Assessing Advances in the Assimilation of Radar Data within a Collaborative Forecasting-Research Environment

John S. Kain[1], Ming Xue[2], Michael C. Coniglio[1], Steven J. Weiss[3], Fanyou Kong[2], Tara L. Jensen[4], Barbara G. Brown[4,5], Jidong Gao[2], Keith Brewster[2], Kevin W. Thomas[2], Yunheng Wang[2], Craig S. Schwartz[5], and Jason J. Levit[3]

[1]NOAA/National Severe Storms Laboratory, Norman, OK
[2]Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, OK
[3]NOAA/NWS/Storm Prediction Center, Norman, OK
[4]Developmental Testbed Center, Boulder, CO
[5]National Center for Atmospheric Research, Boulder, CO

Submitted as a Note to Weather and Forecasting

January 2010

Corresponding Author Address:
John S. Kain
National Severe Storms Laboratory
120 David L. Boren Blvd.
Norman, OK 73072
Email: jack.kain@noaa.gov

Abstract

The impact of radar-data assimilation in real-time, convection-allowing model forecasts was evaluated during the spring seasons of 2008 and 2009 as part of Hazardous Weather Testbed (HWT) Spring Experiment activities. Preliminary results suggest that an early prototype version of a CONUS-scale assimilation system had a positive impact on model forecasts, especially when organized convective activity was ongoing at the initial time. Daily interrogation of output by teams of modelers, forecasters, and verification experts provided additional insight into the value-added characteristics of the radar-assimilation forecasts. This evaluation revealed that the positive effect of the assimilation was 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 resolution 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., Weis 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 data sources, so current CAM-based forecast systems typically use relatively coarse resolution initial and lateral boundary conditions and rely on the "spin-up" process to develop smaller scale circulations such as those associated with convective overturning (Skamarock 2004). The spin-up process typically requires 3-6 h of integration time and it inherently results in errors in forecasts of small-scale phenomena such as thunderstorms.

Intuition 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 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, inflow, hydrometeors, etc. 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 WSR-88D radar network holds considerable promise for initializing CAMs

with storm-scale information. This network provides extensive coverage of reflectivity and radial velocity over the CONUS (the contiguous 48 states in the U.S.), from which storm-scale hydrometeor, kinematic, and thermodynamic properties can be derived. In fact, the use of Doppler radar data for initializing storm-scale numerical weather prediction models had been envisioned since early 1990's (Lilly 1990).

Beginning with the early leadership of Douglas Lilly, scientists at the University of Oklahoma Center for Analysis and Prediction of Storms (CAPS) have been working on assimilation of radar data for many years (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, but in the Spring of 2008, in association with the National Oceanic and Atmospheric Administration Hazardous Weather Testbed (NOAA HWT) Spring Experiment (hereafter SE2008) they for the first time initialized near-CONUS-scale CAM forecasts every evening at 0000 UTC with storm-scale data. These forecasts utilized data from the national network of WSR-88Ds, analyzing them at the model-native 4 and 2 km resolutions using CAPS' three-dimensional variational (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 ARPS modeling system (Xue et al. 2003). For comparison, CAPS also produced a 4-km forecast daily that was initialized without the radar data, but was identically configured in every other way.

This effort 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 run with radar assimilation were compared to corresponding fields from the parallel no-radar forecast, and both were critically judged by comparison with observed radar reflectivity. Subjective impressions from these comparisons were documented as part of the SE2008 record.

After SE2008, traditional verification metrics, such as bias and Gilbert Skill Scores (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 National Developmental Testbed Center (DTC), who had participated in SE2008. The DTC is a center of expertise in the verification of high resolution model forecasts and through these consultations DTC and HWT scientists identified a compelling strategy for realtime 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 during in an operational environment. 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 in which the impact of the CAPS radar-assimilation methodology was assessed from several different perspectives. The purpose of this Note

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, etc. will be outlined and the framework for verification described. This will be followed by a section on results and a final section on the summary and conclusions.

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 through Friday in 2008 and Monday through Thursday in 2009.

a. Model configuration and initialization procedures

CAPS ran a 4-km resolution ensemble prediction system 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 assimilation of radar and other observational data, while the other one did not. This study focuses exclusively on output from these two members.

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 NCEP (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 this background directly as initial condition 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 condition. Specifically, the unique assimilation process in the CN run ingested data from the national network of WSR-88D radars, 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 conventional rawinsonde, wind profiler, METAR (Meteorological Aviation Report) surface observations and Oklahoma Mesonet observations was also included. Furthermore, visible and infrared channel-4 data from GOES satellites were used in the cloud analysis package. Details about this complex assimilation process can be found in Xue et al. (2008; 2009, and references therein).

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

b. Verification data sets

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 comparison of SR with observed reflectivity from WSR-88D radars in both 2008 and 2009. Objective verification in 2008 used accumulated precipitation fields. The verifying data are derived from the Stage II hourly precipitation analyses (Baldwin and Mitchell 1998). In 2009, both accumulated precipitation fields and simulated reflectivity were evaluated but the focus of this

discussion will be on simulated reflectivity. 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 microphysical parameterization that was used during model integration. For this study, 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 from national composite-reflectivity mosaics that are part of the SPC operational data stream. These mosaics are generated by Unisys Corporation at a frequency of five minutes 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 were 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? 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 assessment of the evaluation teams was 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 realtime 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 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 simulated reflectivity 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

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 Meteorological Evaluation Tools (MET) software package (Brown et al. 2007). In addition to numerous traditional verification metrics, such as the Gilbert Skill Score (GSS) and Critical Success Index (CSI), bias, and false alarm ratio, this package includes tools that produce additional verification statistics designed to quantify the correspondence between objects, or features, in forecasts and observations. This Method of Object-based Diagnostic Evaluation (MODE – see Davis et al. 2006; Brown et al. 2007) 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 comparison of forecasts and observations. Meanwhile, work is underway 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 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 altogether by one hour 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 hydrometeors from radar data are quite faithful to the observations - the correspondence between 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 but retains a good semblance of the general envelope of convective activity (cf. Figs. 2 d and e). 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 integration time progresses, the two forecasts begin to look more and more alike, and less like observations (cf. Figs. 2 j, k, and l), but the run with 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. Also, the input radar dataset was incomplete on many days. Nonetheless, and in spite of some inconsistencies in performance from day to day, 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 higher 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 a surprise to many who participated in the daily SE2008 subjective assessments. Apparently, the realtime focus on the most intense convective cores (which typically dissipated within the first hour of the CN forecast) drew attention away from the fact that assimilated mesoscale features, such as MCSs, often showed relatively good continuity after initialization. These larger systems can dominate traditional verification metrics by simple virtue of their size, 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 supported by a number of cases discussed in 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 like observations beyond the first several hours. Specifically, when GSSs and CSIs are re-computed 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), 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. This is apparently because the two runs share very similar forecasts of the mesoscale environment. After the predictability of small convective-scale features initialized by the radar data is lost, the development and evolution of convective activity is strongly modulated by forcing mechanisms in the environment.

3) PRELIMINARY ASSESSMENT FOLLOWING SE2008

The impact of radar data assimilation 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 "non-meteorological" in the first six hours of the forecasts on some days, after which more realistic looking features began to spin up in both the run with radar-data assimilation and the "cold start" run. So the radar 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 (QPF) in the first 3 to 6 h. Beyond this time, and through about 15h, GSS and CSI scores indicated a QPF advantage for the CN runs in spite of the fact that CN and C0 runs looked much more like each other than observations during this period.

b. SE2009

1) SUBJECTIVE ASSESSMENTS

In 2009, the subjective assessments were greatly facilitated by a web-based display 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

[1] This conclusion may be specific to the particular data assimilation strategy and model grid spacing used during SE2008 and SE2009. Due to the constraint of realtime 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. 2008) techniques, smaller scale convective structures introduced through assimilation of WSR-88D data appear to evolve quite realistically through the life cycle of the storms.

.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, showing the observed objects identified by MODE outlined in blue, with the corresponding objects from the two forecasts (CN in the next-to-bottom row, C0 in the bottom) overlaid as color filled areas. The corresponding SR fields are in the second (CN) and third (C0) rows; OR fields are in the top row.

Examination of these displays yielded impressions that were similar to those of 2008. For example, the initialization from 0000 UTC 14 May 2008 indicated that the CAPS assimilation scheme placed hydrometeor fields on the model grid very accurately (cf. Figs. 4 a and b), but the smaller scale features appeared to lose coherence rapidly in the forecast. By the two hour time, the CN SR pattern seemed quite disorganized compared to observations (cf. Figs. 4 f and g), 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 observations, especially on the southern end of the line (Fig. 4 j).

Four hours into the 14 May forecast, the CN forecast still had a diffuse representation of the main convective line compared to observations, but also continued to have very good spatial overlap with the observed system, at least at the 30dBz threshold (Figs. 4 k, l, and n). At this time, it could be argued that SR from the C0 run looked more like observations than did SR from the CN forecast in terms of convective evolution and morphology (cf. Figs. 4k, l, and 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. 4 p, r, and t. In fact, an upstream lag in C0 systems was noted on a number of days during SE2009. The tendency for this 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 GSS and CSI scores favor the CN forecasts well beyond the ~ 6 h integration time when such a systematic difference 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, generation of statistically significant inferences from these metrics is quite challenging due to sample size constraints (i.e. 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. But GSSs from the CN and C0 runs converge at about the 10 h time (Fig. 5). Although the convergence point may be a few hours earlier than 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

difference. In a broad sense, traditional verification metrics from 2008 and 2009 are rather similar, indicating a decrease in the impact of the radar assimilation 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 radar assimilation and the 3-6 h spin-up time for the forecasts without radar assimilation. All measures indicated a clear advantage for the CN runs through about 6h. 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 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 after the 6 h time (Fig. 5).

With verification metrics limited to traditional approaches, the systematic impact of this phase lag remains speculative. However, work is underway 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 convective features that are active at model initialization time. Alternatively, spatially shifted verification scores such as those used in Dawson and Xue (2006) could be applied. MET also has algorithms 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 discussion

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 more than a decade of experience working on this issue and they recently applied their expertise in the development of a CONUS-scale realtime forecasting system, including one at 1-km resolution in 2009 (Xue et al 2009). This system assimilates data from the entire national network of WSR-88D radars. Because of the computational demands associated with a CONUS-scale implementation, this prototype system uses a relatively efficient 3DVAR analysis technique, rather than a more computationally intensive 4DVAR (e.g., Sun and Crook 1998) or ensemble-based approach (e.g., Tong and Xue 2005; Lei et al. 2008; Aksoy et al. 2009) approaches to assimilate cloud-scale data. Furthermore, high-frequency assimilation cycles (e.g., Hu and Xue 2008) making use of multiple volume scans of radar data was not done for the realtime experiments. The lack of balance among state variables and consistency with the prediction model in the single time 3DVAR-based analysis is at least partially responsible for the relatively rapid decrease is the forecast skills of convective-scale features in the first few hours (Hu et al. 2006b, Hu and Xue 2008). Meanwhile, model biases (Kong et al 2009) and sensitivity to microphysics (Xue et al. 2009) can also significantly affect the

forecast performance. Nonetheless, the effort represents a ground-breaking scientific and computational achievement and lays the foundation for more advanced applications of such a prototype and further improvement to the data assimilation methodologies in the future.

Output from this forecasting system incorporating radar assimilation was visually compared to output from forecasts with no assimilation of radar data during HWT Spring Experiments in 2008 and 2009 and the differences were scrutinized in order to assess the impact of the radar-data assimilation. In addition, the impact was measured after the close of the experimental periods using various objective verification metrics.

Compared to the no-radar (C0) forecasts, the visual assessments of the radar (CN) runs 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. After the 6 h time 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 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 observations and the CN forecasts. This phase lag may help to explain why the CN forecasts earned slightly higher aggregate objective verification scores after ~ 6 h when any advantage in overall convective evolution and morphology past this time was not readily discernible in side-by-side subjective assessments.

This study provides valuable guidance for future experimentation related to the radar-assimilation problem, but its impact reaches far beyond that. By bringing together experts not only in numerical modeling, but also in the areas of severe weather forecasting and model verification, this synergistic effort 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 fundamental components of future numerical guidance systems, such as Warn-On-Forecast systems (Stensrud et al. 2009), and early engagement with the forecasting community is essential for their success. 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 strength 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. During SE2009 participants generally had a positive impression of the information provided by 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 parameters and algorithms in the MET/MODE software for use in SE2010.

Acknowledgements 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 Steve Sullivan of the DTC helped with data retrieval and development of web pages for critical displays and 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.

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 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 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 Conference on Weather Analysis and Forecasting/18th Conference on Numerical Weather Prediction*, Amer. Meteor. Soc., Park City, UT. paper 10A.2

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.

Casati, B., L. J. Wilson, D. B. Stephenson, P. Nurmi, A. Ghelli, M. Pocernich, U. Damrath, E. E. Ebert, B. G. Brown, and S. Mason, 2008: Forecast verification: Current status and future directions. *Meteor. Appl.*, **15**, 3–18.

Davis, C., B. Brown, and R. 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. Ocean. Tech.* **21**, 457-469.

Hu, M., M. Xue, 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.

Hu, M., M. Xue, 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.

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.

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

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.

Janjic, 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., S. J. Weiss, D. R. Bright, M. E. Baldwin, J. J. Levit, G. W. Carbin, C. S. Schwartz, M. Weisman, K. K. Droegemeier, D. Weber, and K. W. Thomas, 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. 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., paper J4J.7

Kong, F., M. Xue, M. Xue, K. K. Droegemeier, K. W. Thomas, Y. Wang, J. S. Kain, S. J. Weiss, D. Bright, and J. Du, 2008: Real-time storm-scale ensemble forecast experiment - Analysis of 2008 spring experiment data. *Preprints, 24th Conf. Several Local Storms*, Savannah, GA, Amer. Meteor. Soc., Paper 12.3.

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. *Extended abstract, 13th Conf. of IOAS-AOLS, AMS Meetings 2008*, Paper 6.4.

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

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., G. J. DiMego, T. L. Black, M. B. Ek, B. S. Ferrier, G. A. Gayno, Z. Janjic, Y. Lin, M. E. Pyle, V. C. Wong, W.-S. Wu, and J. Carley, 2009: The NCEP North American mesoscale modeling system: Recent changes and future plans. *Preprints, 23rd Conference on Weather Analysis and Forecasting/19th Conference on Numerical Weather Prediction*, Amer. Meteor. Soc., Omaha, NE. paper 2A.4

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.

Skamarock, W. C., 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.]

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.

Stensrud, D. J. , M. Xue, L. J. Wicker, K. E. Kelleher, M. P. Foster, J. T. Schaefer, R. S. Schneider, S. G. Benjamin, S. S. Weygandt, J. T. Ferree, and J. P. Tuell, 2009: Convective-Scale Warn-on-Forecast System. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1499.

Tao, W.-K.; J. Simpson; D. Baker, S. Braun, M.-D. Chou, B. S. Ferrier, B.; D. Johnson, A. Khain, A.; S. Lang, B. Lynn, C.-L. Shie, D. Starr, D.; C.-H. Sui, Y. Wang, and P. Wetzel, 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., D. J. Seo, K.W. Howard, J. Zhang, D. H. Kitzmiller, M. G. Mullusky, W. F. Krajewski, E. A. Brandes, R. M. Rabin, D. S. Berkowitz, H. E. Brooks, J. A. McGinley, R. J. Kuligowski, and B. G. Brown, 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. Janjic, 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 Conference on Severe Local Storms*, Amer. Meteor. Soc., Savannah, GA. CD-ROM P10.8

Xue, M., K. Brewster, D. Weber, K. W. Thomas, F. Kong, and E. Kemp, 2002: Realtime storm-scale forecast support for IHOP 2002 at CAPS. Preprint, 15th Conf. Num. Wea. Pred. and 19th Conf. Wea. Anal. Forecasting, San Antonio, TX, Amer. Meteor. Soc., 124-126

Xue, M., 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. Physics*, **82**, 139-170.

Xue, M., F. Kong, K. W. Thomas, J. Gao, Y. Wang, K. Brewster, K. K. Droegemeier, J. S. Kain, S. J. Weiss, D. R. Bright, M. C. Coniglio, and J. Du, 2008: CAPS realtime storm-scale ensemble and high-resolution forecasts as part of the NOAA Hazardous Weather Testbed 2008 Spring Experiment. *Preprints, 24th Conf. Severe Local Storms*, Savannah, GA, Amer. Meteor. Soc., Paper 12.2.

Xue, M., F. Kong, K. Thomas, J. Gao, Y. Wang, K. Brewster, K. Droegemeier, J. S. Kain, S. J. Weiss, D. R. Bright, M. C. Coniglio, and J. Du, 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 Conference on Weather Analysis and Forecasting/19th Conference on Numerical Weather Prediction*, Amer. Meteor. Soc., Omaha, NE. paper 16A.2

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.

Figure Captions

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

Fig. 2. Comparison of observed and simulated composite reflectivity for forecast hours 0, 1, 3, and 6, beginning at 0000 UTC 5 June 2008. Observations are in the left-hand column, the CN forecast is in the middle column, and the C0 forecast is in the right column.

Fig. 3. Coverage Bias, CSI, and GSS as a function of time for all forecast hours and all days during SE2008, for the 1 mm hr^{-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.

Fig. 4. Comparison of observed and simulated composite reflectivity for forecast hours 0, 2, 4, and 6, beginning at 0000 UTC 14 May 2009. Observations are in the top row, with the CN and C0 forecasts in the second and third rows, respectively. In the bottom two rows, the location of observed features is 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.

Fig. 5. Gilbert Skill Score as a function of time for all forecasts during SE2009. The dark (light) curves are derived from the run with (without) assimilation of radar data, solid (dashed) lines represent scores using a 40 (20) dBz reflectivity threshold.

Table Caption

Table I. Weather Research and Forecasting (WRF) model configuration used by CAPS during SE2008 and SE2009. ARW denotes the Advanced Research WRF dynamic core (Skamarock et al. 2005); the physical parameterizations include Thompson microphysics (Thompson et al. 2008), MYJ boundary layer/turbulence (Janjic 1994), Goddard short-wave radiation (Tao et al. 2003), RRTM: Rapid Radiative Transfer Model (Mlawer et al. 1997; Iacono et al. 2000). Version 2.2 of the WRF-ARW was used in 2008, version 3.0.1.1 in 2009.

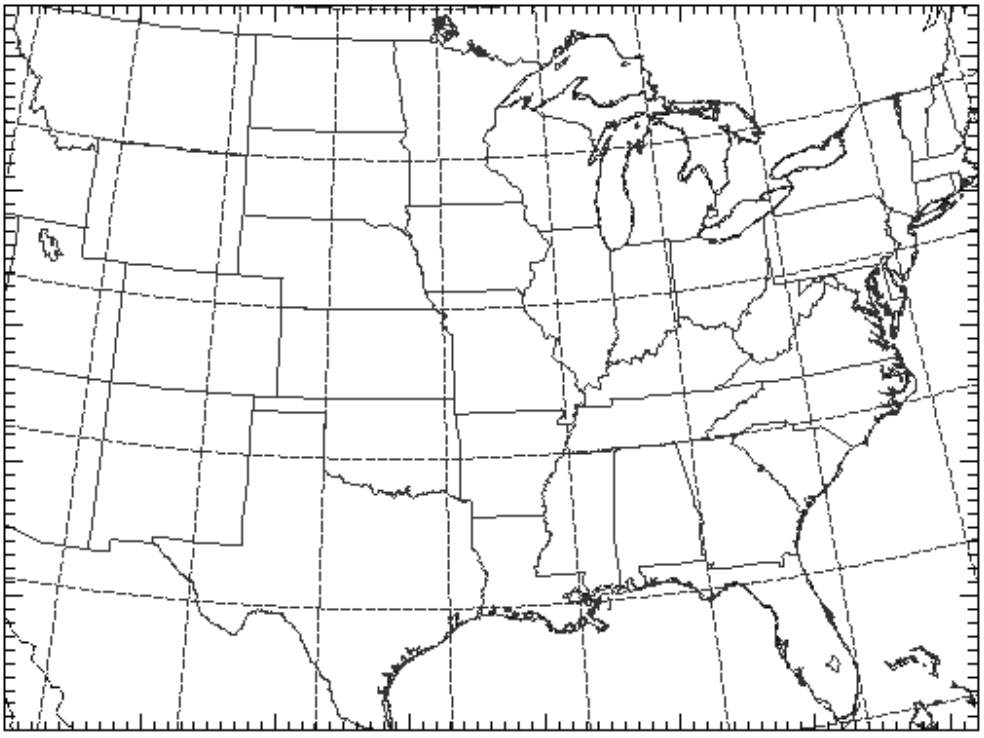

Fig. 1. Model domain for the 4 km 2008 and 2009 CAPS forecasts.

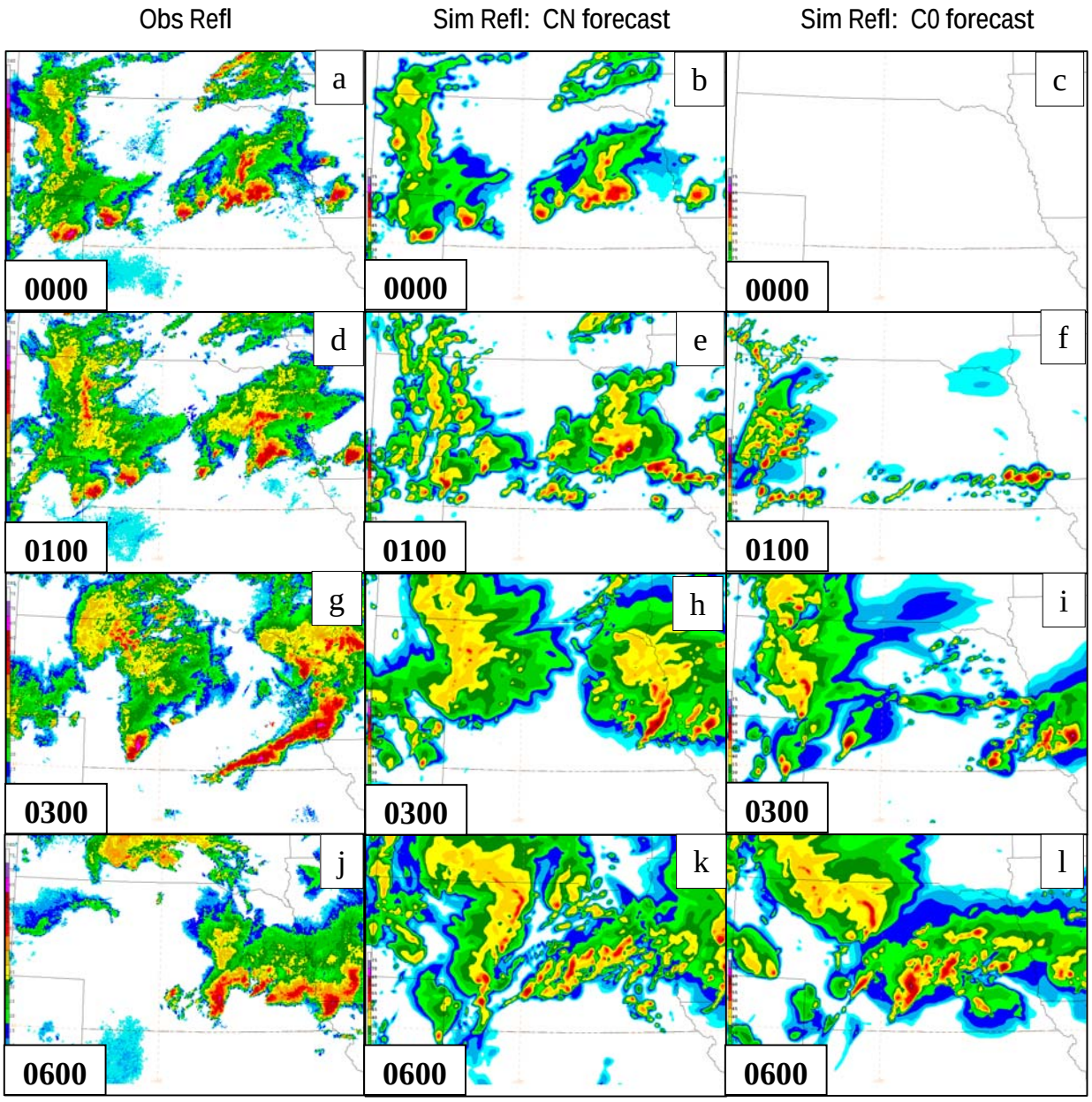

Fig. 2. Comparison of observed and simulated composite reflectivity for forecast hours 0, 1, 3, and 6, beginning at 0000 UTC 5 June 2008. Observations are in the left-hand column, the CN forecast is in the middle column, and the C0 forecast is in the right column.

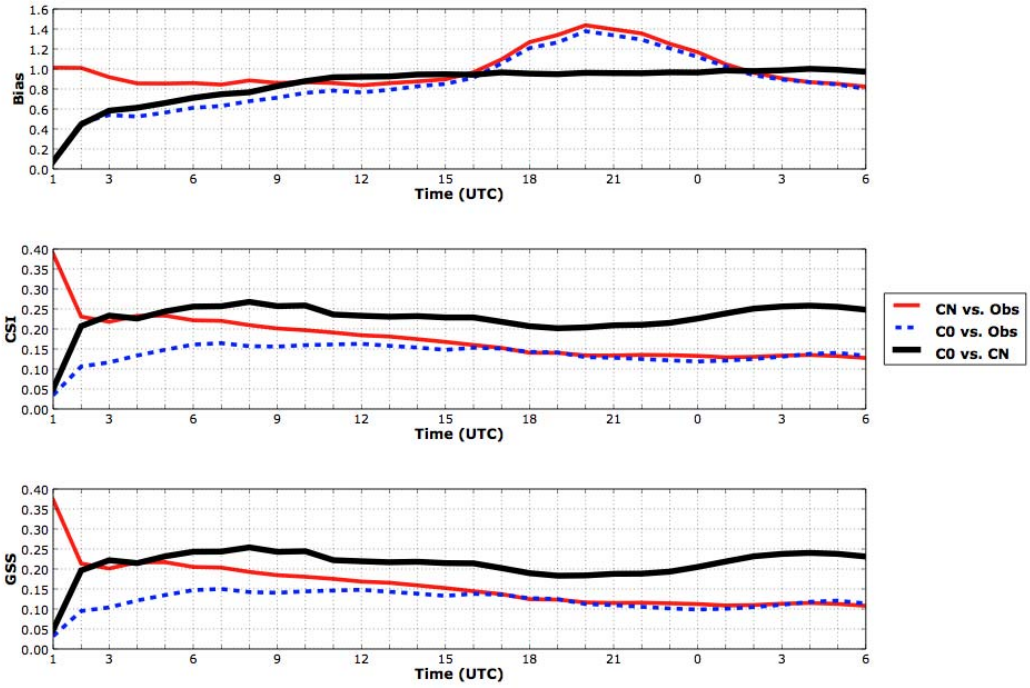

Fig. 3. Coverage Bias, CSI, and GSS as a function of time for all forecast hours and all days during SE2008, for the 1 mm hr^{-1} precipitation-rate threshold. Note that the black curve in each image compares the C0 and CN fields, while the other curves compare the respective model forecasts to Stage II observed precipitation data.

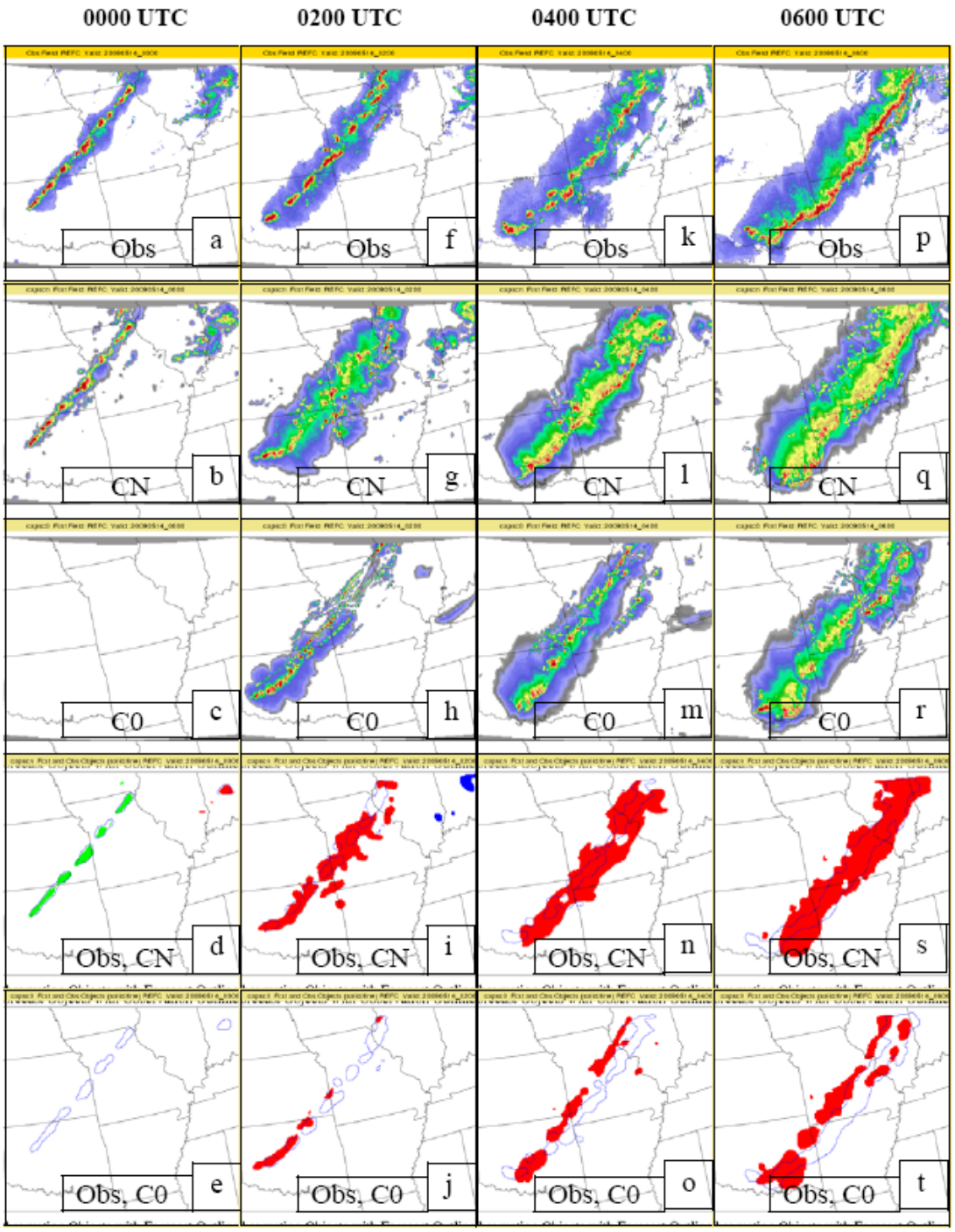

Fig. 4. Comparison of observed and simulated composite reflectivity for forecast hours 0, 2, 4, and 6, beginning at 0000 UTC 14 May 2009. Observations are in the top row, with the CN and C0 forecasts in the second and third rows, respectively. In the bottom two rows, the location of observed features is 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.

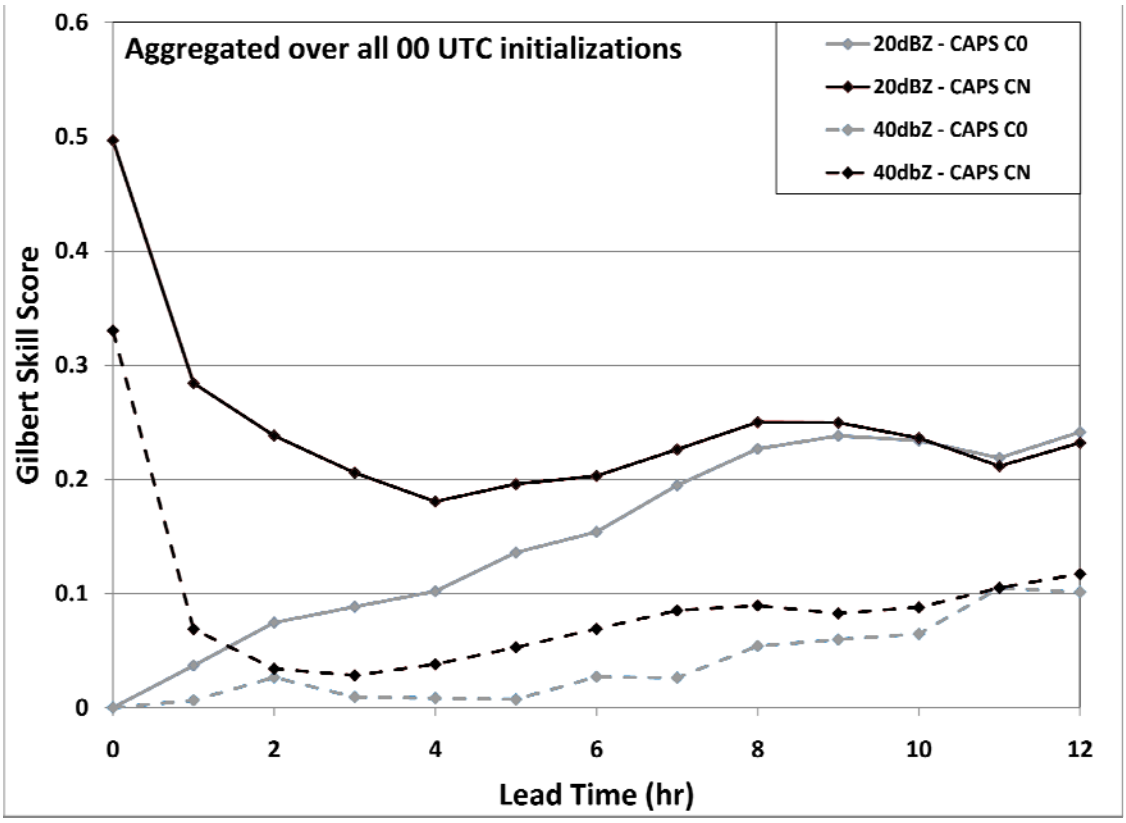

Fig. 5. Gilbert Skill Score as a function of time for all forecasts during SE2009. The dark (light) curves are derived from the run with (without) assimilation of radar data, solid (dashed) lines represent scores using a 40 (20) dBZ reflectivity threshold.

Table I. Weather Research and Forecasting (WRF) model configuration used by CAPS during SE2008 and SE2009. ARW denotes the Advanced Research WRF dynamic core (Skamarock et al. 2005); the physical parameterizations include Thompson microphysics (Thompson et al. 2008), MYJ boundary layer/turbulence (Janjic 1994), Goddard short-wave radiation (Tao et al. 2003), RRTM: Rapid Radiative Transfer Model (Mlawer et al. 1997; Iacono et al. 2000). Version 2.2 of the WRF-ARW was used in 2008, version 3.0.1.1 in 2009.

Dynamic Core	WRF-ARW
Horiz. Grid (km)	4
Vertical Levels	51
Microphysics Param	Thompson
PBL/Turb. Param	MYJ
Radiation (SW/LW)	Goddard/RRTM
Background IC/LBC	12 km NAM