SHORT-RANGE ENSEMBLE FORECASTING (SREF) AT NCEP/EMC

M. Steven Tracton* and Jun Du

National Centers for Environmental Prediction (NCEP), NWS/NOAA, Washington DC

1. INTRODUCTION

Over the past few years ensemble prediction has come to the fore as a major element in defining the future of numerical weather prediction (NWP) and operational weather forecasting. This stems basically from convergence of increasing recognition of the importance of explicitly addressing the intrinsic uncertainties in forecasts with existing/prospective super-computer resources and development of ensemble strategies. The net result of this convergence is an expanding capability to provide quantitative estimates of those uncertainties. It is widely agreed that ensemble based probabilities and/or measures of confidence hold the best potential for enhancing the ability to make user dependent informed decisions. Indeed, the National Weather Service (NWS) is requiring that many forecast products evolve to become probabilistic in nature, especially for quantitative precipitation forecasting (QPF).

NCEP now runs operationally 17-member ensembles each day with the MRF model for medium-range (3-14 days) predictions (Tracton and Kalnay 1993). Results are very encouraging (Toth and Kalnay 1997). This paper focuses upon development of a system and strategies for short-range (0-3 days), regional model based ensembles. The regional models are NCEP's Eta (Black 1994) and Regional Spectral Model (RSM) (Juang and Kanamitsu 1994). In the following we emphasize preliminary assessments of experimental runs with regard to prospective operational applications in context of the NWS "End-to-End" (ETE) forecast process, where user needs and requirements are (should be) the key driver in model development and model strategies. Explicit or implicit in this discussion are reference to the basic problems, issues, etc. associated with SREF. The ultimate near term objective at NCEP/EMC is to develop a prototype short-range ensemble system, strategy, and product suites for operational implementation upon acquisition of the next generation (Class VIII) super computer (late '98/early '99).

2. SREF SYSTEM

The current experimental SREF system consists of 25 members composed of 5 five-member subsets (Fig. 1). Three of these subsets, which have been running since 1995 (Brooks et. al. 1995, Hamill and Colucci 1997, Stensrud et. al. 1997), consist of Eta model runs at 80-km horizontal resolution with global "bred" perturbations (Toth and Kalnay 1997) and with "random" perturbations associated with different "in house" NCEP analyses, and 80-km RSM runs from the global "bred" modes. The two new subsets are with "regional bred" perturbations used to initialize both the (80 km) RSM and Eta. The "regional" breeding *enhances* the

* Corresponding author address: Dr. M. Steven Tracton, NCEP/EMC, 5200 Auth Road, Camp Springs, MD 20746; e-mail: <Steve.Tracton@noaa.gov>.

global bred modes consistent with the resolution of the respective regional model. Runs with 40- and 48-km versions of the RSM and Eta, respectively, are just beginning. Overall, the experimental SREF enables consideration of intermodel comparisons of ensembles using alternative perturbation methodologies, including the effects of resolution and, very importantly, the advantages of multi-model ensembles. These and other questions are subject to verification/validation of the skill and utility of ensembles with regard to such questions as: a) ensemble mean versus control, b) efficacy of spread (divergence amongst ensemble members), c) reliability of probability estimates, d) relative information content of ensembles versus single higher-resolution runs, and e) the utility of output products (e.g, clustering, "spaghetti" charts).

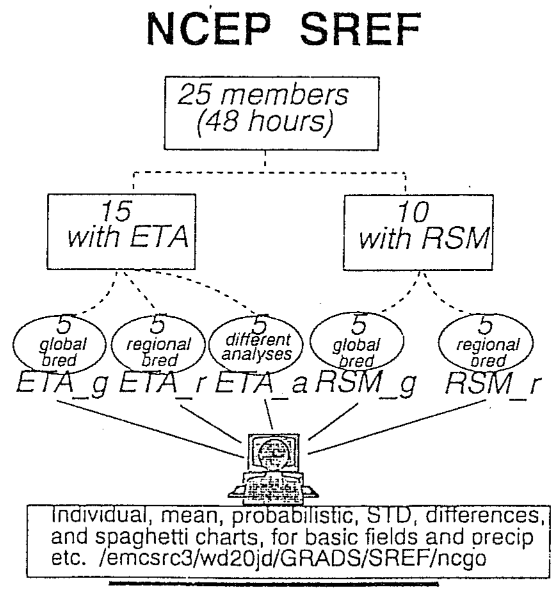

Fig. 1: The current NCEP SREF system.

3. KEY POINTS

The key points to be made in this paper, as demonstrated by preliminary evaluation of case studies and relevant verification statistics, are: a) there are inevitable uncertainties in regional model, short-range (1-2 days) NWP owing to the sensitivity of predictions to intrinsic uncertainties in initial conditions and model formulation, b) SREF is essential for estimating these uncertainties and provide a more complete picture than the control and even single higher-resolution model runs, c) more complete information yields demonstrable enhanced operational utility, d) increased and enhanced diversity of solutions is obtained with multi-model ensembles (**combined** influence of model differences and analysis uncertainties), e) increased and enhanced utility of ensembles is achieved using higher resolution (40 km) as compared to lower resolution (80 km) models, and f) dynamically constrained perturbations ("bred modes") provide much better spread than "randomly" perturbed initial conditions.

4. SOME EXAMPLES

Figure 2 illustrates the advantage of SREF in the QPF arena. In particular note that in this east coastal "bomb" case, the 40-km RSM control (as true for the operational Eta) gives no hint of the observed heavy rainfall maximum over Delaware and New Jersey (cf. Figs. 2a and 2b), while there is a strong signal of same from the 40-km RSM ensemble – in both the ensemble mean forecast (not shown) and the ensemble-based probabilistic forecast (Fig. 2c) (Eta ensemble unavailable). There is no indication of this heavy rainfall center from the 80-km RSM ensemble (not shown). Generally, the ensemble mean precipitation provides higher Threat Score skill (Fig. 3) and less Bias (not shown) than the corresponding operational 48-km Eta and 29-km meso-

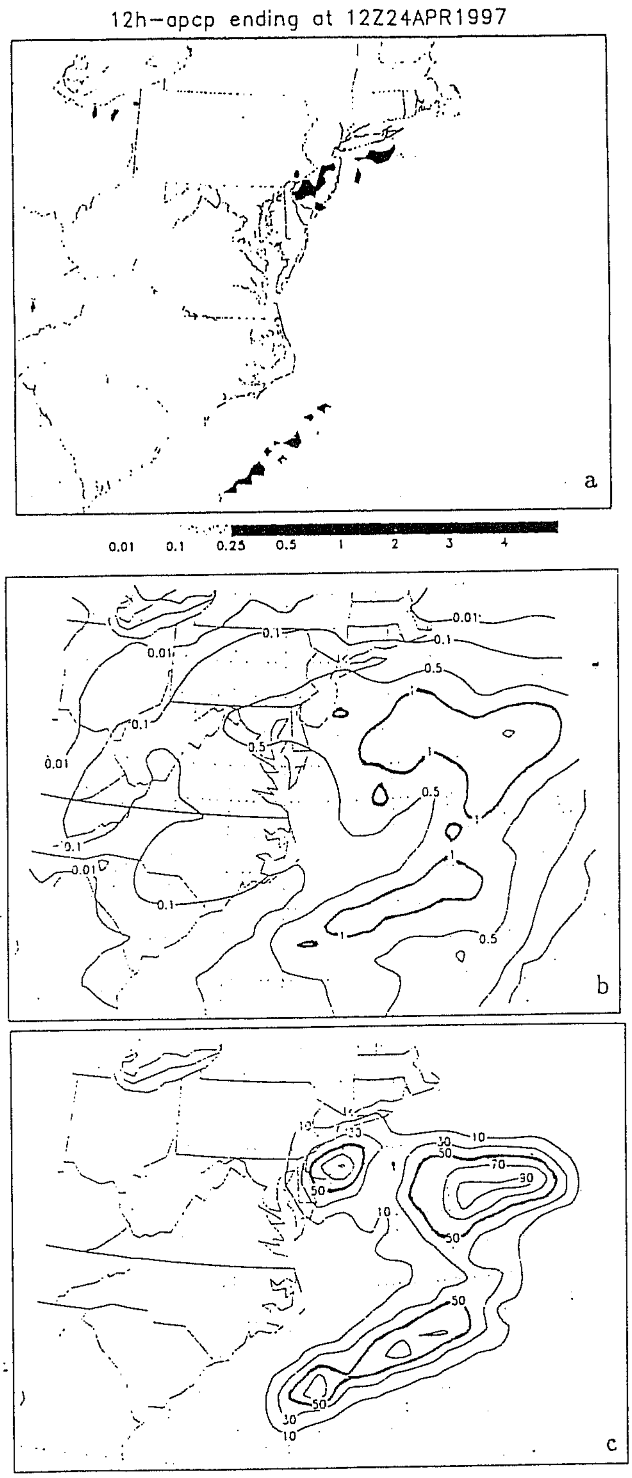

Fig. 2: 12h-accumulated precipitation (inch, 24-36h model forecast period, model initiated from 00 UTC of 04/23/97) in an event of east coast explosive cyclogenesis: (a) observed, (b) control single forecast with 40-km RSM, and (c) probabilistic forecast of exceeding 1.0" based on the 5-member 40-km RSM_r ensemble.

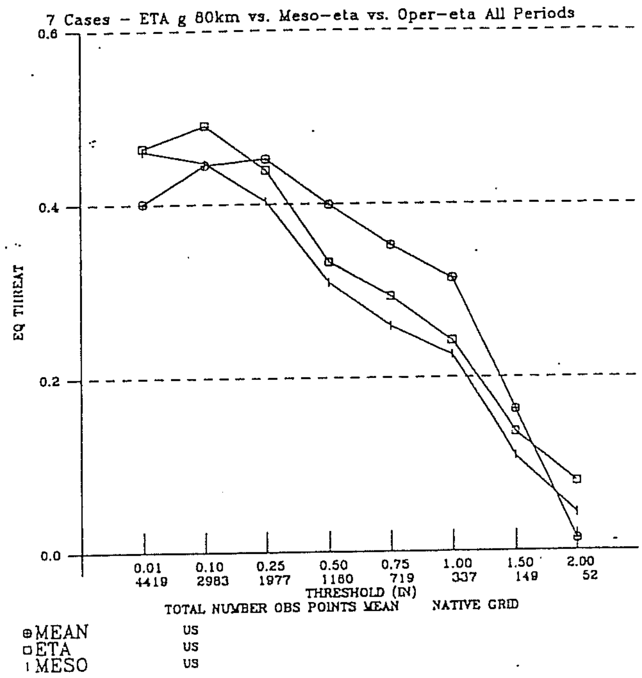

Fig. 3: Rain-gauge based Equitable Threat Scores (ETSs) of 24h-accumulated precipitation at various thresholds. The verification was based on both 24h and 48h forecasting periods. The number of rain gauges used for each threshold is indicated right below the corresponding threshold. "Circle +" is the ensemble mean forecast based on the 5-member 80-km ETA_g ensemble system, "Square" the operational 48-km Eta, and "Pipe" the 29-km meso-Eta. All these results were the averages of 7 explosive cyclone cases.

Eta runs. Also, overall, verification statistics show better cyclone position and central pressure scores with the ensemble mean relative to the relevant single run control forecasts (not shown).

Figure 4 demonstrates that the spread in the ensemble forecasts from "random" perturbations (Fig. 4b) is demonstrably insufficient relative to the bred perturbations[1] (Fig. 4d), even though the magnitude and extent of the "random" perturbations can be larger than that of the "bred" initial conditions

[1] However, spread in precipitation forecasts produced by "randomly" perturbed initial conditions is still quite large (e.g., Du et. al. 1997).

Fig. 4: Standard deviation of sea-level pressure at 00h (a and c with contour interval of 0.2 hpa) and 48h (b and d with contour interval of 0.5 hpa) for the 80-km 5-member Eta ensembles. The upper panel (a-b) is with "random" perturbations (ETA_a), while the lower panel (c-d) with global bred modes (ETA_g). Synoptic case and model initialization time are the same as those in Fig. 2.

(cf. Figs. 4a and 4c).[2]. This conclusion is also true in an overall sense based on 20 cases (Fig. 5). Figure 5 is so-called "Talagrand distribution" or verification rank histogram which indicates percentages of a verifying analysis falling into each bin out of N+1 bins (arranged in an increasing order at each grid point for an N-member ensemble). For an ideal ensemble prediction system (EPS)[3], the verification should be equally likely to occur in each of the N+1 possible bins, i.e. this distribution statistically should be flat. Comparing Fig. 5b with Fig. 5a, it's clear that the ensemble with the "random" perturbations has much smaller spread, showing that most of the time (65%) the verification analysis falls into the lowest or highest extream ranks (i.e. beyond the forecast outlier range, Fig. 5b), than the ensemble with the bred perturbations (Fig. 5a). Note too that the spread information in forecasts is enhanced further with the higher-resolution ensembles (Fig. 6). However, given limited computer resources, there remains the critical issue of the optimal resolution for ensembles versus a single higher-resolution control.

[2] It indicates that the size of the global breeding perturbations is well within realistic range of uncertainty in analysis/initial condition (see also Iyengar et. al. 1996).

[3] An ideal EPS should, at least, meet the following conditions: (1) each member is equally likely, (2) ensemble has an adequate spread, and (3) forecasts have no bias.

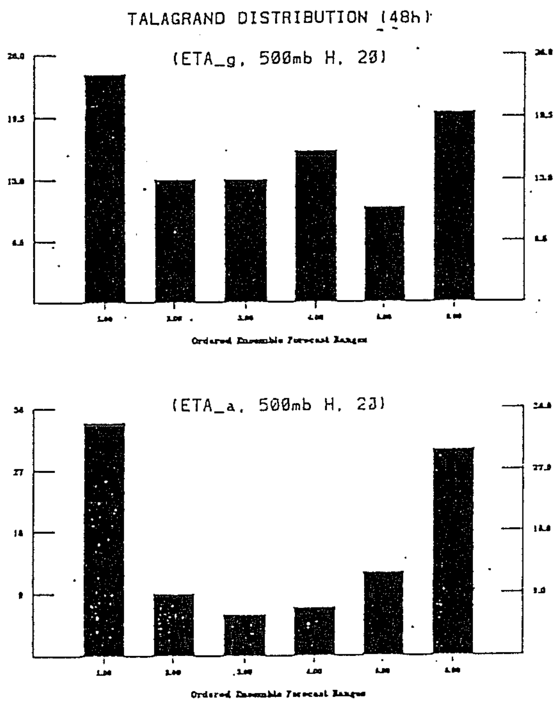

Fig. 5: Talagrand distribution or verification rank histogram for 48h forecast of 500 hpa height with (a) the bred perturbations (ETA_g) and (b) the "random" perturbations (ETA_a). The distribution is an average of 20 cases.

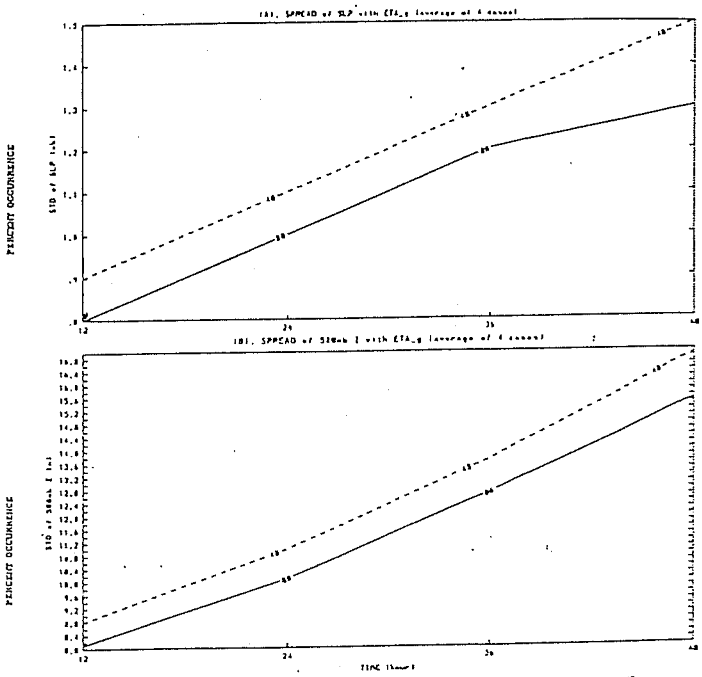

Fig. 6: Ensemble spread (STD) in the 5-member ETA_g ensemble for (a) sea-level pressure and (b) 500 hpa height. The dash curves are for 48km resolution, while the solid curves for 80km resolution. The result is the average of 4 cases.

Figure 7 shows, first, the acute sensitivity of cyclone position and depth to initial condition uncertainty in the RSM and Eta models individually, e.g., cyclone positions can be as far as about 900 km apart from one forecast (p1/RSM in the "SW" cluster) to another (p2/RSM in the "NE" cluster), and, second, the greater information content of a multi-model ensemble (combined RSM and Eta ensembles) as compared to a single model ensemble (or an ensemble based on a single run from different models, not shown), e.g., the probability of the "NE" cluster (in terms of cyclone position) increases from 40% (2/5) to 50% (5/10) near the verification when the Eta ensemble is added to the RSM ensemble. In addition to the increase of information in probabilistic content, a mean forecast from a multi-model ensemble should also be more accurate than that from its component single-model sub-ensembles. Figure 8 shows that by combining the RSM and Eta ensembles the rms error reduction for the ensemble mean forecast (500 hpa height) is quite impressive in comparison to the RSM ensemble alone. However, in these cases, the rms error of the combined ensemble was only comparable to the Eta's. The reason for this is that the RSM predictions have a considerable warm bias, such that majority of times (52%) the analysis falls into the lowest rank (Fig. 9b), while the Eta runs don't have a similarly strong bias (Fig. 9a). This result suggests that forecast bias of each model should be minimized or removed in order to maximize the benefit from a multi-model ensemble.

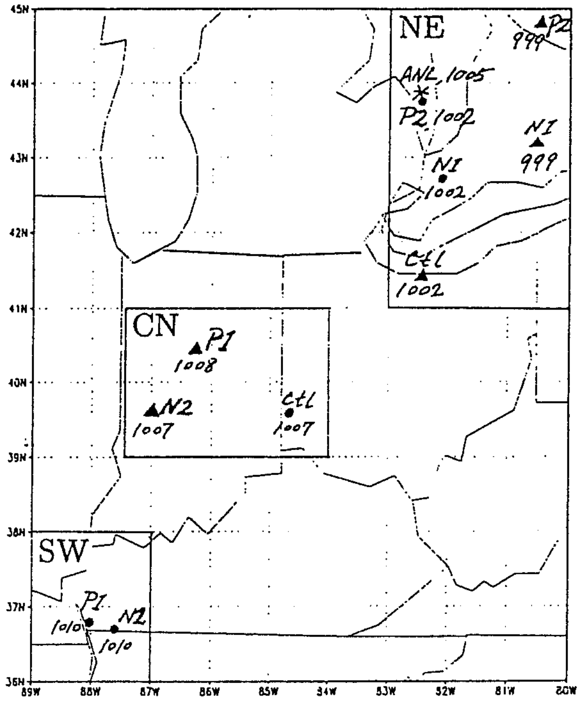

Fig. 7: 48h forecasts of cyclone position and central depth (hpa) at sea level produced by various ensemble members with 80-km RSM_g (dots) and ETA_g (triangles). The asterisk denotes the verification. This cyclone event was associated with a cold front. The models were initiated from 00 UTC of 09/09/97. Note: ctl = control run; n1/p1 (n2/p2) = runs with first (second) negatively/positively perturbed global bred modes.

Figure 10 shows an example of how a simple ensemble averaging process can significantly improve the prediction of a short-wave system over that of a single control forecast. The error in the position of the short wave is reduced in this case by half if the ensemble mean is used (cf. Figs. 10b and 10c). Also, overall verification statistics show smaller rms errors and higher anomaly correlation coefficients for ensemble mean 500 hpa height and sea-level pressure forecasts as compared to those for relevant single control forecasts, especially during the cold season (not shown). This example also indicates that the improved forecast accuracy that results from ensemble averaging is *not* necessarily due to a simple smoothing effect, but

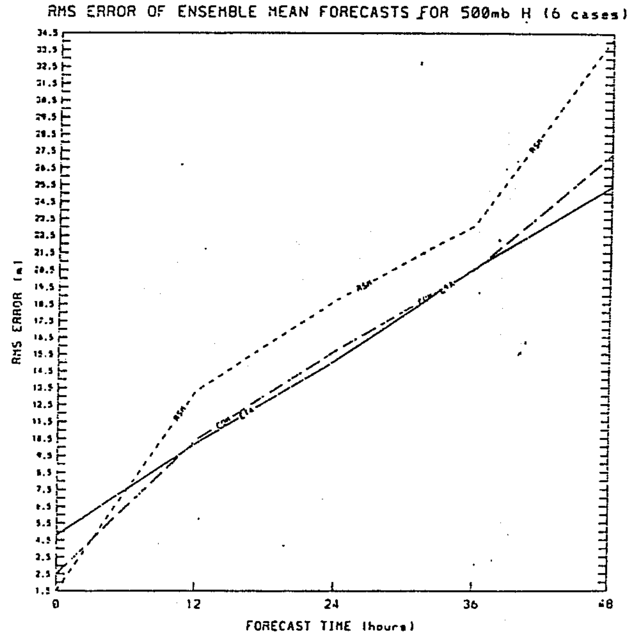

Fig. 8: Rms errors of the ensemble mean forecasts (500 hpa height) for (a) 5-member RSM_g ensemble alone (curve "RSM"), (b) 5-member ETA_g ensemble alone (curve "ETA"), and (c) 10-member RSM_g/ETA_g combined ensemble (curve "COM"). The result is based on 6 cases.

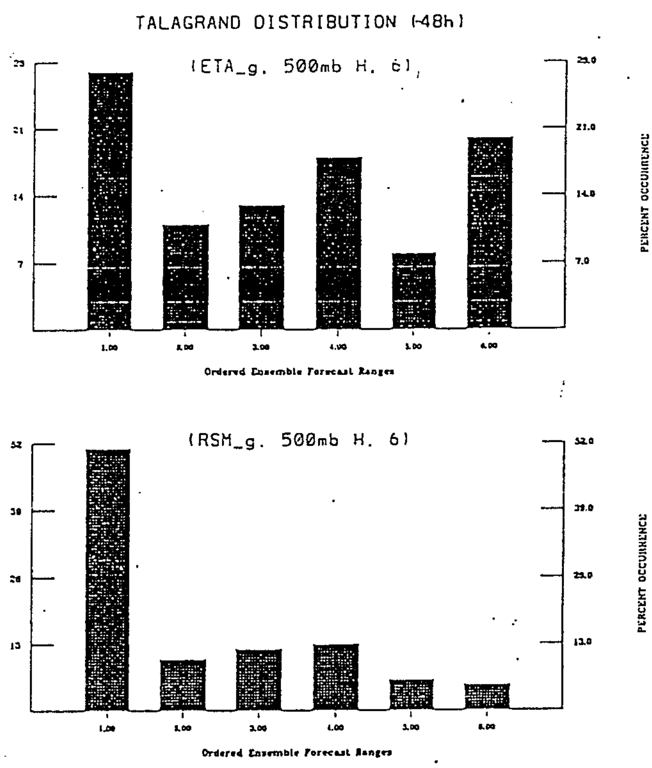

Fig. 9: Talagrand distribution for 48h forecast of 500 hpa height with (a) 5-member ETA_g ensemble and (b) 5-member RSM_g ensemble. The result is based on the same 6 cases as in Fig. 8.

rather to a nonlinear filtering effect that *eliminates unpredictable components* from forecast flow.

5. ACKNOWLEDGMENTS

This research was supported by an appointment to the COMET Post-Doctoral Fellowship Program (Jun Du) sponsored by the National Weather Service/Office of Meteorology and administered by the University Corporation for Atmospheric Research. We would also like to thank Zoltan Toth for his useful suggestions in the process of this project, Henry Juang and Erric Rogers for their help with RSM and Eta models respectively, Tim Marchok for his assistance in setting up our GrADS-based SREF display system, and Perry Shafran for producing the Figure 3.

6. REFERENCES

Brooks, H. E., M. S. Tracton, D. J. Stensrud, G. J. DiMego, and Z. Toth, 1995: Short-range ensemble forecasting (SREF): report from a workshop. *Bul. Amer. Met. Soc.*, **76**, 1617-1624.

Du, J., S. L. Mullen, and F. Sanders, 1997: Short-range ensemble forecasting of quantitative precipitation. *Mon. Wea. Rev.*, **125**, 2427-2459.

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

Iyengar, G., Z. Toth, E. Kalnay, and J. Woollen, 1996: Are the bred vectors representative of analysis errors? Preprints, *11th Conf. on NWP*, Norfolk, VA, Amer. Met. Soc., J64-J66.

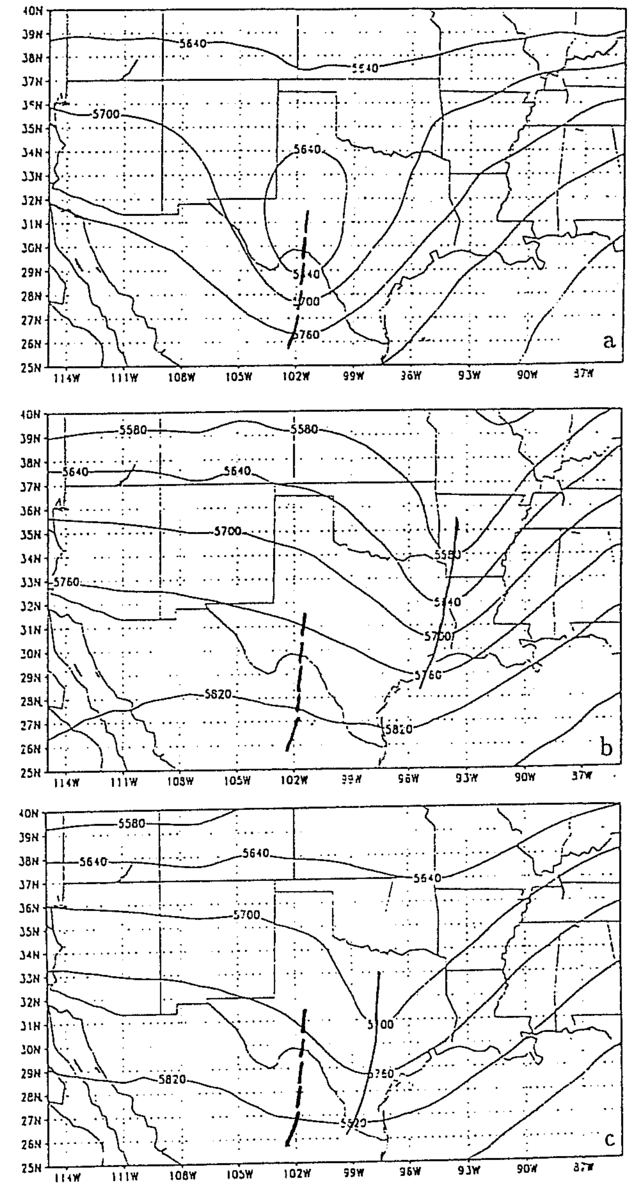

Fig. 10: 48h forecasts of 500 hpa height initiated from 12 UTC of 02/28/97: (a) analysis, (b) 80-km RSM single control run, and (c) ensemble mean forecast based on the 5-member RSM_r ensemble. The broken lines indicate the verification position of the 500 hpa short-wave trough, while the solid lines are the control and mean forecasts, respectively.

Stensrud, D. J., H. E. Brooks, J. Du, M. S. Tracton, and E. Rogers, 1997: Using ensembles for short-range forecasting. submitted to *Mon. Wea. Rev.*.

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297-3319.

Tracton, M. S and E. Kalnay, 1993: Ensemble forecasting at National Meteorological Center: practical aspects. *Wea. Forecasting*, **8**, 378-398.