Ensemble Forecasting at NMC: The Generation of Perturbations

Zoltan Toth*,+ and
Eugenia Kalnay+

Abstract

On 7 December 1992, The National Meteorological Center (NMC) started operational ensemble forecasting. The ensemble forecast configuration implemented provides 14 independent forecasts every day verifying on days 1–10. In this paper we briefly review existing methods for creating perturbations for ensemble forecasting. We point out that a regular analysis cycle is a "breeding ground" for fast-growing modes. Based on this observation, we devise a simple and inexpensive method to generate growing modes of the atmosphere.

The new method, "breeding of growing modes," or BGM, consists of one additional, perturbed short-range forecast, introduced on top of the regular analysis in an analysis cycle. The difference between the control and perturbed six-hour (first guess) forecast is scaled back to the size of the initial perturbation and then reintroduced onto the new atmospheric analysis. Thus, the perturbation evolves along with the time-dependent analysis fields, ensuring that after a few days of cycling the perturbation field consists of a superposition of fast-growing modes corresponding to the contemporaneous atmosphere, akin to local Lyapunov vectors.

The breeding cycle has been designed to model how the growing errors are "bred" and maintained in a conventional analysis cycle through the successive use of short-range forecasts. The bred modes should thus offer a good estimate of possible growing error fields in the analysis. Results from extensive experiments indicate that ensembles of just two BGM forecasts achieve better results than much larger random Monte Carlo or lagged average forecast (LAF) ensembles. Therefore, the operational ensemble configuration at NMC is based on the BGM method to generate efficient initial perturbations.

The only two methods explicitly designed to generate perturbations that contain fast-growing modes corresponding to the evolving atmosphere are the BGM and the method of Lorenz, which is based on the singular modes of the linear tangent model. This method has been adopted operationally at The European Centre for Medium-Range Forecasts (ECMWF) for ensemble forecasting. Both the BGM and the ECMWF methods seem promising, but since it has not yet been possible to compare in detail their operational performance we limit ourselves to pointing out some of their similarities and differences.

1. Why operational ensemble forecasting?

On 7 December 1992, the National Meteorological Center replaced the single 10-day global medium-range forecast (MRF), which was run daily at 0000 UTC, by an ensemble of four 12-day forecasts, plus an extension to 12 days of the aviation 3-day forecast run at 1200 UTC (Tracton and Kalnay 1993). The operational configuration implemented at that time is such that there are 14 forecasts, originating from analyses within the most recent 48 hours, that verify over the same 10-day period. It replaces the previous configuration, where only one operational forecast and one experimental forecast were available for the 6–10-day forecast range. In order not to increase the total use of the CRAY YMP supercomputer, which is already saturated, a compromise had to be found, where the resolution of the MRF was reduced beyond day 6 from triangular truncation T126 (equivalent to a Gaussian grid resolution of 105 Km), to T62 (equivalent to 210 Km). It was found, however, that the reduction of resolution did not significantly affect the quality of the forecasts as long as it was performed after the first five days of the forecast (Tracton and Kalnay 1993).

The replacement of single operational forecasts by an ensemble of operational forecasts reflects explicitly the recognition that the atmosphere is a chaotic system. As pointed out by Lorenz (1963), even an infinitesimally small perturbation (as would be produced, for example, by the "wings of a butterfly") introduced into the state of the atmosphere at a given time will result in an increasingly large change in the evolution of the atmosphere with time, so that after about two or three weeks the trajectories of the perturbed and the original atmosphere would be completely different.

Lorenz's discovery led to the emergence of a new discipline, dynamical systems theory, and to the realization that many apparently deterministic systems, like the atmosphere and its numerical models, are also chaotic: arbitrarily small initial perturbations evolve into large differences with time. As far as real physical systems are concerned, their state can never be

*General Sciences Corporation, Laurel, Maryland
+National Meteorological Center, Washington, D.C.

measured exactly: for example, we know that our analyses of the atmosphere contain errors whose magnitude can only be estimated. The analysis errors are due to errors in measurements and in the first guess, lack of complete data coverage, and approximations in our analysis techniques. Even with a perfect model of the atmosphere, the skill of our forecasts would degrade to zero within a few weeks. However, if we are willing to run an ensemble of forecasts from slightly perturbed initial conditions, then averaging the ensemble can filter out some of the unpredictable components of the forecast, and the spread among the forecasts should provide some guidance on the reliability of the forecasts.

Whether these objectives can be fulfilled by ensemble forecasting depends on two conditions. First, the model that we are using should be realistic. In this paper we will assume that our numerical model is essentially perfect. This is not a serious approximation. As Reynolds et al. (1993) showed, the forecast error in the extratropics is dominated by the error originating from the unstable growth of initial errors, and not by model deficiencies. The second condition is that our ensemble at the initial time should represent the uncertainty in our analysis. In other words, we need to perturb the control analysis with perturbation fields that are representative of the errors present in the control analysis. And it is not only the magnitude of the error that is of importance but also the possible shape of the error field, which varies from day to day.

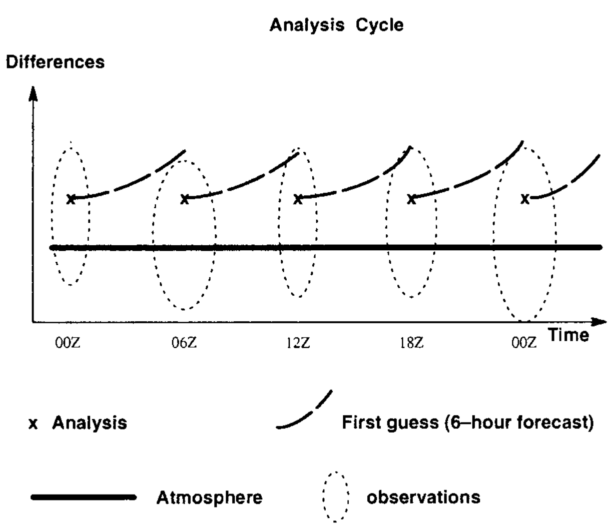

FIG. 1. Schematic of the 6-hour analysis cycle. Indicated on the vertical axis are differences between the true state of the atmosphere (or its observational measurements, burdened with random errors) and the analysis or forecasts of it. Note that the difference between a forecast and the true state of the atmosphere (or the observations) increases with time, due to the growing type of errors in the initial analysis.

Since an estimate of the analysis error is so crucial to the success of ensemble forecasting, we devote section 2 to a discussion of the characteristics of analysis cycle errors. We will see how fast-growing and well-organized errors, which are not random, are introduced and maintained in a conventional analysis cycle. In section 3 we give a brief overview of the methods that have been used so far to create ensemble perturbations, and will discuss to what extent they represent analysis errors. In section 4 we briefly discuss a method that has been designed to model the behavior of the growing errors in the analysis cycle. In this method, fast-growing perturbations are obtained in a "breeding cycle" of growing modes (Toth and Kalnay 1991a). Section 5 gives an overview of the results of ensemble forecasts based on the bred growing modes (BGM) method. A discussion of the results and of other applications for which the breeding method can also be used is given in section 6. A companion paper (Tracton and Kalnay 1993) includes further details of the new NMC ensemble operational configuration, including a discussion of applications.

2. Optimal ensemble perturbations: Growing errors in the analysis

As Epstein (1969) and Leith (1974) showed in their pioneering works, the ensemble mean should give a better forecast than the control forecast as long as the ensemble represents the uncertainty present in our control analysis. This is because the ensemble provides a special, nonlinear filtering that removes some of the growing errors. The question then is what kind of errors we have in our analyses. To answer this question we will take a quick look at how an atmospheric analysis is made.

In a typical 6-hour operational analysis cycle, a global model starts from initial conditions given by a previously completed atmospheric analysis and is integrated for a short (6 hour) forecast. The 6-hour forecast serves as a "first guess" for the next analysis, which is the statistical combination of the first guess with observations collected in a +/– 3-hour window centered at the time of validity of the forecast. This cycle is run four times a day, every day (Fig. 1). When analysis schemes were introduced into meteorology, climatology or persistence were also used as a first guess in the analysis (e.g., Gandin 1963) rather than a short-range forecast. The use of a model forecast as a first guess in the analysis cycle (also denoted four-dimensional data assimilation because of the time dimension introduced by the model) has resulted in major improvements in operational forecast skill in the last 10–15 years (e.g., Kalnay et al. 1990). The analyses (initial conditions) have much smaller overall errors and are far superior to those that could be obtained by using persistence or climatology as a first guess. This is because the model acts as a transporter of information from data-rich to data-poor regions, and therefore provides a good estimate of the actual state of the atmosphere even for those regions or parameters for which there are no observations (e.g., Charney et al. 1969).

The total error in the analysis has therefore been decreased by the introduction of a model first guess. However, the ratio of fast-growing errors to the total error must have increased. We know that the analysis is only a close approximation of the true state of the atmosphere. The errors in the analysis (analysis minus truth) are not known but they must contain both fast-growing, high-energy perturbations, such as baroclinically unstable modes, and slow-growing or nongrowing, low-energy perturbations, such as gravity waves. The random, nongrowing part of the error comes mainly from observational errors. Regarding the growing errors, they originate mainly from the forecasts. When a 6-hour forecast is run from an analysis that has both types of errors, the fast-growing, high-energy perturbations will grow faster, attain relatively large amplitudes, and dominate the error at the end of the forecast.

The use of the new data collected in the +/– 3-hour window will only reduce the size of the error (Fig. 1) but in general will not completely remove the fast-growing errors, which remain present and evolve and amplify again in consecutive analyses. Hence, a high ratio of fast-growing errors in the analysis is introduced and maintained through the successive use of short-range forecasts in the analysis cycle. For this reason, there is a "natural selection of fast-growing errors" in the analysis, and the error growth in the short-range forecasts is generally high. Note that the growing errors in short-range forecasts and hence in the analysis are associated with baroclinically unstable zones, highly dependent on the "flow of the day."

We argued above that there are random, nongrowing components and also well-organized, fast-growing components in the analysis error field. What are the implications for ensemble forecasting? First, one has to note that the growing type of errors are much more important than the nongrowing errors in determining the skill of any specific forecast. The nongrowing errors will project onto growing directions only at a later time, while the growing errors will amplify and dominate the total error field right from the beginning of the forecast. There is a second, more practical

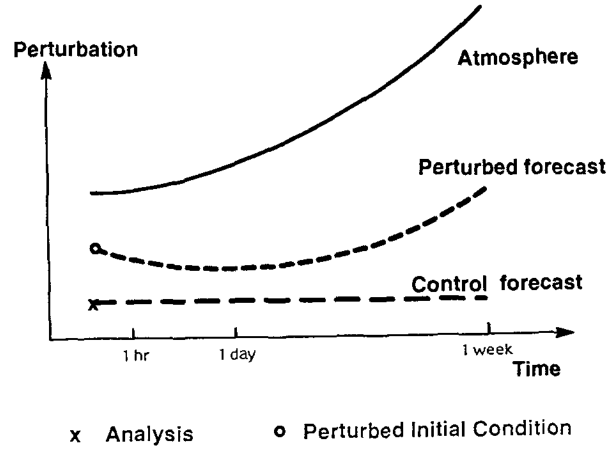

Fig. 2. Schematic illustrating the basic components of ensemble forecasting. Note that the control forecast, starting from a regular analysis, diverges from truth right from the beginning. If the ensemble perturbation is such that the perturbed forecast initially converges to the control, as is the case with a random perturbation indicated here, the goal of ensemble forecasting, that is, including the true evolution in a cloud of ensemble, cannot be achieved.

reason that also suggests that we should ignore the effect of random errors in the analysis. Stochastic, random errors have such a high number of degrees of freedom (dimension) that even if we wanted, we could not sample them properly with our limited number of perturbations. The conclusion is that we must perturb the initial condition along the growing type of errors present in the analysis.

At first sight, perturbing the control analysis with a fast-growing field would seem to go against the goal of improving numerical weather forecasts. To explain why this is necessary we will consider Fig. 2, which shows schematically all the components of an ensemble forecast: a control analysis and a numerical forecast initiating from it; a perturbation upon this analysis and the corresponding perturbed forecast; and finally the verification, that is, the true evolution of the atmosphere. The first thing to point out about the figure is that the analysis differs from the truth (which is never known exactly). Second, the control forecast departs from truth right from the beginning, and at a fast rate. The observed growth rate of the errors is on the order of 1.4–1.6 per day for the 6–72-hour forecast range (Zhao 1993; Reynolds et al. 1993), faster than the typical baroclinic growth. Such initial divergence is an empirical evidence of the existence of fast-growing errors in the analysis. The third point about the figure is that if we introduce a random perturbation upon the control analysis, then the two forecasts (control and perturbed) will initially converge. This is because arbitrary random perturbations in the grid domain project mainly on inertia-gravity waves and these perturbations are generally decaying (e.g., Lacarra and Talagrand 1988). Even if random perturbations are designed to be well balanced, they will grow very slowly or may even decay initially because it is very unlikely that a random perturbation would project onto the few very unstable modes in the high dimensional phase space of the atmosphere.

To rephrase the goal of ensemble forecasting, one can say that *the true evolution of the atmosphere should be a plausible member of an ensemble* (in other words, we must introduce perturbations similar to the actual error in the analysis). It is clear from the schematic Fig. 2 that this goal cannot be achieved with arbitrary, random perturbations. Randomly perturbed forecasts cannot capture the initial, fast divergence between the true atmospheric evolution and the control forecast. After a time of 2–3 days, the perturbed forecast will depart from the control, but by that time the real atmosphere has diverged considerably from both forecasts and we will never again have a chance to "catch up" with it.

In summary, ensemble perturbations must represent the growing errors in the analysis. If we accomplish this, we can also achieve one of the major goals of ensemble forecasting (Leith 1974), that is, to reduce the nonlinear error growth in the ensemble mean forecast. However, we have no *a priori* knowledge of the errors in the analysis. Consequently, we must estimate the growing errors in the analysis and then use these estimates as ensemble perturbations. But there is a caveat here. The growing type of perturbations should be related to actual errors in the analysis. This is because if we introduce a fast-growing perturbation on the analysis in a direction in which the analysis was perfect (had no initial error), then the average of the perturbations would give an inferior ensemble mean forecast, compared to the perfect control forecast, at least during the first days of integration. Similarly, random perturbations that project onto growing modes only at a later time may hurt the ensemble to the extent that these later growing modes are not connected to the initial uncertainty in the analysis.

3. Generation of ensemble perturbations: An overview

Ensemble forecasting has been long used in weather forecasting. In the 1950s analog forecasts were used to make extended- and long-range weather outlooks (see, e.g., Craddock et al. 1962). Here, of course, the size of the perturbations was not under control but rather depended on what nature had to offer during the archival period. Later, Lorenz (1965) and Epstein (1969) introduced the idea of ensemble forecasting in the context of numerical weather prediction. Leith (1974) further experimented with the idea of ensemble forecasting with random perturbations (Monte Carlo forecasting) in a perfect model environment. He showed that if the perturbations correctly describe the uncertainty in the analysis, ensemble forecasting, even with just a few members, has the potential of being quite useful. The major question, as we saw earlier, is how to generate ensemble perturbations that reflect the real initial uncertainty. Below, we will discuss different perturbation methods and, in particular, how they relate to the growing errors in the analysis.

The simplest way to generate perturbations is to use random (Monte Carlo) perturbations. These fields of perturbations can be generated in different ways. Random numbers, for example, can be added to gridpoint values (Tribbia and Baumhefner 1988), spherical coefficients, or empirical orthogonal functions (Schubert and Suarez 1989) of the control analysis. In recent years it has become apparent that using random perturbations is not the best way of making ensemble forecasts. This is because random perturbations (even if they are balanced and representative of the statistical variability of the atmosphere) require a few hours or even days before they organize into dynamically unstable modes that grow on the large-scale "flow of the day" as fast as forecast errors are observed to grow (Fig. 2).

Another method, lagged average forecasting (LAF) was suggested by Hoffman and Kalnay (1983). This scheme, which like the Monte Carlo method has also been widely used, takes advantage of operational forecasts launched before "today" as members of an ensemble. LAF perturbations are basically realistic short-term forecast errors, that is, difference fields between a forecast and an analysis (see schematic Fig. 3). As we argued in the previous section, the analysis contains both the growing and nongrowing type of errors when compared to "truth." By the end of a short-range forecast, the proportion of growing errors will, by definition, be enhanced in the difference between the forecast and our new estimate of truth, the subsequent analysis. This is the reason why LAF perturbations grow faster than Monte Carlo perturbations of the same magnitude. However, LAF ensemble forecasting has the disadvantage that earlier or "older" forecasts have much larger "perturbations" and hence are considerably less skillful than later or "younger" forecasts. This problem can be partly alleviated by either using different weights for different members of the ensemble (Hoffman and Kalnay 1983) or by scaling back the larger errors to a reasonable size (SLAF,[1] Ebisuzaki and Kalnay 1991).

To further increase the growing component in the perturbations, Kalnay and Toth (1991a) used the difference between short-range forecasts (SRFD), started at earlier times but verifying at the initial time of the ensemble (Fig. 3). Here, growing errors further dominate the difference between the forecasts, and, unlike LAF or SLAF, no new random errors are introduced by the latest analysis. Experiments performed by Toth and Kalnay (1991b) showed that there was a clear increase in the growth rate of perturbations from random perturbations to SLAF and from SLAF to SRFD, which was accompanied by an increase in the quality of the ensemble, measured by the skill of the mean of the ensemble forecasts. The faster the initial error growth, the better the perturbation. This is because, as we showed in the previous section, there are fast-growing errors in the analysis. Thus, unless we perturb along fast-growing modes, we have no chance of capturing a trajectory close to the true evolution of the atmosphere.

Lorenz (1965) showed that, in a linear sense, the fastest-growing perturbations for a given period of model integration can be obtained as the eigenmodes of the product A*A with the largest eigenvalues. Here $A(t_1,t_0)$ is the linear model propagator between a time t_0 and t_1, A* is its adjoint, and the eigenmodes of A*A are the singular modes of A optimized for the interval (t_1-t_0). Lorenz used the singular modes to determine the flow-dependent predictability in a simple model. Lacarra and Talagrand (1988), Farrell (1989), and Borges and Hartmann (1992) applied them to study atmospheric instability and short-range forecast errors. More recently, Molteni and Palmer (1992) and Mureau et al. (1992) used Lorenz's method to generate perturbations for ensemble forecasting at the European Centre for Medium-Range Weather Forecasts (ECMWF). The optimization time period chosen at ECMWF is the first 36 hours of the forecast period. Following this procedure they select initial perturbations that grow fastest, in a linear sense, during the first period of the forecast.

At NMC we have developed a new method that attempts to create realistic perturbations that could represent the errors actually present in the analysis cycle. The method, described in the next section, is denoted "breeding of growing modes" or BGM and has been designed to mimic how fast-growing errors are, inadvertently, "bred" in the analysis cycle.

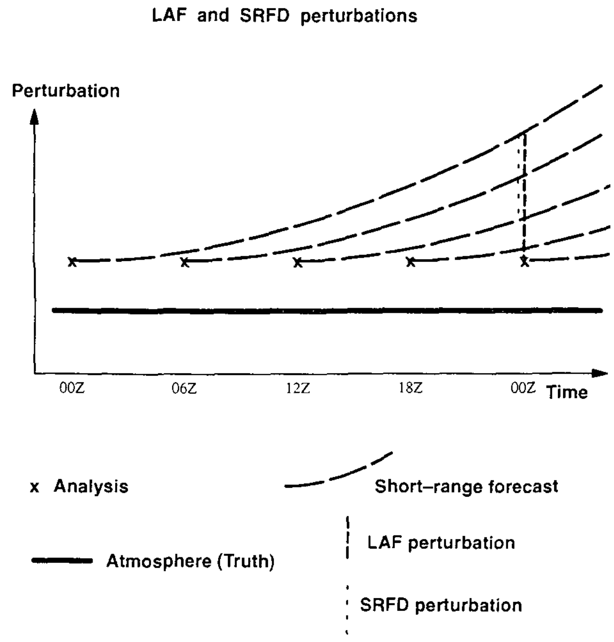

FIG. 3. Schematic of the creation of LAF SRFD perturbations. Note that the LAF perturbation includes not only the short-range forecast errors but also the random errors of the latest analysis, whereas the SRFD perturbation is not affected by the random errors of the latest analysis. This results in a significant reduction of the random errors and therefore in a higher growth rate for the SRFD perturbations.

4. The breeding of fast-growing perturbations

This simple method (Toth and Kalnay 1991a) consists of the following steps: (a) add a small arbitrary perturbation to the atmospheric analysis, (b) integrate the model for 6 hours from both the unperturbed (control) and the perturbed initial condition, (c) subtract the 6-hour control (analysis cycle) forecast from the perturbed forecast, and (d) scale down the difference field so that it has the same size (in an rms sense) as the initial perturbation. This perturbation is now added to the following 6-hour analysis, as in (a), and the process is repeated forward in time (Fig. 4). By construction, this method selects ("breeds") the modes that grow fastest during the cycle, similar to that which occurs in the analysis cycle itself (Fig. 1). Three or 4 days after starting a breeding cycle, the growth rate of the perturbations reaches a saturation value of around

[1]In scaled lagged average forecasting (SLAF), proposed recently by Ebisuzaki and Kalnay (1991), the LAF perturbations are divided by an "age" factor, resulting in similarly sized realistic perturbations. For example, a 12-hour forecast error perturbation could be divided by 2 and used as if it were a 6-hour forecast error. They also suggested that the perturbations could be both added and subtracted from the control (i.e., latest) analysis. In this way, it is possible to create a 17-member ensemble from just the latest 2 days of a 6-hour analysis cycle. Unlike MCF or LAF, this scheme resulted in average forecasts that verified as well or better than the control forecast (the forecast launched from the latest analysis) even after only 12 hours into the forecast.

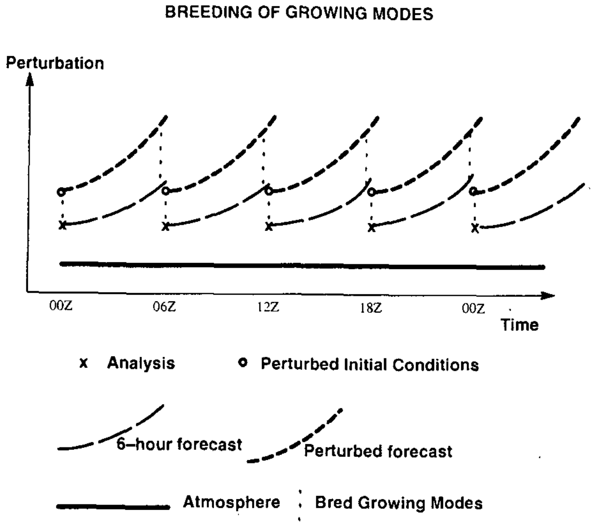

FIG. 4. Schematic of the 6-hour breeding cycle. Note that the breeding cycle depends on the analysis cycle but does not affect it. A small arbitrary perturbation is introduced on the control analysis initially. After a 6-h integration, the difference between the control and perturbed forecasts is scaled back to the size of the initial perturbation and this difference field is now added onto the new analysis. After 3–4 days of cycling, the perturbation is dominated by growing modes due to the "natural selection" of fast-growing perturbations.

1.5 per day, which is comparable to the actual short-range forecast error growth. From this point on, the growth rate does not depend on the initial perturbation. We found that the results are not very sensitive to the length of the integration in each cycle, at least for periods between 6 hours and 2 or 3 days.

Experiments performed with the simple three-variable Lorenz (1963) model show that the breeding method results in growing perturbations very similar to those obtained as eigenmodes of A*A (singular modes of A) whenever perturbations are growing fast. The growth rates obtained by both methods are almost identical. This is because the propagator $A(t_1,t_0)$ is in effect applied once during the interval t_0 to t_1, and, by definition, it is the singular modes (not the eigenmodes) of A that grow fastest during a single application of A. This agrees well with the original conclusions of Lorenz's (1965) study, in which he used a linear tangent version of his 28-variable model to study the evolution of random perturbations carried upon a time-dependent basic flow over a period of 64 days. He concluded that after a few days there will be a strong tendency for a randomly chosen perturbation to assume a shape similar to the fastest-growing singular mode(s). Note that the breeding cycle is like applying a nonlinear perturbation model along a time-evolving flow. The only difference from Lorenz's experiments is that since we use a nonlinear model, we have to scale down our perturbation regularly. Lorenz's arguments suggest that the perturbations in a breeding cycle are in effect linear combinations of the fastest-growing singular modes of the atmosphere. The modes are combined with random weights, which depend on how the initial perturbations project on the fastest modes and are proportional, in a probabilistic sense, to the growth rate of the singular modes.

A comparison of Figs. 1 and 4 suggests that the breeding cycle simulates the behavior of growing modes in the analysis. In both cases, growing errors get scaled down in each cycle. In the analysis cycle, observations are used to reduce the growing errors. In the breeding cycle, since the perturbations are run upon the analysis fields, this can be done by simple rescaling. Thus, the effect of stochastic errors coming from random observational errors in the analysis is eliminated from the breeding cycle. This argument suggests that with the breeding method we can generate fast-growing perturbations that are plausible growing analysis errors, which is the most important condition for effective ensemble forecasting. In fact, we can generate many such perturbations by running multiple breeding cycles, each started with a different initial perturbation. If we follow Lorenz's arguments above, the regional growing modes in each cycle are combinations of the fastest singular modes but with different random weights that depend on how the various initial perturbations project on the various singular modes.

We have subjectively inspected the bred perturbations from independent breeding cycles started with different initial perturbations and found that at any given time, different cycles share roughly half of the well-defined regional modes, although their signs can be either positive or negative. This subjective estimation agrees well with the results obtained by Houtekamer and Derome (1993). In their multiple breeding experiments with a quasigeostrophic model, they found that after 20 days of breeding, 48% of the variance in the different perturbations can be expressed by the first empirical orthogonal function. In the areas where the independent breeding modes differ, the growth rate of the leading singular modes must be comparable.

Lorenz (1965) also noted that there are cases when the fastest-growing singular mode has a growth rate much larger than that of the second one. He went further to assert that "even though the initial error field may be completely unknown, the general configuration of the error field after eight days can be reasonably well estimated, although the sign will of course be in doubt." And indeed we find many cases, especially those associated with fast-developing tropical or extratropical cyclones, in which the bred growing modes reproduce remarkably well the actual errors of the operational forecast at various (12–144 h) lead times. This is another indication that the breeding modes offer a good representation of possible analysis errors.

This discussion suggests that a good estimation of the growing modes in the analysis cycle could be used not only to improve ensemble forecasting, but also to improve the analysis itself. In the analysis schemes currently operational [optimal interpolation, e.g., Lorenc (1982), or statistical spectral interpolation, Parrish and Derber (1992)] it is assumed that the errors in the first guess are geostrophically balanced but are estimated from the time average of many short-range forecast errors and therefore are random. As we saw in section 2, however, the errors in the first guess are not random but are dominated by instabilities associated with the "flow of the day." Therefore these nonrandom, growing modes are an important component of the analysis error that is "ready to grow." Information on these growing modes (their shape and size) could be used to improve the operational analysis. The forecast error covariance "of the day" could be estimated from multiple breeding cycles (D. Parrish 1992, personal communication).

It is worth noting that multiple breeding modes for ensemble forecasting can be generated virtually free of any cost. If we run medium- or extended-range forecasts, each ensemble member can maintain its own independent breeding cycle. The breeding perturbations can be defined as the difference between a member of the ensemble and the control forecast at 1-day lead time (see Fig. 5). This difference is then used as the initial perturbation for the next-day ensemble member, etc. In this configuration, efficient perturbations are generated without any extra computational cost (beyond running the forecasts themselves).

We close this section by pointing out some differences between the two methods that explicitly attempt to create fast-growing ensemble perturbations, that is, the BGM method and the Lorenz method as implemented at ECMWF (Palmer et al. 1992; Molteni and Palmer 1992; Mureau et al. 1992; Buizza 1992). As mentioned in section 3, at ECMWF, three times a week the largest singular modes of a T21, 19-level model are determined for the first 36-hour period of the forecast, in a linear framework. The largest 20 or so singular modes are then combined to serve as final perturbations.

We have suggested that, in agreement with Lorenz (1965), it seems reasonable to assume that the BGM method results in perturbations that also project strongly onto a subset of fast-growing singular modes. What are the differences then between the two methods? The ECMWF method computes the growing modes explicitly for a 36-hour optimization time period starting at the analysis time, while the BGM method estimates the modes that were growing fastest during the interval leading to the initial time. Therefore, the ECMWF method may be optimal in estimating the maximum range of possible forecast errors, while the breeding method may give a better estimate of the actual errors in the initial analysis. The ECMWF method finds individual modes and then combines them to produce perturbation fields, while the BGM method results in complete fields that are a superposition of regional modes. The ECMWF method uses a reduced-resolution linear tangent version of the model, with limited physics, whereas the BGM method uses the nonlinear model at its full resolution with complete physics to compute the perturbations. Therefore, nonlinear saturation of very rapidly growing but low-energy modes (e.g., convection, see Fig. 6) can take place in the breeding cycle (as it does in the analysis cycle) but not in the Lorenz method. Moreover, since the modes obtained by breeding are already balanced, there is no need to introduce nonlinear normal-mode initialization in the BGM method, which has a significant effect in the linear adjoint algorithm (Buizza 1992). Another difference is that the BGM method is essentially cost free, while the Lorenz method requires significant computational resources. Finally, there is no forced time continuity in the ECMWF

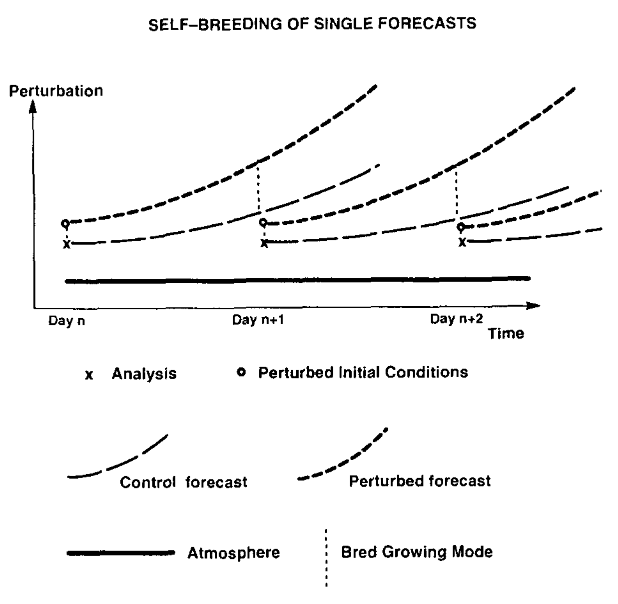

FIG. 5. Schematic showing how an ensemble of perturbed extended-range forecasts can maintain their own breeding cycle. For each ensemble member, an arbitrary perturbation is introduced onto the initial control analysis. The difference between the perturbed and control forecasts at 1-day lead time serves as a new perturbation on the following day. After 3–4 days, the perturbations are dominated by fast-growing modes, just as in the breeding cycle.

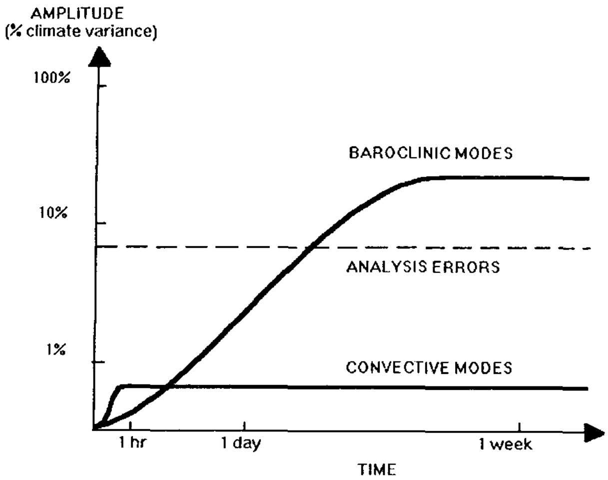

FIG. 6. Schematic of the time evolution of the rms amplitude of high-energy baroclinic modes and low-energy convective modes. Note that although initially growing much faster than the baroclinic modes, convective modes saturate at a substantially lower level. These modes are therefore insignificant for the analysis/ensemble perturbation problem since the errors in the control analysis (dashed line) are much larger than the convective saturation level.

ensemble forecasts in the sense that the selection of today's ensemble members is unrelated to that of yesterday's, so that the order, and even the structure, of the modes may be quite different. The breeding method, on the other hand, is constrained to maintaining considerable consistency in the ensemble forecasts from one day to the next (which may or may not be an advantage).

At this point in time, both centers have had limited operational experience with the two methods, both of which are promising. A detailed comparison of their performance should come in future studies, when enough data and experience has been accumulated at both centers. In the next section we will compare the skill of bred ensembles to that of ensembles generated by other methods.

5. Results obtained with BGM ensemble forecasts

In this section we will evaluate experimental and operational ensemble forecasts with bred perturbations. These ensemble forecasts will also be compared to ensembles generated by other methods. As further discussed in Tracton and Kalnay (1993), ensemble forecasting has three major objectives: (a) improve the skill, by reducing the nonlinear error growth and averaging out unpredictable components, (b) predict the skill, by relating it to the agreement among ensemble forecast members, and (c) provide an objective basis for casting forecasts in a probabilistic form. The latter two objectives, which involve higher moments, are hard to verify without large numbers of forecasts. However, the impact of ensemble prediction on forecast skill can be easily verified and can be used to discriminate among different methods for generating perturbations.

Since we have some evidence that the growing modes obtained by the breeding method are a good representation of the "errors of the day" present in the analysis cycle, we have followed a minimalistic ensemble forecast approach: we generated two-member ensembles using the bred modes as perturbations. The ensemble contains two initial perturbations, obtained by adding to and subtracting from the analysis our growing mode estimate for the same day.

Five-day forecasts were run with the perturbed analyses, and the skill of the mean of the two perturbed forecasts was compared with the skill of the control forecast and also with the skill of "benchmark" ensembles of much larger size generated by other methods (random MCF and SLAF). The size of the initial perturbation was kept constant throughout these experiments at a level of 10% of the climatological variance. All the experiments described below were performed with the T62/18-level version of the NMC global model (Kanamitsu et al. 1991).

The results of experiments performed over several months show, first, that the 5-day forecast skill of the mean of the twin ensemble based on the bred growing modes is higher than the skill of the control in about 80% of the forecasts for both the Northern and the Southern hemispheres. Experiments completed for February and March 1992 show that the average improvements in the anomaly correlation for the 500-hPa streamfunction filed were 2% and 3% for the Northern and Southern hemispheres, respectively. Moreover, the ensemble mean verifies equally or better than the control, even for the short range (1–3 days lead time).

These early results have been confirmed by those obtained since the operational implementation of ensemble forecasting on 7 December 1992. Table 1 shows that the forecasts started at high (T126) and lower resolution (T62) have comparable skill in the 6–10-day range. The skill of the mean of the twin BGM lower-resolution forecasts is more than three percentage points higher in anomaly correlations than either of the control forecasts. It is important to note that running twin BGM forecasts requires twice as many computational resources as running a single lower-resolution forecast, while doubling the horizontal resolution requires 8–10 times more computational time. In this sense, running twin forecasts is a very cost-effective way of enhancing the skill. Moreover, the BGM forecast has substantially more skill than the

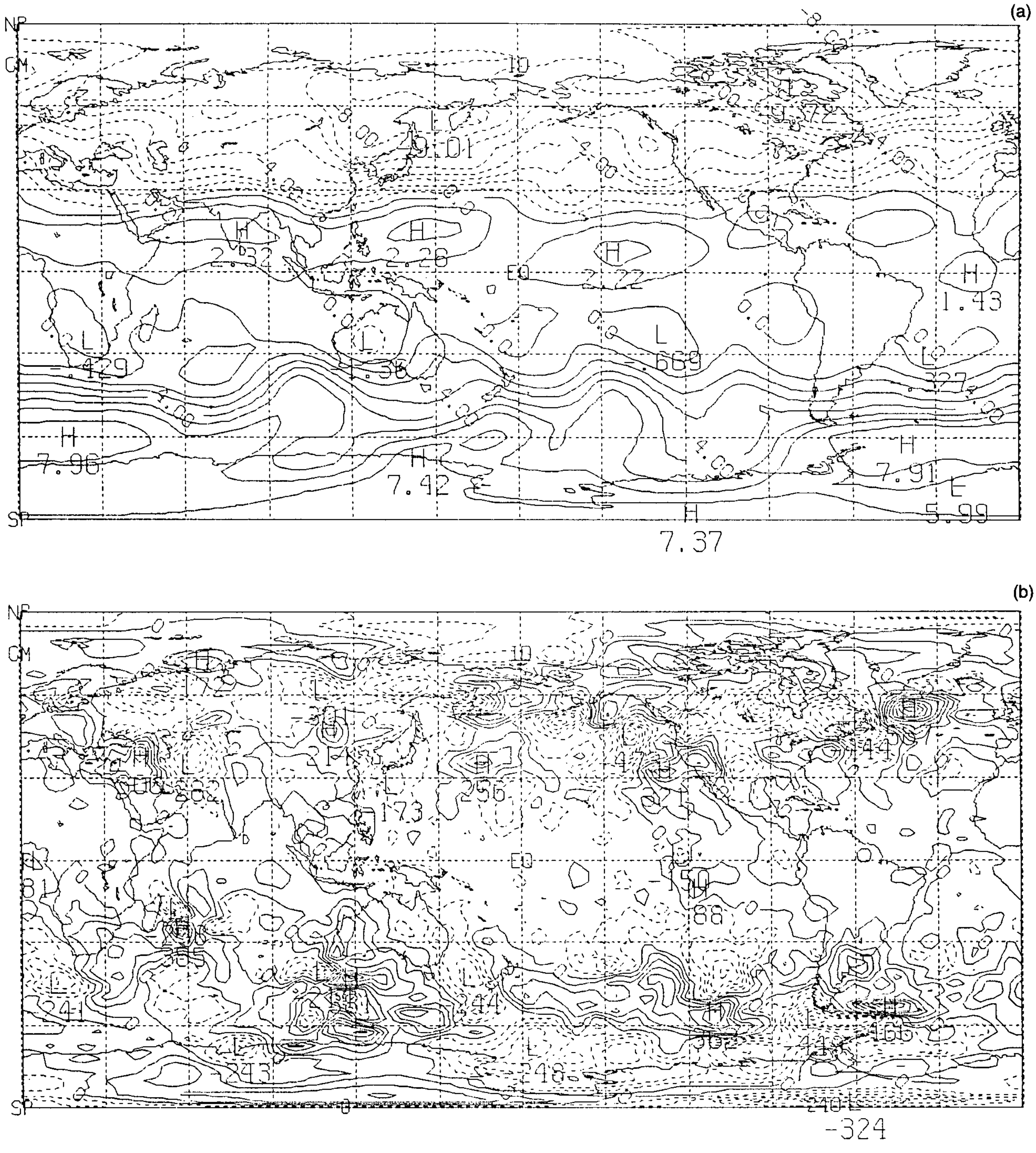

FIG. 7. (a) Analysis corresponding to the initial conditions of 15 February 1992. (b) Bred growing modes for 15 February 1992. Shown are the streamfunction values at sigma layer 9 (around the 500-hPa height level; displayed values are in 10^7 m^2 s^{-1}).

three-member LAF forecasts, which include "older" forecasts with higher error levels.[2]

Second, the skill of the twin ensemble is comparable to or even better than the skill of much larger ensembles using previous methodologies to obtain perturbations. In Table 2, we show the anomaly correlation for a set of control 5-day forecasts and for the mean of the different ensemble forecasts. Both the SLAF and random ensembles contain eight pairs, each of which is generated by adding and subtracting a different perturbation, whereas a single pair is used

[2]Note that the scaled LAF method (SLAF) would not suffer from this handicap but would require additional model integrations.

Table 1. Comparison of the average 6–10-day 500-hPa geopotential height forecast skill (Northern Hemisphere extratropics) obtained using subsets of the operational ensemble forecasts. The verification period is 13 December 1992 through 14 March 1993. The letters AC represents percentage anomaly correlation; HR represents forecasts with the high-resolution T126 analysis and forecast model for the first 6 days, after which the integration is extended to 12 days at T62 resolution. The letters LR represent forecasts with the lower-resolution T62 analysis and forecast model throughout the 12 days of integration. CTL represents control forecasts started from unperturbed 0000 UTC analysis. BGM represents the mean of the twin forecasts started from control analyses perturbed with positive and negative bred growing modes from the breeding cycle. LAF are three-member lagged average forecasts obtained from the latest three 0000 UTC analyses, using uniform weights. The last entry is the average of all the four forecasts performed using the latest 0000 UTC analyses.

Forecast type	Number of forecasts	AC
CTL(HR)	1	55.3
CTL(LR)	1	55.4
BGM(LR)	2	58.7
LAF(HR)	3	53.9
LAF(LR)	3	54.0
CTL(HR)+CTL(LR) +2BGM(LR)	4	60.4

with the BGM. In SLAF, the perturbations are short-range forecast errors (Ebisuzaki and Kalnay 1991), while the random perturbations are combinations of randomly chosen earlier analysis fields (Kalnay and Toth 1991b). The scores, computed and averaged for six independent cases (3 individual days for both the Northern and Southern hemisphere extratropics), show that with only two members the BGM forecasts achieve the same improvement over the control as the SLAF method with 16 forecasts and, in turn, that the SLAF is considerably better than what could be achieved with random perturbations. This latter result is not unexpected, since the SLAF perturbations, like LAF perturbations, are based on short-term forecast errors and, as discussed in section 3, are more likely to contain realistic growing "errors of the day" than random Monte Carlo perturbations.

In order to have a clearer estimate of the difference between a single pair of BGM and a random ensemble, we performed another set of forecasts for 26 February 1992. We ran two independent breeding cycles, started weeks earlier from different random initial perturbations. As indicated before, such independent breeding cycles after a few days end up sharing about half of the growing modes and have different modes elsewhere. We ran 5-day forecasts

Table 2. Verification of the mean of different ensemble forecasts generated by various methods. The anomaly correlation values of 5-day, 500-hPa streamfunction forecasts are averaged for 8, 17, and 22 January 1992 and over the Northern and Southern hemisphere extratropics.

Ensemble type	Control	BGM twin	SLAF	MCF
Number of forecasts	1	2	16	16
Anomaly correlation	65.7	67.3	67.3	66.5

using both breeding cycles and averaged the score of each twin pair. For comparison, we ran 18 twin pairs of random perturbations and averaged the score of each individual pair. We also computed the score for the mean of all 36 random Monte Carlo (MC)forecasts.

As Table 3 shows, a single bred growing mode twin forecast achieves as much improvement over the control forecast as the 18-times larger ensemble of 36 random MC forecasts. Moreover, the average improvement from single random MC twin forecasts is only about one-fourth of that obtained with BGM twins.

To illustrate the typical shapes of a field of growing modes in the BGM scheme and the result of twin ensemble forecasting, we show in Figs. 7–9 an example of a successful BGM forecast. The initial conditions for this case, 0000 UTC 15 February 1992, and the bred growing modes obtained for the same day are shown in Figs. 7a and 7b, respectively. The results are presented in terms of the streamfunction at the midlevel of the atmosphere, which is similar to the 500-hPa geopotential field in the extratropics but resolves the tropical circulation better.

Note that the growing modes are associated with specific features of the atmospheric circulation on that day. For example, a dipole pattern of modes near the

Table 3. Same as Table 2, except that the bred growing mode BGM twin includes the average of the scores of two pairs of independent bred growing mode twin ensembles, while MCF twin gives the average of the scores of 18 randomly perturbed twin forecasts. The MCF gives the score of the mean forecast obtained using all 36 members of the MC ensemble. All results are for 26 February 1992, with the NH and SH scores combined.

Ensemble type	Control	BGM twin	MCF twin	MCF
Number of forecasts	1	2	2	36
Anomaly correlation	72.0	74.5	72.8	74.4

Caspian Sea in Fig. 7b is associated with a traveling low in Fig. 7a in the same area, or the train of short waves off the west coast of North America is accompanied by a cutoff low.

Figure 8a shows the verification analysis corresponding to 0000 UTC 20 February 1992, Fig. 8b the 5-day control (single) forecast, and Fig. 8c the mean of the BGM twin ensemble. Several streamlines have been highlighted in areas where the control and the BGM twin average forecasts differ substantially in both the Northern and the Southern hemispheres. In most areas of disagreement (but not in all) the twin BGM average forecast is closer to the analysis than the control forecast. This subjective statement is substantiated by the objective scores (anomaly correlation), indicated in Table 4. It should be noted that the improved forecast verification is not a result of general smoothing of small-scale features, as is the case at longer lead times and especially when a large number of ensemble members are averaged (or when forecasts are time averaged).

Finally, we note that when we verified over a large number of cases, each member of the BGM ensembles against the corresponding analysis, we found that in 67% of the forecasts at least one member of the twin growing mode ensemble pair was better than the control 5-day forecast, whereas this ratio was only 17% for random Monte Carlo perturbations over the Northern Hemisphere. To further illustrate this point, we show in Fig. 9 the negatively and positively perturbed 5-day forecasts for the above case of 20 February 1992. The corresponding skill scores are listed in Table 4. The high ratio of ensemble members that are superior to the control forecast may be considered surprising but actually could be expected from the results of Lorenz (1965). This is again an indication that our growing mode estimates do project onto the actual error field of the control analysis and thus provide good initial perturbations. It also suggests again that it should be possible to take advantage of the BGM cycle to improve the analysis.

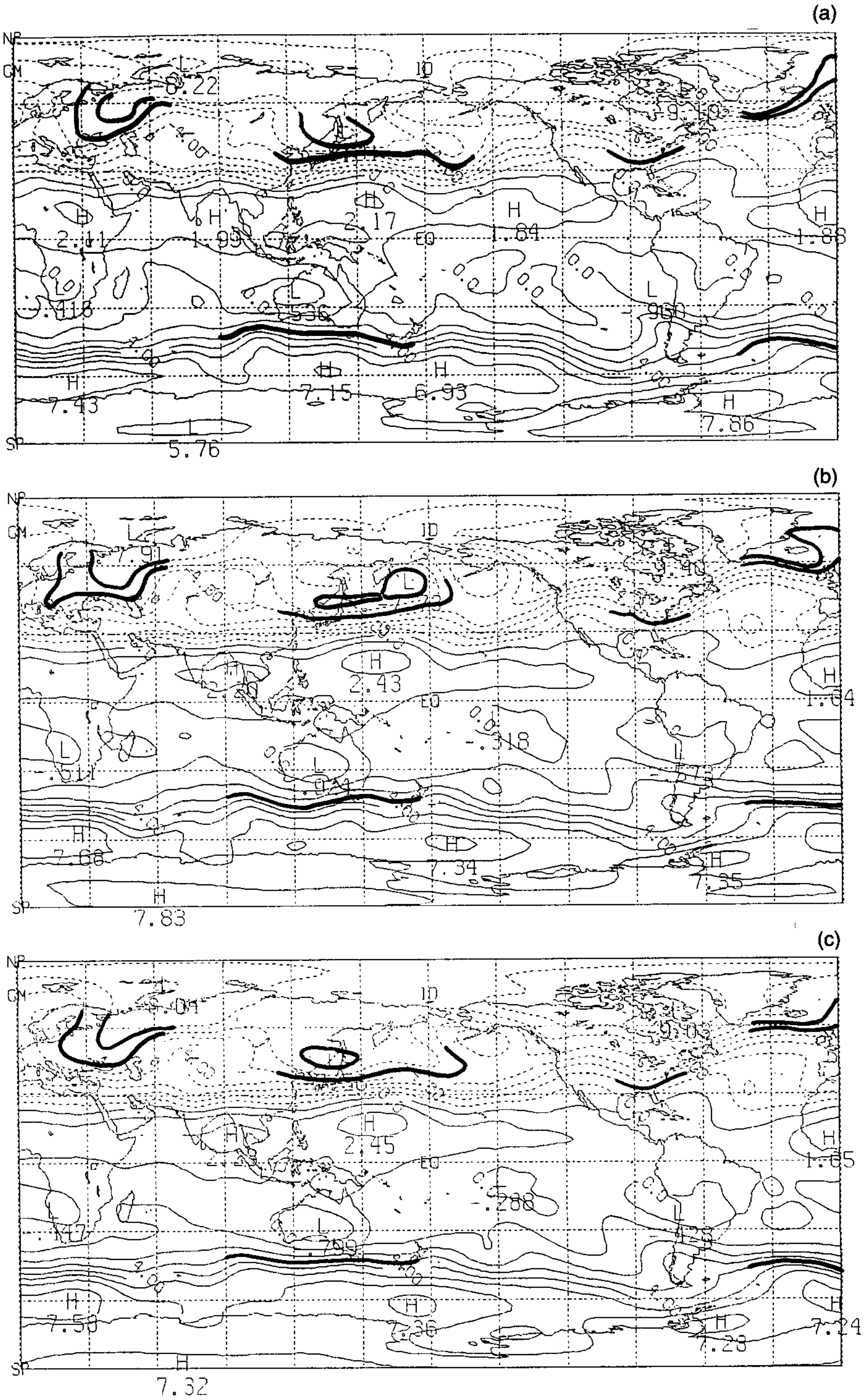

FIG. 8. (a) Verifying analysis for 20 February 1992. (b) Five-day control forecast. (c) Five-day mean forecast from the BGM twin forecasts. For further details, see Fig. 7.

Table 4. Anomaly correlation values for 5-day, 500-hPa streamfunction forecasts verified on 20 February 1992, over the Northern and Southern hemisphere extratropics separately. Next to the score for the control forecast are the scores for the two individual forecasts perturbed by adding and subtracting the BGM modes and also for the mean of the two perturbed forecasts (BGM twin).

Forecast		BGM negative perturbation	Control	BGM positive perturbation	Mean of negative and positive
Anomaly	NH	59	58	68	70
Correlation	SH	71	66	64	70

6. Discussion

In this paper we have briefly reviewed and compared different ensemble forecasting techniques. As Epstein (1969) and Leith (1974) showed, ensemble forecasting can be successful if the perturbations sample the initial uncertainty in the analysis. We have argued that, in addition to random errors that are traditionally assumed to be present in the analysis, there is a substantial component of growing errors. The growing errors are connected to instabilities of the varying flow and they get enhanced and maintained through the successive use of short-range forecasts as first guesses (Fig. 1). It follows then that ensemble perturbations must sample the growing errors in the analysis.

For this purpose we have developed a simple method, denoted breeding of growing modes (BGM). This method can be considered equivalent to running a nonlinear perturbation model with regular rescaling of the perturbation size to keep the perturbation amplitude limited. This effect is achieved by running the full nonlinear model from both the control and a perturbed analysis in a "breeding cycle" that follows the trajectory of the atmosphere through its analyzed fields (Fig. 4). The breeding method itself is a model of how the growing errors, which are highly dependent on the details of the varying large-scale circulation, are maintained or "bred" in a conventional analysis cycle (cf. Figs. 1 and 4). Consequently, the bred perturbations should offer a good representation of possible growing errors in the analysis. In fact, we argued that other, random or nongrowing perturbations should not be included in ensemble perturbations.

The BGM perturbations have been used in the ensemble forecasting scheme that became operational at NMC on 7 December 1992. The details of the ensemble configuration, along with possible applications, are presented in Tracton and Kalnay (1993). Both experimental and operational results show that even small ensembles generated by the breeding method can yield considerable improvement upon the

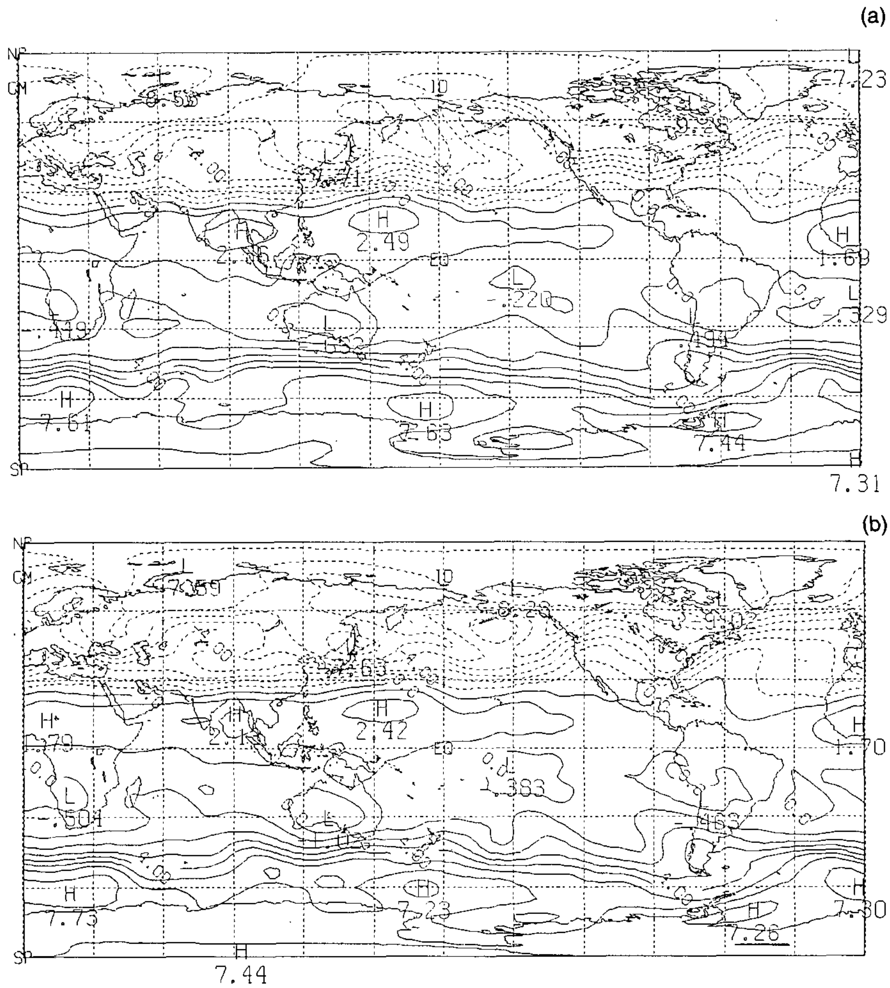

Fig. 9. (a) Five-day forecast from the BGM positive perturbation. (b) Five-day forecast from the BGM negative perturbation. For further details, see Fig. 7

control forecast. To achieve comparable gains one would have to use much larger ensembles of the less effective random Monte Carlo or lagged average forecast techniques.

Considering the large amount of similarity between long-maintained perturbation fields carried along a time-dependent flow and the singular modes of a linear system in Lorenz's (1965) experiments, it seems likely that the growing perturbations in the breeding cycle are superpositions of regional growing modes, each of which is a combination of the fastest-growing singular modes of the full nonlinear system.[3]

At ECMWF, where ensemble forecasting is also operational, Lorenz's linear adjoint algorithm is used to compute directly the singular modes of a linear model and then use them as ensemble perturbations. Since a comparative verification of the two operational systems is not yet available, we can indicate only a few differences between the two methods. First, the ECMWF method explicitly determines the fastest-growing modes for the short-range forecast period while the breeding method estimates the fastest-growing modes during the data assimilation period. Consequently, the breeding method may have a better chance at estimating analysis errors while the ECMWF approach may be better at detecting the possibility of extreme forecast failures. Second, at ECMWF a T21 horizontal resolution linear tangent model and its adjoint, with limited physical parameterizations, is used to compute the growing modes, while at NMC the full-resolution nonlinear model is used for the same purpose. A third difference is that while the breeding method is basically cost free (because medium- or extended-range ensemble forecasts can maintain their own breeding cycles), the method used at ECMWF requires additional computational resources.

It was suggested that the problem of ensemble forecasting is closely related to that of atmospheric analysis, and research in one area could be beneficial to the other. Preliminary analysis experiments with Lorenz's three-variable model and also with the full NMC global system indicate that there are at least two possible ways to use the bred modes to improve the analysis. First, the growing modes can be used to compute the forecast error covariance, a task that is normally very difficult because it depends on the availability of sparse rawinsonde observations, burdened with random errors, to estimate very short-range forecast errors. Since the growing modes are complete three-dimensional fields that describe the shape of the fastest-growing errors that appear in the short-range forecasts, their outer product can be computed without difficulty, and should provide an efficient and accurate parameterization for improving the forecast error covariance used at the next data time (D. Parrish 1992, personal communication). Second, it may be possible to improve the analysis by locally modifying the first guess toward the observations along the direction of the growing modes.

In addition to analysis and ensemble forecasting, the BGM method could also be useful for other applications such as forecast of the skill and stability studies of the observed or modeled atmosphere. The most unstable, fastest-growing modes can be estimated by breeding, either for a sequence of analyses, representing the evolution of the atmosphere, or for a long model run. The growth rate, statistical and dynamical properties, and structure of the most unstable perturbations can be derived from these fields. For example, an examination of the zonally averaged vertical cross section of the growing modes clearly shows that they are able to penetrate the winter stratosphere but not the summer stratosphere, as might be expected from the Charney–Drazin theory.

[3]A. Trevisan and J. Ahlquist (1993, personal communication) have pointed out that, by construction, the bred growing modes are akin to local Lyapunov vectors.

Acknowledgments. We are most grateful to M. Kanamitsu, H. Juang, M. Iredell, and S. Saha for their essential help with both the research and computational aspects of our work and with its operational implementation. We also had many stimulating discussions with D. Parrish, J. Derber, J. Purser, H. van den Dool, W. Ebisuzaki, T. Palmer, J. Ahlquist, J. Anderson, A. Trevisan, and A. Navarra. J. Irwin and other members of the Automation Division were extremely helpful in the operational implementation of the breeding cycle. S. Tracton provided the data for Table 1. The encouragement and support of R. McPherson are much appreciated. The comments of two anonymous reviewers helped to improve the presentation of the manuscript.

References

Borges, M. D., and D. L. Hartmann, 1992: Barotropic instability and optimal perturbations of observed nonzonal flows. *J. Atmos. Sci.*, **49,** 335–354.

Buizza, R., 1992: Unstable perturbations computed using the adjoint technique. Research Department Tech. Memo. No. 189, ECMWF, Reading, England.

Charney, J. G., M. Halem, and R. Jastrow, 1969: Use of incomplete historic data to infer the present state of the atmosphere. *J. Atmos. Sci.* **26,** 1160–1163.

Craddock, J. M., H. Flohn, and J. Namias, 1962: The present status of long-range forecasting in the world. WMO-No. 126. TP. 56, Geneva, 23 pp. [Available from WMO, C.P. No. 2300, CH1211, Geneva, Switzerland.]

Ebisuzaki, W., and E. Kalnay, 1991: Ensemble experiments with a new lagged analysis forecasting scheme. Research Activities in Atmospheric and Oceanic Modelling, Report No. 15, WMO. [Available from WMO, C.P. No. 2300, CH1211, Geneva, Switzerland.]

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

Farrell, B. F., 1989: Optimal excitation of baroclinic waves. *J. Atmos. Sci.* **46,** 1193–1206.

Gandin, L. S., 1963: Objective analysis of meteorological fields. Leningrad, Gidrometeoizdat, in Russian. (English translation, 1965, Israel Program for Scientific Translations, Jerusalem, 242 pp.)

Hoffmann, R. N., and E. Kalnay, 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus,* **35A,** 100–118.

Houtekamer, P., and J. Derome, 1993: Prediction experiments with two-member ensembles. *Mon. Wea. Rev.*, submitted.

Kalnay, E., and Z. Toth, 1991a: Estimation of the growing modes from short range forecast errors. Research Highlights of the NMC Development Division: 1990–1991, 160–165. [Available from the National Meteorological Center, Washington D.C. 20233.]

——, and ——, 1991b: Efficient selection of Monte Carlo forecast ensemble members. Research Highlights of the NMC Development Division: 1990–1991, 157–159. [Available from the National Meteorological Center, Washington D.C. 20233.]

——, M. Kanamitsu and W. E. Baker, 1990: Global numerical weather prediction at the National Meteorological Center. *Bull. Amer. Meteor. Soc.,* **71,** 1410–1428.

Kanamitsu, M., J. C. Alpert, K. A. Campana, P. M. Caplan, D. G. Deaven, M. Iredell, B. Katz, H.-L. Pan, J. Sela, and G. H. White, 1991: Recent changes in the Global Forecast System at NMC. *Wea. Forecasting,* **6,** 425–435.

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.

Lorenc, A., 1982: A global three dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.,* **101,** 701–721.

Lorenz, E. N., 1963: Deterministic non-periodic flow. *J. Atmos. Sci.,* **20,** 130–141.

——, 1965: A study of the predictability of a 28-variable atmospheric model. *Tellus,* **17,** 321–333.

Molteni, F., and T. N. Palmer, 1992: Predictability and finite-time instability of the northern winter circulation. *Quart. J. Roy. Meteor. Soc,* **119,** 269–298.

Mureau, R., F. Molteni, and T. N. Palmer, 1992: Ensemble prediction using dynamically conditioned perturbations. *Quart. J. Roy. Meteor. Soc,* **119,** 299–323.

Palmer, T., F. Molteni, R. Mureau, R. Buizza, P. Chapelet, and J. Tribbia, 1992: Ensemble prediction. ECMWF Research Department Tech. Memo. No 188. {Available from ECMWF, Reading, England.]

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

Reynolds, C., P. J. Webster, and E. Kalnay, 1993: Random error growth in the NMC Medium Range Forecasting model. *Mon. Wea. Rev.*, submitted.

Schubert, S. D., and M. Suarez, 1989: Dynamical predictability in a simple general circulation model: Average error growth. *J. Atmos. Sci.,* **46,** 353–370.

Toth, Z., and E. Kalnay, 1991a: Estimating the growing modes of the atmosphere: The breeding method. Research Highlights of the NMC Development Division: 1990–1991, 439–443. [Available from the National Meteorological Center, Washington D.C. 20233.]

——, and ——, 1991b: Ensemble forecasting: Are two members enough? Research Highlights of the NMC Development Division: 1990–1991, 435–438. [Available from the National Meteorological Center, Washington D.C. 20233.]

Tracton, S., and E. Kalnay, 1993: Ensemble forecasting at NMC: Operational Implementation. *Wea. Forecasting,* submitted.

Tribbia, J. J., and D. P. Baumhefner, 1988: Estimates of the predictability of low-frequency variability with a spectral general circulation model. *J. Atmos. Sci.,* **45,** 2306–2317.

Zhao, Q., 1993: The incorporation and initialization of cloud water/ice in an operational forecast model. Ph.D. thesis, the University of Oklahoma, 210 pp. [Available from the National Meteorological Center, Washington D.C. 20233.]