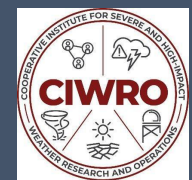


Unified Forecasting System results from recent NOAA/Hazardous Weather Testbed Spring Forecasting Experiments

Unifying Innovations in Forecasting Capabilities Workshop
18-22 July, College Park, MD

Adam J. Clark^{2,4}, Burkely T. Gallo^{1,3}, Louis J. Wicker^{2,4}



¹NOAA/NWS/NCEP Storm Prediction Center, Norman, Oklahoma

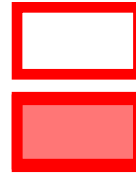
²NOAA/OAR National Severe Storms Laboratory, Norman, Oklahoma

³Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma, Norman, Oklahoma

⁴School of Meteorology, University of Oklahoma, Norman, Oklahoma

UFS-related evaluations 2021-22

2021



UFS-related evaluations

UFS-related evaluations in this talk

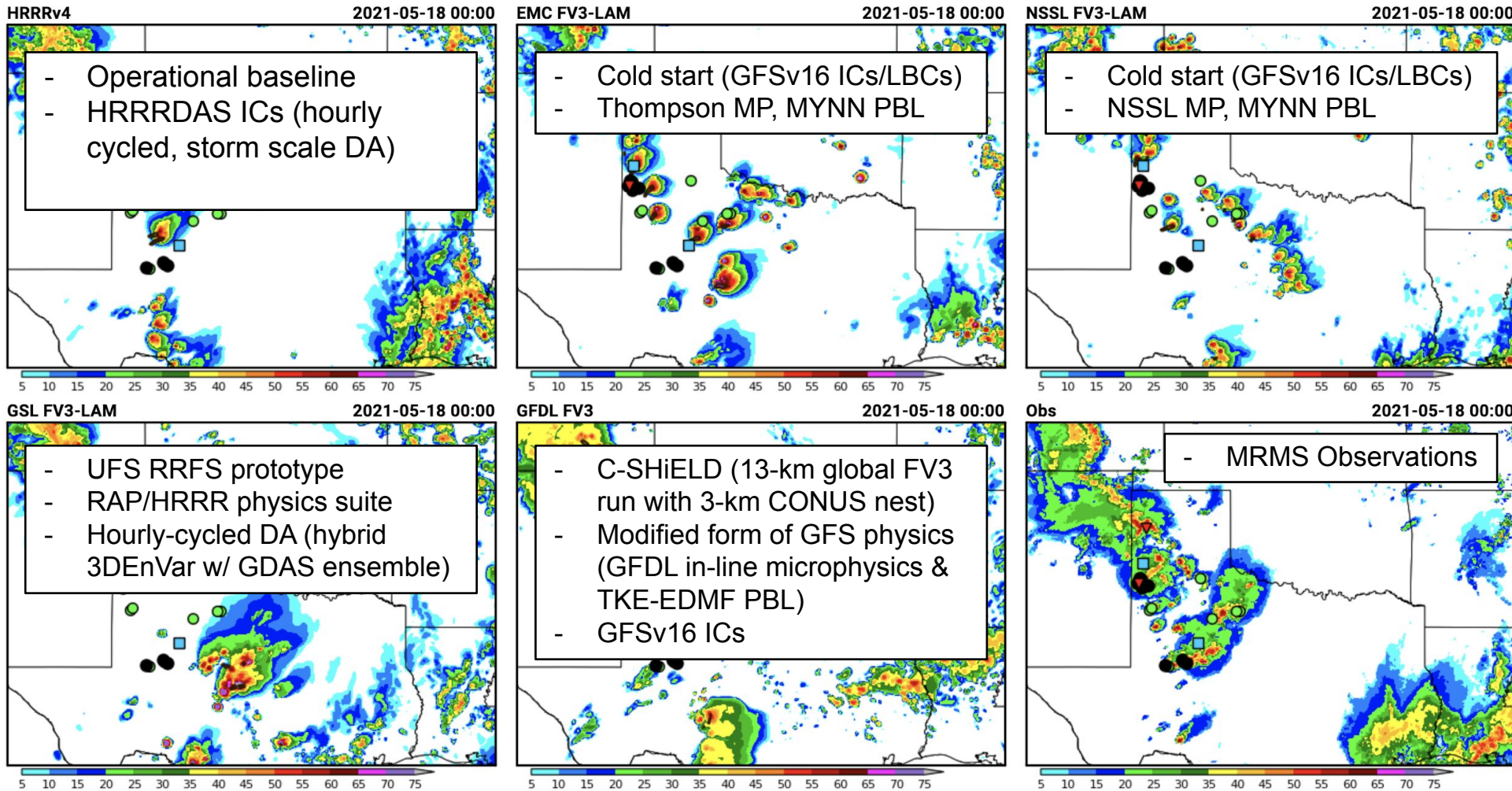
2022

Group A: Calibrated Guidance	Group B: Deterministic CAMs	Group C: CAM Ensembles	Group D: Medley
Days 1 & 2 Tornado	Deterministic Flagships	00Z Ensembles	ISU ML Severe Wind Probabilities
00Z HREF-based Tornado	FV3-LAM Domain Comparison	12Z Ensembles	NCAR ML Convective Mode Probabilities
Days 1 & 2 Hail	FV3-LAM DA Comparisons	Hourly Updating CAM Ensembles	Mesoscale Analysis Background
00Z HREF-based Hail	FV3 Physics Suites	Valid-Time-Shifting DA	Mesoscale Analysis DA frequency
Days 1 & 2 Wind	FV3 Stochastic Physics	WoFS vs. HRRR-TL	Storm Scale Analyses
00Z HREF-based Wind		Deterministic WoFS	GEFS vs. SREF Days 2 & 3
		ML-based Calibrated	

Group A: Calibrated Guidance	Group B: Deterministic CAMs	Group C: CAM Ensembles	Group D: Medley
GEFS-based hazard guidance (Tor, Wind, Hail)	Deterministic Flagships	00Z Ensembles	ISU ML Severe Wind Probabilities
Days 1 & 2 hazard guidance (Tor, Wind, Hail)	RRFSp2e vs. HRRR	RRFSp2e vs. HREF	NCAR ML Convective Mode Probabilities
00Z HREF-based hazard guidance (Tor, Wind, Hail)	FV3-LAM DA Comparisons	FV3-LAM DA Comparisons	Mesoscale Analysis Background
OU-MAP flow-dependent hazard guidance (Tor, Wind, Hail)	FV3 Physics Suites	TTU ensemble Subsetting	Storm Scale Analyses
00Z HRRR NCAR NN hazard guidance (Tor, Wind, Hail)	1-km vs. 3-km NSSL-WRF	WoFS: Number of Members & time-laging	GEFS vs. SREF Days 2 & 3

SFE 2021: Deterministic Flagships

- **Goal:** Gauge progress for severe weather predictions for a single deterministic model from each SFE contributor relative to HRRRv4.

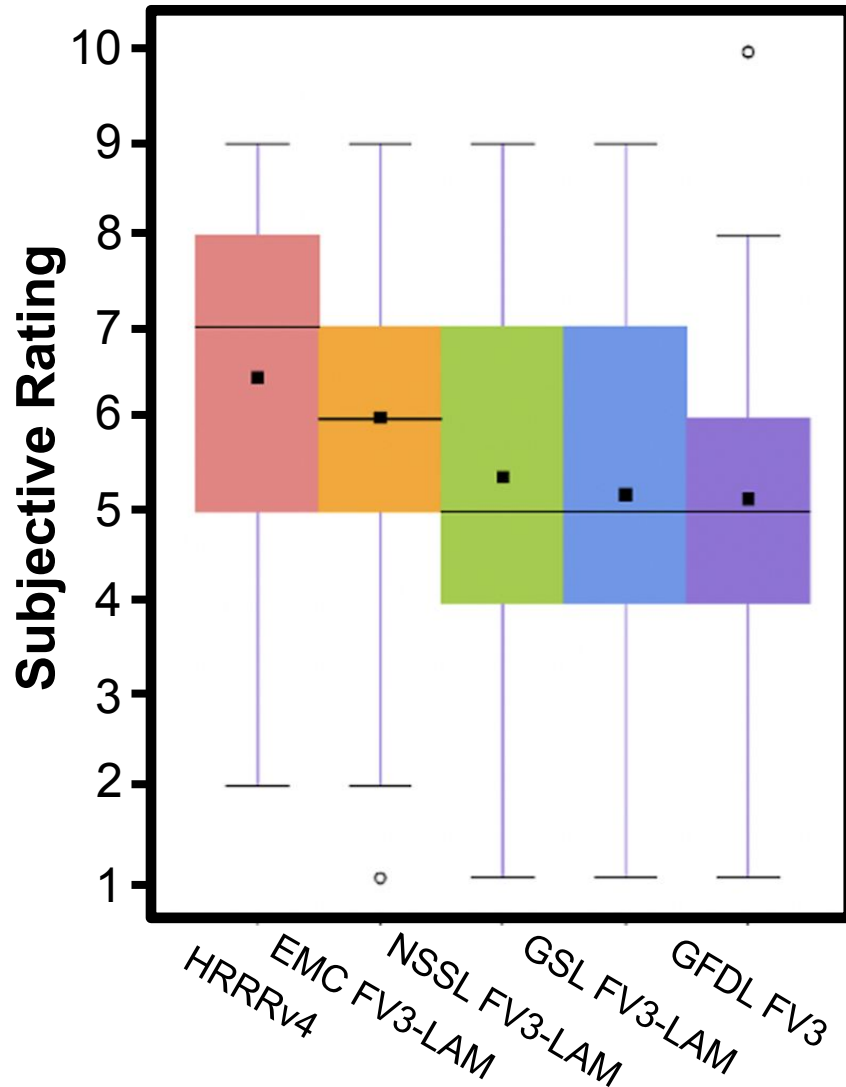


UH and composite reflectivity evaluated at 18, 00, & 06 UTC.

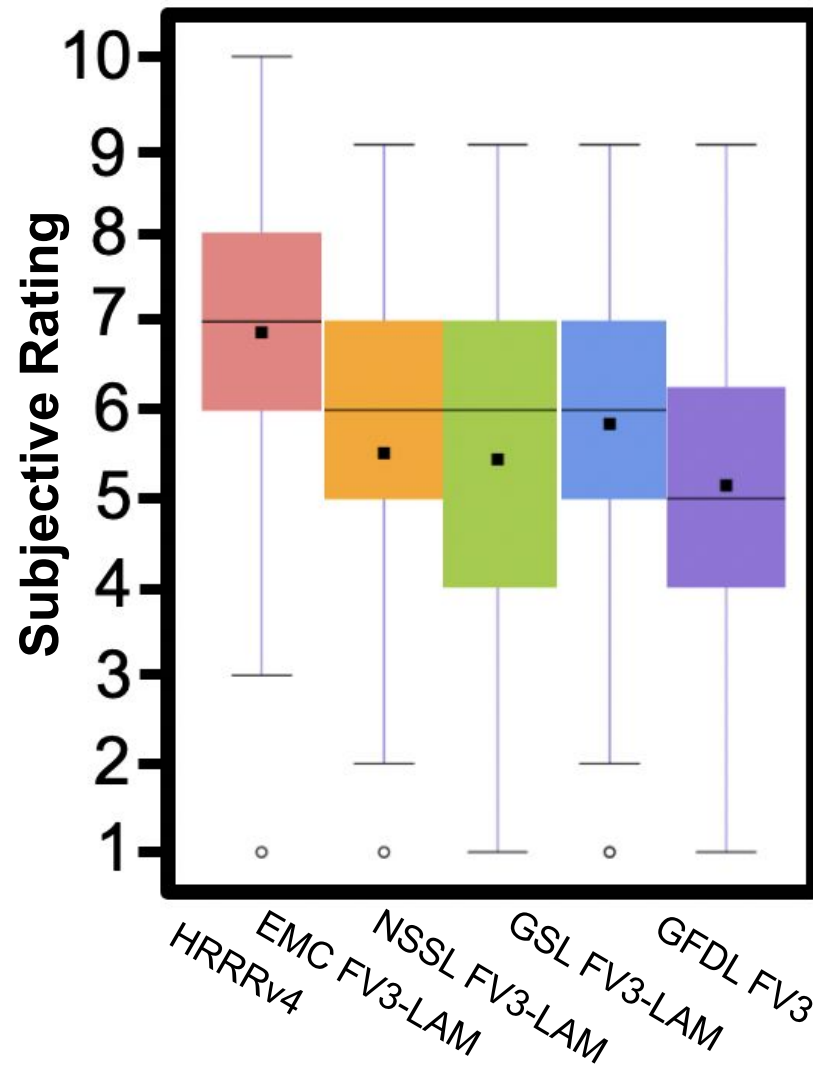
Environmental fields (T, Td, & CAPE) evaluated @ 18 UTC.

SFE 2021: Deterministic Flagships

UH & Reflectivity (0000 UTC)



Environment (1800 UTC; T, Td, & CAPE)

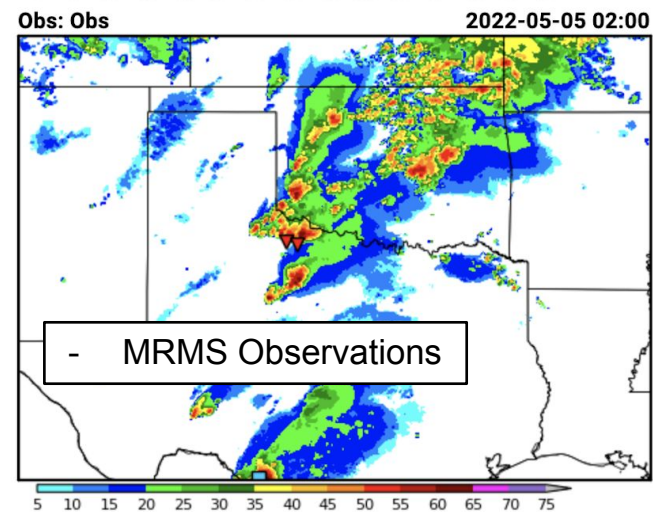
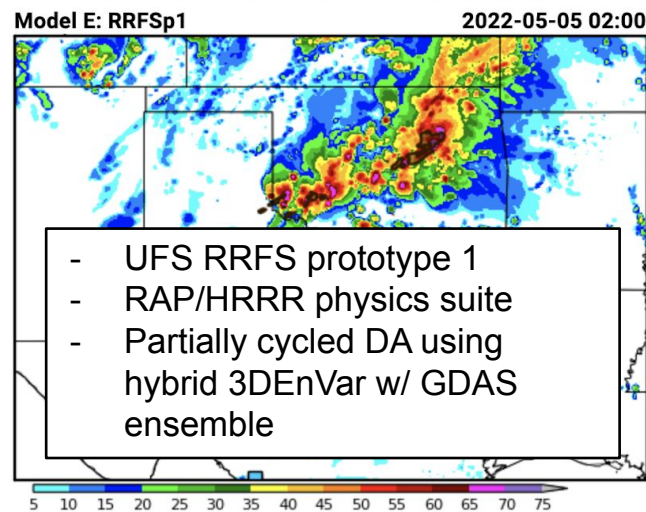
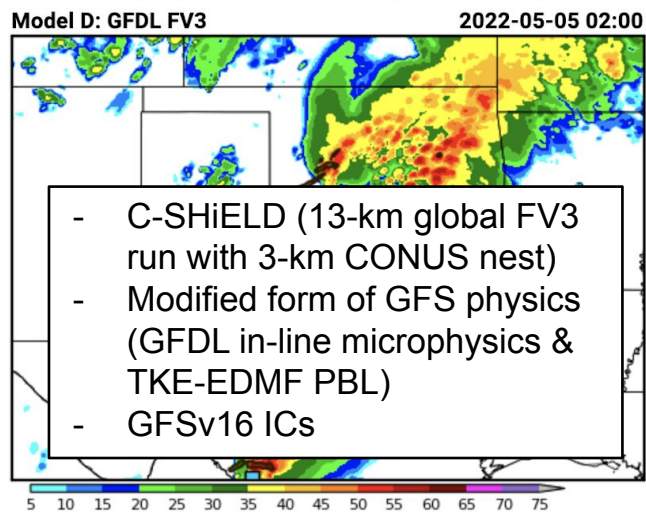
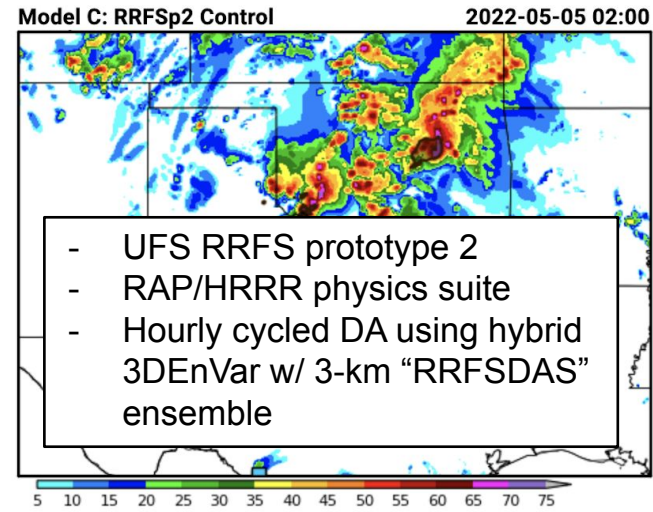
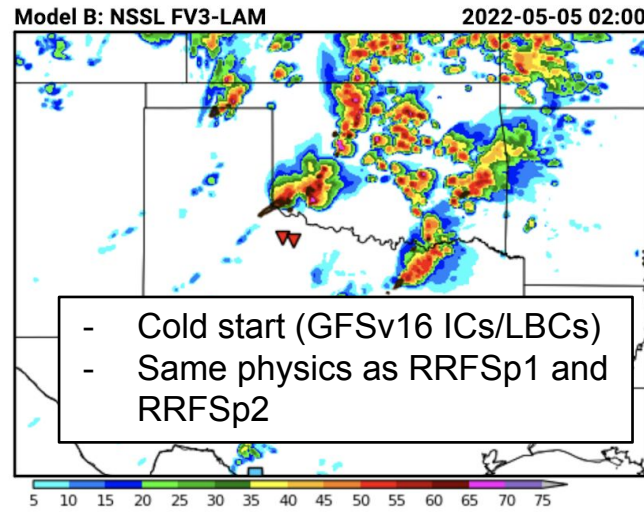
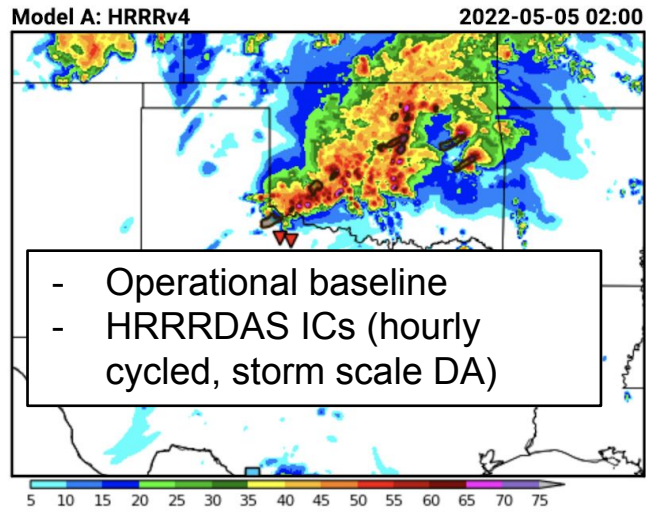


Results

- HRRR clearly performed best.
- All models struggled with SBCAPE magnitude.
- Cool biases in FV3 runs.

SFE 2022: Deterministic Flagships

- **Goal:** Gauge progress for severe weather predictions for a single deterministic model from each SFE contributor relative to HRRRv4.



UH and composite reflectivity evaluated for 12-36 h lead times.

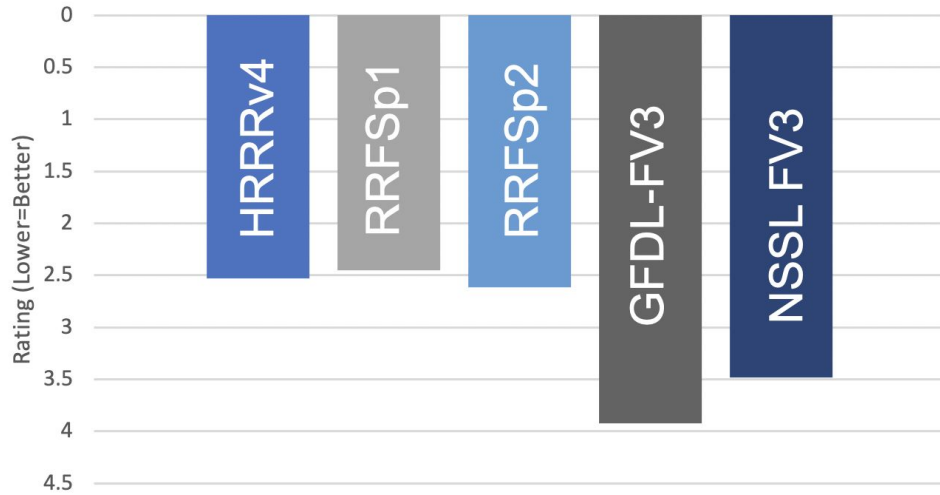
Environmental fields (T, Td, & CAPE) evaluated for 12-36 h lead times.

Changes in SFE2022:

- Comparisons were blind.
- Rankings instead of 1-10 ratings.

SFE 2022: Deterministic Flagships

Mean Ranking for Cases with All Model Data (12)



Results: Mean rankings

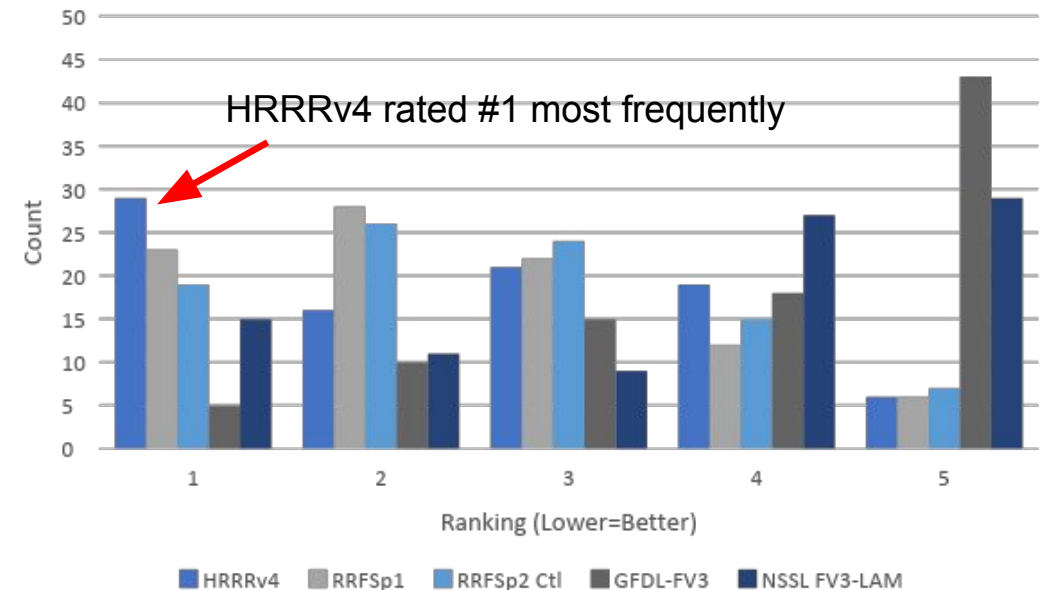
- HRRRv4, RRFSp1, & RRFSp2 very similar
- GFDL and NSSL FV3 clearly the worst

Results: Rankings distribution

- HRRRv4 most frequently rated #1
- GFDL followed by NSSL-FV3 most frequently rated worst



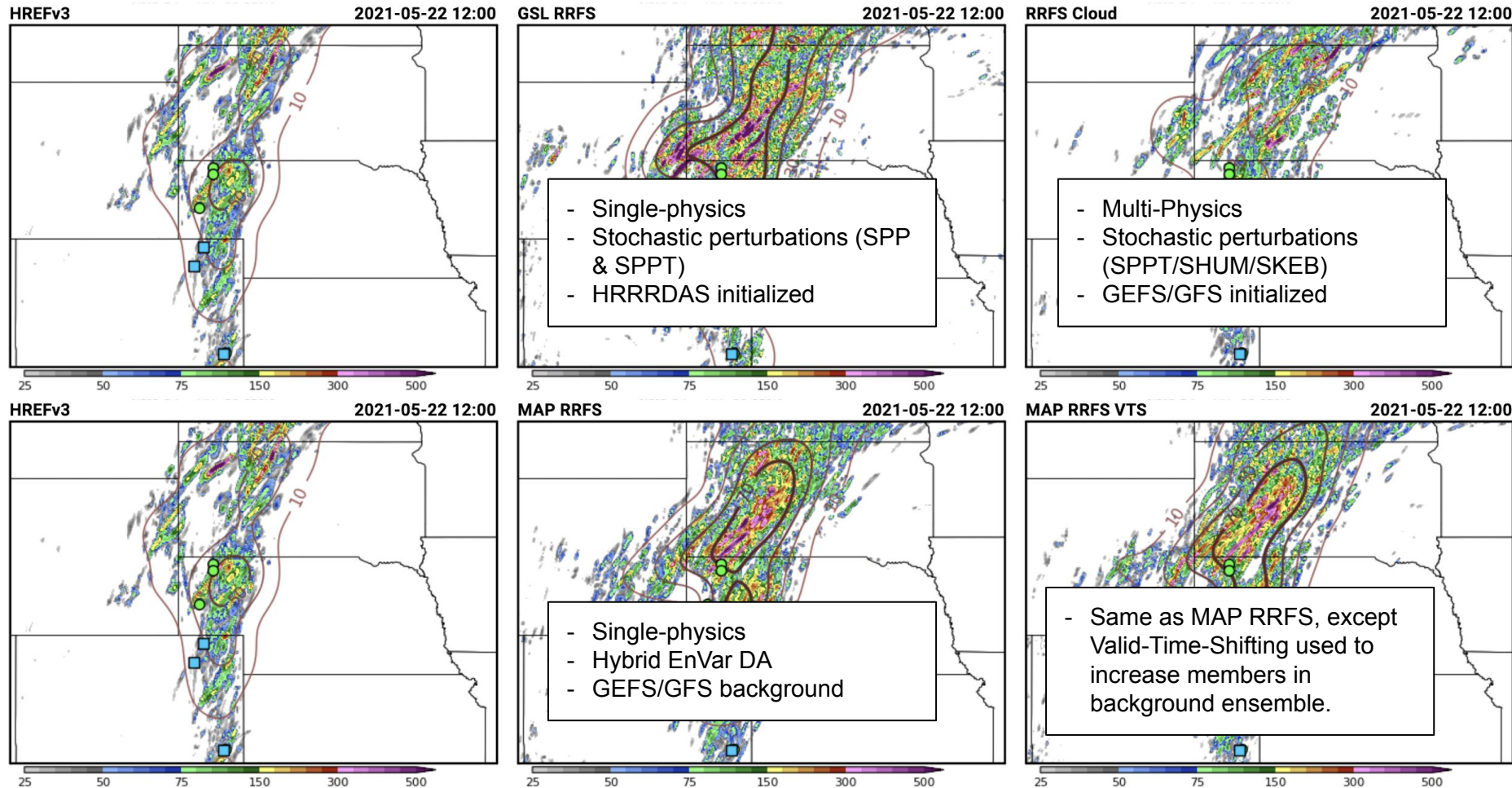
Ranking Distribution for Cases with All Model Data (12)



Tentative conclusion: RRFS prototypes approaching utility of HRRRv4, and relative skill improved from 2021

SFE 2021: 00Z CAM Ensembles

- **Goal:** Gauge progress for severe weather predictions for 00Z initialized CAM ensembles relative to HREF.

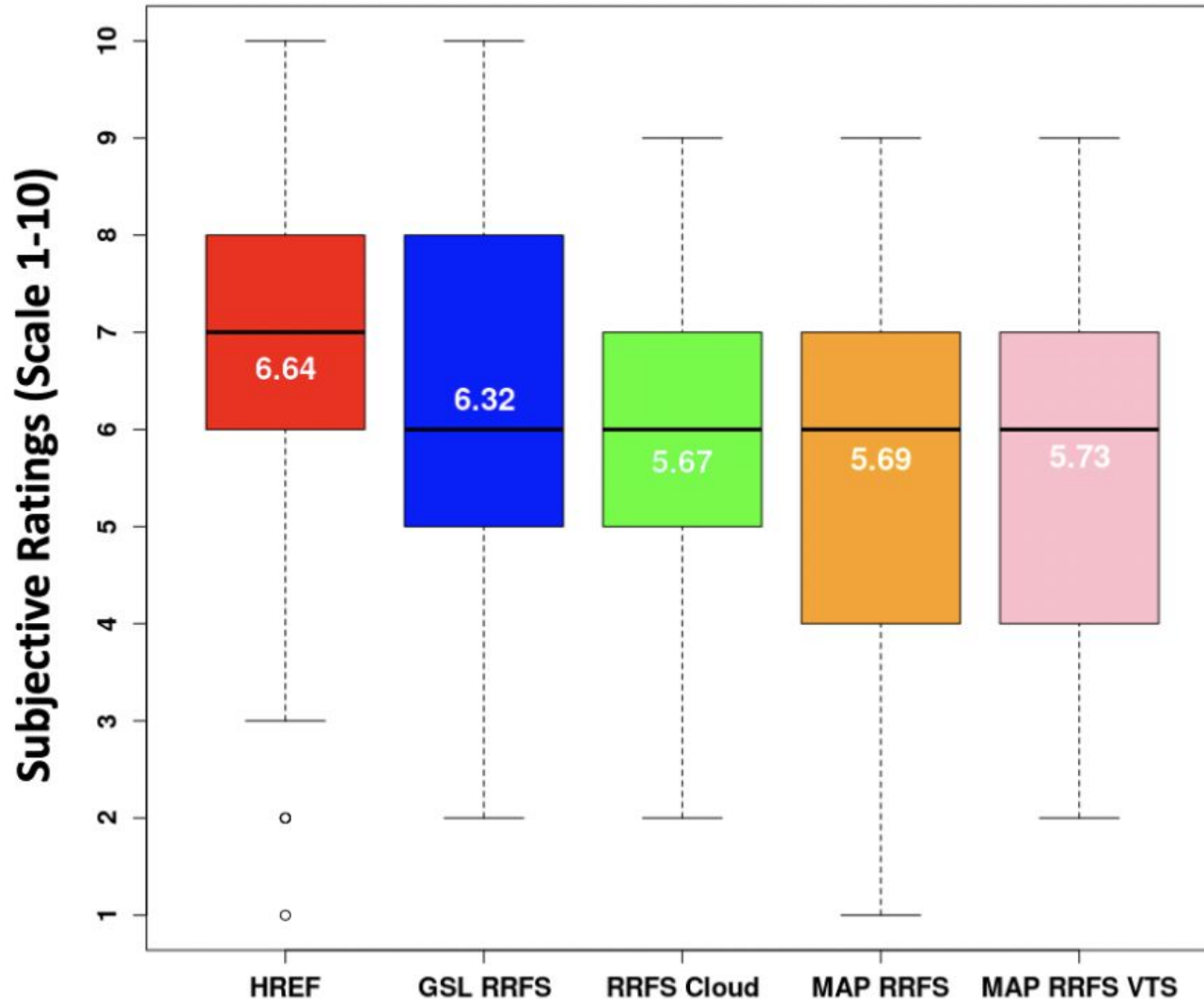


Methods

Compare 24-h summary of 2-5 km UH along w/ LSRs. “Rolling” 4-h UH also available & 40 DbZ reflectivity probs.

Subjectively rate forecast quality on scale of 1-10.

SFE 2021: 00Z CAM Ensembles

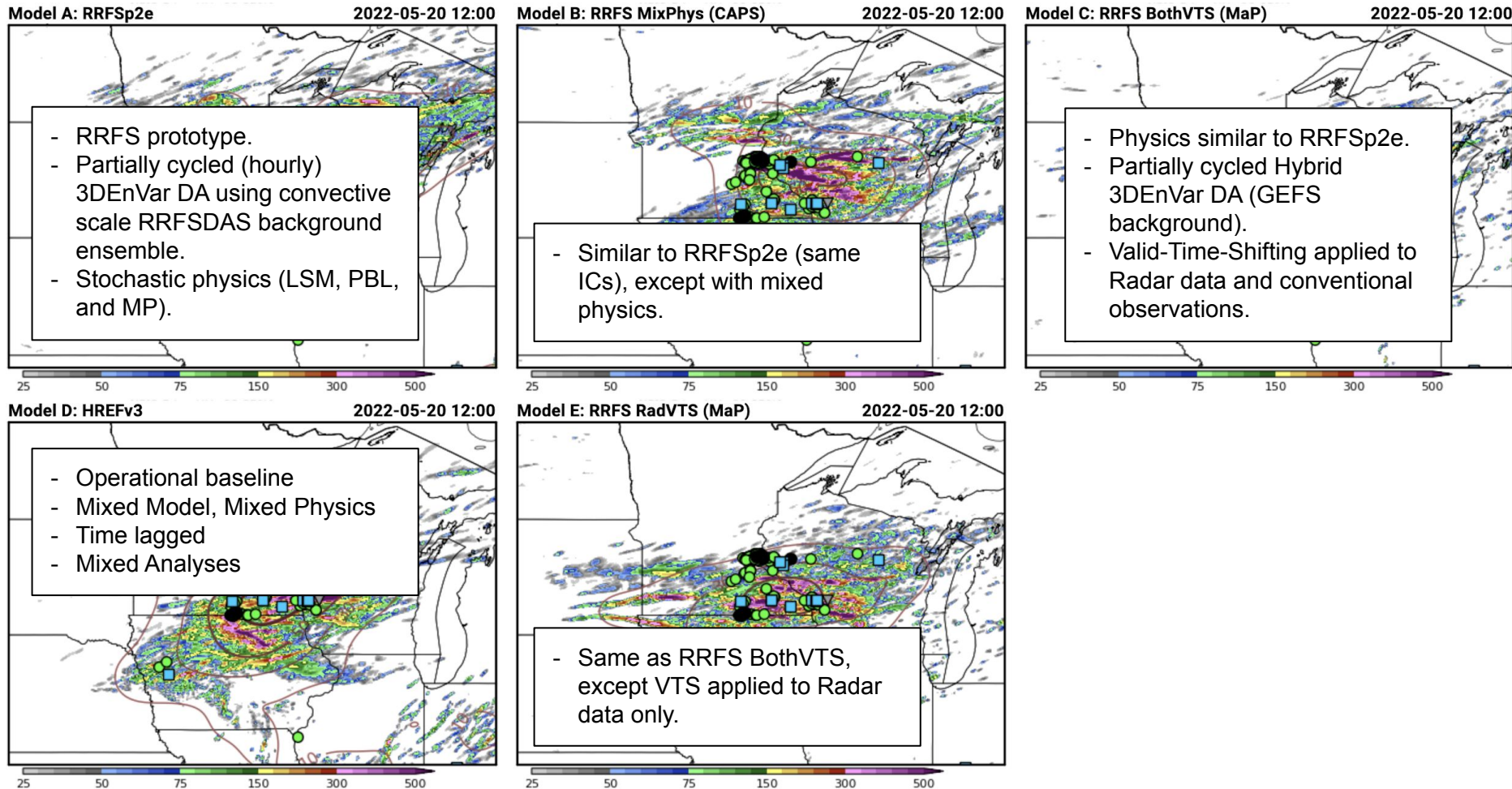


Results

- HREF performs best w/ GSL RRFS (HRRRDAS ICs) a close second.
- Participant comments:
 - HREF and GSL RRFS predicted locations of severe weather very accurately.
 - HREF and RRFS Cloud had largest spread or broadest coverage of probabilities.
 - MAP runs “too aggressive”
- HREF continues to be formidable baseline.

SFE 2022: 00Z CAM Ensembles

- **Goal:** Gauge progress for severe weather predictions for 00Z initialized CAM ensembles relative to HREF.



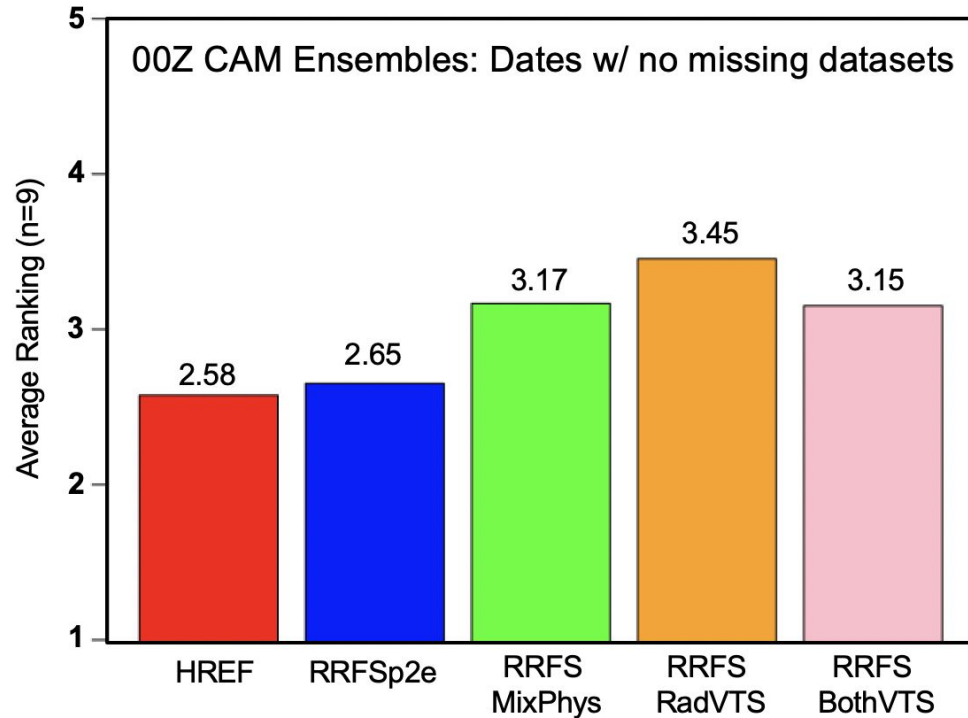
Methods

Compare 24-h summary of 2-5 km UH along w/ LSRs. “Rolling” 4-h UH also available & 40 DbZ reflectivity probs.

Changes from SFE2021

- Comparisons were “blind”
- Rankings instead of ratings.

SFE 2022: 00Z CAM Ensembles

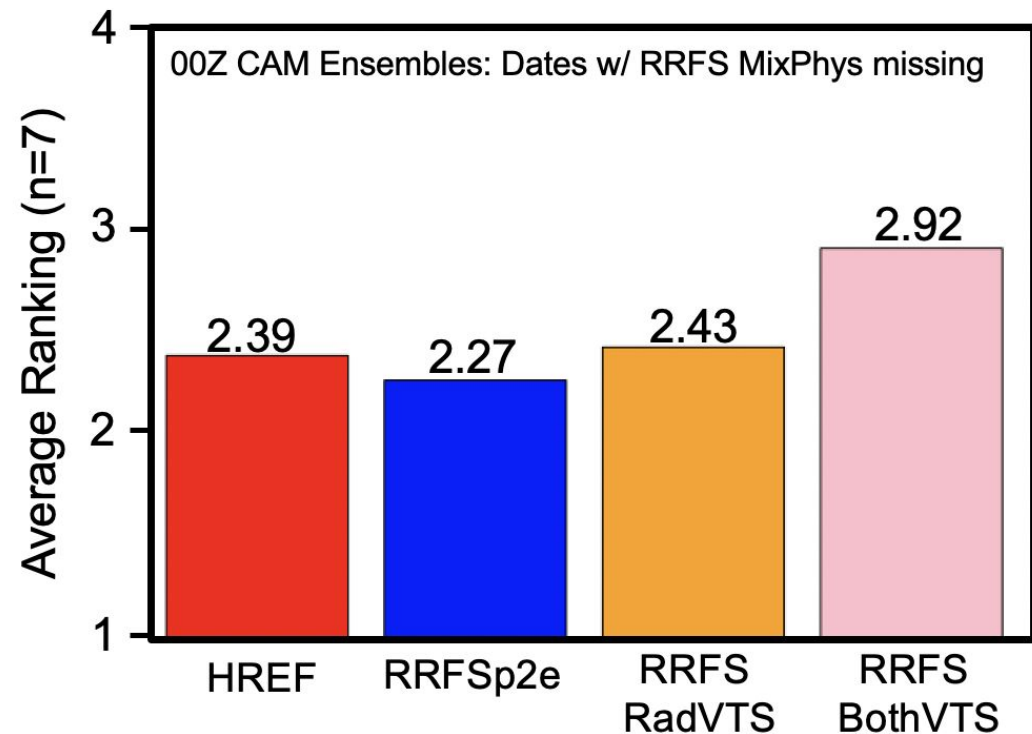


9 cases where all 5 ensembles were available

- HREF and RRFSp2e average ranking very similar with HREF having slight advantage

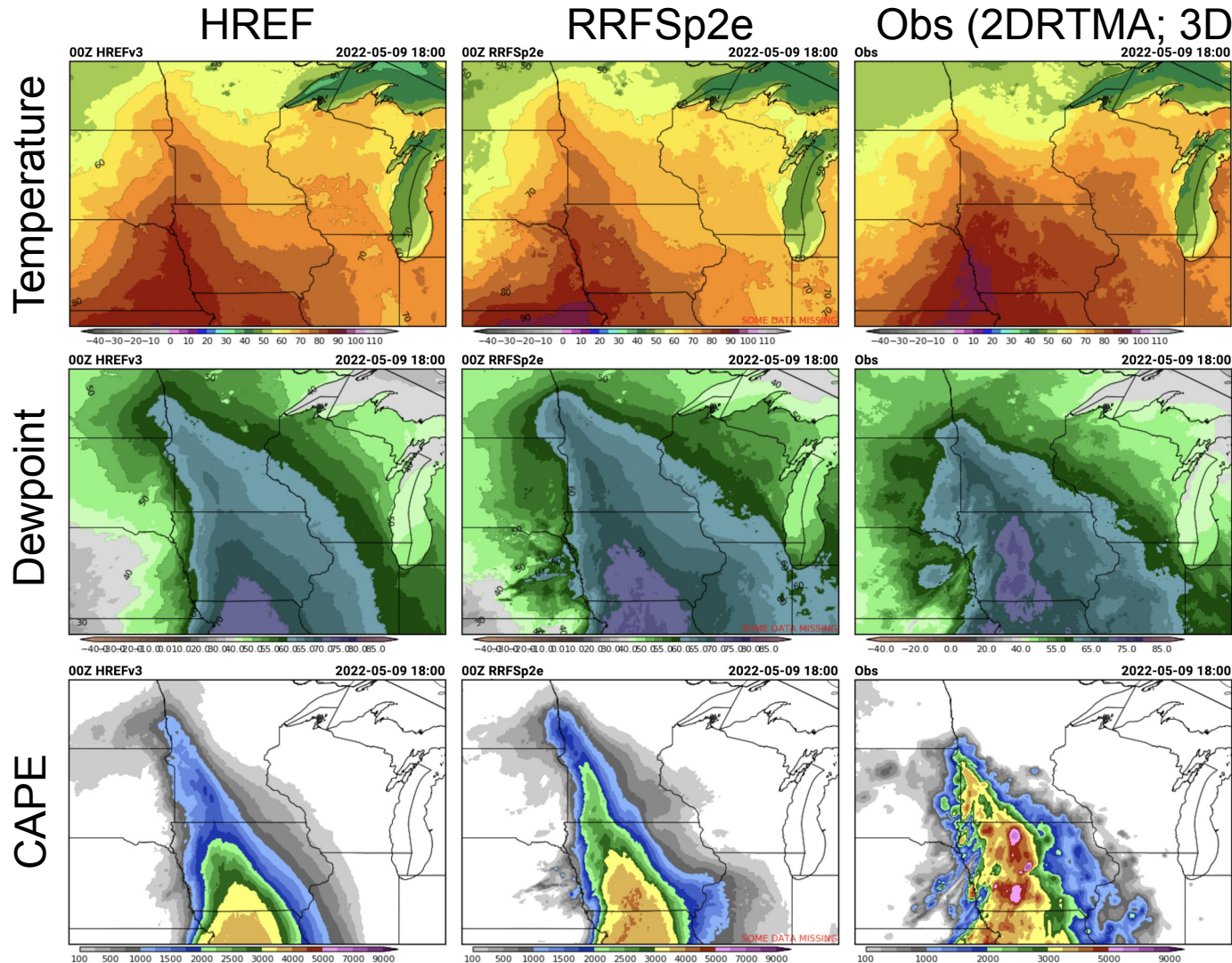
7 cases where only RRFS MixPhys was missing

- HREF and RRFSp2e average ranking very similar with RRFSp2e having slight advantage



SFE 2022: RRFSp2e vs. HREF

- **Goal:** Provide comprehensive evaluation of environment and storm attribute fields in RRFSp2e compared to HREF.



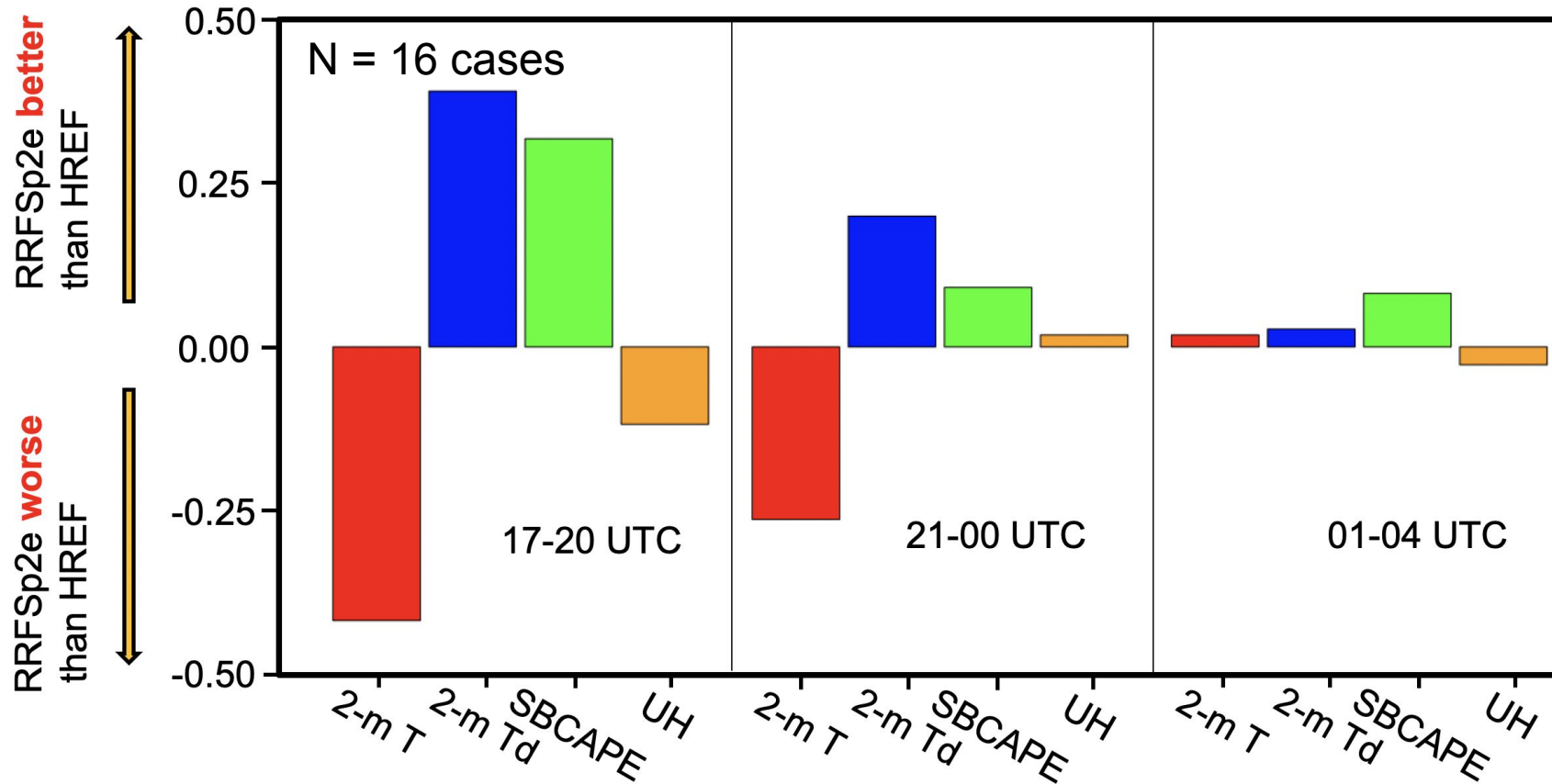
Participants were asked: For the following ensemble mean environmental fields (T, Td, & CAPE) and UH, please rate the performance of RRFSp2e relative to HREF for the time periods 17-20, 21-00, & 01-04Z.

This was the 1st time that ensemble mean environment fields have been examined in the HWT.

SFE 2022: RRFSp2e vs. HREF

Rate RRFSp2e relative to HREF

2 = Much better, 1 = Better, 0 = About the same, -1 = Worse, -2 = Much worse



Temperature: HREF was clearly better; likely result of cool bias in RRFSp2e that was most prevalent earlier in the forecast.

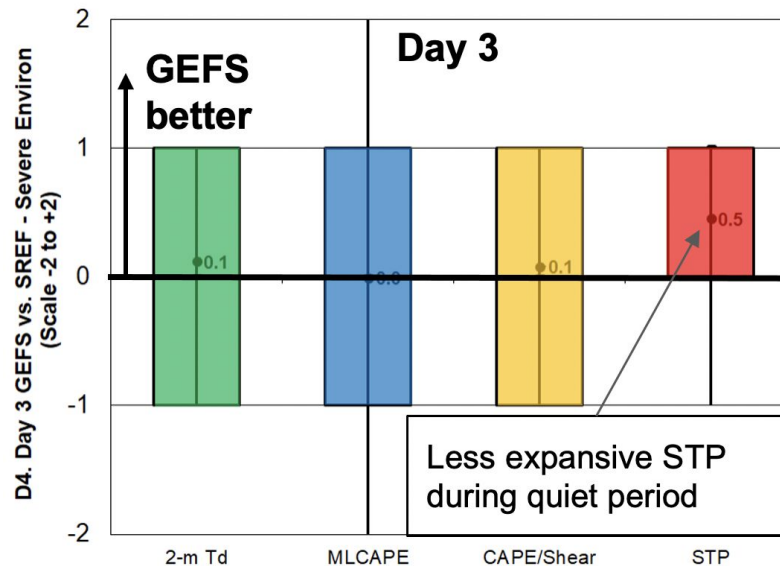
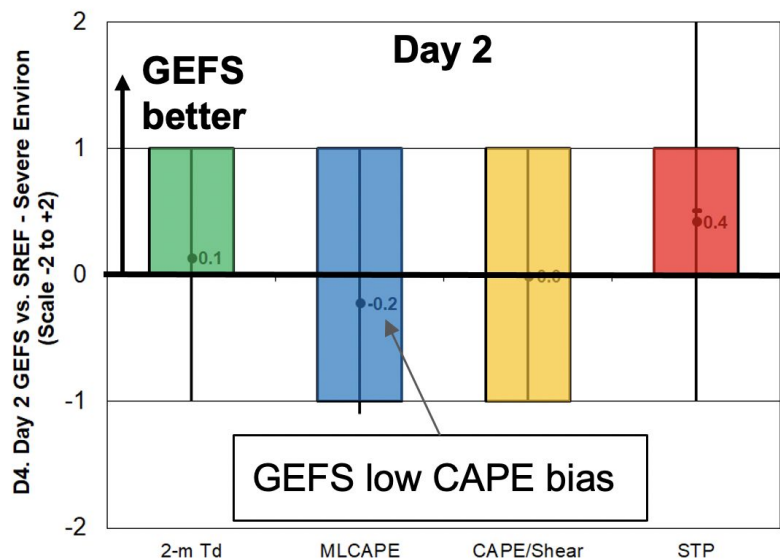
Dewpoint: RRFSp2e was clearly better; likely result of dry bias in HREF that was most prevalent earlier in the forecast.

CAPE: RRFSp2e was clearly better with magnitudes that more closely matched observations.

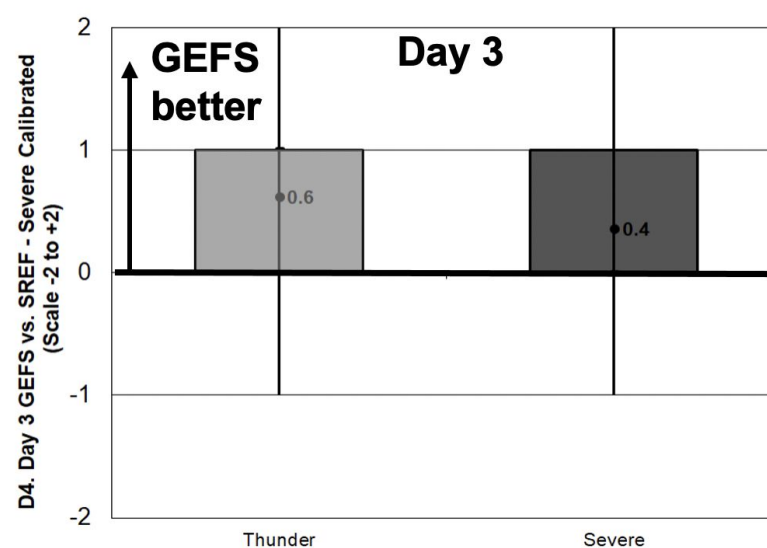
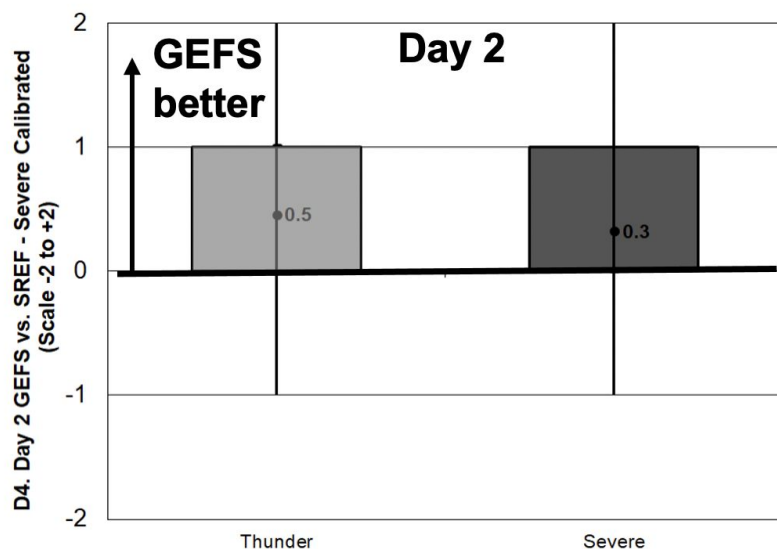
UH: Differences were mostly small.

SFE 2021: GEFS vs. SREF Days 2 & 3

- **Goal:** Evaluate severe weather fields (e.g., Td, MLCAPE, CAPE/Shear combined probabilities, & STP) and calibrated thunder and severe weather probabilities to gauge GEFS readiness to replace SREF.



Overall, GEFS performed as well as SREF for most severe weather fields

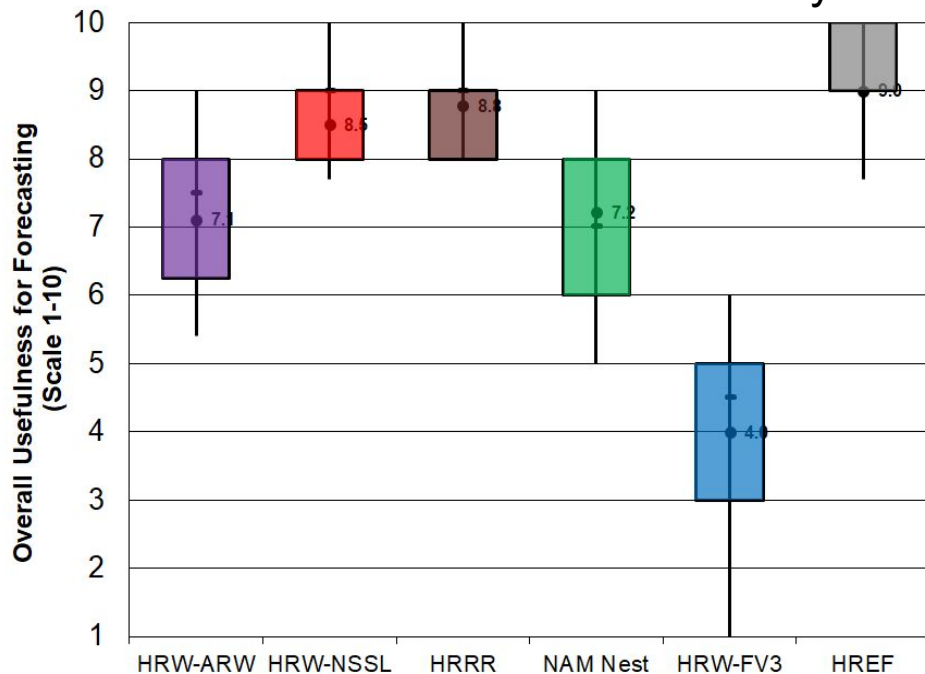


Overall, GEFS performed better than SREF for most severe weather fields

Summary

- It appears that significant progress has been made in the last year.
- HWT evaluations indicate that UFS-based systems are generally performing comparable to the operational systems they will replace or subsume.
- **Important caveats/limitations:** Limited sample & objective statistics not computed yet. Also – initial operational versions of FV3 haven't been received well.
- Warn-on-Forecast team at NSSL is getting unacceptable results with FV3 (big problem with spurious storms at model initialization and unrealistic storm characteristics).

2021 SPC forecaster survey



* "SPC forecasters are extremely concerned and skeptical about moving to an FV3-based CAM ensemble to replace the current HRRR and HREF"

END

