Preprints, 24 th Conf. on Severe Local Storm, Amer. Metor. Soc., 27-31 October 2008, Savannah,

12.3 REAL-TIME STORM-SCALE ENSEMBLE FORECAST EXPERIMENT – ANALYSIS OF 2008 SPRING EXPERIMENT DATA

Fanyou Kong[1]*, Ming Xue[1,2], Kevin W. Thomas[1], Kelvin K. Droegemeier[1,2], Yunheng Wang[1], Keith Brewster[1], Jidong Gao[1], Jack Kain[4], Steven J. Weiss[3], David Bright[3], 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

Funded partly by NOAA CSTAR program and in its second project year, a real-time storm-scale ensemble forecasting experiment has been conducted as part of the NOAA Hazardous Weather Testbed (HWT) 2008 Spring Experiments (Xue et al. 2008; Kong et al. 2008). At 4-km horizontal grid spacing, the WRF-ARW-based ensemble system, developed at the Center for Analysis and Prediction of Storms (CAPS), the University of Oklahoma, runs daily for 30 hours from 14 April through 6 June, for a domain covering most 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.

Ensemble forecast products are created in real time through existing capabilities in the SPC version of the N-AWIPS system for evaluation by researchers and operational forecasters during the experiment. The performance of the ensemble forecasts, in terms of quantitative skill scores, is evaluated to assess the effectiveness of the EFs at storm-scale. This extended abstract presents some post-season analysis results. Several comparisons of 2008 and 2007 ensemble data are also presented. More focus is given to evaluation of post-processing techniques that suit for deterministic QPF and probabilistic QPF (PQPF) derived from the storm-scale ensemble forecasts.

2. EXPERIMENT HIGHLIGHT

As the second year of the three-year project, the 2008 Spring Program began on 14 April 2008 and ended on 6 June. All experimental forecasts were generated with the Weather Research and Forecast (WRF) Advanced Research WRF (ARW) model (V2.2), as in 2007 experiment (Kong et al. 2007; Xue et al. 2007). Several major changes from the 2007 experiment were made for the 2008 season: (1) The model domain was enlarged (Figure 1); (2) Daily 30 h forecasts were initiated at 0000 UTC, using NAM 12 km (218 grid) 00Z analyses as background for initialization with the initial condition perturbations for the ensemble coming from the NCEP Short-Range Ensemble (SREF); (3) available WSR-88D data were assimilated through ARPS 3DVAR and cloud analysis package into all but one members; (4) Eight members were constructed as hybrid with both initial perturbations and physics variations. In 2007, only four members were hybrid, and no radar data was assimilated.

The initial perturbations were extracted from the 3 h forecasts of eight 21Z SREF members and are scaled to their initial perturbation amplitudes. All forecast output at hourly intervals were 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.

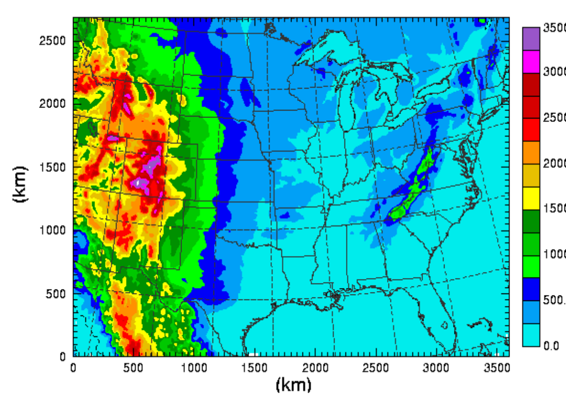

Figure 1. Model domain coverage for 2008 season, with terrain height in color shading.

The daily 30 h ensemble forecasts, for the weekdays from Monday through Friday, started at 0000 UTC and ended at 0600 UTC of the next day. Special weekend runs were arranged if it was requested by SPC based on the severe weather outlook. The ensemble configuration includes ten hybrid members, all of which

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

were run on *BIGBEN*, a NSF TeraGrid resource (Cray XT3) at PSC. Model execution began around 0230 UTC (21:30 local time) and finished in about 6-10 hours (depending on members and convection activities), using about 1000 CPUs, with results being processed as they become available.

Table 1. Ensemble member configuration

member	IC	LBC	Radar data	mp_phy	sw-phy	pbl_phy
cn	00Z ARPSa	00Z NAMf	yes	Thompson	Goddard	MYJ
c0	00Z NAMa	00Z NAMf	no	Thompson	Goddard	MYJ
n1	cn – em_pert	21Z SREF em_n1	yes	Ferrier	Goddard	YSU
p1	cn + em_pert	21Z SREF em_p1	yes	WSM 6-class	Dudhia	MYJ
n2	cn – nmm_pert	21Z SREF nmm_n1	yes	Thompson	Goddard	MYJ
p2	cn + nmm_pert	21Z SREF nmm_p1	yes	WSM 6-class	Dudhia	YSU
n3	cn – etaKF_pert	21Z SREF etaKF_n1	yes	Thompson	Dudhia	YSU
p3	cn + etaKF_pert	21Z SREF etaKF_p1	yes	Ferrier	Dudhia	MYJ
n4	cn – etaBMJ_pert	21Z SREF etaBMJ_n1	yes	WSM 6-class	Goddard	MYJ
p4	cn + etaBMJ_pert	21Z SREF etaBMJ_p1	yes	Thompson	Goddard	YSU

Table 1 outlines the basic configuration for each individual members. *cn* refers to the control member, with radar data analysis, *c0* is the same as *cn* except no radar data. *n1-n4* and *p1-p4* are members with initial perturbation added on top of *cn* initial condition, NAMa and NAMf refer to 12 km NAM analysis and forecast, respectively. ARPSa refers to ARPS 3DVAR analysis using NAMa as background. For the eight perturbed members, the ensemble initial conditions consist of a mixture of bred perturbations coming from the 21Z SREF perturbed members (one pair each from WRF-em, WRF-nmm, eta-KF, and eta-BMJ) and physics variations (grid-scale microphysics, shortwave radiation, land-surface and PBL physics). 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-perturbed members (*cn* and *c0*). For all members, the long-wave radiation schemes are RRTM, the surface physics uses Noah scheme, and no cumulus scheme is used (see WRF manual for detail description for all physics schemes).

In addition to the SPC's N-AWIPS system, CAPS also makes available a webpage demonstrating the EF products (http://www.caps.ou.edu/wx/spc).

3. VERIFICATION OF THE ENSEMBLE SYSTEM

During the course of the experiment, there are a total of 36 days with complete forecasts for all ten members. Post analyses and verifications are carried out over these complete dates to assess the statistical feature and the performance of the ensemble system, unless specified in the text and figures.

Figure 2 shows the domain-mean ensemble spread (defined as standard deviation against ensemble mean) of some fields, averaged over 36 complete forecast dates (covering most of the experiment period from April 16 through June 6).. It can be seen that the hybrid perturbation configuration of the ensemble system exhibits reasonable dispersion for the mass-related fields such as sea level pressure and 500 hPa geopotential height. For QPF related variables such as hourly accumulated precipitation and reflectivity fields diurnal pattern is evident, reflecting the quiet morning hours (low around 10am) and the active late afternoon hours (high around 7pm) for the summer convective storm activity (Figure 2c).

The real-time storm-resolving (or storm-permitting, convection-permitting in some literatures) ensemble forecasting system offers a unique capability of producing quantitative precipitation forecast (QPF), both deterministic and probabilistic, at very high temporal and spatial resolution. For post-season verification purpose, 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

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

domain and used as verification dataset to verify the predicted QPF quantities (1 h accumulated precipitation and composite reflectivity).

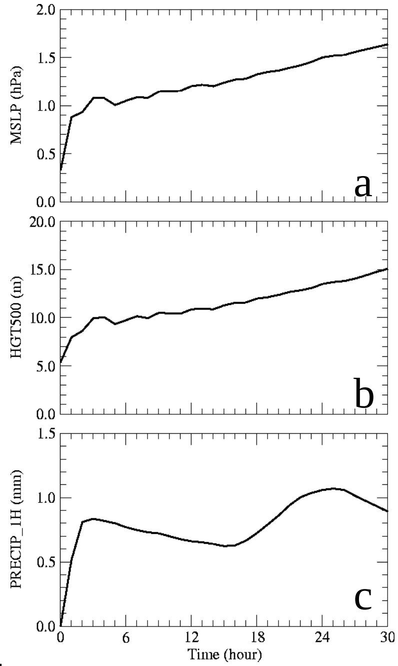

Figure 2. Domain-mean ensemble spread (standard deviation), averaged over 36 forecast dates. (a) mean sea level pressure, (b)) 500 hPa geopotential height, and (c) 1 h accumulated precipitation.

3.1 BIAS score

In 2007 experiment season, systematically high QPF biases (Figure 3b) were observed compared to several other storm-scale deterministic forecasts produced during the same period by NCAR and NSSL (Kain, personal communication). Prior to the 2008 season, several reruns of the 2007 dates were carried out to identify the cause of such high biases. Factors considered include model start time, data for lateral boundary condition, number of vertical levels, NAM data type used (12 km 218 grid vs 40 km 212 grid). It turns out the major culprit is the use of 21Z NAM analysis as the initial condition in 2007 season, combined with the use of 18Z NAM forecast as LBCs (Kong et al. 2007). Figure 4 shows an example of BIAS score comparison from the reruns of May 6, 2007. Dark black line refers to the condition of 2007 season - 21Z NAM analysis (NDAS) for IC and 18Z NAM forecast for LBC – with the highest BIAS score over the convection active hours. The forecast with 00Z NDAS and 00Z LBC (solid red line) has the lowest BIAS score. 18Z LBC also contributes to elevated BIAS (dash red line). The change to initiate ensemble forecasts at 0000 UTC using 00Z NAM analysis and forecast as IC background and LBCs for the 2008 season helps significantly bring down the QPF biases (Figure 3a).

Figure 3 shows the BIAS scores of 1 h accumulated precipitation exceeding 0.1 in (2.54 mm) for individual members and ensemble mean from both 2007 and 2008 seasons. Other thresholds exhibit similar pattern. In spite of the large reduction of BIAS score in 2008 season, some members (including *cn* and *c0*) still have maximum BIAS values of close to 2 during afternoon hours, pointing to over-prediction of precipitation by some members.

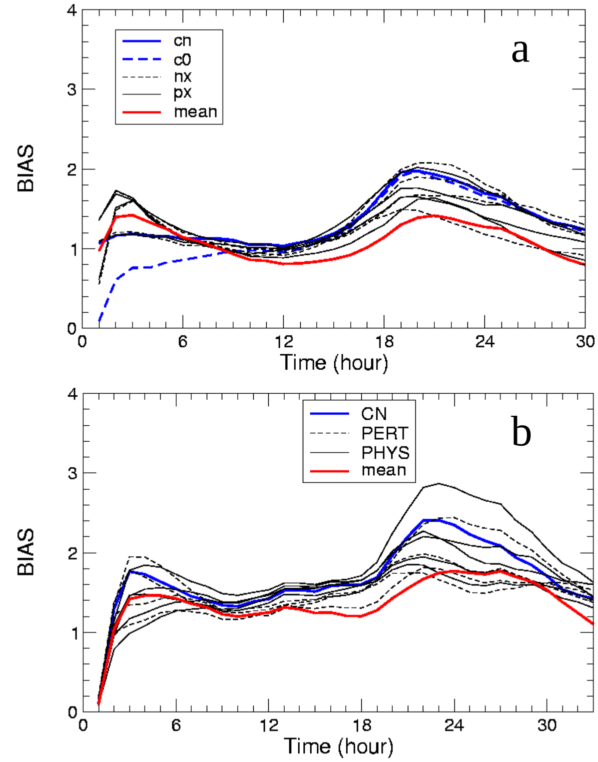

Figure 3. BIAS scores of 1 h accumulated precipitation ≥0.1in averaged over all complete forecast dates from (a) 2008 and (b) 2007seasons. For 2007 season, CN refers to control member, PERT refer to four initial perturbation plus physics perturbation, and PHYS refer to five physics perturbation only members (Kong et al. 2007).

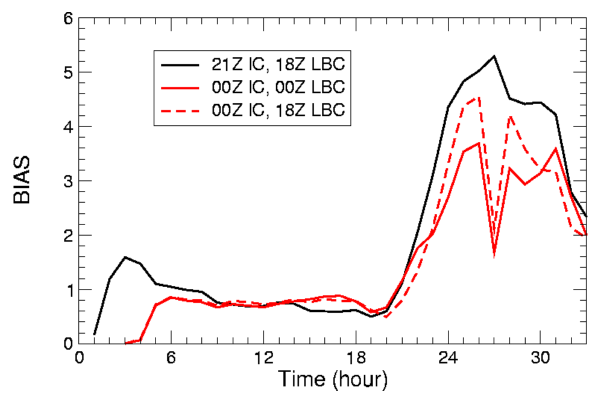

Figure 4. BIAS scores of 1 h accumulated precipitation ≥0.1in for the data of May 6, 2007. IC and LBC refer to times of NAM analysis for initial condition and NAM forecast for lateral boundary condition, respectively.

3.2 ETS scores

The ETS scores are calculated for the criteria of 1 h accumulated precipitation ≥ 0.1 in for all members and the ensemble mean (Figure 5). Compared to 2007 season, the unique feature of 2008 experiment is the addition of radar data for nine out of ten members. Figure 5 indicates that the inclusion of radar data helps boost the ETS scores for the initial hours and the influence lasts for 12 h. c0 member in 2008 and all members in 2007 underscore the nine 2008 members with radar data assimilated for the initial 6 h. The radar data influence fade away after 12 h.

The ensemble means in Figure 5 show clear outscore to all individual members. However, it should be cautious to interpret the phenomena, as described in next subsection and in Figure 6 indicating less usefulness of ensemble means in high temporal (e.g., 1 h accumulation) precipitation forecast.

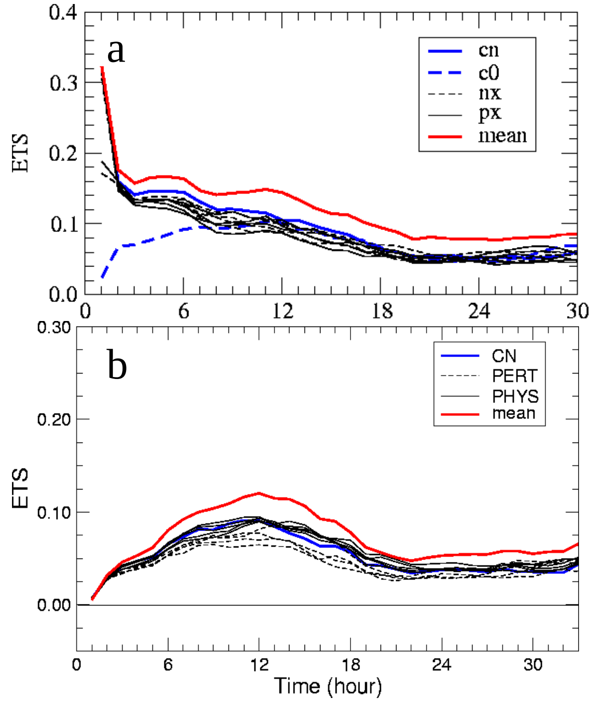

Figure 5. ETS of 1 h accumulated precipitation ≥0.1in averaged over all complete forecast dates from (a) 2008 and (b) 2007seasons.

3.3 Deterministic QPF from ensemble

For many desired meteorological variables such as sea level pressure, 2 m temperature, 10 m wind etc. ensemble mean is a good deterministic quantity derived from ensemble forecasting products that can outscore forecasts using single model runs. However, owing to very high spatial and temporal variance associated with precipitation, especially when produced with very high-resolution model runs like in this case, ensemble mean of precipitation field tends to be excessively broad in area coverage and too weak in magnitude, and thus is not a useful QPF product (Ebert 2001; Kong et al. 2006; Clark et al. 2008). Figure 6a gives an example of ensemble mean of 1 h accumulated precipitation in comparison with observation (Figure 6d).

Probability Matching (PM, hereafter) technique was demonstrated by Ebert (2001) and Clark et al. (2008) to be more useful by assuming that the best spatial representation of rainfall is given by the ensemble mean and that the best frequency distribution of rainfall is given by the ensemble member QPFs. PM products for QPF are produced by first pooling QPF amounts of all ensemble members and over all grid points for a given forecast lead time and sorted from the highest to the lowest to obtain a QPF distribution. The ensemble mean QPF amounts are also sorted from the highest to the lowest. Then the QPF values from the ensemble mean are reassigned using values from the corresponding ranks of the QPF distribution. Given N ensemble members and M total grid numbers of model domain, there are MN elements in QPF distribution versus N in ensemble mean. Ebert (2001) picked every N sorted element from the QPF distribution pool. We denote it as PM Method 1. The problem of this method is that QPF values decrease drastically from higher ranks to lower ranks at the high end, especially when very high-resolution storm-permitting models are involved. That can cause artificially high peak QPF values if the first element of each N segment is picked. Alternately, random picking among each N element can be used. We exercised a new approach by averaging the N elements and assigning the mean to the corresponding rank of ensemble mean (denoted PM Method 2).

Figure 6 shows an example 1 h accumulated precipitation using both methods, with simple ensemble mean and observation side by side for comparison. Both methods demonstrate better spatial coverage and amplitudes than the ensemble mean, with PM Method 2 (111 mm) more close to observation (65 mm) than PM Method 2 (132 mm) in maximum values. The ensemble mean (Figure 6a) has a maximum value of 28 mm.

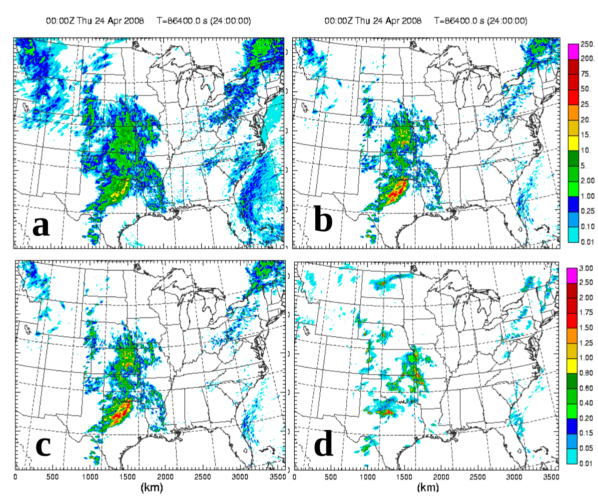

Figure 6. 1-h accumulated precipitation from 24 h ensemble forecast initiated at 0000 UTC April 23, 2008 for (a) ensemble mean, (b) PM in method 1, (c) PM in method 2, and from (d) NSSL observation, valid at 0000 UTC April 24, 2008.

Figure 7 gives a more clear picture on how model predicted precipitation maximum values and area coverage (represented by grid count) in the forms of ensemble mean and PM (both Methods 1 and 2) are compared to observations. In this single day example (April 23, 2008), ensemble means significantly under-estimate precipitation intensity and over-estimate precipitation area. Both PM methods improve the intensity (with Method 2 – PM2 in the figure – more close to observation) and lower precipitation area (in Figure 7b, PM1 and PM2 are identical). Figure 8 shows the same curves (without PM1) as Figure 7 but averaged over all available 36 days, confirmed the general improvement of using PM against simple ensemble mean. The improvement is especially significant for area coverage.

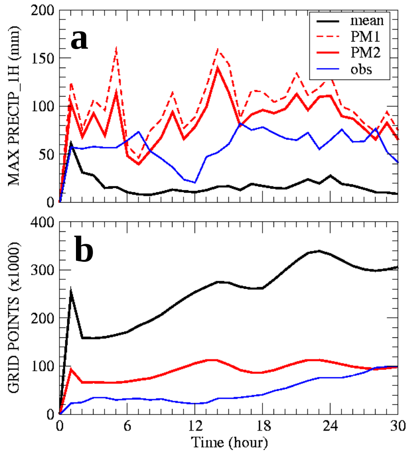

Figure 7. Maximum 1 h accumulated precipitation (a) and grid counts (area coverage) (b) for the 30 h period starting at 0000 UTC April 23, 2008. PM1 and PM2 refer to PM method 1 and 2, respectively.

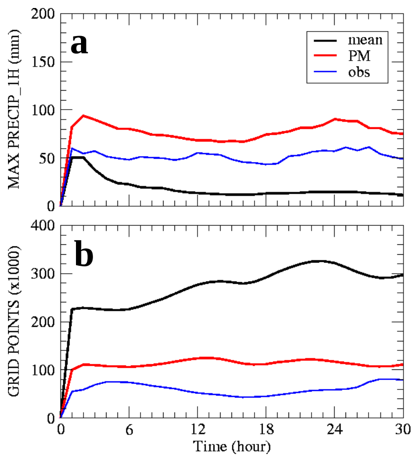

Figure 8. Maximum 1 h accumulated precipitation (a) and grid counts (area coverage) (b), averaged over 36 days. PM refers to PM method 2.

In next several figures, traditional verification scores are compared among PM (method 2), ensemble mean, and individual members. BIAS scores of PM are generally in the middle of individual members and perform much better than ensemble means (Figure 9), with exception of 0.1in threshold. ETS scores of PM are close to ensemble mean for light rain (Figure 10a), with small outscore after 15 h, and clearly outscore ensemble mean for 0.5in rain threshold (Figure 10b), both outscore individual members. Such improvements are not reflected in RMSE scores with PM curve just in between individual members (figure not shown).

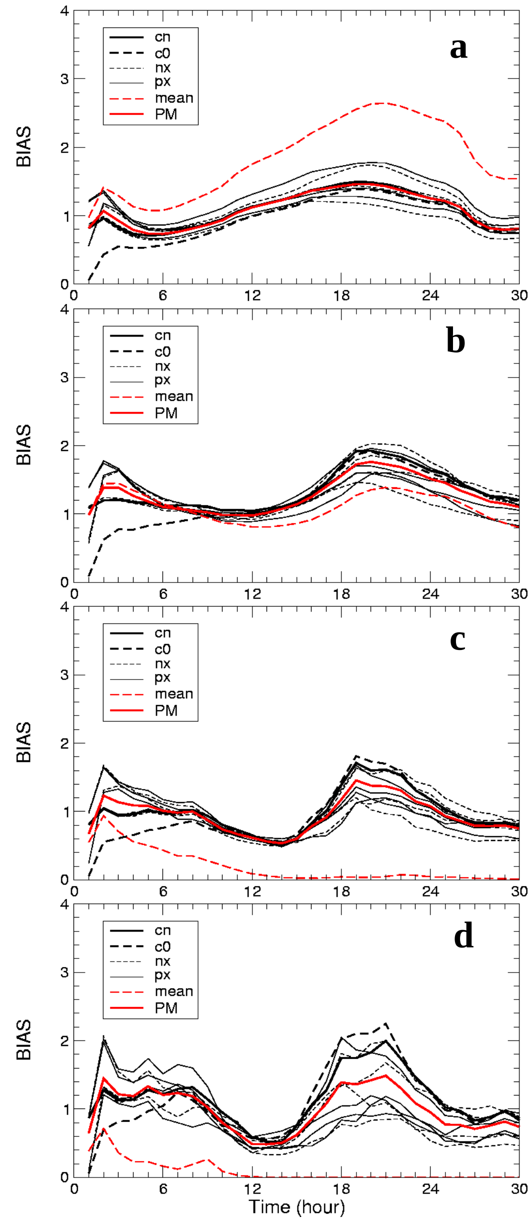

Figure 9. Averaged BIAS scores of 1 h precipitation ≥ 0.01in (a), 0.1in (b), 0.5in (c), and 1.0in (d).

3.3 Probabilistic QPF from ensemble

The ability to generate probabilistic QPF (PQPF) from ensemble members is one of selling point for ensemble forecasting. However, obtaining skillful and reliable PQPF from high-resolution and storm-permitting ensemble forecasts remains a huge challenging task, especially for high temporal PQPF such as 1 h accumulated precipitation as compared to 12 h, 24 h or even longer period of accumulation.

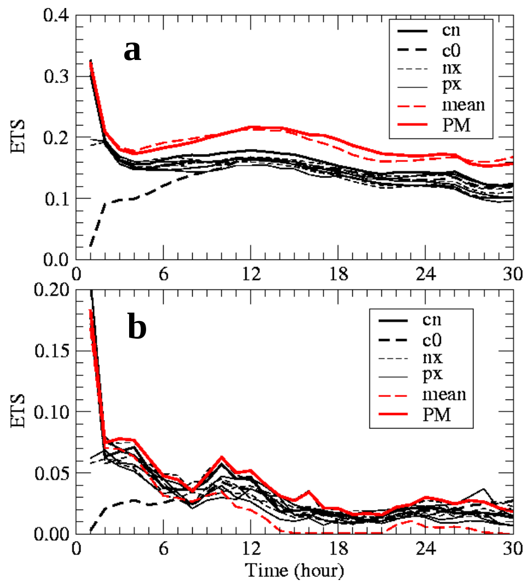

Figure 10. ETS scores of 1 h precipitation ≥ 0.01in (a and 0.5in (b).

During the course of Spring Experiment, various probabilistic QPF forecast variables were derived by using simple relative frequency among ensemble members for specified thresholds. Figure 11 shows an example probability map produced in quasi real-time of the 24 h forecast of 1 h accumulated precipitation exceeding 0.1 in, valid at 0000 UTC April 24, 2008 (Figure 6d is the corresponding observation).

Post-season verification of these simple PQPF are conducted by means of verification rank histogram (Hamill and Colucci 1997) and reliability diagrams. Figure 12 presents 24 h and 30 h verification rank histogram charts for the 1 h accumulated precipitation, averaged over the complete 36 day dataset. Small right-tilting and U-shape can be seen, indicating some degree of overprediction of precipitation and underdispersion. But in general, the verification rank histograms are quite flat.

Figure 13 shows reliability diagrams, with both 2007 and 2008 data presented, of two forecast lead-time and thresholds of 1 h accumulated precipitation. Even though 2008 ensemble data show improvement over 2007, both years feature reliability lines quite away from (below) the diagonal line (perfect reliability), indicating much less ideal PQPF for 1 h accumulated precipitation. Verification of longer period accumulated precipitation PQPFs may show different performance and are the next set of analysis tasks. Bias-correction and various calibration procedures are also called into attention for post-processing of ensemble forecasts to improve skills and reliability/resolution of PQPFs.

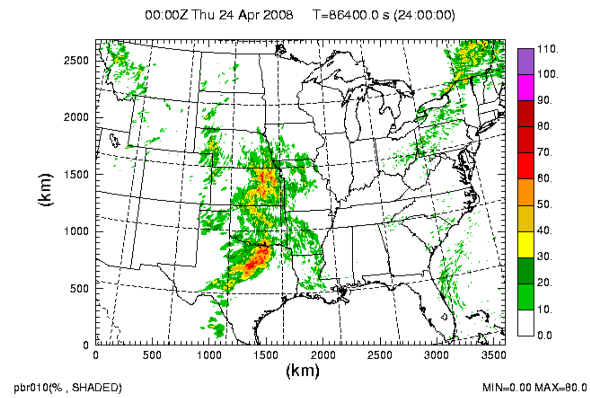

Figure 11. Probability of 1 h accumulated precipitation ≥ 0.1in for the 24 h forecast valid at 0000 UTC April 24, 2008.

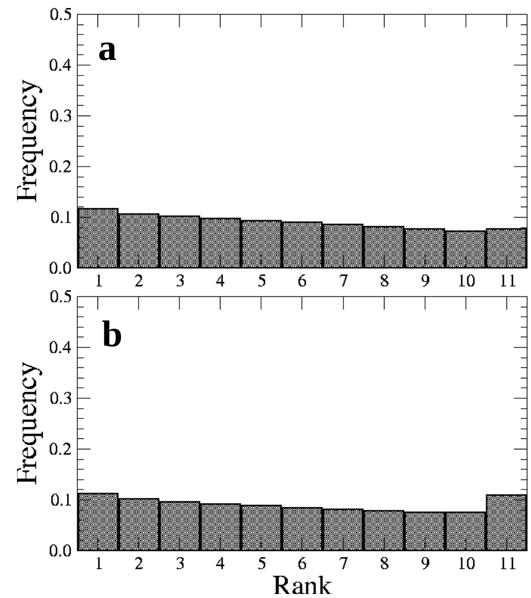

Figure 12. Verification rank histogram of 1 h accumulated precipitation for the forecast hours of (a) 24 h and (b) 30 h.

3.3 PQPF and bias correction

Many researches demonstrated the merit and success in post-processing ensemble forecasting data through bias-correction and various calibration techniques to produce PQPF with improved skills and reliabilities for 24 h accumulated period and longer and over very coarse grids (e.g., Hamill and Colucci 1997; Eckel and Walters 1998). Post-processing high-resolution PQPF for periods shorter than 24 h have not been examined until recent (Stensrud and Yussouf 2007; Yussouf and Stensrud 2008). In the two journal papers last cited, Stensrud and Yussouf (2007, 2008) employed a simple binning technique to remove bias of 3 h accumulated precipitation from a multimodel short-range ensemble forecasting system, using past 12 days as a training period, and produced significantly more skillful and reliable PQPFs than the raw forecast probabilities. However, the inclusion of zero observation in populating QPF bins leads to somewhat noisy precipitation field after adjustment, which in turn leads to noisy probability maps and irregular (noisy?) verification rank histogram charts.

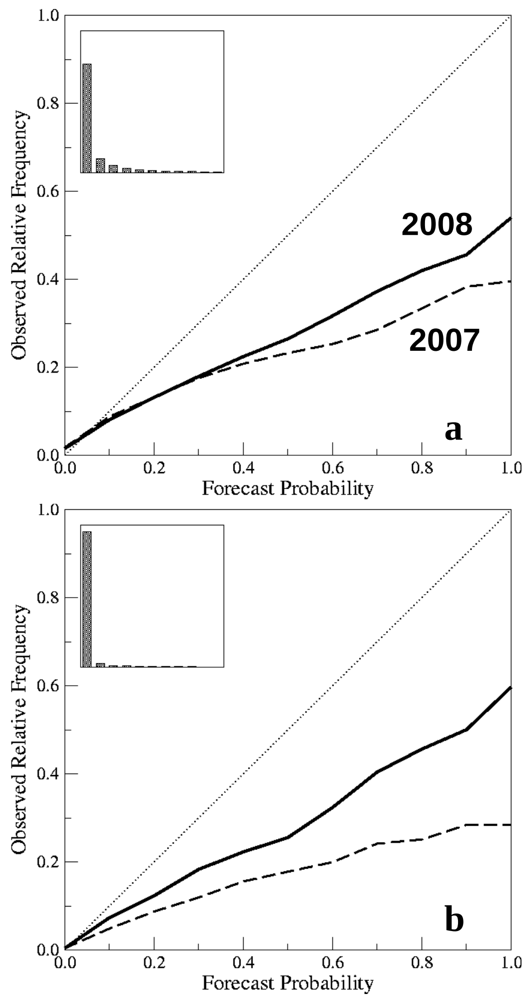

Figure 13. Reliability diagrams for the 1 h accumulated precipitation ≥ 0.01in at 24 h (a) and 0.1in at 12 h. Solid lines are from 2008 season data, and thick dash lines are from 2007 season data.

Recently, Clark et al. (2008) demonstrated that applying a bias correction approach similar to probability matching (PM) (Ebert 2001), except using observation as QPF distribution, to all individual members can result in more flat rank histograms. This approach, however, is not a bias correction technique to produce PQPFs because current observation is required.

Another challenge of QPF bias correction lies in the use of high-resolution storm-permitting model forecasts which produce very high variance precipitation fields with excessively high maxima (see Figure 8). Traditional bias correction concept of removing a mean bias based on a short training period prior to forecast time doesn't address such excessive peak value issue when applying to QPF, since mean bias for QPF, when averaged over domain, is often a very small amount.

Having the correction of excessive precipitation in mind, a new approach is exercised by correcting precipitation biases based on ranks. For each member at each forecast hour, the 1 h accumulated precipitation field is sorted from the highest to the lowest, and the same ranks are averaged over a training period (12 days in this exercise). The observations over the same 12 days period are also sorted and averaged. Figure 14a is a example ranking results for the 24 h forecast, averaged over a 12 day period from April 16 through May 7, 2008.

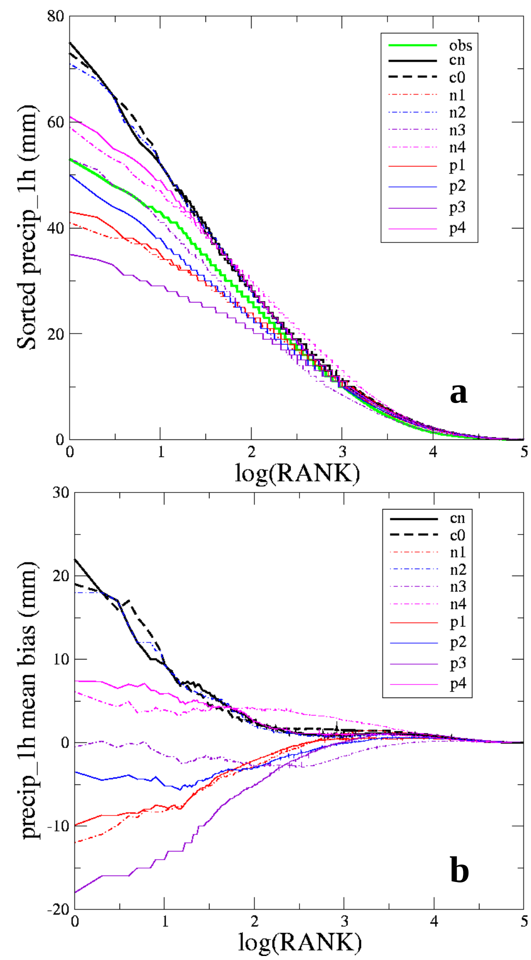

Figure 14. (a) Sorted 1 h accumulated precipitation, and (b) differences between members and observation (bias) for the 24 h forecast, averaged over a 12-day period from April 16 to May 7, 2008.

In Figure 14, be aware of the use of logarithm coordinate for rank (from 1 to M, M is total horizontal grid points), and only a first third of ranks are shown. It can be seen that individual members vary widely around observation for the highest several hundred ranks, with peak biases (difference between members and observations for the first rank – Figure 14b) swing 20 mm on either side. This specific training period shows high positive biases for cn, co, and n2, and high negative bias for p4. The mean biases in Figure 14b are used to correct each member forecast of the following date by applying the mean biases to corresponding ranks of sorted precipitation field of each member at each forecast hour. A total 15 days of bias-corrections are made, from May 8 through May 26, 2008. Ensemble means and PMs are generated from the corrected dataset.

BIAS scores are presented in Figure 15, using the corrected 15 days dataset. Compared with Figure 9, significant improvement can be seen, especially for the two lower thresholds, across all forecast hours.

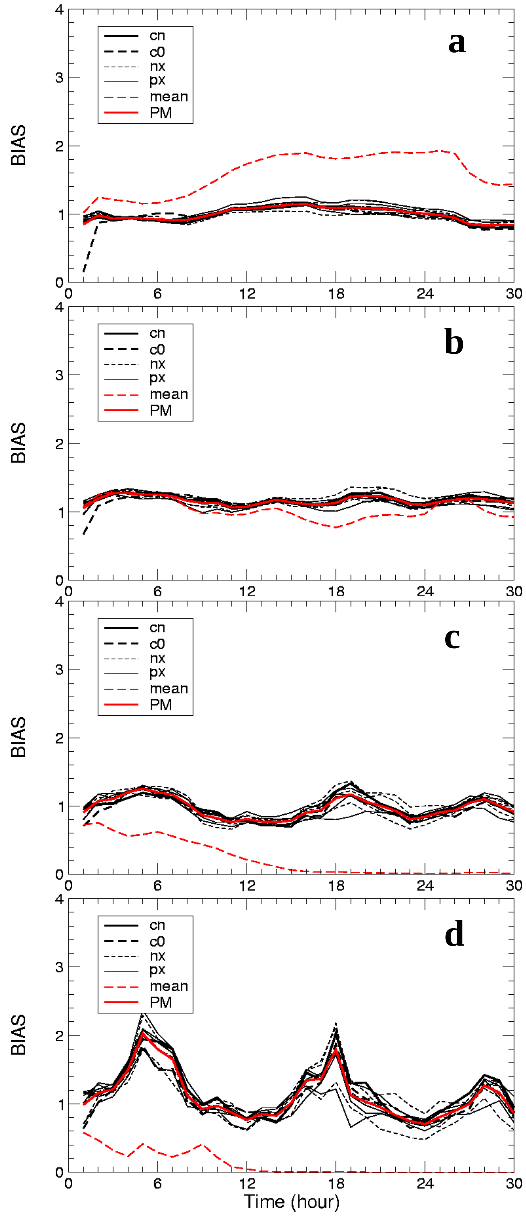

Figure 15. Same as Figure 9, except bias corrected and averaged over 15 days.

Figure 16 shows rank histogram charts of two forecast hours, 18 h and 24 h, both with more flat distributions compared to raw ensemble. The difference is very small for the 6 h and 30 h forecast times (figure not shown).

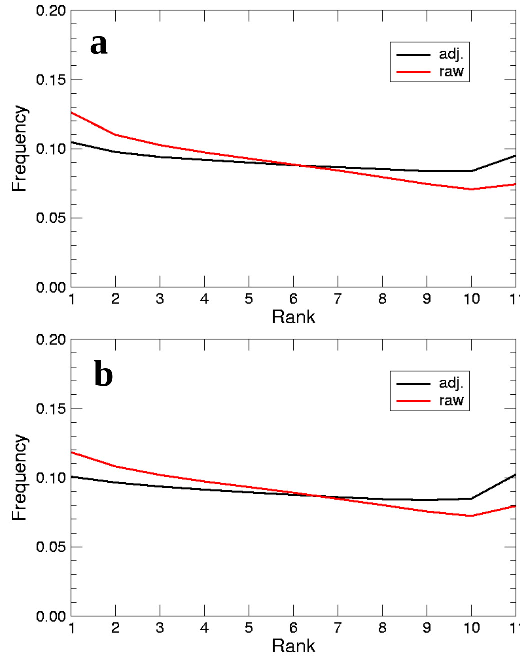

Figure 16. Verification rank histogram of 1 h accumulated precipitation for the forecast hours of (a) 18 h, and (b) 24 h, averaged over 15 days of bias corrected dataset.

In spite of big improvement in BIAS score and some improvement in rank histogram, no improvement is shown in reliability diagrams (Figure not shown).

4. DISCUSSION

The extremely high variance nature of 1 h accumulated precipitation from high-resolution storm-permitting WRF ensemble forecast may intrinsically discourage post-calibration effort to be able to produce reliable and high-skill PQPFs measured by conventional verification metrics. New verification approaches such as object oriented and neighborhood methods (Schwartz et al. 2008) may be a solution.

As a future focus, we will further examine various post-processing techniques including refining the bias correction approach exercised in this paper, and apply to 3 h accumulation or longer period. Comparisons of QPF/PQPFs, in terms of quantitative skill scores, from the high-resolution ensemble forecast in the Spring Experiment with the NCEP operational NAM 12 km and SREF forecasts over the same domain and period will be another valuable effort.

5. ACKNOWLEDGMENTS

This work 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), and by NSSL. 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

Clark, A. J., W. A. Gallus, M. Xue, and F. Kong, 2008: A comparison of precipitation forecast skill between small near-convection-permitting and large convection-parameterizing ensembles. Submitted to *Weather and Forecasting*.

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.

Eckel, F. A., and M. K. Walters, 1998: Calibrated probabilistic quantitative precipitation forecasts based on the MRF ensemble. *Weather and Forecasting*, **13**, 1132-1147.

Hamill, T. M., and S. J. Colucci, 1997: Verification of Eta-RSM short-range ensemble forecast. *Mon. Wea. Rev*., **125**, 1312-1327.

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., 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. *9th WRF User's Workshop*, NCAR Center Green Campus, 23-27 June 2008, Paper 7.3.

Schwartz, C. S., and coauthors, 2008: Toward improved convection-allowing ensembles: model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Preprints, 24th Conf. on Severe Local Storm*, Amer. Meteor. Soc., Savannah, GA, Paper 13A.6.

Stensrud, D. J., and N. Yussouf, 2007: Reliable probabilistic quantitative precipitation forecasts from a short-range ensemble forecasting system. *Weather and Forecasting*, **22**, 3-17.

Yussouf, N., and D. J. Stensrud, 2008: Reliable probabilistic quantitative precipitation forecasts from a short-range ensemble forecasting system during the 2005/06 cool season. *Mon. Wea. Rev*., **136**, 2157-2172.

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.