Prediction of Convective Morphology in Near-Cloud-Permitting WRF Model Simulations

DARREN V. SNIVELY AND WILLIAM A. GALLUS JR.

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

(Manuscript received 16 April 2013, in final form 23 October 2013)

ABSTRACT

The Weather Research and Forecasting (WRF) model’s ability to forecast convective morphological evolution is examined for 37 convective systems. The simulations used Thompson microphysics with 3-km horizontal grid spacing. Ten convective mode classifications were used. An objective score was developed to determine the accuracy of the simulated morphologies considering a normalized duration of each mode simulated and its agreement with observations. Rapid Update Cycle analyses were used to compare larger-scale preinitiation conditions to simulated morphology accuracy, as well as to examine how the WRF model’s skill at predicting these larger-scale conditions influenced its prediction of morphology. Two case studies selected as representative of the most common simulated morphology deficiencies were examined in detail. The model simulated cellular systems relatively well but struggled more with linear systems, particularly bow echoes and squall lines having trailing stratiform rain regions. Morphological evolution was generally better simulated in environments with enhanced deep-layer shear and cooler potential temperatures at the level of maximum \( u \). Weaker deep-layer shear, cooler potential temperatures at the surface, and quickly warming potential temperatures with height increased the likelihood of timing errors. The first case study showed that a warmer cold pool, much larger line-normal shear, and excessive midlevel drying were present in the model run that failed to develop a trailing stratiform region. The second case study showed that weak shear and the absence of a well-developed cold pool may have played a role in the lack of bowing.

1. Introduction

Forecasting convection remains a challenge for meteorologists. Most past efforts toward improving forecasting of convection focused on quantitative precipitation forecasting (QPF) (Olson et al. 1995; Wang and Seaman 1997; Gallus 1999; Alhamed et al. 2002; Gallus and Bresch 2006). In models with grid spacings coarse enough to require the use of a convective scheme, QPF is profoundly sensitive to the convective scheme used (Gallus and Segal 2001; Jankov et al. 2007). Clark et al. (2009) noted convection-allowing models (e.g., 4-km horizontal grid spacing) forecasted timing and location of precipitation better than convective parameterizing models, particularly when mesoscale convective systems (MCSs) occurred. Improved computational resources are allowing such fine grid spacings to be used more often.

The use of convection-allowing grid spacings results in simulated systems having some finescale structures similar to those observed with radar (Kain et al. 2006, 2008). Work is only just beginning to examine how well models simulate morphology (Fowle and Roebber 2003; Done et al. 2004; Weisman et al. 2008; Schumann and Roebber 2010). Done et al. (2004) showed the Weather Research and Forecasting (WRF) model failed to develop stratiform rain regions on multiple occasions. Fowle and Roebber (2003) found forecasts of mode were not as accurate when substantial large-scale forcing features were absent. Weisman et al. (2008) showed higher-resolution models were valuable for predicting convective mode. Additional work is needed to understand how well models depict the evolution of convective morphology since upscale growth from single-cell systems into multicell systems is common (Jirak and Cotton 2007).

One convective mode can evolve into another via several factors. Large-scale environmental parameters (shear, instability, etc.) alone do not differentiate between
convective modes, but localized differences of these parameters in combination with synoptic features help determine mode evolution (French and Parker 2008). James et al. (2006) concluded bow echoes tend to form in environments that are neither extremely dry nor moist in the low levels, as well as when the cold pool overwhelms the shear in a small area instead of along the entire line. The orientation of wind flow and wind shear with respect to surface boundaries has been shown to influence the formation and maintenance of linear systems. In the absence of a synoptic-scale forcing mechanism, such as a surface front, multicellular convection is more likely with higher wind shear (Schumann and Roebber 2010). Rapid evolution from cellular convection into a linear system can occur if midlevel flow is approximately parallel to a surface boundary due to merging cold pools and precipitation regions of the cells (Dial and Racy 2004). If the vertical shear vector is normal to the line of forcing, squall lines are more easily maintained with isolated supercells at the ends (Bluestein and Weisman 2000). Weisman et al. (1988) and Weisman (1992) showed how that shear affects the tilting of updrafts and, in turn, the extent of stratiform rain. How fast a cellular system becomes linear is positively correlated with the amount of deep-layer forcing, even more than the orientation of the wind vector in the cloud layer (Dial et al. 2010). Schumann and Roebber (2010) agreed that increased forcing tends to favor multicellular convection over individual cells because of widespread atmospheric destabilization.

Convective mode classification can help in understanding the behavior of observed and simulated weather systems and potentially in the forecasting of several hazards. Gallus et al. (2008), for instance, found significant severe hail [diameter of 2 in. (1 in. = 2.54 cm) or more] was most common from cellular storms, while the main hazard from linear systems was wind. Smith et al. (2012) showed the convective mode varied with tornado reports based on geography and seasonality, with discrete cells and clusters of cells producing the most tornadoes in the southern Great Plains during the spring while the Ohio Valley experienced a higher number of tornadoes from quasi-linear convective systems (QLCSs). French and Parker (2012) found hail was often reported before an isolated supercell merged with a linear system, whereas wind reports increased after a merger.

The present study examines convective evolution timing, while making use of the aforementioned findings. Wind shear and orientation, moisture, and synoptic forcing are compared between the WRF model simulations and observations. Two detailed case studies are performed to highlight some of the most common discrepancies in the WRF simulations compared to observations. Section 2 explains the convective mode classification scheme and the scoring method used to determine accuracy. Section 3 discusses the convective mode comparisons and the accuracy analysis. Section 4 contains the case studies, and section 5 presents the results and overall conclusions.

2. Methodology

a. Convective mode classification

Morphology classification was performed for 37 events occurring during the warm seasons from 2006 to 2010 primarily in the U.S. Great Plains and the Midwest (Fig. 1). WRF, version 3.1.1, with the Advanced Research dynamics core (ARW; Skamarock et al. 2008) was used to simulate the events with 3-km horizontal grid spacing. The majority of the cases selected in the present research and all the model schemes used match those in Duda and Gallus (2013). Initial and lateral boundary conditions used North American Mesoscale Model (NAM) 12-km output. The WRF default of 40 vertical levels was used. For the majority of the events, the model was initialized at 1200 UTC and integrated for 24 h. Simulated composite reflectivity output in 15-min increments was used to determine the morphology and its evolution.

Convective modes were identified throughout each event’s evolution using nine classifications from Gallus et al. (2008), along with an additional classification developed for the present research. Three modes composed the cellular group: individual cells (IC), clustered cells (CC), and broken line (BL). Five modes represented...
linear systems: those with no stratiform rain region (NS), trailing stratiform (TS), parallel stratiform (PS), leading stratiform (LS), and bow echoes (BE). Another mode was nonlinear (NL). The new classification, mixed complex (MC), was reserved for situations that exhibited characteristics of two or more of the aforementioned convective modes.

Convective initiation was defined as the first instance of 40-dBZ reflectivity, and the system had to maintain 40-dBZ reflectivity in order to retain a classification. The minimum length of the convective portion (reflectivity exceeding 40 dBZ) required for linear systems was 75 km. Stratiform regions were defined as having at least 30-dBZ reflectivity (Hilgendorf and Johnson 1998) over an area at least twice as wide as the adjacent convective lines. A system became linear in nature when a 3:1 length-to-width ratio was attained (Gallus et al. 2008). The event must have shown characteristics of a particular convective mode for a minimum of 2 h to receive that mode’s classification. The centroid of the simulated system must also have initiated within 300 km of the observed system’s centroid in order to be classified. The mosaic radar archive from the University Corporation of Atmospheric Research (UCAR) website (http://locust.mmm.ucar.edu) was used to analyze convective evolution of the observed systems. The same criteria for classification were applied to the simulated events.

### b. Scoring method for WRF mode accuracy

Events were scored based on how well the simulated convective mode matched the radar observations. Time was normalized with convective initiation set to zero, dissipation set to one, and the duration of each convective mode represented by its portion of the event’s lifetime. As an example, in case study 1 (to be discussed later), radar showed the system initiating at 1851 UTC as BL and ending at 1200 UTC as NL, marking 0 and 1 for the time scale (Table 1). The first mode change occurred at 2321 UTC, which is 4.5 h later and roughly one-quarter through the lifetime of the system. If the system did not dissipate by the end of the model run (usually 1200 UTC), the end of the model run was defined as 1.0. That point would also mark the end of the period for observed systems, as is the case in Table 1. If the system moved out of the domain, the time at which the convective mode could no longer be classified was defined as 1.0. Two time normalizations were performed for each event: one based on radar observations and the other on the simulation. The simulated system’s normalized time scale for this case is displayed in Table 2.

If the simulated event’s initiation or dissipation occurred more than 3 h different from that observed, a penalty was introduced through an adjustment to the WRF’s time scale, thus reducing the maximum possible score. For example, if the simulated system initiated earlier or dissipated later than the observed system, the fraction of the life span outside the 3-h grace period was not considered for comparison to the observed system, thus earning a score of 0 (effectively a comparison of “no system” to the simulated mode). If the simulated system initiated more than 3 h later or dissipated more than 3 h earlier than the observed system, the time between the grace period and the dissipation—initiation was added to the overall simulated time span and not considered for comparison so that it again was scored as 0.

The two time scales were merged by partitioning the event’s time scale whenever a convective mode changed in either the observations or the simulation. Table 3

### Table 1. Normalized time scale and scoring for the observed system from case study 1. Mode refers to the convective mode displayed by the radar, time represents the point at which the particular convective mode was first observed, length is the duration of the mode, and normalized time refers to the duration of the mode using the 0 to 1 time scale.

<table>
<thead>
<tr>
<th>Mode</th>
<th>Time (UTC)</th>
<th>Length (h)</th>
<th>Normalized time</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>BL</td>
<td>1851</td>
<td>0</td>
<td>0.18</td>
<td>1</td>
</tr>
<tr>
<td>NS</td>
<td>2321</td>
<td>4.5</td>
<td>0.26</td>
<td>0.82</td>
</tr>
<tr>
<td>TS</td>
<td>0321</td>
<td>8.5</td>
<td>0.55</td>
<td>0.50</td>
</tr>
<tr>
<td>NL</td>
<td>0900</td>
<td>14.0</td>
<td>0.82</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>1200</td>
<td>17.0</td>
<td>1</td>
<td>0.04</td>
</tr>
</tbody>
</table>

### Table 2. As in Table 1, but for the simulated system.

<table>
<thead>
<tr>
<th>Mode</th>
<th>Time (UTC)</th>
<th>Length (h)</th>
<th>Normalized time</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>BL</td>
<td>2000</td>
<td>0</td>
<td>0.04</td>
<td>0</td>
</tr>
<tr>
<td>NS</td>
<td>2330</td>
<td>3.5</td>
<td>0.22</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>1200</td>
<td>16.0</td>
<td>1</td>
<td>0.38</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mode</th>
<th>Time (UTC)</th>
<th>Length (h)</th>
<th>Normalized time</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>BL</td>
<td>2000</td>
<td>0</td>
<td>0.04</td>
<td>0</td>
</tr>
<tr>
<td>NS</td>
<td>2000</td>
<td>0</td>
<td>0.04</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1200</td>
<td>16.0</td>
<td>1</td>
<td>0.38</td>
</tr>
</tbody>
</table>

### Table 3. Combined time scale and scoring method for case study 1: $T_1$ is the time at which a mode is first observed and $T_2$ is the end of the particular mode comparison; $\Delta T$ is the duration of the mode comparison. Boldface values correspond to correct matches (full score per time unit) and italics correspond to group matches (half score per time unit). Radar and model indicate the respective modes displayed during the comparison by the observations and the simulation. Score indicates the score earned during individual time periods, with boldface indicating the total score for the case.

<table>
<thead>
<tr>
<th>$T_1$</th>
<th>$T_2$</th>
<th>$\Delta T$</th>
<th>Radar</th>
<th>Model</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.22</td>
<td>0.22</td>
<td>BL</td>
<td>BL</td>
<td>0.22</td>
</tr>
<tr>
<td>0.22</td>
<td>0.26</td>
<td>0.04</td>
<td>BL</td>
<td>NS</td>
<td>0</td>
</tr>
<tr>
<td>0.26</td>
<td>0.50</td>
<td>0.24</td>
<td>NS</td>
<td>NS</td>
<td>0.24</td>
</tr>
<tr>
<td>0.50</td>
<td>0.82</td>
<td>0.32</td>
<td>TS</td>
<td>NS</td>
<td>0.16</td>
</tr>
<tr>
<td>0.82</td>
<td>1</td>
<td>0.18</td>
<td>NL</td>
<td>NS</td>
<td>0</td>
</tr>
</tbody>
</table>
shows the merged table for the case 1 example. The first convective mode change occurred when the model transitioned from BL to NS after just shy of one-quarter of its life span, and the observed system changed from BL to NS shortly thereafter. This method was followed until normalized time 1.0. Scores were then computed using both general group matches and more detailed morphology matches. For a group match, the WRF-simulated general morphology (cellular, linear, or nonlinear) had to match that observed, even if the specific morphologies differed (e.g., IC versus BL). Group matches were awarded a 0.5 score for the duration they were observed. A correct match occurred when the specific morphologies (the 10 types) matched, and it was awarded a 1.0 score. The portions of the event’s life span where matches were identified were summed to calculate the total accuracy score $S$ using

$$S = \sum_{i=1}^{N} M \Delta t,$$

where $N$ represents the total number of mode comparisons possible (a function of the number of times the mode changed in either the observations or the simulation), $M$ a weight based on the match type (0 for no agreement, 0.5 for a group match, 1.0 for a correct match), and $\Delta t$ the duration of the mode comparison using the normalized time scale. Figure 2 displays all 37 cases’ morphologies and scores.

3. Results

The frequency of each classification for observed and simulated systems is shown in Table 4. Overall, radar observations contained six more convective modes than the model (115 versus 109). The WRF produced the same number of BL (24), PS (1), and LS (0) systems. The largest differences in counts occurred with BE events (13 fewer simulated than observed), TS (6 fewer), NS (5 more), and CC (9 more). In general, the WRF forecasted too many cellular systems (48 versus 37) and too few linear systems (44 versus 58). The WRF model simulated a linear mode at one point in each system’s evolution in 26 of the 34 events (76%) that featured an observed linear system. Of the 21 events that included an observed bow echo, the WRF captured that mode at some point in the systems’ evolution in only eight events (38%). The model also had trouble simulating TS systems, only showing TS in 7 of 19 observed TS events (37%). This lack of TS rain regions is consistent with the results of Done et al. (2004). There were also six instances of the model failing to produce BL events.

a. Match accuracy

Detailed and group matches are outlined in Table 5, and convective mode comparisons are shown in Table 6. Of the 185 comparisons, only 58 (31%) were detailed matches, with 104 (56%) group matches (Table 5). The model was the most accurate in simulating cellular systems, with approximately half of the comparisons for these three classifications yielding detailed matches. BL had the most detailed matches of all classifications with 17 (Table 5). The majority of the BL comparisons that did not match were cases when the model instead showed NS (eight occurrences) or CC modes (six occurrences).
FIG. 2. Simulated (top bar) and observed (bottom bar) mor-phologies for each case having (a) objective scores of at least 0.5 with no major timing error, (b) scores less than 0.5 with no major timing error, and (c) major errors in timing (initiation or dissipation timing differences exceeding 3 h). Shading for convective modes is shown at the end of (c). Dashed borders in (c) represent fraction of time span outside a 3-h grace period for initiation and dissipation. Dashed boxes in (c) indicate the model either initiated more than 3 h early or dissipated more than 3 h later than the observed event. Dashed–dotted boxes indicate the model either initiated more than 3 h late or dissipated more than 3 h earlier than the observed event.
The former discrepancy was usually due to a model timing error (the model was too quick to form a line), and the latter usually was present during the initial mode (the model cells did not initiate in a line). When accounting for group matches, the accuracy increased to over 70% (Table 5), implying the WRF usually did simulate cellular convection but had trouble with the details of cell arrangement.

Linear systems were not simulated as well by the WRF model, with only 16% of bow echoes and 24% of trailing stratiform systems correctly simulated (seven detailed matches each; see Table 5). Common model errors for BE events included simulation of NS (10 occurrences) or TS events (8 occurrences) (Table 6). Those results imply for some cases problems in generating stratiform rain that may play a role in the bowing process and, for other events, problems in simulating the bowing, even when trailing stratiform rainfall was correctly simulated. When a TS system was observed, the model simulated an NS event 28% of the time and a BL event 17%, again implying a frequent systematic problem in the generation of stratiform rain, along with some tendency to struggle with upscale organization into lines. Nearly 60% of BE and TS comparisons resulted in group matches, suggesting the model usually did simulate a linear system when BE or TS events were observed. Observed NS events featured multiple occurrences of simulated cellular systems, which suggest the model may have more difficulty producing linear systems when no stratiform rain or bowing exists in the observed system (Table 6).

Approximately one-third of the observed nonlinear systems were simulated by the model, and for the remaining NL events, the model modes were evenly split between cellular and linear modes. Simulations of mixed-complex systems were especially poor, with only one out of the six observed events correctly forecasted.

The normalized time-scale scoring method resulted in an average score of 0.49 for the 37 cases, with 15 cases receiving a score of at least 0.50. The best score was a perfect 1.0 for one case of a TS existing for the entire time. Of note, the next highest score occurred for a case evolving through four different modes. The worst score was 0.03 for a case with a major delay in initiation and a simulation of IC for the entire time while a linear system was observed. Nine cases violated the timing criterion with an average score of 0.32. The average accuracy scores for the cases where the WRF model completely failed to simulate an observed BE or TS event were 0.42 and 0.44, respectively. The WRF correctly simulated the first convective mode in 13 of the 15 cases with a score of 0.50 or greater and 12 of the 22 cases with a score of less than 0.50.

b. Relationship of mode accuracy to environmental conditions

The Wilcoxon rank-sum test was applied to the cases' mode accuracy scores and observed prestorm environmental parameters to determine the statistical significance of differences in the large-scale conditions present for cases simulated accurately and those not. The test showed cases with scores under 0.50 tended to have significantly warmer potential temperatures at the level of maximum $\theta_e$ and significantly lower 0–6-km bulk shear values than cases with scores over 0.50, with at least 90% confidence (Table 7). The average deep-layer

---

**Table 4. Frequency of convective modes for the observed and simulated events with the difference (model − radar).**

<table>
<thead>
<tr>
<th>Modeled modes</th>
<th>IC</th>
<th>CC</th>
<th>BL</th>
<th>BE</th>
<th>TS</th>
<th>NS</th>
<th>LS</th>
<th>NL</th>
<th>MC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radar</td>
<td>6</td>
<td>7</td>
<td>24</td>
<td>21</td>
<td>19</td>
<td>17</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Model</td>
<td>8</td>
<td>16</td>
<td>24</td>
<td>8</td>
<td>13</td>
<td>22</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Difference</td>
<td>2</td>
<td>9</td>
<td>0</td>
<td>-13</td>
<td>-6</td>
<td>5</td>
<td>0</td>
<td>-1</td>
<td>-2</td>
</tr>
</tbody>
</table>

---

**Table 5. Number of correct and group matches for each mode comparison.**

<table>
<thead>
<tr>
<th>Observed mode</th>
<th>No. of occurrences</th>
<th>Correct matches</th>
<th>Group matches</th>
</tr>
</thead>
<tbody>
<tr>
<td>IC</td>
<td>7</td>
<td>3</td>
<td>43%</td>
</tr>
<tr>
<td>CC</td>
<td>11</td>
<td>5</td>
<td>45%</td>
</tr>
<tr>
<td>BL</td>
<td>34</td>
<td>17</td>
<td>50%</td>
</tr>
<tr>
<td>BE</td>
<td>44</td>
<td>7</td>
<td>16%</td>
</tr>
<tr>
<td>TS</td>
<td>29</td>
<td>7</td>
<td>24%</td>
</tr>
<tr>
<td>PS</td>
<td>1</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>NS</td>
<td>24</td>
<td>8</td>
<td>33%</td>
</tr>
<tr>
<td>LS</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>29</td>
<td>10</td>
<td>34%</td>
</tr>
<tr>
<td>MC</td>
<td>6</td>
<td>1</td>
<td>17%</td>
</tr>
<tr>
<td>Total</td>
<td>185</td>
<td>58</td>
<td>31%</td>
</tr>
</tbody>
</table>

---

**Table 6. Distribution of simulated convective mode (top of table) as a function of observed mode (left side). The boldface values represent detailed matches.**

<table>
<thead>
<tr>
<th>Modeled modes</th>
<th>IC</th>
<th>CC</th>
<th>BL</th>
<th>BE</th>
<th>TS</th>
<th>PS</th>
<th>NS</th>
<th>LS</th>
<th>NL</th>
<th>MC</th>
</tr>
</thead>
<tbody>
<tr>
<td>IC</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CC</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BL</td>
<td>1</td>
<td>6</td>
<td>17</td>
<td>2</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BE</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>1</td>
<td>10</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TS</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>8</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>NS</td>
<td>4</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td></td>
<td>8</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Table 7. Number of convective modes for the observed and simulated events with the difference (model − radar).**

<table>
<thead>
<tr>
<th>Modeled modes</th>
<th>IC</th>
<th>CC</th>
<th>BL</th>
<th>BE</th>
<th>TS</th>
<th>NS</th>
<th>LS</th>
<th>NL</th>
<th>MC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radar</td>
<td>6</td>
<td>7</td>
<td>24</td>
<td>21</td>
<td>19</td>
<td>17</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Model</td>
<td>8</td>
<td>16</td>
<td>24</td>
<td>8</td>
<td>13</td>
<td>22</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Difference</td>
<td>2</td>
<td>9</td>
<td>0</td>
<td>-13</td>
<td>-6</td>
<td>5</td>
<td>0</td>
<td>-1</td>
<td>-2</td>
</tr>
</tbody>
</table>

---

Unauthenticated | Downloaded 10/23/23 01:03 PM UTC
Table 7. Averages of environmental parameters for all 37 cases, cases with accuracy scores of greater than 0.5, cases with scores less than 0.5, cases that met the timing criterion, and cases that did not meet the timing criterion. SBCAPE and MUCAPE represent surface-based and most unstable CAPE values, respectively, and ΔCAPE represents the difference (MUCAPE – SBCAPE). SBCIN represents surface-based convective inhibition. The bulk shear values are in the 0–3- and 0–6-km AGL layers. Here, θ_{sfc} and θ_{max} represent the potential temperature near the surface (0–30-hPa AGL average) and at the level of maximum equivalent potential temperature, respectively; Δθ represents the difference between the two θ values; Δp represents the pressure difference between the surface and the level of maximum equivalent potential temperature; and θ_{lap} represents the change of potential temperature per unit pressure in the layer specified by Δp. Boldface values represent statistically significant differences either between cases with scores greater than 0.5 vs less than 0.5, or cases that met the temporal criterion vs those that did not at the 90% confidence level or greater.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>All cases</th>
<th>Score &gt; 0.5</th>
<th>Score &lt; 0.5</th>
<th>Met timing criterion</th>
<th>Did not meet timing criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBCAPE (J kg⁻¹⁻)</td>
<td>1860</td>
<td>1936</td>
<td>1802</td>
<td>1783</td>
<td>2100</td>
</tr>
<tr>
<td>MUCAPE (J kg⁻¹⁻)</td>
<td>2422</td>
<td>2251</td>
<td>2553</td>
<td><strong>2192</strong></td>
<td><strong>3139</strong></td>
</tr>
<tr>
<td>ΔCAPE (J kg⁻¹⁻)</td>
<td>562</td>
<td>315</td>
<td>751</td>
<td>409</td>
<td>1039</td>
</tr>
<tr>
<td>SBCIN (J kg⁻¹⁻)</td>
<td>-119</td>
<td>-101</td>
<td>-134</td>
<td>-130</td>
<td>-86</td>
</tr>
<tr>
<td>0–3-km shear (m s⁻¹⁻)</td>
<td>11.7</td>
<td>11.9</td>
<td>11.6</td>
<td>11.9</td>
<td>11.3</td>
</tr>
<tr>
<td>0–6-km shear (m s⁻¹⁻)</td>
<td>19.8</td>
<td><strong>22.4</strong></td>
<td><strong>17.8</strong></td>
<td><strong>21.9</strong></td>
<td><strong>12.9</strong></td>
</tr>
<tr>
<td>θ_{sfc} (K)</td>
<td>303.5</td>
<td>302.6</td>
<td>304.2</td>
<td>303.4</td>
<td>300.8</td>
</tr>
<tr>
<td>θ_{max} (K)</td>
<td>307.1</td>
<td><strong>305.2</strong></td>
<td><strong>308.5</strong></td>
<td>307.2</td>
<td>306.7</td>
</tr>
<tr>
<td>Δθ (K)</td>
<td>3.6</td>
<td>2.6</td>
<td>4.3</td>
<td><strong>2.8</strong></td>
<td><strong>5.9</strong></td>
</tr>
<tr>
<td>Δp (hPa)</td>
<td>65</td>
<td>48</td>
<td>77</td>
<td>62</td>
<td>73</td>
</tr>
<tr>
<td>θ_{lap} (K hPa⁻¹⁻)</td>
<td>5.8</td>
<td>5.7</td>
<td>5.8</td>
<td><strong>5.2</strong></td>
<td><strong>7.7</strong></td>
</tr>
</tbody>
</table>

Shear for the cases where the model performed relatively well was 22.4 compared to 17.8 m s⁻¹ for cases simulated more poorly. Assuming the amount of shear reflects the large-scale forcing, this result implies mode might be better predicted for cases with stronger largescale forcing than weaker forcing, a result found to be true for QPF in convection-parameterized model runs (Jankov and Gallus 2004). The average potential temperature at the level of maximum θ_e was 3.3 K warmer in the low-scoring cases than the better-scoring cases, and surface-based convective inhibition (CIN) was greater. The maximum equivalent potential temperature, on average, was lower for cases with scores greater than 0.5. In addition, the level of maximum θ_e was more elevated above the surface (bigger Δp) for cases with scores less than 0.5. These results imply the forecasting of mode is worse when the height of the inversion is greater, as likely would be the case for elevated convection. No relationship was found between accuracy scores and surface-based CAPE or 0–3-km bulk shear.

Comparisons were also made between cases that did not meet the timing criterion and those that stayed within the 3-h initiation or dissipation thresholds. Cases that violated the timing criterion had significantly lower (99% confidence) 0–6-km bulk shear values compared to cases that met the criterion, with the average values being 12.9 and 21.9 m s⁻¹, respectively (Table 7). The rank-sum test also showed moderate evidence that cases that did not meet the threshold had lower potential temperatures at the surface and steeper potential temperature lapse rates (stronger inversions). Here, the potential temperature lapse rate is defined as the increase of potential temperature with decreasing pressure (−∂θ/∂p) from the surface to the level of maximum θ_e. The cases that violated the temporal criterion featured on average a lower surface temperature and a steeper lapse rate, indicating greater stability near the surface prior to convective initiation. These cases also had an average most unstable CAPE value nearly 1000 J kg⁻¹ higher than the cases that met the criterion (95% confidence). However, the increase in the number of statistically significant results for the timing issue cases may be due to the small sample size for these cases.

c. Comparison of environmental parameters between the WRF simulations and RUC analyses

Overall, the WRF-simulated events featured higher bulk shear values than were present in the RUC analyses, and levels of maximum θ_e that were closer to the surface, resulting in shallower inversion layers and less surface-based CIN (Table 8). On average, the WRF model underestimated the difference between SBCAPE and MUCAPE, primarily because it showed less MUCAPE. The average SBCAPE and MUCAPE values were slightly higher in the WRF simulations compared to the RUC data for cases with scores of at least 0.5, while the average values were lower for the cases scoring more poorly. Interestingly, the average absolute error was larger for the cases scoring 0.5 and greater. For cases that did not meet the timing criterion, the underprediction of MUCAPE was accentuated in the WRF simulations. The average value for those cases was approximately 900 J kg⁻¹ lower than that in the RUC analyses, and the average absolute error was nearly 1500 J kg⁻¹. The WRF
simulations also showed much lower SBCAPE and MUCAPE values than in the RUC for cases that featured a missed BE event. Surface-based CIN (SBCIN) was usually underestimated in the WRF simulations, but cases that did not meet the timing criterion had a larger average CIN in the WRF than in the RUC.

Bulk shear (3 and 6 km) was often larger in the WRF simulations than in the RUC analyses, and the average absolute error was roughly the same between the cases scoring above and below 0.5. Similar to the RUC analyses, the WRF model exhibited, on average, much less shear in cases that did not meet the timing criterion. The WRF model usually simulated higher surface potential temperatures than the RUC; however, cases with accuracy scores of at least 0.5 had on average higher theta values at the level of maximum \( \theta_e \), while the poorer-scoring cases and cases that violated the timing criterion had cooler theta values.

These results suggest the model predicts convective evolution more accurately when observations show deep-layer shear is relatively high. If shear is relatively weak, or if the model greatly underestimates shear, initiation or dissipation more than 3 h apart from the observed event is more likely to occur. The large difference between most unstable and surface-based CAPE and between potential temperature at the surface and aloft for the low-scoring cases, along with a higher level of maximum equivalent potential temperature in the RUC analyses, indicates a stronger stable layer near the ground, suggesting that some of these events likely include elevated convection. The model also appears to be more prone to delayed initiation or early dissipation when elevated convection exists. The model is likely to portray a system’s evolution better when the WRF overestimates CAPE and simulates well the level of maximum \( \theta_e \). In general, however, Table 8 does not show a clear signal for morphology to be better predicted when these larger-scale parameters are better predicted.

### Table 8. As in Table 7, but for the average difference between the WRF-simulated parameters and the RUC values (positive numbers indicate WRF values larger than RUC values). Boldface represents statistical significance at 90% confidence.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>All cases</th>
<th>Score &gt; 0.5</th>
<th>Score &lt; 0.5</th>
<th>Met timing criterion</th>
<th>Did not meet timing criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBCAPE (J kg(^{-1}))</td>
<td>15</td>
<td>250</td>
<td>-164</td>
<td>97</td>
<td>-240</td>
</tr>
<tr>
<td>MUCAPE (J kg(^{-1}))</td>
<td>-205</td>
<td>321</td>
<td>-605</td>
<td>-28</td>
<td>-292</td>
</tr>
<tr>
<td>( \Delta \text{CAPE} ) (J kg(^{-1}))</td>
<td>-220</td>
<td>71</td>
<td>-441</td>
<td>-69</td>
<td>-688</td>
</tr>
<tr>
<td>SBCIN (J kg(^{-1}))</td>
<td>48</td>
<td>62</td>
<td>37</td>
<td>79</td>
<td>-51</td>
</tr>
<tr>
<td>0–3-km shear (m s(^{-1}))</td>
<td>3.7</td>
<td>3.7</td>
<td>3.7</td>
<td>4.3</td>
<td>1.7</td>
</tr>
<tr>
<td>0–6-km shear (m s(^{-1}))</td>
<td>1.7</td>
<td>-0.8</td>
<td>3.6</td>
<td>0.9</td>
<td>3.9</td>
</tr>
<tr>
<td>( \theta_{\text{crit}} ) (K)</td>
<td>2.2</td>
<td>2.6</td>
<td>1.8</td>
<td>2.4</td>
<td>1.4</td>
</tr>
<tr>
<td>( \theta_{\text{max}} ) (K)</td>
<td>0.2</td>
<td>1.2</td>
<td>-0.5</td>
<td>0.8</td>
<td>-1.6</td>
</tr>
<tr>
<td>( \Delta \theta ) (K)</td>
<td>-1.9</td>
<td>-1.4</td>
<td>-2.3</td>
<td>-1.6</td>
<td>-3.0</td>
</tr>
<tr>
<td>( \Delta p ) (hPa)</td>
<td>-17</td>
<td>-16</td>
<td>-19</td>
<td>-17</td>
<td>-19</td>
</tr>
<tr>
<td>( \theta_{\text{ape}} ) (K hPa(^{-1}))</td>
<td>-5.9</td>
<td>-6.8</td>
<td>-5.2</td>
<td>-6.0</td>
<td>-5.4</td>
</tr>
</tbody>
</table>

### 4. Case studies

Two case studies were performed for events representing the two most common types of errors observed (missed BE and TS events), to investigate in more detail the potential causes of the differences in convective evolution between the WRF simulations and observations. Synoptic conditions and mesoscale environments were explored in detail when a convective mode changed either in the observations or in the simulations. In addition, because it has been shown that precipitation accumulation among ensemble members can be a function of the microphysical scheme used (e.g., Jankov et al. 2005; Schwartz et al. 2010) and thus morphology might vary as well, microphysical sensitivity tests were also performed.

Gallus and Pfeifer (2008) noted reflectivity was underestimated in the convective portion of a squall line when the Lin scheme was used, and the Thompson and WRF single-moment six-class (WSM6) schemes overestimated reflectivity within the stratiform anvil. Morrison et al. (2009) found notable differences in trailing stratiform rain production between the one- and two-moment microphysical schemes. In the tests performed in the present study, the Thompson et al. (2008) microphysics scheme was replaced by the single-moment WSM6 (Hong and Lin 2006) and Lin et al. (1983) schemes, as well as the double-moment Morrison et al. (2005) scheme. These microphysics schemes are widely used by researchers running the WRF model. Because forecasts are also sensitive to initial and lateral boundary conditions (Jankov et al. 2007; Weisman et al. 2008), an additional test was performed for each case study where in the control run, Global Forecasting System (GFS) output was used for initial and lateral boundary conditions instead of NAM.

#### a. Case study 1: 23–24 May 2006

For the 23–24 May 2006 TS system, the WRF model simulated NS, representative of 42% of all TS events.
The model received an accuracy score of 0.62 for this event, mainly because it did capture a linear mode (Table 3), albeit without stratiform rain. RUC analyses at 1800 UTC showed little difference between surface-based CAPE and most unstable CAPE (131 J kg⁻¹), and the maximum equivalent potential temperature was only 47 hPa above the surface, which could suggest this event began as surface-based convection. Low-level instability and appreciable deep-layer shear (20.7 m s⁻¹) were present, values consistent with the good-scoring cases (Table 7).

An occluding cold front stretched from a low pressure center in eastern Montana southward through Kansas and likely contributed to convective initiation for this event (not shown). The upper-level low at 500 hPa associated with the system deepened during the period, with heights falling from 5700 m at 0000 UTC 24 May to 5640 m at 1200 UTC (not shown).

The BL system initiated at 1900 UTC 23 May in north-central Nebraska and south-central South Dakota ahead of the frontal system (Fig. 3a). The model predicted convective initiation well (Fig. 2b) by simulating the correct mode less than 100 km to the east just 1 h later than the observed event (2000 UTC). Near-storm and prestorm environmental shears were approximately 15–20 m s⁻¹ (Fig. 4).

At 2330 UTC, both the observed and simulated systems evolved to NS (Fig. 3). The RUC analysis showed an increase in deep-layer shear in central Nebraska that the WRF model did not predict, but the shear vectors were mainly parallel to the boundary for both systems, which might have helped consolidate the cells into a line (Fig. 4). The model also correctly depicted the placement of low-level moisture and negative moisture advection at 850 hPa directly behind the system (not shown).

However, at 0330 UTC 24 May, the observed system became TS, while the WRF simulation continued to depict NS (Fig. 3). The highest values of 0–6-km bulk shear in the RUC analysis were located in Kansas, and most of the near-storm environmental shear in eastern Nebraska was 15–20 m s⁻¹ (Fig. 4e). The WRF showed a shear maximum of 30 m s⁻¹ along the line, but values quickly dropped off behind the line to 15 m s⁻¹ (Fig. 4f). Negative moisture advection at 850 hPa lagged well behind the system in the RUC analysis, but the negative advection was modeled farther east by the WRF output (Fig. 5). This pattern was also evident in the 850-hPa relative humidity in the WRF model, with values as low as 30% in eastern Nebraska behind the convective line (Fig. 5b). At the surface, the RUC analysis showed a fully developed cold pool with potential temperature perturbations of approximately 6 K and an obvious divergent pattern in the 10-m winds (Fig. 6a). However, in the WRF model output, the diverging surface winds and temperature perturbations were confined to a much smaller area (Fig. 6b).

A vertical cross section of the component of storm-relative wind normal to the line in the WRF simulation was taken to examine the cold pool–wind shear interaction (Fig. 7a). The erect updraft with little tilt is consistent with Rotunno et al. (1988) when a vorticity balance between the cold pool and low-level shear exists. The front-to-rear flow at anvil level suggests hydrometeors should be transported to the rear of the system (Smull and Houze 1985), and drier air should be advected in from the rear-inflow jet. The jet had maximum storm-relative flow of approximately 15 m s⁻¹ near 500 hPa, consistent with “strong rear inflow” cases defined in Smull and Houze (1987). The layer of positive storm-relative flow was approximately 5 km thick directly behind the main updraft, 2 km thicker than the mean strong-rear-inflow profile for lines with trailing stratiform precipitation in Smull and Houze (1987). A simulated atmospheric sounding taken at the point of maximum front-to-rear flow showed a very dry layer in the midlevels of the atmosphere (Fig. 7b). Hydrometeors falling into this region would most likely evaporate or sublimate before reaching the surface, thus playing a role in the lack of a trailing stratiform rain region in the WRF model.

Dial et al. (2010) differentiated between systems that developed trailing stratiform rain regions within 3 h of initiation and those that did not by looking at the 2–6-km shear normal to the boundary. That method was followed similarly in this study, but by using points spaced 50 km apart ahead of the main frontogenetical zone. The simulation had an average of 5.2 m s⁻¹ of shear normal to the boundary, while RUC analyses showed an average of 1.2 m s⁻¹. Dial et al. (2010) found systems with trailing stratiform rain regions had smaller shear values, and thus this result is consistent with the lack of a trailing stratiform region in the WRF model, while one was observed.

By 0900 UTC 24 May, the observed system transitioned to NL, while the WRF simulated NS (Figs. 3g,h). At this time, the RUC analysis showed a broad region of 15–20 m s⁻¹ of deep-layer shear with a local maximum of 20–25 m s⁻¹ over central Iowa (Fig. 4g). The WRF output depicted a maximum of 25–30 m s⁻¹ of shear along the leading edge of the line and smaller values extending west toward the Missouri River (Fig. 4h). The broad area of higher shear in the WRF model output than what was indicated in observational data might have been a factor in maintaining an organized simulated system.

The microphysical scheme sensitivity tests in the present case study resulted in little change in reflectivity
FIG. 3. Reflectivity (dBZ) during convective mode transitions for the 23–24 May 2006 event. Observed reflectivity from UCAR image archive at (a) 1900, (c) 2330, (e) 0330, and (g) 0900 UTC. Simulated reflectivity at (b) 2000, (d) 2330, (f) 0330, and (h) 0900 UTC. The first green hue correlates to 30 dBZ, and the first yellow hue correlates to 40 dBZ (see color bar).
FIG. 4. Bulk shear (m s$^{-1}$) from 0–6 km during convective mode transitions for the 23–24 May 2006 event. RUC analyses of 0–6-km bulk shear at (a) 1900, (c) 0000, (e) 0400, and (g) 0900 UTC. WRF output at (b) 2000, (d) 2330, (f) 0330, and (h) 0900 UTC. Contour interval is 5 m s$^{-1}$ with blue shades having magnitudes of 15 m s$^{-1}$ and greater.
and convective mode for the first several hours compared to the original simulation (not shown). All three additional runs simulated BL at approximately 2000 UTC evolving to NS by 2300 UTC. However, the test simulations did produce a system with equal width and length of high reflectivity values (>60 dBZ) near 0600 UTC, prompting an NL classification, a result different from the control run (NS) and matching observations at these later times. All simulations failed to simulate TS, suggesting this problem was either not primarily related to a deficiency in any particular microphysical scheme, or that the same deficiency was present in all four schemes tested.

Because forecasts are sensitive to initial and lateral boundary conditions (Jankov et al. 2007; Weisman et al. 2008), one test was performed using GFS data instead of NAM for initialization and lateral boundary conditions in the control configuration. The GFS-initialized simulation did indeed produce a TS region for the northern half of the line at 0000 UTC, and maintained the stratiform region for several hours. Analyses at model initialization (1200 UTC 23 May) showed that 0–6- and 2–6-km shear and surface-based CAPE and CIN were similar between the RUC and WRF runs using GFS data (not shown). The WRF run using NAM data underestimated 0–6-km bulk shear and CAPE and overestimated 2–6-km shear in the region of the convective initiation. Upon transition to TS in the GFS-initialized simulation, the 2–6-km line-normal shear ahead of the main frontogenetical forcing for the northern portion of the line was half the magnitude of that in the NAM-initialized WRF simulation, and thus closer to the observations. The line of intense reflectivity (50+ dBZ) was also more continuous than in the NAM-initialized run, reflecting a more unstable environment present at model initialization in the GFS-initialized run. The stronger convection in the GFS-initialized simulation may have altered the flow in the near-storm environment, weakening the line-normal component of the

---

**Fig. 5.** (a) RUC analysis at 0400 UTC and (b) WRF output at 0330 UTC of 850-hPa relative humidity (%) on 24 May 2006. (c) RUC analysis at 0400 UTC and (d) WRF output at 0330 UTC of 850-hPa moisture advection ($\times 10^{-5}$ g kg$^{-1}$ s$^{-1}$).
midlayer shear, so as to better allow the development of a TS system.


In the 26–27 May 2006 case, the WRF model initiated IC that grew into NS, but the radar showed BE for much of the system’s life. The modeled line also transitioned to a NL late in its life span while the radar continued to display BE. The WRF model received a score of 0.24 for this event due to errors seen in Table 9. Convection initiated with nearly 3500 J kg\(^{-1}\) of surface-based CAPE and 3900 J kg\(^{-1}\) of most unstable CAPE. Deep-layer shear was below the 37-case average at 18 m s\(^{-1}\), and there was very little change in potential temperature (2 K) from the surface to the level of maximum equivalent potential temperature, which was 90 hPa above ground level.

A surface low pressure center located in southeastern Colorado was moving northeast into the central Great Plains (not shown). A warm front slowly moved north across Kansas and Missouri and was dissipating. Rawinsondes from 0000 UTC 27 May, around the time of the convective initiation of interest, indicated a 500-hPa trough over the Pacific Northwest and a ridge centered over the western Great Lakes, placing the study area in southwest flow. A remnant MCS was located across north-central Kansas and south-central Nebraska at the beginning of the study period, and it traveled southeast before dissipating near Topeka, Kansas, around 1900 UTC. The simulation also showed the ongoing MCS, but it allowed the convection to stay intact well into Missouri.

The first cell that would quickly become a cluster was observed on radar at 2230 UTC in northwest Kansas, roughly 200 km northeast of the surface low (Fig. 8a). Radar also showed some convective cells in eastern Colorado, but that convection was not associated with the upscale evolution into the bow echo. Cellular convection was under way 2 h earlier in the simulation along the Front Range of Colorado, and the ongoing MCS was located in east-central Kansas, still attaining reflectivity values of 65 dBZ (Fig. 8b). The modeled convection in Colorado initiated in an environment with cooler dewpoint temperatures (10°C) than the observed convection in northwest Kansas (20°C).

By 0300 UTC 27 May, the observed CC had grown upscale into BE (Fig. 8c). The WRF model, however, maintained IC into northwest Kansas and southwest Nebraska. A few of the simulated cells merged, while others remained discrete during the time frame. Because any merging lasted less than 2 h, the classification of the system remained IC. The modeled system moved east-northeast into southern Nebraska, consistent with a southwest 2–6-km mean wind. This trajectory helped steer the cell into an area with negligible forcing and frontogenesis (Fig. 9a). The primary band of frontogenesis associated with the warm front was also located too far south in the simulation. The RUC analysis showed the frontogenesis along the Nebraska–Kansas border and stretching into the Kansas City, Missouri, area (Fig. 9b). The RUC 2–6-km mean flow also had a larger westerly component than what was indicated in the WRF output, which allowed the observed convection to ride along the warm front and to develop into BE.

By 0500 UTC, convection had quickly developed in eastern Nebraska north of the ongoing BE system. The new cells merged into BE approximately 90 min later to expand the line from Sioux City, Iowa, to Salina, Kansas. Trailing stratiform rain was also observed with the bow echo. The simulation quickly developed new cells in eastern Nebraska at 0600 UTC, and they merged into
a line within an hour; thus, the system was classified as NS (Fig. 8d). Extensive drying was evident in the layer from the surface to 750 hPa in addition to a small inversion directly above the surface (Fig. 10). Despite the stable conditions near the surface, most unstable CAPE values reached 1300 J kg\(^{-1}\). Higher potential temperatures were advection into the region from central and western Nebraska by 20 m s\(^{-1}\) 700-hPa flow. Much of western Iowa was under much weaker 700-hPa flow (5–10 m s\(^{-1}\)). Positive moisture advection in combination

![Fig. 7. WRF output at 0330 UTC 24 May 2006 for (a) line-normal storm-relative flow (m s\(^{-1}\)). (b) Skew T-logp diagram at the point of maximum front to rear flow.](image-url)
The microphysical scheme sensitivity tests led to more variation in convective mode and the distribution of precipitation than was present in the first case (not shown). The Lin scheme simulated individual cells at initiation, similar to the original run, but the cells dissipated 3 h before the NS formed. The WSM6 scheme produced CC at initiation, which matched the observed system, but it was located in Colorado instead of Kansas.

The Morrison scheme, which produced the most convection among the four schemes, and the WSM6 scheme transitioned cellular systems into NL events. Overall, all four microphysical schemes failed to produce the observed BE.

In the initialization sensitivity test, less convection was produced when GFS data were used, but the convective modes primarily remained the same. IC initiated and grew to NS later in the evolution without becoming a BE. Convective initiation was later in the GFS-initialized run compared to that using NAM (2200 UTC versus 2030 UTC), and the transition to NS occurred at 1000 UTC 27 May compared to 0600 UTC for the NAM-initialized run. No NL event was observed with the GFS-initialized run, possibly due to the late transition to NS. The GFS-initialized simulation featured much less frontogenetical forcing with the warm front and ongoing convection compared to the NAM-initialized run and the RUC analyses, in addition to smaller precipitable water values in the foothills of the Rocky Mountains.

### Table 9. As in Table 3, but for the 26–27 May 2006 case.

<table>
<thead>
<tr>
<th>$T_1$</th>
<th>$T_2$</th>
<th>$\Delta T$</th>
<th>Radar</th>
<th>Model</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.31</td>
<td>0.31</td>
<td>CC</td>
<td>IC</td>
<td>0.16</td>
</tr>
<tr>
<td>0.31</td>
<td>0.61</td>
<td>0.30</td>
<td>BE</td>
<td>IC</td>
<td>0</td>
</tr>
<tr>
<td>0.61</td>
<td>0.77</td>
<td>0.16</td>
<td>BE</td>
<td>NS</td>
<td>0.08</td>
</tr>
<tr>
<td>0.77</td>
<td>1</td>
<td>0.23</td>
<td>BE</td>
<td>NL</td>
<td>0</td>
</tr>
</tbody>
</table>

5. Conclusions

The present study investigated simulated convective morphologies and evolution for 37 warm season events. The classifications used to designate the convective modes for the events included three cellular, five linear, and a nonlinear mode used by Gallus et al. (2008) in addition to a mixed-complex mode added for the present study. Overall, the model produced more cellular modes than were observed, specifically CC, and too few linear modes, especially BE and TS.

A method using normalized time scales was devised to gauge the model’s accuracy in predicting convective mode with respect to radar observations. A “match” occurred if the simulated mode was the same as the observed mode at the same normalized time in the evolution, or if the two modes were in the same classification group. Of the 185 mode comparisons made, 58 were correct matches, and 104 were group matches. The model was least accurate in matching observed BE and TS events, and most accurate matching cellular modes. The total time in the event’s evolution in which a match was observed defined the event score. The average accuracy score for the 37 cases was 0.49, with 15 cases scoring at least 0.50. The model was also penalized if convective initiation or dissipation occurred more than 3 h different from that observed. Nine cases violated the timing criterion, and their average accuracy score was 0.32.

Statistical significance testing showed stronger 0–6-km bulk shear and cooler potential temperatures aloft before convective initiation tended to result in higher-accuracy
scores for the model. Meanwhile, weaker deep-layer shear, cooler surface potential temperatures, greater potential temperature lapse rates, and large differences between surface-based and most unstable CAPE were associated with cases having major timing errors in the simulations. The temperature-related parameters imply elevated convective situations were more difficult to simulate accurately. When similar parameters in the WRF were compared to the observations, little signal was present that increased skill in representing the mesoscale environment led to improved skill in simulating morphology. Instead, accuracy scores were generally better when the model overestimated CAPE. The model especially struggled to simulate morphology accurately when CAPE values were underestimated, especially MUCAPE. Poorly forecasted morphological evolution also tended to be associated with overforecasts of 0–6-km bulk shear.

Fig. 8. Reflectivity (dBZ) during convective mode transitions for the 26–27 May 2006 event. Observed reflectivity from UCAR imagery archive at (a) 2230, (c) 0300, and (e) 0830 UTC. Simulated reflectivity at (b) 2030, (d) 0600, and (f) 0830 UTC. The first green hue correlates to 30 dBZ, and the first yellow hue correlates to 40 dBZ (see color bar).
Two case studies were performed to investigate in detail the poor simulation of TS and BE events. The first case study, 23 May 2006, featured an observed TS system that traversed the northern Great Plains throughout the evening and overnight hours. The simulation successfully predicted the first two convective modes (BL and NS), but failed to develop the observed stratiform rain. It was shown the amount of deep-layer shear may have been a major factor in the simulation’s failure, as well as excessive drying in the midlayers of the atmosphere.

In the second event, 26 May 2006, CC evolved into BE while the simulation portrayed IC briefly becoming NS before transitioning to NL. The isolated nature of the
cells did not produce a sufficiently organized cold pool needed for mature linear development, and midlevel flow steered the convection away from frontogenetical forcing. Weak shear, little CAPE in the near-storm environment, the lack of a developed cold pool, and weak front-to-rear flow within the simulated system might have brought a quick end to the event’s linearity.

Microphysical sensitivity tests for the case studies showed problems simulating convective mode remained no matter what scheme was used. Initial condition sensitivity tests showed some variation with the convective modes, and more importantly, different data did allow the WRF to simulate the observed mode in the first case study. An underprediction of CAPE in the run using NAM data for model initialization, among other factors, might have contributed to the lack of stratiform rain development in the 23 May 2006 event. Mode varied less between the simulations in the second case study, but substantial differences in the amount of convection were present. Less frontogenetical forcing at the surface and smaller precipitable water values in the GFS data inhibited much of the cellular convection the NAM-initialized run produced.

The findings from the present study suggest a link between some observed environmental parameters before initiation and the accuracy of the model’s depiction of convective evolution. While these factors, such as shear and instability on the synoptic scale, are not solely responsible for the convective mode evolution, they can influence storm-scale processes that determine mode changes. It is unclear how much of the differences in these larger-scale processes in the WRF runs are a result of the initial and lateral boundary conditions, and how much of the differences are due to model errors within the WRF. Future work should examine the role of these errors in more detail. Future work also should examine the role of the planetary boundary layer scheme on morphological evolution since the model values of surface-based and most unstable CAPE often differed substantially from the RUC analyses. Sensitivity to model vertical resolution in the low levels of the atmosphere should be explored as well, as the case studies indicated the WRF model had difficulty producing adequate cold pools, which may be responsible for problems in simulating stratiform regions. Additional case studies could address other notable issues in the simulations, such as

**FIG. 10.** Skew $T$–log$p$ diagram from the WRF output at 0600 UTC 27 May near the NS event.
the WRF simulating CC events when BL were observed, which occurred in six cases, and the particular difficulty with MC events, as only one of six observed MC systems was correctly simulated.

Acknowledgments. The authors thank Jeffrey Duda for providing continuous assistance with the WRF, GrADS, and scripting, and Jon Hobbs for assistance with statistical significance testing. Daryl Herzmann and Dave Flory helped with computational resources. The constructive comments of three anonymous reviewers helped to improve the manuscript. Some RUC data were obtained from the Atmospheric Radiation Measurement (ARM) program sponsored by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Climate and Environmental Sciences Division. This work was supported by NSF Grant ATM0848200, with funds from the American Recovery and Reinvestment Act of 2009, and also NSF Grant ATM1222383. Some data processing was performed using Unidata software.

REFERENCES


