3B.5 SOME PRACTICAL CONSIDERATIONS FOR THE FIRST GENERATION OF OPERATIONAL CONVECTION-ALLOWING NWP: HOW MUCH RESOLUTION IS ENOUGH?

John S. Kain*[1], Steven J. Weiss[2], David R. Bright[2], Michael E. Baldwin[3], Jason J. Levit[2], Gregory W. Carbin[2], Craig S. Schwartz[4], Morris Weisman[5], Kelvin K. Droegemeier[6], Daniel Weber[6], Kevin W. Thomas[6],

[1]NOAA/OAR/National Severe Storms Laboratory
[2]NOAA/NWS/Storm Prediction Center
[3]Department of Earth and Atmospheric Sciences, Purdue University
[4]Department of Meteorology, The Pennsylvania State University
[5]National Center for Atmospheric Research
[6]Center for Analysis and Prediction of Storms, University of Oklahoma

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 U. S., now called the North American Mesoscale model (NAM) (Black 1994; Janjic 1994). Scientists at NCEP's (National Center for Environmental Prediction) Environmental Modeling Center (EMC) have decreased the NAM's grid spacing decrease 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 levelled 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 when grid spacing is greater than about 10 km (see Molinari and Dudek 1986; Kalb 1987; Zhang et al. 1988; Gallus and Segal 2001; Belair and Mailhot 2001; Liu et al. 2001), but parameterization (CP) is typically avoided at higher resolution. This avoidance is appropriate because the conceptual basis for CP becomes increasingly ambiguous as grid spacing decreases further (Molinari and Dudek 1992; Arakawa 2004); it is also practical because the available parameterizations simply do not work very well at higher resolution.

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, other problems arise when grid resolution is too coarse to represent convective processes well and CP is disabled . For example, numerous studies have shown that convective overturning tends to develop and evolve too slowly when it is poorly resolved in a model, and that updraft and

* *Corresponding author address:* Jack Kain, National Severe Storms Laboratory, 120 David L. Boren Blvd., Norman, OK 73072; e-mail: jack.kain@noaa.gov

downdraft mass fluxes, along with precipitation rates, are too strong when the convective process matures with this model configuration (e.g., Molinari and Dudek 1986; Moncrieff and Klinker 1997). Furthermore, the likelihood of the failure mode (no convective development) increases without CP when grid resolution is too coarse to represent the processes responsible for convective initiation (e.g., Liu et al. 2001). 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.

In a test of sensitivity to grid spacing, Weisman et al. (1997) found that a spacing of 4 km is adequate to represent the mesoscale structures associated strong mid-latitude squall lines. Inspired by this result, the National Center of Atmospheric Research (NCAR) generated realtime, large domain (about 2/3 CONUS) 4-km WRF-model forecasts in support of the BAMEX project (Davis et al. 2004) during the late spring and summer of 2003. These 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).

These results motivated a broader series of realtime high-resolution forecasts, generated daily in the spring and early summer of 2004 and evaluated in the 2004 SPC/NSSL (Storm Prediction Center/National Severe Storms Laboratory) Spring Experiment[1]. In this experiment, large-domain, approximately 4-km forecasts were generated by NCEP/EMC and CAPS (the Center for Analysis and Prediction of Storms), 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) for convective-system mode (morphology), and comparable guidance for convective initiation and evolution (Kain et al. 2006). Thus, the Done et al. (2004) results with regard to morphology were corroborated and concerns about the failure mode at $\delta x = 4$ km (no convective initiation) were alleviated.

In 2005, a third round of high-resolution forecasts were conducted, again in collaboration with the annual SPC/NSSL Spring Experiment. Primary contributors were the same as in 2004 - NCAR, EMC, and CAPS - but the emphasis of the model evaluation was somewhat different. In particular, rather than compare the high-resolution models to operational models with CP, 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 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 increasing grid resolution from 4 km to 2 km. This latter focus was enabled by CAPS scientists, who ran with grid spacing of just 2 km in 2005, while NCAR and EMC continued to work at about 4 km spacing.

The sensitivity to grid resolution has important practical implications because doubling the resolution requires at least a ten-fold increase in computer power (in addition to

1. 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. Additional details about the HWT and each annual experiment can be found at URL http://www.nssl.noaa.gov/hwt

much greater demands on storage and dissemination 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?

Bryan et al. (2003) argued that convection-allowing models cannot be considered convection-*resolving* until grid spacing approaches 100 m, primarily because they are not designed to parameterize in-cloud turbulence, including the entrainment process, an inherent component of convective overturning. Clearly, that kind of resolution is out of reach for many NWP generations to come. On a more modest scale, Petch et al. (2002) suggested that grid increments below 1 km were necessary to predict accurately the timing and intensity of convective activity forced by surface heating over land. They hypothesized that such fine resolution was needed to resolve the boundary layer eddies that were responsible for convective initiation. Meanwhile, numerous studies by CAPS scientists have shown that explicit simulations of convective supercells are very sensitive to grid spacing in the range from 0.25 km to 2 km (i.e., Droegemeier et al. 1994; Droegemeier et al. 1996; Lilly et al. 1998; Adlerman and Droegemeier 2002). Thus, it seems clear that while 4 km spacing may be adequate to resolve and perhaps predict organized convective systems well, much higher resolution may be required to predict details of the convective process well.

For severe weather forecasters, such as those at the SPC, the CAPS studies have particular relevance because one of the forecasters' biggest challenges is the successful prediction of supercells - storms that produce a disproportionate share of severe weather compared to other modes of convection. Consequently, part of this study addresses the question of supercell forecasts: Can the 4 km and 2 km forecasts provide explicit guidance for the prediction of supercells? Can the 2 km forecasts provide better guidance than their 4 km counterparts?

With these considerations in mind, this study provides an important step in addressing practical considerations related to operational forecasts with convection-allowing models, focusing on severe convection and a comparison of model performance at 2 km and 4 km grid spacing. The Advanced Research WRF model (WRF-ARW; see Skamarock et al. 2005) was used for all forecasts discussed in this paper, although the WRF-NMM (Janjic et al. 2005) 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. Details of the model configurations and experimental methods are provided in the next section, followed by results, then a summary and discussion.

2. METHODOLOGY

During the 2005 Spring Experiment, the different versions of the WRF model were run on a daily basis (Mon.-Fri.) at remote locations and output was collected at the SPC. There the output was utilized to produce experimental forecasts for severe 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 seven weeks. Objective analysis of the data was conducted after the experiment ended.

2.1 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-resolution 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 grid-point dimensions were 1500 X 1320, which, given the realtime production schedule, presented extraordinary computational and data-management challenges. One of these challenges was in the area of initial conditions. CAPS programmers found that prohibitive time delays were introduced when the standard WRF initialization (interpolation) package was used with their grid, so they developed and used a new, more computationally efficient routine. NCAR scientists continued to use the standard package. The different approaches resulted in initial atmospheric conditions that were broadly similar, but visibly different in their finer-scale details.

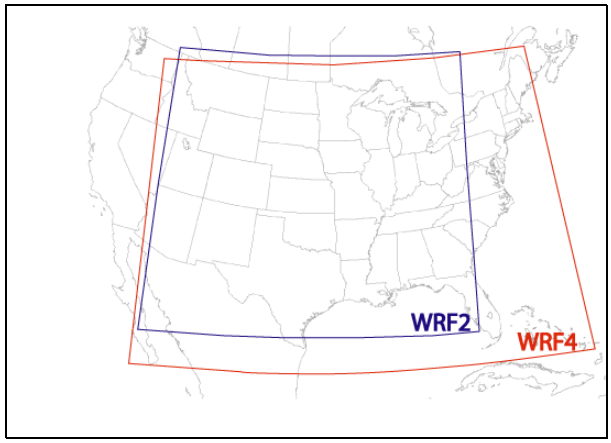

Fig. 1. Model integration domains for the CAPS (WRF2) and NCAR (WRF4) forecasts.

Initial soil-moisture fields were also slightly different. NCAR scientists initialized soil moisture using HRLDAS (High Resolution Land Data Assimilation System - see Chen et al. 2004), an off-line 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 impacted next-day forecasts. The focus of this study is on a

	NCAR	CAPS
Dynamic Core	ARW	ARW
Horiz. Grid Spacing (km)	4.0	2.0
Vertical Levels	35	51
PBL/Turb. Param.	YSU	YSU
Microphysical Param.	WSM6	WSM6
Radiation Param. (SW/LW)	Dudhia/RRTM	Dudhia/RRTM
Initial Conditions	00Z 40 km Eta	00Z 40 km Eta

Table 1. Model configurations used for the high resolution forecasts. ARW: Advanced Research WRF (Skamarock et al. 2005); 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).

comparison of the general character and statistical properties of the two forecasts. In this respect, it is assumed that spatial resolution is the dominant factor that leads to differences in the two forecasts.

2.2 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 five minutes or less on a grid with 2 km spacing between points. The mosaics incorporate the lowest elevation cut (0.5 degree elevation angle) of the most recent reflectivity data from each of the 142 continental United States WSR-88D radars as received via the National Weather Service Radar Product Central Collection Dissemination Services (RPCCDS). 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 by 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 degrees north and 98 degrees west with standard latitude values of 38 degrees and 45 degrees north.

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 co-located, the maximum value takes precedence. Data values from sites operating in clear-air mode are converted to corresponding precipitation mode data level values. Site data exceeding an age limit are excluded. Hereafter, the observed reflectivity fields are referred to as BREF.

[Anecdotally, SPC forecasters have noted that individual storms tend to appear larger in the national mosaic than they do in reflectivity data from the individual WSR-88D sites. If this perception reflects reality, the areal coverage of BREF storms may be artificially enhanced in the mosaic]

2.3 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 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

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

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

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

and ρ_l is the density of liquid water (1000 kg liquid m^{-3}), ρ_a is the local density of dry air

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

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

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

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

where $N_{0s} = 2\times10^7\ m^{-4}$, $N_{0g} = 4\times10^6\ m^{-4}$, ρ_s is the assumed density of snow (100 kg m^{-3}), ρ_g is the assumed density of graupel (400 kg m^{-3}),

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

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

where q_s is the mixing ratio of snow (kg snow kg^{-1} dry air) and q_s is the mixing ratio of graupel. The total reflectivity can then be obtained by simply adding Z_{er}, Z_{es}, and Z_{eg},

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

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

$$SRF(dBZ) = 10\log(Z_e) \quad . \qquad (8)$$

During the 2005 Spring Experiment, 2-D 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 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 degree elevation angle. Thus it senses hydrometeors at relatively low altitudes close to the transmitter, but higher altitudes at greater distances. In contrast, the *SRF* field is based on precipitation particles at a fixed altitude. There is 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.

2.4 Observed Precipitation

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

2.5 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 and Droegemeier 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

$$H = \vec{V} \cdot \nabla \times \vec{V}. \qquad (9)$$

The horizontal component of environmental helicity, commonly expressed in 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 (ref). 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

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

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

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

where ζ is the vertical component of the relative vorticity. Note that ζ can be negative when updrafts are rotating anti-cyclonically; further, *w* is negative in downdrafts, which can also rotate cyclonically or anti-cyclonically. During the 2005 Spring Program, each of the possible combinations of *w* and ζ were examined separately, but in this study the focus is on cyclonically rotating updrafts where both *w* and ζ are positive. Since the primary interest is on storm rotation in the lower to middle troposphere, (3) was integrated vertically from z_0 = 2 km to z_t = 5 km AGL using a midpoint approximation. Specifically, data was available every 1 km AGL, so equation (3) was computed using

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

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.

2.6 Subjective Evaluation

As in previous SPC/NSSL Spring Experiments (e.g., see Weiss et al. 2004; Kain et al. 2003a), 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 models were the focus of the afternoon time period. The strategy for subjective model

evaluation was to provide a short written description of relevant differences between model forecasts and verifying observations, followed by a rating of model performance on a scale of 1 to 10 (see Kain et al. 2003b). Participants were asked to comment specifically on perceived differences between the 2 km 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 example, Fig. 2 shows the forecast/evaluation domain for 24 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 convective initiation and the first few hours of mesoscale convective development and evolution.

For this paper, subjective verification results will be based on the equivalent precipitation-reflectivity field. Although many other fields are relevant and were examined during the Spring Experiment, the equivalent reflectivity field is likely more revealing than any other single output field (see below).

2.7 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 domain (Fig. 3). Reflectivity, precipitation, and UH output fields were examined. 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

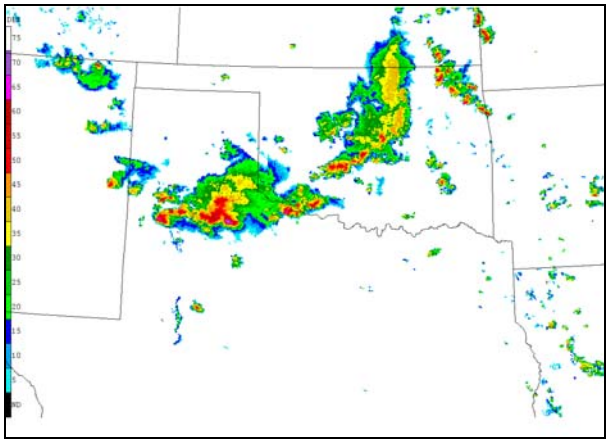

Fig. 2. Forecast/Evaluation domain for May 31, 2005.

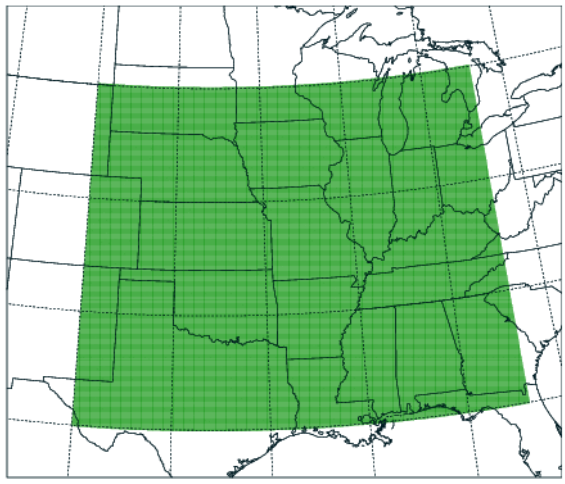

Fig. 3. Domain for the calculation of model climatology.

experiment (33 days for reflectivity and UH data, 31 days for precipitation). The hourly means were plotted as a time series to reveal important differences in the diurnal cycle of the two models and the corresponding observations. In addition, overall size distributions were plotted to provide information about the relative level of detail in the two model forecasts.

In addition, the equitable threat (ET) score (see Baldwin and Kain 2006) 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 ET score 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 rainfall forecasts from different models (e.g., Mesinger 1996), especially those with mesoscale resolution and parameterized convection. 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 features), resulting in little or no overlap between 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 in the context of this study.

3. RESULTS

3.1. Subjective Assessment

A subjective assessment begins with a simple visual comparison, though it must be acknowledged that this comparison is conditioned by our previous experience in examining and interpreting reflectivity fields from weather radar and precipitation forecasts from models. In other words, our visual assessment is influenced by what we are accustomed to seeing. Examination of two representative cases and selected time periods will suffice for our purposes.

3.1.1 31 May 2005

The first case considered is the forecast initialized at 0000 UTC 31 May. Convective activity was well underway by the 23 h forecast time in both models and in observations (left-hand side of 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 TX Panhandle into northwestern OK, apparently linked to more scattered multicellular convection in west-central and north-central Oklahoma. 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 convective activity differs considerably 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 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 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 seven hours later, at the 30 h forecast time (right-hand side of 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 Oklahoma and into north-central

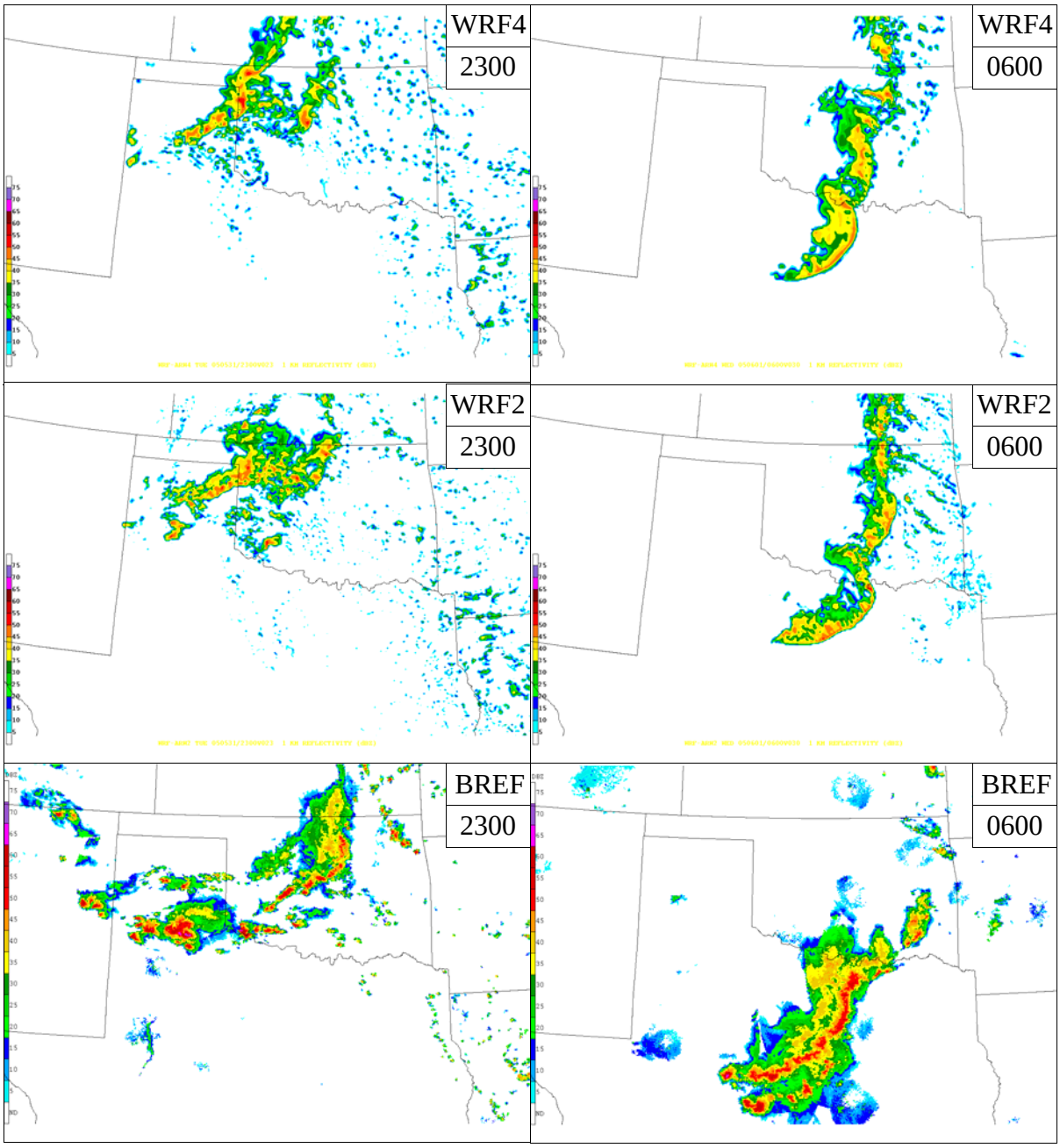

Fig. 4. SRF and corresponding BREF for selected times, associated with the model forecasts initialized 0000 UTC May 31, 2005.

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 focused farther south (bottom-right panel in Fig. 4). There is little convection in Kansas and Oklahoma and a large bowing MCS (mesoscale convective system) propagating into central Texas, with isolated 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 observed activity.

3.1.2 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 models and observations. At 2200 UTC on 2 June, deep convection was developing rapidly in northeastern Colorado. Both models predicted a dominant isolated cell at this time and these forecasts corresponded well with observations (left-side panels in Fig. 5). Furthermore, model diagnostics revealed that both simulated storms exhibited characteristics commonly associated with supercells, including a clearly discernible inflow notch and strong low-to-mid level rotation (not shown), 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 quite remarkable initially, but their correspondence with reality 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 reality. 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 meso-β scale structures were very similar. Above all, the model forecasts looked more like each other than they did like observations.

3.1.3 Mean subjective ratings

Similarity and differences between the forecasts can also be gaged 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. The mean subjective ratings are shown in Fig. 6. Although there are slight differences in the mean values, these differences are not statisti-

cally significant.

3.2 Model Climatology

The mean characteristics of model output reveal important differences and similarities in the model behavior, as well as a comparison with observations. 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.

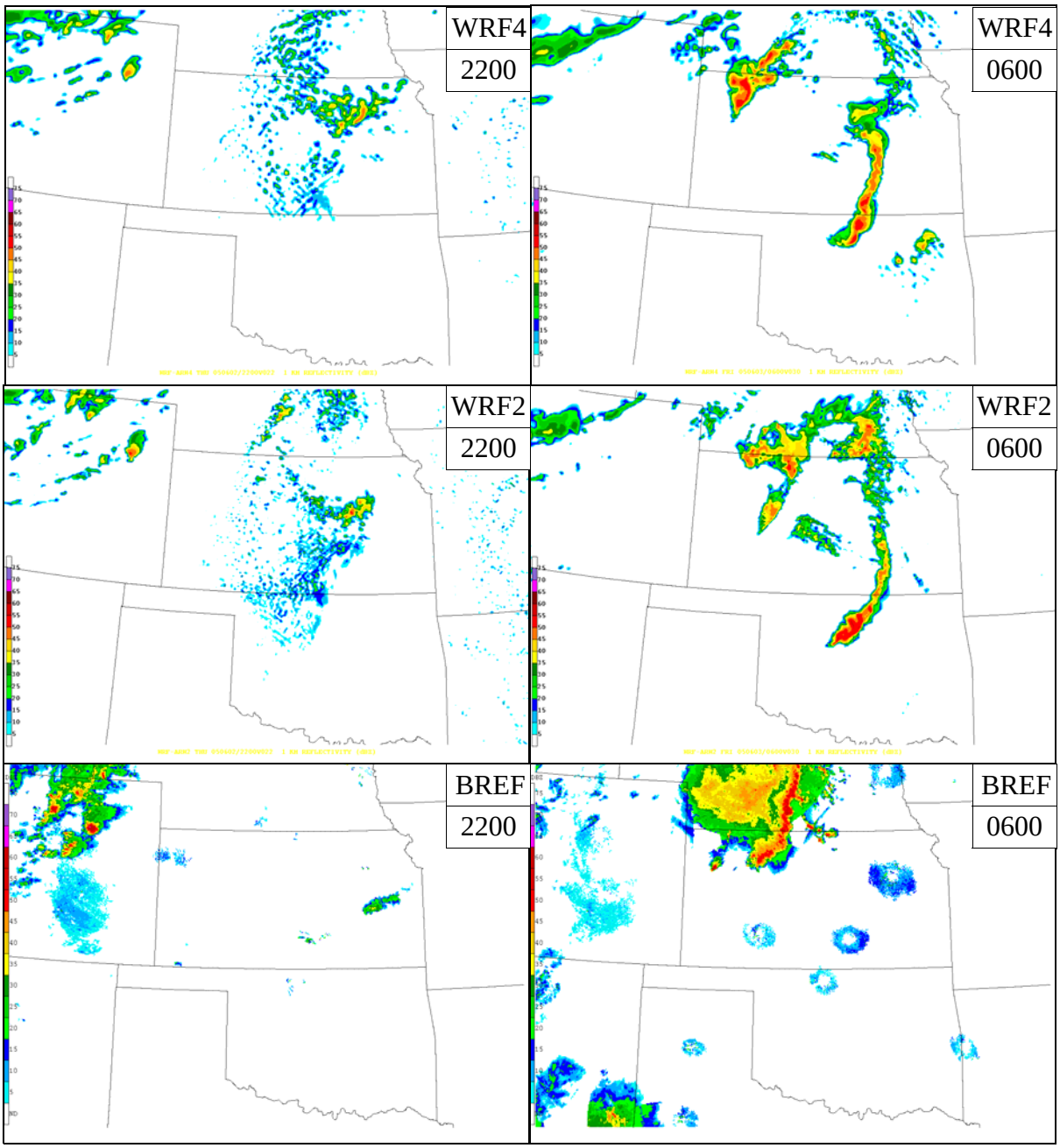

Fig. 5. SRF and corresponding BREF for selected times, associated with the model forecasts initialized 0000 UTC May 31, 2005.

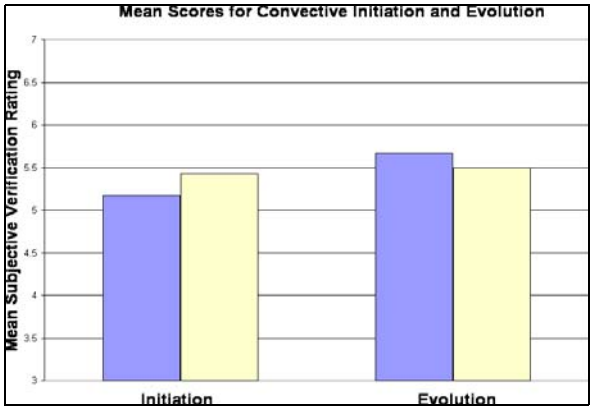

Fig. 6. Mean subjective ratings for convective initiation and evolution forecasts from all three high resolution models. Note that none of the differences are significant at the 95% level.

3.2.1 Reflectivity

The SRF fields from convection-allowing models have proven to be quite appealing to forecasters, likely because of their resemblance to widely utilized observed reflectivity fields and because can be useful for identifying mesoscale structures and processes in the model atmosphere.

3.2.1.1 bulk coverage characteristics

3.2.1.1.1 equitable threat scores

ET scores were used to provide a measure of the degree of over lap between the SRF fields and BREF. For comparison purposes, ET scores 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. Computation of these scores required interpolation of all reflectivity fields to a common grid. This grid 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 includes all days on which both model runs were available.

When ET scores were used to compare the SRFs to observed base reflectivity, the 2 and the 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 towards 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 ground truth, the maximum value was almost 0.35 and scores remain much higher at all thresholds than when observed reflectivity was the verifying field. These scores seem to confirm the subjective impressions based on only two events: The 2 and 4 km reflectivity fields were much more similar to each other than to observations.

3.2.1.1.2 coverage bias

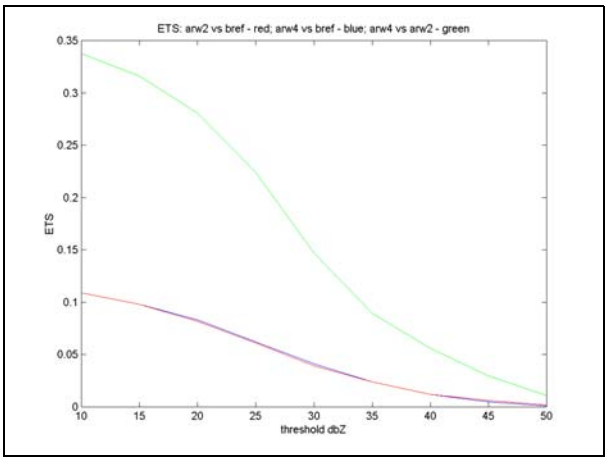

Fig. 7. Equitable threat scores as a function of SRF/BREF threshold. Note that the two lower curves (red and blue) indicate model performance (degree of overlap) relative to observations while the higher curve (green) indicates the degree of overlap between the ARW2 and ARW4.

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 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 "spin-up" 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

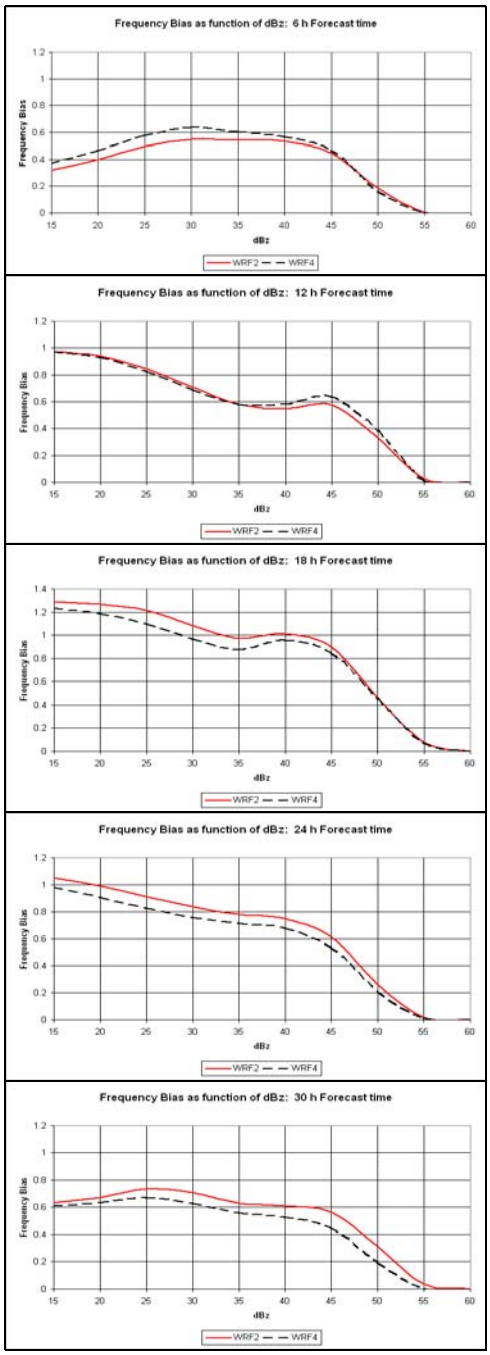

Fig. 8. Model Climatology: Frequency bias of reflectivity (SRF for the models, BREF for observations) as a function of reflectivity threshold, valid at selected forecast times and averaged over all days during the Spring Experiment.

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 observations, especially at the 40 dBz threshold. They generate maximum coverage around 0000 UTC (24 h forecast). In contrast, observed reflectivity areas continue to increase in size well beyond 0000 UTC, with a clear nocturnal maximum.

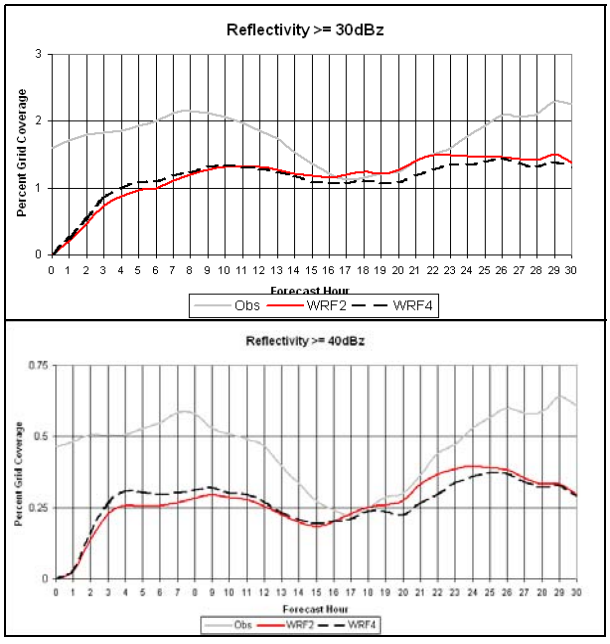

Fig. 9. Model Climatology: Areal coverage of reflectivity (SRF for the models, BREF for observations) as a function of time, averaged over all days during the Spring Experiment.

3.2.1.2 Coverage of individual entities

Although the aggregate areal coverage of SRF features is similar for the WRF2 and WRF4 forecasts, the corresponding numbers of individual features are quite different (Fig. 10). After the initial 3-4 h spin-up and through the initial overnight period, the WRF2 forecasts have about twice has many distinct SRF entities as those from the WRF4. This ratio increases to about 3:1 during the daytime heating cycle, when peak values are reached, then it retreats back to about 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

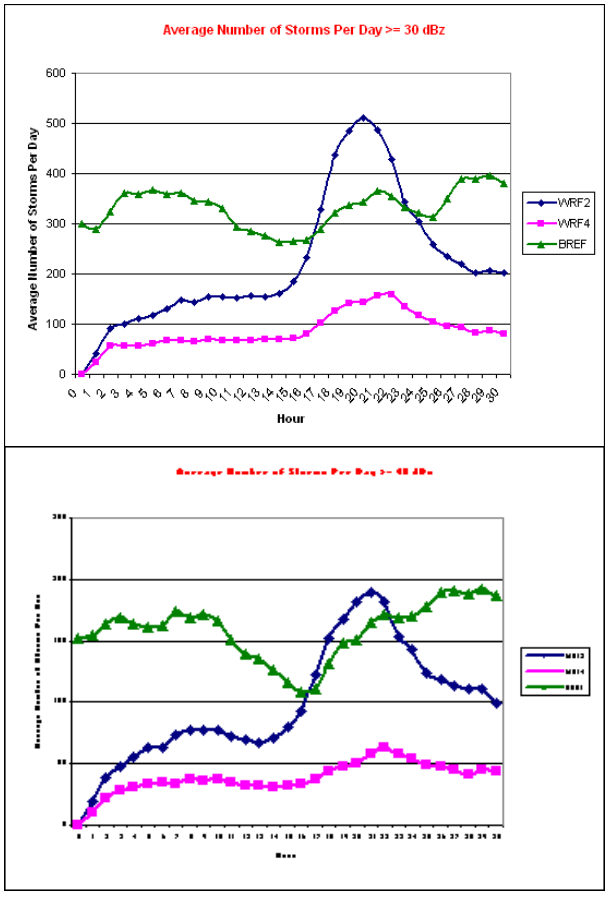

Fig. 10. Model Climatology: Average number of individual reflectivity entities (SRF for the models, BREF for observations) as a function of time for 30 dbz (top) and 40 dbz (bottom) reflectivity thresholds, based on all days during the Spring Experiment.

trends lag the WRF2 trends by about an hour.

The diurnal cycle of 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 cycle is dominated by a single peak in late afternoon.

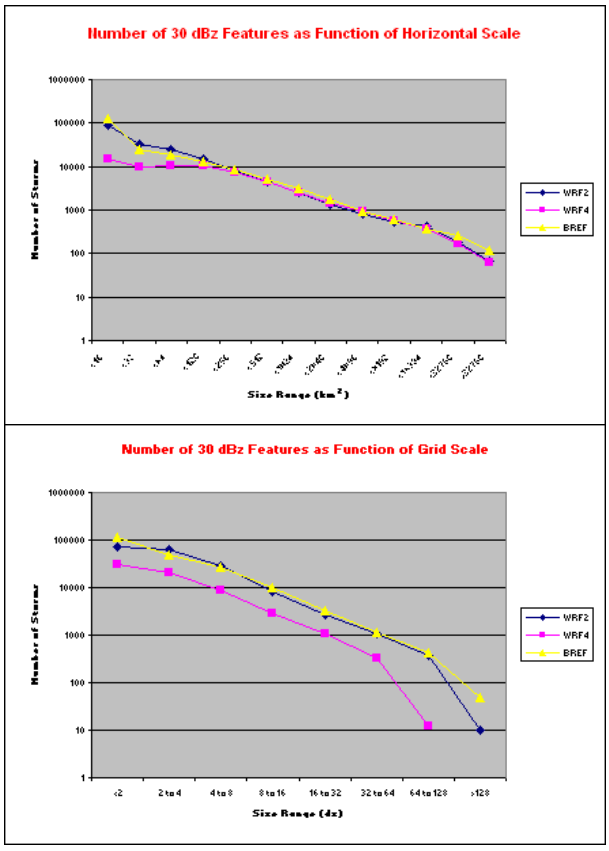

Fig. 11. Model Climatology: Number of individual 30 dbz reflectivity entities (SRF for the models, BREF for observations) as a function of size. The top figure shows the distribution as a function of the absolute areal coverage of individual entities (km2) while the bottom figure show the distribution as a function of model (and observations) grid spacing. The data includes the 12-30h forecast period and all days of the Spring Experiment.

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 12h are excluded). The WRF2 generates many more small-scale features than the WRF4, but the numbers converge at a size of about 200 km^2, 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 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 dynamical equations represent a spectrum of convective overturning processes that is consistent at these two levels of spatial resolution.

It is interesting to see that the distribution of BREF entities is quite similar to the WRF2. It is likely that spectral decomposition and anti-aliasing algorithms that are applied to the observational data are similar in effect to the dynamical equations in the model.

3.2.2 Precipitation

Diurnal trends in areal coverage of the WRF2 and WRF4 precipitation fields 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 b). 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 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 mid afternoon (16-20 h forecast time), and they trend towards 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 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 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. 2007). It may be related to the model, the fact that convection is poorly resolved, and/or to other factors. But it is somewhat surprising that the opposite bias is found in the reflectivity field. It is worth reiterating (as first emphasized in section 2c) that quantitative comparisons of simulated and observed reflectivity fields should be made with caution. The SRF field is quite useful for

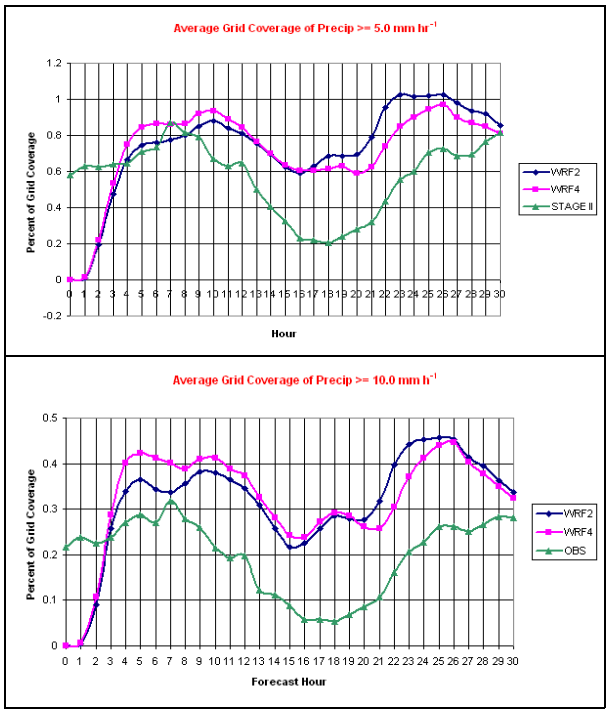

Fig. 12. Model Climatology: Areal coverage of hourly precipitation rate exceeding the 5 (top) and 10 (bottom) mm h^{-1} threshold as a function of time, averaged over all days during the Spring Experiment.

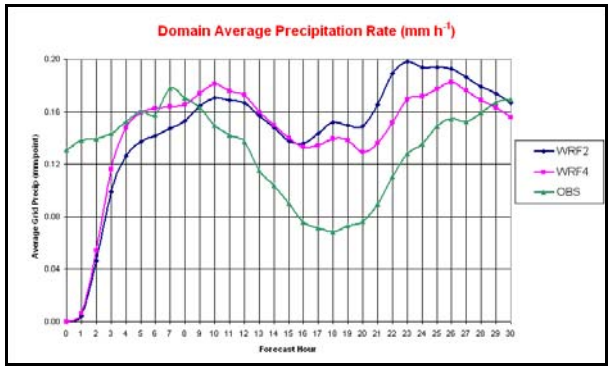

Fig. 13. Model Climatology: Domain average precipitation rates as a function of time, averaged over all days during the Spring Experiment.

direct subjective comparisons with BREF, providing clues about important mesoscale circulations, etc., but these two fields are inherently different in many ways.

3.2.3 Supercells/mesocyclones

As a unique class of thunderstorms, supercells are responsible for a disproportionate share of observed severe weather. Consequently, forecasters at the SPC would be very interested in any guidance product that could provide skillful predictions of these long-lived rotating thunderstorms, also known as mesocyclones. Currently, they use a variety of diagnostic tools to assess the likelihood of these phenomena, based on environmental parameters predicted by NWP models. In particular, they use these tools to highlight areas in which various combinations of shear and instability would favor the supercell mode, contingent upon the development of storms. But current operational models do not predict the storms themselves. In contrast, the WRF4 and WRF2 configurations used in this study produce explicit predictions of mesocyclones.

During the 2005 Spring Experiment, two algorithms were used to detect mesocyclones in the model output. Specifically, both of these algorithms were designed to measure co-located, vertically coherent patterns of rotating up- and downdrafts in the lower to middle troposphere. Since these algorithms provided similar results, the focus here is limited to the Updraft Helicity (UH) algorithm, described in section 2.

The UH output from the WRF2 was subjectively calibrated by selecting a threshold value that produced a reasonable number of mesocyclone "alerts" in the model output. In particular, a value was selected so that the number of rotating storms identified in the hourly snapshots from the WRF2 was comparable to the number of mesocyclones identified at the top of the hour by the NWS's Mesoscyclone Detection Algorithm (MDA - see Stumpf et al. 1998). This value was 50 m^2s^{-2}. Although the MDAs 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.

During the Experiment, the actual number of grid points exceeding the 50 m^2s^{-2} threshold value was typically so small that individual entities were barely discernible on displays of the regional domains. Consequently, simulated storms that exceeded the threshold value were flagged with 50-mile wide open circles. For example, Fig. 14 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 MDAs issued during the preceding hour are indicated by filled red 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 mesocyclones were possible in this environment. Thus, it is instructive to examine the climatology of model-predicted supercells, and

in 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 values >= 50 m^2s^{-2} peaks at about 0000 UTC in both sets of forecasts (Fig. 15). 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 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 the same as the difference in amplitude between the two curves in Fig. 15. 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 areal coverage of mesocyclones using 2 km and 4 km grids.

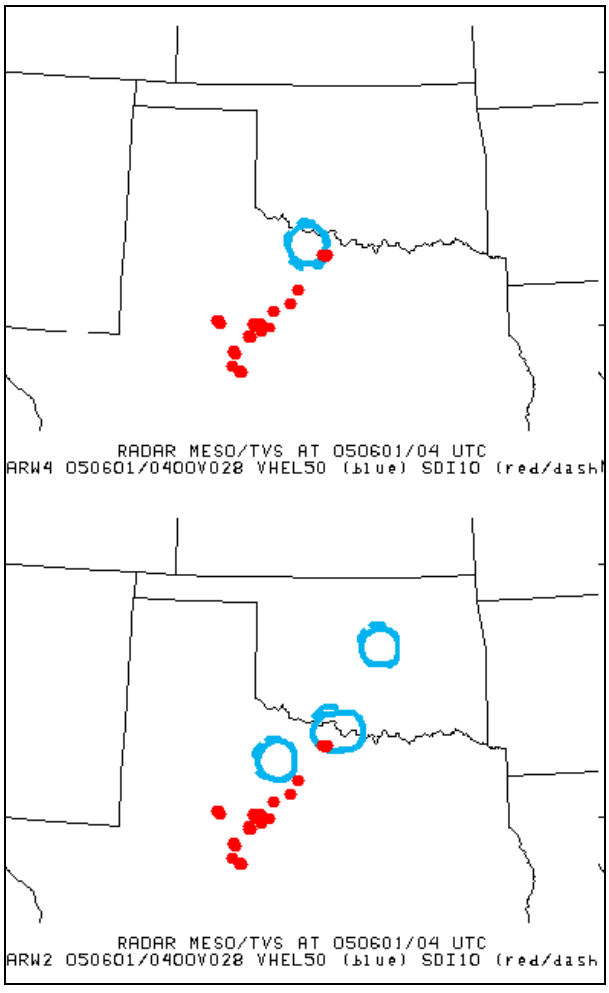

Fig. 14. UH entities from the WRF4 (top) and WRF2 (bottom) forecasts indicated by large open circles, with NWS MDA from the preceeding hour indicated by filled smaller circles; 28 h forecasts valid 0400 UTC 1 June 2005.

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 climatology 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. 16, 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

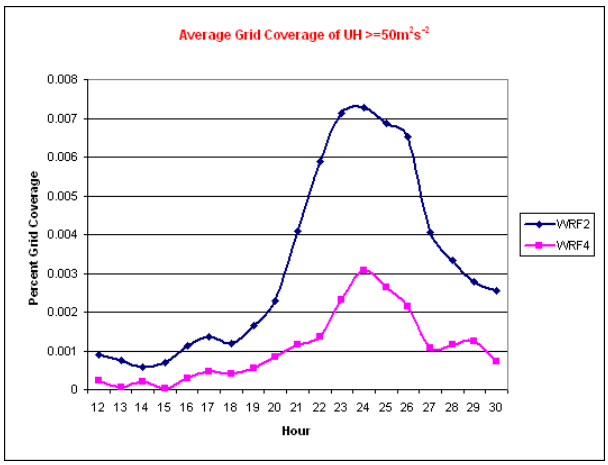

Fig. 15. Model Climatology: Areal coverage of UH>= 50 m^2s^{-2} as a function of time, averaged over all days during the Spring Experiment.

WRF4 before the search for contiguous entities, climatologies become more similar (Fig. 16, bottom). Furthermore, recognizing that the stronger UH entities are very small in scale and that the 2 km configuration inherently produces more small-scale features (e.g., see Figs. 10, 11), one can deduce that a second scaling factor is needed to account for the inherent differences in effective resolution at 2 and 4 km. Although the specific magnitude of this second factor is not immediately obvious, it appears that application of appropriate factors might allow the 4 km configuration to provide comparable information regarding the numbers of mesocyclones as well as their areal coverage.

It is important to emphasize that this assessment is based on average behaviors and model climatology, rather than strict deterministic prediction of supercells. The typical rotating thunderstorms that the WRF2 produced during the Spring Experiment generally had a more realistic appearance in the simulated reflectivity field than their WRF4 counterparts. Furthermore, it is likely that the 2 km storms contained more realistic representations of supercell dynamics. But neither configuration engendered much skill in predicting convective scale features such as supercells, at least with the initialization procedures used during 2005, so the practical implications of resolving supercell dynamics may not be important at this stage. Perhaps when data assimilation strategies and other improvements to the models allow for more accurate predictions of convective-scale forcing and feedbacks, horizontal resolution in the 2-4 km range may become a more important factor in predicting mesocyclones.

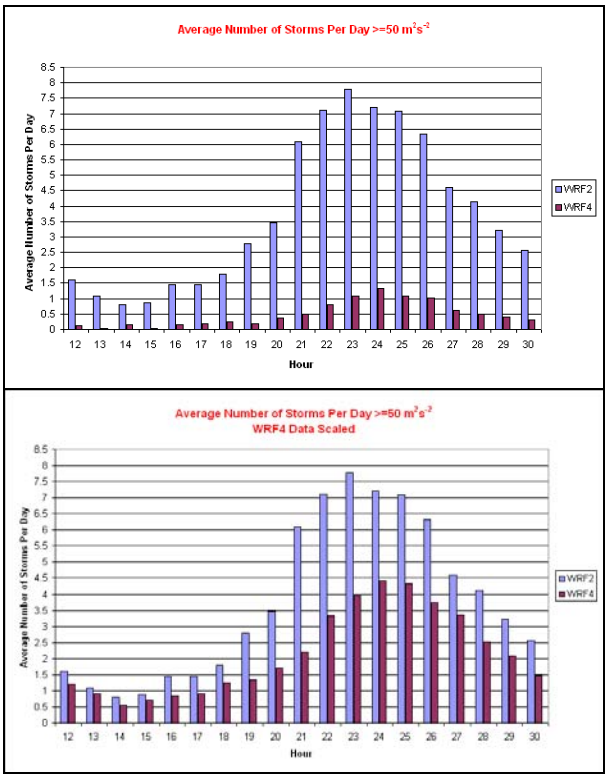

Fig. 16. Model Climatology: Average number of individual UH entities exceeding the 50 m^2s^{-2} threshold as a function of time, based on all days during the Spring Experiment. In the bottom figure, the UH calculation includes a scaling factor of about ~2.6 based on individual scaling factors of about 2.0 for the vorticity and 1.3 for the w factors in the calculation.

4. SUMMARY

The horizontal resolution of operational NWP models in the U. S. increased rapidly during the 1990s but leveled off at about 10 km grid spacing around the turn of the century. Recent studies suggest that we may be ready for another phase of resolution increases, beginning with a jump to about 4 km grid spacing, where convective parameterization can be eliminated. But there remains some uncertainty as to whether 4 km spacing provides fine enough resolution. 4 km may be adequate, but should we wait until we can achieve, say, 2 km spacing before abandoning convective parameterization and relying on explicit prediction of convective clouds? This paper is intended to provide guidance on this question based a series of parallel runs of the

WRF-ARW model at 2 km and 4 km grid spacing from the multi-week 2005 NOAA HWT Spring Experiment.

Subjective assessments during the Spring Experiment, in tandem with more objective analyses afterwards, suggest that mesoscale structures and the overall evolution of convective systems were quite similar in the WRF2 and WRF4 forecasts. The WRF2 forecasts typically showed finer scale structure that seemed more consistent in a general sense with the level of detail present in observed reflectivity images (which was also mapped with 2 km grid spacing), but the models appeared to have little skill in predicting these finer scale details, except perhaps in a climatological sense. The broad conclusion from the subjective evaluation was that the WRF2 and WRF4 output fields showed comparable levels of skill in predicting overall convective initiation, evolution, and morphology.

Acknowledgements

Special thanks and appreciation is extended to many people for their creative insights and assistance in 2005 Spring Experiment preparations/planning, programming, and data flow issues. Without the combined efforts of many SPC and NSSL staff, the Spring Program 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 especially acknowledge the outstanding efforts and expertise of Kelvin Droegemeier, Dan Weber, Kevin Thomas, Yunheng Wang and Keith Brewster for developing and contributing the CAPS 2 km ARW model, the Pittsburgh Supercomputing Center and David O'Neal for providing technical support and computer facilities to run the 2 km ARW; Morris Weisman and Wei Wang for providing the NCAR 4 km WRF runs and offering technical support and advice on the ARW system; Zavisa Janjic, Tom Black, Matt Pyle, and Geoff DiMego for developing and contributing the EMC 4.5 km WRF runs.

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

We thanks Lou Wicker and Bob Davies-Jones for deriving 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 parame-

ters. *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.

Baldwin, M. E., and M. S. Wandishin, 2002: Determining the resolved spatial scales of Eta model precipitation forecasts. *Preprints, 15th Conference on Numerical Weather Prediction*, San Antonio, TX, Amer. Meteor. Soc., 85-88.

Baldwin, M. E., and J. S. Kain, 2006: Sensitivity of several performance measures to displacement error, bias, and event frequency. *Wea. Forecasting*, **21**, 636-648.

Bélair, 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 Conference on Numerical Weather Prediction*, Amer. Meteor. Soc., Seattle, WA. CD-ROM, paper 22.3

Davis, C., and Coauthors, 2004: The bow echo and MCV experiment (BAMEX): Observations and opportunities. *Bull. Amer. Meteor. Soc.*, **85**, 1075–1093.

Done, J., C. Davis, and C. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the Weather Research and Forecasting (WRF) model. *Atmos. Sci. Let.*, published online 8 September 2004, p. 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.

Droegemeier, K. K., 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 Conference on Numerical Weather Prediction*, American Meteorological Society, Portland, 376-379.

Droegemeier, K. K., G. Bassett, D. K. Lilly, and M. Xue, 1996: Does helicity really play a role in supercell longevity? *Preprints, 18th Conf. on Severe Local Storms*, 15-20 Jan., Amer. Meteor. Soc., San Francisco, CA, 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, 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, 14873-14890.

Janjic, Z.I., 1994: The step-mountain eta coor-

dinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927-945.

Janjic, Z. I., T. L. Black, M. Pyle, H.-Y. Chuang, E. Rogers and G. DiMego, 2005: An Evolutionary Approach to Nonhydrostatic Modeling. Available at URL: http://www.wrf-model.org/wrfadmin/publications/Chuang_Janjic_NWP50yearsfinalshort.pdf

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., P. R. Janish, S. J. Weiss, M. E. Baldwin, R. S. Schneider, and H. E. Brooks, 2003a: Collaboration between forecasters and research scientists at the NSSL and SPC: The Spring Program. *Bull. Amer. Meteor. Soc.*, **84**, 1797-1806.

Kain, J. S., M. E. Baldwin, and S. J. Weiss, P. R. Janish, M. P. Kay, and G. Carbin, 2003b: Subjective verification of numerical models as a component of a broader interaction between research and operations. *Wea. Forecasting*, **18**, 847-860.

Kain, J. S., 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 Conference on Weather Analysis and Forecasting/17th Conference on Numerical Weather Prediction, Washington, D. C., Amer. Meteor. Soc., CD-ROM, 2A.5.

Kain, J. S., S. J. Weiss, 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, 11th Conference on Mesoscale Processes*, Albuquerque, NM, Amer. Meteor. Soc., CD-ROM, J4J.7

Lilly, D.K., 1960: On the theory of disturbances in a conditionally unstable atmosphere. *Mon. Wea. Rev.*, **88**, 1-17.

Lilly, D. K., G. M. Bassett, K. K. Droegemeier, and P. Bartello, 1998: Stratified turbulence in the atmospheric mesoscales. *Theoretical and Comp. Fluid Dyn*, **11**, 139-153.

Liu, C., M. W. Moncrieff, 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**, 16663-16682.

Molinari, J.M., and M. Dudek, 1986: Implicit versus explicit convective heating in numerical weather prediction models. *Mon. Wea. Rev.*, **120**, 326-344.

Molinari, J.M., and M. Dudek, 1992: Parameterization of convective precipitation in mesoscale numerical models: A critical review. *Mon. Wea. Rev.*, **120**, 326-344.

Moncrieff, M.W. and E. Klinker, 1997: Mesoscale cloud systems in the tropical Western Pacific as a process in general cir-

culation models: A TOGA-COARE case study. *Quart. J. Roy. Met. Soc.*, **123**, 805-823.

Moncrieff, M. W., and C. Liu, 2006: Convective organization in prediction models at ~10 km grid spacing. Submitted to *J. Atmos. Sci.*

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., 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.

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., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, 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].

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.

Weisman, M. L., W. C. Skamarock, and J. B. Klemp, 1997: The resolution dependence of explicitly modeled convective systems. *Mon. Wea. Rev.*, **125**, 527-548.

Weisman, M. L., C. Davis, W. Wang, K. W. Manning, and J. B. Klemp, 2007: Experiences with 0-36 h Explicit convective forecasts with the WRF-ARW model. Conditionally accepted for publication in *Wea. Forecasting*.

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 Conference on Severe Local Storms*, Hyannis, MA, Amer. Meteor. Soc., CD-ROM, paper 17.1

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-β-scale numerical model. *Quart. J. Roy. Meteor. Soc.*, **114**, 31-60.