

## Some Practical Considerations Regarding Horizontal Resolution in the First Generation of Operational Convection-Allowing NWP

JOHN S. KAIN,\* STEVEN J. WEISS,<sup>+</sup> DAVID R. BRIGHT,<sup>+</sup> MICHAEL E. BALDWIN,<sup>#</sup> JASON J. LEVIT,<sup>+</sup> GREGORY W. CARBIN,<sup>+</sup> CRAIG S. SCHWARTZ,<sup>@</sup> MORRIS L. WEISMAN,<sup>&</sup> KELVIN K. DROEGEMEIER,<sup>@,\*\*</sup> DANIEL B. WEBER,<sup>\*\*</sup> AND KEVIN W. THOMAS<sup>\*\*</sup>

\*NOAA/OAR/National Severe Storms Laboratory, Norman, Oklahoma

<sup>+</sup>NOAA/NWS/Storm Prediction Center, Norman, Oklahoma

<sup>#</sup>Department of Earth and Atmospheric Sciences, Purdue University, West Lafayette, Indiana

<sup>@</sup>School of Meteorology, University of Oklahoma, Norman, Oklahoma

<sup>&</sup>National Center for Atmospheric Research, Boulder, Colorado

<sup>\*\*</sup>Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma

(Manuscript received 15 December 2007, in final form 5 March 2008)

### ABSTRACT

During the 2005 NOAA Hazardous Weather Testbed Spring Experiment two different high-resolution configurations of the Weather Research and Forecasting-Advanced Research WRF (WRF-ARW) model were used to produce 30-h forecasts 5 days a week for a total of 7 weeks. These configurations used the same physical parameterizations and the same input dataset for the initial and boundary conditions, differing primarily in their spatial resolution. The first set of runs used 4-km horizontal grid spacing with 35 vertical levels while the second used 2-km grid spacing and 51 vertical levels.

Output from these daily forecasts is analyzed to assess the numerical forecast sensitivity to spatial resolution in the upper end of the convection-allowing range of grid spacing. The focus is on the central United States and the time period 18–30 h after model initialization. The analysis is based on a combination of visual comparison, systematic subjective verification conducted during the Spring Experiment, and objective metrics based largely on the mean diurnal cycle of the simulated reflectivity and precipitation fields. Additional insight is gained by examining the size distributions of the individual reflectivity and precipitation entities, and by comparing forecasts of mesocyclone occurrence in the two sets of forecasts.

In general, the 2-km forecasts provide more detailed presentations of convective activity, but there appears to be little, if any, forecast skill on the scales where the added details emerge. On the scales where both model configurations show higher levels of skill—the scale of mesoscale convective features—the numerical forecasts appear to provide comparable utility as guidance for severe weather forecasters. These results suggest that, for the geographical, phenomenological, and temporal parameters of this study, any added value provided by decreasing the grid increment from 4 to 2 km (with commensurate adjustments to the vertical resolution) may not be worth the considerable increases in computational expense.

### 1. Introduction

As computer resources have increased in recent years, operational modeling centers have responded by introducing numerical weather prediction (NWP) models with progressively higher resolution. This trend has been evident with both deterministic models and, in more recent years, with ensemble systems. As an example, consider the primary 1–3-day operational model

in the United States, now called the North American Mesoscale model (NAM; Black 1994; Janjić 1994). Scientists at the National Centers for Environmental Prediction's Environmental Modeling Center (NCEP/EMC) have decreased the NAM's grid spacing from 80 km in 1993 to 48 km in 1995 to 32 km in 1998, 22 km in 2000, and 12 km in 2001.

The downward trend has leveled off since 2001 as NCEP scientists have evaluated different options to make optimal use of current and future computing resources. A major concern in this regard is the representation of deep moist convection. Deep convection is generally parameterized and not explicitly predicted

---

Corresponding author address: John S. Kain, NSSL, 120 David L. Boren Blvd., Norman, OK 73072.  
E-mail: jack.kain@noaa.gov

when the horizontal grid spacing is greater than about 10 km (see Molinari and Dudek 1986; Kalb 1987; Zhang et al. 1988; Gallus and Segal 2001; Bélair and Mailhot 2001; Liu et al. 2001; Roberts 2003), but convective parameterization (CP) is typically avoided at higher resolution. This avoidance is appropriate because the conceptual basis for CP becomes increasingly ambiguous as the grid spacing decreases further (Molinari and Dudek 1992; Arakawa 2004).

As early as the late 1970s, it was argued that it would be preferable to allow convective overturning to proceed as an explicitly resolved process in a model, so that there could be a broad and continuous spectrum of interactions between convective and larger-scale processes (Rosenthal 1978). While this is conceptually appealing, disabling CP can have undesirable consequences with coarse resolution, ranging from clearly unacceptable errors with grid lengths of several tens of kilometers (e.g., Molinari and Dudek 1986; Jung and Arakawa 2004) to less egregious, but still unrealistic, performance in the range of 5–10-km grid spacing (e.g., Weisman et al. 1997). In general, convective overturning tends to develop and evolve too slowly when it is poorly resolved in a model, and updraft and downdraft mass fluxes, along with precipitation rates, are too strong when the convective process matures (e.g., Weisman et al. 1997; Roberts 2005). Furthermore, the likelihood of the failure mode (no convective development) increases without CP when the grid resolution is too coarse to represent the processes responsible for convective initiation (e.g., Liu et al. 2001; Petch et al. 2002).

As horizontal grid spacing is taken below about 5 km, there is still some debate about whether CP is needed and furthermore, when CP is excluded, just how much resolution is needed for faithful simulations of convection. With 4-km grid spacing, Deng and Stauffer (2006) and Lean et al. (2008) showed that quantitative precipitation forecasts (QPFs) could be improved in some cases by applying CP. Even at more than double that resolution, important components of the convective process may still be poorly represented. For example, Petch et al. (2002) suggested that horizontal grid increments below 1 km are necessary to predict accurately the timing and intensity of convective activity forced by surface heating over land. They hypothesized that such fine resolution is needed to resolve the boundary layer eddies that are responsible for convective initiation. Similarly, a series of studies at the University of Oklahoma's Center for Analysis and Prediction of Storms (CAPS) showed that explicit simulations of convective supercells are very sensitive to grid spacing in the range from 0.25 to 2 km (i.e., Droegemeier et al. 1994;

Droegemeier et al. 1996; Lilly et al. 1998; Adlerman and Droegemeier 2002). Furthermore, Bryan et al. (2003) and Petch (2006) argued that convection-allowing models cannot be considered convection-resolving until horizontal grid spacing approaches 100 m, primarily because such fine grid increments are necessary to begin resolving in-cloud turbulence, including the entrainment process, an inherent component of convective overturning.

Clearly, that scale of resolution is out of reach for many NWP generations to come. However, the near-term outlook for operational NWP may not be as bleak as some of these studies suggest. In another set of idealized tests, Weisman et al. (1997) found that the mesoscale structures associated with strong midlatitude squall lines could be represented adequately using a horizontal grid spacing of 4 km, without CP, even if the convective-scale details are lacking. Inspired by this result, the National Center for Atmospheric Research (NCAR) generated real-time, large-domain [about 2/3 of the contiguous United States (CONUS)] 4-km forecasts in support of the Bow Echo and Mesoscale Convective Vortex Experiment (BAMEX; Davis et al. 2004) during the late spring and summer of 2003. They used the new Weather Research and Forecasting model (WRF; Skamarock et al. 2005) (with CP switched off), and the forecasts were quite successful. In particular, they provided skillful guidance for convective-system morphology that was not available from operational models that used CP (Done et al. 2004; Weisman et al. 2008).

These results motivated a broader series of real-time high-resolution forecasts, generated daily in the spring and early summer of 2004 and evaluated in the 2004 Storm Prediction Center/National Severe Storms Laboratory (SPC/NSSL) Spring Experiment.<sup>1</sup> (Additional details about the HWT and each annual experiment can be found online at <http://www.nssl.noaa.gov/hwt/>.) In this experiment, large-domain, approximately 4-km grid-spacing forecasts were generated by NCEP/EMC and the Center for Analysis and Prediction of Storms (CAPS), as well as NCAR. Systematic, consensus-oriented, subjective evaluation procedures in the Spring Experiment revealed that, on average, the 4-km forecasts provided better guidance than the primary 1–3-day operational model from the time (i.e., NCEP's Eta Model, using 12-km horizontal grid spacing) for convective-system mode (morphology), and comparable guidance for convective initiation and evolution

---

<sup>1</sup> This experiment, formerly called the SPC/NSSL Spring Program, has been conducted annually since 2000 as an activity of the NOAA Hazardous Weather Testbed (HWT), during the peak severe weather season, from mid-April through early June.

(Kain et al. 2006). Although this experiment did not evaluate the accumulation of precipitation from a quantitative perspective, it corroborated the Done et al. (2004) results with regard to convective morphology and it alleviated concerns about the failure mode at  $\delta x = 4$  km (no convective initiation).

In 2005, a third round of high-resolution forecasts was conducted, again in collaboration with the annual SPC/NSSL Spring Experiment. Primary contributors were the same as in 2004—NCAR, EMC, and CAPS—but this time the emphasis of the model evaluation was somewhat different. In particular, rather than compare the high-resolution models to CP-configured operational models, the evaluations concentrated on sensitivities to model configuration, with the aim of providing insight for the development of optimal (and affordable) configurations for the next generation of operational convection-allowing models. One focus was on comparison of the NCAR and EMC forecasts, particularly the distinctive sounding structures (vertical profiles) associated with different boundary layer parameterizations in these configurations (Kain et al. 2005). A second emphasis, and the topic of this paper, was on the impact of decreasing horizontal grid spacing from 4 to 2 km. This latter focus was enabled by CAPS scientists, who used a grid spacing of just 2 km in 2005, while NCAR and EMC continued to generate forecasts using approximately 4-km spacing.

The sensitivity to grid resolution has important practical implications because doubling the resolution requires at least a tenfold increase in computer power (in addition to much greater demands on storage, dissemination, and display of output). While there seems to be general agreement that more resolution is better, an important question, especially for the operational modeling community, is: How much better? How much do we gain in forecasting utility when we increase computing costs by an order of magnitude? The challenge for both the operational and research communities is to find a grid resolution at which CP can be “turned off,” the negative impacts associated with poor resolution of convective processes can be reduced to a tolerable level, and numerical forecasts can be generated in a timely enough manner to remain useful for operational forecasting.

Several operational NWP centers have been experimenting with at least semioperational convection-allowing forecast systems in recent years. For example, Steppeler et al. (2003) discuss plans for implementation of a 2.5-km grid-spacing model at Deutscher Wetterdienst (DWD) in Germany, while Speer and Leslie (2002) show favorable results from a prototype 2-km grid-spacing model for Australia’s Bureau of Meteorol-

ogy. Narita and Ohmori (2007) and Lean et al. (2008) discuss similar efforts with the Japan Meteorological Agency’s (JMA) nonhydrostatic mesoscale model and the U.K. Met Office’s Unified Model, respectively, although both of these centers retain some form of parameterized convection even at horizontal grid spacings of  $\sim 5$  km. In the United States, NCEP/EMC has continued daily experimental forecasts of the  $\sim 4$  km WRF-Nonhydrostatic Mesoscale Model (WRF-NMM; see Janjić et al. 2007) that they had begun as part of the 2004 SPC/NSSL Spring Experiment, providing this model output to SPC operations on a daily basis. In the fall of 2007, they began their first operational forecasts with convection-allowing domains of this size ( $\sim 3/4$  CONUS), supplementing their normal 12-km NAM forecasts with 4-km WRF-NMM and 5.1-km WRF-Advanced Research WRF (WRF-ARW; see Skamarock et al. 2005) forecasts (G. DiMego, NCEP/EMC, 2007, personal communication). Convective parameterization is not used in the latter two configurations.

This study provides an important evaluation of practical considerations related to operational forecasts of disruptive convective weather using models with horizontal grid spacing less than 5 km and no CP. In particular, it focuses on a comparison of 18–30-h WRF model convective forecasts at 2- and 4-km grid spacing. The WRF-ARW was used for all forecasts discussed in this paper, although the WRF-NMM was also used in the Spring Experiment. The specific objective is to evaluate the sensitivity of daily forecasts to resolution, using both subjective and objective metrics. The evaluation is based primarily on simulated reflectivity and accumulated precipitation fields. However, a secondary focus is based on model predictions of supercells (or thunderstorms containing mesocyclones). These storms produce a disproportionate share of severe weather compared to other modes of convection, and anticipating their development is one of the biggest challenges for severe weather forecasters, such as those at the SPC. Thus, part of this study addresses the question of supercell forecasts: Can the 4- and 2-km forecasts provide explicit guidance for the occurrence of supercells? Can the 2-km forecasts provide better guidance than their 4-km counterparts? Eventually, we hope to extend this inquiry beyond supercells to the explicit prediction of other phenomena associated with severe weather, such as bow echoes, hail signatures, and book-end vortices, and even some of the phenomena themselves, for example, strong surface-layer winds, hail, etc., but this extension is beyond the scope of this paper.

Details of the model configurations, experimental

TABLE 1. Model configurations used for the high-resolution forecasts: YSU, Yonsei University (Noh et al. 2003); WSM6, WRF single-moment, 6-class microphysics (Hong et al. 2004); Dudhia (Dudhia 1989); RRTM: Rapid Radiative Transfer Model (Mlawer et al. 1997; Iacono et al. 2000).

	NCAR	CAPS
Dynamic core	ARW	ARW
Horizontal grid spacing (km)	4.0	2.0
Vertical levels	35	51
PBL/turbulence parameterization	YSU	YSU
Microphysical parameterization	WSM6	WSM6
Radiation parameterization (SW-LW)	Dudhia/RRTM	Dudhia/RRTM
Initial conditions	0000 UTC 40-km Eta	0000 UTC 40-km Eta

methods, and evaluation parameters are provided in the next section. This is followed by results and discussion, then summary and conclusions.

## 2. Methodology

During the 2005 Spring Experiment, the different versions of the WRF model were initialized at 0000 UTC on a daily basis (Monday–Friday, 18 April–3 June) at remote locations and output was collected at the SPC. There, the output was used as guidance for experimental forecasts of severe convective weather and it was examined in detail using systematic subjective evaluation methods. Both the forecasts and the model evaluation efforts were conducted by groups of 6–10 scientists and forecasters, with a new group rotating in at the start of each week. The experiment continued for 7 weeks. Objective analysis of the data was conducted after the experiment ended.

### a. Model configurations

The two model configurations used for this study are summarized in Table 1. The first was run at NCAR, using 4-km horizontal grid spacing with 35 vertical levels (hereafter WRF4). The second was run by CAPS at the Pittsburgh Supercomputing Center, with 2-km grid spacing and 51 vertical levels (hereafter WRF2). Both configurations were initialized by interpolating 0000 UTC initial conditions from the Eta Model (Black 1994) to the high-resolution grids.

Daily production of the 2-km grid spacing forecasts was a ground-breaking achievement in itself. Even though the CAPS domain was somewhat smaller than that used by NCAR (Fig. 1), the horizontal gridpoint dimensions were  $1500 \times 1320$ , which, given the real-time production schedule, presented extraordinary computational and data-management challenges. One of these challenges was in the area of the initial conditions. CAPS programmers found that prohibitive time delays were introduced when the standard WRF interpolation and initialization (WRFISI) package was used

with their grid, so they developed and used a new, more computationally efficient routine. NCAR scientists continued to use WRFISI. The different approaches resulted in initial atmospheric conditions that were broadly similar, but visibly different in their finer-scale details.

Initial soil moisture fields were also slightly different. NCAR scientists initialized soil moisture using the High Resolution Land Data Assimilation System (HRLDAS; see Chen et al. 2004), an offline soil model that incorporates observed surface variables, precipitation, and radiation data (W. Wang, NCAR, 2005, personal communication), while CAPS scientists used their new initialization routine to interpolate the Eta Model's soil moisture field to their high-resolution grid. It is not clear if, or to what degree, the slight differences in initial conditions or domain sizes impacted next-day forecasts. The focus of this study is on a comparison of the general character and statistical properties of the two forecasts. In this respect, it is assumed that the spatial resolution, including both the horizontal and vertical components, is the dominant factor that leads to differences in the two forecasts.

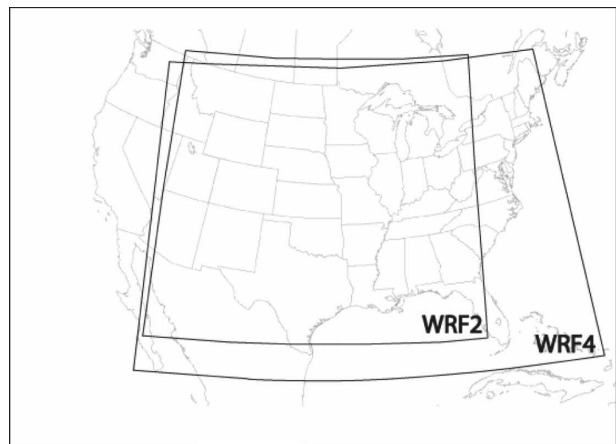


FIG. 1. Model integration domains for the CAPS (WRF2) and NCAR (WRF4) forecasts.

### b. Observed radar reflectivity

The observed radar reflectivity images used in this study come from national base-reflectivity mosaics that are part of the SPC's operational datastream. These images are generated by Unisys Corporation at a frequency of 5 min or less on a grid with 2-km spacing between points. The mosaics incorporate the lowest elevation cut (0.5° elevation angle) of the most recent reflectivity data from each of the 142 continental U.S. Weather Surveillance Radars-1988 Doppler (WSR-88Ds) as received via the National Weather Service Radar Product Central Collection Dissemination Services (RPC-CDS). Multiple proprietary techniques are used to remove anomalous propagation and residual ground clutter from the data.

The base-reflectivity data positions are transformed from underlying radial (i.e., azimuth–range) format to corresponding latitude–longitude locations. The latitude–longitude-based values are then mapped to a national 2 km × 2 km grid with zonal and meridional extents of 5120 and 3584 km, respectively. Each radar bin is assigned to the grid box having the nearest latitude–longitude. A Lambert conformal conic projection is employed with the grid centered at 38°N and 98°W, with standard latitude values of 38° and 45°N.

The mosaics are presented in the same 16 data levels of dBZ that correspond to the WSR-88D precipitation mode data level scale. Where multiple radar bins are collocated, the maximum value takes precedence. Data values from sites operating in clear-air mode are converted to the corresponding precipitation mode data level values. Site data exceeding an age limit are excluded. Hereafter, the observed reflectivity fields are referred to as BREF (for base reflectivity).

### c. Simulated radar reflectivity

A surrogate for observed reflectivity can be computed, based on the concentration of precipitation-sized hydrometeors predicted by a model. As with observed reflectivity, this derived field (hereafter called the simulated reflectivity factor, SRF) can be quite useful for monitoring the intensity, movement, and areal coverage of precipitation features (see Koch et al. 2005).

For the experimental runs used during the 2005 Spring Experiment, hydrometeors were generated by the WRF Single-Moment 6-Class Microphysics (WSM6) microphysical parameterization (Hong et al. 2004) in the WRF model. This parameterization carries three categories of precipitation-sized hydrometeors—rain, snow, and graupel—as prognostic variables. The

SRF is computed based on separate contributions from each category.

Following Koch et al. (2005), the equivalent reflectivity for rain,  $Z_{er}$ , is computed as

$$Z_{er} = 720N_{0r}\lambda_r^{-7} \times 10^{18}, \quad (1)$$

where  $N_0$  is the intercept parameter, assumed to have a constant value of  $8 \times 10^6 \text{ m}^{-4}$  in WSM6, and  $\lambda$  is the slope factor, defined by

$$\lambda_r = \left( \frac{\pi N_0 \rho_l}{\rho_a q_r} \right)^{1/4}, \quad (2)$$

and  $\rho_l$  is the density of liquid water ( $1000 \text{ kg liquid m}^{-3}$ ),  $\rho_a$  is the local density of dry air ( $\text{kg air m}^{-3}$ ), and  $q_r$  is the rainwater mixing ratio ( $\text{kg liquid kg}^{-1}$  dry air). The factor  $10^{18}$  is included to convert from units of  $\text{m}^3$  to the more commonly expressed units of  $\text{mm}^6 \text{ m}^{-3}$ .

In a similar fashion, the equivalent reflectivities for snow and graupel,  $Z_{es}$  and  $Z_{eg}$ , respectively, are given by

$$Z_{es} = 161.3N_{0s}\lambda_s^{-7} \left( \frac{\rho_s}{\rho_l} \right)^2 \times 10^{18}, \quad (3)$$

and

$$Z_{eg} = 161.3N_{0g}\lambda_g^{-7} \left( \frac{\rho_g}{\rho_l} \right)^2 \times 10^{18}, \quad (4)$$

where  $N_{0s} = 2 \times 10^7 \text{ m}^{-4}$ ,  $N_{0g} = 4 \times 10^6 \text{ m}^{-4}$ ,  $\rho_s$  is the assumed density of snow ( $100 \text{ kg m}^{-3}$ ),  $\rho_g$  is the assumed density of graupel ( $400 \text{ kg m}^{-3}$ ),

$$\lambda_s = \left( \frac{\pi N_{0s} \rho_s}{\rho_a q_s} \right)^{1/4}, \quad (5)$$

and

$$\lambda_g = \left( \frac{\pi N_{0g} \rho_g}{\rho_a q_g} \right)^{1/4}, \quad (6)$$

where  $q_s$  is the mixing ratio of snow ( $\text{kg snow kg}^{-1}$  dry air) and  $q_g$  is the mixing ratio of graupel.

The total reflectivity can then be obtained by simply adding  $Z_{er}$ ,  $Z_{es}$ , and  $Z_{eg}$ ,

$$Z_e = (Z_{er} + Z_{es} + Z_{eg}). \quad (7)$$

The SRF, expressed in dBZ, is a logarithmic form given by

$$\text{SRF}(\text{dBZ}) = 10 \log(Z_e). \quad (8)$$

During the 2005 Spring Experiment, 2D fields of model-derived SRF were examined at individual vertical levels and in a composite form, the latter given by the maximum value of SRF at any level. Evaluation of the different products suggested that the SRF at 1 km AGL (above ground level) compared most favorably to

the observed base reflectivity and it revealed detailed storm structures better than the composite or higher-altitude SRF products. Consequently, only the 1-km AGL SRF is considered in this paper. Hereafter, it is denoted simply as SRF.

It is important to recognize that SRF is best regarded as a surrogate for observed reflectivity, rather than a mathematical equivalent. There is no unique quantitative relationship between hydrometeors, observed reflectivity, and SRF; a given value of reflectivity or SRF can come from many different combinations of different hydrometeors. Furthermore, the simulated and observed hydrometeor fields considered here are sampled in very different ways. Observed base reflectivity is derived from a radar beam transmitted at a 0.5°-elevation angle. Thus, it senses hydrometeors at relatively low altitudes close to the transmitter, but progressively higher altitudes at greater distances. In contrast, the SRF field is based on precipitation particles at a fixed altitude. There are a host of other inconsistencies that preclude a strict quantitative assessment of SRF using observed reflectivity. Yet, these data sources convey remarkably similar information to the human analyst and, as will be shown below, comparison of these fields using objective verification metrics is also quite revealing.

#### d. Observed precipitation

Hourly precipitation output from the WRF configurations is compared to observations from the stage II precipitation archive produced at NCEP (Baldwin and Mitchell 1998). The stage II data are a mosaic of hourly rainfall observations on a 4-km grid. The mosaic is generated using optimal estimates based on both radar and rain gauge data (Seo 1998).

#### e. Mesocyclone detection

Two different algorithms were used to detect mesocyclones in hourly model output during the Spring Program. The first was based on a layer-averaged correlation between vertical velocity and vertical vorticity (e.g., Weisman and Klemp 1984; Droegeimer et al. 1993). The second was based on direct computation of the local product of these two fields, again averaged over a vertical layer. Through a subjective calibration process, threshold values were established for both of these methods so that the number of storms exceeding these values in the 2-km forecasts was in reasonable agreement with the number of observed rotating storms. Subjective assessments from the program suggested that these two algorithms were equally useful; they often flagged the same storms in the models and generally provided very similar information.

For this study, only the second algorithm will be considered. This approach is favored here because its dependence on grid spacing appears to be relatively straightforward, facilitating a direct comparison between the 2- and 4-km forecasts. This algorithm is based on the concept of helicity,  $H$ , which is a scalar measure of the potential for helical flow (i.e., the pattern of a corkscrew) to develop in a moving fluid. It is defined by

$$H = \mathbf{V} \cdot \nabla \times \mathbf{V}. \quad (9)$$

The horizontal component of the environmental helicity, commonly expressed within a storm-relative framework, is easily computed from current operational model output and is often used by forecasters to assess the potential for rotating thunderstorms, or supercells (e.g., Johns and Doswell 1992). In models that resolve convective updrafts explicitly, such as those considered here, rotating storms can be detected directly by measuring the vertical component of helicity,  $H_v$ , given by

$$H_v = w \left( \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right). \quad (10)$$

In this study,  $H_v$  is integrated over a layer to yield a measure of the updraft helicity, UH, given by

$$\text{UH} = \int_{z_0}^{z_t} w \zeta \, dz, \quad (11)$$

where  $\zeta$  is the vertical component of the relative vorticity. Note that  $\zeta$  can be negative when updrafts are rotating anticyclonically; further,  $w$  is negative in downdrafts, and downdrafts can also rotate cyclonically or anticyclonically. During the 2005 Spring Experiment, each of the possible combinations of  $w$  and  $\zeta$  was examined separately, but in this study the focus is on cyclonically rotating updrafts where both  $w$  and  $\zeta$  are positive. Since the primary interest is on storm rotation in the lower to middle troposphere, (11) was integrated vertically from  $z_0 = 2$  km to  $z_t = 5$  km AGL using a midpoint approximation. Specifically, data were available every 1 km AGL, so Eq. (11) was computed using

$$\begin{aligned} \text{UH} &\approx w \sum_{z=2000\text{m}}^{z=5000\text{m}} \overline{w \zeta \Delta z} \\ &= (\overline{w \zeta}_{2,3} + \overline{w \zeta}_{3,4} + \overline{w \zeta}_{4,5}) \times 1000, \end{aligned} \quad (12)$$

where the overbar indicates a layer average and the subscripts indicate the bottom and top of the layer in kilometers. All UH values were smoothed using a standard nine-point smoother.

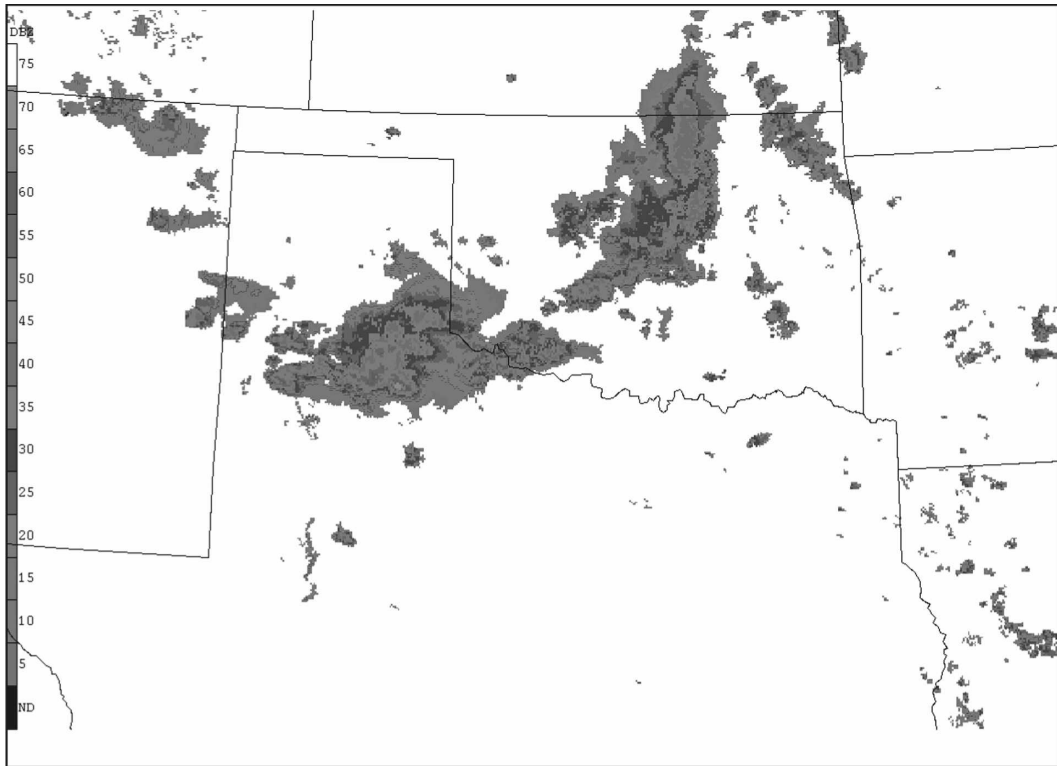


FIG. 2. Forecast–evaluation domain for 31 May 2005.

#### f. Subjective evaluation

As in previous SPC/NSSL Spring Experiments (e.g., see Weiss et al. 2004; Kain et al. 2003b), daily activities in 2005 were roughly evenly divided between experimental forecasting exercises and interrogation and evaluation of model output; the first half of the day was devoted to human forecasts while the WRF numerical output was the focus of the afternoon time period. The strategy for subjective model evaluation was to provide a short written description of the relevant differences between the model forecasts and the verifying observations, followed by a rating of model performance on a scale of 1 to 10 (see Kain et al. 2003a). Participants were asked to comment specifically on perceived differences between the 2- and 4-km WRF output.

It is important to emphasize that the rating process was not a trivial matter. It involved a systematic assessment of model output and comparison with corresponding observations. The assessment typically involved lively discussion within a group of 6–10 expert forecasters and researcher scientists. The specific rating was obtained by consensus of the group.

Both the experimental forecasts and the model evaluations were conducted over limited regional domains and relatively short (6 h) time periods. For ex-

ample, Fig. 2 shows the forecast–evaluation domain for 31 May. The size and aspect ratio of this domain were held constant throughout the experiment, but the window was relocated every day to focus on the area of greatest threat for severe convective weather for that day. Likewise, the focused 6-h time frame was shifted within the 18–30-h forecast period based on the expected timing of the convective initiation.

For this paper, subjective verification results are based on comparisons of BREF and SRF (1 km AGL) from the different WRF configurations. Although many other fields are relevant and were examined during the Spring Experiment, SRF is perhaps more revealing than any other single output field (see below) and radar imagery is widely used by forecasters and researchers to observe and analyze convective storm systems.

#### g. Model climatology

The mean characteristics, or climatology, of key model output fields were measured using a combination of simple statistical methods. These analyses involved hourly data extracted from a fixed common domain (Fig. 3). SRF, precipitation, and UH output fields were examined and compared to relevant observations

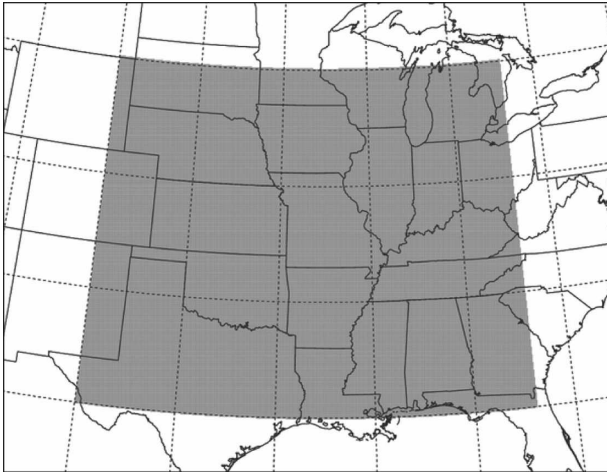


FIG. 3. Domain (shaded) for the calculation of model climatology.

when possible. The analyses were performed on the native grid of each output field to avoid interpolation of data and to preserve individual small-scale features in each field. They were designed to measure the areal coverage of various features both in a bulk (domain wide) sense and in terms of the numbers and size distributions of individual features. Mean characteristics were computed on an hourly basis, averaging over all days of the experiment (33 days for reflectivity and UH data, 31 days for precipitation). The hourly means were plotted as a time series to reveal important characteristics in the diurnal cycle of the two models and the corresponding observations. In addition, the overall size distributions were plotted for the SRF fields to provide information about the relative level of detail in the two model forecasts.

In a separate analysis, the equitable threat score (ETS) was used to measure the degree of overlap between corresponding reflectivity fields. This score requires all data to be on the same grid, so reflectivity fields were interpolated to a common 4-km grid for this task. The ETS has a maximum value of 1 and a minimum value of  $-1/3$ . It has been relied upon quite heavily at NCEP in recent years to compare the skill of quantitative precipitation forecasts from different models (e.g., Mesinger 1996), especially those with mesoscale resolution and parameterized convection. However, it is less useful when higher-resolution forecasts of convective phenomena need to be compared, especially if these forecasts contain a predominance of small-scale, high-amplitude features such as convective cells. Even with the best forecasts, the scale of these features is typically smaller than the displacement error (specific location compared to corresponding observed fea-

tures), resulting in little or no overlap between the forecast and observations, and a very low ET score (e.g., see Baldwin and Wandishin 2002; Baldwin and Kain 2006). Nonetheless, the ET score methodology can provide some useful information within the context of this study, especially since the 2- and 4-km forecasts are similarly affected by the deficiencies of the ET concept.

### 3. Results

#### a. Subjective assessment

A subjective assessment begins with a simple visual comparison. As an example of the process, the relevant attributes from two representative events are highlighted below.

##### 1) 31 MAY 2005

The first case considered is the forecast initialized at 0000 UTC 31 May. Convective activity was well under way by the 23-h forecast time in both models and in observations (left-hand side in Fig. 4). The two model forecasts are qualitatively similar. The primary feature in both forecasts is an area of convective activity characterized by a loosely organized, quasi-linear convective line extending from the central part of the Texas Panhandle into northwestern Oklahoma, apparently linked to more scattered multicellular convection in west-central and north-central Oklahoma at 2300 UTC. These general characteristics are consistent enough between the two model runs to give one confidence that this feature represents a common meteorological phenomenon, even though the general outline of the convective activity differs somewhat from one run to the other. Nearby, the 2-km run generates two isolated areas of convective activity that appear to be lacking in the WRF4, one in southwestern Oklahoma, the other in the south-central part of the Texas Panhandle. In general, the WRF2 features appear to be more complex with a higher degree of local variability.

Comparison with the observed base reflectivity is less favorable. One could say that the models successfully predicted convective activity in parts of Oklahoma and the Texas Panhandle, but the correspondence between the major features (i.e., convective clusters) in models forecasts and observations is ambiguous. Simulated reflectivity patterns from the two models are much more like each other than either one is like the observations.

The same is true 7 h later, at the 30-h forecast time (right-hand side in Fig. 4). Mesoscale organizational structures in the model forecasts are remarkably similar at this time. Both show a line of convective activity extending from eastern Kansas southward through



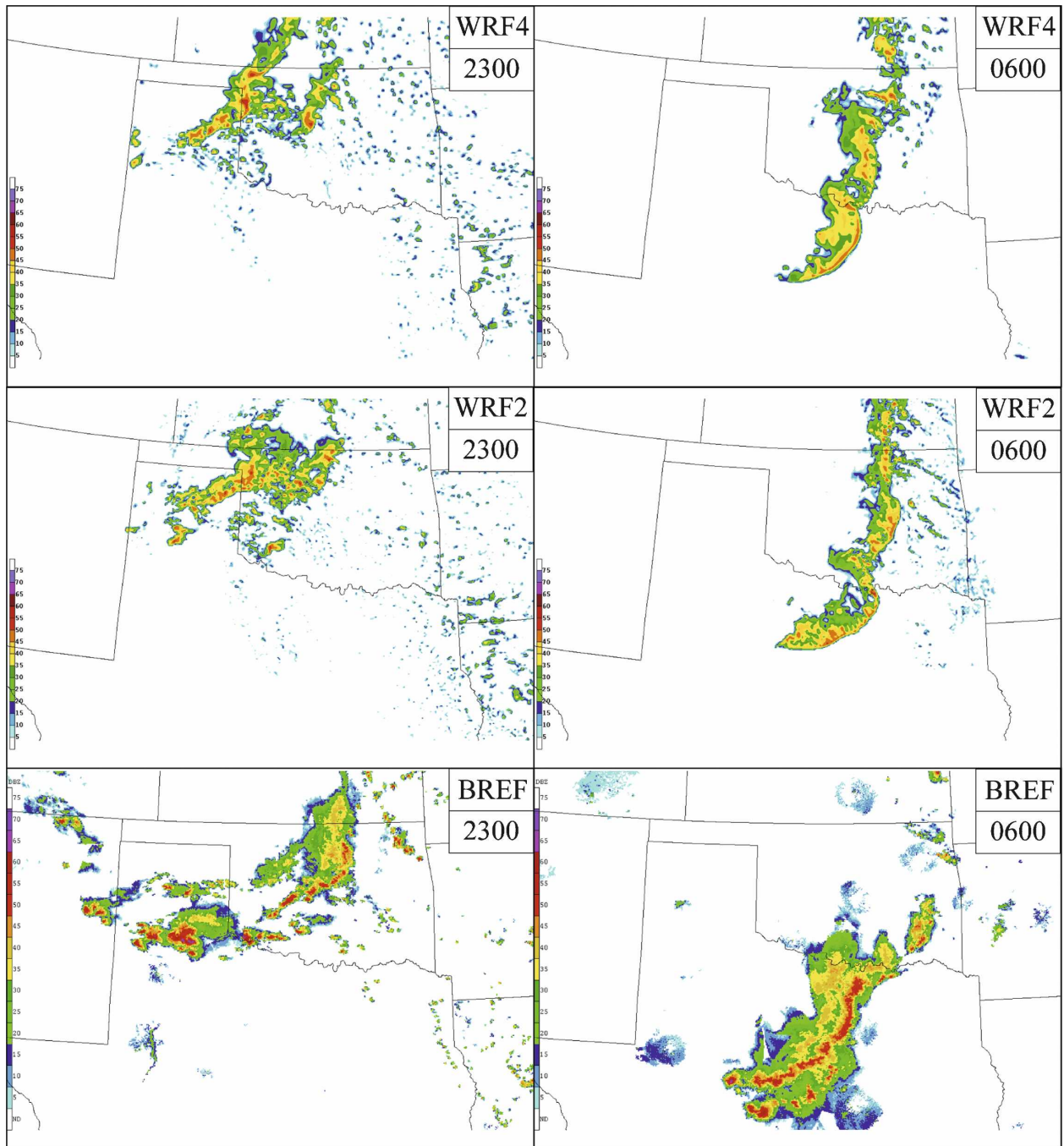


FIG. 4. SRF and corresponding BREF for selected times, associated with the model forecasts initialized at 0000 UTC 31 May 2005.

Oklahoma and into north-central Texas, with a clearly discernible bowing segment south of the Oklahoma–Texas border and some evidence of a bowing structure in Oklahoma as well. Again, finer-scale structures appear in the 2-km run, but the mesoscale organization of convection is quite consistent between the two forecasts. In contrast, observed convective activity is fo-

cused farther south (bottom-right panel in Fig. 4). There is little convection in Kansas and Oklahoma and a large bowing mesoscale convective system (MCS) propagating into central Texas, with isolated intense convective cells ahead of the southern half of this line. Coverage of convection in the two model forecasts overlaps to a much higher degree than either forecast

overlaps with the observed activity, but both model configurations correctly suggest the predominance of the quasi-linear convective mode.

## 2) 2 JUNE 2005

Examination of model forecasts and observations from 2 days later leaves much the same impression, although initial convective development was very similar in both the models and observations. At 2200 UTC on 2 June, deep convection was developing rapidly in northeastern Colorado. Although both model configurations appeared to underforecast the number and intensity of storms at this time, they both predicted a dominant isolated cell over the region, corresponding fairly well with observations (left-hand panels in Fig. 5). Furthermore, model diagnostics revealed that both of these isolated model storms exhibited high UH values and characteristics commonly associated with supercells, including a clearly discernible inflow notch and strong low-to-midlevel rotation (discussed in section 3c), while the corresponding observed storm had the radar presentation of a supercell, and it was flagged by the operational NSSL Mesocyclone Detection Algorithm (Stumpf et al. 1998) as a likely mesocyclone.

Both forecasts appeared to be remarkably good initially, but their correspondence with the observations diminished considerably as convective development continued. By 0600 3 June, the 2- and 4-km forecasts still looked quite similar to one another, but they differed significantly from the observations. Both models had developed a spurious convective system over southwestern Kansas around 0000 UTC and this system had evolved in both forecasts into a bow-shaped line extending from east-central Kansas to north-central Oklahoma. Meanwhile they both had a second system along the west-central Kansas–Nebraska border that appeared to correspond to an observed system, but the observed system was much larger and had a different configuration than either forecasted system.

Again, the 2-km reflectivity fields appeared to have more detailed finer-scale structure than the 4-km forecasts, but the meso- $\beta$ -scale structures were very similar. Above all, the model forecasts looked more like each other than they did like the observations.

## 3) MEAN SUBJECTIVE RATINGS

Similarity and differences between the forecasts can also be gauged by comparing the subjective ratings for the 2- and 4-km forecasts. During the 2005 Spring Program, forecast evaluation teams rated the skill of all model forecasts in categories of convective initiation and the evolution of mesoscale convective features.

Specifically, the teams were instructed to assess the correspondence with observations in terms of “timing and location” for convective initiation and “direction and speed of system movement, areal coverage, configuration and orientation of mesoscale features, and perceived convective mode” for evolution. Subjective ratings from both the WRF2 and WRF4 were available in real time for only 24 of the 33 forecast days. The mean subjective ratings in each category for these 24 days are shown in Fig. 6. Although there are slight differences in the mean values, these differences are not statistically significant, based on paired  $t$  tests. However, it is noteworthy that of the 24 days, WRF2 initiation forecasts were rated higher than those of WRF4 on 6 days, while the reverse was true on only 1 day (identical ratings were assigned on the remaining 17 days), perhaps providing an indication that higher resolution is advantageous for prediction of convective initiation, as previous studies have suggested.

### b. Model climatology

The mean characteristics of model output reveal important differences and similarities in the behavior of the two configurations. Equitable threat score is used to provide information related to the mean degree of overlap of reflectivity features while various other measures of mean areal coverage are used to compare coverage biases, differences in diurnal cycle, and the level of detail in individual features.

#### 1) REFLECTIVITY

The SRF fields from convection-allowing models have proven to be quite useful to forecasters, likely because of their resemblance to widely utilized observed reflectivity fields and because they can be useful for identifying mesoscale structures and processes in the model atmosphere (e.g., Koch et al. 2005).

##### (i) Bulk coverage characteristics

###### A. EQUITABLE THREAT SCORES

ETSs were used to provide a measure of the degree of overlap between the SRF fields and BREF. For comparison purposes, ETSs were also used to measure the overlap between the two model forecasts by using the WRF2 SRF as the verifying field for the WRF4 SRF. The common grid used for these computations had 4-km spacing and covered the largest possible common area—approximately the domain covered by the WRF2 runs—and the verification time window was the 18–30-h forecast period. The data include all days on which both model runs were available.

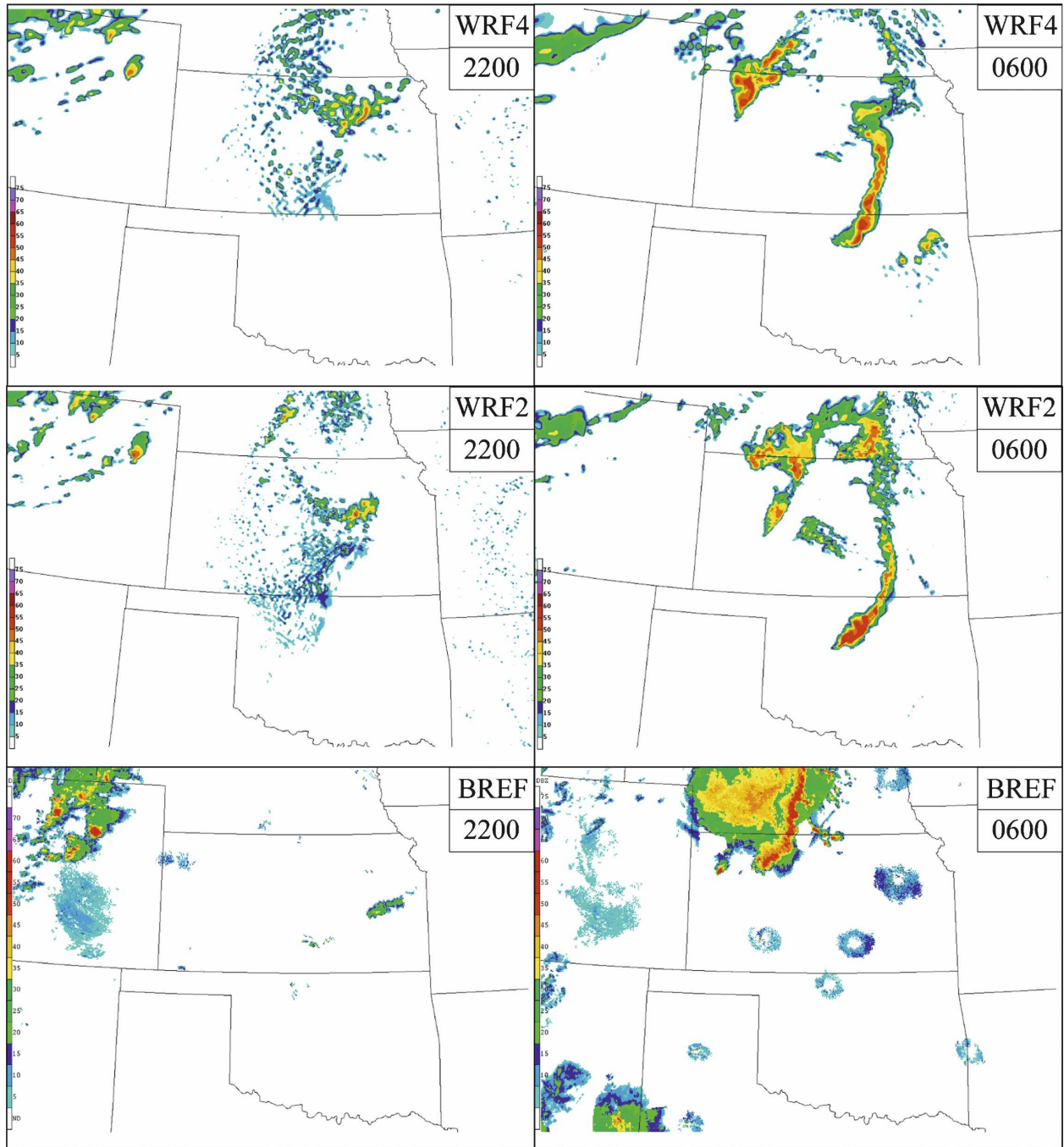


FIG. 5. SRF and corresponding BREF for selected times, associated with the model forecasts initialized at 0000 UTC 2 Jun 2005.

When ETs were used to compare the SRFs to the observed base reflectivity, the 2- and 4-km scores were nearly identical and quite low. In particular, they had maximum values of just over 0.1 at the lowest (10 dBZ) threshold and gradually trended toward zero at higher reflectivity thresholds (Fig. 7), indicating poor overlap between simulated and observed reflectivity features.

However, the degree of overlap between the two model forecasts was considerably higher. When the ET score was computed for the 4-km forecasts using the 2-km runs as “truth,” the maximum value was almost 0.35 and scores remained much higher at all thresholds than when the observed reflectivity was the verifying field. These scores confirm the subjective impressions based

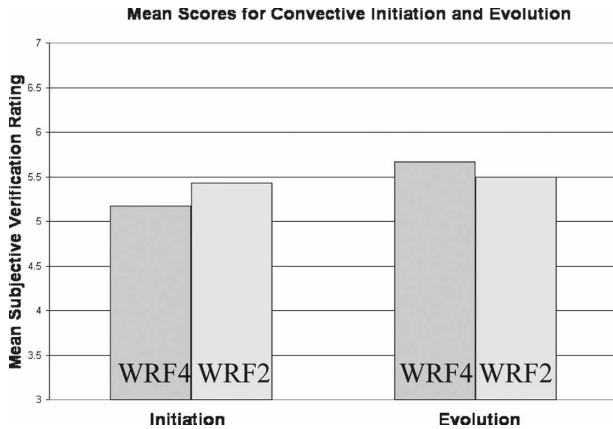


FIG. 6. Mean subjective ratings for convective initiation and evolution forecasts from the WRF4 and WRF2. Note that none of the differences are significant at the 95% level.

on only the two events presented earlier: The 2- and 4-km reflectivity fields were much more similar to each other than to the observations.

B. COVERAGE BIAS

Coverage biases (forecast area divided by observed area:  $A_f/A_o$ , perfect score = 1) varied as a function of time and reflectivity threshold. First, consider the bias plotted as a function of dBZ threshold at selected times. As shown in Fig. 8, the reflectivity biases were generally less than 1 (coverage underpredicted relative to BREF). They tended to have a maximum value at lower dBZ values, decreasing slowly as a function of increasing

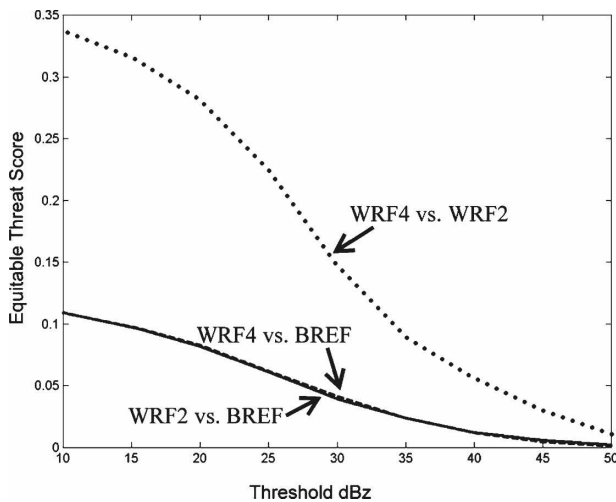


FIG. 7. ETSs as a function of SRF–BREF threshold during the 18–30-h forecast period. Note that the two lower curves indicate model performance (degree of overlap) relative to observations while the top curve indicates the degree of overlap between the WRF2 and WRF4.

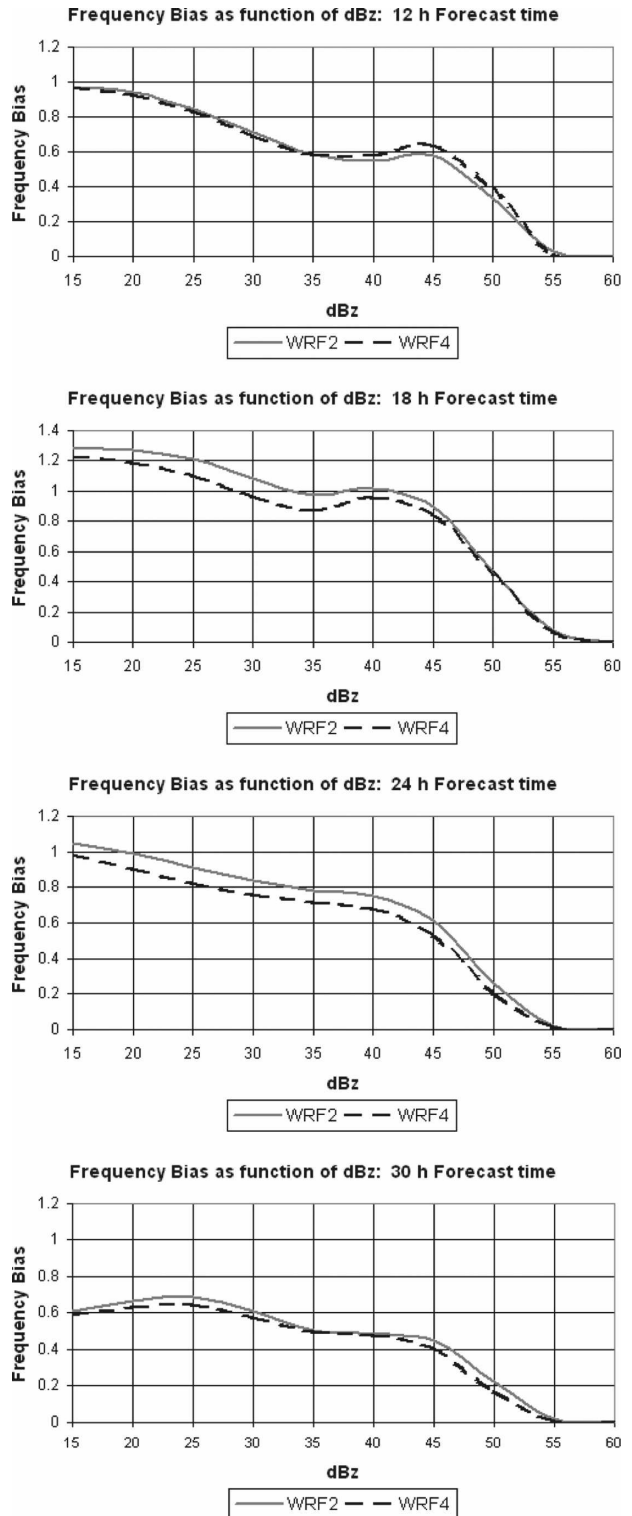


FIG. 8. Model climatology: Frequency bias of reflectivity (SRF for the models, BREF for the observations) as a function of reflectivity threshold, valid at selected forecast times and averaged over all days during the Spring Experiment.

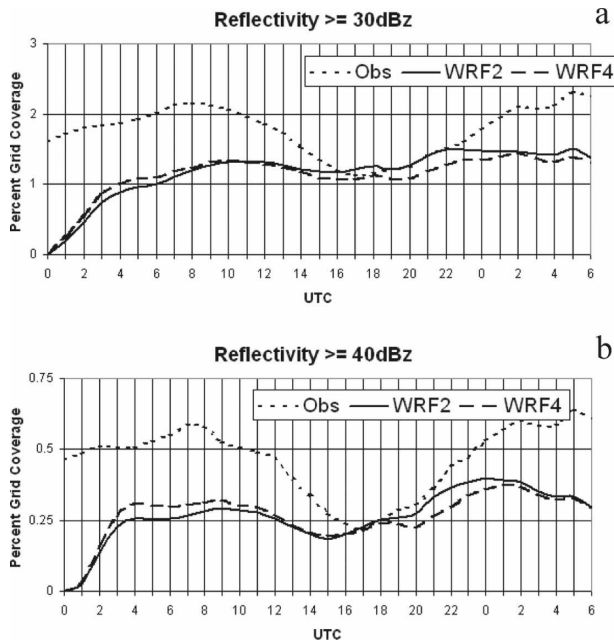


FIG. 9. Model climatology: Areal coverage of reflectivity (SRF for the models, BREF for the observations) as a function of time exceeding the (a) 30- and (b) 40-dBZ thresholds, averaged over all days during the Spring Experiment.

threshold up to 40–45 dBZ, then dropping off sharply toward higher thresholds. This general trend was evident at all times and, in general, the behaviors associated with the different WRF configurations were quite similar.

Now, consider the fractional coverage from each source plotted as a function of forecast hour for selected dBZ thresholds (Fig. 9). These plots reveal several important characteristics of the data:

- *Similarities—differences between the two model forecasts:* In general, the different WRF configurations produce similar trends in reflectivity coverage. With both model configurations, SRF coverage increases rapidly over the first 3–4 h (the “spinup” period), reaches a broad maximum before 1200 UTC (12-h forecast), a late morning minimum, and a second maximum value around 0000 UTC (24-h forecast). The WRF4 tends to produce higher coverage than the WRF2 during the first 12–15 h, but lesser coverage thereafter. There is some suggestion that WRF4 trends lag WRF2 trends by about 1 h, but this is difficult to confirm with only 1-h temporal resolution.
- *Differences between the models and BREF:*
  - *Amplitude of the diurnal cycle—*Observed reflectivity appears to have a much higher-amplitude response to the diurnal heating cycle than the SRF from the model. For example, at the 40-dBZ

threshold (bottom panel in Fig. 9), the maximum BREF coverage is about 3 times the minimum value, but the ratio is only about 2:1 for the models. As a consequence, while reflectivity coverage bias is close to the optimal value of 1 during the afternoon heating cycle, it is much less than 1 both before and after this period. The same relationships hold at the 30-dBZ threshold, although the amplitudes of all cycles are smaller.

- *Time of minimum and maximum values—*The model configurations appear to produce the minimum reflectivity coverage about 2 h before the observations, especially at the 40-dBZ threshold. They generate maximum coverage around 0000 UTC (24-h forecast). In contrast, the observed reflectivity areas continue to increase in size well beyond 0000 UTC, with a clear nocturnal maximum.

### (ii) Coverage of individual entities

Although the aggregate areal coverage of the SRF features is similar for the WRF2 and WRF4 forecasts, the corresponding numbers of individual SRF entities are quite different. Individual entities are identified by searching for contiguously adjacent grid points that exceed a specified SRF threshold, with no smoothing of the gridpoint data. After the initial 3–4-h spinup and through the initial overnight period, the WRF2 forecasts have about twice as many distinct SRF entities, on average, as those from the WRF4 (Fig. 10). This ratio increases to about 3:1 during the daytime heating cycle, when peak values are reached, then it retreats back toward 2:1 during the second overnight. This pattern is similar for both the 30- and 40-dBZ thresholds. As with the bulk coverage statistics, it appears that WRF4 trends lag the WRF2 trends by about an hour.

The diurnal cycle of the individual entities in the BREF data is similar to BREF’s bulk coverage, showing a minimum value in late morning and relatively broad maxima at night. The mean numbers of features tend to be lower than the WRF2 during the afternoon heating cycle, but higher at all other times—and much higher than the WRF4 entities at all times. While the diurnal cycle of BREF entities appears to have two peaks—one in late afternoon and one overnight—the model cycles are dominated by a single peak in the late afternoon.

Additional insight can be gained by examining the mean size distributions for all entities in each of these datasets. These distributions include all distinct reflectivity features from all days during the 12–30-h forecast period (i.e., the first 12 h are excluded). The WRF2 generates many more small-scale features than the WRF4, but the numbers converge at a size of about

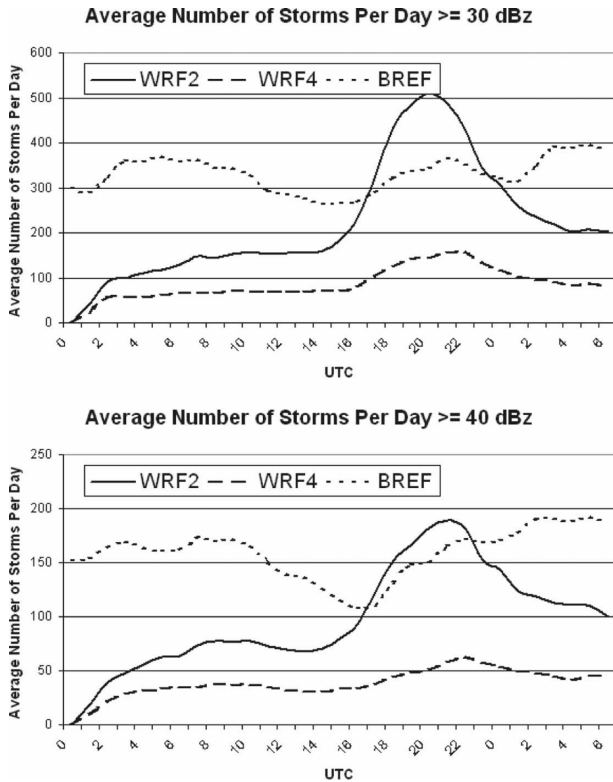


FIG. 10. Model climatology: Average number of individual reflectivity entities (SRF for the models, BREF for the observations) as a function of time for (top) 30- and (bottom) 40-dBZ reflectivity thresholds, based on all days during the Spring Experiment.

200 km<sup>2</sup>, with roughly equal numbers of larger entities (Fig. 11, top). This essentially confirms our subjective impression that the higher-resolution model forecasts have more detailed small-scale structure. In this range of resolution, convective instability tends to be released on the smallest resolvable scales of the model. If we plot the size distributions as a function of grid dimensions rather than raw areal coverage, we can see consistency in this regard between the WRF2 and WRF4. For example, Fig. 11 (bottom) shows plots of size distributions in which the size of the entities is expressed in terms of the number of grid points spanning the diameter of each feature (assuming a circular geometry). The distributions are approximately parallel, suggesting that the model numerical algorithms allow a spectrum of convective overturning processes that is consistent with resolution limitations regardless of grid length.

It is interesting to see that the distribution of BREF entities is quite similar to the WRF2. This suggests that filtering techniques that are applied to the observational data are similar in effect to the small-scale dissipation mechanisms used in the WRF model (see related discussion in Skamarock 2004).

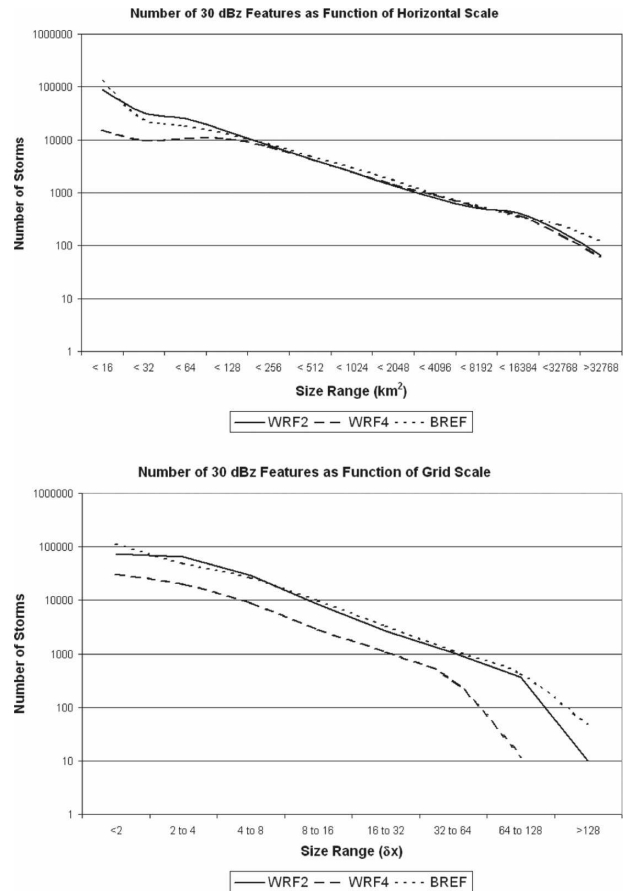


FIG. 11. Model climatology: Number of individual 30-dBZ reflectivity entities (SRF for the models, BREF for the observations) as a function of size. (a) The distribution as a function of the absolute areal coverage of individual entities (km<sup>2</sup>) and (b) the distribution as a function of model (and observations) grid spacing. The data include the 12–30-h forecast period and all days of the Spring Experiment.

## 2) PRECIPITATION

Diurnal trends in areal coverage of the WRF2 and WRF4 precipitation fields (based on 1-h accumulation) are similar in many ways to the corresponding cycles of the SRF fields (cf. the curves associated with the forecast fields in Figs. 9 and 12), as expected. For example, at a given precipitation threshold, the WRF4 runs tend to produce greater coverage through the initial overnight period and into the next morning, while the WRF2 coverage is higher thereafter (Figs. 12a and 12b). Likewise, higher-amplitude cycles are associated with higher thresholds. Again, there is some suggestion that the WRF4 cycle lags the WRF2 cycle by about an hour.

The diurnal cycle of measured precipitation is also quite similar to its corresponding representation in the reflectivity field (cf. the BREF and stage II curves in

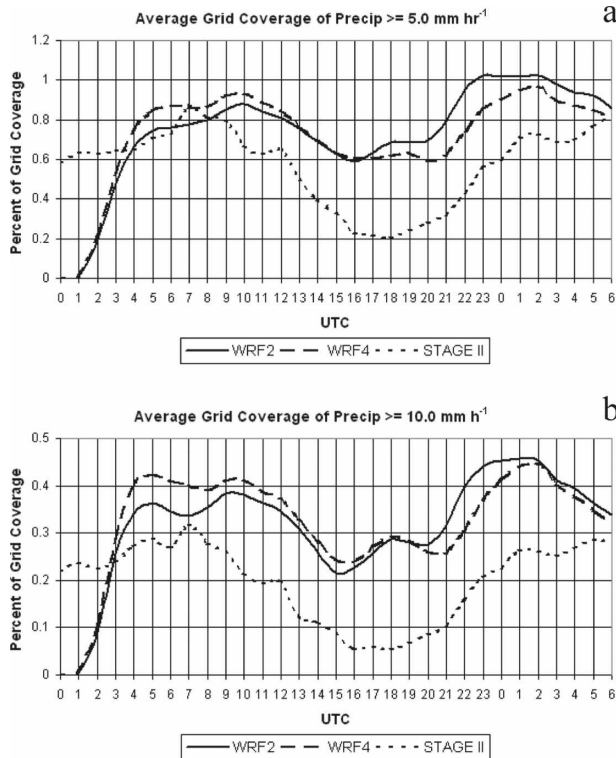


FIG. 12. Model climatology: Areal coverage of precipitation rate as a function of time exceeding the (a) 5 and (b) 10 mm h<sup>-1</sup> thresholds, averaged over all days during the Spring Experiment.

Figs. 9 and 12). All of the BREF and stage II curves reach an initial maximum value in the early morning hours (corresponding to the 7–9-h forecast time), a minimum in the early to midafternoon (16–20-h forecast time), and then show a trend toward a second maximum at the end of the data period during the next night.

While many similarities can be found between Figs. 9 and 12 (the diurnal cycles of reflectivity and precipitation, respectively), there is a glaring difference: coverage of the observed reflectivity field is generally greater than the simulated reflectivity (model bias <1 for reflectivity; see Fig. 9), while coverage of the observed precipitation field is generally less than the simulated precipitation (bias >1 for precipitation; see Fig. 12). The latter bias is corroborated by domain-average precipitation rates (Fig. 13). The models overpredict the total precipitation, especially during the late morning to early afternoon hours (16–20-h forecast time) when the model produces about twice as much precipitation as observed. A high bias in precipitation has been noted in other similarly configured WRF forecasts and the cause for this is under investigation (Weisman et al. 2008). Recent implementation in WRF-ARW of a positive-definite advection routine is expected to mitigate this

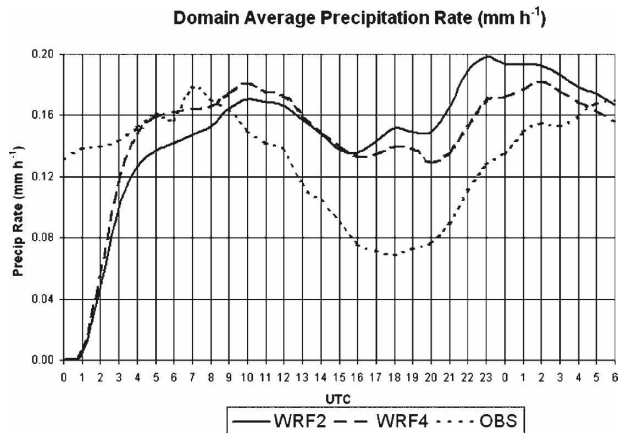


FIG. 13. Model climatology: Domain-average precipitation rates as a function of time, averaged over all days during the Spring Experiment.

bias in future applications (W. Skamarock, NCAR, 2007, personal communication).

The inconsistency between the precipitation and reflectivity biases seems surprising at first, but it is worth reiterating (see section 2c) that SRF and BREF are fundamentally different quantities. Although they provide very similar qualitative information regarding mesoscale circulations and precipitation structures (see Koch et al. 2005), quantitative comparisons between these fields must be made with caution. On the other hand, there is much less ambiguity in comparisons of observed and predicted precipitation accumulations. Thus, the diagnosis of a high bias in precipitation appears to be physically consistent and quite robust, while the low bias in reflectivity involves more quantitative uncertainty.

### c. Supercells/mesocyclones

The defining characteristic of supercells (or mesocyclones) is a persistent deep rotating updraft (e.g., Johns and Doswell 1992). Currently, forecasters at the SPC use a variety of empirical diagnostic tools to assess the likelihood of this unique class of storms, based on environmental parameters predicted by NWP models. In particular, they use these tools to highlight areas in which various combinations of vertical shear and instability would favor the supercell mode, contingent upon the development of storms. But current operational models with parameterized convection do not predict the storms themselves. In contrast, the WRF4 and WRF2 configurations used in this study do generate explicitly resolved rotating convective cells that resemble observed supercells in many ways. Because model output is available only once per hour, the



dataset available for this study is poorly suited for evaluating characteristics such as cell longevity, storm splitting, deviate propagation, etc.—characteristics that are commonly associated with observations of supercells—but it does allow us to detect deep rotating updrafts in the model output. For example, if one zooms in on the northeastern Colorado storms on the left-hand side of Fig. 5, mesocyclone structures begin to emerge in both the WRF2 and WRF4 output fields (Fig. 14). In particular, the reflectivity patterns in some of the simulated storms show characteristics of the classic hook-echo pattern associated with supercells, with an inflow notch in the southeastern quadrant of the storms. Collocated with this notch is a deep rotating updraft in the lower to middle troposphere, indicated by the localized maximum in UH (see section 2e for a detailed description of this diagnostic parameter).

In this case, the model-predicted mesocyclones corresponded remarkably well in both time and space to the observed storms that had supercell characteristics (cf. SRF to BREF in Fig. 14), but such close correspondence was rare during the experiment. Nonetheless, the model forecasts still appeared to have some skill and reliability in predicting the regional frequency of mesocyclone development, after a subjective calibration process was completed. The calibration involved establishing a UH exceedance threshold that produced a reasonable number of mesocyclone “alerts” in the model output. In particular, a threshold value was selected so that the number of rotating storms identified in the hourly snapshots from the WRF2 was roughly comparable to the number of mesocyclones identified at the top of the hour by the National Weather Service’s Next Generation Doppler Radar (NEXRAD) based Mesocyclone Detection Algorithm (MDA; see Stumpf et al. 1998). This UH threshold was  $50 \text{ m}^2 \text{ s}^{-2}$ . Although the MDA alerts provided a useful baseline and rough calibration for a “first look” at daily explicit supercell predictions, they were not deemed to be suitable for a robust quantitative evaluation of model skill because they are not applied consistently at all local forecast offices and suffer from other known biases. Ongoing work in the radar community may allow for more rigorous verification of UH alerts in the future, but currently there is no suitable dataset for observational verification of supercells. Nonetheless, it is useful to compare UH alerts from the two model configurations.

During the experiment, individual high-UH entities were so small they were barely discernible on displays of the regional domains. Consequently, simulated storms that exceeded the threshold value were flagged with 50-mi-wide open circles. For example, Fig. 15

shows where simulated mesocyclones were identified in the 28-h forecast valid at 0400 UTC 1 June 2005 in both the WRF4 (top) and WRF2 (bottom). [Note that this is from the same event shown in Fig. 4.] The locations of MDA alerts issued during the preceding hour are indicated by filled circles. In this case, both configurations of the model correctly predicted mesocyclones, but there was an apparent northward displacement error, consistent with the reflectivity fields shown in Fig. 4.

Forecast displacement errors of this magnitude were common during the experiment. Nonetheless, forecasts like this were still considered quite useful because they showed that discrete thunderstorms containing mesocyclones were distinctly possible in this environment. Thus, it is instructive to examine the climatology of model-predicted supercells, and within the context of this study, it is important to compare the climatologies of supercells in the WRF4 and WRF2 forecasts.

The mean fractional coverage of grid points with UH value  $s \geq 50 \text{ m}^2 \text{ s}^{-2}$  peaks at about 0000 UTC in both sets of forecasts (Fig. 16). Coverage in the WRF2 forecasts is considerably larger, but this is not surprising because each of the components of UH (vertical vorticity and vertical velocity) is expected to scale with grid spacing. Following Adlerman and Droegemeier (2002), one way to determine an appropriate scaling factor is to take an average of the peak values of each component over all hours and days. Using this approach, the vertical vorticity term in the UH calculation scales by a factor of about 2.0 in going from 4- to 2-km grid spacing, while the vertical velocity term scales by about 1.3, giving a combined factor of about 2.6. Interestingly, this is approximately proportionate to the difference in amplitude between the two curves in Fig. 16. Thus, it appears that it may be relatively simple to scale the UH threshold as a function of grid spacing to produce essentially the same prediction of the areal coverage of mesocyclones using 2- and 4-km grids.

Of course, areal coverage is not necessarily the field that we want. Our subjectively determined threshold value was based on the number of individual UH entities predicted by the model rather than the areal coverage. When the climatologies of the numbers of UH entities are quantified rather than their areal coverage, the disparity in amplitude between the WRF2 and WRF4 forecasts becomes greater (Fig. 17, top). However, much of this difference can again be reconciled by simple scaling arguments, at least in a qualitative sense. If the factor of 2.6 is applied to the UH field from the WRF4 before the search for contiguous entities, the climatologies become more similar (Fig. 17, bottom). Furthermore, recognizing that the stronger UH entities are very small in scale and that the 2-km configuration



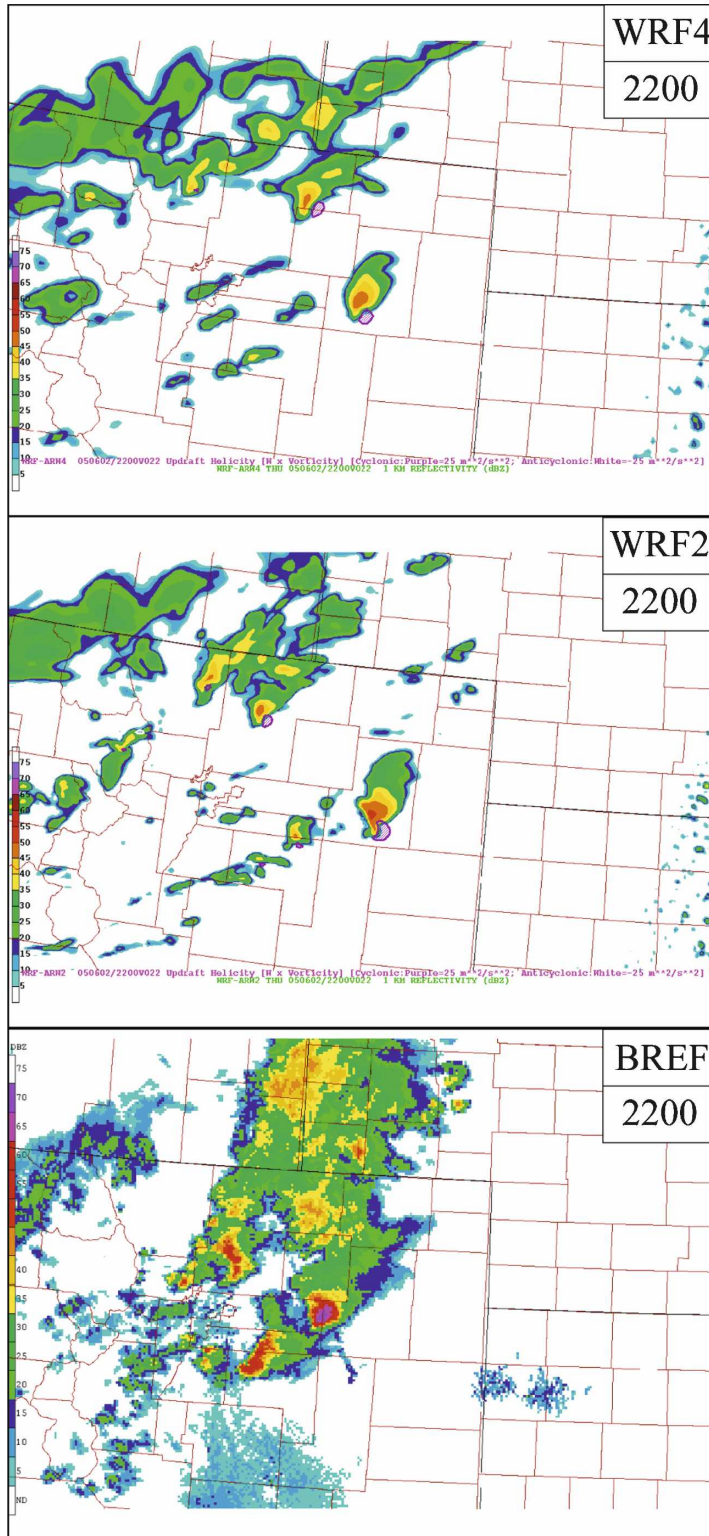


FIG. 14. As in Fig. 5 but zoomed-in on northeastern CO and with purple hatching in the top two panels indicating areas where  $UH \geq 25 \text{ m}^2 \text{ s}^{-2}$ .

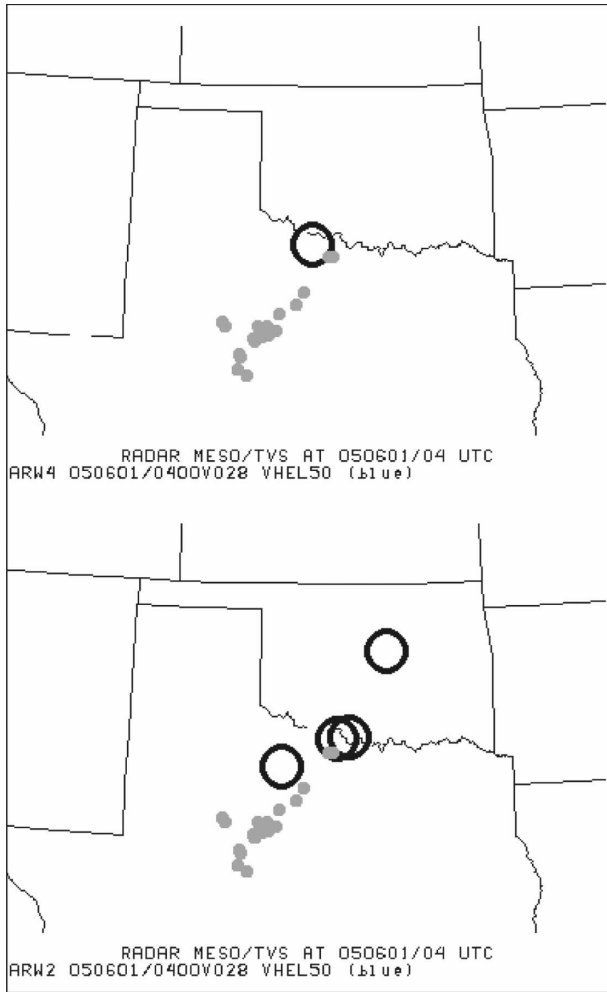


FIG. 15. The UH entities from the (a) WRF4 and (b) WRF2 forecasts indicated by large open circles, with NWS MDA alerts from the preceding hour indicated by filled smaller circles; 28-h forecasts valid at 0400 UTC 1 Jun 2005.

inherently produces more small-scale features (e.g., see Figs. 10 and 11), one can deduce that a second scaling factor is needed to account for the inherent differences in the effective resolution at 2 and 4 km. Although the specific magnitude of this second factor is not immediately obvious, it appears that the application of the appropriate factors might allow the 4-km configuration to provide comparable information regarding the numbers of mesocyclones as well as their areal coverage.

**4. Summary and discussion**

During the 2005 Hazardous Weather Testbed (HWT) Spring Experiment, a series of convection-allowing, large-domain, numerical forecasts were generated using the WRF model. In this study, two sets of

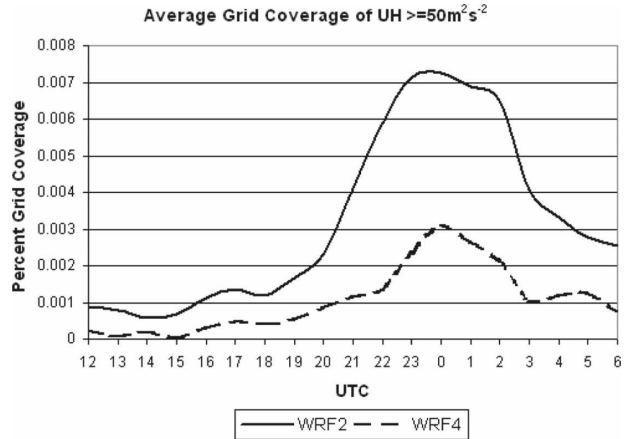


FIG. 16. Model climatology: Areal coverage of  $UH \geq 50 \text{ m}^2 \text{ s}^{-2}$  as a function of time, averaged over all days during the Spring Experiment.

these forecasts were compared to assess the impact of grid resolution. Specifically, forecasts using 4-km horizontal grid spacing with 35 vertical levels were compared to corresponding predictions using 2-km grid

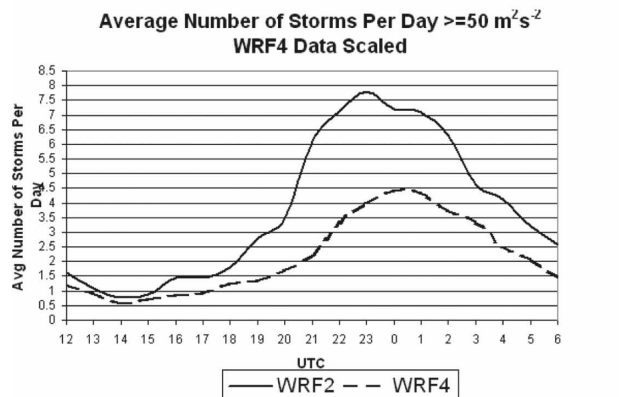
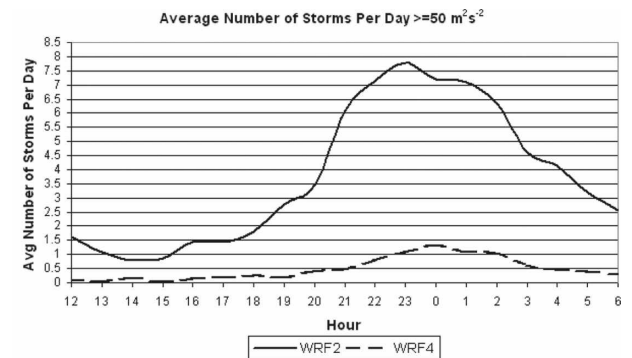


FIG. 17. Model climatology: Average number of individual UH entities exceeding the  $50 \text{ m}^2 \text{ s}^{-2}$  threshold as a function of time, based on all days during the Spring Experiment. In the bottom panel, the UH calculation for the WRF4 data includes a scaling factor of about 2.6 based on individual scaling factors of about 2.0 for the vorticity and 1.3 for the  $w$  terms in the calculation.

spacing with 51 levels in the vertical. The comparison began by highlighting the evolution of simulated reflectivity fields for two representative convective events, then providing a summary of subjective perceptions based on similar fields that were examined on all days during the Spring Experiment. The subjective assessments were substantiated and additional insight into WRF model behavior was gained through objective measures of model climatology, including comparisons of model reflectivity and accumulated precipitation with their observational counterparts. Finally, a preliminary investigation into model predictions of deep rotating updrafts, a coarsely represented analog to supercells in the real atmosphere, was presented.

In the two highlighted events the simulated reflectivity fields produced by the different model configurations, and the mesoscale organizational structures implied by the reflectivity fields, were generally very similar to each other, but somewhat different from the corresponding observed reflectivity. In particular, both model configurations showed skill in predicting the dominant mesoscale convective organization (i.e., convective mode) and general convective evolution in these events, but there were substantial errors in the placement, configuration, and mesoscale details of major features. The WRF2 provided reflectivity fields with finer-scale structure than did those from the WRF4, but it was not clear that there was any value to this additional finescale structure in this context, because of overriding errors on larger scales.

Subjective verification statistics from the Spring Experiment did not reveal any significant differences in mean ratings of WRF2 and WRF4 forecasts in terms of convective initiation and overall evolution. These ratings were based on the simulated reflectivity field. Although the average ratings were essentially the same, WRF2 forecasts for initiation were rated higher on 6 days, compared to just 1 day for WRF4 (they were rated identically on the remaining days), perhaps suggesting that the 2-km grid spacing could provide some benefit for convective initiation forecasts.

Objective assessments of simulated reflectivity and 1-h accumulated precipitation fields, and their observational counterparts, were revealing in several ways:

- They corroborated the subjective ratings. For example, equitable threat scores and depictions of the diurnal cycle of storm coverage showed that, in the mean, the precipitation and reflectivity fields from the two model configurations were much more similar to each other than to the observations.
- They provided useful information about how both WRF configurations represent the diurnal cycle of

convection. For example, both configurations appeared to underestimate the amplitude of the diurnal cycle and predict the minima and maxima too early; there was some indication that the WRF4 cycle lagged behind the WRF2 by about an hour and both configurations failed to predict the observed nocturnal maximum in precipitation.

- They revealed systematic biases in the simulated reflectivity fields. SRF coverage generally suffered from a low bias; that is, the coverage at a given dBZ threshold was typically less than the corresponding BREF coverage. In contrast, the more direct comparison of model-generated versus observed precipitation revealed a high bias, consistent with previous studies. Furthermore, the SRF bias rapidly approached a value of zero (no coverage) above about 45 dBZ.

In addition, when reflectivity entities were counted and categorized according to size, it was confirmed that the WRF2 generated many more individual small-scale reflectivity features than did the WRF4, in spite of the fact that the total coverage at a given reflectivity threshold was very similar. When the entity sizes were plotted as a function of the number of grid cells, rather than the absolute dimensions, the WRF4 and WRF2 distributions were approximately parallel. This implies that the size distribution of entities across all allowable scales was approximately the same regardless of grid spacing, even though the spectrum of entity sizes expanded to smaller *absolute* scales as the resolution was increased [see Skamarock (2004) for a related discussion on kinetic energy spectra as a function of grid spacing].

A new diagnostic field, updraft helicity (UH), was used as a surrogate for thunderstorms containing mesocyclones, or supercells, in the model forecasts. Experimental forecasting activities during the Spring Experiment indicated that, on many days, the models generated storms containing localized UH maxima under environmental conditions that produced observed supercells. Thus, subjective assessments indicated that the UH field had some utility as a forecast tool. WRF2 forecasts generated greater numbers of storms containing identifiable UH maxima, and UH values tended to be higher in WRF2 storms, as expected from simple scaling arguments. However, in a qualitative sense the different configurations appeared to have comparable levels of reliability and skill in predicting the occurrence of one (or more) identifiable mesocyclone under environmental conditions that are commonly associated with observed supercells.

Before concluding, it is important to reiterate the context of this study. It is part of an investigation of the

utility of convection-allowing output from the WRF model as guidance for human forecasts of *severe convection over the relatively flat terrain of the central United States 18–30 h after model initialization*. In this time frame, both subjective and objective assessments revealed systematic differences between the WRF2 and WRF4 forecasts that appeared to be associated with the disparity in grid resolution. In general, the WRF2 configuration produced more detailed structures and more numerous small-scale features, including more mesocyclones; there was some suggestion that convective activity was initiated slightly earlier in the WRF2 forecasts. Yet, in this context the practical value of both configurations remained anchored in their ability to provide reliable guidance about the mesoscale organization and evolution of deep convection, information that is lacking in coarser-resolution models that parameterize deep convection. The different configurations appeared to have comparable levels of skill in this regard. The more detailed structures produced by the WRF2 were intriguing, but appeared to add little, if any, practical value as forecast guidance. Thus, for forecast applications of this type, decreasing grid spacing from 4 to 2 km may not be worth the  $\sim$ tenfold increase in computational expense.

*Acknowledgments.* Special thanks and appreciation are extended to many people for their creative insights and assistance with 2005 Spring Experiment preparations/planning, programming, and data flow issues. Without the combined efforts of many SPC and NSSL staff members, the Spring Experiment could not be conducted. In particular, special thanks to Phillip Bothwell (SPC) for providing access to radar and severe storm report verification data; Gregg Grosshans (SPC) for establishing model data flow and configuring the experimental forecasts for transmission and archival, and for developing and organizing model display files; and Jay Liang (SPC) and Doug Rhue (SPC) for assistance in configuring and upgrading hardware–software and display systems in the Science Support Area. Linda Crank (SPC), Peggy Stogsdill (SPC), and Sandra Allen (NSSL), ably assisted with logistical and budget support activities.

We are grateful to Yunheng Wang and Keith Brewster of CAPS, David O’Neal of the Pittsburgh Supercomputing Center, and Wei Wang of NCAR for their exceptional efforts to run and provide output from the high-resolution WRF model. We are also indebted to Zavis Janjić, Tom Black, Matt Pyle, and Geoff DiMego of EMC for providing a separate set of WRF model forecasts that have been critically important for other studies related to the 2005 Spring Experiment.

We further wish to recognize the full support of SPC and NSSL management, and the numerous contributions and insights provided by the many participants who clearly demonstrated the value of collaborative experiments involving the research, academic, and forecasting communities, and whose presence and enthusiasm resulted in a positive learning experience for everyone.

We thank Unisys Corporation for supplying observed radar reflectivity fields and also Mark Logan of Unisys for providing a brief description of the process used to generate the operational radar images.

We very much appreciate the contributions of Lou Wicker and Bob Davies-Jones of NSSL, who derived various algorithms to detect mesocyclones in the high-resolution model output.

#### REFERENCES

- Adlerman, E. J., and K. K. Droegemeier, 2002: The sensitivity of numerically simulated cyclic mesocyclogenesis to variations in model physical and computational parameters. *Mon. Wea. Rev.*, **130**, 2671–2691.
- Arakawa, A., 2004: The cumulus parameterization problem: Past, present, and future. *J. Climate*, **17**, 2493–2525.
- Baldwin, M. E., and K. E. Mitchell, 1998: Progress on the NCEP hourly multi-sensor U.S. precipitation analysis for operations and GCIP research. Preprints, *Second Symp. on Integrated Observing Systems*, Phoenix, AZ, Amer. Meteor. Soc., 10–11.
- , and M. S. Wandishin, 2002: Determining the resolved spatial scales of Eta model precipitation forecasts. Preprints, *15th Conf. on Numerical Weather Prediction*, San Antonio, TX, Amer. Meteor. Soc., 85–88.
- , and J. S. Kain, 2006: Sensitivity of several performance measures to displacement error, bias, and event frequency. *Wea. Forecasting*, **21**, 636–648.
- Bélaïr, S., and J. Mailhot, 2001: Impact of horizontal resolution on the numerical simulation of a midlatitude squall line: Implicit versus explicit condensation. *Mon. Wea. Rev.*, **129**, 2362–2376.
- Black, T. L., 1994: The new NMC mesoscale Eta Model: Description and forecast examples. *Wea. Forecasting*, **9**, 265–278.
- Bryan, G. H., J. C. Wyngaard, and J. M. Fritsch, 2003: Resolution requirements for the simulation of deep moist convection. *Mon. Wea. Rev.*, **131**, 2394–2416.
- Chen, F., K. W. Manning, D. N. Yates, M. A. LeMone, S. B. Trier, R. Cuenca, and D. Niyogi, 2004: Development of high resolution land data assimilation system and its application to WRF. Preprints, *16th Conf. on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., 22.3. [Available online at <http://ams.confex.com/ams/pdfpapers/67333.pdf>.]
- Davis, C., and Coauthors, 2004: The Bow Echo and MCV Experiment (BAMEX): Observations and opportunities. *Bull. Amer. Meteor. Soc.*, **85**, 1075–1093.
- Deng, A., and D. R. Stauffer, 2006: On improving 4-km mesoscale model simulations. *J. Appl. Meteor. Climatol.*, **45**, 361–381.
- Done, J., C. Davis, and M. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the Weather Research and Forecasting (WRF) model. *Atmos. Sci. Lett.*, **5** (6), 110–117.

- Droegemeier, K. K., S. M. Lazarus, and R. P. Davies-Jones, 1993: The influence of helicity on numerically simulated convective storms. *Mon. Wea. Rev.*, **121**, 2005–2029.
- , G. Bassett, and M. Xue, 1994: Very high-resolution, uniform-grid simulations of deep convection on a massively parallel processor: Implications for small-scale predictability. Preprints, *10th Conf. on Numerical Weather Prediction*, Portland, OR, Amer. Meteor. Soc., 376–379.
- , —, D. K. Lilly, and M. Xue, 1996: Does helicity really play a role in supercell longevity? Preprints, *18th Conf. on Severe Local Storms*, San Francisco, CA, Amer. Meteor. Soc., 205–209.
- Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077–3107.
- Gallus, W. A., Jr., and M. Segal, 2001: Impact of improved initialization of mesoscale features on convective system rainfall in 10-km Eta simulations. *Wea. Forecasting*, **16**, 680–696.
- Hong, S.-Y., J. Dudhia, and S.-H. Chen, 2004: A revised approach to ice-microphysical processes for the bulk parameterization of cloud and precipitation. *Mon. Wea. Rev.*, **132**, 103–120.
- Iacono, M. J., E. J. Mlawer, S. A. Clough, and J.-J. Morcrette, 2000: Impact of an improved longwave radiation model, RRTM, on the energy budget and thermodynamic properties of the NCAR Community Climate Model, CCM3. *J. Geophys. Res.*, **105**, 14 873–14 890.
- Janjić, Z. I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927–945.
- , T. L. Black, M. Pyle, H.-Y. Chuang, E. Rogers, and G. DiMego, cited 2007: An evolutionary approach to non-hydrostatic modeling. [Available online at [http://www.wrf-model.org/wrfadmin/publications/Chuang\\_Janjić\\_NWP50yearsfinalshort.pdf](http://www.wrf-model.org/wrfadmin/publications/Chuang_Janjić_NWP50yearsfinalshort.pdf).]
- Johns, R. H., and C. A. Doswell III, 1992: Severe local storms forecasting. *Wea. Forecasting*, **7**, 588–612.
- Jung, J.-H., and A. Arakawa, 2004: The resolution dependence of model physics: Illustrations from nonhydrostatic model experiments. *J. Atmos. Sci.*, **61**, 88–102.
- Kain, J. S., M. E. Baldwin, S. J. Weiss, P. R. Janish, M. P. Kay, and G. Carbin, 2003a: Subjective verification of numerical models as a component of a broader interaction between research and operations. *Wea. Forecasting*, **18**, 847–860.
- , P. R. Janish, S. J. Weiss, M. E. Baldwin, R. S. Schneider, and H. E. Brooks, 2003b: Collaboration between forecasters and research scientists at the NSSL and SPC: The Spring Program. *Bull. Amer. Meteor. Soc.*, **84**, 1797–1806.
- , S. J. Weiss, M. E. Baldwin, G. W. Carbin, D. A. Bright, J. J. Levit, and J. A. Hart, 2005: Evaluating high-resolution configurations of the WRF model that are used to forecast severe convective weather: The 2005 SPC/NSSL Spring Program. Preprints, *21th Conf. on Weather Analysis and Forecasting/17th Conf. on Numerical Weather Prediction*, Washington, DC, Amer. Meteor. Soc., 2A.5. [Available online at <http://ams.confex.com/ams/pdfpapers/94843.pdf>.]
- , —, J. J. Levit, M. E. Baldwin, and D. R. Bright, 2006: Examination of convection-allowing configurations of the WRF model for the prediction of severe convective weather: The SPC/NSSL Spring Program 2004. *Wea. Forecasting*, **21**, 167–181.
- Kalb, M. W., 1987: The role of convective parameterization in the simulation of a Gulf coast precipitation system. *Mon. Wea. Rev.*, **115**, 214–234.
- Koch, S. E., B. Ferrier, M. Stolinga, E. Szoke, S. J. Weiss, and J. S. Kain, 2005: The use of simulated radar reflectivity fields in the diagnosis of mesoscale phenomena from high-resolution WRF model forecasts. Preprints, *12th Conf. on Mesoscale Processes*, Albuquerque, NM, Amer. Meteor. Soc., J4J.7. [Available online at <http://ams.confex.com/ams/pdfpapers/97032.pdf>.]
- Lean, H. W., P. A. Clark, M. Dixon, N. M. Roberts, A. Fitch, R. Forbes, and C. Halliwell, 2008: Characteristics of high-resolution versions of the Met Office Unified Model for forecasting convection over the United Kingdom. *Mon. Wea. Rev.*, **136**, 3408–3424.
- Lilly, D. K., G. M. Bassett, K. K. Droegemeier, and P. Bartello, 1998: Stratified turbulence in the atmospheric mesoscales. *Theor. Comput. Fluid Dyn.*, **11**, 139–153.
- Liu, C., M. W. Moncrieff, and W. W. Grabowski, 2001: Explicit and parameterized realizations of convective cloud systems in TOGA COARE. *Mon. Wea. Rev.*, **129**, 1689–1703.
- Mesinger, F., 1996: Improvements in quantitative precipitation forecasts with the Eta regional model at the National Centers for Environmental Prediction: The 48-km upgrade. *Bull. Amer. Meteor. Soc.*, **77**, 2637–2650.
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16 663–16 682.
- Molinari, J. M., and M. Dudek, 1986: Implicit versus explicit convective heating in numerical weather prediction models. *Mon. Wea. Rev.*, **114**, 326–344.
- , and —, 1992: Parameterization of convective precipitation in mesoscale numerical models: A critical review. *Mon. Wea. Rev.*, **120**, 326–344.
- Narita, M., and S. Ohmore, 2007: Improving precipitation forecasts by the operational nonhydrostatic mesoscale model with the Kain–Fritsch convective parameterization and cloud microphysics. Preprints, *12th Conf. on Mesoscale Processes*, Waterville Valley, NH, Amer. Meteor. Soc., 3.7. [Available online at <http://ams.confex.com/ams/pdfpapers/126017.pdf>.]
- Noh, Y., W. G. Cheon, S.-Y. Hong, and S. Raasch, 2003: Improvement of the K-profile model for the planetary boundary layer based on large eddy simulation data. *Bound.-Layer Meteor.*, **107**, 401–427.
- Petch, J. C., 2006: Sensitivity studies of developing convection in a cloud-resolving model. *Quart. J. Roy. Meteor. Soc.*, **132**, 345–358.
- , A. R. Brown, and M. E. B. Gray, 2002: The impact of horizontal resolution on the simulations of convective development over land. *Quart. J. Roy. Meteor. Soc.*, **128**, 2031–2044.
- Roberts, N. M., 2003: Results from high resolution modeling of convective events. U.K. Met Office Tech. Rep. 402, 47 pp. [Available online at [http://www.metoffice.gov.uk/research/nwp/publications/papers/technical\\_reports/2003/FRTR402/FRTR402.pdf](http://www.metoffice.gov.uk/research/nwp/publications/papers/technical_reports/2003/FRTR402/FRTR402.pdf).]
- , 2005: An investigation of the ability of a storm scale configuration of the Met Office NWP model to predict flood-producing rainfall. U.K. Met Office Tech. Rep. 455, 80 pp. [Available online at [http://www.metoffice.gov.uk/research/nwp/publications/papers/technical\\_reports/2005/FRTR455/FRTR455.pdf](http://www.metoffice.gov.uk/research/nwp/publications/papers/technical_reports/2005/FRTR455/FRTR455.pdf).]
- Rosenthal, S. L., 1978: Numerical simulation of tropical cyclone development with latent heat release by the resolvable scales. I: Model description and preliminary results. *J. Atmos. Sci.*, **35**, 258–271.

- Seo, D. J., 1998: Real-time estimation of rainfall fields using radar rainfall and rain gauge data. *J. Hydrol.*, **208**, 37–52.
- Skamarock, W. C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Wea. Rev.*, **132**, 3019–3032.
- , J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, 2005: A description of the Advanced Research WRF version 2. NCAR Tech Note NCAR/TN-468+STR, 88 pp. [Available from UCAR Communications, P. O. Box 3000, Boulder, CO 80307.]
- Speer, M. S., and L. M. Leslie, 2002: The prediction of two cases of severe convection: Implications for forecast guidance. *Meteor. Atmos. Phys.*, **80**, 165–174.
- Steppeler, J., G. Doms, U. Schattler, H. W. Bitzer, A. Gassmann, and U. Damrath, 2003: Meso-gamma scale forecasts using the nonhydrostatic model LM. *Meteor. Atmos. Phys.*, **82**, 75–96.
- Stumpf, G. J., A. Witt, E. D. Mitchell, P. L. Spencer, J. T. Johnson, M. D. Eilts, K. W. Thomas, and D. W. Burgess, 1998: The National Severe Storms Laboratory Mesocyclone Detection Algorithm for the WSR-88D. *Wea. Forecasting*, **13**, 304–326.
- Weisman, M. L., and J. B. Klemp, 1984: The structure and classification of numerically simulated convective storms in directionally varying wind shears. *Mon. Wea. Rev.*, **112**, 2479–2498.
- , W. C. Skamarock, and J. B. Klemp, 1997: The resolution dependence of explicitly modeled convective systems. *Mon. Wea. Rev.*, **125**, 527–548.
- , C. Davis, W. Wang, K. W. Manning, and J. B. Klemp, 2008: Experiences with 0–36-h explicit convective forecasts with the WRF-ARW model. *Wea. Forecasting*, **23**, 407–437.
- Weiss, S. J., J. S. Kain, J. J. Levit, M. E. Baldwin, and D. R. Bright, 2004: Examination of several different versions of the WRF model for the prediction of severe convective weather: The SPC/NSSL Spring Program 2004. Preprints, *22nd Conf. on Severe Local Storms*, Hyannis, MA, Amer. Meteor. Soc., 17.1. [Available online at <http://ams.confex.com/ams/pdfpapers/82052.pdf>.]
- Zhang, D.-L., E.-Y. Hsie, and M. W. Moncrieff, 1988: A comparison of explicit and implicit prediction of convective and stratiform precipitating weather systems with a meso- $\beta$ -scale numerical model. *Quart. J. Roy. Meteor. Soc.*, **114**, 31–60.