The Influence of Vertical Advection Discretization in the WRF-ARW Model on Capping Inversion Representation in Warm-Season, Thunderstorm-Supporting Environments

DAVID S. NEVIUS[a] AND CLARK EVANS

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

ABSTRACT

Previous studies have suggested that the Advanced Research version of the Weather Research and Forecasting (WRF-ARW) Model is unable, in its default configuration, to adequately resolve the capping inversions that are commonly found in the warm-season, thunderstorm-supporting environments of the central United States. Since capping inversions typically form in environments of synoptic-scale subsidence, this study tests the hypothesis that this degradation results, in part, from implicit numerical damping of shorter-wavelength features associated with the model-default third-order-accurate vertical advection finite-differencing scheme. To aid in testing this hypothesis, two short-range, deterministic, convection-allowing model forecasts, one using the default third-order-accurate vertical advection finite-differencing scheme and another using a fourth-order-accurate differencing scheme (which lacks implicit damping but is numerically dispersive), are conducted for 25 days during the 2017 NOAA Hazardous Weather Testbed Spring Forecasting Experiment. Model-derived vertical profiles at lead times of 11 and 23 h are validated against available rawinsonde observations released in regions located in the Storm Prediction Center's 0600 UTC day 1 convection outlook's "general thunderstorm" forecast area. The fourth-order-accurate vertical advection finite-differencing scheme is shown to not result in statistically significant improvements to model-forecast capping inversions or, more generally, the vertical thermodynamic profile in the lower troposphere. Instead, the fourth-order-accurate differencing scheme primarily impacts the representation of longer-wavelength features already reasonably well resolved by the model. The analysis does, however, provide quantitative evidence over a large sample that, on average, the WRF-ARW model forecasts capping inversions that are too weak, with negative buoyancy spread out over too deep of a vertical layer, compared to observations.

1. Introduction

Vertical temperature and moisture profiles govern the buoyancy distribution [e.g., convective available potential energy (CAPE) and convective inhibition (CIN) as vertically integrated metrics] and are thus important necessary but insufficient predictors of thunderstorm initiation and severity. Previous studies have shown that the Advanced Research version of the Weather Research and Forecasting (WRF-ARW; Skamarock et al. 2008; Powers et al. 2017) numerical model is limited in its ability to reliably predict vertical profile structure in known thunderstorm-supporting environments (e.g., Burlingame et al. 2017; Cohen et al. 2015, 2017; Coniglio et al. 2013; Jirak et al. 2015; Kain et al. 2017), particularly those associated with

capping or subsidence inversions (e.g., Farrell and Carlson 1989; Lanicci and Warner 1991). Since WRF-ARW is used as the basis of NOAA's current-generation convection-allowing modeling system (e.g., Benjamin et al. 2016), this limitation can impact both deterministic and probabilistic forecasts of convective events, related phenomena such as dryline propagation, and even particulate concentrations and air quality. The governing motivation behind this research is to determine if altering the vertical advection finite-differencing formulation in WRF-ARW improves thermodynamic profiles, with a specific focus on capping and subsidence inversion representation in warm-season, thunderstorm-supporting environments.

Each spring as part of the Hazardous Weather Testbed (HWT), NOAA conducts the Spring Forecasting Experiment (SFE; e.g., Kain et al. 2003; Clark et al. 2012; Gallo et al. 2017), working with the National Weather Service (NWS) and National Severe Storms Laboratory (NSSL) to promote collaboration between research and

---

[a] Current affiliation: Delta Airlines, Savannah, Georgia.

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

operations and evaluate new products for operations. During the 2015 HWT SFE (Gallo et al. 2017), participants were asked the following question: "Compare forecast soundings in regions with elevated mixed layers (EMLs) from the NSSL-WRF [e.g., Coffer et al. 2013] and 2.2 km Operational UM (Unified Model [Walters et al. 2014, Wood et al. 2014]) at sites where observed raob [rawinsonde observation] data is available. With a focus on sounding structure in the PBL [planetary boundary layer] and depiction of any capping inversions, which model has the best forecast sounding?" From the 89 responses, 67% answered the UM was better than the NSSL-WRF, 10% answered that the UM was worse than the NSSL-WRF, and 21% stated they performed about the same (Jirak et al. 2015). Therefore, one of the findings from the 2015 SFE is that strong vertical gradients in temperature and moisture associated with capping inversions are better resolved in the UM compared to the NSSL-WRF (Jirak et al. 2015; Gallo et al. 2017). Qualitatively, similar findings were obtained from a similar evaluation in the 2014 SFE: the NSSL-WRF tended to have a smoothed representation of capping inversions compared to the UM (Kain et al. 2017). This smoothed representation, associated with inversions that are too weak and negative buoyancy that is often spread over too deep of a vertical layer (e.g., Coniglio et al. 2013), influences the vertical negative buoyancy distribution for ascending near-surface-based parcels and thus the vertical displacement required for these near-surface parcels to ascend to their respective LFCs. Previous studies have shown, however, that increasing the WRF-ARW's number of model vertical levels or varying the PBL parameterization are both insufficient to resolve this issue (Coniglio et al. 2013; Kain et al. 2017; Burlingame et al. 2017), with the former likely resulting in part from the coarse vertical resolution of the model initial conditions.

Gridpoint-based models such as WRF-ARW use finite-difference approximations of varying complexity to compute partial derivative terms in the model's governing equations. The accuracy of a given finite-difference approximation depends on the lowest-ordered partial derivative term truncated from the Taylor series expansion used to obtain the approximation; the higher the order of the lowest-order truncated partial derivative term, the more accurate the approximation. Further, finite-difference approximations introduce numerical artifacts into the model solution that must be mitigated in some fashion. Among these numerical artifacts, two that are particularly relevant in this study are *implicit damping* and *numerical dispersion* (e.g., Warner 2011). Implicit damping represents the scale-selective damping of modeled features exclusively as a function of the chosen finite-difference scheme(s), that is, not that associated with an explicit damping or diffusion term. In general, odd-order-accurate finite-difference schemes are associated with implicit damping, with shorter wavelengths dampened to greater extent at each model time step than longer wavelengths. Further, higher-order-accurate finite-difference schemes are more scale selective in this damping. By contrast, numerical dispersion represents the scale-selective departure of a wave's phase speed from its true phase speed. As any atmospheric variable can be represented by the sum of an infinite number of waves, each of different wavelength and amplitude, numerical dispersion can appear to manifest as a change in amplitude of a given *feature* such as a capping inversion, but it does not change the amplitude of the waves that compose that feature. In general, even-order-accurate finite-difference schemes are associated with numerical dispersion, the properties of which vary between the finite-difference schemes used.

The WRF-ARW model-default finite-difference schemes are odd-order accurate for horizontal (fifth order) and vertical (third order) advection (Skamarock et al. 2008), whereas the operational Rapid Refresh (RAP) and High Resolution Rapid Refresh (HRRR) models (Benjamin et al. 2016) are fifth-order accurate for both horizontal and vertical advection. These odd-order formulations include the next-higher even-ordered scheme plus a residual term that acts as an implicit damping operator of the same order (Skamarock et al. 2008). Consequently, these finite-difference schemes implicitly dampen short-wavelength features (e.g., Fig. 1 in Wicker and Skamarock 2002). In contrast, even-order-accurate formulations do not implicitly dampen and are slightly more accurate than the lower odd-order-accurate formulation in one-dimensional advection applications (Wicker and Skamarock 2002). However, the even-order-accurate formulations are numerically dispersive and have stricter numerical stability criteria. As capping inversions typically form in environments of and are maintained by large-scale subsidence (Carlson and Ludlam 1968; Lanicci and Warner 1991), it is hypothesized that the implicit damping of short wavelengths by the default third-order-accurate vertical advection finite-difference scheme in WRF-ARW is the primary contributor to degraded modeled capping inversion representation. Noting that the UM uses a semi-Lagrangian formulation, which does not implicitly dampen for advection, to solve the nonhydrostatic, fully compressible equations of motion (Walters et al. 2014; Wood et al. 2014), it is speculated that improved capping inversion representation in the UM relative to WRF-ARW is due to the former's absence of implicit numerical damping.

In this study, identically configured deterministic numerical simulations (except for the vertical advection finite-difference approximation) are conducted for a representative set of warm-season thunderstorm events to test the hypothesis that degraded WRF-ARW-forecast capping inversions are due to implicit numerical damping associated with the third-order-accurate vertical advection finite-difference approximation. In one simulation for each event, the default third-order-accurate vertical advection formulation is used; in the other, the next-highest-order-accurate (fourth order, which does not implicitly dampen but is numerically dispersive) vertical advection formulation is used. Despite the stricter stability criteria associated with the fourth-order-accurate formulation, no change is made to the model time step to maintain numerical stability. The numerical stability criteria for one-dimensional linear vertical advection in Cartesian, height-based coordinates using the WRF-ARW model-default third-order Runge–Kutta time-differencing scheme are

$$\frac{w\Delta t}{\Delta z} \leq 1.61 \quad \text{and} \quad (1a)$$

$$\frac{w\Delta t}{\Delta z} \leq 1.26, \quad (1b)$$

for the third- and fourth-order-accurate finite-differencing formulations, respectively (Wicker and Skamarock 2002), where w is the vertical velocity (m s^{-1}), Δt is the model time step (s), and Δz is the vertical grid spacing (m). Using the WRF-ARW-recommended time step of $6\Delta x$ (where Δx is the horizontal grid spacing in km, such that $\Delta t = 24$ s; Skamarock et al. 2008), the above stability criteria are satisfied so long as the ratio between the w and Δz remains sufficiently small. For an approximate vertical grid spacing of 500 m, which resembles that of the middle troposphere in the simulations in this study, the maximum-allowable magnitudes of w range from 26.25 m s^{-1} (fourth order) to 33.54 m s^{-1} (third order). These are sufficiently high to ensure the numerical stability of the third- and fourth-order-accurate schemes in this study, and linear instability is not observed in the simulations conducted herein.

The structure of this paper is as follows. Section 2 describes the model configuration, sounding identification, and evaluation used in this work. Section 3 discusses the key results, focusing on vertical temperature and moisture profiles as well as derived thermodynamic variables both located and not located where an observed capping inversion is found. A summary and discussion of the results from and key implications of the research are provided in section 4.
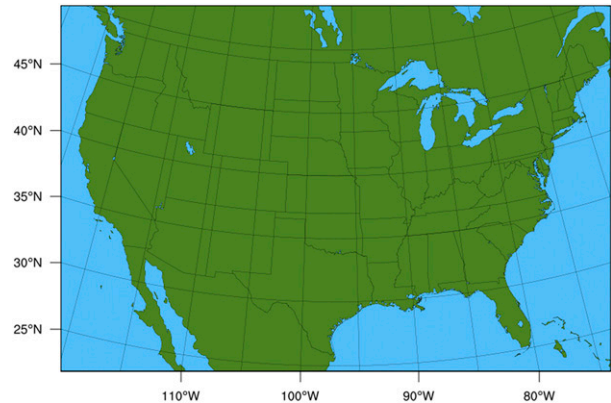

FIG. 1. The domain used for all numerical model simulations, encompassing the conterminous United States.

2. Methods

a. Real-data model simulation description

This study uses WRF-ARW version 3.8.1 to conduct daily (0000 UTC initialization) model simulations between 3 and 31 May 2017, except 15, 17, 19, and 21 May, due to hardware limitations, using third- and fourth-order-accurate finite-difference formulations for vertical advection. The simulation that uses the default third-order-accurate finite-difference formulation is termed the control and the simulation using the fourth-order-accurate finite-difference formulation is termed the fourth-order simulation.

Apart from the vertical advection finite-difference formulation, the NSSL-WRF configuration (Coffer et al. 2013) of WRF-ARW is used for both simulations as it provides a benchmark, well-tested configuration (e.g., Weiss et al. 2007; Kain et al. 2010; Clark et al. 2012; Gallo et al. 2017) for convection-allowing model forecasts. This configuration uses the Mellor–Yamada–Janjić (MYJ) PBL (Janjić 1994), WRF single-moment 6-class microphysics (Hong and Lim 2006), Rapid Radiative Transfer Model (RRTM) longwave radiation (Mlawer et al. 1997), Dudhia shortwave radiation (Dudhia 1989), and Noah land surface (Chen and Dudhia 2001; Tewari et al. 2004) parameterizations. In addition, the model uses positive-definite moisture advection (Skamarock and Weisman 2009), $\Delta x = 4$ km over a 1200×800 conterminous United States domain (Fig. 1), 35 vertical levels (with 10 below the $\sigma = 0.8$ terrain-following coordinate surface, approximately corresponding to 800 hPa over flat terrain at sea level), a time step of 24 s, and a forecast length of 36 h. Explicit diffusion is applied on model coordinate surfaces only in the horizontal; the PBL parameterization handles all vertical diffusion. The WRF-ARW Model also offers an additional sixth-order explicit numerical diffusion formulation
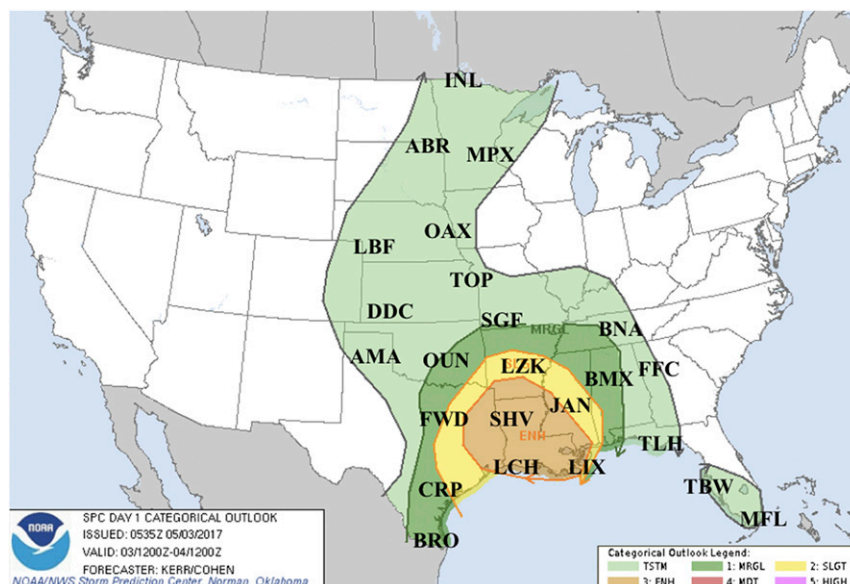
FIG. 2. SPC day 1 convective outlook issued at 0535 UTC 3 May 2017 for the period 1200 UTC 3 May–1200 UTC 4 May 2017 (shaded per the legend at bottom right: TSTM, thunderstorm; MRGL, marginal; SLGT, slight; ENH, enhanced; MDT, moderate; and HIGH, high-risk areas), and the rawinsonde locations that are verified for the 11- and 23-h lead times for this case.

(Knievel et al. 2007), but this is not used in this study. Initial and lateral boundary conditions for all simulations are obtained from 40-km North American Mesoscale Forecast System (NAM; Janjić and Gall 2012; NCEP 2018) model data.

b. Vertical profile identification

In verifying model-derived vertical thermodynamic profiles, routine 0000 and 1200 UTC rawinsonde observations, which are typically launched approximately 1 h before observation time, are considered as the best-available "truth." Model forecasts from 11 h are verified against 1200 UTC rawinsonde profiles, whereas model forecasts from 23 h are verified against 0000 UTC rawinsonde profiles; these represent the forecast hours closest to the times that the PBL is typically sampled by the rawinsondes. At 1200 UTC, nocturnal features such as the low-level jet, a residual layer, and a radiation inversion rooted at the surface are often present, with vertical wind shear as the primary driver of turbulent vertical mixing. At 0000 UTC, buoyancy and mechanically driven turbulent vertical mixing and their effects on PBL thermodynamic properties are both present.

All observed and model vertical profiles are linearly interpolated onto a common grid above ground level with an interval of 100 m. This approximately matches the model vertical grid spacing near the surface. Given the demonstrated minimal sensitivity to vertical grid spacing in previous studies (Kain et al. 2017; Burlingame

et al. 2017), this choice is not expected to significantly impact the results presented herein. It is possible that an advanced interpolation technique such as cubic-spline
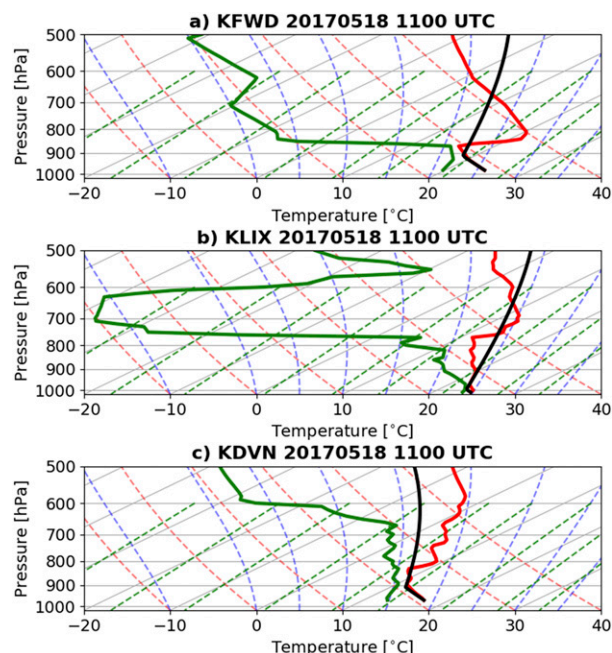

FIG. 3. Observed (a) KFWD, (b) KLIX, and (c) KDVN skew T–log$p$ diagrams [red line, temperature (°C); green line, dewpoint temperature (°C); black line, parcel ascent curve for a surface-based parcel] valid at the approximate release time of 1100 UTC 18 May 2017.
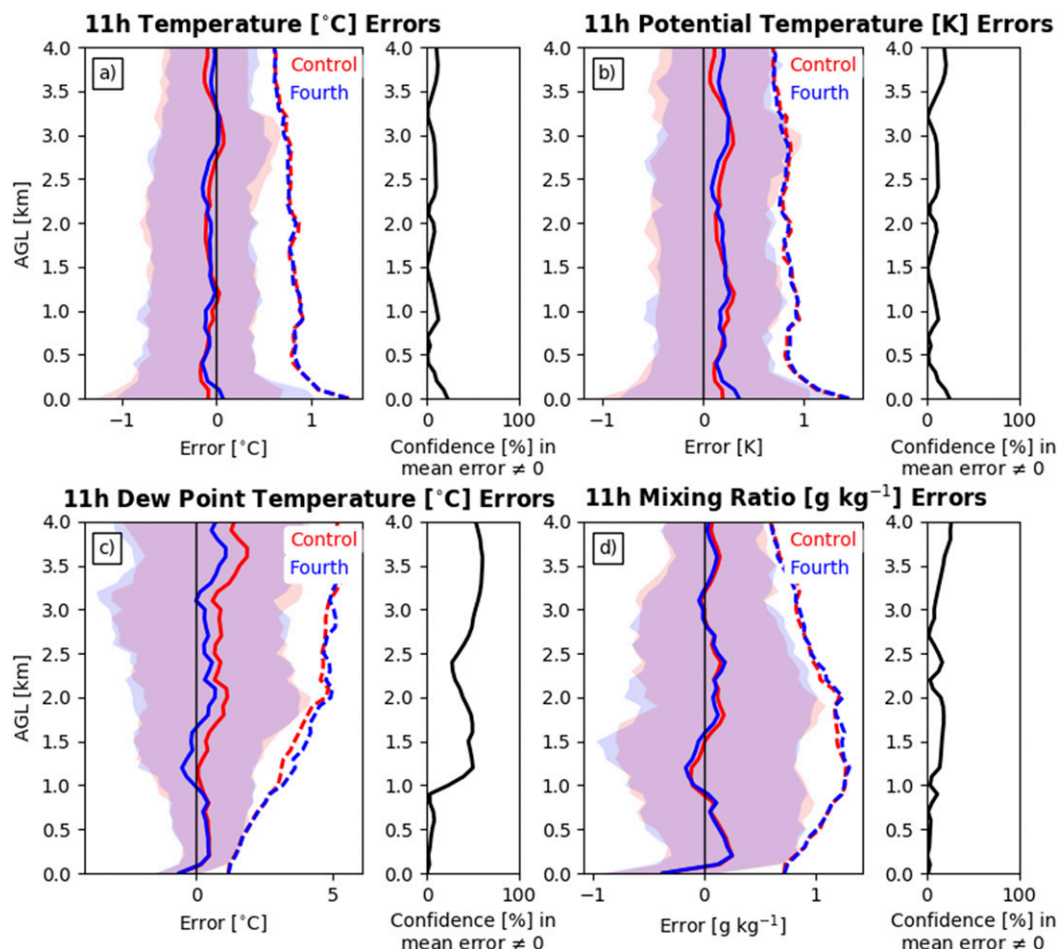
FIG. 4. (a) The left panel shows bias (solid) and mean absolute error (dashed) between 0 and 4 km AGL for the control (red) and fourth-order (blue) samples for the 11-h forecast temperature (°C). Shading represents the distribution between the 25th and 75th percentiles. The right panel is the vertical profile of confidence in the bias difference being nonzero between the control and fourth-order samples. (b)–(d) As in (a), but for potential temperature (K), dewpoint temperature (°C), and mixing ratio (g kg^{-1}), respectively. Note the varying x-axis values between the leftmost insets in each panel.

interpolation would add detail (realistic or otherwise) to both the observed and model-derived vertical profiles, but this has not been tested and is left for future study.

Only rawinsondes released at locations within the Storm Prediction Center's 0600 UTC day 1 convective outlook (valid for the subsequent 1200–1200 UTC period) for a given event are considered to isolate those observations located in thunderstorm-supporting environments. For example, verifying 3 May 2017 forecasts only include radiosonde locations within SPC's 3 May 2017 day 1 convective outlook at 0600 UTC for the period 1200 UTC 3 May–1200 UTC 4 May 2017 (Fig. 2). A total of 1665 rawinsonde profiles meet these criteria, of which 72.7% are from locations east of the Rocky Mountains. Note that most days in May 2017 are associated with relatively small numbers of tornado, severe wind, and

severe hail local storm reports, such that the sample primarily consists of rawinsonde profiles on days and in regions of primarily nonsevere thunderstorm potential.

Precipitation, whether stratiform or convective in nature, exerts a significant control on vertical thermodynamic profiles (e.g., PBL moistening and cooling, middle- to upper-tropospheric drying and warming). In convection-allowing model forecasts, these processes invoke the microphysical parameterization, and the clouds associated with the precipitation will influence the shortwave and longwave radiation budgets. To isolate the vertical profiles (observed and/or modeled) obtained in regions absent of precipitation and thus to reduce the degrees of freedom in the analysis, a method similar to that of Coniglio et al. (2013) is followed. A rawinsonde is not used in the evaluation if there is observed rainfall or
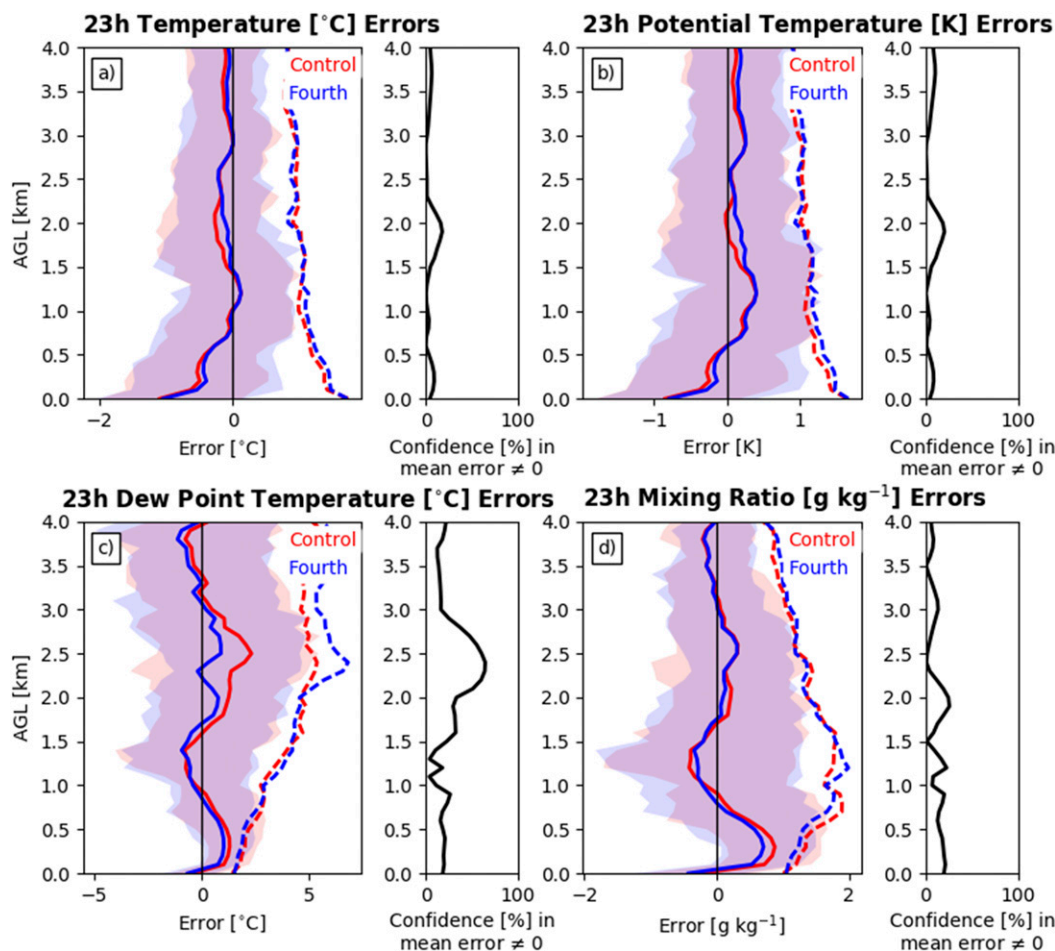
FIG. 5. As in Fig. 4, but for the 23-h forecast.

simulated rainfall, from either simulation, of $\geq 0.5\,mm\,h^{-1}$ at any grid point within 40 km in any 1 h during the 3 h prior to the rawinsonde release time. For observed rainfall, the Stage IV (Lin and Mitchell 2005) hourly multisensor precipitation analysis for the continental United States is used; simulated rainfall is derived by each model simulation. Removing soundings in or near regions of observed and/or modeled precipitation results in a total of 809 rawinsonde observations remaining for further evaluation.

c. Capping inversion identification

As an emphasis of this study is on evaluating the simulated capping inversion representation between the control and fourth-order simulations, a method for identifying observed capping inversions is required. To do so, a subset of the criteria outlined by Farrell and Carlson (1989) is used. A capping inversion is typically associated with a substantial increase in temperature and a decrease in dewpoint temperature through the

inversion layer, with temperature and dewpoint temperature profiles that follow a dry adiabat and mixing ratio isopleth, respectively, immediately above the inversion layer. Thus, an observed rawinsonde profile is said to contain a capping inversion if the temperature increases by any amount and the dewpoint temperature decreases by $\geq 2°C$ within any contiguous 200-m layer between the surface and 600 hPa. These criteria isolate capping inversions from other types of inversions (e.g., tropopause, frontal, nocturnal/radiation) that might be present. Of the original 809 rawinsonde observations, 383 are associated with capping inversions (275 at 1200 UTC and 108 at 0000 UTC). For those soundings that contain a capping inversion, cap strength is determined by calculating the parcel buoyancy minimum B_{min}, or the minimum value of the surface-based lifted parcel's temperature minus the environment's temperature (Trier et al. 2014).

To illustrate the effectiveness of the inversion identification method, observed soundings from 18 May 2017 are depicted in Fig. 3. The sounding from Dallas–Fort

Worth, Texas (KFWD), depicts a strong capping inversion ($B_{min} = -6°C$; Fig. 3a). Temperature rapidly increases with height and dewpoint temperature rapidly decreases with height over a shallow vertical layer near 850 hPa, atop of which substantial negative buoyancy is present over the layer from ~825 hPa to ~690 hPa. At Slidell, Louisiana (KLIX), a weaker capping inversion is present ($B_{min} = -1.5°C$; Fig. 3b). Although a substantial decrease in dewpoint temperature is observed at approximately 750 hPa, only a small increase in temperature is observed at this level. No capping inversion is present at Davenport, Iowa (KDVN; Fig. 3c). Although temperature increases with height over a shallow layer centered near 810 hPa, there is an insufficient decrease in dewpoint temperature with height within this layer to qualify as a capping inversion using the criteria outlined above.

Because of variability between soundings, it is difficult to objectively and accurately identify every rawinsonde profile with (and without) a capping inversion using this or any methodology. However, subjective analysis of inversion classifications suggest that the method utilized in this study effectively and accurately identifies most soundings with a capping inversion of any strength and nearly all soundings with well-defined capping inversions (e.g., Fig. 3).

d. Sounding analysis methods

The MetPy (May et al. 2017) package is used to calculate all thermodynamic variables from both the observed and model sounding profiles, including lifted condensation level (LCL), level of free convection (LFC), equilibrium level (EL), B_{min}, mixed-layer CAPE and CIN (MLCAPE and MLCIN; here, a 100-hPa-deep mixed layer is assumed), most-unstable CAPE and CIN (MUCAPE and MUCIN; here, the most-unstable parcel is that having the largest equivalent potential temperature over the lowest 300 hPa above the surface), and surface-based CAPE and CIN (SBCAPE and SBCIN).

Though MetPy does not use the virtual temperature correction (Doswell and Rasmussen 1994) for computing derived thermodynamic parameters, the internally consistent method used to compute these parameters for both modeled and observed soundings suggests that the evaluation presented herein is *qualitatively* robust, although it is acknowledged that *quantitative* accuracy is not assured.

In this study, B_{min} is computed as the parcel buoyancy minimum over all negative buoyancy layers that reside beneath an LFC; in effect, this limits B_{min} to the lower- to middle troposphere, consistent with Trier et al. (2014). Furthermore, MetPy by default includes negative buoyancy between a parcel's LCL and first LFC when integrating CIN for any lifted parcel, though both observed and modeled rawinsonde profiles can have multiple LFCs and thus multiple ELs, and only returns the first LFC
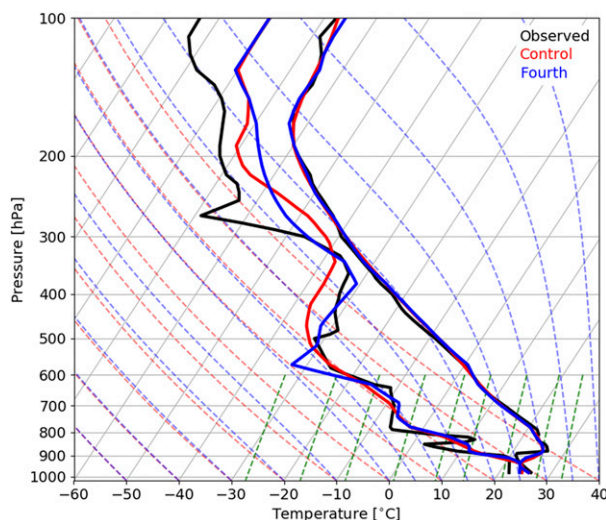

FIG. 6. Skew T–log$p$ diagram for KFWD valid at 1100 UTC 27 May 2017, showing an example of the control and fourth-order simulations artificially smoothing a capping inversion. The red line represents the control simulation sounding, the blue line represents the fourth-order simulation, and the black line is the observed sounding. The left and right traces depict dewpoint temperature and temperature, respectively. See Table 1 for each profile's derived thermodynamic variables.

when multiple LFCs exist. Thus, slight code modifications are made to integrate CIN over all negative buoyancy layers and return the LFC that resides above the identified B_{min} (which may, but need not necessarily, be the LFC closest to the ground).

For vertical profiles of temperature, potential temperature, dewpoint temperature, and mixing ratio, mean bias and mean absolute error (both defined as modeled minus observed) are computed over the lowest 4 km above ground level (AGL). Since the bias distributions for each sample are approximately normally distributed about their means at all altitudes [not shown; evaluated using SciPy following D'Agostino (1971) and D'Agostino and Pearson (1973)], a two-tailed Student's t test is used to evaluate the null hypothesis that the difference in the mean bias between the control and fourth-order simulations for a given variable and altitude is equal to zero. A given null hypothesis is rejected, and the differences between the means of the control and fourth-order simulation are said to be statistically significant, if and only if the confidence [here, $100 \times (1 - p)$, where p is the two-tailed decimal probability of the null hypothesis being false] exceeds 95%. Observed vertical profiles are treated as independent observations despite closely located observations at a given observation time likely not being truly independent of each other. Because of the assumption of independence, the sample size used in computing the Student's t-test statistic is the full sample size
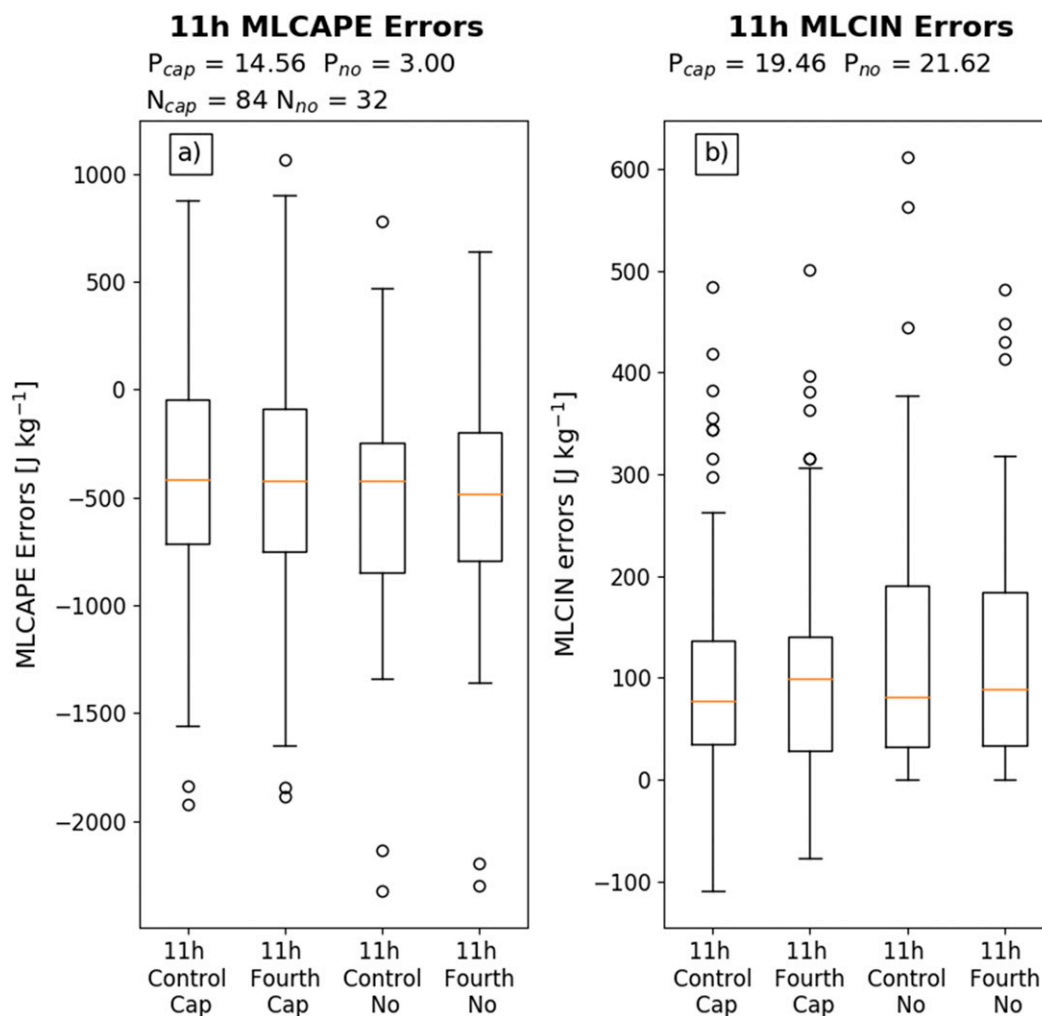
FIG. 7. Box-and-whisker diagrams of forecast (a) MLCAPE and (b) MLCIN for the 11-h forecast lead time for the control (Control) and fourth-order (Fourth) samples. The center orange line represents the mean of the sample, with the box enclosing the distribution between the 25th and 75th percentiles. The whiskers represent the maximum and minimum values, excluding outliers denoted in circles. The cases denoted with "cap" represent the sample that includes capping inversions, and cases denoted with "no" represent the sample of soundings without a capping inversion present. At the top P represents the p value for each respective sample, and N represents the sample size. In this study, CIN is assigned as a positive number. Cases with zero positive buoyancy are removed from all samples. Note the varying y-axis values between panels.

rather than a smaller effective sample size that attempts to account for nonindependent observations, such as in Coniglio et al. (2013). As will be seen, most samples are sufficiently large that this does not meaningfully impact the results presented herein. Using a smaller sample size would decrease the likelihood that a given difference is statistically significant at a specified confidence level.

e. Single-column model experiments

To gain further insight into the relative contributions of the vertical advection finite-difference formulation and PBL parameterization to the WRF-ARW modeled

capping inversion representation, a set of eight simulations for each of 30 strong capping inversion cases is conducted using the WRF-ARW single-column model (Hacker et al. 2007; Hacker and Rostkier-Edelstein 2007). The single-column model is configured without horizontal advection but with vertical advection and realistic physics (viz., the surface layer, land surface, radiation, and planetary boundary layer).

Single-column model simulations are run for the 30 strongest observed capping inversions, identified using B_{min} (section 2c), at 0000 UTC that pass the precipitation filtering criteria (section 2b). Simulations are initialized
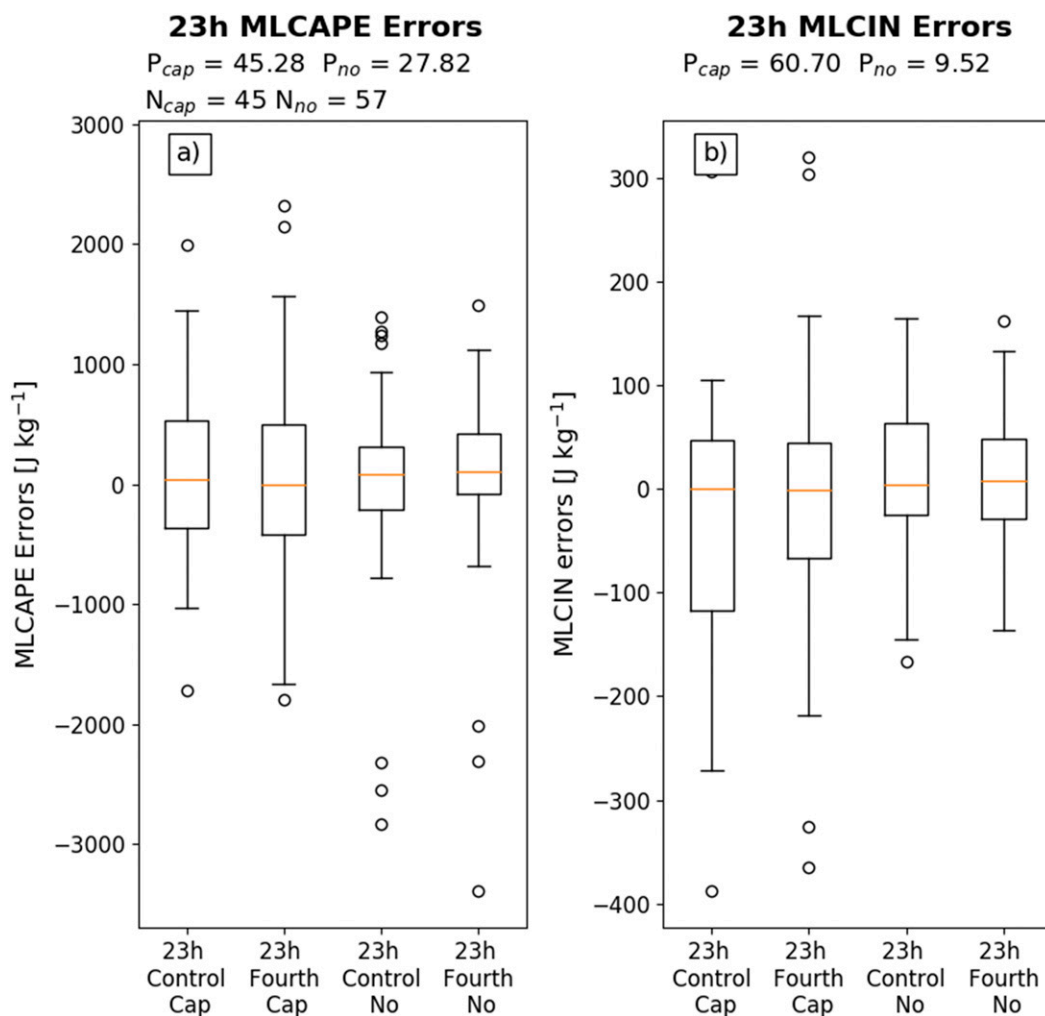
FIG. 8. As in Fig. 7, but for the 23-h forecast.

using the observed potential temperature, water vapor mixing ratio, and horizontal wind rawinsonde data. All land surface parameters (soil temperature and moisture, land-use index, soil type, vegetation fraction, canopy water) and the vertical velocity used to force the model are obtained from 0-h WRF-ARW model analyses at the model grid point closest to the observation location. Simulations are initialized at 0000 UTC and extend forward 23 h; the vertical velocity, soil temperature, and soil moisture forcing are set to remain constant over the simulation period. Physical parameterizations are identical to those used in the full-physics model simulations except for the PBL parameterization, as described below. Finally, the single-column model uses 35 vertical levels from the surface and 20 km; the number of vertical levels matches that of the full-physics model simulations, and the model top of 20 km approximates the 50-hPa model top of the full-physics model simulations.

Nine simulations are run for each of the 30 strongest observed capping inversions. In four, the MYJ PBL parameterization is used, and the vertical advection finite-difference formulation is varied among the first-, third-, fourth-, fifth-, and sixth-order-accurate formulations. Since the single-column model is hard coded to approximate vertical advection using first-order-accurate backward finite differences, the single-column model code is modified to implement the WRF-ARW third- through sixth-order-accurate vertical advection finite-difference schemes. In the remaining four simulations, the third-order-accurate vertical advection finite-difference formulation is used, and the PBL and associated surface layer parameterizations are varied between the Yonsei University (YSU; Hong et al. 2006), Asymmetric Convective Model 2 (ACM2; Pleim 2007), Mellor–Yamada–Nakanishi–Niino level 2.5 closure (MYNN; Nakanishi and Niino 2009), and quasi-normal scale elimination (QNSE;

11h MUCAPE Errors
$P_{cap} = 5.41 \quad P_{no} = 12.88$
$N_{cap} = 112 \quad N_{no} = 52$

11h MUCIN Errors
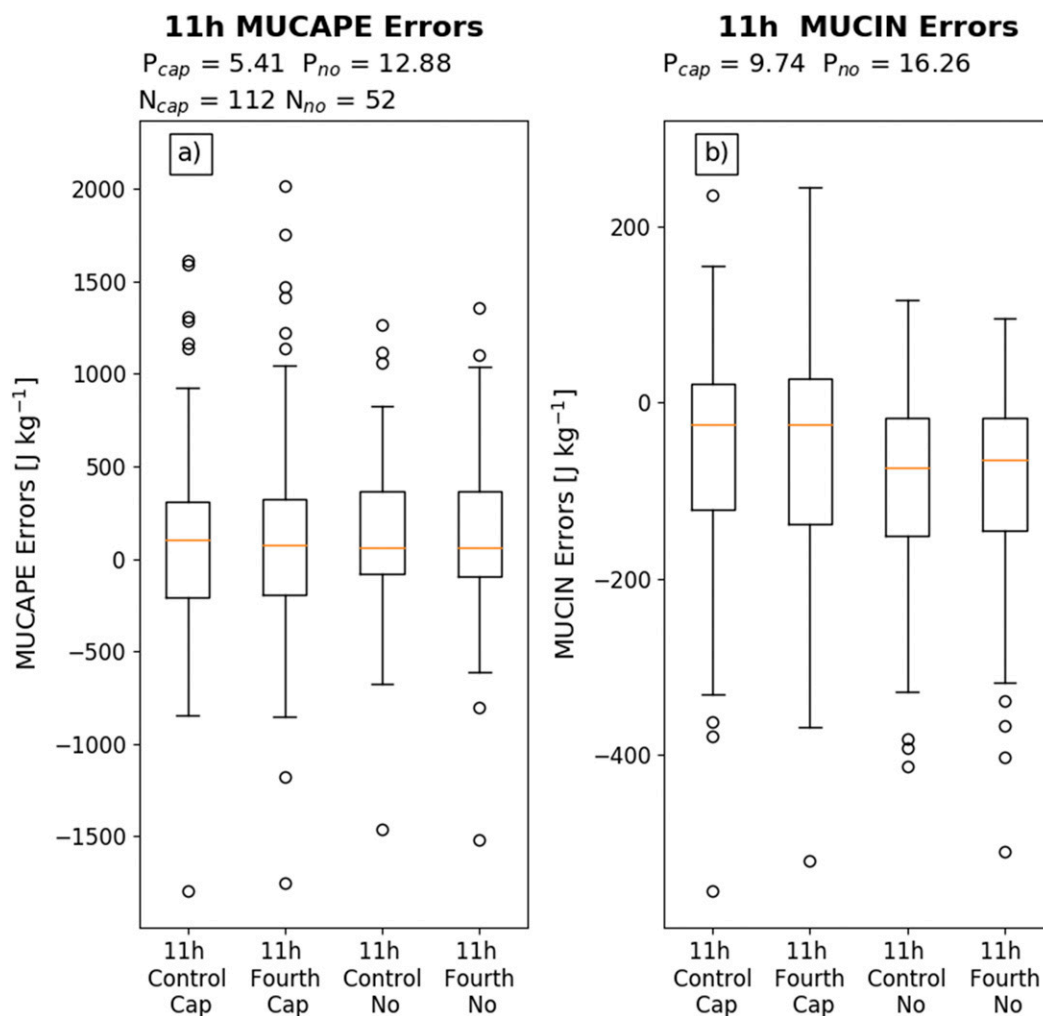$P_{cap} = 9.74 \quad P_{no} = 16.26$

FIG. 9. As in Fig. 7, but for MUCAPE and MUCIN for the 11-h forecast.

Sukoriansky et al. 2005) parameterizations. Of these, the first two represent upward mixing using nonlocal closures (i.e., turbulent eddies can span the depth of the boundary layer) and the last two and MYJ represent upward mixing using local closures (i.e., turbulence is limited to adjacent vertical levels). Only YSU represents downward mixing via nonlocal closure.

Given this methodology, it is important to emphasize the purposes of the single-column model experiments. Since the single-column model simulations are initialized with an observed sounding valid at 0000 UTC that contains a capping inversion, and in the absence of spatially or temporally varying forcing during the single-column model simulations, the single-column model simulations are unable to faithfully mimic the observed atmospheric evolution in capping inversion situations. Rather, the single-column model experiments document

the capping inversion's evolution in simplified environments (viz., weak lower- to middle-tropospheric synoptic-scale ascent) characteristic of those in which capping inversions are found over a full diurnal cycle. The simplified framework enables a more direct comparison of the effects of parameterized turbulent vertical mixing and vertical advection finite-difference formulation on the modeled capping inversion representation than in full-physics simulations, while also being associated with an infinitesimal computational cost.

3. Results

a. Temperature profiles

For the 11-h (morning) forecast, the mean bias of temperature in the lowest 4 km AGL is close to zero for both the control and fourth-order samples
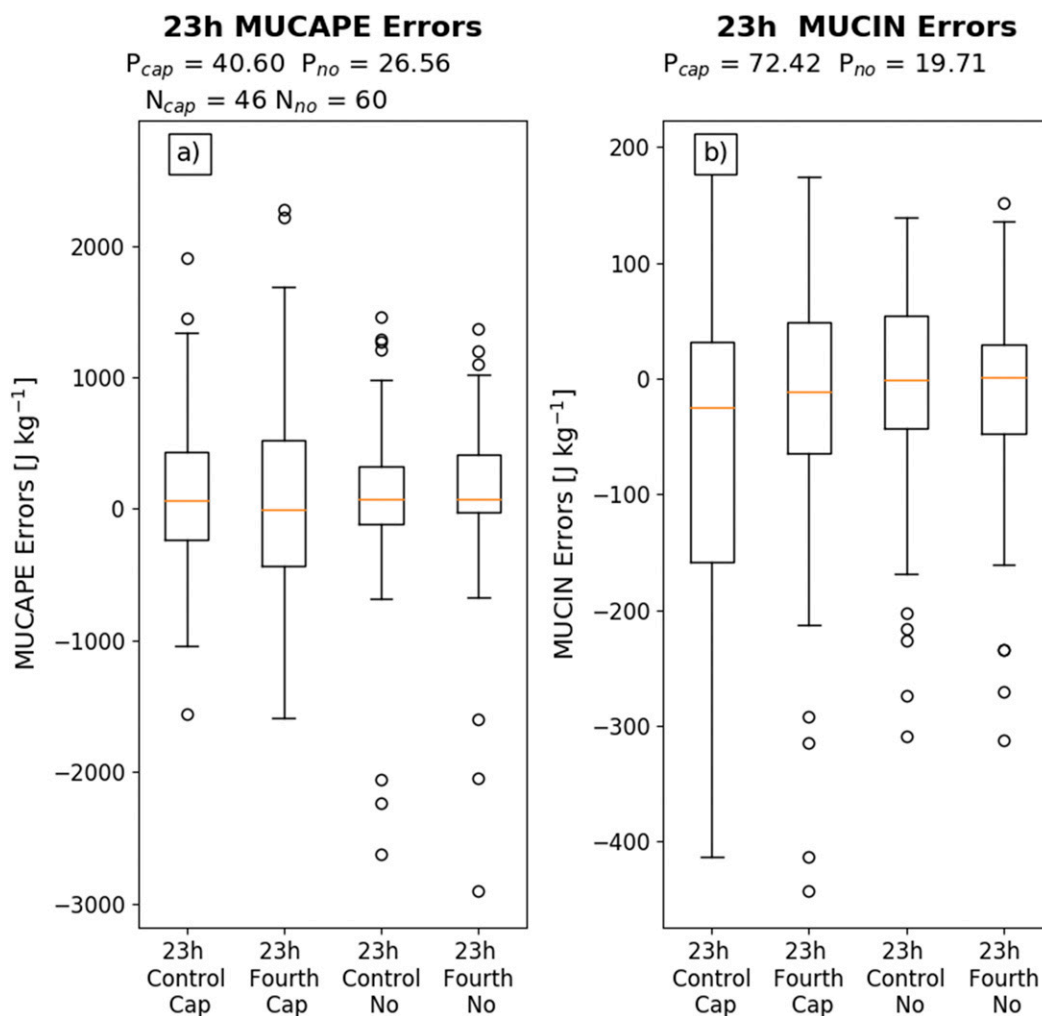
FIG. 10. As in Fig. 7, but for MUCAPE and MUCIN for the 23-h forecast.

(Fig. 4a), whereas there is a slight positive potential temperature bias in both samples (Fig. 4b). The vertical bias profiles for both temperature and potential temperature have nearly identical shape to each other, with the latter offset by about $+0.2\,\mathrm{K}$ from the former. WRF-ARW uses a staggered vertical grid, with the lowest model level on which the mass variables (including temperature and potential temperature) offset slightly above the surface, here, from $\sigma = 1$ to $\sigma = 0.9965$. As the raw model-level output is used in the analysis, the surface pressure is slightly lower in the model forecasts than in the observations, resulting in the small positive offset in the potential temperature analysis. A small positive potential temperature bias at 11 h is also identified by Coniglio et al. (2013), which they hypothesize results from a positive bias in the NAM initial conditions. However, numerous updates made to the NAM model between 2011 and 2012

(in Coniglio et al. 2013) and 2017 render it uncertain at best as to whether the root cause is the same in the present study.

For the 23-h (late afternoon) forecast, a cold or negative bias below 0.75 km AGL and a warm or positive bias between 0.75 and 2 km AGL are noted for both temperature and potential temperature in both the control and fourth-order samples (Figs. 5a,b). These vertical bias profiles are consistent with the tendency of the MYJ PBL parameterization to undermix in daytime convective boundary layers (e.g., Hu et al. 2010; Coniglio et al. 2013; Clark et al. 2015; Cohen et al. 2015; Burlingame et al. 2017). The mean absolute errors are roughly 1°C and 1 K for temperature and potential temperature, respectively, at all altitudes at both 11 and 23 h (Figs. 4 and 5a,b). The statistical confidence in the differences in bias between the control and fourth-order samples for both 11- and 23-h temperature and potential temperature forecasts is

11h SBCAPE Errors

$P_{cap} = 16.92 \quad P_{no} = 4.14$
$N_{cap} = 79 \quad N_{no} = 27$

11h SBCIN Errors
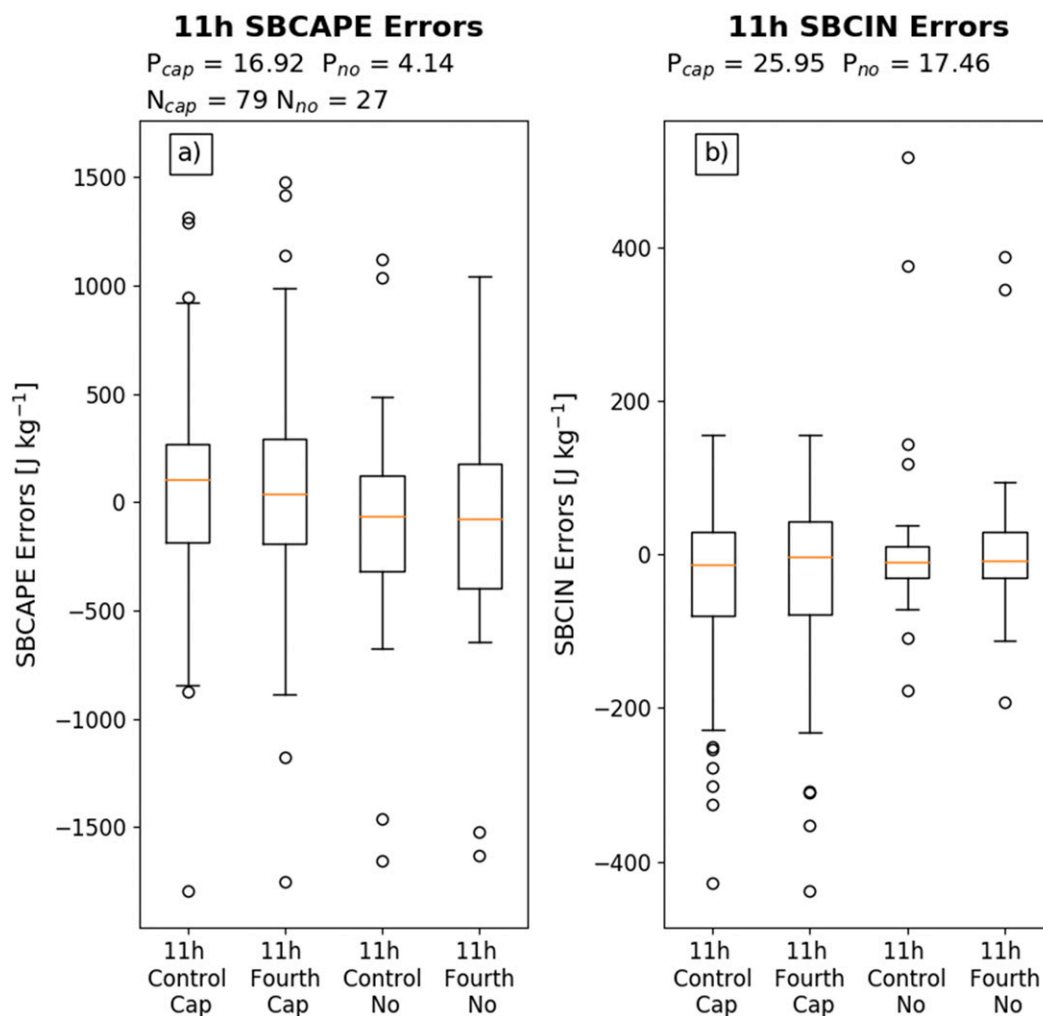
$P_{cap} = 25.95 \quad P_{no} = 17.46$

FIG. 11. As in Fig. 7, but for SBCAPE and SBCIN for the 11-h forecast.

too low at all altitudes for the null hypothesis of nonzero differences to be rejected.

b. Moisture profiles

For the 11-h forecast, the mean dewpoint temperature bias is slightly positive at nearly all altitudes above the surface layer, particularly for the control sample (Fig. 4c), whereas the mean water vapor mixing ratio bias is slightly positive between the top of the surface layer and 1 km AGL and slightly negative between 1 and 1.5 km AGL (Fig. 4d). Though the model simulations to this analysis time encompass the period 0000–1100 UTC, which is largely during the local nighttime in the United States in May, the water vapor mixing ratio bias above the surface layer is consistent with the tendency of the MYJ PBL parameterization to undermix in daytime convective boundary layers. It is believed that this bias profile largely results from a ~1°C lower-tropospheric moist bias in the

NAM initial conditions (which are generated from a 6-h NAM forecast using the MYJ PBL parameterization; Evans et al. 2018) that is maintained in the subsequent forecast.

The dewpoint temperature mean absolute error increases with height, reaching 5°C by 2 km AGL, whereas the mixing ratio mean absolute error is largest between 1 and 2 km AGL, peaking at a magnitude of $1.2\,\mathrm{g\,kg^{-1}}$ (Figs. 4c,d). Large dewpoint temperature mean absolute error results from differences in how shallow dry layers atop the PBL are represented in each model simulation and the observations (e.g., Fig. 6); it is not accompanied by large mixing ratio mean absolute errors since the absolute value of the water vapor mixing ratio decreases rapidly with increasing height. Similar results are noted at 23 h (Figs. 5c,d), with the MYJ parameterization's characteristic undermixing particularly evident at this time (during the local daytime

23h SBCAPE Errors
$P_{cap} = 46.04$ $P_{no} = 27.86$
$N_{cap} = 45$ $N_{no} = 58$

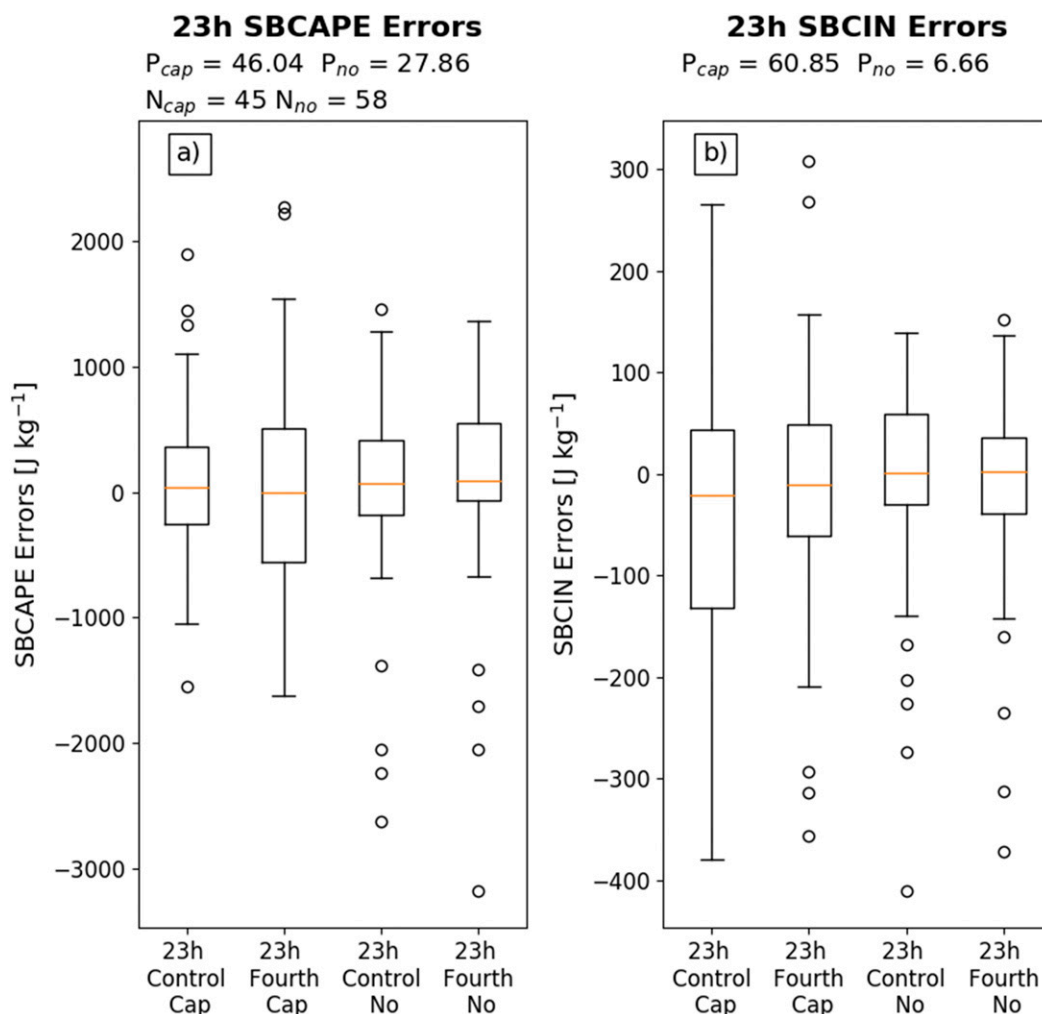## 23h SBCIN Errors
$P_{cap} = 60.85$ $P_{no} = 6.66$

FIG. 12. As in Fig. 7, but for SBCAPE and SBCIN for the 23-h forecast.

hours) as compared to 11 h. However, the statistical confidence in the differences in bias between the control and fourth-order samples for both 11- and 23-h dewpoint temperature and water vapor mixing ratio forecasts is too low at all altitudes for the null hypothesis of nonzero differences to be rejected.

c. Derived sounding variables

Mixed-layer parcels (Fig. 7) in the morning tend to be more stable than observed for both the control and fourth-order samples, with mean MLCAPE errors of approximately $-500 \, \mathrm{J \, kg^{-1}}$ and mean MLCIN errors of approximately $80 \, \mathrm{J \, kg^{-1}}$. A similar stable bias at 11 h is noted by Coniglio et al. (2013), a result that is shown in their study to be insensitive to the PBL parameterization. A mixed-layer parcel is defined in this study and in Coniglio et al. (2013) as one having the mean characteristics of the lowest 100 hPa of the profile, independent

of whether there is a mixed layer. Thus, it is hypothesized that the mixed-layer parcel stable bias results from errors in the modeled representation of the nocturnal inversion that influence the mixed-layer parcel stability calculation. Lending partial credence to this hypothesis, note that mean MLCAPE and MLCIN errors for both the control and fourth-order samples are approximately $0 \, \mathrm{J \, kg^{-1}}$ in the 23-h forecast (Fig. 8). Both distributions are similar between the capping inversion and no capping inversion samples (Figs. 7 and 8). There is no statistically significant improvement at either 11 or 23 h in forecast MLCAPE or MLCIN in the fourth-order simulation relative to the control, whether or not a capping inversion is present.

Similar evaluations are conducted for other commonly considered parcels, namely those that are surface based or are the most unstable (here defined as the parcel in the lowest 300 hPa AGL with the maximum
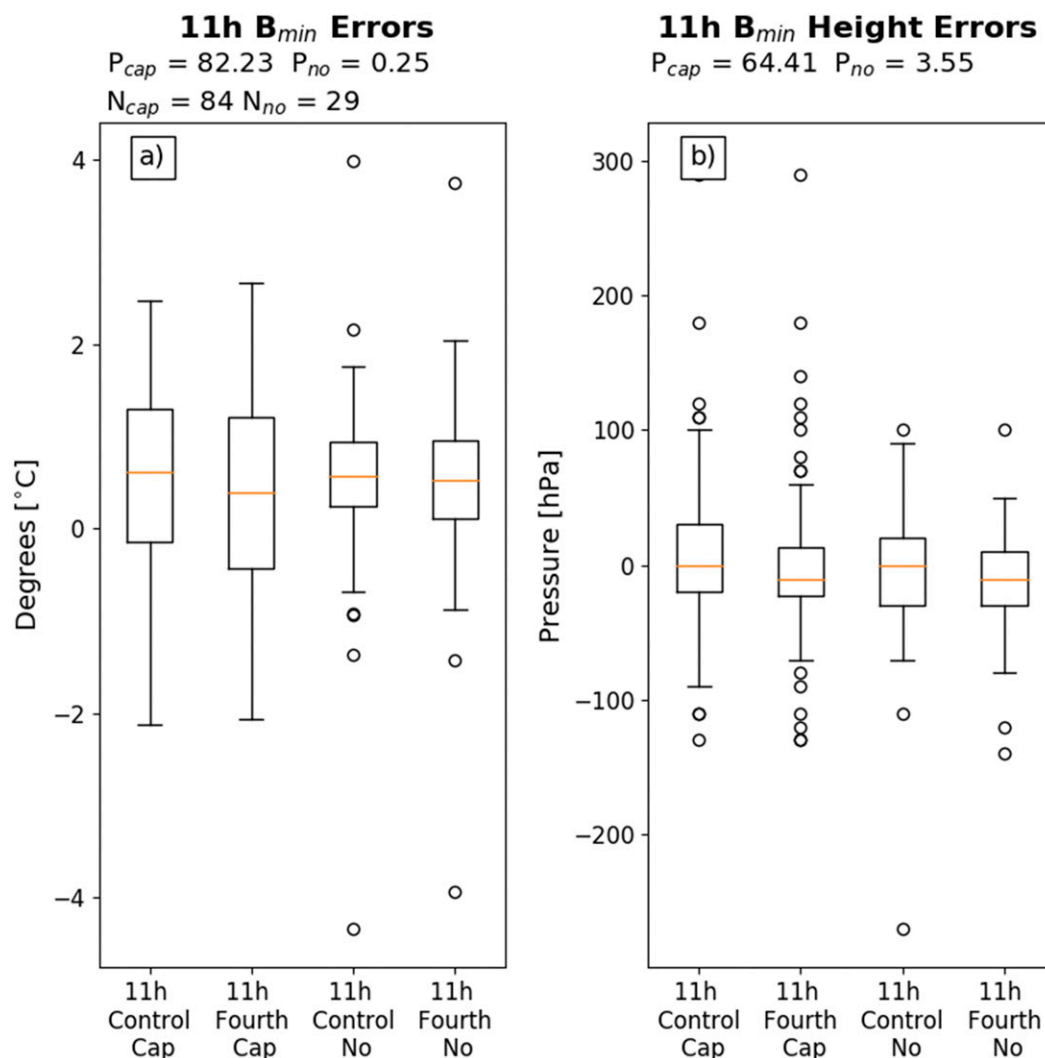
FIG. 13. As in Fig. 7, but for $B_{\min}$ and the $B_{\min}$ height error for the 11-h forecast. Note that it is still possible to have negative buoyancy but not meet the criteria for a capping inversion. As $B_{\min}$ is defined as the minimum of the surface-based parcel temperature minus the environment temperature, $B_{\min}$ is a negative value (°C) when SBCIN is nonzero; thus, positive $B_{\min}$ errors indicate a given model sample is less stable, and negative $B_{\min}$ errors indicate a given model sample is more stable than the observations.

equivalent potential temperature). At 11 and 23 h, most-unstable parcels are, on average, slightly less stable than the observations for both the control and fourth-order samples (Figs. 9 and 10). The only meaningful difference between the capping inversion and no capping inversion samples is seen at 11 h with MUCIN, at which time the mean unstable bias is greater in the no capping inversion sample than in the capping inversion sample (Fig. 9b). Mean errors in surface-based CAPE and CIN at 11 and 23 h, for both the control and fourth-order samples and capping inversion and noncapping inversion cases, are approximately $0 \, \text{J kg}^{-1}$ (Figs. 11 and 12). An exception is found for SBCAPE at 11 h for capping inversion cases, where it is slightly unstable biased (Fig. 11a).

As with mixed-layer CAPE and CIN, differences in most-unstable and surface-based CAPE and CIN between the control and fourth-order samples are not statistically significant.

Parcel minimum buoyancy $B_{\min}$ and the pressure level at which it is found are used to evaluate model forecasts of capping inversion strength (or, more generally, minimum buoyancy) and height. The control and fourth-order samples each underpredict the minimum negative buoyancy at 11 h, with mean errors of approximately 0.75°–1°C for both the capping inversion and no capping inversion cases (Fig. 13a); however, the mean error in the level at which $B_{\min}$ is found is approximately 0 hPa for both samples and all cases (Fig. 13b). At 23 h, the

23h B$_{min}$ Errors
P$_{cap}$ = 77.92 P$_{no}$ = 11.30
N$_{cap}$ = 58 N$_{no}$ = 100

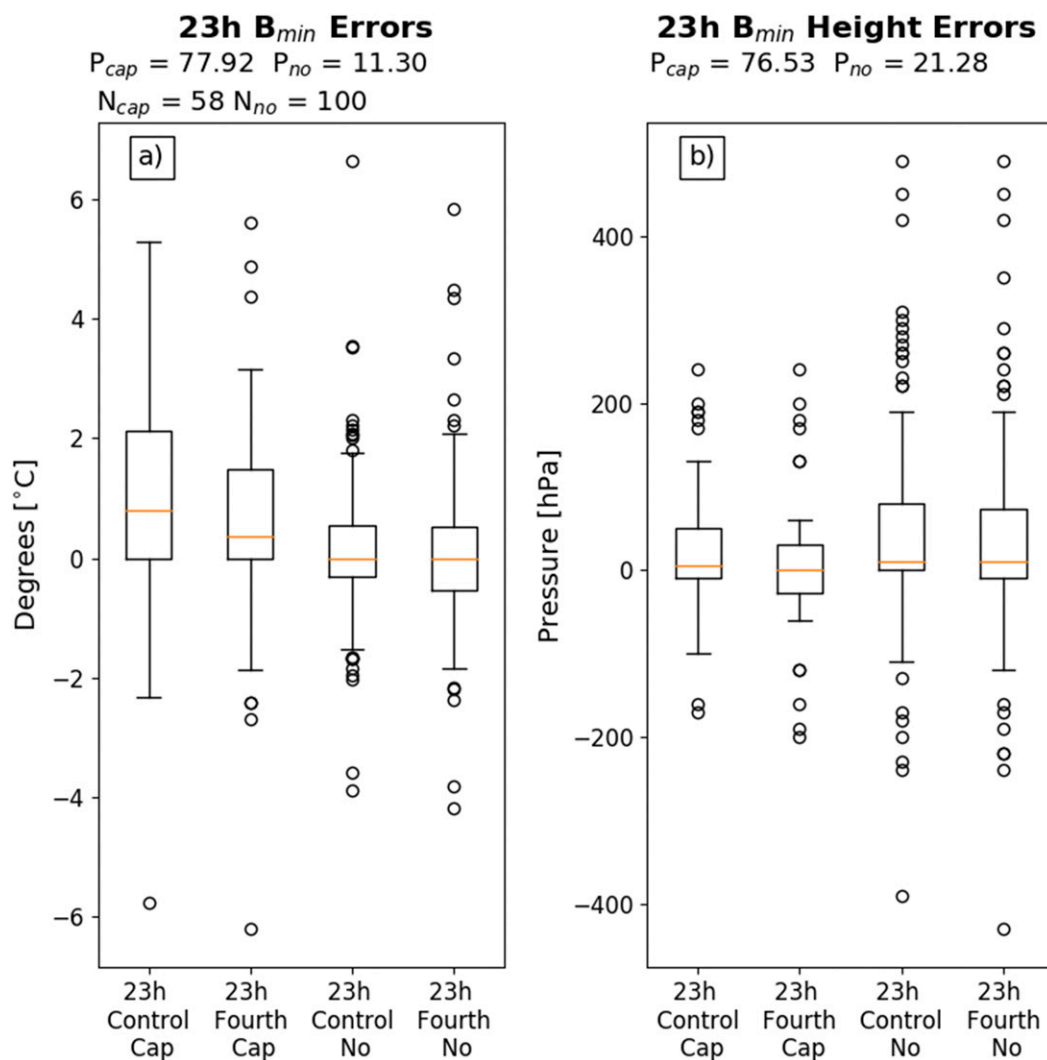## 23h B$_{min}$ Height Errors
P$_{cap}$ = 76.53 P$_{no}$ = 21.28

FIG. 14. As in Fig. 13, but for the 23-h forecast.

control and fourth-order samples again underpredict minimum negative buoyancy for capping inversion cases (Fig. 14a). Together with the 11-h result, this indicates that, in the sample mean, the WRF-ARW Model underpredicts the cap strength by approximately 0.5°–1°C. The mean B$_{min}$ error at 23 h for the no capping inversion cases is approximately 0°C, and the mean error in the level at which B$_{min}$ is found is approximately 0 hPa for both the no-capping and capping inversion cases (Fig. 14). Of all the variables considered in this study, the confidence in rejecting the null hypothesis of zero difference between the control and fourth-order samples is highest for B$_{min}$, and generally in the fourth-order sample's favor, but remains insignificant at the specified 95% confidence level.

Despite underpredicting the capping inversion or negative buoyancy strength, on average both the control

and fourth-order samples are generally associated with a deeper vertical layer over which negative buoyancy is found. This is particularly evident for the noninversion cases at 11 h, with near-zero mean surface-based LCL error (Fig. 15a) and a −50-hPa mean surface-based LFC error (i.e., the surface-based LFC is 50 hPa too high above the ground; Fig. 16a), and for both inversion and noninversion cases as 23 h, with positive mean surface-based LCL errors (i.e., the surface-based LCL is too close to the ground; Fig. 15b) and near-zero mean surface-based LFC errors (Fig. 16b). Given that mean SBCIN errors are near zero (Figs. 11b and 12b), this implies a deeper negative buoyancy layer in both the control and fourth-order samples relative to observations that, in turn, influences the vertical displacement (but not necessarily the vertically integrated forcing) required to lift a parcel to its LFC. As with all other variables considered,
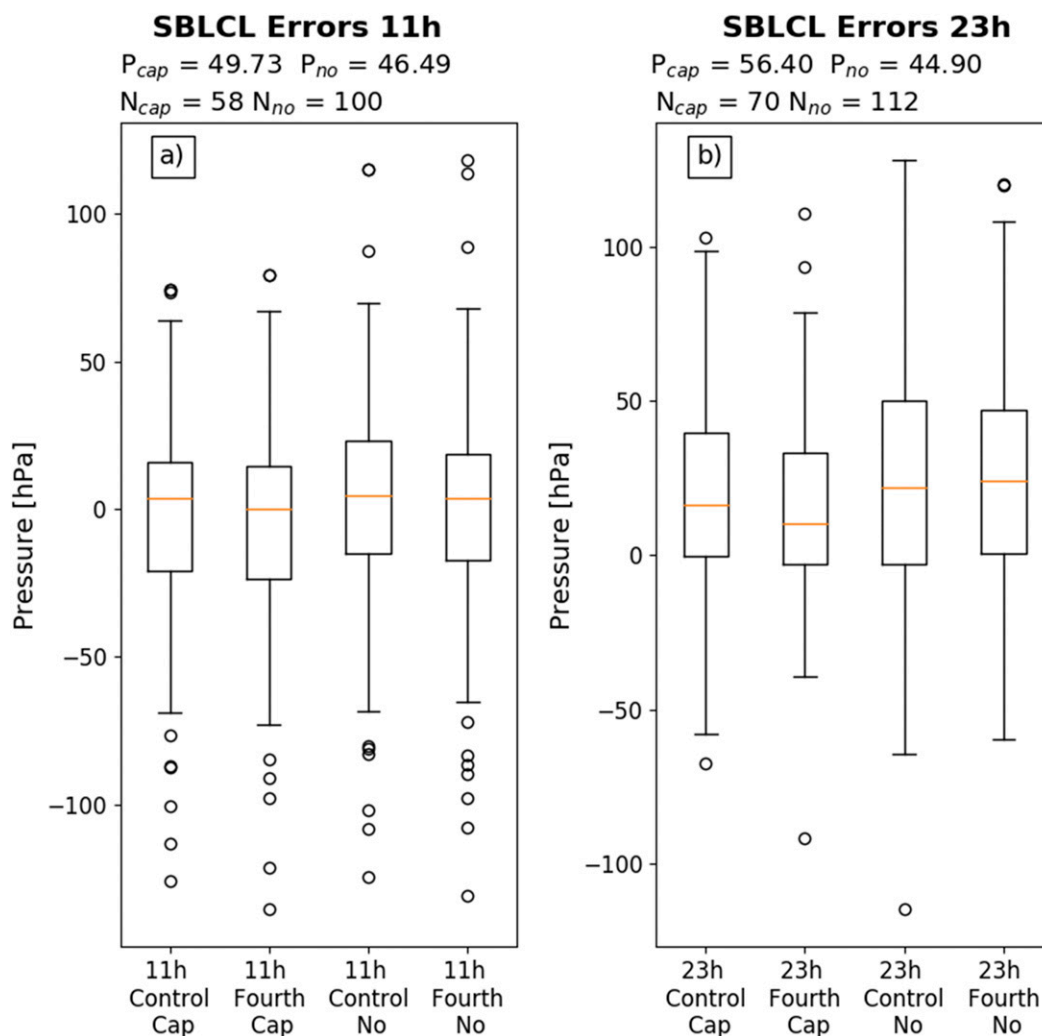
FIG. 15. As in Fig. 7, but for surface-based LCL (SBLCL) for the (a) 11- and (b) 23-h forecasts. Note that the *y* axes vary between panels.

however, the differences in surface-based LCL and LFC between the control and fourth-order samples are not statistically significant to at least 95% confidence.

d. Kinetic energy spectrum analysis

Given the results presented to this point, it is natural to ask why the fourth-order-accurate vertical advection finite-difference formulation does *not* improve the representation of model-derived vertical profiles and capping inversions, and what impact does the fourth-order-accurate formulation have on the model simulations? To attempt to answer these questions, a kinetic energy (KE) spectrum analysis is conducted for both sets of simulations following Skamarock (2004) and Skamarock et al. (2014). KE spectra are computed following Skamarock (2004) and include hourly forecast data between forecast hours 12 and 36 (after the

model spinup period) at all vertical levels and all grid points more than 15 points away from the lateral boundaries to mitigate the influence of the specified lateral boundary conditions. Spectra are first computed separately for each simulation between 3 and 31 May, from which the mean and standard deviation are computed. Note that the KE spectra are sensitive to implicit numerical damping but not to numerical dispersion, as numerical dispersion is only associated with departures of a wave's phase speed from its true velocity but not a change in the amplitude of or the energy carried by a given wave.

At long wavelengths ($\lambda \geq 1000$ km), the two spectra are nearly identical, with each exhibiting the expected k^{-3} dependence (Fig. 17). As the wavelength becomes smaller, the two spectra diverge, with higher KE in the fourth-order simulation set than the control. The fourth-order simulation KE spectrum exhibits the expected
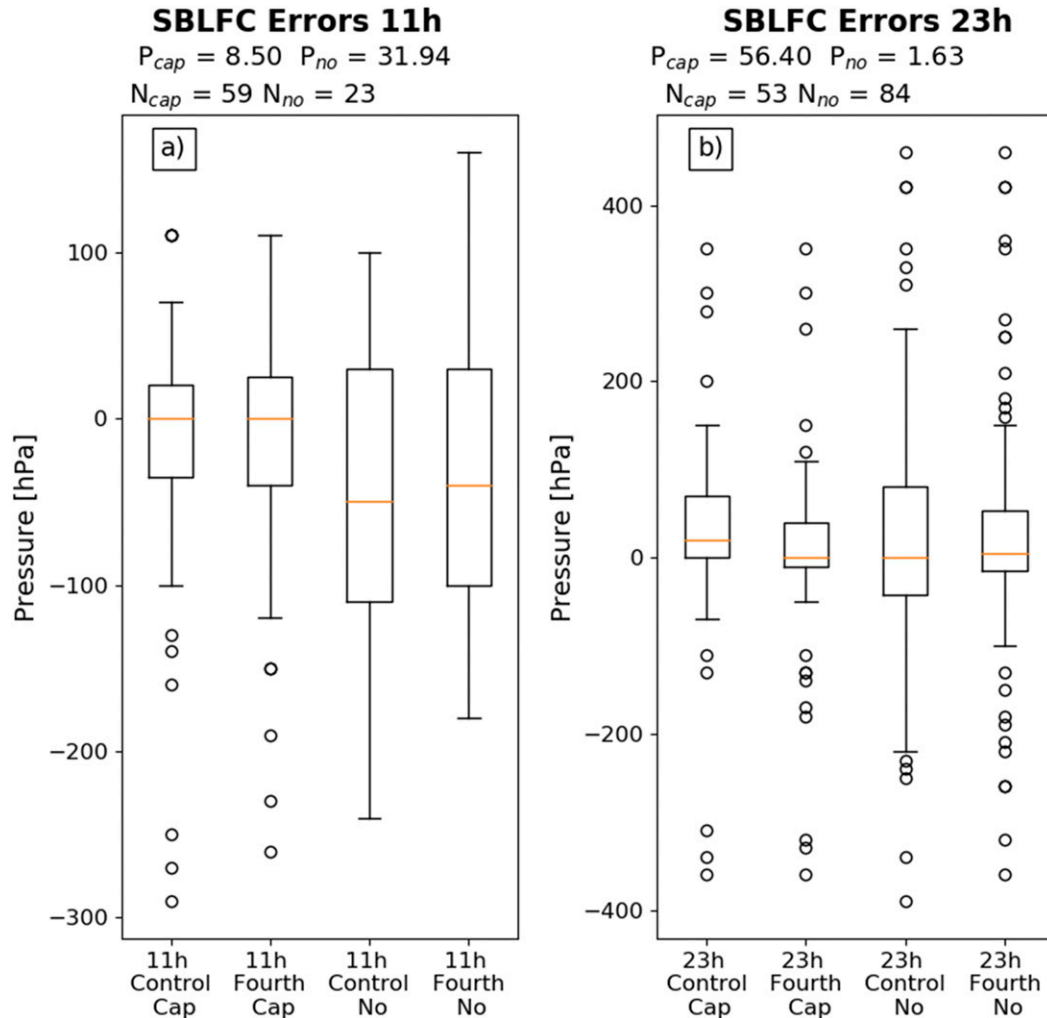
FIG. 16. As in Fig. 7, but for surface-based LFC (SBLFC) for the (a) 11- and (b) 23-h forecasts. Note that the y axes vary between panels and that the LFC is defined in this study as that occurring above the level at which B_{min} is located.

$k^{-5/3}$ dependence, whereas the control KE spectrum decays at a slightly faster rate due to the implicit damping associated with the third-order-accurate vertical advection finite-difference formulation. The extent to which the improved spectral match relative to observations contributes to higher forecast skill in the fourth order than the control simulations is left for future work, however. At wavelengths smaller than the model's effective resolution, or the wavelength at which the model's KE spectrum begins to substantially depart from the observations (Skamarock 2004; for WRF-ARW, this is $\sim7\Delta x$, or ~28 km in the simulations in this study), KE in the fourth-order simulations is rapidly diminished such that it matches that of the control simulations by the shortest-representable wavelength ($2\Delta x$). This is hypothesized to result from implicit damping associated with the fifth-order-accurate horizontal advection finite-difference

formulation and the third-order Runge–Kutta temporal finite-difference formulation, as well as explicit horizontal damping employed in all model simulations. As a result, the fourth-order-accurate vertical advection finite-difference formulation does not improve the representation of short-wavelength features such as capping inversions in the WRF-ARW Model. More generally, it is concluded from this analysis that the vertical advection finite-difference approximation is not the primary contributor to degraded modeled capping inversion representation, with other sources of implicit and explicit damping having a comparatively greater influence.

e. Single-column model simulation results

In the single-column model simulations, model-forecast capping inversions are almost always found at a higher altitude after 23 h than in the model initialization (e.g.,
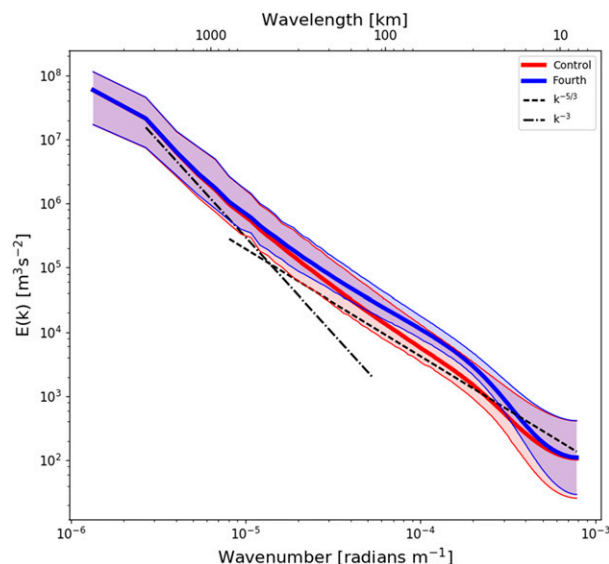
FIG. 17. The KE spectra for the control (red) and the fourth-order (blue) simulations. Solid lines represent the mean spectrum for each set of simulations as a function of wavenumber and wavelength, whereas shaded areas represent the distributions for each simulation set between the 0th and 100th percentiles. The dashed and dotted–dashed lines represent the hypothetical spectra with slopes of $k^{-5/3}$ and k^{-3}, respectively, representing the meso- and smaller-scale and large-scale dependences, respectively, of the atmospheric KE spectrum.

Fig. 18, which depicts output for four representative cases simulated with the single-column model). This is because most of the capping inversions considered herein are located in environments of weak lower- to middle-tropospheric synoptic-scale ascent, which acts over the 23-h simulation duration to lift and weaken the capping inversion to an extent that depends on the strength of both the inversion and ascent. In approximately 50% of cases, the prolonged ascent is sufficient to result in air parcels being lifted to saturation (e.g., Fig. 18d). This is an unrealistically large fraction of cases and results from the absence of spatially and temporally varying forcing used within this single-column model framework.

Model-forecast capping inversions are nearly identical for the third- through sixth-order-accurate vertical advection finite-differencing schemes (Fig. 18), with only small differences noted between the forecasts from these schemes and those of the first-order-accurate vertical advection finite-differencing scheme in selected cases (e.g., Fig. 18a). Larger variation in modeled capping inversion structure is noted in the PBL parameterization sensitivity simulations and is primarily manifest in the implied vertical extent of parameterized turbulent vertical mixing in the simulations (Fig. 18). This holds even in the cases where air parcels are lifted to saturation

(e.g., Fig. 18d). In many cases, the implied vertical extent of the parameterized turbulent vertical mixing is consistent with previous studies that used full-physics models (Coniglio et al. 2013; Burlingame et al. 2017), with deeper mixing by the nonlocal ACM2 and YSU parameterizations and shallower mixing by the local MYJ and MYNN parameterizations (e.g., Figs. 18b,c).

4. Summary and discussion

This study tested the hypothesis that degraded capping inversion representation in the WRF-ARW Model largely results from implicit damping associated with the default third-order-accurate vertical advection finite-difference formulation. The primary dataset used to test this hypothesis was output from two parallel deterministic, convection-allowing numerical simulations conducted for a set of 25 days in May 2017: one using a third-order-accurate vertical advection finite-difference formulation and one using a fourth-order-accurate vertical advection finite-difference approximation, which is numerically dispersive but does not implicitly dampen. Model-derived thermodynamic profiles and stability parameters were evaluated against 1200 UTC (11 h) and 0000 UTC (23 h) routine rawinsonde observations collected in known thunderstorm-supporting environments unaffected by recent precipitation in observations or either model forecast. For model-derived thermodynamic stability parameters, the evaluation was conducted separately for rawinsonde subsets with and without observed capping inversions. Two-tailed Student's t tests were used to test the guiding hypothesis and evaluate the extent to which differences between the third- and fourth-order-accurate finite-difference approximations were statistically significant to ≥95% confidence.

There was no statistical significance in the differences in sample-mean bias between the control and fourth-order simulations for any parameter, variable, or vertical level considered. As a result, the hypothesis that the fourth-order-accurate vertical advection finite-difference formulation would result in improved capping inversion forecasts relative to the default third-order-accurate vertical advection finite-difference formulation was rejected. This result was supported by single-column model simulations that demonstrated a negligible difference in simulated capping inversion structure between the higher-order-accurate vertical advection finite-differencing schemes tested. A greater control on simulated inversion structure, particularly inversion height, by PBL parameterization was found in the single-column model simulations. There are two possible conclusions that can be drawn from these findings: that implicit damping (tied to the odd-order-accurate
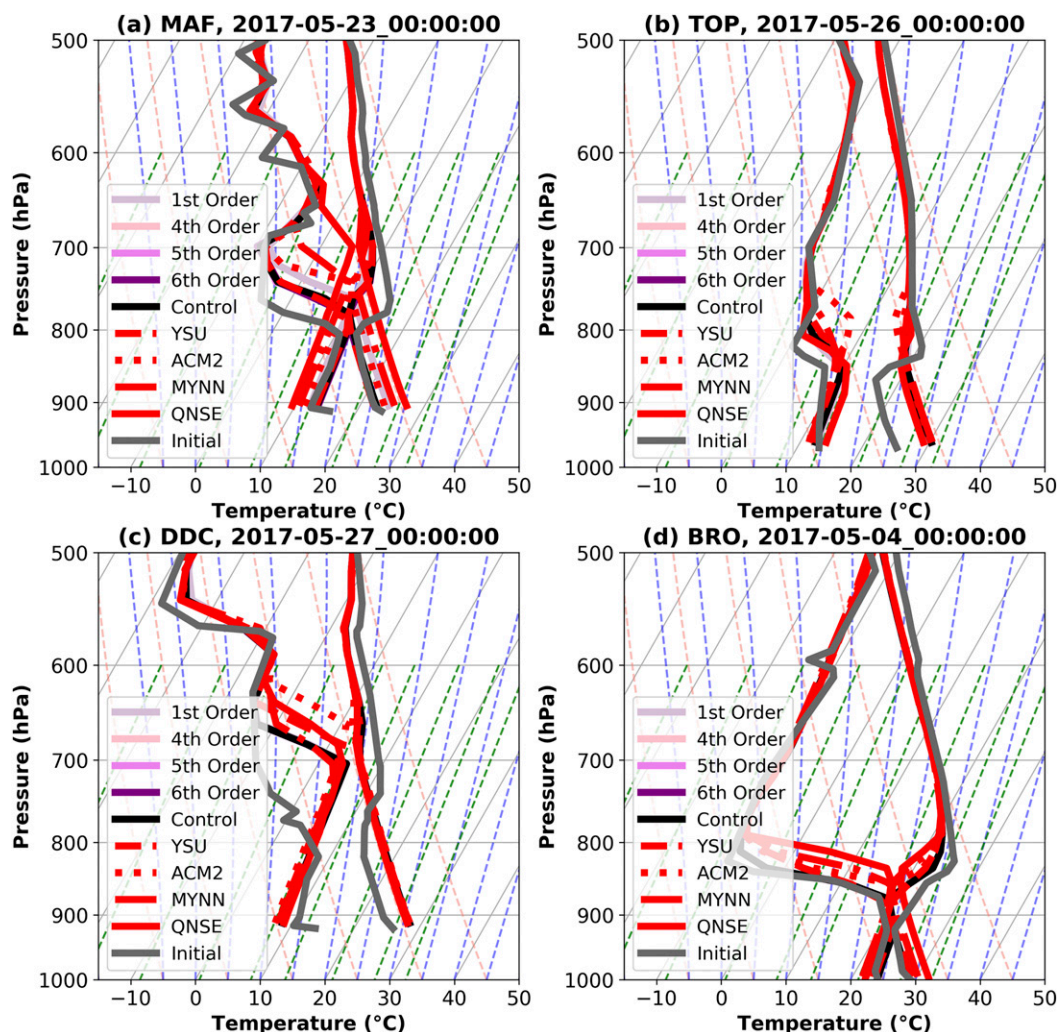
FIG. 18. (a) Skew T-log$p$ diagrams, between 1000 and 500 hPa, of temperature (boldface right lines) and dewpoint temperature (boldface left lines), at 23 h for the single-column model simulations initialized at 0000 UTC 23 May 2017 using the observed Midland, TX (MAF), vertical temperature and dewpoint temperature profiles (gray lines). Control simulation output is given by the boldface black lines. Output from the first-, fourth-, fifth-, and sixth-order-accurate vertical advection finite-differencing simulations are depicted by the light purple, peach, pink, and purple lines, respectively. Output from the YSU, ACM2, MYNN, and QNSE PBL parameterization simulations are depicted by the long-dashed, short-dashed, dashed–dotted, and solid red lines, respectively. Dashed light orange, light blue, and light green lines indicate isotherms, dry adiabats, and moist/saturated adiabats, respectively. (b) As in (a), but for Topeka, KS (TOP), initialized at 0000 UTC 26 May 2017. (c) As in (a), but for Dodge City, KS (DDC), initialized at 0000 UTC 27 May 2017. (d) As in (a), but for Brownsville, TX (BRO), initialized at 0000 UTC 4 May 2017.

finite-differencing schemes) and numerical dispersion (tied to the even-order-accurate finite-differencing schemes) have similar deleterious impacts on simulated capping inversion representation, or that the vertical advection finite-difference formulation in the WRF-ARW Model (over the range of formulations tested) is not the primary cause of degraded capping inversion representation. The latter conclusion is supported in part by a KE spectrum analysis, which implied that other damping mechanisms are more

likely contributors to degraded capping inversion representation, although further study is necessary to increase the confidence in this conclusion.

This study demonstrated that WRF-ARW Model capping inversions are too weak on average, with negative buoyancy spread out over too deep of a vertical layer versus observations for both capping inversion and noncapping inversion cases, particularly during the local late afternoon hours. An example is provided in Fig. 6,

TABLE 1. Thermodynamic variables for each profile found in Fig. 6. Note that the LFC is defined as that occurring above the level at which B_{min} is located. SBEL is the surface-based EL.

	Observed	Control	Fourth order
MLCAPE (J kg^{-1})	3330.60	4373.79	4051.20
MLCIN (J kg^{-1})	335.12	269.32	299.75
MUCAPE (J kg^{-1})	3062.44	4373.79	4111.16
MUCIN (J kg^{-1})	385.86	269.32	293.90
SBCAPE (J kg^{-1})	3062.44	4373.79	4051.20
SBCIN (J kg^{-1})	385.86	269.34	299.75
B_{min} (°C)	−7.36	−5.57	−5.79
B_{min} height (hPa)	860.00	850.00	820.00
SBLCL (hPa)	921.99	962.30	955.44
SBLFC (hPa)	670.00	710.00	700.00
SBEL (hPa)	164.13	153.73	154.39

depicting a case where the cap strength is under-predicted by approximately 1.5°C in both the control and fourth-order simulations (Table 1) that results from a modeled capping inversion that is too smooth. Further, though the ~250-hPa depths of the negative buoyancy layer are similar between each model simulation and observations, the layer over which the negative buoyancy is found is ~40 hPa closer to the surface in each model simulation (Table 1). Finally, the tendency of the MYJ PBL parameterization to undermix in the daytime convective boundary layer, thus resulting in a cool, moist bias near the surface and a warm, dry bias near the top of the PBL, was also corroborated by the analysis.

It is worthwhile to consider the operational implications of the results of this study, particularly since the operational RAP and HRRR models also use the WRF-ARW Model (albeit with a larger number of vertical levels and a higher-order vertical advection finite-differencing scheme, among other differences). Despite model configuration and initialization differences, an analysis of 11-h RAP and HRRR model forecasts of the capping inversions considered in this study suggests that both operational models struggle to adequately resolve sharper, stronger capping inversions of the sort depicted in Fig. 6 (not shown). Although no studies have yet been conducted to specifically evaluate RAP and HRRR model capping inversion forecasts, this finding is generally supported by evaluations of HRRR model forecasts of the sharp vertical temperature gradients atop shallow cold air masses in the lee of sloped terrain (James et al. 2018). However, short-range HRRR model forecasts of mixed-phase precipitation events are generally reasonable in their depiction of deeper inversions in the associated cold-season environments (Ikeda et al. 2017). However, further work is needed to quantify the extent to which the RAP and HRRR models adequately

resolve capping inversions, as well as to understand the implications for other forecast elements such as convection initiation (e.g., if observed CIN beneath weak capping inversions is not present in model forecasts due to an overly smooth simulated inversion), dryline propagation, and air quality.

This and previous studies documented that model vertical resolution and PBL parameterization are not main contributors to degraded WRF-ARW modeled capping inversions, whereas this study further implies that the vertical advection formulation may also not be a primary contributor to degraded WRF-ARW modeled capping inversions. Thus, we are left to ask, what *are* the main contributors to degraded WRF-ARW modeled capping inversions? A cursory analysis of NAM, RAP, and HRRR 0-h analyzed capping inversions suggests that each model's analysis is unable to adequately resolve observed capping inversions with large vertical temperature gradients at their base (not shown), suggesting that each model's cycled initialization (which updates a short-range model forecast) may contribute to subsequent forecast capping inversion errors. Additionally, numerical models using semi-Lagrangian methods to represent three-dimensional advection do not suffer from implicit numerical damping associated with finite-difference formulations for these terms in the equations of motion. It is likely that this contributes to improved capping inversion representations in the UM relative to the WRF-ARW Model. However, the UM uses comparatively weaker three-dimensional diffusion than WRF-ARW (D. Walters 2018, personal communication), which could also contribute to this finding. Among models that use finite-difference formulations to solve for vertical advection, it has been hypothesized that models using height-based terrain-following coordinates better depict capping inversions than those using pressure-based terrain-following coordinates (C. Alexander 2018, personal communication). Although the reasons why this may be true are unclear, it has been suggested that the increased pressure thickness of the warm elevated mixed layer that typically resides atop a capping inversion may increase the spacing between adjacent levels across the inversion in pressure-based relative to height-based terrain-following coordinate models (M. Coniglio 2018, personal communication). Further investigation is needed to test these hypotheses, however, and quantify the respective contributions of all hypothesized contributors to degraded WRF-ARW Model capping inversions.

author was supported by a teaching assistantship during the conduct of this research. This work benefited from the second author's participation in the 2015, 2017, and 2018 NOAA Hazardous Weather Testbed Spring Forecasting Experiments and input from Curtis Alexander, Adam Clark, Michael Coniglio, Israel Jirak, Jonathan Kahl, Sergey Kravtsov, Paul Roebber, David Walters, and Steven Weiss. We also acknowledge helpful feedback and insight from Chief Editor Gary Lackmann and three anonymous reviewers.

REFERENCES

Benjamin, S. G., 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.

Carlson, T. N., and F. H. Ludlam, 1968: Conditions for the formation of severe local storms. *Tellus*, **20**, 203–226, https://doi.org/10.3402/tellusa.v20i2.10002.

Chen, F., and J. Dudhia, 2001: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Wea. Rev.*, **129**, 569–585, https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2.

Clark, A. J., and Coauthors, 2012: An overview of the 2010 Hazardous Weather Testbed Experimental Forecast Program Spring Experiment. *Bull. Amer. Meteor. Soc.*, **93**, 55–74, https://doi.org/10.1175/BAMS-D-11-00040.1.

——, 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 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 environments. *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.

D'Agostino, R. B., 1971: An omnibus test of normality for moderate and large sample size. *Biometrika*, **58**, 341–348, https://doi.org/10.1093/biomet/58.2.341.

——, and E. S. Pearson, 1973: Tests for departure from normality. Empirical results for the distributions of b2 and √b. *Biometrika*, **60**, 613–622, https://doi.org/10.1093/biomet/60.3.613.

Doswell, C. A., III, and E. N. Rasmussen, 1994: The effect of neglecting the virtual temperature correction on CAPE calculations. *Wea. Forecasting*, **9**, 625–629, https://doi.org/10.1175/1520-0434(1994)009<0625:TEONTV>2.0.CO;2.

Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077–3107, https://doi.org/10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2.

Evans, C., S. J. Weiss, I. L. Jirak, A. R. Dean, and D. S. Nevius, 2018: An evaluation of paired regional/convection-allowing forecast vertical thermodynamic profiles in warm-season, thunderstorm-supporting environments. *Wea. Forecasting*, **33**, 1547–1566, https://doi.org/10.1175/WAF-D-18-0124.1.

Farrell, R. J., and T. N. Carlson, 1989: Evidence for the role of the lid and underrunning in an outbreak of tornadic thunderstorms. *Mon. Wea. Rev.*, **117**, 857–871, https://doi.org/10.1175/1520-0493(1989)117<0857:EFTROT>2.0.CO;2.

Gallo, B. T., and Coauthors, 2017: Breaking new ground in severe weather prediction: The 2015 NOAA/Hazardous Weather Testbed Spring Forecasting Experiment. *Wea. Forecasting*, **32**, 1541–1568, https://doi.org/10.1175/WAF-D-16-0178.1.

Hacker, J. P., and D. Rostkier-Edelstein, 2007: PBL state estimation with surface observations, a column model, and an ensemble filter. *Mon. Wea. Rev.*, **135**, 2958–2972, https://doi.org/10.1175/MWR3443.1.

——, J. L. Anderson, and M. Pagowski, 2007: Improved vertical covariance estimates for ensemble-filter assimilation of near-surface observations. *Mon. Wea. Rev.*, **135**, 1021–1036, https://doi.org/10.1175/MWR3333.1.

Hong, S.-Y., and J.-O. J. Lim, 2006: The WRF single-moment 6-class microphysics scheme (WSM6). *J. Korean Meteor. Soc.*, **42**, 129–151.

——, Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341, https://doi.org/10.1175/MWR3199.1.

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.

Ikeda, K., M. Steiner, and G. Thompson, 2017: Examination of mixed-phase precipitation forecasts from the High-Resolution Rapid Refresh model using surface observations and sounding data. *Wea. Forecasting*, **32**, 949–967, https://doi.org/10.1175/WAF-D-16-0171.1.

James, E. P., C. R. Alexander, and J. M. Brown, 2018: Testing the impact of additional vertical levels for the 3km High-Resolution Rapid Refresh. *25th Conf. on Numerical Weather Prediction*, Amer. Meteor. Soc., Denver, CO, 12B.3, https://ams.confex.com/ams/29WAF25NWP/webprogram/Paper344731.html.

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.

——, 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.

Jirak, I., A. Clark, J. Correia, K. Knopfmeier, C. Melick, B. Twiest, M. Coniglio, and S. Weiss, 2015: Spring Forecasting Experiment 2015: Preliminary findings and results. NOAA/NSSL/Storm Prediction Center, 32 pp., http://hwt.nssl.noaa.gov/Spring_2015/HWT_SFE_2015_Prelim_Findings_Final.pdf.

Kain, J. S., P. R. Janish, S. J. Weiss, M. E. Baldwin, R. S. Schneider, and H. E. Brooks, 2003: Collaboration between forecasters and research scientists at the NSSL and SPC: The Spring Program. Bull. Amer. Meteor. Soc., 84, 1797–1806, https://doi.org/10.1175/BAMS-84-12-1797.

——, 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.

Knievel, J. C., G. H. Bryan, and J. P. Hacker, 2007: Explicit numerical diffusion in the WRF Model. Mon. Wea. Rev., 135, 3808–3824, https://doi.org/10.1175/2007MWR2100.1.

Ladwig, W., 2017: Wrf-python (version 1.1.1). NCAR, https://doi.org/10.5065/D6W094P1.

Lanicci, J. M., and T. T. Warner, 1991: A synoptic climatology of the elevated mixed-layer inversion over the southern Great Plains in spring. Part I: Structure, dynamics, and seasonal evolution. Wea. Forecasting, 6, 181–197, https://doi.org/10.1175/1520-0434(1991)006<0181:ASCOTE>2.0.CO;2.

Lin, Y., and K. E. Mitchell, 2005: The NCEP Stage II/IV hourly precipitation analyses: Development and applications. 19th Conf. on Hydrology, San Diego, CA, Amer. Meteor. Soc., 1.2, https://ams.confex.com/ams/Annual2005/techprogram/paper_83847.htm.

May, R. M., S. C. Arms, P. Marsh, E. Bruning, and J. R. Leeman, 2017: MetPy: A Python package for meteorological data. Unidata, https://doi.org/10.5065/D6WW7G29.

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, 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.

NCEP, 2018: North American Mesoscale (NAM) analysis and forecast system characteristics. National Centers for Environmental Prediction, 9 pp., http://www.emc.ncep.noaa.gov/mmb/mmbpll/misc/NAM_2017.pdf.

Pleim, J. E., 2007: A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: Model description and testing. J. Appl. Meteor. Climatol., 46, 1383–1395, https://doi.org/10.1175/JAM2539.1.

Powers, J. G., J. B. Klemp, W. C. Skamarock, C. A. Davis, J. Dudhia, D. O. Gill, J. L. Coen, and D. J. Gochis, 2017: The Weather Research and Forecasting Model: Overview, system efforts, and future directions. Bull. Amer. Meteor. Soc., 98, 1717–1737, https://doi.org/10.1175/BAMS-D-15-00308.1.

Skamarock, W. C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. Mon. Wea. Rev., 132, 3019–3032, https://doi.org/10.1175/MWR2830.1.

——, and M. L. Weisman, 2009: The impact of positive-definite moisture transport on NWP precipitation forecasts. Mon. Wea. Rev., 137, 488–494, https://doi.org/10.1175/2008MWR2583.1.

——, 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.

——, S.-H. Park, J. B. Klemp, and C. Snyder, 2014: Atmospheric kinetic energy spectra from global high-resolution non-hydrostatic simulations. J. Atmos. Sci., 71, 4369–4381, https://doi.org/10.1175/JAS-D-14-0114.1.

Sukoriansky, S., B. Galperian, and V. Perov, 2005: Application of a new spectral theory of stable stratified turbulence to the atmospheric boundary layer over sea ice. Bound.-Layer Meteor., 117, 231–257, https://doi.org/10.1007/s10546-004-6848-4.

Tewari, M. F., and Coauthors, 2004: Implementation and verification of the unified Noah land surface model in the WRF Model. 20th Conf. on Weather Analysis and Forecasting/16th Conf. on Numerical Weather Prediction, Amer. Meteor. Soc., Seattle, WA, 17, https://ams.confex.com/ams/84Annual/techprogram/paper_69061.htm.

Trier, S. B., C. A. Davis, D. A. Ahijevych, and K. W. Manning, 2014: Use of the parcel buoyancy minimum (B_{min}) to diagnose simulated thermodynamic destabilization. Part I: Methodology and case studies of MCS initiation environments. Mon. Wea. Rev., 142, 945–966, https://doi.org/10.1175/MWR-D-13-00272.1.

Walters, D. N., and Coauthors, 2014: The Met Office Unified Model global atmosphere 4.0 and JULES global land 4.0 configurations. Geosci. Model Dev., 7, 361–386, https://doi.org/10.5194/gmd-7-361-2014.

Warner, T. T., 2011: Numerical Weather and Climate Prediction. Cambridge University Press, 550 pp.

Weiss, S., J. Kain, D. Bright, J. Levit, G. Carbin, and M. Coniglio, 2007: Experimental Forecast Program Spring Experiment 2007: Program overview and operations plan. NOAA/NSSL/Storm Prediction Center, 62 pp., https://hwt.nssl.noaa.gov/sfe/2007/opsplan/Spring_Experiment_2007_ops_plan_May13.pdf.

Wicker, L. J., and W. C. Skamarock, 2002: Time-splitting methods for elastic models using forward time schemes. Mon. Wea. Rev., 130, 2088–2097, https://doi.org/10.1175/1520-0493(2002)130<2088:TSMFEM>2.0.CO;2.

Wood, N., and Coauthors, 2014: An inherently mass-conserving semi-Lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. Quart. J. Roy. Meteor. Soc., 140, 1505–1520, https://doi.org/10.1002/qj.2235.