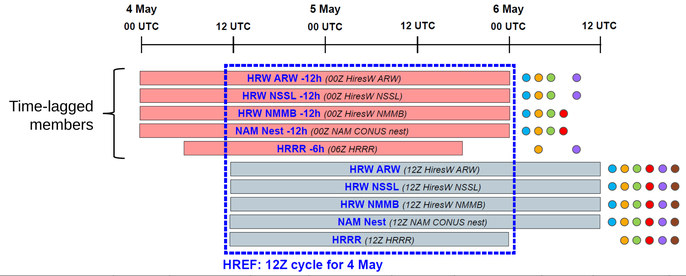
Evaluating Potential Future Configurations of the High Resolution Ensemble Forecast System

Burkely T. Gallo[1,2], Brett Roberts[1,2,3], Israel L. Jirak[2], Adam J. Clark[4], Christina P. Kalb[4], and Tara Jensen[4]

[1]Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma, [2]NOAA/NWS/NCEP Storm Prediction Center, Norman, Oklahoma, [3]NOAA/OAR National Severe Storms Laboratory, Norman, Oklahoma, [4]Developmental Testbed Center, Boulder, Colorado

Introduction

The High Resolution Ensemble Forecast (HREF) is a multi-model, multi-physics, multi-initial condition convection-allowing model (CAM) ensemble. The eight-member HREFv2 was formally implemented at NCEP in 2017 and has been a valuable operational resource, particularly for convective forecasting. The HREFv2 contains time-lagged members for another source of diversity.



HREF Configuration	HRW ARW	HRW ARW -12h	HRW NSSL	HRW NSSL -12h	HRW NMMB	HRW NMMB -12h	NAM Nest	NAM Nest -12h	HRRR	HRRR -6h
HREFv2 [8]	✓	✓	✓	✓	✓	✓	✓	✓		
HREFv2+HRRR [10]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
HREFv2+HRRR (No HRRR-TL) [9]	✓	✓	✓	✓	✓	✓	✓	✓	✓	
HREFv2+HRRR (No ARW-TL) [7]	✓		✓		✓	✓	✓	✓	✓	
HREFv2+HRRR (No NMMB-TL) [8]	✓	✓	✓	✓	✓		✓		✓	✓
HREFv2+HRRR (No TL) [5]	✓		✓		✓		✓		✓	

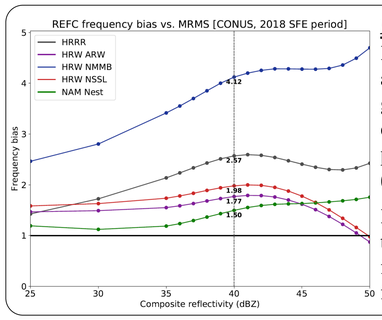
As of July 2018, the operational HRRR now runs out to 36-h lead time at 00Z, 06Z, 12Z, and 18Z, making its incorporation into the HREF technically feasible. In this study, as decision support for the next HREF update, fields of interest to convective forecasting are verified for **six candidate HREF membership configurations** (shown above).

Verification is performed on **00Z HREF** runs for **4/30/18-6/1/18** *[except 5/13 and 5/26-5/28]*

Composite Reflectivity

The **40-dBZ composite radar reflectivity (REFC)** threshold is frequently associated with convective storms, making it suitable for evaluating model skill in convective forecasting applications. The observational dataset used for verification was **MRMS Merged Reflectivity QC Composite**. 00Z HREF forecasts for 13-30h (13Z-06Z) were verified.

To evaluate ensemble forecasts of reflectivity, **neighborhood maximum ensemble probability (NMEP)** fields were generated for REFC ≥40 dBZ for each candidate configuration. The neighborhood is an 80x80 km square, and a Gaussian smoother (σ = 40 km) was applied to the grid point NMEP field prior to verification.



Member biases:

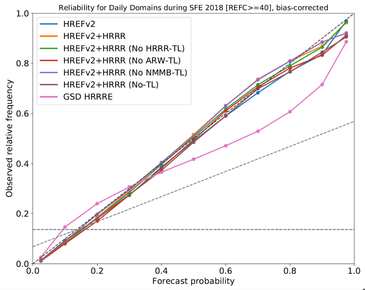
In order to produce NMEP forecasts evenly weighted among ensemble members, each deterministic CAM's grid point frequency bias for REFC ≥40 dBZ was calculated over the CONUS during the evaluation period. All CAMs were high biased, ranging from 1.50 (NAM Nest) to 4.32 (HRW NMMB).

In subsequent plots, *bias-corrected* refers to choosing the REFC threshold for each member that yields a frequency bias of ~1; *uncorrected* uses 40 dBZ as the forecast threshold for all members.

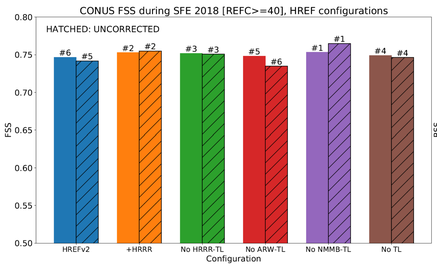
Probabilistic forecast reliability:

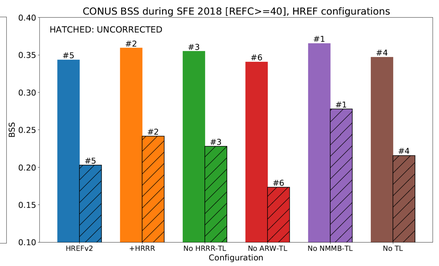
A reliability diagram (right) is shown for the bias-corrected forecasts for each candidate configuration; the GSD HRRRE is also shown for context.

All HREF configurations exhibit excellent reliability characteristics; minor differences at high probabilities slightly favor the no-NMMB-TL configuration. The HRRRE shows comparatively less resolution, suggesting the multi-dimensional diversity of HREF's membership may be successfully generating useful spread with respect to convective evolution.

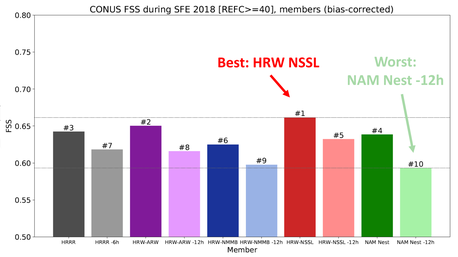


Fractions Skill Score (FSS) and Brier Skill Score (BSS) were calculated on the configuration 40-dBZ NMEP forecasts over the CONUS for the evaluation period. The configuration rankings were nearly identical for both metrics and for the bias-corrected and uncorrected forecasts. The No-NMMB-TL consistently performed best, suggesting **removing NMMB-TL members may help** improve HREF reflectivity forecasts. The baseline HREFv2 was near the bottom, suggesting the new **HRRR members are adding value** to the ensemble.





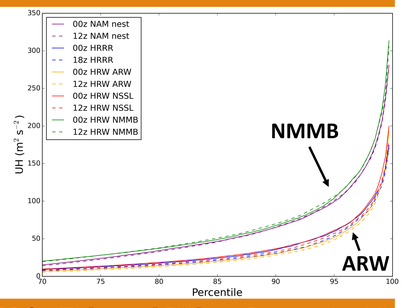
FSS among individual member CAM forecasts varied more widely than among the ensemble configurations. While the HRRR was not the best performing CAM during the evaluation period, it may often be adding **useful spread** to the baseline HREFv2 forecasts, yielding higher FSS scores for the HREF configurations with HRRR members.



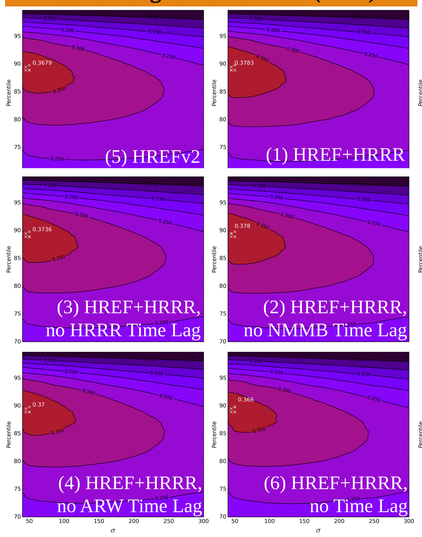
Surrogate Severe

Surrogate severe fields were calculated for each member if a UH threshold was exceeded. Averaging the member fields produced the final field for each configuration.

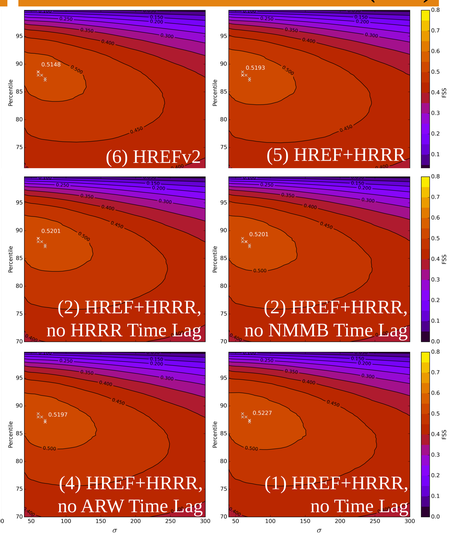
Surrogate severe fields were generated using 53 σ values and 100 percentiles of updraft helicity (UH). Using percentiles rather than thresholds accounted for model climatology; **members with NMMB cores tend to have higher UH**.



FSS for Slight Risk or Less *(n=15)*



FSS for Enhanced Risk or Greater *(n=14)*



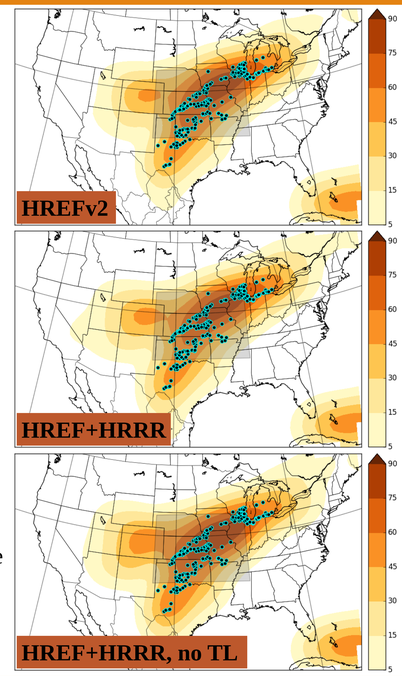
Case Study: 2 May 2018

On 2 May 2018, 239 reports occurred across the central United States, including 19 tornado reports, 173 wind reports, and 118 hail reports. These plots emulate the product a forecaster would use during their forecast process, with overlaid reports.

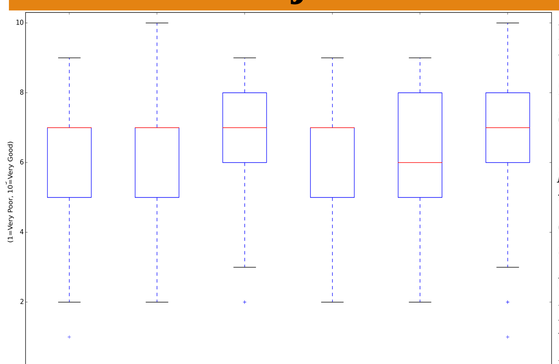
The base HREFv2 system highlighted the corridor of reports well, although it (and all configurations) missed the relative minimum between the reports around Chicago and the reports across the central and southern great plains. Slight false alarm occurred across CO in all ensemble configurations.

The addition of the HRRR to the base HREF configuration shifted the highest probabilities slightly northward, and increased the westward extent of high probabilities.

Eliminating the time-lagged members shifted the axis of high probabilities southward, suggesting that the **later model runs were focusing on the southern convection**. Higher confidence could also have resulted from having fewer members.



Subjective Evaluation



Participants were asked to "*Subjectively rate (1-10) the 4-h ensemble HMF products (UH, Updraft Speed, & 10-m Winds) during the 16-12Z period.*" **Subjective ratings were similar between ensemble configurations.** Contrary to objective results, the no NMMB TL ensemble performed slightly worse than the others. The highest mean rating was for the no-TL ensemble configuration.

Conclusions

- Six possible configurations for the high-resolution ensemble forecast system were tested during the 2018 Spring Forecasting Experiment.
- Skill scores for REFC ≥40 dBZ generally favored removing NMMB time-lagged members and adding new HRRR members, though configuration score differences were small in magnitude.
- Time-lagged members add more value to surrogate severe fields on days where SPC has issued a slight risk or less; on days with a higher risk categorization, the configurations without time-lagged members perform slightly better than the other configurations.
- HREF configurations **performed quite similarly** during SFE 2018, but the HRRR shows potential to improve HREF convective forecasts.