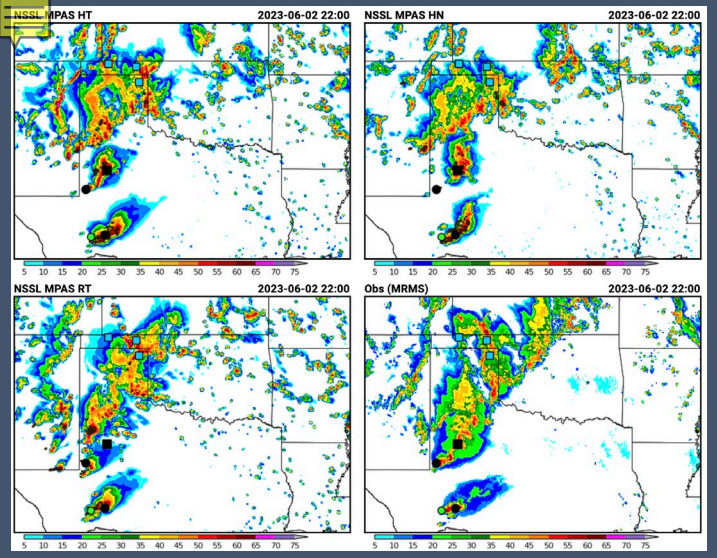
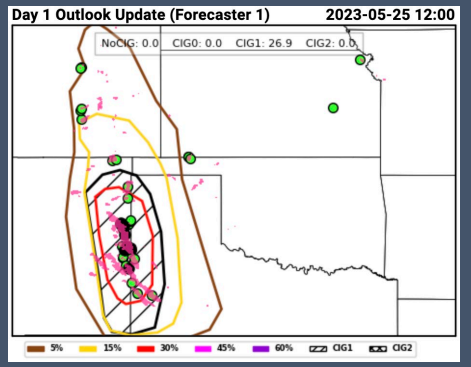
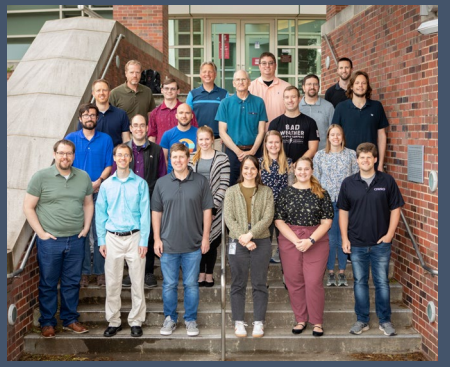






Evaluations of three regional MPAS configurations for severe weather forecasting during the 2023 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment

Unifying Innovations in Forecasting Capabilities Workshop
24-28 July 2023, Boulder, CO

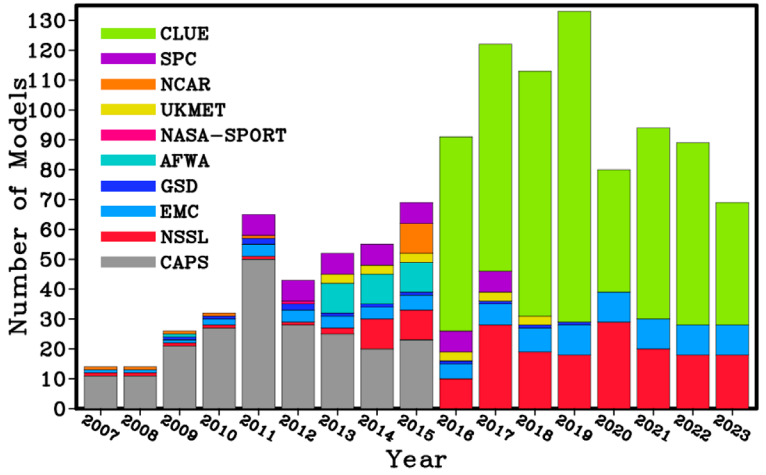
Adam J. Clark[2,4], David Harrison[1,3], Kent Knopfmeier[2,3], Yunheng Wang[2,3], Larissa Reames[2,3], Israel Jirak[1], Louis Wicker[2,4], Pamela Heinselman[2,4], David Dowell[5], Craig Schwartz[6], Michael Duda[6], William Skamarock[6], and Patrick Burke[2]

[1]NOAA/NWS/NCEP Storm Prediction Center, Norman, Oklahoma
[2]NOAA/OAR National Severe Storms Laboratory, Norman, Oklahoma
[3]Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma, Norman, Oklahoma
[4]School of Meteorology, University of Oklahoma, Norman, Oklahoma
[5]NOAA/OAR Global Systems Laboratory, Boulder, Colorado
[6]National Center for Atmospheric Research, Boulder, Colorado

2023 NOAA/HWT Spring Forecasting Experiment

- SFEs are five-week experiments jointly organized and facilitated by SPC and NSSL
- The 2023 SFE was *hybrid* with 50 remote and 77 in-person participants.

HWT SFE Model Contributions



SFE Goals include:

- Testing & evaluation of emerging technologies for severe weather prediction
- Accelerating R2O; developing & strengthening O2R pathways
- Facilitating experiments to: ***optimize deterministic and ensemble CAMs informing Unified Forecast System Development*** [using Community Leveraged Ensemble (CLUE) framework].

Model for Prediction Across Scales (MPAS)

- SFE 2023 included regional MPAS configurations run by NSSL in collaboration with NCAR. Why?
- NSSL seeking next-generation model core to replace WRF in the Warn-on-Forecast System (WoFS) .
- Extensive testing with FV3 found too many spurious storms during model spin-up, inability to recover from early imbalances, & unrealistic storm characteristics.
- We need a model that accommodates (1) further refinements in grid-spacing (i.e., $\leq$ 1-km), (2) advances in DA, and (3) fits within NOAA's UFS framework. First step is testing regional MPAS at Day 1 lead times (i.e., 0-36 h).

NSSL's regional MPAS configurations

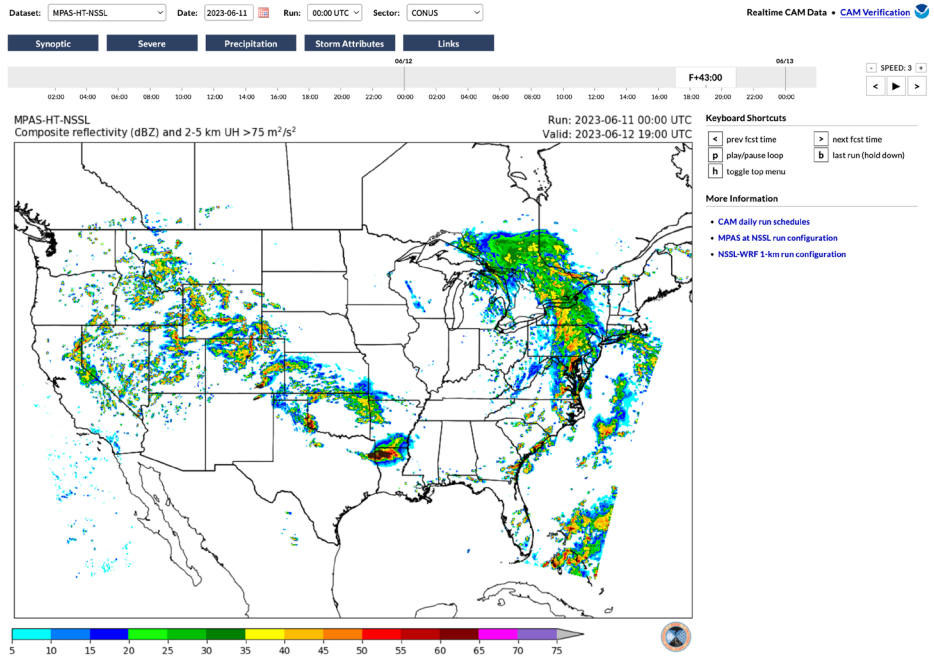
SFE 2023 CLUE Subsets

Clue Subset	# of mems	IC/LBC perts	Mixed Physics	Data Assimilation	Dynamical Core	Agency	Init. Times (UTC)	Forecast Length (h)	Domain
RRFS	10	EnKF	no	Hybrid 3DEnVar	FV3	EMC/GSL	00-23	60/18	CONUS
RRFSphys	9	EnKF	yes	Hybrid 3DEnVar	FV3	EMC/GSL	00-23	60/18	CONUS
NSSL1	1	none	no	HRRR ICs	ARW	NSSL	00	36	2/3 CONUS
NSSL-MPAS	3	none	no	HRRR or RRFS ICs	MPAS	NSSL	00	48	CONUS
GFDL-FV3	1	none	no	GFS cold start	FV3	GFDL	00	126	CONUS
NASA-FV3	1	none	no	GEOS-DA	FV3	NASA	00	120	CONUS
NCAR-FV3	10	GEFS	no	GEFS cold start	FV3	NCAR	00	192	CONUS
NCAR-MPAS	5	GEFS	no	GEFS cold start	MPAS	NCAR	00	132	CONUS

NSSL MPAS

NCAR MPAS

cams.nssl.noaa.gov



NSSL-MPAS Configurations

Table 5 Specifications for the NSSL-MPAS CLUE members. These members use 3-km grid-spacing covering the CONUS and are driven by the HRRR or RRFS. The last two letters of each member denote the ICs and microphysics ("HN" = HRRR-NSSL, "HT" = HRRR-Thompson, and "RT" = RRFS-Thompson).

Member: NSSL-MPAS	ICs	LBCs	Microphysics	PBL	LSM	Radiation	Dynamical Core
NSSL-MPAS-HN	HRRR	HRRR	NSSL	MYNN	RUC	RRTMG	MPAS
NSSL-MPAS-HT	HRRR	HRRR	Thompson	MYNN	RUC	RRTMG	MPAS
NSSL-MPAS-RT	RRFS	RRFS	Thompson	MYNN	RUC	RRTMG	MPAS

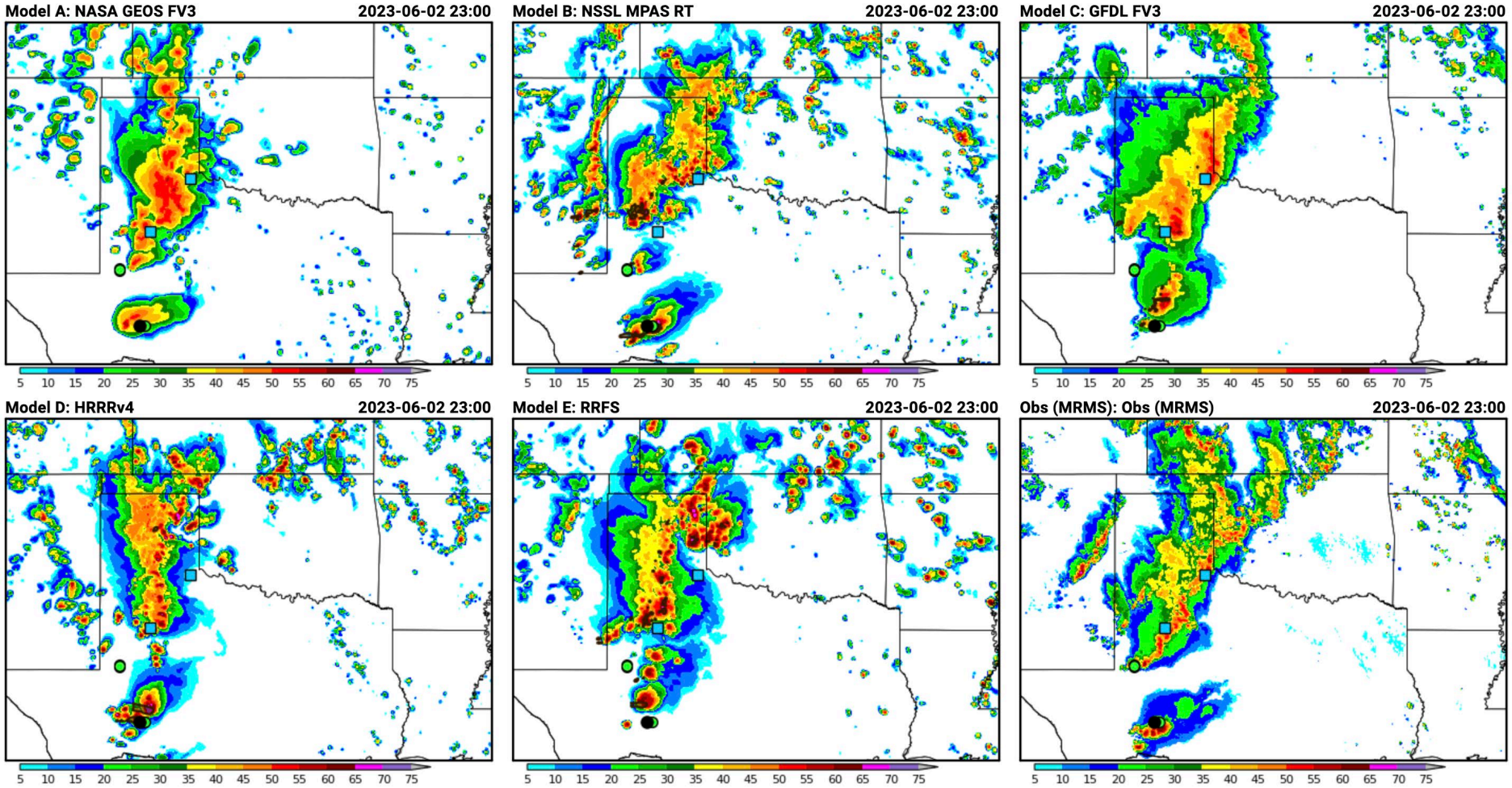
SFE 2023 Evaluations

MPAS-related evaluations

(D)eterministic CAMS	CAM (E)nsembles	(C)alibrated Guidance	(A)nalyses	Funded (P)rojects
D1: Day 1 Deterministic Flagships	E1: 00Z RRFS vs. HREF	C1-3: Calibrated Tornado Guidance	A1: Mesoscale Analysis Background	P1: ISU ML Severe Wind Probabilities
D2: Day 2 Deterministic Flagships	E2: 12Z Day 1 RRFS Physics & Time-Lagging vs. HREF	C4-7: Calibrated Hail Guidance	A2: Storm Scale Analysis	P2: WoFS Loken ML Guidance
D3: RRFS vs. HRRR	E3: 12Z Day 2 RRFS Physics & Time-Lagging vs. HREF	C8-10: Calibrated Wind Guidance		
D4: RRFS vs. HRRR Data Assimilation	E4: Medium-Range Lead Time/Core/Members	C11: Medium Range 00Z GEFS Total Severe		
D5: NSSL MPAS configurations		C12: Day 1 HRRR Neural Network Hazard Guidance		
D6: NSSL1 vs. HRRR				

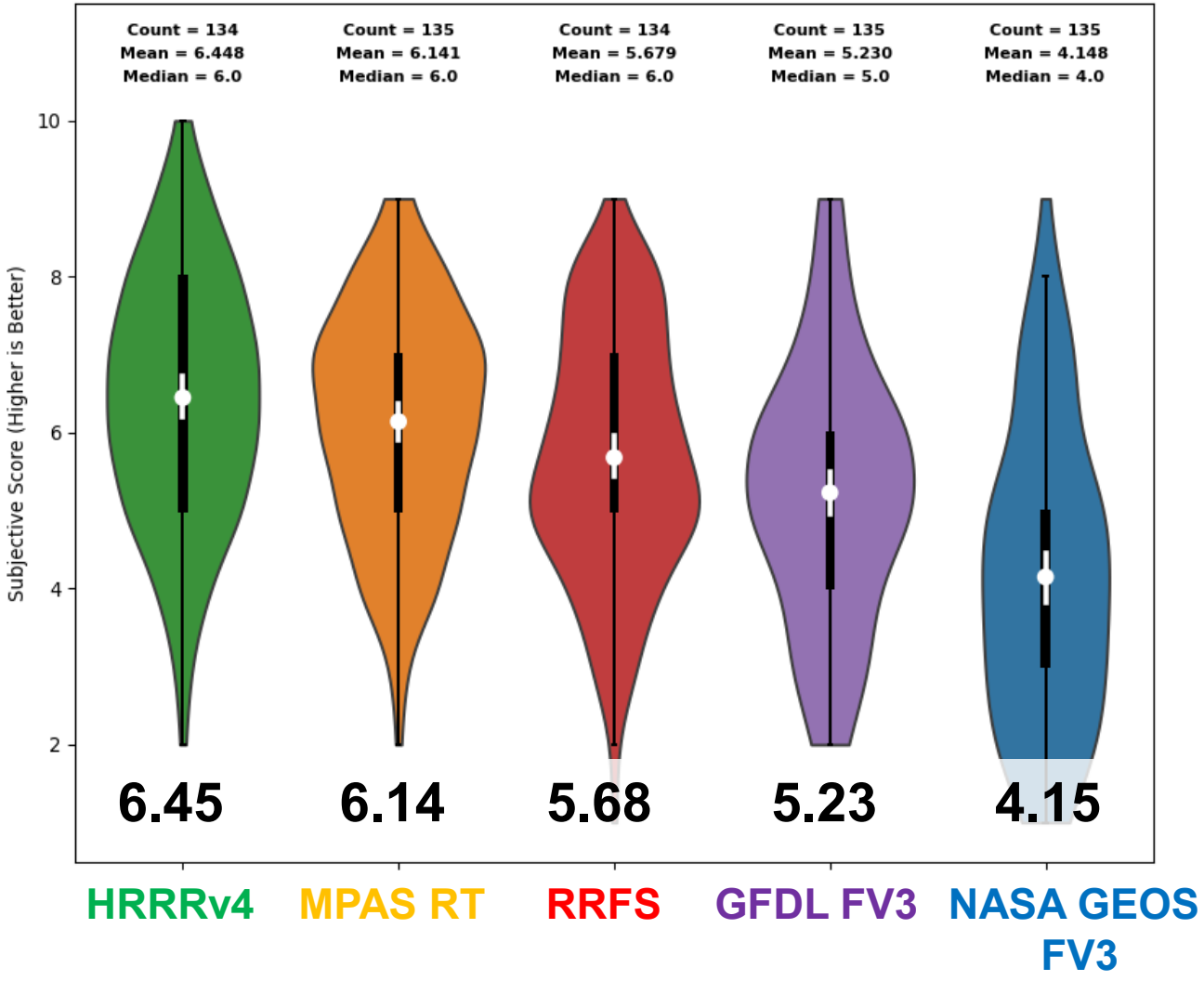
Day 1 00Z Deterministic Flagships

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



- UH & composite reflectivity evaluated over forecast hours 12-36. Consider timing of CI, convective mode, displacement errors, etc.
- **Comparisons were blinded**, and models revealed after ratings were assigned.
- MPAS RT used here because forecasts went to 60 h, and we were most confident in its performance.

Deterministic Flagships: Composite Reflectivity & UH



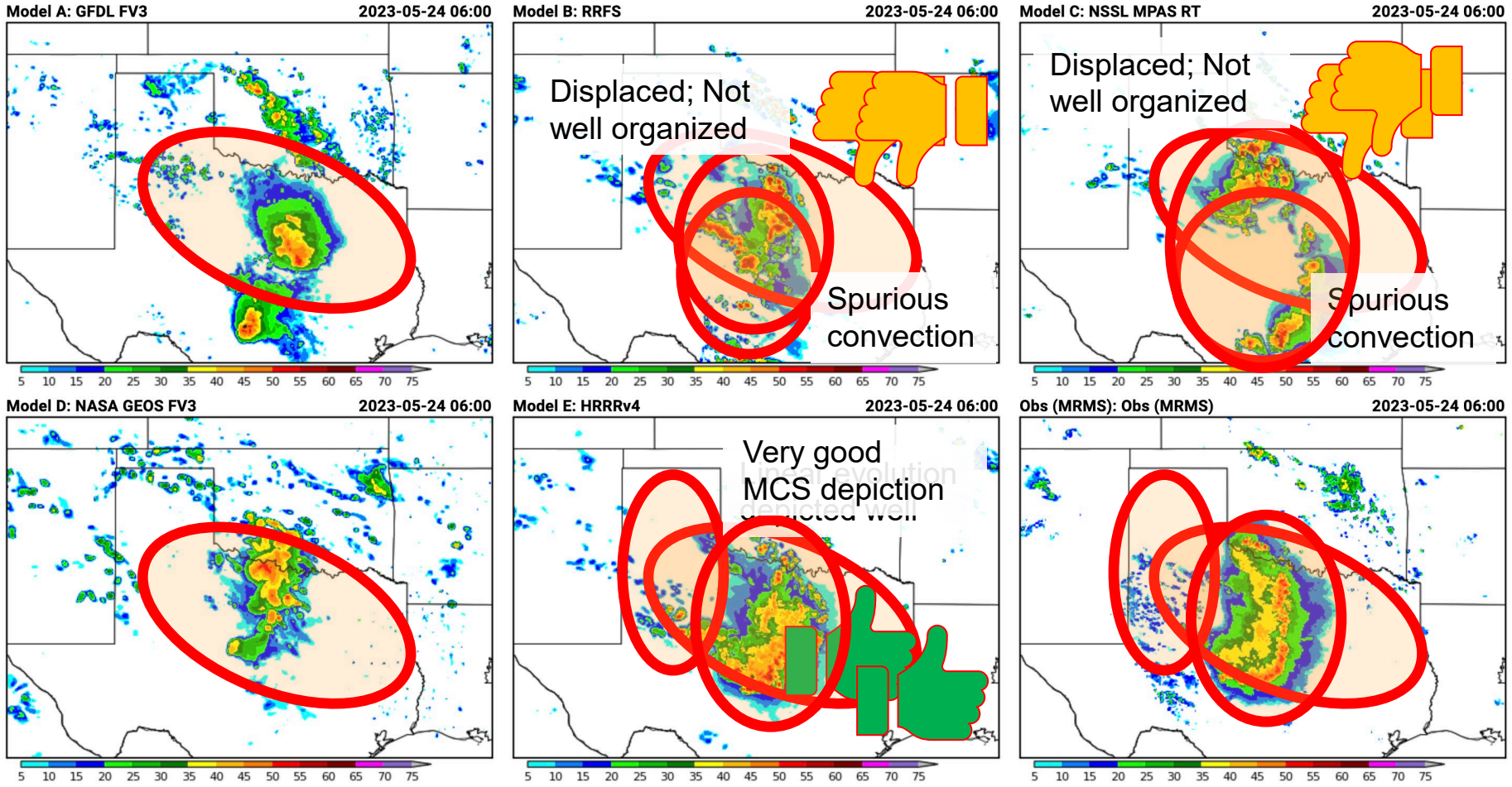
HRRRv4 was clear top performer & differences in averages ratings were significant (paired student's t-test).

MPAS RT was clear runner-up & differences w.r.t lower rated models were significant (including RRFS).

2022 results showed similar performance in HRRRv4 and RRFS, so the 2023 results are a stark difference. What happened?

(1) RRFS radar DA implemented in 2023. Clear problems with intensity bias and spurious convection near initialization time. NSSL found similar issues with FV3 in WoFS.
(2) Quiet weather regimes with weakly forced events. Intensity biases in FV3 seem more apparent in these regimes.

Example Case: 23 May 2023



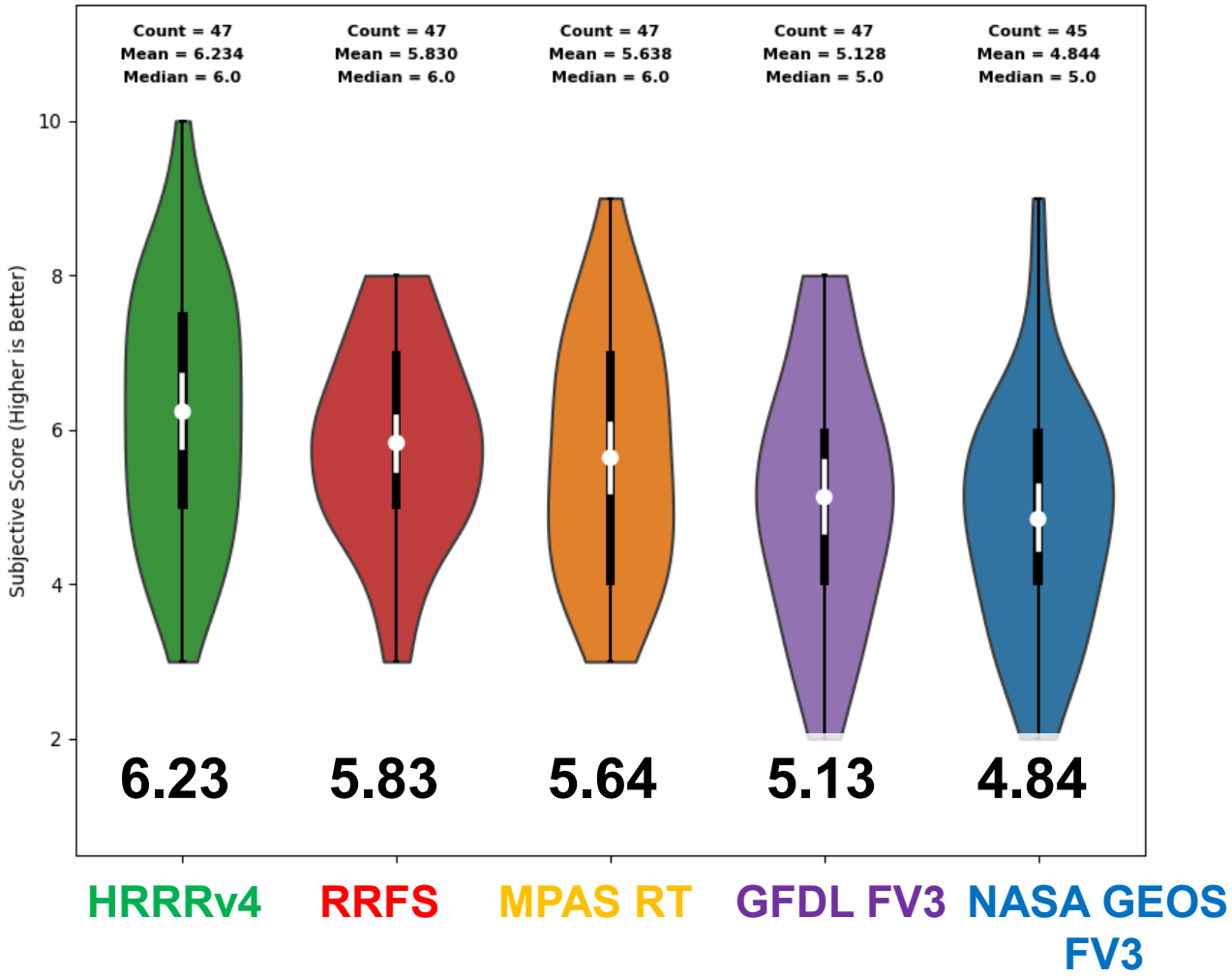
f00-12: Dissipating MCS/MCV depicted by all models. HRRR, RRFS, and MPAS have more realistic structure.

f24: Convection grows rapidly upscale into linear system. HRRR is only model that depicts linear evolution well. MPAS and RRFS have spurious convection in central TX, which improperly affects downstream environment.

f30: HRRR is clearly superior, but a little too intense. RRFS and MPAS displaced and not well organized.

Average Scores:
HRRR: 7.7
MPAS: 5
RRFS: 5.2
NASA: 4.8
GFDL: 5.2

Deterministic Flagships: 2-m Temperature



- HRRRv4 rated highest & RRFS was runner up. MPAS RT rated 3rd.

We asked: "What characteristics of the forecast were most important to you when rating the 2-m T fields?"

Sample of answers:

"boundary placement and evolution, convective feedback and outflow progression, actual temperature accuracy"

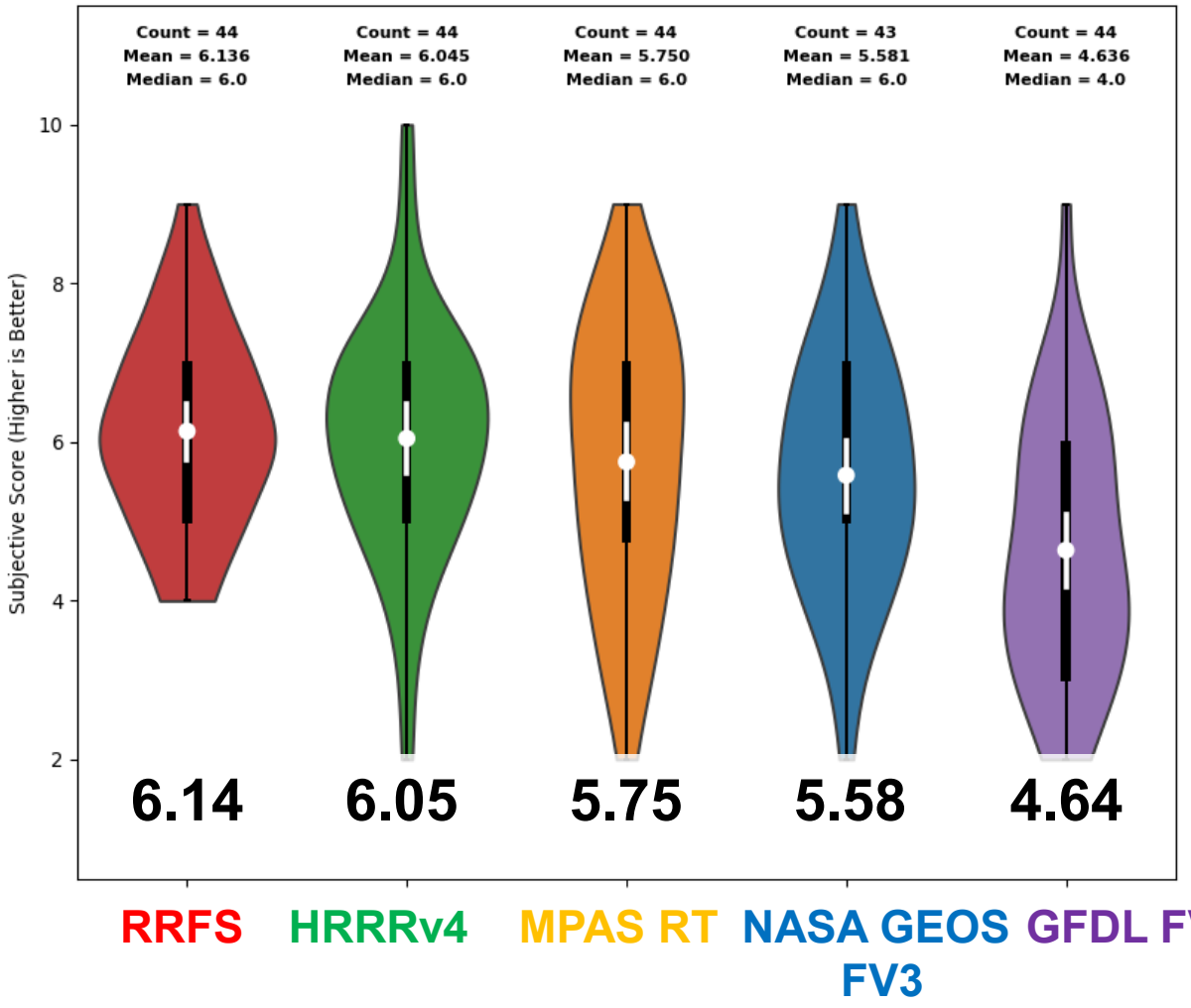
"Ahead of CI ... all were too warm in focus areas."

"All models where way too warm in OK, highlighting how they all struggled with the morning MCV and convection in its wake ..."

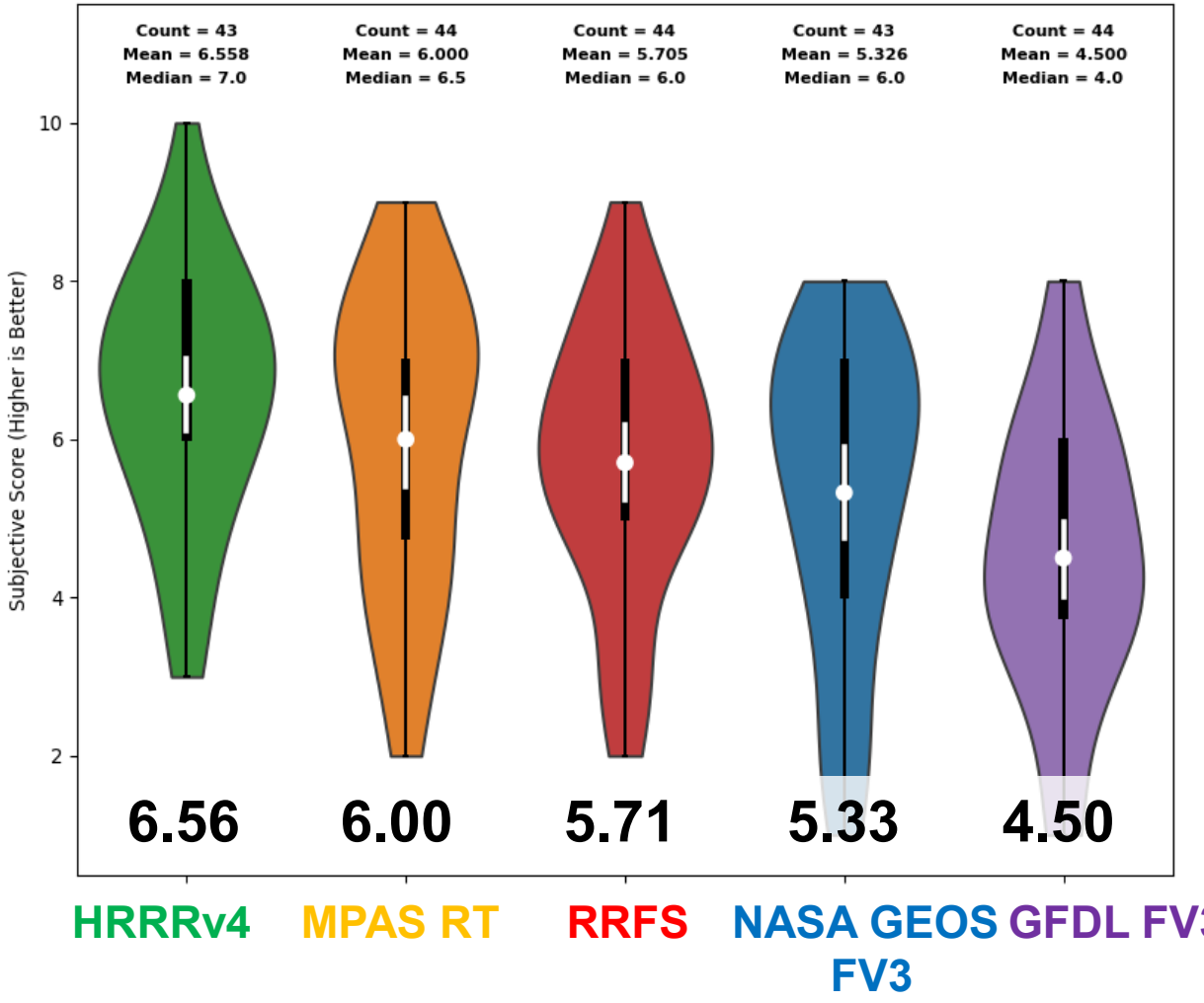
"Most important was the magnitude of the temperatures because on dryline days and severe weather days like this one that is important to get right and can impact how the model initiates convection."

Deterministic Flagships: 2-m Td & surface-based CAPE

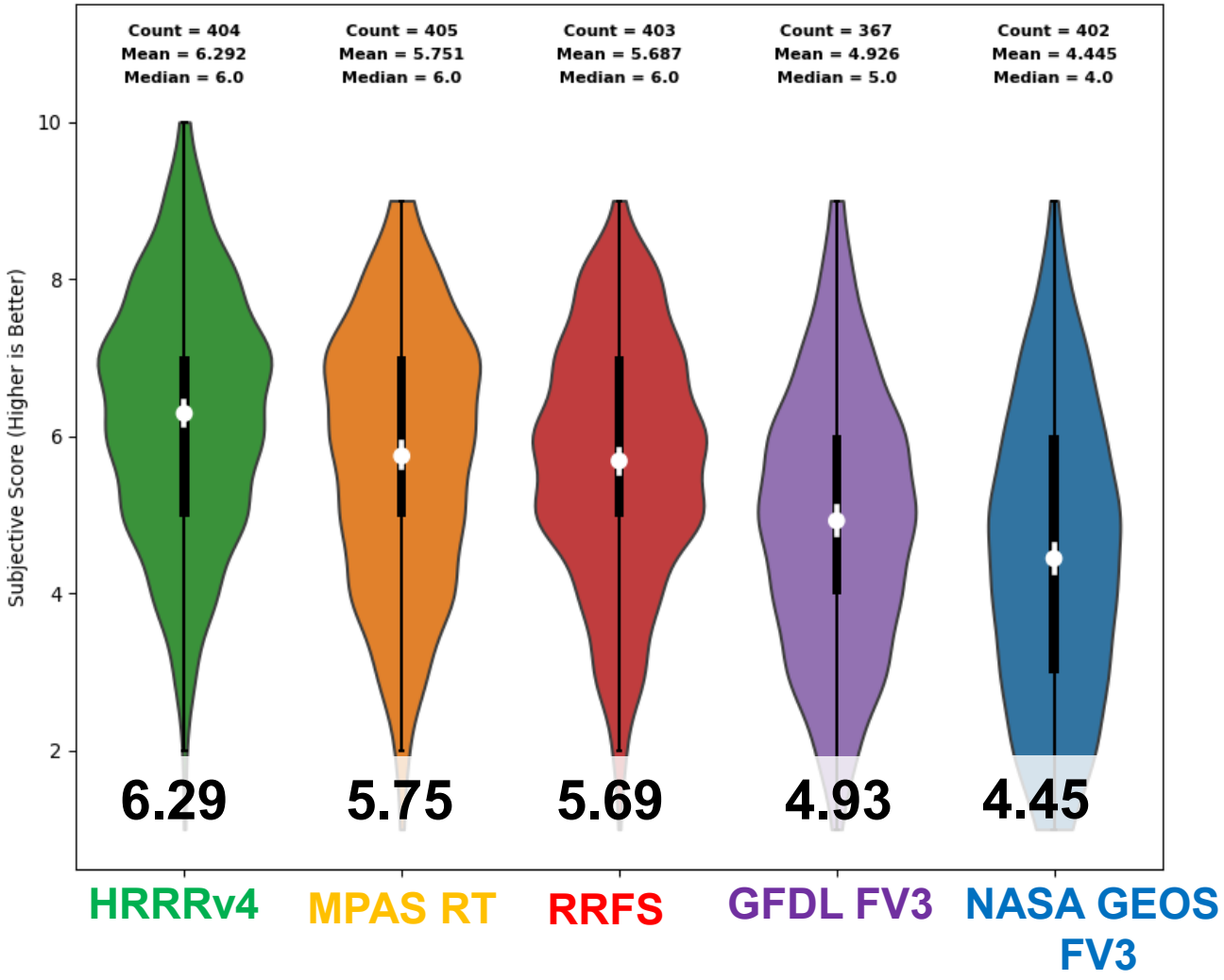
2-m Dewpoint



Surface-based CAPE



Deterministic Flagships: Combined Results



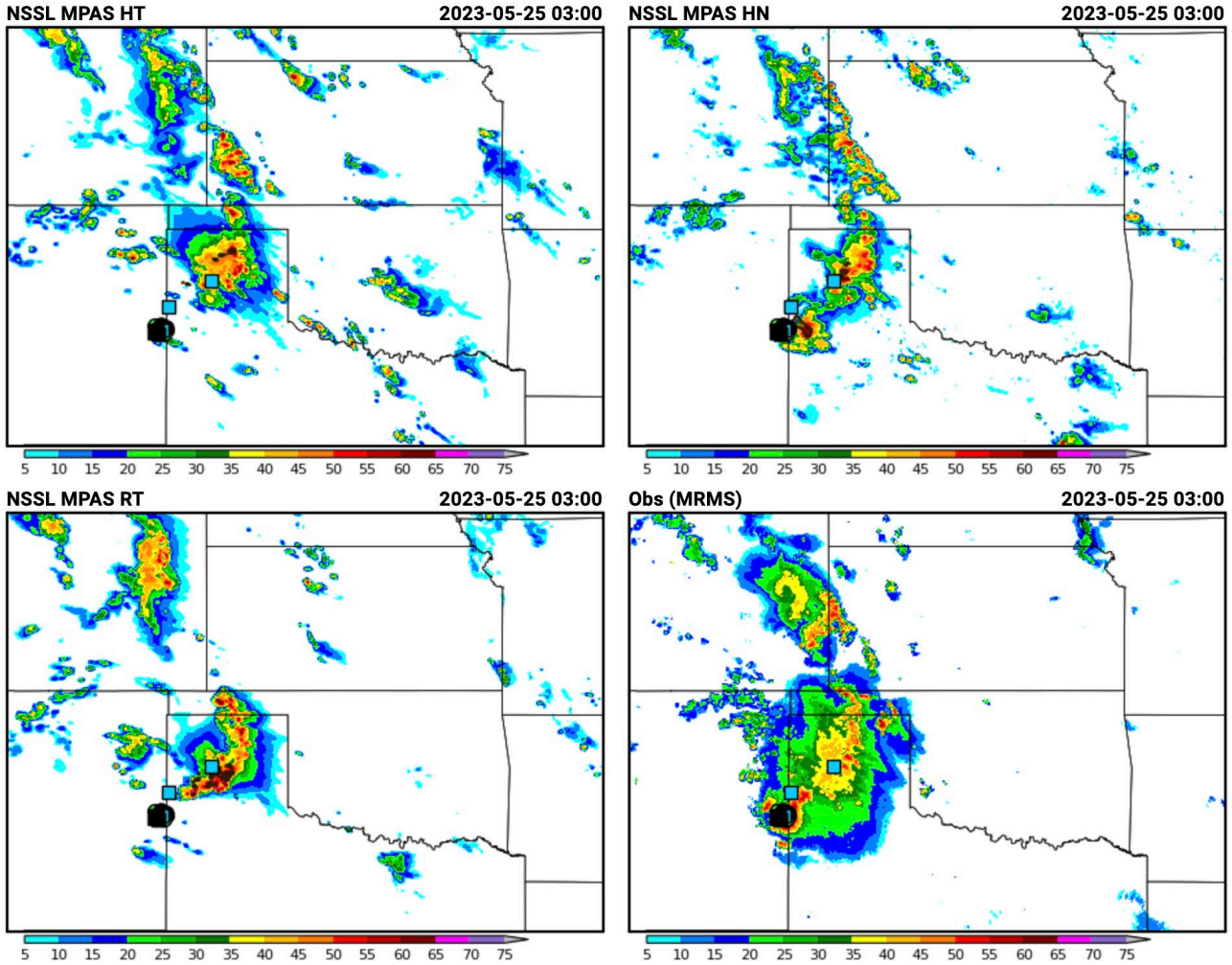
HRRRv4 clearly performed best.

MPAS is runner-up and performs slightly better than RRFS, from which MPAS was initialized.

GFDL and NASA GEOS performed relatively poorly.

NSSL MPAS Configurations

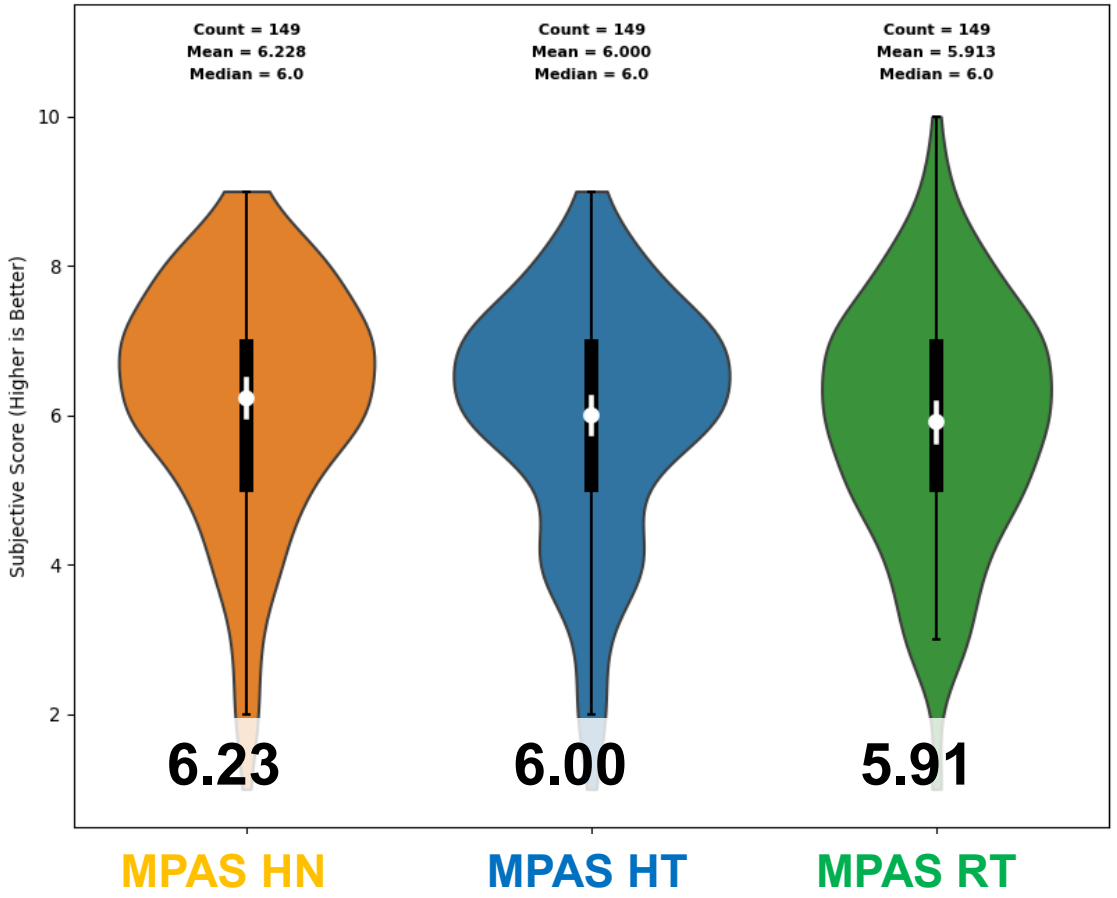
- **Research Questions:** How does the MPAS initialization dataset (RRFS vs. HRRR) and microphysics scheme (Thompson vs. NSSL) impact MPAS performance?



- UH and composite reflectivity evaluated over forecast hours 01-36. Consider timing of CI, convective mode, displacement errors, etc.
- Recall, MPAS RT was evaluated in the flagship comparison. Can the HRRR initialized runs perform better?

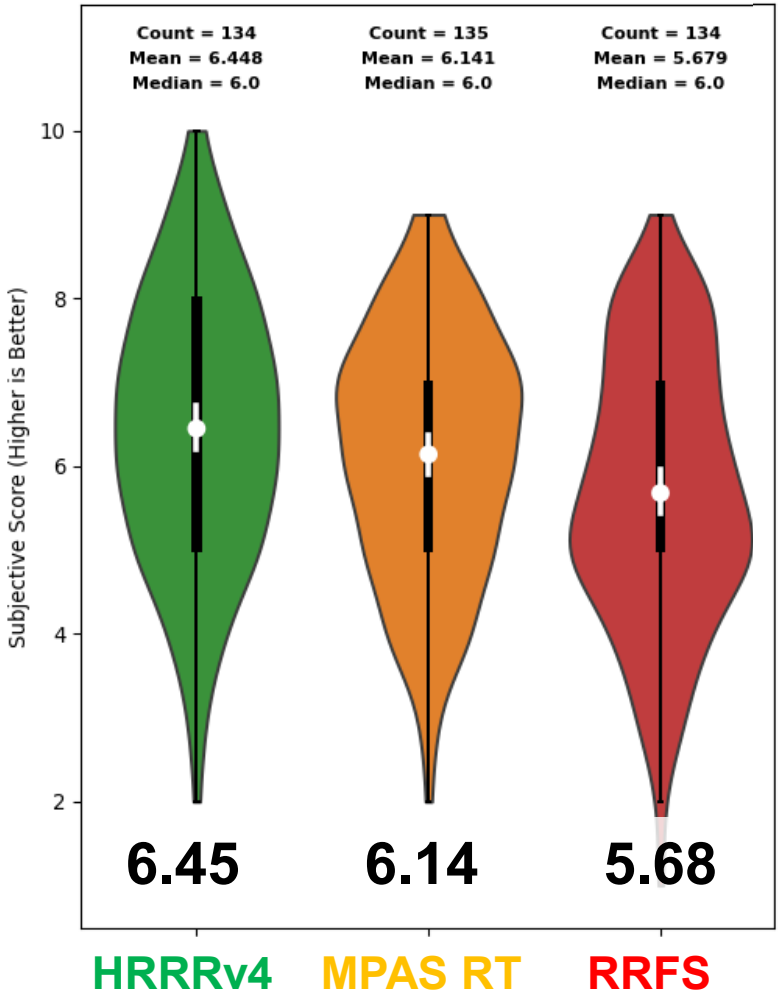
NSSL MPAS Configurations

- NSSL microphysics outperforms Thompson
- HRRR initializations outperform RRFS



- MPAS HN may have similar performance to HRRRv4

With MPAS HN favored over MPAS RT, the MPAS HN ratings are approaching the HRRR.



What is going on?

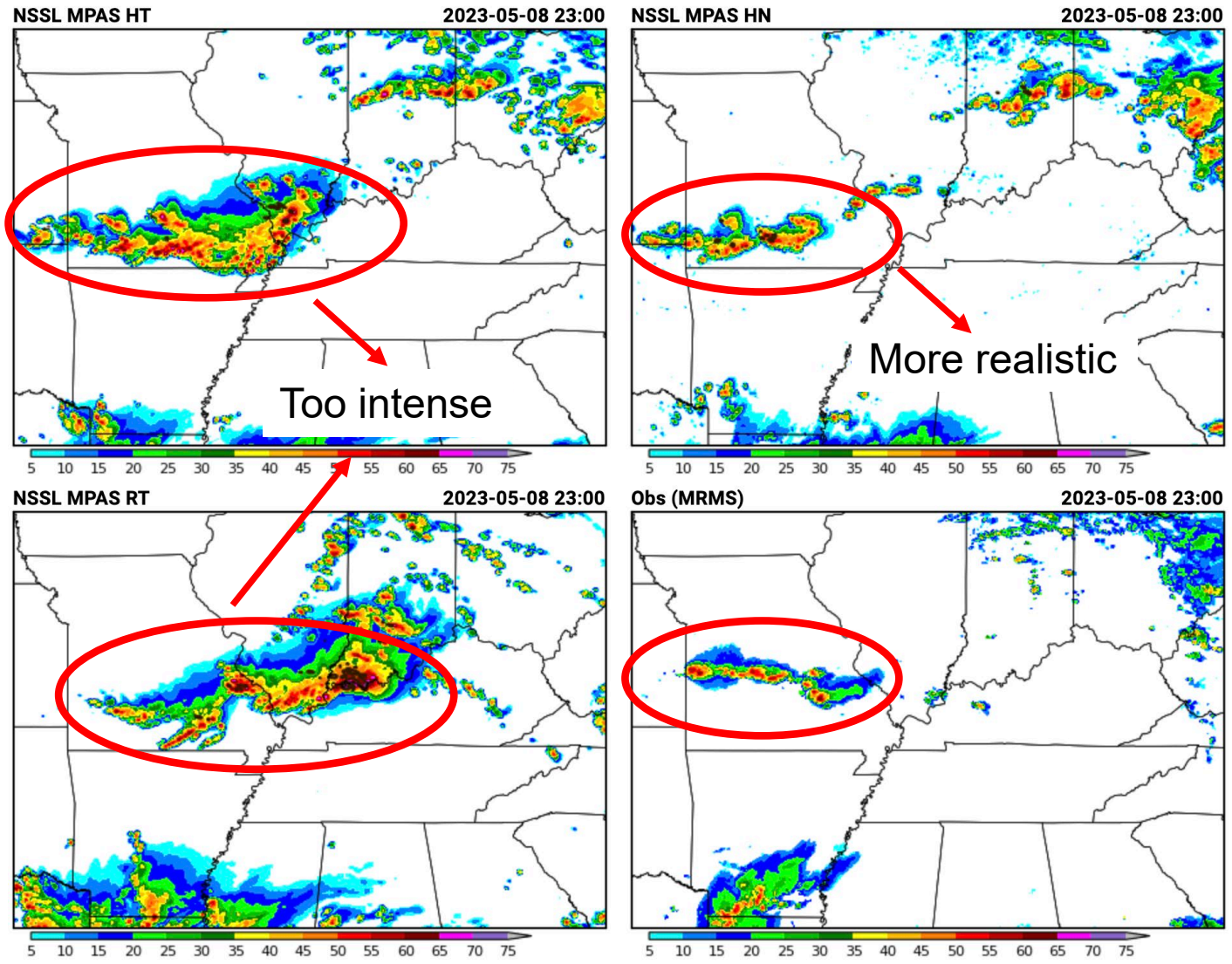
Microphysics: General theme from comments was that MPAS-HN better depicted intensity, model, and convective evolution.

"*The HN performed better with placement, mode, and intensity later in the period*."

"*The HN was almost always most faithful to the actual storm evolution.* The HT often slightly overdid convection, while the RT appeared to underdo nocturnal convection and overdo diurnal convection."

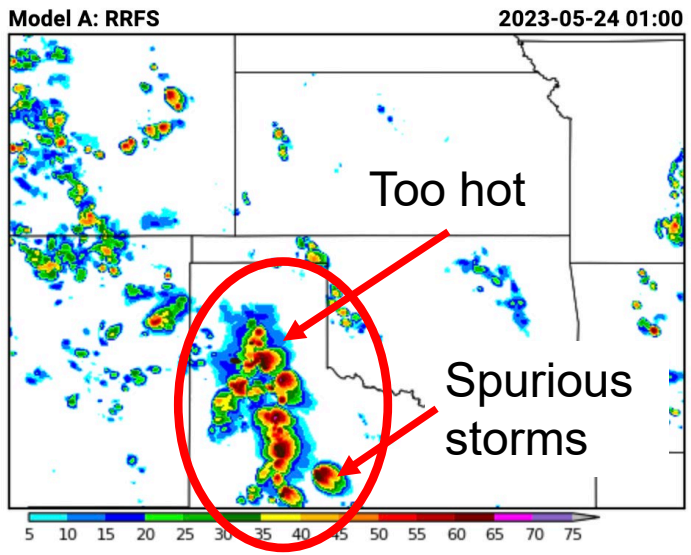
"*HN preferred for organisation and structure*. RT and HT both offered hot reflectivity, RT especially so, while both also offered excess structure and organisation (although they had the right idea) - both holding on to storms too long into the post-00Z period."

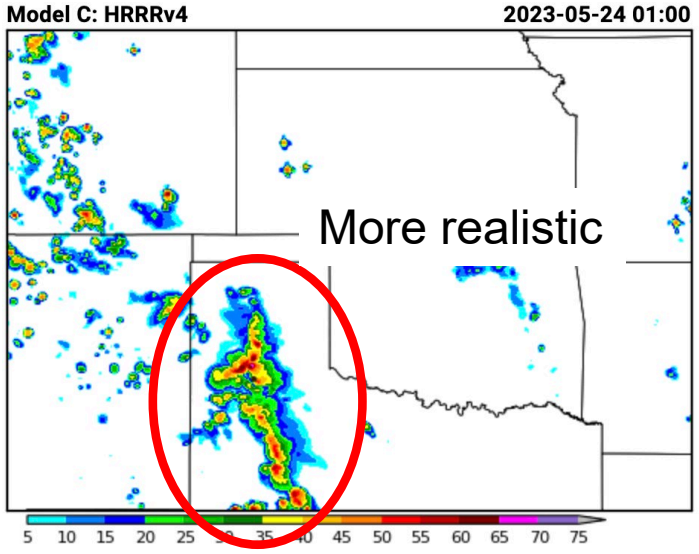
"RT actually handled the previous overnight better in Arkansas. But for the main event in Kansas during the day these were all fairly impressive, but especially HT and HN which had a clear bowing structure. *HN better depicted the narrow stratiform region with the primary bow*, but both HT and HN looked very very good".

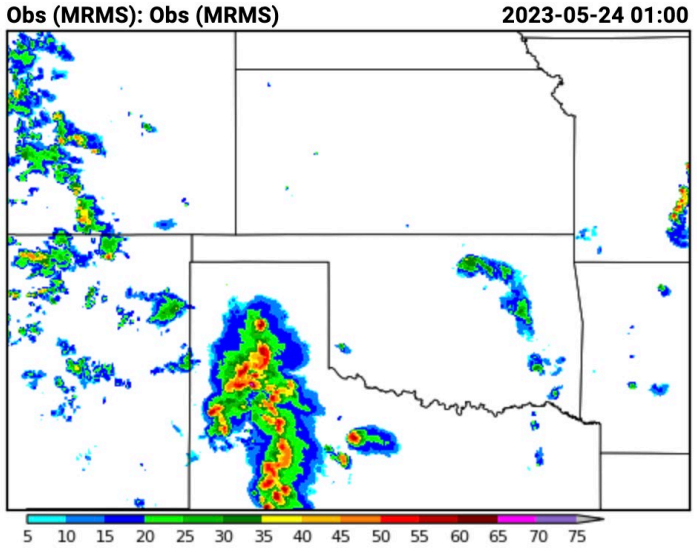


What is going on?

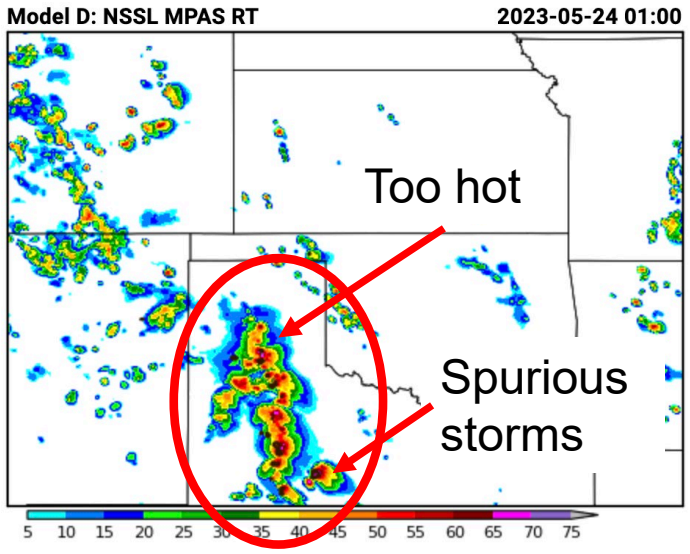
Initial Conditions: It is clear that MPAS RT runs are inheriting errors in RRFS initial conditions.

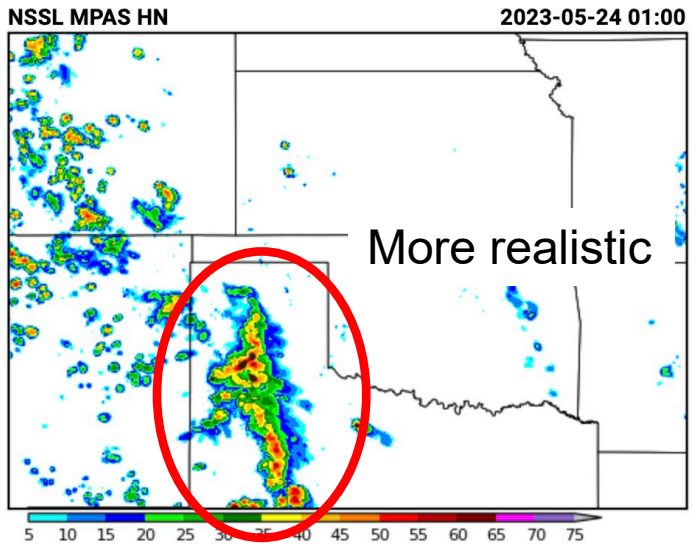


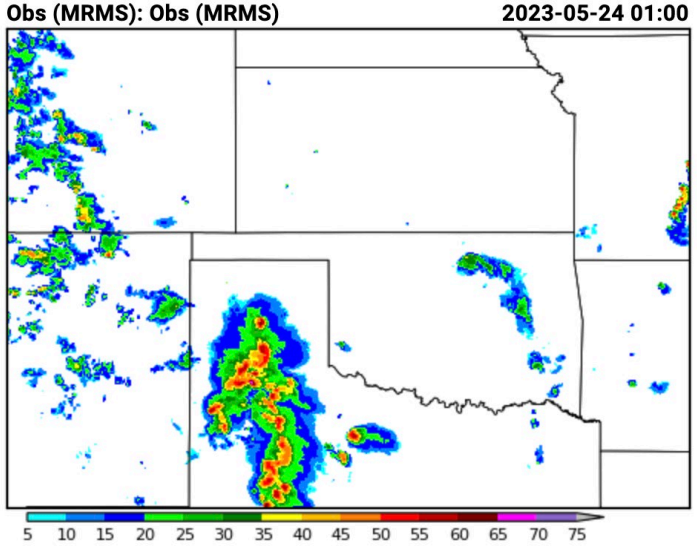




Initialization datasets



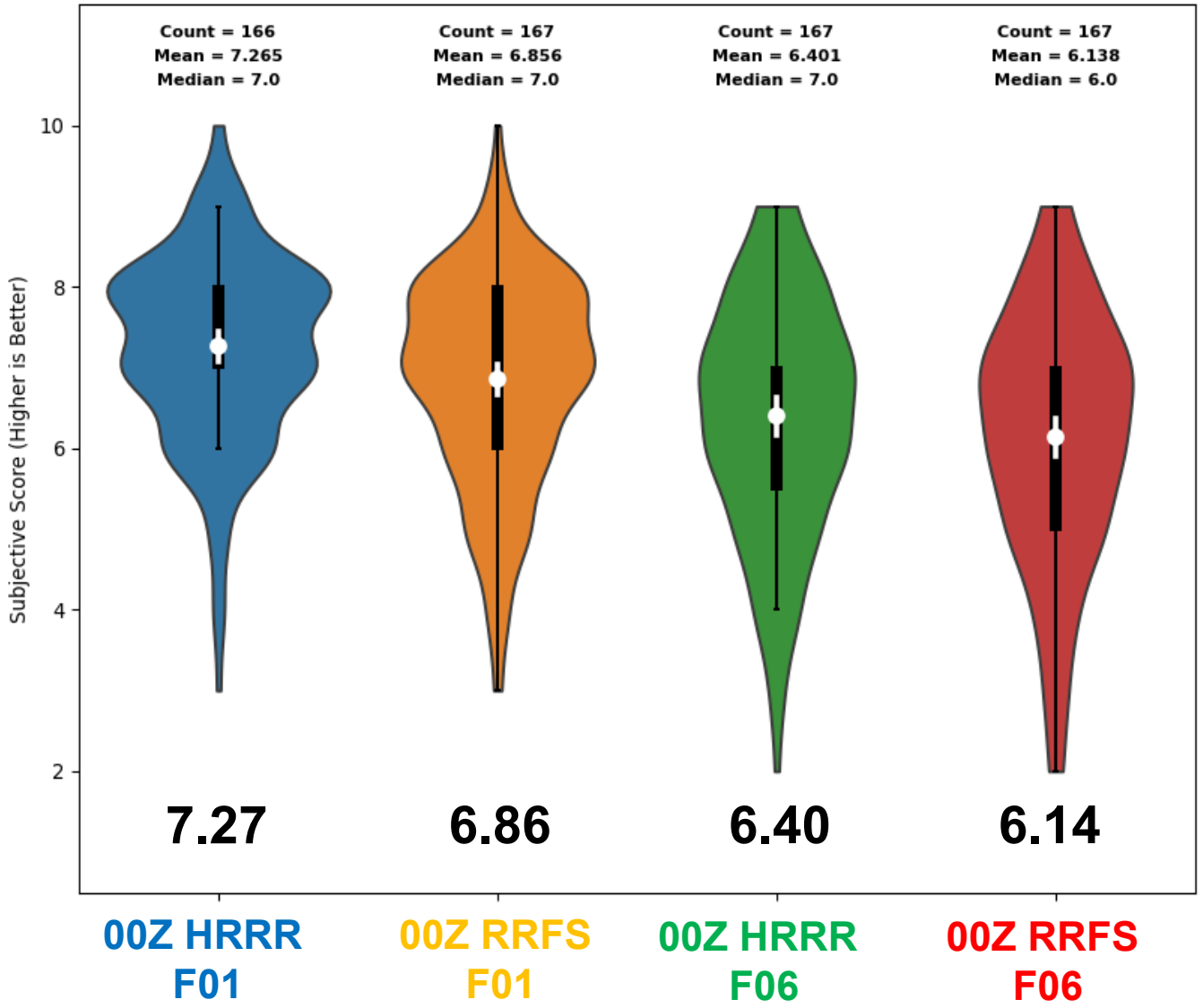




Corresponding MPAS runs

To drive this home...

Evaluation D4 examined forecast hours 1 & 6 in RRFS vs. HRRR to assess DA impact on forecast quality.



- Clear advantage in HRRR.

Sample of comments:

"*RRFS initializations were 'hot*'"

"Both models were *a bit more cellular than the obs, especially the RRFS*. *The RRFS also had way more storms than the HRRR* and obs had."

"Overall, the two models produced similar depictions of convective evolution, although *the RRFS had a higher reflectivity bias, including several areas of spurious storms*."

"*RRFS echoes are too intense*, especially compared to the HRRR4 and the observations."

"Sig. difference between RRFS and HRRRv4 - *RRFS misleads with spurious convection in the south* (TN et al.); difference between the two models quite pronounced. Interestingly, difference between the two DA cycles less pronoucned."

"*RRFS has stronger and more storms at each time than HRRR, and consistently has erroneous convection in TN and stronger convection than realized*. HRRR does much better at the intensity and location of storms, particularly one hour after initialization times."

"*RRFS starts hot at initialization and seems to carry some of this "excess" convection over into the start of the forecast.* HRRR 21z seemed to do well with overall evolution...00z HRRR not quite as good."

Summary and Conclusions

- NSSL's MPAS configurations performed very well and the 00Z MPAS HN (HRRR/NSSL) subjective ratings approached those of the HRRR during the Day 1 period (i.e., f12-f36).

- MPAS RT outperforms the RRFS from which it was initialized. RRFS has major issues with storms that are too intense and spurious convection, which are most apparent at initialization, diurnal peak, and the uncapped warm sector in weakly forced environments.

- **Not shown:** For CAM-based extended range prediction, MPAS (NCAR configuration) received significantly higher ratings than FV3 (C-SHiELD configuration). Note: The only similarity between these runs was the initial conditions from the GEFS.

- What's next?

- NSSL will test an MPAS-based WoFS system this Fall. We will also implement some fixes to improve the MPAS environment forecasts.

- **Take Away:** Tremendous progress was made with regional MPAS over a very short development period of about 6 months. Big credit to developers at NCAR and NOAA management that has supported this work.

NSSL's Warn-on-Forecast group is hiring!

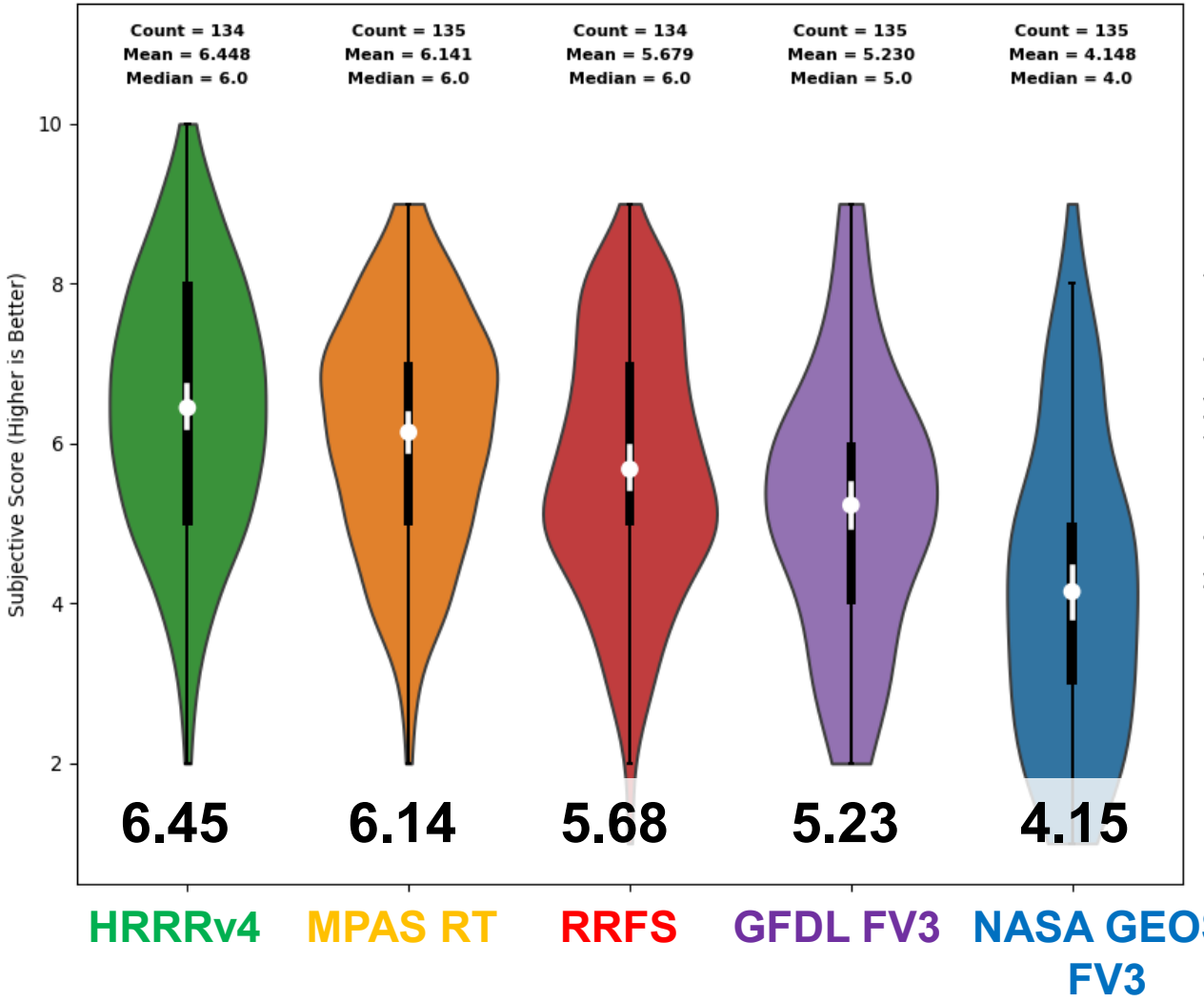
- Opening for MS/PHD to work on a 3D convective reanalysis project.

- See: https://ciwro.ou.edu/careers

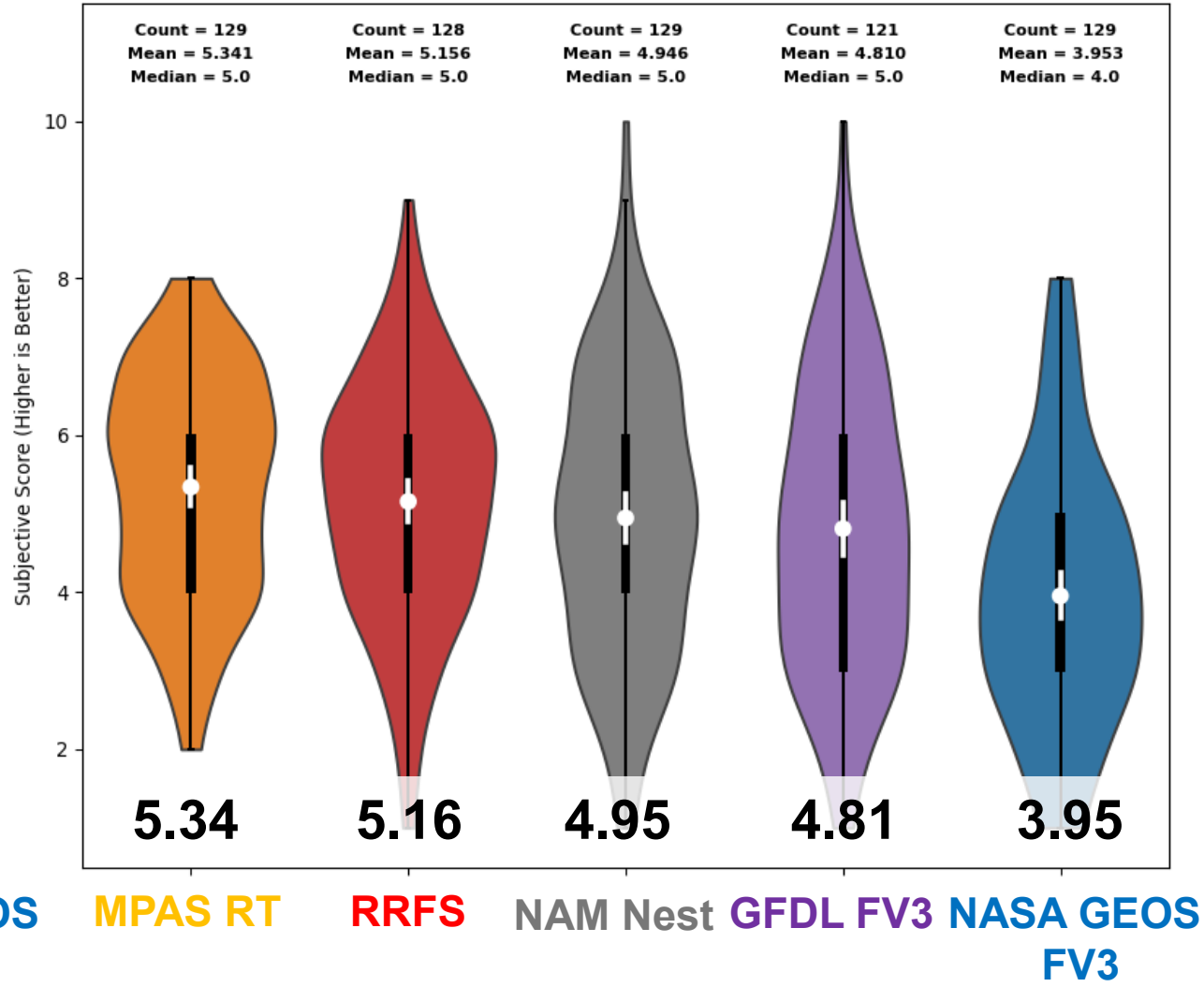
END

Days 1 & 2 Deterministic Flagships: Composite Reflectivity & UH

Day 1

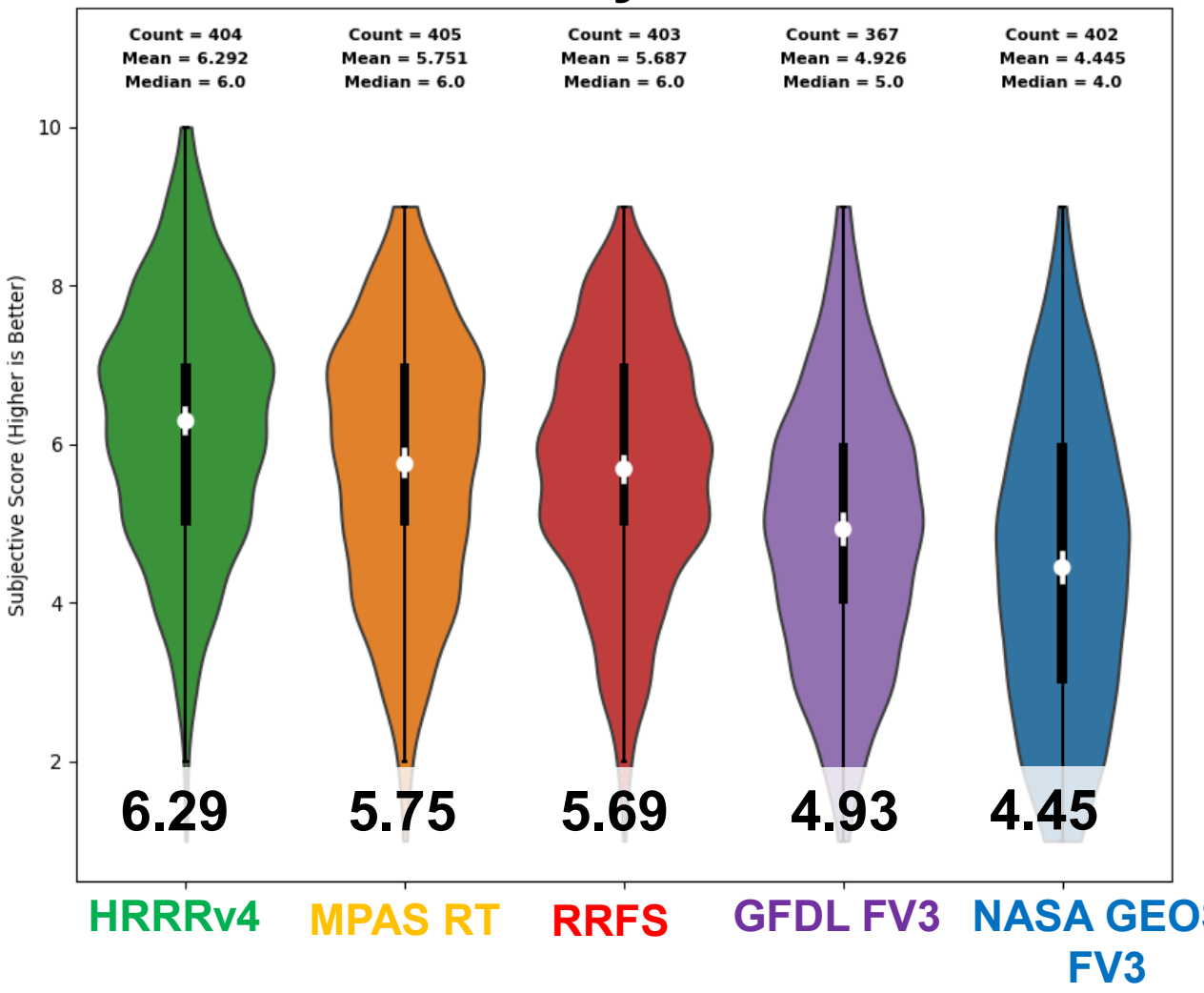


Day 2

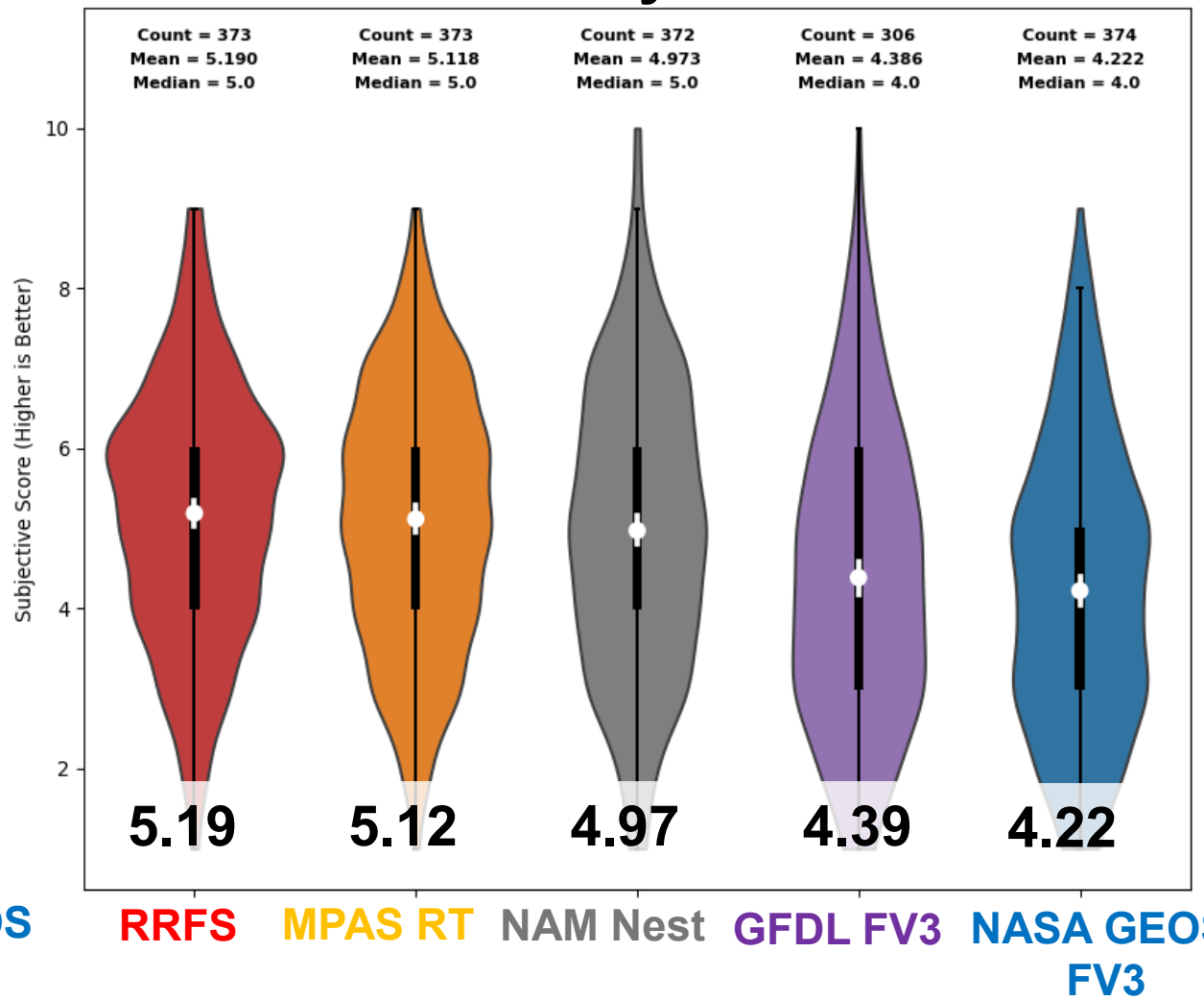


Deterministic Flagships: Combined Results

Day 1

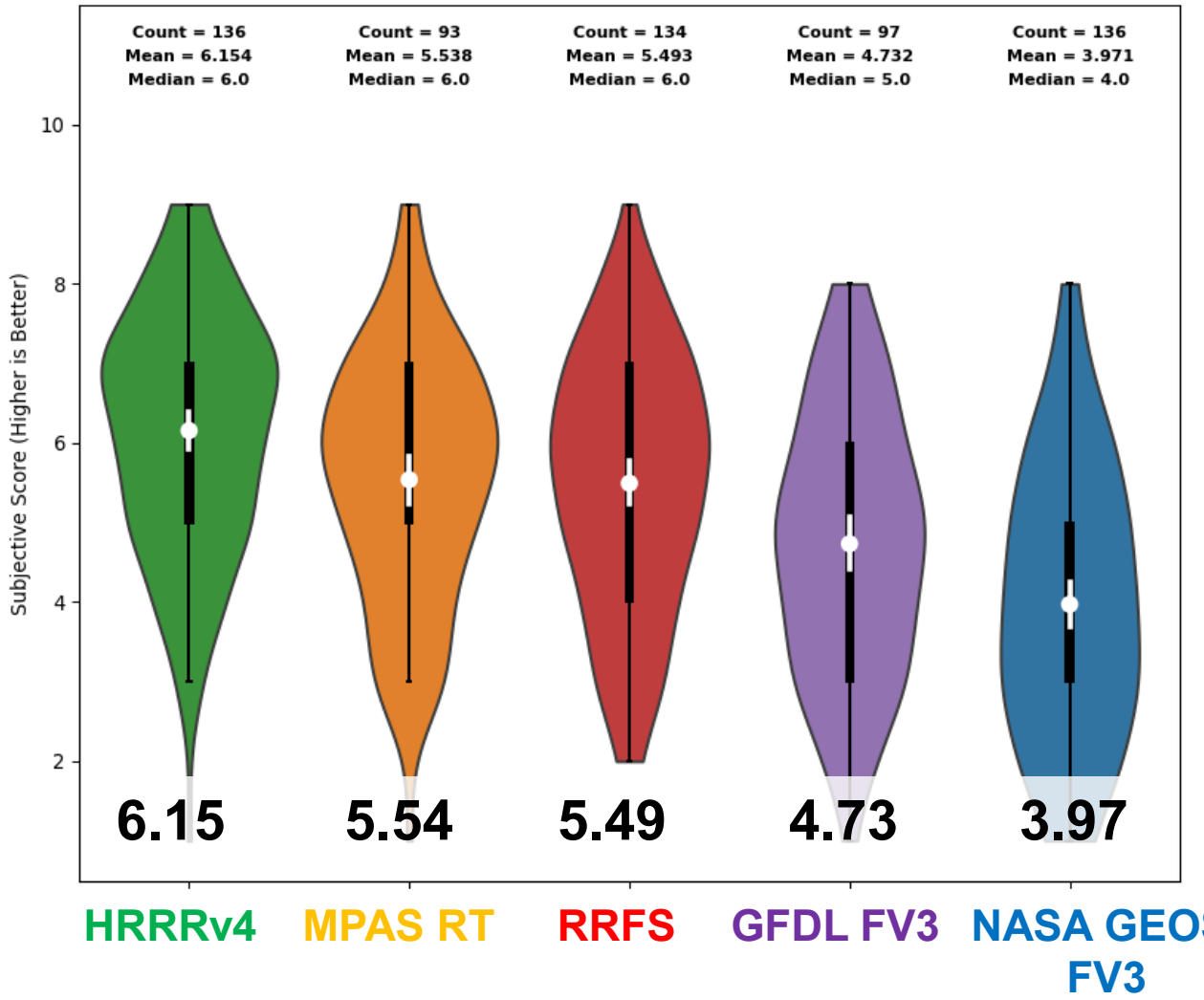


Day 2

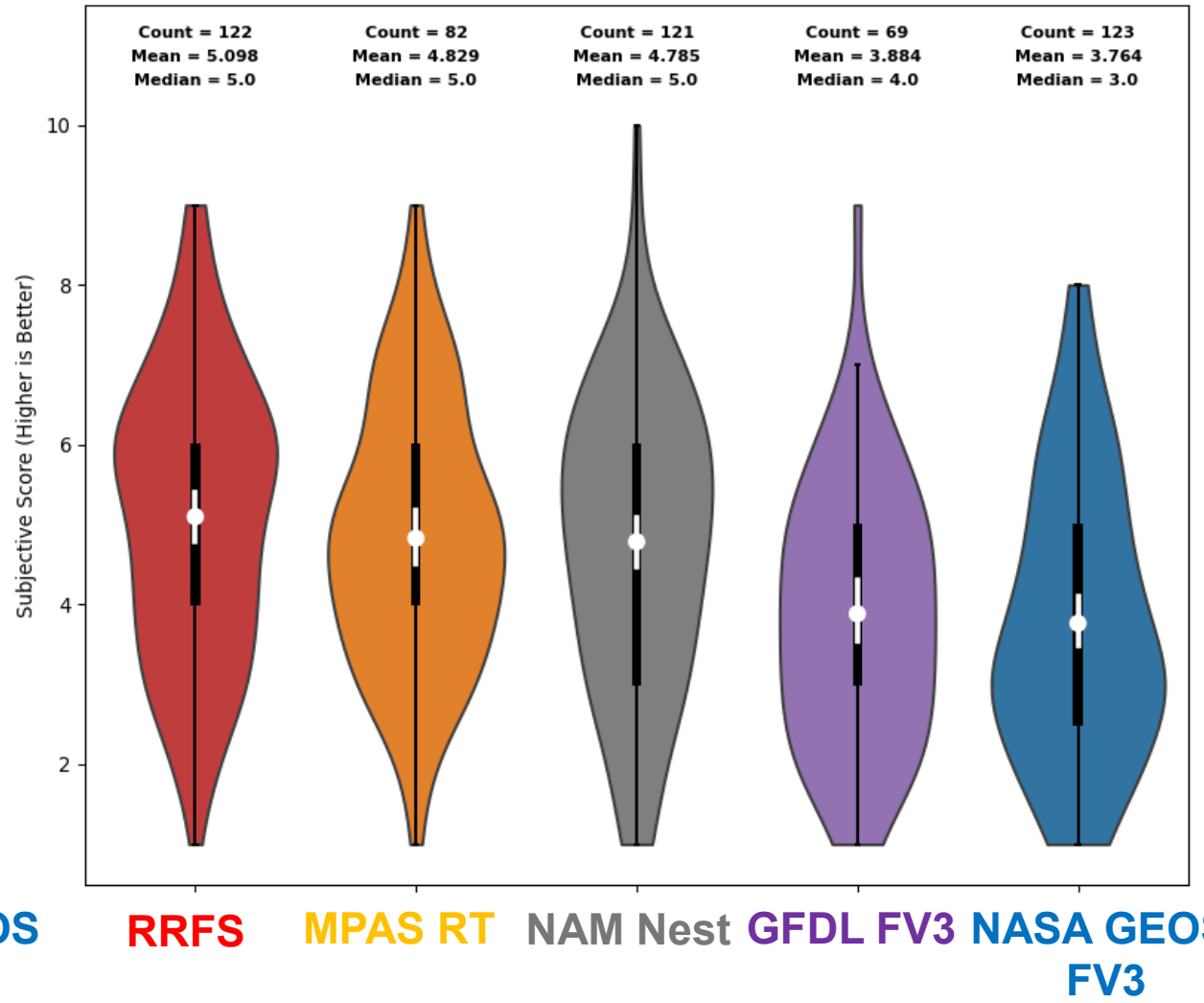


Days 1 & 2 Deterministic Flagships: QPF

Day 1

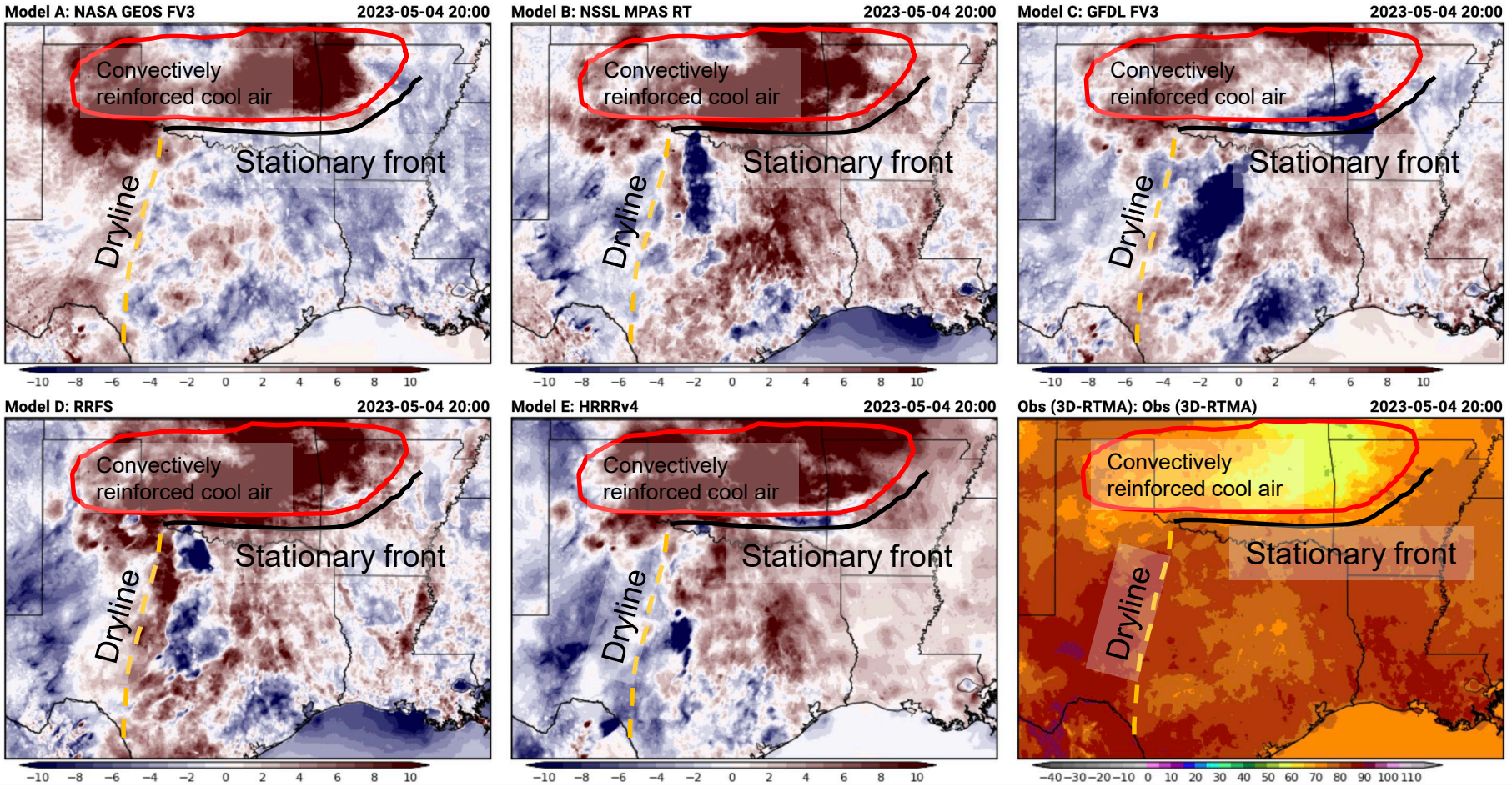


Day 2



Example Case: 4 May 2023

2000 UTC



Large temperature errors over Oklahoma. What is going on?

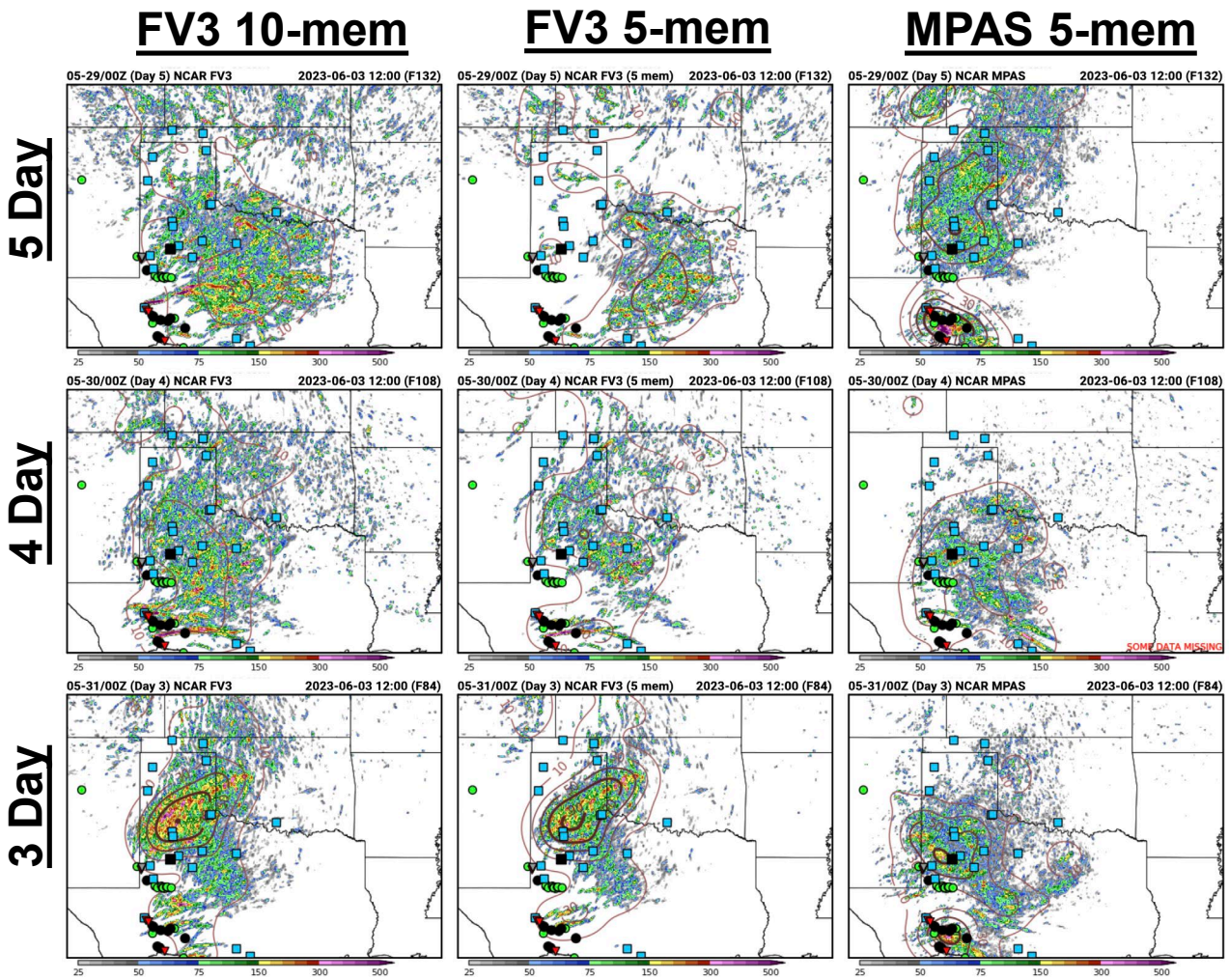
Cloud cover and cooling from earlier convection not well depicted.

MPAS, RRFS, & HRRR too warm in pre-CI environment over Texas.

Objective temperature metrics (e.g., RMSE & bias) won't tell whole story, especially when errors are "feature relative".

Extended Range FV3 and MPAS Comparisons

- **Research questions:** What is the maximum lead time at which CAM ensembles have value? What are the differences in forecast quality and characteristics between the FV3 and MPAS model cores for lead times of 3-5 days?



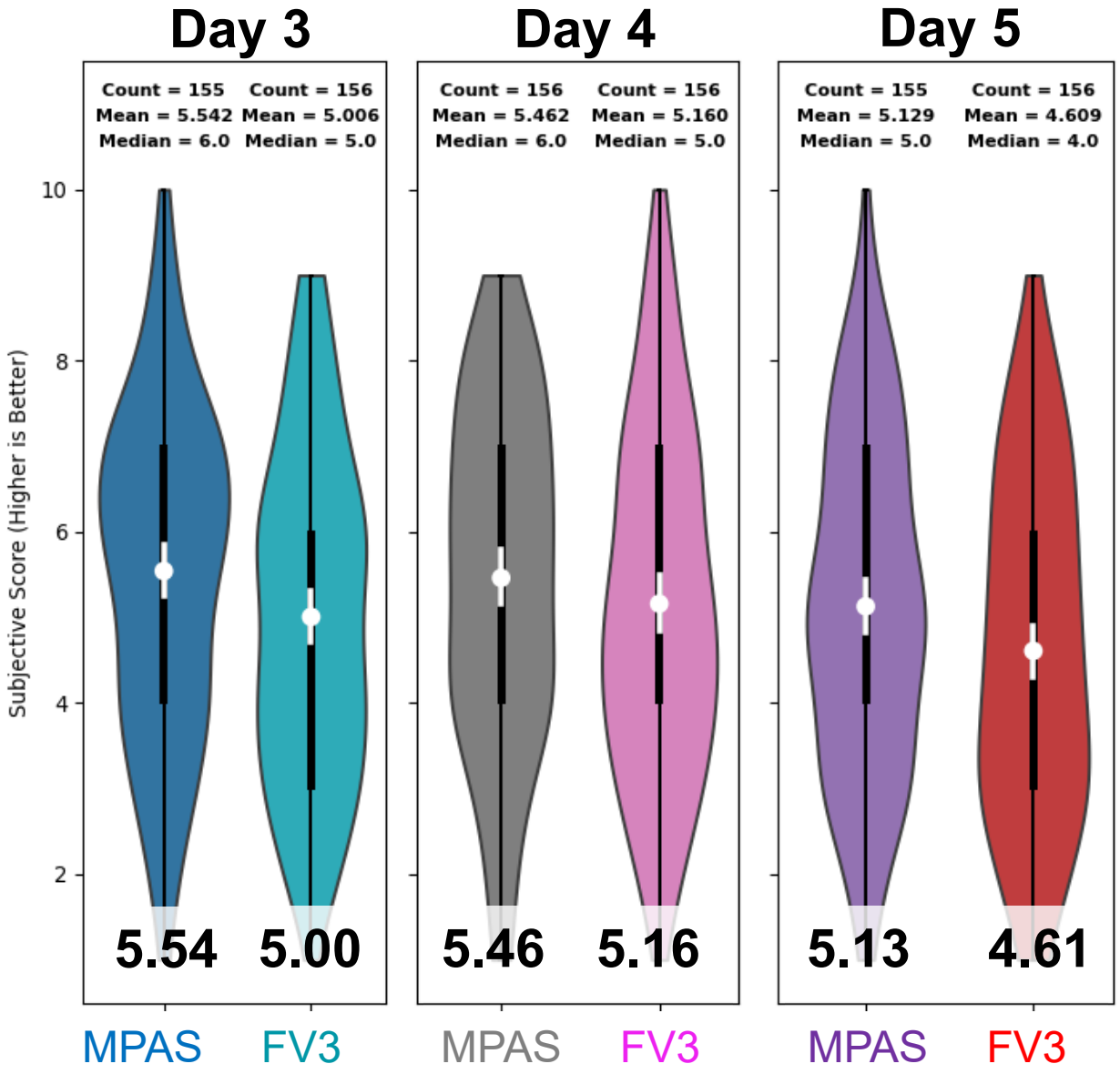
24-h aggregate storm-attribute products compared to storm reports to assess quality of guidance for severe weather forecasting.

Days 3-7 examined, but MPAS forecasts only went to 5 days.

FV3 had 10 members, while MPAS only had 5. To equitably compare, 5-member FV3 is compared to 5-member MPAS.

The 5 members in each ensemble are initialized from the same 5 GEFS members.

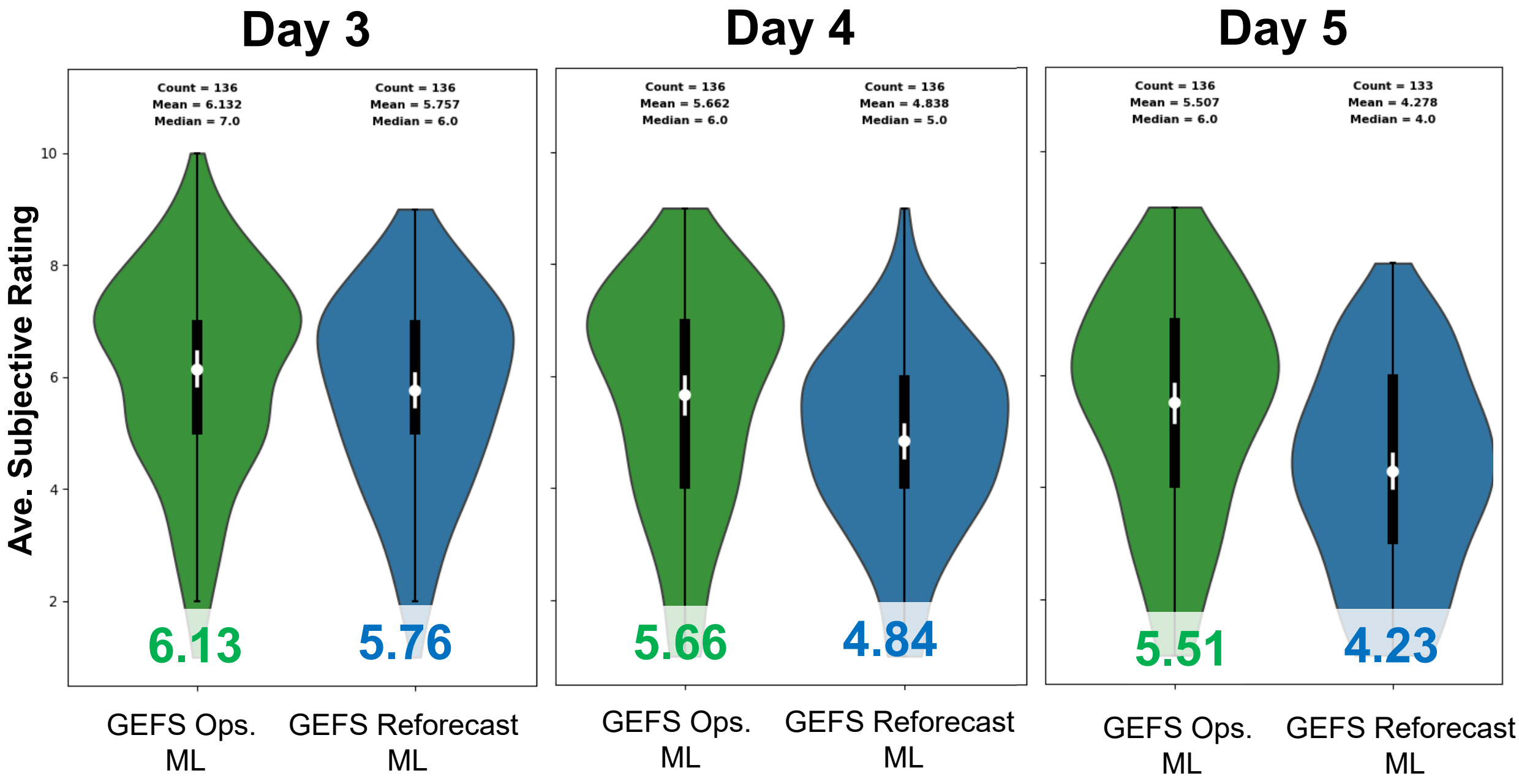
Extended Range FV3 and MPAS Comparisons



There was oftentimes clearly value in the extended range CAM ensembles past day 5.

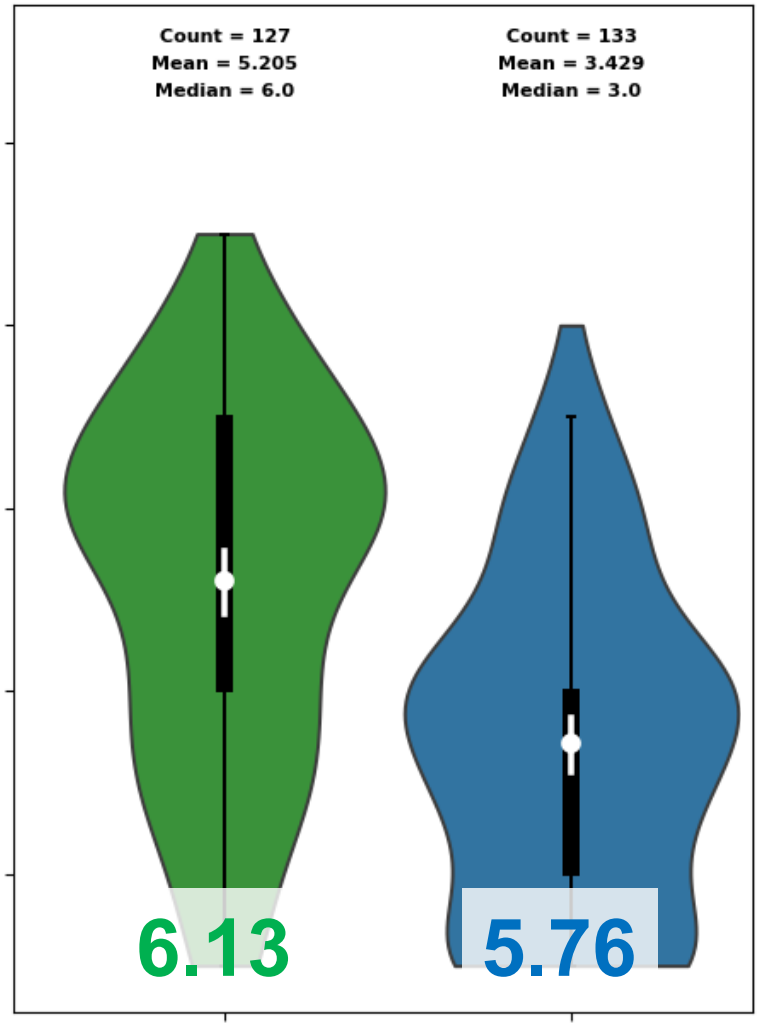
Average subjective ratings in MPAS were significantly higher than FV3.

For both FV3 and MPAS, there were run-to-run inconsistencies in performance (i.e., sometimes the Day 5 would be better than Day 3 or 4).

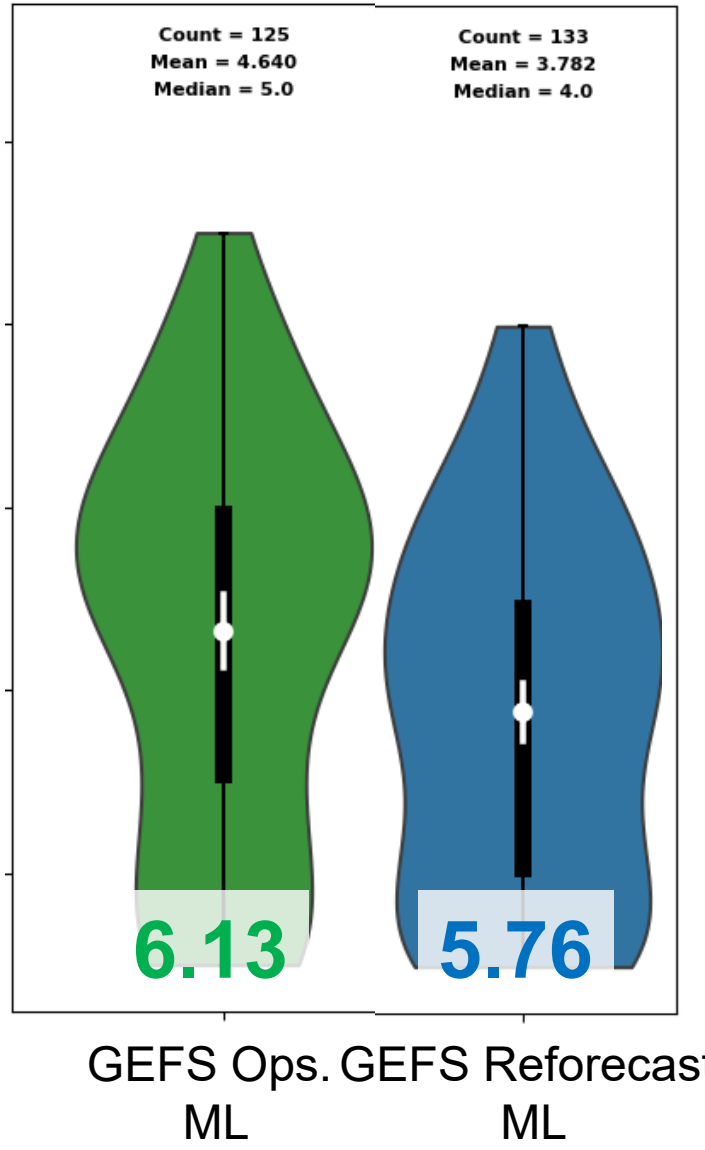

Day 3
Day 4
Day 5
Ave. Subjective Rating
10
8
6
4
2
Count = 136
Mean = 6.132
Median = 7.0
Count = 136
Mean = 5.757
Median = 6.0
Count = 136
Mean = 5.662
Median = 6.0
Count = 136
Mean = 4.838
Median = 5.0
Count = 136
Mean = 5.507
Median = 6.0
Count = 133
Mean = 4.278
Median = 4.0
6.13
5.76
5.66
4.84
5.51
4.23
GEFS Ops. ML
GEFS Reforecast ML
GEFS Ops. ML
GEFS Reforecast ML
GEFS Ops. ML
GEFS Reforecast ML

Ave. Subjective Rating

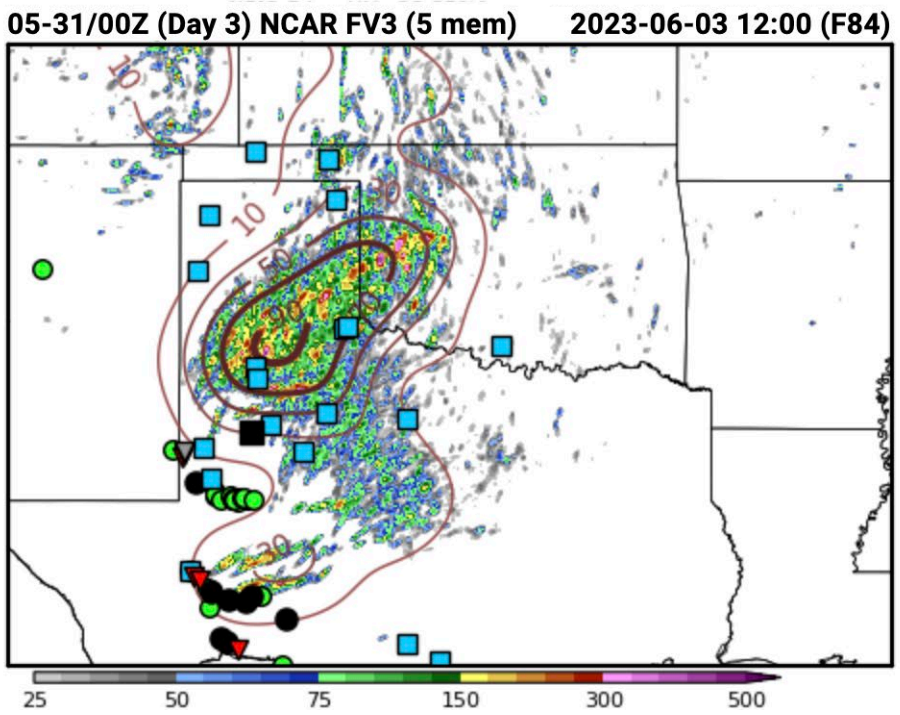
Day 3

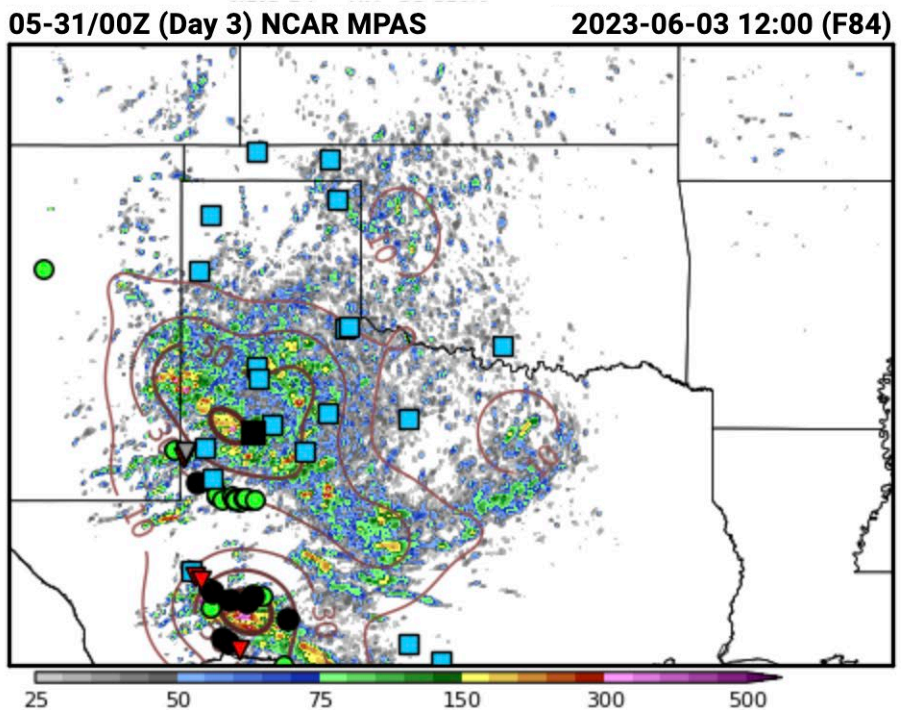


Day 7



Example Case: 2 June 2023



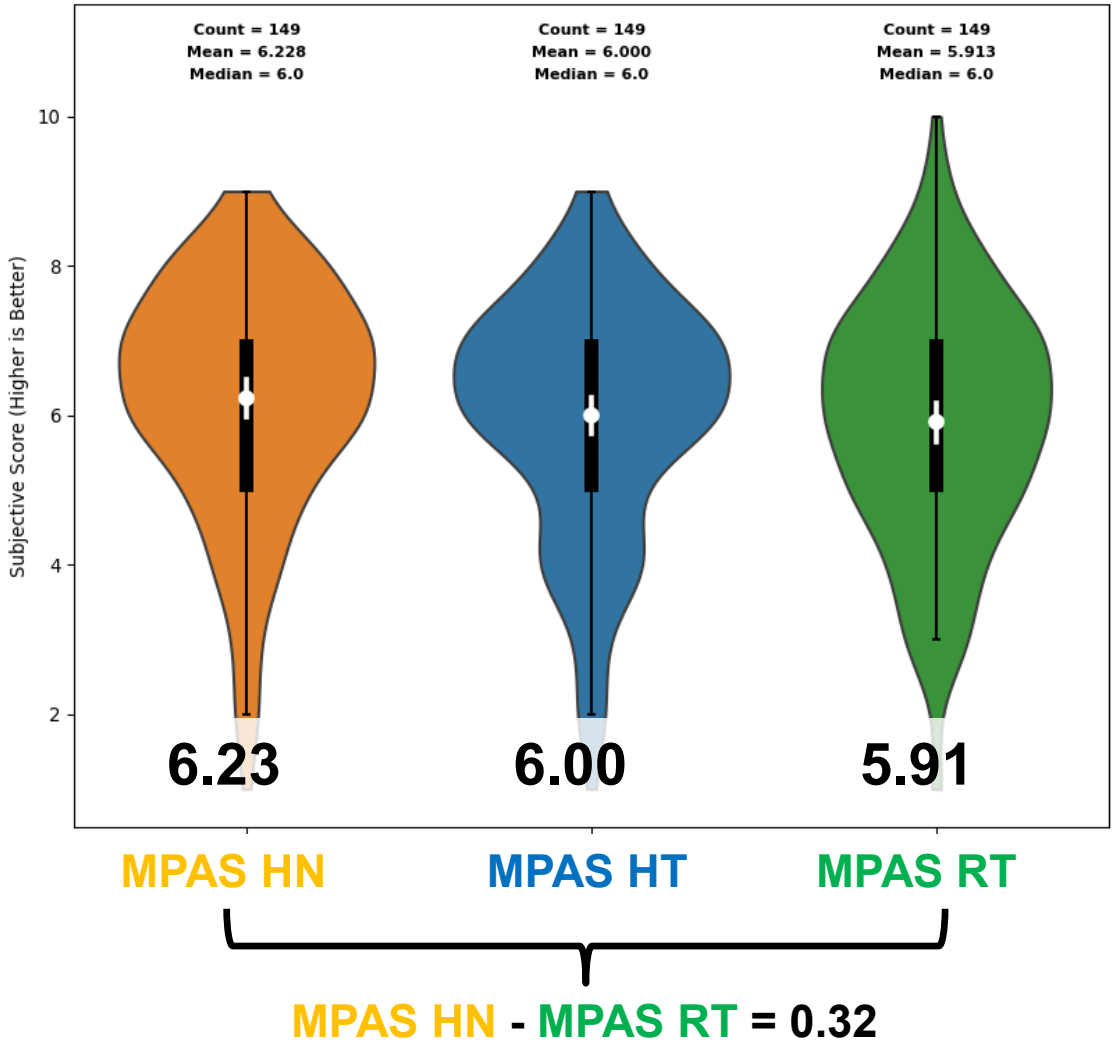


Day 4

MPAS is more consistent from Day 4 to Day 3 in the cluster of supercells to the south. FV3 struggles with the evolution of TX, and Days 5 almost everything else.

NSSL MPAS Configurations

- NSSL microphysics outperforms Thompson
- HRRR initializations outperform RRFS



- MPAS HN may have similar performance to HRRRv4

Add this difference to MPAS RT from the flagships comparison & MPAS HN scores about the same as HRRR (6.14+0.32 = 6.46).