Using Ensembles for Short-Range Forecasting

DAVID J. STENSRUD AND HAROLD E. BROOKS

NOAA/ERL/National Severe Storms Laboratory, Norman, Oklahoma

JUN DU, M. STEVEN TRACTON, AND ERIC ROGERS

NOAA/NWS/National Centers for Environmental Prediction, Camp Springs, Maryland

ABSTRACT

Numerical forecasts from a pilot program on short-range ensemble forecasting at the National Centers for Environmental Prediction are examined. The ensemble consists of 10 forecasts made using the 80-km Eta Model and 5 forecasts from the regional spectral model. Results indicate that the accuracy of the ensemble mean is comparable to that from the 29-km Meso Eta Model for both mandatory level data and the 36-h forecast cyclone position. Calculations of spread indicate that at 36 and 48 h the spread from initial conditions created using the breeding of growing modes technique is larger than the spread from initial conditions created using different analyses. However, the accuracy of the forecast cyclone position from these two initialization techniques is nearly identical. Results further indicate that using two different numerical models assists in increasing the ensemble spread significantly.

There is little correlation between the spread in the ensemble members and the accuracy of the ensemble mean for the prediction of cyclone location. Since information on forecast uncertainty is needed in many applications, and is one of the reasons to use an ensemble approach, the lack of a correlation between spread and forecast uncertainty presents a challenge to the production of short-range ensemble forecasts.

Even though the ensemble dispersion is not found to be an indication of forecast uncertainty, significant spread can occur within the forecasts over a relatively short time period. Examples are shown to illustrate how small uncertainties in the model initial conditions can lead to large differences in numerical forecasts from an identical numerical model.

1. Introduction

Numerical weather prediction is one of the foundations upon which operational forecasters rely to produce accurate and timely weather forecasts for both public and private uses. Although these computer generated predictions have never been perfect, they offer a picture of the evolution of the atmosphere over some predetermined time window. One of the many tasks a forecaster faces each day is to determine when a numerical forecast is going awry and then to compensate for these errors in the forecast products produced.

Although numerical predictions have improved dramatically in the last 30 yr (Bonner 1989), many sources of error remain. These sources can be separated into contributions from model error, reflecting the unresolved forcings present in the atmosphere but included imperfectly in the model, and errors in observing the true initial state of the atmosphere, including instrument

error, sampling error, and initialization error (Tribbia and Baumhefner 1988). Examples of the sensitivities of numerical weather prediction models to uncertainties in both model initial conditions and model physics abound. For explosive cyclogenesis, it is recognized that the forecast skill is limited as much by initial condition error as by model error (Reed et al. 1988; Sanders and Auciello 1989; Kuo and Low-Nam 1990; Mullen and Baumhefner 1989, 1994). One example of a potential error is that the vertical distribution of latent heat release can have a significant influence on forecast cyclone deepening rates (Gyakum 1983), and this vertical distribution depends upon the model initial and forecast vertical thermodynamic structure, the physical processes included in convective parameterization schemes (Kain and Fritsch 1990), and the specification of rain and ice processes in explicit microphysical schemes (Kogan and Martin 1994).

In simulations of convection, Kain and Fritsch (1992) find that the development and evolution of a squall line is affected greatly by the form of the function that determines where and when the model parameterized convection is activated. There are a number of realistic

Corresponding author address: Dr. David J. Stensrud, NSSL, 1313 Halley Circle, Norman, OK 73069.
E-mail: David.Stensrud@nssl.noaa.gov

alternatives to use for this function, all of which produce distinctly different evolutions of the squall line. There also exist different physical parameterization schemes for the same atmospheric process. Wang and Seaman (1997) show that different convective parameterization schemes produce different evolutions of convective activity in the model simulations and that no one scheme is better than all the others when different measures of skill are compared. Stensrud and Fritsch (1994a,b) show sensitivities to the presence, or absence, of mesoscale details in the model initial conditions. Regions of convective development, and subsequent heavy rainfall, shift by several hundred kilometers depending upon whether or not a mesoscale-sized convective outflow is included in the model initial condition. Zou and Kuo (1996) illustrate similar sensitivities in the model precipitation simulations to mesoscale features in the initial conditions, but suggest that four-dimensional data assimilation of rainfall data can improve the model initial conditions. However, even assuming it is possible to create a perfect initial condition, model error still remains. One approach that attempts to use these model and observational uncertainties advantageously is the use of forecast ensembles. Since weather forecasts always contain a degree of uncertainty, a numerical forecast that explicitly represents this uncertainty would be of great value to many users of weather information (Murphy 1977, 1993).

Ensemble forecasting has the explicit goal of predicting the probability of future weather events or conditions (Epstein 1969; Leith 1974), since it is well known that a single deterministic forecast is sensitive to small uncertainties in the initial condition (Lorenz 1963). As discussed by Leith (1974), an ensemble consists of equally likely analyses of the atmospheric initial state that encompass the unknown ''true'' state of the atmosphere. The mean position of this cloud of analyses in phase space represents the best estimate of the true state of the atmosphere in a least-square-error sense, and the spread between the individual analyses is an estimate of analysis uncertainty. Thus, this set of initial analyses represents an estimate of the probability density function (PDF) for the true initial state. A set of forecasts is then produced using a deterministic model to predict the future state of the atmosphere from each of the analyses; this set of forecasts represents a random sample of the PDF of the atmospheric state at this future time. The mean of the forecast ensemble is the best estimate of the true state of the atmosphere in a least-square-error sense, assuming the model is perfect. However, the forecasts continue to diverge with time, owing to the inherent nonlinearity of the atmosphere, and eventually the mean separation between the forecasts equals the mean separation of randomly chosen atmospheric states. At this point in time the forecast skill is equal to that from climatology.

As discussed by Mullen and Baumhefner (1994), the goal of ensemble forecasting is to specify the evolution of the atmospheric PDF as completely as possible. Unfortunately, while the idea of ensembles is simple, the implementation of an ensemble strategy is not. It is often assumed that by using a set of different initial conditions in which each initial condition is constrained to match to the basic observational data, while also differing from all the other initial conditions in an estimate of the observational error, the PDF of the true initial state can be obtained. This process usually is accomplished by using the control analysis as the best estimate of the true initial state and adding perturbations to the control, thereby virtually guaranteeing that the mean of the ensemble initial states is the control analysis. Several methods exist for defining these perturbations, including Monte Carlo (Mullen and Baumhefner 1988), the breeding of growing modes (Toth and Kalnay 1993), and singular vectors (Molteni et al. 1996), although the best method to use for operational ensembles is not yet known (see Anderson 1996).

One problem common to all these approaches is the general lack of knowledge of analysis error. Daley and Mayer (1986) have estimated global analysis error, but the case-to-case variability in the uncertainty of the true atmospheric state could be much larger or smaller than the estimated global analysis error (Mullen and Baumhefner 1989). As the scales of phenomena are decreased, our understanding of analysis error also decreases. On the mesoscale, we lack the observations necessary to determine even the mean analysis error, although special field programs have highlighted the fine-scale structure of the atmosphere that is missed by the present observational network (e.g., Hoecker 1963; Johnson and Hamilton 1988; Schneider 1990; Ziegler and Hane 1993). This lack of knowledge is especially troublesome when trying to define a PDF for an ensemble of mesoscale model initial conditions, which is currently under exploration (Stensrud et al. 1998).

Another problem is the computer resources needed for ensemble forecasting. Although Du et al. (1997) have shown that as few as 8–10 ensemble members can improve precipitation forecasts for a case of explosive cyclogenesis, it is desirable to have a much larger number of ensemble members to produce a better estimate of the PDF for the atmosphere and obtain a more accurate estimate of the spread (see Epstein 1969). Producing a larger number of model forecasts is computationally expensive and present computer resources at most operational centers are not sufficient to create more than 10–50 ensemble members. Using a relatively small number of ensemble members makes the methods by which the initial conditions are perturbed even more important, since it is then necessary to sample as much of the PDF as possible with very few points.

Even with the difficulties involved, ensemble strategies have been found to be of benefit to operational forecasting. Ensemble techniques have been used for global medium-range (5–15 day) forecasting and have demonstrated the usefulness of ensemble predictions

over a single control forecast valid at the same time (Dalcher et al. 1988; Murphy 1990; Brankovic et al. 1990; Tracton and Kalnay 1993; Toth and Kalnay 1993; Molteni et al. 1996; Buizza 1997). Medium-range ensembles also have shown utility in the forecasting of extreme weather events during the cool season over the Mediterranean region (Petroliagis et al. 1997), in which the ensemble probabilities are used to provide a measure of confidence in a higher-resolution numerical forecast valid at the same time. Similar results indicating that the spread between the ensemble members can be used to predict forecast skill are shown by Kalnay and Dalcher (1987) for 500-hPa height patterns. Although the spread in the ensemble is less than the difference between the forecast and the validating analysis as shown by Buizza (1995), indicating that the methods for generating the ensemble do not necessarily capture the PDF for the true atmosphere, the usefulness of the ensemble approach for medium-range forecasting is clear. The success of global model-based ensemble forecasting techniques at the medium range, and evidence from case studies of ensemble systems used for short-range prediction (Mullen and Baumhefner 1988, 1989; Du et al. 1997), led to the development of a pilot program at the National Centers for Environmental Prediction (NCEP) in 1994 to explore the usefulness of regional short-range ensemble forecasting (SREF) (Brooks et al. 1995).

Initial results from the SREF pilot program suggest that the ensemble approach can provide value for probabilistic quantitative precipitation forecasts (PQPFs). While Hamill and Colucci (1997) find that the NCEP ensemble forecasts of precipitation are underdispersive, the precipitation forecasts can be postprocessed to correct this problem and yield an adjusted ensemble with more desirable statistical properties. They further show that the ensemble can generate forecasts that have similar or lower error than forecasts from the 29-km Meso Eta Model. In a subsequent paper, Hamill and Colucci (1998) show that the accuracy of the adjusted PQPFs from the NCEP ensemble are more skillful than the Nested Grid Model's model output statistics for all precipitation categories except the basic probability of measurable precipitation. However, the ensemble output shows no ability to predict the forecast skill of precipitation for the cases examined.

The goal of the present paper is to further examine the output from the NCEP SREF pilot program, but to focus upon different parameters and also illustrate the spread seen in some of the ensemble forecasts. We briefly summarize the approach to developing the ensemble member initial conditions in section 2, and then proceed to a verification of the SREF output in section 3. Several illustrations of ensemble spread are shown in section 4, followed by a final discussion in section 5.

2. Creation of the ensemble member initial conditions

The configuration of the ensemble members and choice of numerical weather prediction model for this pilot program on SREF were determined at a workshop held at NCEP during July 1994 (Brooks et al. 1995). The pilot program was designed in an attempt to address most of the scientific questions and concerns of the workshop members, while being constrained by the computational and human resources of NCEP. The result is a 15-member ensemble, with 10 of the members from the 38-level, 80-km Eta Model (Janjić 1994) and 5 members from the 28-level, 80-km regional spectral model (RSM) (Juang and Kanamitsu 1994). Six of the Eta Model initial conditions are from different analyses that are interpolated to the Eta Model grid. These analyses are from a static eta optimum interpolation analysis (Rogers et al. 1995), the Nested Grid Model regional analysis (DiMego et al. 1992), the Eta Model data assimilation system (Rogers et al. 1996), the 3D-variational analysis (Parrish et al. 1996), the medium-range forecast model (MRF) control forecast (Parrish and Derber 1992), and the aviation run of the MRF (Parrish and Derber 1992). The remaining four Eta Model initial conditions are from two positive and two negative bred perturbations (Toth and Kalnay 1993). The five RSM initial conditions are from the MRF control forecast and two positive and two negative bred perturbations. Other details of this ensemble configuration can be found in Hamill and Colucci (1997).

The six different analyses for the Eta Model initial conditions to some extent represent analysis uncertainty, since the analyses are created using different techniques, although also with slightly different input datasets owing to the varying data cutoff times that depend upon the operational model suite execution structure. The breeding of growing modes (BGM) technique is another approach to generating ensemble members in which the numerical forecast model is used to determine the fastest growing modes that are then inserted back into the model initial conditions (Toth and Kalnay 1993). A key difference between the BGM approach and using different analyses to vary the initial conditions is that the breeding approach imposes a dynamic constraint on the initial conditions, whereas the different analyses represent estimates of truth from the available input data as seen through the lenses of different analysis techniques. However, several of the analysis techniques produce initial conditions that are very similar to each other, suggesting that the uncertainty in the initial state of the atmosphere is likely underestimated using this approach. Although the NCEP SREF initial condition perturbations vary in magnitude, with the bred modes typically having larger root-mean-square differences (RMSD) in 500-hPa heights and smaller RMSD in 850-hPa temperatures than the analyses (Hamill and Colucci 1998), the ensemble mean of these fields corresponds fairly closely to the Eta Model initial condition from the aviation run of the MRF (see Hamill and Colucci 1998). In some sense, this ensemble can be viewed as perturbations surrounding the aviation run initial condition

where this initial condition represents the best estimate of the true atmospheric state.

A total of 81 cases now exist of the ensemble forecasts out to 48 h with model fields available at 6-h intervals for the Eta Model and 12-h intervals for the RSM. Some cases do not contain all 15 of the ensemble members, owing to model and data transmission problems. In addition, 6-h output from the 50-level, 29-km Meso Eta Model (Black 1994) out to 36 h also are archived to compare the relative benefits of a single, higher-resolution model forecast to the multiple run, low-resolution ensemble.[1] The versions of the eta and RSM used in this study are identical to those used operationally at NCEP, and have changed over time as the operational models have been upgraded. We begin by verifying the ensemble mean of basic atmospheric variables at mandatory pressure levels with output from the Meso Eta Model and observations.

3. Verification of the ensemble and Meso Eta forecasts

One approach to documenting the usefulness of SREF is to compare the ensemble mean with observations and a single higher-resolution model forecast. The ensemble mean should provide a better forecast than any individual ensemble member owing to errors in the individual forecasts canceling when averaged (Epstein 1969; Leith 1974). However, an important question to address is whether or not the ensemble mean of a low-resolution model can produce forecasts of model parameters that are as accurate as those from a higher-resolution model forecast.

a. Standard atmospheric variables

Using the ensemble mean of the 10 80-km eta forecasts valid at the same time, the bias, mean absolute error (MAE), and root-mean-square error (rmse) of temperature, relative humidity, geopotential height, wind speed, and wind direction are calculated from 20 forecasts started at 1200 UTC. Model data are interpolated to the location of the rawinsonde sites without correcting for balloon drift or duration of ascent. These same verification parameters also are calculated from the 29-km mesoscale version of the Eta Model as archived on a 40-km grid. The cases chosen are the first 20 cases in which both the 10 member eta ensemble data and the Meso Eta Model data are available. We have neglected

the RSM data in order to simplify our interpretations, since the identical numerical model is then used for both the 80-km and 29-km Eta Model forecasts. The verification parameters are calculated at 850, 700, 500, 300, 200, and 100 hPa for all rawinsonde observations available within North America at the 12-, 24-, and 36-h forecast times. The values of relative humidity are not verified above 300 hPa owing to difficulties with correctly measuring humidity at low temperatures (Elliott and Gaffen 1991; Wade 1994). For each time period and parameter over 1200 observations are included in the calculations.

Results indicate that the ensemble mean compares favorably with the Meso Eta Model for all parameters and pressure levels (Table 1). To estimate the probability that the ensemble mean forecast is more accurate than the Meso Eta Model, or vice versa, we compare the domain-average errors of each variable from both models over all case days using the Wilcoxon signed-rank test (Wilks 1995). For each variable and output time, we obtain 20 pairs of the value of the domain-average error from which we calculate the absolute value of the error differences. These difference values are ranked from 1 to 20 in ascending order and the rank values summed for values with the same sign of the error difference. The statistical significance of the difference of the two distributions is tested by calculating the specific probability that the sum of the ranks will occur according to the null distribution. Results indicate that most of the differences between the ensemble mean and the Meso Eta Model are significant at the 95% level (Table 1), with the ensemble mean more accurate than the Meso Eta Model for just over half of the variables chosen for examination. Although only the MAE is shown, the overall comparison between the ensemble mean and the Meso Eta yields a nearly identical outcome when using the rmse.

These comparisons indicate that the mean of the small 10-member ensemble is producing a level of forecast accuracy of the basic atmospheric parameters better than or equivalent to the 29-km Meso Eta Model for most parameters, a model with over twice the horizontal resolution and 1.3 times the vertical resolution of the 80-km Eta Model used to generate the ensemble runs. This result parallels that found in ensembles of medium-range forecasts, namely that the loss of skill from a reduction in model resolution can be recovered by using an ensemble approach (Tracton and Kalnay 1993). However, in terms of CPU time, the 10-member eta ensemble is 1.45 times more expensive than the single run of the Meso Eta Model, and the 15-member eta and RSM ensemble is twice as expensive as the single run of the Meso Eta Model. Therefore, one hopes that some added information could be gained from the ensemble that is not provided by the Meso Eta Model.

For the medium-range forecast problem, one added benefit gained from using an ensemble approach is that the dispersion between members of an ensemble can be

---

[1] The Meso Eta Model forecasts are actually started at 0300 and 1500 UTC after a 3-h data assimilation period. Therefore, we are comparing the 9-h, 21-h, and 33-h Meso Eta Model forecasts with the 12-h, 24-h, and 36-h ensemble forecasts. To avoid confusion we have chosen to ignore the 3-h time difference in these forecasts and consider the Meso Eta Model forecasts to be 12-h, 24-h, and 36-h forecasts.

TABLE 1. Mean absolute errors in temperature (K), RH (%), geopotential height (m), wind speed (m s⁻¹), and wind direction (degrees) calculated from the 29-km Meso Eta Model (meso) and the 10-member 80-km Eta Model ensemble mean (ens) at 12 h, 24 h, and 36 h into the model forecasts for various mandatory pressure levels (850, 700, 500, 300, 200, and 100 kPa). Values in bold print indicate that the differences are significant at the 95% level.

	12 h meso	12 h ens	24 h meso	24 h ens	36 h meso	36 h ens
Temperature MAE (K)						
850	**1.29**	1.39	1.5	**1.48**	1.71	1.79
700	1.08	**1.03**	1.16	1.16	1.29	**1.26**
500	0.93	**0.85**	0.97	**0.9**	1.16	**1.09**
300	1.04	**0.92**	1.1	**1**	1.23	**1.19**
200	1.38	**1.3**	1.59	**1.35**	1.73	**1.64**
100	1.66	**1.28**	1.59	**1.31**	1.68	**1.46**
Relative humidity MAE (%)						
850	**12.68**	13.03	14.08	13.94	**14**	14.36
700	14.33	**13.87**	15.08	15.25	14.9	15
500	15.03	**14.09**	16.4	**15.78**	17.11	**16.48**
300	14.42	**13.24**	14.62	14.32	15.82	**15.06**
Geopotential height MAE (m)						
850	10.77	**10.2**	**13.37**	16.17	14.58	16.1
700	10.07	**9.98**	**13.26**	15.99	14.3	15.69
500	**11.16**	12.71	**14.19**	17.85	**15.88**	18.36
300	**16.4**	19.11	26.73	24.23	23.86	26.87
200	**21.28**	23.41	33.02	28.19	29.35	31.85
100	30.92	29.52	30.42	**28.66**	**34.14**	35.93
Wind speed MAE (m s⁻¹)						
850	**2.15**	2.26	**2.48**	2.57	2.48	2.48
700	2.48	**2.34**	2.58	**2.55**	2.8	**2.75**
500	2.65	**2.55**	2.94	**2.82**	3.24	**3.19**
300	3.57	**3.45**	4.15	**3.92**	4.49	**4.45**
200	4.02	**3.75**	4.46	**4.18**	4.87	**4.5**
100	3.43	**2.92**	3.35	**3.05**	3.45	3.51
Wind direction MAE (degrees)						
850	31.88	**31.15**	**33.02**	34.88	**33.42**	35.51
700	24.04	**23.67**	**25.03**	25.96	26.58	28.96
500	**16.26**	18.35	**17.38**	19.11	**19.29**	20.92
300	**14.62**	15.48	**14.52**	15.96	**17.57**	18.76
200	**13.31**	15.37	**14.42**	16.31	**16.79**	17.74
100	19.12	19	**17.84**	18.32	**19.77**	21
Lapse rate MAE (×10⁻³ K m⁻¹)						
700–500	0.533	**0.500**	0.601	**0.547**	0.625	**0.578**
700–300	0.243	**0.228**	0.268	**0.257**	0.287	**0.270**

used to predict the skill of the weather forecasts (Kalnay and Dalcher 1987; Buizza 1995). Results from Hamill and Colucci (1998) indicate that this is not true for the SREF dataset. Their results show little correlation between the ensemble variability and the skill of the precipitation forecast in the 13 cases they examined. To further explore the SREF data, we examine forecasts of cyclone location as done for explosively developing cyclones by Junker et al. (1989), since the SREF pilot program is focused upon the prediction of sensible weather events and cyclones are important contributors to sensible weather.

b. Cyclone position

As another way to evaluate the information content of the ensembles, we examine the locations of cyclones over North America from the ensemble mean and the Meso Eta Model. Cyclones that are over the continental United States in the model initial conditions and remain within the model domain during the entire model forecast are documented. However, before comparing results from the ensemble mean and the Meso Eta Model, it is important to determine the most appropriate method for defining the ensemble mean of the cyclone position. The ensemble mean low position can be defined in two ways: 1) the cyclone position defined by the ensemble mean sea level pressure field (MSLP), and 2) the cyclone position defined by the mean position of the individual cyclones (MPIC) from each of the ensemble forecasts.

To maximize the number of cases, output from only the 48-h time of the 10 80-km Eta Model forecasts are used, providing a total of 44 cases from August 1995 through September 1997. The cyclone centers are located to a precision of 1° lat from the model data, and from the NCEP surface analyses before June 1997 and an objective analysis of the sea level pressure for the two cases that occurred after June 1997. If a cyclone

TABLE 2. MAE in cyclone position (° lat) at 36 h from 33 cases using the 29-km Meso Eta Model (Meso Eta), the MPIC from the 15-member eta and RSM ensemble, and the 80-km eta model ensemble member that uses the operational initial condition (OPNL).

Model	MAE
Meso Eta	2.19
MPIC	2.20
OPNL	2.58

TABLE 3. Spread calculated from forecasts at 36 h using the BGM technique and from the different in-house NCEP analyses defined as the mean distance (in ° lat) between each of the individual ensemble members and their MPIC. Nine ensemble member initial conditions are created using the BGM technique, and six are created from different analyses. For the Eta vs RSM spread comparison, both calculations use five ensemble members: one from the MRF control initial condition plus the two positive and two negative bred perturbations.

Ensemble members	Spread
All	1.62
Analyses	0.95
Bred modes	1.80
Eta bred modes	1.41
RSM bred modes	1.53

does not occur in a model forecast, or if the forecast is unavailable, then the MPIC and MSLP are calculated only from the remaining ensemble members.[2] The maximum absolute difference between the MPIC and MSLP from all the cases is 3.9° lat, whereas the mean absolute difference is 0.76° lat, indicating that these two methods produce very different determinations of cyclone location. Results indicate that the MPIC has a MAE of 2.36° lat, the MSLP has a MAE of 2.64° lat, and, for comparison, the operational version of the 80-km Eta Model has a MAE of 3.13° lat. Using the Wilcoxon signed-rank test, the differences in the MAEs of the MPIC and MSLP are estimated to be at the 77% significance level, whereas the differences in the MAEs of the MPIC and the operational version of the 80-km Eta Model are estimated to be at the 99% significance level.

Part of the difficulty in evaluating the ability of a model to forecast specific types of events is that, even for synoptic-scale features, the number of times these events occur, and for which good verification data are available, is much less than the number of forecasts. This requires a very large dataset in order to get statistically significant results. Unfortunately, the 81 ensemble forecast case days presently available are not always sufficient to provide high levels of statistical significance for specific types of weather events that are important to providing a good evaluation of the forecast potential for short-range ensembles. Therefore, the numbers presented here must be viewed with a degree of caution. Nevertheless, it appears that the MPIC may be a better estimate of the forecast cyclone position than the MSLP and we choose to use the MPIC to define the ensemble cyclone position for a comparison of the ensemble mean with the Meso Eta Model.

A total of 33 cyclone cases are documented in both the 15-member eta and RSM ensemble and the Meso Eta Model 36-h forecasts for which verification data are available. Results indicate that the MAEs of the ensemble mean and Meso Eta Model cyclone locations are virtually identical (Table 2) and both have smaller

MAEs than that of the operational version of the 80-km Eta Model. Whereas the differences between the Meso Eta and ensemble cyclone locations are not significant, both forecasts of cyclone location are more accurate than the operational version of the 80-km Eta Model at the 90% significance level. Therefore, both higher horizontal resolution and the use of ensembles can improve forecasts of cyclone location. The virtually identical position errors of the ensemble mean and the Meso Eta suggest that the the loss of accuracy from a reduction in model resolution can be recovered by using an ensemble approach.

The cyclone position dataset also can be used to evaluate the amount of dispersion produced by the six different analyses versus the nine bred modes. Hamill and Colucci (1998) show that the bred modes on average have 25% larger domain-averaged 500-hPa height perturbations than the different analyses, whereas the analyses have 18% larger 850-hPa temperature perturbations than the bred modes. A natural question is which method, as applied in this NCEP pilot program, is more effective in producing dispersion in the forecasts. This is an important question to answer, since ensemble forecasts typically are underdispersive (Buizza 1997). If we define spread as the mean distance between all individual ensemble member cyclone positions and their MPIC, then the results from the 36-h forecasts indicate that the bred modes produce 0.85° lat more spread than the different analyses (Table 3) (see Tracton et al. 1998). This difference is significant at the 99% level. Although the difference in spread between the different analyses used in the Eta Model and only the Eta Model bred modes is smaller (0.46° lat), it is still statistically significant. The spread calculated from only the RSM bred modes is not statistically different from the spread calculated from the Eta Model bred modes. However, the spread found with the bred modes at 36 h increases by 28% when using the two different numerical models in the ensemble, and this difference is significant at the 98% level. This result strongly suggests that model differences are important contributors to spread in this ensemble, in agreement with Tracton et al. (1998).
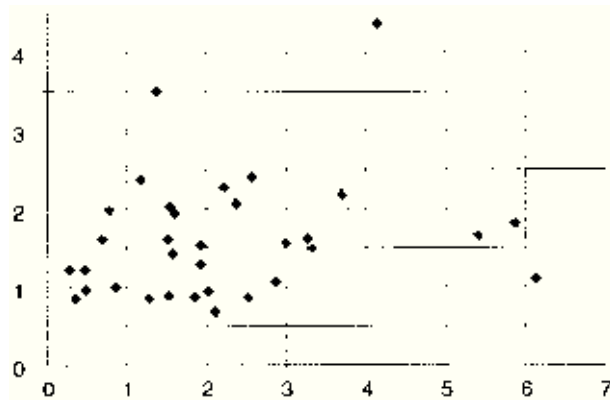
---

[2] This leads to an overestimate of forecast skill, since a priori one does not know if a low exists at a given forecast time or not. Only a few of the forecasts used in these calculations show uncertainty with regard to the presence of a cyclone, suggesting that for the cases chosen, the overestimate of forecast skill is small. However, for larger datasets this may be a concern.

FIG. 1. Plot of the MPIC error (° lat) vs the spread from the ensemble (° lat) defined as the mean distance between all ensemble members and the MPIC.
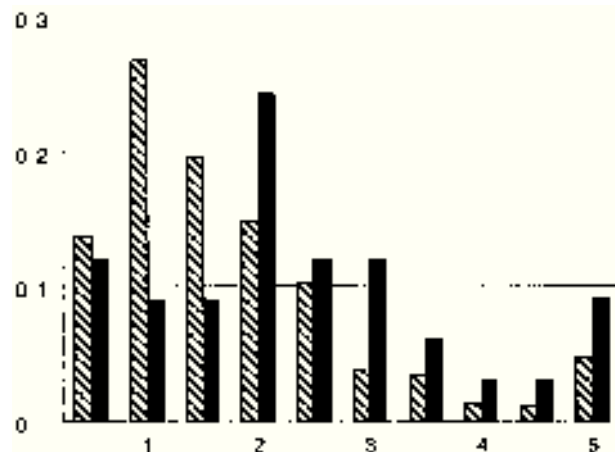

FIG. 2. Histogram of the relative frequency of ensemble (diagonal line) and observed (solid) cyclone distances from the MPIC at 36 h binned into 0.5° lat categories. All values greater than 5° lat are included in the 5° bin.

To examine how the spread changes with time, the cyclone position data from only the Eta Model ensemble members at the 48-h time are examined. The spread is calculated from five different analyses and the four bred modes combined with a control forecast, yielding five forecasts each. Results indicate that the spread from the different analyses is 1.49° lat and from the bred modes is 1.84° lat. The spread has increased 30% over the past 12 h from the bred modes, whereas it has increased over 57% from the analysis differences. Therefore, by 48 h the bred modes produce only 0.35° lat more spread than the different analyses, a smaller difference than found at 36 h. These results suggest that the BGM technique produces more spread in this ensemble than does analysis differences, although the amount of increased spread attributed to the BGM perturbations appears to decrease between 36 and 48 h. This result also parallels that found by Tracton et al. (1998), who investigated 500-hPa geopotential height and sea level pressure fields.

One of the hopes of ensemble forecasting is that the spread, or dispersion, of the ensemble provides information on the uncertainty of the forecast. Forecasts with larger spread should be less certain than forecasts with smaller spread (see Hamill and Colucci 1998). This relationship has been seen to some extent in medium-range forecasts of 500-hPa heights (Kalnay and Dalcher 1987). One way to examine the relationship between spread and forecast accuracy is to calculate the correlation coefficient for the two parameters. But there are at least two ways to define forecast accuracy: the MPIC position error and the control forecast position error. Here we define the control forecast as the Meso Eta Model forecast. When the MPIC position errors are compared to the spread (Fig. 1) there is very little correlation ($r = 0.21$), indicating that the ensemble is not able to predict the forecast skill of cyclone locations. There is a slightly higher correlation ($r = 0.36$) between the Meso Eta Model cyclone location errors and the ensemble spread, but even using this small data sample it appears that this ensemble cannot predict forecast skill for cyclone location. Indeed, if we examine the error in cy-

clone location of the MPIC from the Eta Model bred modes and the Eta Model analyses, we find that the errors are 2.48 and 2.47° lat, respectively. Therefore, the increased dispersion produced by the Eta Model bred modes does not correspond to a decrease in accuracy for cyclone position at 36 h. This result parallels that found by Hamill and Colucci (1998), who find that the ensemble variability cannot be used to predict the forecast variable specificity of the ensemble probability distribution from day-to-day and location-to-location.

Part of the inability of the ensemble to predict forecast skill may be due to the underdispersion of the ensemble. If the distance between each ensemble member and the MPIC is calculated and divided into 0.5° lat bins, and then compared against the distance between the MPIC and the observed cyclone location, we find that 60% of the ensemble members are within 2° lat of the MPIC, whereas only 30% of the observed cyclones occur within the same range (Fig. 2). This result is consistent with the analysis of Persson (1996), who shows that the relationship between spread and error decreases if the ensemble members are correlated with one another. Therefore, this small dataset strongly suggests that the model forecasts are underdispersive and one of the challenges of ensemble forecasting is creating an ensemble system that is more dispersive and less correlated while also maintaining or improving the accuracy of the ensemble mean. Persson (1996) indicates that as the numerical model improves and as the ensemble members become more uncorrelated, the covariance between spread and error should increase. An ensemble system with larger spread has recently been created at NCEP using an improved numerical model (Tracton et al. 1998), but it is too early to determine if the relationship between spread and error in this ensemble system has improved.
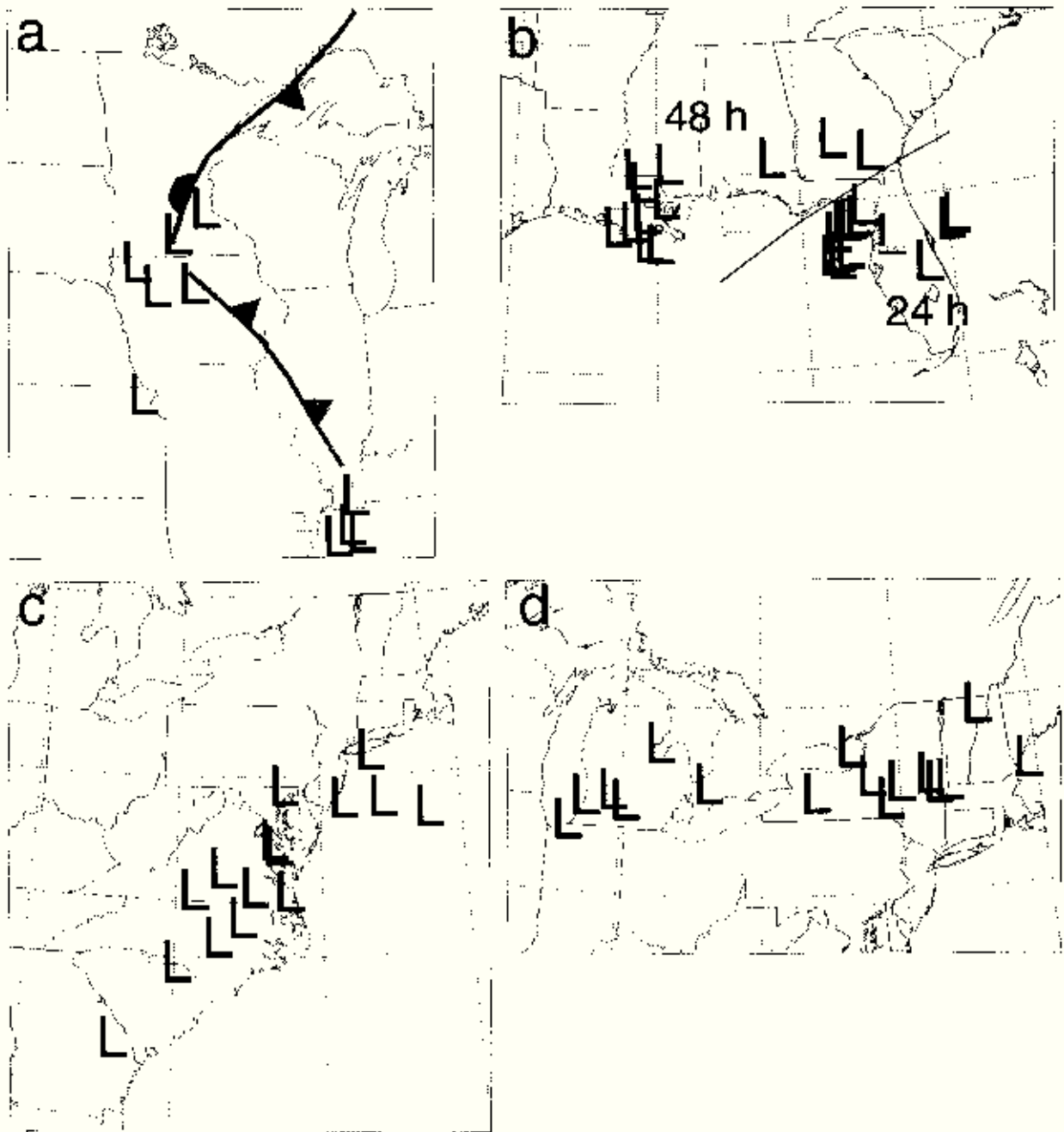
FIG. 3. Plots of cyclone positions from (a) 48-h forecasts valid 1200 UTC 23 Sep 1994, (b) 24-h and 48-h forecasts valid 1200 UTC 3 Aug 1995, (c) 48-h forecasts valid 1200 UTC 19 Dec 1996, and (d) 48-h forecasts valid 1200 UTC 19 Jul 1996. Locations of cyclones from individual ensemble members indicated by a black L and the observed location is indicated by a gray L.

4. Ensemble variability

A number of cases have been identified to illustrate the significant amount of dispersion that occurs in the ensemble over relatively short time periods. These are shown to illustrate that even though, by some measures, the spread in the ensemble members may decrease initially before the most unstable modes begin to grow (Lacarra and Talagrand 1988), significant differences between ensemble members can be found in sensible weather parameters very soon after the start of the forecasts. This is illustrated schematically by Toth and Kalnay (1993), who show that the low-energy convective modes grow fastest initially and then saturate at a smaller amplitude than the baroclinic modes that grow more
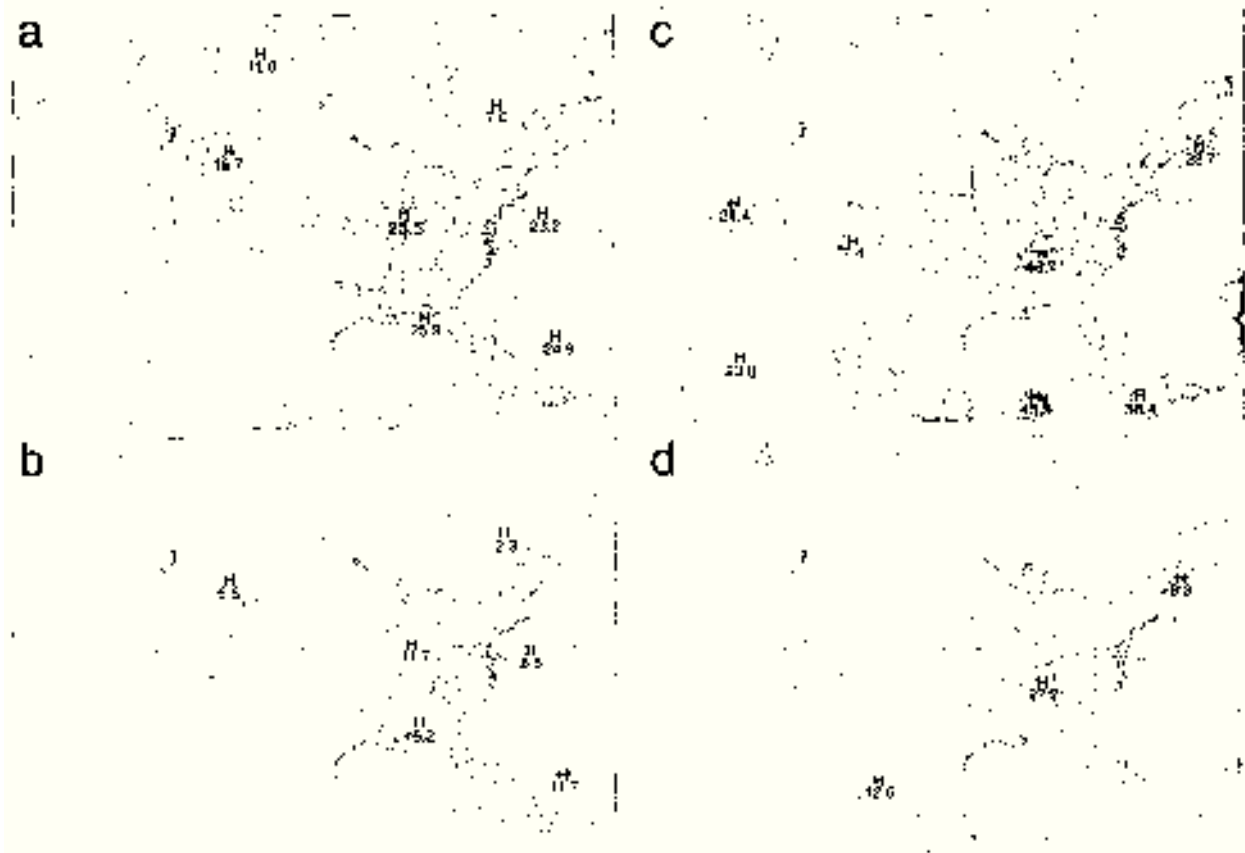
FIG. 4. Forecast 12-h (a) maximum rainfall (mm) and (b) minimum rainfall valid 0000 UTC 29 May 1996, and (c) maximum rainfall and (d) minimum rainfall valid 1200 UTC 9 Sep 1997. Rainfall extrema calculated at each individual grid point from all ensemble members.

slowly. However, for SREF it is not clear that these convective modes are unimportant, since precipitation is one of the most critical short-range forecast parameters. Since the production of short-range ensembles has occurred only recently, a more qualitative look at the variety of the forecasts produced, and the different types of events captured, may be helpful to illustrate the dispersion that is created within this small ensemble.

a. Cyclones

Four different cases are chosen to illustrate the dispersion that occurs in cyclone location at 36 and 48 h into the ensemble forecasts (Fig. 3). Most of the cases examined do not show a large spread in cyclone location at the 24-h forecast time; the spread begins to become more clearly evident at 36 h and is larger at 48 h. Admittedly, these cases illustrate the maximum dispersion seen in the ensemble dataset and are not in general representative of the ensemble behavior on every day where the spread on average is less (see Fig. 2). However, it is important to note that spread can occur in all seasons and may indicate more than just different placements of the cyclones. On 21 September 1994 (Fig. 3a), output from the ensemble member that positioned the

low in northeastern Kansas indicates that this cyclone has occluded by moving into the cold air behind the frontal boundary. Therefore, some of the differences between the ensemble members are produced by more than just differences in phase; they may also indicate changes in evolution.

b. Rainfall

As indicated in the study by Du et al. (1997), there can be large variability in the rainfall totals from ensemble member to ensemble member. This behavior is seen consistently in the NCEP short-range ensemble output. This variability may be illustrated most clearly by calculating the maximum and minimum 12-h rainfall totals at each grid point in the 80-km Eta Model domain and plotting the resulting rainfall fields (Fig. 4). The minimum rainfall field encompasses where all the ensemble members produce rainfall over this 12-h time period, whereas the maximum rainfall field encompasses the area over which any ensemble member produces rainfall. Typically the extrema in the minimum rainfall field are 2–3 times smaller than the extrema in the maximum rainfall field, regardless of synoptic setting. It is clear that the ensemble members disagree significantly
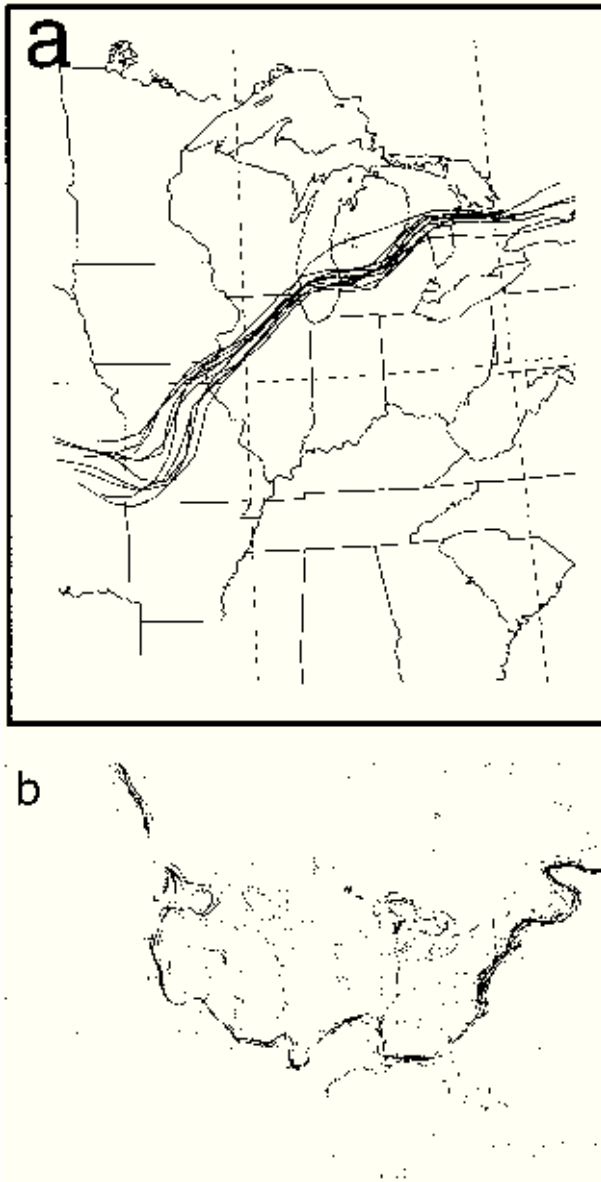
FIG. 5. Location of the 2-m freezing line valid (a) 0000 UTC 28 Nov 1995 and (b) 1200 UTC 18 Nov 1997. Shaded areas in (a) denote regions that received snowfall of over 4 in. for this event.

on which regions receive rainfall. However, Hamill and Colucci (1998) indicate that this spread is not related to forecast skill.

c. Freezing line

Forecasts of heavy snowfall are often assisted by a good prediction of the freezing line, since the heaviest snowfall typically occurs just to the cold side of the freezing line. Output from the ensemble shows that significant differences in the freezing line position can occur within the first 12 h of the model forecasts. Differences of up to 250 km are seen in the placement of the freezing line during a heavy snowfall event in Michigan (Fig. 5a). Larger differences in placement are seen in the November 1997 case (Fig. 5b) where the northernmost location of the freezing line in the central United States varies from Texas to South Dakota, whereas in other parts of the United States the ensemble members are more in agreement.

d. Convective available potential energy

One parameter that is used to assist in forecasting the general type of convective activity is the convective available potential energy (CAPE). Forecasts of CAPE also show significant variability at the earliest forecast times (Fig. 6). Locations of local extrema vary from ensemble member to ensemble member, and the extrema of the maximum and minimum fields typically differ by a factor of 1.5. As with the rainfall fields, the areal extent of the region of positive CAPE varies significantly, particularly outside of the warm sectors of cyclones.

e. Secondary cyclogenesis

Seven cases of secondary type B cyclogenesis (Miller 1946) off the east coast of the United States are also found within the ensemble dataset. Secondary cyclogenesis events represent approximately 15% of the damaging coastal storms of the eastern United States (Mather et al. 1964) and present a significant forecast concern. In two of the cases, all the ensemble members fail to show any evidence of the developing coastal low. Of the remaining five cases, one is chosen to illustrate that the spread in cyclone location can be caused by changes in cyclone development (Fig. 7). The ensemble data for this case suggest two very different alternative scenarios at 48 h, and many combinations of these two scenarios. The first scenario is that the primary cyclone near Lake Erie is the only cyclone along the east coast (Fig. 7a). The second scenario is that the secondary cyclogenesis process is strong enough that there is only a weak indication of the parent cyclone over Lake Erie (Fig. 7b). The other scenarios show combinations of these two outcomes, with both cyclones clearly shown in the model output for five of the remaining ensemble members (Fig. 7c). The location of the center of the parent cyclone varies from northern Wisconsin through northern Michigan to Lake Erie, depending upon which forecasts are examined.

These examples illustrate that significant dispersion does occur within the short-range ensemble data, even though there is good evidence that this ensemble dataset is underdispersive using the measures of spread that have been examined. More evaluation is required to ascertain if there is any probabilistic information content in this spread that could be used to assist in the forecast process, although this likely will require a larger dataset. In-depth case studies of some of these events are warranted.
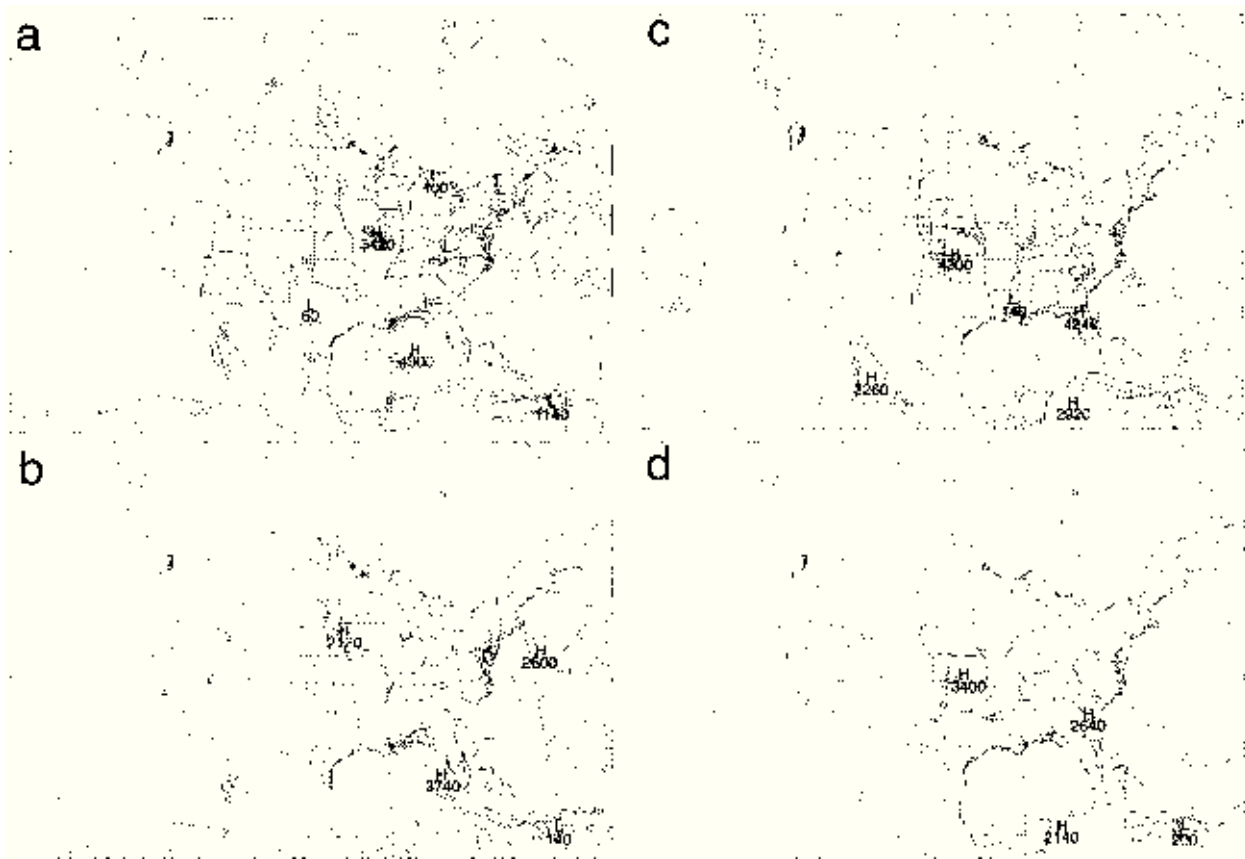
FIG. 6. Forecast 12-h (a) maximum CAPE (J kg^{-1}) and (b) minimum CAPE valid 0000 UTC 18 Jul 1996, and (c) maximum CAPE and (d) minimum CAPE valid 0000 UTC 8 May 1996. Values calculated at each individual grid point from all ensemble members.

5. Discussion

We have shown that the ensemble means of basic atmospheric variables, such as temperature, relative humidity, geopotential height, and wind speed and direction, are as accurate as the Meso Eta Model at mandatory pressure levels. The same conclusion is reached when examining the locations of model forecast cyclones over North America at 36 h using data from 33 cases. This indicates that the loss of accuracy from a reduction in model resolution can be recovered by using an ensemble approach. Furthermore, the BGM technique produces more spread in the ensemble than does using different analyses at both the 36-h and 48-h forecast times. However, the accuracy of the cyclone position forecasts from these two initialization techniques is nearly identical. The spread is increased significantly when both the RSM and Eta Model ensemble members are used together, suggesting that model differences are important contributors to ensemble spread. Yet our results indicate that for cyclone position there is little correlation between the spread and the error in the ensemble mean cyclone location, as also found by Hamill and Colucci (1998) for PQPF, suggesting that

this particular ensemble is unable to forecast the forecast skill.

Admittedly, this study and those of Hamill and Colucci (1997, 1998) have only touched upon the data from the 81 ensemble cases that are archived. More research needs to be done with this and other SREF datasets in order to evaluate the potential for ensembles to assist in the short-range forecast problem. Our results suggest that one of the most important issues to be addressed is how to create ensemble members that realistically sample the analysis error without sacrificing the accuracy of the ensemble information. The NCEP is presently testing the BGM technique (Toth and Kalnay 1993) within the context of RSM and eta-based higher-resolution regional model ensembles. Initial results are encouraging, largely predicated on the increased spread relative to the ensemble system used in this study (Tracton et al. 1998). But other techniques, such as using dynamic singular vectors (Buizza and Palmer 1995) or Monte Carlo techniques (Mullen and Baumhefner 1994; Du et al. 1997) also should be examined to determine which method, or combination of methods, provides the most skill and spread. Low-order models may provide
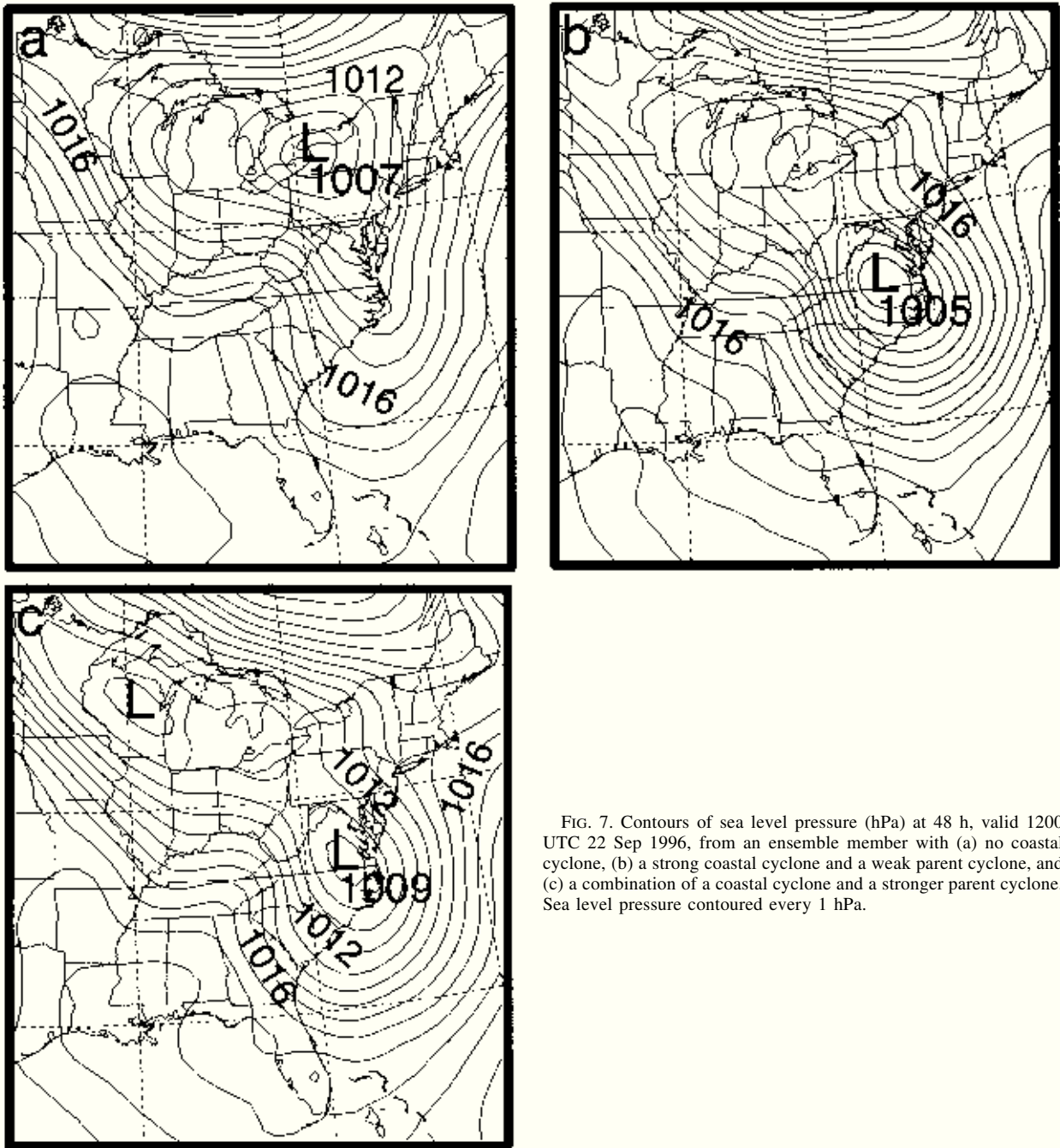
FIG. 7. Contours of sea level pressure (hPa) at 48 h, valid 1200 UTC 22 Sep 1996, from an ensemble member with (a) no coastal cyclone, (b) a strong coastal cyclone and a weak parent cyclone, and (c) a combination of a coastal cyclone and a stronger parent cyclone. Sea level pressure contoured every 1 hPa.

guidance on this difficult problem (Anderson 1996). In addition, as suggested by Stensrud and Fritsch (1994b) and supported by our results and those of Du et al. (1997), it is important to explore whether developing ensemble members with different models or model physical parameterization schemes can contribute to the development of an ensemble that samples the atmospheric probability density function.

REFERENCES

Anderson, J. L., 1996: Selection of initial conditions for ensemble forecasts in a simple perfect model framework. *J. Atmos. Sci.,* **53,** 22–36.

Black, T., 1994: The new NMC mesoscale Eta Model: Description and forecast examples. *Wea. Forecasting,* **9,** 265–278.

Bonner, W. D., 1989: NMC overview: Recent progress and future plans. *Wea. Forecasting,* **4,** 275–285.

Brankovic, C., T. N. Palmer, F. Molteni, and U. Cubasch, 1990: Extended-range predictions with the ECMWF models: Time lagged ensemble forecasting. *Quart. J. Roy. Meteor. Soc.,* **116,** 867–912.

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

Buizza, R., 1995: Optimal perturbation time evolution and sensitivity of ensemble prediction to perturbation amplitude. *Quart. J. Roy. Meteor. Soc.,* **121,** 1705–1738.

——, 1997: Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system. *Mon. Wea. Rev.,* **125,** 99–119.

——, and T. N. Palmer, 1995: The singular vector structure of the atmospheric general circulation. *J. Atmos. Sci.,* **52,** 1647–1681.

Dalcher, A., E. Kalnay, and R. Hoffman, 1988: Medium range lagged average forecasts. *Mon. Wea. Rev.,* **116,** 402–416.

Daley, R., and T. Mayer, 1986: Estimates of global analysis error from the global weather experiment observational network. *Mon. Wea. Rev.,* **114,** 1642–1653.

DiMego, G. J., and Coauthors, 1992: Changes to NMC's regional analysis and forecast system. *Wea. Forecasting,* **7,** 185–198.

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

Elliott, W. P., and D. J. Gaffen, 1991: On the utility of radiosonde humidity archives for climate studies. *Bull. Amer. Meteor. Soc.,* **72,** 1507–1520.

Epstein, E. S., 1969: Stochastic dynamic prediction. *Tellus,* **21,** 739–759.

Gyakum, J. R., 1983: On the evolution of the QE II storm. Part II: Dynamic and thermodynamic structure. *Mon. Wea. Rev.,* **111,** 1156–1173.

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

——, and ——, 1998: Evaluation of Eta-RSM ensemble probabilistic precipitation forecasts. *Mon. Wea. Rev.,* **126,** 711–724.

Hoecker, W. H., Jr., 1963: Three southerly low-level jet streams delineated by the Weather Bureau special pibal network of 1961. *Mon. Wea. Rev.,* **91,** 573–582.

Janjić, Z. I., 1994: The step-mountain Eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.,* **122,** 927–945.

Johnson, R. H., and P. J. Hamilton, 1988: The relationship of surface pressure features to the precipitation and air flow structure of an intense midlatitude squall line. *Mon. Wea. Rev.,* **116,** 1444–1472.

Juang, H.-M., and M. Kanamitsu, 1994: The NMC nested regional spectral model. *Mon. Wea. Rev.,* **122,** 3–26.

Junker, N. W., J. E. Hoke, and R. H. Grumm, 1989: Performance of NMC's regional models. *Wea. Forecasting,* **4,** 368–390.

Kain, J. S., and J. M. Fritsch, 1990: A one-dimensional entraining/detraining plume model and its application in convective parameterization. *J. Atmos. Sci.,* **47,** 2784–2802.

——, and ——, 1992: The role of the convective ''trigger function''

in numerical forecasts of mesoscale convective systems. *Meteor. Atmos. Phys.,* **49,** 93–106.

Kalnay, E., and A. Dalcher, 1987: Forecasting forecast skill. *Mon. Wea. Rev.,* **115,** 349–356.

Kogan, Y. L., and W. J. Martin, 1994: Parameterization of bulk condensation in numerical cloud models. *J. Atmos. Sci.,* **51,** 1728–1739.

Kuo, Y.-H., and S. Low-Nam, 1990: Prediction of nine explosive cyclones over the western Atlantic with a regional model. *Mon. Wea. Rev.,* **118,** 3–25.

Lacarra, J. F., and O. Talagrand, 1988: Short range evolution of small perturbations in a barotropic model. *Tellus,* **40A,** 81–95.

Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.,* **102,** 409–418.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.,* **20,** 131–140.

Mather, J. R., H. Adams, and G. A. Yoshikoa, 1964: Coastal storms of the eastern United States. *J. Appl. Meteor.,* **3,** 693–706.

Miller, J. E., 1946: Cyclogenesis in the Atlantic coastal region of the United States. *J. Meteor.,* **3,** 31–44.

Molteni, R., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.,* **122,** 73–119.

Mullen, S. L., and D. P. Baumhefner, 1988: Sensitivity to numerical simulations of explosive oceanic cyclogenesis to changes in physical parameterizations. *Mon. Wea. Rev.,* **116,** 2289–2329.

——, and ——, 1989: The impact of initial condition uncertainty on numerical simulations of large-scale explosive cyclogenesis. *Mon. Wea. Rev.,* **117,** 2800–2821.

——, and ——, 1994: Monte Carlo simulations of explosive cyclogenesis. *Mon. Wea. Rev.,* **122,** 1548–1567.

Murphy, A. H., 1977: The value of climatological, categorical, and probabilistic forecasts in the cost-loss ratio situation. *Mon. Wea. Rev.,* **105,** 803–816.

——, 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting,* **8,** 281–293.

Murphy, J. M., 1990: Assessment of the practical utility of extended-range ensemble forecasts. *Quart. J. Roy. Meteor. Soc.,* **116,** 89–125.

Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation system. *Mon. Wea. Rev.,* **120,** 1747–1763.

——, J. Purser, E. Rogers, and Y. Lin, 1996: The regional 3D-variational analysis for the eta model. Preprints, *11th Conf. on Numerical Weather Prediction,* Norfolk, VA, Amer. Meteor. Soc., 454–455.

Persson, A., 1996: Forecast error and inconsistency in medium range weather prediction. Preprints, *13th Conf. on Probability and Statistics,* San Francisco, CA, Amer. Meteor. Soc., 253–259.

Petroliagis, T., R. Buizza, A. Lanzinger, and T. N. Palmer, 1997: Potential use of the ECMWF ensemble prediction system in cases of extreme weather events. *Meteor. Appl.,* **4,** 69–84.

Reed, R. J., A. J. Simmons, M. D. Albright, and P. Unden, 1988: The role of latent heat release in explosive cyclogenesis: Three examples based on ECMWF operational forecasts. *Wea. Forecasting,* **3,** 217–229.

Rogers, E., D. G. Deaven, and G. J. DiMego, 1995: The regional analysis system for the operational ''early'' eta model: Original 80-km configuration and recent changes. *Wea. Forecasting,* **10,** 810–825.

——, T. L. Black, D. G. Deaven, and G. J. DiMego, 1996: Changes to operational ''early'' eta analysis forecast system at the National Centers for Environmental Prediction. *Wea. Forecasting,* **11,** 391–413.

Sanders, F., and E. P. Auciello, 1989: Skill in prediction of explosive cyclogenesis over the western North Atlantic Ocean, 1987–1988: A forecast checklist and NMC dynamical models. *Wea. Forecasting,* **4,** 157–172.

Schneider, R. S., 1990: Large-amplitude mesoscale wave disturbances

within the intense midwest extratropical cyclone of 15 December 1987. *Wea. Forecasting,* **5,** 533–558.

Stensrud, D. J., and J. M. Fritsch, 1994a: Mesoscale convective systems in weakly forced large-scale environments. Part II: Generation of a mesoscale initial condition. *Mon. Wea. Rev.,* **122,** 2068–2083.

——, and ——, 1994b: Mesoscale convective systems in weakly forced large-scale environments. Part III: Numerical simulations and implications for operational forecasting. *Mon. Wea. Rev.,* **122,** 2084–2104.

——, J.-W. Bao, and T. T. Warner, 1998: Ensemble forecasting of mesoscale convective systems. Preprints, *12th Conf. on Numerical Weather Prediction,* Phoenix, AZ, Amer. Meteor. Soc., 265–268.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.,* **74,** 2317–2330.

Tracton, S., and E. Kalnay, 1993: Ensemble forecasting at NMC: Operational implementation. *Wea. Forecasting,* **8,** 379–398.

——, J. Du, Z. Toth, and H. Juang, 1998: Short-range ensemble forecasting (SREF) at NCEP/EMC. Preprints, *12th Conf. on Numerical Weather Prediction.* Phoenix, AZ, Amer. Meteor. Soc., 269–272.

Tribbia, J. J., and D. P. Baumhefner, 1988: The reliability of improvements in deterministic short-range forecasts in the presence of initial state and modeling deficiencies. *Mon. Wea. Rev.,* **116,** 2276–2288.

Wade, C. G., 1994: An evaluation of problems affecting the measurement of low relative humidity on the United States radiosonde. *J. Atmos. Oceanic Technol.,* **11,** 687–700.

Wang, W., and N. L. Seaman, 1997: A comparison study of convective parameterization schemes in a mesoscale model. *Mon. Wea. Rev.,* **125,** 252–278.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction.* Academic Press, 467 pp.

Ziegler, C., and C. Hane, 1993: An observational study of the dryline. *Mon. Wea. Rev.,* **121,** 1134–1151.

Zou, X., and Y.-H. Kuo, 1996: Rainfall assimilation through an optimal control of initial and boundary conditions in a limited-area mesoscale model. *Mon. Wea. Rev.,* **124,** 2859–2882.