Regional and Seasonal Biases in Convection-Allowing Model Forecasts of Near-Surface Temperature and Moisture

ANDREW R. WADE,[a,b] ISRAEL L. JIRAK,[b] AND ANTHONY W. LYZA[a,c]

[a] *Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma, Norman, Oklahoma*
[b] *NOAA/NCEP/Storm Prediction Center, Norman, Oklahoma*
[c] *NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma*

(Manuscript received 20 July 2023, in final form 2 October 2023, accepted 3 October 2023)

ABSTRACT: This study investigates regional, seasonal biases in convection-allowing model forecasts of near-surface temperature and dewpoint in areas of particular importance to forecasts of severe local storms. One method compares model forecasts with objective analyses of observed conditions in the inflow sectors of reported tornadoes. A second method captures a broader sample of environments, comparing model forecasts with surface observations under certain warm-sector criteria. Both methods reveal a cold bias across all models tested in Southeast U.S. cool-season warm sectors. This is an operationally important bias given the thermodynamic sensitivity of instability-limited severe weather that is common in the Southeast cool season. There is not a clear bias across models in the Great Plains warm season, but instead more varied behavior with differing model physics.

SIGNIFICANCE STATEMENT: The severity of thunderstorms and the types of hazards they produce depend in part on the low-level temperature and moisture in the near-storm environment. It is important for numerical forecast models to accurately represent these fields in forecasts of severe weather events. We show that the most widely used short-term, high-resolution forecast models have a consistent cold bias of about 1 K (up to 2 K in certain cases) in storm environments in the southeastern U.S. cool season. Human forecasters must recognize and adjust for this bias, and future model development should aim to improve it.

KEYWORDS: Cloud-resolving models; Model errors; Model evaluation/performance; Numerical weather prediction/forecasting; Regional models

1. Introduction

The High-Resolution Ensemble Forecast (HREF) system (Roberts et al. 2019), comprising 10 current and time-lagged solutions from five member models, is the primary convection-allowing model (CAM) guidance in operational use in the United States. HREF forecasts verify favorably in comparison with other CAM ensembles (Clark et al. 2021) and benefit from its model diversity (Roberts et al. 2020). For models whose primary purpose is explicit representation of realistic thunderstorms, forecast properties of potential convective inflow are much more operationally important than those of air masses that clearly do not support deep convection. Yet traditional model verification, even for CAMs, usually does not specifically isolate mesoscale environments that may support convective storms, with a few exceptions comparing individual models with severe weather-related observations (Coniglio 2012; Coniglio et al. 2013).

From 2021 to 2023, the lead author maintained a real-time comparison of near-surface HREF forecast fields with Real-Time Mesoscale Analysis (RTMA; De Pondeca et al. 2011; Morris et al. 2020) fields in support of National Weather Service (NWS) Storm Prediction Center (SPC) operations. Inspection of these error fields during cool-season severe weather events in the southeastern U.S. anecdotally suggested a cool bias in temperature 2 m above ground level (AGL) in warm-sector or inflow environments. Figure 1 demonstrates this possible bias in three different cases during the 2021/22 cool season, including at the onset of a high-impact wintertime tornado outbreak on 10 December 2021 (Fig. 1a). [Figure 1 here uses the Unrestricted Mesoscale Analysis (URMA), similar to RTMA but preferred for retrospective verification because of its inclusion of higher-latency observations than RTMA.] While both warm and cold errors are present in these examples, the areas of interest are the mesoscale inflow sectors to the immediate south and southeast of where tornadoes occurred over the subsequent hours. These cold errors motivate a systematic investigation of many more cases. Quantifying any such bias is particularly important for this region and season because of the prevalence of high-shear, low-CAPE severe weather (Sherburn and Parker 2014) and accompanying convective forecast sensitivity to minor thermodynamic errors. Cohen et al. (2017) demonstrated the considerable sensitivity of select Southeast cool-season events to planetary boundary layer (PBL) parameterizations alone; the set of current and future operational CAMs varies widely not only in PBL schemes, but other parameterizations, initial and boundary conditions, and model cores.

The National Oceanic and Atmospheric Administration's long-term plan to simplify its operational forecast model suite includes the new Rapid Refresh Forecast System (RRFS) eventually superseding the HREF system (Alexander et al. 2023). As

Corresponding author: Andrew R. Wade, andrew.wade@noaa.gov

DOI: 10.1175/WAF-D-23-0120.1

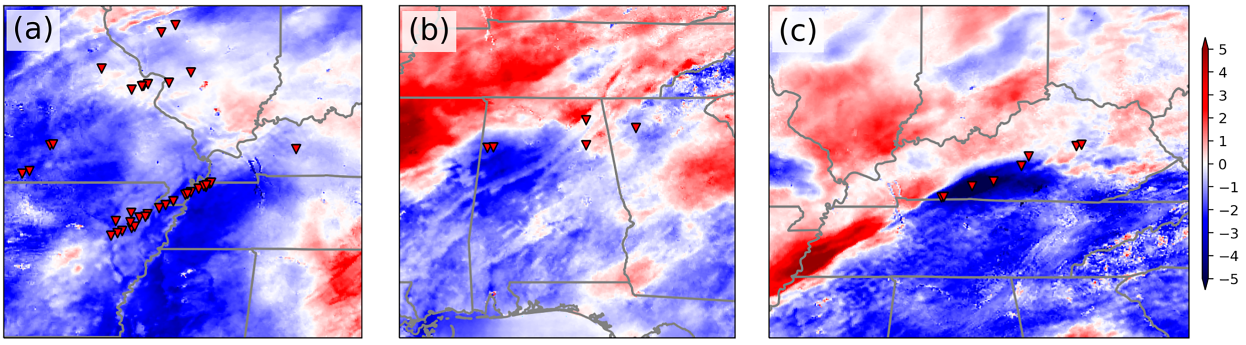

FIG. 1. HREF mean 2-m temperature errors (K) vs URMA (color fill) and tornadoes over the subsequent 4 h (red triangles), valid at (a) 2300 UTC 10 Dec 2021, (b) 0000 UTC 30 Dec 2021, and (c) 1700 UTC 1 Jan 2022.

of publication time, the first operational implementation of RRFS is expected in fiscal year 2025. In this study, we examine the currently operational HREF members as well as an RRFS prototype, where available, for near-surface temperature and moisture biases, using two distinct methods of targeting potential storm inflow environments.

2. Data and methods

a. Tornado-centered errors versus URMA

In the first method, tornado reports are obtained from the National Centers for Environmental Information's *Storm Data*. These reports are spatiotemporally filtered, retaining the highest-rated tornadoes first and working downward, so that none fall within 250 km and 3 h of each other, increasing independence of the mesoscale environments sampled and preventing a small number of highly reported events from dominating the sample. From HREF member forecasts valid 0–1 h before the report time and initialized 6–18 h prior to that time at either 0000 or 1200 UTC, 2-m temperature and dewpoint fields are extracted in 400×400 km^2 regions centered on each tornado report. Then, URMA fields valid at the same time are regridded via bilinear interpolation from the 2.5-km URMA grid to the 3-km HREF grid. Resulting difference fields (forecast minus URMA) for each member model are averaged across cases for the Southeast cool season (15 October–15 March, autumn 2019–spring 2023; ~20 months) and for the Great Plains warm season (15 March–15 October, spring 2019–autumn 2022; ~28 months), with geographic domains defined in Fig. 2. This method has the benefit of directly capturing the environments of most importance to these CAMs' utility in forecasting severe local storms.

b. Warm-sector errors versus ASOS observations

The former method is somewhat limited by the sample size available for the two newer members in version 3 of the HREF: the High-Resolution Rapid Refresh (HRRR; Dowell et al. 2022; James et al. 2022), version 4, and the High-Resolution Window FV3. Furthermore, in the Southeast cool season, the tornado-centered method may be biased toward errors of a particular sign. Warm sectors that are cooler than forecast may not produce tornadoes at all in CAPE-limited winter regimes, thereby escaping the sample. Both of these limitations are avoided by verifying forecasts against surface observations within warm sectors, regardless of whether severe storms are observed nearby. Seven Automated Surface Observing System (ASOS) sites spread across each region are used for verification: KBNA, KHSV, KJAN, KMEM, KMOB, KTCL, and KTUP for the Southeast and KABI, KABR, KBGD, KDDC, KLBF, KLNK, and KOUN for the Great Plains (Fig. 2). Observations at 0000, 0600, 1200, 1800, and 2100 UTC over the seasons defined in the previous section are retrieved and ad hoc "warm sector" criteria are applied. For the Southeast cool season, temperature must be at least 16°C, dewpoint at least 13°C, and the southerly component of 10-m wind at least 3 m s^{-1}. While these criteria are arbitrary, sensitivity tests varying the requisite temperature and dewpoint by 1°C in each direction found no systematic effect on the error distributions. For the Great Plains warm season, temperature must be at least 21°C, dewpoint at least 16°C, and the

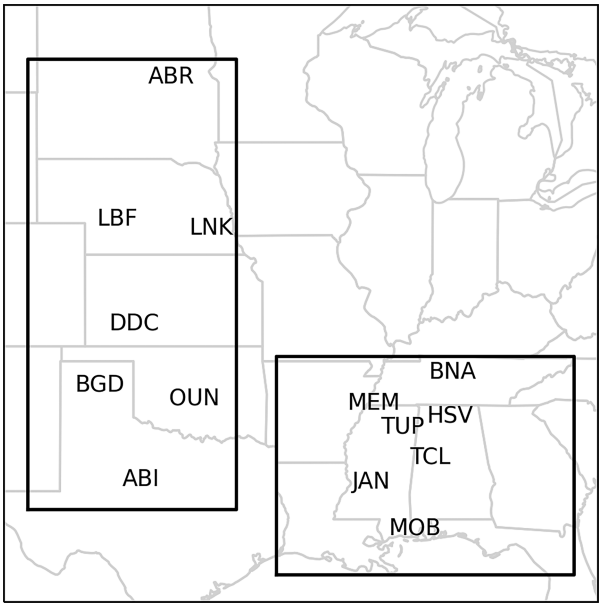

FIG. 2. Great Plains and Southeast domains (outlined boxes) of tornado reports used for the method detailed in section 2a, and the locations of ASOS sites (text labels) used for the method detailed in section 2b.

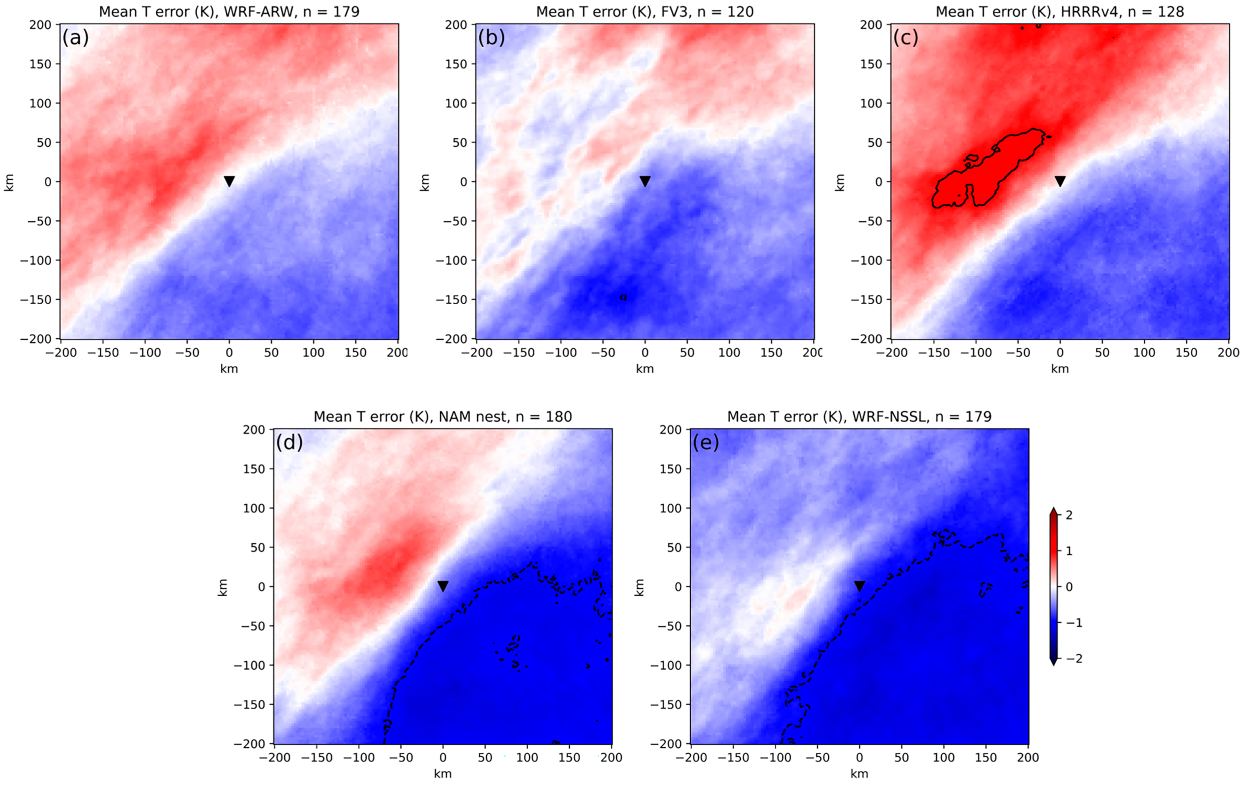

FIG. 3. Mean temperature errors (K) vs URMA for each HREF member, centered on filtered tornado reports (black triangle at center) in the Southeast cool season, via the method described in section 2a. Black contours are every 1 K, solid for positive values and dashed for negative values.

southerly component of 10-m wind at least 3 m s^{-1}. An additional criterion of HRRR-analyzed mixed-layer CAPE (not available in the two-dimensional URMA) of at least 100 J kg^{-1} is also applied for Great Plains warm-season observations, since the temperature, dewpoint, and wind criteria are near climatology for much of this season, but this does not notably change the results. Observations meeting these criteria are then compared with 6–12-h HREF member forecasts (i.e., not including HREF's time-lagged members) valid at the nearest gridpoint at the same time. This relatively short lead time demonstrates how quickly biases can emerge and emulates forecasters' routine use of 1200 UTC initializations to forecast for the 1800–0000 UTC period.

3. Results

a. *Tornado-centered errors versus URMA*

In the tornado-centered mean error fields, the region of interest is not the location of the tornado itself but the broad inflow sector to the east and south. All HREF member models have a cold bias in that sector in the Southeast cool season (Fig. 3). Two members, the North American Mesoscale (NAM) model convection-allowing nest and the National Severe Storms Laboratory's (NSSL) version (Kain et al. 2010) of the Weather Research and Forecasting (WRF; Skamarock et al. 2008) Model, are colder than the others. These two members are the only two that use the Mellor–Yamada–Janjić (MYJ; Janjić 1994) planetary boundary layer (PBL) parameterization, a local scheme with cool, moist biases (Hu et al. 2010). These members' mean inflow sectors contain large areas of mean errors colder than −1 K. However, undermixing is not the only source of these errors, and perhaps not even the primary one; the cold bias is only somewhat reduced in members using other PBL schemes (including a nonlocal scheme in the WRF-ARW) and the near-surface moisture field is essentially unbiased (to slightly dry) in the inflow sector for most members (Fig. 4). We also tested the effect of masking grid points in individual cases where the model produced precipitation 0–3 h before the valid time, to remove any cold errors resulting from small errors in storm placement. This had little effect on inflow-sector cold biases. Furthermore, this technique might bias results by limiting any cold errors resulting from modeled precipitation in observed precipitation-free inflow, but not any warm errors resulting from observed precipitation in modeled precipitation-free inflow. Only the original, unmasked version is shown here.

In contrast to the nearly identical biases present in all members in the Southeast cool season, the Great Plains warm season produces a variety of distinct model behaviors in both temperature (Fig. 5) and dewpoint (Fig. 6) fields. While some of the local biases, particularly in moisture forecasts, are larger, they vary considerably across members and do not all have the same sign. This wider range of behaviors is more intuitive for a mixed-

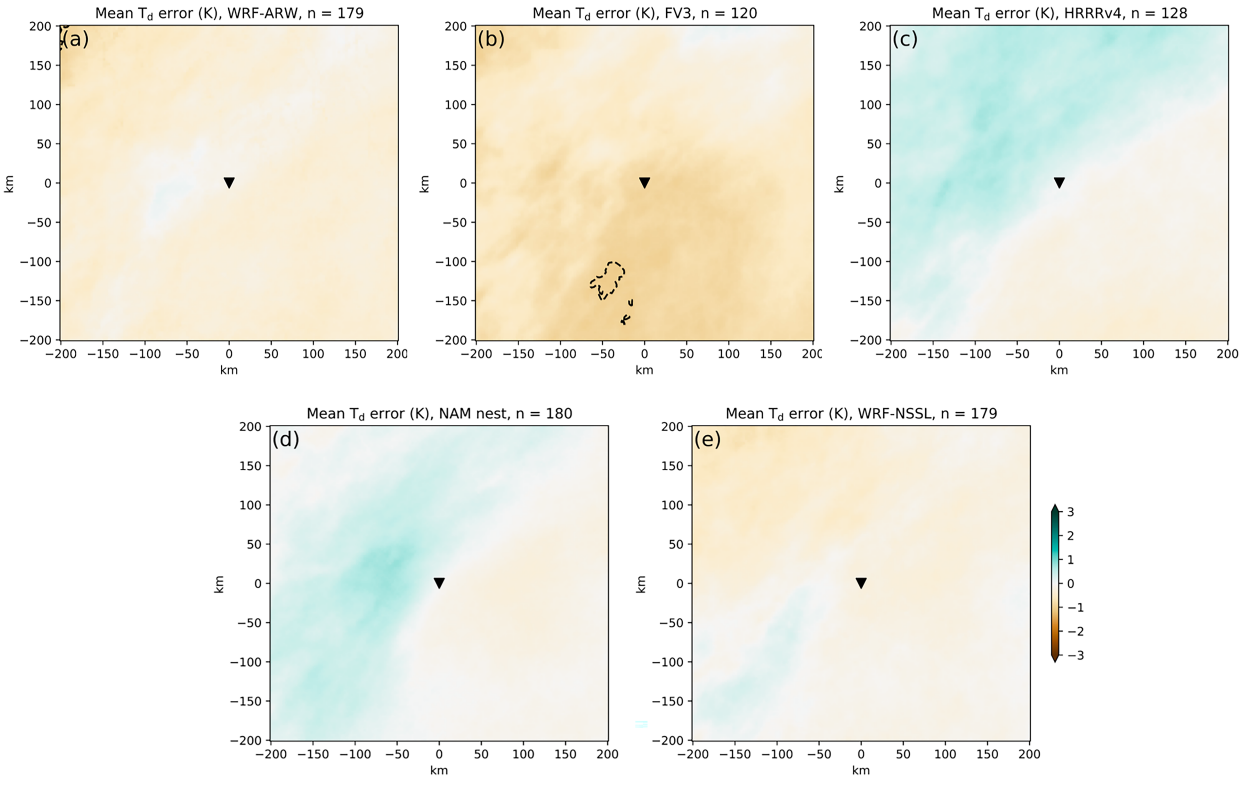

FIG. 4. As in Fig. 3, but for dewpoint temperature.

physics ensemble. Some degree of at least localized cold mean errors still appears in a majority of members, but these are more spatially heterogeneous than the cold errors in the Southeast cool-season inflow sectors.

b. *Warm-sector errors versus ASOS*

Although targeting a much broader set of environments and verifying with point observations instead of an objective analysis, the ASOS-based method produces a result similar to the tornado-centered method. For Southeast cool-season warm sectors, a cold bias on the order of 1 K exists across HREF members (Fig. 7), with the two members that use the MYJ scheme (the NSSL WRF and the NAM nest) somewhat colder. Dewpoint errors (Fig. 8) deviate slightly more from the tornado-centered results with moist biases evident in the members that use MYJ. The larger sample allows stratification by time of day. The 1200 UTC errors (Fig. 9) are nearly uniform across members, but 2100 UTC errors (Fig. 10) reveal a sizable gap between the NSSL WRF and NAM nest, which have an ~2-K cold bias near peak heating, and the other members. However, members not using MYJ retain a cold bias even at 2100 UTC. Errors may also be stratified by observing sites' reported low-level sky cover categories, with no meaningful differences (not shown here); the cold bias persists from clear skies or few clouds all the way to overcast conditions. Operational users of these models should also note that the temperature errors become slightly worse, with distributions shifted roughly 0.5 K colder, in the day-2 time frame at forecast hours 24–36 (Fig. 11).

Meanwhile, results from the Great Plains warm season indicate small temperature biases (Fig. 12) with no shared bias across HREF members. Dewpoint errors (Fig. 13) are consistent with the tornado-centered framework: WRF-ARW, FV3, and HRRR forecasts are dry, and the two members using the MYJ scheme are neutral. Despite the additional criterion of at least 100 J kg^{-1} MLCAPE, the Great Plains warm-season results are probably still less specific to convective inflow environments than the Southeast cool-season results, since the warm-sector criteria used here are near climatology for much of the Great Plains during the warm season.

c. *Looking ahead to RRFS*

Three caveats accompany the RRFS data presented here. First, the sample is necessarily limited, as the data have only recently been routinely available. Second, active RRFS development continued both during and after the study period, and the model that produced these fields will differ somewhat from the eventual operational model. Third, the RRFS solutions evaluated here come from only the control member of what will be a full RRFS ensemble, although we refer to the single member here as simply "RRFS." Nevertheless, we hope that evaluating the currently available data offers forecasters some foreknowledge during the transition to a new modeling framework.

For the RRFS period from the spring of 2022 to the early spring of 2023, verification against warm-sector ASOS observations allows a more robust sample than tornado-centered verification. This method suggests that RRFS is fairly HRRR-

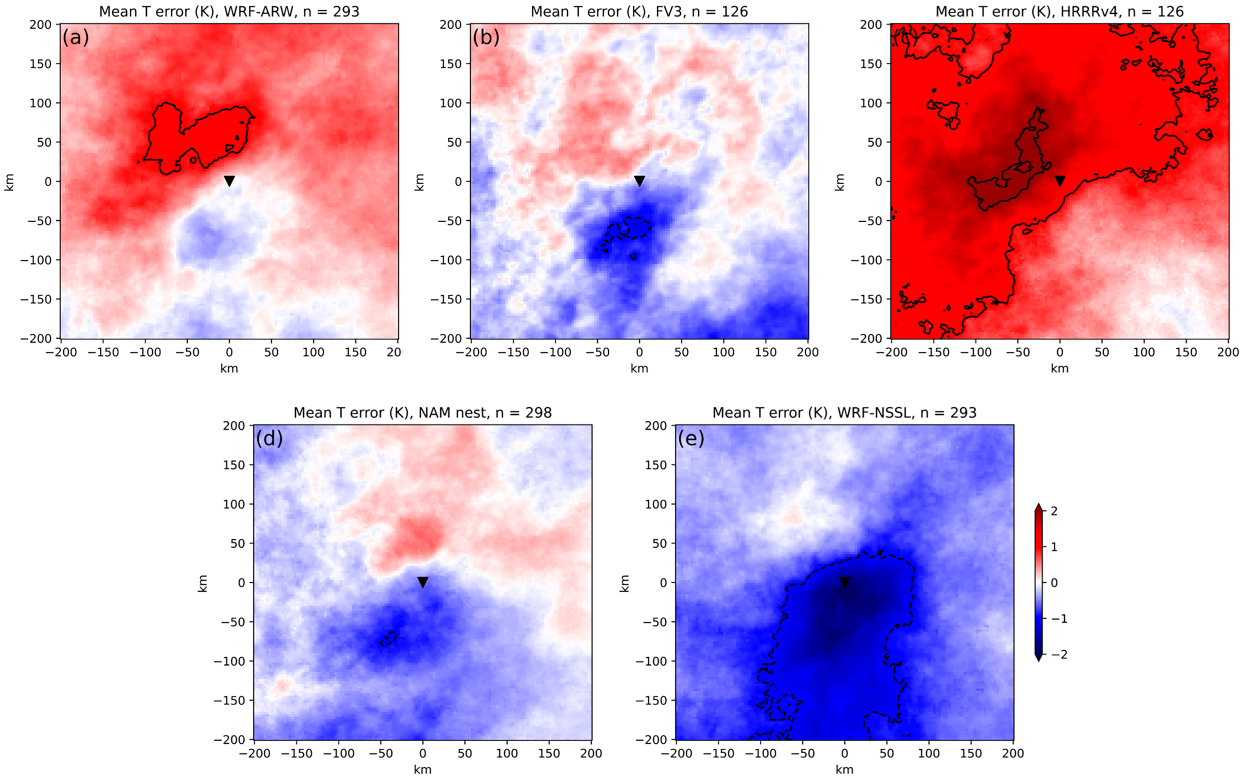

FIG. 5. As in Fig. 3, but for the Great Plains warm season.

like in its handling of near-surface warm-sector thermodynamics. Its distribution of Southeast cool-season temperature errors (Fig. 14) is similar to those of HRRR and WRF-ARW, but its accompanying moist tendency (Fig. 15) is more like HRRR than WRF-ARW. This is a reasonable finding since RRFS and HRRR share the MYNN PBL scheme (Nakanishi and Niino 2006, 2009), among several other parameterizations.

Great Plains temperature and dewpoint errors for the 2022 warm season at times with RRFS available (Figs. 16 and 17) should be viewed with caution. The temperature error distributions for the non-RRFS models differ from those in the previous warm season (Fig. 12) where the sample is much larger, unlimited by intermittent RRFS availability. Nevertheless, RRFS appeared to perform fairly well in warm-sector environments in the 2022 Great Plains warm season, with minimal temperature bias and a dry tendency comparable to the FV3 member. Because of the very small sample, tornado-centered mean errors for RRFS are not shown here for either region/season.

4. Conclusions

Two mostly independent verification approaches agree that the HREF membership suffers from a near-surface cold bias in Southeast cool-season severe weather environments. While many forecasters responsible for issuing forecasts for the region are already anecdotally aware that this bias exists, confirmation of mean/median errors of roughly −1 K (−2 K if using the MYJ scheme during diurnal heating) provides a quantitative baseline for approaching model guidance in these environments. For further context, an idealized low-CAPE thermodynamic profile emulating that of Sherburn and Parker (2019) and containing 470 J kg^{-1} surface-based CAPE yields only 326 J kg^{-1} if the parcel temperature is reduced by 1 K, so that a cold error of 1 K in such a scenario would reduce the forecast CAPE by 30% of the amount observed. By comparison, there is not a clear bias across all HREF members in Great Plains warm-season warm sectors or tornado environments.

After internal discussions with SPC and the NWS Environmental Modeling Center (EMC), the reasons for the Southeast cool-season bias remain unclear. The HREF membership comprises multiple model cores and initial conditions, four PBL schemes, and four microphysics schemes. While the MYJ PBL scheme may exacerbate the problem in two members, none of these physics choices can offer a complete explanation. The HREF archive does not contain enough vertical levels to assess characteristics of modeled cloud layers in comparison with upper-air observations. However, any systematic error in cloud thickness should lead to a diurnal cycle of 2-m temperature bias, not a persistent cold bias at both 1200 and 2100 UTC. Preliminary verification of RRFS suggests that it shares this bias but falls toward the less biased end of the established HREF range in these scenarios. With the current operational HREF as well as with the future RRFS, forecasters should account for the likelihood of cool errors in these regimes.

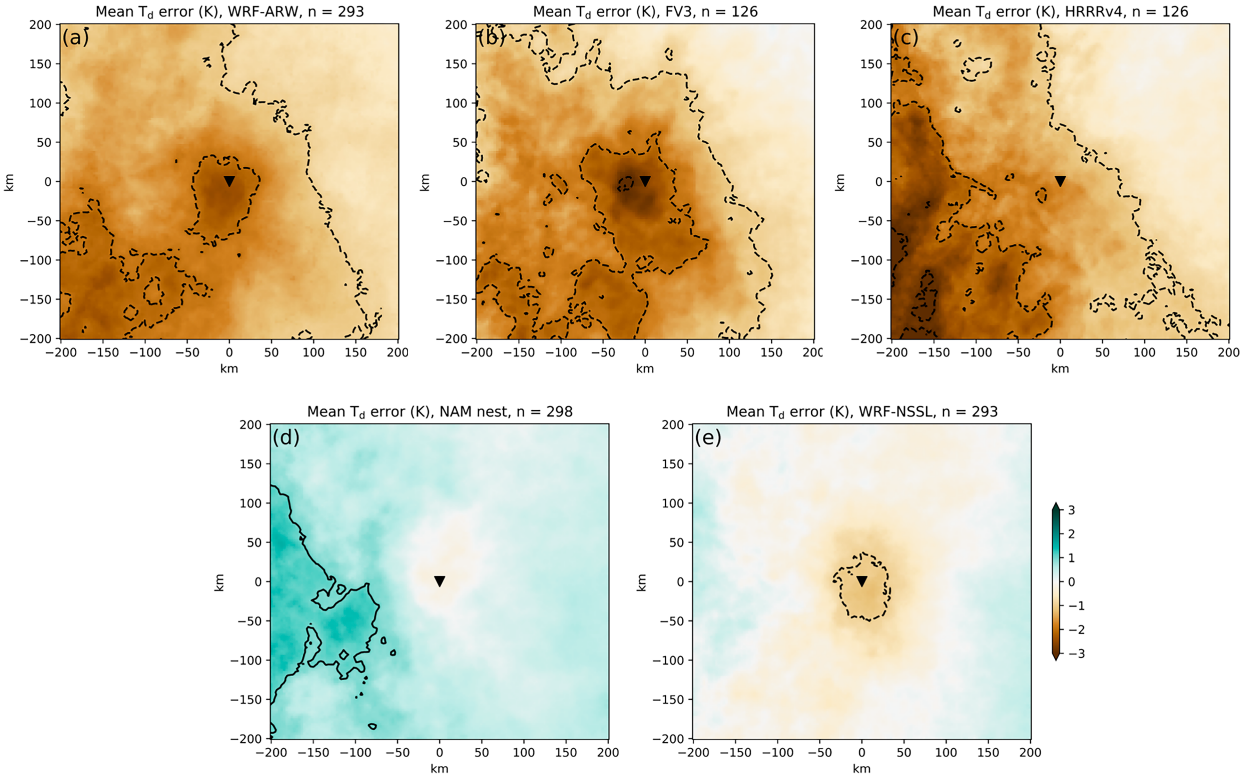

FIG. 6. As in Fig. 3, but for dewpoint temperature in the Great Plains warm season.

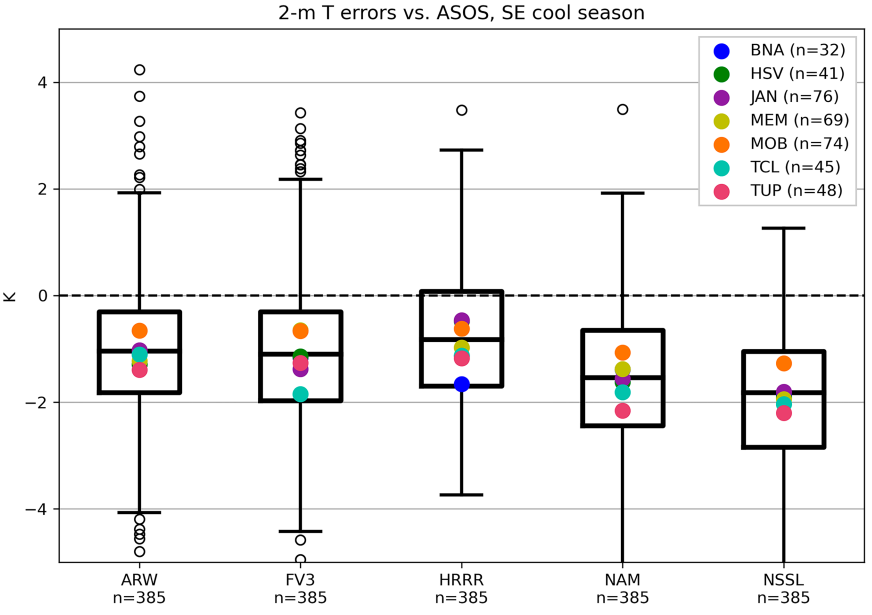

FIG. 7. Boxplots of southeastern U.S. warm-sector 2-m temperature errors (K) for each HREF member vs ASOS observations, via the method described in section 2b, during the 2021/22 cool season. Center lines represent median errors, and whiskers extend to the 10th and 90th percentiles. Colored circles represent median errors at individual ASOS sites as labeled. Members with MYJ PBL scheme are the two rightmost: NAM nest and NSSL-WRF.

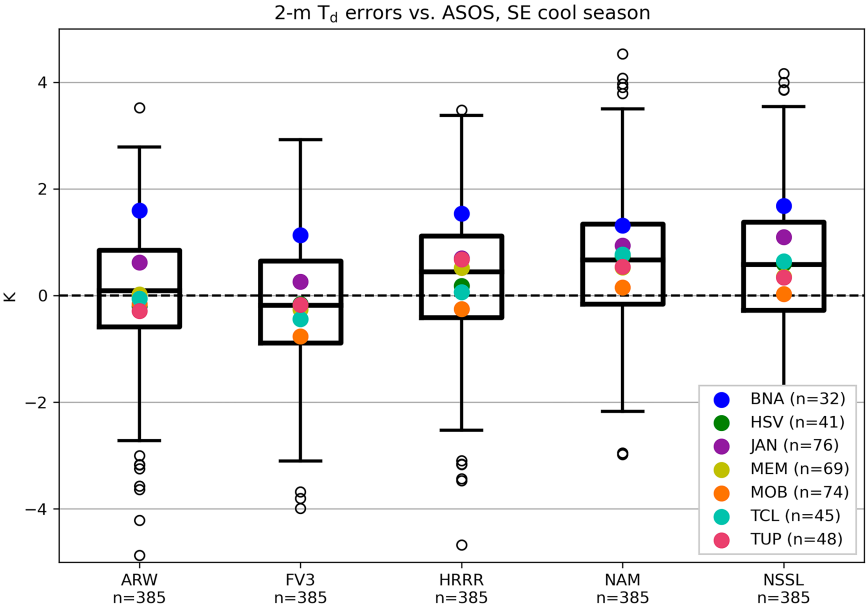

FIG. 8. As in Fig. 7, but for dewpoint temperature.

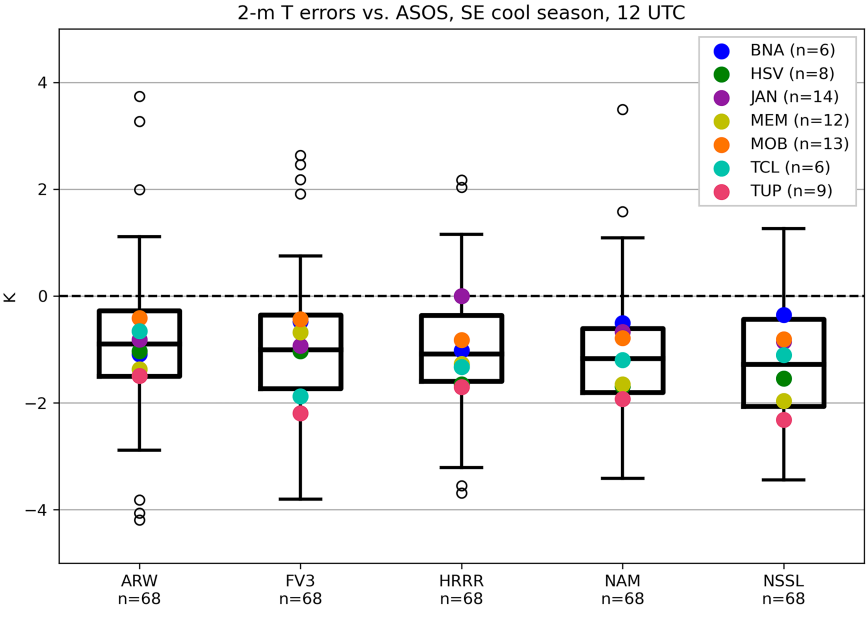

FIG. 9. As in Fig. 7, but using only 1200 UTC observations.

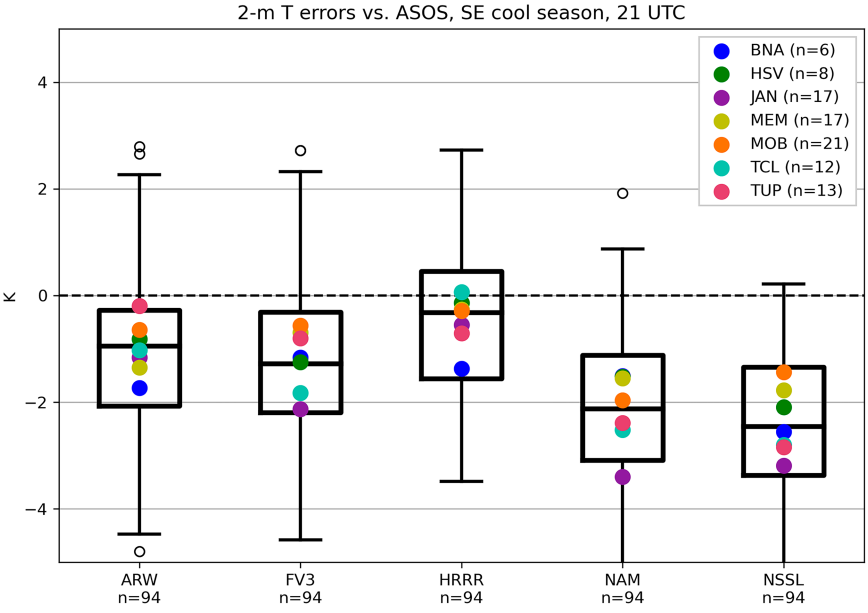

FIG. 10. As in Fig. 7, but using only 2100 UTC observations.

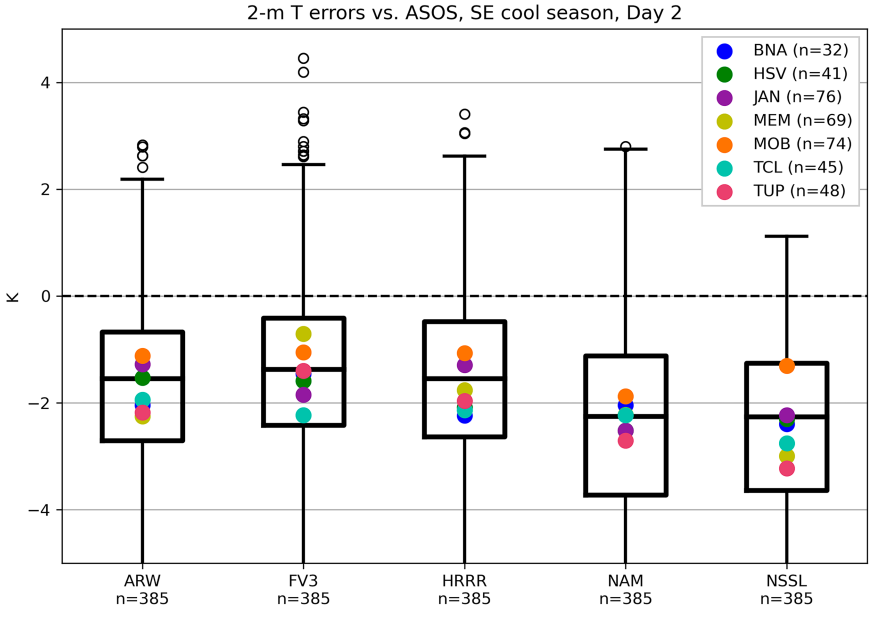

FIG. 11. As in Fig. 7, but for forecast hours 24–36.

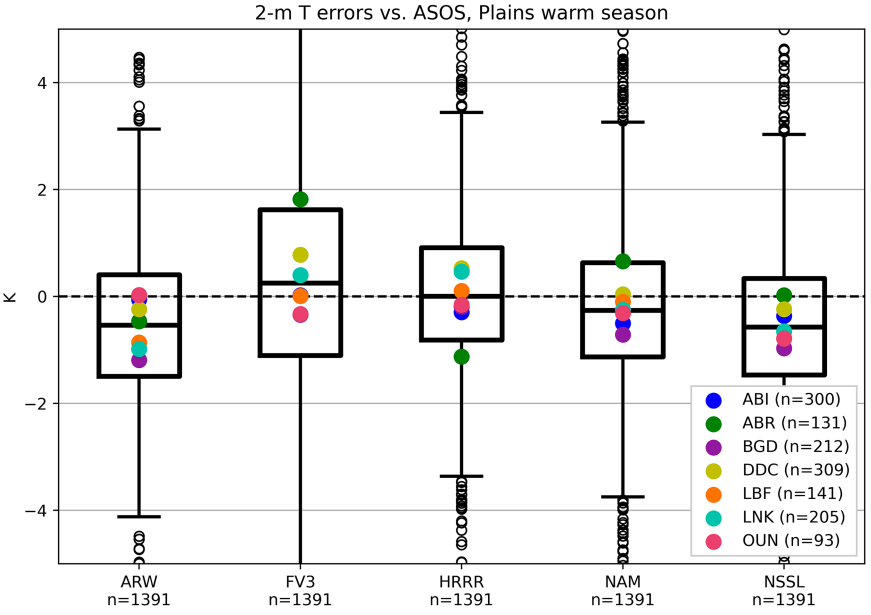

FIG. 12. As in Fig. 7, but for the 2021 Great Plains warm season.

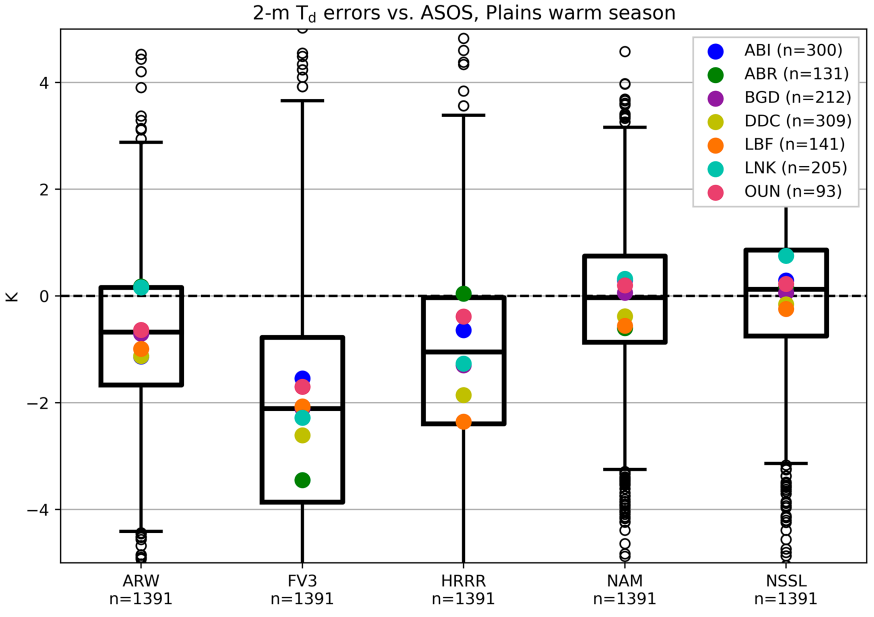

FIG. 13. As in Fig. 7, but for dewpoint temperature in the 2021 Great Plains warm season.

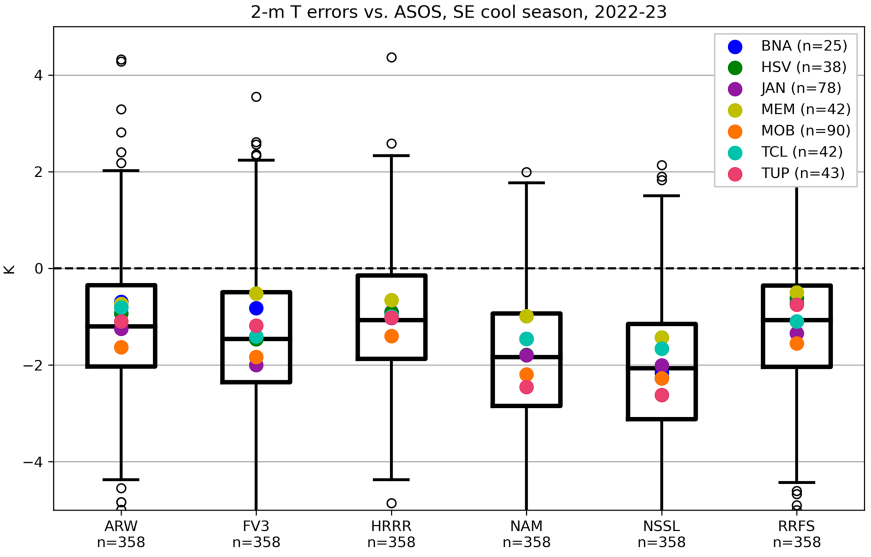

FIG. 14. As in Fig. 7, but for the cool season of 2022/23 at times with RRFS available.

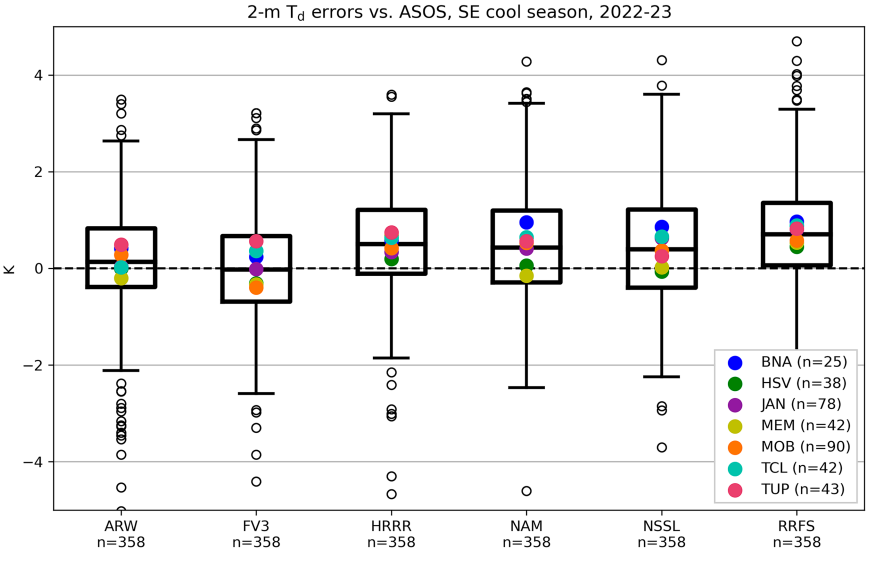

FIG. 15. As in Fig. 7, but for dewpoint temperature in the cool season of 2022/23 at times with RRFS available.

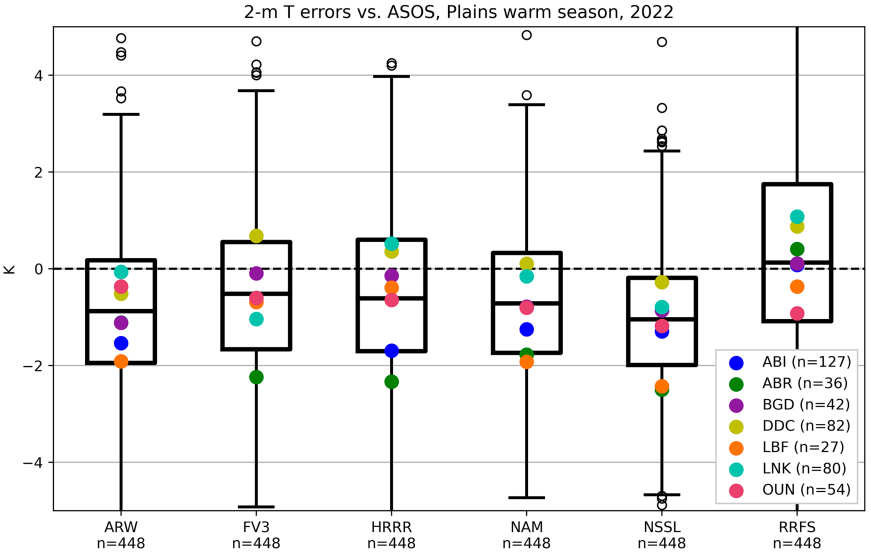

FIG. 16. As in Fig. 7, but for the Great Plains in the warm season of 2022 at times with RRFS available.

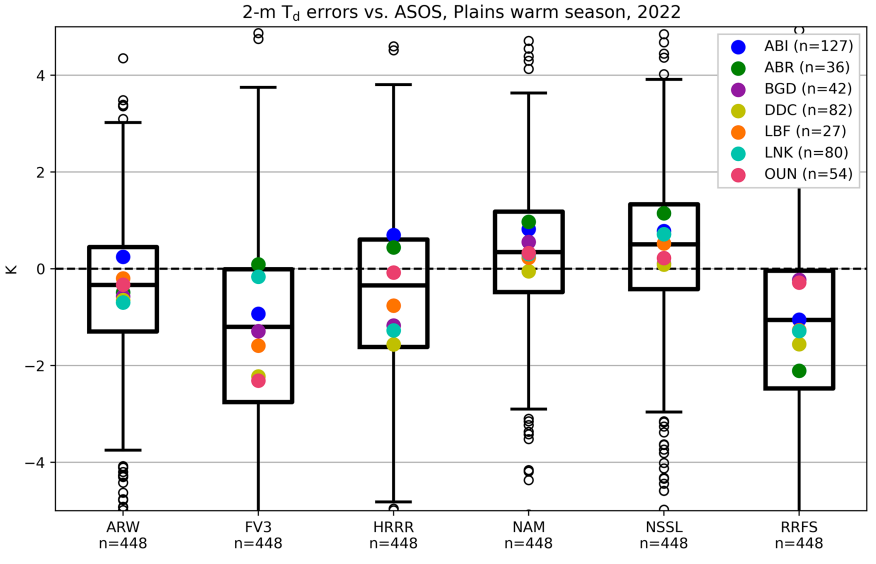

FIG. 17. As in Fig. 7, but for dewpoint temperature in the Great Plains warm season of 2022 at times with RRFS available.

Acknowledgments. Lead author Wade was supported by NOAA–University of Oklahoma Cooperative Agreement NA16OAR4320115. The authors appreciate EMC's collaboration in providing both RRFS data and valuable feedback for evaluations, and the contributions of anonymous reviewers especially to improving the choice of observational data used.

Data availability statement. The HREF and RRFS model data used in this study are archived internally at the National Severe Storms Laboratory (NSSL) and are available upon request. Plotted HREF products are also viewable online (https://www.spc.noaa.gov/exper/href/). NCEI *Storm Data* files used for identifying tornado locations are available online (https://www.ncdc.noaa.gov/stormevents/). ASOS data used for verification in warm sectors are available online (https://mesonet.agron.iastate.edu/request/download.phtml).

REFERENCES

Alexander, C., J. Carley, and M. Pyle, 2023: The Rapid Refresh Forecast System: Looking beyond the first operational version. *28th Conf. on Numerical Weather Prediction*, Madison, WI, Amer. Meteor. Soc., 7.5, https://ams.confex.com/ams/WAFNWPMS/meetingapp.cgi/Paper/425613.

Clark, A., and Coauthors, 2021: Spring Forecasting Experiment 2021: Preliminary findings and results. NOAA Hazardous Weather Testbed, 86 pp., https://hwt.nssl.noaa.gov/sfe/2021/docs/HWT_SFE_2021_Prelim_Findings_FINAL.pdf.

Cohen, A. E., S. M. Cavallo, M. C. Coniglio, H. E. Brooks, and I. L. Jirak, 2017: Evaluation of multiple planetary boundary layer parameterization schemes in southeast U.S. cold season severe thunderstorm environments. *Wea. Forecasting*, **32**, 1857–1884, https://doi.org/10.1175/WAF-D-16-0193.1.

Coniglio, M. C., 2012: Verification of RUC 0–1-h forecasts and SPC mesoscale analyses using VORTEX2 soundings. *Wea. Forecasting*, **27**, 667–683, https://doi.org/10.1175/WAF-D-11-00096.1.

——, J. Correia Jr., P. T. Marsh, and F. Kong, 2013: Verification of convection-allowing WRF model forecasts of the planetary boundary layer using sounding observations. *Wea. Forecasting*, **28**, 842–862, https://doi.org/10.1175/WAF-D-12-00103.1.

De Pondeca, M. S. F. V., and Coauthors, 2011: The real-time mesoscale analysis at NOAA's National Centers for Environmental Prediction: Current status and development. *Wea. Forecasting*, **26**, 593–612, https://doi.org/10.1175/WAF-D-10-05037.1.

Dowell, D. C., and Coauthors, 2022: The High-Resolution Rapid Refresh (HRRR): An hourly updating convection-allowing forecast model. Part I: Motivation and system description. *Wea. Forecasting*, **37**, 1371–1395, https://doi.org/10.1175/WAF-D-21-0151.1.

Hu, X.-M., J. W. Nielsen-Gammon, and F. Zhang, 2010: Evaluation of three planetary boundary layer schemes in the WRF Model. *J. Appl. Meteor. Climatol.*, **49**, 1831–1844, https://doi.org/10.1175/2010JAMC2432.1.

James, E. P., and Coauthors, 2022: The High-Resolution Rapid Refresh (HRRR): An hourly updating convection-allowing forecast model. Part II: Forecast performance. *Wea. Forecasting*, **37**, 1397–1417, https://doi.org/10.1175/WAF-D-21-0130.1.

Janjić, Z. I., 1994: The step-mountain Eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927–945, https://doi.org/10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2.

Kain, J. S., and Coauthors, 2010: Assessing advances in the assimilation of radar data and other mesoscale observations within a collaborative forecasting-research environment. *Wea. Forecasting*, **25**, 1510–1521, https://doi.org/10.1175/2010WAF2222405.1.

Morris, M. T., J. R. Carley, E. Colon, A. Gibbs, M. S. F. V. De Pondeca, and S. Levine, 2020: A quality assessment of the Real-Time Mesoscale Analysis (RTMA) for aviation. *Wea. Forecasting*, **35**, 977–996, https://doi.org/10.1175/WAF-D-19-0201.1.

Nakanishi, M., and H. Niino, 2006: An improved Mellor–Yamada level-3 model: Its numerical stability and application to a regional prediction of advection fog. *Bound.-Layer Meteor.*, **119**, 397–407, https://doi.org/10.1007/s10546-005-9030-8.

——, and ——, 2009: Development of an improved turbulence closure model for the atmospheric boundary layer. *J. Meteor. Soc. Japan*, **87**, 895–912, https://doi.org/10.2151/jmsj.87.895.

Roberts, B., I. L. Jirak, A. J. Clark, S. J. Weiss, and J. S. Kain, 2019: Postprocessing and visualization techniques for convection-allowing ensembles. *Bull. Amer. Meteor. Soc.*, **100**, 1245–1258, https://doi.org/10.1175/BAMS-D-18-0041.1.

——, B. T. Gallo, I. L. Jirak, A. J. Clark, D. C. Dowell, X. Wang, and Y. Wang, 2020: What does a convection-allowing ensemble of opportunity buy us in forecasting thunderstorms? *Wea. Forecasting*, **35**, 2293–2316, https://doi.org/10.1175/WAF-D-20-0069.1.

Sherburn, K. D., and M. D. Parker, 2014: Climatology and ingredients of significant severe convection in high shear, low-CAPE environments. *Wea. Forecasting*, **29**, 854–877, https://doi.org/10.1175/WAF-D-13-00041.1.

——, and ——, 2019: The development of severe vortices within simulated high-shear, low-CAPE convection. *Mon. Wea. Rev.*, **147**, 2189–2216, https://doi.org/10.1175/MWR-D-18-0246.1.

Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., https://doi.org/10.5065/D68S4MVH.