

# SPC Ensemble Review: Performance and Needs

---

**ISRAEL JIRAK**

SCIENCE AND OPERATIONS OFFICER

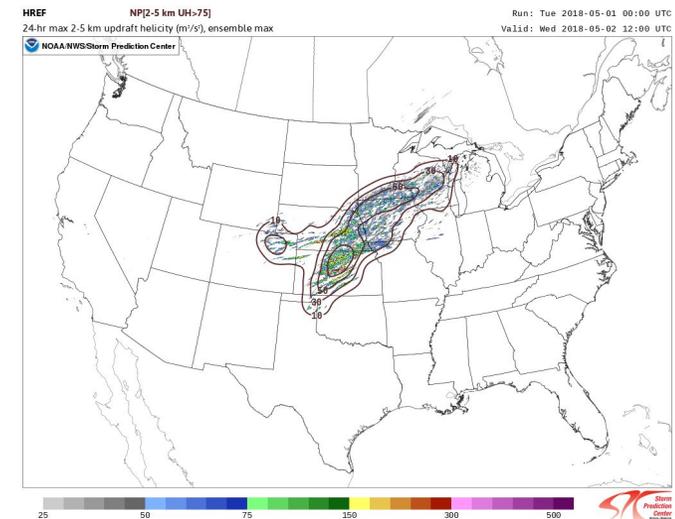
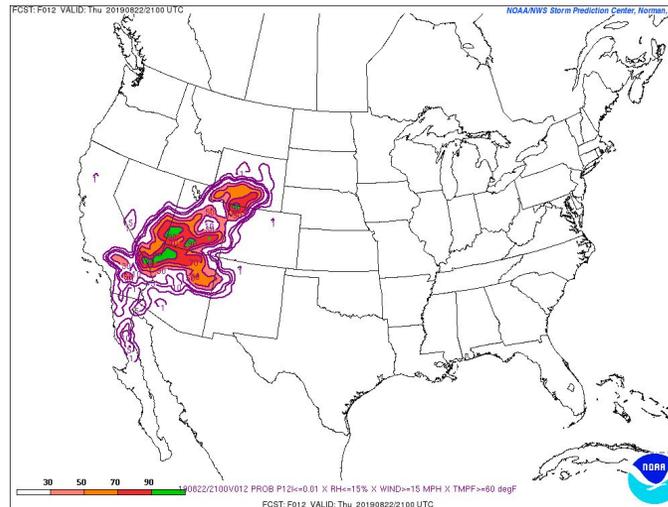
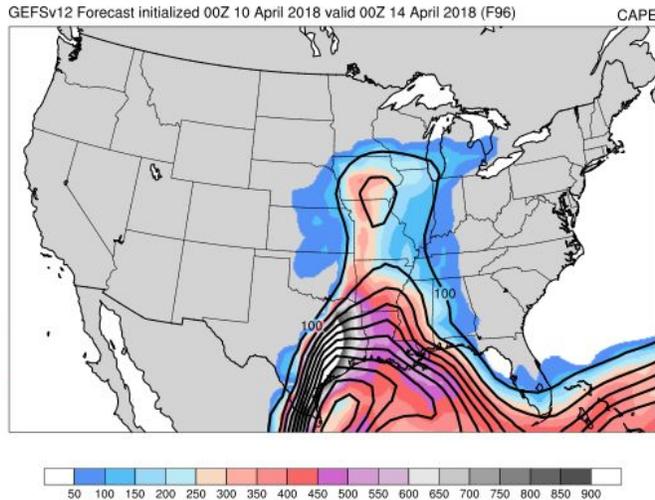
NOAA/STORM PREDICTION CENTER



# SPC Ensemble Review: Outline

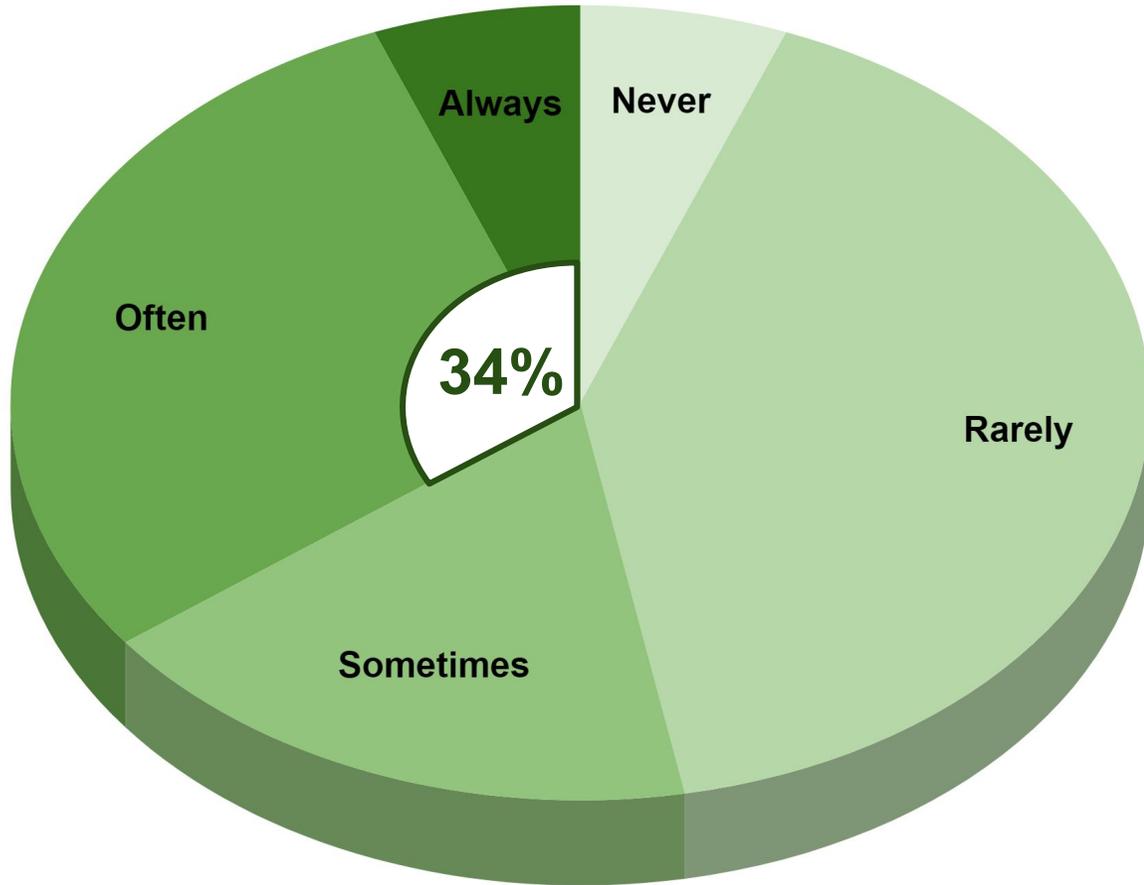
- Medium-Range Guidance (Days 3-8): GEFS
- Days 2 & 3 Guidance: SREF, GEFS
- Convection-Allowing Model (CAM) Guidance: HREF
- Ensemble Needs

SPC  
Forecaster  
Perspective  
on Ensemble  
Use & Quality  
- **2021 Survey**

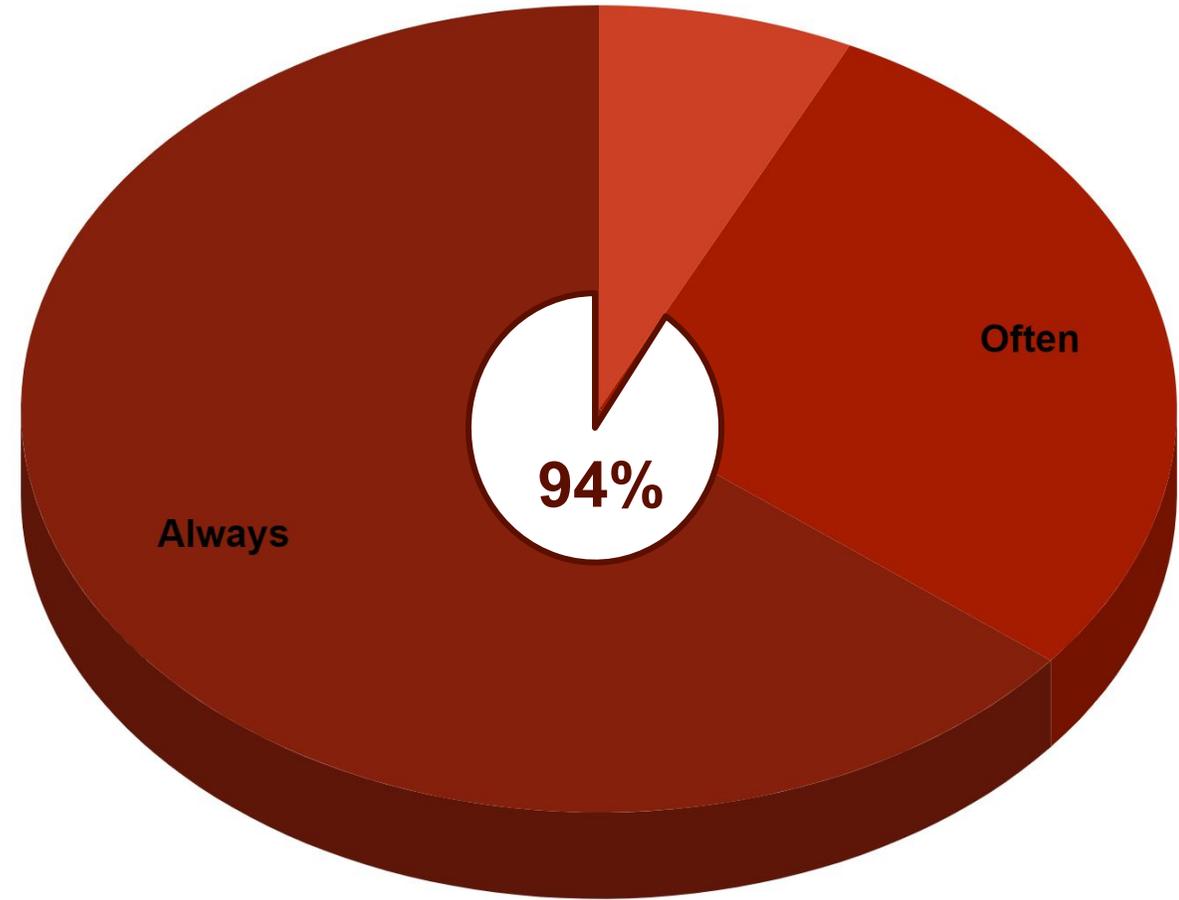


# SPC Ensemble Review: Medium Range

*How often do you use these **deterministic models** in the forecast process?*



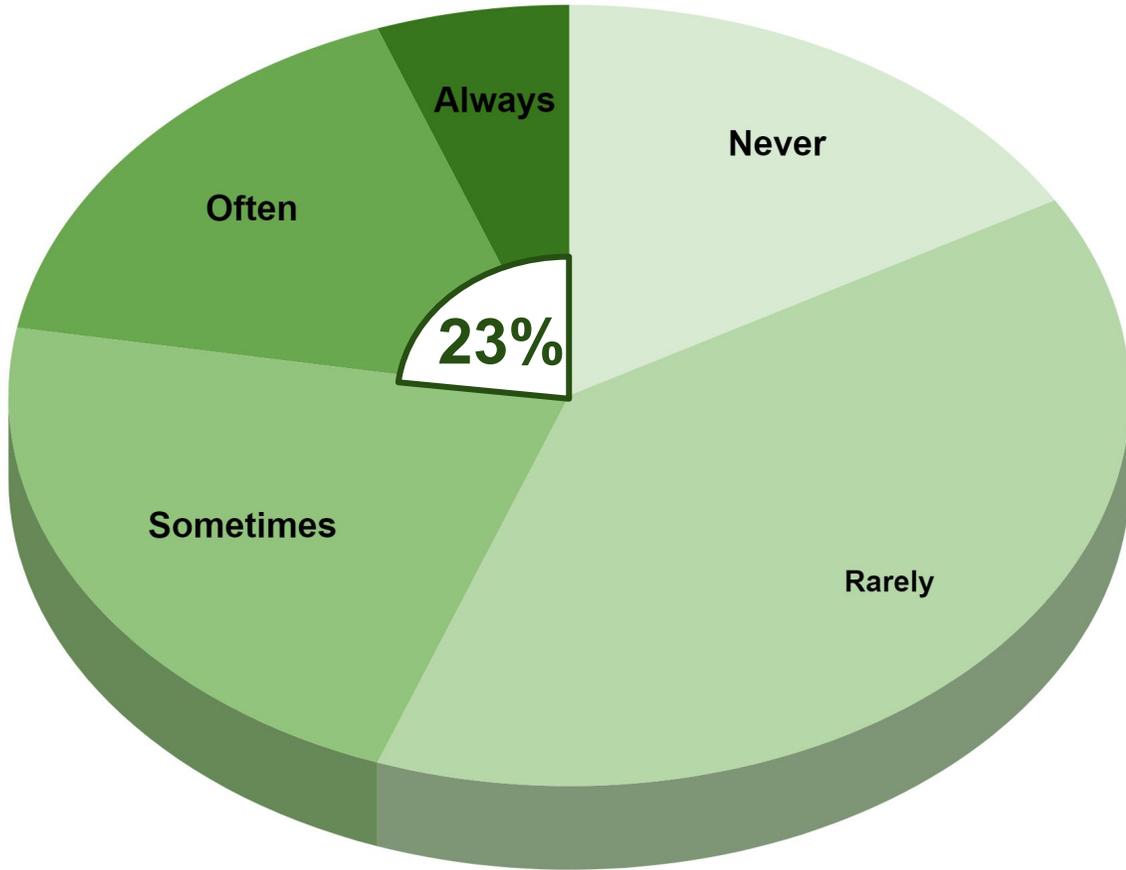
GFS



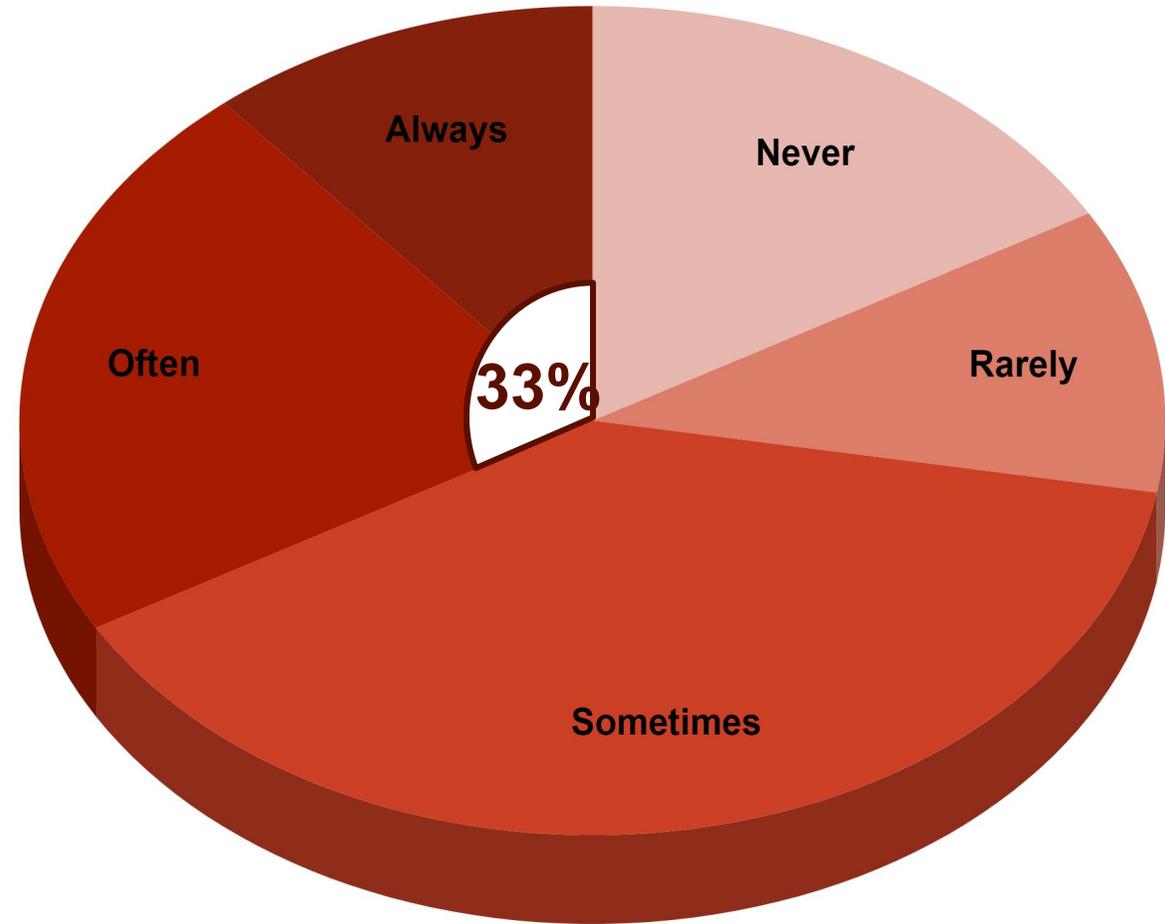
ECMWF

# SPC Ensemble Review: Medium Range

*How often do you use these ensembles in the forecast process?*



GEFS

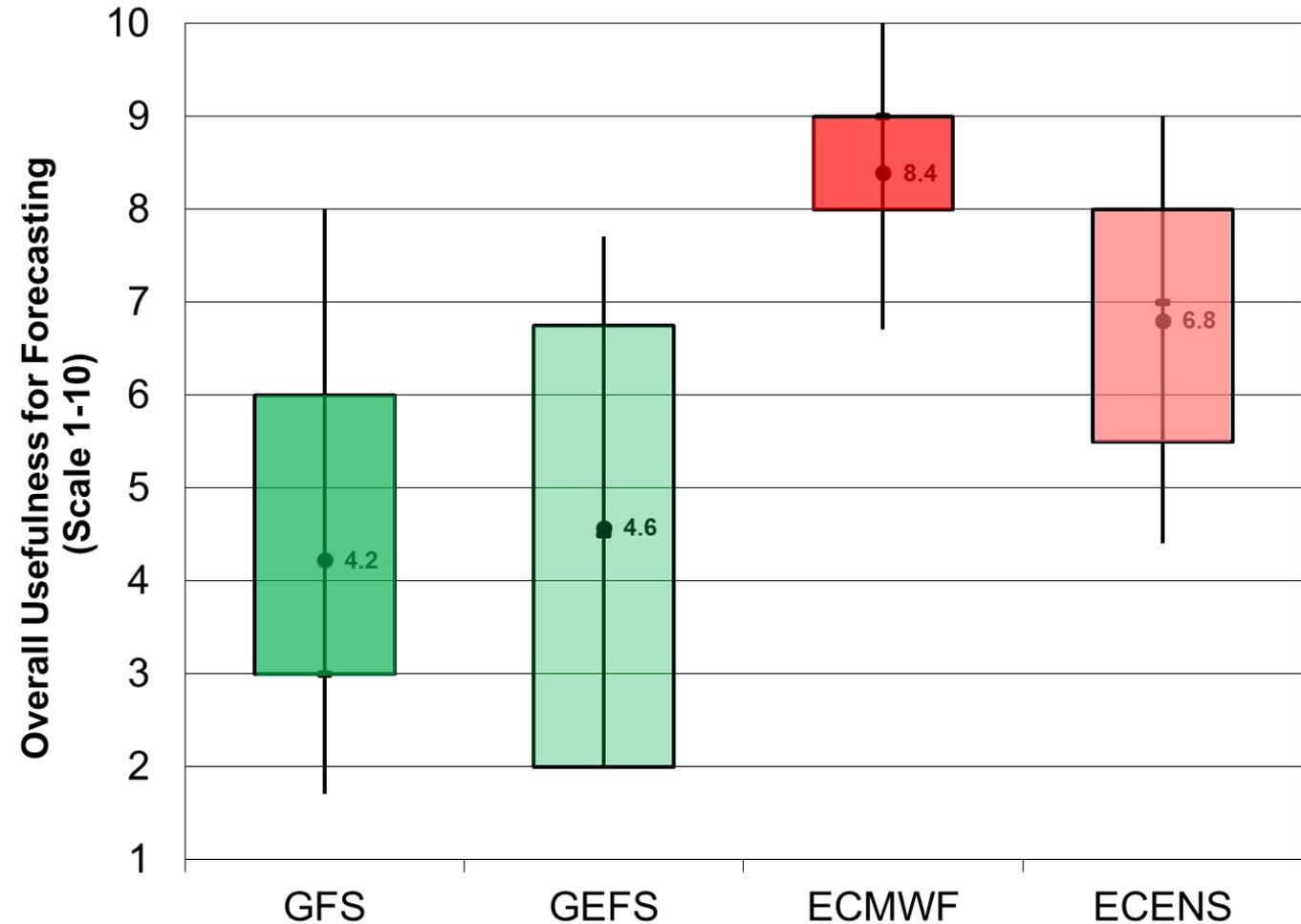


ECENS

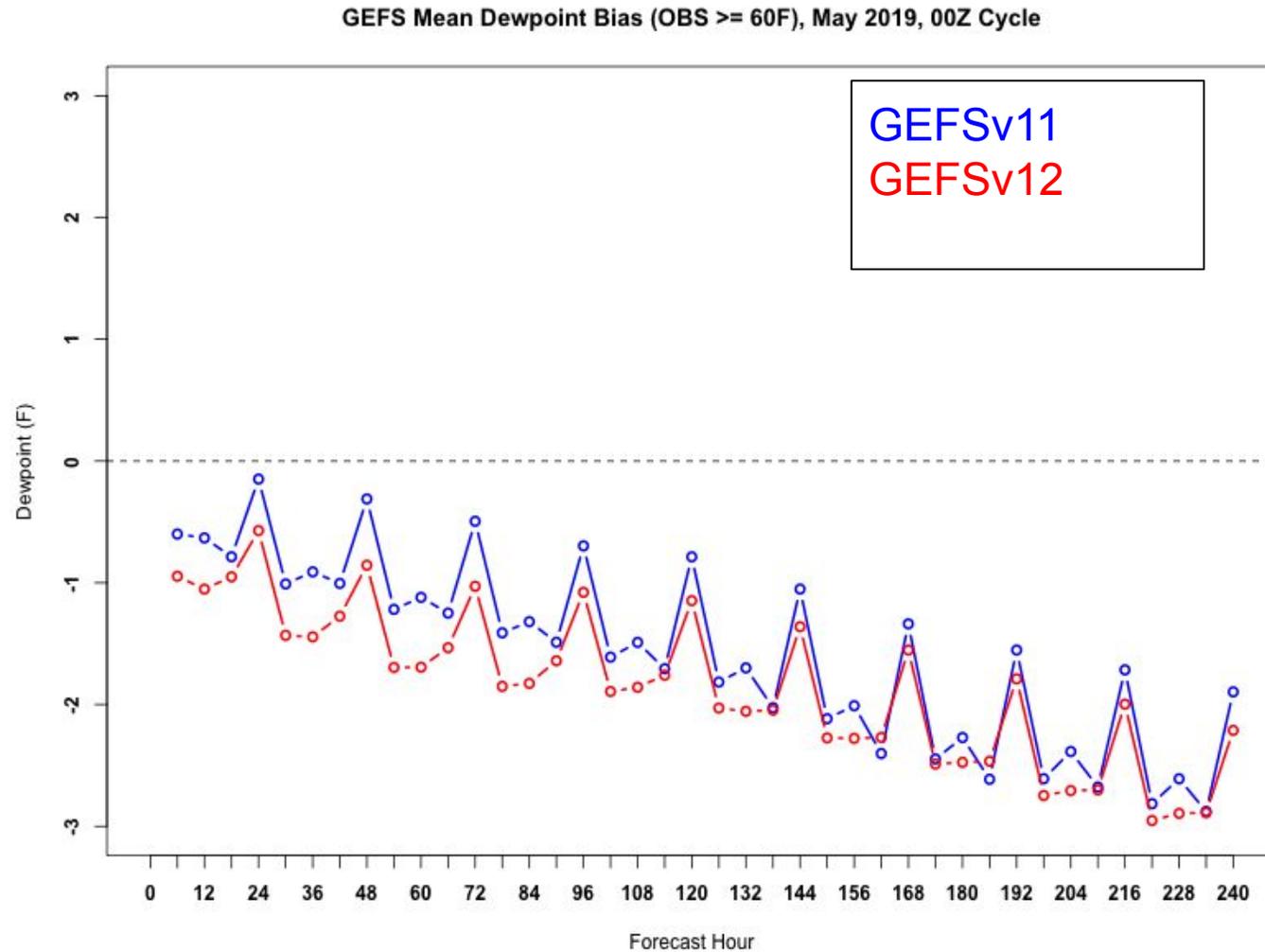
# SPC Ensemble Review: Medium Range

*How useful are these **models & ensembles** for supporting your overall forecasting needs?*

- *Challenge:* While acknowledging the benefits of an ensemble approach to forecasting, SPC forecasters perceive the usefulness and quality of current global ensembles to be less than that of their deterministic counterparts.
- One can speculate on reasons (resolution, visualization, etc.), but perhaps primary function and utility of ensembles should be for creating derived/calibrated products.

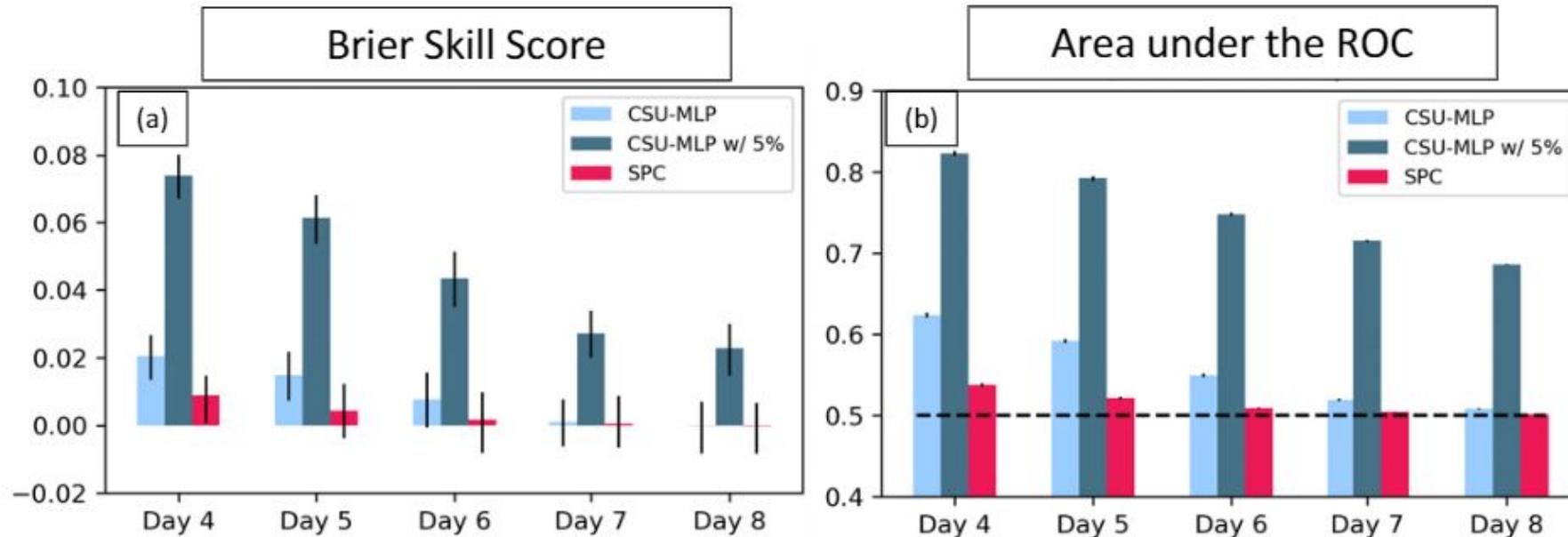


- In the GEFSv12 evaluation by SPC, it was noted that the **low 2-m dewpoint and instability bias** was **slightly worse** during the spring over the CONUS, especially in environments favorable for supporting severe weather
- The systematic biases inherited from the GFS overwhelm the GEFS signal for some high-impact events: progressive nature of shortwave troughs and overmixing in the PBL
- Despite LoAs between EMC & SPC in 2015 & 2020 to address the GFS PBL issues, progress has not been reported



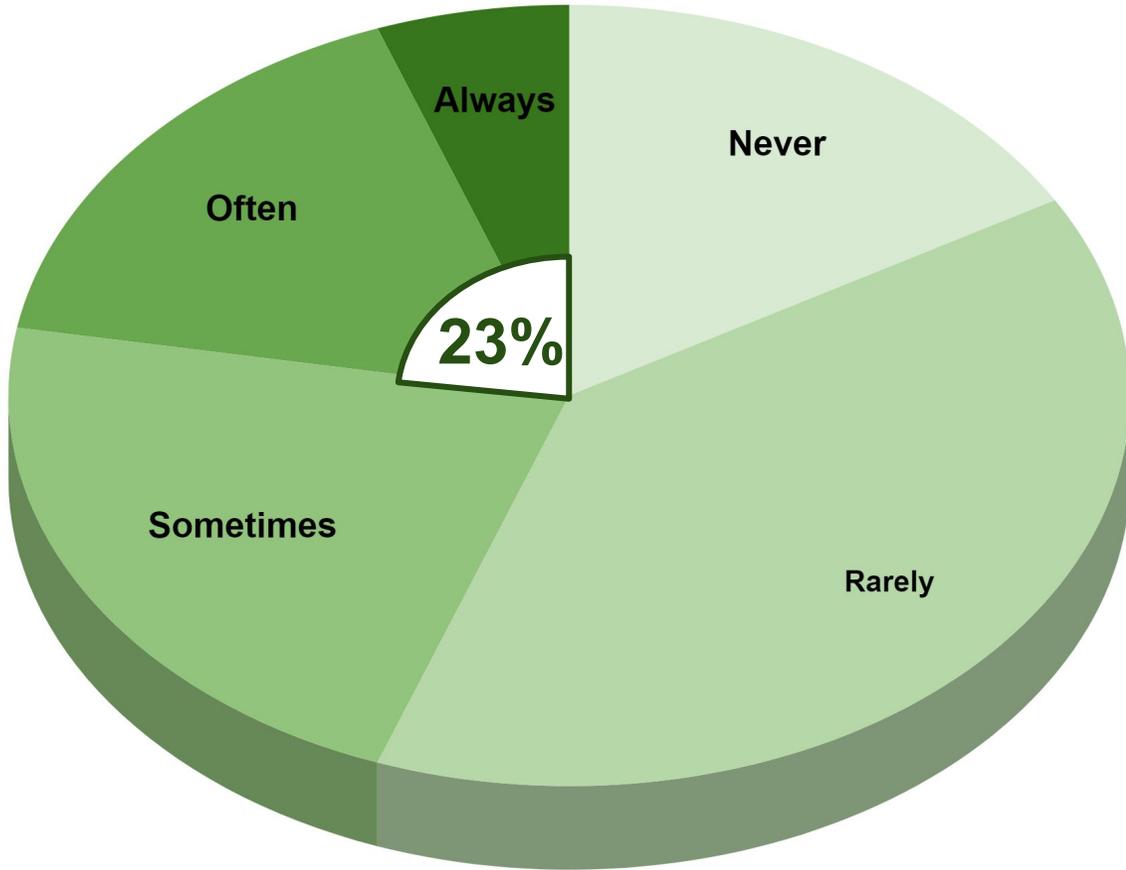
# SPC Ensemble Review: Medium Range

- Despite these issues, the GEFs **reforecast dataset** and **machine learning** can be leveraged to create skillful and reliable probabilistic medium-range severe weather guidance
- The CSU-MLP guidance has been thoroughly evaluated by SPC forecasters with favorable results and is commonly used in operations as a confidence builder, especially for Days 3-5
- This guidance is being considered for operational implementation on WCOS2 in SPC Post

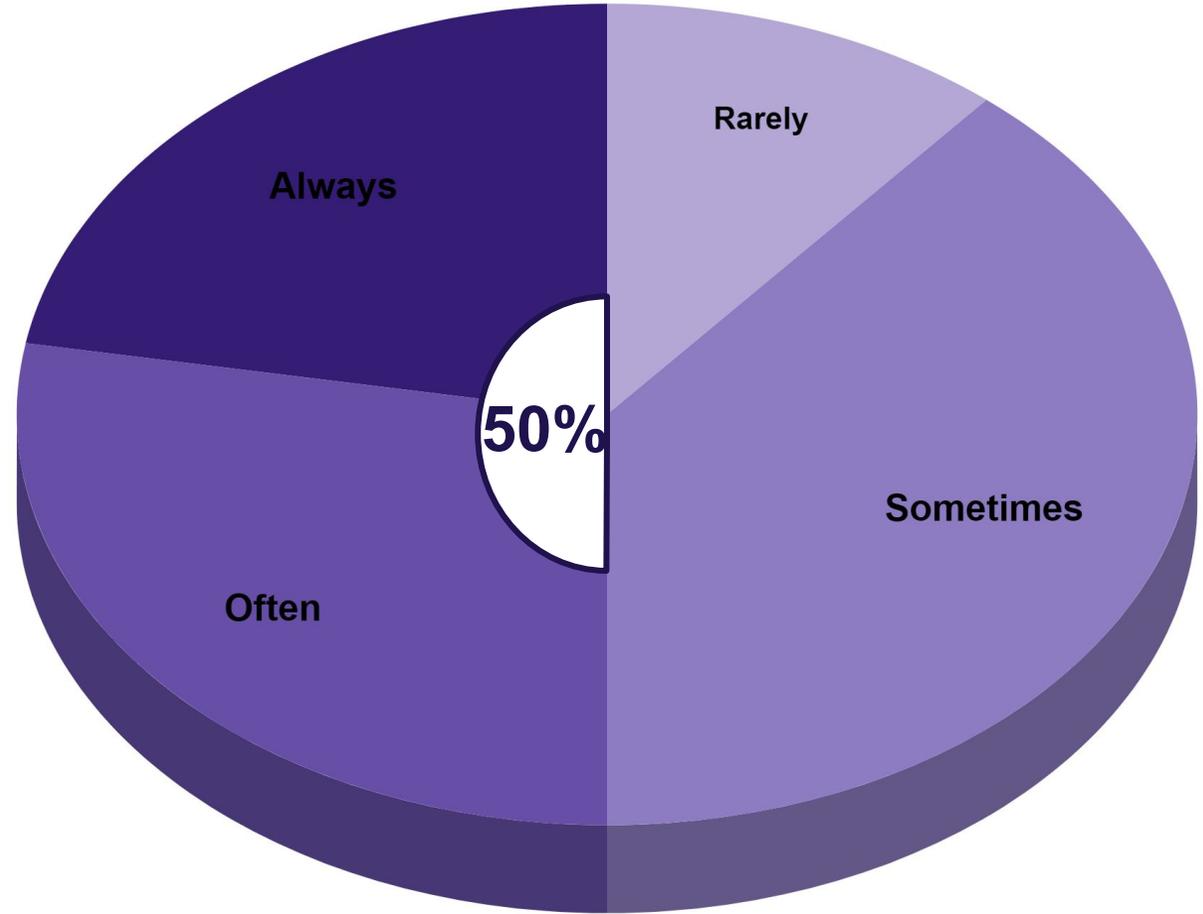


# SPC Ensemble Review: Days 2-3

*How often do you use these ensembles in the forecast process?*



GEFS

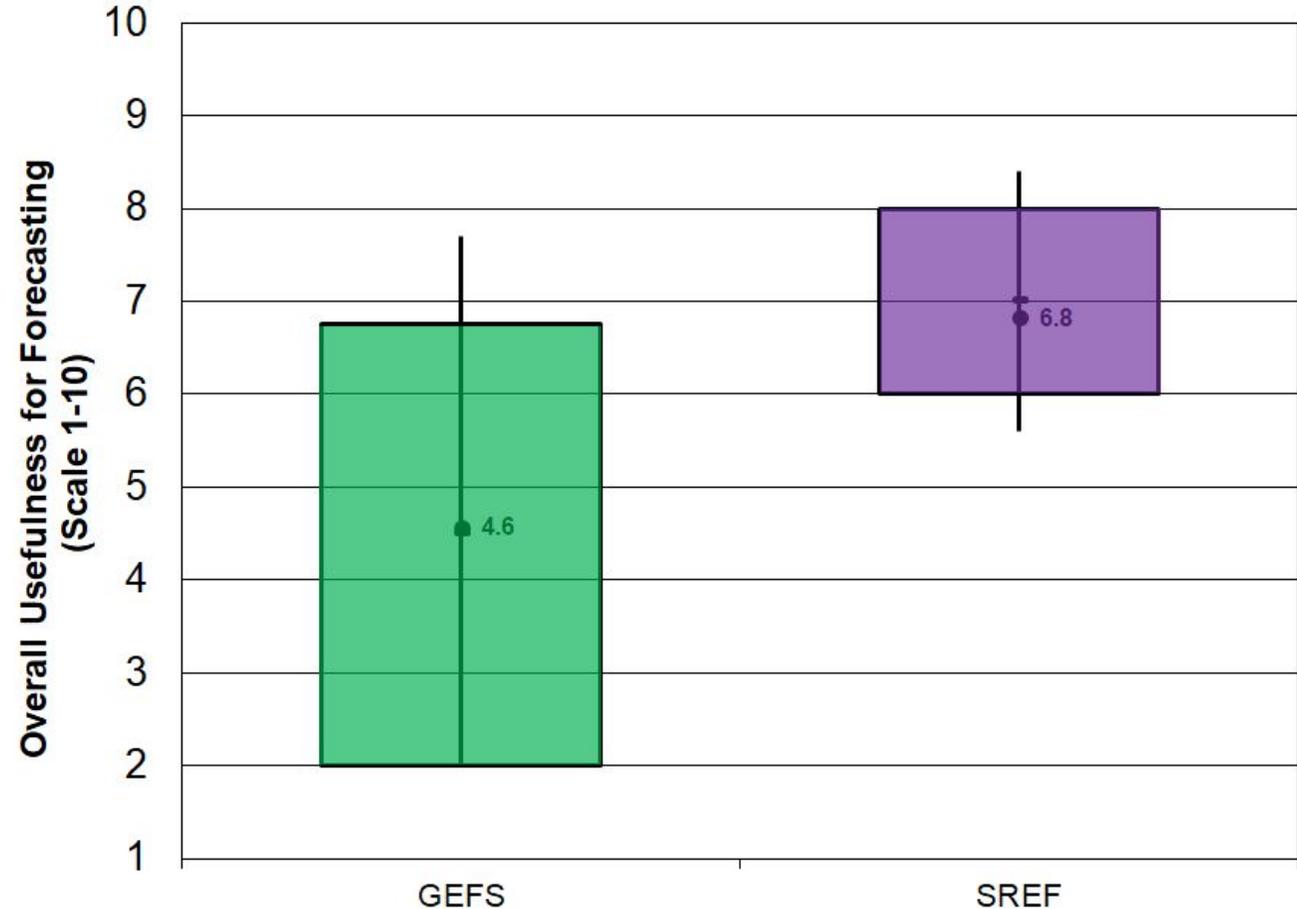


SREF

# SPC Ensemble Review: Days 2-3

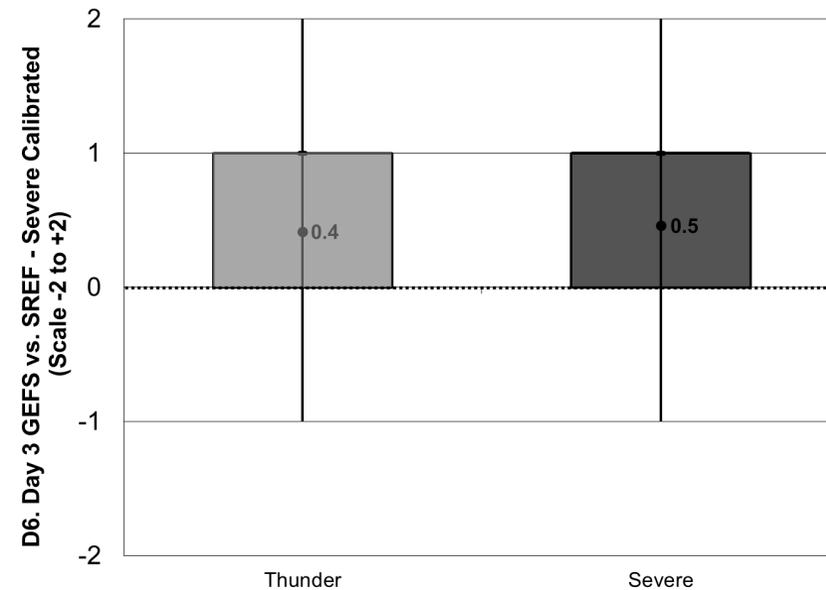
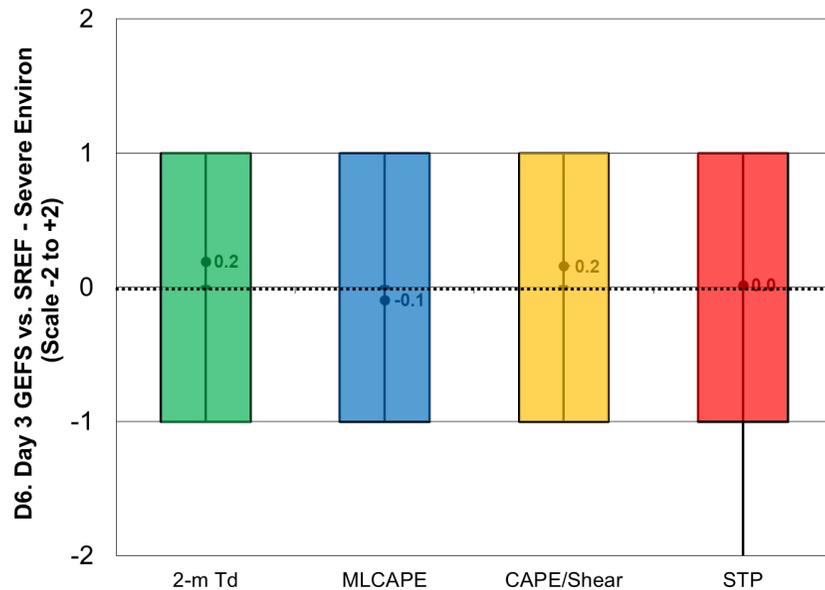
*How useful are these **ensembles** for supporting your overall forecasting needs?*

- Despite the SREF being frozen since 2015, the SPC forecasters still perceive the SREF being more useful than the GEFS for their forecasting applications
- There are probably two reasons for this perception:
  - The raw ensemble GEFS fields have significant biases in the PBL
  - More calibrated products have been developed for the SREF at SPC than for the GEFS



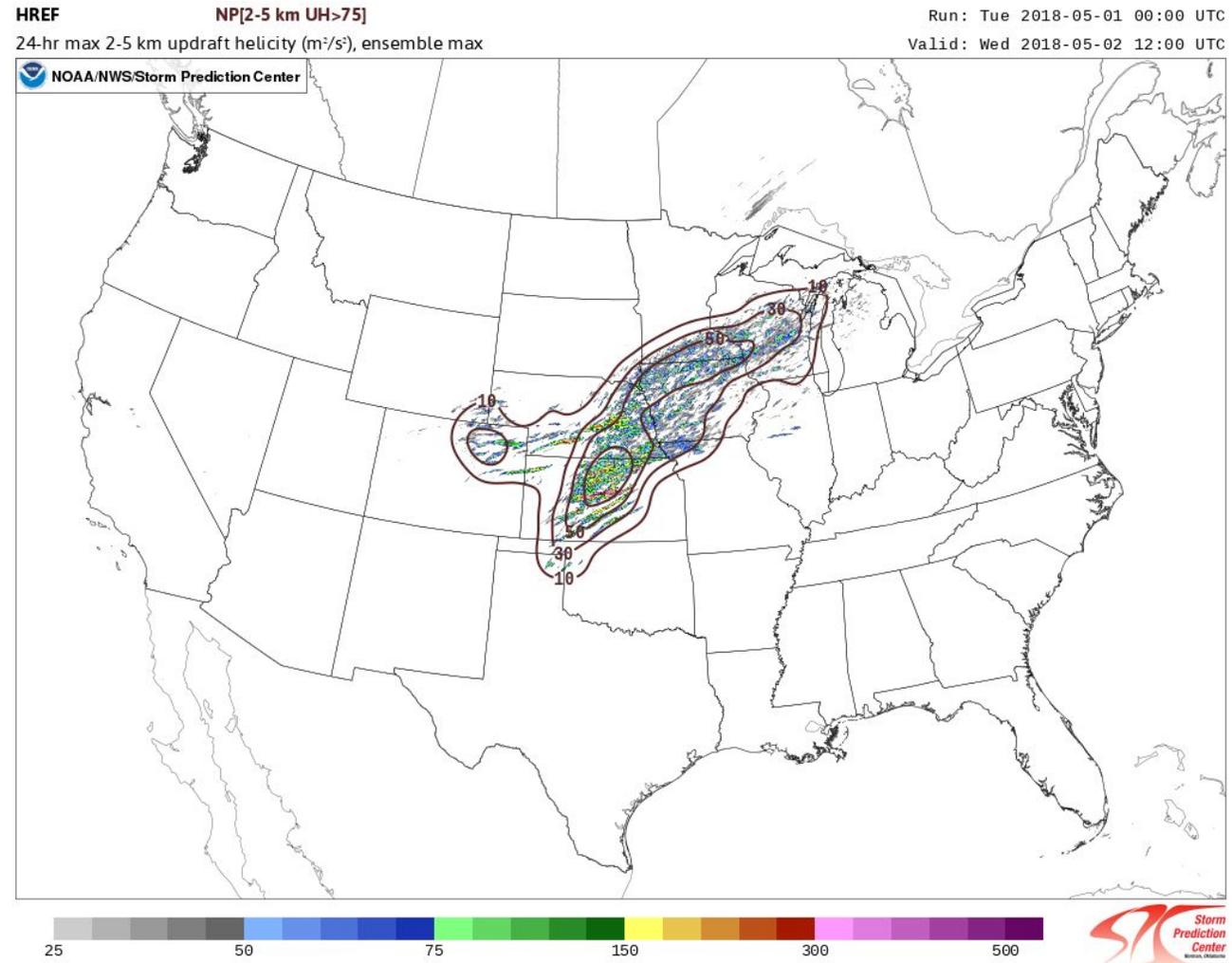
# SPC Ensemble Review: Days 2-3

- The GEFS was compared to the SREF for severe weather forecasting applications for Days 2 & 3 in the HWT Spring Forecasting Experiments in 2021 & 2022
- Despite some comments about the low instability bias and overconfident solutions, the GEFS was rated similarly to the SREF overall for raw ensemble fields
- Using similar algorithms for **calibrated thunder & severe probabilities**, the GEFS products (leveraging the 20-y reforecast dataset) were generally **favored** over the SREF products



\*\*\*Flashback from the last Ensemble Workshop\*\*\*

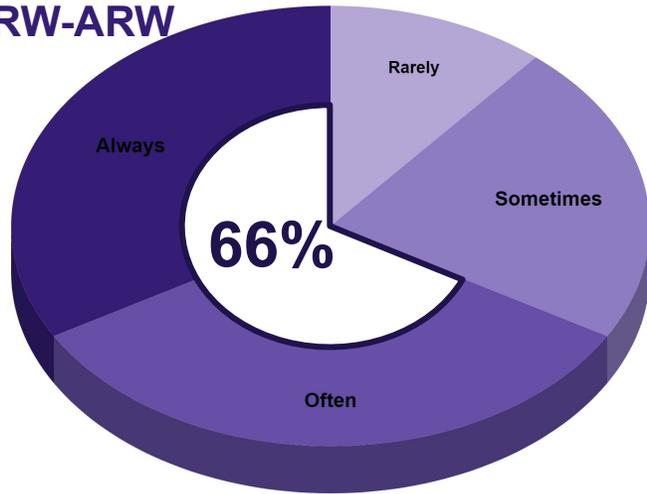
- The HREFv2 was implemented on 1 November 2017; an operational instantiation of the SSEO (a recommendation from the last Ensemble Workshop!)
- **Monumental implementation;** convection-allowing aspect provides probabilistic information on thunderstorm timing, mode, and intensity
- Heavily and widely used in NWS operations for forecasting, outlooks, & IDSS



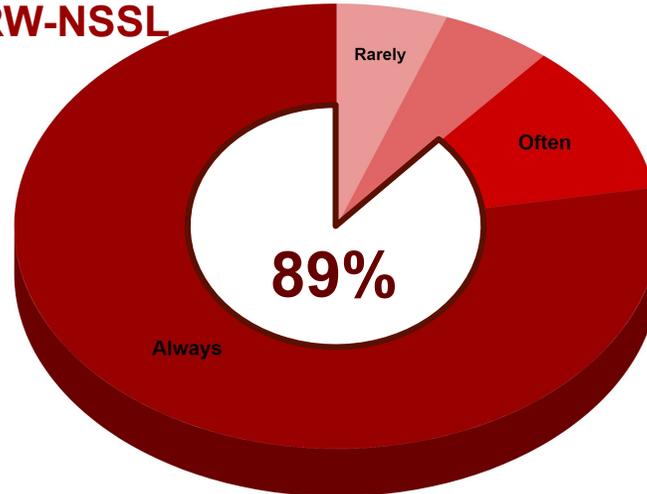
# SPC Ensemble Review: CAMs - Days 1-2

*How often do you use these models & ensembles in the forecast process?*

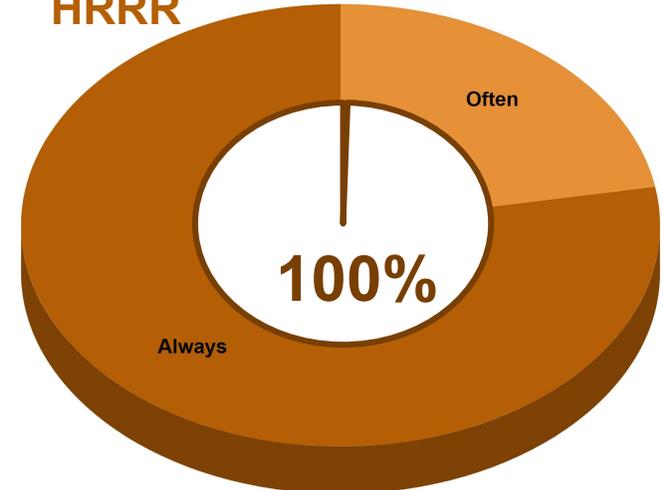
HRW-ARW



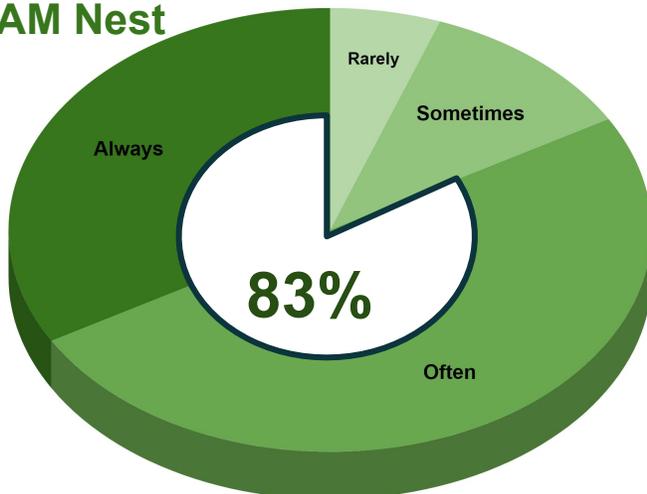
HRW-NSSL



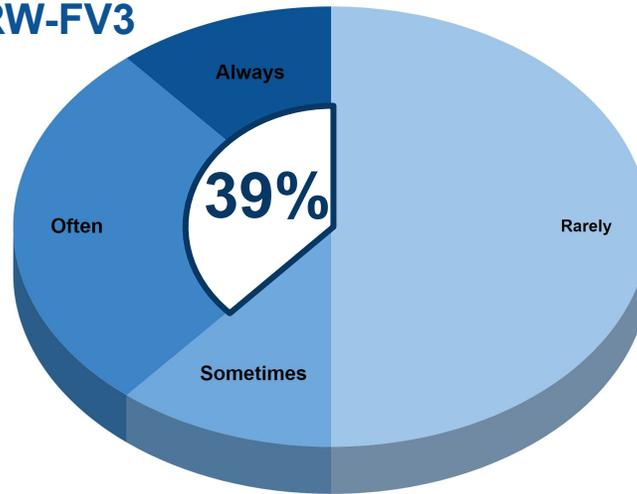
HRRR



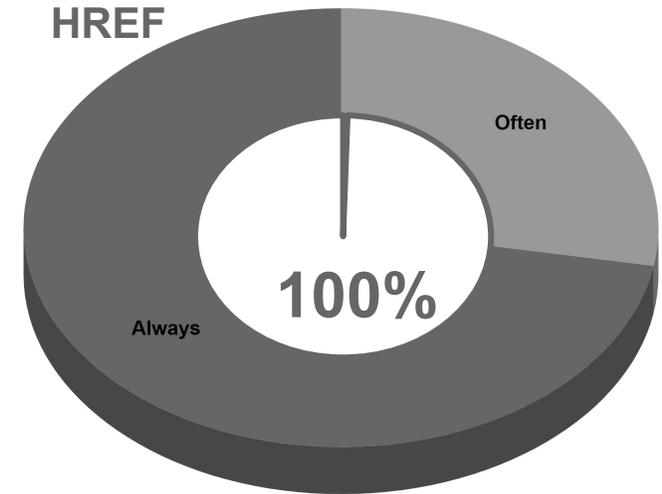
NAM Nest



HRW-FV3



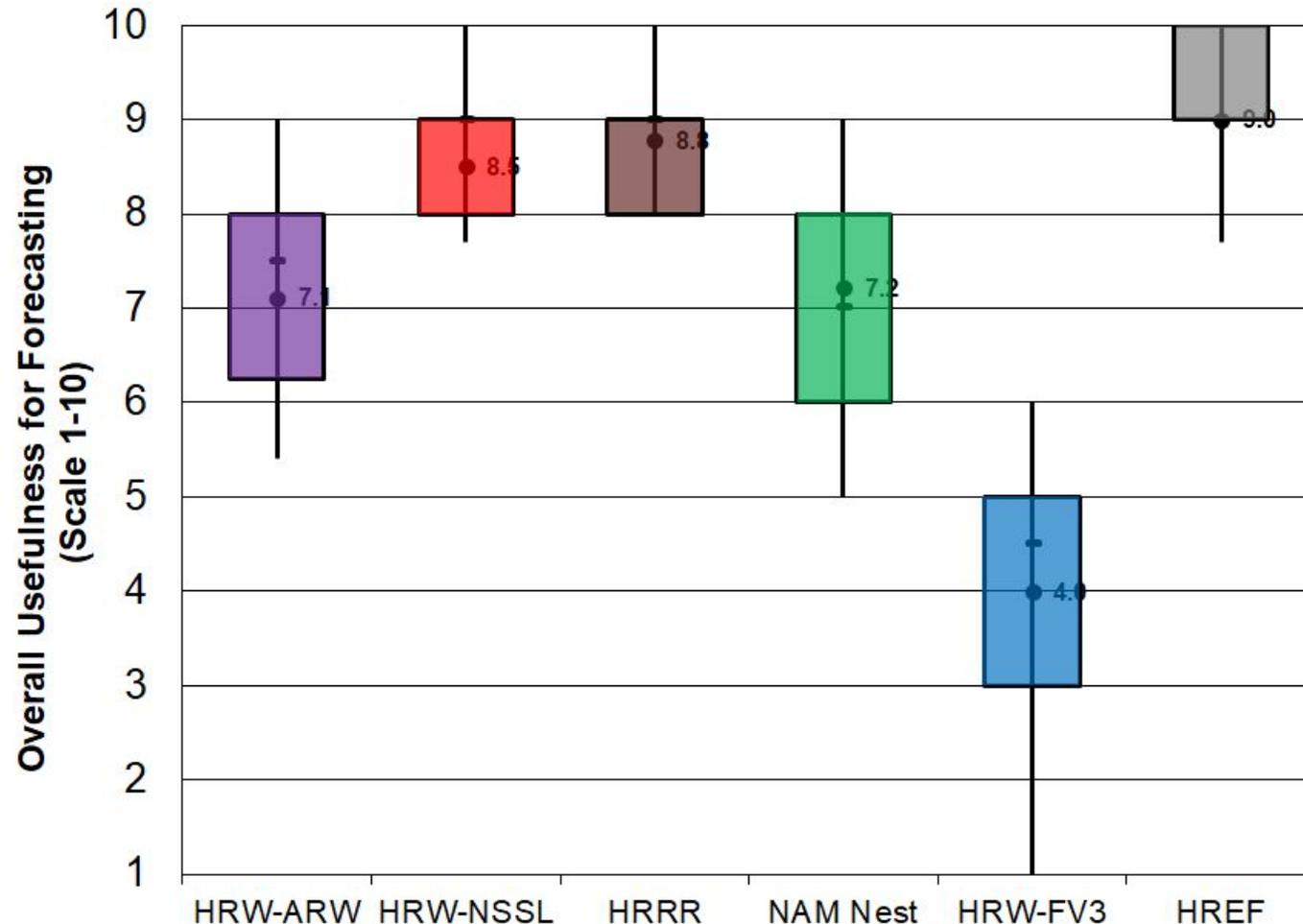
HREF



# SPC Ensemble Review: CAMs - Days 1-2

*How useful are these **models & ensembles** for supporting your overall forecasting needs?*

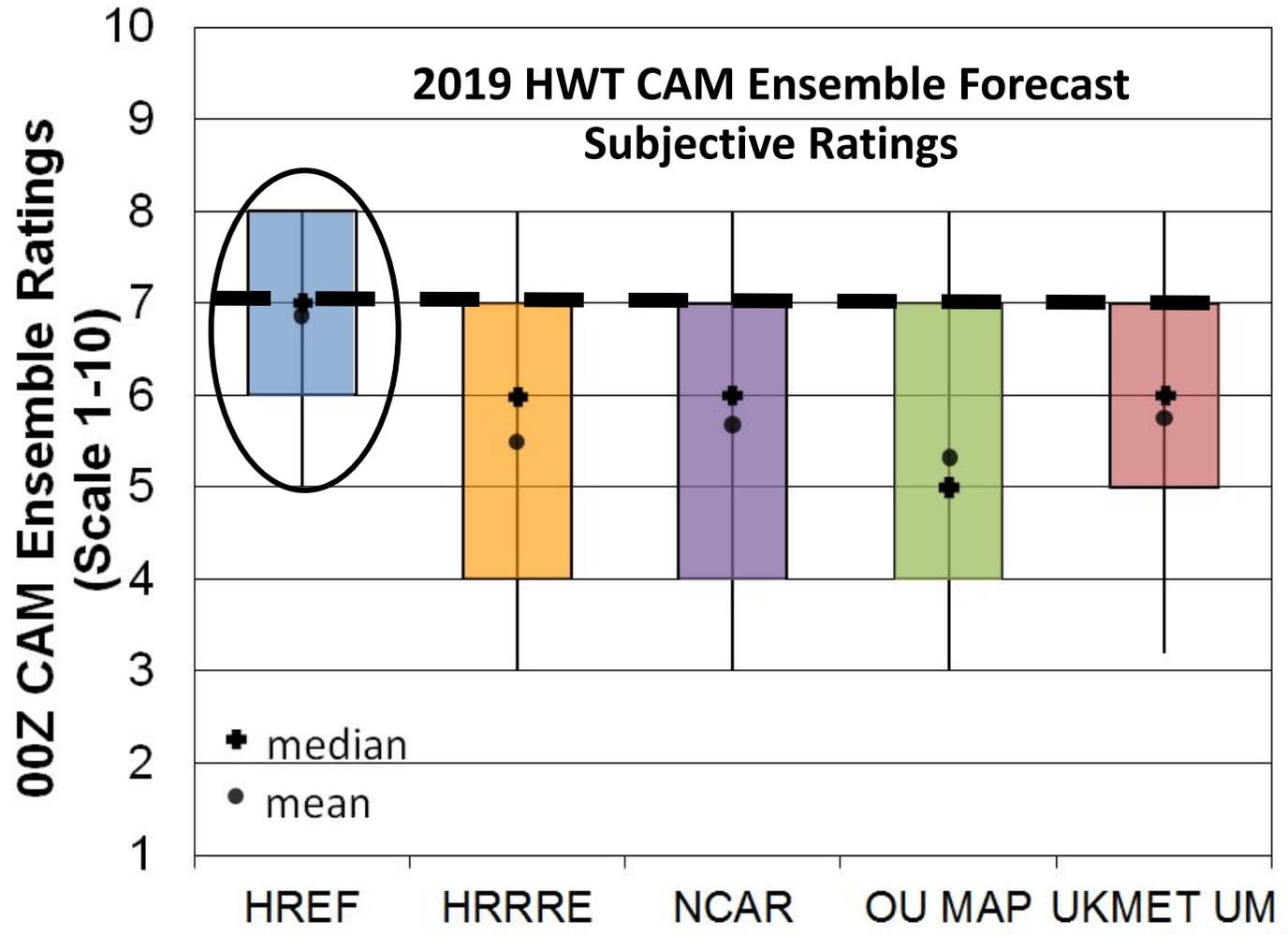
- Unlike the coarser ensembles, the CAM ensemble (HREF) was rated as being more useful than any individual HREF member
- HRW-NSSL and HRRR are top-tier CAMs used most frequently
- HRW-ARW and NAM Nest are respectable next-tier CAMs
- HRW-FV3 subjectively rated well below other operational CAMs
  - Uses GFS PBL scheme
  - SPC forecasters concerned of the quality of FV3-based CAM ensemble



# SPC Ensemble Review: CAMs - Days 1-2

\*\*\*Flashback from the last Ensemble Workshop\*\*\*

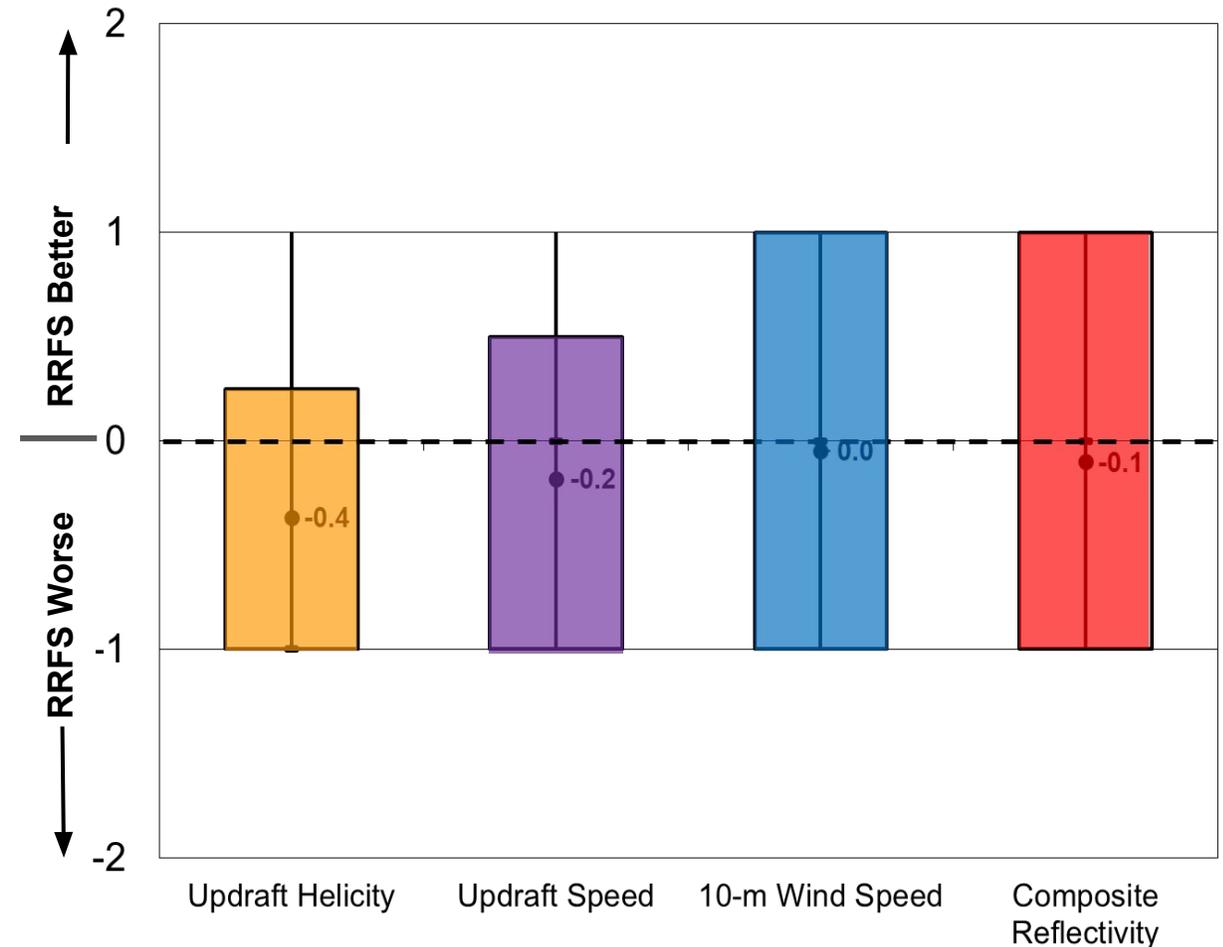
- The HREF is a very useful and skillful CAM ensemble that **sets a high operational NWS standard for next-generation CAM ensembles**
- The US is the world leader in NWP at the CAM scale, so we should acknowledge that achievement, learn from/leverage those successes, and build on our strengths



# SPC Ensemble Review: CAMs - Days 1-2

## Comparison of RRFS to HREF during 2023 HWT Spring Forecasting Experiment

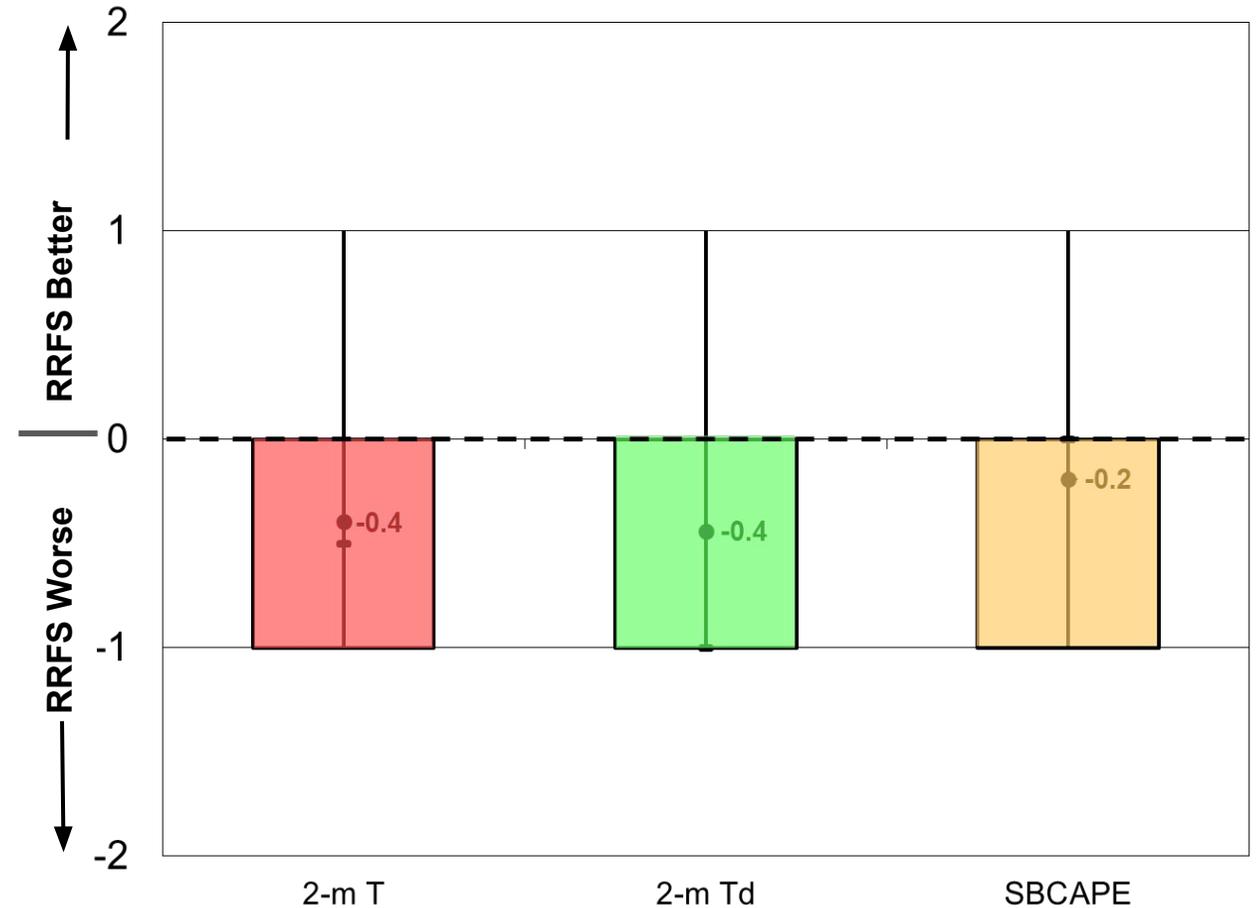
- Directly compared the 00Z single-physics RRFS ensemble to the HREF for Day 1 period
- For the storm-attribute fields, the HREF has slightly higher ratings for updraft helicity & updraft speed
- The rating distributions are more neutral (about the same) for 10-m wind speed and composite reflectivity.



# SPC Ensemble Review: CAMs - Days 1-2

*Comparison of RRFS to HREF during 2023 HWT Spring Forecasting Experiment*

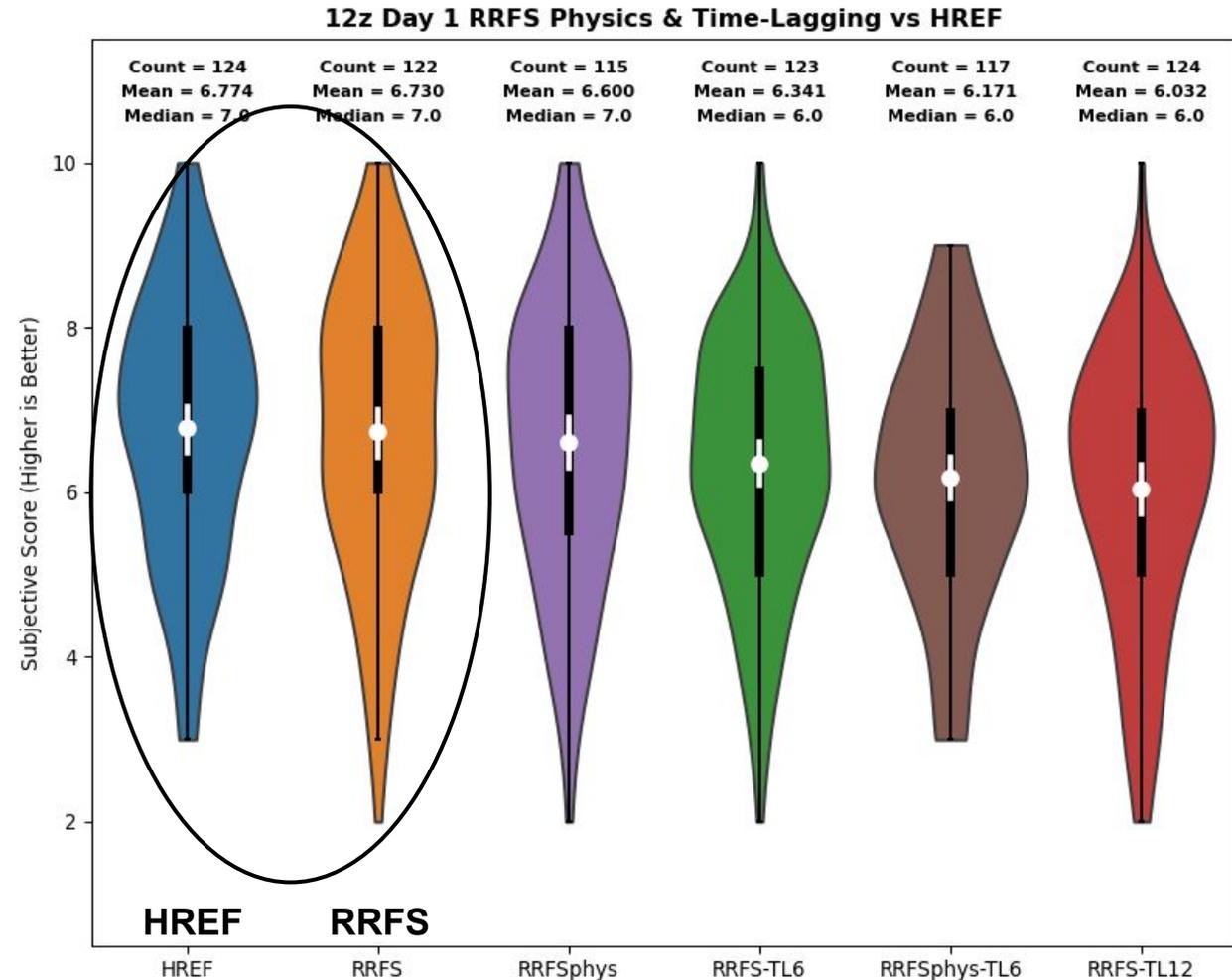
- Directly compared the 00Z single-physics RRFS ensemble to the HREF for Day 1 period
- For the environment fields between 16-20Z, the rating distributions are shifted toward RRFS being slightly worse
- SBCAPE forecasts were closer to “about the same”
- Later times (not shown) are similar with a subtle shift toward being about the same



# SPC Ensemble Review: CAMs - Days 1-2

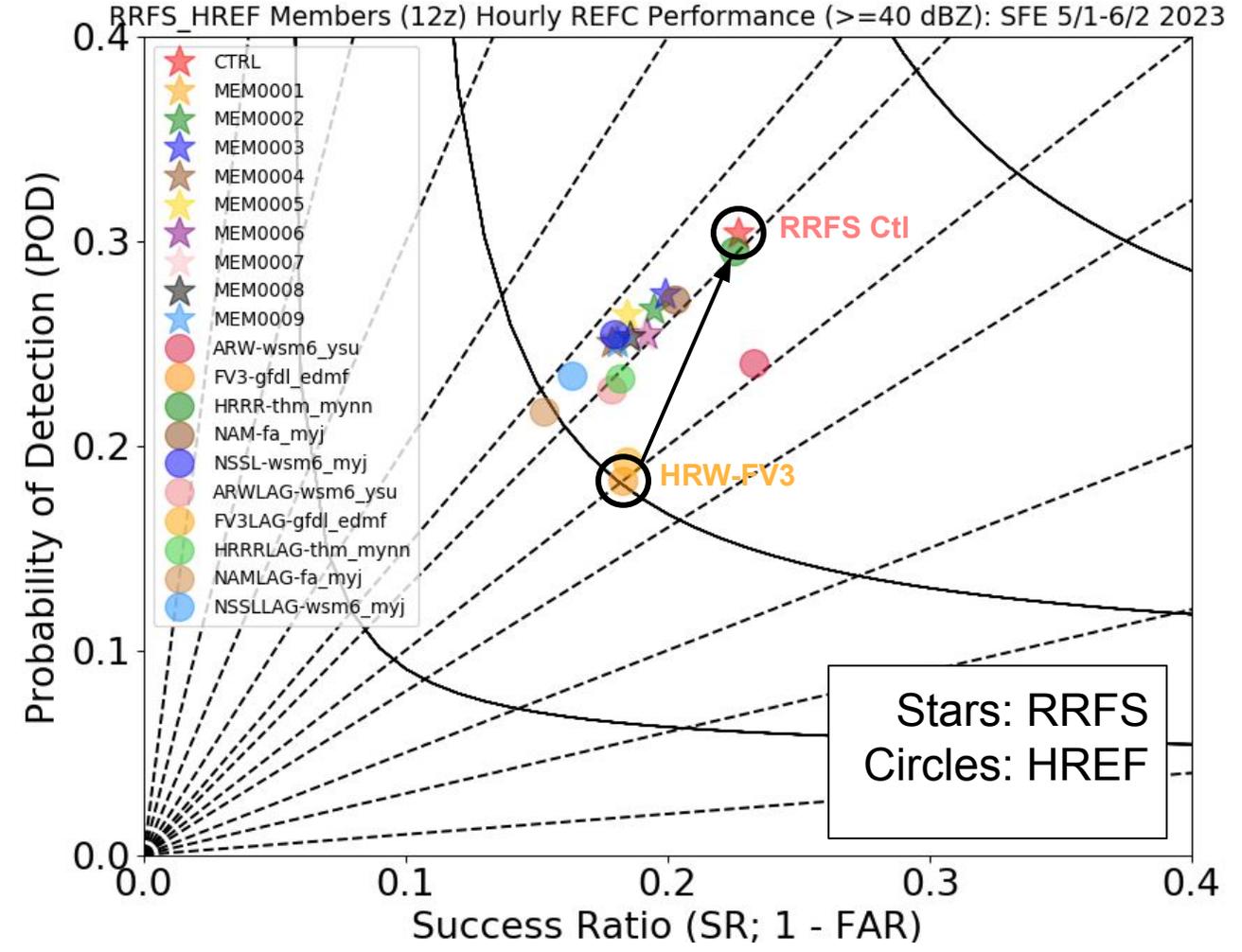
*Comparison of RRFS to HREF during 2023 HWT Spring Forecasting Experiment*

- For the 12Z cycle, evaluated multiple configurations of RRFS ensemble for comparison with HREF on Day 1
- Similar rating distributions for the 12Z HREF and single-physics RRFS covering the Day 1 period
- This is an encouraging result and ***evidence of the progress made***

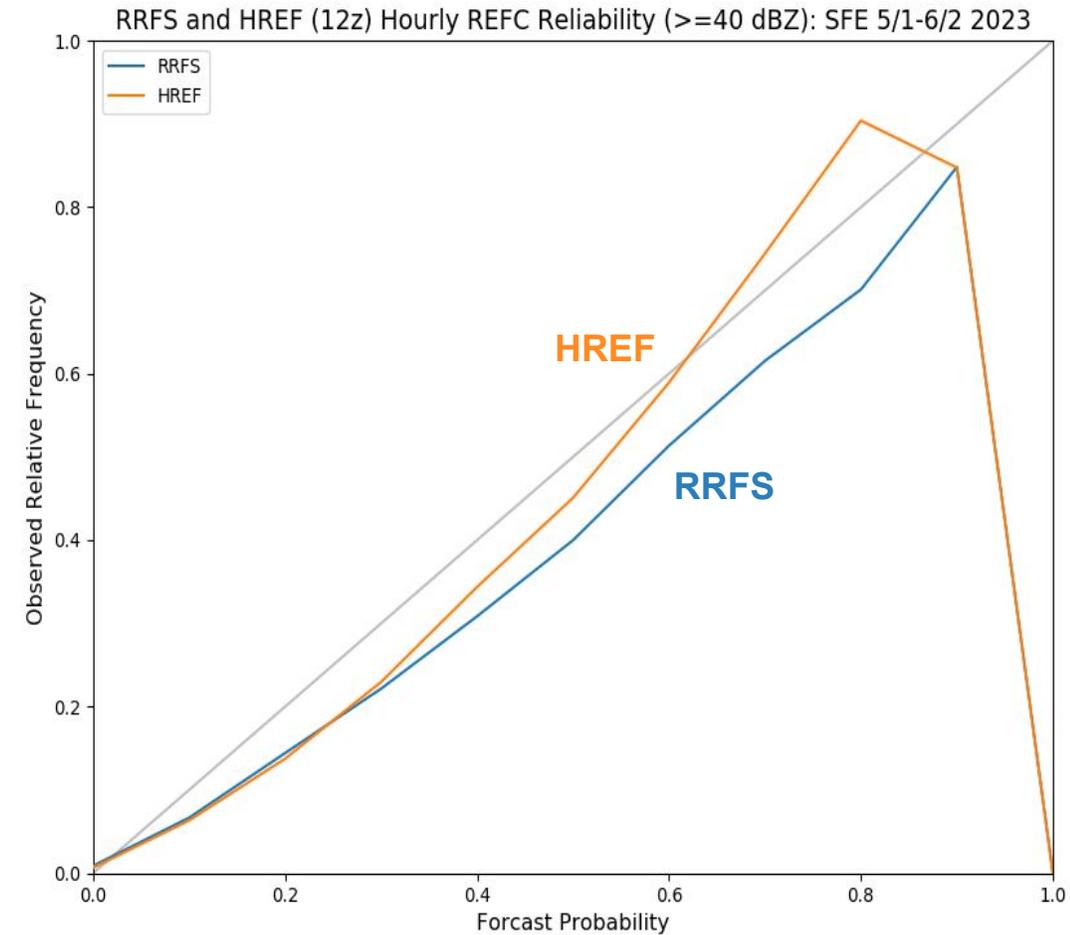
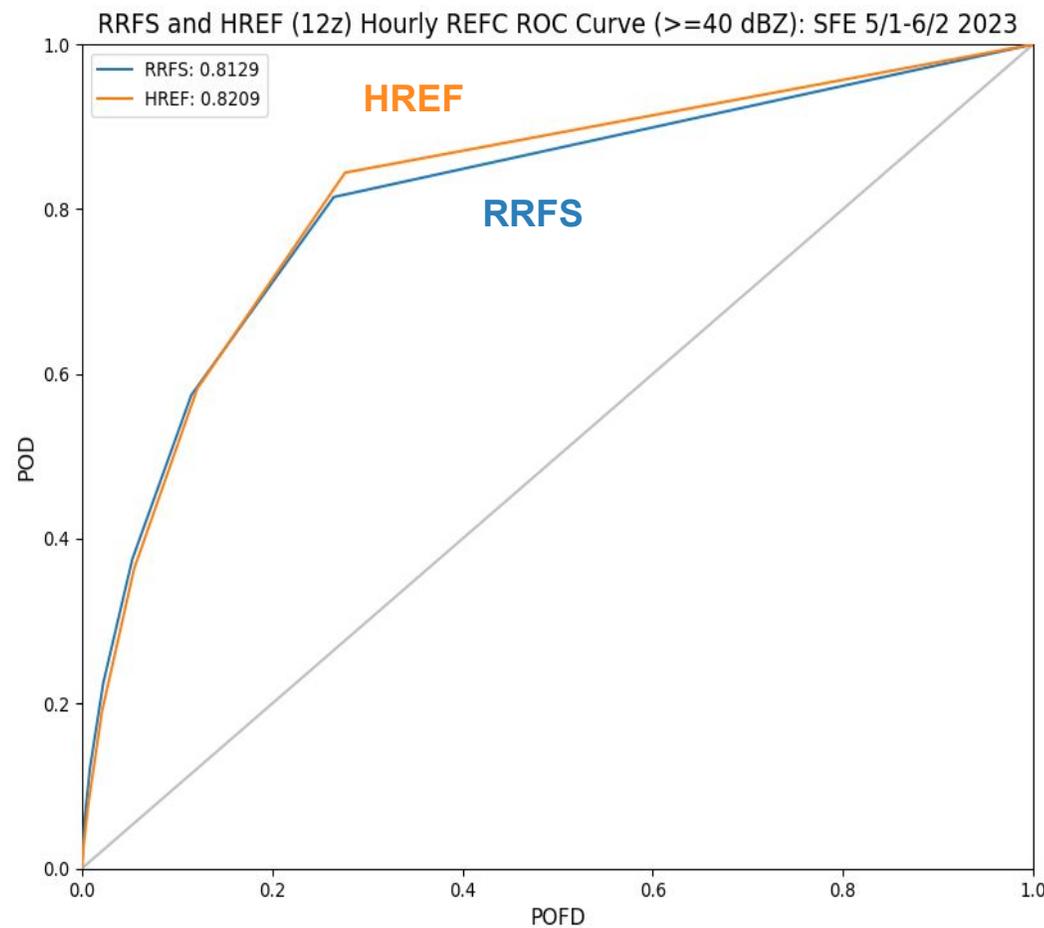


## Comparison of RRFS to HREF during 2023 HWT Spring Forecasting Experiment

- It is worth acknowledging the significant development efforts and improvements to get the FV3 core to where it is today for convective-scale forecasting
- This is a true worst-to-first story going from the HRW-FV3 to the current RRFS control member, which uses improved physics schemes
- Notice that perturbed RRFS members (stars) perform much worse than the control member



# SPC Ensemble Review: CAMs - Days 1-2

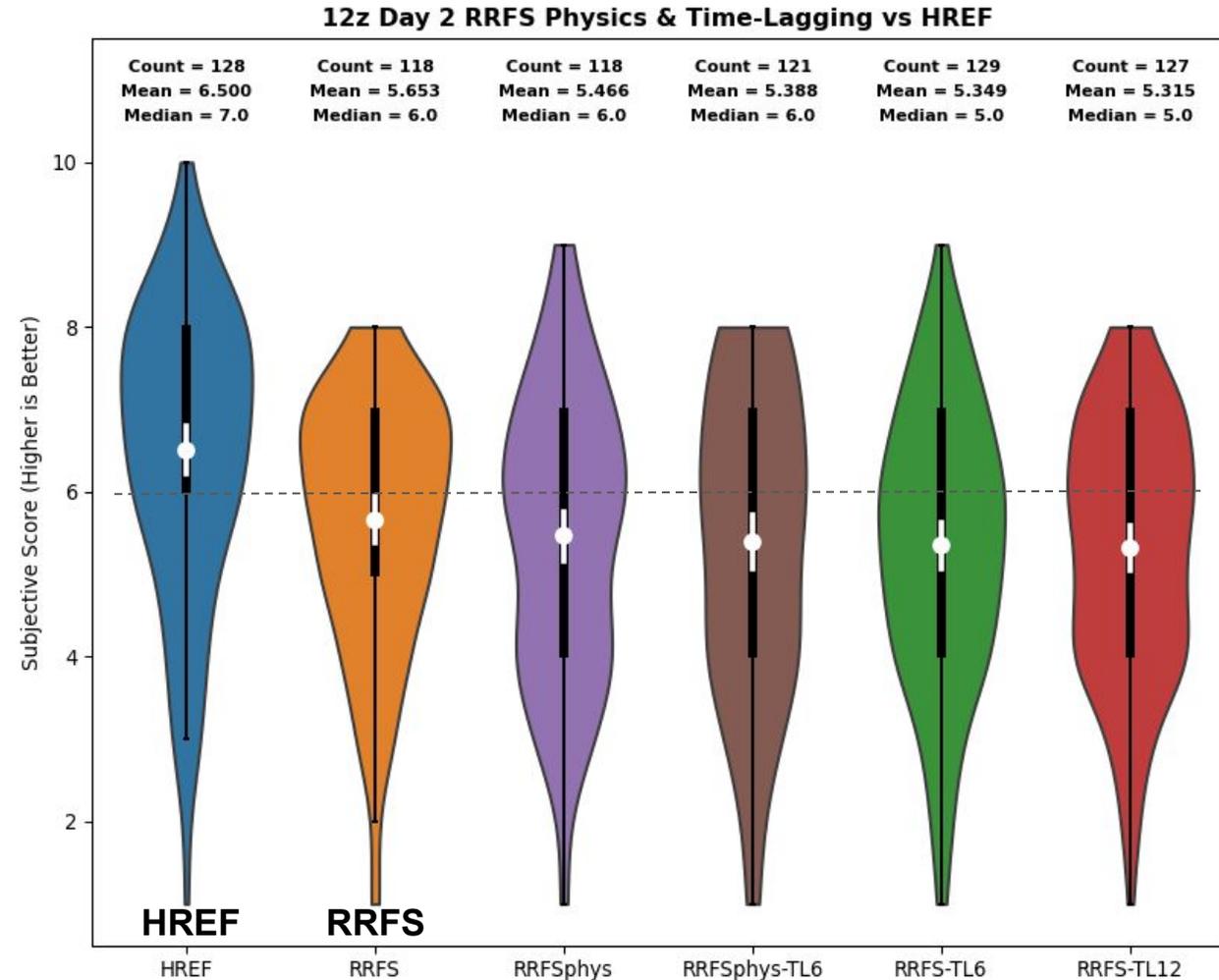


- Objective statistics still favor the HREF over the RRFS in terms of ROC AUC and reliability for reflectivity ( $\geq 40$  dBZ) valid for the convective day for the 12Z cycle

# SPC Ensemble Review: CAMs - Days 1-2

## Comparison of RRFS to HREF during 2023 HWT Spring Forecasting Experiment

- One of the more surprising and strongest results occurred with **Day 2** forecasts (f24-f48)
- HREF has much higher rated Day 2 forecasts overall compared to RRFS
- All of the RRFS configurations (physics, time-lagged) had a similar distribution of ratings for Day 2



- **Medium-Range Guidance (Days 3-8): GEFS**
  - Collaborate with SPC, MEG, others on *improving PBL thermodynamics*
  - Generate more *tailored, calibrated products* leveraging reforecast dataset
- **Days 2 & 3 Guidance: SREF, GEFS**
  - Develop a plan with the user community to *address the gap in useful mesoscale guidance for forecast range of 48-84 hours* after retirement of SREF and NAM (e.g., improve GEFS PBL, extend range of CAM ensemble forecasts, etc.)
- **CAMs - Days 1-2: HREF**
  - *Think outside of traditional ensemble design* (e.g., control member time lagging, multiple cores, etc.) for RRFS, as HREF is difficult to beat subjectively and objectively by a single-model ensemble, given the diversity of HREF composition.
  - Develop a strategy for *multi-year reforecast datasets for RRFS*. Several skillful experimental and operational calibrated guidance products have been developed from the HREF now that there are five years of archived data.