

12.2 CAPS REALTIME STORM-SCALE ENSEMBLE AND HIGH-RESOLUTION FORECASTS AS PART OF THE NOAA HAZARDOUS WEATHER TESTBED 2008 SPRING EXPERIMENT

Ming Xue¹, Fanyou Kong¹, Kevin W. Thomas¹, Jidong Gao¹, Yunheng Wang¹, Keith Brewster¹,
Kelvin K. Droegemeier¹, John Kain², Steve Weiss³, David Bright³, Mike Coniglio⁴ and Jun Du⁵

¹Center for Analysis and Prediction of Storms, University of Oklahoma

²National Severe Storms Laboratory, NOAA

³Storm Prediction Center/NCEP, NOAA

Norman Oklahoma

⁴Cooperative Institute for Mesoscale Meteorological Studies, Norman Oklahoma

⁵Environmental Modeling Center/NCEP, NOAA, Maryland

1. Introduction

Accurate prediction of convective-scale hazardous weather continues to be a major challenge, because of the small spatial and short temporal scales of the associated weather systems, and the inherent nonlinearity of their dynamics and physics. So far, the resolutions of typical operational numerical weather prediction (NWP) models remain too low to resolve explicitly convective-scale systems, which constitutes one of the biggest sources of uncertainty and inaccuracy of quantitative precipitation forecast. These and other uncertainties as well as the high-nonlinearity of the weather systems at such scales render probabilistic forecast information afforded by high-resolution ensemble forecasting systems especially valuable to weather forecasters and decision makers.

Under the support of the NOAA CSTAR (Collaborative Science, Technology, and Applied Research) Program with leverage on the support of the NSF Large ITR LEAD (Linked Environment for Atmospheric Discovery) project, the Center for Analysis and Prediction of Storms (CAPS) at the University of Oklahoma is carrying out a three year project, in collaborations with the NOAA Hazardous Weather Testbed (HWT, see, e.g., Weiss et al. 2007) in Norman Oklahoma, to develop, conduct, and evaluate realtime high-resolution ensemble and deterministic forecasts for convective-scale hazardous weather. The realtime forecasts, together with retrospective analyses using the real time data, aim to address scientific issues including: (1) the values and cost-benefits of storm-scale ensemble versus coarser-resolution short-range ensembles and even-higher-resolution deterministic forecast; (2) suitable perturbation methods for storm-scale ensemble, physics perturbations, and multi-model ensemble; (3) proper handling and use of lateral and lower boundary perturbations; (4) the value and impact of assimilating high-resolution data including those from WSR-88D radars;

(5) the most effective ensemble forecast products for the storm scales; and (6) the impact of such unique products on realtime forecasting and warning.

In the spring of 2007, the first year of the project, daily 33-hour 10-member 4-km-resolution storm-scale ensemble forecasts (SSEF) were produced as contributions to the HWT Spring Experiment (Weiss et al. 2007). At the same time, a single 33-hour 2-km deterministic forecast was produced over the same domain that covers two thirds of the continental US (CONUS). The forecast configurations and preliminary analyses of the results were reported in Xue et al. (2007) and Kong (2007). Such a large data set consisting of ensemble and corresponding higher-resolution deterministic forecasts over a near-CONUS-sized domain, at convection-allowing/resolving resolutions and over an extended period provide an unprecedented opportunity for detailed post-realtime analyses and for investigating many of the forecasting issues raised earlier. In fact, several papers/manuscripts have resulted from the use of this data set (e.g., Clark et al. 2008; Schwartz et al. 2008a; Schwartz et al. 2008b), and additional papers are under preparation.

For the spring 2007 forecasts, initial conditions (ICs) were obtained by directly interpolating NCEP NAM analyses at 2100 UTC, and the lateral boundary conditions (LBCs) were derived from the 1800 UTC NAM forecasts, with perturbations derived from the 2100 UTC SREF (Short-range Ensemble Forecast, Du et al. 2006) forecasts added to the LBCs of four of the ensemble members (Kong et al. 2007; Xue et al. 2007). For the purpose of isolating the effects of different microphysics and PBL parameterization schemes, 5 of the 10 ensemble members used the same initial and boundary conditions as the control member, while other members contained both physics and IC and LBC perturbations. This configuration allowed for the investigation on physics sensitivity (Schwartz et al. 2008b) while the subsequent analyses also showed clear underdispersion among the physics-perturbation-only members (Kong et al. 2007; Kong et al. 2008).

In the spring of 2008, CAPS continued to provide

¹ Corresponding author address:

Dr. Ming Xue, Center for Analysis and Prediction of Storms,
University of Oklahoma, Norman OK 73072; e-mail: mxue@ou.edu

4-km ensemble and 2-km high-resolution forecasts to the HWT Spring Experiment. Learning from the experiences and results of 2007, the forecast configurations were improved for 2008. All of the ten 4-km ensemble members included initial and boundary condition perturbations as well as physics perturbations and the forecasts were initialized at 0000 UTC instead. The 0000 UTC initial conditions benefited from observations at this synoptic time and the forecasts were able to use the 0000 UTC NAM forecasts for boundary conditions. Comparison tests showed that forecasts using this configuration are significantly better. With the later start time, forecasts were run for 30 instead of 33 hours, ending at 0006 UTC of the second day.

The most important enhancement to the forecasts of 2008 is the assimilation of level-2 radial velocity and reflectivity data from over 120 operational WSR-88D radars into all except for one model run. Further, the model domain is enlarged to beyond the eastern coast of CONUS, and the domain used by both 4 km and 2 km forecasts is shown in Fig. 1.

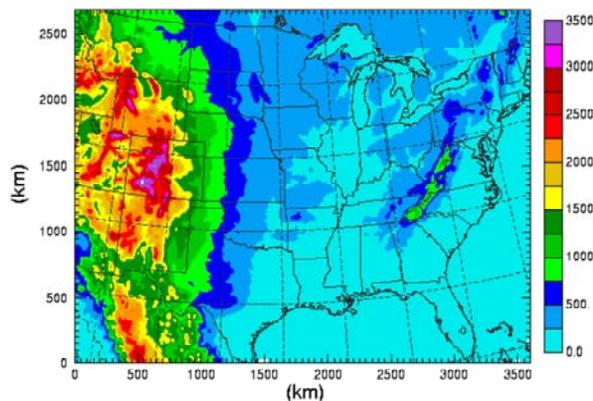


Fig. 1. The domain of the 2-km HR and 4-km storm-scale ensemble forecasting (SSEF) models used in 2008. Color shaded contours for terrain elevation are shown.

This paper reports on the experimental design and discusses certain logistic issues. Examples of ensemble and deterministic forecasts will be presented, and compared against radar observations. A companion paper by Kong et al. (2008) presents a more detailed evaluation of the ensemble forecasts. Coniglio et al. (2008) further discuss the values of such forecasts in terms of convective-scale features.

2. Forecast Configurations

During the spring of 2008, the CAPS forecasts were produced from April 14 through June 6 typically on Saturday through Thursday of each week for evaluations during the ensuing weekdays at HWT. As in 2007, the Weather Research and Forecast (WRF) ARW-core (V2.2) was used as the forecast model. Major changes from the 2007 experiment include: (1) The model do-

main is enlarged (Fig. 1); (2) Daily 30 h forecasts are initiated at 0000 UTC, using NAM 12 km (218 grid) 0000 UTC analyses as the background for additional storm-scale analysis including radar data, and with the initial condition perturbations for the ensemble extracted from the SREF; (3) Available WSR-88D data are assimilated through ARPS 3DVAR and cloud analysis package into all but one member; (4) Eight members contain both IC/LBC and physics perturbations. The IC perturbations are extracted from the 3 h forecasts of eight 2100 UTC SREF members and are scaled to their initial perturbation amplitudes.

Similar to 2007, special software developed at CAPS was used to initialize the WRF forecasts. The software is based on the pre-processing programs of the ARPS (Xue et al. 2003) modeling system for the creation of ICs and LBCs of regular and ensemble forecasts. These programs were configured for an ARPS grid that is identical to the WRF grid in the horizontal, with the same map projection, and all the gridded data preparations, including the data analysis using the ARPS 3DVAR system (Gao et al. 2003), were done on the ARPS grid. The final fields were interpolated (in the vertical) to the WRF grid and converted to WRF-ARW variables using program ARPS2WRF. The WRF forecast outputs were converted to the ARPS grid using WRF2ARPS for most of the post-processing by ARPS programs. All pre- and post-processing programs, including ARPS 3DVAR, were parallelized using MPI, with significant optimization with respect to data I/O, etc. The 4-km and 2-km grids contained $903 \times 675 \times 53$ grid and $1803 \times 1347 \times 53$ grid points respectively, and the first layer about ground was about 20 meter thick and the vertically stretched grid was configured using a hyperbolic tangent function based on the ARPS.

For the first time ever, level-2 radial velocity and reflectivity data together from all WSR-88D radars within a near-CONUS-sized domain were directly assimilated into convection-allowing/resolving forecast models. At CAPS, the level-II and level-III (used when level-II data were not available) WSR-88D radar data were ingested through the LDM software. They were automatically quality controlled and mapped to the 2 and 4-km ARPS grid columns using program 88D2ARPS. Conventional rawinsonde, profiler, SAO (Surface Aviation Observation) and Oklahoma Mesonet data were also ingested. The visible and infrared channel-4 data of GOES satellites were used in the ARPS complex cloud analysis package together with radar reflectivity data. The parallel version of the ARPS 3DVAR system was run for both the 4 and 2 km grids, using the above observations except for rawinsonde data. The latter is because the rawinsonde data were already included in the background analyses of NAM. The primary goal of our analysis was to introduce new information carried in high-resolution observations,

especially that in the surface mesonet and radar data.

As is the typical practice with the use of ARPS 3DVAR (e.g., Hu et al. 2006b), a multi-pass procedure was used. The profiler and surface observations were first analyzed in the first pass, using a horizontal background error de-correlation scale of 300 km. The radar radial velocity data are then analyzed in the second pass with a 30 km de-correlation scale. The variational analysis was then followed by a complex cloud analysis step that incorporated the reflectivity data, satellite cloud observations and surface network cloudbase data. Wind and temperature information in the variational analysis was utilized in the cloud analysis procedure. The key effects of the cloud analysis include the adjustments to the temperature and moisture fields inside the clouds, and the analyses of cloud and hydrometeor fields based on the observations and the analysis background (Brewster 2002). This procedure and its effect have been documented in a number of papers (e.g., Xue et al. 2003; Hu et al. 2006a; Sheng et al. 2006; Hu and Xue 2007a) within the ARPS framework and applied to WRF prediction in Hu and Xue (2007b). It was first used in realtime forecasts during 2002 in support of the IHOP field experiment, on grids of 27, 9 and 3 km horizontal resolutions (Xue et al. 2002). It is the first time for it to be used in realtime at the national scale and at convection allowing/resolving resolutions, and in combination with a 3DVAR procedure. Because of the memory constraint on each processor node, the radar radial velocity data at every other column, e.g., at 4 km horizontal spacings, were used in the 2 km 3DVAR analysis. The same spacing was used for the 4 km analyses.

Three dimensional gridded outputs were written to disk every hour, and all were archived in the mass storage system of the Pittsburgh Supercomputing Center (PSC) and are available for post-realtime analysis. The data archived were about 1 terabytes a day. In addition

to the hourly 3-D output, 2-D composite reflectivity fields from 2-km (CN2), the 4-km control (CN) and the 4-km run without radar data (C0) (Table 1) were written out every 5 min and animations were produced in quasi-real-time and posted on the web together with corresponding animations produced from the NSSL mosaic reflectivity interpolated to the same grid (e.g., <http://www.caps.ou.edu/~fkong/spring08/2008052100.html>).

A total of about 1500 CPUs of a Cray XT-3 super-computer at the PSC were reserved for up to 8 hours each night for the forecasts. The 2-km 3DVAR analysis and WRF forecasts used 720 processors, and the 4-km analysis and each 4-km forecast used 80 processors. The 4-km forecasts took 5 to 8 hours to complete (depending on physics options) while the 2-km forecasts took about 8 hours.

All forecasts used the RRTM shortwave and Goddard long-wave radiation schemes and the NOAH land surface model. No cumulus parameterization was used. The subgrid-scale turbulence mixing was turned on without explicit computational mixing. The microphysics and PBL schemes are varied among the ensemble members while the 4-km control member used the Thompson microphysics scheme predicting two moments for cloud ice together with the Mellor-Yamada-Jancic (MYJ) PBL scheme. The physics options for other ensemble members are given in Table 1. The 2-km forecast (CN2) used the same options as the 4-km control members (CN and C0), except for the microphysics, which was the single-moment 6-category WRF scheme (WSM6). This scheme was used in 2007 for both 2-km and 4-km control, and was left unchanged in 2008 for the 2-km forecast, even though the intention was to use the same options as the 4-km control. For this reason, the 2-km forecasts shared the same physics options as 4-km member N3, but without IC or LBC perturbations.

Table 1. CAPS 2008 4-km storm-scale ensemble forecast (SSEF) and the 2-km forecast configurations

Member	IC	LBC	Radar	Microphysics	Shortwave Radiation	PBL Scheme
CN	00Z ARPS Analysis	00Z NAMf	yes	Thompson	Goddard	MYJ
C0	00Z NAMa	00Z NAMf	no	Thompson	Goddard	MYJ
N1	CN – arw_pert	21Z SREF arw_n1	yes	Ferrier	Goddard	YSU
P1	CN + arw_pert	21Z SREF arw_p1	yes	WSM6	Dudhia	MYJ
N2	CN – nmm_pert	21Z SREF nmm_n1	yes	Thompson	Goddard	MYJ
P2	CN + nmm_pert	21Z SREF nmm_p1	yes	WSM6	Dudhia	YSU
N3	CN – etaKF_pert	21Z SREF etaKF_n1	yes	Thompson	Dudhia	YSU
P3	CN + etaKF_pert	21Z SREF etaKF_n1	yes	Ferrier	Dudhia	MYJ
N3	CN – etaBMJ_pert	21Z SREF etaBMJ_n1	yes	WSM6	Goddard	MYJ
P4	CN + etaBMJ_pert	21Z SREF etaBMJ_n1	yes	Thompson	Goddard	YSU
CN2 (2km)	00Z ARPS Analysis	00Z NAMf	yes	WSM6	Goddard	MYJ

Two WRF-ARW, two WRF-NMM and four Eta members in the NCEP SREF were used to construct the perturbed ICs and LBCs for our storm-scale ensemble forecast (SSEF) system (Table 1). Ensemble members N1 and P1, N2 and P2 are pairs that used negative and positive IC and LBC perturbations from the WRF-ARW and WRF-NMM pairs of the SREF while N3, P3, N4 and P4 used the perturbations from the EtaKF and EtaBMJ members of SREF. Such perturbations were extracted from the SREF fields by taking the difference between the ensemble and control members then interpolating and adding the per-

turbation fields to the unperturbed IC and LBC. More details on the ensemble system design can be found in Kong et al. (2008).

Selected 2-D fields and soundings were extracted from the 3-D gridded output, and shipped to HWT for direct ingest into the N-AWIPS systems and for interactive manipulation and display by the forecast and evaluation teams. Additional post-processing and product generation from the ensemble output were also performed within the N-AWIPS. The 2-D data set, including hourly precipitation and composite reflectivity fields, are convenient and often sufficient for verification studies.

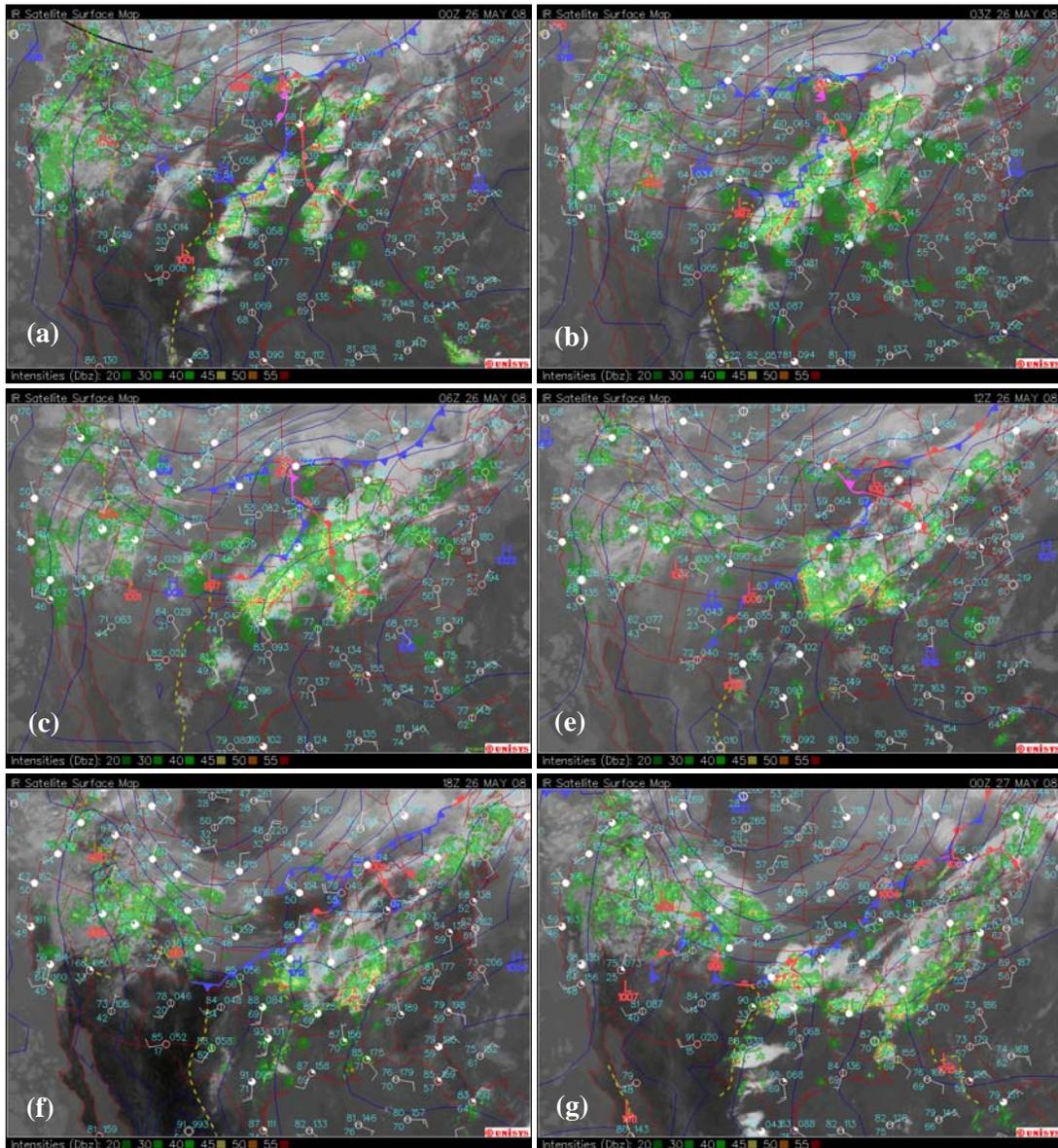


Fig. 2. Surface weather analyses with IR satellite imagery and radar reflectivity overlaid, for (a) 0000, (b) 0300, (c) 0600, (d) 1200, (e) 1800 UTC, 26 May 2008, and (f) 0000 UTC, 27 May 2008. Maps courtesy of Unisys through NCAR.

Parallel to the N-AWIPS system, graphical plotting and ensemble post-processing were also performed by CAPS, with hourly graphical products generated as soon as the model outputs were available and posted on the web at <http://www.caps.ou.edu/wx/spc>. These graphical products were produced using ARPSPLT, run in MPI mode, after WRF outputs were converted to the ARPS grid via WRF2ARPS.

During the 2008 HWT Spring Experiment, additional forecasts at 3 to 4 km resolutions were produced by NCAR, NSSL and EMC/NCEP for the same forecast periods. These forecasts acted as additional members of a larger ensemble, and the diversity of these forecasts provided additional probabilistic information for evaluation at HWT.

3. Forecast Examples and Subjective Evaluations

A more systematic evaluation of the performance of the 4-km ensemble predictions, in terms of the precipitation forecast skills, and the system's ability to produce reliable probability forecast, can be found in companion paper Kong et al. (2008). In this paper, we choose to demonstrate of the performance of the ensemble prediction system and the accompanying 2-km forecasts, and evaluate the impact of assimilating radar data, by presenting two cases. In the first case, quasi-linear convection existed at the initial condition time which evolved into a squall line that propagated away from a quasi-stationary front the initiated it. The squall line went through merger with another group of pre-existing cells. The case also contained secondary initiations triggered by outflow boundaries, and new initiation along the front and dryline on the second day. While the evolution of convective features in this case is rather complex, the forecasts with radar data captured many of the observed features while missing some of the others, and the positive impact of radar data are seen to last at least 24 hours. In the second case, an intense squall line developed ahead a propagating cold front and lasted for more than 36 hours. The squall line propagated ahead of the cold front, at a slightly faster speed but maintained its proximity to the front during the first 12 hours after initiation. The 30 hour forecasts started at 0000 UTC, 5 May and the forecasts initialized 24 hours later and valid at the same time are remarkably similar. The assimilation of radar data had little impact in this strongly forced case.

a. The 26 May 2007 case

At 0000 UTC, May 26, 2008, a low was centered over Minnesota, and a weak, quasi-stationary, cold front extended from the low center southwestward to the western Kansas border, where it intercepted a dryline that extended southward near eastern New Mexico border into northern Mexico. Fully developed

quasi-linear convection existed about 100 km east of the cold front in the Kansas region and at a further distance east of the dryline in western Texas (Fig. 2a). Over the next three hours, the convection evolved into a solid squall line that extended all the way from the Great Lakes into Texas panhandle (Fig. 2b). This squall line propagated eastward and maintained its identity up to 0000 UTC, May 27, when it was found over eastern Mississippi, northern Alabama and eastern Tennessee (Fig. 2g). At around 1800 UTC, the line went through a merge with another line ahead of it (Fig. 2f). In fact, this line was reminiscent of the convection that existed at the beginning of the period over south-central Missouri and north-central Arkansas (Fig. 2a). This line was eventually caught up by and absorbed into the fast propagating squall line. After the merge, the squall line strengthened between 2000 and 2300 UTC, and was still well defined by 0000 UTC, May 27 (Fig. 2g).

During the period, the cold front was more or less stationary; therefore the squall line was mostly self-propagating, driven by its own cold pool. The initial frontal forcing was therefore lost during its later propagation stage. This line quickly dissipated after 0000 UTC, May 27 (now shown), as it approached the high pressure ridge extending over the southeastern states.

An examination of the Central Plains through Mississippi Valley regions shows other convective activities during the period. By 1200 UTC, May 26, an almost north-south-oriented solid line of convection existed in eastern Kansas, between the main squall line and the cold front, along the westsouthwest edge of the clouds shown in satellite IR imagery (Fig. 2e). Animations of the images indicate that this line developed along the outflow boundary left by the first squall line and it maintained its line structure until 2200 UTC (Fig. 2f). After that time, this line decayed while new convection developed ahead of the quasi-stationary front that ran through the northeast and southwest corners of Kansas, and east of the dryline (Fig. 2g).

Between 2000 and 2100 UTC, a line of convective cells running from the northeastern corner of Arkansas through the central portion of the Oklahoma-Arkansas border developed, apparently along southeastward propagating outflow boundary originated from the convective line between the main squall line and the cold front. This line of cells is clearly seen at 0000 UTC, May 27 which extended into western Tennessee (Fig. 2g). Within Fig. 2, some of the outflow boundaries can be found as marked by the dashed yellow lines. Obviously, the evolution of convection during this 24 hour period was rather complex and many activities were modulated by the cold pools and gust fronts and their interactions. Such a situation is harder to predict than cases where strong synoptic scale features, such as a cold front, play a more controlling role.

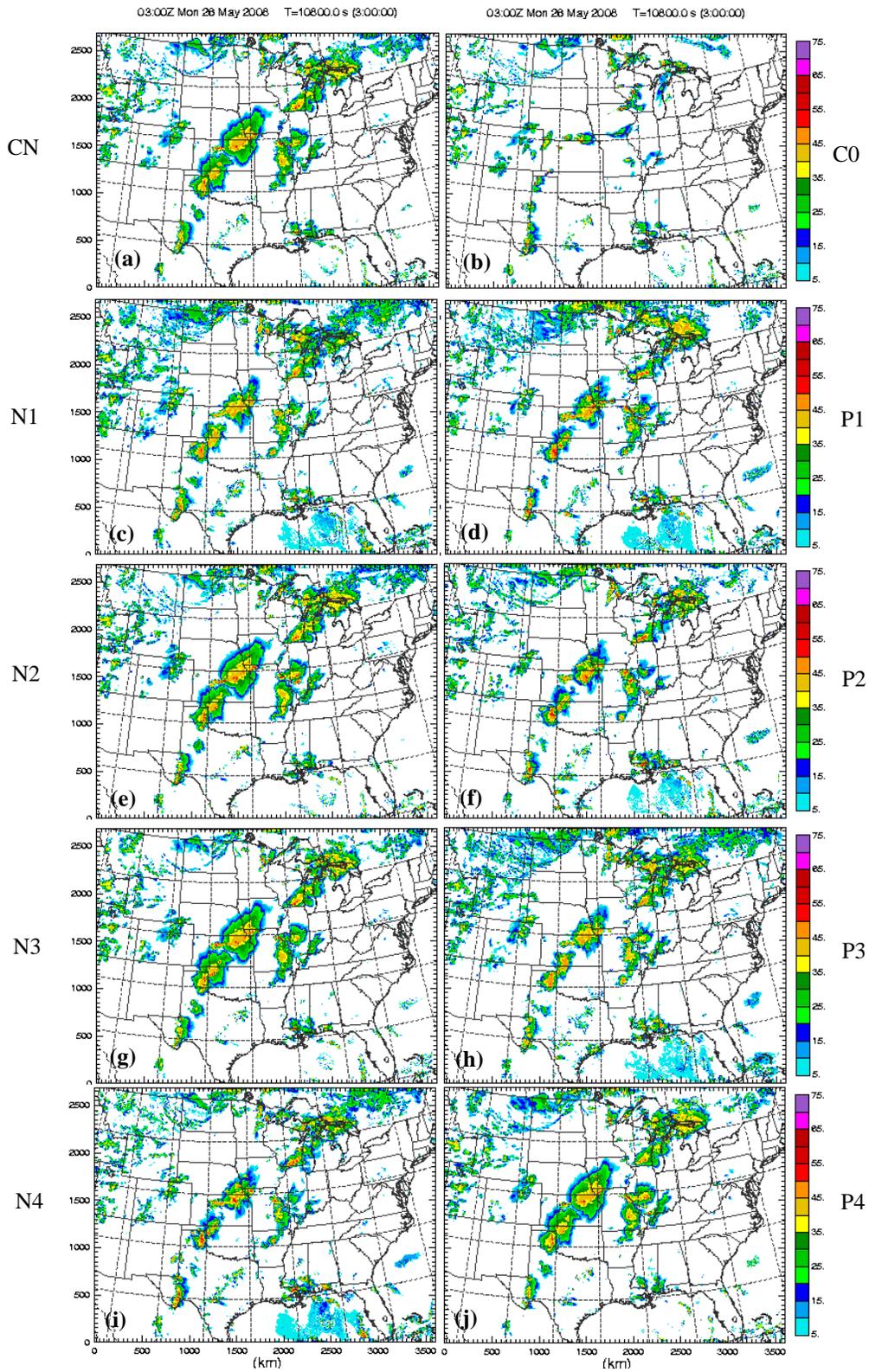


Fig. 3. Postage stamp view of 3-hour forecast composite reflectivity (Z_c) fields from ten members of the 4-km ensemble, as labeled, valid at 0000 UTC, 26 May 2008.

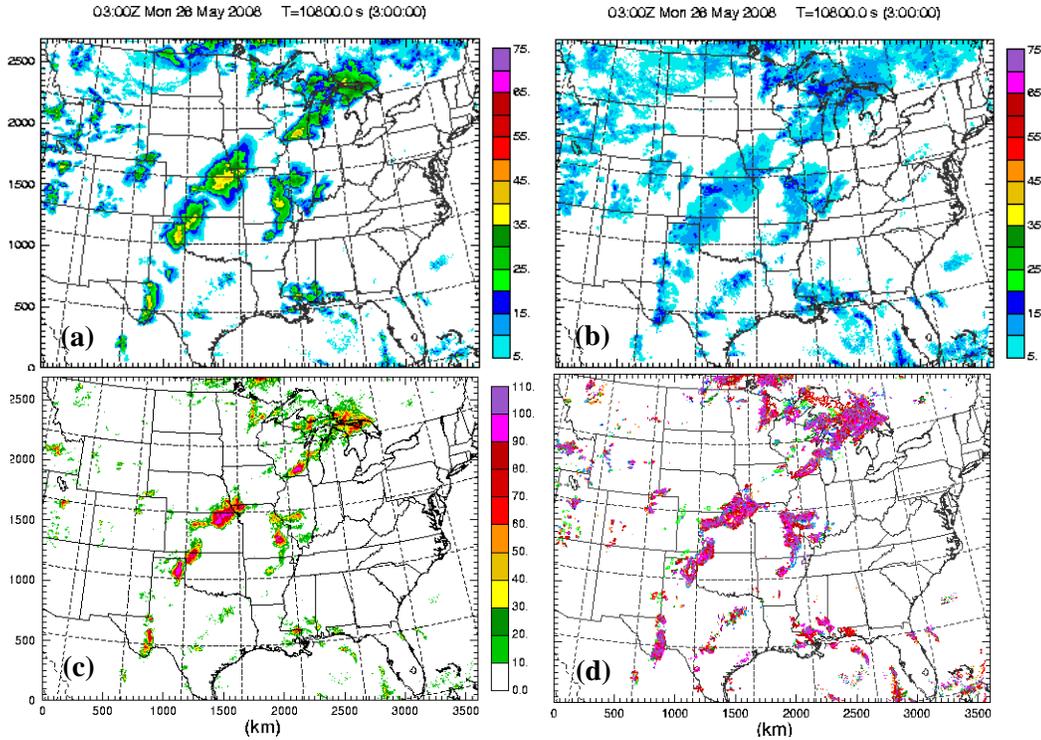


Fig. 4. Three-hour forecast ensemble mean Z_c (a), forecast ensemble spread (b), and ensemble-derived probability of Z_c exceeding 35 dBZ (c), and the 'spaghetti' plot of 35 dBZ Z_c contours, valid at 0300 UTC, 26 May 2008.

Next we discuss the forecasts by the 4-km ensemble and the 2-km high-resolution grid. Fig. 3 shows a postage stamp view of the 3-hour forecast composite (column maximum) reflectivity, Z_c , from the ensemble, valid at 0300 UTC, May 26. It can be seen that in all the forecasts that assimilated radar data, the convective line at this time is well predicted (c.f., Fig. 2b), and the variability in the forecast reflectivity across those ensemble members is relatively small. This is supported by the ensemble products shown in Fig. 4, where the ensemble average shows a very similar pattern as the individual members, though with smaller peak values due to averaging (Fig. 4a). The ensemble spread is generally smaller than 15 dBZ (Fig. 4b), and the simple frequency-based probability shows values between 70 and 80% in the strong echo regions along the line (Fig. 4c) while the spaghetti plots of 35 dBZ contours are mostly clustered together (Fig. 4d). These all indicate that the convective storms analyzed into the initial condition by the ARPS 3DVAR and cloud analysis procedure are retained well during the first three hours of forecast and the differences in the initial mesoscale and synoptic scale environment as introduced through SREF perturbations and the difference in model physics have had only small quantitative effects by this time.

Among the ensemble members, C0 that was directly initialized using NAM analysis, i.e., without

going through the storm-scale analysis and incorporating radar data, is an outlier. At this time, some weaker cells are found to have been initiated along the western Texas border where the dryline is found. Further north, cells are initiated along the cold front, while the pre-existing pre-frontal line of convection is missing in the forecast. Obviously, without knowledge of this pre-existing line that had been triggered by the cold front several hours earlier, the model was trying to initiate new convection along the frontal and dryline forcing zones. In the forecasts that included radar data, the model continued to evolve the existing storms.

The 12, 18 and 24 hour forecasts from CN and C0, and from CN2 are shown in Fig. 5 through Fig. 8. At 1200 UTC or the time of 12 hour forecast, most consistently predicted is the cluster of convection over southern Illinois close to St. Louis, which also existed at the initial time (e.g., Fig. 5b,c,f). The convection at the southwest end of the main squall line is also consistently forecast, especially in CN (Fig. 5c). The associated cells appear a little too strong in CN2. None of the forecasts captured the newly initiated northnorthwest-southsoutheast-oriented line that was initiated along the outflow boundary, however (Fig. 5e), indicating more difficulty in predicting secondary development (as a consequence of earlier convective activities). Without considering this line, the frequency pattern shown in

Fig. 5e actually resembles the observed reflectivity pattern in that region (Fig. 5a) quite well.

At 1800 UTC (Fig. 6), the CN2 and CN forecasts are showing correctly the main line catching up to the initially eastern region of convection that is now located near the Indian-Kentucky border. The main line propagated a little slower in CN2 (Fig. 6b) and more so in CN (Fig. 6c). In C0 that did not have radar data, the main squall line is actually missing; the model predicted a line that developed after the initial time along the cold front which is seen running from southeast Kansas through southern Illinois. This line is not the correct main squall line.

In almost all forecasts (Fig. 6e), a spuriously

strong line of convection developed in southeastern Texas around 1800 UTC, with that in CN2 being the strongest (Fig. 6b). The observed convection in the same region was weak at 1800 UTC (Fig. 6a) and dissipated over the next few hours (Fig. 7a). The spurious line maintained its intensity in most of the forecasts over the next 6 hours (Fig. 7 and Fig. 8), and affected the realism of the model solutions. Note that the convective cells near the eastern border of Mississippi in Fig. 7a and Fig. 8a developed in-situ instead of propagating into the region from western Texas, which occurred in CN2 (Fig. 8b). The reason for this discrepancy is worth investigating.

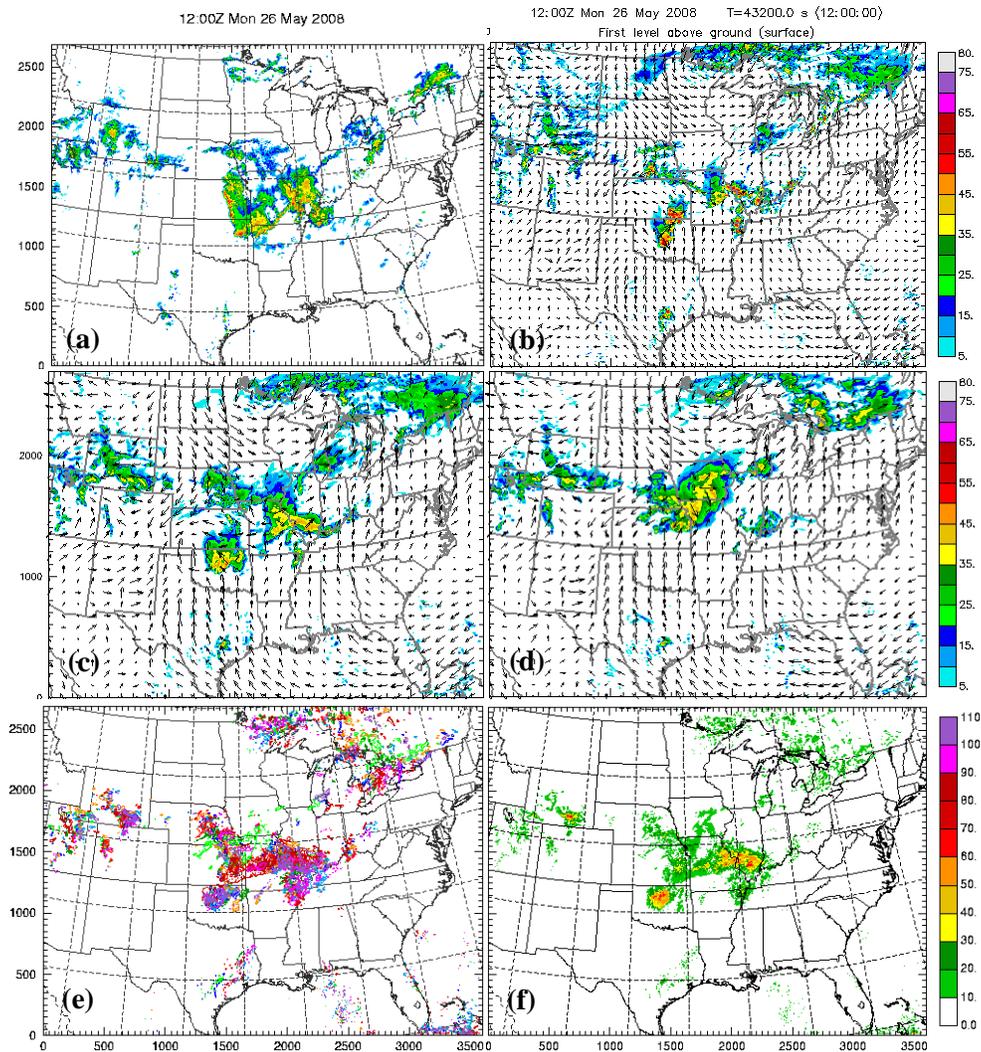


Fig. 5. Observed Z_c (a), 12-hour forecasts of (b) CN2, (c) CN, (d) C0, and (e) the 'spaghetti' plot of 35 dBZ Z_c contours, and (f) ensemble-derived probability of Z_c exceeding 35 dBZ, valid at 1200 UTC, May 26, 2008.

We point out here that the second development of a line of cells noted earlier in the observations across Arkansas is also captured in the 2 km solution (Fig. 7b and Fig. 8b), though with some orientation and timing error. The eastern end of it is too far north and the model initiation is about one hour too early (Fig. 7a and Fig. 7b). This line is captured in none of the 4 km ensemble members (Fig. 7c,d, Fig. 8c,d,e). In this respect, the 2 km forecast did the best.

Over the next 6 hours after 0000 UTC, the convective storms that propagated into the eastern U.S. all dissipated, in both the observations and in the model. New cells were initiated along the stationary cold front and dryline again during the late afternoon hours (before 0000 UTC), in about three regions at central Kansas, southwestern Oklahoma and southwestern Texas (Fig. 7a, Fig. 8a). Such initiations are captured nicely

in CN2 and CN, with CN2 doing a better job, especially with additional cells at north-central Oklahoma. The forecast of C0 remains poor at this time, indicating that the negative effect of lacking storm-scale radar information in the initial condition has carried over all the way through the 24 hours of forecast.

In summary, by assimilating radar data at a time when mature pre-front convection existed, both 2 and 4 km control forecasts significantly outperformed that of the 4 km forecast without radar data. The forecasts captured many of the structures and behaviors of the observed storms, including the propagation and evolution of two main groups of convection, and the initiation of a secondary line. But at the same time, the model missed a different secondary line development and spuriously predicted the intensification of another line.

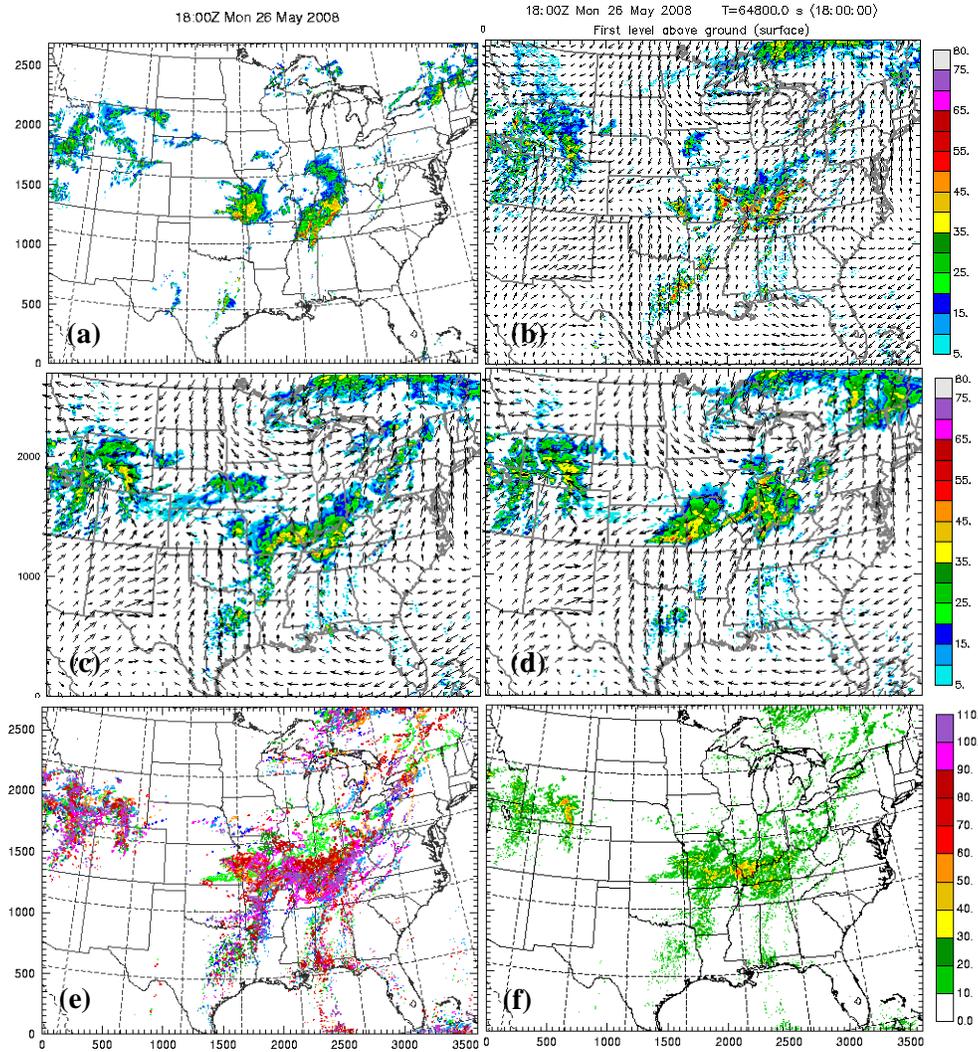


Fig. 6. As Fig. 5 but for observations and 18-hour forecast time valid at 1800 UTC, May 26, 2008.

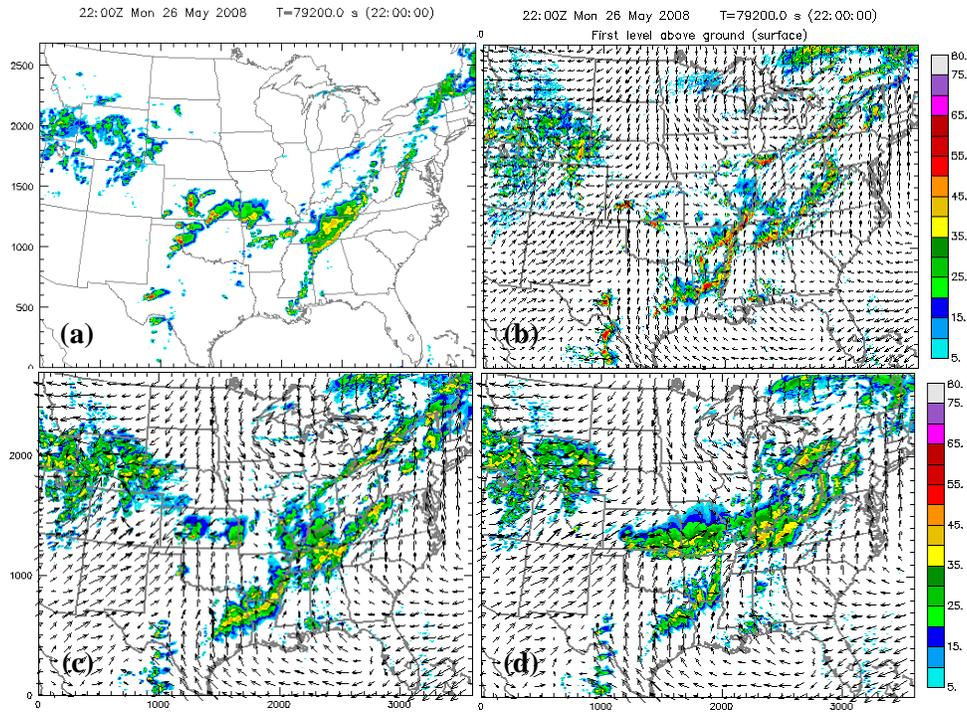


Fig. 7. As Fig. 5 but for observations and 22-hour forecast time valid at 1800 UTC, May 26, 2008, and without 'spaghetti' plot probability fields.

Due to the narrowness of the convective lines and cells, and spatial errors that can easily be larger than the width of the line and cells, especially at longer forecast ranges, quantitative verification scores will suffer when point-wise measures are used. The qualitative improvement in the forecasts with radar data and with the use of higher resolution may not be reflected in such verification scores. This points to the importance of more sophisticated verification methods and also the need for subjective evaluation of model forecasts. Fortunately, more researchers are working in this area.

b. The 5-6 June, 2008 case

For the 5-6 June 2008 case, we examine forecasts produced on two consecutive days. In fact, the forecasts of 6 June 2008 were added at the end of spring 2008 forecast experiment due to interesting weather. In this case, a convective line was initiated ahead of a strong cold front at about 1800 UTC, 5 June. It evolved into a strong squall line that lasted more than 36 hours; it maintained its proximity to the cold front during the first 12 hours after initiation (Fig. 9).

The 21 hour forecasts from CN2, CN and C0 are shown in Fig. 10, together with the observations. By this time, the observed squall line is well established (Fig. 10a) while that in the model forecasts are being organized. The line in CN2 is best established at this time (Fig. 10b) while those in CN and C0 are some-

what weaker, indicating some delay in the model initiation of this squall line. The forecasts of CN and C0 are very similar, indicating little effect of assimilating radar data 21 hours earlier. Actually, there was no precipitation in the Kansas, Oklahoma and Texas region at 0000 UTC, 5 June, while precipitation present at that time either dissipated or moved out of the domain before 2100 UTC.

By 0000 UTC, 6 June, an intense solid squall line had been established in the observations (Fig. 11) that ran from the northwest corner of Kansas through the southwest corner of Oklahoma. This is well forecast by the 2 km grid (Fig. 10b), and by CN and C0 (Fig. 10c,d). The southern portion of the line is too far east though, with a displacement error of about 100 km. The position forecasts by the ensemble members are very similar, as shown by the spaghetti (Fig. 10e) and ensemble spread plots (Fig. 10f). The maximum spread along the line is generally less than 15 dBZ even though all forecasts predicted significant reflectivity values along the line.

The predicted squall line continued to propagate eastward. By 0600 UTC, 6 June, the position error is actually smaller in the three forecasts (Fig. 12b,c,d). The 2 km forecast has the best intensity and structure, and a weak segment along the line in south-central Oklahoma is also captured in the forecasts. Overall, these forecasts of a squall line that was initiated 18 hours into the forecast are excellent in terms of its posi-

tion and timing, although again due to the narrowness of the line and the differences in detailed structures, point-wise verification against high-resolution radar-derived reflectivity or precipitation fields will not necessarily yield high scores.

The 6-hour forecasts started at 0000 UTC, 6 June, valid at the same time as those in Fig. 12 are shown in Fig. 13. The forecast squall lines in CN2, CN and C0 are all accurately positioned although that in CN2 appears too strong in terms of reflectivity while that in C0 appears too weak. The overall forecasts are remarkably similar to those initialized 24 hours earlier, and the differences appear to be within the range of uncertainties within the forecast ensemble. Clearly, for this strongly forced case, the predictability of the squall line is rather long.

To see the spin up process of the squall line in the forecast that did not include radar data in the initial

condition, i.e., in C0, 1-hour forecasts valid at 0100 UTC are shown in Fig. 14. It can be seen the line is already half way spun up in the first hour. This is perhaps not surprising given the very strong convergence in the wind fields along the squall line.

The three 24-hour forecasts initialized at 0000 UTC, 6 June are shown in Fig. 15. Again the general structures of the squall line are predicted well in all three forecasts, with that in CN2 being somewhat stronger than observed and those in CN and C0 somewhat weaker in terms of reflectivity. In all three cases, the line propagated too fast, resulting 100-150 km position error at this 24 hour forecast range. By 0600 UTC, 7 June, only the southern portion of the squall line remained and it turned into a westsouthwest – eastnortheast orientation (not shown). All forecasts position the line too far north, by 100 – 300 km.

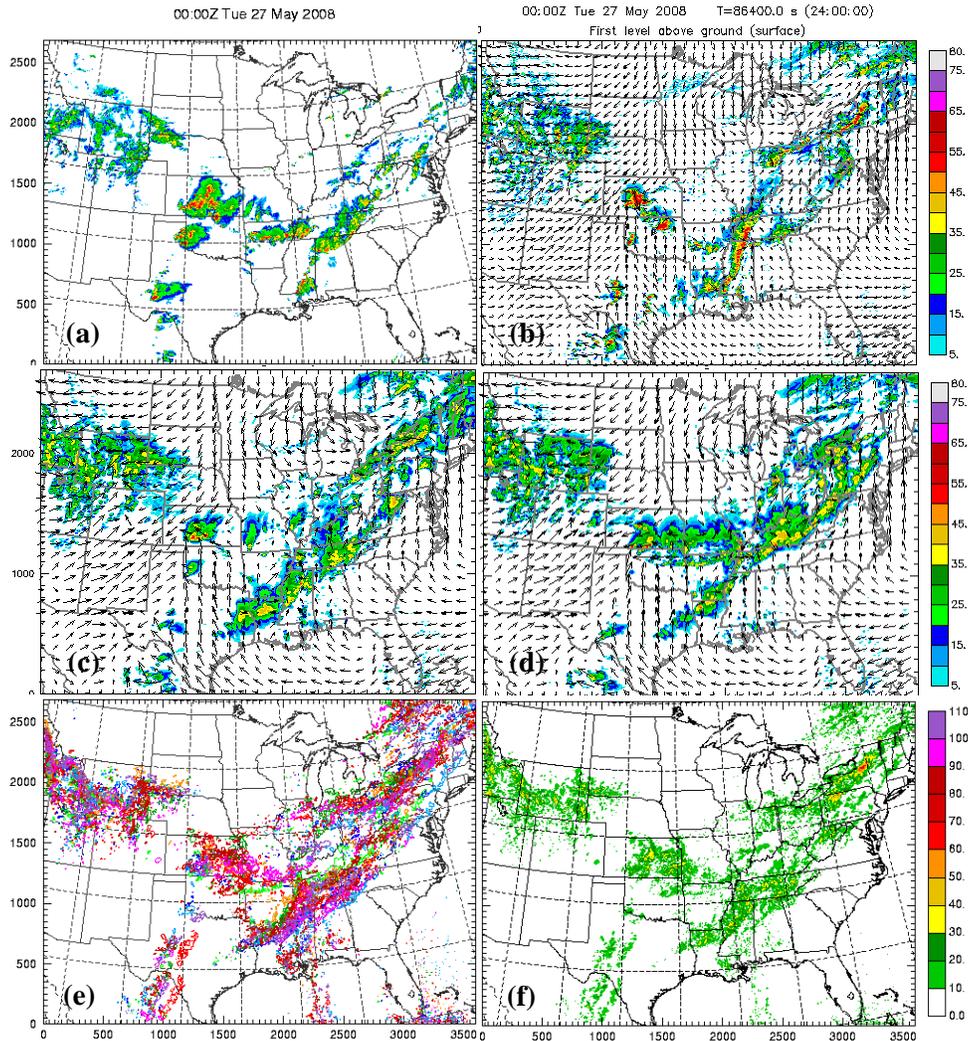


Fig. 8. As Fig. 5 but for observations and 24-hour forecast time valid at 000 UTC, May 27, 2008.

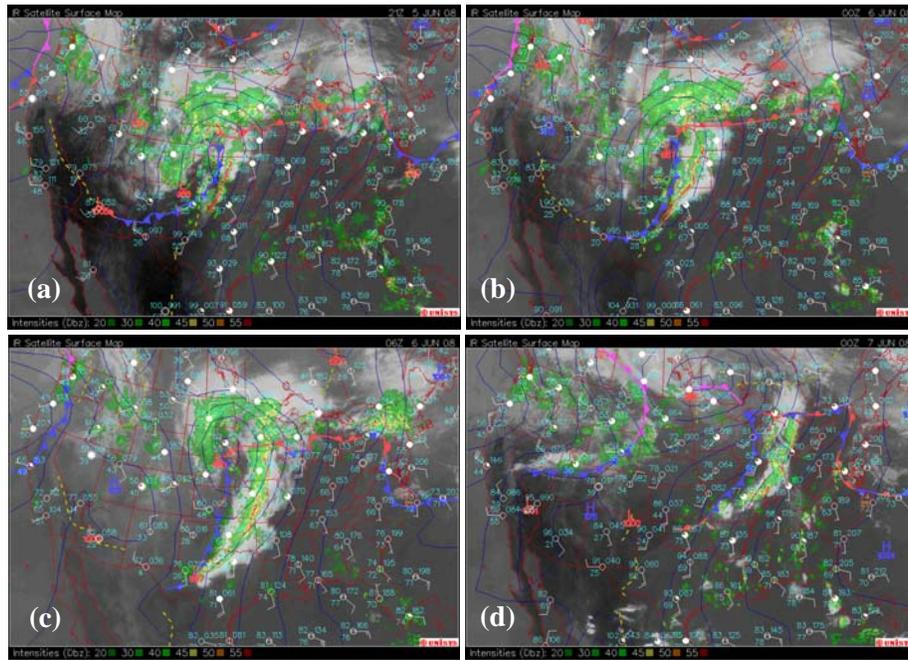


Fig. 9. Surface weather analyses with IR satellite imagery and radar reflectivity overlaid, for (a) 2100 UTC, June 5, 2008 and, (b) 0000 and (c) 0600 UTC, June 6, and (d) 0000 UTC, June 7, 2008. Maps courtesy of Unisys through NCAR.

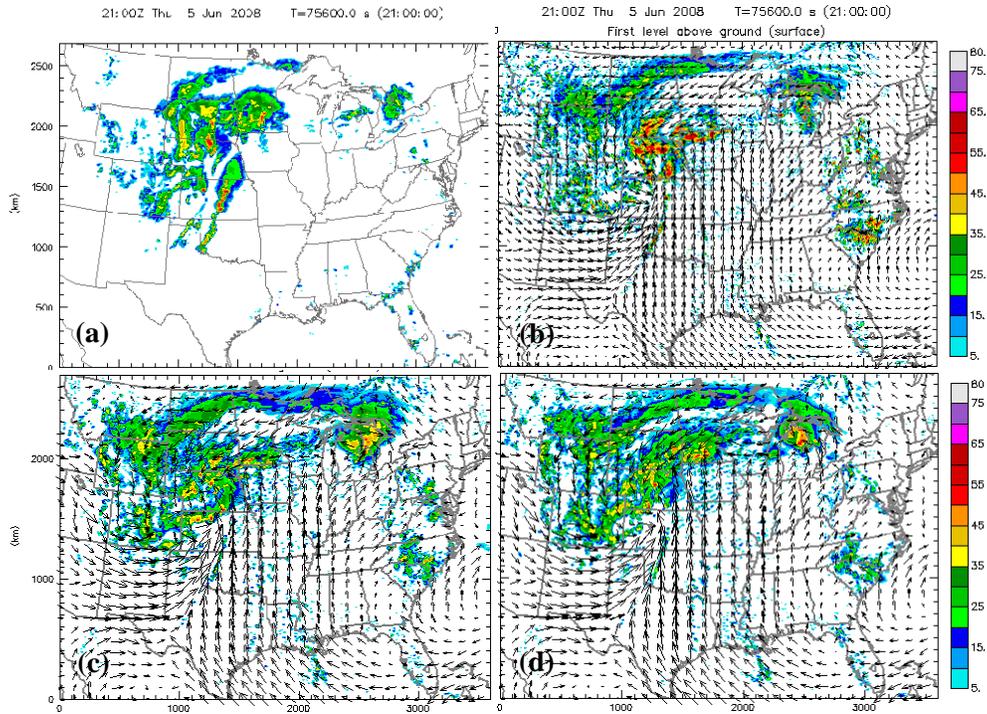


Fig. 10. Observed Z_c (a), 21-hour forecasts of (b) CN2, (c) CN, and (d) C0, and (e) valid at 2100 UTC, June 5, 2008.

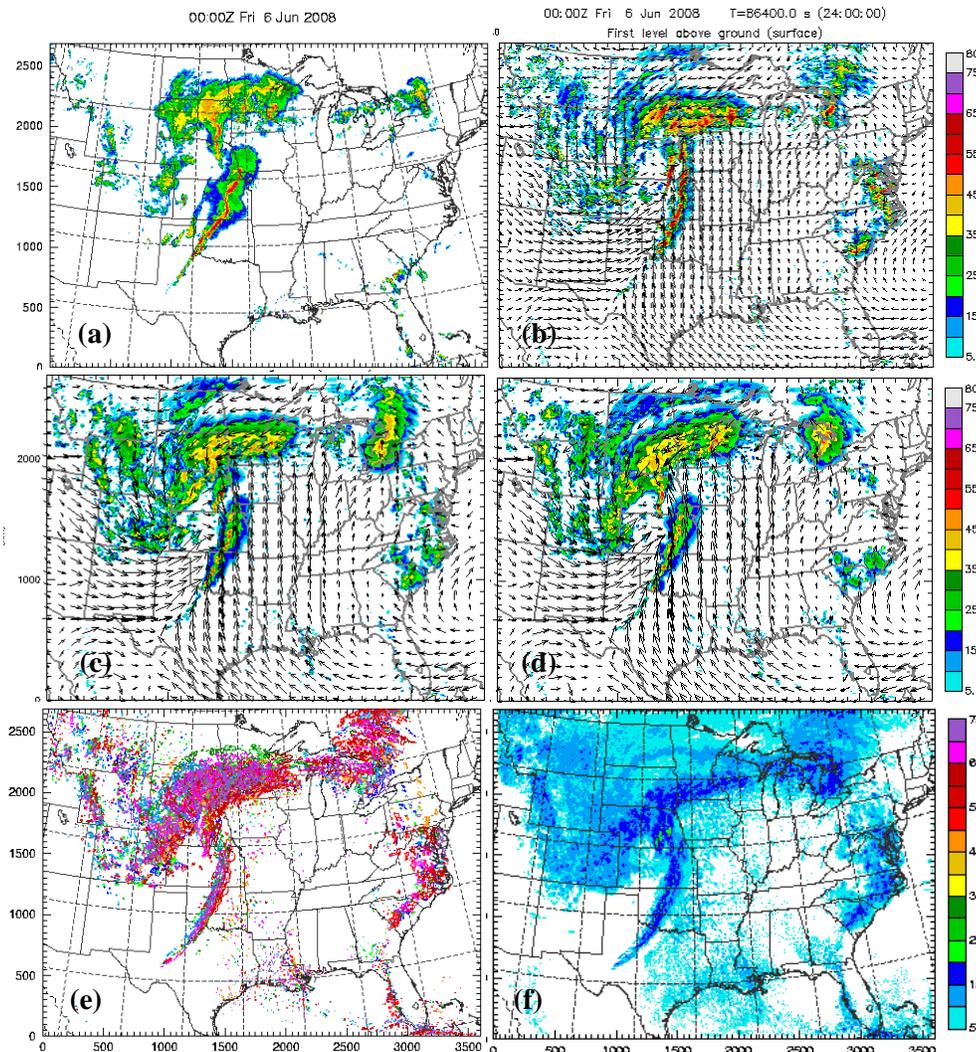


Fig. 11. As Fig. 10 but for 24-hour forecasts valid at 0000 UTC, June 6, 2008 plus (e) the 'spaghetti' plot of 35 dBZ Zc contours, and (f) the ensemble spread valid at the same time.

In summary, we presented and discussed two rather different examples where the predictability of the weather systems and the sensitivity of the forecasts to initial conditions and resolutions are very different. Whether the convective system is controlled mainly by large-scale features or by the smaller-scale internal dynamics of the convective systems is clearly the main distinguishing factor. For these reasons, our investigation on and the understanding of sensitivity, predictability and data impact should take into account of the differences in the weather systems. Subjective evalua-

tion and process studies remain necessary for the ultimate goal of improving weather prediction.

We note that the second case discussed here is somewhat similar to the 23-24 May 2007 case discussed in Xue et al. (2007), from the 2007 spring experiment, in terms of the synoptic scale control. Finally, unlike during 2007, the composite reflectivity fields shown here were calculated using formulas consistent with the microphysics schemes used in the individual forecast members.

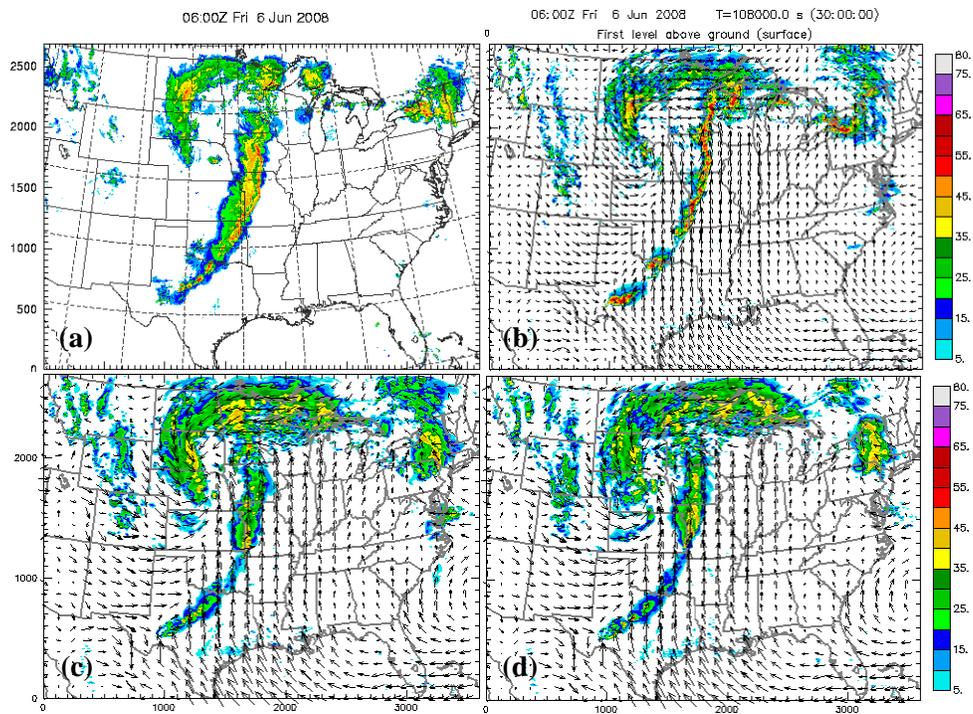


Fig. 12. As Fig. 10 but for 30-hour forecasts valid at 0600 UTC, June 6, 2008.

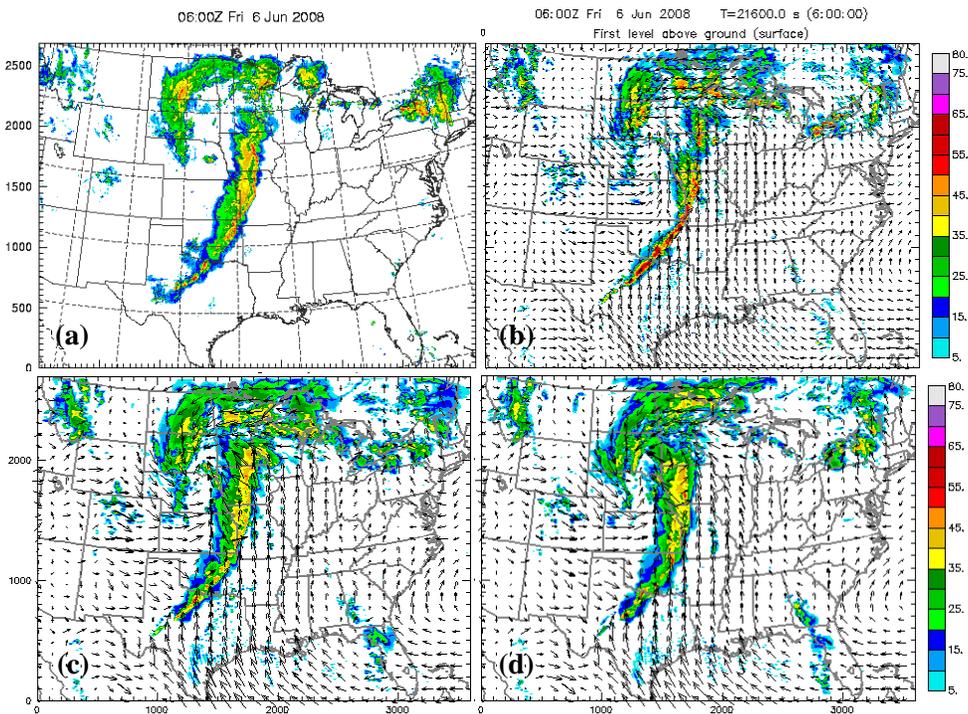


Fig. 13. As Fig. 12 for 6-hour forecasts starting from valid as the same time.

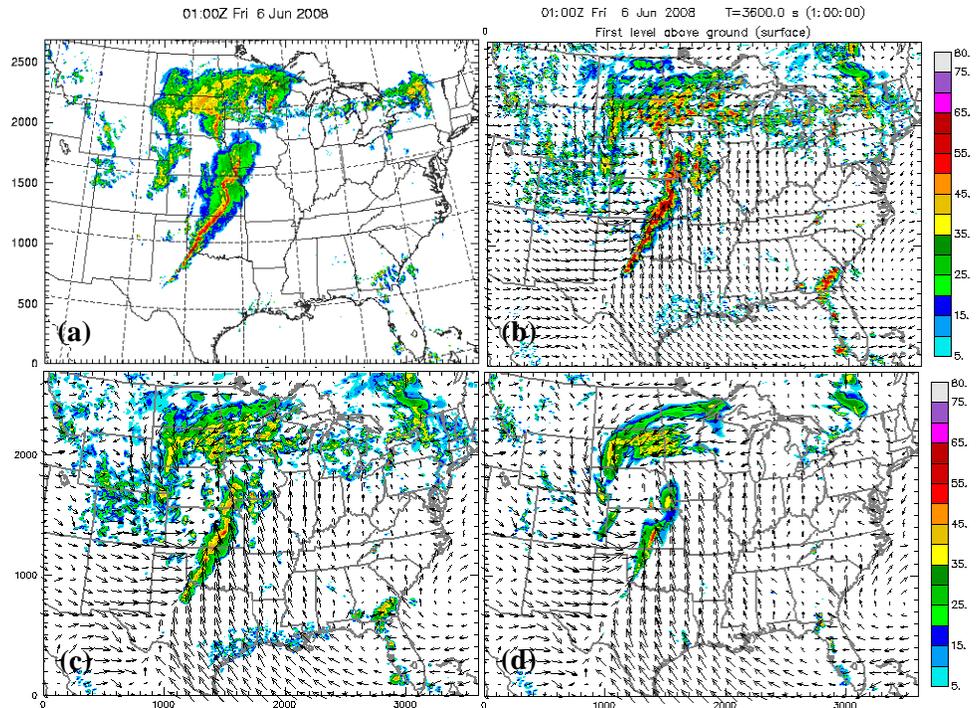


Fig. 14. As Fig. 13 but for 1-hour forecasts valid at 0100 UTC, June 6, 2008.

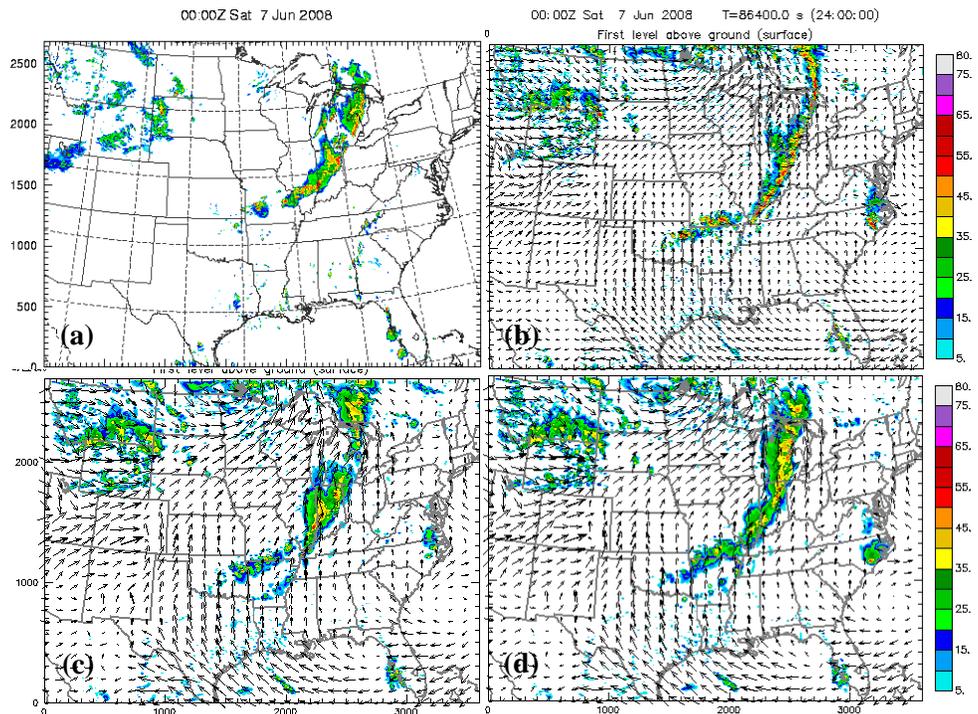


Fig. 15. As Fig. 13 but for 24-hour forecasts valid at 0000 UTC, June 7, 2008.

4. Precipitation Verification

Despite the limitations of standard verification scores noted earlier, they may still provide some general assessment on the forecast quality. While more approaches have been taken in companion studies (e.g., Clark et al. 2008; Kong et al. 2008; Schwartz et al. 2008a; Schwartz et al. 2008b), we will show here only standard Equitable Threat Scores (ETSs) and bias scores, mainly for the purpose of examining the value of radar data assimilation.

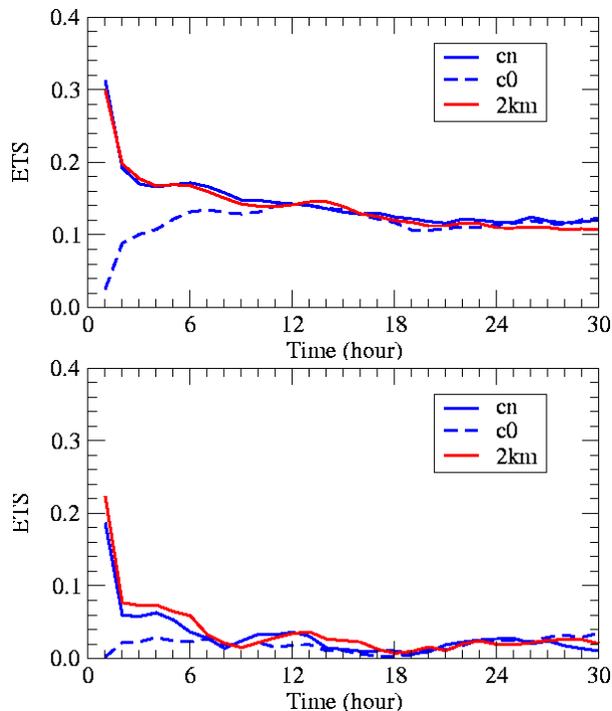


Fig. 16. Equitable threat scores (ETS) of hourly precipitation at 0.01 inch (upper panel) and 0.5 inch (lower panel) thresholds for the 4-km control (cn, solid blue), 4-km run without radar data (c0, dashed blue), and the 2km forecasts (2km, solid red), averaged over the last 15 days of 2008 CAPS forecasts.

Fig. 16 shows the ETSs for the 0.01 and 0.5 inch/hour precipitation thresholds, for CN, C0 and CN2, averaged over the last 15 days of forecasts from 2008. Immediately evident is that the forecasts (CN and CN2) starting from radar initialized initial conditions start with high ETSs that decrease with the forecast range while the forecasts without radar data (C0) start with a zero ETS and it takes more than 9 hours to their scores to reach a similar level as those of radar cases. The differences after 12 hours become negligibly small, which can be due to both the limitations of the verification method and the predictability of convective scale features. The 1-km-resolution radar-based precipitation estimates produced by the National Severe Storms Lab

were used for the verification, and the detailed structures present in the high-resolution data do not favor higher resolution forecasts of the 2 km grid, which may be a reason why the 2 km scores are lower at some of the forecast ranges. The higher precipitation forecast bias at 2 km resolution shown in Fig. 17 can be another reason. High precipitation bias is a common problem with such convection-allowing-resolution forecasts, and the exact source of the error remains unknown.

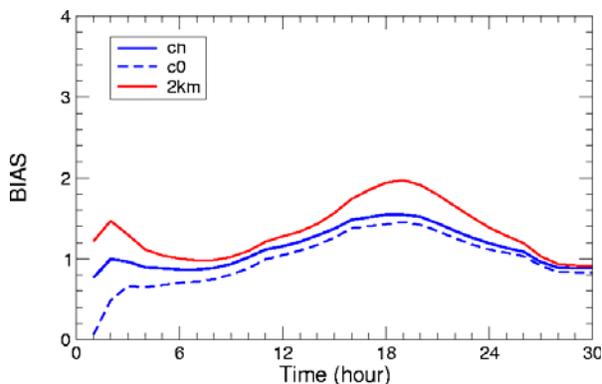


Fig. 17. As Fig. 16 but for the precipitation bias at 0.01 inch threshold.

5. Future Plan

The output from the ensemble and high-resolution forecasts, saved at hourly intervals, for more than 6 weeks from spring 2008, as well as a similar data set from 2007, provide us an unprecedented opportunity for investigating many aspects of convective-scale prediction. Quantitative as well as qualitative evaluations of the forecasts will continue and collaborations exploiting this valuable data set are welcome.

The CAPS forecast experiment will continue in the springs of 2009 and 2010, as part of the HWT and in support of the VORTEX-II field experiment. For the spring of 2009, the planned enhancements include the addition of WRF-NMM model and improved assimilation of radar data. Our current plan is to double the ensemble size, with half of the members using WRF-ARW and the other half using WRF-NMM. We may also add a few members using the ARPS model. In the future years, pending availability of computational resources, we will attempt 1 km realtime forecasts over the CONUS domain, assimilating radar data at the same resolution. It is our hypothesis that a 1 km resolution will be able to better take advantage of the high-resolution radar data, to much better resolve the internal structures of convective storms, and provide more accurate guidance for high-impact parameters such as strong low-level winds and rotation. It will also help answer the question as to how much resolution is enough (Kain et al. 2007; Schwartz et al. 2008a).

Acknowledgement: This research was mainly supported by a grant to CAPS from the NOAA CSTAR program. Supplementary support was also provided by NSF ITR project LEAD (ATM-0331594). Suggestions and input from NCAR scientists Drs. Jimmy Dudhia, Morris Weisman, Greg Thompson and Wei Wang on the WRF model configurations were every helpful. The realtime predictions were performed at the Pittsburgh Supercomputing Center (PSC). David O’Neal of PSC provided ever present technical and logistic support that was essential to the success of the forecast experiment. Support from Dr. James Kimpel of NSSL, Joseph Schaeffer of SPC and Geoff DiMego of NCEP/NMC are also greatly appreciated.

References

- Brewster, K., 2002: Recent advances in the diabatic initialization of a non-hydrostatic numerical model. *Preprints, 15th Conf on Numerical Weather Prediction and 21st Conf on Severe Local Storms*, San Antonio, TX, Amer. Meteor. Soc., J6.3.
- Clark, A. J., W. A. Gallus, Jr., M. Xue, and F. Kong, 2008: A comparison of precipitation forecast skill between small near-convection-permitting and large convection-parameterizing ensembles. *Wea. and Forecasting*, Submitted.
- Coniglio, M. C., J. S. Kain, S. J. Weiss, D. R. Bright, J. J. Levit, G. W. Carbin, K. W. Thomas, F. Kong, M. Xue, M. L. Weisman, and M. E. Pyle, 2008: Evaluation of WRF model output for severe-weather forecasting from the 2008 NOAA Hazardous Weather Testbed Spring Experiment. *24th Conf. Several Local Storms*, Savannah, GA, Amer. Meteor. Soc., 12.4.
- Du, J., J. McQueen, G. DiMego, Z. Toth, D. Jovic, B. Zhou, and H. Chuang, 2006: New dimension of NCEP Short-Range Ensemble Forecasting (SREF) system: Inclusion of WRF members. *Preprint, WMO Expert Team Meeting on Ensemble Prediction System*, Exeter, UK, 5pp.
- Gao, J.-D., M. Xue, K. Brewster, and K. K. Droegemeier, 2003: A 3DVAR method for Doppler radar wind assimilation with recursive filter. *31st Conf. Radar Meteor.*, Seattle, WA, Amer. Meteor. Soc.
- 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.
- 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.
- 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, 2007: Some practical considerations for the first generation of operational convection-allowing NWP: How much resolution is enough? *Preprints, 22nd Con. Wea. Ana. Forecasting/18th Conf. Num. Wea. Pred.*, Park City, UT, Amer. Meteor. Soc., CD-ROM 3B5.
- Kong, F., M. Xue, D. Bright, M. C. Coniglio, K. W. Thomas, Y. Wang, D. Weber, J. S. Kain, S. J. Weiss, and J. Du, 2007: Preliminary analysis on the real-time storm-scale ensemble forecasts produced as a part of the NOAA hazardous weather testbed 2007 spring experiment. *22nd Conf. Wea. Anal. Forecasting/18th Conf. Num. Wea. Pred.*, Salt Lake City, Utah, Amer. Meteor. Soc., CDROM 3B.2.
- 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 forecasting during the 2008 spring experiment. *24th Conf. Several Local Storms*, Savannah, GA, Amer. Meteor. Soc., 12.3.
- Schwartz, C., J. Kain, S. Weiss, M. Xue, D. Bright, F. Kong, K. Thomas, J. Levit, and M. Coniglio, 2008a: Next-day convection-allowing WRF model guidance: A second look at 2 vs. 4 km grid spacing. *24th Conf. Several Local Storms*, Savannah, GA, Amer. Meteor. Soc., P10.5.
- Schwartz, C. S., J. S. Kain, S. J. Weiss, M. Xue, D. R. Bright, F. Kong, K. W. Thomas, J. J. Levit, M. C. Coniglio, and M. S. Wandishin, 2008b: Toward improved convection-allowing ensembles: model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *24th Conf. Several Local Storms*, Savannah, GA, Amer. Meteor. Soc., 13A.6.
- 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. *Meteo. Atmos. Phys.*, **94**, 167-183.
- Weiss, S. J., J. S. Kain, D. R. Bright, J. J. Levit, G. W. Carbin, M. E. Pyle, Z. I. Janjic, B. S. Ferrier, J. Du, M. L. Weisman, and M. Xue, 2007: The NOAA Hazardous Weather Testbed: Collaborative testing of ensemble and convection-allowing WRF models

- and subsequent transfer to operations at the Storm Prediction Center. *22nd Conf. Wea. Anal. Forecasting/18th Conf. Num. Wea. Pred.*, Salt Lake City, Utah, Amer. Meteor. Soc., CDROM 6B.4.
- 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., 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., F. Kong, D. Weber, K. W. Thomas, Y. Wang, K. Brewster, K. K. Droegemeier, J. S. K. S. J. Weiss, D. R. Bright, M. S. Wandishin, M. C. Coniglio, and J. Du, 2007: CAPS realtime storm-scale ensemble and high-resolution forecasts as part of the NOAA Hazardous Weather Testbed 2007 spring experiment. *22nd Conf. Wea. Anal. Forecasting/18th Conf. Num. Wea. Pred.*, Salt Lake City, Utah, Amer. Meteor. Soc., CDROM 3B.1.