Influences of Horizontal Grid Spacing and Microphysics on WRF Forecasts of Convective Morphology Evolution for Nocturnal MCSs in Weakly Forced Environments

JONATHAN E. THIELEN AND WILLIAM A. GALLUS JR.

Department of Geological and Atmospheric Sciences, Iowa State University, Ames, Iowa

(Manuscript received 21 December 2018, in final form 10 June 2019)

ABSTRACT

Nocturnal mesoscale convective systems (MCSs) are important phenomena because of their contributions to warm-season precipitation and association with severe hazards. Past studies have shown that their morphology remains poorly forecast in current convection-allowing models operating at 3–4-km horizontal grid spacing. A total of 10 MCS cases occurring in weakly forced environments were simulated using the Weather Research and Forecasting (WRF) Model at 3- and 1-km horizontal grid spacings to investigate the impact of increased resolution on forecasts of convective morphology and its evolution. These simulations were conducted using four microphysics schemes to account for additional sensitivities to the microphysical parameterization. The observed and corresponding simulated systems were manually classified into detailed cellular and linear modes, and the overall morphology depiction and the forecast accuracy of each model configuration were evaluated. In agreement with past studies, WRF was found to underpredict the occurrence of linear modes and overpredict cellular modes at 3-km horizontal grid spacing with all microphysics schemes tested. When grid spacing was reduced to 1 km, the proportion of linear systems increased. However, the increase was insufficient to match observations throughout the evolution of the systems, and the accuracy scores showed no statistically significant improvement. This suggests that the additional linear modes may have occurred in the wrong subtypes, wrong systems, and/or at the wrong times. Accuracy scores were also shown to decrease with forecast length, with the primary decrease in score generally occurring during upscale growth in the early nocturnal period.

1. Introduction

Mesoscale convective systems (MCSs) play a crucial role climatologically in precipitation across the central United States. These systems account for roughly 30%–70% of the precipitation that occurs during the April–September period (warm season) in this region (Ashley et al. 2003) and are therefore key phenomena of interest when seeking to improve the quantitative precipitation forecast (QPF) skill of models (Fritsch et al. 1986). While this rainfall is essential to agricultural production in the central United States, MCSs are also associated with numerous and widespread severe weather risks such as flooding, hail, wind, and tornadoes (Ashley and Mote 2005; Jirak and Cotton 2007; Schoen and Ashley 2011; Smith et al. 2012). These severe hazards are particularly notable because many MCSs occur at night (Haberlie and Ashley 2019) when the general public is less alert and potentially more susceptible to harm. Historically, these systems have been poorly forecast (Jirak and Cotton 2007), which has led to intense observational and modeling studies in recent years (e.g., Bryan and Morrison 2012; Lebo and Morrison 2015; Squitieri and Gallus 2016a,b; Campbell et al. 2017; Geerts et al. 2017; Schumacher and Peters 2017; Haberlie and Ashley 2018; Carlberg et al. 2018). As a result, noticeable progress has been made in the prediction of MCSs through the improvement of convection-allowing models (CAMs) (Gallo et al. 2017); however, there remain substantial areas for improvement, particularly in the depiction of finer-scale details of system morphology (Snively and Gallus 2014, hereafter SG14; Gallo et al. 2017; Carlberg et al. 2018).

Corresponding author: Jonathan E. Thielen, jthielen@iastate.edu

DOI: 10.1175/WAF-D-18-0210.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).
The development and life cycle of nocturnal MCSs are closely associated with the low-level jet (LLJ) that frequently develops at night over the Great Plains during the spring and summer months (Pitchford and London 1962; Bonner 1968; Augustine and Caracena 1994; French and Parker 2010). Squitieri and Gallus (2016a) noted that, based on differences in synoptic background conditions, LLJs can be divided into two categories: Type C, with cyclonic upper-level flow and coupling with an upper-level jet streak, and Type A, with anticyclonic upper-level flow and no coupling. In Squitieri and Gallus (2016a), Type-C LLJs are considered the cases with strong synoptic forcing since the coupling with upper-level flow allows for the greatest large-scale vertical motion, whereas Type-A LLJs are labeled as weakly forced. However, other forms of strong forcing may still exist in Type-A situations, such as warm advection (Maddox 1983) or deformation leading to frontogenesis (Augustine and Caracena 1994), but these mechanisms are restricted to the lower troposphere. Still, important differences between Type-C and Type-A cases were found with respect to associations of LLJ and MCS forecast errors, and the Type-A cases were identified to be in particular need of further study (Squitieri and Gallus 2016b, hereafter SG16b).

When studying MCSs, their morphology—shape, organization, and structure—becomes an important factor to consider, as it reflects the dynamical processes of the systems and is strongly associated with potential hazards that the system could produce (Houze et al. 1990; Bluestein and Jain 1985). In particular, cellular types are most strongly associated with hail and tornadoes, whereas linear modes give all types of severe weather depending on the exact system mode, but with wind and flooding threats being of particular note (Gallus et al. 2008, hereafter G08; Smith et al. 2012). While the exact details of convective morphology classification procedures differ between studies (e.g., Fowle and Roebber 2003; Done et al. 2004; Grams et al. 2006; Duda and Gallus 2010; Smith et al. 2012), they most commonly rely upon subjective analysis of radar imagery to determine organizational patterns (G08; Smith et al. 2012; Haberlie and Ashley 2018).

Because different convective modes can be associated with different severe weather threats, it is important to understand how well models simulate modes (Fowle and Roebber 2003; Done et al. 2004; Grams et al. 2006) and the evolution of modes (SG14). As the horizontal grid spacing used within CAMs decreases, the types of detailed structures simulated in convective systems become increasingly similar to the kind observed on radar (Clark et al. 2012), allowing simulated reflectivity output from the models to be used to forecast convective morphology. However, past studies have called into question the accuracy of such forecasts and the potential benefits of resolving these detailed convective structures in CAM output (Kain et al. 2008; Schwartz et al. 2009; Clark et al. 2012). For example, SG14 found that the Advanced Research version of WRF, when running at 3-km horizontal grid spacing, performed well in simulating cellular systems, but performed more poorly with linear systems, especially bow echoes and squall lines with trailing stratiform rain regions that were markedly underforecast. Their finding regarding poor simulation of stratiform rain regions agrees with past modeling studies (Fovell and Ogura 1988; McCumber et al. 1991; Lang et al. 2003; Gallus and Pfeifer 2008; Morrison et al. 2009b).

Although the use of typical CAM horizontal grid spacings of a few kilometers can lead to simulated convective structures similar to those observed, albeit not without some inaccuracies, some studies have indicated that important processes related to convection may not be adequately simulated until grid spacing is reduced further (e.g., Weisman et al. 1997; Bryan et al. 2003; Potvin and Flora 2015). For instance, the horizontal grid spacing used has been found to have a statistically significant influence on the cold pools of modeled systems, and thereby their morphologies (Bryan and Morrison 2012; Squitieri and Gallus 2019). However, some studies have urged caution with reducing grid spacing to certain levels. For instance, Schumacher (2015) demonstrated that, for the destructive tornado/flash flood case of 31 May–1 June 2013 in central Oklahoma, a 4-km grid spacing simulation performed best, and that increased-resolution runs experienced degraded performance. This was due to the planetary boundary layer schemes operating with grid scales of the same order as the turbulent motions, a situation Wyngaard (2004) terms the “terra incognita.” In accord with this, Ching et al. (2014) explained that largest eddies in the planetary boundary layer, convectively induced secondary circulations (with horizontal wavelengths on the order of 2–10 km), cannot be reliably simulated at these grid scales. Work investigating the accuracy of simulations of mesoscale convection at these reduced grid spacings has been ongoing (e.g., Schwartz et al. 2009, 2017; Squitieri and Gallus 2019), and this present study further examines the influence of grid spacing on MCS morphology.

In addition, since stratiform regions are dependent upon both the advection of hydrometeors away from the intense convective cores and to hydrometeor production and growth mechanisms in the mesoscale updrafts within the stratiform regions (Rutledge and Houze 1987; Gallus and Johnson 1995a,b; Parker and
Johnson 2004; Parker 2007), it seems likely that convective morphology and its evolution would be sensitive to the microphysical parameterizations used (Bryan and Morrison 2012; Cintineo et al. 2014; Clark et al. 2014; Grasso et al. 2014). For instance, in a study of a German squall line, even though all microphysical schemes tested were found to overpredict reflectivity intensity but underpredict areal coverage of lower values, especially in the stratiform rain regions, the degrees differed (Gallus and Pfeifer 2008). Also, ice microphysics are well known to exert a substantial influence on the simulated stratiform regions (e.g., Fovell and Ogura 1988; Liu et al. 1997; Gilmore et al. 2004). Additionally, alterations of the details of graupel behavior within a single microphysics parameterization scheme have been shown to cause changes in the cold pool characteristics and bowing behavior of a modeled convective system, thus indicating higher potential for forecast errors as a result of the microphysical sensitivity (Adams-Selin et al. 2013). It is likely that the sensitivities may be greater for MCSs occurring in weakly forced environments since the evolution may be more sensitive to subtle differences in smaller-scale forcing mechanisms (SG16b).

The present study examines the impacts that reducing horizontal grid spacing from 3 to 1 km and utilizing four different microphysics schemes have on the skill of the WRF model in predicting convective morphology and stratiform region evolution for nocturnal MCSs in weakly forced environments. We follow the classification procedures outlined by G08 and SG14 and then evaluate the performance of the model simulations of these nocturnal MCSs using comparisons of mode distributions and a quantitative accuracy score. Section 2 explains the types of cases, morphology classification scheme, scores, and model configurations used in the study. Analysis and results then follow in section 3, with section 4 presenting the conclusions, summary, and directions for future work.

2. Methodology

An initial sample of five nocturnal MCS events with Type-A LLJs present was randomly selected from the 15 Type-A cases of SG16b occurring during 2010–13. Five additional nocturnal Type-A cases were added from the Plains Elevated Convection At Night (PECAN) field project occurring 1 June–15 July 2015 (Geerts et al. 2017) to allow for additional comparisons with the more dense dataset of radar and in situ measurements from the project in future work (Fig. 1). All 10 of the cases produced some severe weather, with damaging wind being the most common storm report. In accord with Squitieri and Gallus (2016a), a “nocturnal MCS” refers to an MCS that reached maturity in the overnight hours (0200–1100 UTC). Type-A LLJs were identified as in Squitieri and Gallus (2016a); they occur when a southerly LLJ is present at 900 hPa, but weak anticyclonic flow exists aloft with little or no coupling of the LLJ with an upper-level jet at 200 hPa (Uccellini and Johnson 1979). Therefore, Type-A cases can be thought of as having weak synoptic

![Fig. 1. Overview of the evolution of observed system mode for the 10 cases of the study. At each hour, the observed mode at that hour is shown by a colored interval centered at that hour. Local storm reports associated with the system are indicated above the mode interval at the time of occurrence (tornado: red triangle, hail: green diamond, wind: blue square).](image-url)
forcing. The identification of Type-A LLJs was performed via streamline analysis using RAP/RUC analyses (NOAA/National Climatic Data Center 2015b; Benjamin et al. 2016).

The present study utilized version 3.6.1 of the WRF-ARW (Skamarock and Klemp 2008) with runs integrated 24 h. The runs were initialized at 1200 UTC prior to system development, because, for CAMs initialized only from coarser model initial and lateral boundary conditions, it typically takes a number of hours of forward integration before coherent precipitation systems develop in the model (e.g., Kain et al. 2010). These runs were initialized using 12-km NAM forecast output (NOAA/National Climatic Data Center 2015a), with boundary conditions from the NAM output on 6-hourly intervals, in accord with SG16b. A variable size and relocatable 1-km domain was placed with one-way nesting within a 3-km domain, as described in further detail below, with a vertical grid consisting of 50 manually specified eta levels, with 25 below and 25 above 850 hPa, as in Squitieri and Gallus (2016a). The runs used in this study employed the MYJ planetary boundary layer (Janjic´ 1994), Dudhia shortwave radiation (Dudhia 1989), and RRTM longwave radiation (Mlawer et al. 1997) schemes. The radar reflectivity factor was turned on to allow for direct evaluation of simulated reflectivity from the hourly output produced by the model.

Due to the relatively limited computational resources available for this high-resolution study and the wide geographic distribution of the MCSs studied, variable-size, relocatable domains were chosen for each of the 10 cases, with the inner 1-km domain of the nested configuration chosen to fit observed system tracks with a buffer of roughly 100 km along the lateral edges (Fig. 2). The respective 3-km parent domains were centered at the same points, but with twice the extent in the north–south and east–west directions. It is the output from these respective 3- and 1-km domains that are compared in the present study. While some degree of error from these constrained lateral boundary conditions is unavoidable (Warner et al. 1997), the sizes of the inner 1-km domains used in this study are comparable to sizes of the 3-km domains of SG14, and the sizes of the outer 3-km domains are generally as large or larger than the 3-km domain of SG16b, so that the errors involved should be comparable to those of past studies.

Four different microphysics schemes were utilized to evaluate the impact of horizontal grid spacing variability over a range of microphysical schemes. These included the partially double-moment Thompson (Thompson et al. 2008), single-moment WSM6 (Hong and Lim 2006), and fully double-moment Morrison (Morrison et al. 2009a) schemes as present in WRF version 3.6.1, along with a modified version of WSM6 where graupel fall speed parameters were altered (intercept $n_{OG} = 4 \times 10^2 m^{-4}$ and density $G = 900 kg m^{-3}$) to make graupel behave more like hail (Adams-Selin et al. 2013).

To determine the morphology of the simulated convective systems, the composite (or column maximum) reflectivity from the model output was used to manually classify the convective mode according to the 10-category scheme of SG14 (which itself is based on the 9-category scheme of G08, Fig. 3). This scheme differentiates systems into three broad types: cellular, linear, and nonlinear (NL). The cellular category is further refined into three cellular subtypes: isolated cells (IC), clusters of cells (CC), and broken lines (BL). While other studies (e.g., Bluestein and Jain 1985) have considered broken lines to be a linear mode, G08 considered it a cellular mode because the severe weather reports associated with them are “more dependent on storm-scale dynamics than on mesoscale organization.” There are also five subtypes of linear systems: lines without a stratiform precipitation region (NS), bow echoes (BE), and lines with leading (LS), parallel (PS), and trailing (TS) stratiform regions (Parker and Johnson 2000). The 10th mode, mixed complex (MC), was reserved for systems that could not be adequately classified into one of the other nine categories due to simultaneously exhibiting characteristics of cellular, nonlinear, and linear modes. Observed systems were classified by the same manual method using the column-maximum reflectivity from the hourly GridRad 3D gridded NEXRAD product (Bowman and Homeyer 2017). While subjectivity is inherent in any storm mode
classification due to its nature as a manual process discriminating between closely related categories (Smith et al. 2012), the guidelines of past studies using this same procedure (G08; Duda and Gallus 2010; SG14; Carlberg et al. 2018) were closely followed to maintain as much consistency and reliability in the classification as possible. The first instance of 40-dBZ reflectivity over an area of at least $6 \times 6$ km defined convective initiation, and the interval over which this criterion was maintained defined the duration of the system. In this way, the system duration includes the MCS stage and any pre- or post-MCS stage in the system evolution from initiation to dissipation (or end of the analysis period at 1200 UTC), in accord with SG14. To receive a linear classification, a system must have attained a convective ($\geq 40$ dBZ) region length of 75 km with a 3:1 length-to-width ratio. The threshold for stratiform regions was reflectivity of at least 30 dBZ existing over an area at least twice as wide as the adjacent convective line. To ensure temporal continuity, the characteristics of a mode must be present for at least 2 h to be classified as that mode. To fit this criterion, if hour-to-hour variations in characteristics were present over some interval, the single convective mode most representative of the system over that interval was chosen (Fig. 4). As in SG14, the centroid of modeled systems must have been within 300 km of the corresponding observed system to be classified. If the centroids of more than one discrete system in a model run lied within this region, the modeled system in closest proximity to the observed system was selected as the matching system.

To evaluate the model depictions of convective morphology, the distributions of mode occurrence in the 10-case

**Fig. 3.** The nine primary convective modes used (schematic taken from G08). These modes include the three cellular modes [isolated cells (IC), clusters of cells (CC), and broken lines (BL)], five linear modes [lines without a stratiform precipitation region (NS), bow echoes (BE), and lines with leading (LS), parallel (PS), and trailing (TS) stratiform regions], and a nonlinear mode (NL). As in SG14, a 10th mode, mixed complex (MC) was added.
sample were compared between the model configurations and observations, including creation of heat maps showing mode associations between 3- and 1-km forecasts. Simultaneous Fisher’s exact tests (Fisher 1935) were performed to determine if the distributions of modes for any of the eight model configurations differed in a statistically significant way from the observed distribution of modes or if any of the microphysics scheme or grid spacing changes resulted in statistically significant differences in distributions of modes (exact tests were used in place of $\chi^2$ tests for homogeneity due to low bin counts for some modes). Since a large number of simultaneous tests were conducted, Bonferroni correction was applied to each family of tests to control for the Type I error rate (i.e., to control for the probability of concluding a false positive) (Mendenhall and Sincich 2007). To evaluate how the distribution of modes changes over time, the proportions of modes occurring at each time throughout the duration of the system along a normalized time scale were compared (Figs. 5a,b). In addition, the convective mode accuracy score of SG14 was used to assess the forecast performance of the model runs. This score ranges from 0.0 to 1.0 for each model run, and was determined by aligning the evolution of morphology seen in observations and the model run onto a normalized 0.0–1.0 time scale, and then making an evaluation at each normalized time interval (Fig. 5c). If the forecasted mode was a perfect

![Observed Composite Reflectivity - June 2-3, 2015](image)

**FIG. 4.** Example of the evolution of observed system mode for the 2–3 Jun 2015 case that demonstrates the effect of the temporal continuity constraint. The BL mode present at 2300 and 0000 UTC transitions to BE by 0200 UTC (which is maintained at 0300 UTC). During the transition, at 0100 UTC, the system most closely resembles an NS mode. However, due to the requirement that the characteristics of a mode must be present for at least two consecutive hours, the system cannot receive the NS label at 0100 UTC. It instead receives the BE label, since it more closely resembles a BE rather than BL system at that time.

![Graphs and charts](image)

**FIG. 5.** Illustration of the time normalization and accuracy scoring procedure of SG14 as used in this study. (a) The observed and modeled (Thompson 3 km run) evolutions of mode are shown over real time. (b) These modes are shown scaled to the normalized 0.0–1.0 time scale. (c) Time intervals (on the normalized scale) where an exact match occurred (shaded dark gray) receive a score component of 1.0 and those where a category match occurred (shaded light gray) receive a score component of 0.5. The accuracy score in this example would be 0.503.
match, a 1.0 score was awarded for that time. If the forecasted mode was a category match, in that it correctly identified cellular, linear, or nonlinear types but did not match exactly a subtype, a score of 0.5 was instead awarded. A nonmatch received a 0.0 score. In the present study (with the WSM6 modified microphysics) occurred with such a timing error. The scores over the system’s evolution were then summed to calculate the total accuracy score $S$ as

$$S = \sum_{i=1}^{N} M_i \Delta t_i,$$

where $N$ is the number of time intervals for comparison and $M_i$ and $\Delta t_i$ are the score weight and normalized time interval length for the $i$th interval, respectively. After collecting the results from each of the 10 cases, bootstrapped paired $t$ tests (Mendenhall and Sincich 2007) were used to determine statistical significance of differences in mean score between 3- and 1-km grid spacing for each microphysics scheme. Through numerical resampling, bootstrapping provides a more robust estimate of uncertainty in these paired $t$ tests when compared to a standard paired $t$ test given the relatively small sample size of 10 cases. Also, the individual score weights (as a function of normalized time) from each of the cases were aggregated over each model configuration to assess the overall model accuracy in forecasting convective morphology over the evolution of the system.

Additionally, two quantitative morphological parameters (areal coverage and average reflectivity) were analyzed for systems with stratiform or bow echo modes to supplement the categorical analysis of convective morphology. To obtain these parameters, image regions describing the systems were extracted using scikit-image (van der Walt et al. 2014) according to the automated segmentation procedure of Haberlie and Ashley (2018). In this procedure, all convective regions ($\geq 40$ dBZ) containing intense cells ($\geq 50$ dBZ in at least one pixel) were identified and merged according to a convective search radius (48 km in the present study) to form “MCS cores.” Then, all adjacent stratiform regions

![Fig. 6. Overview of the three steps of “MCS slice” extraction: identifying intense cells, merging into “MCS cores” according to the convective search radius, and merging with adjacent stratiform regions into “MCS slices” according to the stratiform search radius. [Figure reproduced from Fig. 3 of Haberlie and Ashley (2018).]
were joined according to a stratiform search radius (192 km in the present study) to form the “MCS slices” (Fig. 6). For both search radii parameters, the largest values examined in Haberlie and Ashley (2018) were used to ensure the full extent of each manually identified system was selected by the automated procedure. In addition, even though the manual classification procedure of this study used a minimum stratiform reflectivity threshold of 30 dBZ, the stratiform threshold of $\geq 20$ dBZ was maintained in this quantitative analysis for consistency with Haberlie and Ashley (2018). The extracted MCS slices obtained in this way from composite reflectivity data were then used to determine the morphological parameters of both the observed and modeled systems. To ensure consistency of system area, the GridRad data were regridded from 0.02° latitude × 0.02° longitude grid to a 2-km Lambert Conformal grid using nearest-neighbor interpolation. Also, to assess differences in vertical structure among the model configurations, reflectivity was analyzed on vertical levels at each 1 km in altitude (above sea level) from 1 to 10 km so that MCS slice objects could be extracted for comparisons of system area (such an analysis could not be conducted with observed reflectivity due to gaps in radar coverage at the lowest and highest levels).

3. Results

a. Distributions of modes

Distributions of predicted convective modes from the four WRF configurations at both 3- and 1-km horizontal grid spacing are compared to the observed mode distribution in Fig. 7. In the model runs with 3-km grid spacing, linear modes were underpredicted with TS being most underpredicted, in agreement with past studies at the same grid spacing (SG14; Carlberg et al. 2018). With all four microphysics schemes, this underprediction in linear modes had a corresponding overprediction in CC, with WSM6 also overpredicting BL. Within the 3-km models, Thompson had the highest proportion of TS (while still being less than half of the observed proportion), and Morrison had the highest proportion of BE (although still slightly less than observed). In comparison, past work found BE to be the mode most severely underpredicted in 3-km WRF simulations (SG14). Also, the WSM6 scheme with graupel modified to behave more like hail predicted almost exclusively cellular modes at 3-km grid spacing, with only 3.3% of predicted modes being a linear mode (always NS), in contrast with 47% of observed systems being linear. Overall, Thompson failed to show NL and NS,
TABLE 1. The $p$ values from Monte Carlo-simulated Fisher’s exact tests (106 replicates) on $10 \times 2$ mode comparison tables used to test differences of categorical distributions of modes. Significance assessed at the $\alpha = 0.05$ confidence level with Bonferroni correction for this family of tests (individual tests assessed at 0.05/12 = 0.0042 level), with significant results italicized (all differences in distributions found statistically significant).

<table>
<thead>
<tr>
<th>Comparison between</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations Thompson 3 km</td>
<td>$10^{-6}$</td>
</tr>
<tr>
<td>Observations WSM6 3 km</td>
<td>$10^{-6}$</td>
</tr>
<tr>
<td>Observations WSM6 Modified 3 km</td>
<td>$10^{-6}$</td>
</tr>
<tr>
<td>Observations Morrison 3 km</td>
<td>$10^{-6}$</td>
</tr>
<tr>
<td>Observations Thompson 1 km</td>
<td>$10^{-6}$</td>
</tr>
<tr>
<td>Observations WSM6 1 km</td>
<td>$10^{-6}$</td>
</tr>
<tr>
<td>Observations WSM6 Modified 1 km</td>
<td>$10^{-6}$</td>
</tr>
<tr>
<td>Observations Morrison 1 km</td>
<td>0.001 211</td>
</tr>
<tr>
<td>Thompson 3 km Other</td>
<td>0.000 235</td>
</tr>
<tr>
<td>WSM6 3 km</td>
<td>0.000 041</td>
</tr>
<tr>
<td>WSM6 Modified 3 km</td>
<td>0.000 002</td>
</tr>
<tr>
<td>Morrison 3 km Other</td>
<td>0.000 214</td>
</tr>
</tbody>
</table>

Whereas both WSM6 schemes failed to produce any lines with stratiform precipitation.

When grid spacing was reduced from 3 to 1 km, all schemes produced an increase in the occurrence of linear modes, which generally lessened but did not completely remove the underprediction of linear modes. It should be mentioned, however, that the total number of simulated BE events improved markedly in all configurations except those using Morrison (which already had only a small underestimate with 3-km grid spacing). Because the BE mode is frequently associated with a large number of severe wind and hail reports (Przybylinski 1995; G08), the better agreement in the number of BE events when 1-km grid spacing was used versus 3-km grid spacing is initially encouraging. However, this result is based solely on proportion of occurrence, not forecast accuracy (which is explored further in section 3c). Thompson, WSM6, and WSM6 Modified saw a roughly corresponding decrease in IC and CC occurrence, and Morrison saw a corresponding decrease in IC and CC; however, the overprediction of CC remained in all schemes. Although BL was the most common cellular morphology observed, in both the 1 and 3-km runs CC was the most common cellular morphology, except for the 3-km WSM6 runs when CC was equal to BL. Thus, the reduction of grid spacing did not eliminate this shortcoming. The occurrence of NL also increased in all schemes, particularly in WSM6 where the proportion more than doubled from 7.2% to 17.8%. This proportion, along with that for NL events in the Morrison runs, was nearly double the observed proportion. WSM6 Modified at a grid spacing of 1 km predicted many more linear modes than at 3 km, whereas both WSM6 schemes failed to produce any linear modes, thus agreeing better with the other configurations. However, both Thompson and Morrison still had more TS and BE than either WSM6 or WSM6 Modified, thereby more closely resembling the observed proportions. Fisher’s exact tests comparing the categorical distributions of modes indicated that, for each of the eight model configurations, respectively, the modeled distribution of modes differed in a statistically significant way from the observed distribution of modes, and, for each of the four microphysics schemes, the 3-km grid spacing modeled distribution differed significantly from the matching 1-km distribution (Table 1). Also, at each of the two grid spacings, the distributions across the microphysics schemes differed in a statistically significant way (Table 2). These results provide statistical confidence that all of the distributions of

TABLE 2. As in Table 1, but on $10 \times 4$ mode comparison tables used to test differences of categorical distributions of modes between the four microphysics schemes at a given grid spacing. Significance assessed at the $\alpha = 0.05$ confidence level with Bonferroni correction for this family of tests (individual tests assessed at 0.05/2 = 0.025 level), with significant results italicized (both differences in distributions found statistically significant).

<table>
<thead>
<tr>
<th>Comparison at grid spacing of</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 km</td>
<td>$10^{-6}$</td>
</tr>
<tr>
<td>1 km</td>
<td>$10^{-6}$</td>
</tr>
</tbody>
</table>

it was unable to predict any linear modes, thus agreeing better with the other configurations. However, both Thompson and Morrison still had more TS and BE than either WSM6 or WSM6 Modified, thereby more closely resembling the observed proportions. Fisher’s exact tests comparing the categorical distributions of modes indicated that, for each of the eight model configurations, respectively, the modeled distribution of modes differed in a statistically significant way from the observed distribution of modes, and, for each of the four microphysics schemes, the 3-km grid spacing modeled distribution differed significantly from the matching 1-km distribution (Table 1). Also, at each of the two grid spacings, the distributions across the microphysics schemes differed in a statistically significant way (Table 2). These results provide statistical confidence that all of the distributions of

FIG. 8. Association heat map of convective mode predicted by the 3-km grid spacing model configurations (y axis) compared with the corresponding 1-km-predicted mode (x axis). Exact matches are shown along the diagonal from lower left to upper right, with mismatches off the diagonal. Counts of hourly depicted modes are given in each bin by number and shading when such a match or mismatch occurred. Column totals are located on the top row and row totals are located on the right column. Here, and in the following figures, “—” denotes the absence of a classifiable system in the model.
modes truly differ, rather than differing simply due to random chance.

These changes in convective mode depiction from 3- to 1-km grid spacing can be investigated in greater detail through a heat map of hour-by-hour mode associations (Fig. 8). While this model-to-model comparison is limited in that it cannot not provide insight into the accuracy of the model forecasts compared to observations, it is still useful in contrasting the model behavior between grid spacings. In the heat map, high counts of exact matches along the diagonal indicate that many systems maintained the same classification at both grid spacings, however, numerous off-diagonal counts are also apparent. First, many systems that were classified as cellular in the 3-km model runs were classified as linear in the corresponding 1-km runs, whereas no systems that were linear at 3 km became cellular at 1 km. This indicates that cellular systems becoming linear

---

**Fig. 9.** Observed, 3-km grid spacing modeled, and 1-km grid spacing modeled composite reflectivity for three example cases (with mode labels given in parentheses): (a) 0200 UTC 3 Jun 2015, with Thompson microphysics (observed: BE, 3 km: BL, 1 km: TS); (b) 0300 UTC 13 Jul 2015, with WSM6 Modified microphysics (observed: BE, 3 km: BL, 1 km: BE); and (c) 1100 UTC 11 Jun 2010, with Morrison microphysics (observed: PS, 3 km: CC, 1 km: CC).
as grid spacing was reduced were responsible for the observed increase in linear mode occurrence with increasing horizontal resolution. Of particular note is the high count of the BL to NS bin, indicating broken lines “filling in” with convective-level ($\geq 40$ dBZ) values of reflectivity at reduced grid spacing (Fig. 9a). It is possible that this filling in is related to the better resolution of the strong temperature gradients and upward forcing along cold pool boundaries in the 1-km runs (Squitieri and Gallus 2019). It is also worth noting that despite the fact that BE modes are most similar to TS, with the only usual difference being the convective line curvature that is most often due to a strengthened rear-inflow jet (Weisman 1993; Przybylinski 1995; Wakimoto et al. 2006), the increase in BE modes at 1-km grid spacing (Fig. 7) was not primarily due to more TS modes becoming BE. In fact, more BE modes present at 3 km became TS at 1 km than vice versa. Instead, much of the increase in the occurrence in BE was due to NS, BL, and CC modes at 3 km becoming BE at 1 km (a total of 27 of 42 systems that became BE at 1 km) (e.g., Fig. 9b). And so, the main impact of the higher-resolution grid does not appear to be a strengthening of the rear-inflow jet often present in TS systems to cause them to bow, but instead a more substantial change in the structure of other types of events to create more continuous, bowing lines of intense convection, often with stratiform rain. Finally, when considering the general increase in the occurrence of NL, the high count of the CC to NL bin indicates that this increase in occurrence was generally due to clusters of cells similarly filling in with convection between storm cell centroids.

b. Evolution of modes over time

As previously discussed, the overall proportions of mode occurrence differed at a statistically significant level between observations and the eight model configurations of this study (Fig. 7, Tables 1 and 2).
However, the timing of the mode occurrence is also critical when considering the evolution of a convective system during its life cycle. Using normalized time to depict system evolution (Fig. 10), it can be seen that in both the 3- and 1-km model runs aggregated across microphysics schemes, the overprediction of CC was maintained through time, with 3-km runs having a particularly large overprediction in the 0.5–1.0 normalized time range (which corresponds to the mid- to late nocturnal period) (Fig. 9c). Also, although small in proportion relative to CC, NS was also maintained too long throughout the system evolution, with the issue being more prevalent in 1-km runs than in 3-km runs. Additionally, while BL was depicted with sufficiently high proportions in the prenocturnal (roughly 0.0–0.25) and early nocturnal (roughly 0.25–0.5) periods, the models maintained too high of proportions into the later periods, whereas the observed occurrence of BL dropped off. With TS, which was generally underpredicted by the models, high proportions were
observed in the middle and later time periods, whereas the proportions were small with both model grid spacings. In the middle time period, 1-km runs actually had a slight overprediction of BE. Later on, an underprediction was present in the late nocturnal period that resulted in the slight net underprediction shown earlier in Fig. 7. The change from 3 to 1 km noticeably improved the prediction of NL systems during the latter time periods, in agreement with observations. In summary, the 3-km models in aggregate poorly captured the observed temporal changes in mode occurrence and only picked up on some of the upscale growth that tended to occur in the early nocturnal period for these systems. The 1-km models performed better but still had errors in CC overprediction throughout and linear system underprediction in later periods.

Important variations in trends of mode occurrence also existed between the microphysics schemes used. As shown in Fig. 11, in 3-km grid spacing runs, linear modes happened most often in the Morrison run, especially in the later times of the system evolution. In the later periods when the models predicted linear modes during upscale growth, Thompson emphasized stratiform modes, WSM6 more often NS, and Morrison a
split between NS, TS, and BE. Only Morrison at 3 km showed the diversity of modes observed throughout the system evolution (Fig. 7 top). When grid spacing was reduced to 1 km, more variation in modes developed (Fig. 12). All microphysics schemes had some BE occurrence, particularly during upscale growth in the early nocturnal period. Thompson, WSM6, and WSM6 Modified maintained an overprediction of cellular types during later times; however, Morrison produced a lower proportion of cellular modes and a higher proportion of TS in the mid- to late nocturnal period that agreed better with observations. Given this, the 1-km Morrison runs achieved the greatest similarity with observations, except in the overprediction of NL. In general, it seems some of the greatest variability in the middle and late stages of system evolution was present at 3 km among the microphysics schemes in the depiction of NL, BE and TS modes, and the largest increases in occurrence as grid spacing was reduced occurred with these modes.

c. Morphology forecast accuracy score

While it is important to assess whether or not a convection-allowing model accurately captures the proportions of mode occurrence throughout time, it is even more important for the model to be accurate in any particular forecast. The morphology accuracy score of SG14 aims to capture how well the morphology forecast performs considering the evolution of morphology. When tabulated for the 10 cases and eight model configurations of this study, the morphology accuracy score indicated high run-to-run variation in morphology forecast accuracy with roughly as many cases seeing improvement as those seeing worsening when model resolution was increased (Fig. 13). On average, these morphology forecasts were less accurate than those of SG14, which obtained a mean score of 0.49 for 3-km Thompson simulations that were not restricted to nocturnal systems in weakly forced regimes. Even the best performing configuration, Morrison 1 km, achieved a mean score of 0.486. The large variation in scores among runs also resulted in there being no statistically significant differences in mean accuracy score between 3-km model runs and 1-km runs for any of the four microphysics schemes tested in this study (Table 3). In a practical sense, since roughly as many simulations had increases in accuracy score between 3- and 1-km grid spacings as those that had decreases, this means that any systematic change in accuracy (if it exists) was drowned out by the high run-to-run variability of the simulations. Furthermore, the fact the accuracy scores did not consistently improve despite an increase in linear mode occurrence in 1-km runs suggests that the additional linear modes often occurred in the wrong subtypes, wrong systems, and/or at the wrong times.

![Fig. 13. SG14 morphology accuracy score for 3- (shown on left) and 1-km (shown on right) runs in each case for each of the four microphysics schemes tested: (a) Thompson, (b) WSM6, (c) WSM6 Modified, and (d) Morrison. Lines indicate correspondence between scores for 3- and 1-km grid spacing for each case, but do not indicate a linear change in score as resolution increases.](image-url)

<table>
<thead>
<tr>
<th>Microphysics</th>
<th>Mean score (1-km grid spacing)</th>
<th>Mean score (3-km grid spacing)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thompson</td>
<td>0.426</td>
<td>0.387</td>
<td>0.511 12</td>
</tr>
<tr>
<td>WSM6</td>
<td>0.401</td>
<td>0.430</td>
<td>0.612 04</td>
</tr>
<tr>
<td>WSM6 Modified</td>
<td>0.399</td>
<td>0.371</td>
<td>0.687 78</td>
</tr>
<tr>
<td>Morrison</td>
<td>0.486</td>
<td>0.428</td>
<td>0.441 03</td>
</tr>
</tbody>
</table>
When evaluating the morphology accuracy score across time, the score for all model configurations, on average, was highest and was more accurate than a category match in the prenocturnal period (Fig. 14), but decreased from around normalized time 0.2 to 0.6. This decrease corresponds with the models’ failure to accurately capture the modes that occur during the upscale growth of the observed systems in the early nocturnal period. After reaching a general minimum at around normalized time 0.6 (when overprediction of CC peaked), the average model scores remained below 0.5 (the category match level) with high model-to-model spread. No consistent behavior was present among microphysics schemes as grid spacing was reduced from 3 to 1 km.

d. Quantitative morphology analysis

Given that bow echoes and linear modes with stratiform regions are among the most poorly forecasted modes (SG14; Carlberg et al. 2018) and are associated with some of the highest occurrences of severe hazards (G08), further investigation into the details of the errors of morphology depiction is needed for these modes. One important parameter related to system morphology is the total area covered by an MCS, which for BE and stratiform linear modes is most often dominated by the extent of the stratiform region. As shown in Fig. 15, Thompson (with both 3- and 1-km grid spacings) had the greatest spread of system areas among the model configurations and the largest systems areas, including some systems with greater extent than the largest observed systems. When grid spacing was refined from 3 to 1 km, the primary change was the increase in counts with each microphysics scheme; otherwise, the unimodal distributions were maintained with similar spreads. At 3 km, WSM6 Modified failed to have any BE or stratiform modes, whereas, at 1 km, the distribution of areas was strongly right skewed, implying that the modeled systems were small even if they did have stratiform regions. When compared to observations when matching systems occurred (Fig. 16), the majority of BE/stratiform systems simulated by both the 3- and 1-km Thompson configurations overpredicted the system areal coverage. In contrast, with WSM6 at both 3 and 1 km and with WSM6 Modified at 1 km, a negative bias in areal coverage was observed, and median errors were near zero in both Morrison configurations.

Comparison of predicted and observed average composite reflectivity in BE/linear stratiform systems is shown in Fig. 17. The errors were most often small and negative for Thompson and positive for WSM6 and Morrison at 3-km grid spacing. In the 1-km model runs where there were more predicted BE/linear stratiform systems to compare to observations, there was greater spread in errors for each configuration, but the negative bias remained in Thompson and the positive bias likewise in WSM6 and Morrison. The modified WSM6 configuration at 1 km had less positively biased average reflectivity than the original WSM6 configuration.

FIG. 14. Evolution of the SG14 morphology accuracy score over normalized time for each of the eight model configurations and the aggregated 3- and 1-km simulation scores. Jagged trend lines from the time normalization procedure are smoothed using a Gaussian filter with $\sigma = 15$ for clarity of presentation.
Overall, these biases remained relatively small on average (and within the at most 5-dBZ error expected from GridRad due to its binning and weighting procedure, see Homeyer and Bowman 2017), indicating that, when a model predicted a BE/stratiform linear mode, the system-average reflectivity magnitude was reasonably accurate.

Vertical variations in system extent reveal details related to convective system structure and help identify the origins of the differences in morphology classifications between the model configurations (Fig. 18). With both 3- and 1-km grid spacing, the Thompson runs had a large increase in system area with height with maxima at 9 or 10 km in altitude. These maxima had areas over twice those of the corresponding areas at lower levels (1–4 km). In comparison, the systems in the Morrison runs had maximum extent at 7 or 8 km in altitude, with less of a change in extent from lower levels, and WSM6 had a more uniform distribution with height, with systems smaller on average than those in the Morrison configurations at each level. Consistent with the overall system areas derived from composite reflectivity, the 1-km WSM6 Modified configuration had very small system areas at all levels. Thompson runs experienced relatively large increases in area at almost all levels as grid spacing was refined, even aloft where the areas were already larger than in other schemes at 3 km. WSM6 and Morrison experienced a small decrease in areas for the lowest layers with increases aloft, especially in WSM6 that experienced the greatest increase, over 70% at 10 km, found with any of the microphysics schemes that had produced stratiform regions in 3-km runs. Some stratiform regions did develop in the WSM6 Modified runs, unlike in the 3-km runs, but they remained far smaller than observed or with any of the other microphysics schemes (Fig. 16b).

By examining the microphysical composition of the modeled systems, these variations in system extent with height can, in part, be traced back to differences

---

**Fig. 15.** Counts of system area (in 1000 km²) as depicted in composite reflectivity for all system hours classified as a bow echo or stratiform (TS, PS, or LS) mode. The top row compares WRF runs with 3-km horizontal grid spacing to observations on the left; the bottom row does the same for runs with 1-km horizontal grid spacing. Runs using the Thompson microphysics scheme are shown in column 2, WSM6 in column 3, Modified WSM6 in column 4, and Morrison in the rightmost column.
in the representation of hydrometeors between the model configurations. Vertical cross sections of the most prevalent hydrometeor species for a representative hour from the 13 July 2015 case are shown in Fig. 19. At both 3- and 1-km grid spacing, the Thompson configurations exhibited a narrow region where graupel dominates above the melting layer of the system in the convective line, with large areas of snow ahead and behind the convective line aloft. These large areas of snow being lofted far from the convective region correspond to the previously observed large system areas aloft (Figs. 18a,e) and low-biased system average reflectivity (Fig. 17). In comparison, for both WSM6 and Morrison, a wider region of the modeled system has graupel dominating above the melting layer and much smaller regions of snow. However, Morrison exhibited these regions with wider extent than WSM6, and they were slightly wider in the 1-km runs compared to the 3-km runs. Less can be said about WSM6 Modified in this case due to the inability to have a consistent cross-section placement between the substantially different modeled systems at 3 and 1 km. However, it can still be seen that, in contrast to the other three microphysics schemes that had almost exclusively rain as the dominant species below the melting layer, a portion of each system in the WSM6 Modified scheme has graupel as the dominant species extending to the surface, which is consistent

FIG. 16. Errors in system area (in 1000 km²) for (a) 3- and (b) 1-km model configurations. Microphysics schemes are shown from top to bottom. Counts of system hours included in the distributions (those with stratiform or bow echo modes and matching observed systems) are given on the right, with the count out of all system hours with stratiform or bow echo modes observed.

FIG. 17. As in Fig. 16, but for errors of system-average reflectivity factor (dBZ). The averaging was performed over all grid points contained within the system slice extracted from the model/reprojected GridRad analysis.
4. Conclusions

This study utilized WRF simulations of 10 nocturnal MCSs cases in weakly forced environments to investigate the influences of smaller horizontal grid spacing from 3 to 1 km and microphysics diversity on the prediction of convective morphology. Four different microphysics schemes were used in the WRF simulations: Thompson, WSM6, WSM6 modified with hail-like graupel, and Morrison. With all four microphysics schemes, linear modes were found to be underpredicted when compared to observations, and cellular modes [especially clusters of cells (CC)] overpredicted at 3-km grid spacing, in agreement with past studies (SG14; Carlberg et al. 2018). While the decrease in grid spacing to 1 km resulted in mode distributions that differed in a statistically significant way with those of the corresponding 3-km simulations, with all configurations exhibiting increased occurrences of linear systems, the resolution increase did not always improve the accuracy of the prediction of linear modes.

In particular, the CC overprediction and linear mode underprediction were found to be maintained throughout system evolution in all model configurations. The increase in linear modes at 1-km grid spacing could be traced back to increased coverage of higher ($\geq 40 \text{ dBZ}$) reflectivity that resulted in several broken line (BL) systems at 3 km becoming nonstratiform line (NS) at 1 km or, similarly, CC becoming nonlinear (NL). Also, more bow echo (BE) modes occurred at 1-km grid spacing, especially during upscale growth in the early to midnocturnal period. Differences in morphology depiction between the microphysics schemes were also noted: model runs with Thompson microphysics tended to favor stratiform modes for linear systems, whereas WSM6 favored NS and Morrison had a more equal distribution. The modified WSM6 configuration performed especially poorly at 3-km grid spacing with only 3.3% of predicted modes being linear compared to 47%
of observed systems. This result improved at 1-km grid spacing with a distribution somewhat more similar to those of the other microphysics schemes.

Beyond the mode distribution comparisons, forecast accuracy scores were also tabulated to assess the ability of the model to correctly predict MCS morphology. High run-to-run variation was noted with these scores, and no statistically significant differences occurred in score between the 3- and 1-km WRF forecasts. This implies that the aforementioned increases in linear mode occurrence did not necessarily result in improved morphology forecasts in all cases, suggesting they may have occurred with the wrong systems, with the wrong subtypes, or at the wrong times. Also, the verification scores decreased with increased forecast length, particularly during the usual period of upscale growth, and the spread of scores among the model configurations increased in the late nocturnal period.

Fig. 19. Vertical cross sections of the most prevalent hydrometeor species indicated in model output for the 0300 UTC 13 Jul 2015 case. Columns correspond to grid spacing (3 km on left, 1 km on right) and rows correspond to microphysics scheme (from top to bottom: Thompson, WSM6, Modified WSM6, and Morrison). “Most prevalent” in this context refers to the species with the highest mixing ratio, and these labels are masked to regions where simulated reflectivity is at least 20 dB [in accord with the threshold of Haberlie and Ashley (2018)]. Corresponding composite reflectivity images with cross-section paths are inset in the upper-left-hand corner of each panel.
No consistent changes occurred in the trends of the accuracy score as grid spacing was refined from 3 to 1 km.

To supplement the analysis of convective mode forecasts, areal coverage and system average reflectivity values were compared for the modeled and observed systems that contained stratiform regions. Corresponding to the increased occurrence of stratiform modes, the Thompson and Morrison configurations produced larger system areas (and thereby increased stratiform region extent) compared to either WSM6 scheme. The large stratiform extent of systems in the Thompson runs came primarily from snow in the high levels (9 or 10 km), with an increasing profile of average system area with height, whereas the WSM6 and Morrison profiles exhibited less change. However, at 3-km grid spacing, the WSM6 runs had maximum system area at the lowest level, whereas the system area in Morrison runs increased in the middle and upper altitudes. Additionally, when BE/stratiform modes did occur with WSM6 Modified (only at 1-km grid spacing), the system areas remained very small at all vertical levels.

To expand upon the findings of this study, several areas of future work should be explored. First, this study was limited in its sample size (10 cases) due to the manual classification procedure and computationally intensive simulations used. Future work in this area should explore automated classification techniques, such as those made possible with machine learning, to permit the use of much larger datasets beyond those that can be feasibly analyzed by hand. Additionally, some past studies have shown that simulated convection is sensitive not only to microphysics schemes but also to the planetary boundary layer (PBL) parameterization used (Cohen et al. 2015). Thus, future work should investigate how convective morphology varies when different PBL schemes are used as horizontal grid spacing is refined. In addition, with 3 km likely to remain a common grid spacing in convection-allowing models run operationally, it should be examined in more detail why such large differences exist between microphysics schemes in the depiction of NL, BE, and TS modes at middle and late stages in system life cycle when 3-km grid spacing is used, and why these modes become so much more common in 1-km runs than in 3-km runs. Finally, many other quantitative measures of morphology exist, such as those offered by the community MET-MODE system (Wolff et al. 2014). By further quantifying additional aspects of the shape, size, and structure of convective systems, these additional morphological parameters can shed further light on the performance of model forecasts of convective morphology and may prove useful in classifying such systems in an automated fashion.

Acknowledgments. This work was supported by National Science Foundation Grant AGS1624947 and an REU Supplement to it. The authors thank Brian Squitieri for providing reference WRF simulations and suggesting the presentation of accuracy score evolution over system normalized time. The authors also thank Alex Haberlie for providing the original code for automated MCS extraction and parameter calculation. Thanks are also due to Ken Bowman and Cameron Homeyer for providing assistance with GridRad data and to Dave Flory and Daryl Herzmann for assistance in the use of computational resources. Local storm report data were provided by the Iowa Environmental Mesonet (https://mesonet.agron.iastate.edu). Additional software packages instrumental to this work include NumPy, SciPy, Matplotlib, Cartopy, xarray, netCDF4-python, wrf-python, rpy2, and MetPy. Finally, the authors would like to sincerely thank the three anonymous reviewers whose constructive comments helped to substantially improve this manuscript.

REFERENCES


