Extended Abstract, 25th Conference on Severe Local Storms
Amer. Meteor. Soc., Denver, Colorado, October 11-14

7B.3 CAPS REALTIME STORM-SCALE ENSEMBLE AND CONVECTION-RESOLVING HIGH-RESOLUTION FORECASTS FOR THE NOAA HAZARDOUS WEATHER TESTBED 2010 SPRING EXPERIMENT

Ming Xue[1,2], Fanyou Kong[1], Kevin W. Thomas[1], Yunheng Wang[1], Keith Brewster[1], Jidong Gao[1], Xuguang Wang[1,2], Steve Weiss[4], Adam Clark[3], John Kain[3], Mike Coniglio[3], Jun Du[5], Tara Jensen[6] and Ying-Hwa Kuo[6]

[1]Center for Analysis and Prediction of Storms and [2]School of Meteorology
University of Oklahoma
[3]National Severe Storms Laboratory, NOAA
[4]Storm Prediction Center/NCEP, NOAA
Norman Oklahoma
[5]Environmental Modeling Center/NCEP, NOAA, Maryland
[6]National Developmental Testbed Center, Boulder Colorado

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 other funding, the Center for Analysis and Prediction of Storms (CAPS) at the University of Oklahoma has been carrying out a multi-year project since 2007 (Xue et al. 2007; Kong et al. 2007; Xue et al. 2008; Kong et al. 2008; Xue et al. 2009; Kong et al. 2009), to develop, conduct, and evaluate realtime high-resolution ensemble and deterministic forecasts for convective-scale weather, as a contribution to the NOAA Hazardous Weather Testbed (HWT, see, e.g., Weiss et al. 2007). The realtime forecasts, together with retrospective analyses using the real time data, aim to address scientific issues including: (1) the values and cost-benefit of convection-allowing-resolution ensemble versus coarser-resolution short-range ensembles and even-higher-resolution convection-resolving deterministic forecast; (2) optimal design of a storm-scale ensemble system including the initial condition perturbation methods, physics perturbations, and the use of multiple models; (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 post-processing and forecast products for the convective storm scales, including ensemble calibration; and (6) the value and impact of such unique products for forecasting guidance and warning.

This paper will briefly summarize the 2007-2009 CAPS spring forecast experiments and highlight key findings. It will then report on the experimental design, logistic issues and preliminary analysis results from the spring 2010 forecast experiment. A number of other papers examining the CAPS's storm-scale ensemble forecasts (SSEF) data from different perspectives will also be presented at this conference.

2. CAPS forecasts of 2007-2009

For the spring 2007 forecasts, 33-hour 10-member 4-km-resolution ensemble forecasts and a single 2-km deterministic forecast initialized at 2100 UTC were produced daily for a forecast domain covering two thirds of the continental US (CONUS). The control initial conditions (ICs) were obtained by directly interpolating NCEP NAM analyses at 2100 UTC. The IC and lateral boundary condition (LBC) perturbations were derived from the 2100 UTC SREF (Short-range Ensemble Forecast, Du et al. 2006) forecasts (Xue et al. 2007; Kong et al. 2007). Five 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. 2010) while the subsequent analyses also showed clear under-dispersion among the physics-perturbation-only members (Kong et al. 2007; Kong et al. 2008). Realtime production of ensemble forecasts for a significant portion of a continent at a convection-allowing resolution was the first in the world, and the analysis of the data set have resulted in several additional referred publications (Schwartz et al. 2009; Clark et al. 2009; Clark et al. 2010a, b).

In the spring of 2008, the ensemble size remained at 10 but all members included IC and LBC perturbations as well as physics perturbations. The forecasts were initialized at 0000 UTC instead and lasted for 30 hours. Comparison tests showed that forecasts starting from 0000 UTC had smaller positive biases in precipitation.

The most significant enhancement to the forecasts of 2008 was the assimilation of level-2 radial velocity and reflectivity data from over 120 operational WSR-88D Doppler weather radars into all except for one model run, using a parallelized version of the ARPS 3DVAR (Gao et al. 2003) data analysis system that includes a complex cloud analysis procedure (Hu et al. 2006). Furthermore, the model domain was enlarged to reach beyond the eastern coast of the US continent. In both 2007 and 2008, version 2.2 of the Advanced Research Weather Research and Forecast (ARW hereafter, Skamarock et al. 2005) model was used for the forecast.

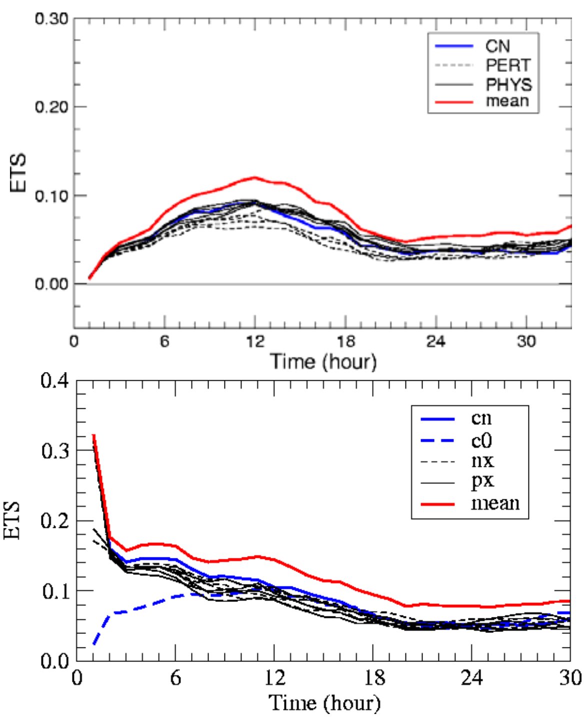

Fig. 1. Equitable threat scores (ETS) of hourly precipitation ≥ 0.1inch for the 4-km ensemble forecasts from 2007 (upper panel) and 2008 (lower panel). cn refers to the control forecast which in 2008 included radar data. c0 is the control forecast without radar. Red lines is the simple ensemble mean (from Kong et al. 2008).

The results of the 2008 experiment showed, for the first time, based on forecasts over an extended period, that radar data can eliminate the typical spin-up problem associated with the precipitation forecast of almost all existing operational NWP models; the precipitation forecasts assimilating radar data are improved for up to 12 hours over those that do not assimilate radar data, with the positive impact being the greatest in the first 6 hours (Xue et al. 2008) (lower panel of Fig. 1). Without radar data, all forecasts in 2007 required a significant length of spin-up time, just as the non-radar member (c0) of 2008. The ensemble mean shows better scores for both years but this is true for low precipitation only because of the smearing effect of simple ensemble mean with precipitation forecasts. Probability-matched ensemble mean is a better quality which is used in 2010 (more on this later).

The positive impact of radar data assimilation within the CAPS forecasts is also discussed in Kain et al. (2010) and Berenguer et al. (2010). Further, Berenguer et al. (2010) compared the CAPS 2008 4-km control forecasts with those from the McGill University MAPLE nowcasting system, and those from two other lower-resolution models and showed that CAPS's radar-assimilating WRF forecasts out-perform all other forecasts in general. Examining the 2008 data, Coniglio et al. (2010) assessed the forecasts of the pre-convective and near-storm environments by the CAPS 4-km ensemble and found that mean forecasts from the ensemble were substantially more accurate than forecasts from deterministic forecasts.

In the spring of 2009, the CAPS ensemble was expanded to include two additional models, the WRF-NMM (Janjic 2003) and ARPS (Xue et al. 2000). Further, the resolution of the deterministic forecast was increased to 1 km. 3DVAR analyses of full-volume radar data in the entire model domain were again performed, at the native 4-km and 1-km resolutions (Xue et al. 2009; Kong et al. 2009). The 4-km ensemble had 10 WRF-ARW members, 8 WRF-NMM members, and 2 ARPS members. The 1-km deterministic forecast used WRF-ARW configured the same way as the 4-km ARW control member. Each of the models had two 'control' forecasts in which the ICs and LBCs were unperturbed. One of the forecasts included radar data assimilation and one did not. All other members included radar data.

Fig. 2 show the ETS scores of 3-hourly accumulated precipitation exceeding 0.5 inch threshold for the CAPS forecasts of 2008 through 2010, together with the NCEP operational 12-km NAM forecasts.

PM in the plots refers to the probability-matched ensemble mean. Since precipitation is spatially discontinuous, simple ensemble mean tends to severely underestimate the peak values due to the smoothing effect of the ensemble averaging, a probability matching (PM) technique proposed by Ebert (2001) is used here. The procedure assumes that the best spatial representation of rainfall is given by the ensemble mean and the best frequency distribution is given by the precipitation amounts forecast by the ensemble members. The ensemble mean obtained from the PM procedure can help correct for large biases in areal rainfall coverage and underestimation with the standard ensemble mean, and results in a precipitation field with a much more realistic distribution (Clark et al. 2009).

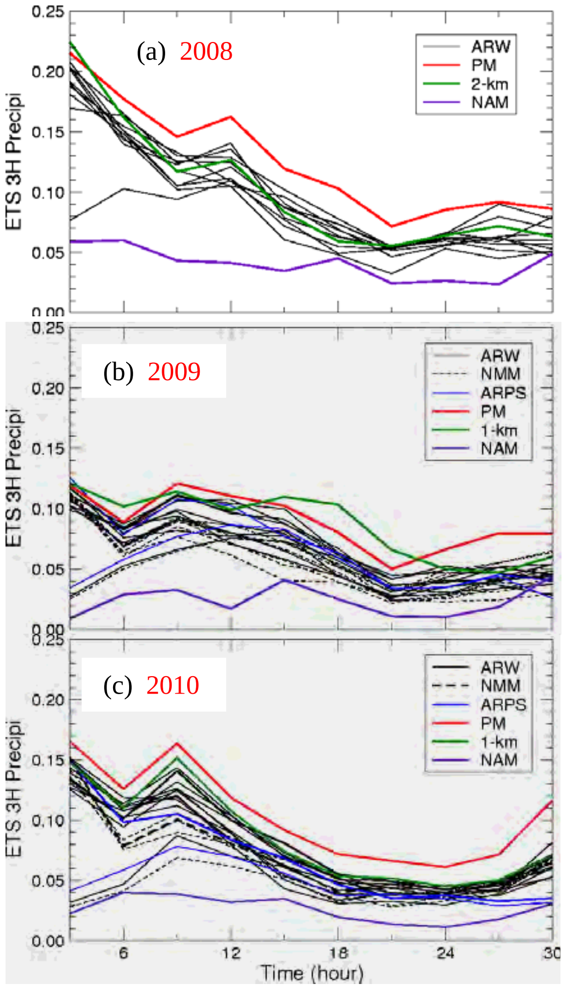

Fig. 2. ETS scores of 3-hourly accumulated precipitation ≥ 0.5 inch for the CAPS forecasts of three years, together with the NCEP operational 12-km NAM forecasts. PM refers to probability-matched ensemble mean (see text). The control members (cn) of each model are in bold lines in 2009 and 2010.

Several conclusions can be drawn from the plots in Fig. 2. The radar-assimilating 4-km members clearly outperform the non-radar members, and the positive impact lasts between 12 and 18 hours or even longer, with the largest impact being in the first 6 to 9 hours. The probability-matched ensemble mean has higher scores than each of the individual ensemble members, and is higher than the 2-km forecast in 2008 and the 1-km forecast in 2010, suggesting that for light precipitation at least a 4-km ensemble can produce better deterministic forecasts than the more expensive 1 or 2 km high-resolution forecasts. The 1-km forecasts are generally better than the individual 4-km forecasts, and in 2009 they exceed the skill scores of PM forecast at some of the forecast hours. The higher-resolution convection-allowing/resolving forecasts are exclusively better than the 12-km operational NAM forecasts, with much larger differences being achieved when radar data are performed. These results are consistent with independent evaluations by the Developmental Testbed Center (Jensen et al. 2010). The results of 2008 and 2009 also suggest systematic performance differences between WRF-ARW and WRF-NMM cores. The two models shared a common set of initial and boundary conditions, and in some cases common physics schemes.

3. Spring 2010 CAPS Forecast Configurations

In the spring of 2010, CAPS continued its participation in the NOAA HWT Spring Experiment by providing realtime 4-km multi-model ensemble and 1-km high-resolution deterministic forecasts, and for supporting the VORTEX-2 field experiment. In 2010, in addition to the joint organizers of the HWT, the Storm Prediction Center (SPC) and the National Severe Storm Laboratory (NSSL), the HWT Spring Experiment was also actively participated by the Aviation Weather Center (AWC), the Hydrometeorological Prediction Center (HPC) and the Developmental Testbed Center (DTC).

In spring 2010, the CAPS SSEF system was further enhanced. Major changes include: 1) The forecast domain was expanded to cover the entire continental United States, increasing total computing grid points by ~40% compared to 2009 domain; 2) The total number of ensemble members was increased to 26, consisting of again three dynamics cores (WRF-ARW, WRF-NMM, ARPS); 3) New microphysics and PBL schemes available in WRF V3.1 were included in some ARW members; 4) A comprehensive set of ensemble post-processed products were generated and made available in near-real-time to HWT.

More specifically, the CAPS 2010 spring forecasts started on 26 April 2010 and ended on 18 June, encompassing the HWT 2010 Spring Experiment that is officially between 17 May and 18 June. This experiment period was shifted into mid-June to cover both the severe storm season of the central Great Plains and aviation interests in early summer. Forecasts were produced Monday through Friday, initialized at 0000 UTC (1900 CDT) of each day and made available early morning for evaluation at HWT. The 4-km ensemble contained 19 ARW, 5 NMM and 2 ARPS members for a total of 26. Six of the ARW members had differences only in the physics parameterization schemes, and were designed to facilitate the evaluation of several microphysics and PBL schemes newly available in WRF-ARW. Three ARW members contained initial condition perturbations of different scales to allow for the study of multi-scale error growth and predictability. For the perturbed boundary members, forecasts from consistent NCEP SREF members were used to provide the lateral boundary conditions at a higher frequency (hourly) than before. The upgraded ensemble-transform-initialized

SREF provided initial condition perturbations for most SSEF members.

As in 2008 and 2009, level-2 radial velocity and reflectivity data from over 140 operational WSR-88D radars were analyzed using the ARPS 3DVAR together with its cloud analysis package. The results served as the control initial conditions for the three models. Thirty-hour forecasts were produced once per day.

Parallel to the 4-km ensemble, a 1-km high-resolution deterministic forecast was produced once a day, for the same CONUS domain. Radar and other high-resolution (e.g., Oklahoma Mesonet) data were also analyzed, on the native 1-km grid. A National TeraGrid supercomputer operated by National Institute for Computational Science (NICS) at the University of Tennessee, a Cray XT-4 supercomputer with 18,000+ CPU cores, was used in a dedicate mode for about 6 hours a day for producing these forecasts and for post-processing data output. Hourly 3D model outputs were produced and archived while some fields were output every 5 minutes for high-frequency animations.

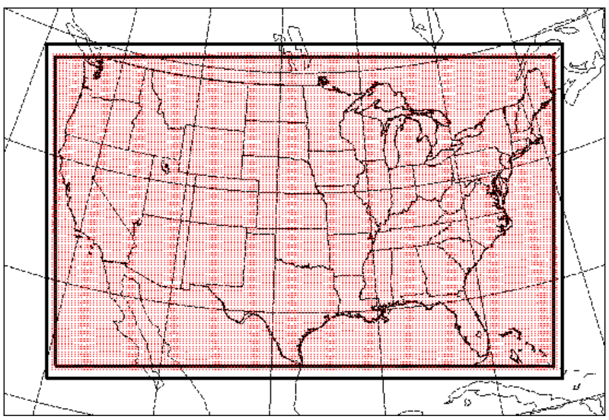

Fig. 3. The 2010 CAPS Spring Forecast Experiment domains. The thick inner box represents the ARW and ARPS forecast domains for the 1-km convection-resolving and 4-km ensemble forecasts. The red-dotted area represents the NMM forecast domain as part of the ensemble. The ARPS 3DVAR analyses are performed on the larger grid enclosed by the outer black rectangle. The difference in the map projection used by the NMM grid requires this special setup. Forecast verifications are performed on the inner common grid.

Fig. 3 shows the coverage area of the model domains. Because NMM only supports the rotated latitude-longitude E-grid, its domain cannot be made to match the common ARPS and ARW grid. As a result, three different domains have to be used in the analysis and forecast process. The ARW and ARPS members have the same forecast domain bounded by the inner thick box, which also serves as the common verification domain for all models. The NMM forecast domain (red-dotted area) is slightly larger and encompasses the ARW and ARPS domain, and it is further encompassed by an even larger domain (bounded by the thick outer box) in which 3DVAR analysis is performed. Similar was done in 2009, but for somewhat smaller domains.

Special software codes were developed at CAPS to convert between the NMM E-grid and the ARPS/WRF C-grid in order to utilize a single 3DVAR/Cloud Analysis over a larger outer domain that encompasses both forecast domains (Fig. 3). The conversion software was upgraded to be compatible with the new WRF version (V3.1.1) used in the 2010 season.

The procedures for creating initial and boundary conditions for each of three models and their members are similar to those used in 2009. Xue et al. (2009) provides a detailed description on them. The paper also provides the details on the ARPS 3DVAR and cloud analysis configurations. All pre- and post-processing programs were optimized to handle very large computational grids and huge volumes of data; all of them support parallelization through the MPI programming interface.

Tables 1-3 list the configurations for each individual members of each model group (arw, nmm, and arps). cn refers to the control member, with radar data analysis, c0 is the same as cn except for no radar data is analyzed in. m3 – m19 are members with either initial perturbation or physics perturbation or both added on top of cn initial condition. NAMa and NAMf refer to 12-km NAM analysis and forecast, respectively. ARPSa refers to ARPS 3DVAR and cloud analysis using NAMa as the background. For the perturbed members arw_m5~m14 and nmm_m3~m5, the ensemble initial conditions consist of a mixture of bred and Ensemble Transform (ET) perturbations coming from the 21Z SREF perturbed members (4 WRF-em (ARW), 4 WRF-nmm (NMM), 2 ETA-KF, 2 ETA-BMJ, and 1 RSM-SAS) and physics variations (grid-scale microphysics, land-surface model (LSM), and PBL physics).

New in 2010 Spring Experiment is the addition of three random perturbation members (arw_m3~m5) and five extra physics-perturbation-only members (arw_m15~m19). Two types of random perturbations are added, one is completely random Gaussian perturbations and another is Gaussian perturbations smoothed by a recursive filter to have convective scale horizontal correlations. The physics-perturbation-only members are added to help assessing impacts from different, especially new microphysics and PBL schemes. The lateral boundary conditions come from the corresponding 21Z SREF forecasts directly for those perturbed members and from the 00Z 12-km NAM forecast for the non-SREF-perturbed members. For the ARPS model group, the only members are cn and c0, as in 2009 season. We note here that all ensemble-derived products of 2010 used the 15 members highlighted orange in Tables 1-3. They exclude members without radar data, members that had physics perturbations only, and members that introduced additional initial random perturbations.

Compared to the earlier years, several new physics options available in WRF V3.1.1 were used. They include a new version of the Thompson microphysics scheme that includes the prediction of rain number concentration, the WRF Double-Moment 6-class (WDM6) scheme that also predicts two moments of rain, the Morrison double-moment scheme (Morrison et al. 2005) that predicts the number concentrations of ice, snow, rain and graupel. The new PBL schemes tested include the Nakanishi and Niino improved version of Mellor-Yamada (MYNN) Level-2.5 scheme (Nakanishi and Niino 2004, 2006), and the QNSE, based on the quasi-normal scale elimination turbulence model (Sukoriansky et al. 2005; Sukoriansky et al. 2006). Experiments arw_m15 through arw_m19, together with arw_cn, form a set of 6 experiments that share the same initial and boundary conditions within the same dynamic core. They will allow for detailed studies on the behaviors of these rather new schemes over a period containing many cases and a variety of weather conditions. A discussion on other schemes used in the ensemble can be found in Xue et al. (2009).

Table 1. Configurations for each individual member with WRF-ARW core. NAMa and NAMf refer to the 12-km NAM analysis and forecast, respectively. ARPSa refers to ARPS 3DVAR and cloud analysis. Members highlighted orange are used in producing probabilistic ensemble products.

member	IC	BC	Radar	microphysics	LSM	PBL
arw_cn	00Z ARPSa	00Z NAMf	yes	Thompson	Noah	MYJ
arw_c0	00Z NAMa	00Z NAMf	no	Thompson	Noah	MYJ
arw_m3	arw_cn + random pert	00Z NAMf	yes	Thompson	Noah	MYJ
arw_m4	arw_cn + RF-smoothed pert	00Z NAMf	yes	Thompson	Noah	MYJ
arw_m5	arw_cn + em-p1 + recur pert	21Z SREF em-p1	yes	Morrison	RUC	YSU
arw_m6	arw_cn + em-p1_pert	21Z SREF em-p1	yes	Morrison	RUC	YSU
arw_m7	arw_cn + em-p2_pert	21Z SREF em-p2	yes	Thompson	Noah	QNSE
arw_m8	arw_cn – nmm-p1_pert	21Z SREF nmm-p1	yes	WSM6	RUC	QNSE
arw_m9	arw_cn + nmm-p2_pert	21Z SREF nmm-p2	yes	WDM6	Noah	MYNN
arw_m10	arw_cn + rsmSAS-n1_pert	21Z SREF rsmSAS-n1	yes	Ferrier	RUC	YSU
arw_m11	arw_cn – etaKF-n1_pert	21Z SREF etaKF-n1	yes	Ferrier	Noah	YSU
arw_m12	arw_cn + etaKF-p1_pert	21Z SREF etaKF-p1	yes	WDM6	RUC	QNSE
arw_m13	arw_cn – etaBMJ-n1_pert	21Z SREF etaBMJ-n1	yes	WSM6	Noah	MYNN
arw_m14	arw_cn + etaBMJ-p1_pert	21Z SREF etaBMJ-p1	yes	Thompson	RUC	MYNN
arw_m15	00Z ARPSa	00Z NAMf	yes	WDM6	Noah	MYJ
arw_m16	00Z ARPSa	00Z NAMf	yes	WSM6	Noah	MYJ
arw_m17	00Z ARPSa	00Z NAMf	yes	Morrison	Noah	MYJ
arw_m18	00Z ARPSa	00Z NAMf	yes	Thompson	Noah	QNSE
arw_m19	00Z ARPSa	00Z NAMf	yes	Thompson	Noah	MYNN

* For all members: ra_lw_physics= RRTM; ra_sw_physics=Goddard; cu_physics= NONE

Table 2. Configurations for each individual member with WRF-NMM core

member	IC	BC	Radar	mp_phy	lw_phy	sw-phy	sf_phy
nmm_cn	00Z ARPSa	00Z NAMf	yes	Ferrier	GFDL	GFDL	Noah
nmm_c0	00Z NAMa	00Z NAMf	no	Ferrier	GFDL	GFDL	Noah
nmm_m3	nmm_cn + nmm-n1_pert	21Z SREF nmm-n1	yes	Thompson	RRTM	Dudhia	Noah
nmm_m4	nmm_cn + nmm-n2_pert	21Z SREF nmm-n2	yes	WSM 6-class	RRTM	Dudhia	RUC
nmm_m5	nmm_cn + em-n1_pert	21Z SREF em-n1	yes	Ferrier	GFDL	GFDL	RUC

* For all members: cu_physics= NONE; pbl_physics= MYJ.

Table 3. Configurations for each individual member with ARPS

member	IC	BC	Radar	Microphysics	radiation	PBL	Turb	sf_phy
arps_cn	00Z ARPSa	00Z NAMf	yes	Lin	Chou/Suarez	TKE	3D TKE	2-layer
arps_c0	00Z NAMa	00Z NAMf	no	Lin	Chou/Suarez	TKE	3D TKE	2-layer

* For all members: no cumulus parameterization

4. Forecast products

Selected 2D fields from each ensemble member were written in GEMPAK format and directly transferred into SPC's N-AWIPS system to be evaluated by the Spring Experiment participants at HWT during the weekdays. In addition, CAPS also made available a realtime webpage for the 2010 Spring Experiment products (http://forecast.caps.ou.edu/), with highlights to an ensemble post-processed product page[1] and an animation page[2] with 5-min interval composite reflectivity movies for the ARW control members (arw_cn and arw_c0) and 1-km forecasts.

New in 2010 Spring Experiment regarding forecast product is the generation, in near realtime, of a large set of post-processed ensemble products from 15 of the 26 4-km members (orange-colored in Table 1-3) that represent multi-model, IC perturbation, and physics variation ensemble with influence of radar data assimilation. The products include ensemble maximum and mean, probability matched mean (Ebert 2001; Clark et al. 2009; Kong et al 2008), ensemble exceedance probability, and neighborhood probability. Variables processed include forecast reflectivity, 1-, 3-, and 6-h accumulated precipitation, 2-m temperature and dew point, 10-m wind, 3-6 km updraft/downdraft velocities, echo top exceeding 18 dBZ, updraft helicity, 0-1 km and 0-6 km wind share, vertically integrated graupel/hail content, and some convective storm related indices (CAPE, CIN, LCL). Other variables diagnosed include Bunkers right-moving storm motion vector and speed, Supercell Composite Parameter (SCP), Significant tornado Parameter (STP) (Bunkers et al. 2000; Thompson et al. 2002; Thompson et al. 2004).

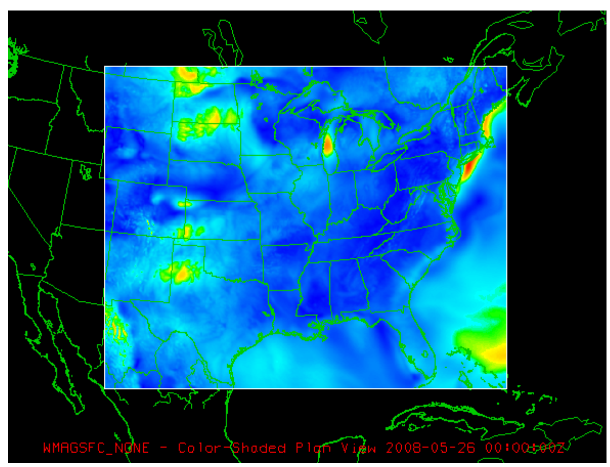

Fig. 4. Subdomain (inset) for HWT GEMPAK data set.

Both individual member and ensemble products were made available to HWT, AWS and DTC. The latter is also available to HPC. In order to minimize the data flow, the GEMPAK dataset, including individual member and ensemble product, was limited to a subdomain that covers the eastern 2/3 of CONUS (Fig. 4). CAPS saved the full domain 2D data set in HDF4 format.

[1] http://www.caps.ou.edu/~fkong/sub_atm/spring10.html
[2] http://forecast.caps.ou.edu/ywang/animation

5. Forecast examples

To provide a subjective impression first on the 4-km ensemble and 1-km high-resolution forecasts, we present forecast examples for several heavy precipitation and/or severe weather cases in this section.

a. The June 14, 2010 Oklahoma City flooding case

During the morning hours of June 14, 2010, up to 10 inch of rain fell over part of OKC. Slow-moving squall lines were involved, with training cells that passed over the Oklahoma City area over a period of many hours. Fig. 5 presents the 14 hour forecasts, from the 1-km grid (Fig. 5a), the SSEF-derived probability-matched ensemble mean (Fig. 5c), the uncalibrated probability of hourly precipitation exceeding 0.5 inch (Fig. 5d), and the corresponding observed composite reflectivity (Fig. 5b). Clearly, the line of cells in southwest-northeast orientation through central Oklahoma is well predicted by both the 1-km grid and the 4-km ensemble. The probability of hourly precipitation exceeding 0.5 inch does show a maximum over the Oklahoma City area, but the probability value appears too low. Calibration is clearly needed for reliability probability forecasts.

Fig. 6 and Fig. 7 compare the simple and probability-matched SSEF ensemble means of 6-hour accumulated precipitation with the corresponding ensemble means from the operational SREF model that has an average resolution of about 30 km. Also shown are the NAM forecasts and quantitative prediction estimate (QPE). The 6-hour periods covered are 1200-1800 UTC and 1800-0000 UTC, respectively, spanning the period of heavy precipitation in Oklahoma.

From the figures, one can see that the general precipitation pattern and magnitude are best predicted by SSEF PM mean, with the maximum being around 6 inch, very close the observed value. The simple mean under-estimates the peak value as expected. The SREF means in both forms severely underestimate the maximum precipitation, while NAM performs only slightly better. These are not surprising because of their lower spatial resolution and inability of explicitly represent convective cells. During the period of 1800 to 0000 UTC, SSEF PM mean again performs the best. The general forecast precipitation region matches the observation exceptionally well, and the heavy precipitation regions match the structure and orientation of observations also quite well.

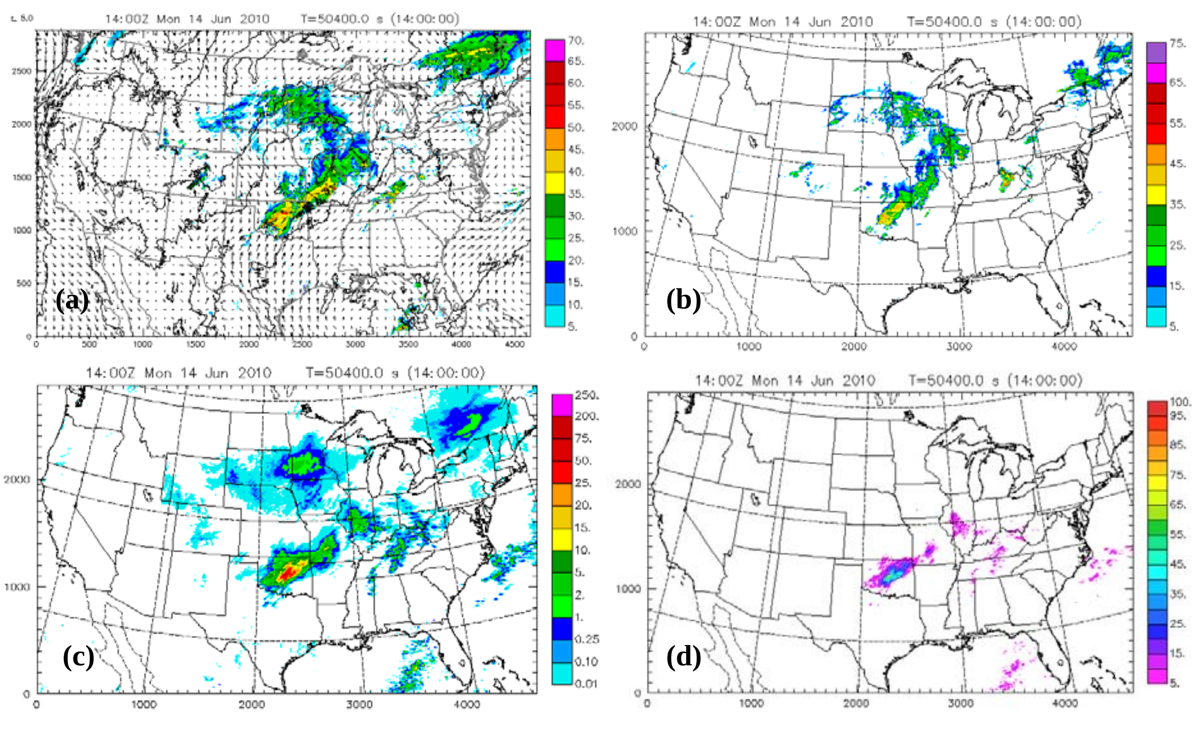

Fig. 5. 14-hour forecast composite reflectivity of the 1-km grid (a), corresponding observed radar reflectivity (b), probability-matched hourly precipitation (max=125 mm) derived from the 4-km ensemble (c), and uncalibrated probability of hourly precipitation exceeding 0.5 inch. Forecasts are 14 hour long and valid at 1400 UTC June 14, 2010.

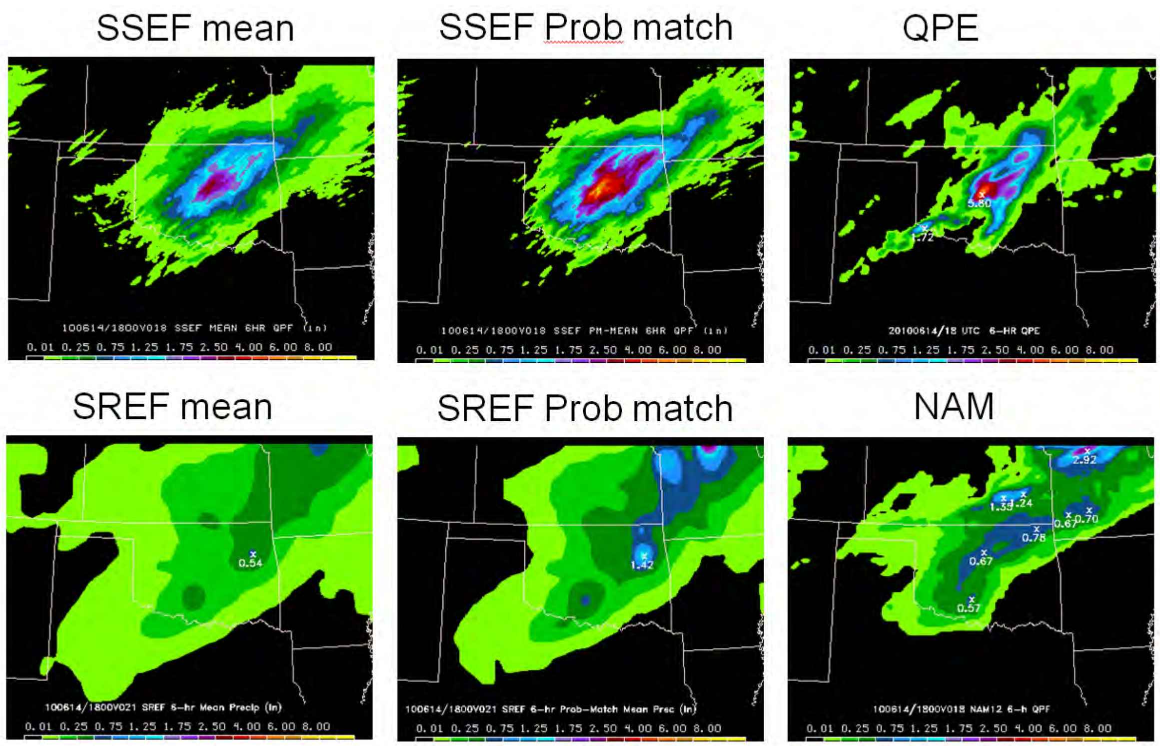

Fig. 6. Various forms of six-hour accumulated precipitation. In the left column are the ensemble mean of the CAPS SSEF forecasts and the NCEP SREF forecasts, while in the central column are the corresponding probability-matched ensemble means. In the right column are the QPE (observed precipitation), and the NCEP NAM forecast. They are valid at 1800 UTC, June 14, 2010, corresponding to 18 hour forecast time.

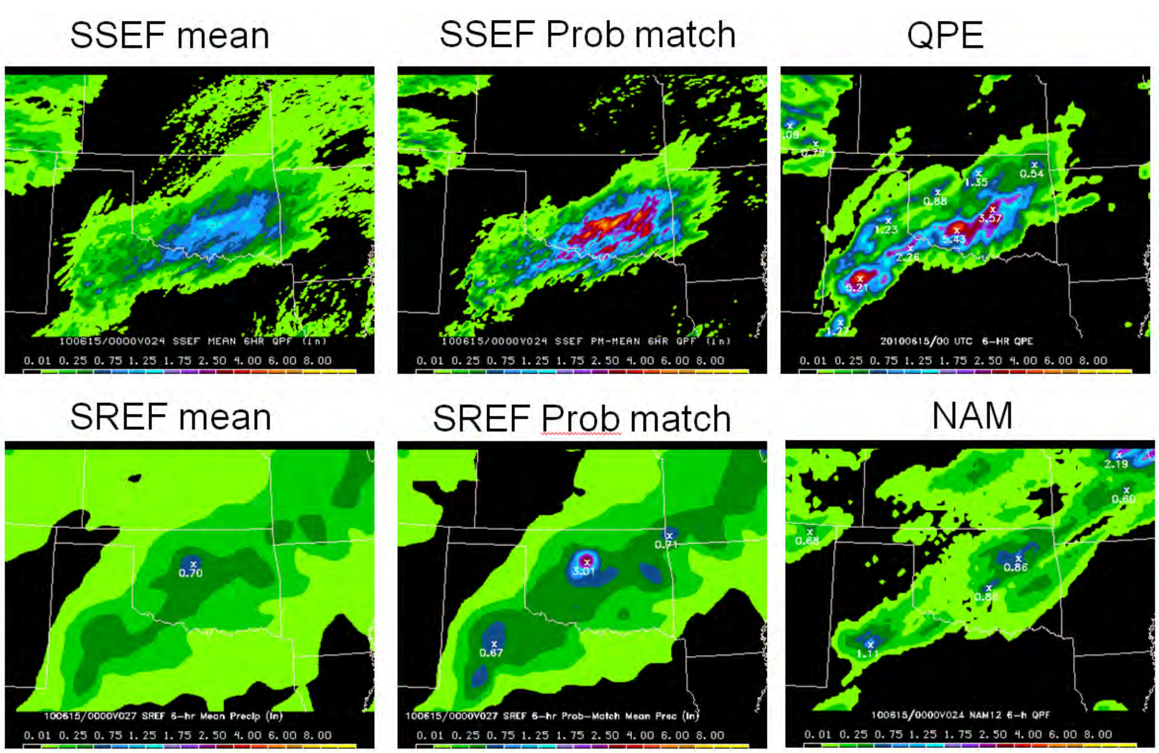

Fig. 7. As in Fig. 6 but valid at 0000 UTC, June 15, 2010, corresponding to 24 hour forecast time.

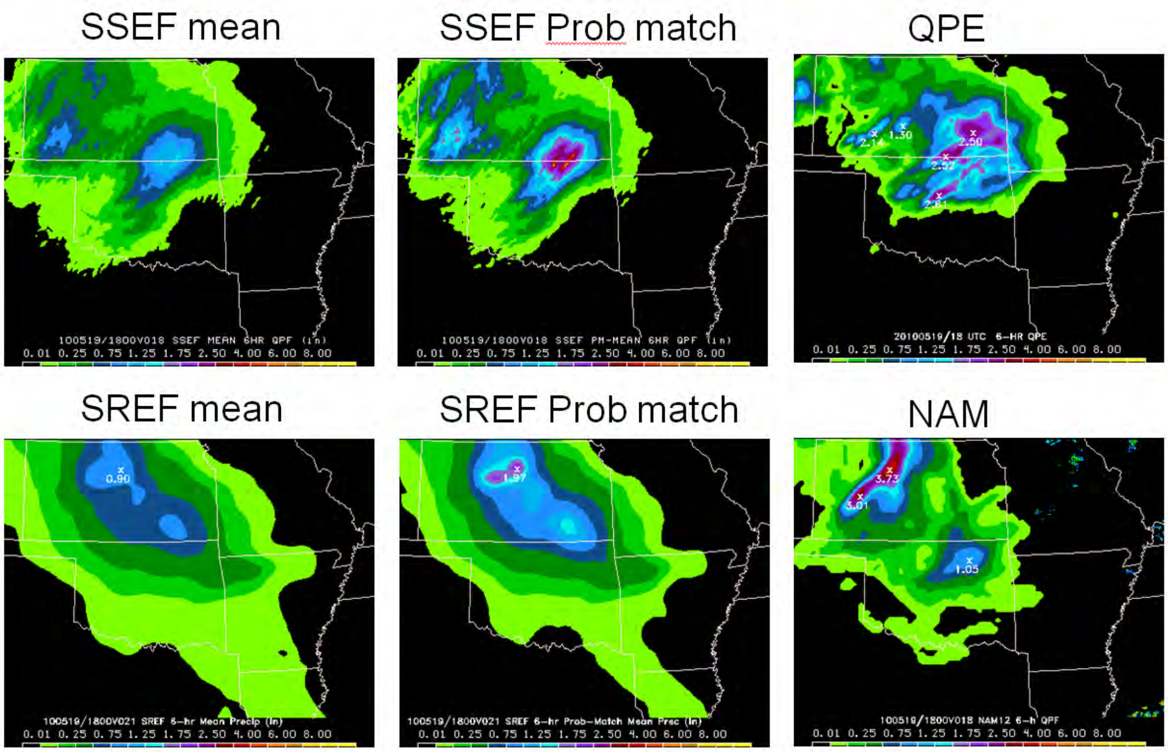

Fig. 8. As in Fig. 6 but valid at 1800 UTC, May 19, 2010, corresponding to 18 hour forecast time.

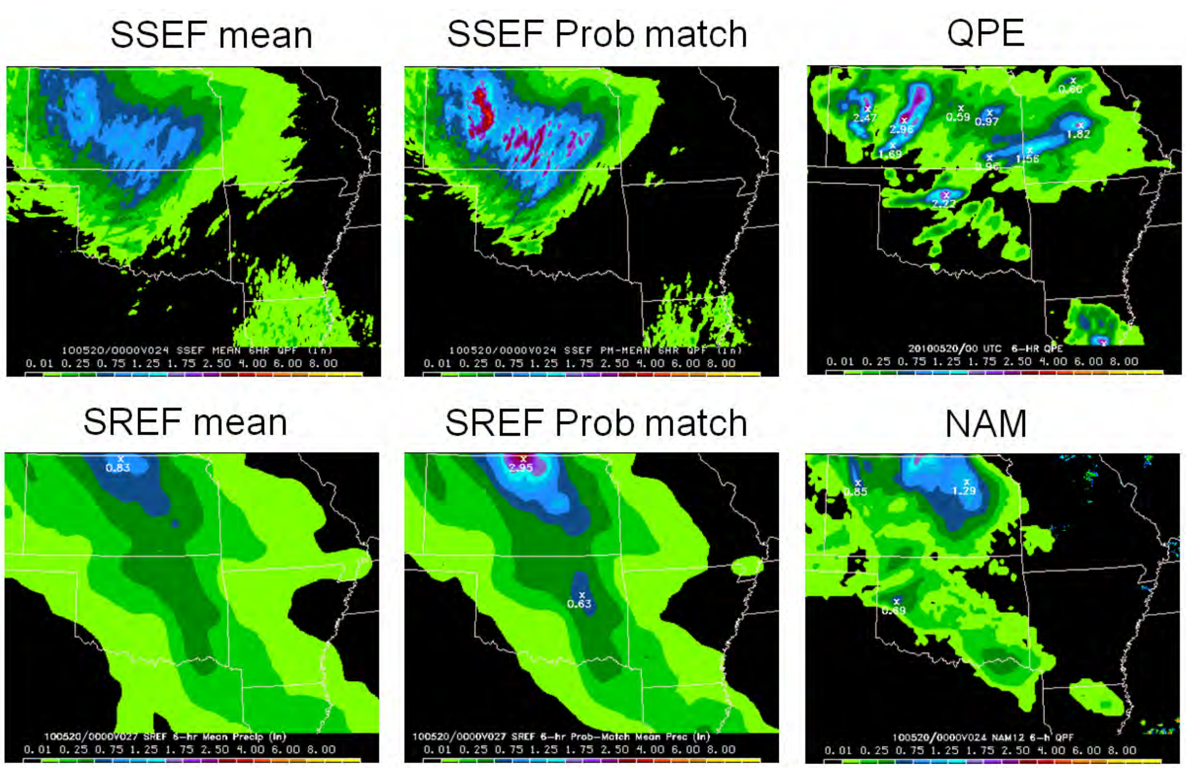

Fig. 9. As in Fig. 6 but valid at 0000 UTC, May 20, 2010, corresponding to 24 hour forecast time.

b. Heavy precipitation case of May 19-20

In late May 19 through early May 20, moderately heavy precipitation fell in the southern and eastern portion of Kansas and in the northern part of Oklahoma. Similar statements can be made about the forecasts for the two 6-hour periods as in the previous case. Again the SSEF PM means provide the best depiction of the precipitation distribution, structure and magnitude. Between 1200 and 1800 UTC, SSEF produced heavy precipitation in southeast Kansas and near the northern Oklahoma border, organized in a southwest-northeast orientation, similar to the observed. The SREF, however, predicts a general precipitation pattern that is northwest-southeast oriented, with the heavy precipitation core located too far northwest in northern Kansas. The same problem occurs with the NAM forecast, whose maximum precipitation is better than SREF.

In the second 6-hour period, the primary observed heavy precipitation region is in central-western Kansas. This particular region is captured in SSEF PM forecast, though the location is displaced somewhat to the west. SREF predicts a heavy precipitation center dislocated north at the Kansas-Nebraska border, similarly does NAM. In general, the SSEF forecast again looks the best.

c. Heavy precipitation case of June 8-9, 2010

For the three 6-hour periods covering the 12-18 hour, 18-24 hour and 24-30 hour precipitation forecasts (Fig. 10 - Fig. 12), we can see that during the first 6 hours (Fig. 10) SSEF PM predicts the heavy prediction pattern the best, although it over-predicts the amount and coverage. The heavy prediction pattern of NAM has a wrong structure and orientation, and the maximum center in SREF is displaced north. During the second 6 hours (Fig. 11), for the main precipitation feature in eastern Kansas and western Missouri, the SSEF PM captures the general structure but the amount is too low. The NAM forecast appears to agree better in that region. Both SREF PM and NAM over-predict the precipitation in north Illinois.

In the last 6 hours (Fig. 12), the SSEF PM captures the wide east-west precipitation band in Kansas and Missouri the best in most aspects. Both SREF and NAM under-predict the precipitation in this region.

The subject evaluations of the three example cases presented in this section clearly demonstrate superior forecasts produced by the convection-allowing storm-scale ensemble, compared to the operational short-range ensemble and 12-km operational north-America mesoscale model NAM; these findings are also supported by quantitative evaluation scores presented in section 2, and in next section. Since these forecasting capabilities are already available, major push should be made by both operational and research communities to implement similar capabilities operationally. The positive economic impact will be huge.

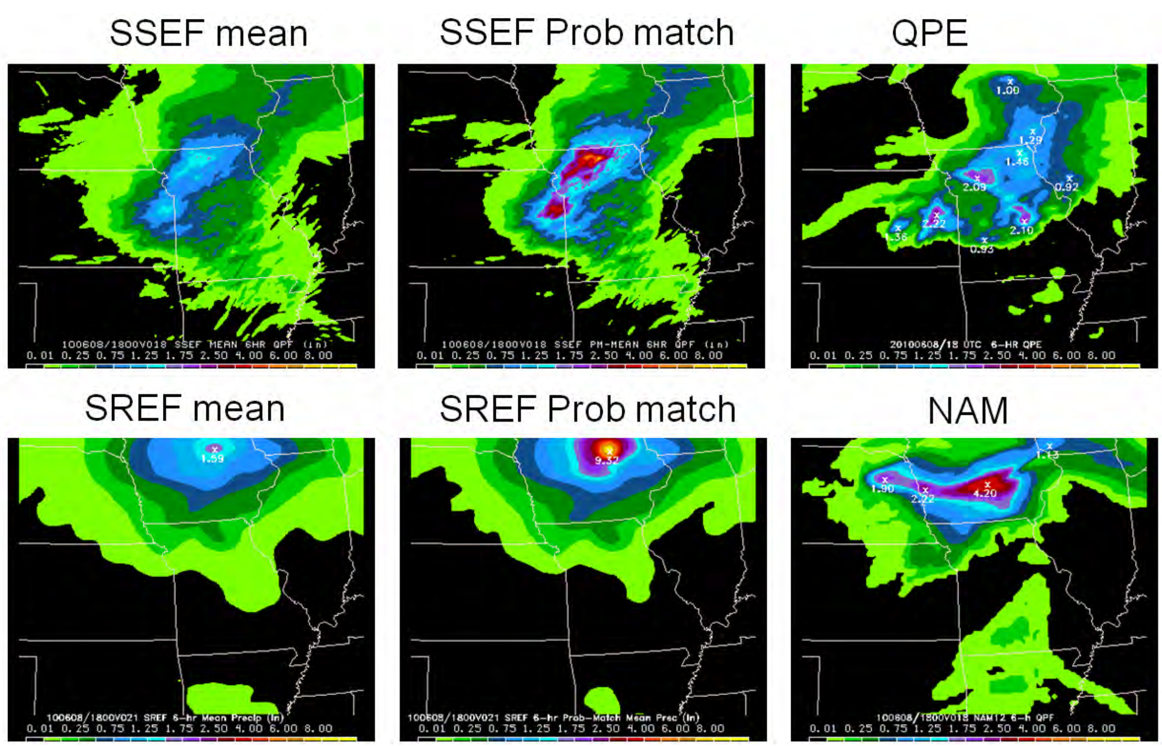

Fig. 10. As in Fig. 6 but valid at 1800 UTC, June 8, 2010, corresponding to 18 hour forecast time.

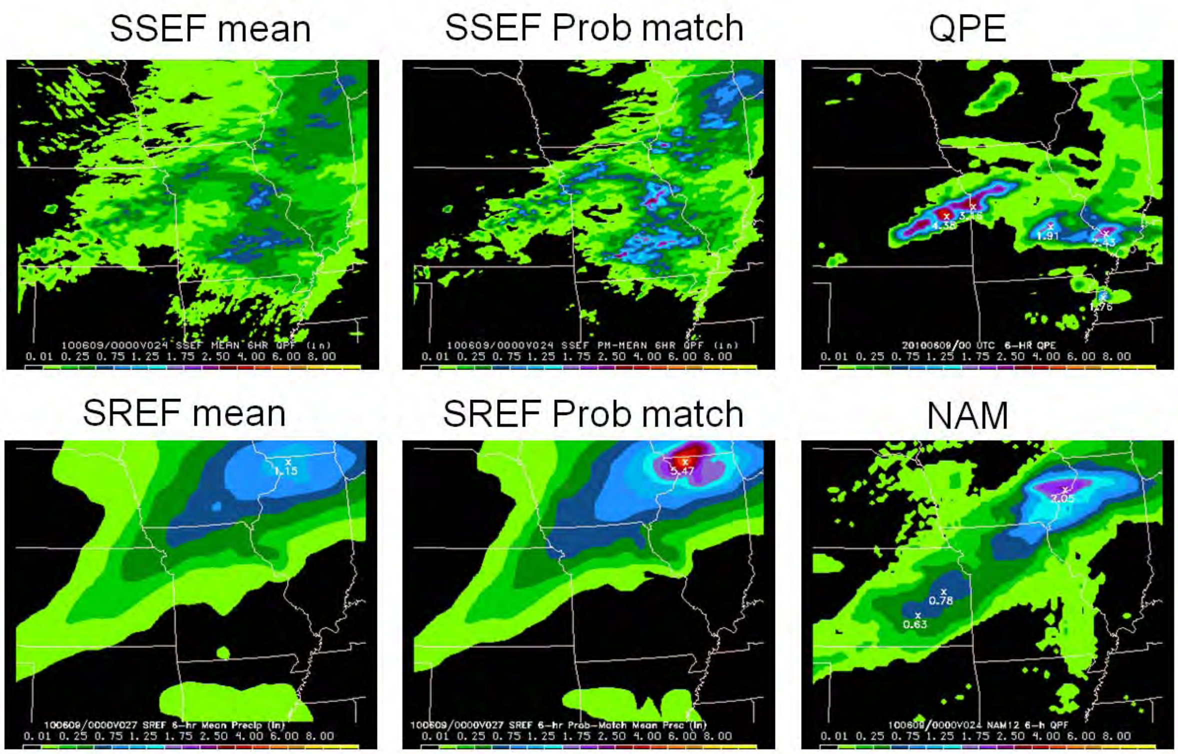

Fig. 11. As in Fig. 6 but valid at 0000 UTC, June 9, 2010, corresponding to 24 hour forecast time.

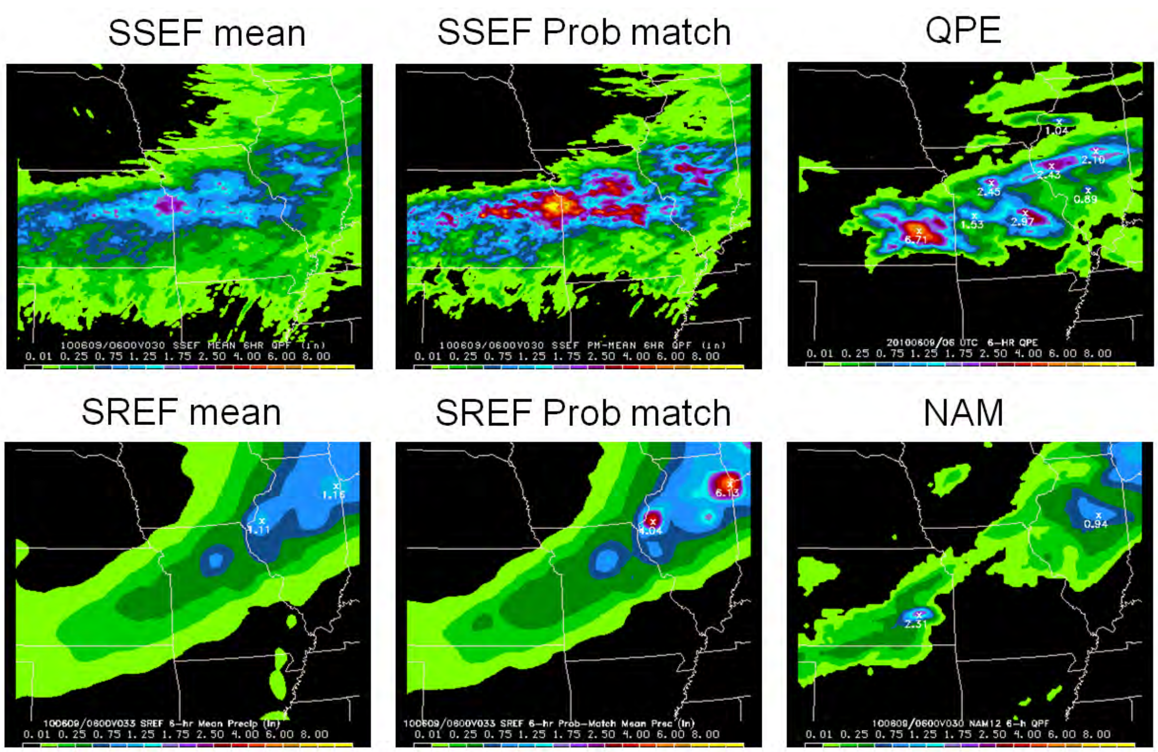

Fig. 12. As in Fig. 6 but valid at 0600 UTC, June 9, 2010, corresponding to 30 hour forecast time.

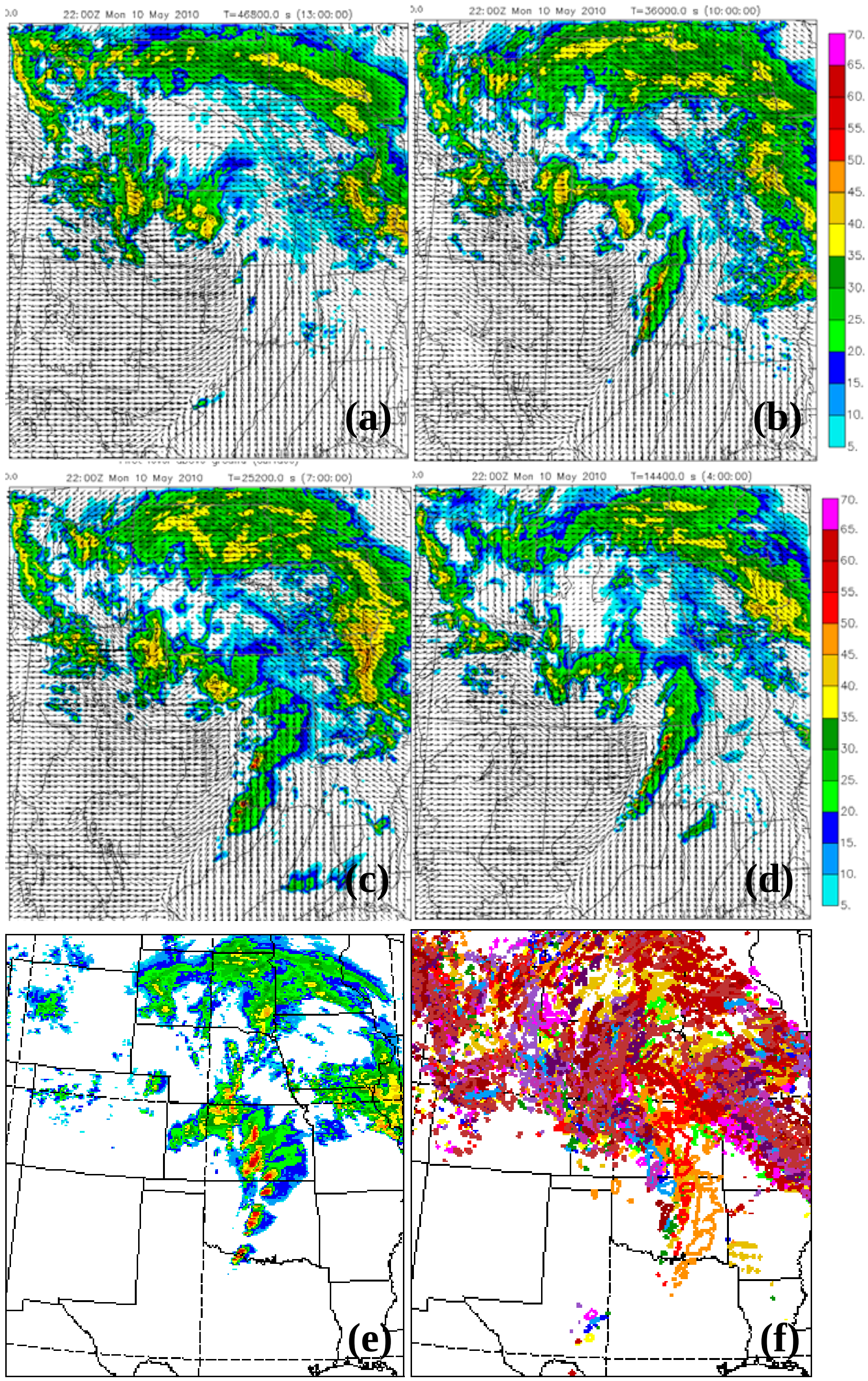

Fig. 13. Panels *(a)-(d): WRF-ARW forecasts for the central U.S. VORTEX domain, initialized at (a) 0900, (b) 1200, (c)1500, and (d) 1800 UTC, corresponding to 13, 10, 7 and 4 hour forecasts, all valid at 2200 UTC, 10 May 2010, the time of a major tornado outbreak of Oklahoma and Kansas. Panel (e) shows the observed composite radar reflectivity and panel (f) shows the spaghetti plots of the 35 dBZ composite reflectivity contours, valid at the same time.*

d. The 10 May 2010 Oklahoma-Kansas tornado outbreak.

A major tornado outbreak occurred on May 10, 2010 affecting large areas of Oklahoma, Kansas, and Missouri, with the bulk of the activity in central and eastern Oklahoma. Over 60 tornadoes, with up to EF4 intensity, affected large parts of Oklahoma and adjacent parts of southern Kansas and Missouri, with the most destructive tornadoes causing severe damage in southern suburbs of the Oklahoma City metropolitan area and just east of Norman, Oklahoma, where the fatalities were reported from both tornado tracks.

An intense trough with dry line activity moved across the southern Plains, especially Oklahoma and Kansas, on May 10. A high risk of severe weather was issued by the Storm Prediction Center, covering portions of Oklahoma and Kansas from the interchange of I-35, I-40 and I-44 in the Oklahoma City area, northeastward into southeastern Kansas and eastern Oklahoma; these areas were in the warm sector. Very strong deep layer wind shear aided by a strong 500 hPa jet streak and mixed-layer CAPE values well in excess of 3000 J kg^{-1}, aided in highly unstable conditions capable of explosive thunderstorm development and the likelihood of strong, long-track tornadoes with any severe thunderstorms.

Among many tornadoes the formed on that day, an intense tornado developed in the southern part of Norman, very close to the National Weather Center that houses CAPS, NSSL and SPC, at 5:32 p.m. CDT (2232 UTC), and initially producing EF1 damages. The tornado later gained EF4 intensity and left a 35 km long damage track on the ground.

Being in the afternoon of the second day relative to the initialization time, most of the CAPS SSEF forecasts did not capture the line of strong cells extending north-south through central Oklahoma but most did forecast those cells in south-central Kansas (Fig. 13e and Fig. 13f). The members that did predict the storm cells within Oklahoma include arw_m7, arw_m9, arw_m10, arw_m11, mmm_cn and mmm_m5, with arw_m10 predicting the strongest cells but displacing them to eastern Oklahoma (Fig. 13f, orange contours). Considering that these forecasts are 22 hours long, the fact that some members were able to predict cells of supercell nature (with high updraft helicity – not shown) is encouraging. In this case, the control members of ARW (arw_cn and arw_c0) only predicted weak cells with reflectivity no exceeding 30 dBz in Oklahoma; this points to the need for an ensemble to capture climatologically low-probability event. The 1-km ARW (control) forecast predicted stronger isolated cells through central Oklahoma although they were somewhat short-lived; the radar-data-assimilating 0000 UTC 11 May 2010 forecasts did capture many of the cells from that time on (not shown), thanks to the radar data assimilation. These results also suggest the challenges associated with forecasting severe weather outbreak nearly one day in advance, when the prediction of the tornadic cells depends strongly on the prediction of the storm environment. Such predictions are apparently sensitive to both initial condition and physics perturbations.

As an effort to support the VORTEX-2 field experiment, CAPS was also running a set of smaller domain WRF forecasts initialized at 0900, 1200, 1500, 1800 and 2100 UTC every day, that were configured the same way as arw_cn and arw_c0, at 4-km resolution, except for the domain size and initialization time. Fig. 13 shows that a line of cells was captured well at 2200 UTC, in the forecasts initialized at 1200, 1500 and 1800 UTC (Fig. 13b-d), but not in the forecast of 0900 UTC (Fig. 13a). It would be interesting to compare the 3-hour forecast from 0900 UTC valid at 1200 UTC, with the 1200 UTC initial condition to identify the key differences that caused the large differences in forecast. At 1200 UTC, convection was minimal in central and north Oklahoma so radar data assimilation should not have made a large impact. Most likely, the 1200 UTC and later forecasts benefited from observations at this synoptic time. These results also point to the need for frequently updated forecasts, a direction we are moving towards in future experiments (see the final section).

6. Preliminary objective evaluations

In section 2, we have presented ETS scores for 3-hour accumulated precipitation for a given threshold, including for the forecasts of 2010. Consistent with the results of earlier years, the probability-matched ensemble mean derived from the radar-assimilation members has the highest scores. This can be seen in the hourly precipitation ETS scores also (Fig. 14). The ETS score differences for the three different models are largest between 3 and 9 hours. The WRF-ARW scores appear to be among the highest group while those of WRF-NMM among the lowest group, with the ARPS radar assimilating control in-between the two groups. After 18 hours, their scores become indistinguishable but by this time, because of the difficulty in predicting precipitation overlap with observations, the ETS score is no longer an effective measure. We note here that for the forecasts of first 2-3 hours, there appear to be 2-3 outliers among the radar assimilating members that have particularly low ETS scores. Two of them are the ARW members with Ferrier microphysics. We suspect there exists inconsistency between the ARPS cloud analysis package used to initialize the cloud hydrometeor fields and the Ferrier microphysics. Further investigation is needed.

The lower panel of Fig. 14 plots the frequency bias of hourly precipitation exceeding 0.1 inch. As was noted with CAPS forecasts of past springs (Kong et al. 2008; Kong et al. 2009), significant positive bias usually exists with the 4-km ensemble. During the first 6 hours, positive biases with the radar-assimilating members are between 50 and 100% at their peak, while the non-radar members have low bias during the first 3 hours because they need time to spin up precipitation. Between 6 and 15 hours, for most members, the positive bias is actually not bad, generally below 30%. Two members appear to be outliers between 3 and 18 hours, giving high biases close to 100%. These two are the NMM members using the RRTM-Dudhia radiation. How radiation affects the precipitation bias so much requires investigation.

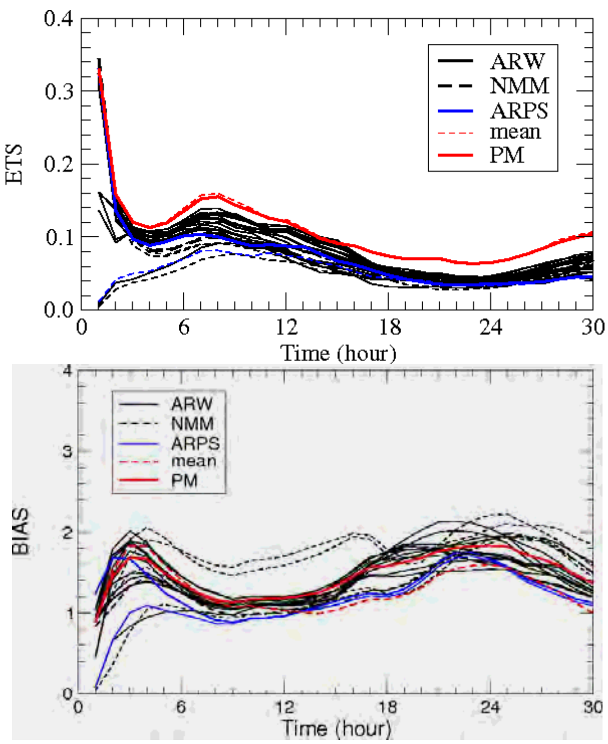

Fig. 14. (Upper panel) ETS scores of hourly precipitation exceeding 0.1 inch, and (lower panel) the frequency biases for the same precipitation threshold, for the 2010 CAPS forecasts.

As an ensemble forecasting system, the probalistic prediction skills of the system need to be carefully evaluated. Many probalistic skill scores can be examined to evaluate different properties of the ensemble system. This part of the work is ongoing, and we present here only a few.

Fig. 15 shows the verification rank histogram of hourly precitation forecasts valid at 24 hours, averaged over all cases of 2010. The relative flatness of the histogram suggests that the dispersion of the ensemble is reasonable while the skewness of the histogram indicates significant over-prediciton, which is consistent with the positve biases observed earlier. Kong et al. (2008) applied a rank-based bias correction procedure that reduced the skewness of the histogram but did not found improvement to reliability.

The probabilistic quantitative precipitation forecast (PQPFs) can be evaluated using the area under the relative operating characteristic curve (ROC area, Mason 1982). As discussed in Clark et al. (2010c), the ROC area measures ability to distinguish between events and non-events and is closely related to the economic value of a forecast system (Richardson 2000). The ROC area is calculated by computing the area under a curve constructed by plotting the probability of detection (POD) against the probability of false detection (POFD) for specified ranges of PQPFs. The range of ROC area is 0 to 1, with 1 a perfect forecast and areas greater than 0.5 having positive skill. A ROC area of 0.7 is generally considered the lower limit of a useful forecast (Buizza 1997).

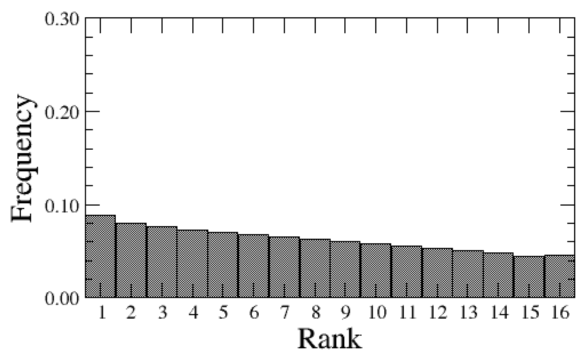

Fig. 15. Verification rank histogram of 1 h accumulated precipitataion for the forecast hour 24, from the CAPS 2010 SSEF.

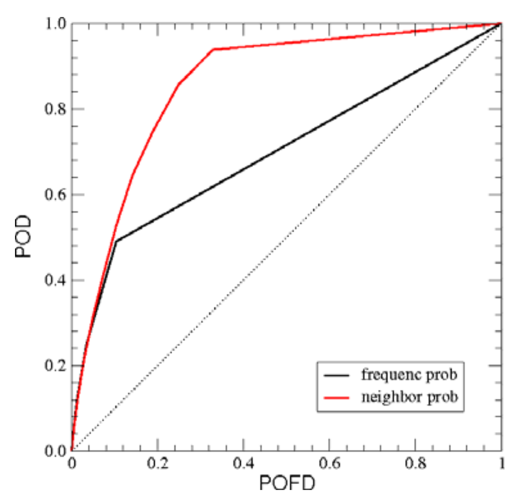

Fig. 16. ROC (Relative Operation Characteristics) curves for hourly precipitation exceed 0.25 inch, valid at 24 hours. The red curves are for

neighborhood probability where a neighborhood radius of 40 km was used.

Recognizing that small displacement errors in high-resolution precipitation forecasts can cause serious decrease in grid point-based skill scores, "neighborhood" methods have been developed that relax the requirement that the model output and corresponding observations match exactly in order for a forecast to be considered correct (Theis et al. 2005). These "neighborhood" approaches have also been used to generate probabilistic information from deterministic grids and applied to the CAPS 2007 data (Schwartz et al. 2010). Theis et al. (2005) suggested that a neighborhood approach could be combined with traditional methods of producing probabilistic forecasts – this approach is tested here. In the 2010 season, neighborhood probabilities were calculated by using a 40 km neighborhood radius and ROC curves are plotted in Fig. 16 and Fig. 17 based on this neighborhood probability (red lines) as well as the simple frequency-based probability. It can be seen that the use of neighborhood probability increases the ROC area by about 0.15. Without the use of neighborhood probability, the ROC area is between 0.7 and 0.75, indicating useful forecast already. The use of neighborhood probability does not improve the Brier skill scores or reliability, however (not shown). Further investigation is needed. Effective bias removal and ensemble calibration are most likely needed first.

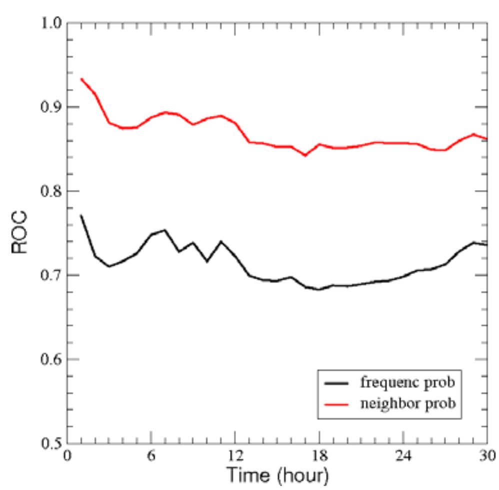

Fig. 17. ROC areas for hourly precipitation exceed 0.25 inch, throughout the forecast period.

7. Future plan

The CAPS SSEF experiment will continue over the next few years, and additional funding support is being sought to help optimize the design of SSEF and to allow for the research and development needed for producing calibrated storm-scale probabilistic products. Because of significant precipitation biases found in the forecasts of all previous years, bias removal is essential for producing reliable probabilistic precipitation forecasts. Bias removal methods suitable for convection-resolving forecasts have to be developed. Because of the need for more cases from the same ensemble system for bias removal and ensemble calibration, we plan to keep the 15 core ensemble members that same as those of 2010 next year. Additional forecasts may be run for testing new capabilities and schemes, and for investigating other issues.

In 2011, the Navy's COAMPS model will also be added to the multi-model ensemble framework, while in future years, more frequent forecasts using advanced ensemble Kalman filter (EnKF) data assimilation and initial perturbation methods are among the important goals. Closer collaborations will also occur between CAPS, HWT and DTC, and with other NCEP Centers including HPC and AWC, as has already occurred in 2010. Funding is also being sought to produce synthetic satellite images from the ensemble for GOES infrared channels, and for future GOES-R ABI instrument channels, in collaboration with CIMSS at the University of Wisconsin and CIRA at the Colorado State University, and the NESDIS, and the products will be evaluated at the HWT as part of the GOES-R Proving Ground activities. The availability of an ensemble of storm-scale forecasts involving multiple models with multiple microphysics parameterization schemes offer an unprecedented opportunity for producing an ensemble of synthetic satellite data that would allow for the study of radiation transfer model (RTM) and microphysics interaction and sensitivity, evaluation of the utility and potential of future GOES-R products, improvement to the RTMs in the presence of cloud and precipitation, and the evaluation of convection-resolving forecasting performance and development of probabilistic forecasting products in terms of satellite observables. The experience learned with the radiative transfer models can help improve satellite data assimilation. With the recent development of an efficient parallel version of a radar-assimilating EnKF system at CAPS, and the potential availability of even larger supercomputers, we believe continental-scale realtime ensemble data assimilation and forecasting at a convection-allowing resolution is quite possible within several years.

Acknowledgement: This research was mainly supported by a grant to CAPS from the NOAA CSTAR program. Supplementary support was also provided by NSF AGS-0802888, and by other grants at CAPS. Suggestions and input from NCAR scientists Drs. Jimy Dudhia, Morris Weisman, Greg Thompson and Wei Wang on the WRF model configurations were very helpful. The CONUS-domain forecasts presented here were produced at the National Institute of Computational Science (NICS) at the University of Tennessee, and at the Pittsburgh Supercomputing Center (PSC, for the previous years). Facilitation and support from Bruce Loftis and Phi Andrew of NICS are deeply appreciated, so are the incredible technical support provide by a NICS support team, including Kwai Wong, Troy Baer, Alan Blatecky, Barry Schneider, and Ed Seidel. David Oneal of PSC provided valuable support for forecasts produced at PSC. A supercomputer of the Oklahoma Supercomputing Center for Research and Education (OSCER) was used for running the VORTEX-2 domain forecasts; excellent support from OSCER system administrators is equally appreciated. Graphics in Figs. 6 through 12 were taken from the HWT's realtime forecast website – additional people contributed to the development of the website.

References

Berenguer, M., M. Surcel, I. Zawadzki, M. Xue, and F. Kong, 2010: The diurnal cycle of precipitation from continental radar mosaics and numerical weather prediction models. Part II: Intercomparison between numerical models and with nowcasting. *Mon. Wea. Rev.***,** Submitted.

Buizza, R., 1997: Potential Forecast Skill of Ensemble Prediction and Spread and Skill Distributions of the ECMWF Ensemble Prediction System. *Mon. Wea. Rev.*, **125,** 99-119.

Bunkers, M. J., B. A. Klimowski, J. W. Zeitler, R. L. Thompson, and M. L. Weisman, 2000: Predicting supercell motion using a new hodograph technique. *Wea. Forecasting*, **15,** 61-79.

Clark, A. J., W. A. Gallus, Jr., M. Xue, and F. Kong, 2009: A comparison of precipitation forecast skill between small convection-permitting and large convection-parameterizing ensembles. *Wea. Forecasting*, **24,** 1121-1140.

Clark, A. J., W. A. Gallus, Jr., M. Xue, and F. Kong, 2010a: Growth of spread in convection-allowing and convection-parameterizing ensembles *Wea. Forecasting*, **25,** 594-612.

Clark, A. J., W. A. Gallus, Jr., M. Xue, and F. Kong, 2010b: Convection-allowing and convection-parameterizing ensemble forecasts of a mesoscale convective vortex and associated severe weather. *Wea. Forecasting*, **25,** 1052-1081.

Clark, A. J., J. S. Kain, D. J. Stensrud, M. Xue, F. Kong, M. C. Coniglio, K. W. Thomas, Y. Wang, K. Brewster, J. Gao, S. J. Weiss, D. Bright, and J. Du, 2010c: Probabilistic precipitation forecast skill as a function of ensemble size and spatial scale in a convection-allowing ensemble. *Mon. Wea. Rev.***,** Submitted.

Coniglio, M. C., K. L. Elmore, J. S. Kain, S. Weiss, M. Xue, and M. L. Weisman, 2010: Evaluation of WRF model output for severe-weather forecasting from the 2008 NOAA Hazardous Weather Testbed Spring Experiment. *Wea. Forecasting*, **25,** 408-427.

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.

Ebert, E. E., 2001: Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Mon. Wea. Rev.*, **129,** 2461-2480.

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., M. Xue, and K. Brewster, 2006: 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.

Janjic, Z., 2003: A nonhydrostatic model based on a new approach. *Meteo. Atmos. Phys.*, **82,** 271-286.

Jensen, T., S. Weiss, J. Kain, M. Harrold, M. Xue, F. Kong, B. Brown, P. Marsh, M. Coniglio, and R. Schneider, 2010: The Developmental Testbed Center (DTC) Objective Evaluation Performed During the Hazardous Weather Testbed (HWT) 2010 Spring Experiment. *Extended Abstract, 11th WRF Users' Workshop*, NCAR, Boulder CO.

Kain, J. S., M. Xue, M. C. Coniglio, S. J. Weiss, F. Kong, T. L. Jensen, B. G. Brown, J. Gao, K. Brewster, K. W. Thomas, Y. Wang, C. S. Schwartz, and J. J. Levit, 2010: Assessing advances in the assimilation of radar data within a collaborative forecasting-research environment. *Wea. Forecasting***,** In press.

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 forecast experiment - Analysis of 2008 spring experiment data. *24th Conf. Several Local Storms*, Savannah, GA, Ameri. Meteor. Soc., Paper 12.3.

Kong, F., M. Xue, K. Thomas, Y. Wang, K. A. Brewster, J. Gao, K. K. Droegemeier, J. S. Kain, S. J. Weiss, D. R. Bright, M. C. Coniglio, and J. Du, 2009: A real-time storm-scale ensemble forecast system: 2009 Spring Experiment. *23nd Conf. Wea. Anal. Forecasting/19th Conf. Num. Wea. Pred.*, Omaha, Nebraska, Amer. Meteor. Soc., Paper 16A.3.

Mason, I. B., 1982: A model for the assessment of weather forecasts. *Aust. Meteor. Mag.*, **30,** 291-303.

Morrison, H., J. A. Curry, and V. I. Khvorostyanov, 2005: A new double-moment microphysics parameterization for application in cloud and climate models. Part I: Description. *J. Atmos. Sci.*, **62,** 1665-1677.

Nakanishi, M. and H. Niino, 2004: An Improved Mellor--Yamada Level-3 model with condensation physics: Its design and verification. *Bound - Layer Meteor.*, **112,** 1-31.

Nakanishi, M. and H. Niino, 2006: An Improved Mellor--Yamada Level-3 Model: Its numerical stability and application to a regional prediction of advection fog. *Bound - Layer Meteor.*, **119,** 397-407.

Richardson, D. S., 2000: Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Quart. J. Roy. Meteor. Soc.*, **127,** 2473-2489.

Schwartz, C., J. Kain, S. Weiss, M. Xue, D. Bright, F. Kong, K. Thomas, J. Levit, and M. Coniglio, 2009: Next-day convection-allowing WRF model guidance: A second look at 2 vs. 4 km grid spacing. *Mon. Wea. Rev.*, **137,** 3351-3372.

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, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea Forecasting*, **25,** 263-280.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. D. Powers, 2005: A Description of the Advanced Research WRF Version 2, 88 pp.

Sukoriansky, S., B. Galperin, and I. Staroselsky, 2005: A quasi-normal scale elimination model of turbulent flows with stable stratification. *Phys. Fluids,* **17,** 085107–1–28.

Sukoriansky, S., B. Galperin, and V. Perov, 2006: A quasi-normal scale elimination model of turbulence and its application to stably stratified flows. *Nonlinear Processes Geophys.*, **13,** 9-22.

Theis, S. E., A. Hense, and U. Damrath, 2005: Probabilistic precipitation forecasts from a deterministic model: A pragmatic approach. *Meteor. Appl.*, **12,** 257-268.

Thompson, R. L., R. Edwards, and J. A. Hart, 2002: Evaluation and interpretation of the supercell composite and significant tornado parameters at the Storm Prediction Center. *Preprints, 21st Conf. on Severe Local Storms,* Amer. Meteor. Soc., J11-J14.

Thompson, R. L., R. Edwards, and C. M. Mead, 2004: An update to the supercell composite and significant tornado parameters. *Preprints, 22nd Conf. on Severe Local Storms,* Amer. Meteor. Soc., P8.1.

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., K. K. Droegemeier, and V. Wong, 2000: The Advanced Regional Prediction System (ARPS) - A multiscale nonhydrostatic atmospheric simulation and prediction tool. Part I: Model dynamics and verification. *Meteor. Atmos. Physics*, **75,** 161-193.

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.*, Amer. Meteor. Soc., CDROM 3B.1.

Xue, M., F. Kong, K. W. Thomas, J. Gao, Y. Wang, K. Brewster, K. K. Droegemeier, J. Kain, S. Weiss, D. Bright, M. 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. *24th Conf. Several Local Storms*, Savannah, GA, Ameri. Meteor. Soc., Paper 12.2.

Xue, M., F. Kong, K. W. Thomas, J. Gao, Y. Wang, K. Brewster, K. K. Droegemeier, X. Wang, J. Kain, S. Weiss, D. Bright, M. Coniglio, and J. Du, 2009: CAPS realtime multi-model convection-allowing ensemble and 1-km convection-resolving forecasts for the NOAA Hazardous Weather Testbed 2009 Spring Experiment. *23nd Conf. Wea. Anal. Forecasting/19th Conf. Num. Wea. Pred.*, Omaha, NB, Amer. Meteor. Soc., Paper 16A.2.