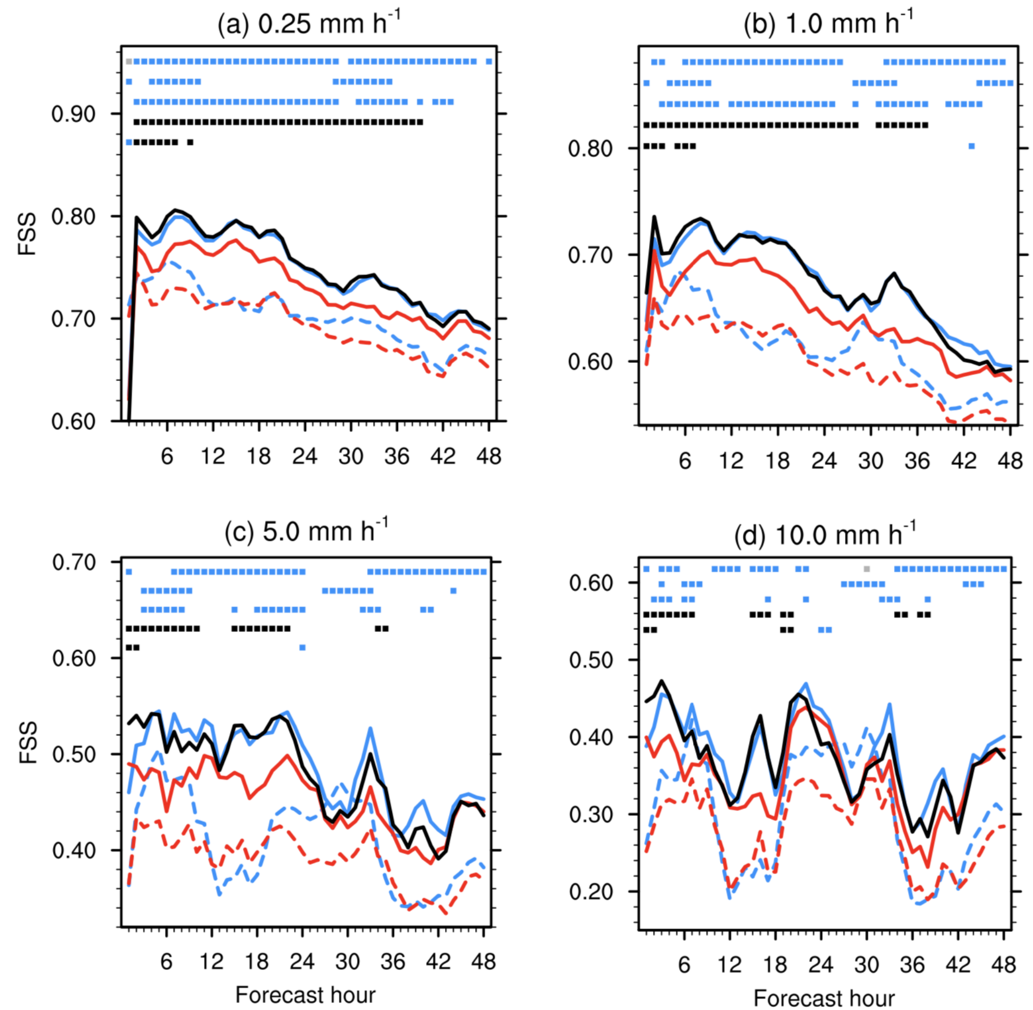


SREF has greater perturbation variance than **EnKF** at large scales, likely due to systematic errors from the different dynamic core and physics perturbations.

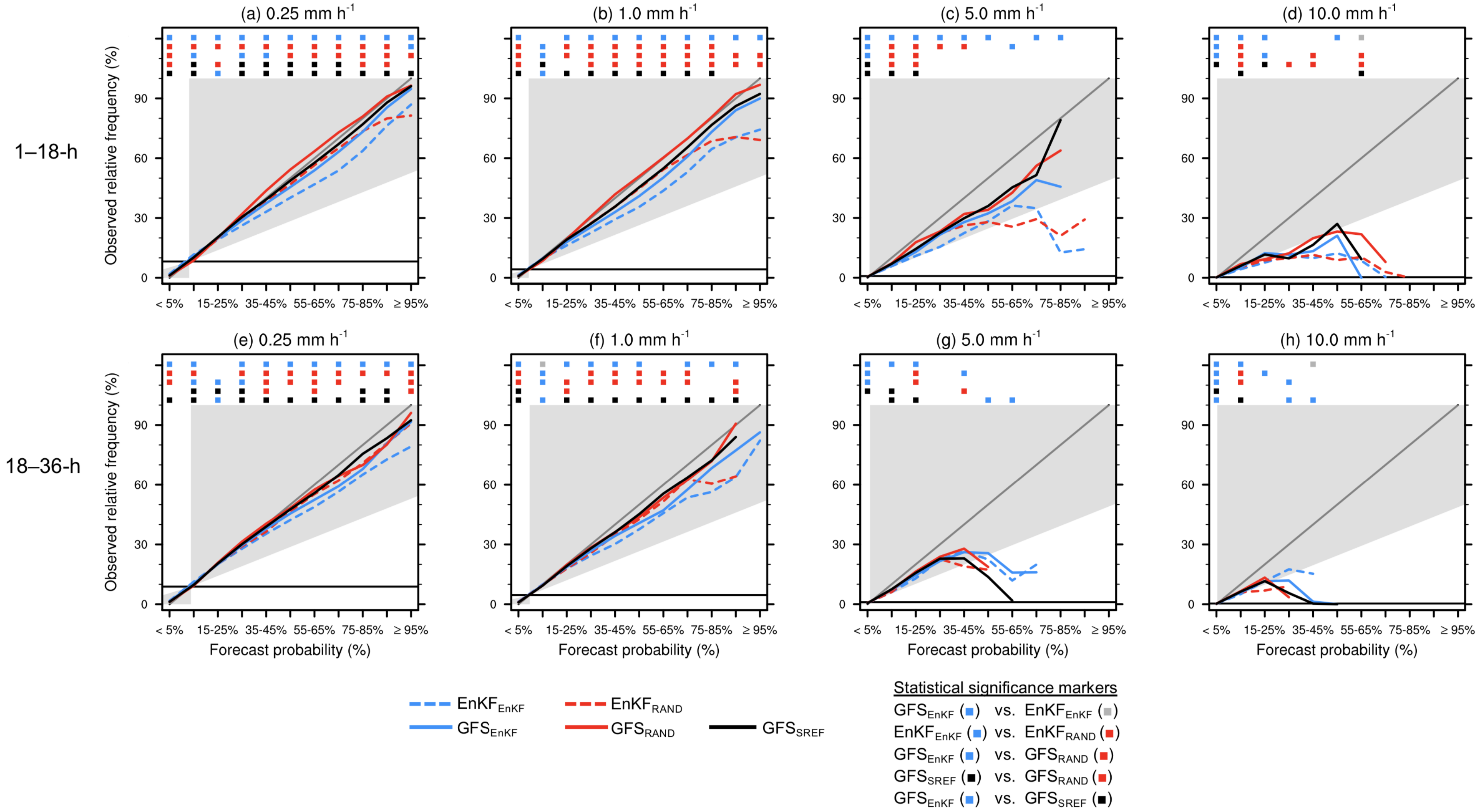
By 12 hrs, **Random CV** perturbations grow faster than **SREF** and **EnKF** perts at **all** scales. However, the forecasts with only Random CV perturbations are **less skillful.**

Initial perturbation ensembles – skill by perturbation type

- Fractions skill score – higher values means greater skill
- Largest skill benefit comes from using a more skillful mean analysis
- SREF and EnKF perturbations were similarly skillful when using the same mean analysis
- Random perturbations generally degraded the forecast skill, regardless of the perturbation source



Initial perturbation ensembles – skill by perturbation type



- Forecasts varied in dispersion from EnKF (least dispersive). SREF, to random perturbations (most dispersive)
- Improving mean forecast trajectory (dashed vs. solid) boosts reliability more than perturbation approach

Wavelet analysis for spread/skill relationship

Looking at the ratio of mean squared error and spread to find the 'central scale' (compare spread spatial scale to error spatial scale)

For every grid point (*i,j*):

$$C_{i,j} = \sum_k \left(\lambda_k w_{i,j,k}{}^2 \right) / \sum_k w_{i,j,k}{}^2$$

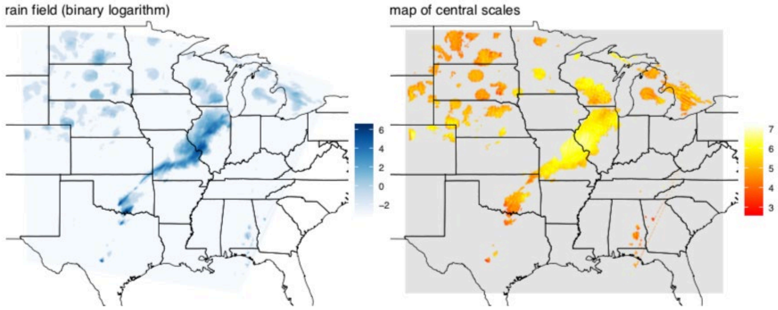
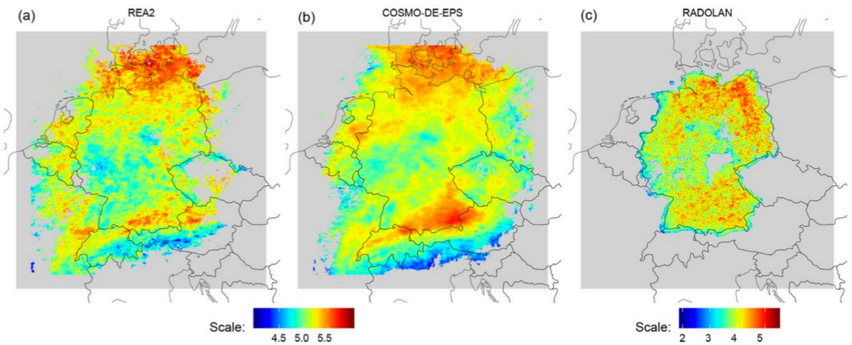


Figure 3. Logarithmized rain field and corresponding map of central scales from the stage II reanalysis on 26-04-2005 as used by Ahijevych et al. (2009) and contained in the SpatialVx-package. The field has been cut and padded with zeroes to 512×512, scales were calculated using the least asymmetric D_4 wavelet, only locations with non-zero rain are shown.

Buschow et al. (2019)



Buschow and Friedrichs (2020)

Seeking flow-dependent spread that is similar in spatial scales to the RMS error spatial scales

Wavelet analysis for spread/skill relationship

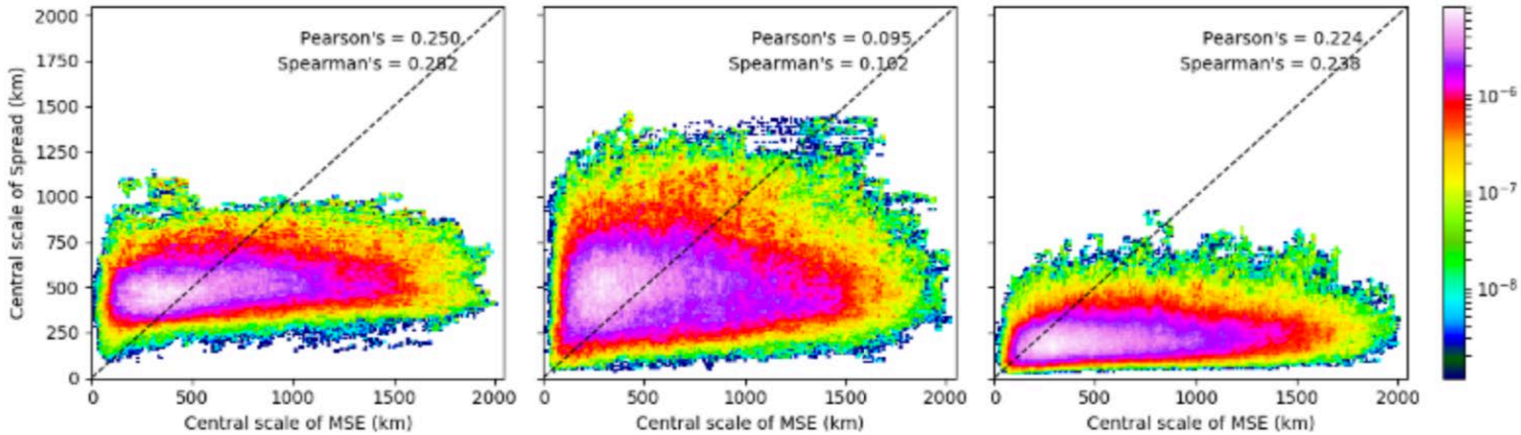


FIG. 4. Joint probability density distributions of the central scales of MSE and spread for the three initial perturbation types (a) ENS-RAND, (b) ENS-SREF, and (c) ENS-DART.

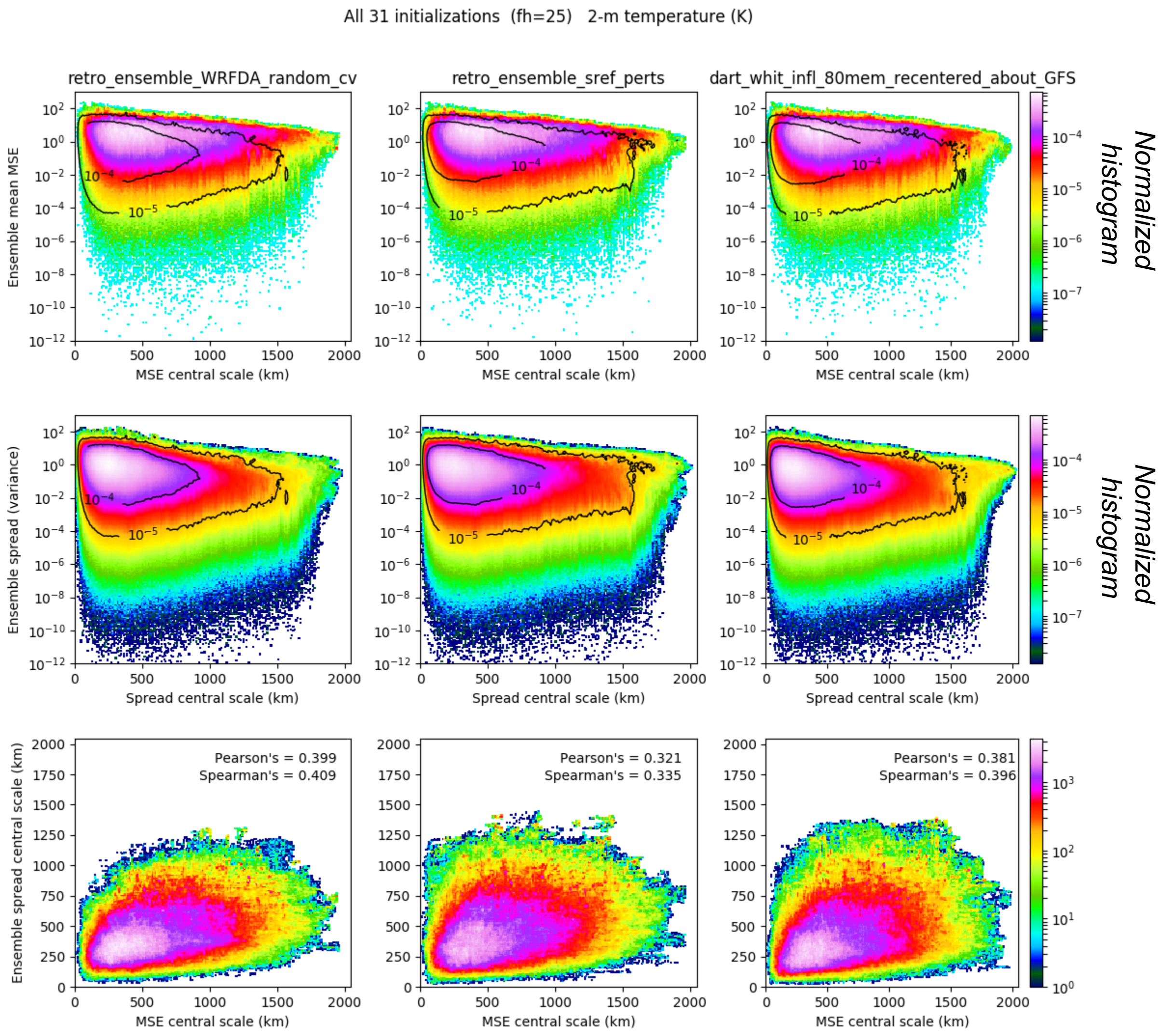
Generally poor correlations between initial perturbations and errors for all types – errors dominated by systematic errors not captured in perturbations. EnKF has small range of initial central scale spread.

Wavelet analysis for spread/skill relationship

2-m temperature

All member spread power spectra

Mean central scale of member spread power vs. central scale of ensemble MSE

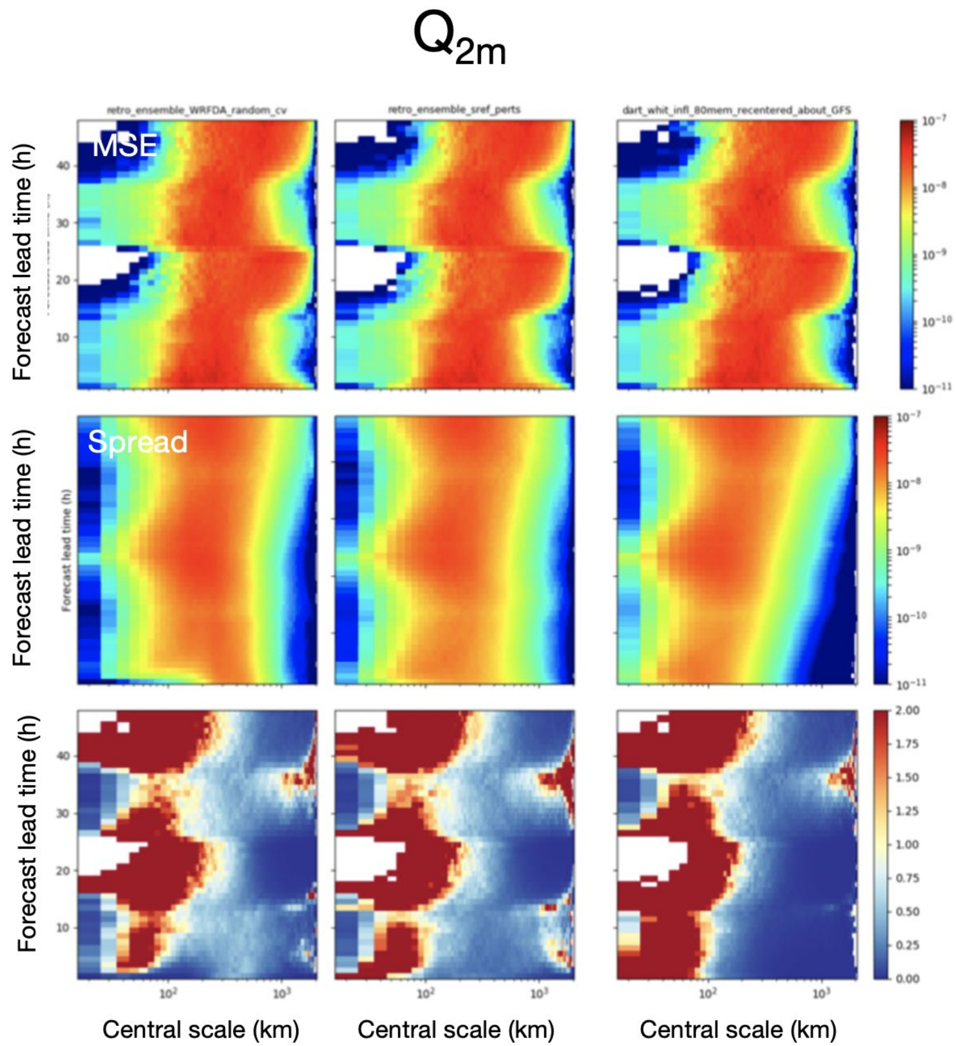


Wavelet analysis for spread/skill consistency

Skill

Spread

Consistency ratio



<u>Ensemble struggles</u>:
- Diurnal variability
- Under-dispersive at larger scales, over-dispersive at smaller scales

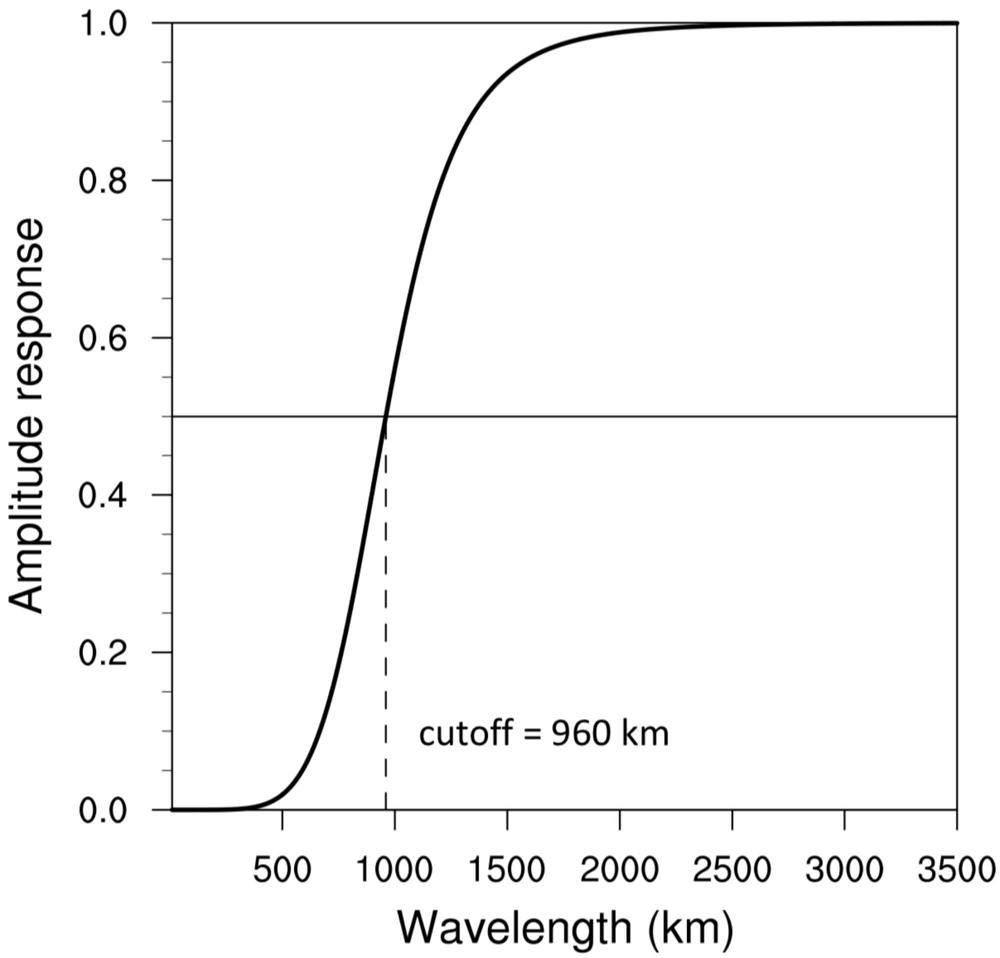
<u>Positives:</u>
EnKF perturbations show consistent error growth with increasing lead time

High resolution ensemble analysis

- Explore whether finer horizontal grid spacing for the ensemble analysis leads to more skillful subsequent forecasts
  - Explicit representation of convection
  - Eliminate downscale errors
- Massive change in computational demands:
  - 3-km 80-member ensemble analysis over full CONUS
  - Beyond available NCAR resources for real-time full experiment
  - Hourly cycling, compare 15- and 3-km grid spacing analyses to initialize forecasts
- Additional test on impact of radar reflectivity assimilation
  - Little value beyond the first few hours, not shown here

Schwartz et al. (2020)

High resolution ensemble analysis



Test blending with GEFS at large scales for initial conditions

Speculation that analysis quality degrades within regional model with continuous cycling

High-resolution ensemble analysis and blending

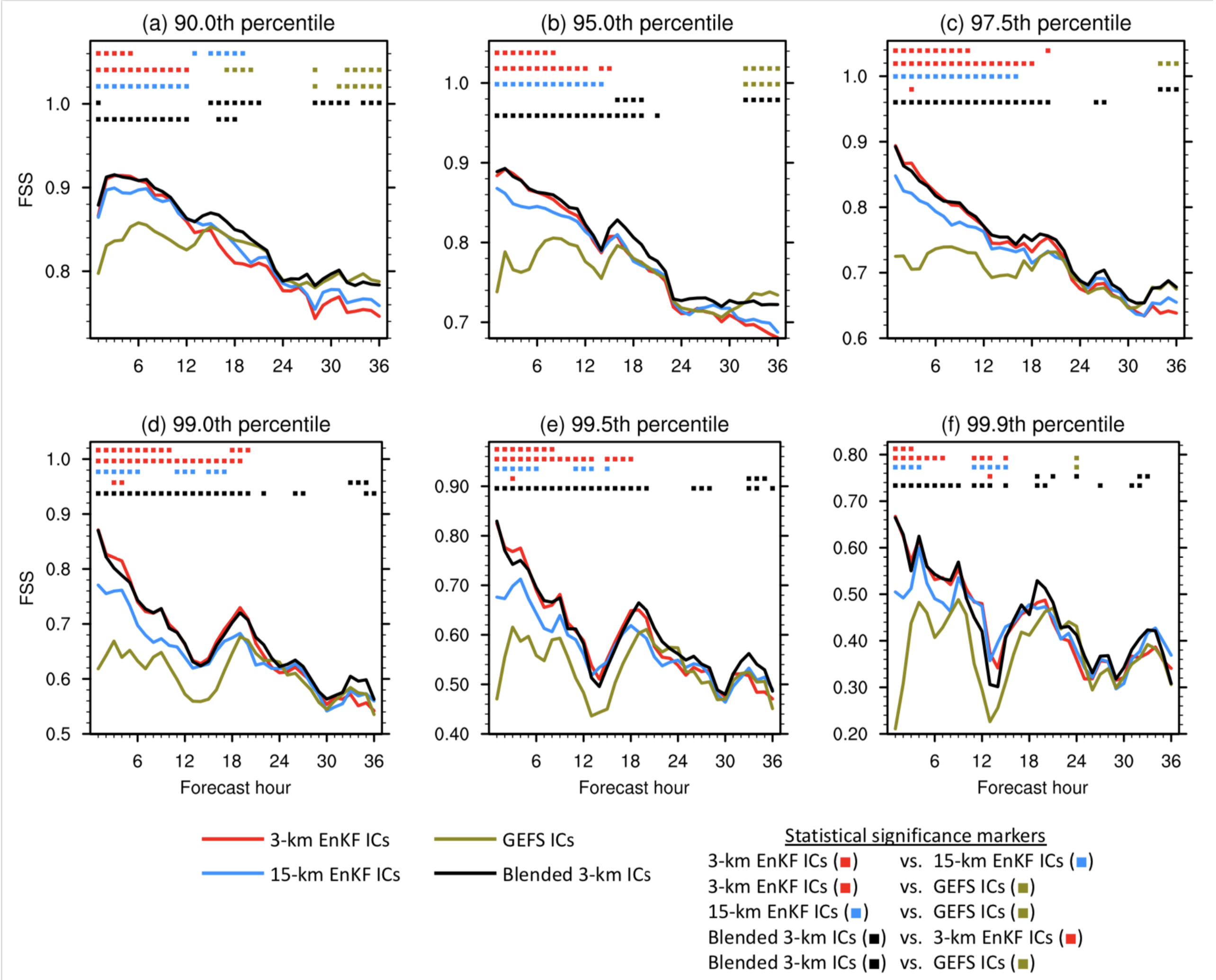
FSS – larger values indicate greater forecast skill

GEFS – more skillful at long lead (> 24 hrs)

15-km EnKF – more skillful at short lead (< 18 hrs)

3-km EnKF – more skillful than 15-km through first 12 hours

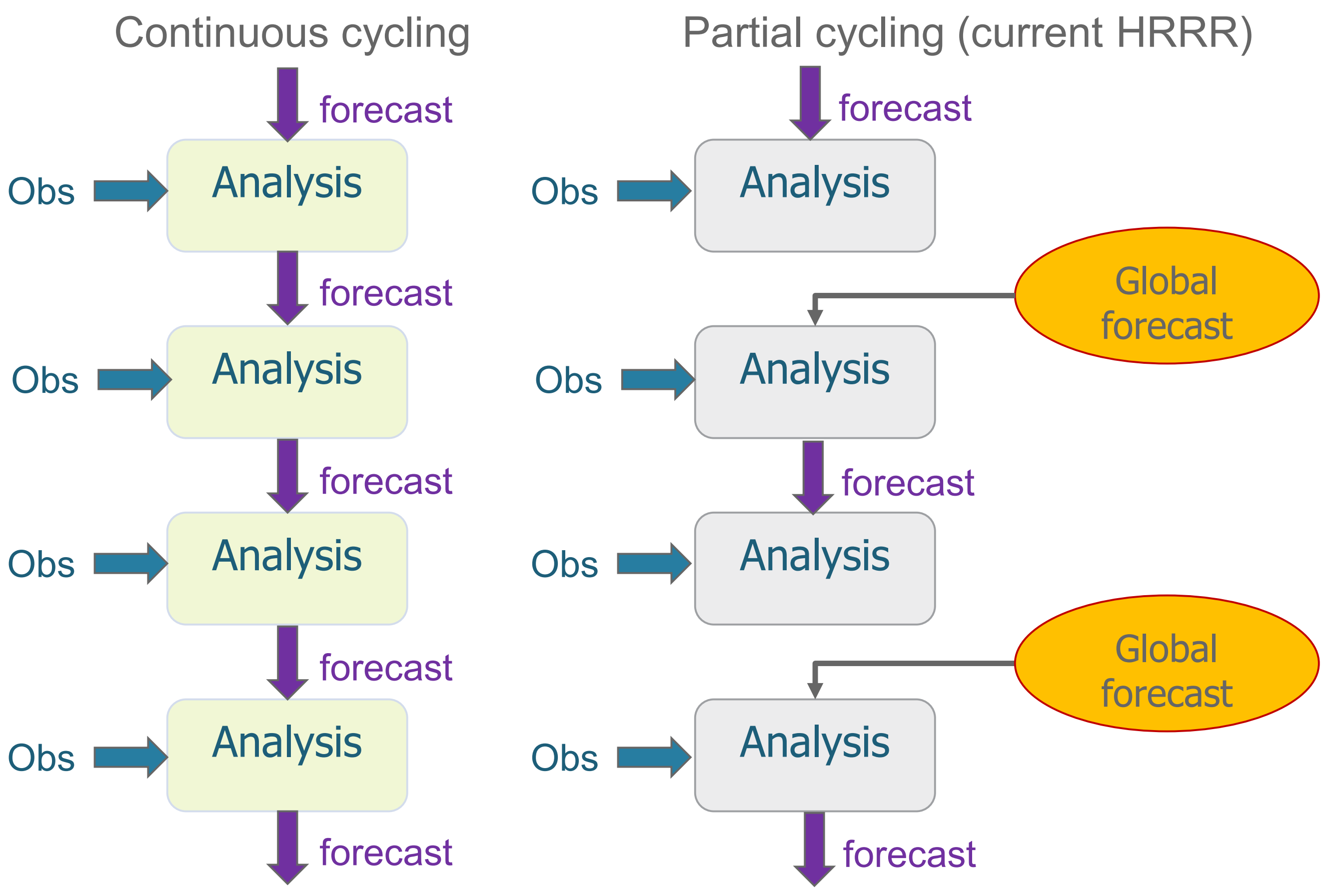
Blended – combine large-scale GEFS and 3-km, improved short and long lead skill



Schwartz et al. (2020)

Added value from convection-permitting analysis extends ~ 12 hours into the forecast

Partial versus Continuous Cycling



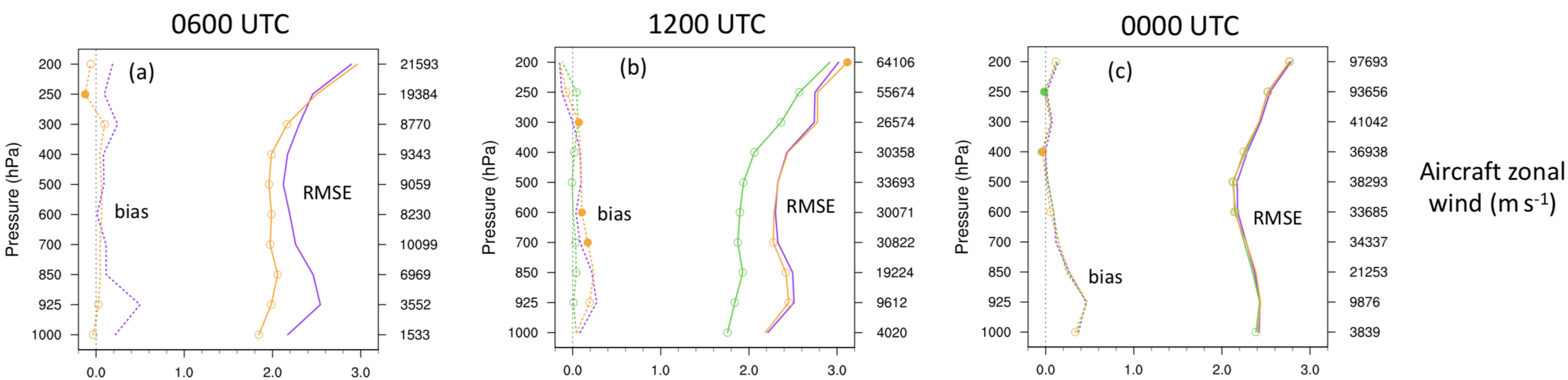
Partial cycling reduces drift (bias) owing to:

- Physics climate
- Observation availability
- Regional model boundary errors

Additional workflow burden

Partial versus Continuous Cycling

- Hourly cycling, springtime CONUS domain
- Partial cycling – external model conditions have very little bias and smaller RMS error
- After 12-18 hours of cycling, little apparent difference in bias and RMS error near observation sites

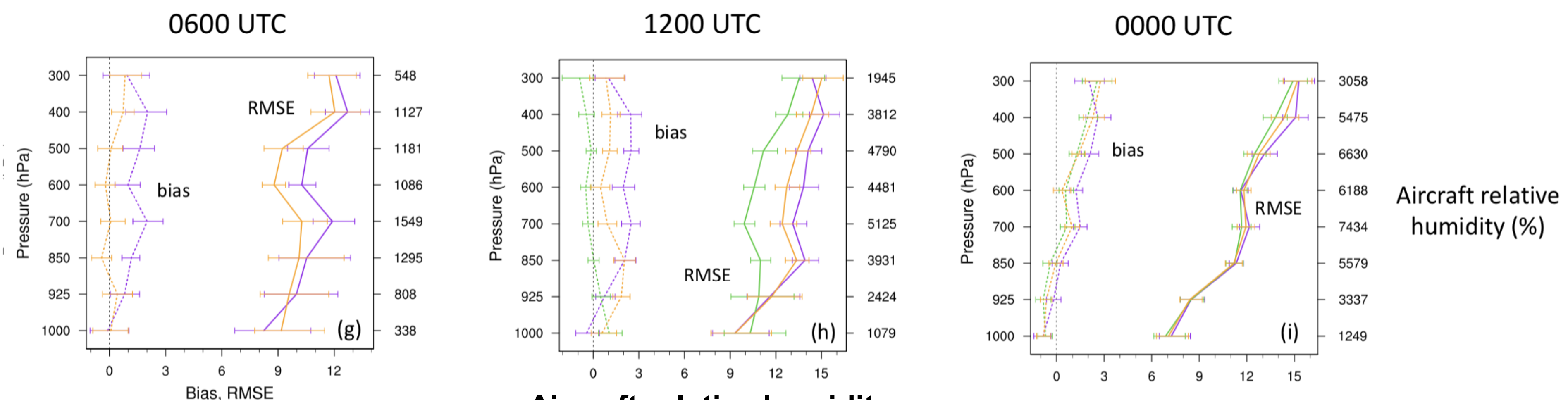


Aircraft zonal wind

Continuously cycled (CC), 06 UTC initialized partial cycle and 12 UTC initialized partial cycle from GEFS

Partial versus Continuous Cycling

- Hourly cycling, springtime CONUS domain
- Partial cycling – external model conditions have very little bias and smaller RMS error
- After 12-18 hours of cycling, little apparent difference in bias and RMS error near observation sites



Aircraft relative humidity

Continuously cycled (CC), 06 UTC initialized partial cycle and 12 UTC initialized partial cycle from GEFS

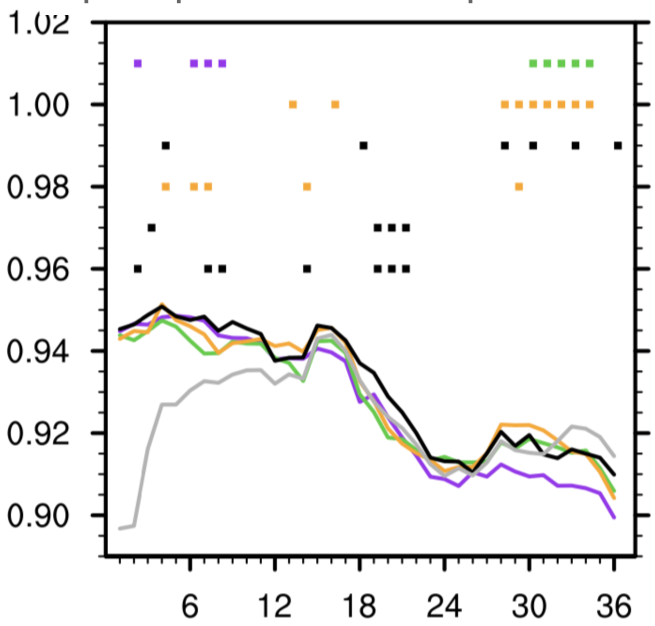
Partial versus Continuous Cycling

Continous cycling forecast skill degrades beyond 24 hours

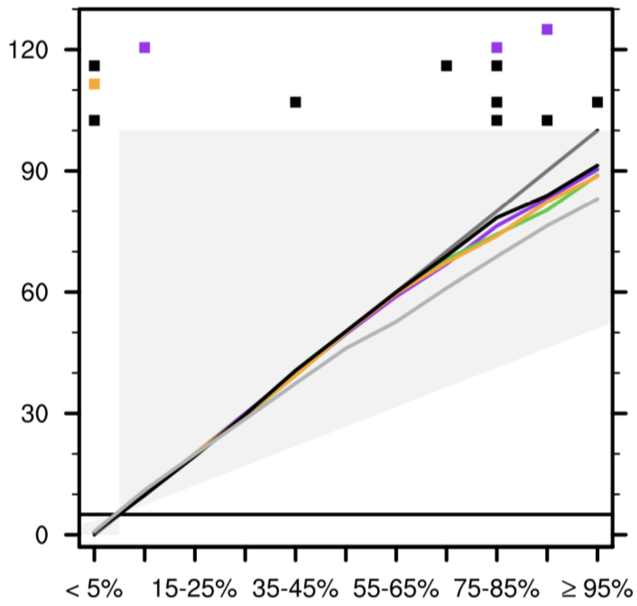
Partial cycling and blending yield similar performance

Recommendation:
Employ blending in place of partial cycling for comparable results but simplified workflow

Area under the ROC curve for precipitation at 95th percentile

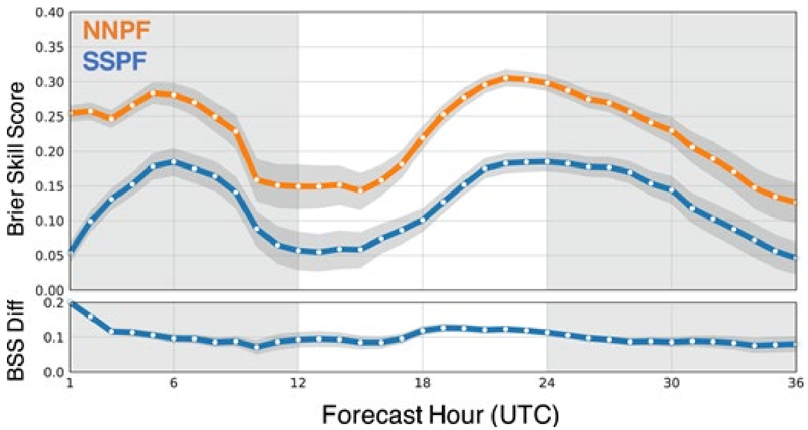


Reliability diagram for precipitation at 95th percentile

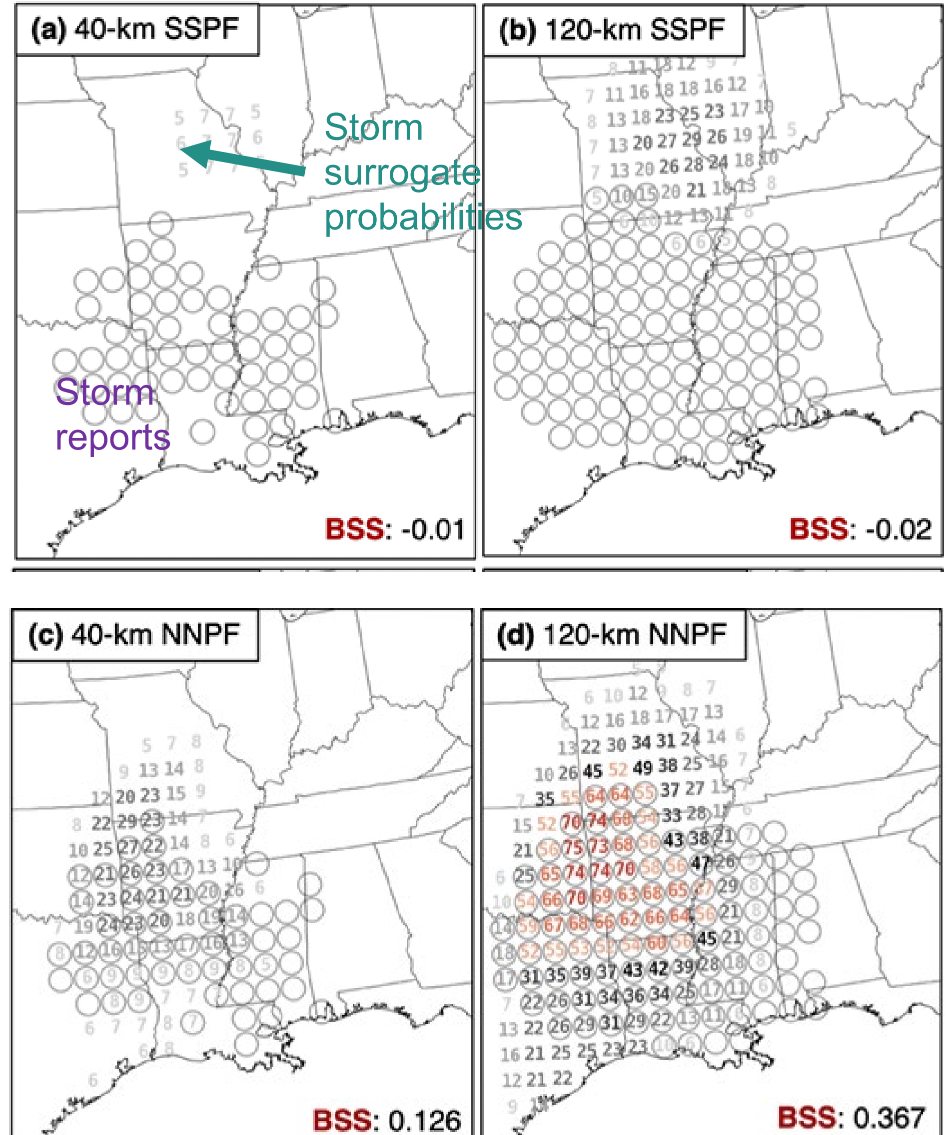


Making the most of what we have – AI post-processing

Instead of relying solely on explicit prediction and surrogates, ML allows for environmental conditions and other factors to be included, improving predictive skill by post-processing the same forecast

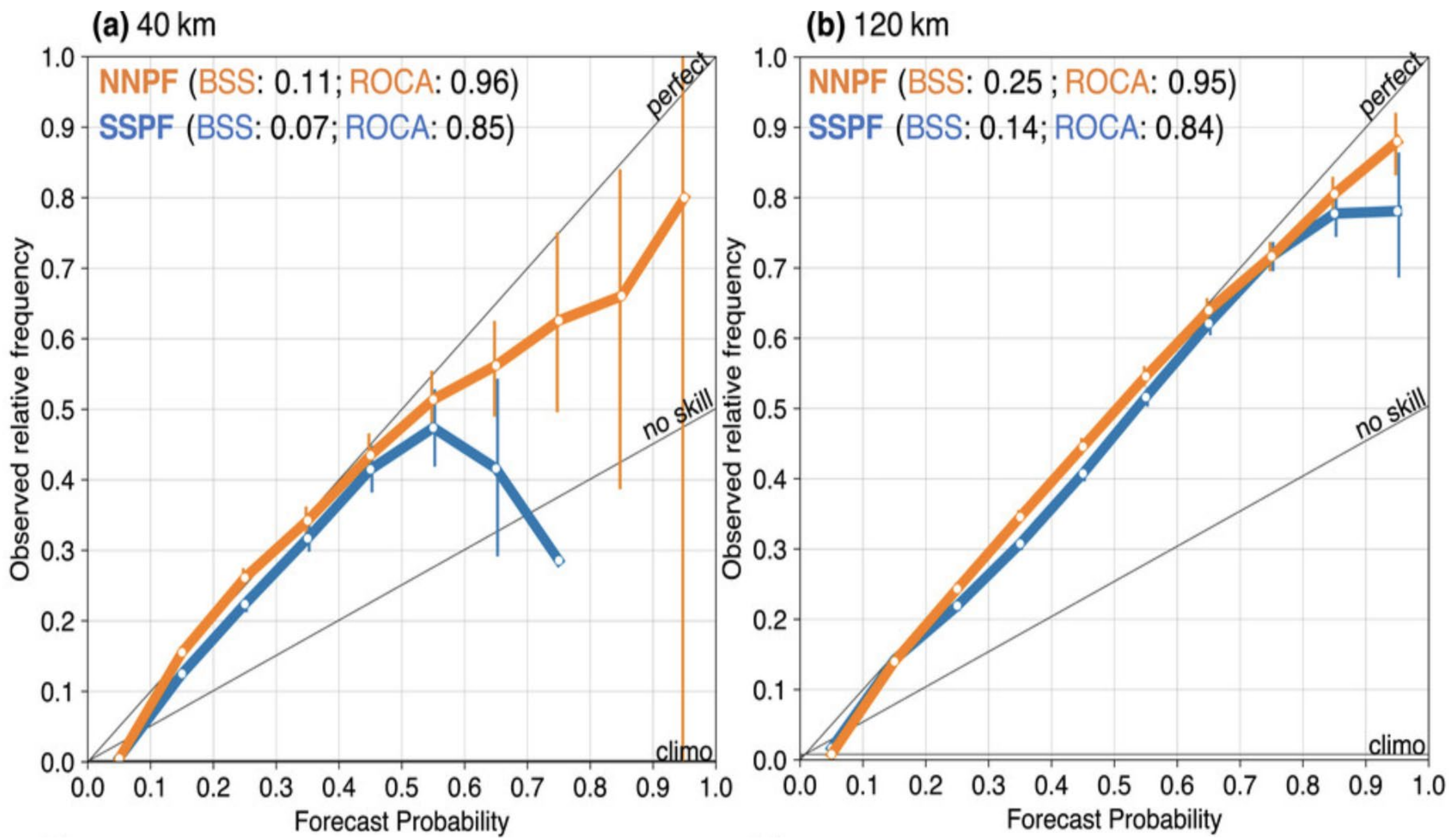


Neural network probability forecast
Storm surrogate probability forecast



Sobash et al. 2020

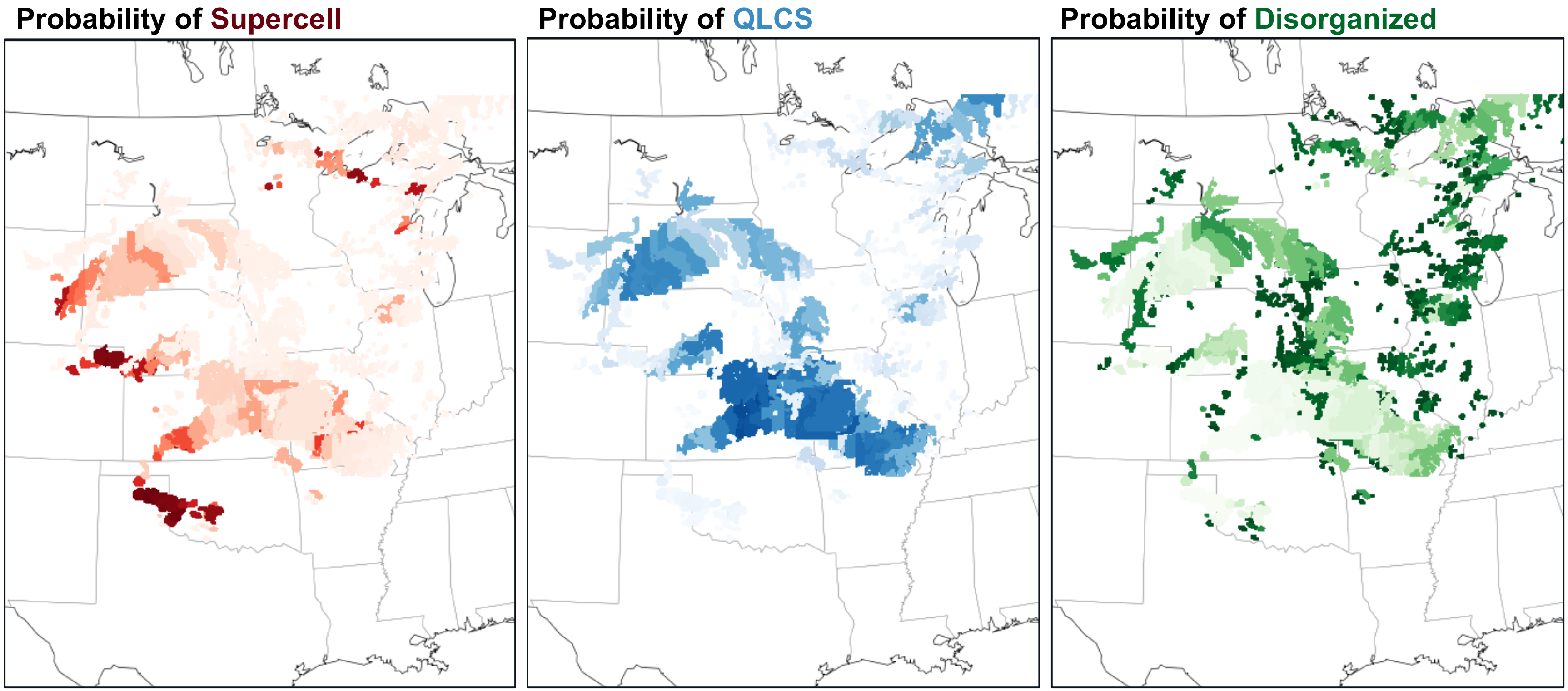
Making the most of what we have – AI post-processing



Neural network probability forecast
Storm surrogate probability forecast

Making the most of what we have – AI post-processing

Building an ML-based system to objectively identify convective mode in CAM output



Forecast initialized 00 UTC 24 May 2016, valid 12 UTC 24 May 2016 – 12 UTC 25 May 2016

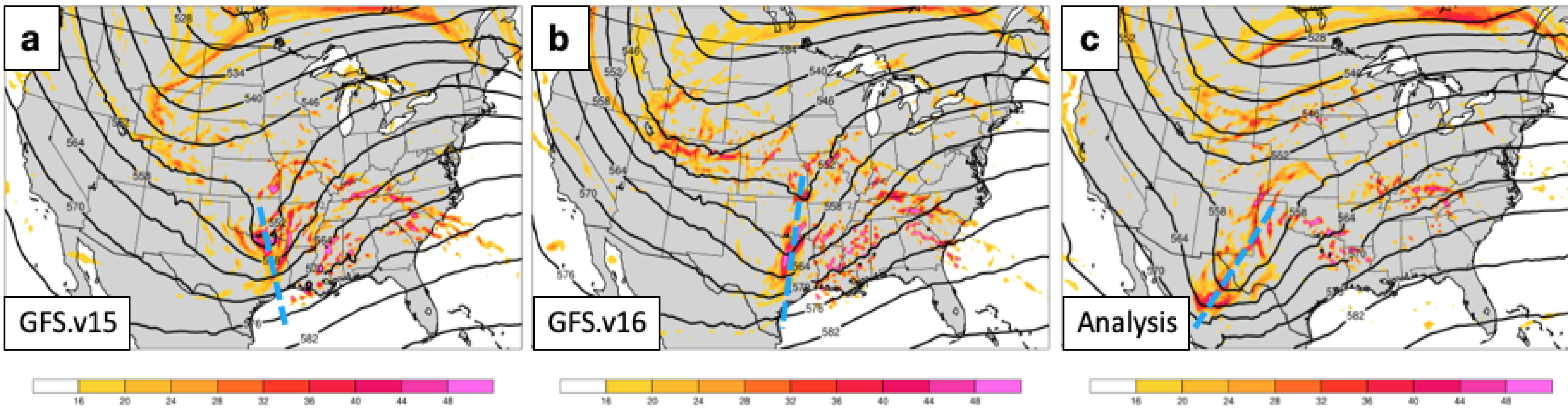
Predictions using CNNnew (including S2 into QLCS category) – higher probabilities indicated by darker shading

New research activities led by Ryan Sobash with HWT (and soon JTTI) support

Looking ahead: Tendency diagnostics for *conditional* model error

Diagnosing synoptic progressiveness forecast errors within the UFS MRWA

May Wong, Craig Schwartz, and Glen Romine of NCAR, Alicia Bentley and Geoffrey Manikin NOAA/EMC



500-hPa geopotential height (dam; contours) and absolute vorticity (x 10^{-5} s^{-1}; fill) for a) GFS.v15 and b) GFS.v16 initialized at 1200 UTC 08 April 2020 and valid 1200 UTC 12 April 2020

- Progressiveness may be associated with re-connecting and detaching cutoff lows
- We will develop object-based diagnostics to investigate physics behavior of cutoffs

Summary – moving toward intrinsic predictability

- Improving model skill is a faster pathway to better initial conditions and subsequently better forecasts
  - Higher resolution (explicit) background analysis
  - Else, address regional model shortcomings with blended analysis
- Reliable forecasts are equally challenging
  - Dependence on the characteristics of initial ensemble perturbations
- Improved post-processing (AI) can lead to better skill and reliability
- In the works:
  - Moving toward understanding conditional forecast error diagnosis