J9.2 Assessment of Probabilistic Short-Term CAM Forecasts for Informing SPC Products in the Watch-to-Warning Timeframe

Israel L. Jirak[1*], Mike Baldwin[1,2], Tim Supinie[1], Joey Picca[1,2], Patrick T. Marsh[1], Rich Thompson[1], and Russell S. Schneider[1]
[1]NOAA/NWS/NCEP/Storm Prediction Center, Norman, OK
[2]Cooperative Institute for Severe and High-Impact Weather Research and Operations, Norman, OK

1. INTRODUCTION AND MOTIVATION

Over the past several years, the Storm Prediction Center (SPC) has focused on providing information on meso-β corridors through Mesoscale Discussions (e.g., Fig. 1) in the watch-to-warning time frame. The SPC's long-term goal is to improve quantitative Impact Decision Support Services (qIDSS) by enhancing short-term probabilistic forecasts for specific corridors and population centers as a severe-weather event unfolds (i.e., "Day 0"). To support this effort, a primary objective of the 2026 Hazardous Weather Testbed (HWT) Spring Forecasting Experiment (SFE) was to determine the skill of current convection-allowing models (CAMs) and CAM ensembles in identifying these severe weather corridors at 0-6 hour lead times.

This paper provides the methodology for conducting this evaluation in the 2026 HWT SFE, initial subjective and objective results of short-term CAM ensemble performance for convective applications, and a discussion of the operational implications at SPC and the broader National Weather Service.

2. METHODOLOGY

An evaluation during the 2026 HWT SFE was designed to gauge the performance of the prototype Warn-on-Forecast System (WoFS) relative to operationally available hourly systems, specifically a time-lagged ensemble comprised of the HRRR and RRFSv1 (HRRR/RRFS-TLE).

This evaluation examined 0-6 hour forecasts initialized at two different times for storm-attribute fields, including simulated composite reflectivity (CREF), updraft helicity (UH), and 10-m winds. Participants subjectively rated the 2100 and 0000 UTC cycles of both WoFS and the HRRR/RRFS-TLE on a scale of 1-10 (with 10 being best) for hourly composite reflectivity and 4-h storm-attribute fields.

CAM Ensemble Configuration Details:

- **WoFS VLAB:** 18 members (GSI-EnKF), utilizing WRF-ARW with various PBL schemes.
- **HRRR/RRFS-TLE:** 10 members total; 1 on-time and 4 time-lagged HRRR members (t to t-4) utilizing WRF-ARW, and 5 time-lagged RRFS members (t-1 to t-5) utilizing FV3.

* *Corresponding author address*: Dr. Israel L. Jirak, NOAA/NWS/NCEP/Storm Prediction Center, 120 David L. Boren Blvd., Norman, OK 73072; e-mail: Israel.Jirak@noaa.gov

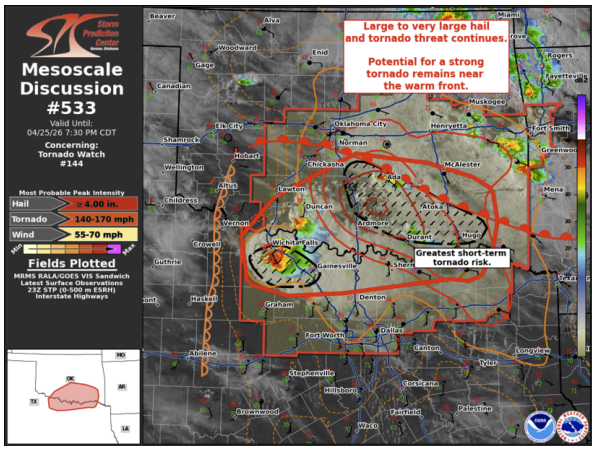


Figure 1. Example of an operational Mesoscale Discussion issued by SPC on 25 April 2026, highlighting multiple meso-β corridors of severe-weather risk.

This evaluation was conducted over the WoFS domain of the day, which was typically placed over the region with the greatest severe risk across the CONUS. Owing to a delay in getting the WoFS data into the interactive evaluation viewer, only the last 14 days of the SFE (beginning with 5 May) are included in the subjective evaluation. Lastly, given the short-term nature of these forecasts, a 21-km radius and Gaussian smoother were utilized to calculate the neighborhood probability fields for both ensembles to explore higher spatial resolution than current SPC Convective Outlooks (i.e., within 25 miles).

3. RESULTS

Some examples from the subjective evaluation are provided in this section along with the cumulative subjective evaluation results. Following the SFE, objective verification statistics were also calculated for composite reflectivity forecasts to see how well they aligned with the subjective ratings.

3.1 Subjective Evaluation

One of the first days on which this evaluation occurred was 6 May, which turned out to be one of the more active tornado days of the SFE. On this Enhanced-Risk day, eight tornadoes occurred across southern Mississippi, including a long-track (~80-mile) EF3. The 2100 UTC WoFS forecast generally had high updraft helicity (UH) probabilities in southern Mississippi (Fig. 2), but they were slightly offset from the strongest tornadoes of the day. Additionally, WoFS had

false-alarm areas of high UH probabilities across northern Alabama and west-central Louisiana, where no NWS warnings or severe weather occurred. Meanwhile, the 2100 UTC HRRR/RRFS-TLE had maximum UH probabilities centered on the long-track EF3 in southern Mississippi and much less false-alarm area across the domain (Fig. 2). Six out of seven SFE evaluators favored the HRRR/RRFS-TLE over WoFS for this event.

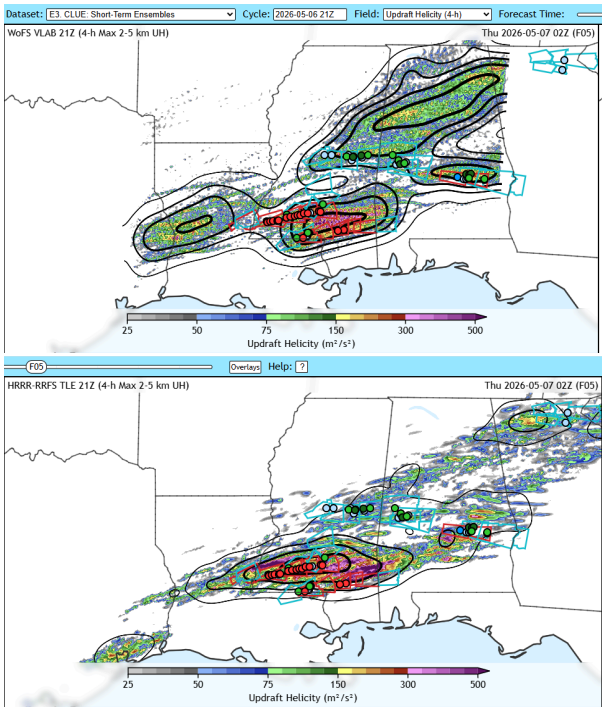


Figure 2. Four-hour neighborhood probabilities of 2-5 km AGL UH for the 2100 UTC WoFS (top panel) and HRRR/RRFS-TLE (bottom panel) valid at 0200 UTC on 7 May 2026. The neighborhood probabilities of UH exceeding the 99.85th percentile are contoured (10, 30, 50, 70, & 90%) while the ensemble maximum UH values are shaded. The preliminary local storm reports (tornado - red, wind - blue, and hail - green) and NWS warning polygons during this 4-h period are overlaid.

Another active severe weather day during the SFE was a Moderate Risk event across the central Plains on 18 May. There were 19 tornadoes across northeast Kansas, southwest Nebraska, northwest Missouri, and southwest Iowa. By evening, the convection grew upscale into a damaging QLCS across the Kansas City metropolitan area. The 2100 UTC WoFS struggled with the evolution of the QLCS, as the propagation was too slow on the north end in Iowa and not extensive enough on the southwest end in south-central Kansas (Fig. 3). Furthermore, WoFS had higher false-alarm probabilities across north-central Oklahoma, where convection did not form. The HRRR/RRFS-TLE did a better job of capturing the axis, orientation, and shape of the QLCS with its reflectivity probabilities from central Iowa extending southwestward into south-central Kansas (Fig. 3). For this event, five out of seven SFE participants favored the HRRR/RRFS-TLE forecast.

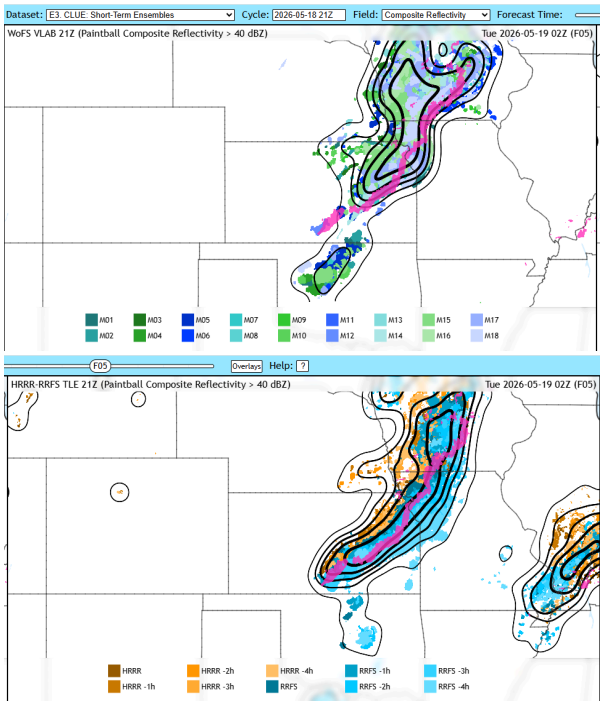


Figure 3. Neighborhood probabilities of composite reflectivity exceeding 40 dBZ for the 2100 UTC WoFS (top panel) and HRRR/RRFS-TLE (bottom panel) valid at 0200 UTC on 19 May 2026. The neighborhood probabilities exceeding 40 dBZ are contoured (10, 30, 50, 70, & 90%) while the 40 dBZ paintballs from all ensemble members are shaded. The observed MRMS composite reflectivity exceeding 40 dBZ is overlaid in magenta.

As a final example, there was a Slight Risk on 27 May across West Virginia and Virginia with numerous wind damage reports across the region. For this event, the WoFS forecast was favored over the HRRR/RRFS-TLE forecast by all seven of the SFE evaluators. Overall, WoFS forecast a better evolution of the convection across Virginia with the highest probabilities centered on the storms by evening (Fig. 4). The HRRR/RRFS-TLE was less focused and disorganized with convection across the region and struggled with the location and evolution of the strongest storms. The RRFS especially developed overly large and intense storms, as well as a number of spurious storms.

Over the course of the SFE, the subjective ratings from participants were compiled for this evaluation. Overall, the HRRR/RRFS-TLE was favored over WoFS with higher mean ratings than WoFS across all evaluated cycles and fields (Fig. 5). In fact, the 2100 UTC HRRR/RRFS-TLE actually had higher mean and median subjective ratings than the 0000 UTC WoFS (i.e., initialized three hours later) for UH. For the same cycle, the HRRR/RRFS-TLE had slightly less than a half-point edge in the mean for composite reflectivity and more than a half-point advantage for UH. These differences over the course of a multi-week experiment are noteworthy.

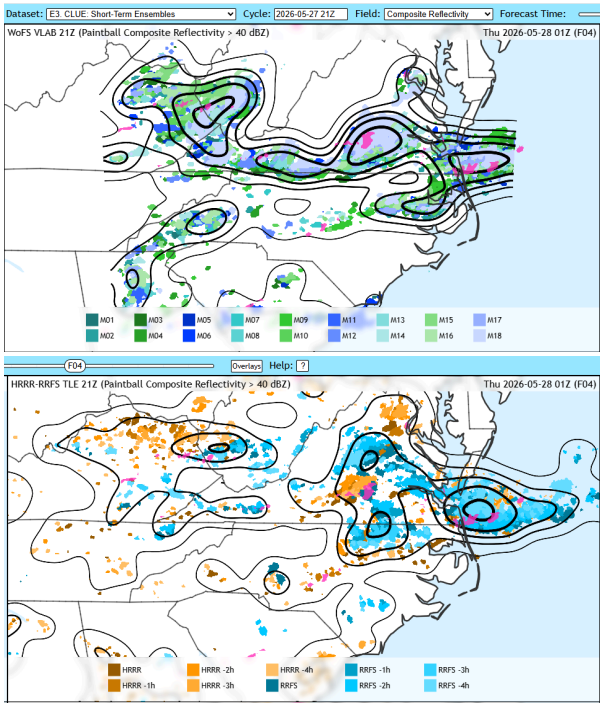


Figure 4. Same as Fig. 3, except for a 4-h forecast from 2100 UTC valid at 0100 UTC on 28 May 2026.

Another approach to examining these subjective ratings was to compare the individual participant ratings for each run of the WoFS and HRRR/RRFS-TLE. If a *majority* of participants rated one ensemble higher than the other, that ensemble forecast was considered to be favored on that day. When comparing the systems day-by-day, the HRRR/RRFS-TLE was favored by the majority of participants on most days for both the 2100 and 0000 UTC runs and for composite reflectivity (CREF) and UH::

- **2100 UTC CREF:** HRRR/RRFS-TLE favored on 10/14 days (71%).
- **0000 UTC CREF:** HRRR/RRFS-TLE favored on 9/14 days (64%).
- **2100 UTC UH:** HRRR/RRFS-TLE favored on 11/14 days (79%).
- **0000 UTC UH:** HRRR/RRFS-TLE favored on 12/14 days (86%).

3.2 Objective Evaluation

Following the SFE, objective verification of the neighborhood maximum ensemble probability (NMEP) fields for composite reflectivity exceeding 40 dBZ was performed for these CAM ensembles to see how well the objective metrics aligned with the subjective ratings. The NMEP fields are smoothed with a 21-km Gaussian filter and verified against MRMS composite reflectivity exceeding 40 dBZ dilated to a 21-km neighborhood. While both ensembles were similarly overconfident in forecast probabilities of 40 dBZ reflectivity in a reliability diagram (not shown), some interesting differences were apparent in the performance diagram. In the first 3 hours of the forecast, WoFS performs noticeably better than the HRRR/RRFS-TLE at higher probabilities (>40%) and slightly worse (primarily owing to lower POD) at lower probabilities (Fig. 6). However, the results from forecast hours 4–6 look significantly different. At these lead times, the HRRR/RRFS-TLE performs better than the WoFS at all probabilities with the performance curve shifted to the upper right part of the diagram relative to the WoFS. Again, the difference largely lies with higher POD for the HRRR/RRFS-TLE at a given probability threshold compared to the WoFS. These results are consistent with the Skinner et al. (2025) paper, which compared WoFS to the deterministic HRRR. In that paper, it is acknowledged that the primary advantage of WoFS over the HRRR is the probabilistic information provided from the WoFS ensemble, so maybe it should not be surprising that probabilistic information from a multi-model, time-lagged ensemble is competitive with the WoFS in the short-term prediction of deep convection.

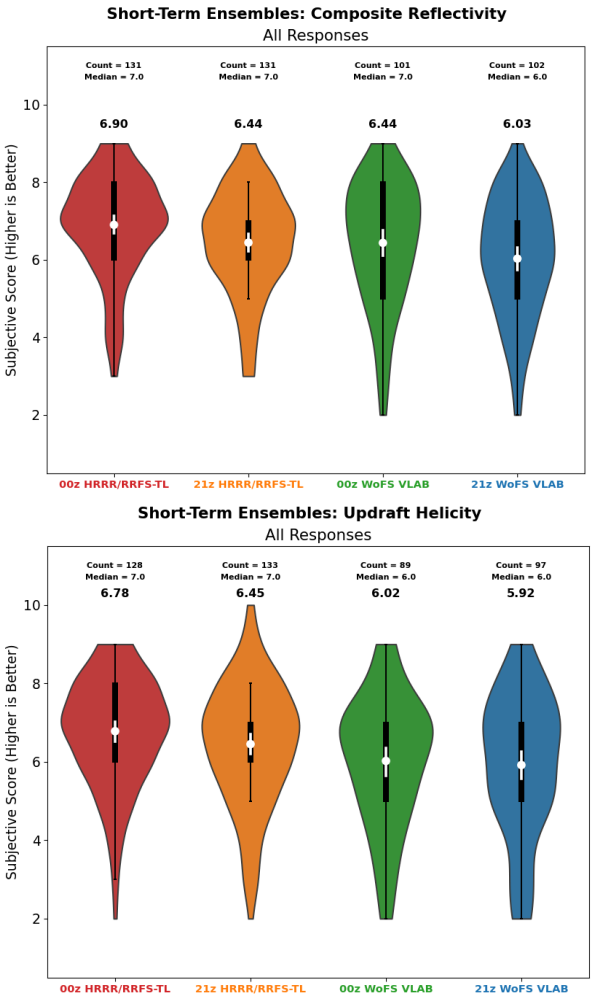


Figure 5. Distributions of subjective ratings (1 to 10) by SFE participants of composite reflectivity (top panel) and updraft helicity (bottom panel) for the 0000 UTC HRRR/RRFS-TLE (red), 2100 UTC HRRR/RRFS-TLE (orange), 0000 UTC WoFS (green), and 2100 UTC WoFS (blue) valid for forecast hours 1–6 over the WoFS domain during the SFE.

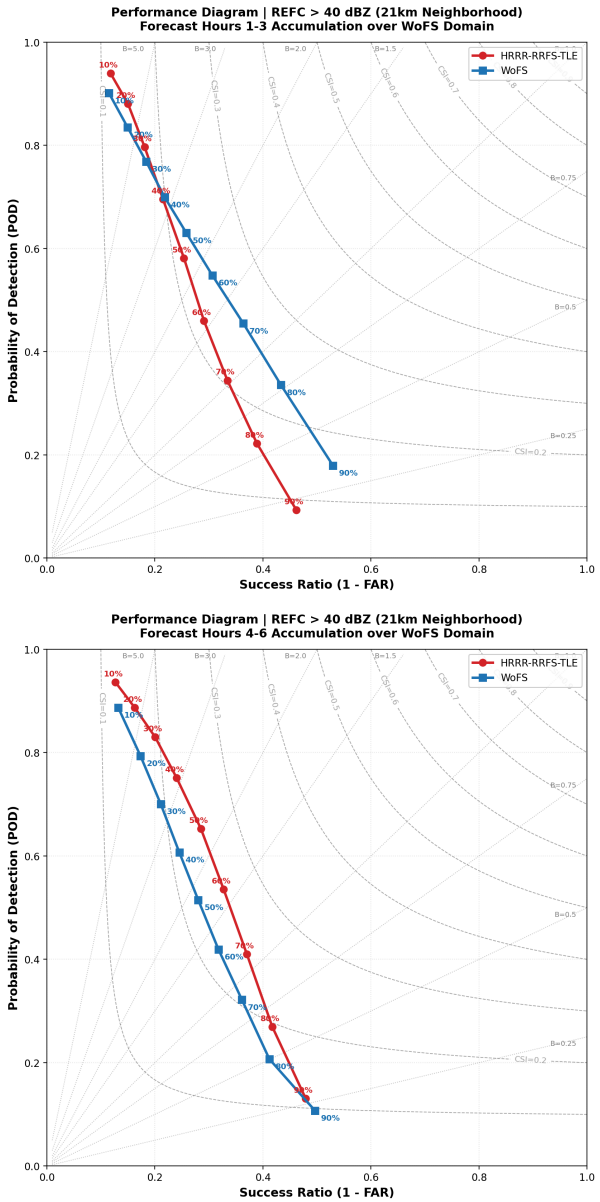


Figure 6. Performance diagrams for NMEP forecasts of CREF ≥ 40 dBZ for forecast hours 1–3 (top panel) and forecast hours 4–6 (bottom panel) for the 2100 UTC HRRR/RRFS-TLE (red) and WoFS (blue) over the WoFS domain during the SFE.

When calculating the spatial Pearson correlation between the NMEP forecasts exceeding 40 dBZ and observations, both ensembles had similar correlation to observations in the first three hours. However, the HRRR/RRFS-TLE exhibited higher correlation to observations for forecast hours 4-6, performing better at these longer forecast lead times, which is in agreement with the performance diagram results.

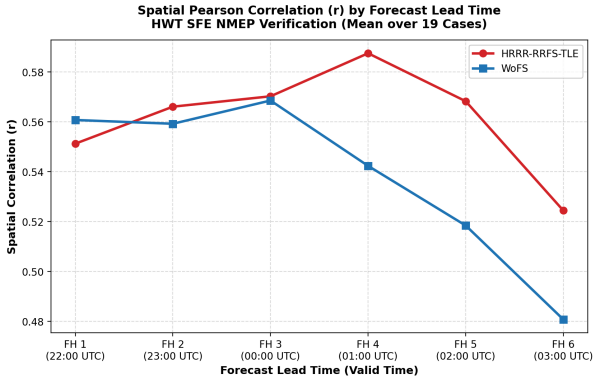


Figure 6. Spatial Pearson correlation for NMEP forecasts of CREF ≥40 dBZ to MRMS observations by forecast hour for the 2100 UTC HRRR/RRFS-TLE (red) and WoFS (blue) over the WoFS domain during the SFE.

4. SUMMARY AND OPERATIONAL IMPLICATIONS

Results from the 2026 HWT SFE indicate that the HRRR/RRFS-TLE is a viable and skillful ensemble for short-term severe weather prediction. While the performance of the HRRR/RRFS-TLE relative to the WoFS might initially seem surprising, HWT SFE results from over a decade consistently suggest that a multi-model, time-lagged control member design is an effective and practical approach for CAM ensembles.

Because the HRRR/RRFS-TLE will be available operationally 24x7 across the CONUS, SPC plans to provide a public website for this guidance after the RRFSv1 is implemented in the fall of 2026. Meanwhile, WoFS remains a useful tool for assessing individual storm attributes and evolution for approximately 3 hours after convective initiation (CI), largely due to its 15-minute data assimilation cycling. Circling back to the motivation for this study, current short-term CAM guidance appears highly useful for supporting the SPC "Day 0" operational concept (e.g., 6 May in Fig. 7).

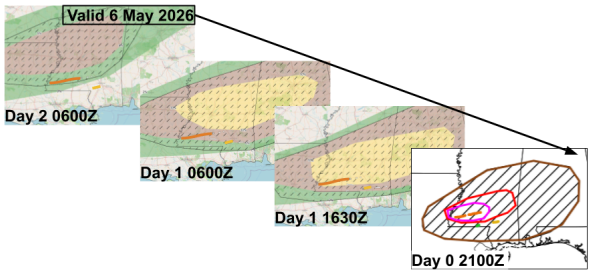


Figure 7. Evolution of operational SPC tornado outlooks to the prototype short-term outlook from the HWT SFE valid on 6 May 2026: 0600 UTC Day 2, 0600 UTC Day 1, 1630 UTC Day, & 2100 UTC Day 0. The two significant tornadoes are overlaid.

REFERENCES

Skinner, P. S., and coauthors, 2025: A comparison of short-term convective forecasts between the Warn-on-Forecast System (WoFS) and High-Resolution Rapid Refresh (HRRR). *Wea. Forecasting*, **40**, 1839–1858, https://doi.org/10.1175/WAF-D-24-0238.1.