24th Conference on Weather and Forecasting/20th Conference on Numerical Weather Prediction. AMS, Seattle, WA, January 23-27, 2011.

457 EVALUATION OF CAPS MULTI-MODEL STORM-SCALE ENSEMBLE FORECAST FOR THE NOAA HWT 2010 SPRING EXPERIMENT

Fanyou Kong[1]*, Ming Xue[1,2], Kevin W. Thomas[1], Yunheng Wang[1],
Keith Brewster[1], Xuguang Wang[1,2], Jidong Gao[1],
Steven J. Weiss[3], Adam Clark[4], Jack Kain[4], Michael C. Coniglio[3], and Jun Du[5]
[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]NOAA National Severe Storm Laboratory, Norman, OK 73072
[5]NOAA/NWS/NCEP, Camp Springs, MD 20746

1. INTRODUCTION

As part of the NOAA Hazardous Weather Testbed (HWT) 2010 Spring Experiment, the Center for Analysis and Prediction of Storm (CAPS) once again produced multi-model storm-scale ensemble forecasts (SSEF) in realtime from 26 April through 18 June. Compared to previous experiment seasons (Kong et al. 2007, 2008, 2009; Xue et al. 2007, 2008, 2009), the 2010 forecast domain was expanded to cover the full continental United States (Figure 1) and the number of 4-km ensemble members was increased from 20 to 26. The 2010 SSEF included 19 WRF-ARW, 5 WRF-NMM, and 2 ARPS members; and WRFV3.1.1 was used for both dynamic cores. New in 2010 Spring Experiment was the generation of a large set of post-processed ensemble products from a 15-member sub-ensemble that consists of the multi-physics, IC and LBC perturbation, and radar analysis members. Post-processed products include ensemble mean and maximum, probability matching mean, frequency-based ensemble probability, and neighborhood probability of accumulated precipitation, reflectivity, environmental fields such as surface temperature, dew point, maximum downdraft speed and maximum 10-m wind speed, CAPE-shear parameters, and storm-attribute parameters such as updraft helicity, updraft speed, and integrated graupel. For all members but three (one from each model), full-resolution radar data from the nationwide WSR-88D radar network (both reflectivity and radial wind) were analyzed into the ICs using the ARPS 3DVAR and cloud analysis package (Gao et al. 2004; Hu et al. 2006). 30-h forecasts, initiated at 0000UTC, were produced daily on weekdays from April 26 to June 18 on a Cray XT4 supercomputer with over 18,000 cores at the National Institute of Computing Science (NICS) of the University of Tennessee. The post-processed ensemble products were made available in real-time in GEMPAK format data files and on the web for the SPC, NSSL, DTC, and HPC, and evaluated daily by HWT participants.

A total of 36 days of complete ensemble forecasts were produced during the experiment period. Using the NSSL 1km resolution QPE data as verification dataset, the SSEF QPF and probabilistic QPF performance has been evaluated using various traditional verification metrics and compared to the operational 12-km NAM forecasts and the 32-km Short-Range Ensemble Forecast (SREF) products.

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

2. EXPERIMENT HIGHLIGHT

The CAPS 2010 Spring Program started on 26 April 2010 and ended on 18 June, encompassing the NOAA HWT 2010 Spring Experiment that is officially between 17 May and 18 June. This experiment period is shifted into mid-June to accommodate Aviation Weather Testbed activity. Three numerical weather models are used to produce a 26 member 30 h ensemble forecast during weekdays, initialized at 0000 UTC, covering a full-CONUS domain (Figure 1) at 4 km horizontal grid spacing. Nineteen members are produced using the Weather Research and Forecast (WRF) Advanced Research WRF core (ARW), five members are produced using the WRF Nonhydrostatic Mesoscale Model core (NMM), and two members are produced using the Advanced Regional Prediction System (ARPS). Both WRF cores are the V3.1.1 release.

All forecasts use NAM 12 km (218 grid) 00Z analyses as background for initialization with the initial condition perturbations for the ensemble members coming from the NCEP Short-Range Ensemble Forecast (SREF). Doppler radar radial wind and reflectivity data from over 140 available WSR-88D stations within the domain are assimilated through ARPS 3DVAR and Cloud Analysis package into all but three members (one from each model group).

All 4 km ensemble forecast, along with one single deterministic 1 km forecast, are performed on *Athena*, a Cray XT4 supercomputer system with 18000+ computing cores, at the NSF sponsored National Institute of Computational Sciences (NICS) in the University of Tennessee. This allows the entire forecasts – 26 ensemble runs at 4 km grid and 1 deterministic run at 1 km grid – to use the entire machine in dedicated mode overnight, bringing the total forecast walltime down to within 6 h. Hourly model outputs are archived on the mass storage HPSS at NICS.

Figure 1 shows the coverage area of the model domains.

Since NMM uses rotated E-grid while both ARW and ARPS use C-grid, special software codes were developed at CAPS to convert between the two grids in order to utilize a single 3DVAR/Cloud Analysis over a larger outer domain that encompasses both forecast

domains (Figure 1) by converting the analysis to the forecast domains, and to convert NMM forecast to a common verification domain that is the same as the ARW and ARPS forecast domain. These special codes were upgraded to be compatible with the new WRF version (V3.1.1) used in 2010 season.

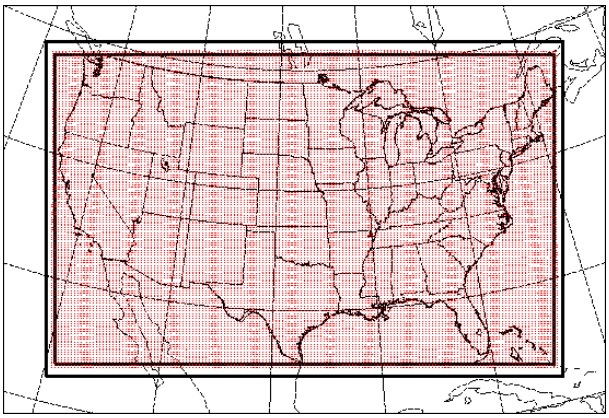

Figure 1. Computational domains for the 2010 Spring Experiment. The outer thick rectangular box represents the domain for performing 3DVAR/Cloud Analysis (1200×780). The red dot area represents the WRF-NMM forecast domain (790×999). The inner thick box is the domain for WRF-ARW and ARPS forecast and also for common verification (1160×720).

Tables 1-3 highlight 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 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 Gaussian type perturbation and another is Gaussian perturbation plus a recursive filter (RF) with convective storm scale length. The physics-perturbation-only members are added to help assessing impacts from different 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 the 2009 season.

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

member	IC	BC	Radar data	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 + RF smoothed 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 data	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 data	Microphysics	radiation	PBL	Turb	sf_phy
arps_cn	00Z ARPSa	00Z NAMf	yes	Lin	Chou/Suarez	TKE	3D TKE	Force-restore
arps_c0	00Z NAMa	00Z NAMf	no	Lin	Chou/Suarez	TKE	3D TKE	Force-restore

* For all members: no cumulus parameterization

3. ENSEMBLE PRODUCTS & EXAMPLES

Post-processed ensemble products from a subset of 15 out of 26 ensemble members (red colored in Table 1-3), representing multi-model, IC perturbation, and physics variation ensemble with influence of radar data assimilation, are generated in near-realtime. The products include ensemble maximum and mean, probability matching 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 grapeul/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, 2004).

Both individual member and ensemble products are 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, represents only a sub-domain that covers the eastern 2/3 of CONUS (Figure 2). CAPS keeps a complete set of 2D variables in full domain in HDF4 format, extracted from the realtime forecasts.

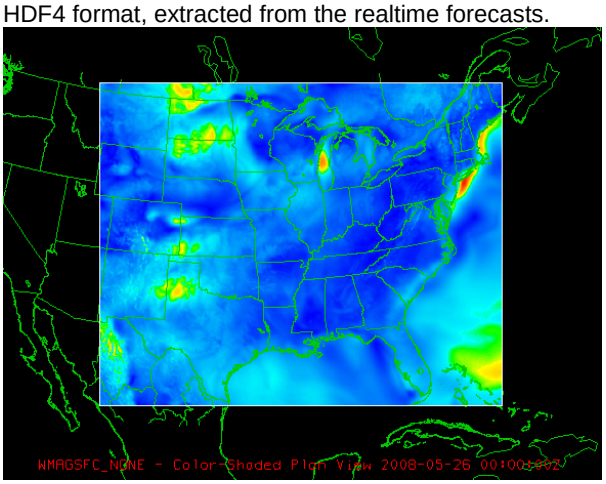

Figure 2. SPC/NSSL sub-domain for the HWT GEMPAK dataset (860x690).

Xue et al. (2010) presents detail description for the CAPS involvement in the NOAA 2010 HWT Spring Experiment, with many example forecast cases. Figure 3 is an example of May 10 Oklahoma Tornado case, showing some post-processed ensemble products that can provide valuable forecast guidance long before the tornado occurrence. 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 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.

In Figure 3a, the red circle highlights the tornado bearing storm system around the tornado occurrence time at 22 UTC May 10, 2010. Most ensemble members, including the ARW control member (cn), missed the line of supercells that produced tornados, except only a few (Figure 3c). However, good indications of environment favoring tornadic activities are predicted 22 h in advance from ensemble products such as the neighborhood probability of hourly maximum updraft helicity – a product of 3-5 km updraft velocity and vertical vorticity – exceeding 25 m^2/s^2 (Figure 3b), and the probability of Significant Tornado Parameter (STP) – a composite parameter calculated based on 0-6 km bulk wind difference, 0-1 km storm relative helicity, surface-based CAPE, LCL and CIN – exceeding 3 (Figure 3d). Majority of reported F2 and above tornados were associated with STP exceeding 1 (Thompson et al. 2002, 2004).

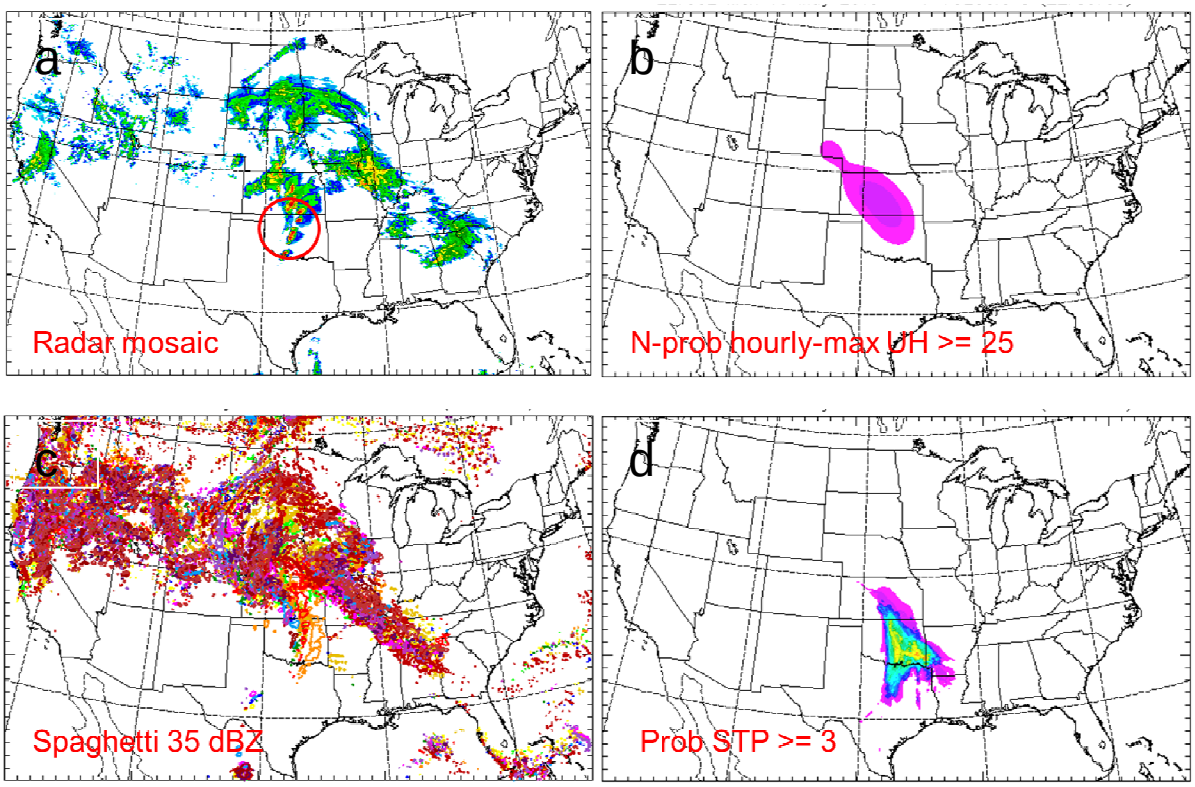

Figure 3. Observed composite reflectivity mosaic, valid at 22 UTC May 10, 2010, and the 22 h forecast of the neighborhood probability of the hourly maximum updraft helicity exceeding 25 m^2/s^2 (a), spaghetti contours off composite reflectivity equal to 35 dBZ (c), and the probability of STP exceeding 3 (d).

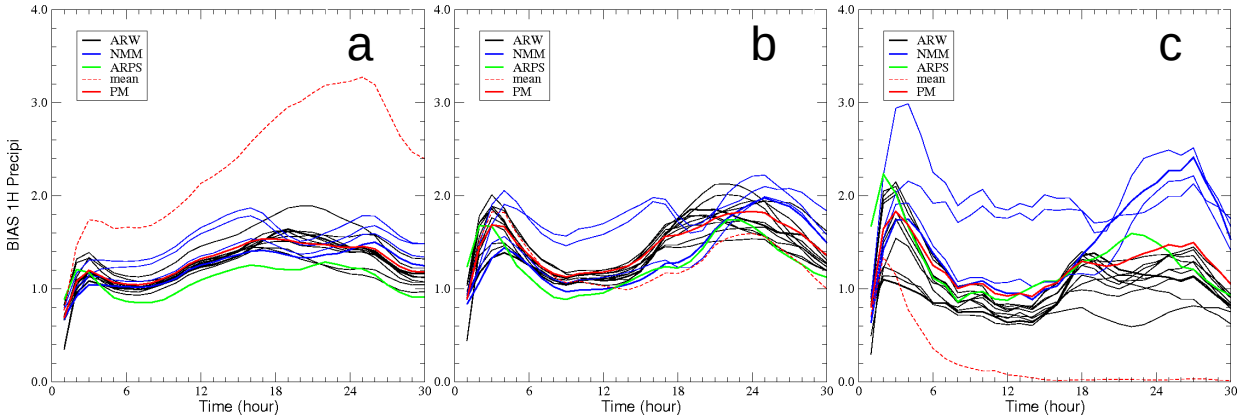

Figure 4. Frequency biases of 1-h accumulated precipitation with the thresholds of (a) 0.01 inch, (b) 0.1 inch, and (c) 0.5 inch, averaged over 36 days. PM refers to the probability matched mean.

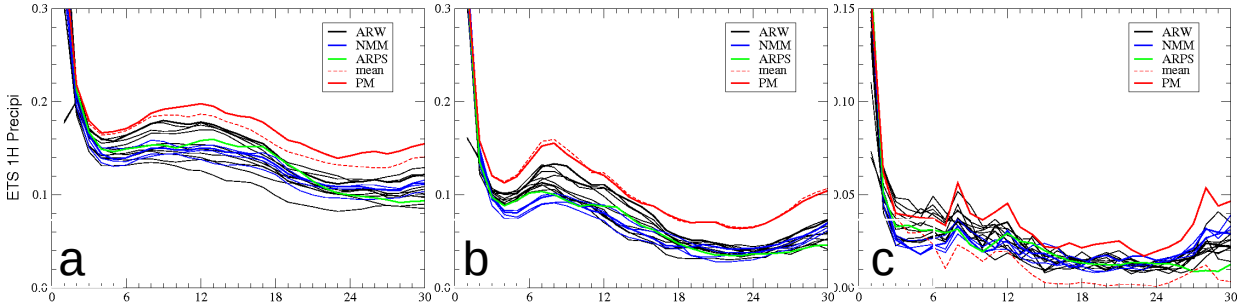

Figure 5. Equitable Threat Scores (ETS) of 1-h accumulated precipitation with the thresholds of (a) 0.01 inch, (b) 0.1 inch, and (c) 0.5 inch, averaged over 36 days.

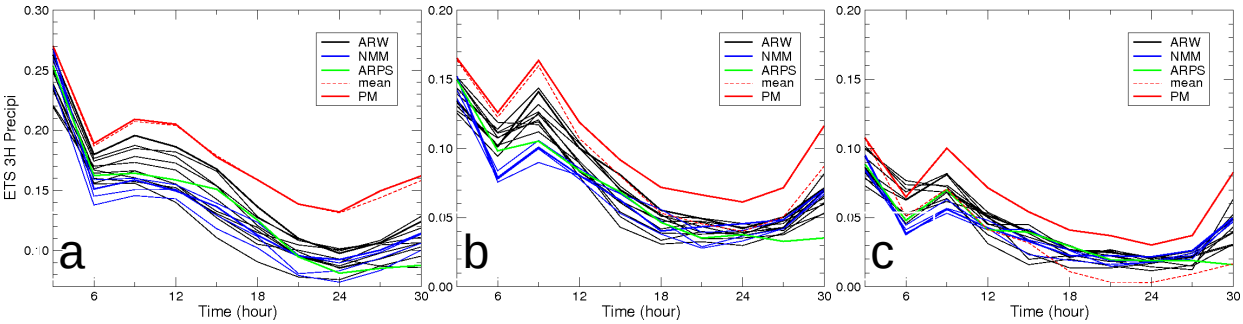

Figure 6. Equitable Threat Scores (ETS) of 3-h accumulated precipitation with the thresholds of (a) 0.1 inch, (b) 0.5 inch, and (c) 1.0 inch, averaged over 36 days.

4. STATISTICAL EVALUATION OF QPF

4.1 BIAS and ETS

The NSSL NMQ precipitation estimation product (QPE) has been used in this study to verify the QPF and PQPF from the CAPS SSEF. The verifications are performed over the entire verification domain (see Figure 1). Figure 4 shows the frequency biases (BIAS) of the 1-h accumulated precipitation for the thresholds of 0.01, 0.1, and 0.5 inch, respectively, for the 15-member sub-ensemble, averaged over the 36 days that have complete ensemble data through the experiment period. As found in previous experiment years (Kong et al. 2007, 2008), the ensemble means have the highest positive bias for the light rain threshold (over 200% around 24 h) and have very significant underforecast for the heavy rain threshold through most of forecast duration. Among models, while NMM members tend to overforecast precipitation areas, with the two highest lines between 0-18 h associated with members using non-operational radiation options (RRTM-Dudhia), the ARPS member generally has the least BIAS, except for the heavy rain thresholds.

Figure 5 shows the Equitable Threat Scores (ETS) of the 1-h accumulated precipitation for the same

thresholds as in Figure 4. The probability matched means (PM) outscore all individual members, as well as the simple ensemble mean except for the 0.1 inch threshold where the two are comparable. The later scores the worst for the heavy rain threshold. Figure 6 shows the ETS scores for the 3-h accumulated precipitation, which again exhibits that PMs have the best performance with respect to the deterministic QPF sense measured by the traditional ETS metrics.

4.2 BS and ROC

The areas under the Relative Operating Characteristic curves (ROC) measure the ability of probabilistic forecasts to distinguish between event and non-event, and are insensitive to forecast biases (Mason 1982; Clark et al. 2010). 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). Figure 7 shows the ROC areas of the frequency-based ensemble probability forecast of the 1-h accumulated precipitation. It can be seen that all three threshold curves are above 0.6, while the 0.01 and 0.25 inch curves stay above 0.7.

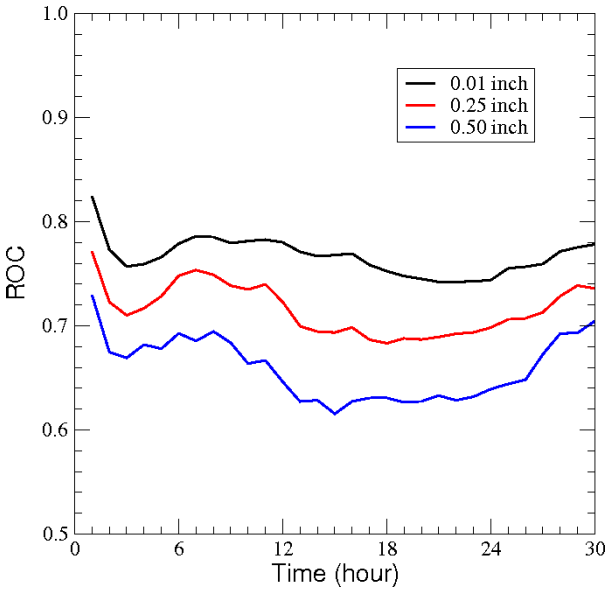

Figure 7. ROC areas of the frequency-based probability of the 1-h accumulated precipitation for various thresholds.

Recognizing that small displacement errors in high-resolution precipitation forecasts can cause serious decrease in grid point-based skill scores, in the 2010 season neighborhood probabilities of certain QPF fields were also calculated by using a 40 km radius of influence (ROI) with a 2D Gaussian smoother using a σ value of 10. Figure 8 compares the simple frequency based probability (black lines) and the neighborhood probability (red lines) in terms of ROC. It can be seen that the use of neighborhood probability increases the ROC area by about 0.15. Even without the use of neighborhood probability, the ROC area is well above 0.7 for the threshold, indicating useful forecast already.

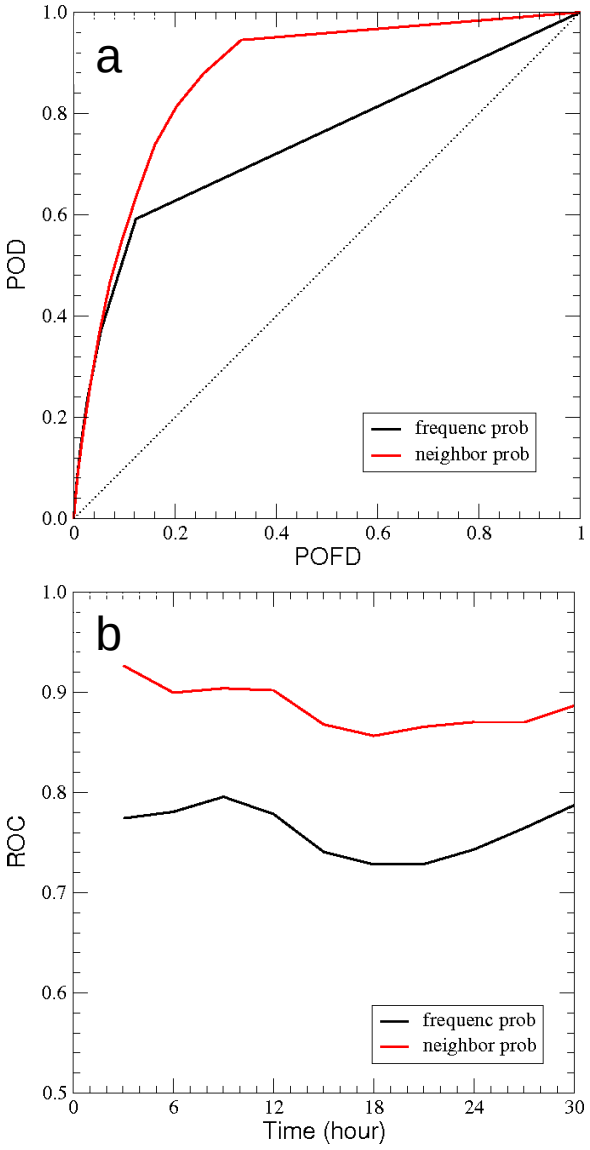

Figure 8. ROC curves of frequency (black line) and neighborhood (red line) probabilities of 3-h accumulated precipitation exceeding 0.5 inch at 24 h (a) and the ROC areas (b).

ROC is a measure of the resolution for a probabilistic forecast, not the reliability. The use of neighborhood probability does not improve the Brier skill scores and reliability. In fact, extra biases are introduced through the neighborhood and smoothing process, further affecting reliability. Figure 9 compares the Briers scores between the simple frequency probability and the

neighborhood probability of the 3-h accumulated precipitation exceeding 0.5 inch, showing worse scores for the neighborhood probability throughout the forecast duration. Discussions among co-authors indicate the need to apply the same neighborhood technique (without Gaussian smoothing) to the QPE fields before doing verification. Further investigation is needed. This suggests the importance of effective bias removal and ensemble calibration in order to produce highly skillful probabilistic QPFs from SSEF.

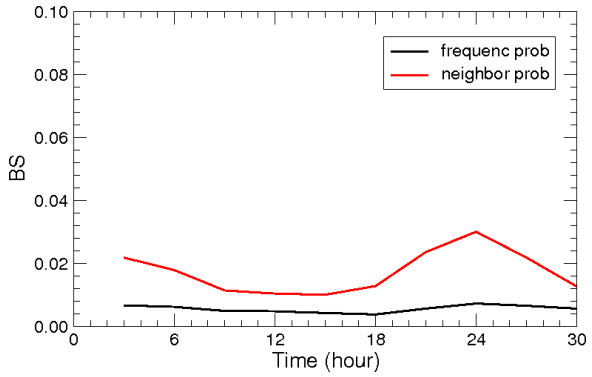

Figure 9. Briers score of the frequency (black line0 and neighborhood (red line) probabilities of the 3-h accumulated precipitation exceeding 0.5 inch.

4.3 SSEF vs NAM and SREF

It is interesting to directly compare the CAPS storm-scale ensemble forecast (SSEF) QPF/PQPF, without any bias removal and calibration, against those from the operationally available NCEP 12 km NAM (deterministic) and the coarser resolution short-range ensemble forecast (SREF, probabilistic). Precipitation forecasts from NAM and SREF, over the same 36 days as in SSEF dataset, are first interpolated to the 4km verification grid (see Figure 1). For SREF, only the 11 members that are used for providing LBC perturbations for SSEF are included in the comparison.

Figure 10 plots the ETS scores for the 3-h accumulated precipitation from the two ARW control members (ARW_CN with radar, ARW_C0 without radar), the probability matched means from the 15-member SSEF (SSEF_PM) and the 11-member SREF sub-ensemble (SREF_PM), and the NCEP 12 km NAM. The following conclusions can be drawn from Figure 10:

- Probability matched means from SSEF outperform NAM by a wide margin, also outperform SREF_PM by a wide margin for the two higher thresholds
- The 4 km ARW_CN member also outperforms NAM, and outperforms SREF_PM except at the light rain threshold
- Radar analysis contributes to higher ETS scores (ARW_CN vs. ARW_C0) throughout the entire forecast duration, mostly evident between 0-18 h

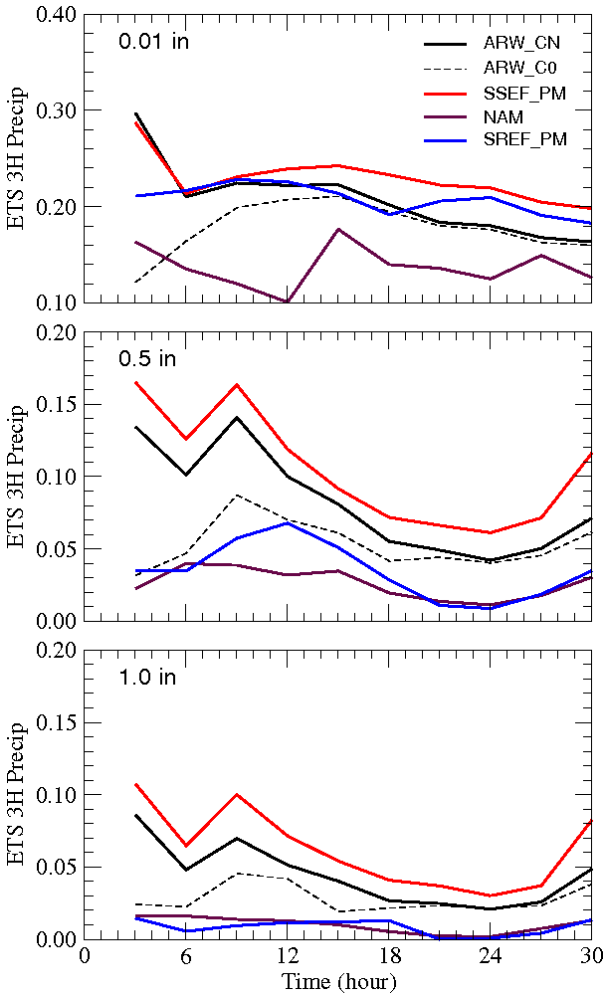

Figure 10. ETS scores of the 3-h accumulated precipitation, averaged over 36 days.

- Combined with Figure 6, it can been seen that SSEF_PM outperforms all individual SSEF (4 km storm-scale ensemble) members for all thresholds

It needs to point out that the SREF forecasts have 3 more hours in lead time than the SSEF and NAM forecasts, due to SREF's staggered initiation times. Nevertheless, for the two higher thresholds (0.5 and 1.0 inch), the SREF_PM, though with coarser resolution, (at 32 km) have comparable performance to the 12 km NAM.

To evaluate the probabilistic QPF, the ROC curves at 24 h and ROC areas of the 3-h accumulated precipitation exceeding 0.5 inch are presented in Figure 11. It can be seen that at this threshold the operational SREF sub-ensemble has ROC areas between 0.53 and 0.62, higher than the skillful (0.5) level but lower than the lower limit of 0.7 that is thought to be a useful probabilistic forecast. The high-resolution SSEF, on the other hand, has a much higher ROC areas, ranging between 0.74 an 0.8, indicating a probabilistic QPF with

greatly improved economical values. Cautions should be taken in this comparison, however, since the larger ensemble number of SSEF than the SREF sub-ensemble (15 vs. 11) may unfairly give SSEF an advantage when computing the ROC curves based on frequency probabilities. Full SREF ensemble will be verified and compared in next season.

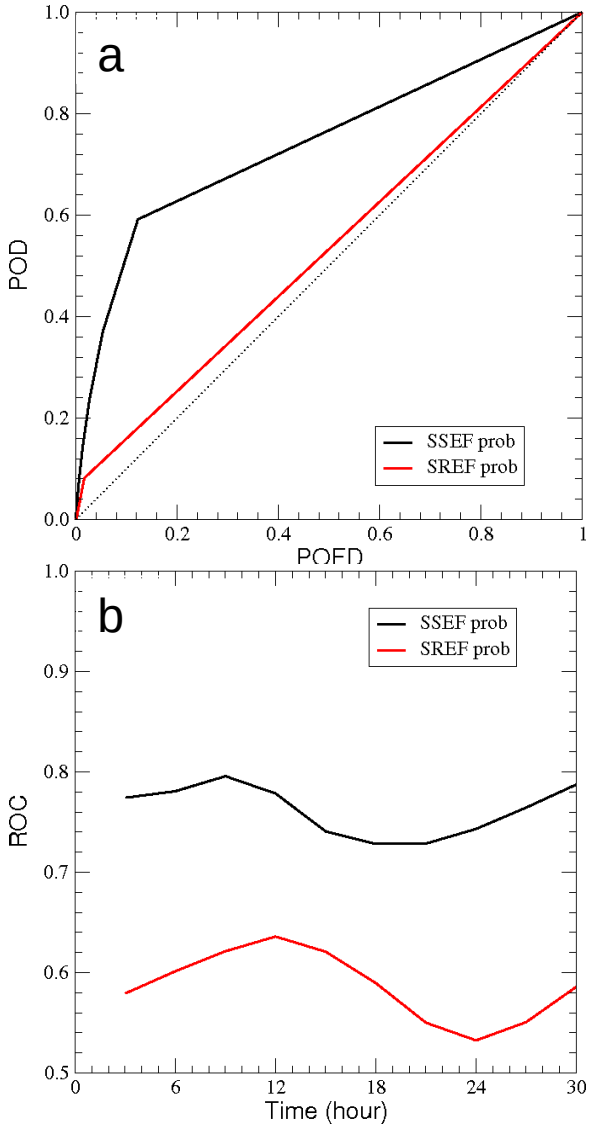

Figure 11. comparison of ROC curves of frequency-based probabilities of 3-h accumulated precipitation exceeding 0.5 inch at 24 h (a) and the ROC areas throughout the forecast duration (b), between SSEF (black lines) and SREF (red lines).

5. ACKNOWLEDGMENTS

This research was mainly supported by a grant to CAPS from the NOAA CSTAR program. Supplementary support was also provided by NSF ATM-0802888, and other NSF grants to CAPS. The realtime forecasts presented here were produced at the National Institute of Computational Science (NICS) at the University of Tennessee. A supercomputer of the Oklahoma Supercomputing Center for Research and Education (OSCER) was used for some of the ensemble post-processing.

6. REFERENCES

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.

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

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, 2010: Probabilistic precipitation forecast skill as a function of ensemble size and spatial scale in a convection-allowing ensemble. Mon. Wea. Rev. conditionally accepted,

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. and Forecasting*, **15**, 61–79.

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., M. Xue, K. Brewster, and K. K. Droegemeier 2004: A three-dimensional varia-tional data assimilation method with recursive filter for single-Doppler radar, J. Atmos. Oceanic. Technol. 21, 457-469.

Hu, M., M. Xue, J. Gao and K. Brewster, 2006: 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.

Kong, F., and co-authors, 2007: Preliminary analysis on the real-time storm-scale ensemble forecasts produced as a part of the NOAA Hazardous Weather Testbed 2007 Spring Experiment. *Preprints, 22th Conf. on Weather Analysis and Forecasting and 18th Conf. on Numerical Weather Prediction* Amer. Meteor. Soc., Park City, UT, 3B.2.

Kong, F., and co-authors, 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.W. Thomas, J. Gao, Y. Wang, K. Brewster, K.K. Droegemeier, J. Kain, S. Weiss, D. Bright, M. Coniglio, and J. Du, 2009: A real-time storm-scale ensemble forecast system: 2009 spring experiment.. 23rd Conf. Wea. Anal. Forecasting/ 19th Conf. Num. Wea. Pred., Omaha, NB, Amer. Meteor. Soc., Paper 16A.3

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

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, San Antonio, TX, 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, Hyannis MA., Amer. Meteor. Soc., P8.1.

Xue, M., and co-authors , 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. Prediction*., Amer. Meteor. Soc., Park City, UT, Paper 3B.1.

Xue, M., and co-authors, 2008: CAPS Realtime Storm-scale Ensemble and High-resolution Forecasts as Part of the NOAA Hazardous Weather Testbed 2008 Spring Experiment. *Preprints*, *24th Conf. on Severe Local Storm*, Amer. Meteor. Soc., Savannah, GA, 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 Experiement. 23rd Conf. Wea. Anal. Forecasting/19th Conf. Num. Wea. Pred., Omaha, NB, Amer. Meteor. Soc., Paper 16A.

Xue, M., F. Kong, K.W. Thomas, Y. Wang, K. Brewster, J. Gao, X. Wang, S. Weiss, A. Clark, J. Kain, M. Coniglio, J. Du, T. Jensen, H-Y. Kuo, 2010: CAPS realtime storm-scale ensemble and convection-resolving forecasts for the NOAA Hazardous Weather Testbed 2010 Spring Experiment. 25th Conf. on Severe Local Storms, Denver, CO, Amer. Meteor. Soc., Paper 7B3.