Preprints, 22th Conf. on Weather Analysis and Forecasting
and 18th Conf. on Numerical Weather Prediction
Amer. Meteor. Soc., Park City, UT, 25-29 June 2007

3B.2 PRELIMINARY ANALYSIS ON THE REAL-TIME STORM-SCALE ENSEMBLE FORECASTS PRODUCED AS A PART OF THE NOAA HAZARDOUS WEATHER TESTBED 2007 SPRING EXPERIMENT

Fanyou Kong[1]*, Ming Xue[1,2], Kelvin K. Droegemeier[1,2], David Bright[3], Michael C. Coniglio[4],
Kevin W. Thomas[1], Yunheng Wang[1],
Dan Weber[1], Jack Kain[5], Steven J. Weiss[3], and Jun Du[6]
[1]Center for Analysis and Prediction of Storms, and [2]School of Meteorology,
University of Oklahoma, Norman, OK 73072
[3]NOAA/NMS/NCEP Storm Prediction Center
[4]CIMMS/University of Oklahoma
[5]NOAA National Severe Storm Laboratory, Norman, OK 73072
[6]NOAA/NWS/NCEP, Camp Springs, MD 20746

1. INTRODUCTION

The prediction of convective-scale (or storm-scale) hazardous weather is very important from both meteorological and public service/societal impact perspectives. However, accurate prediction of such weather continues to be a major challenge. Various sources of uncertainties associated with numerical weather prediction (NWP) as well as the high-nonlinearity of the weather systems at the convective-scales render probabilistic forecast information afforded by high-resolution ensemble forecasting systems especially valuable to operational forecasters.

Ensemble forecasting has proven valuable in medium-range global model forecasts (6-10 days) (Kalnay 2003). Short-range ensemble forecasting (SREF, ~40 km resolution, 1-3 days) with limited-area models has been underway for some time (Brooks et al. 1995; Du and Tracton 2001; Hamill et al. 2000; Hou et al. 2001). Owning to the difficulties associated with accurate prediction of severe convective storm weather, there are fast growing interests in storm-scale ensemble forecasting (SSEF) studies (e.g.,; Elmore et al. 2003; Walsar et al. 2004; Kong et al. 2006, 2007). Nevertheless, the effectiveness of the stochastic-dynamic approach (*aka* ensemble forecasting) on the storm-scale has yet to be fully explored.

A real-time storm-scale ensemble forecasting experiment is currently underway as part of the NOAA Hazardous Weather Testbed (HWT) 2007 Spring Experiment. At 4-km horizontal grid spacing, the WRF-ARW-based ensemble system, developed at the Center for Analysis and Prediction of Storms (CAPS), runs daily for 33 hours from mid April through early June, for a domain covering the eastern 2/3 of the continental U.S (Figure 1). This pilot system consists of ten hybrid perturbation members that consist of a combination of perturbed initial conditions and various microphysics and PBL physics parameterization schemes. Close collaborations among forecasters and scientists from CAPS, the Storm Prediction Center (SPC), the Aviation Weather Center (AWC), the Hydrometeorological Prediction Center (HPC), the Environmental Modeling Center (EMC/NCEP), the National Severe Storms Laboratory (NSSL), the NWS Norman Weather Forecast Office (WFO), the NWS Southern Region Headquarters, and the Pittsburgh Supercomputing Center (PSC) make this unprecedented experiment happen.

In addition to traditional ensemble products widely used in large-scale and mesoscale ensemble forecasting systems, such as the mean, spread, and probability of selected forecast fields, emphases are given to the generation and assessment of products specific to storm-scale, cloud-resolving ensemble forecasts. Such products include but are not limited to: probability of storm type (e.g., linear vs. cellular), large hail probability, icing potential (high super-cooled water content probability), damaging wind gusts at surface, reflectivity exceedance, updraft rotation, and supercell thunderstorm detection in the form of probability or joint probability for Supercell Composite Parameter, Significant Tornado Parameter, Supercell Detection Index, and Updraft Helicity. Many of these products are created in real time through existing capabilities in the SPC version of the N-AWIPS system for the use and evaluation by researchers and operational forecasters during the experiment. The statistical consistency of the ensemble system, in terms of spread-error relation, will be assessed using the entire two-months of data after the experiment. The performance of the ensemble forecasts, in terms of quantitative skill scores, is evaluated and will be compared with the NCEP operational SREF and 12 km NAM forecasts, and a CAPS 2-km WRF forecast over the same domain and period. Skill scores for sub-groups of the ensemble will also be examined to assess the effectiveness of initial condition and physics perturbations.

This extended abstract mainly describes the basic design of the storm-scale WRF-ARW ensemble forecasting system, and presents some preliminary forecast results and analysis. The experiment is still underway and comprehensive post analysis and findings from the study will be presented at the conference and for journal publication.

2. EXPERIMENT HIGHLIGHT

As the first year of the three-year project, the 2007 Spring Program began on 15 April 2007 and will end on

**Corresponding Author Address*: Dr. Fanyou Kong, Center for Analysis and Prediction of Storms, Univ. of Oklahoma, Norman, OK 73072; e-mail: fkong@ou.edu

8 June. All experimental forecasts are generated with the Weather Research and Forecast (WRF) Advanced Research WRF (ARW) model. The NAM analyses available on the 12 km grid are used for initialization with the initial condition perturbations for the ensemble coming from the NCEP Short-Range Ensemble (SREF). Local and WSR-88D data will be added in the future years. All model code, initial conditions, and forecast output at hourly intervals are archived at the Pittsburgh Supercomputing Center (PSC) Mass storage facility and will later be made available to the national weather community. Figure 1 shows the coverage area of the model domain, with terrain height info in color shading.

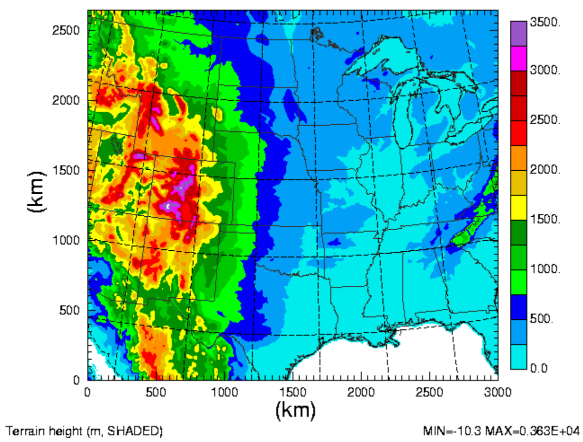

Figure 1. Model domain coverage, with terrain height in color shading.

The daily 33 h ensemble forecast starts at 2100 UTC and ends at 0600 UTC of the third day. The ensemble configuration includes ten hybrid members, all of which are run on *BIGBEN*, a NSF TeraGrid resource (Cray XT3) at PSC. The forecasts are initialized using the 21 UTC 12 km NAM (218 grid) analysis and the SREF WRF member output each day. Model execution begins around 0130 UTC (20:30 local time) and finish in about 6-10 hours (depending on members and convection activities), using about 770 CPUs, with results being processed as they become available. This allows model output to inform the morning (1200 UTC) SPC convective outlook as well as provide forecast information about convective activity during the subsequent afternoon through the following morning.

The ensemble initial conditions consist of a mixture of bred perturbations coming from 2100 UTC SREF WRF perturbed members and physics variations (grid-scale microphysics, land-surface and PBL physics), along with a control run. The lateral boundary conditions come from the corresponding 2100 UTC SREF WRF forecasts directly for those perturbed members and from the 1800 UTC 12 km NAM forecast for the non-perturbed members (control and physics variation members).

Table 1 outlines the basic configuration for each individual members. CN refers to the control member, N1, N2, P1, P2 are four initial perturbation members, PH1 – PH5 are members with only physics variations compared to CN. NAMa and NAMf refer to 12 km NAM analysis and forecast, respectively.

Table 1. Ensemble member configuration

member	IC	LBC	mp_physics	sf_sfclay_physics	bl_pbl_physics
CN	21Z NAMa	18Z NAMf	WSM 6-class	Janjic Eta	MYJ
N1	CN – em_pert	21Z SREF em-n1	Ferrier	Janjic Eta	MYJ
P1	CN + em_pert	21Z SREF em-p1	Thompson	Janjic Eta	MYJ
N2	CN – nmm_pert	21Z SREF nmm-n1	Thompson	Monin-Obukhov	YSU
P2	CN + nmm_pert	21Z SREF nmm-p1	WSM 6-class	Monin-Obukhov	YSU
PH1	21Z NAMa	18Z NAMf	Thompson	Janjic Eta	MYJ
PH2	21Z NAMa	18Z NAMf	Ferrier	Janjic Eta	MYJ
PH3	21Z NAMa	18Z NAMf	WSM 6-class	Monin-Obukhov	YSU
PH4	21Z NAMa	18Z NAMf	Thompson	Monin-Obukhov	YSU
PH5	21Z NAMa	18Z NAMf	Ferrier	Monin-Obukhov	YSU

Four SREF WRF members are used to provide initial perturbations and LBCs for the four perturbed members, two from WRF-ARW subgroup and two from WRF-NMM subgroup.

For all members, the short- and long-wave radiation schemes are RRTM and Goddard, respectively, and the surface physics uses Noah scheme (see WRF manual for detail description for all physics schemes).

Selected 2D fields including sea level pressure, 10 m winds, 2 m temperature and dew point, composite reflectivity, 1 h accumulated precipitation etc. are written out both in GEMPAK and binary formats. The GEMPAK data are directly transferred from PSC to SPC N-AWIPS system. The binary data, as well as all model raw output data (in ARPS format) are archived in the mass storage facility at PSC. CAPS makes available a separate webpage demonstrating the real-time ensemble forecast products (http://www.caps.ou.edu/wx/spc).

3. PERFORMANCE OF THE ENSEMBLE SYSTEM

This ensemble system, as illustrated in Table 1, can be analyzed as three sub-ensembles: A full ten-member ensemble (FULL), a four-member perturbed ensemble (PERT), and a six-member physics variation ensemble (PHYS) which includes CN.

3.1 Ensemble spread

Figure 2 plots the domain-mean ensemble spread (defined as standard deviation against ensemble mean) of some 2D fields, averaged over 35 forecast dates (covering most of the experiment period from April 18 through June 4). Several findings are drawn from Figure 2: 1) For the mass-related fields (Fig 2a,d), spreads increase monotonously with forecast lead time; 2) The QPF related spreads (Fig 2g,h) exhibit a clear diurnal pattern, reflecting the quiet morning hours (around 10am) and the active late afternoon hours (around 7pm) for the summer convective storm activity. Near surface and low level quantities (Fig 2b,e,f,i) also show similar diurnal pattern. In spite of the diurnal variation, these spread curves still show a general trend of increase with forecast lead time; 3) The initial perturbation members (PERT) have the biggest contributions to the spread of non-QPF fields, while the physics-variation-only members (PHYS) contribution far less, especially for the mean sea level pressure and 500 hPa geopotential height (Fig 2a,d). For the QPF fields, both PERT and PHYS sub-ensembles have roughly equal spread contributions (Fig 2d,g).

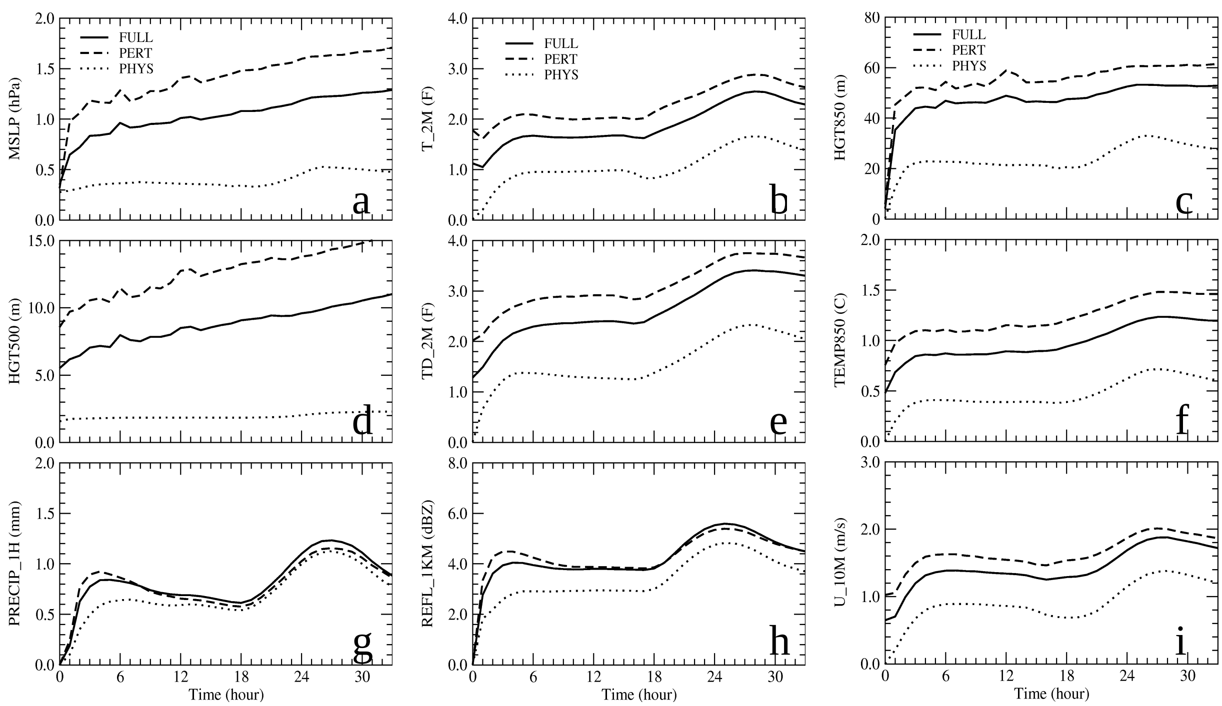

Figure 2. Domain-mean ensemble spread (standard deviation), averaged from April 18 through June 4 for the full 10-member ensemble (FULL, solid lines), the 4-member perturbed (PERT) members (dash lines), and the 6-member physics variation (PHYS) members (dot lines). (a) mean sea level pressure, (b) 2 m temperature, (c) 850 hPa and (d) 500 hPa geopotential height, (e) 2 m dew point temperature, (f) 850 hPa temperature, (g) 1 h accumulated precipitation, (h) reflectivity at 1 km AGL, and (i) 10 m wind (u-component).

3.2 RMSE of ensemble mean QPF

The real-time storm-resolving (or storm-permitting, convection-permitting in some literatures) ensemble forecasting system offers a unique capability of producing QPF (both deterministic and probabilistic) at very high temporal and spatial resolution. The experimental fine grid (1 km) national radar mosaic and QPE products generated by the NSSL/NMQ project[1] are first interpolated to the 4 km grid model domain and used to verify the predicted QPF quantities (1 h accumulated precipitation and composite reflectivity). Due to the lack of very high-resolution (at or near 4 km grid spacing) analysis for other forecast fields such as near surface wind and temperature, verification of such quantities remains a big challenge.

As the Spring Experiment is still underway at the writing of this extended abstract and the extremely slow data retrieval process for such a gigantic dataset, Figure 3 only shows an example of five-day average (May 20-24) ensemble mean RMSEs for the predicted 1 h accumulated precipitation and composite reflectivity fields. The same diurnal pattern as shown in Figure 2 is evident for both fields.

The post analysis of the complete dataset will allow us to assess the statistical consistency of the ensemble system by examining the spread-error relationship.

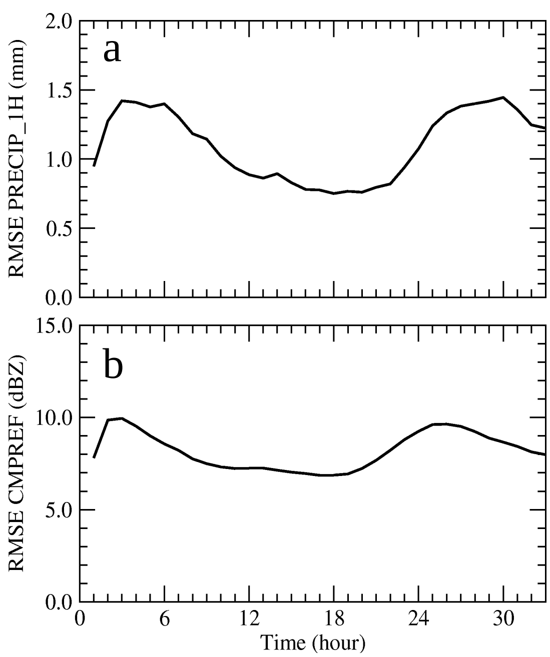

Figure 3. Ensemble mean RMSE of (a) 1 h accumulated precipitation and (b) composite reflectivity, averaged over five forecast dates (May 20-24), for the full ten-member ensemble.

3.3 Brier score

As the RMSE measure of the probabilistic forecast, Brier score (BS) is widely used in probabilistic QPF verification. Figure 4 shows an example one-day (May 22) BS curves with three thresholds (1 h precipitation ≥ 0.01, 0.1, and 0.5 in, respectively). Although a single day sample is far from adequate to draw statistically meaningful findings, the big decrease of BS value in Figure 4 with increasing threshold is misleading. It would just reflect the diminish number of grid points that bears heavy precipitation.

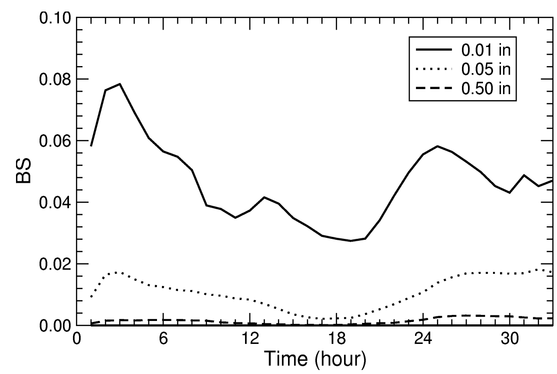

Figure 4. Brier score (BS) of the 1 h accumulated precipitation for the forecast date on May 22, 2007.

3.4 Verification rank histogram

Figure 5 is the verification rank histogram for the 33 h forecast 1 h accumulated precipitation for the single forecast date of May 22, 2007. Since it is calculated using only one date and by sampling all grid points, without taking consideration of sample independence, it is not meant to provide statistically meaningful reading of ensemble characteristics (Hamill 2001). Nevertheless, it still helps us getting a general understanding of whether the ensemble is underdispersive. Figure 5 shows a quite uniform distribution.

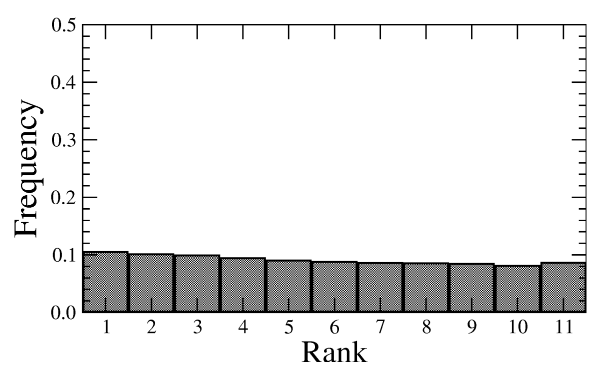

Figure 5. Verification rank histogram for the 33 h forecast 1 h accumulated precipitation for the forecast date on May 22, 2007.

4. ENSEMBLE FORECAST PRODUCTS - EXAMPLES

[1] http://www.nmq.nssl.noaa.gov/

In this section, we present some example results from the real-time storm-scale ensemble experiment.

4.1 April 23-25, 2007

The ensemble forecast was initiated at 2100 UTC April 23, 2007. Some products at two forecast lead times, 18 h and 33 h, along with the observed radar mosaic images are shown in Figures 6 – 10.

The probabilistic forecasts produced from ensemble members are more useful than deterministic forecast products, owing to the chaotic nature of weather systems. Figures 7 and 9 are such probability maps showing the relative frequency of the number of members that predict composite reflectivity equal to or larger than 35 dBZ. The comparison between the probability maps and the observed radar composite reflectivity mosaic at the same validate times (Figures 6 and 8) is not straightforward - the former gives probability values in percentage for the given condition (here >=35dBZ), while the latter directly gives reflectivity intensity values in dBZ.

For this case date, both the 18 h and 33 h probabilistic forecasts for composite reflectivity greater than or equal to 35 dBZ very well captured the major storm systems, as represented by high values of composite reflectivity in Figures 6 and 8, with high probabilities generally corresponding to high reflectivity areas..

Comparing the 10-member ensemble "postage-stamp" charts (Figure 10) to the observed radar mosaic in Figures 6 and 8, the high-resolution (at 4 km grid spacing) WRF ensemble forecasting system well predicted the observed storm system for up to 33 hours, both in location and in much detail. The ten individual ensemble members present different storm details and intensities, though the overall system looks quite similar in the postage-stamp charts. The differences represent uncertainties associated with model errors and limitations on storm-scale predictability. The traditional single member forecast (*aka* deterministic forecast) only produces one forecast, good or bad.

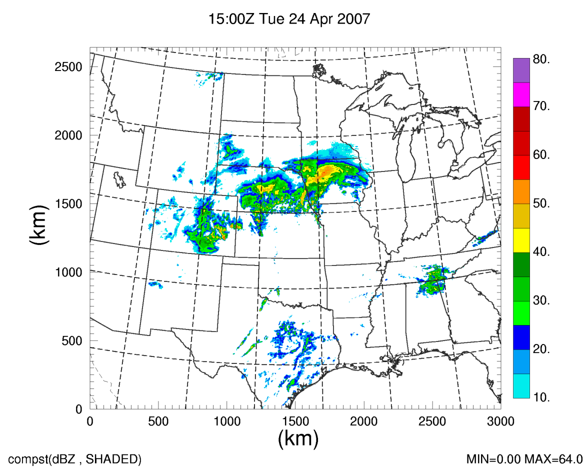

Figure 6. Observed composite radar reflectivity mosaic, valid at 1500 UTC April 24, 2007.

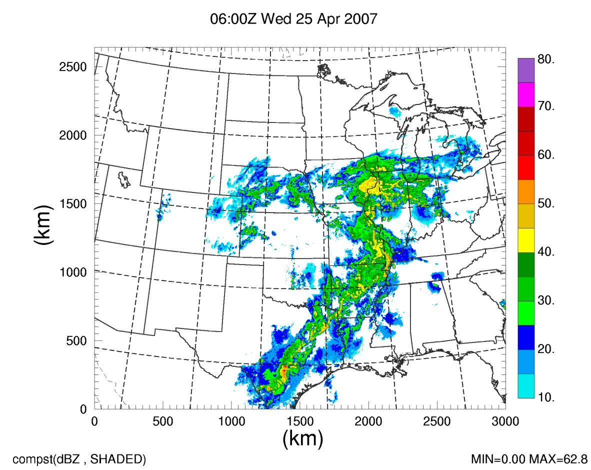

Figure 8. Observed composite radar reflectivity mosaic, valid at 0600 UTC April 25, 2007.

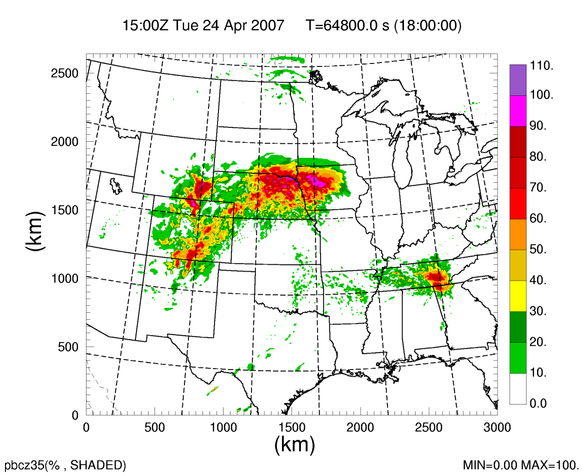

Figure 7. Probability of the 18 h forecast composite reflectivity ≥35 dBZ, valid at 1500 UTC April 24, 2007.

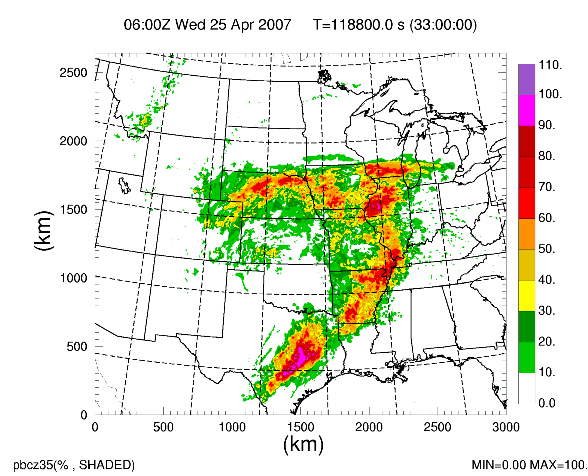

Figure 9. Probability of the 33 h forecast composite reflectivity ≥35 dBZ, valid at 0600 UTC April 25, 2007.

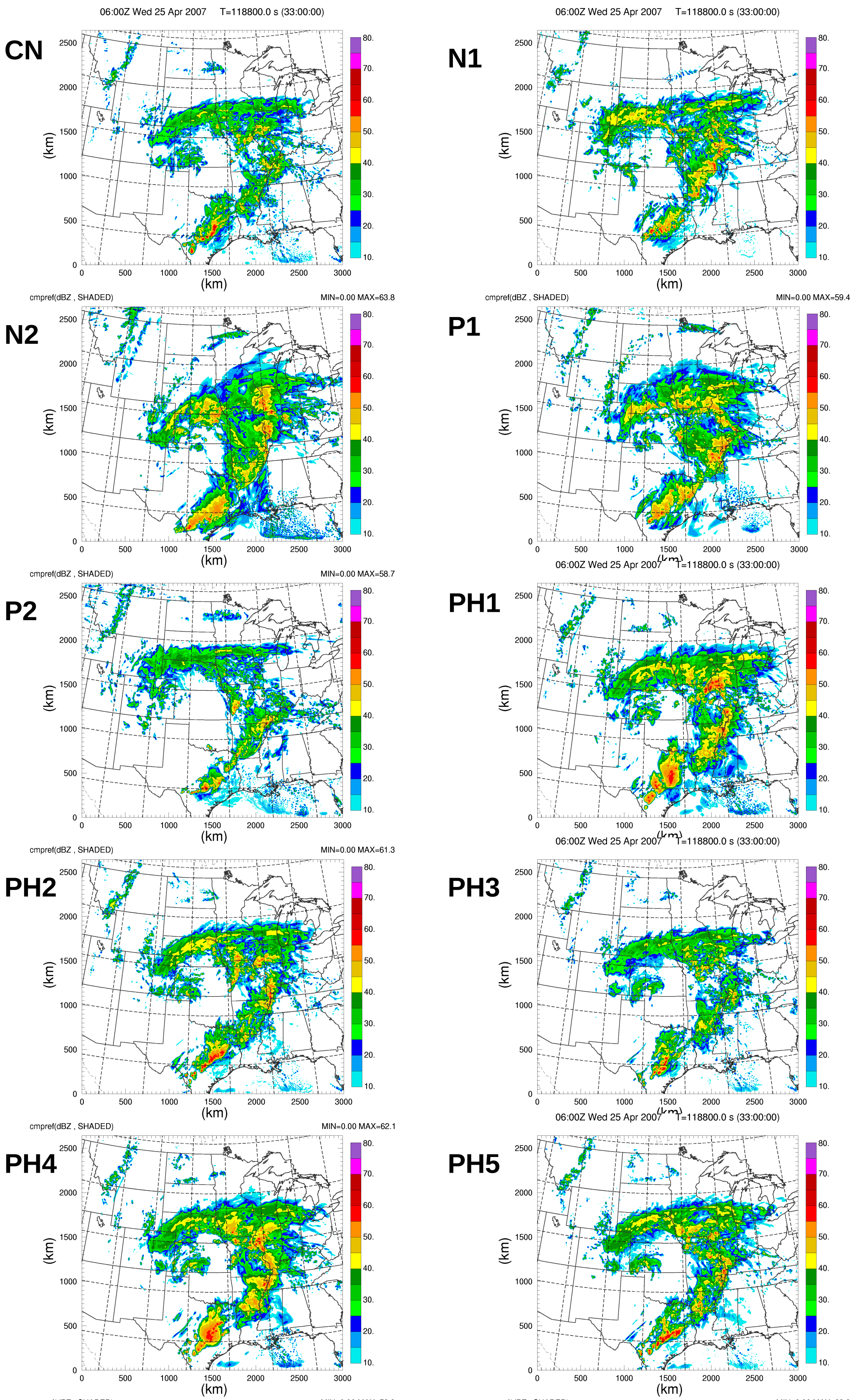

Figure 10. Postage-stamp chart of the 33 h forecast composite reflectivity from individual members.

5. ACKNOWLEDGMENTS

This research is mainly funded by a grant to CAPS from the NOAA CSTAR program. Supplementary support is also provided by the NSF ITR project LEAD (ATM-0331594). The numerical forecasting is performed at the Pittsburgh Supercomputing Center. David O'Neal of PSC provided enormous technical and logistic support for the experiment. The authors also thank Drs. Jimy Dudhia, Morris Weisman, Greg Thompson and Wei Wang for their very helpful suggestions and input on the WRF model configurations.

REFERENCES

Brooks, H. E., M. S. Tracton, D. J. Stensrud, G. Dimego, and Z. Toth, 1995: Short-range ensemble forecasting: Report from a workshop (25-27 July 1994). *Bull. Amer. Meteor. Soc.*, **76**, 1617-1624.

Du, J., and M. S. Tracton, 2001: Implementation of a real-time short-range ensemble forecasting system at NCEP: An update. *Preprints, 9th Conf. on Mesoscale Processes*, Amer. Meteor. Soc., 355-356.

Elmore, K. L., S. J. Weiss, and P. C. Banacos, 2003: Operational ensemble cloud model forecasts: Some preliminary results. *Wea. And Forecasting*, **18**, 953-964.

Hamill, T. M., S. L. Mullen, C. Snyder, and Z. Toth, 2000: Ensemble forecasting in the short to medium range: Report from a workshop. *Bull. Amer. Meteor. Soc.*, **81**, 2653-2664.

Hamill, T. M., 2001: Interpretation of Rank Histograms for Verifying Ensemble Forecasts. *Mon. Wea. Rev.*, **129**, 550-560.

Hou, D., E. Kalnay, and K. K. Droegemeier, 2001: Objective verification of the SAMEX'98 ensemble forecasts. *Mon. Wea. Rev.*, **129**, 73-91.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge Press, 341pp.

Kong, F., K. K. Droegemeier, and N. L. Hickmon, 2006: Multiresolution ensemble forecasts of an observed tornadic thunderstorm system. Part I: Comparison of coarse- and fine-grid experiments. *Mon. Wea. Rev.*, **134**, 807-833.

Kong, F., K. K. Droegemeier, and N. L. Hickmon, 2007: Multiresolution ensemble forecasts of an observed tornadic thunderstorm system. Part II: Storm-scale experiments. *Mon. Wea. Rev.*, **135**, 759-782.

Walsar, A., D. Luthi, and C. Schar, 2004: Predictability of precipitation in a cloud-resolving model. *Mon. Wea. Rev.*, **132**, 560-577.