Comparison of Next-Day Convection-Allowing Forecasts of Storm motion on 1- and 4-km Grids

MICHAEL A. VANDENBERG
School of Meteorology, University of Oklahoma, Norman, Oklahoma

MICHAEL C. CONIGLIO
NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

ADAM J. CLARK
Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, and NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

(Manuscript received 17 January 2014, in final form 3 April 2014)

ABSTRACT

This study compares next-day forecasts of storm motion from convection-allowing models with 1- and 4-km grid spacing. A tracking algorithm is used to determine the motion of discrete storms in both the model forecasts and an analysis of radar observations. The distributions of both the raw storm motions and the deviations of these motions from the environmental flow are examined to determine the overall biases of the 1- and 4-km forecasts and how they compare to the observed storm motions. The mean storm speeds for the 1-km forecasts are significantly closer to the observed mean than those for the 4-km forecasts when viewed relative to the environmental flow/shear, but mostly for the shorter-lived storms. For storm directions, the 1-km forecast storms move similarly to the 4-km forecast storms on average. However, for the raw storm motions and those relative to the 0–6-km shear, results suggest that the 1-km forecasts may alleviate some of a clockwise (rightward) bias of the 4-km forecasts, particularly for those that do not deviate strongly from the 0–6-km shear vector. This improvement in a clockwise bias also is seen for the longer-lived storms, but is not seen when viewing the storm motions relative to the 850–300-hPa mean wind or Bunkers motion vector. These results suggest that a reduction from 4- to 1-km grid spacing can potentially improve forecasts of storm motion, but further analysis of closer storm analogs are needed to confirm these results and to explore specific hypotheses for their differences.

1. Introduction and motivation

Numerical weather prediction (NWP) models run with horizontal grid spacing $\leq 4$ km without convective parameterization are valuable weather forecasting tools. Compared to models that use coarser grid spacing with parameterized convection, convection-allowing models have been shown to produce more accurate precipitation forecasts (Weisman et al. 1997; Mass et al. 2002; Schwartz et al. 2009; Clark et al. 2009, 2012) and much more realistic depictions of severe thunderstorms, like bow echoes, hook echoes, and deviant storm motions (Weisman et al. 1997; Done et al. 2004; Kain et al. 2006; Weisman et al. 2008; Kain et al. 2010). However, because convection-allowing grid spacing is relatively coarse compared to the typical size of convective storms ($\sim 10$ km in diameter), the size of modeled storms is frequently overestimated (Weisman et al. 1997; Lean et al. 2008; Bryan and Morrison 2012; Johnson et al. 2013). This is because the model attempts to depict storms on the smallest resolvable scale (Bryan et al. 2003). Therefore, deep convective storms are generally depicted better on 1-km grids compared to 4-km grids owing to a better representation of nonhydrostatic processes associated.
with smaller updraft sizes (Weisman et al. 1997; Bryan and Morrison 2012), but few studies have found potential benefits to using a grid spacing <3–4 km for severe weather forecasting tools.

Studies showing benefits of 1-km grid spacing over 4-km grid spacing tend to be for general precipitation forecasts over topographically diverse regions (Colle and Mass 2000; Colle et al. 2005; Lean et al. 2008; Schwartz 2014). Across the United Kingdom, Roberts and Lean (2008) and Lean et al. (2008) found that a 1-km model performed better than coarser grid spacing for the timing of convective initiation, but precipitation forecasts were only improved for lower rainfall thresholds, which brings the utility of the results for convective weather forecasting into question. Over the United States, Kain et al. (2008) and Schwartz et al. (2009) compared the diurnal cycle of precipitation, simulated reflectivity, size distributions of distinct convective cells, and mesocyclone occurrence from 2- and 4-km NWP model forecasts of convection mostly east of the Rocky Mountains and west of the Appalachians. Subjectively, the 2-km forecasts produced more detailed convective structures, though there did not appear to be significant forecast skill for those objects resolvable on the finer grid. For mesoscale features and precipitation forecasts, both of these studies found that the 2- and 4-km forecasts had similar levels of skill for next-day forecasts. Ultimately, it was concluded that any additional value provided by using a grid spacing of 2 km as forecast guidance for these forecasts ranges may not be worth the increased computational expense. Similarly, Clark et al. (2012) discussed the finding that 1-km model forecasts do not appear to provide more accurate depictions of convective evolution than do 4-km model forecasts as determined from side-by-side subjective comparisons of hourly simulated reflectivity forecasts compared to the observed reflectivity over regional areas.

While Kain et al. (2008), Lean et al. (2008), and Schwartz et al. (2009) provided valuable insight into how forecasts of convection are affected by grid spacing, these studies focused primarily on forecasts of accumulated precipitation—a measure of the aggregate effects of convection—and have given relatively little attention to the potential benefits of smaller grid spacing on the properties of individual convective cells. Verifying model forecasts using accumulated precipitation may mask the benefits of the increased resolution on the depiction of individual storm characteristics. Furthermore, Bryan and Morrison (2012) found that surface precipitation from a squall line simulated on 1- and 4-km grids was similar because of compensating effects (both total condensation and total evaporation are higher on a 1-km grid).

In the comparisons mentioned above, other than storm size, there were no attempts to objectively verify object attributes of storms. A promising step in this direction was taken by Johnson et al. (2013), in which the general size, shape, and orientation of storms were verified. They generally found better forecasts of convection from the 1-km grid, but only for those storms that are not resolvable on the 4-km grid (i.e., the smallest storms). There are many storm attributes other than size, shape, and orientation that may be helpful for severe weather forecasting applications, like storm mode (discrete versus linear), storm rotation, or storm motion, that have yet to be verified objectively over many events.

The motion of storms is important to depict accurately both for the direct benefits to forecasts of storm evolution (Johns and Doswell 1992) and because the ability of storms to acquire rotation from environmental streamwise vorticity (or storm-relative environmental helicity, SREH) is directly related to storm motion (Davies-Jones 1984). Generally, as motion away from the shear vector increases, so does SREH, and larger values of SREH have been clearly related to a greater potential for supercells and tornadic storms (Markowski et al. 2003; Thompson et al. 2007). Therefore, if model forecasts of convection suggest (accurately) that storms will move more deviantly than anticipated, then this may alert forecasters that deviant storms may realize more streamwise vorticity and become more severe than anticipated.

In this study, storm motion is used to verify model forecasts with 1- and 4-km horizontal grid spacing. Storm motion is a storm (object) attribute that has yet to be examined in a comparison of 1- and 4-km models. Furthermore, storm motion is used as a metric because it may be a proxy for the potential significant improvement in resolving individual convective storms on a 1-km grid through the propagation component of storm motion. The propagation component of storm motion is directly related to processes internal to the storms associated with the well-known midlevel shear/updraft mechanisms outlined in Weisman and Klemp (1982, 1984) and Rotunno and Klemp (1982, 1985). Storms tend to move away from the mean winds, often to the right of the mean winds for severe discrete convective cells (supercells) (Browning 1964) because of these mechanisms. As the cold pool of air generated by the storm’s downdrafts expands, shear–cold pool interactions and variations in storm-relative inflow may add complexity to the propagation component of supercells (Corfidi 2003; Weisman and Rotunno 2004). So if the 1-km grid is resolving storms better, then these internal storm processes that affect storm motion may be resolved better, and this improvement may be reflected by better predictions of storm motion. It should
be noted that a comparison of storms depicted on 1- and 4-km model grids could be done by examining directly the internal storm characteristics, like rotation, dynamic pressure perturbations, or cold pools. However, this would involve considerably more effort than examining the storm motion. Perhaps more importantly, there is no observational counterpart against which to compare the strength of dynamic pressure perturbations or mesocyclone/cold pool strength over a large number of cases.\(^2\)

The goal of this study is to determine if there are differences in the overall biases of individual storm motions using the same model configured for real-time prediction of convection, but with 1- and 4-km grid spacing. As outlined in sections 2 and 3, this is accomplished through an analysis of storm speeds and directions in the observations and model forecasts and those relative to the near-storm environmental flow. The goal of this study is not to identify the specific internal processes on the 1-km grid that may be handled differently than on a 4-km grid, but to first determine if there is a difference in storm motion forecasts when using a convection-allowing grid finer than 4 km. If differences in storm motion are identified, then further studies would be needed to identify the particular processes responsible for those improvements, some of which are hypothesized in section 3.

2. Methodology

a. Numerical model experiments

Our sample dataset was chosen from storm events observed on 22 separate days over regional areas of the central United States in 2010 (Table 1). In 2010, the University of Oklahoma’s Center for Analysis and Prediction of Storms (CAPS) produced several Weather Research and Forecasting Advanced Research Core \(version 3.1.1\) (WRF-ARW 3.1.1; Skamarock et al. (2008)) model forecasts on convection-allowing grids covering the contiguous United States to support part two of the Verification of the Origins of Rotation in Tornadoes Experiment (VORTEX2) and the weekday operations of the National Oceanic and Atmospheric Administration (NOAA) Hazardous Weather Testbed (HWT) Spring Forecasting Experiment (SFE) [see Clark et al. (2012) for more information on this SFE]. Among the suite of model forecasts produced by CAPS, the two that varied only by their grid spacing (1 and 4 km) are the subject of this study. All other aspects of the simulations were identical, including the initialization procedure, domain sizes, number of vertical levels (51), and physics options from WRF-ARW 3.1.1 that included Thompson cloud microphysics (Thompson et al. 2004), Goddard shortwave

\(^2\) Radar-based estimates of observed storm rotation are just now becoming sound enough to begin using them in the verification of convection-allowing models (Miller et al. 2013; Clark et al. 2013), but they have been used for defining rotation tracks, not their intensity.
radiation (Chou and Suarez 1994), Rapid Radiative Transfer Model (RRTM) longwave radiation (Mlawer et al. 1997), Noah land surface physics (Niu et al. 2011), and Mellor–Yamada–Janjić turbulent mixing (Mellor and Yamada 1982; Janjić 1994). These forecasts were initialized at 0000 UTC on each weekday between 1 May and 18 June and produced 30-h forecasts (to 0600 UTC the following day). The reader is referred to the online supplement for Clark et al. (2012, which is available online at http://dx.doi.org/10.1175/BAMS-D-11-00040.2) for more details on these model experiments.

b. Processing of simulated and observed reflectivity

An analysis of observed composite (column maximum) radar reflectivity is provided on a 0.01° × 0.01° grid by the National Severe Storms Laboratory (Zhang et al. 2011) and is smoothed to the 4-km model grid using a two-dimensional Gaussian filter. Likewise, simulated composite reflectivity forecasts from the 1-km forecasts are smoothed to the 4-km grid prior to their verification. The smoothed 1-km forecasts are created by averaging all the reflectivity values from the 1-km forecast within a 4 × 4 grid cell area that encompasses the same area on the 4-km grid (see Fig. 1 for an example of simulated reflectivity before and after the smoothing). The philosophy for smoothing the radar analysis to the 4-km model grid is to have an analysis that is comparable to the coarsest model grid in the evaluation, similar to the verification approach of Clark et al. (2009). Using radar observations analyzed on a 1-km grid would unfairly handicap the 4-km model forecasts as the latter would not be able to resolve a significant portion of the convection on the 1-km observation grid. Using an analysis of radar observations on a 1-km grid to verify the 1-km model forecasts and a separate analysis of the radar observations on a 4-km grid to verify the 4-km model forecasts is a reasonable approach [as was done in Johnson et al. (2013)], but this would introduce two different “truths” into the comparison. The verification method used here is an effective way to assess the value that 1-km forecasts may add to the 4-km forecasts by using an analysis and forecasts of convective features with similar spatial scales.

c. Storm-tracking algorithm

One of the most important components of this research is the storm-tracking algorithm written by the third author. Composite reflectivities of 5 min are used as input into the tracking algorithm to define storm tracks for the WRF-ARW 4-km model forecasts (W4), the smoothed 1-km model forecasts (S1), and the observations. The storms tracked are mostly discrete (often supercells) but are sometimes well-defined cells embedded within lighter precipitation associated with convective systems.

The first step in the algorithm is to identify storm objects, defined as contiguous regions of reflectivity exceeding a certain threshold, at each 5-min output time. To decide on an appropriate reflectivity threshold, plots of cell tracks for three thresholds—45, 50, and 55 dBZ—were overlaid with the observed and simulated reflectivity images. The algorithm with a 55-dBZ threshold did not identify cells early in their life cycle, particularly for the model forecasts, which have a low bias for high reflectivity (Stratman et al. 2013). A 55-dBZ threshold also did not allow the algorithm to identify weaker cells that nonetheless were very clearly isolated supercells. On the other hand, while a 45-dBZ threshold worked well for those events containing moderately weak, widely spaced cells, this threshold tracked far too many cells that were embedded in larger mesoscale convective systems (MCSs). Ultimately, it was determined that a 50-dBZ threshold resulted in a reasonable number of tracked cells based on a visual inspection of the tracks and the reflectivity in the majority of cases.

After storm objects are identified for each time, the objects have to be associated across time to determine tracks and to account for storm splits and mergers. In the simplest case (i.e., no splits or mergers), a single object present at one time overlaps with, or is adjacent to, a single object at the next time, and these objects are considered part of the same track. Once the algorithm begins identifying a storm track, it searches through each subsequent time until the end of the track is found. A storm split occurs when more than one future object overlaps with, or is adjacent to, a current object. In this case, the largest future object is considered a continuation of the track of the present object, while the smaller overlapping/adjacent future objects are considered the potential starting points of new tracks. A storm merger occurs when more than two objects at the present time overlap with, or are adjacent to, one object at the future time. In this case, the track of the largest object at the present time is continued to the future time, while the tracks of any smaller objects at the present time stop. Once a track has been defined, the objects composing the track cannot be considered part of any other track.

To evaluate the tracking algorithm, for each of the 22 events, two observed cell tracks are defined manually, independent of the algorithm, and then compared to tracks produced by the algorithm (examples of comparisons between the manual and algorithm-produced tracks are shown in Fig. 2 for four discrete cell tracks over two events). Though the algorithm clearly excels at tracking discrete cells, manual quality control was performed for many tracks because the algorithm occasionally has difficulty defining tracks when one cell dies near another cell that is forming and for cells that are
near or part of convective systems. The most glaringly erroneous (i.e., erratic) tracks are simply deleted from the dataset, while many of the remaining tracks are smoothed subjectively (Fig. 3). For example, the tracks with number labels in Figs. 3b–d are those that were removed because the algorithm tracked cells that grew upscale into a convective system and tried to continue the initial track despite the cell’s loss of a discrete identity within a convective line. Furthermore, if storms occurred in a particular forcing regime in the real atmosphere but they did not occur in that regime in one or the other model forecasts (or vice versa), then these storm tracks were removed from the analysis. For example, the two storm tracks in east-central Kansas in Fig. 3a occurred very near the surface position of a warm front and appeared to be sustained by surface-based air in the southerly low-level flow for at least a portion of their lifetime, whereas the other storms in northwest Kansas
FIG. 2. Manual tracks (black) overlaid on a portion of the raw algorithm-produced tracks (gray) for the (a) 18 May and (b) 2 Jun events. Plots like these were made for all 22 events to examine the ability of the tracking algorithm to define storm tracks.
and northwest Missouri were more clearly elevated above the warm front. Neither of the two model forecasts predicted the surface-based storms in east-central Kansas and thus these observed storm tracks were removed from the dataset. Figure 3 illustrates that, overall, only the more well-defined discrete storm tracks are retained.

d. Analysis of storm motion

The period used for tracking the cells spans from 0600 to 0600 UTC the following day (6–30-h forecasts). Storm events that began between 0000 (the start of the model forecast) and 0600 UTC the same day are not used to avoid issues with model spinup (Kain et al. 2010; Stratman...
et al. 2013). Regional verification domains, typically sized about 450 km × 450 km (Figs. 2 and 3), are defined for each case to focus on individual storm systems. For example, the cluster of discrete supercells along a dryline in Oklahoma and southern Kansas ahead of a strong upper-level short-wave trough on 10 May 2010 (Fig. 1) is an example of what is considered a storm system (with the regional domain for this day enclosed by the dashed lines in Fig. 3). On days with multiple storm systems to choose from, preference is given to those systems that produced the most storms. Furthermore, it is reiterated that only storm systems in which storms occurred in both of the model forecasts and in the observations were included in the analysis.

The shortest track that is considered is 1 h. The above procedures result in 259 tracks for the radar observations, 238 tracks for W4, and 208 tracks for S1 to be analyzed. It should be noted that the raw 1-km forecasts produced more discrete storms than did the 4-km forecasts but they were usually smaller (as in Lean et al. (2008) and Johnson et al. (2013)). The averaging from the 1- to 4-km grids effectively smooths some of these smaller storms enough so that they are no longer identified at the 50-dBZ threshold used to define storm objects. So, the storms analyzed from the 1-km forecasts tend to be the larger storms that it produces.

Average storm directions and speeds are computed for each track. To assess the statistical significance of differences between the model forecasts and observations, a two-tailed Wilcoxon–Mann–Whitney rank sum test (Wilks 2006, 156–161) is used with the number of tracks for each set assumed to be the number of independent samples. Throughout the paper, the $p$ value resulting from applying this test is stated and estimates the probability that the location (the nonparametric analog of the mean) is the same between the two distributions in question (low $p$ values correspond to higher significance that the means are different) (Wilks 2006, 156–161).

3. Results and analysis

a. Comparison of modeled and observed tracks

In this section, the raw storm speeds and directions are examined, while in section 3b, the storm speeds and directions relative to the background environment are examined. The raw average storm speeds and directions are first binned into 5 m s$^{-1}$ and 20° intervals, respectively. Without considering the environmental flow, no large differences are found for the overall mean storm speeds (Fig. 4a). The mean observed storm speed of 13.5 m s$^{-1}$ is slightly overestimated by both W4 and S1 with mean speeds of 14.0 m s$^{-1}$ ($p = 0.21$) and 14.6 m s$^{-1}$ ($p = 0.12$), respectively (Fig. 4a). However, inspection of the speed distribution shows that S1 follows the observed storm speed distribution closer than W4. The relative frequency of speeds for S1 is either nearly the same or closer to the observed distribution than W4 for all of the bins (Fig. 4a). In other words, W4 has an overall mean speed that is closer to the observed mean,
but the distribution of speeds for W4 is narrower and has higher amplitude near the mean than does the distribution for S1.

The distribution of storm directions for the observations and model forecasts is more complex (Fig. 4b). In some bins, W4 is closer to the relative frequency of the observations while in others S1 is closer to the observed relative frequency. The mean direction for W4 (245°) is larger than the mean for the observations (239°) ($p = 0.03$). A mean error of +6° for W4 and a distribution for W4 that is shifted to the right compared to the observed distribution (Fig. 4b) indicates a clockwise (rightward) bias for storm directions in W4. This clockwise shift in W4 is most apparent in the 180°–200° and 280°–300° bins (cf. the blue and black lines in Fig. 4b). For S1, the mean direction is ~243° and is closer to that for the observations ($p = 0.25$). Although the mean differences between W4 and S1 are not very large, the distribution for S1 may be alleviating some of this clockwise bias in W4, as seen by frequencies for S1 that are closer to the observed frequency in the 180°–200°, 240°–260°, and 280°–300° bins.

b. Storm track–environment deviations

The analysis of storm tracks is taken further by defining the tracks relative to the kinematic environment. When comparing different model forecasts of storm motion to observations, it is important to take the differences in the background environments into account as environmental differences can contribute to differences in storm motion (i.e., if a simulated storm moves at 270° and the real storm moves at 290°, it is important to make sure the difference in direction cannot simply be attributed to a different flow or shear in the environment). Hourly Rapid Update Cycle (RUC) analyses on a 20-km grid (Benjamin et al. 2010) are used to define the background environmental flow for the observations. Four variables related to the motion of storms are computed: the 0–6-km shear vector, the 850–300-hPa mean wind vector, the Bunkers motion for right-moving supercells (Bunkers et al. 2000), and the Rasmussen and Blanchard (1998) supercell motion.

Data storage constraints limited the domain-wide WRF-ARW model data to the surface and to 850, 700, 600, 500, and 250 hPa. For calculation of the RUC-analyzed environmental flow vectors listed in the previous paragraph, only data from these six levels from the RUC pressure-level data are used to allow for an equitable comparison to the WRF-ARW model environments. The average vectors within near-storm (40 km × 40 km) regions (boxes) are used to represent the near-storm conditions. For each cell, these boxes are translated along with the cell every hour with the cell placed in the top-left corner of the box; the box is then translated along the cell motion vector so that the possible contamination of the ambient environment by the storms is minimized. The 40 km × 40 km boxes are different between S1, W4, and the observations because storms were in different locations. RUC and WRF-ARW forecast fields nearest to the hour of the 5-min time step of the track are used. It should be clarified that the purpose of calculating these flow vectors is not to find as accurate a storm motion predictor as possible, but simply to define vectors that can be used to compute storm motion in coordinates relative to the nearby kinematic environment. While the use of only six levels may smooth details of the hodograph and may result in a suboptimal environment-based predictor of storm motion (as pointed out by an anonymous reviewer), the use of six levels is sufficient to characterize the mean near-storm flow/shear and how it varies horizontally across the regional domains for the purpose of normalizing the storm motions to their near-storm environment.

The larger meso- and synoptic-scale environments of the WRF-ARW model forecasts are quite similar to the RUC-analyzed environments over the regional domains (not shown). This is expected because convection-allowing models typically predict larger-scale mass fields quite well in the first 24–30 h of the forecasts (Weisman et al. 2008; Coniglio et al. 2010). Some differences do exist in both the thermodynamic and kinematic environments of the storms, but the focus here is on the kinematic environments. Along with the small but potentially nonnegligible differences in the kinematic environments between the WRF-ARW model forecasts and the RUC analyses, the forecasted and observed cells do not occur in exactly the same locations in the regional domains (see the example in Fig. 1), so the locations sampled by the near-storm domains vary from the observations and model forecasts because of the horizontal variability in the environment.

---

3 All directions referred to hereafter are for their standard meteorological convention (a storm moving from west to east has a 270° direction).
To ensure that the differences in the kinematic environments that interact with the storms do not significantly impact the comparison, deviations of the modeled storm tracks from the WRF-ARW environment vectors are then compared to deviations of the observed storm tracks from the RUC-analyzed environment vectors. In this comparison, the smaller the difference in the deviations of the storm motions from their environment between the model forecasts and the observations, the more accurate the forecast. In the convention used here a positive speed deviation indicates that the storm moved faster than the environment vector magnitude, while a positive direction deviation indicates that the storm moves clockwise (to the right) from the environment vector direction.

1) STORM SPEED DEVIATIONS

First, when viewed relative to the 0–6-km shear vector magnitude (Fig. 5a), the mean speed deviation for S1 (−8.7 m s$^{-1}$) is closer to the observed mean (−9.7 m s$^{-1}$) than that for W4 (−7.6 m s$^{-1}$); the mean speed deviation for W4 is larger than the mean speed deviation for the observations at the 99% level ($p < 0.01$). Furthermore, the $p$ value for comparing the W4 mean speed deviations to the S1 mean speed deviations also is small ($p = 0.03$). The $p$ value for S1 compared to the observations is much higher and not statistically significant (0.68). Therefore, the S1 storms may be alleviating some of the fast speed bias in W4 as seen by the shift in the peak of the distribution to the left, closer to the observed shape distribution.

Similar conclusions are made when viewing the deviations relative to the 850–300-hPa mean wind speed (Fig. 5c) and to the Bunkers motion magnitude (Fig. 5e) (the Rasmussen and Blanchard supercell motion estimate is not shown as the results are similar to those for the Bunkers motion estimates). For the 850–300-hPa mean wind speed deviations, the overall means between the observations and S1 are nearly the same while that for W4 is much larger than that for the observations ($p = 0.004$). Although the overall S1 and W4 means for the Bunkers motion magnitude are closer to each other than the observations (Fig. 5e), the magnitude distribution for S1 clearly follows the shape of the distribution for the observed Bunkers motion magnitude deviations more closely than that for W4 in the range from −5 to +5 m s$^{-1}$, in which about 80% of the cases occur. A few outliers in very large Bunkers motion speed magnitudes for S1 (>15 m s$^{-1}$) skew the S1 mean toward positive values.

Further analysis shows that the short-lived (<2 h) storms contribute the most to this improvement in storm speeds for S1 over W4 (about 65% of the storms fall into this category) (Fig. 6a). As illustrated by the speed deviations from the 0–6-km shear vector (Fig. 6), both the mean speeds and the distributions for the longer-lived storms show little difference overall between W4 and S1 (Fig. 6c). Deviations from the 850–300-hPa mean wind speed and Bunkers motion magnitude show a similar result (not shown). This indicates that the storms in W4 move faster than the observed storms mostly for the short-lived storms, while the S1 and W4 storms tend to move at a speed similar to that for the observed storms for the longer-lived storms. The qualitative similarity for the speed-deviation results among the environmental flow variables gives confidence in the robustness of these results for the speed deviations.

2) STORM DIRECTION DEVIATIONS

For all of the environmental flow variables, the overall mean direction deviations for W4 and S1 are similar to each other (no differences with $p$ values smaller than 0.15 are found). The 0–6-km shear vector direction deviations are close to zero (Fig. 5b) whereas those for the 850–300-hPa mean wind direction are $\sim +15^\circ$ (Fig. 5d), indicating storms tend to move clockwise, or to the right of this vector, and those for the Bunkers motion direction are $\sim -15^\circ$ (Fig. 5f), indicating storms tend to move counterclockwise or to the left of this vector. However, unlike for the speed deviations, the interpretation of the distribution of the direction deviations varies with the underlying environmental flow variable.

Some differences in the distributions for S1 and W4 are found for the 0–6-km shear vector deviations (Fig. 5b). In the bins from −30$^\circ$ to +30$^\circ$ (where about 60% of the total tracks occur), the distribution of the S1 direction deviations follows the observed direction deviations closer than that for W4. These findings generally agree with Fig. 4b, which suggested that S1 may reduce the clockwise bias seen in W4. However, S1 overestimates the frequency of storms that deviate strongly to the right of the 0–6-km shear vector ($> +30^\circ$), and also underestimates the frequency of storms that deviate moderately to the left of the 0–6-km shear vector (in the bin from −30$^\circ$ to −45$^\circ$). The overall mean direction deviation for S1 is more clockwise than the overall mean direction deviation for W4 (shown by the vertical dashed lines in Fig. 5b) because this rightward bias in the tails skews the overall mean toward larger deviations. However, in the range of direction deviations from −30 to +30$^\circ$, the mean deviations for the observed storms is −0.5$^\circ$, whereas the mean deviations for S1 and W4 are 0.1$^\circ$ and 4.4$^\circ$, respectively. The $p$ value for the mean differences between W4 and the observations is very small ($p = 0.006$), as is the $p$ value for the mean differences between W4 and S1 ($p = 0.02$). Conversely, the $p$ value for the comparison of S1 to the observations is much larger ($p = 0.74$). This...
shows that in the range of small to moderate deviations from the 0–6-km shear vector direction (from $-30^\circ$ to $+30^\circ$), where the majority (about 60%) of the storm tracks occur, W4 has a rightward bias that is not seen in S1. However, the S1 and W4 direction deviation distributions for the 850–300-hPa mean wind and Bunkers motion are more like each other than the observations for most of the bins, and both display a rightward bias compared to the observations. This rightward bias is seen for both right and left movers (Figs. 5d,f). In other words, the improvement in S1 over W4 in the bins from $-30^\circ$ to $+30^\circ$ seen for the 0–6-km shear vector direction deviations is not apparent for the 850–300-hPa mean wind direction and Bunkers motion direction. The reasons why direction deviations from these two variables are different than those for the 0–6-km shear direction.
deviations are not clear. Distributions of the raw 850–300-hPa mean wind and Bunkers motion directions within the near-storm domains are very similar in the WRF-ARW forecasts and RUC analyses and do not reveal any indications of why deviations from one vector would display characteristics that are different from another. It could be argued that the 0–6-km shear vector is the best way to define the environmental flow within a framework for storm motion because this vector points in roughly the direction that discrete storms can be expected to move, which is shown in Fig. 5b by a mean deviation near 0° for the observed and forecasted storms. Furthermore, storms that split and move apart (left and right movers) are generally defined relative to the direction of the 0–6-km shear vector (Klemp and Wilhelmson 1978; Weisman and Klemp 1982). However, further analysis is needed to determine if the differences between S1 and W4 in the range from −30° to +30° for the 0–6-km shear deviations are robust because the results for the 850–300-hPa mean wind direction deviations and for the Bunkers motion direction deviations suggest a different interpretation. However, there is higher confidence overall that both S1 and W4 tend to display a rightward bias compared to the observations when considering both the raw and environment-based analyses together.

c. Further analysis

The reasons why the W4 storms tend to move faster than the short-lived S1 storms and observed storms are not clear. Compared to 4-km grids, 1-km grids have been applied to...
shown to produce stronger cold pools, higher cloud tops, more condensation, more evaporation, and stronger upward mass fluxes (Weisman et al. 1997; Bryan and Morrison 2012). However, it is not immediately clear how any of these factors may translate into slower-moving storms. Exploring these reasons would require delving into the 5-min model output in much more detail than intended for this study, but one possibility is related to a hypothesis mentioned earlier. Storms that deviate to the right of the low- to midlevel shear tend to move slower than the shear magnitude because of a rightward propagation component induced by a nonhydrostatic pressure perturbation on its right flank that opposes the low- to midlevel winds (Rotunno and Klemp 1982). Similarly, storms that deviate to the left of the low- to midlevel shear tend to move faster than the shear magnitude as a leftward propagation component induced by a nonhydrostatic pressure perturbation on its left flank augments the low- to midlevel winds. Therefore, if these processes are important and S1 storms are moving slower overall than the W4 storms, then the S1 storms should also be moving clockwise (to the right) compared to the W4 storms. Indeed, the S1 storms that show the greatest difference in speed from W4 (those that last <2 h) also show a ~4° clockwise (rightward) bias compared to the W4 storms (~5.6° for S1 and ~9.7° for W4; see Fig. 6b), although the statistical significance of this difference is not large ($p = 0.56$). The observed deviation for these short-lived storms is ~11.5°, indicating that W4 is closer to the observed direction for these shorter-lived storms, so even though the speeds for the short-lived storms may be predicted better by S1, the directions may not be predicted better.

Again, the S1 storms may be slower than the W4 storms, mostly for the shorter-lived storms. As also discussed earlier, for the storms that do not deviate from the low- to midlevel shear more than 30°, W4 may have a rightward bias compared to the S1 and the observations (when viewing the results for both the raw storm directions and the deviations from the 0–6-km shear vector). If only the longer-lived (>2 h) storms in the bin from ~30° to +30° are examined, W4 also is found to have a rightward bias compared to S1 and the observations (Fig. 6d), with a deviation of 3.2° for S1 compared to 9.4° for W4 ($p = 0.03$). Regarding these longer-lived storms, it is reasonable to presume that the longer a storm lives, the more cold-pool processes might influence its evolution. Likewise, the less a storm deviates from the low- to midlevel shear direction, the less likely it is to be a supercell and influenced by the updraft/shear propagation mechanism. Therefore, for the longer-lived storms in this bin from ~30° to +30°, cold-pool processes may have a greater influence than internal updraft/shear effects the longer a storm lives, particularly if it is non-supercellular. Indeed, storms with substantial cold pools that move at a small angle to the 0–6-km shear are in a favorable configuration for forward propagation of storms (Corfidi 2003; Cohen et al. 2007). Then, if cold pools become stronger in the 1-km forecasts than in the 4-km forecasts, as shown to be the case in Bryan and Morrison (2012), the stronger cold pools could be a reason why S1 tends to deviate less from the low- to midlevel shear than the W4 forecasts for the longer-lived storms in the range from ~30° to 30°. Stronger cold pools could favor downshear propagation relative to any right-flank propagation. The fact that the longer-lived storms in the range from ~30° to +30° in W4 have a rightward bias compared to observations to a high significance (9.4° for W4 compared to 1.6° for the observations; $p = 0.005$) suggests that W4 may be overdoing the updraft/shear effects compared to the S1 forecasts for the longer-lived storms that do not deviate more than 30° from the 0–6-km shear direction. Inspection of the reflectivity images and hourly temperature fields on the lowest model level from S1 and W4 suggests that cold-pool generation and forward propagation of cells is more prevalent in S1 than in W4, which supports this claim. However, again, a more detailed analysis of storm structures and processes is required to explore these hypotheses in depth.

4. Summary and conclusions

In this study, the ability of WRF-ARW model simulations at grid spacings of 1 and 4 km to predict observed storm motion over the plains of the United States during the 2010 convective season is examined. The motion of storms is used as the means to compare the model forecasts primarily because it may be an efficient way to gauge if convective storms are resolved differently on a 1-km grid than on a relatively coarse 4-km grid as storm motion is somewhat dependent on internal storm processes. Individual convective cells are tracked in both an analysis of radar observations and in the simulated reflectivity fields from the WRF-ARW model forecasts every 5 min using an object-based tracking algorithm. The observed reflectivity and simulated reflectivity from the 1-km model forecasts are smoothed to the 4-km grid prior to tracking the cells to ensure that approximately the same spatial scales are represented in the analyses. These storm tracks are compared both directly and relative to the kinematic environment to remove influences of small but potentially nonnegligible differences in kinematic environments among the model forecasts and observations.
Examination of the raw storm tracks did not reveal any large differences in the 1- and 4-km forecasts except for a possible alleviation of a fast speed and rightward bias in the 4-km forecasts in the middle of each distribution. When the storm tracks are viewed relative to the 0–6-km shear, mean wind, and Bunkers motion in the near-storm environment, it is found that the 1-km model forecasts of storm speed more closely approximate the speed of the observed storms when compared to the 4-km model, especially for the shorter-lived storms. Furthermore, when considering those tracks that only deviate <30° from the 0–6-km shear vector, the 1-km forecasts have a significantly smaller mean direction error compared to the 4-km forecasts. In this range of direction deviations, the 1-km forecasts may indicate an alleviation of a clockwise (rightward) bias seen in the 4-km forecasts for both left and right movers. However, the overall mean direction errors for the 1-km model are larger than those for the 4-km model, because of the larger number of events in the tails of the direction distribution for the 1-km model. Furthermore, the better direction forecasts evident in the 0–6-km shear deviations are not seen in the 850–300-hPa mean wind and Bunkers motion deviations, which casts some doubt on the robustness of the environment-relative results for the storm direction comparisons.

It is possible that the shorter-lived storms (about 65% of the total dataset) in the 1-km model forecasts are slower and more clockwise oriented compared to the 4-km forecasts because S1 handles the propagation of the shorter-lived discrete storms differently. However, these same S1 storms that move slower and to the right of the W4 storms also tend to move too far to the right of the observed shorter-lived storms, indicating that S1 may be overestimating (underestimating) right-moving (left moving) supercell modes when cold pools are not too strong. On the other hand, the 1-km forecasts have a counterclockwise (left) bias compared to W4, and the 1-km forecasts are significantly closer to the observed storms, when considering the longer-lived storms that do not deviate more than 30° from the 0–6-km shear vector direction. The longevity of the storms and the smaller direction deviations from the shear suggest more of a cold-pool influence in these storms. The rightward (and slow) bias seen in W4 for these storms suggests it is overdoing the updraft/shear influence on storm propagation in these situations when cold pools and downshear forward propagation along them are important (as is the case when storms do not deviate more than 30° from the 0–6-km shear).

The analyses presented here consider every track within regional areas regardless of where the modeled tracks are positioned relative to the observed tracks. Because the same storm systems are analyzed on each day, this provides a good overview of the biases of storm motion forecasts compared to the real storms within the same storm systems. However, it is of interest to compare specific tracks that could be considered the “same” storm among at least two of the datasets so that more direct error comparisons can be made and the hypotheses stated above could be addressed more directly. To do this, an attempt was made to find analogous tracks by looking for individual storm tracks that formed within roughly an hour and 100 km of each other, and formed in nearly the same location relative to features in the environment (drylines, instability axes, etc.). While there are many tracks that fit these criteria (by virtue of restricting the analysis to the same storm systems), it is rarely evident that any individual storm in one model forecast or the observations is the “same” storm in the other model forecast. Occurrences of these directly analogous storm tracks are relatively rare likely because the range of forecasts considered (15–36 h) is relatively long for convection-allowing grid scales. On these times scales, despite meso- and synoptic-scale environments that are generally similar, small-scale errors have already grown rapidly to larger scales (Hohenegger and Schär 2007; Zhang et al. 2007) and realizations of the convective event on the smallest resolvable scales are different among the model forecasts. Therefore, an analysis of the few analogous tracks that were identified is not presented here. It is suggested that forecasts over shorter time scales (0–12 h) should be used for a comparison of analogous tracks, preferably those that start from an initial set of conditions derived from the cycled assimilation of radar data, in which the storms analyzed in the initial conditions continue smoothly into the forecasts [see Stensrud et al. (2013) for experiments of this type]. The forecasts used for this analysis presented herein are not sufficient to do this as the three-dimensional variational data assimilation (3DVAR)/cloud analysis procedure that was used for the WRF-ARW forecasts (Clark et al. 2012) has difficulty retaining convective-scale information more than a few hours into the forecast [see the discussions in Kain et al. (2010) and Stratman et al. (2013)]. Short-term forecasts of directly analogous storms on 1- and 4-km grids could then be used to more closely examine specific reasons for the potential improvement in storm motion forecasts seen on 1-km grids that are only hypothesized herein. Likewise, the benefits of using 1-km forecast grids on individual storms are likely to be seen over shorter time scales than those used in this study.

Acknowledgments. This work was made possible by a Presidential Early Career Award for Scientists and Engineers and through the hard work and dedication of
REFERENCES


