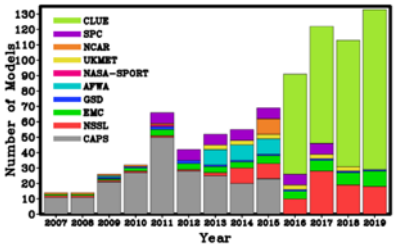
What Does a Convection-Allowing Ensemble of Opportunity Buy Us in Forecasting Thunderstorms?

Brett Roberts[123], Burkely T. Gallo[13], Israel L. Jirak[1], Adam J. Clark[25], David C. Dowell[4], Xuguang Wang[5], Yongming Wang[5]

UFS Users' Workshop, 27 July 2020

[1] NOAA/NWS/NCEP Storm Prediction Center, Norman, OK
[2] NOAA/OAR/National Severe Storms Laboratory, Norman, OK
[3] Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, OK
[4] NOAA/OAR/ESRL Global Systems Laboratory, Boulder, CO
[5] School of Meteorology, University of Oklahoma, Norman, OK

Background and Motivation



- **Community Leveraged Unified Ensemble (CLUE, 2016-present)** provides an unprecedented opportunity for systematic comparisons of CAM ensemble systems

- Spring Forecasting Experiments (SFEs) have given us subjective feedback on CAM ensembles
  - Primary focus is CAM ensemble forecasts of **convective evolution** and **severe hazards**

- Recent SFEs have noted some consistent themes from subjective comments/ratings:
  - Diverse “ensembles of opportunity” (SSEO 2011-17; HREF 2018-present), which aggregate output from CAM runs with disparate configurations, tend to score higher than formal/unified ensembles
  - A common impression is that formal ensembles are **underdispersive**, while SSEO/HREF are more dispersive and tend to show spatially broader neighborhood probability fields for conv. fields

- We want to quantify this subjective gap between HREF and unified ensembles in the CLUE

Datasets

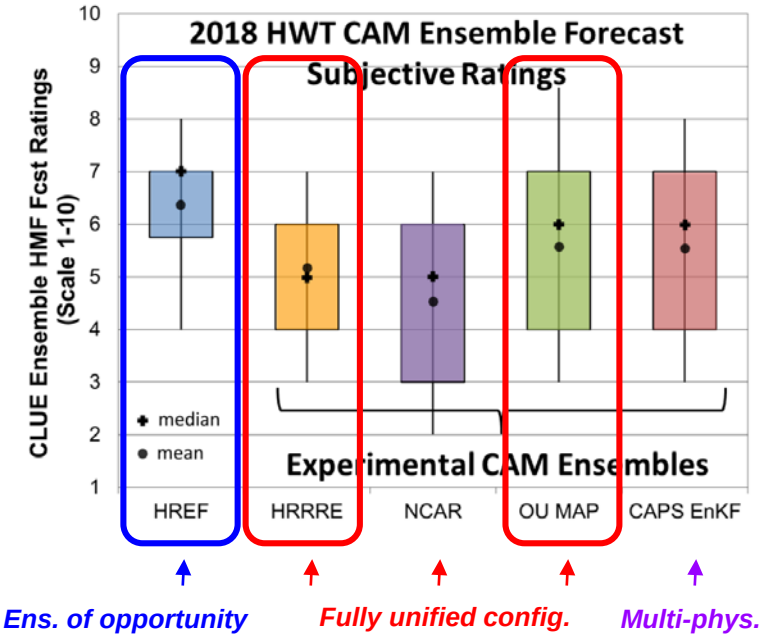
Nightly 00 UTC runs of 3 CAM ensembles examined over the set of days in **SFE2018** with all data available (**21 days total**):

- **HREFv2.1 (N=10;** multi-core/physics/IC parents & time lagging)
- **HRRRE (N=9;** unified configuration; perturbed ICs/LBCs)
- **OU MAP (N=10;** unified configuration; perturbed ICs/LBCs)

HRRRE and OU MAP essentially share the “CLUE” configuration: *WRF-ARW, Thompson microphysics, MYNN PBL, RUC LSM*

HREFv2.1 member configuration details:

Member	Core	ICs	LBCs	Microphysics	PBL
HRRR	WRF-ARW	RAP -1h	RAP -1h	Thompson	MYNN
HRRR -6h	WRF-ARW	RAP -1h	RAP -1h	Thompson	MYNN
HRW ARW	WRF-ARW	RAP	GFS -6h	WSM6	YSU
HRW ARW -12h	WRF-ARW	RAP	GFS -6h	WSM6	YSU
HRW NMMB	NMMB	RAP	GFS -6h	Ferrier-Aligo	MYJ
HRW NMMB -12h	NMMB	RAP	GFS -6h	Ferrier-Aligo	MYJ
HRW NSSL	WRF-ARW	NAM	NAM -6h	WSM6	MYJ
HRW NSSL -12h	WRF-ARW	NAM	NAM -6h	WSM6	MYJ
NAM CONUS Nest	NMMB	NAM	NAM	Ferrier-Aligo	MYJ
NAM CONUS Nest -12h	NMMB	NAM	NAM	Ferrier-Aligo	MYJ



HREF rated best overall by SFE participants

OU MAP rated best among unified ensembles

HRRRE rated next-best among unified ensembles

Verification fields

We are concerned with CAM ensemble forecasts of convective evolution and severe hazards…
In other words, roughly what SPC must forecast in their outlooks. So we will verify two fields:

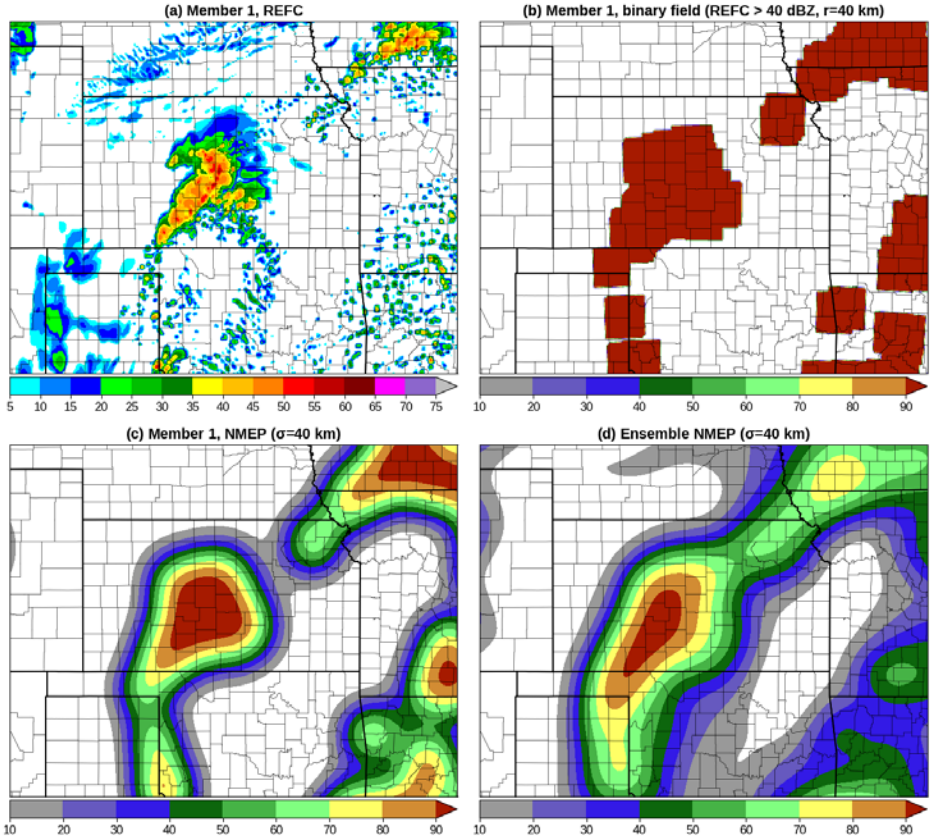
1. **Composite radar reflectivity (CREF)**
    a. Succinctly captures convective evolution and coverage
    b. High-quality truth dataset available to verify against: MRMS national mosaic
    c. Instantaneous hourly fields are verified, so timing errors are heavily penalized

2. **2-5 km AGL Updraft helicity (UH) -- a.k.a. “surrogate severe” forecasts**
    a. Highlights intense, rotating storms responsible for most severe hazards
    b. The surrogate severe verification methodology will be used (LSRs as truth)
    c. Verified over the entire 24-h convective day, so timing errors are largely tolerated

These fields are central in the SFE when subjective impressions are formed, and are arguably the most important CAM fields examined by SPC forecasters.

Composite Reflectivity (CREF) methodology

- All verification is performed on **neighborhood maximum ensemble probabilities (NMEPs)** for the **40-dBZ CREF** threshold
  - 40-km NH “radius” (box half-length)
  - NMEP smoothed w/ 40-km Gaussian
  - **Percentile-based member bias correction**
- We verify **hourly snapshots** of CREF for forecast hours 13-30 (13Z-06Z daily)



Q: ***Are there storms in the right areas for each snapshot?***

Brier Skill Score for CREF >40 dBZ NMEPs

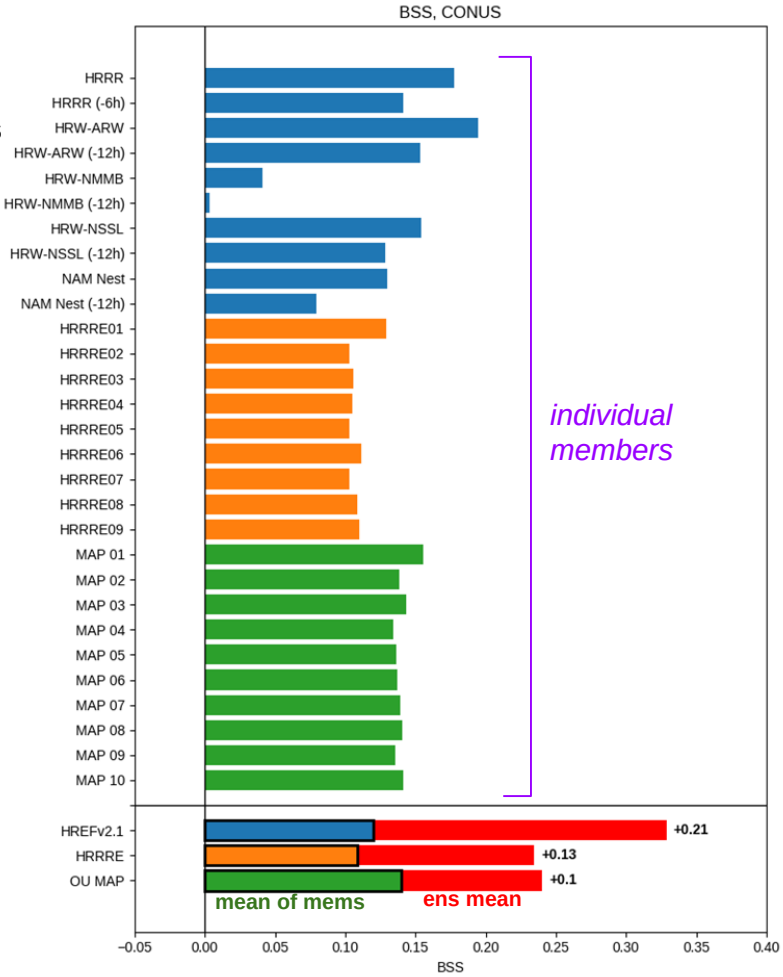
Member ranking:
1. OU MAP
2. HREFv2.1
3. HRRRE

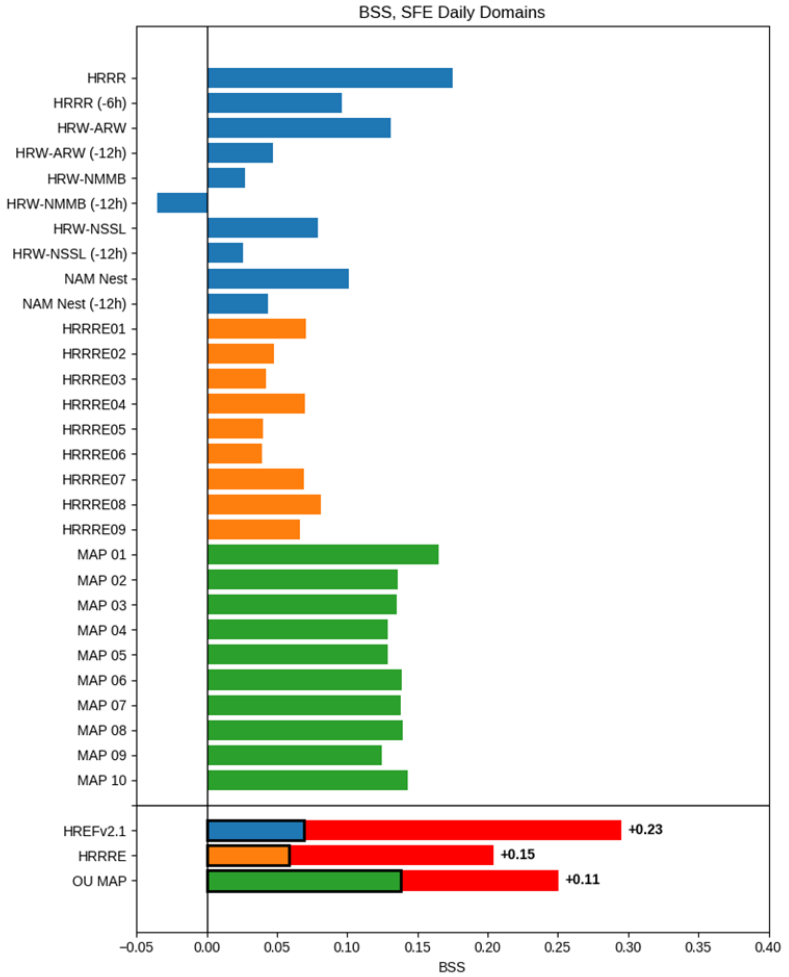
Ensemble ranking:
1. HREFv2.1
2. OU MAP
3. HRRRE

Red bars: **BSS_{gained}**

= $BSS_{ens\text{-}mean}$ - $\overline{BSS_{mems}}$

HREFv2.1 is making the most of its members, as indicated by BSS_{gained}.

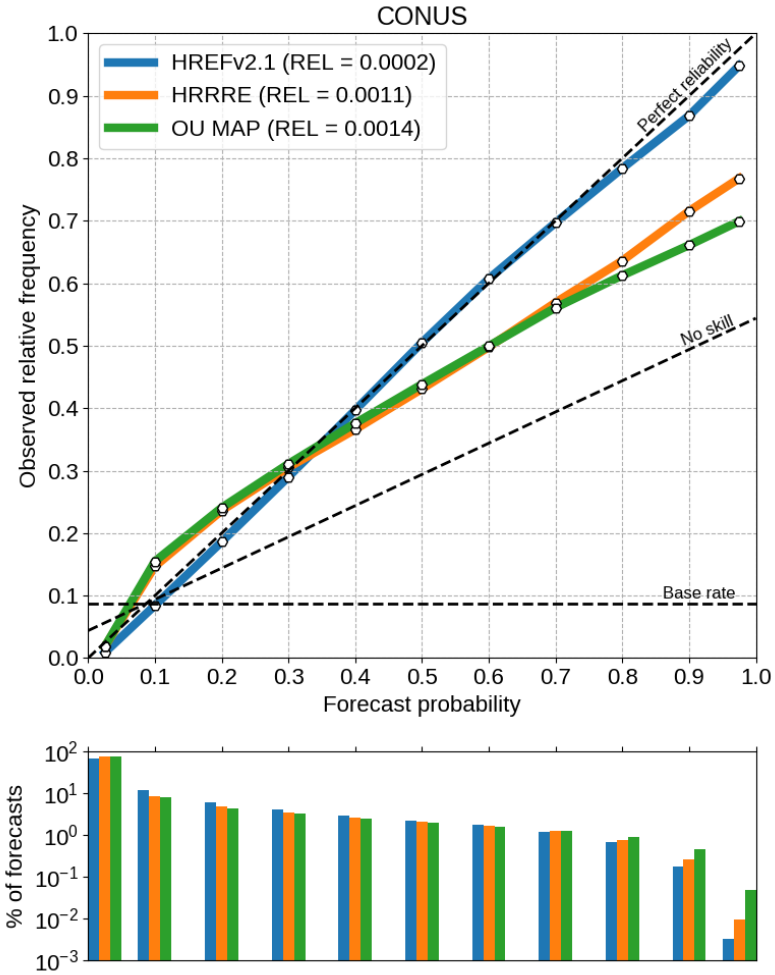


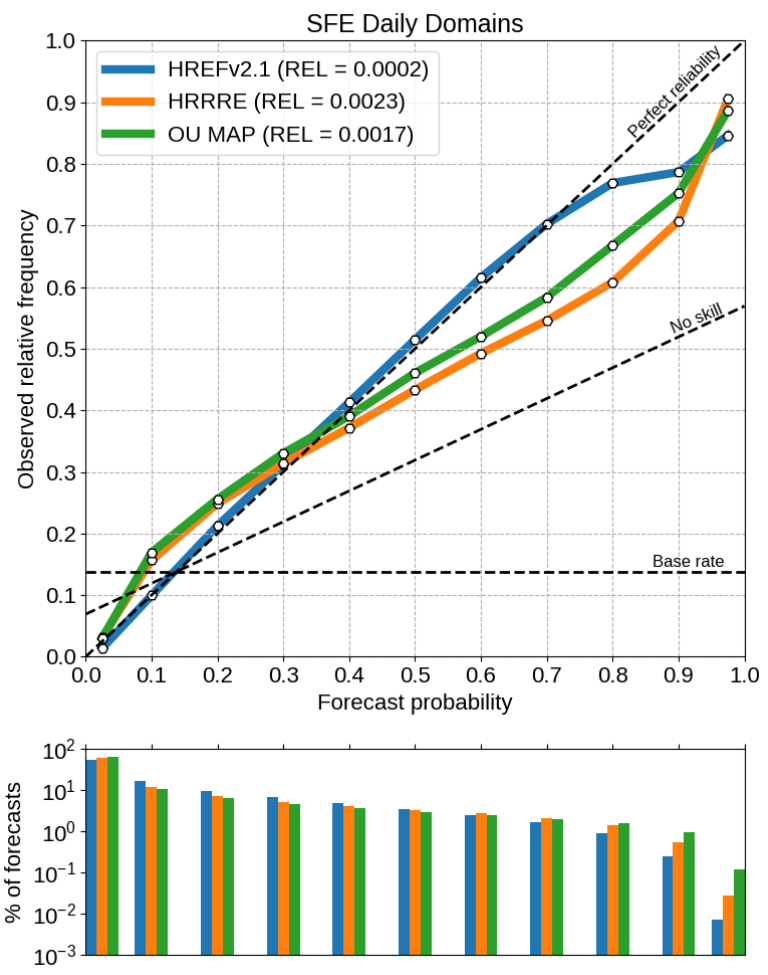


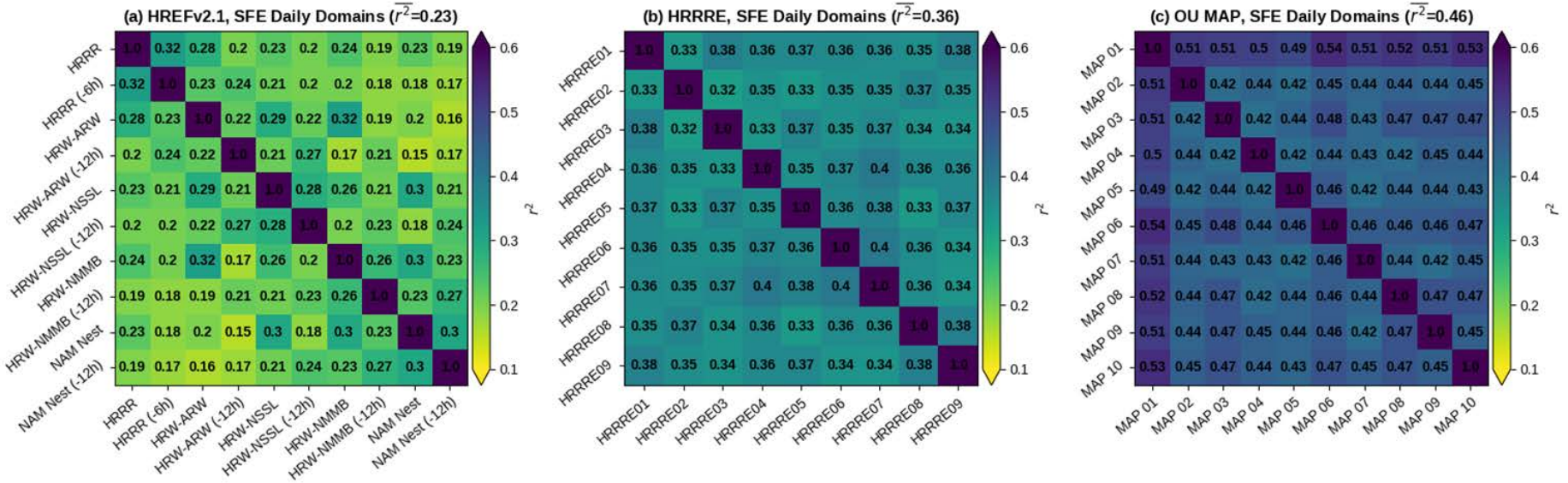
Attributes diagrams for CREF >40 dBZ NMEPs

HREFv2.1 shows superior reliability, except at very high probabilities within daily domains.

HRRRE and OU MAP show classic underdispersive signal.







Correlation matrices for CREF >40 dBZ NMEPs

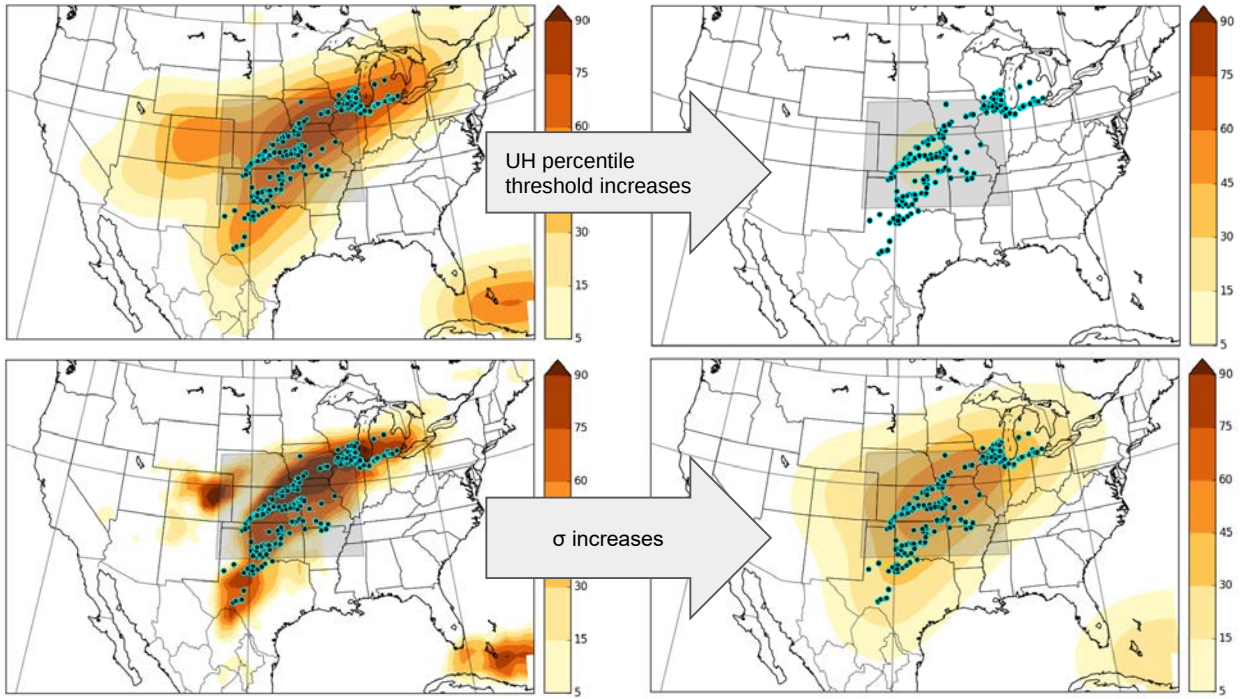
Mean correlation between members with respect to CREF placement:

1. OU MAP (0.46)
2. HRRRE (0.36)
3. HREFv2.1 (0.23)

Surrogate Severe methodology

SS forecasts are fundamentally similar to CREF NMEPs, except:

1. Computed on 80-km grid
2. One field for max UH over the full 24-h convective day (12Z-12Z)
3. We compute for numerous percentile thresholds and smoothing (**σ**) values to see how skill varies in that parameter space



Q: ***Did the CAM (or ensemble) do well at forecasting where intense, rotating storms tracked throughout the convective day?***

Surrogate Severe FSS

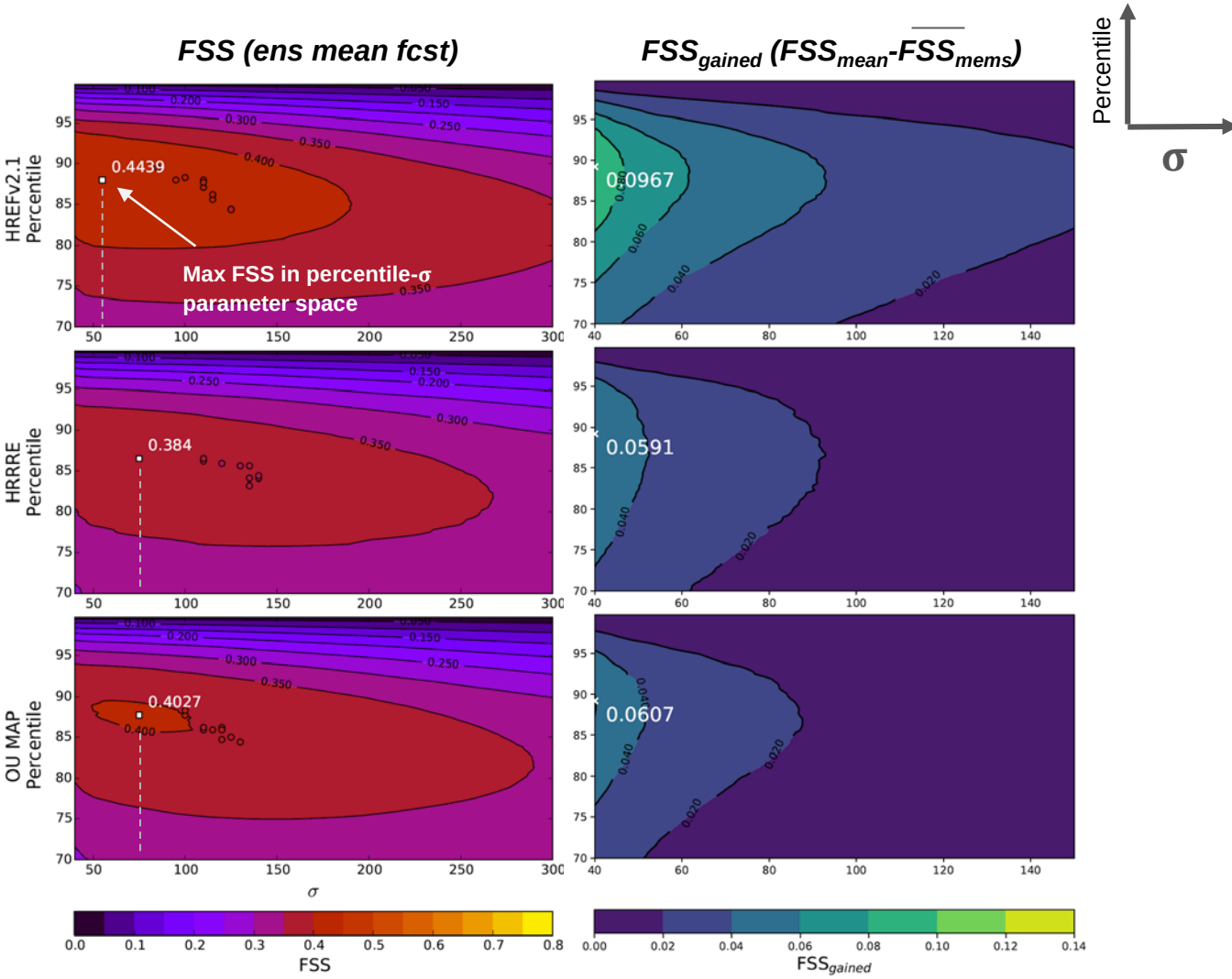
Ranking of ensemble mean forecasts matches CREF:

1. HREFv2.1
2. OU MAP
3. HRRRE

UH percentiles maximizing skill differ some, and **HREF maximizes skill with less smoothing than HRRRE and OU MAP.**

Ensemble SS forecasts outperform individual members more as smoothing decreases.

HREF again shows a more pronounced improvement over its members than HRRRE or OU MAP.



Summary and Conclusions

- Overall, HREF was best at forecasting convective evolution during SFE 2018, followed by OU MAP, and then HRRRE.
  - This matches subjective participant ratings

- Skill differences between systems are meaningful at smaller scales ($\sigma \sim 40$ km), but for surrogate severe, they largely wash out after applying aggressive smoothing ($\sigma > 100$ km) to the NMEPs.

- Objective statistics confirm HREF member probability fields are less correlated with one another, resulting in better spread-skill relationship and reliability than OU MAP or HRRRE.

Summary and Conclusions

- **Key takeaway:** HREF membership design *seems* to be sampling model uncertainty effectively. How difficult will it be for stochastic perturbations in unified ensembles to replicate this as they mature?

- **Caveat #1:** we looked at the “next-day problem.” Relevance of our spread findings to “WoF problem” (short lead times, small spatial scales) is TBD.

- **Caveat #2:** we verified SFE2018 data here; preliminary stats suggest HRRRE spread improved in SFE2019, so gap may be closing some.