

# An Evaluation of Paired Regional/Convection-Allowing Forecast Vertical Thermodynamic Profiles in Warm-Season, Thunderstorm-Supporting Environments

CLARK EVANS

*Atmospheric Science Program, University of Wisconsin–Milwaukee, Milwaukee, Wisconsin*

STEVEN J. WEISS,<sup>a</sup> ISRAEL L. JIRAK, AND ANDREW R. DEAN

*NOAA/NWS/Storm Prediction Center, Norman, Oklahoma*

DAVID S. NEVIUS<sup>b</sup>

*Atmospheric Science Program, University of Wisconsin–Milwaukee, Milwaukee, Wisconsin*

(Manuscript received 24 July 2018, in final form 25 September 2018)

## ABSTRACT

This study evaluates forecast vertical thermodynamic profiles and derived thermodynamic parameters from two regional/convection-allowing model pairs, the North American Mesoscale Forecast System and the North American Mesoscale Nest model pair and the Rapid Refresh and High Resolution Rapid Refresh model pair, in warm-season, thunderstorm-supporting environments. Differences in bias and mean absolute error between the regional and convection-allowing models in each of the two pairs, while often statistically significant, are practically small for the variables, parameters, and vertical levels considered, such that the smaller-scale variability resolved by convection-allowing models does not degrade their forecast skill. Model biases shared by the regional and convection-allowing models in each pair are documented, particularly the substantial cool and moist biases in the planetary boundary layer arising from the Mellor–Yamada–Janjić planetary boundary layer parameterization used by the North American Mesoscale model and the Nest version as well as the middle-tropospheric moist bias shared by the Rapid Refresh and High Resolution Rapid Refresh models. Bias and mean absolute errors typically have larger magnitudes in the evening, when buoyancy is a significant contributor to turbulent vertical mixing, than in the morning. Vertical thermodynamic profile biases extend over a deep vertical layer in the western United States given strong sensible heating of the underlying surface. The results suggest that convection-allowing models can fulfill the use cases typically and historically met by regional models in operations at forecast entities such as the Storm Prediction Center, a fruitful finding given the proposed elimination of regional models with the Next-Generation Global Prediction System initiative.

## 1. Introduction

Forecasters at the Storm Prediction Center (SPC) and other private/public forecast entities extensively use model-derived vertical profiles of temperature, moisture, and wind to help predict when, where, and whether thunderstorms will develop and, if they do form, their potential severity and hazards (e.g., tornadoes, hail, wind, and heavy rain). At the SPC, output from regional

models, here defined as limited-area models with horizontal grid spacing of 10–40 km, are typically used to derive forecast vertical profiles and provide forecasters with a conceptual understanding of the meso- to synoptic-scale environment in which thunderstorms may form and evolve. Though regional models must use a deep cumulus parameterization to represent thunderstorms, their historically higher resolution and comparative timeliness relative to global models have fostered their extensive use in SPC operations since the 1990s. Global models also suffer from other shortcomings that further limit their forecast utility. For example, vertical profiles available to SPC that are derived from the European Centre for Medium-Range Weather

---

<sup>a</sup> Retired.

<sup>b</sup> Current affiliation: Delta Air Lines, Savannah, Georgia.

---

*Corresponding author:* Dr. Clark Evans, evans36@uwm.edu

Forecasts's global model are limited to mandatory levels only. Further, the National Centers for Environmental Prediction's Global Forecast System ([Environmental Modeling Center 2016](#)) has been known to forecast overly deep turbulent vertical mixing during the warm season in the conterminous United States (e.g., [Weiss et al. 2015](#)).

In recent years, convection-allowing models (CAMs) with sufficiently fine horizontal grid spacing to be able to crudely resolve thunderstorms ( $\Delta x \leq 4$  km) have become operational (e.g., supported 24/7, routinely updated in a timely manner, and widely available to forecasters). CAMs allow for the explicit prediction of storm mode (e.g., linear versus discrete; [Fowle and Roebber 2003](#); [Done et al. 2004](#); [Weisman et al. 2008](#)) and, through the extraction of surrogate storm attributes or use of postprocessing or parameterizations, approximations of severity (e.g., [Kain et al. 2010](#); [Sobash et al. 2011, 2016a,b](#); [Gallo et al. 2016](#); [Adams-Selin and Ziegler 2016](#); [Gagne et al. 2017](#)). In this sense, CAMs provide additional explicit storm information that forecasters can use to refine the conceptual model for the event established by their assessment of the larger-scale and mesoscale environments.

In addition to being able to crudely resolve thunderstorms, CAMs are also able to crudely resolve meso- $\gamma$ - to meso- $\beta$ -scale variability in the preinitiation and near-storm environments. One example of such variability is manifest through horizontal convective rolls in the daytime planetary boundary layer (e.g., [Malkus and Riehl 1964](#); [LeMone 1973](#)). Observations in pre-convective environments indicate that substantial variability in lower-tropospheric vertical thermodynamic profile structure and derived thermodynamic stability parameters exists across the several-kilometer wavelength of a single horizontal convective roll (e.g., [Weckwerth et al. 1996](#)). In the representative example in [Fig. 1](#), CAM-derived vertical thermodynamic profiles located less than 10 km apart differ in 2-m dewpoint temperature by 2.2°C, surface-based convective available potential energy by 920 J kg<sup>-1</sup>, and surface-based convective inhibition by 63 J kg<sup>-1</sup>, all consistent with observations (e.g., [Fig. 9 in Weckwerth et al. 1996](#)). Although this variability exists in nature and is important for convection initiation ([Weckwerth et al. 1999](#); [Weckwerth and Parsons 2006](#); [Arnott et al. 2006](#)) and storm-scale processes (e.g., [Nowotarski et al. 2015](#)), it occurs on scales with limited predictability (e.g., [Lorenz 1969](#); [Madaus and Hakim 2016](#)). Further, point variability inherent to vertical profiles from CAMs may not be representative of the larger-scale environment of interest.

To date, there has been limited formal evaluation of vertical thermodynamic profiles and derived

thermodynamic parameters from convection-allowing and regional models over uniform samples. Studies that evaluated vertical thermodynamic profile, derived thermodynamic profile, and/or related feature forecasts from both regional and convection-allowing models over consistent samples ([Coniglio et al. 2013](#); [Coffer et al. 2013](#); [Clark et al. 2015](#)) generally compared operational regional model forecasts against experimental or quasi-operational CAM forecasts that used different dynamical cores, initializations, and/or physical parameterizations. Other studies that evaluated CAM-forecast vertical thermodynamic profiles and derived thermodynamic parameters (e.g., [Cohen et al. 2015, 2017](#); [Burlingame et al. 2017](#); [Kain et al. 2017](#); [Nevius and Evans 2018](#)) generally did not concurrently evaluate regional model performance. Furthermore, although model performance evaluations are routinely performed by operational modeling centers to evaluate model updates (e.g., [Alexander et al. 2017](#)), these do not always include detailed examination of vertical thermodynamic profiles (particularly outside of the mandatory levels) or derived thermodynamic parameters, nor do they generally concurrently evaluate distinct modeling systems.

This study is motivated by a simple question: can operational CAMs satisfy all thunderstorm-related SPC use cases, namely predicting the larger-scale and mesoscale environments in which thunderstorms may form and, if they form, their mode and severity? The focus of this work lies with the thunderstorm environment: how skillful are operational CAMs compared to their closely related (e.g., using a nearly identical configuration apart from horizontal grid spacing and cumulus parameterization) operational parent regional model counterparts in predicting vertical thermodynamic profiles and derived thermodynamic parameters in known thunderstorm-supporting environments? This research tests the hypothesis that CAMs are equally skillful to regional models with making such predictions. Underlying this hypothesis is an expectation that meso- $\beta$ - to meso- $\gamma$ -scale variability resolved by CAMs will average to zero over a sufficiently large sample and a hope that more firmly entering the modeling terra incognita ([Wyngaard 2004](#)) alone will not degrade model skill.

The remainder of this manuscript is structured as follows. [Section 2](#) documents the methods used in this study, including details about the regional and convection-allowing models evaluated, the observed cases that compose the verification sample, and the verification metrics used. Results, including breakdowns by model pair, forecast lead time, time of day, geographic location, and the observed thermodynamic environment, are presented in [section 3](#). A brief summary of the major

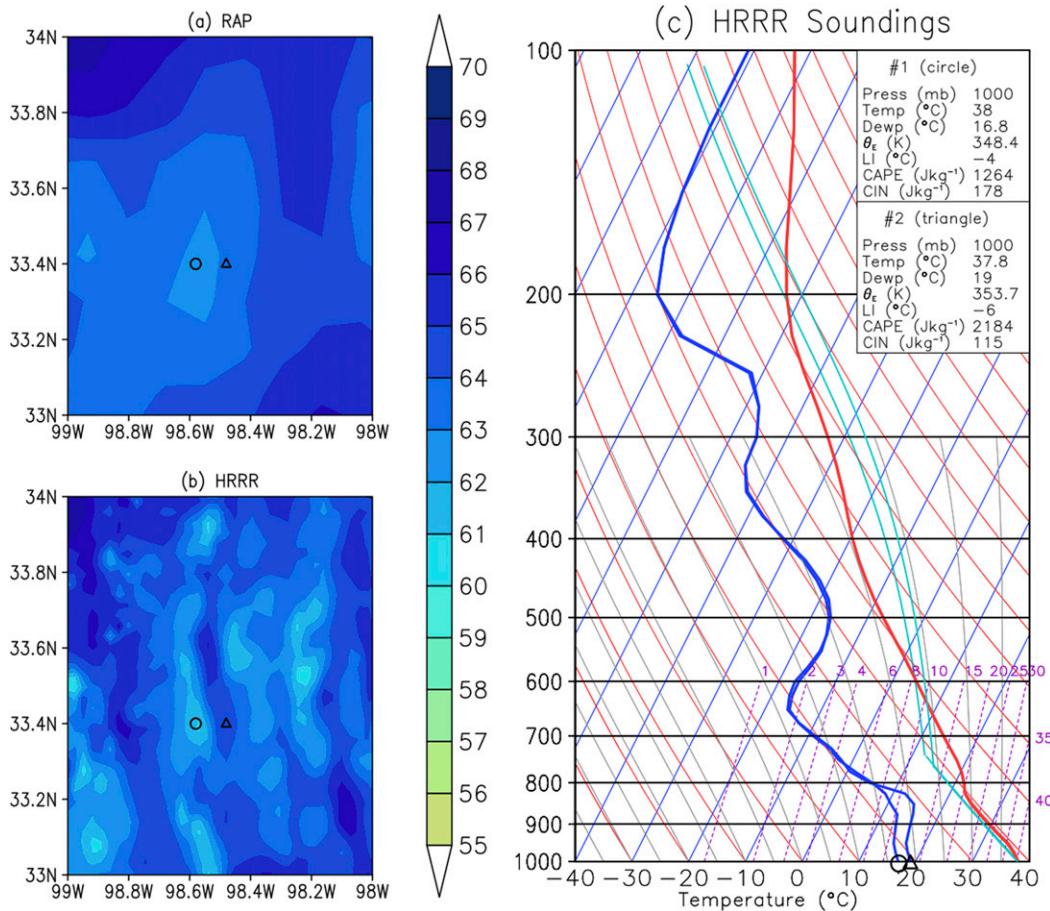


FIG. 1. (a) RAP 3-h model forecast, valid 2100 UTC 31 May 2018, of 2-m dewpoint temperature (°F; shaded) across 33°–34°N and 98°–99°W (north-central Texas). (b) As in (a), but from the HRRR model. The open circle (33.4°N, 98.58°W) and triangle (33.4°N, 98.48°W) depict the locations of the (c) HRRR forecast skew  $T$ – $\log p$  diagrams. Insets in (c) depict selected parameters for an ascending surface-based parcel (light blue curves) at each location, and the forecast temperature and dewpoint temperature vertical profiles are depicted by the solid red and dark blue lines, respectively.

findings and a discussion of their implications for operational forecasters and model developers are presented in section 4.

## 2. Methodology

### a. Model selection and configuration

Vertical thermodynamic profiles and derived thermodynamic parameters are evaluated for two closely related operational regional and convection-allowing model pairs: the North American Mesoscale Forecast System (NAM; e.g., Janjić and Gall 2012; Environmental Modeling Center 2017 and references therein) model and its conterminous U.S. convection-allowing nest (NAM Nest), and the Rapid Refresh (RAP; Benjamin et al. 2016) and High Resolution Rapid Refresh (HRRR; Smith et al. 2008, Benjamin et al. 2016) models.

These selections are made to leverage operational guidance actively used by forecasters while minimizing the differences between the considered models.

Configurations of the NAM and NAM Nest as applicable during the period considered in this study (May 2017; section 2b) are documented by the Environmental Modeling Center (2017) and references therein. Common configuration aspects are summarized in Table 1. The NAM uses a horizontal grid spacing of 12 km with 60 hybrid sigma-pressure vertical levels over a domain covering most of the western Northern Hemisphere, whereas the NAM Nest uses a finer horizontal grid spacing (3 km) over a domain encompassing only the conterminous United States, southern Canada, northern Mexico, and adjacent waters. Deep and shallow cumulus are parameterized in the NAM by the Betts–Miller–Janjić deep and shallow cumulus parameterization

TABLE 1. Common model configuration aspects between the NAM and NAM Nest (center column) and between RAP and HRRR (right column) model pairs.

	NAM and NAM Nest	RAP and HRRR
Dynamical core	Nonhydrostatic multiscale model on the B grid (Janjić and Gall 2012)	Advanced Research version of the Weather Research and Forecasting Model, version 3.6 (Skamarock et al. 2008)
Planetary boundary layer parameterization	MYJ (Janjić 2001)	MYNN (Nakanishi and Niino 2004, 2009; Benjamin et al. 2016)
Microphysics parameterization	Ferrier–Aligo (Aligo et al. 2018)	Aerosol aware (Thompson and Eidhammer 2014)
Shortwave and longwave radiation parameterization	Rapid Radiative Transfer Model (Mlawer et al. 1997; tuned specifically for the NAM)	Rapid Radiative Transfer Model for General Circulation Models (Iacono et al. 2008)
Land surface model	Noah (Ek et al. 2003)	Rapid Update Cycle (Smirnova et al. 2016)

(Janjić 1994), whereas no cumulus parameterization is used by the NAM Nest. Initial conditions for both models are drawn from a partially cycled (hourly over a 6-h period) hybrid ensemble–three-dimensional variational data assimilation system (Wang 2010); however, assimilation is performed on each of the model domains, such that the NAM Nest initialization at a given time is not just a downscaled NAM initialization. NAM lateral boundary conditions are provided from the 6-h old Global Forecast System model forecast and are updated every 3 h; NAM Nest lateral boundary conditions are provided from the current 12-km NAM forecast and are updated every model time step.

Configurations of the RAP (version 3) and HRRR (version 2) applicable during the period considered in this study are documented by Benjamin et al. (2016). Common configuration aspects are summarized in Table 1. The RAP uses a horizontal grid spacing of approximately 13 km with 50 terrain-following sigma vertical levels over a domain that covers most of the western Northern Hemisphere; the HRRR uses a finer horizontal grid spacing (3 km) over a domain encompassing the conterminous United States, far southern Canada, far northern Mexico, and adjacent waters. Deep and shallow cumulus are parameterized in the RAP by the scale-aware Grell and Freitas (2014) parameterization, whereas no cumulus parameterization is used by the HRRR. RAP initial conditions are drawn from a partially cycled (hourly over a 6–18-h period starting at 0300 and 1500 UTC daily) hybrid ensemble–three-dimensional variational data assimilation system (Wang 2010) and digital filter initialization (Peckham et al. 2016). The HRRR is initialized from the 1-h-prior RAP analysis, with the first hour used as a preforecast period intended to assimilate radar reflectivity and selected other observations every 15 min on the 3-km model grid. RAP lateral boundary conditions are obtained from

the most-recent Global Forecast System model forecast, whereas HRRR lateral boundary conditions are provided from the current RAP forecast; both are updated every 3 h.

#### b. Sounding selection and processing

Modeled vertical thermodynamic profiles and derived thermodynamic parameters from the NAM/NAM Nest and RAP/HRRR model pairs are evaluated against routine 0000 and 1200 UTC rawinsonde observations from the conterminous United States during the period 3–31 May 2017. Though rawinsonde observations at other times (e.g., 1800 UTC) exist, the number of these (~50) is too small to allow for robust evaluation of model performance for these observations. Despite some shortcomings (e.g., sensor response times in environments characterized by sharp vertical variations in meteorological fields), rawinsonde observations are believed to be the best-available “truth,” particularly above the surface. Only model analyses and forecasts from the 0000 and 1200 UTC forecast cycles, which include routine rawinsonde observations in their initial conditions (as described in Benjamin et al. 2004, 2016), are verified. Since the focus of this study is on the short range (e.g., in the time frame of a Storm Prediction Center day 1 outlook), this results in verification at three forecast times (0, 11, and 23 h) for the NAM and NAM Nest and two forecast times (0 and 11 h) for the RAP and HRRR. The choice of 11 and 23 h follows Coniglio et al. (2013) and is motivated by the relative proximity of the rawinsonde release time to these times. The period 3–31 May 2017 is selected as representative of early warm-season thermodynamic environments, or those in which buoyancy exerts a substantial influence on vertical turbulent mixing within the daytime planetary boundary layer.

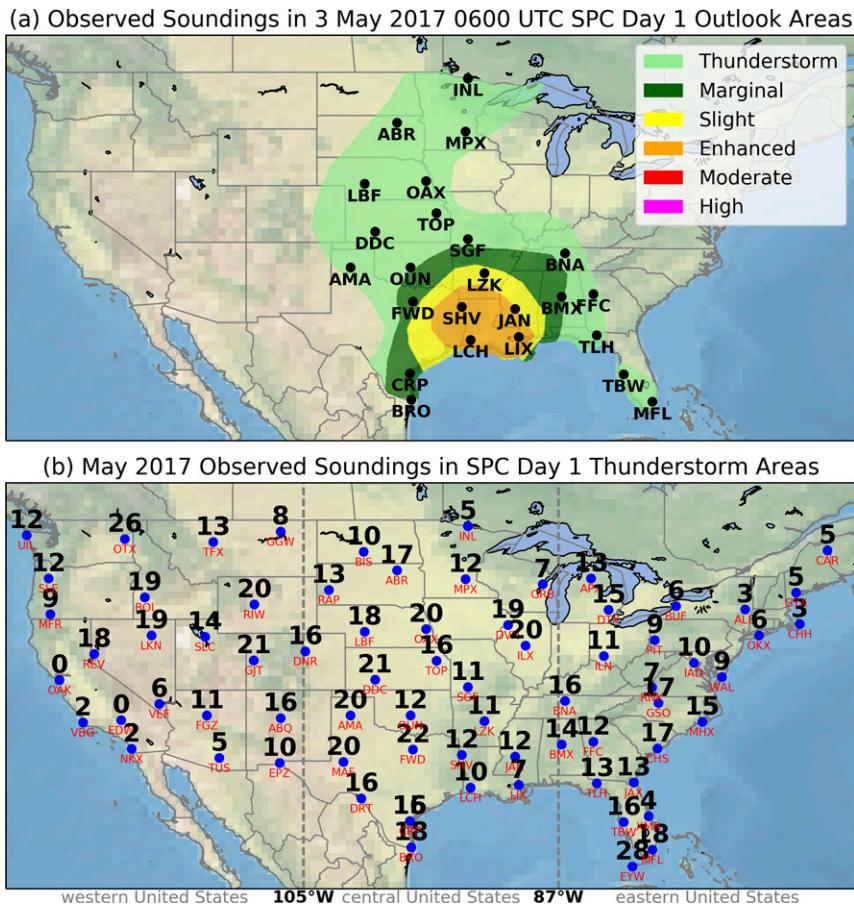


FIG. 2. (a) SPC day 1 convective outlook, issued 0600 UTC 3 May 2017 for the period 1200 UTC 3 May–1200 UTC 4 May 2017. Rawinsonde sites located in any of the shaded categorical risk areas are denoted by black dots with their three-letter site identifier listed beneath the dot. (b) Number of rawinsonde observations by site taken between 1200 UTC 3 May and 0000 UTC 1 Jun 2017, located within a given day’s SPC day 1 convective outlook, and not experiencing  $\geq 1$  mm precipitation within any of the 3 h prior to observation time. The 105°W and 87°W meridians are indicated by dashed gray lines and separate the western, central, and eastern U.S. rawinsonde locations for later data partitioning.

Prior to analysis, the full dataset of all 0000 and 1200 UTC rawinsonde observations from all conterminous U.S. sounding locations is reduced in size (filtered) based on rawinsonde location (or, more precisely, thermodynamic environment) and precipitation occurrence at the location. First, only those observations located in thunderstorm-supporting environments are retained. This is done to focus on one class of environments of importance to forecasters and is the result of an expectation that forecast characteristics will not be identical to those in more stable (e.g., postfrontal and/or strongly subsiding) environments. This classification is based on the Storm Prediction Center day 1 convective outlook “general thunderstorm” area from the 0600 UTC forecast issuance valid for the period from 1200 UTC on that day to 1200 UTC the next day (the

“convective day”). This forecast represents a synthesis of available model guidance, as modified by forecaster assessment and expertise, and is meant to identify a 10% probability of thunderstorms occurring within 12 mi (~19 km) of a location at some point during the convective day. An example of this filtering for the first case in the dataset is depicted in Fig. 2a.

Next, observed and model-derived profiles are excluded based upon whether precipitation was present within the 1–3 h prior to the time of the profile. This is done to remove cases in which precipitation is occurring or has recently occurred, and thus has substantially modified the local vertical thermodynamic profile compared to the preconvective environment, from the evaluation sample. Observed profiles are excluded if there is at least 1 mm of precipitation in any one of the

3 h prior to the observation release time within a 40-km radius of the observation location. This methodology closely follows Coniglio et al. (2013) and results in a sample of 910 observations, the geographic distribution of which is depicted in Fig. 2b. Due to data availability limitations, forecast profiles at 11 and 23 h are excluded if there is *any* precipitation in the 1 h prior to forecast time at the grid point nearest the observation location. An observation is excluded from the evaluation for a given model pair if precipitation is forecast by either model in the pair; this ensures a uniform sample at each forecast time. The resulting sample sizes are 755, 698, and 775 for the 11-h NAM/NAM Nest, 23-h NAM/NAM Nest, and 11-h RAP/HRRR, respectively.

As alluded to earlier in this section, both forecast vertical thermodynamic profiles and derived thermodynamic parameters are evaluated against observations. For the former, vertical profiles of temperature ( $^{\circ}\text{C}$ ) and dewpoint temperature ( $^{\circ}\text{C}$ ) are evaluated. Potential temperature and specific humidity exhibit similar performance characteristics (not shown) and are not included in the present evaluation. The derived thermodynamic parameters considered include convective available potential energy (CAPE), convective inhibition (CIN), lifted condensation level (LCL), and level of free convection (LFC), each for three lifted parcels: surface based, most unstable (defined as the parcel in the lowest 300 hPa above ground level with the highest value of equivalent potential temperature), and mixed layer (defined as the parcel with the average atmospheric properties of the lowest 100 hPa above ground level). As the most-unstable lifted parcel originates from the surface in approximately two-thirds of the most-unstable parcels considered, the most-unstable parameter composites are limited to observed profiles in which the most-unstable CAPE exceeds surface-based CAPE (i.e., cases with elevated instability that is larger than surface-based instability). Further, note that the observed and both forecast profiles within a given model pair and at a given forecast time must have nonzero CAPE (and thus an LFC) for a given parcel to be included in that parameter's evaluation.

Observed and forecast vertical thermodynamic profiles and derived thermodynamic parameters are processed using the National Center Sounding and Hodograph Analysis and Research Program (N-SHARP; Hart and Korotky 1991, with more recently updated information available at <https://www.spc.noaa.gov/exper/soundings/help/index.html>). Because the pressure altitude at which rawinsonde observations originate varies between observation locations from approximately 1000 hPa in the eastern United States to approximately 850 hPa in the Intermountain West, all processed vertical

thermodynamic profiles are subsequently interpolated onto a common grid between 0 and 5 km above ground level with a vertical grid spacing of 100 m. The interpolated profiles are used only for vertical thermodynamic profile evaluation; derived thermodynamic parameters are obtained exclusively from the N-SHARP-processed datasets.

### c. Verification statistics

The primary verification measures used in this study are bias (or mean error; model minus observations) and mean absolute error (MAE; the mean of the absolute error values). For vertical thermodynamic profiles, bias and MAE are calculated at each altitude on the interpolated grid; the root-mean-square error, which penalizes large errors more than does MAE, is also calculated but provides similar qualitative insight to MAE (not shown) and is thus not included. For each variable (and altitude, for the vertical thermodynamic profiles) considered, the two-tailed, nonparametric Wilcoxon signed-rank test (Wilcoxon 1945; Wilks 2011, section 5.3.1) is used to test whether the paired model bias distributions are significantly different from each other to  $\geq 95\%$  confidence.

## 3. Results

### a. Vertical thermodynamic profiles

#### 1) GENERAL OVERVIEW

The full-sample bias and MAE results for the NAM and NAM Nest at 0, 11, and 23 h and the RAP and HRRR at 0 and 11 h are depicted in Figs. 3 and 4, respectively. Temperature bias and MAE are generally small ( $<1^{\circ}\text{C}$ ) at all altitudes and lead times for all models considered. Error distributions are narrow, with similar width between the convection-allowing and regional models in each model pair. By contrast, dewpoint temperature biases are nonzero, with differences in the vertical bias profiles between the model pairs. Dewpoint temperature MAE is increasingly large with both increasing altitude and forecast lead time for all models. Dewpoint temperature error distributions are substantially wider than for temperature, with similar widths between the regional and convection-allowing models in each model pair at most altitudes and lead times. For both fields, bias is consistent in sign, and both bias and MAE grow in magnitude, from model initialization to later forecast times. The comparatively large dewpoint temperature bias and MAE are particularly important to keep in mind as CAPE magnitude (for any considered lifted parcel) is approximately twice as sensitive to

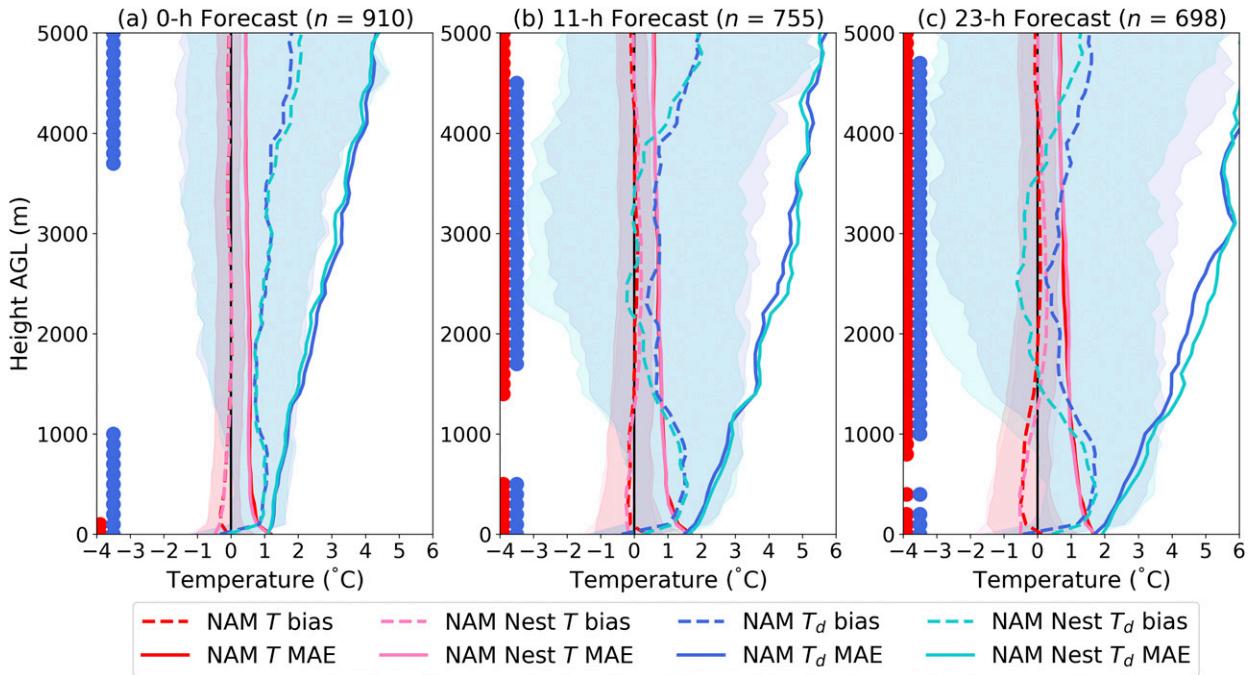


FIG. 3. (a) Vertical profiles of sample-mean bias ( $^{\circ}\text{C}$ ; dashed lines; defined as model minus observations) and MAE ( $^{\circ}\text{C}$ ; solid lines) between 0 and 5 km AGL for NAM and NAM Nest 0-h temperature (red and pink lines, respectively) and dewpoint temperature (dark blue and light blue lines, respectively) analyses. Shading depicts the interquartile ranges of the error distributions for each variable and model. Solid red and blue dots indicate vertical levels at which the temperature and dewpoint temperature error distributions, respectively, between the NAM and NAM Nest are significantly different to at least 95% confidence, as assessed using the two-tailed, nonparametric Wilcoxon signed-rank test. The number of observations contributing to each sample is depicted above each panel. (b) As in (a), but for 11-h forecasts. (c) As in (a), but for 23-h forecasts.

dewpoint temperature variability as to temperature variability.

The preceding paragraph highlights three features meriting further mention: comparatively narrow error distributions for temperature versus dewpoint temperature, an increasingly large dewpoint temperature MAE with height, and maintenance of the general bias and MAE structures (i.e., bias signs and bias/MAE vertical profile shapes) from model initialization through to later forecast times. It is hypothesized that the narrower temperature error distributions result primarily from the smaller assumed rawinsonde temperature observation errors (instrument precision, observation representativeness, and retrieval algorithm uncertainty, where applicable) in each model’s data assimilation system (Hu et al. 2017). Next, the increasingly large dewpoint temperature MAE with height is believed to primarily result from errors in model analyses and forecasts of sharp vertical moisture gradients atop the planetary boundary layer (Fig. 5). Broadly, such errors can be classified as magnitude errors (Fig. 5a) and/or vertical displacement errors (Fig. 5b). Finally, the persistence of the bias and MAE structures from model

initialization to later forecast times is believed to result at least in part from the partial cycling (described in section 2a; in essence, each model uses the previous forecast as its first guess for the updated analysis) used to generate each model’s initial conditions; as a result, like-signed model biases can be present not just in the model forecast but also at the initialization time.

In addition to the overall characteristics summarized above, distinct structures to bias and MAE for temperature and dewpoint temperature exist for the NAM/NAM Nest pair as compared to the RAP/HRRR pair. For example, below 1.5 km AGL, both the NAM and NAM Nest exhibit a large positive (i.e., moist) dewpoint temperature and small negative (i.e., cool) temperature bias at all verification times, the magnitude of which increases with time (Fig. 3). This is likely a result of the well-known afternoon to-evening undermixing (i.e., too shallow depth of turbulent vertical mixing in the planetary boundary layer) bias of the MYJ planetary boundary layer parameterization in warm-season environments (e.g., Hu et al. 2010; Coniglio et al. 2013; Clark et al. 2015). Atop the composite planetary boundary layer, between 1.5 and 4 km AGL, NAM forecasts at

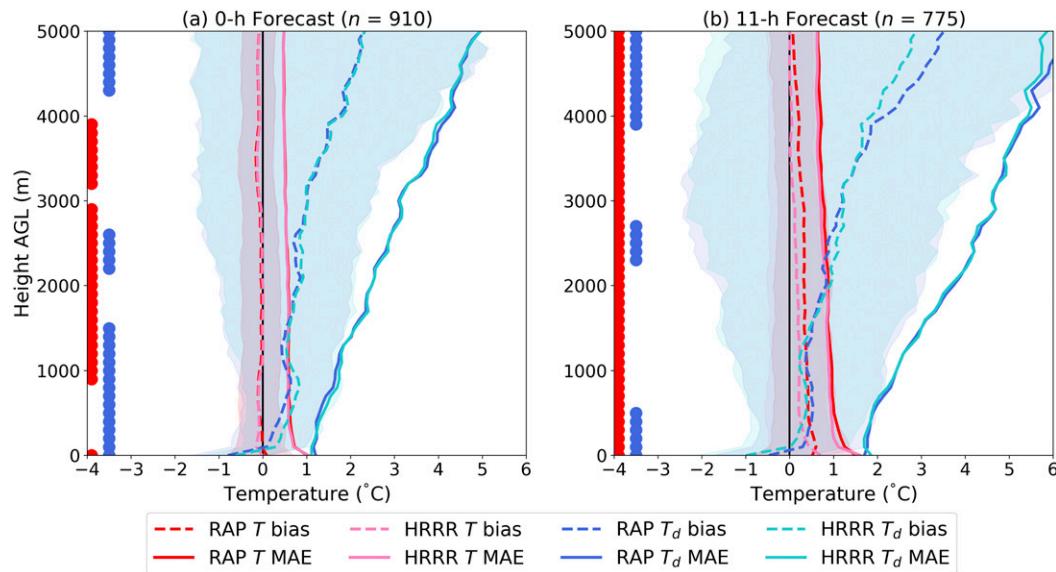


FIG. 4. As in Figs. 3a,b, but for the RAP (temperature in red, dewpoint temperature in dark blue) and HRRR (temperature in pink, dewpoint temperature in light blue) models.

11 and 23 h maintain a small moist bias while the NAM Nest forecasts are near zero to slightly dry biased (Figs. 3b,c). However, the NAM forecasts have a slightly smaller MAE over this layer, as the smaller NAM Nest biases are associated with a wider error distribution on the negative tail.

Likewise, distinct structures to bias and MAE for temperature and dewpoint temperature exist for the RAP/HRRR pair as compared to the NAM/NAM Nest pair. After model initialization, a small warm bias with largest values near the surface develops in the first 11 h in both the RAP and HRRR (Fig. 4). The warm bias is smaller in the HRRR than in the RAP, but the corresponding HRRR MAE is only smaller in the lowest 0.5 km AGL. Both models are also dry biased in the surface layer and slightly moist biased between the surface and 1.5 km AGL (Fig. 4). The near-surface warm bias is not consistent with previous investigations of the performance of the Mellor–Yamada–Nakanishi–Niino (MYNN) planetary boundary layer parameterization in warm-season environments, though the lower-tropospheric moist bias is roughly consistent with the results of such studies (Coniglio et al. 2013; Clark et al. 2015; Burlingame et al. 2017). That said, it should be noted that the MYNN version used by the RAP and HRRR is substantially modified as compared to that used in earlier studies, with the modifications intended to prevent negative turbulent kinetic energy, improve the mixing-length formulation, and couple parameterized shallow cumulus clouds to shortwave and longwave radiation parameterizations (Benjamin et al. 2016, their

appendix B), and the results are generally consistent with the RAP version 3 evaluations of Alexander et al. (2017). Finally, both the RAP and HRRR are associated with large moist biases at and above 3 km AGL at both times, with smaller bias and MAE for the HRRR above 4 km AGL at 11 h (Fig. 4b). As shown later, the large moist biases are found primarily in the western United States.

## 2) VARIATION BY VERIFICATION TIME, LOCATION, AND THERMODYNAMIC ENVIRONMENT

Additional insight regarding model performance can be gained by partitioning the data. In this study, three such partitions are considered: verification time (0000 vs 1200 UTC), geographic region (western, central, and eastern United States; Fig. 2), and thermodynamic environment (e.g., buoyancy magnitude). For simplicity, and given the general consistency in the results at each time considered, the results presented in this section focus on the 11-h forecast time.

When stratifying by verification time, the 0000 and 1200 UTC samples for each model are broadly consistent with their respective full-period samples (cf. Figs. 3, 4, and 6). In general, biases and MAEs are larger at 0000 UTC, when turbulent vertical mixing is driven by both buoyancy and vertical wind shear, than at 1200 UTC, when vertical wind shear is the near-exclusive contributor to turbulent vertical mixing. The undermixing bias of the NAM and NAM Nest forecasts is evident primarily at 0000 UTC, particularly in the NAM

### Albuquerque, NM (ABQ), 11-h forecasts valid 12 May 2017 2300 UTC

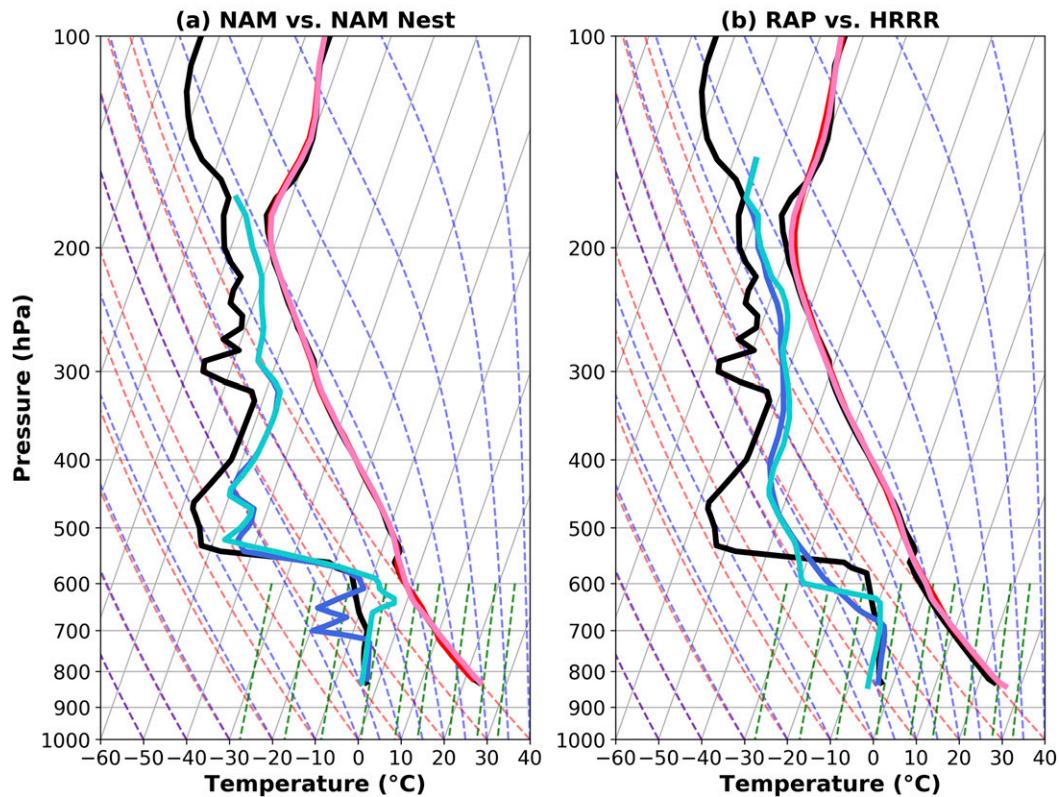


FIG. 5. (a) Skew  $T$ -log $p$  diagram for the observed 0000 UTC 13 May 2017 Albuquerque, NM (ABQ), temperature ( $^{\circ}\text{C}$ ; rightmost solid black line) and dewpoint temperature ( $^{\circ}\text{C}$ ; leftmost solid black line). Overlaid are the 1200 UTC 12 May 2017 NAM (temperature in red, dewpoint temperature in dark blue) and NAM Nest (temperature in pink, dewpoint temperature in light blue) 11-h forecast temperature and dewpoint temperature at the model grid point closest to ABQ. (b) As in (a), but for the RAP (temperature in red, dewpoint temperature in dark blue) and HRRR (temperature in pink, dewpoint temperature in light blue) model pair.

Nest (Fig. 6a), likely the result of the local-closure (i.e., using known quantities only at adjacent vertical grid points to parameterize turbulent vertical mixing) MYJ planetary boundary layer parameterization. It is also at 0000 UTC where the NAM's smaller dewpoint temperature MAE in the 1.5–4 km AGL layer is particularly evident (Fig. 6a). For the RAP and the HRRR, 0000 UTC is also the time at which the slightly smaller dewpoint temperature biases below 1.5 km AGL and above 4 km AGL for the HRRR are found (Fig. 6c). The smaller warm temperature bias in the HRRR relative to the RAP is primarily the result of observations at 1200 UTC (cf. Figs. 6c,d). The physical interpretation of these last three insights is not clear, however.

When stratifying by region, the central and, to lesser extent, eastern United States samples for each model are broadly consistent with their respective full-period samples (cf. Figs. 3, 4, and 7a–d), although some differences from the full-period samples are noted (e.g., a dry

bias between 1 and 2.5 km AGL in the NAM Nest in the eastern United States, smaller moist biases above 3 km AGL in the RAP and HRRR in the eastern and central United States). Substantially different results are obtained for the western United States samples, however (cf. Figs. 3, 4, and 7e,f). For the NAM and NAM Nest, cool and moist biases are particularly large and extend over a much deeper vertical layer in the western United States. Given the observation distribution in this region (Fig. 2), this is speculated to result from deeper, more intense mixing over the elevated terrain of the Intermountain West that is not well forecast by either model. The comparatively large moist bias also contributes to a comparatively large dewpoint temperature MAE below 2 km AGL for both models (Fig. 7e). For the RAP and the HRRR, the near-surface warm bias is largest in the western United States, though the bias remains smaller in the HRRR as compared to the RAP. More significantly, the large moist bias above 3 km AGL

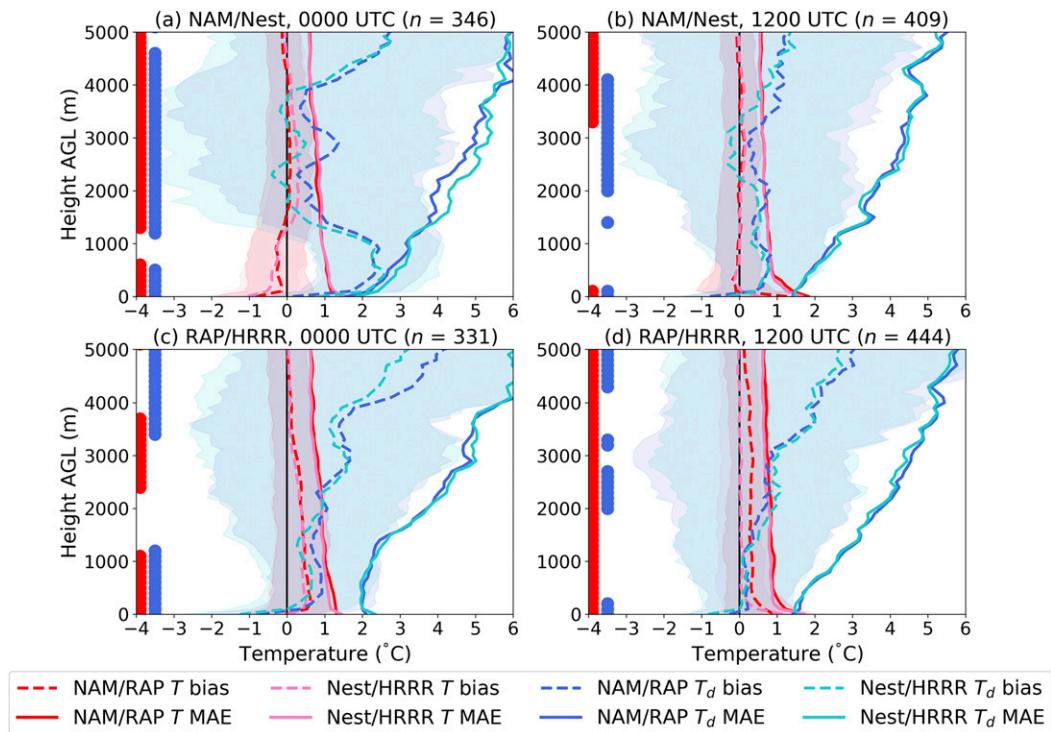


FIG. 6. (a) As in Fig. 3a, but only evaluating 11-h forecast profiles against 0000 UTC rawinsonde observations. (b) As in (a), but for 1200 UTC rawinsonde observations. (c),(d) As in (a) and (b), but for the RAP and HRRR model pair.

in both models is particularly evident in the western United States (Fig. 7f), which is consistent with the systematic model moist bias above 500 hPa (which is 3–5 km AGL in the Rocky Mountains) reported by Alexander et al. (2017).

There are notable differences from the full samples for both model pairs when the data are stratified by thermodynamic environment, based here on the observed mixed-layer CAPE with a threshold value of  $1000 \text{ J kg}^{-1}$  (cf. Figs. 3, 4, and 8). For the NAM and NAM Nest, both models are characterized by undermixing, which is more substantial in the NAM Nest than in the NAM, in the lowest 1–3 km AGL (Fig. 8a). Over this layer, bias magnitude and MAE for both temperature and dewpoint temperature are significantly smaller in the NAM compared to the NAM Nest. For the RAP and HRRR, both models are characterized by dry and warm biases between 0.5 and 2.5 km AGL but near-zero biases below (Fig. 8b). Dewpoint temperature biases above 3 km AGL are small in both models, likely because few of the observations in this large-CAPE composite are in the West. Unlike the NAM and NAM Nest, however, RAP and HRRR model performance is comparable below 2.5 km AGL in the more unstable environments considered in this subsample.

### b. Derived thermodynamic parameters

As compared to the vertical thermodynamic profile evaluation, which provides information about error characteristics for model prognostic variables as a function of height AGL, the derived thermodynamic parameter evaluation provides a vertically integrated (whether explicit or implicit in nature) perspective on model error characteristics for selected buoyancy-related parameters. The full-sample bias and MAE for most-unstable, surface-based, and mixed-layer CAPE, CIN, LCL, and LFC at 11 h (0 and 23 h are similar to those at 11 h and are not shown) for the NAM and NAM Nest are depicted in Fig. 9 and for the RAP and HRRR in Fig. 10. As a reminder, note that most-unstable parcel samples only include those observations where most-unstable CAPE exceeds surface-based CAPE; thus, the most-unstable parcels defined in this section are preferentially characterized by elevated instability.

As with the vertical thermodynamic profiles, differences between the NAM and NAM Nest for the derived thermodynamic parameters considered herein are small (Fig. 9). However, unlike the vertical thermodynamic profiles, most derived thermodynamic parameter differences are not statistically significant at the 95%

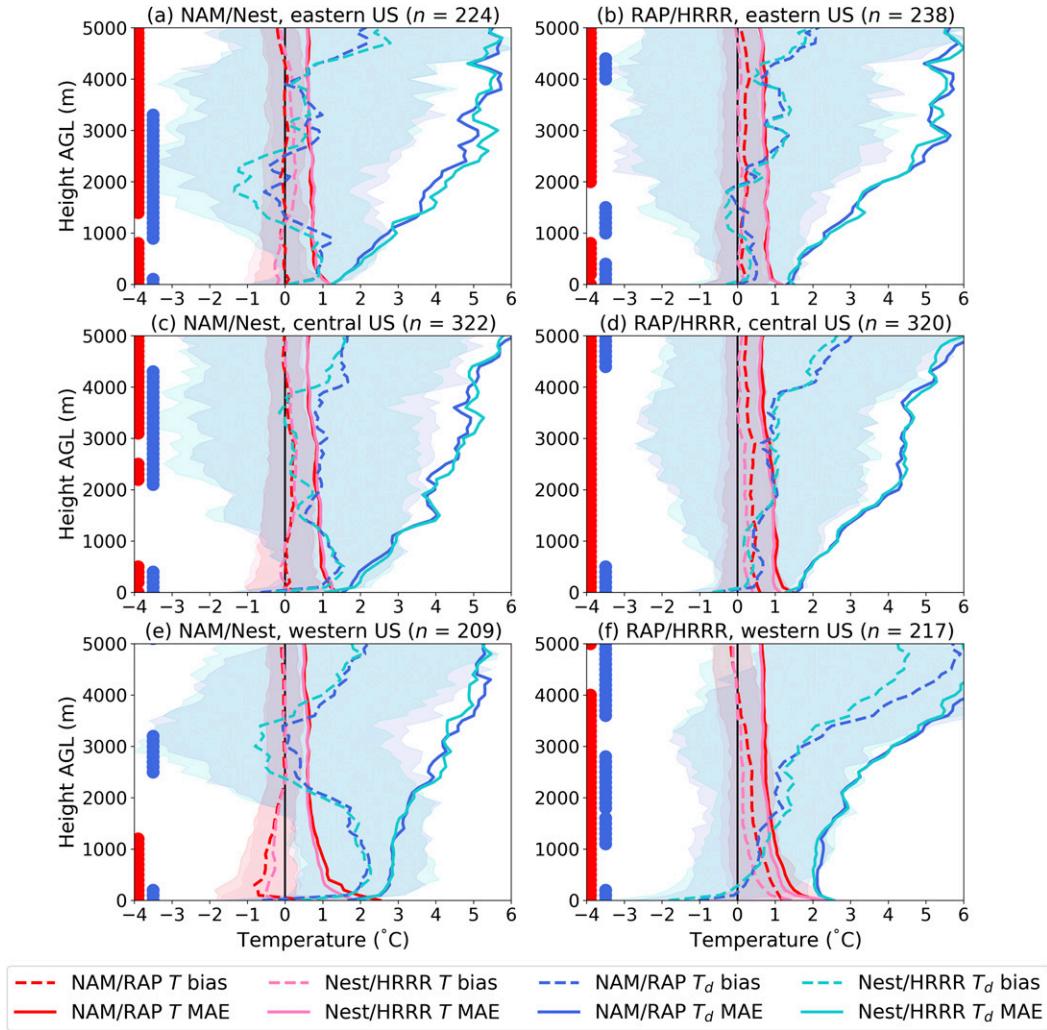


FIG. 7. (a) As in Fig. 3a, but only evaluating 11-h forecast profiles against rawinsonde observations from the eastern United States (as defined in Fig. 2b). (b) As in (a), but for the RAP and HRRR model pair. (c),(d) As in (a) and (b), but for the central United States. (e),(f) As in (a) and (b), but for the western United States.

significance level. Each error distribution is approximately normal, albeit with large variation about the mean and median errors. In their sample means, most-unstable parcels are unstable biased, surface-based parcels are near-zero biased, and mixed-layer parcels are unstable biased, the latter of which likely stems at least in part from a substantial near-surface moist bias; Fig. 3b) characteristic of the MYJ planetary boundary layer parameterization. MAE is generally lowest for surface-based parcels for all parameters, consistent with the near-zero surface temperature and dewpoint temperature mean biases for both models (Fig. 3b).

Likewise, derived thermodynamic parameter differences between the RAP and HRRR are small, albeit with many that are statistically significant at the 95% significance level, particularly for surface-based parcels

(Fig. 10). When considering only those parameters for which statistically significant differences exist, the RAP MAE and bias are smaller for surface-based parcels and the HRRR MAE and bias are smaller for most-unstable and mixed-layer parcels (Fig. 10). However, the *practical* (or operational) significance of these differences is argued to be subjectively low given the small absolute error differences and approximately normal error distributions with similar widths between models. The MAE values, as well as error distribution widths and shapes, are similar for the RAP and HRRR and the NAM and NAM Nest for all parameters and parcels considered (cf. Figs. 9 and 10). On average, both surface-based and most-unstable parcels are slightly stable biased in both models for all variables except CIN, which is believed to result from the models' dry bias at the

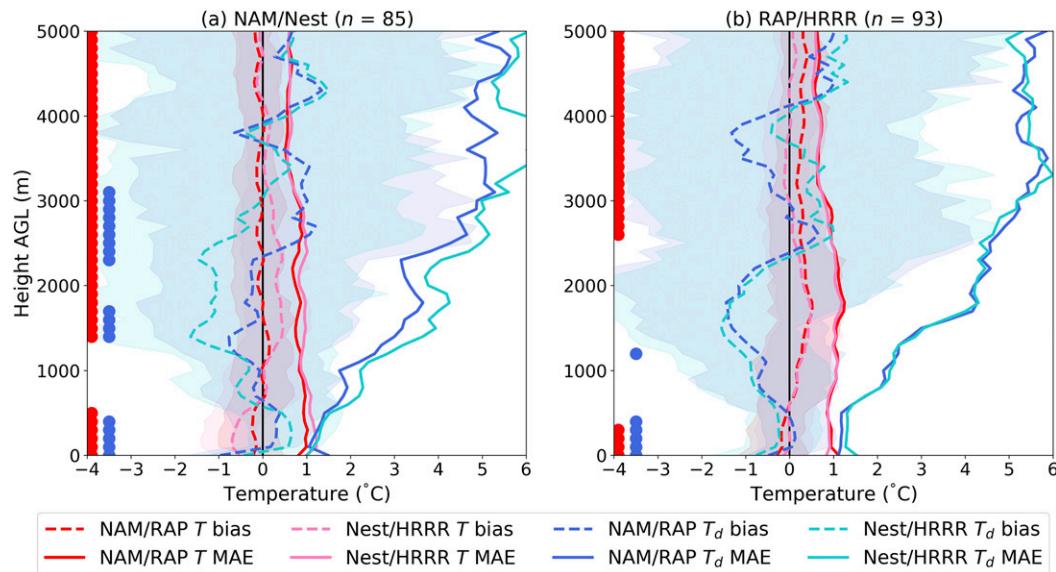


FIG. 8. (a) As in Fig. 3a, but only evaluating 11-h forecast profiles against rawinsonde observations in which at least  $1000 \text{ J kg}^{-1}$  mixed-layer CAPE is present. (b) As in (a), but for the RAP and HRRR model pair.

surface (Fig. 4b). Conversely, mixed-layer parcels are slightly unstable biased for all variables, which is hypothesized to result from the warm- and moist-biased planetary boundary layer for both models (Fig. 4b). In contrast to the NAM and NAM Nest, the mixed-layer CAPE MAE is smaller than the most-unstable and surface-based CAPE MAE, likely due to the smaller moist bias in the planetary boundary layer for the RAP and HRRR (cf. Figs. 3b and 4b).

As with the vertical thermodynamic profiles, additional insight into model performance for derived thermodynamic parameters can be gained by partitioning the data. For derived parameters, two partitions are considered: verification time and geographic region. For simplicity and given the overall importance of CAPE in assessing the convective potential, as well as the general consistency in the results between the parameters considered (not shown), the results presented here focus only on most-unstable, surface-based, and mixed-layer CAPE.

When stratifying by verification time, MAE for the most-unstable and mixed-layer parcels for all models considered is smaller at 1200 UTC than at 0000 UTC (Fig. 11). This is consistent with the vertical thermodynamic profile verification time stratifications, in which temperature and dewpoint temperature MAE values are typically smaller at 1200 UTC than at 0000 UTC for all models (Fig. 6). For the NAM and NAM Nest, the mixed-layer CAPE bias is smaller and the error distribution is narrower at 1200 UTC than at 0000 UTC (Figs. 11a,b), consistent with the smaller near-surface

moist bias in both models in the mechanical-turbulence-driven planetary boundary layer (Fig. 6b). The positive most-unstable CAPE bias in the full sample (Fig. 9a) largely results from observations verifying at 0000 UTC (Fig. 11a), when elevated instability is infrequent and the associated sample is small. For the RAP and HRRR, the most-unstable CAPE error distributions are substantially less normally distributed, with a number of large positive errors, at 0000 UTC as compared to 1200 UTC (Figs. 11c,d), although it should be noted that the most-unstable parcel sample at 0000 UTC is small.

In the geographic region stratification, similar performance characteristics (e.g., mean bias values, error distribution widths and positions, and MAE values) are noted for both the eastern and central United States subcomposites, as seen in the full samples for all models considered (cf. Figs. 9, 10, and 12a–d). The most substantial differences are noted in the western United States, in which surface-based and mixed-layer parcel MAEs are smaller and the respective error distributions are narrower than for the corresponding full samples (cf. Figs. 9, 10, and 12e,f). This may be related to smaller CAPE magnitudes in the more deeply mixed environments of the Intermountain West. For instance, the 90th percentiles of the observed, postprecipitation filtering, mixed-layer CAPE distributions are 2630, 2929, and  $690 \text{ J kg}^{-1}$  for the eastern, central, and western United States, respectively, in the NAM/NAM Nest samples and 2522, 3293, and  $823 \text{ J kg}^{-1}$  in the RAP/HRRR samples. However, note that the sample sizes in the western United States are smaller than for other

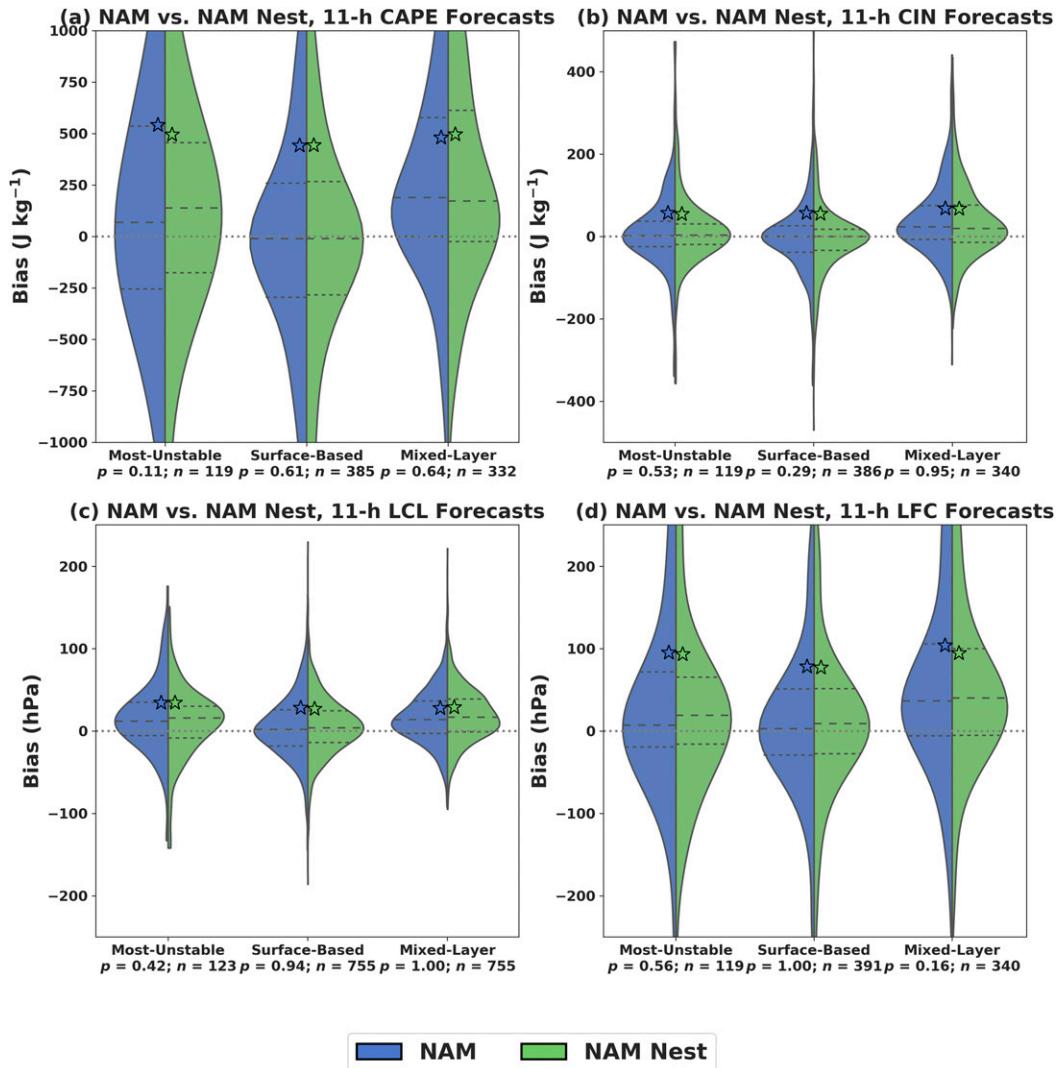


FIG. 9. (a) Kernel density estimates of the bias (here defined as modeled minus observed) in the most-unstable, surface-based, and mixed-layer CAPE ( $J kg^{-1}$ ) for the NAM (blue) and NAM Nest (green) 11-h forecasts. From bottom (smallest) to top (largest), the dashed lines in each distribution represent its 25th, 50th, and 75th percentiles, respectively, and stars indicate the MAE. The number of observations  $n$  in each sample and the  $p$  value from the two-tailed, nonparametric Wilcoxon signed-rank test that the bias distributions are significantly different from each other (at the 95% significance level) are indicated along the  $x$  axis. (b) As in (a), but for CIN. (c) As in (a), but for LCL (hPa). (d) As in (a), but for LFC (hPa). For CAPE, a parcel must have nonzero CAPE to be included; for LFC, a parcel must have a defined LFC to be included. Note that the most-unstable composite for each variable only includes observations where the most-unstable CAPE exceeds the surface-based CAPE (i.e., only parcels with elevated instability).

regions, particularly for the most-unstable parcel samples, and therefore the uncertainty in this inference is somewhat high.

#### 4. Conclusions

Utilizing a sample of observed and model-derived soundings from May 2017 across the conterminous United States, this study tests the hypothesis that CAMs

are equally skillful as their regional model parents in forecasting vertical thermodynamic profiles and derived thermodynamic parameters in early warm-season, thunderstorm-supporting environments at short lead times (0–23 h). In a practical sense, and often also in a statistical sense, CAMs are equally skillful as their closely related regional counterparts in forecasting vertical thermodynamic profile structure and derived thermodynamic parameters in the preconvective environment. In the cases

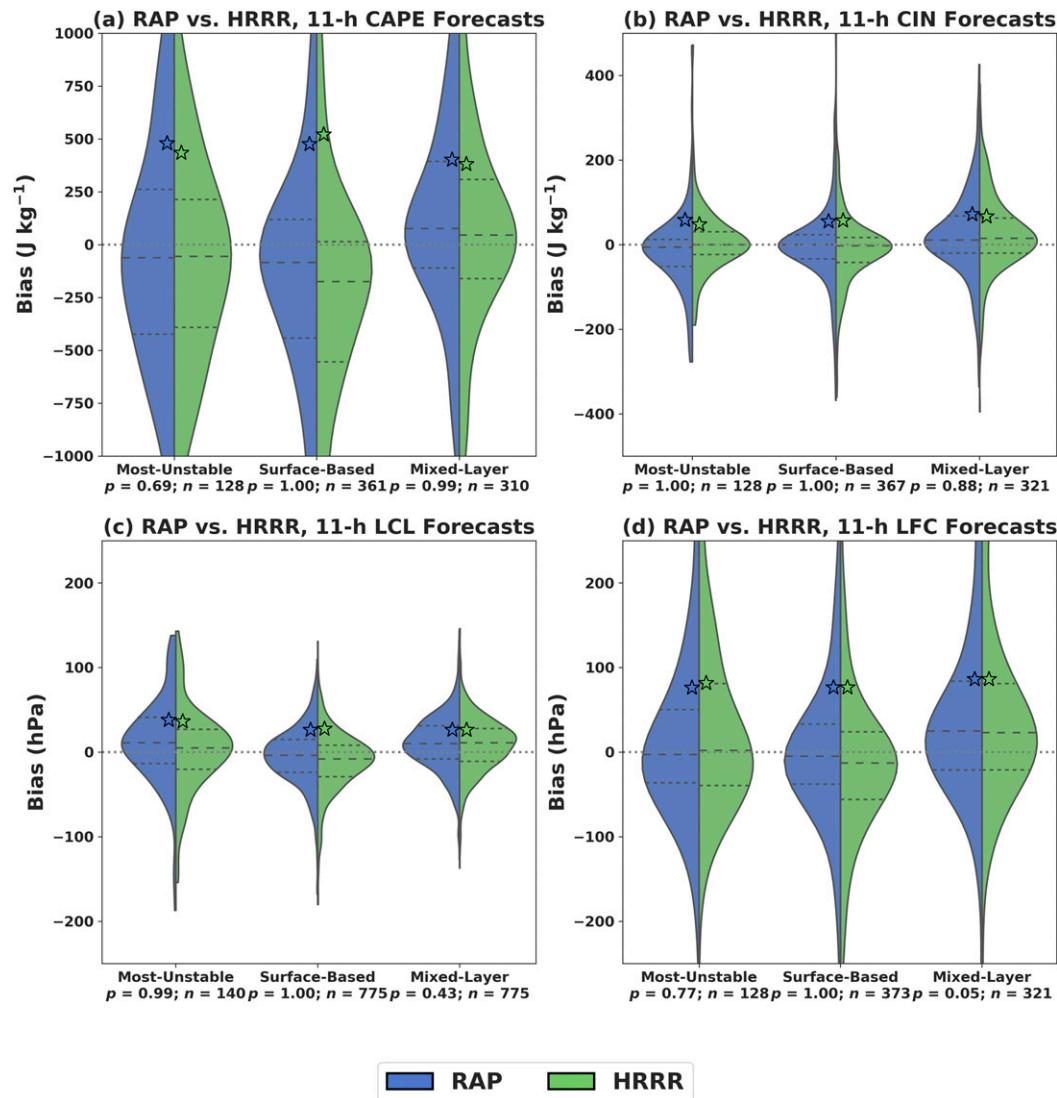


FIG. 10. As in Fig. 9, but for the RAP (blue) and HRRR (green) model pair.

where statistically significant forecast skill differences exist, the NAM is superior to the NAM Nest and the HRRR is superior to the RAP for vertical thermodynamic profiles. For derived thermodynamic parameters, the RAP is superior to the HRRR for surface-based parcels and the HRRR is superior to the RAP for both most-unstable and mixed-layer parcels. However, it should be emphasized that absolute differences between the regional parent and the CAM in a pair are small even when statistically significant, such that the differences are deemed to be negligible from a *practical* or operational perspective (see section 2 in Nicholls 2000 for a discussion of this as it relates to the atmospheric sciences). Further, error distributions are of approximately equal width between the convection-allowing

and regional models in a pair. This implies that the small-scale variability represented by the CAMs is not systematically larger than the regional model forecast errors over each sample. This is perhaps not surprising since the verifying rawinsonde observations also sample local variability (rather than variability on the scales resolved by regional models), although balloon drift with altitude, especially in environments characterized by strong winds aloft, can increase uncertainty about the spatial extent sampled by the rawinsonde. Overall, these findings lend confidence to using the NAM Nest and HRRR to evaluate the forecast meso-scale environment, whether on their native grids or after applying a short-wavelength filter to emphasize the larger-scale environment.

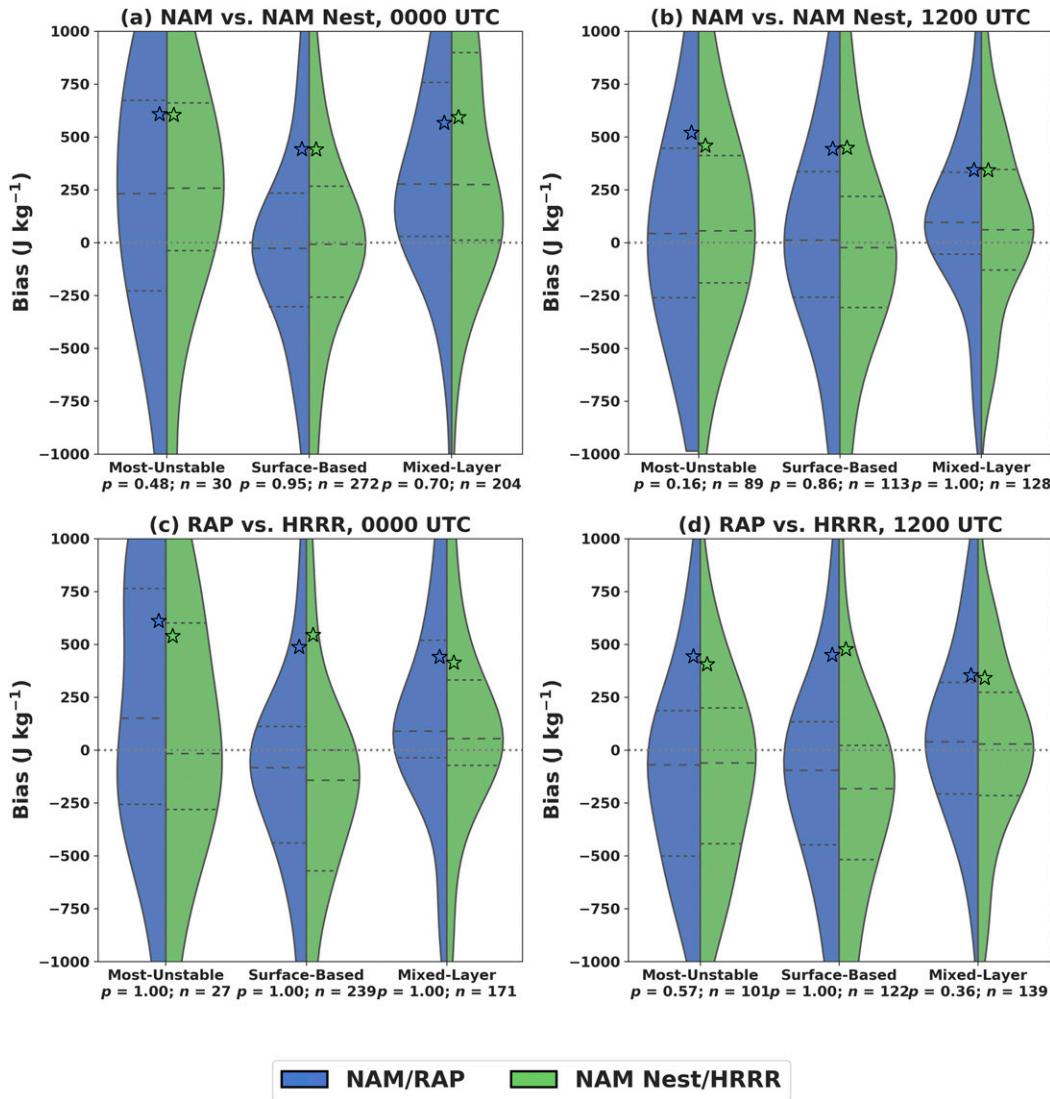


FIG. 11. (a) As in Fig. 9a, but only evaluating against 0000 UTC rawinsonde observations. (b) As in (a), but only evaluating against 1200 UTC rawinsonde observations. (c),(d) As in (a) and (b), but for the RAP and HRRR model pair.

The evaluation presented in this study also provides insight into model bias characteristics in the warm-season, thunderstorm-supporting environments considered herein. For the NAM and NAM Nest, the undermixing bias of the MYJ planetary boundary layer parameterization results in cool- and moist-biased planetary boundary layers and unstable-biased mixed-layer parcels. These biases are particularly evident at 0000 UTC, when buoyancy is a significant contributor to vertical mixing, and in the western United States, where strong sensible heating contributes to stronger and deeper vertical mixing over the typically arid Intermountain West. For the RAP and HRRR, warm and moist biases extend from just above the

surface to at least 5 km AGL, resulting in a small unstable bias for mixed-layer parcels. The moist bias is particularly large above 4 km AGL in the western United States, which is consistent with internal model evaluations. In environments with moderate to large mixed-layer CAPE, both the RAP and HRRR are characterized by a substantial dry bias between 0.75 and 2.5 km AGL, which could influence entrainment within model convection. Most-unstable and surface-based parcels from both models are slightly stable biased, likely due to the dry bias of each model in the surface layer. Together, these insights can be used by forecasters to subjectively bias correct, and thus add value to, model-derived forecasts.

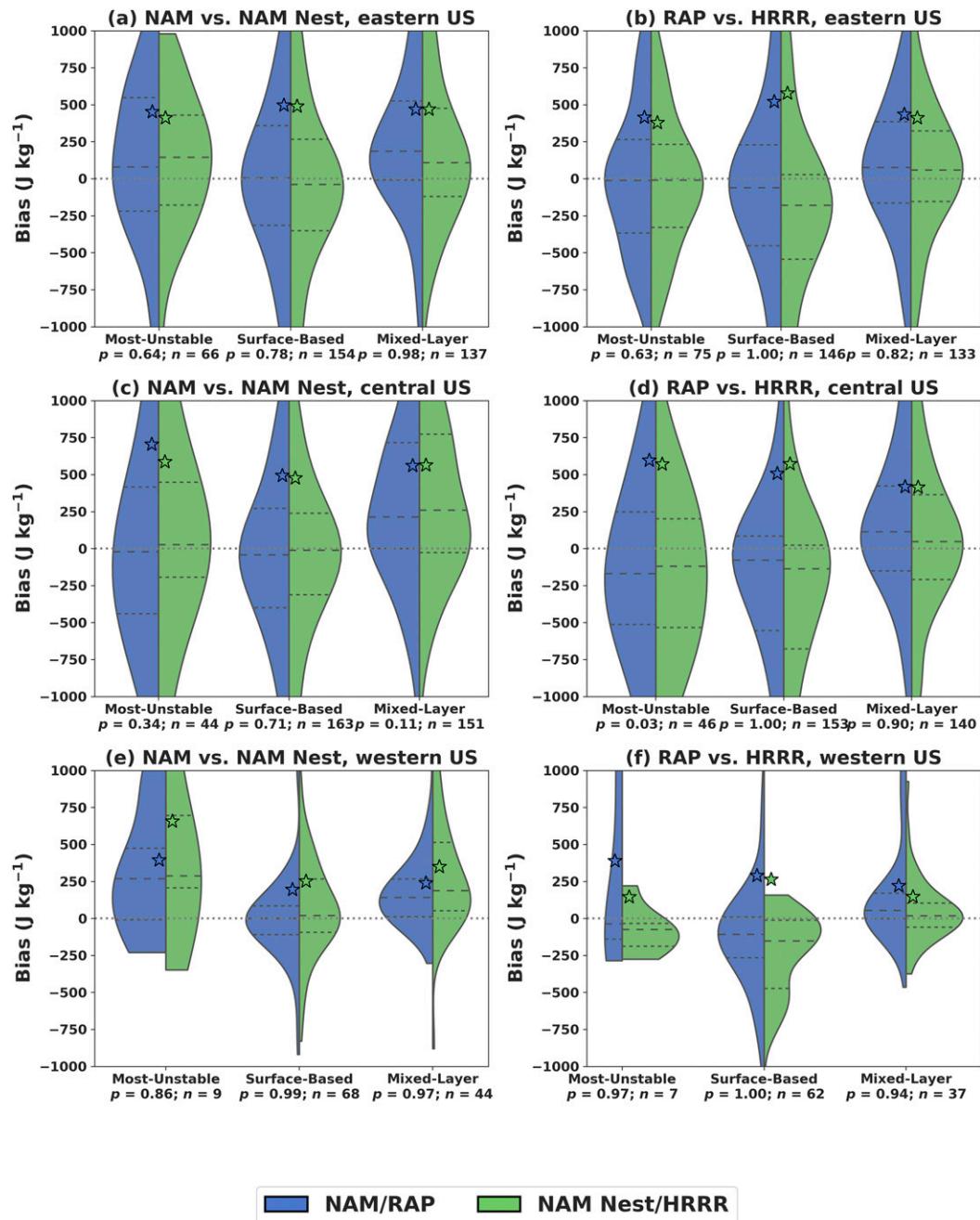


FIG. 12. (a) As in Fig. 9a, but only evaluating against rawinsonde observations from the eastern United States (as defined in Fig. 2b). (b) As in (a), but for the RAP and HRRR model pair. (c),(d) As in (a) and (b), but for the central United States. (e),(f) As in (a) and (b), except for the western United States.

There are several caveats to the results worth noting. Although the NAM and NAM Nest remain identically configured to the versions considered in this study as of the time of this writing, the operational RAP and HRRR versions received significant upgrades in July 2018. As part of these upgrades are changes to the MYNN parameterization, including implementation of

the scale-aware eddy diffusivity-mass flux formulation of Sušelj et al. (2012), and changes to how pseudo-observations in the planetary boundary layer (used to extend the influence of surface observations upward through the planetary boundary layer in well-mixed environments; Benjamin et al. 2016, James and Benjamin 2017) are assimilated. Initial tests indicate that these

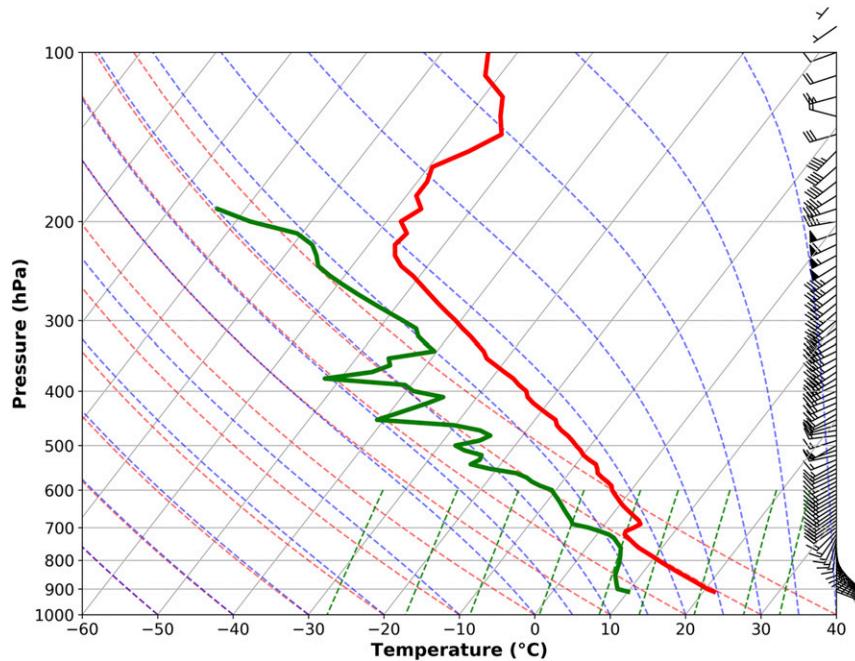


FIG. 13. Skew  $T$ - $\log p$  diagram for the observed 0000 UTC 10 May 2017 North Platte, NE (LBF), temperature ( $^{\circ}\text{C}$ ; solid red), dewpoint temperature ( $^{\circ}\text{C}$ ; solid green), and wind [black barbs; half barb, 5 kt; full barb, 10 kt; pennant, 50 kt ( $1 \text{ kt} \approx 0.51 \text{ m s}^{-1}$ )].

changes result in small but statistically significant reductions in root-mean-square error for vertical thermodynamic environments in warm-season environments (Alexander et al. 2017, 2018), but further examination is necessary to quantify the extent to which the results presented herein may be different in the upgraded modeling system. Second, this study focused on early warm-season, thunderstorm-supporting environments as represented by the May 2017 data sample. Different results may be obtained for other environments of interest, such as those associated with high fire potential, frozen precipitation, or cold-season severe weather events. Related to this, this study only considered one month of data in a single year. Although the models analyzed are continually evolving, such that their characteristics are likely to differ at least slightly from year to year, consideration of a larger data sample is needed to ensure the generality of the results. Third, SPC evaluation of observed rawinsonde profiles has identified an occasional discontinuity in dewpoint temperature between the surface and the overlying atmosphere (e.g., Fig. 13). This could result from a mismatch between the observed surface conditions appended to the lowest rawinsonde level and those sampled by the rawinsonde instrument itself, although this warrants further investigation. Regardless of its cause, this discontinuity deleteriously influences surface-based and mixed-layer derived thermodynamic parameter calculations. It is

uncertain how many observed rawinsonde profiles in the May 2017 sample considered in this study exhibit such a discontinuity, and no attempt is made to account for it in the evaluations presented herein; thus, some caution is urged in interpreting the precise bias and MAE values for each model in the surface layer. Finally, this study only evaluated vertical thermodynamic profiles as well as a small selection of derived parameters. Different insights may be obtained from an evaluation of kinematic fields or specific features [e.g., boundary placement, as in Clark et al. (2015), or organized convective systems, whether the latter are parameterized or explicitly forecast] in the regional and convection-allowing model output.

Over the next five years through the Next Generation Global Prediction System initiative, NCEP plans to unify its modeling system into a single-model framework using the Finite Volume on the Cubed Sphere (FV3; Lin 2004; Putman and Lin 2007; Harris and Lin 2013) that is capable of seamless, skillful prediction from the storm scale to the climate scale (National Weather Service 2017). As compared to the current operational modeling suite, only global and convection-allowing models will likely remain, and regional models such as the NAM and RAP are planned to be phased out. The results of this study suggest that existing CAMs can fulfill regional model roles in SPC operations; however, it is presently unknown whether this will hold for the FV3 dynamical

core. Further, given that the first global FV3 operational model implementation will utilize the eddy diffusivity–mass flux turbulent vertical mixing parameterization (Han et al. 2016) used by the Global Forecast System, which is known to forecast overly deep turbulent vertical mixing during the warm-season over land (Weiss et al. 2015), it is unknown whether the initial operational version of the global FV3 model will be able to fulfill all regional model roles in SPC operations. Additional research is planned to address these questions.

*Acknowledgments.* This research is the result of two week-long visits by the first author to the Storm Prediction Center supported by a UCAR Cooperative Programs for the Advancement of Earth System Science NCEP Research Collaboration Opportunities award. We acknowledge fruitful conversations about this research with Michael Coniglio (NSSL), James Correia Jr. (SPC), Christopher Karstens (SPC), and Patrick Skinner (NSSL). Feedback from three anonymous reviewers helped to improve this manuscript.

#### REFERENCES

- Adams-Selin, R. D., and C. L. Ziegler, 2016: Forecasting hail using a one-dimensional hail growth model within WRF. *Mon. Wea. Rev.*, **144**, 4919–4939, <https://doi.org/10.1175/MWR-D-16-0027.1>.
- Alexander, C., and Coauthors, 2017: WRF-ARW research to operations update: The Rapid-Refresh (RAP) version 4, High-Resolution Rapid Refresh (HRRR) version 3 and convection-allowing ensemble prediction. *18th WRF User's Workshop*, Boulder, CO, UCAR–NCAR, 2.5, [https://ruc.noaa.gov/ruc/ppt\\_pres/Alexander\\_WRFworkshop\\_2017\\_Final.pdf](https://ruc.noaa.gov/ruc/ppt_pres/Alexander_WRFworkshop_2017_Final.pdf).
- , and Coauthors, 2018: The operational future of the Rapid-Refresh (RAP) and High-Resolution Rapid Refresh (HRRR). *Joint WRF/MPAS User's Workshop*, Boulder, CO, UCAR–NCAR, 8.2.
- Aligo, E. A., B. Ferrier, and J. R. Carley, 2018: Modified NAM microphysics for forecasts of deep convective storms. *Mon. Wea. Rev.*, <https://doi.org/10.1175/MWR-D-17-0277.1>, in press.
- Arnott, N. R., Y. P. Richardson, J. M. Wurman, and E. M. Rasmussen, 2006: Relationship between a weakening cold front, misocyclones, and cloud development on 10 June 2002 during IHOP. *Mon. Wea. Rev.*, **134**, 311–335, <https://doi.org/10.1175/MWR3065.1>.
- Benjamin, S. G., and Coauthors, 2004: An hourly assimilation–forecast cycle: The RUC. *Mon. Wea. Rev.*, **132**, 495–518, [https://doi.org/10.1175/1520-0493\(2004\)132<0495:AHACTR>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<0495:AHACTR>2.0.CO;2).
- , and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, <https://doi.org/10.1175/MWR-D-15-0242.1>.
- Burlingame, B. M., C. Evans, and P. J. Roebber, 2017: The influence of PBL parameterization on the practical predictability of convection initiation during the Mesoscale Predictability Experiment (MPEX). *Wea. Forecasting*, **32**, 1161–1183, <https://doi.org/10.1175/WAF-D-16-0174.1>.
- Clark, A. J., M. C. Coniglio, B. E. Coffer, G. Thompson, M. Xue, and F. Kong, 2015: Sensitivity of 24-h forecast dryline position and structure to boundary layer parameterizations in convection-allowing WRF Model simulations. *Wea. Forecasting*, **30**, 613–638, <https://doi.org/10.1175/WAF-D-14-00078.1>.
- Coffer, B. E., L. C. Maudlin, P. G. Veals, and A. J. Clark, 2013: Dryline position errors in experimental convection-allowing NSSL-WRF Model forecasts and the operational NAM. *Wea. Forecasting*, **28**, 746–761, <https://doi.org/10.1175/WAF-D-12-00092.1>.
- Cohen, A. E., S. M. Cavallo, M. C. Coniglio, and H. E. Brooks, 2015: A review of planetary boundary layer parameterization schemes and their sensitivity in simulating southeastern U.S. cold season severe weather events. *Wea. Forecasting*, **30**, 591–612, <https://doi.org/10.1175/WAF-D-14-00105.1>.
- , —, —, —, 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., 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>.
- Done, J., C. A. Davis, and M. L. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the Weather Research and Forecasting (WRF) Model. *Atmos. Sci. Lett.*, **5**, 110–117, <https://doi.org/10.1002/asl.72>.
- Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.*, **108**, 8851, <https://doi.org/10.1029/2002JD003296>.
- Environmental Modeling Center, 2016: The Global Forecast System (GFS)—Global Spectral Model (GSM), version 13.0.2. NWS/EMC, <http://www.emc.ncep.noaa.gov/GFS/doc.php>.
- , 2017: North American Mesoscale (NAM) analysis and forecast system characteristics. NWS/EMC, 9 pp., [http://www.emc.ncep.noaa.gov/mmb/mmbpll/misc/NAM\\_2017.pdf](http://www.emc.ncep.noaa.gov/mmb/mmbpll/misc/NAM_2017.pdf).
- Fowle, M. A., and P. J. Roebber, 2003: Short-range (0–48 h) numerical prediction of convective occurrence, mode, and location. *Wea. Forecasting*, **18**, 782–794, [https://doi.org/10.1175/1520-0434\(2003\)018<0782:SHNPOC>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<0782:SHNPOC>2.0.CO;2).
- Gagne, D. J., II, A. McGovern, S. E. Haupt, R. A. Sobash, J. K. Williams, and M. Xue, 2017: Storm-based probabilistic hail forecasting with machine learning applied to convection-allowing ensembles. *Wea. Forecasting*, **32**, 1819–1840, <https://doi.org/10.1175/WAF-D-17-0010.1>.
- Gallo, B. T., A. J. Clark, and S. R. Dembek, 2016: Forecasting tornadoes using convection-permitting ensembles. *Wea. Forecasting*, **31**, 273–295, <https://doi.org/10.1175/WAF-D-15-0134.1>.
- Grell, G. A., and S. R. Freitas, 2014: A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmos. Chem. Phys.*, **14**, 5233–5250, <https://doi.org/10.5194/acp-14-5233-2014>.
- Han, J., M. L. Witek, J. Teixeira, R. Sun, H.-L. Pan, J. K. Fletcher, and C. S. Bretherton, 2016: Implementation in the NCEP GFS of a hybrid eddy-diffusivity mass-flux (EDMF) boundary layer parameterization with dissipative heating and modified stable boundary layer mixing. *Wea. Forecasting*, **31**, 341–352, <https://doi.org/10.1175/WAF-D-15-0053.1>.
- Harris, L. M., and S.-J. Lin, 2013: A two-way nested global-regional dynamical core on the cubed sphere grid. *Mon. Wea. Rev.*, **141**, 283–306, <https://doi.org/10.1175/MWR-D-11-00201.1>.

- Hart, J. A., and W. Korotky, 1991: The SHARP workstation v1.50 users guide. NOAA/National Weather Service, 30 pp.
- Hu, M., H. Shao, D. Start, K. Newman, C. Zhou, G. Ge, and X. Zhang, 2017: Gridpoint statistical interpolation user's guide, version 3.6. Developmental Testbed Center, 150 pp., [https://dtcenter.org/com-GSI/users/docs/users\\_guide/GSIUserGuide\\_v3.6.pdf](https://dtcenter.org/com-GSI/users/docs/users_guide/GSIUserGuide_v3.6.pdf).
- 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**, 1831–1843, <https://doi.org/10.1175/2010JAMC2432.1>.
- Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13103, <https://doi.org/10.1029/2008JD009944>.
- James, E. P., and S. G. Benjamin, 2017: Observation system experiments with the hourly updating Rapid Refresh model using GSI hybrid ensemble-variational data assimilation. *Mon. Wea. Rev.*, **145**, 2897–2918, <https://doi.org/10.1175/MWR-D-16-0398.1>.
- 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](https://doi.org/10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2).
- , 2001: Nonsingular implementation of the Mellor–Yamada level 2.5 scheme in the NCEP Meso model. NCEP Office Note 437, 61 pp., <http://www.emc.ncep.noaa.gov/officenotes/newernotes/on437.pdf>.
- , and R. L. Gall, 2012: Scientific documentation of the NCEP nonhydrostatic multiscale model on the B grid (NMMB). Part 1 Dynamics. NCAR Tech. Note NCAR/TN-489+STR, 75 pp., <https://doi.org/10.5065/D6WH2MZX>.
- Kain, J. S., S. R. Dembek, S. J. Weiss, J. L. Case, J. J. Levit, and R. A. Sobash, 2010: Extracting unique information from high-resolution forecast models: Monitoring selected fields and phenomena every time step. *Wea. Forecasting*, **25**, 1536–1542, <https://doi.org/10.1175/2010WAF2222430.1>.
- , and Coauthors, 2017: Collaborative efforts between the United States and United Kingdom to advance prediction of high-impact weather. *Bull. Amer. Meteor. Soc.*, **98**, 937–948, <https://doi.org/10.1175/BAMS-D-15-00199.1>.
- LeMone, M. A., 1973: The structure and dynamics of horizontal roll vortices in the planetary boundary layer. *J. Atmos. Sci.*, **30**, 1077–1091, [https://doi.org/10.1175/1520-0469\(1973\)030<1077:TSADOH>2.0.CO;2](https://doi.org/10.1175/1520-0469(1973)030<1077:TSADOH>2.0.CO;2).
- Lin, S.-J., 2004: A “vertically Lagrangian” finite-volume dynamical core for global models. *Mon. Wea. Rev.*, **132**, 2293–2307, [https://doi.org/10.1175/1520-0493\(2004\)132<2293:AVLFDC>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<2293:AVLFDC>2.0.CO;2).
- Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307, <https://doi.org/10.3402/tellusa.v21i3.10086>.
- Madaus, L. E., and G. J. Hakim, 2016: Observable surface anomalies preceding simulated isolated convective initiation. *Mon. Wea. Rev.*, **144**, 2265–2284, <https://doi.org/10.1175/MWR-D-15-0332.1>.
- Malkus, J. S., and H. Riehl, 1964: Cloud structure and distributions over the tropical Pacific Ocean. *Tellus*, **16**, 275–287, <https://doi.org/10.1111/j.2153-3490.1964.tb00167.x>.
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16 663–16 682, <https://doi.org/10.1029/97JD00237>.
- Nakanishi, M., and H. Niino, 2004: An improved Mellor–Yamada level-3 model with condensation physics: Its design and verification. *Bound.-Layer Meteor.*, **112**, 1–31, <https://doi.org/10.1023/B:BOUN.0000020164.04146.98>.
- , 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>.
- National Weather Service, 2017: Strategic Implementation Plan (SIP) for evolution of NNGPS to a national unified modeling system. NWS, 139 pp., <https://www.weather.gov/media/sti/nggps/Presentations%202017/Implementation/SIP-FY18-20-v4.pdf>.
- Neivius, D. S., and C. Evans, 2018: The influence of vertical advection discretization in the WRF-ARW Model on capping inversion representation in warm-season, thunderstorm-supporting environments. *Wea. Forecasting*, **33**, 1639–1660, <https://doi.org/10.1175/WAF-D-18-0103.1>.
- Nicholls, N., 2000: The insignificance of significance testing. *Bull. Amer. Meteor. Soc.*, **81**, 981–986, [https://doi.org/10.1175/1520-0477\(2001\)082<0981:CAATIO>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<0981:CAATIO>2.3.CO;2).
- Nowotarski, C. J., P. M. Markowski, Y. P. Richardson, and G. H. Bryan, 2015: Supercell low-level mesocyclones in simulations with a sheared convective boundary layer. *Mon. Wea. Rev.*, **142**, 3955–3976, <https://doi.org/10.1175/MWR-D-13-00349.1>.
- Peckham, S. E., T. G. Smirnova, S. G. Benjamin, J. M. Brown, and J. S. Kenyon, 2016: Implementation of a digital filter initialization in the WRF Model and its application in the Rapid Refresh. *Mon. Wea. Rev.*, **144**, 99–106, <https://doi.org/10.1175/MWR-D-15-0219.1>.
- Putman, W. M., and S.-J. Lin, 2007: Finite-volume transport on various cubed-sphere grids. *J. Comput. Phys.*, **227**, 55–78, <https://doi.org/10.1016/j.jcp.2007.07.022>.
- Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., <http://dx.doi.org/10.5065/D68S4MVH>.
- Smirnova, T. G., J. M. Brown, S. G. Benjamin, and J. S. Kenyon, 2016: Modifications to the Rapid Update Cycle land surface model (RUC LSM) available in the Weather Research and Forecasting (WRF) Model. *Mon. Wea. Rev.*, **144**, 1851–1865, <https://doi.org/10.1175/MWR-D-15-0198.1>.
- Smith, T. L., S. G. Benjamin, J. M. Brown, S. Weygandt, T. Smirnova, and B. Schwartz, 2008: Convection forecasts from the hourly updated, 3-km High Resolution Rapid Refresh (HRRR) model. *24th Conf. on Severe Local Storms*, Savannah, GA, Amer. Meteor. Soc., 11.1, [https://ams.confex.com/ams/24SLS/techprogram/paper\\_142055.htm](https://ams.confex.com/ams/24SLS/techprogram/paper_142055.htm).
- Sobash, R. A., J. S. Kain, D. R. Bright, A. R. Dean, M. C. Coniglio, and S. J. Weiss, 2011: Probabilistic forecast guidance for severe thunderstorms based on the identification of extreme phenomena in convection-allowing model forecasts. *Wea. Forecasting*, **26**, 714–728, <https://doi.org/10.1175/WAF-D-10-05046.1>.
- , G. S. Romine, C. S. Schwartz, D. J. Gagne II, and M. L. Weisman, 2016a: Explicit forecasts of rotation from convection-allowing models for next-day tornado prediction. *Wea. Forecasting*, **31**, 1591–1614, <https://doi.org/10.1175/WAF-D-16-0073.1>.
- , C. S. Schwartz, G. S. Romine, K. R. Fossell, and M. L. Weisman, 2016b: Severe weather prediction using storm

- surrogates from an ensemble forecasting system. *Wea. Forecasting*, **31**, 255–271, <https://doi.org/10.1175/WAF-D-15-0138.1>.
- Sušelj, K., J. Teixeira, and G. Matheou, 2012: Eddy diffusivity/mass flux and shallow cumulus boundary layer: An updraft PDF multiple mass flux scheme. *J. Atmos. Sci.*, **69**, 1513–1533, <https://doi.org/10.1175/JAS-D-11-090.1>.
- Thompson, G., and T. Eidhammer, 2014: A study of aerosol impacts on clouds and precipitation development in a large winter cyclone. *J. Atmos. Sci.*, **71**, 3636–3658, <https://doi.org/10.1175/JAS-D-13-0305.1>.
- Wang, X., 2010: Incorporating ensemble covariance in the grid-point statistical interpolation variational minimization: A mathematical framework. *Mon. Wea. Rev.*, **138**, 2990–2995, <https://doi.org/10.1175/2010MWR3245.1>.
- Weckwerth, T. M., and D. B. Parsons, 2006: A review of convection initiation and motivation for IHOP\_2002. *Mon. Wea. Rev.*, **134**, 5–22, <https://doi.org/10.1175/MWR3067.1>.
- , J. W. Wilson, and R. M. Wakimoto, 1996: Thermodynamic variability within the convective boundary layer due to horizontal convective rolls. *Mon. Wea. Rev.*, **124**, 769–784, [https://doi.org/10.1175/1520-0493\(1996\)124<0769:TVWTCB>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<0769:TVWTCB>2.0.CO;2).
- , T. W. Horst, and J. W. Wilson, 1999: An observational study of the evolution of horizontal convective rolls. *Mon. Wea. Rev.*, **127**, 2160–2179, [https://doi.org/10.1175/1520-0493\(1999\)127<2160:AOSOTE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<2160:AOSOTE>2.0.CO;2).
- Weisman, M. L., C. Davis, W. Wang, K. W. Manning, and J. B. Klemp, 2008: Experiences with 0–36-h explicit convective forecasts with the WRF-ARW Model. *Wea. Forecasting*, **23**, 407–437, <https://doi.org/10.1175/2007WAF2007005.1>.
- Weiss, S., R. Schneider, I. Jirak, C. Melick, A. Dean, and P. Marsh, 2015: Storm Prediction Center NCEP Production Suite Review, 7 December 2015. *2015 NCEP Production Suite Review*, College Park, MD, NOAA/SPC, [http://www.atmos.albany.edu/daes/atmclasses/atm401/spring\\_2016/ppts\\_pdfs/NCEP-PSR\\_Weiss\\_SPC\\_Dec15.pptx](http://www.atmos.albany.edu/daes/atmclasses/atm401/spring_2016/ppts_pdfs/NCEP-PSR_Weiss_SPC_Dec15.pptx).
- Wilcoxon, F., 1945: Individual comparisons by ranking methods. *Biom. Bull.*, **1**, 80–83, <https://doi.org/10.2307/3001968>.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Elsevier, 676 pp.
- Wyngaard, J. C., 2004: Toward numerical modeling in the “terra incognita.” *J. Atmos. Sci.*, **61**, 1816–1826, [https://doi.org/10.1175/1520-0469\(2004\)061<1816:TNMITT>2.0.CO;2](https://doi.org/10.1175/1520-0469(2004)061<1816:TNMITT>2.0.CO;2).