10.2 EVALUATION OF THE RAPID REFRESH FORECAST SYSTEM DURING THE 2023 NOAA HAZARDOUS WEATHER TESTBED SPRING FORECASTING EXPERIMENT

Jacob T. Vancil*
Cooperative Institute for Severe and High-Impact Weather Research and Operations (CIWRO), Norman, OK

Israel L. Jirak
Storm Prediction Center, Norman, OK

1. INTRODUCTION

Every spring, the Experimental Forecast Program (EFP) of the NOAA/Hazardous Weather Testbed (HWT) conducts the Spring Forecasting Experiment (SFE). This collaborative experiment is organized by the Storm Prediction Center (SPC) and the National Severe Storms Laboratory (NSSL). The 2023 edition of the HWT SFE was the first hybrid (in-person and virtual) SFE and took place in the National Weather Center in Norman, Oklahoma. The 2023 SFE occurred on weekdays between 1 May and 2 June 2023 (24 days).

SFE activities occurred over a seven-hour schedule each day. The first two-hour block each Tuesday through Friday consisted of the prior day (deterministic and ensemble) convective allowing model (CAM) and calibrated guidance evaluations. A suite of new and improved experimental CAM guidance was contributed by a large group of SFE collaborators. As in prior SFEs, all contributed CAMs were a part of an ensemble framework called the Community Leveraged Unified Ensemble (CLUE; Clark et al. 2018). Each year the CLUE is constructed by using common model specifications (e.g., grid-spacing, domain size, post-processing, etc.) so that CAM output from each contributor can be used in various controlled experiments. Additionally, the High-Resolution Ensemble Forecast system version 3 (HREFv3) and High-Resolution Rapid Refresh Version 4 (HRRRv4) were evaluated as operational modeling baselines.

Annual goals of the SFE include to accelerate research-to-operation (R2O) activities, promote operationally relevant research, and explore the the performance of CAM systems. Due to the proposed timeline of the Unified Forecast System (UFS) for future forecast model retirements and implementation, the 2023 SFE had specific research objectives. A few of these objectives included, (1) evaluate the Rapid Refresh Forecast System (RRFS) against the HREFv3, (2) evaluate the deterministic control member of the RRFS against the HRRRv4, and (3) evaluate a mixed-physics configuration of the RRFS (RRFS_mphys) against the single-physics RRFS. An emphasis was also placed on accessing applications towards severe weather forecasting for the RRFS and HREFv3/HRRRv4 systems.

Daily participant subjective ratings (1-10; with 10 being best) of the RRFS, HREFv3, HRRR, and RRFS_mphys (00 and 12 UTC cycles) were accumulated over all 24 days of the 2023 SFE. This study looks add quantitative verification metrics, in addition to the subjective ratings, to evaluate the 12 UTC performance over the entire 2023 SFE.

2. DATA AND METHODS

2.1 RAPID REFRESH FORECAST SYSTEM

The RRFS was created as part of the NOAA UFS initiative. The UFS is a community that includes researchers, developers, and users from NOAA, federal agencies, academia, and the private sector. This community was tasked with developing, improving, and implementing a new simplified suite of weather prediction systems for NOAA. NOAA's current suite of forecasting models consist of many independent forecast systems, each of which must be maintained and improved. The simplification of the model suite to a single system could increase the efficiency of future system maintenance and development.

*Corresponding author: Jacob T. Vancil, jacob.vancil@noaa.gov

Members: RRFS	ICs	LBCs	Micro-physics	PBL/SFC	LSM	Radiation	Shallow Cumulus	Dynamical Core
RRFS (ctl)	RRFS hybrid 3DEnVar	GFS	Thompson	MYNN/MYNN	RUC	RRTMG	n/a	FV3
RRFS01	enkf1	GEFS m1	Thompson	MYNN/MYNN	RUC	RRTMG	n/a	FV3
RRFS02	enkf2	GEFS m2	Thompson	MYNN/MYNN	RUC	RRTMG	n/a	FV3
RRFS03	enkf3	GEFS m3	Thompson	MYNN/MYNN	RUC	RRTMG	n/a	FV3
RRFS04	enkf4	GEFS m4	Thompson	MYNN/MYNN	RUC	RRTMG	n/a	FV3
RRFS05	enkf5	GEFS m5	Thompson	MYNN/MYNN	RUC	RRTMG	n/a	FV3
RRFS06	enkf6	GEFS m6	Thompson	MYNN/MYNN	RUC	RRTMG	n/a	FV3
RRFS07	enkf7	GEFS m7	Thompson	MYNN/MYNN	RUC	RRTMG	n/a	FV3
RRFS08	enkf8	GEFS m8	Thompson	MYNN/MYNN	RUC	RRTMG	n/a	FV3
RRFS09	enkf9	GEFS m9	Thompson	MYNN/MYNN	RUC	RRTMG	n/a	FV3

Fig. 1. SFE 2023 version of the RRFS ensemble member configurations.

The RRFS ensemble uses a single dynamical core (FV3) and a single physics suite among its ten-members. Further RRFS member configurations (IC's, LBC's, etc.) are shown (Fig. 1). A major difference between the RRFS and HREFv3 ensembles is a single- (RRFS) versus mixed-physics (HREFv3) approach in the ensemble member configurations. With the UFS's intent for the RRFS to replace the currently operational HREFv3, evaluation of the RRFS during the 2023 SFE was a major objective.

Due to the tendency of single-core, single-physics ensembles to lack sufficient forecast spread, an experimental mixed-physics RRFS ensemble (RRFS_mphys) was also evaluated in the 2023 SFE. Like the RRFS, the RRFS_mphys ensemble used a single dynamical core (FV3) and consisted of ten members. The combinations of micro-physics and PBL/SFC schemes used for each RRFS_mphys member are also shown (Fig. 2). This study performed verification of the composite reflectivity (REFC) variable from both the RRFS and RRFS_mphys ensembles.

Members: RRFSphys	ICs	LBCs	Micro-physics	PBL/SFC	LSM	Radiation	Shallow Cumulus	Dynamical Core
RRFS (ctl)	RRFS hybrid 3DEnVar	GFS	Thompson	MYNN/MYNN	RUC	RRTMG	n/a	FV3
RRFSphys01	enkf1	GEFS m1	Thompson	H-EDMF/GFS	RUC	RRTMG	saSAS Shal	FV3
RRFSphys02	enkf2	GEFS m2	Thompson	TKE-EDMF/GFS	RUC	RRTMG	saSAS Shal	FV3
RRFSphys03	enkf3	GEFS m3	Thompson	MYNN/MYNN	RUC	RRTMG	saSAS Shal	FV3
RRFSphys04	enkf4	GEFS m4	Thompson	TKE-EDMF/GFS	RUC	RRTMG	saSAS Shal	FV3
RRFSphys05	enkf5	GEFS m5	NSSL	MYNN/MYNN	RUC	RRTMG	saSAS Shal	FV3
RRFSphys06	enkf6	GEFS m6	NSSL	H-EDMF/GFS	RUC	RRTMG	saSAS Shal	FV3
RRFSphys07	enkf7	GEFS m7	NSSL	TKE-EDMF/GFS	RUC	RRTMG	saSAS Shal	FV3
RRFSphys08	enkf8	GEFS m8	NSSL	MYNN/MYNN	RUC	RRTMG	saSAS Shal	FV3
RRFSphys09	enkf9	GEFS m9	NSSL	TKE-EDMF/GFS	RUC	RRTMG	saSAS Shal	FV3

Fig. 2. SFE 2023 version of the RRFS mixed-physics ensemble member configurations.

2.2 HIGH-RESOLUTION ENSEMBLE FORECAST SYSTEM

The HREFv3 is a ten-member, multi-dynamical core, mixed-physics, and time-lagged ensemble. The five non-time-lagged members consist of the High-Resolution Window Advanced Research version of the Weather Research and Forecast Model (HRW ARW), the HRW NSSL model, the HRW North American Mesoscale Forecast System (NAM), the HRW Finite Volume Cubed Sphere (FV3) model, and the High-Resolution Rapid Refresh (HRRR) model. The remaining five members consist of 12-h time-lagged duplicates of the HRW members, and the 6-h time-lagged initialization of the HRRR. Further detail on each member's configuration is also provided (Fig. 3). This study performed verification of the composite reflectivity (REFC) variable from the HREFv3.

HREFv3	ICs	LBCs	Microphysics	PBL	dx (km)	Vertical Levels	HREF hours
HRRRv4	HRRRDAS	RAP -1h	Thompson	MYNN	3.0	50	0 – 48
HRRRv4 -6h	HRRRDAS	RAP -1h	Thompson	MYNN	3.0	50	0 – 42
HRW ARW	RAP	GFS -6h	WSM6	YSU	3.2	50	0 – 48
HRW ARW -12h	RAP	GFS -6h	WSM6	YSU	3.2	50	0 – 36
HRW FV3	GFS	GFS -6h	GFDL	EDMF	3	50	0 – 60
HRW FV3 -12h	GFS	GFS-6h	GFDL	EDMF	3	50	0 – 48
HRW NSSL	NAM	NAM -6h	WSM6	MYJ	3.2	40	0 – 48
HRW NSSL -12h	NAM	NAM -6h	WSM6	MYJ	3.2	40	0 – 36
NAM CONUS Nest	NAM	NAM	Ferrier-Aligo	MYJ	3.0	60	0 – 60
NAM CONUS Nest -12h	NAM	NAM	Ferrier-Aligo	MYJ	3.0	60	0 – 48

Fig. 3. SFE 2023 version of the HREFv3 ensemble member configurations.

2.3 MUTI-RADAR/MULTI-SENSOR SYSTEM

The Multi-Radar/Multi-Sensor System (MRMS), developed at the NSSL, is a fully automated system that quickly integrates data from multiple radars, surface and upper air observations, lightning detection systems, satellite observations, and forecast models. MRMS products aid in the detection of severe weather hazards (tornado, wind, and hail), precipitation estimations, convection, and several other products (Zhang et al. 2011). MRMS hourly merged composite reflectivity (quality controlled) was used as the observational dataset for the REFC verification. A 40-dBZ REFC threshold is frequently associated with convective storms, making it suitable to be used in this study for evaluating ensemble skill in severe weather forecasting applications.

2.4 METPLUS

The Model Evaluation Tools (MET) software was developed by the Developmental Testbed Center (DTC) with support from the 557th Weather Wind of the United States Air Force, NOAA, and the National Center for Atmospheric Research (NCAR). MET is designed to be a

customizable suite of verification tools. MET's goal is to provide a framework for reproducible verification methods and results. METplus, contains the core framework and tools of MET with additional python wrappers to improve automation and functionality. METplus will also be integrated into the UFS framework as an important tool for verification. This study specifically used MET V10.0 and METplus V5.0.

For this study, deterministic models were evaluated using the METplus GridStat tool and ensemble systems were evaluated using the METplus EnsembleStat tool. All verification metrics were calculated for each hour of the day-one convective period (12-12 UTC) throughout the entire SFE (24 days). Verification metrics were then accumulated over each hour to produce statistics over the entire SFE. The full period statistics were calculated by using the raw hourly 2x2 contingency table results.

Deterministic models were evaluated using a binary 40-km circular neighborhood with a >=40-dBZ REFC threshold. Ensemble systems were also evaluated using a 40-km circular neighborhood with the same >=40-dBZ REFC threshold. A neighborhood maximum ensemble probability (NMEP) of REFC was used for the verification of each ensemble system. Both deterministic and ensemble forecast systems used the MRMS merged composite reflectivity as the observational dataset.

3. RESULTS

3.1 RRFS AND HREFV3 ENSEMBLE VERIFICATION

Examining the performance diagram of each RRFS and HREFv3 member (12 UTC cycle) reveals interesting REFC performance characteristics. The HRRR member (of the HREFv3) and the RRFS_CTRL member (of the RRFS) are clear outperformers compared to all other respective ensemble members. The remaining members of the RRFS tended to have alike performance results. A slight outperformance of RRFS members compared to HREFv3 members is also seen (Fig 4). However, the HREFv3 ensemble contains five time-lagged members which would be expected to under-perform the non-time-lagged RRFS members. When excluding the HREFv3 time-lagged members, performance between the RRFS and HREFv3 members appear much closer.

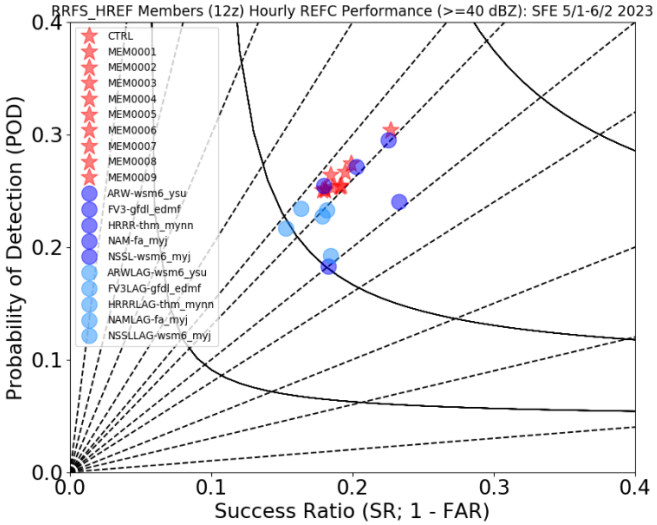

Fig. 4. RRFS and HREFv3 performance diagram over the entire 2023 SFE period. RRFS members are shown by the red stars. HREFv3 members are shown by the blue circles. Time-lagged HREFv3 members are shown as the ligh-blue circles.

Further evaluation of the RRFS and HREF shows reliability differences between the two ensembles. Both ensembles appear to equally over-forecast up until the 30% forecast REFC probability level. The HREFv3 starts to become more reliable at >30% forecast probabilities and is perfectly reliable at roughly 60% forecast probabilities. At higher forecast probabilities (>=60%) the HREFv3 tended to under-forecast REFC probabilities. The RRFS ensemble seemed to over-forecast across all forecast probability levels (Fig. 5).

Lastly, the receiver-operator-curve (ROC) and area under the curve (AUC) metrics were evaluated for the RRFS and the HREFv3. A slight advantage for the HREFv3 is shown with a higher AUC value when compared to the RRFS. However, much of the HREFv3 outperformance can be attributed to a slightly higher probability of detection (POD) of the 10% forecast REFC probability bin (Fig. 6).

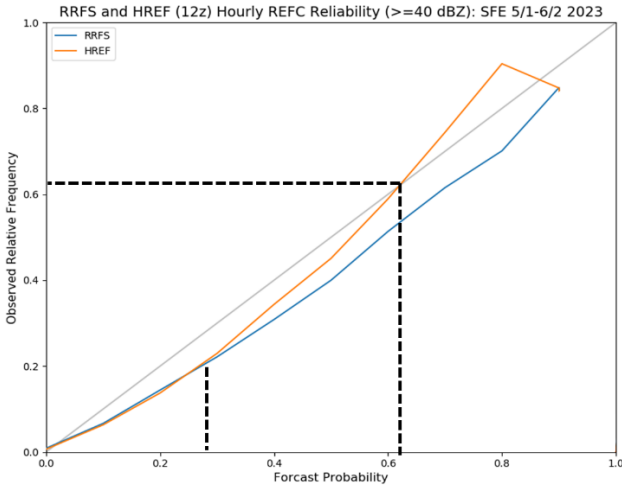

Fig. 5. RRFS and HREFv3 reliability diagram. The RRFS is shown with the blue line. The HREFv3 is shown with the orange line. Dotted lines highlight important areas of the plot.

The objective REFC verification metrics discussed appear to support the subjective 2023 SFE ensemble ratings. As previously discussed, the objective REFC verification of the RRFS and HREFv3 (12 UTC cycle) appear similar. Likewise, participant ratings of the RRFS and

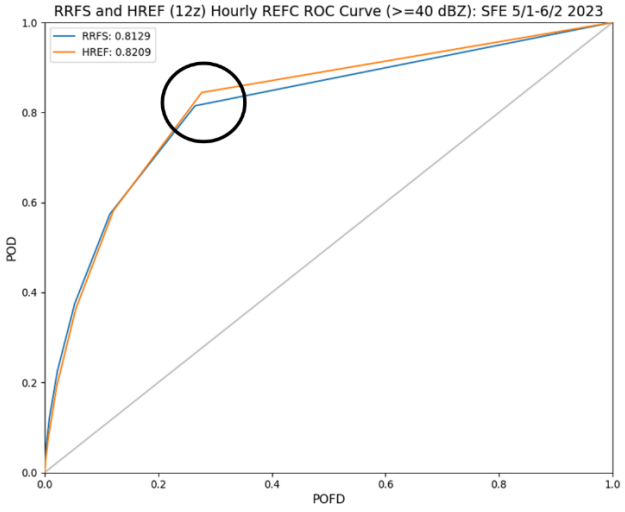

Fig. 6. RRFS and HREFv3 ROC/AUC diagram. The RRFS is shown with the blue line. The HREFv3 is shown with the orange line. The black circle highlights the 10% forecast probability bin ROC point.

HREFv3 (12 UTC cycle) over the entire 2023 SFE period were also comparable. The HREFv3 was rated slightly higher in mean subjective rating compared to the RRFS. However, the 25th and 75th percentile of ratings appear similar. A slight difference between the RRFS and the HREFv3 exists in the range of ratings given, where the RRFS ratings seem to be more spread out across the rating options (Fig. 7).

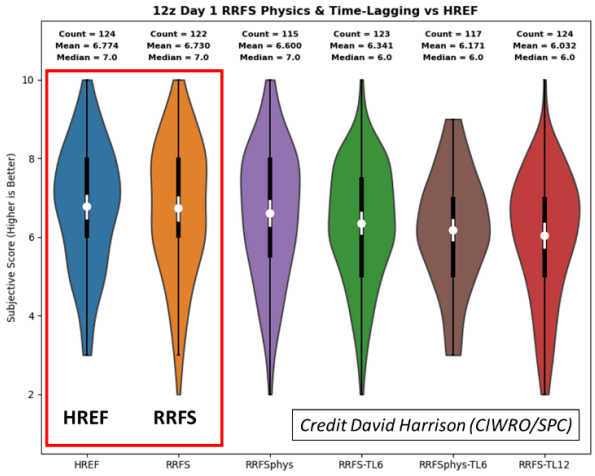

Fig. 7. The subjective ratings of the day 1 12 UTC cycle RRFS and HREFv3 from the 2023 SFE. The RRFS is shown as the orange violin. The HREFv3 is shown as the blue violin. The red square highlight only the RRFS and HREFv3.

3.2 RRFS MULTI-PHYSICS ENSEMBLE VERIFICATION

The RRFS_mphys performance was evaluated in comparison to the RRFS. As seen in the performance of the RRFS members, the RRFS_mphys members also tended to cluster around similar performance values. The RRFS members tended to outperform RRFS_mphys members. It was also shown that members using the Thompson/MYNN micro-physics/PBL schemes outperformed all other combinations (Fig. 8). The use of the GFS-EDMF PBL schemes in six of the ten RRFS_mphys members appears to be a clear disadvantage to the entire RRFS_mphys ensemble. The CTRL member, shared between both ensembles, clearly outperforms the remaining nine members of the RRFS and RRFS_mphys ensembles.

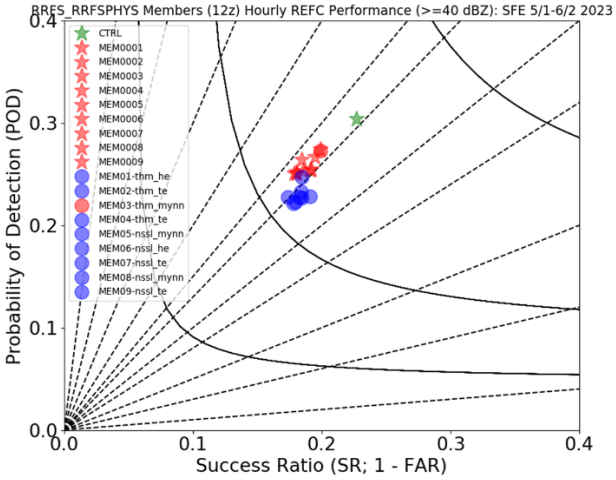

Fig. 8. RRFS and RRFS_mphys performance diagram over the entire 2023 SFE period. RRFS members are shown by the stars. RRFS_mphys members are shown by the circles. The green star is the RRFS_CTRL member shared between the ensembles. Any members using the Thompson-MYNN schemes are shown in red. Any members using any other combination of physics/PBL schemes are shown in blue.

Further evaluation of the RRFS and the RRFS_mphys ensembles reveals similar reliability characteristics between the two ensembles. The RRFS_mphys and the RRFS tend to over-forecast at all forecast REFC probability levels (Fig. 9). The ROC and AUC metrics for both ensembles were also similar (not shown).

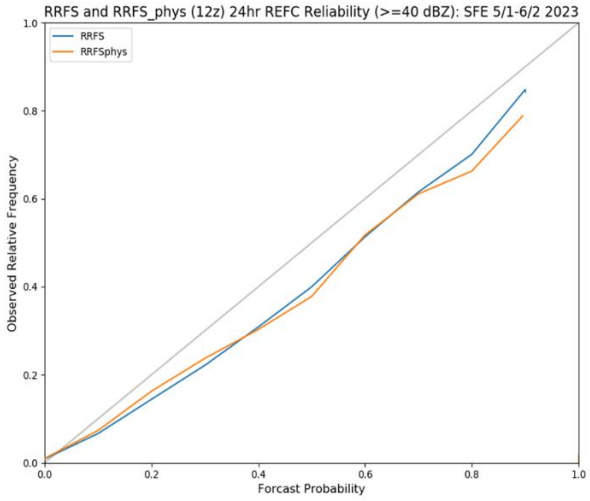

Fig. 9. RRFS and RRFS_mphys reliability diagram. The RRFS is shown with the blue line. The RRFS_mphys is shown with the orange line.

The subjective RRFS and RRFS_mphys ratings (12 UTC cycles) from the 2023 SFE supported the objective metrics just discussed. The RRFS ensemble was rated nearly identically to the RRFS_mphys by SFE participants across the entire analysis period. Both ensembles had very similar distributions of subjective ratings given by the participants (Fig. 10).

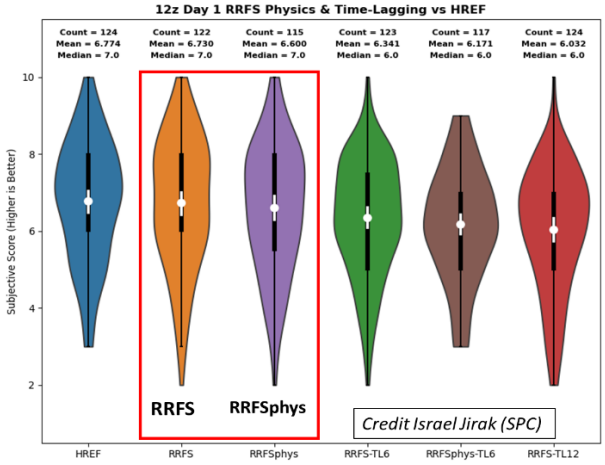

Fig. 10. The subjective ratings of the day 1 12 UTC cycle RRFS and RRFS_mphys from the 2023 SFE. The RRFS is shown as the blue violin. The RRFS_mphys is shown as the purple violin. The red square highlights the RRFS and RRFS_mphys.

4. DISCUSSION

The REFC verification of the HREFv3, RRFS, and RRFS_mphys forecast ensembles (12 UTC cycles) provided interesting results. Analysis of the HREFv3 and RRFS performance diagram may reveal a slight edge to the RRFS members. However, the clustering of the performance of the RRFS members seem to negatively impact the overall ensemble reliability. As noted from SFE participants and facilitators, the RRFS tended to produce much higher REFC ensemble probabilities across individual forecast hours. The higher probabilities appear to be a symptom of the lack of forecast spread among the RRFS members. This leads to an over-confident, over-forecast of REFC probabilities as shown in the reliability diagram. The HREFv3 had greater REFC forecast spread among its members. SFE participants also mentioned higher REFC ensemble probabilities (>70%) were not as common in the HREFv3 as the RRFS.

Performance between the RRFS and RRFS_mphys ensembles were very similar during the 2023 SFE. Most verification metrics examined for each ensemble behaved similarly. A slight edge in forecast skill of the RRFS members to the RRFS_mphys members is shown. However, all members of each ensemble tended to cluster around similar performance values. As discussed earlier, the clustering in performance is evidence of the lack of REFC forecast spread among the ensemble members. The lack of forecast spread among the RRFS_mphys ensemble came as a surprise given the combinations of micro-physics and PBL schemes used across the members. However, the RRFS_mphys members mix of configurations are not nearly as diverse as the HREFv3 members. The RRFS_mphys members still used the same dynamical core (FV3) and is evenly split between using the Thompson and NSSL microphysics schemes. Perhaps a more diverse mix of model configurations, for the RRFS_mphys members, can be evaluated in future SFE's.

This study only evaluated the 12 UTC cycles of the HREFv3, RRFS, and RRFS_mphys ensembles. Though it should be noted, the 2023 SFE also evaluated the 00 UTC cycles of these ensembles. The results of the SFE subjective ratings (for 00 UTC cycle ensembles) showed interesting results between the HREFv3 and the RRFS. While the 12 UTC cycle subjective ratings of the HREFv3 and the RRFS were very similar, the 00 UTC cycle subjective ratings revealed larger differences. For the 00 UTC ratings, the HREFv3 was rated notably higher than the RRFS (not shown). With ensemble performance varying between cycle runs not being an ideal trait, further research into the 00 UTC cycle of the RRFS is needed.

5. SUMMARY

The 2023 SFE successfully provided subjective verification of the HREFv3, RRFS, and RRFS_mphys ensemble systems. This study objectively accessed these ensembles on applications towards severe weather forecasting. Composite reflectivity verification metrics of each ensemble and each respective ensemble member (12 UTC cycle) were provided. While individual members of the RRFS may slightly outperform individual members of the HREFv3, the lack of forecast spread among the RRFS members negatively impacted the ensemble performance. The lack of forecast spread of the RRFS members results in overconfident REFC probability forecasts. While the HREFv3 was shown to be more reliable than the RRFS at forecast REFC probabilities of >30%. Likewise, the RRFS tends to produce more frequent higher REFC probabilities (>70%) then the HREFv3. From use in severe weather forecasting applications during the 2023 SFE, the 12 UTC cycle RRFS and HREFv3 performed similarly. However, given the reliability results of the RRFS and the HREFv3, an adjustment to the over-forecast REFC bias of the RRFS would be needed if used for operational severe weather forecasting.

The objective verification of these ensemble systems is essential for the advancement of the UFS goals. The annual SFE also provides a key role in accessing experimental ensembles in real-world severe hazard forecasting applications. The combination of the SFE subjective ratings and this study's objective verification metrics aim to provide feedback and potential concerns over future model retirements and possible implementation timelines. Further evaluation of these ensemble forecast systems will be necessary to advance and improve future model development.

REFERENCES

Clark, A. J., et al., 2018: The Community Leveraged Unified Ensemble (CLUE) in the 2016 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment. Bull. Amer. Meteor. Soc., 99, 1433–1448, https://doi.org/10.1175/BAMS-D-16-0309.1.

Zhang, J., and Coauthors, 2011: National Mosaic and Multi-Sensor QPE (NMQ) system: Description, results, and future plans. Bull. Amer. Meteor. Soc., 92, 1321–1338, https://doi.org/10.1175/2011BAMS-D-11-00047.1.

Prestopnik, J., J. Opatz, J. Halley Gotway, T. Jensen, J. Vigh, M. Row, C. Kalb, H. Fisher, L. Goodrich, D. Adriaansen, M. Win-Gildenmeister, G. McCabe, J. Frimel, L. Blank, T. Arbetter, 2022: The METplus Version 5.0.2 User's Guide. Developmental-Testbed Center, https://github.com/dtcenter/METplus/releases.

Clark, A., Jirak, I., et. al., 2023: Spring Forecasting Experiment 2023 conducted by the Experimental Forecast Program of the NOAA/Hazardous Weather Testbed – Program Overview and Operations Plan. NOAA/NWS/NCEP Storm Prediction Center, NOAA/OAR National Severe Storms Laboratory, Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma School of Meteorology, Institute for Public Policy Research and Analysis.