Assessing Advances in the Assimilation of Radar Data and Other Mesoscale Observations within a Collaborative Forecasting–Research Environment

JOHN S. KAIN,* MING XUE,+ MICHAEL C. CONIGLIO,* STEVEN J. WEISS,# FANYOU KONG,+ TARA L. JENSEN,@ BARBARA G. BROWN,@ JIDONG GAO,+ KEITH BREWSTER,+ KEVIN W. THOMAS,+ YUNHENG WANG,+ CRAIG S. SCHWARTZ,@ AND JASON J. LEVIT#

* *NOAA/National Severe Storms Laboratory, Norman, Oklahoma*
\+ *Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma*
\# *NOAA/NWS/Storm Prediction Center, Norman, Oklahoma*
@ *National Center for Atmospheric Research, Boulder, Colorado*

(Manuscript received 26 January 2010, in final form 6 May 2010)

ABSTRACT

The impacts of assimilating radar data and other mesoscale observations in real-time, convection-allowing model forecasts were evaluated during the spring seasons of 2008 and 2009 as part of the Hazardous Weather Test Bed Spring Experiment activities. In tests of a prototype continental U.S.-scale forecast system, focusing primarily on regions with active deep convection at the initial time, assimilation of these observations had a positive impact. Daily interrogation of output by teams of modelers, forecasters, and verification experts provided additional insights into the value-added characteristics of the unique assimilation forecasts. This evaluation revealed that the positive effects of the assimilation were greatest during the first 3–6 h of each forecast, appeared to be most pronounced with larger convective systems, and may have been related to a phase lag that sometimes developed when the convective-scale information was not assimilated. These preliminary results are currently being evaluated further using advanced objective verification techniques.

1. Introduction

Numerical weather prediction has advanced to higher and higher resolutions since its inception and in recent years numerous operational prediction centers have implemented convection-allowing configurations (i.e., configurations with fine enough resolution to obviate convective parameterization) in their forecast models (e.g., Weiss et al. 2008; Dixon et al. 2009). This latest generation of operational models appears to have numerous advantages over traditional approaches, most notably an improved ability to provide information on convective morphology and evolution (e.g., Done et al. 2004; Kain et al. 2008; Weisman et al. 2008). Yet, the resolution of convection-allowing models (CAMs) is considerably higher than that provided by traditional observational data sources, so current CAM-based forecast systems typically use relatively coarse-resolution initial and lateral boundary conditions (ICs/LBCs) and rely on the ''spinup'' process to develop smaller-scale circulations, such as those associated with convective overturning (Skamarock 2004). The spinup process typically requires 3–6 h of integration time and it inherently results in errors in forecasts of small-scale phenomena such as thunderstorms.

Logical reasoning suggests that these errors could be ameliorated significantly with robust initialization procedures that include observations on scales that are commensurate with model resolution and the size of the weather phenomena we are interested in predicting. For example, when thunderstorms are present at the initial time, numerical forecasts should benefit greatly if the storm-scale circulations, hydrometeors, and other characteristics are represented well in the initial conditions. This information is not available from conventional observations, but it is retrievable from various remote sensing platforms. Among these platforms, the Weather Surveillance Radar-1988 Doppler (WSR-88D) network holds considerable promise for initializing CAMs with storm-scale information. This network provides extensive

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

DOI: 10.1175/2010WAF2222405.1

coverage of reflectivity and radial velocity over the contiguous 48 states in the United States (CONUS), from which storm-scale hydrometeor, kinematic, and thermodynamic properties can be estimated.

The use of Doppler radar data for initializing storm-scale numerical weather prediction models has been envisioned since the early 1990s (Lilly 1990). Scientists at the University of Oklahoma Center for Analysis and Prediction of Storms (CAPS) have been pioneers in making this vision a reality (e.g., Xue et al. 2002; Xue et al. 2003; Hu et al. 2006a,b; Dawson and Xue 2006; Sheng et al. 2006; Hu and Xue 2007a,b). Their early efforts involved single radars or regional data and relatively small forecast domains, and most of their experiments were conducted without the time constraints imposed by scheduled daily forecasts. But in the spring of 2008 they implemented a near-CONUS-scale assimilation of Doppler velocity and reflectivity data, and they transitioned to a real-time, semioperational framework in order to provide model guidance for daily experimental forecasting exercises in the National Oceanic and Atmospheric Administration (NOAA) Hazardous Weather Test bed (NOAA HWT) Spring Experiment (hereafter SE2008). Specifically, every evening they used radar data from the national network of WSR-88Ds to supplement conventional observational datasets. They analyzed the data on model-native 4- and 2-km grid intervals using CAPS's three-dimensional variational data assimilation (3DVAR; Xue et al. 2003; Gao et al. 2004) and cloud analysis scheme (Xue et al. 2008; Kong et al. 2008) that is part of the Advanced Regional Prediction System (ARPS) modeling system (Xue et al. 2003). They used this process to initialize 9 members of a 10-member, 4-km grid spacing, forecast ensemble at 0000 UTC. The last ensemble member was configured identically to one of the other nine, but was initialized without this special assimilation process. Comparison of forecasts from these two ''control'' members forms the basis for this study.

This assimilation experiment was remarkable because of its large scale and the fact that it adhered to quasi-operational time constraints—the high-resolution forecasts were ready to be used as forecast guidance by 1200 UTC—but it was also unique because the output was scrutinized by teams of expert researchers and forecasters every day during SE2008. Simulated reflectivity fields from the parallel assimilation–no-assimilation forecasts were compared, and both were critically judged based on differences with observed radar reflectivity fields. Subjective impressions from these comparisons were documented as part of the SE2008 record.

After SE2008, traditional verification metrics, such as bias and Gilbert skill score (GSS, also known as the equitable threat score), were applied to the data to supplement the subjective comparisons (e.g., Xue et al. 2008; Kong et al. 2008). The traditional metrics provided useful additional information, but these measures have known deficiencies that are magnified as model resolution increases (e.g., Baldwin and Kain 2006; Casati et al. 2008). Thus, in preparation for the 2009 Spring Experiment (SE2009), during which CAPS ran a similar set of forecasts, scientists from the HWT consulted with colleagues from the Developmental Test Bed Center (DTC), who had participated in SE2008. The DTC is a center of expertise in the testing and evaluation of high-resolution model forecasts. Through these consultations, DTC and HWT scientists identified a compelling strategy for real-time verification efforts in 2009. This strategy engaged HWT scientists and SE2009 participants in a subjective evaluation of the potential value of emerging new verification tools within the context of simulated forecasting operations. The collaboration also brought to bear innovative verification metrics that may be more appropriate for these unique model output datasets.

This collaboration between scientists and forecasters from the HWT, CAPS, and DTC provided a unique framework within which the impacts of the CAPS 3DVAR and cloud-analysis methodology were assessed from several different perspectives. The purpose of this article is to highlight the preliminary findings of this assessment and the value of this collaborative arrangement. In the next section, technical details about model configurations and initialization procedures will be outlined and the framework for verification described. This will be followed by a section on results and a final section containing a summary and discussion.

2. Methodology

Data generation, dissemination, and verification occurred in conjunction with SE2008 and SE2009, from mid-April through early June each year. A full suite of experimental activities were conducted daily: Monday–Friday in 2008 and Monday–Thursday in 2009.

a. Model configuration and initialization procedures

CAPS ran an ensemble prediction system with 4-km grid spacing and provided the output to the Spring Experiment in both 2008 and 2009 (Xue et al. 2008, 2009). Each year, two members of the ensemble were configured identically but initialized differently. Specifically, the initialization of one of these members included the assimilation of radar and other observational data, while the other one did not. This study focuses exclusively on output from these two members. The reader is referred to Xue et al. (2008, 2009) for details regarding the full ensemble configuration and performance.

TABLE 1. The Weather Research and Forecasting Model (WRF) configuration used by CAPS during SE2008 and SE2009. ARW denotes the Advanced Research WRF dynamic core (Skamarock et al. 2005); the physical parameterizations include the Thompson microphysics (Thompson et al. 2008), the Mellor–Yamada–Janjić (MYJ) boundary layer/turbulence (Janjić 1994), Goddard shortwave radiation (Tao et al. 2003), and the Rapid Radiative Transfer Model (RRTM; Mlawer et al. 1997; Iacono et al. 2000). Version 2.2 (3.0.1.1) of the WRF-ARW was used in 2008 (2009).

Description	Configuration
Dynamic core	WRF-ARW
Horizontal grid spacing (km)	4
Vertical levels	51
Microphysical parameterization	Thompson
PBL/turbulence parameterization	MYJ
Radiation (SW–LW) parameterization	Goddard/RRTM
Background ICs/LBCs	12-km NAM

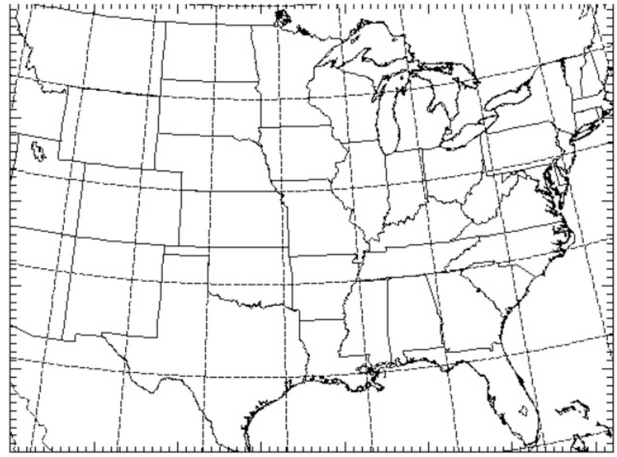

FIG. 1. Model domain for the 4-km 2008 and 2009 CAPS forecasts.

The configurations of these two members are highlighted in Table 1 and the forecast domain is shown in Fig. 1. For both years, background fields were generated by interpolating the National Centers for Environmental Prediction (NCEP) North American Mesoscale (NAM; Rogers et al. 2009) model 0000 UTC analysis (native 12-km grid) to the 4-km high-resolution grid. One of these members (hereafter C0) used the background fields directly as the initial conditions while the other member (hereafter CN) incorporated additional observational datasets in a storm-scale analysis, including assimilation of radar reflectivity and velocity data in the initial conditions. Specifically, the unique assimilation process in the CN run ingested data from the national network of WSR-88D, typically using the level II dataset, but occasionally using the level III data [see Hu et al. (2006a) for a discussion of these two datasets] when level II was not available. Information from numerous other data sources was also ingested, including conventional rawinsonde, wind profiler, aviation routine weather report (METAR) surface observations, and Oklahoma Mesonet observations, as well as visible and infrared channel-4 data from Geostationary Operational Environmental Satellites (GOES). Although these other observations undoubtedly play an important role, radar data are likely to have a dominant impact in the vicinity of ongoing deep convection (e.g., Xue et al. 2003; Hu et al. 2006a; Hu and Xue 2007a), where this study is consistently focused. Additional details about the complex assimilation process can be found in Xue et al. (2008, 2009, and references therein).

In spite of its complexity, it is important to note that the assimilation process used in these CAPS real-time experiments was considerably simpler than the procedures that CAPS scientists routinely use in ongoing research. In particular, the SE2008 and SE2009 routines used a relatively efficient 3DVAR analysis strategy, rather than more computationally intensive four-dimensional version (4DVAR; e.g., Sun and Crook 1998) or ensemble-based (e.g., Tong and Xue 2005; Lei et al. 2009; Aksoy et al. 2009) approaches to assimilate cloud-scale data. Furthermore, a single-time assimilation step was used, allowing only one volume scan of data from each radar. This was computationally efficient but likely provided minimal opportunity to achieve balance and dynamical consistency among state variables compared to a procedure that uses multiple assimilation cycles and volume scans (e.g., see Hu and Xue 2007a). Although adoption of this streamlined approach was necessary to satisfy the real-time constraints of SE2008 and SE2009, its implementation on a CONUS scale still represents a remarkable computational achievement and establishes the framework for more sophisticated assimilation procedures that will be implemented in future operational settings as computational resources permit.

Both the CN and C0 members were integrated out to 30 h, but the focus here will be primarily on the first 6–12 h. This is the period when the assimilated radar data are expected to have the most impact (Zhang et al. 2007).

b. Verification datasets

Simulated reflectivity (SR) from CAMs has proven to be a very useful diagnostic output field because it provides important clues about a variety of circulations and processes in a model forecast (e.g., Xue et al. 2003; Koch et al. 2005). Thus, the subjective evaluations were based primarily on the comparison of SR with the observed reflectivity from WSR-88D radars in both 2008 and 2009. Objective verification in 2008 used accumulated precipitation fields, with stage II hourly precipitation analyses (Lin and Mitchell 2005) as the verifying dataset. In 2009, both accumulated precipitation fields and SR were evaluated but the focus of this discussion will be on SR. The remainder of this section describes the fields in detail.

The SR was computed from the three-dimensional hydrometeor field as described in Kain et al. (2008), with all relevant parameters (such as those describing particle size distributions) set to the values used by the Thompson et al. (2008) microphysical parameterization that was used during model integration. For this study, the composite SR was used, meaning that gridded values represent the largest computed simulated reflectivity at any level in each vertical column.

In 2008, the observed reflectivity (OR) data were derived from national composite-reflectivity mosaics that are part of the Storm Prediction Center (SPC) operational data stream. These mosaics are generated by Unisys Corporation at a frequency of 5 min or less on a 2-km grid. In 2009, the OR data came from the National Severe Storms Laboratory (NSSL) national 1-km radar mosaic (Vasiloff et al. 2007).

c. *HWT Spring Experiment activities*

During both SE2008 and SE2009, hourly SR output from the C0 and CN forecasts was visually compared to corresponding OR fields. Whenever possible, this comparison was focused on a region of active thunderstorms at the model initialization time by zooming in on a regional domain so that features ranging from individual convective cores to mesoscale convective systems (MCSs) could be examined within a single image. The comparison was a group activity, conducted by teams of expert forecasters and researchers. The teams were asked to focus on 1) the continuity and evolution of reflectivity features assimilated into the CN forecast (i.e., did the model forecast appear to ‘‘hit the ground running’’ with a smooth and realistic-looking evolution of precipitation features on all scales, from those of individual convective cells to MCSs?) and 2) how long it appeared to take for coherent precipitation structures in the C0 run to ‘‘spin up’’ (i.e., develop similar levels of detail and apparent realism in the reflectivity field as compared to the CN forecast). The evaluation was performed by manually stepping through hourly SR and OR fields, displayed side by side on a large elevated plasma screen. The assessments of the evaluation teams were formally documented each day, along with any related discussion.

d. *DTC activities and collaboration*

During SE2008, NSSL and SPC scientists worked with colleagues from the DTC to develop a framework for incorporating real-time objective verification procedures into daily HWT spring experiment activities. These procedures were implemented for the first time during SE2009. As part of this process, model SR and precipitation fields were extracted from output at the HWT and transferred to the DTC, along with verifying reflectivity and precipitation fields from the NSSL national radar reflectivity and quantitative precipitation estimate (QPE) mosaics (Vasiloff et al. 2007). These datasets were ingested at the DTC, and several different types of verification statistics were computed, focusing on a specified (moveable) regional domain where active weather was expected at the model initialization time (0000 UTC) each day. Graphical displays of the statistical results were posted to an internal Web page along with selected output fields, such as SR, in time for subjective assessments and critical examination by forecast teams during the SE2009 daily activities. This group evaluation was led by a DTC scientist each day, as the DTC rotated several scientists through SE2009 on a weekly basis. The group was instructed to focus on assessing 1) the degree to which the objective verification metrics corroborated subjective impressions and 2) the potential utility of the various objective metrics in an operational environment.

Verification procedures at the DTC used the Model Evaluation Tools (MET) software package (Brown et al. 2007). MET includes numerous traditional verification metrics, such as coverage bias, probability of detection, false alarm ratio, and GSS. Furthermore, it contains new verification tools designed to quantify the correspondence between objects, or features, in forecasts and observations, known as the Method of Object-based Diagnostic Evaluation (MODE; Davis et al. 2006; Brown et al. 2007). The MODE software relies on user-specified parameters to identify similar features (such as precipitation elements) in forecasts and observations. In this study, the matching features are overlaid to enhance visual comparisons of forecasts and observations. Meanwhile, work is under way to utilize MODE with the SE datasets to quantify the degree of correspondence between features in terms of attributes such as size, distance, orientation, internal structure, etc., but this effort is still at an early stage of development and will be reported upon in a separate paper.

3. Results

a. *SE2008*

1) SUBJECTIVE ASSESSMENTS

On many days evaluation teams noted a ‘‘lack of continuity’’ during the first 1–3 h of the CN forecasts. Specifically, they noted that the most intense convective cells, the location of which appeared to be initialized very well, were often substantially weakened if not gone

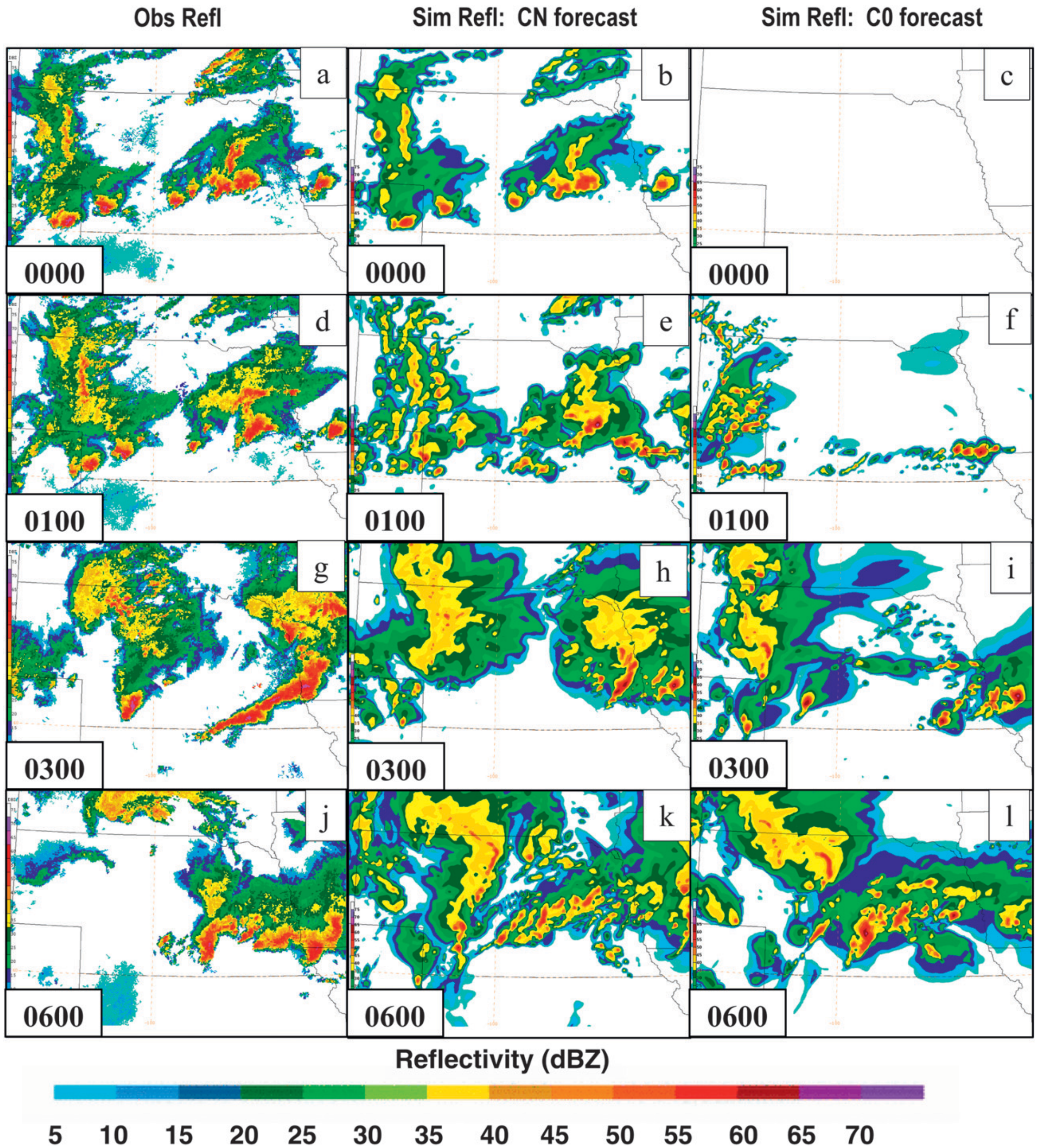

FIG. 2. Comparison of observed and simulated composite reflectivities for forecast hours 0000, 0100, 0300, and 0600 UTC, beginning at 0000 UTC 5 Jun 2008. Shown are (left) the observations, (middle) the CN forecast, and (right) the C0 forecast.

altogether by 1 h into the forecast. They also noted that spurious (no correspondence with observations), transient convection ''bloomed'' nearby on some days. In addition, they indicated that, on most days, the C0 run spun up convection within 3–6 h, noting that beyond this time period SR fields from the CN and C0 runs began to resemble each other more so than either one resembled the corresponding OR field.

These general characteristics are exemplified quite well by the set of forecasts that begins with the 0000 UTC 5 June 2008 initialization. Comparison of Figs. 2a and 2b suggests that the CAPS procedures for assimilating

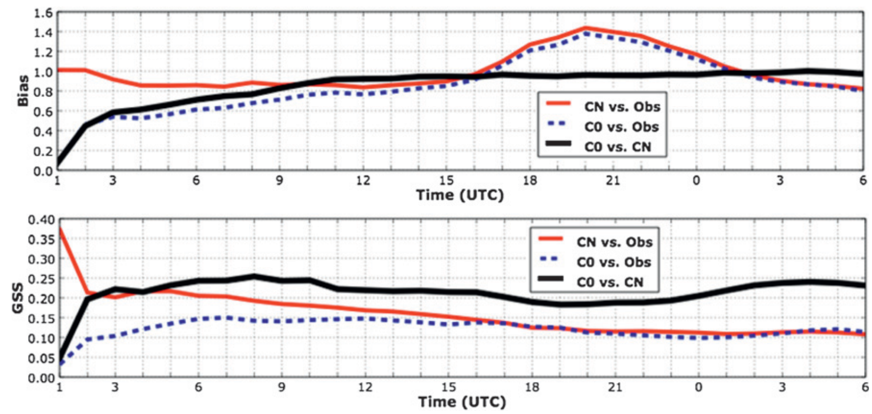

FIG. 3. Precipitation coverage biases and GSSs as a function of time for all forecast hours and all days during SE2008, for the 1 mm h^{-1} precipitation rate threshold. Note that the black curve in each image compares the C0 and CN fields, while the other curves compare the respective model forecasts to stage II observed precipitation data. Models were initialized at 0000 UTC.

hydrometeors from radar data are quite faithful to the observations; the correspondence between the observed and assimilated reflectivity structures is remarkably good. However, within the first hour the model forecast begins to lose correspondence with the intense convective cells, although it retains a good semblance of the general envelope of convective activity (cf. Figs. 2d and 2e). Meanwhile, the C0 forecast develops some isolated storms in southeastern Nebraska during this first hour and begins to form a cohesive convective system in the Nebraska panhandle and northeastern Colorado (Fig. 2f). As the integration time progresses, the two forecasts begin to look more and more alike, and less like the observations (cf. Figs. 2j, 2k, and 2l), but the run with the data assimilation clearly appears to get a ''head start'' in its representation of two coherent MCSs.

2) OBJECTIVE VERIFICATION

During much of the SE2008 period, being the first time to run their system at such a large scale, CAPS scientists were making adjustments to their data assimilation system. In addition, the input radar dataset was incomplete on many days. Nonetheless, and in spite of some inconsistencies in performance from day to day, the objective verification metrics indicate that the net impact of the data assimilation was positive. Aggregate verification of hourly precipitation data from all days suggests that the assimilation procedures led to better precipitation skill scores for the CN forecasts through at least the first 12 h of integration (Fig. 3).

This result, particularly the extended duration of the positive impact, came as somewhat of a surprise to many who participated in the daily SE2008 subjective assessments. Apparently, the real-time focus on the most intense convective cores (which typically dissipated within the first hour of the CN forecast) drew attention away from the assimilated mesoscale features, such as MCSs, which often showed relatively good continuity after initialization. These larger systems can dominate traditional verification metrics by the simple virtue of their size (e.g., Baldwin and Kain 2006), and the objective statistics suggested that the measurable benefits of the CAPS radar assimilation for mesoscale features extended beyond the time when such benefits were obvious to the naked eye. This view is consistent with results from individual cases as discussed by Xue et al. (2008, 2009).

An alternative application of the same objective metrics can be used to substantiate our subjective impression that the two forecasts tend to look more like each other than they do like the observations beyond the first several hours. Specifically, when GSSs are recomputed for the C0 dataset, but this time using the CN data as the verifying ''observation'' instead of the stage II data (black curves in Fig. 3), the scores are higher than those associated with either the CN or C0 data, as verified against the stage II data, beyond the 4-h time period. Furthermore, analogous computation of bias scores yields values approaching 1.0 by 9–12 h with little change for the remainder of the 30-h forecast, so the areal coverage of the precipitation from the CN and C0 forecasts becomes practically the same after the 12 h. These results are likely due to the fact that the two runs should share very similar forecasts of forcing mechanisms in the mesoscale environment. After the predictability of the small convective-scale features initialized by the radar data is lost, the development and evolution of the convective activity is strongly modulated by these forcing mechanisms.

3) Preliminary assessment following SE2008

The impacts of the 3DVAR and cloud analysis procedures appeared to be scale dependent. Visual examination indicated that convective-scale information was lost very quickly after the model integration started and, in fact, SE2008 participants noted that smaller-scale reflectivity structures in the CN forecasts looked ''nonmeteorological'' in the first 6 h of the forecasts on some days, after which more realistic looking features began to spin up in both the run with advanced assimilation and the ''cold start'' run. So the special assimilation appeared to add relatively little in the way of reliable guidance for the occurrence and/or behavior of discrete convective cells during the first 6 h or so of integration.[1] However, it clearly provided added value for regional quantitative precipitation forecasts (QPFs) in the first 3–6 h. Beyond this time, and through about 12 h, GSSs indicated a QPF advantage for the CN runs in spite of the fact that the CN and C0 runs looked much more like each other than the observations during this period.

b. SE2009

1) Subjective assessments

In 2009, the subjective assessments were greatly facilitated by a Web-based display application developed at the DTC. As in SE2008, this display placed hourly time-matched composite reflectivity images (SR and OR) side by side, but in 2009 these images were supplemented with simple graphical displays derived using MODE that contained overlays of corresponding SR and OR objects. For example, Fig. 4 displays images in a format similar to the one used in SE2009. The supplemental graphics are in the bottom two rows.

Examination of these displays yielded impressions that were similar to those of 2008. For example, the initialization from 0000 UTC 14 May 2009 indicated that the CAPS assimilation scheme produced an SR field that matched OR very well (cf. Figs. 4a and 4b), but the smaller-scale features appeared to lose coherence rapidly in the forecast. Two hours into the 14 May forecast, the CN SR pattern seemed quite disorganized compared to the observations (cf. Figs. 4f and 4g), yet the object field, which was based on a reflectivity threshold of 30 dBZ, indicated substantial overlap between the SR and OR fields (Fig. 4i). Meanwhile, deep convection was beginning to spin up in the C0 forecast (Fig. 4h), with the location axis in fairly good agreement with the observations, especially on the southern end of the line (Fig. 4j).

Four hours into the 14 May forecast, the CN forecast still had a diffuse representation of the main convective line compared to the observations, but also continued to have very good spatial overlap with the observed system, at least at the 30-dBZ threshold (Figs. 4k, 4l, and 4n). At this time, it could be argued that SR from the C0 run looked more like the observations than did SR from the CN forecast in terms of convective evolution and morphology (cf. Figs. 4k–m), yet a subtle but important trend is revealed by the overlay in Fig. 4o. Namely, the C0-predicted convective system appears to lag behind the corresponding observed and CN systems.

This relative upstream lag persists for the lifetime of this convective system, as suggested by Figs. 4p, 4r, and 4t. In fact, an upstream lag in C0 systems was noted on a number of days during SE2009. The tendency for this pattern of behavior would certainly help the CN forecasts retain an advantage over the C0 forecasts in terms of traditional verification metrics, especially for narrow features like this convective line. This may explain why aggregate GSSs favor the CN forecasts well beyond the ~6-h integration time when a systematic difference in the placement of major features is difficult to discern from simple side-by-side visual inspections of the SR and OR fields.

2) Objective verification

Object-based verification metrics, such as those available in MODE, are capable of diagnosing systematic biases, such as the upstream lag proposed above. However, the generation of statistically significant inferences from these metrics is quite challenging due to sample size constraints (i.e., the number of identified objects is small over a 4–6-week time period) and is the subject of ongoing work. Nonetheless, application of traditional verification metrics, this time based on reflectivity fields rather than accumulated precipitation, yields some interesting results.

As expected, GSSs start out at a high level in the CN runs, with the magnitude depending on the reflectivity threshold (Fig. 5). But GSSs from the CN and C0 runs converge at about 10 h into the forecast. Although the convergence point may be a few hours earlier than was seen with the previous (i.e., 2008) dataset (which was verified on the basis of accumulated precipitation), it is not clear whether there is any real significance to this

[1] This conclusion may be specifically related to the particular data assimilation strategies and model grid spacings used during SE2008 and SE2009. Due to the constraint of real-time operations, only a single-time 3DVAR analysis was performed, and the forecasts discussed here used relatively coarse 4-km grid spacing. In case studies conducted by CAPS using higher-resolution and cycled 3DVAR (e.g., Hu et al. 2006a) or ensemble Kalman filter (e.g., Lei et al. 2009) techniques, smaller-scale convective structures introduced through the assimilation of WSR-88D data appear to evolve quite realistically through the life cycle of the storms.

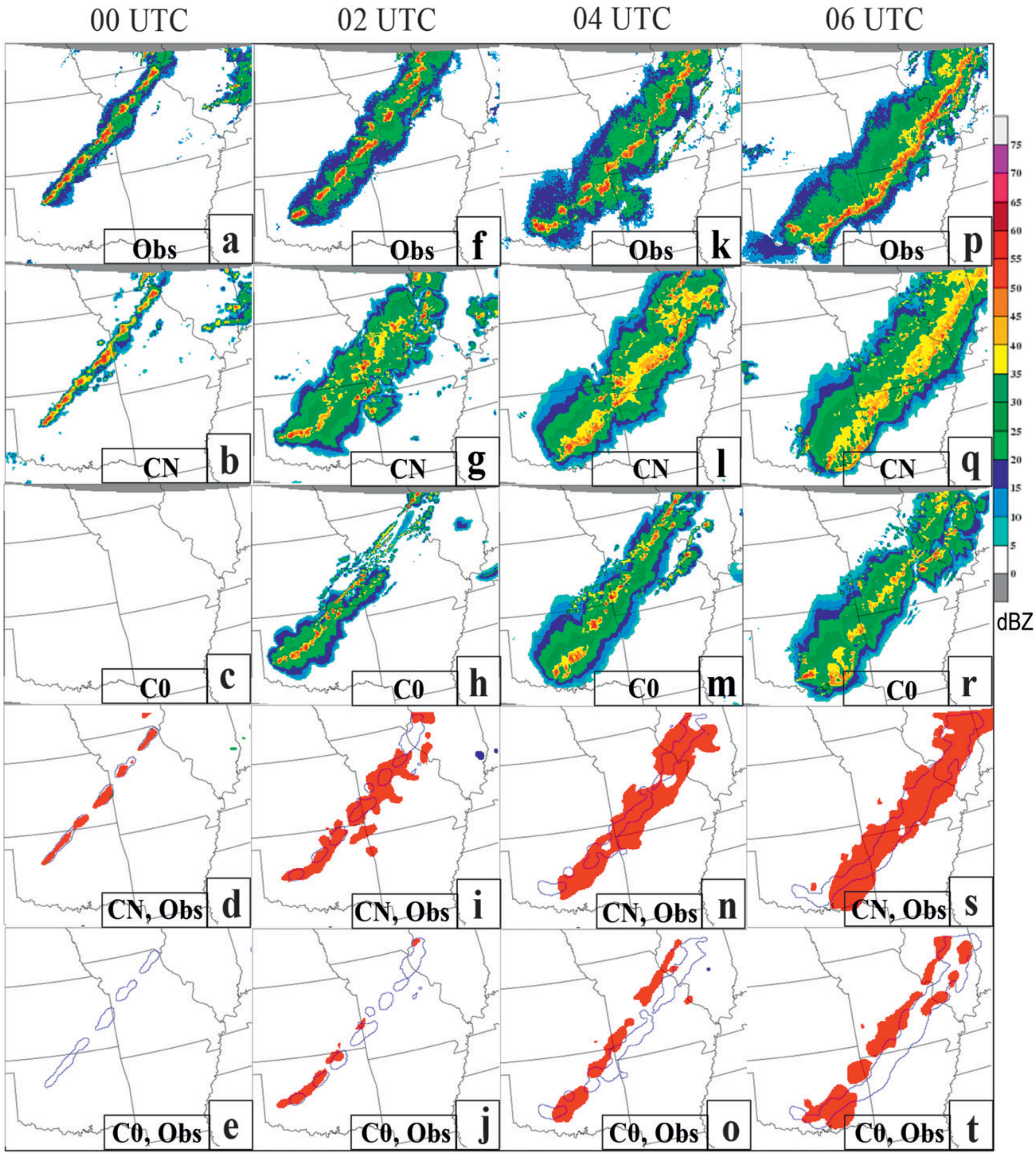

FIG. 4. Comparison of observed and simulated composite reflectivities for forecast hours 0000, 0200, 0400, and 0600 UTC, beginning at 0000 UTC 14 May 2009. From top to bottom the first three rows show the observations and the CN and C0 forecasts, respectively. In the bottom two rows, the locations of observed features are outlined in blue and the features predicted by the CN (fourth row) and C0 (fifth row) are overlaid and color filled. Objects filled with a single color have been grouped (or merged) by the MODE software as a collection of related objects. Each time period is analyzed independently by MODE; the 30-dBZ reflectivity threshold is used to define object outlines.

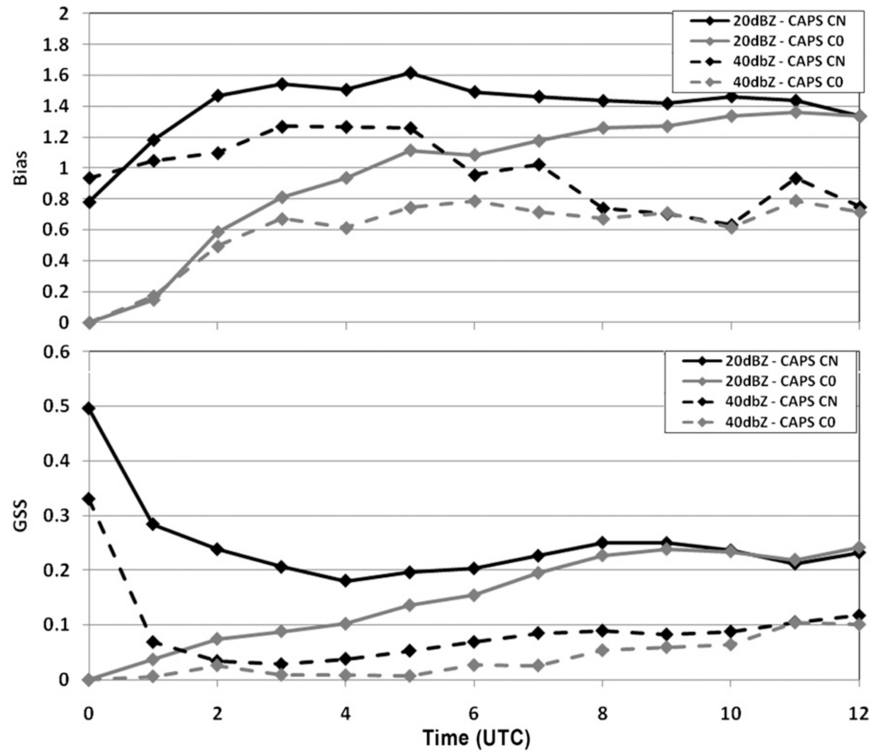

FIG. 5. Reflectivity coverage biases and GSSs as a function of time for all forecast hours and all days during SE2009. The dark (light) curves are derived from the run with (without) the assimilation of radar data, solid (dashed) lines represent scores using a 40- (20-) dBZ reflectivity threshold. Models were initialized at 0000 UTC.

difference. In a broad sense, traditional verification metrics from 2008 and 2009 are rather similar, indicating a decrease in the impacts of the advanced assimilation procedures after 6–12 h.

3) COMBINED ASSESSMENT FOLLOWING SE2008 AND SE2009

Visual examination of daily forecasts during SE2009 substantiated the results from SE2008, including the scale dependence of the value added with advanced assimilation (prediction of larger-scale systems seems to benefit more than prediction of small-scale convective elements) and the 3–6-h spinup time for the forecasts without the assimilation. All measures indicated a clear advantage for the CN runs through about 6 h. Beyond this time, side-by-side visual inspection did not reveal a consistent advantage for either the CN or C0 forecasts in terms of convective evolution and morphology, yet simple overlays of the SR and OR entities revealed a phase lag in the C0 forecasts on many days. This phase lag, similar in character to the phase shifts identified by Dawson and Xue (2006), may help explain the measured differences between CN and C0 GSSs 6 h into the forecast (Fig. 5).

With verification metrics limited to traditional approaches, the systematic impacts of this phase lag remain speculative. However, work is under way to develop more robust ways of applying MET/MODE to this dataset. A better characterization of the errors in both the C0 and CN forecasts is the expected outcome of this effort. For example, one of the MODE diagnostics is the centroid distance between objects. The algorithm for this metric can be modified easily to relate this distance to phase speed and direction, allowing us to confirm or deny our hypothesis of a systematic C0 phase lag in the prediction of convective features that are active in observations at the model initialization time. Alternatively, spatially shifted verification scores such as those used in Dawson and Xue (2006) could be applied. MET also has algorithms designed to determine forecast skill as a function of spatial scale (e.g., Casati et al. 2004), which should be useful in formally assessing the scale dependence of any skill differential.

4. Summary and conclusions

The assimilation of radar data into numerical weather prediction models holds great promise for improving forecasts, but it is also one of the most challenging problems facing numerical modelers. This problem has many different facets, including theoretical, computational, and practical constraints. Consequently, it is being addressed from many different angles (see, e.g., Stensrud et al. 2009 and references therein). Scientists at CAPS have been working on this issue for more than a decade and they recently applied their expertise in the development of a prototype CONUS-scale real-time forecasting system. This system has been deployed in real time with grid spacing as fine as 1 km (Xue et al. 2009), but this study focuses on an application involving two members of an ensemble in which the grid spacing was set to 4 km. The CAPS 3DVAR and cloud analysis scheme was used to enhance the 12-km NAM-based initial conditions in one of these members, but not in the other. Otherwise, they were configured identically. Thus, the impacts of the CAPS assimilation package were readily assessed by comparing the output from these two members.

This comparison was done on a daily basis during SE2008 and SE2009 to provide a subjective assessment of the impacts. In addition, the differences were measured after the close of the experimental periods using various objective verification metrics.

Compared to the no-assimilation (C0) forecasts, the visual assessments of the runs with enhanced initial conditions (CN) indicated a positive impact from the CAPS assimilation system during the first 6 h of integration—the time period during which precipitation features were spinning up in the C0 runs. Beyond six hours into the forecast the evaluation teams focusing on visual side-by-side comparisons often found it difficult to discern which forecast was better. However, simple overlay techniques introduced by DTC scientists seemed to indicate a small, yet systematic, phase lag in the C0 forecasts, similar to the phase shift identified in the MCS simulations by Dawson and Xue (2006). That is, not only did precipitation systems require 3–6 h of integration time to spin up when radar and other observational data were not assimilated, these features were often displaced slightly upstream compared to corresponding features in the observations and the CN forecasts. The reason for this phase lag is not clear, although it may be related to delayed and/or inadequate cold-pool production in the C0 runs, a factor that was shown to be operative when assimilated data were withheld in the Dawson and Xue (2006) case. Regardless of its source, the presence of the phase lag may help to explain why the CN forecasts earned slightly higher aggregate objective verification scores after ~6 h when any advantage in the overall convective evolution and morphology past this time was not readily discernible in side-by-side subjective assessments.

Visual scrutiny of the CN forecasts also suggested that the predictability time scales for individual convective cells that were introduced through the assimilation procedure were quite short. Although small-scale predictability is inherently limited, this limitation was likely exacerbated by the streamlined data assimilation strategy (single-time 3DVAR analysis) and relatively coarse grid resolution (4-km grid spacing) necessitated by the time constraints of the simulated operational context of the experiments described here. Ongoing research conducted by CAPS using higher-resolution and cycled 3DVAR (e.g., Hu et al. 2006a) or ensemble Kalman filter (e.g., Lei et al. 2009) techniques to assimilate WSR-88D data appear to yield longer predictability time scales for initialized convective-scale features. As additional computer resources become available and data assimilation expertise advances further, it seems likely that the predictability of initialized small-scale features will increase.

In spite of the limitations imposed by real-time daily production, the prototype forecast system implemented by CAPS during these experiments stands as a groundbreaking scientific and computational achievement. It lays the foundation for more advanced applications and further improvements to data assimilation methodologies, and it provided cutting-edge scientific data for the 2008 and 2009 Spring Experiments.

The availability of such datasets also catalyzed advances in other areas during SE2008 and SE2009. By bringing together experts not only in numerical modeling, but also in the areas of severe weather forecasting and model verification, these synergistic experiments resulted in progress on several fronts. For example, SE2008 and SE2009 exposed severe weather forecasters to new technologies and scientific concepts, including high-resolution deterministic models, ensembles, and advanced techniques in data assimilation. These technologies are envisioned as being fundamental components of future numerical guidance systems, such as Warn-On-Forecast (Stensrud et al. 2009), and early engagement with the forecasting community is essential for the success of these systems. In particular, forecaster feedback can be extremely helpful to model developers, who may have a limited understanding of how model output is used in the operational forecasting process. At the same time, human forecasters must frequently adapt to new innovations in modeling, and the success of this adaptation is facilitated by advanced knowledge of anticipated changes.

Consultation with forecasters is also critically important for developers of verification metrics, such as DTC scientists. For example, object-based verification metrics are designed to replicate an expert user's (e.g., a human forecaster's) assessment of the strengths and weaknesses of numerical guidance. Feedback from users is critical to the success of such a metric, just as it is to the optimization of a numerical model (Kain et al. 2003). During SE2009 participants generally had a positive impression of the information provided by the daily MET/MODE output, but they also identified a number of potential weaknesses in the algorithms used by this software package. Most notably, they found that unambiguous identification of corresponding objects in observations and model forecasts can be very challenging, particularly in convective regimes where reflectivity patterns are quite complex. Feedback from SE2009 participants is currently being used to refine the parameters and algorithms in the MET/MODE software for use in SE2010.

Acknowledgments. Dedicated work by many individuals led to the success of SE2008 and SE2009. At the SPC, HWT operations were made possible by technical support from Jay Liang, Gregg Grosshans, David Bright, Greg Carbin, and Joe Byerly. At the NSSL, Brett Morrow, Brad Sagowitz, Steve Fletcher, and Jeff Horn also provided valuable technical support. The CAPS forecasts were partially supported by the NOAA/CSTAR program and were performed at the Pittsburgh Supercomputing Center supported by the National Science Foundation. Supplementary support was provided by NSF ITR project LEAD, led by Kelvin Droegemeier (ATM-0331594), and by NSF Grant ATM-0802888. David O'Neal of PSC is thanked for his computing system support. The Developmental Testbed Center is funded by the National Oceanic and Atmospheric Administration, Air Force Weather Agency, and the National Center for Atmospheric Research. John Halley-Gotway and Steve Sullivan of the DTC helped with data retrieval and development of Web pages for critical displays. David Ahijevych, Tressa Fowler, and Jamie Wolff, also from the DTC, led discussions related to forecast verification during SE2009 daily activities. Ryan Sobash of NSSL and OU developed critical Web displays used at HWT for SE2009. Constructive reviews of this manuscript were provided by Dave Stensrud of NSSL, Matt Bunkers of the National Weather Service, and two anonymous reviewers.

REFERENCES

Aksoy, A., D. C. Dowell, and C. Snyder, 2009: A multicase comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analyses. *Mon. Wea. Rev.,* **137,** 1805–1824.

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.

Brown, B. G., R. G. Bullock, J. H. Gotway, D. Ahijevych, C. A. Davis, E. Gilleland, and L. Holland, 2007: Application of the MODE object-based verification tool for the evaluation of model precipitation fields. Preprints, *22nd Conf. on Weather Analysis and Forecasting/18th Conf. on Numerical Weather Prediction,* Park City, UT, Amer. Meteor. Soc., 10A.2. [Available online at http://ams.confex.com/ams/pdfpapers/124856.pdf.]

Casati, B., G. Ross, and D. B. Stephenson, 2004: A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteor. Appl.,* **11,** 141–154.

——, and Coauthors, 2008: Forecast verification: Current status and future directions. *Meteor. Appl.,* **15,** 3–18.

Davis, C. A., B. G. Brown, and R. G. Bullock, 2006: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.,* **134,** 1772–1784.

Dawson, D. T., II, and M. Xue, 2006: Numerical forecasts of the 15–16 June 2002 southern plains severe MCS: Impact of mesoscale data and cloud analysis. *Mon. Wea. Rev.,* **134,** 1607–1629.

Dixon, M., Z. Li, H. Lean, N. Roberts, and S. Ballard, 2009: Impact of data assimilation on forecasting convection over the United Kingdom using a high-resolution version of the Met Office Unified Model. *Mon. Wea. Rev.,* **137,** 1562–1584.

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.

Gao, J.-D., M. Xue, K. Brewster, and K. K. Droegemeier, 2004: A three-dimensional variational data analysis method with recursive filter for Doppler radars. *J. Atmos. Oceanic Technol.,* **21,** 457–469.

Hu, M., and M. Xue, 2007a: Impact of configurations of rapid intermittent assimilation of WSR-88D radar data for the 8 May 2003 Oklahoma City tornadic thunderstorm case. *Mon. Wea. Rev.,* **135,** 507–525.

——, and ——, 2007b: Implementation and evaluation of cloud analysis with WSR-88D reflectivity data for GSI and WRF-ARW. *Geophys. Res. Lett.,* **34,** L07808, doi:10.1029/2006GL028847.

——, ——, and K. Brewster, 2006a: 3DVAR and cloud analysis with WSR-88D level-II data for the prediction of Fort Worth tornadic thunderstorms. Part I: Cloud analysis and its impact. *Mon. Wea. Rev.,* **134,** 675–698.

——, ——, J. Gao, and K. Brewster, 2006b: 3DVAR and cloud analysis with WSR-88D level-II data for the prediction of Fort Worth tornadic thunderstorms. Part II: Impact of radial velocity analysis via 3DVAR. *Mon. Wea. Rev.,* **134,** 699–721.

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.

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

——, and Coauthors, 2008: Some practical considerations for the first generation of operational convection-allowing NWP: How much resolution is enough? *Wea. Forecasting,* **23,** 931–952.

Koch, S. E., B. S. Ferrier, M. Stoelinga, 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 Conf. on Mesoscale Processes,* Albuquerque, NM, Amer. Meteor. Soc., J4J.7. [Available online at http://ams.confex.com/ams/pdfpapers/97032.pdf.]

Kong, F., and Coauthors, 2008: Real-time storm-scale ensemble forecast experiment—Analysis of 2008 spring experiment data. Preprints, *24th Conf. on Severe Local Storms,* Savannah, GA, Amer. Meteor. Soc., 12.3. [Available online at http://ams.confex.com/ams/pdfpapers/141827.pdf.]

Lei, T., M. Xue, and T. Yu, 2009: Multi-scale analysis and prediction of the 8 May 2003 Oklahoma City tornadic supercell storm assimilating radar and surface network data using EnKF. Preprints, *13th Conf. on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface (IOAS-AOLS),* Phoenix, AZ, Amer. Meteor. Soc., 6.4. [Available online at http://ams.confex.com/ams/89annual/techprogram/paper_150404.htm.]

Lilly, D. K., 1990: Numerical prediction of thunderstorms—Has its time come? *Quart. J. Roy. Meteor. Soc.,* **116,** 779–798.

Lin, Y., and K. E. Mitchell, 2005: The NCEP Stage II/IV hourly precipitation analyses: Development and applications. Preprints, *19th Conf. on Hydrology,* San Diego, CA, Amer. Meteor. Soc., 1.2. [Available online at http://ams.confex.com/ams/pdfpapers/83847.pdf.]

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.

Rogers, E., and Coauthors, 2009: The NCEP North American mesoscale modeling system: Recent changes and future plans. Preprints, *23rd Conf. on Weather Analysis and Forecasting/19th Conf. on Numerical Weather Prediction,* Omaha, NE, Amer. Meteor. Soc., 2A.4. [Available online at http://ams.confex.com/ams/pdfpapers/154114.pdf.]

Sheng, C., S. Gao, and M. Xue, 2006: Short-term prediction of a heavy precipitation event by assimilating Chinese CINRAD radar reflectivity data using complex cloud analysis. *Meteor. Atmos. Phys.,* **94,** 167–183.

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.]

Stensrud, D. J., and Coauthors, 2009: Convective-scale warn-on-forecast system. *Bull. Amer. Meteor. Soc.,* **90,** 1487–1499.

Sun, J., and N. A. Crook, 1998: Dynamical and microphysical retrieval from Doppler radar observations using a cloud model and its adjoint. Part II: Retrieval experiments of an observed Florida convective storm. *J. Atmos. Sci.,* **55,** 835–852.

Tao, W.-K., and Coauthors, 2003: Microphysics, radiation, and surface processes in the Goddard Cumulus Ensemble (GCE) model. *Meteor. Atmos. Phys.,* **82,** 97–137.

Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.,* **136,** 5095–5115.

Tong, M., and M. Xue, 2005: Ensemble Kalman filter assimilation of Doppler radar data with a compressible nonhydrostatic model: OSS experiments. *Mon. Wea. Rev.,* **133,** 1789–1807.

Vasiloff, S. V., and Coauthors, 2007: Improving QPE and very short term QPF: An initiative for a community-wide integrated approach. *Bull. Amer. Meteor. Soc.,* **88,** 1899–1911.

Weisman, M. L., 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., M. E. Pyle, Z. Janjić, D. R. Bright, and G. J. DiMego, 2008: The operational high resolution window WRF model runs at NCEP: Advantages of multiple model runs for severe convective weather forecasting. Preprints, *24th Conf. on Severe Local Storms,* Savannah, GA, Amer. Meteor. Soc., P10.8. [Available online at http://ams.confex.com/ams/pdfpapers/142192.pdf.]

Xue, M., K. Brewster, D. B. Weber, K. W. Thomas, F. Kong, and E. Kemp, 2002: Realtime storm-scale forecast support for IHOP 2002 at CAPS. Preprints, *15th Conf. Numerical Weather Prediction/19th Conf. on Weather Analysis and Forecasting,* San Antonio, TX, Amer. Meteor. Soc., 124–126.

——, D.-H. Wang, J.-D. Gao, K. Brewster, and K. K. Droegemeier, 2003: The Advanced Regional Prediction System (ARPS), storm-scale numerical weather prediction and data assimilation. *Meteor. Atmos. Phys.,* **82,** 139–170.

——, and Coauthors, 2008: CAPS realtime storm-scale ensemble and high-resolution forecasts as part of the NOAA Hazardous Weather Testbed 2008 Spring Experiment. Preprints, *24th Conf. on Severe Local Storms,* Savannah, GA, Amer. Meteor. Soc., 12.2. [Available online at http://ams.confex.com/ams/pdfpapers/142036.pdf.]

——, and Coauthors, 2009: CAPS realtime 4 km multi-model convection-allowing ensemble and 1 km convection-resolving forecasts for the NOAA Hazardous Weather Testbed 2009 Spring Experiment. Preprints, *23rd Conf. on Weather Analysis and Forecasting/19th Conf. on Numerical Weather Prediction,* Omaha, NE, Amer. Meteor. Soc., 16A.2. [Available online at http://ams.confex.com/ams/pdfpapers/154323.pdf.]

Zhang, F., N. Bei, R. Rotunno, C. Snyder, and C. C. Epifanio, 2007: Mesoscale predictability of moist baroclinic waves: Convection-permitting experiments and multistage error growth dynamics. *J. Atmos. Sci.,* **64,** 3579–3594.