U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
NATIONAL METEOROLOGICAL CENTER

OFFICE NOTE 407

ENSEMBLE FORECASTING AT NMC AND THE BREEDING METHOD

Z. Toth, and E. Kalnay
National Meteorological Center

7 April 1995

ENSEMBLE FORECASTING AT NMC AND THE BREEDING METHOD

Zoltan Toth[1] and Eugenia Kalnay

National Meteorological Center

[1]General Sciences Corporation, Laurel, MD

April 7, 1995

NMC Office Note 407

[1]Corresponding author address: Z. Toth, NMC Development Division, 5200 Auth Rd., Rm. 204, Camp Springs, MD 20746
e–mail: wd20zt@sun1.wwb.noaa.gov and wd23ek@sun1.wwb.noaa.gov

ABSTRACT

The breeding method has been used to generate perturbations for ensemble forecasting at NMC since December 1992. At that time a single breeding cycle with a pair of bred forecasts was implemented. A combination of bred perturbations and lagged forecasts provided a daily set of 14 global forecasts valid to 10 days. In March 1994, the ensemble was expanded to 7 independent breeding cycles on the new Cray C90 supercomputer, and the forecasts extended to 16 days. This provides 46 independent global forecasts valid for two weeks every day.

For efficient ensemble forecasting, the initial perturbations to the control analysis should adequately sample the space of possible analysis errors. We point out that the analysis cycle is like a breeding cycle: it acts as a nonlinear perturbation model upon the evolution of the real atmosphere. The perturbation (i.e., the analysis error), carried forward in the first guess forecasts, is "scaled down" at regular intervals by the use of observations. Because of this, growing errors associated with the evolving state of the atmosphere develop within the analysis cycle and dominate subsequent forecast error growth.

The breeding method simulates the development of growing errors in the analysis cycle. A difference field between two nonlinear forecasts is carried forward (and scaled down at regular intervals) upon the evolving atmospheric analysis fields. By construction, the bred modes are superpositions of the leading local (time dependent) Lyapunov vectors (LLVs) of the atmosphere. An important property of the leading LLVs is that all random perturbations assume their structure after a transient period. When several independent breeding cycles are performed, the phases and amplitudes of individual (and regional) leading LLVs are random, which ensures quasi–orthogonality among the global bred modes from independent breeding cycles.

Off–line experimental runs with a 10–member ensemble (5 independent breeding cycles) show that the ensemble mean is superior to an optimally smoothed control and to randomly generated ensemble forecasts, and compares favorably with the medium range double horizontal resolution control. Moreover, a potentially useful relationship between ensemble spread and forecast error is also found both in the spatial and time domain. The improvement in skill of 0.04–0.11 in AC in forecasts at and beyond 7 days, together with the potential for estimation of the skill, suggest that this system will be a useful operational forecast tool. The results and methodology discussed should be applicable to the new operational ensemble configuration, where 17 independent forecasts are performed every day.

The two methods used so far to produce operational ensemble forecasts, i.e., breeding and the adjoint (or "optimal perturbations") technique applied at ECMWF, have several significant differences, but they both attempt to estimate the subspace of fast growing perturbations. The bred modes are estimates of fastest sustainable growth and as such they represent probable growing analysis errors. The optimal perturbations, on the other hand, estimate vectors with fastest *transient growth* and are less likely to occur in analysis error fields. A major practical difference between the two methods for ensemble

forecasting is that breeding is much simpler and far less expensive than the adjoint technique.

1. INTRODUCTION

It has long been accepted that running an ensemble of numerical forecasts from slightly perturbed initial conditions can have a beneficial impact on the skill of the forecast by means of ensemble averaging (e.g., Leith, 1974). Beyond providing a better estimate of the first moment of possible future states, the ensemble members also offer the possibility of estimating higher moments such as the forecast spread, which can be used as an indicator of expected skill, and, ultimately, the full probability distribution. Theoretically, the probability of future states can also be computed through the Liouville equations (e.g., Ehrendorfer, 1993), if the initial probability distribution is assumed to be known. However, computational and other problems make the use of these equations unfeasible for numerical weather prediction in the foreseeable future. The only current practical solution to estimating forecast probabilities is through ensemble forecasting.

One of the crucial aspects of an ensemble strategy is the generation of initial perturbations. These perturbations should realistically represent the span of possible errors in our control analysis. But since the number of ensemble forecast members is strongly limited by computational costs, it is essential that this limited number of perturbations optimally sample the initial error probability distribution.

In this paper we discuss some of the properties of the breeding method used for ensemble forecasting at NMC since 7 December 1992. At that time a system with a single breeding cycle was introduced, and a combination of bred perturbations and up to two–day lagged forecasts provided 14 global predictions valid to 10 days every day (Tracton and Kalnay, 1993, Toth and Kalnay, 1993). In March 1994, the ensemble was expanded to 7 independent breeding cycles on the new Cray C90 supercomputer, and the forecasts were extended to 16 days. This configuration provides now 46 independent global forecasts valid for two weeks every day. The results presented here were obtained in the process of investigating optimal strategies for the breeding method, many of which have been incorporated into operations.

In sections 2 and 3 we discuss basic questions related to ensemble forecasting. In sections 4 and 5 the characteristics and several technical aspects of the breeding method used at NMC for generating initial ensemble perturbations are pesented. Section 6 is devoted to experimental results. A short review about the operational implementation, further discussions and conclusions are found in sections 7 and 8.

2. ENSEMBLE FORECASTING AND NONLINEAR FILTERING

Leith (1974) showed that averaging the ensemble forecasts yields a mean forecast superior to the control forecast, as long as the ensemble perturbations are *representative of the initial probability distribution* of the basic flow around the control analysis. We illustrate why this is the case by means of a very simple error growth example.

Consider a traveling extratropical low. At the initial time, we assume that the center of the low is analyzed with a small error E_0. We assume that the error will grow

exponentially at first, and that later nonlinear effects will lead to error saturation. We can therefore use Lorenz' (1982) simple error growth model:

$$dv/dt = av(1-v) \qquad (1)$$

where $v(t)$ is the algebraic forecast error measured at the center of the system at time t and a is the linear growth rate. We can create a simple ensemble by adding and subtracting a perturbation P from the control analysis. These perturbed analyses will have an error of E_0+P and E_0-P, respectively. If the perturbation size is smaller than $2E_0$, one of these perturbed analyses will be closer to the true atmospheric solution than the control analysis, though we do not know *a priori* which one it is. If the perturbed initial conditions are plugged into the error equation (1), it is easy to see that the average of the two perturbed forecasts has a smaller error than the control at any forecast time t:

$$v_{con}(t) > (v_{pos}(t) + v_{neg}(t))/2 \qquad (2)$$

where $v_{pos}(t)$ and $v_{neg}(t)$ are the errors for the two perturbed forecasts. In Fig. 1, we show an example of the effect of ensemble averaging in this simple model.

We can generalize the above simple example by assuming that we measure the error $v(t)$ over the whole domain of a synoptic system. In this case, the initial error is a vector $\boldsymbol{E_0}$ of magnitude E_0, whose direction represents a particular spatial distribution pattern. Let us assume that the error growth with time is still given by (1). If the initial perturbation is chosen along the initial error pattern, i.e., if $\boldsymbol{P_0}$ is parallel to $\boldsymbol{E_0}$, then equation (2) is still valid. Ensemble averaging again provides a nonlinear filter that removes part of the growing error. As we will see later, much of this improvement is a characteristic of ensemble averaging and cannot be reproduced by simple spatial filtering.

On the other hand, if $\boldsymbol{P_0}$ is a growing perturbation *orthogonal* to $\boldsymbol{E_0}$, ensemble averaging will result in a worse forecast than the control, which has *no* error along $\boldsymbol{P_0}$. The ensemble average will diverge from the control forecast due to the different nonlinear evolution of the $+\boldsymbol{P_0}$ and $-\boldsymbol{P_0}$ perturbations, whose growth is also represented by equations (1) and (2), and therefore it will have a larger total error than the control. This example, although admittedly very simplistic, makes a strong case for the use of perturbations that are realistic analysis errors as initial ensemble perturbations: Growing errors that are not present in the analysis as errors may be counterproductive for ensemble forecasting, by increasing the error in the ensemble average.

We mention in passing that ideally one would like to use a large ensemble to represent, with different probabilities, all possible states of the atmosphere, given the control analysis. In this case the ensemble mean would provide at all lead times the best estimate possible for the future state of the atmosphere. In practice, however, only a small number of ensemble forecasts can be run. As Leith (1974) showed, hedging the forecasts toward climatology can give an additional improvement in some measures of forecast skill beyond that attained by ensemble averaging. In this paper (except in section 6.4 where the effect of spatial smoothing is studied) we restrict our attention to the impact of ensemble averaging.

3. ERRORS IN THE ANALYSIS

It is clear that with the initial ensemble perturbations we must represent accurately the probability distribution of the state of the atmosphere about our best estimate of the true state of the atmosphere, the latest control analysis. The shape of this probability distribution will depend on what kind of errors we may have in the control analysis. The more likely an error pattern, the higher probability we should assign to the control analysis plus and minus that particular error pattern. This calls for a careful examination of possible analysis errors.

3.1 *Random and Growing Errors*

A typical operational analysis performed with Optimal Interpolation or Spectral Statistical Interpolation (see, e.g., Lorenc, 1982, Parrish and Derber, 1992), is a weighted average of (1) observational measurements and (2) a short–range dynamical forecast (first guess), started from the preceeding analysis. It has been long recognized that the resulting analysis is affected by random errors present in observations. Recently it was also pointed out that the repeated use of a model forecast as a first guess has a profound dynamical effect on the errors in the analysis (Toth and Kalnay, 1993; Kalnay and Toth, 1994). The analysis cycle can be considered as the running of a nonlinear perturbation model upon the true state of the atmosphere. The perturbation amplitude (i.e., the analysis error) is kept small by periodic "rescaling", performed at each analysis time, through the use of limited observational data.

In such a nonlinear perturbation setup, it is inevitable that the random errors introduced at each analysis time will project onto growing modes of the atmospheric flow at later times. This is because the growing components of the error, by definition, rapidly amplify while the decaying components quickly lose their amplitude in the short–range, first guess forecast (see section 4.1.) And since the observations underdetermine the state of the atmosphere, these dynamically developing errors cannot be removed at the next analysis time: their amplitude can only be reduced (see, e.g., Fig. 9 of Bouttier, 1994.) So at the start of the next short–range forecast in the analysis cycle, dynamically developed errors are present in the initial conditions ready to amplify again. In other words, the analysis contains both random errors introduced by the most recent observations, and growing errors associated with the instabilities of the evolving flow, dynamically generated (from earlier introduced random errors) by the repeated use of the forecast first guess.

3.2 *Which Type of Error is Important?*

If we could follow the development of the errors present in the analysis, we would see that random errors, introduced just at the latest analysis time by observational inaccuracies, will decay initially before projecting, after one or two days, onto growing modes of the evolving basic flow. Such an initial decay was apparent in early experiments in atmospheric predictability (see, e.g., Fig. 4 in Smagorinsky, 1969.) Even if the random errors are balanced, they will still initially grow very slowly or decay. By contrast, "growing

errors" will, by definition, amplify rapidly, so that they are primarily responsible for short–range error growth. This means that even though the growing errors constitute only a portion of the total control analysis error field, their contribution is dominant in the forecast error development. Therefore one should focus on the growing errors when creating ensemble perturbations.

Another difference between random and growing errors is that the dimension of the space of possible random perturbations is extremely large, of the order of the number of degrees of freedom of the model, whereas the dimension of the phase space of fast growing perturbations is very much limited by the local (in the phase space) dynamics of the atmosphere. Since the full phase space of the random component of errors cannot be sampled well, random perturbations may actually *degrade* the quality of the ensemble by projecting, at a later time, on growing modes that were not initially present as analysis errors. This explains why ensemble forecasting can never be as successful with random perturbations as with estimates of fast growing errors that are possibly present in the control analysis (see section 6.8.)

4. LOCAL LYAPUNOV VECTORS AND THEIR ESTIMATION THROUGH BREEDING

Since the important, growing component of the analysis error occupies only a relatively small subdomain in the phase space, and it depends on the basic flow, it is possible to compute estimates of possible growing analysis errors through dynamical methods.

4.1 *The Breeding Method*

For this purpose, Toth and Kalnay (1993) proposed a method called *breeding of the growing modes* of the atmosphere (BGM). This procedure consists of the following simple steps: (a) add a very small, *arbitrary* perturbation to the atmospheric analysis (initial state) at a given day t_0 (or to any other basic state, such as a long model run), (b) integrate the model from both the perturbed and unperturbed initial conditions for a short period t_1–t_0 (e.g. one day), (c) subtract one forecast from the other, and (d) scale down the difference field so that it has the same norm (e.g., rms or rotational kinetic energy) as the initial perturbation. This perturbation is now (e) added to the analysis corresponding to the following day t_1, and the process (b)–(e) is repeated forward in time. Note that once the initial perturbation is introduced in step (a), the development of the perturbation field is dynamically determined by the evolving atmospheric flow.

By construction, this method "breeds" the modes that grow fastest on the trajectory taken by the evolving atmosphere in the phase space. One can decompose the initial perturbation $P(t_0)$ into growing and decaying components. Let us consider the development of a small perturbation on top of a nonlinear model trajectory (i. e., the difference between two nonlinear forecasts.) At the end of a short–range integration, by definition, the relative contribution of the growing component will be larger, while that of the decaying component smaller than at initial time. And after a few cycles, the decaying component will become negligible.

Note the similarity between the breeding method and the analysis cycle: in both cases, a nonlinear perturbation model is run with regular rescaling. In case of breeding, the perturbation is run over the analyzed states. The perturbations are defined with respect to the analysis and then rescaling is done in a deterministic fashion, so that stochastic (or decaying) components are eliminated from the perturbations, as discussed above. The resulting perturbations are determined purely by the dynamics of the system. On the other hand, the analysis cycle is run based on observed data. The perturbations here can be defined as the difference (error) between the analysis/first guess and the true state of the atmosphere (which is unknown to us.) In the first guess short range forecast, the growing components of this error will still amplify. However, at the next analysis time observational data will be used to reduce the difference between the analysis and the true state of the atmosphere. The observed data contains random noise which will be periodically reintroduced into the analysis. Consequently the errors present in the analysis, beyond the growing error connected to the use of short range forecasts as first guess fields, also contain a random or stochastic component.

4.2 *Lyapunov vectors*

Theoretically, the bred perturbations are related to the local Lyapunov vectors (LLVs) of the atmosphere. The Lyapunov exponents (λ_i) have been widely used for characterizing the behavior of simple dynamical systems:

$$\lambda_i = \lim_{t\to\infty} 1/t \log_2 [p_i(t)/p_i(0)] \qquad (3)$$

where p is a *linear* perturbation spanning the phase space of the system with orthogonal vectors. Note that while the first Lyapunov exponent is uniquely defined at least for Hamiltonian systems, the rest of the spectrum is derived via a periodic reorthogonalization of the perturbation vectors (see, e. g., Benettin et al., 1980) and hence will depend on the frequency of reorthogonalization. The λ_i's can be computed either for the whole attractor (global Lyapunov exponents) or can be interpreted point–wise, where the growth ratio is evaluated for an infinitesimal time interval at t (local Lyapunov exponents, see, e. g., Trevisan and Legnani, 1995.) The leading Lyapunov exponents are associated with predictability properties of dynamical systems, namely how fast nearby trajectories diverge (or converge) on the attractor. Most importantly, if a system has at least one positive global Lyapunov exponent, its behavior is chaotic, i.e. arbitrarily close points on the attractor will eventually separate into unrelated points (Wolf et al., 1985.)

When the Lyapunov exponents are interpreted locally, each of them can be associated with a perturbation vector, l_i. The first of these vectors, with the largest exponent, can be uniquely determined and represents a perturbation at time t_0, into which *any random perturbation* introduced infinite time earlier, develops linearly. The importance of this property of LLVs in meteorology was first recognized by Lorenz (1965), who found in his experiments with a simple linear perturbation model that initially random perturbations had a strong similarity after eight days of integration. Indeed, our breeding experiments with a state–of–the–art general circulation model indicate that one needs only a few days of integration (3–4 days) in order to get a good estimate of the leading local Lyapunov vectors

of the atmosphere. These LLVs are the vectors that grew asymptotically fastest during a time period *leading* to the analysis. Hence they are likely to dominate growing analysis errors and, because of their sustainable growth, also the forecast errors.

4.3 *Extension of Lyapunov characteristics into the nonlinear domain*

There is an extensive body of literature on the global, and more recently, on the local Lyapunov exponents of simple dynamical models. These studies, however, use a *linear tangent* model approach and are concerned only about error growth in a linear sense. In some studies, a regular rescaling of the perturbations, also used in the breeding method, has been applied. Rescaling in these linear methods, however, is used to avoid computer overflow, not to prevent nonlinear saturation (see, e.g., Benettin et al., 1976; Shimada and Nagashima, 1979.) New aspects of the breeding method as proposed by Toth and Kalnay (1993) are that perturbations are developed for a (1) complex physical system in a (2) nonlinear framework at a (3) high horizontal and vertical resolution, and that it is (4) the perturbation vectors (and not only the exponents) that are studied and used for real world practical applications.

Nonlinearity plays a crucial role in complex systems where a host of different physical processes, associated with widely different growth rates and nonlinear saturation levels occur. A traditional linear approach may find the strongest instability of the system (say, convection) but this may be associated with processes with a very low nonlinear saturation level. For finite perturbation amplitudes these modes would thus be irrelevant and may even decay. Hence the bred modes can be considered as an extension of the notion of LLVs into the nonlinear perturbation domain. Note that the perturbation amplitude is the only free parameter in the BGM method and that the bred modes, just as the linear LLVs, are not sensitive to the type of norm used for rescaling.

Another new aspect of the breeding method is that local Lyapunov vectors, not only global Lyapunov exponents, are estimated. This results in the first computation of the leading local Lyapunov vectors of the atmosphere with a comprehensive nonlinear perturbation model including all physical parameterizations.

4.4 Multiple Breeding Cycles

When a breeding cycle is started, an arbitrary initial perturbation field is added upon the control analysis. After three or four days of breeding, most of the originally decaying components in the perturbation disappear and the perturbation growth rate reaches an asymptotic value around 1.6 per day (with a perturbation amplitude of 1% in total climatological rms variance.) After this time, the perturbations that remain are those that could produce the largest growth over the preceeding 3 days or so, given the initial perturbations. As seen from Fig. 2, the growth rate in a breeding cycle depends on the amplitude of perturbations but is always larger than that obtained with other perturbation methods such as Monte Carlo, Scaled Lagged Averaged Forecasting (Ebisuzaki and Kalnay, 1991) or difference fields between short range forecasts verifying at the same time.

The growth rate and the shape (not shown) of the perturbations are largely independent of the rms amplitude in the range of about 1–10% of the natural rms variability. However, if the perturbation amplitude is reduced to less than 0.1% rms variance, then the growth rate increases enormously, with an amplification factor well above 5 per day. This is because the fastest growing modes in the model atmosphere are, in fact, convective modes (see Fig. 3), not baroclinic modes. The convective modes, however, saturate at much smaller amplitudes than the estimated size of the analysis errors (5–10% of the rms of the natural variability). The modes associated with convection are also present at larger perturbation amplitudes but are not detectable because they saturate at amplitudes much smaller than those of baroclinic instabilities. This also explains why convective modes are not dominant analysis errors.

Since breeding is a nonlinear process, the perturbations in the 1–10% rms variance range, though primarily determined by the dynamics of the system, also depend to some extent on the perturbation at previous times, namely on how those perturbations project on certain growing modes, and on the small scale forcing convection provides to th elarger scales. This forcing (see Fig. 3) is largely stochastic with respect to the baroclinic processes that dominate perturbation development in the amplitude range of 1–10% rms variance. If we start independent breeding cycles with different arbitrary initial perturbations, we find that after a transient period of about 3 days, the perturbations in the different cycles are quite similar (except for their phase, and to some extent, their amplitude, which are arbitrary) but only over roughly half of the global domain. Table 1 shows the results obtained using twenty independent breeding cycles. The local shape of the perturbations were compared to those observed in the perturbation #17 over three selected regions in the Northern and the Southern Hemispheres (see Fig. 4.) A +, – or a blank indicate whether the same perturbation was observed with the same or opposite sign, or whether a different perturbation was observed[2].

In the areas where the perturbations are very similar, the largest Lyapunov exponent must have a value much larger than the succesive Lyapunov exponents. Over the rest of the domain, different modes appear in the independent cycles, suggesting that the first few Lyapunov vectors have similar growth rates, and the appearance of one or another in any cycle depends on the details of perturbation evolution in that cycle a few days earlier and also on the details of instantaneous stochastic forcing (convection).

From the above experiments it is clear that each global perturbation pattern is a superposition of a number of regional features or modes, perhaps of the order of 10–20 in each hemisphere, which, in turn, are primarily associated with baroclinically unstable regions of the evolving basic flow. And as the basic flow has many degrees of freedom, so does the global perturbation field. It follows that the phase and amplitude of the regional modes (and in case there are competing modes with similar growth rates, the modes themselves) in one area are independent of those in remote areas. This insures that the

[2]When the same comparison is made with bred perturbations valid on different dates, even as close as 2 days apart, there is almost no correspondence among the modes, showing that the growing modes crucially depend on the basic flow and its recent evolution.

bred global perturbations from independent cycles are *quasi–orthogonal*, without imposing any constraints.

In summary, a bred global perturbation is a superposition of regional modes, each of which is a combination of the leading local (in phase space) Lyapunov vectors in that area of the atmosphere. The weights on the individual local Lyapunov vectors are randomly assigned by the arbitrary initial perturbation and the stochastic small scale forcing but are, in a statistical (ensemble average) sense proportional to the Lyapunov exponents themselves. The bred perturbations are therefore not unique in a strict sense but only in a statistical, ensemble average sense. And the more independent breeding cycles we have, the better we can span the space of possible fast growing analysis errors. Nonlinear breeding hence can be considered as a generalization of the notion of Lyapunov vectors for complex nonlinear systems. Because of nonlinear interactions and stochastic forcing by convection, and because of the existence of many regional modes, different breeding cycles do not converge to a single leading LLV but rather span the subspace of fast growing perturbations that dominate error growth at the amplitude of the size of the perturbations.

4.5 Optimal perturbations and Lyapunov vectors

There is another method to determine fast growing modes of dynamical systems. This linear method uses the linear tangent and adjoint of a full model to compute the initial perturbations that grow fastest over a specified period, measured with a given norm (Lorenz, 1965). In its application to ensemble forecasting at ECMWF (see, e.g., Molteni et al., 1995), the fastest growing perturbations are determined for a 36–hour forecast trajectory created by the full model. The optimal vectors (which are also called the singular vectors of the linear propagator) are those that amplify most over the optimization period, given the norm and other possible constraints. We briefly compare the LLVs with the optimal vectors and their respective use for ensemble forecasting at NMC and ECMWF.

4.5.a Theoretical considerations

In a linear framework, all perturbations introduced into a dynamical system will with time rotate into the direction of the LLV. This explains why the leading LLVs play such a crucial role in linear perturbation development. The local Lyapunov exponent associated with the leading LLV characterizes the fastest "sustainable" growth; faster growth can exist only for short periods of time. It follows that the LLV does not depend on any norm or other specifications; it is a general property of any dynamical system.

The leading optimal vectors represent specific directions in the phase space in which perturbation growth is higher than that associated with the LLVs. The super–Lyapunov (or super–exponential) growth, however, cannot be sustained for long since it results from a one–time rotation of the initial vector into the direction of the leading LLVs. The optimal vector can be decomposed into two components: one along, and one perpendicular to the leading LLV. At any moment during the development of the optimal vector the instantaneous growth along *the perpendicular* direction is lower than that *along the LLV*. The super–Lyapunov growth is associated with a dramatic rotation of the perpendicular vector to-

ward the LLV which results in an "apparent" extra growth (Szunyogh et al., 1995).

In contrast with the LLVs, the optimal vectors strongly depend on the specific choice of norm, optimization period and possible other constraints chosen for their definition (Vukicevic and Errico, 1990), indicating that they are much more particular features of a system than the LLVs.

4.5.b Sources of super–Lyapunov growth

Super–Lyapunov (or "super–exponential") growth has been documented in several modeling studies (e. g., Schubert and Suarez, 1989; Royer et al., 1993.) As pointed out by Lacarra and Talagrand (1988) and Farrell (1988), super–Lyapunov growth can result from an initial optimal vector being introduced into a system. Recently, however, other explanations have also been suggested. First, the local Lyapunov exponent varies on the attractor and often assumes values well above the global Lyapunov exponent (which is an integral quantity over the whole attractor), thus leading to very fast growth from time to time (see, e. g., Vannitsem and Nicolis, 1994.) Moreover, Trevisan (1993) showed that the logarithm of arithmetically averaged different local growth rate values results in an apparent super–exponential average growth. This super–Lyapunov growth is a result of first averaging different local growth rates and then taking the logarithm of the average (and it disappears when another average error definition is used in which the local logarithm values are averaged directly.) Note that even in the absence of perturbations along the optimal vectors, the arithmetic averaging of the local Lyapunov exponents yields a super–Lyapunov growth rate. As Nicolis et al. (1995) showed for a three–variable Lorenz (1984) model, this apparent super–Lyapunov growth can be readily explained by the impact of the variability of the LLVs on the attractor within the averaging process.

4.5.c LLVs and optimal vectors as analysis errors

For the application of the LLVs or optimal vectors in ensemble forecasting one has to consider whether they represent possible (and probable) analysis/forecast errors. Kalnay and Toth (1994) showed that the bred modes are present in the analysis as errors with a considerable amplitude. This should be expected since any kind of error introduced into the analysis cycle, unless completely removed, would rotate into the direction of the bred modes, due to the dynamics of perturbatin development described above. We recall that the bred modes are an extension of the linear LLVs, toward which all perturbations rotate within a few days.

The optimal vectors, however, would not similarly arise through the dynamics of the system as analysis errors. They can be introduced into a system only through special forcing, which, in case of the analysis cycle, requires a special kind of distribution of random observational errors. And since the number of degrees of freedom in the atmosphere (and in its general circulation models) is very high, and the number of fast growing optimal vectors is very limited, we believe that it is unlikely that a random error pattern would have a strong projection on the leading optimal vectors. The results of Ehrendorfer and Errico (1995), who considered an extended spectrum of optimal vectors in a limited area model, seem to support this conjecture. They found that only a small fraction of the leading optimal

vectors have growth rates higher than one. It follows that random perturbations would project poorly on the leading optimal vectors. This is also true for the leading LLVs. However, in a perturbation model like the analysis cycle random perturbations, by definition, would naturally rotate into the direction of the leading LLVs but not toward the leading optimal vectors (which can arise only through special forcing.)

The statement that the optimal vectors are not likely analysis errors may seem to contradict the results of Rabier et al., 1994. They searched for the initial vector that maximizes perturbation growth projecting at final time onto the 48–hour forecast error pattern. They found that the initial vector that maximizes growth in that special direction has characteristics similar to the optimal vectors. Note, however, that Rabier et al. (1994) solve a problem very similar to (but more specific than) the general (or unconstrained) optimization problem of Molteni et al (1995): The final perturbation vector (48–hour error pattern) acts as an additional constraint in the same optimization problem. Given the dependence of the optimization on the choice of norm, it is not surprising that the resulting initial "sensitivity" patterns of Rabier et al. (1994) have a similar structure to the optimal initial perturbations used for ensemble forecasting at ECMWF (Molteni et al., 1995).

We note again that the sensitivity patterns (and also the optimal vectors) are not unique in the sense that they depend on the norm used in their definition. In other words, there is no unique way of going back in time in a dissipative chaotic system like the atmosphere. In fact, very different initial vectors may yield similar end perturbations, as discussed already above with respect to the leading LLV (see also Zupanski, 1995.) The influence of the norm on optimal/sensitivity patterns is also apparent in the results of Oortwijn and Barkmeijer (1995). Oortwijn and Barkmeijer (1995) used an optimization technique similar to that of Rabier et al. (1994) except that their final target perturbation patterns were anomalies associated with weather regimes. They found that the initial perturbations optimally triggering those anomalies were, to a large extent, combinations of the leading *unconstrained* optimal vectors. Their results suggest that no matter what the exact target perturbation pattern is at final time, once the norm is fixed, the initial optimized perturbations will have a large projection on the leading unconstrained optimal vectors that can produce the fastest temporary growth through the special rotation of their component vectors in phase space.

In short, since the optimal/sensitivity patterns are not unique and they depend on the norm used, the question whether they project on *actual* analysis/forecast errors can only be tested through the use of high quality *observational data*.

4.5.d Practical aspects

There are several practical differences between the breeding and the optimal vectors methods as used at NMC and ECMWF respectively. (1) Computational efficiency: the adjoint technique is very expensive whereas breeding is essentially cost–free, apart from running the ensemble forecasts themselves (see Fig. 5.) (2) Spatial resolution: breeding can be performed at full resolution while for the optimal perturbations technique this is computationally impractical. (3) Breeding is performed with the full nonlinear model with physics

while the optimization is currently done with a tangent linear system with limited physical parameterizations. (4) Spatial coverage: the fastest optimal perturbations cover only a small fraction of the geographical domain with relatively large amplitudes and only over the NH (see Fig. 3.d,e,f in Molteni et al., 1995), while with breeding, the fastest growing regional modes are automatically determined for the whole globe (including the tropics and the Southern Hemisphere), and not only for those regions with highest growth rate. (5) The optimal vectors are determined for a forecast trajectory, so that to the extent the forecast is not perfect, the modes determined will be sub–optimal. In contrast, the LLVs, which provide the fastest sustainable growth for the future[3], depend only on analyzed data.

5. ENSEMBLE PERTURBATIONS

From the discussion about LLVs above, one could draw the conclusion that it does not really matter what initial perturbations are used for medium– or extended–range predictions since all linear perturbations turn into very similar vectors after a few days of integration. However, one should keep in mind that ensemble forecasts, just as the control forecast, are *nonlinear* integrations. With a perturbation size similar to the estimated size of errors in the analysis, nonlinearity becomes important after about two days, and earlier than that in fast developing synoptic systems. It follows that in the ensemble perturbations, we still need to represent realistically the initial uncertainty in the analysis, otherwise, as discussed in section 2, our ensemble will be suboptimal.

In this section we discuss several additional technical points about the breeding method which were investigated in the process of implementing operational ensemble forecasting at NMC.

5.1 Regional rescaling

The breeding method was originally used at NMC with hemispherically determined rescaling factors (Toth and Kalnay, 1993). Depending on the hemispheric rms magnitude of the perturbation, a constant factor was applied over each hemisphere and a linearly interpolated value was used in the tropics in the rescaling. While this method is very good to study the instabilities of the atmosphere as they are represented in our numerical models, it may not be optimal for ensemble forecasting. The perturbations should reflect not only the shape, but also the size of analysis errors. Consequently, we want to have larger regional perturbation amplitudes in regions sparsely observed, and vice versa. With hemispherically fixed rescaling, the perturbation amplitudes will be largest in the areas of strongest instabilities. While these areas are generally over the poorly observed oceans, they do not necessarily correspond to the regionally dependent uncertainty in the analysis.

To estimate the geographically dependent uncertainty in the analysis, we ran two independent analysis cycles for a 30–day period in April–May, 1992. The cycles were

[3]Note that dynamical systems behave continuously in time. The LLV reflects this time continuity: Whatever vector provided the maximum sustainable growth over the segment of trajectory leading to current time, will provide the maximum growth for the future, too.

identical except that in one of them the first guess field was an ensemble average of two first guesses, perturbed by bred modes with positive and negative signs. The two analyses gradually diverged from each other until, a few days later, the difference saturated. Beyond this time, we took the average of rms difference fields between corresponding pairs of analyses. The ~500 hPa average difference field in the streamfunction, scaled so that the global average is one, and smoothed with a Gaussian filter on a sphere (Jim Purser[4], pers. comm., 1993) equivalent to T6–T7 (about 2000 km) resolution, is shown in Fig. 6. Over the Northern Hemisphere, the dominant features of the analysis uncertainty field are the minima over North America and Eurasia, especially over the eastern part of the continents, and the high values over the Pacific ocean. This corresponds well to the good rawinsonde coverage over the continents. Due to the use of dynamical first guess, the information from the observations is "transported" eastward, resulting in minima over the eastern part of the continents.

While there is a hint of a similar behavior in the Southern Hemisphere east of Africa and over eastern Australia, there is more zonal symmetry, and the amplitude of the uncertainty increases poleward. Such behaviour is also consistent with the uniform observational coverage provided by satellite temperatures and by the fast growth of perturbations in the strongly baroclinic southern high latitudes.

Note that with the above procedure, we can estimate the amplitude of growing errors in the analysis, which, as we discussed in section 3.2, are assumed to be more important in ensemble forecasting. Optimal Interpolation (OI) could also be used to estimate the distribution of the analysis errors (Gandin, 1963), but such estimate is very dependent on the assumed error covariances for the forecast and the observations. In addition, the OI estimate would not properly account for the growing component of the error. Therefore we believe the OI estimate would be less reliable than the empirical procedure we have used here.

In a breeding cycle specifically modified for ensemble perturbations, we determine the scaling factor as a function of horizontal location. The perturbation amplitude is measured and rescaled regionally in a smooth fashion, to a level corresponding to the values shown in Fig. 6. At points where the perturbation amplitude (globally scaled to 1) is below that in Fig. 6, no rescaling is applied. So a perturbation traveling into a poorly observed oceanic area is allowed to grow freely, while those reaching a well observed area are scaled back to the size of the estimated analysis error. Since the regional rescaling is done in a smooth fashion, most of the balance naturally present in the bred perturbations is preserved. With regional rescaling we still retain the capability of changing the overall global or hemispheric amplitude but the smoothed relative geographical distribution is left intact.

Medium–range ensemble forecasts performed with the breeding method modified for regional rescaling showed an improvement in skill over the Southern Hemisphere and

[4]By using this spectral filter we avoid the aliasing problem associated with simple truncation in wavenumber space. Different filtering characteristics are described in terms of "equivalent" triangular truncation.

the tropics (compared to the hemispherically rescaled perturbations) while there was no change over the Northern extratropics. We also tested applying the regional rescaling outside of the breeding cycle, to modify only the initial ensemble perturbations, but found that larger changes were necessary after each cycle and that the forecast results were not as good.

5.2 Centering the ensemble around the control analysis

Since our best estimate of the true state of the atmosphere is the control analysis, we must center the ensemble perturbations around this field. This can be easily done by adding and subtracting the same perturbation to the control analysis (e.g., Ebisuzaki and Kalnay, 1991). In this setup, *2n* perturbations are derived from *n* independent breeding cycles (or from other orthogonal vectors.) However, a case can be made for using each perturbation only once, thus possibly improving sampling (J. Purser, pers. comm., 1992.) We tested this hypothesis by averaging *2n independent* perturbations and then removing their average from each individual perturbation vector. The resulting medium–range ensemble integrations, however, had a significantly inferior forecast skill as compared to the identically sized paired ensemble setup: the improvement upon the control forecast obtained with the centered single perturbations was less than two thirds of that obtained with the ensemble of positive and negative pairs of perturbations (see Table 2.) This result underlines the inherently nonlinear nature of ensemble forecasting. The implication is that the nonlinear ensemble filtering mechanism discussed in Section 2 is not as effective if the perturbations, though centered initially in a linear sense, are not paired.

6. ENSEMBLE FORECASTING RESULTS

In this section we will give an overview of ensemble forecasting experiments performed in order to test possible operational configurations. All experiments were done with a T62/18 levels version of the NMC Medium Range Forecast (MRF) model (Kanamitsu et al., 1991). The period used in these experiments is the 40 days between 6 May 1992 and 14 June 1992 (or a subperiod of it, where noted). Unless mentioned otherwise, 10–member ensemble forecasts are evaluated. The initial ensemble perturbations were derived from 5 independent breeding cycles with regional rescaling. To center the ensemble mean on the control analysis at initial time, each of the five perturbations was both added to and subtracted from the analysis. The quality of the ensemble forecasts is estimated using two measures: the skill of the ensemble mean forecast and the spread of the ensemble.

6.1 Measures of ensemble quality

At any lead time, members of the ensemble can be averaged. The mean ensemble forecast is then verified against the corresponding analysis much the same way as the control forecast. As a measure of skill, we use the forecast/analysis pattern anomaly correlation (PAC) measured over three separate belts over the globe: the northern and

southern hemisphere extratropics (20–80° latitude belts) and the tropics (+/– 20° latitude). All scores are computed for the streamfunction field at a sigma layer close to the 500 hPa height level. To compute the anomalies, the observed climatology is used. RMS errors were also computed but are not reported here because they led to identical conclusions as the PACs. Forecast PACs for different types of ensembles are compared to those for the control forecast to see if they represent an improvement due to nonlinear ensemble filtering.

The spread of the ensemble is determined as the average of the difference fields between the individual ensemble forecasts and the ensemble mean. The difference at each grid point is defined as the square root of the kinetic energy in the difference (or error) field. The spatial distribution of the spread is considered as a prediction for the spatial distribution of the actual error in the control forecast, which is measured in the same way, in units of square root of kinetic energy. After setting the mean of both the forecast spread and observed error fields to zero, their correlation is computed (spread/error PAC.) Spread/error PACs are computed only in the T3 – T15 range of equivalent spatial resolution using the spectral filter mentioned above (Purser, pers. comm.). Time correlations between spread and error statistics are also computed (in which case the spatial mean of the spread and error fields is not removed.)

6.2 Size of the initial perturbation

In the section on regional rescaling (5.1), we indicated that the overall size of the initial perturbations is an important parameter that has to be chosen to reflect the size of initial error in the analysis. An estimate of the analysis error can be derived from optimal interpolation analysis techniques (see, e.g., Gandin, 1963, Buizza, 1994.) However, since these estimates are subject to the statistical approximations made within the analysis scheme, we attempted to optimize the overall perturbation size experimentally by verifying ensemble means for ensembles initiated with different initial amplitudes for the bred perturbations. The perturbation size is measured on the ~500 hPa streamfunction field. We note that the wintertime NH natural rms variability of the streamfunction field is around 8500000 m^2/sec (while it is around 80 m for geopotential height.)

To estimate the optimal size of the initial perturbations, we performed tests with different values between 3 and 20% of the NH winter variability for the NH and 6 and 40% for the SH respectively, and recorded the skill score for the mean of the different ensembles. Since at T62 resolution much of the small perturbations develop linearly in the first 24 hours time range, the ensemble mean of perturbations equal to or less than 10% of the rms variance (standard deviation) is not appreciably different from the control at one day. Though at this short lead time the skill of the ensemble mean cannot be directly used to determine the optimal perturbation size[5], it is important to note that perturbed forecasts with 10% initial "error" for the NH and 20% for the SH diverged from the control as much as the

[5]The signal is hard to detect because the errors in the verifying analysis are not much smaller than the short–range forecast errors. Had we used observational data for verifications instead of analysis fields, we may have been able to find a signal even at very short range.

control forecast diverged from the verifying analysis (not shown), suggesting that the optimal perturbation size is around this magnitude. This agrees well with other estimates for the error in global analysis fields. Kalnay et al. (1993) found that the difference between independent 500 hPa height analyses from various centers is between 7 and 16 m for the NH and between 12 and 18 m for the SH. P. Caplan (1994, personal communication) estimated differences in the same range, with the SH uncertainty being about double of that for the NH. These estimates, along with other information such as improvement in forecast skill suggest that the quality of our atmospheric analysis has been considerably improved since the mid 1980's when Daley and Mayer (1986) estimated the global analysis error to be between 15 and 20 m at 500 hPa.

In Table 3 we show the results of using different perturbation sizes for day 3 to 9, comparing them with perturbations of size 10% for the NH and 20% for the SH. We find that, for the NH, at day 3 an amplitude of 7.5% is slightly better than 10%, whereas at day 9, 12.5% is better. This increase in the optimal initial size with forecast length is also observed in the SH: at day 3 a size of about 25% is better, whereas at day 9 a size of 30% is more effective in increasing the skill of the ensemble average.

In a perfect model environment, the optimal perturbation size should not depend on lead time. However, our models are imperfect, which means that forecast errors are growing not only due to the initial difference but also due to model deficiencies (Reynolds et al., 1994). Part of the model generated errors project on growing modes and act like amplifying errors due to the initial uncertainty, whereas others appear as a forecast bias. The model errors that project onto growing modes can be dealt with, to some extent, as an extra amplitude term in the initial error field, explaining why the optimal perturbation amplitude increases slightly with increasing lead time.

Based on the above results we have fixed the initial amplitude of perturbations in the remainder of this study at 12.5/25% rms standard deviation for the NH/SH respectively. Note that this amplitude is larger than optimal for short lead times but is around optimal for the medium and extended range.

6.3 Ensemble mean forecasts

Fig. 7 shows the PAC scores for the control and ensemble mean forecasts for the experimental period. First we should note that ensemble averaging has a greater impact over the winter (in this case the SH) than over the summer hemisphere. This is the case probably because in winter, baroclinic disturbances are the sole major source of instabilities. These instabilities have a relatively long life cycle (few days) and a large saturation amplitude. Consequently, baroclinic instabilities are directly responsible for a large portion of wintertime forecast errors. And since at T62 resolution these instabilities are well resolved, the ensemble based on these perturbations is very effective in filtering out part of the forecast error that is due to initial error uncertainty. In contrast, the circulation in the summer is more "local" in nature, both in space and time. This is also reflected in the fact that the summer circulation has more spatial degrees of freedom (see, e. g., Fraedrich et al., 1995.) Beyond large scale dynamics, it is also strongly influenced by convection,

which has a shorter life time and smaller saturation amplitude. It follows that a larger portion of the total error is left unexplained by baroclinic instabilities. As a consequence our ensembles based primarily on baroclinic instabilities cannot provide as much improvement in skill in the summer as they can in the winter.

As can be seen from Fig. 7 the skill for the control and ensemble mean at day 1 are practically identical when verified against the control analysis (see also footnote 5). However, beginning day 2 the ensemble mean develops an advantage over the control forecast that becomes appreciable by day 3 and reaches a substantial 0.07–0.11 by day 9. If we consider 0.5 PAC as the minimum level of useful skill, ensemble forecasting extends predictability by a day or so, out to 8 days over the NH and 7 days over the SH and the tropics. Note that the improvements from ensemble averaging are as large in the tropics as they are over the summer hemisphere extratropics.

The gain from ensemble forecasting in the medium and extended range compares favorably with the increase obtained by doubling the horizontal resolution: At day 5, the difference between the scores obtained using the NMC operational T126 model and a nearly identical, "parallel" T62 system is slightly below 0.02, averaged over 32 months of operations. The gain obtained by ensemble averaging with 10 members over the 40–day experimental period is substantially larger, although both procedures take about the same computer time. We should point out that increasing the resolution of global NWP models has a clear benefit during the first few days of a forecast (Tracton and Kalnay, 1993). Running ensembles at a lower resolution, however, has a substantial advantage for the range beyond 5 days, where nonlinearities become important. We mention here that ensemble forecasting can also be beneficial for the shorter range, as long as the nonlinear aspects of the flow are relatively well modeled and analyzed (see Brooks et al., 1995.)

6.4 Ensemble averaging vs. spatial smoothing

It might be argued that the gain in skill from ensemble averaging may be dominated by smoothing resulting from averaging the different perturbed forecasts. Fig. 8a shows the verifying analysis for a 9–day forecast started from 30 May 1992. A comparison of the control forecast (8b) with the 10–member ensemble average forecast (8c), and their corresponding errors (8d and 8e respectively) suggests that ensemble averaging does indeed have a smoothing effect. It is more appropriate to call this effect "filtering", since it depends on the flow, particularly upon the varying degree of similarity amongst the ensemble members. Ensemble averaging results in a selective smoothing of those features that cannot be forecast with certainty. Consider, for example, the forecasts in Fig. 8 over North America. The trough over the SE US is well predicted by the control and is hardly changed by the 10–member ensemble mean. The Southern portion of the trough predicted over the West coast, however, did not verify. The ensemble mean filtered out part of this system, resulting in smaller overall errors in this region. Undoubtedly, there are several other areas/cases where the changes in the ensemble mean do not verify well but overall, it still provides an improvement over the single control forecast.

To quantify how much of the improvement due to ensemble averaging is connected to simple spatial smoothing (as compared to nonlinear filtering), we performed experiments where both the control and the ensemble mean forecasts were spatially smoothed till they reached their maximum PAC verification scores. The results, presented in Table 4, show that not much smoothing is needed to maximize the scores in the extratropics. Even at 9 days lead time, a truncation of T20 has to be retained in the control while, as expected, the ensemble average requires somewhat less smoothing. In the tropics (not shown) no amount of smoothing improves the scores. The main result here is that the ensemble average retains a considerable advantage (more than 60%) over the control even after both fields had been optimally smoothed.

6.5 Forecast of the spatial distribution of the errors

Ensemble forecasting should offer more than an improved best estimate of the evolution of the atmosphere (ensemble mean forecast.) It should also provide the means to estimate higher moments, and ultimately the full probability distribution of the forecasts. A first step in achieving this goal is the derivation of an estimate of forecast reliability in the spatial domain. Ideally, we would like to know in which areas errors are more likely. We have used the spatially smoothed ensemble spread of the kinetic energy introduced in section 6.1 for estimating the magnitude of the expected forecast errors. Figs. 8f and g show, for the same 9–day forecast example of the previous subsection, the spatial distribution of the kinetic energy of the error and of the ensemble spread, respectively. Several important aspects of the error field are indicated quite realistically in the ensemble spread field. Note, for example, that the absolute maxima in the error field over the two extratropics is well predicted by the ensemble over South of Australia and over Eastern Asia. Several error features turn out to be well predicted in the subtropics and tropics, as well: See, for example, the correspondence beween the actual and predicted large errors over Western Sahara and East of the Hawaiian Islands.

The spread/error PAC scores based on the ensemble forecasts are displayed in Fig.9. The fact that the spread/error PAC is low at short lead times is due to the presence of random errors in the initial conditions and verifying analyses (see also Barker, 1991 and Wobus and Kalnay, 1995). Since there is a strong zonally symmetric component in the error fields, we computed the PAC of the spread/error both with (not shown) and without the zonal mean included. The spread/skill spatial correlation is about 0.4 without the zonal mean and is above 0.7 with the zonal mean included. This result is encouraging, suggesting that ensemble forecasting can result in skillful predictions of the spatial distribution of the errors.

6.6 Forecast of the temporal variations in skill

The ensemble forecasts can also be used to predict the variations of forecast skill (or the reliability of forecasts) in the time domain. This has been a subject of considerable research because of its importance for medium and extended–range forecasts (e.g., Branstator, 1986; Kalnay and Dalcher, 1987; Palmer and Tibaldi, 1988.) If we can

determine *a priori* which forecasts are going to be most skillful, the utility of extended–range forecasts can be considerably enhanced (e.g., Tracton et al, 1989). Here we will test the relationship in time between ensemble spread and error, both expressed in terms of kinetic energy as discussed above, for the two extratropics and for the tropics. The two time series were correlated for the test period of 40 consecutive daily forecasts started on 6 May 1992, see Fig. 10. For the NH the correlation, except for the first two days, is around the 0.6–0.7 level. (The low correlation at day one may be due to the presence of random errors in the analysis.) The correlation for the tropics (except at days 1 and 2) is lower than that in the NH, nevertheless it still exceeds the mark of statistical significance (.31 at the 5% level) at most lead times. Over the SH the correlation values reach the level of statistical significance only at days 2 and 9.

Except for days one and two, the scores over the NH are somewhat higher than those obtained in the operational system designed by Wobus and Kalnay (1995) to predict forecast skill on a regional basis, based on control forecasts from different centers, and also larger than those reported by Barker (1991) in a T21, 2–layer QG perfect model environment. Although the results presented in the last two subsections on forecast reliability in the spatial and temporal domain are subject to more sampling variability than those on ensemble mean predictions, they indicate the ability of the bred ensemble to successfully predict higher moments of the forecast probability distribution.

6.7 Size of the ensemble

It was Leith (1974) who first considered the question of how many ensemble members are needed to improve the skill of the control forecast by ensemble averaging. Using a simple model he found that eight members are enough to realize most of the gain attainable through ensemble averaging. Houtekamer and Derome (1995), also using a perfect model environment but with a 3–layer, T21 resolution quasi–geostrophic model, basically confirmed Leith's results. Barker (1991), using a setup similar to that of Houtekamer and Derome, examined the effect of ensemble size on the temporal correlation between ensemble spread and control skill. We now consider the same question using a setup equivalent to the operational NMC ensemble system.

In Fig. 11, the skill of the ensemble mean, the skill in forecasting the spatial error pattern, and the temporal correlation between ensemble spread and control error are displayed as a function of ensemble size between 1 and 40 members. The gain from enlarging the ensemble is most obvious when going from 2 to 4 and then to 10–member ensembles, a result in agreement with earlier studies. Regarding forecast skill, only minimal improvement is obtained beyond 20 members. However, in the temporal and spatial relationship between spread and error the improvement continues to increase even up to 40 members. From the shape of these curves it seems there is still a lot to be gained from increasing the size of the ensemble beyond 40 members. Certainly it is clear from the figures that for higher forecast moments it is necessary to have many more members in order to reduce the sampling problem[6].

6.8 *Comparison of bred vs. random initial perturbations*

Finally, we compare the effectiveness of random and bred initial perturbations. The random perturbations are created by linearly combining difference fields between randomly selected analyses with random weights. Toth and Kalnay (1993) showed that 2–member bred ensembles outperformed similarly sized ensembles with random initial perturbations in terms of ensemble mean scores. As indicated in Table 5, this is also true for 10–member ensembles. The difference in the performance of ensembles with random and bred perturbations is even larger for the error forecasts. The advantage of bred perturbations is more pronounced over the winter hemisphere, where baroclinic instabilities have possibly a greater contribution to initial errors.

7. OPERATIONAL IMPLEMENTATION

The initial operational ensemble configuration implemented at NMC in December 1992 consisted one pair of bred perturbed forecasts, one T126 and a T62 control forecast, plus a 12 hour delayed control forecast (Tracton and Kalnay, 1993). All forecasts initiated in the most recent 48 hours were included, making an ensemble of 14 valid for 10 days. Based on the experimental results presented in the previous sections, and following the installation at NMC of a new Cray C90 supercomputer, the ensemble forecasting system was upgraded on 30 March 1994. In addition to the T62 and T126 control forecasts, 5 bred pairs of forecasts are run at 00Z and two pairs at 12Z, and all the forecasts are extended to 16 days. The new configuration amounts to 17 individual ensemble members every day. When the forecasts from the last two days are also considered, the total number of ensemble members valid for two weeks is 46 (see Fig. 12 and also Tracton, 1994).

Based on the results of section 6.2, the size of the initial perturbations is set at 12.5% and 25% of the total rms variance in the NH and SH respectively. (During SH summer, the perturbation size there is reduced to 12.5% rms variance.) In the regional rescaling procedure, the kinetic energy of the flow (rather than the previously used rms streamfunction norm) is applied. We use 24–hour breeding cycles, and the bred perturbations are determined as the difference between two perturbed ensemble forecasts at 24–hour lead time. It should be noted that with this procedure the generation of bred perturbations is performed at no cost beyond that of running the ensemble forecasts themselves (see Fig. 5.) Preliminary results from the new operational ensemble system support the experimental results reported in this paper. A detailed evaluation of the new system is in progress and will be reported separately.

8. CONCLUSIONS

[6]Tracton (personal communication, 1993) also pointed out that in the 14–member ensemble system that was operational earlier at NMC, the average of 4 of the latest forecasts was enough to attain most of the skill improvement, but that the reliability estimates improved further when all 14 forecasts were considered in the 6–10 days forecast range.

In this paper different aspects of ensemble forecasting were examined. First, it was emphasized that the perturbations applied to the control initial state of the atmosphere (analysis) must be chosen in the directions of possible growing errors in the analysis. Using Lorenz' error growth equation we showed that if the perturbations project on the analysis errors, averaging pairs of perturbed forecasts results in a nonlinear filtering of forecast errors. On the other hand, if the initial perturbations do not project on the errors in the analysis, the same nonlinear processes can also result in increased errors. We argued that the analysis errors are composed not only by random errors as assumed in the operational statistical interpolation methods, but also by fast growing "errors of the day" introduced by the successive use of dynamical short–range forecasts as first guess fields within the analysis cycle.

The growing errors in the analysis cycle develop as perturbations upon the evolving true state of the atmosphere. The perturbations (i.e., the analysis errors), carried forward in the first guess forecasts, are scaled down at regular intervals by the use of observations. However, because of the inhomogeneous distribution of observations it seems possible for some errors in the analysis to grow without suppression by observational data. Examples of this can be found over the oceans and the SH, where radiosonde data is scarce. Due to this process, growing errors associated with the changing state of the atmosphere develop within the analysis cycle and dominate subsequent forecast error growth.

We argued that these errors or perturbations can be well estimated by the method of "breeding growing modes", which simulates the development of growing errors in the analysis cycle. In a breeding cycle, the difference field between two nonlinear forecasts is carried forward (and scaled down at regular intervals) upon the evolving atmospheric analysis fields. By construction, the bred modes are superpositions of the leading local ime dependent) Lyapunov vectors (LLVs) of the atmosphere. Breeding cycles with different initial perturbations converge after a few days to a subspace of perturbations that comprises the leading local (in phase space) Lyapunov vectors of the atmosphere. The unique role played by the leading LLVs in analysis/forecasting was emphasized by pointing out that all linear perturbations, after a transient period of 3–4 days, assume the shape of leading LLVs. In any bred perturbation the weight on the individual leading LLVs is random, determined by how the preceeding perturbation projects on the different LLVs at that time and also on the impact of smaller scale stochastic forcing (convection) on the dominant baroclinic processes. It was shown that perturbations from independent breeding cycles are quasi–orthogonal without the introduction of any constraint. For these reasons the bred perturbations lend themselves as good candidates to be used as initial ensemble perturbations.

The growing component of the regionally varying uncertainty in the analysis was measured as the difference between parallel analysis cycles. The average difference field is then used as a mask in the regular rescaling process of the bred modes to insure that the initial ensemble perturbations have a spatial distribution of amplitudes similar to that of the

analysis errors. Each bred perturbation is both added to and subtracted from the control analysis.

Results from 10–member experimental ensembles indicate that for short range forecasts the optimal size of the initial perturbations is about the same as the estimated size of analysis errors. For longer forecasts, the optimal size is somewhat larger, presumably because of the presence of model deficiencies which generate additional forecast errors.

We showed that for medium range forecasting the mean of the bred ensemble has skill superior to that of (1) a double horizontal resolution control; (2) a control smoothed optimally; and (3) an ensemble initiated with random perturbations. We also pointed out that ensemble averaging removes the unpredictable components of the flow while leaving the predictable part virtually intact. These results attest that the bred ensemble mean offers an economic way for improving the control forecast and thus can replace the control as our best estimate of the future state of the atmosphere.

Higher moments of the probability distribution of future states should also be estimated through ensemble forecasting. For limited samples, we showed that bred ensemble spread correlated well with forecast error both in space and in time, surpassing again the performance of randomly generated ensembles. This information about the reliability of the forecasts is especially critical at longer lead times where model performance is known to be case dependent.

The improvement in forecast skill at and beyond 7 days lead times (0.04–0.11 in AC), together with a robust estimate of forecast reliability (~0.6 temporal correlation between ensemble spread and forecast errors) indicates that the bred ensemble system has the potential of extending "weather outlooks" into the second forecast week. To capitalize on this potential, NMC started on 30 March 1994 an ensemble forecasting system with 14 perturbed (bred) and 3 control forecasts, each extending out to 16 days in lead time. When all forecasts initiated within the past 48 hours are considered, there is a 46–member ensemble valid for two weeks available every day.

There has been much discussion recently both at scientific meetings and in the literature about the properties and relative merits of bred modes and optimal vectors. Despite the many differences listed in section 4 we would like to emphasize that both types of perturbations represent a subspace of possible growing perturbations. The adjoint technique used to compute the optimal vectors can be supplemented with different constraints that will make the optimal perturbations more desirable (and hence slower growing or "sub–optimal") as initial ensemble perturbations. For example, Errico and Ehrendorfer (1995) used a norm that emphasizes the larger scales and ensures balanced perturbations. Houtekamer (1995) further modified the technique to obtain modes that are statistically representative of analysis errors in terms of baroclinic shear and spatial error magnitude. The spectral wavenumber distribution of the initial optimal perturbations can also be restricted to larger scales (Hartmann et al., 1995) which has an effect of reducing baroclinicity that is otherwise uncharacteristic of analysis errors. Furthermore, work is in progress at different centers to include more physical processes in the tangent linear and

adjoint formulations of NWP models (e. g., Zupanski and Mesinger, 1995.)

Note that the bred modes are computed using the full physics package of NWP models at the highest required horizontal resolution. Without the introduction of any special constrain, they are well balanced and correspond well with the estimated vertical structure of analysis errors (Houtekamer and Derome, 1995). In view of all the above, it will probably remain a subjective decision which technique one applies operationally for the generation of initial ensemble perturbations. One may add here that if practical aspects such as simplicity and computational costs are considered, the breeding method has a clear advantage over competing methodologies.

At ECMWF the adjoint perturbation technique is used while at FNOC pre–implementation tests are being carried out with the breeding method (M. A. Rennick, personal communication, 1994.) At the Atmospheric Environment Service of Canada experiments have been carried out with an ensemble system in which, beyond the initial atmospheric conditions, the initial surface parameters, as well as some model parameters are also perturbed (P. Houtekamer, personal communication, 1995.) The perturbed atmospheric initial conditions are derived from running independent analysis cycles, in each of which randomly generated "measurement errors" are added to the real observational data. The independent analysis cycles can be considered as breeding cycles, where, beyond the growing modes, random analysis errors are also well represented in a statistical sense.

The different perturbation techniques have various potential advantages. Their impact on the quality of ensemble forecasts can be evaluated only after a careful comparison of experimental results. We conclude by noting that a combination of ensemble forecasts from different numerical prediction centers may give further improvement to the quality of an ensemble (Harrison et al., 1995). The benefits from having a larger number of forecasts, and using different analysis schemes, forecast models and perturbation techniques may all contribute to the success of numerical weather prediction.

Acknowledgements. We greatly appreciate the many discussions and help offered by M. Kanamitsu, H. Juang, M. Iredell, S. Saha, J. Purser, J. Derber, S. Tracton, J. Irwin and P. Caplan at different stages of this work. Jim Purser kindly provided us with the spherical filter used in sections 5 and 6 while Steve Tracton suggested the use of the schematic in Fig. 5. Joe Gerrity, Steve Tracton and Peter Houtekamer offered valuable comments on an earlier version of this manuscript.

REFERENCES

Barker, T. W., 1991: The relationship between spread and forecast error in extended–range forecasts. *J. Climate*, **4**, 733–742.

Benettin, G., L. Galgani, and J. M. Strelcyn, 1976: Kolmogorov entropy and numerical experiments. *Physical Review A*, **14**, 2338–2345.

Benettin, G., L. Galgani, A. Giorgilli, and J. M. Strelcyn, 1980: Lyapunov characteristic exponents for smooth dynamical systems and for Hamiltonian systems; a method for computing all of them. *Meccanica*, **15**, 9.

Branstator, G., 1986: The variability in skill of 72–hour global–scale NMC forecasts. Mon. Wea. Rev., 114, 2628–2639.

Bouttier, F., 1994: A dynamical estimation of forecast error covariances in an assimilation system. *Mon. Wea. Rev.*, **122**, 2376–2390.

Brooks, H. E., M. S. Tracton, D. J. Stensrud, G. DiMego, and Z. Toth, 1995: Short–Range Ensemble Forecasting (SREF): Report from a workshop. Bull. Amer. Meteorol. Soc., in print.

Buizza, R., 1994: Localization of optimal perturbations using a projection operator. *Q. J. R. Meteorol. Soc.*, **120**, 1647–1681.

Dalcher, A., and E. Kalnay, 1987: Error growth and predictability in operational ECMWF forecasts. *Tellus*, **39A**, 474–491.

Daley, R., and T. Mayer, 1986: Estimate of global analysis error from the Global Weather Experiment Observational Network. *Mon. Wea. Rev.*, **114**, 1642–1653.

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.}

Ehrendorfer, M., and R. M. Errico, 1995: Mesoscale predictability and the spectrum of optimal perturbations. *J. Atmos. Sci.*, under review.

Farrel, B., 1988: Optimal excitation of neutral Rossby waves. *J. Atmos. Sci.*, **45**, 163–172.

Fraedrich, K., Ch. Ziehmann, and F. Sielmann, 1995: Estimates of spatial degrees of freedom. *J. Cliamte*, **8**, 361–369.

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

Harrison, M. S. J., D. S. Richardson, K. Robertson, and A. Woodcock, 1995: Medium–range ensembles using both the ECMWF T63 and Unified models – An initial report. Technical Report No. 153, UK Met. Office. [Available from: Forecasting Research Division, Meteorological Office, London Road, Bracknell, Berkshire RG12 2SZ, UK.]

Hartmann, D. L., R. Buizza, and T. N. Palmer, 1995: Singular vectors: The effect of spatial scale on linear growth of perturbations. J. Atmos. Sci, under review.

Houtekamer, P. L., 1995: The construction of optimal perturbations. Mon. Wea. Rev., in press.

Houtekamer, P. L., and J. Derome, 1995: Methods for ensemble prediction. *Mon. Wea. Rev.*, in press.

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

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, S. Saha, G. White, J. Woolen, M. Chelliah, J. Janowiak, K. C. Mo, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne, E. Kung and D. Salstein, 1993: The NMC/NCAR CDAS/Reanalysis Project. NMC Office Note 401. Available: NOAA/NMC, World Weather Building, 5200 Auth Rd., Camp Springs, MD 20746.

Kalnay, E., and Z. Toth, 1994: Removing growing errors in the analysis. Proceedings of the Tenth Conference on Numerical Weather Prediction, July 18–22, 1994, Portland, OR. AMS, p. 212–215.

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.*, **110**, 701–721.

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

Lorenz, E. N., 1982: Atmospheric predictability experiments with a large numerical model. *Tellus*, **34**, 505–513.

Lorenz, E. N., 1984: Irregularity. A fundamental property of the atmosphere. *Tellus*, **36A**, 98–110.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1995: The ECMWF ensemble system: Methodology and validation. *Q. J. R. Meteorol. Soc.*, under review.

Nicolis, C., S. Vannitsem, and J.–F. Royer, 1995: Short range predictability of the atmosphere: Mechanisms for superexponential error growth. *Q. J. R. Meteorol. Soc.*, in print.

Oortwijn,, J, and J. Barkmeijer, 1995: Perturbations which optimally trigger weather regimes. J. Atmos. Sci., under review.

Palmer, T. N., and S. Tibaldi, 1988: On the prediction of forecast skill. Mon. Wea. Rev., 116, 2453–2480.

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

Rabier, F., E. Klinker, P. Courtier, and A. Hollingsworth, 1994: Sensitivity of two–day forecast errors over the Northern Hemisphere to initial conditions.ECMWF Research Department Technical Memorandum No. 203. Available from: ECMWF, Shienfield Park, Reading, Berkshire RG2 9AX, England.

Reynolds, C. A., P. J. Webster, and E. Kalnay, 1994: Random error growth in NMC's global forecasts. *Mon. Wea. Rev.*, **122**, 1281–1305.

Royer, J.–F., R. Stroe, and C. Nicolis, 1993: Error growth in a simple atmospheric model: Comparison between the logistic and Gompertz laws. *C. R. Acad. Sci. Paris*, **t. 316, Serie II**, 193–200 (In French with English summary.)

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

Shimada, I., and T. Nagashima., 1979: A numerical approach to ergodic problem of dissipative dynamical systems. *Progress of Theoretical Physics*, **61**, 1605–1615.

Smagorinsky, J., 1969: Problems and promises of deterministic extended range forecasting. *Bull. Am. Meteorol. Soc.*, **50**, 286–311.

Szunyogh, I., E. Kalnay, and Z. Toth, 1995: A comparison of Lyapunov vectors and optimal vectors in a low resolution GCM. J. Atmos. Sci., to be submitted.

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

Tracton, M. S., 1994: Operational ensemble prediction – The NMC experience. Preprints of the Tenth Conference on Numerical Weather Prediction, July 18–22, 1994, Portland, Oregon, p. 206–208. Available from: AMS, 45 Beacon Str., Boston, MA 02108–3693.

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

Tracton, M. S., K. Mo, W. Chen, E. Kalnay, R. Kistler, and G. White, 1989: Dynamical extended range forecasting (DERF) at the National Meteorological Center. *Mon. Wea. Rev.*, **117**, 2230–2247.

Trevisan, A., 1993: Impact of transient error growth on global average predictability measures. *J. Atmos. Sci.*, **50**, 1016–1028.

Trevisan, A., and R. Legnani, 1995: Transient error growth and local predictability: a study in the Lorenz system. *Tellus*, **A47**, 103–117.

Vannitsem, S., and C. Nicolis, 1994: Predictability experiments on a simplified thermal convection model: The role of spatial scales. *Journal of Geophysical Research*, **99D**, 10,377–10,385.

Vukicevic, T., and R. M. Errico, 1990: The influence of artificial and physical factors upon predictability estimates using a complex limited–area model. *Mon. Wea. Rev.*, **118**, 1460–1482.

Wobus, R. L., and E. Kalnay, 1995: Three years of operational prediction of forecast skill at NMC. *Mon. Wea. Rev.*, **123**, in press.

Wolf, A., J. B. Swift, H. L. Swinney, and J. A. Vastano, 1985: Determining Lyapunov exponents from a time series. *Physica*, **16D**, 285–317.

Zupanski, D. and F. Mesinger, 1995: Four–dimensional variational assimilation of precipitation data. *Mon. Wea. Rev.*, **123**, 1112–1127.

Zupanski, Z., 1995: An iterative approximation to the sensitivity analysis in nonlinear mathematical programming. Mon. Wea. Rev., under review.

Figure Captions

Fig. 1: Example indicating the gain from esemble averaging in a one–dimensional example using Lorenz' error growth equation. The solid curve is the error of the control forecast, the dashed curves are the error of the perturbed forecasts, and the dotted curve is the error of the ensemble mean. E(0) is the initial error of the control forecast and P is the amplitude of the twin perturbations.

Fig. 2: Daily amplification of bred perturbations with different initial perturbation sizes over the Northern (solid) and Southern (dashed) hemisphere extratropics, computed for the period February 23–27, 1992. The range of amplification factors for different random (Monte Carlo) balanced perturbations is shown as a vertical dotted line. Average amplification factors for difference fields between different long short–range forecasts verifying at the initial time of perturbed forecast integrations are also shown with a star (NH) and a plus sign (SH).

Fig.3: An example of bred perturbations at relatively small amplitudes: 500 hPa streamfunction perturbation on 15 February, 1992 with a perturbation amplitude of 0.015% total rms variance (equivalent to ~0.012m in 500 hPa height.). The perturbations at this amplitude are highly nonlinear and are primarily associated with convection.

Fig.4: 500 hPa streamfunction perturbation fields from three independent breeding cycles (with hemispherically constant rescaling) for 23 May 1992. The three cycles were started with independent initial perturbations six days earlier. The six marked boxes (see panel c) correspond to the areas considered in Table 1. Panels a, b, c correspond to breeding cycles br8, br12 and br17 in Table 1, respectively.

Fig. 5: Schematic of a self contained breeding pair of ensemble forecasts. Note that breeding is part of the extended ensemble forecasts and that the creation of efficient initial ensemble perturbations requires no additional computing resources beyond that needed to run the forecasts themselves.

Fig. 6: Relative regional uncertainty (for 500 hPa streamfunction) present in the control analysis as determined from the rms difference between two analyses from independently run NMC analysis cycles in April–May 1992. The analysis cycles were practically identical except that the initial first guesses differed slightly. The values shown are smoothed and the overall global mean is scaled to one.

Fig 7: Forecast skill (pattern anomaly correlation) of a 10–member ensemble mean (solid curve) as a function of lead time for the (a) Northern and (b) Southern extratropics and for the (c) tropics, for 1992 May 6 – June 14. The score for a single control forecast is also shown (dotted curve).

Fig. 8: 10–member, 9–day lead time ensemble forecast started on 30 May 1992. Shown are the 500 hPa streamfunction fields for (a) verifying analysis; (b) control forecast;

(c) ensemble mean forecast; (d) control error; (e) ensemble mean error; (f) as in (d) but in rms and smoothed; (g) forecast of the error (f): as in (e) but in rms and smoothed.

Fig. 9: Pattern correlation between predicted and actual error in the forecasts, averaged for the period 6 May–14 June 1992. For further details, see text.

Fig. 10: Time correlation of predicted and actual forecast errors, for 6 May–14 June 1992. The correlation values at the 0.1 and 0.001 statistical significance levels are 0.264 and 0.501, respectively. For further details, see text.

Fig. 11: Forecast skill (solid line), and evaluation of the prediction of the spatial distribution (dashed lines) and temporal variations (dotted line) in the forecast skill for 1992 May 6 – June 14, as a function of ensemble size, for the NH (a), SH (b), and for the tropics (c). For further details, see text.

Fig. 12: Schematic of the configuration of the operational ensemble forecasting system at NMC. Each horizontal line represents a numerical forecast. High resolution, T126 forecasts are marked with heavy lines while the other forecasts are run at T62 resolution. Note that at 00Z there are two control forecasts, one started at T126 resolution and then truncated to T62 at day 7, while the other started at a T62 truncated resolution. At 12Z, the high resolution control is truncated after 3 days of integration. Pairs of perturbed forecasts based on the breeding method are marked as B1–B7. For the extended range, forecasts originating in the most recent 48 hours are also used (and are shown on the figure.)

TABLE 1. Subjective comparison of perturbations from twenty independent breeding cycles, 1992 May 23. Regional modes in three areas over both the Northern and Southern Hemispheres, marked with boxes in Fig. 4c and numbered correspondingly from left to right, are compared. If a mode in another cycle is very similar to that in cycle 17, a plus or minus sign appears, depending on the sign of the mode.

Table 1

	NH #1	NH #2	NH #3		SH #1	SH #2	SH #3
br 1			+		−	−	
br 2	−	+	+		−	+	+
br 3	+	−	−				+
br 4		+	+			−	
br 5			−		+	−	
br 6	−						
br 7			−		+	+	
br 8	+		+		−	+	+
br 9		−	−		+	+	
br 10	−	−	−			−	
br 11	−	−	−		+	+	+
br 12	−	+	+				+
br 13	+		+				+
br 14					−		−
br 15	+	−			−		+
br 16	+					−	−
br 17	+	+	+		+	+	+
br 18	−				+	+	
br 19		+	+				
br 20		+			−		+
Total	12	11	14		11	12	11

TABLE 2. Comparison of ensembles generated by single bred perturbations (i. e. centering individual perturbations on control analysis, singles) and those generated by positive–negative pairs of perturbations (pairs) for 1992 May 23–28, with 5/10% initial perturbations for the NH/SH. (a) PAC skill scores at day 6; (b) Comparative verification as a function of lead time, NH/SH combined.

a)

	Control	Singles	Pairs
NH	.680	.687	.692
SH	.510	.536	.552
NH/SH Combined	.595	.611	.622

b)

LEAD TIME (days)	PAIRS–SINGLES (WINS)
1	8 – 4
3	7 – 5
4	7 – 5
5	9 – 3
6	10 – 2

TABLE 3. The effect of the size of initial perturbations on the performence of 10–member ensembles for 1992 May 23 – June 6. At the different lead times, PAC scores are computed for the mean forecast from different ensembles. Shown is the relative performance of each perturbation size with respect to 10% (NH, panel a) and 20% (SH, panel b) perturbation size, in terms of PAC wins vs. losses (W/L) and average improvement (AI). The best results are highlighted in bold.

a) NORTHERN HEMISPHERE

PERTURB. SIZE (% rms variance)	DAY 3 W/L	AI	DAY 6 W/L	AI	DAY 9 W/L	AI
5	5 – 6	–	6 – 6	–	4 – 8	–
7.5	**7 – 5**	+	**7 – 5**	/	4 – 8	–
12.5	2 – 9	–	**5 – 7**	+	**6 –6**	+
15	2 –10	–	4 – 8	–	4 – 8	–
20	2 –10	–	1 –11	–	6 – 6	–

a) SOUTHERN HEMISPHERE

PERTURB. SIZE (% rms variance)	DAY 3 W/L	AI	DAY 6 W/L	AI	DAY 9 W/L	AI
10	2 –10	–	5 – 7	–	3 – 9	–
15	4 – 8	–	6 – 6	–	3 – 9	–
25	**7 – 5**	+	**8 – 4**	+	8 – 4	+
30	7 – 5	+	**7 – 4**	+	**7 – 5**	+
40	3 – 9	–	5 – 7	–	8 – 4	+

TABLE 4. The effect of optimal spatial smoothing on the control and 10–member ensemble mean forecasts for the period 1992 May 23 – June 03, with 10/20% initial perturbations for the NH/SH. For further details, see text.

LEAD TIME (Days)	OPTIMAL SMOOTHING (~Triangular truncation)		ENSEMBLE ADVANTAGE OVER CONTROL RETAINED	
	CONTROL	ENSEMBLE	PAC	% Total
5	T30	T40	.02	62.5
7	T25	T35	.033	63.8
9	T20	T30	.042	60.5

TABLE 5. Comparison of the control forecasts and 10–member, randomly generated and bred ensembles for 1992 May 23 – June 6, 10/20% initial rms amplitude, NH and SH results combined, for days 5 (D5) and 9 (D9) lead times. PAC skill score for T126 control is estimated based on average difference between high and low resolution controls for three years. For further details, see text.

FORECAST	FCST SKILL (PAC)		SPATIAL ERROR FCST SKILL (PAC)	
	D5	D9	D5	D9
Control, T62	.659	.354	–	–
Control, T126	.677	–	–	–
Random	.680	.404	.399	.373
Bred	.691	.424	.480	.407

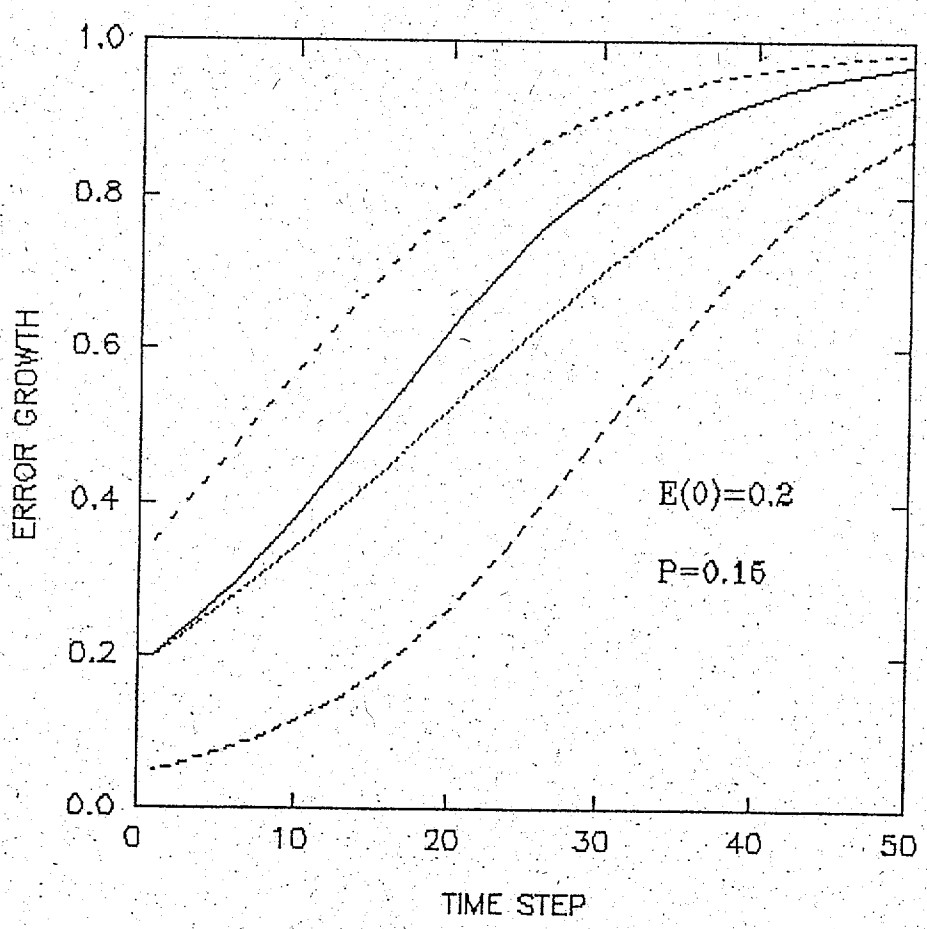

1.0
0.8
0.6
0.4
0.2
0.0
ERROR GROWTH
0
10
20
30
40
50
TIME STEP
E(0)=0.2
P=0.15

Fig. 1

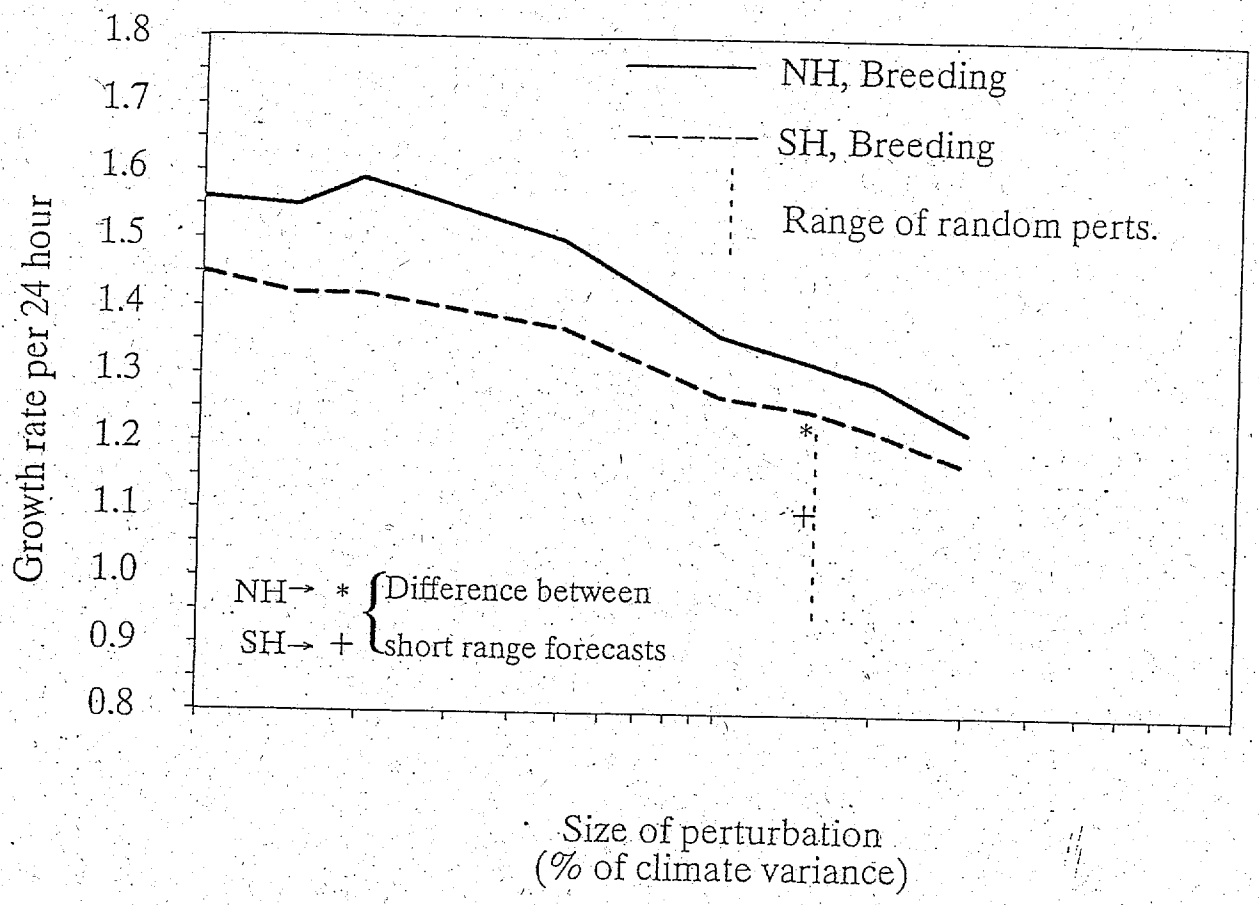

1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
0.8
Growth rate per 24 hour
NH, Breeding
SH, Breeding
Range of random perts.
NH→ *
SH→ +
Difference between
short range forecasts
Size of perturbation
(% of climate variance)

Fig. 2

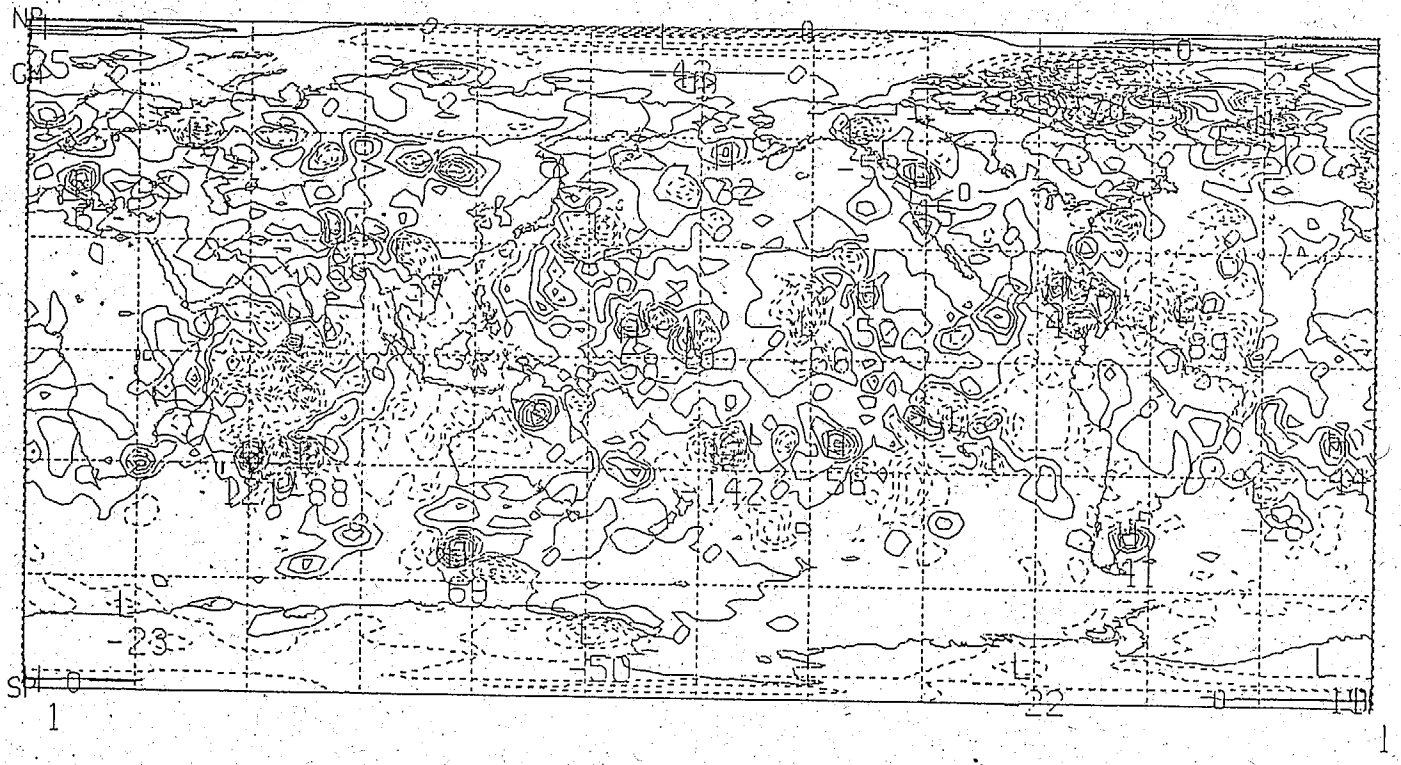


Fig. 3

CONTOUR FROM -0.14000E-01 TO 0.12000E-01 CONTOUR INTERVAL OF 0.10000E-02 PT(3,3)= -0.10974E-02 LABELS SCALED BY 10000.

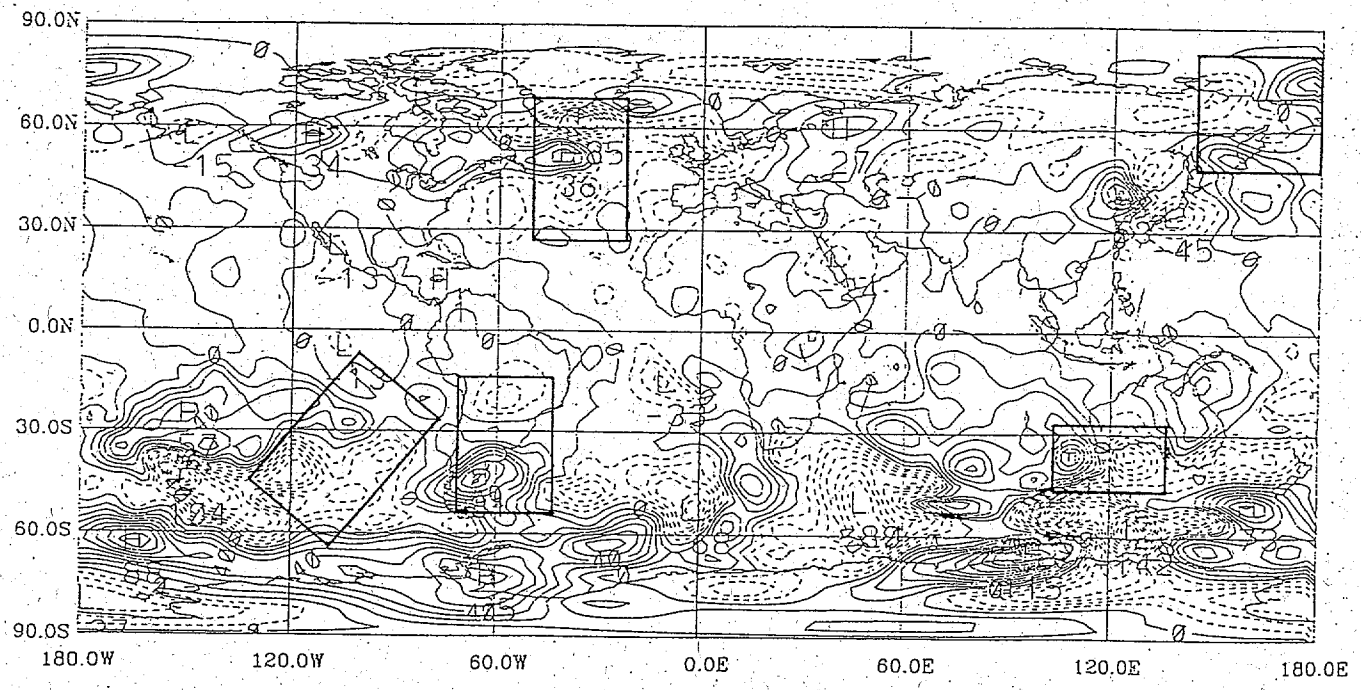


Fig. 4a

CONTOUR FROM -0.14000E+08 TO 0.11000E+08 CONTOUR INTERVAL OF 0.10000E+07 PT(3,3)= -0.21905E+07 LABELS SCALED BY 0.10000E-04

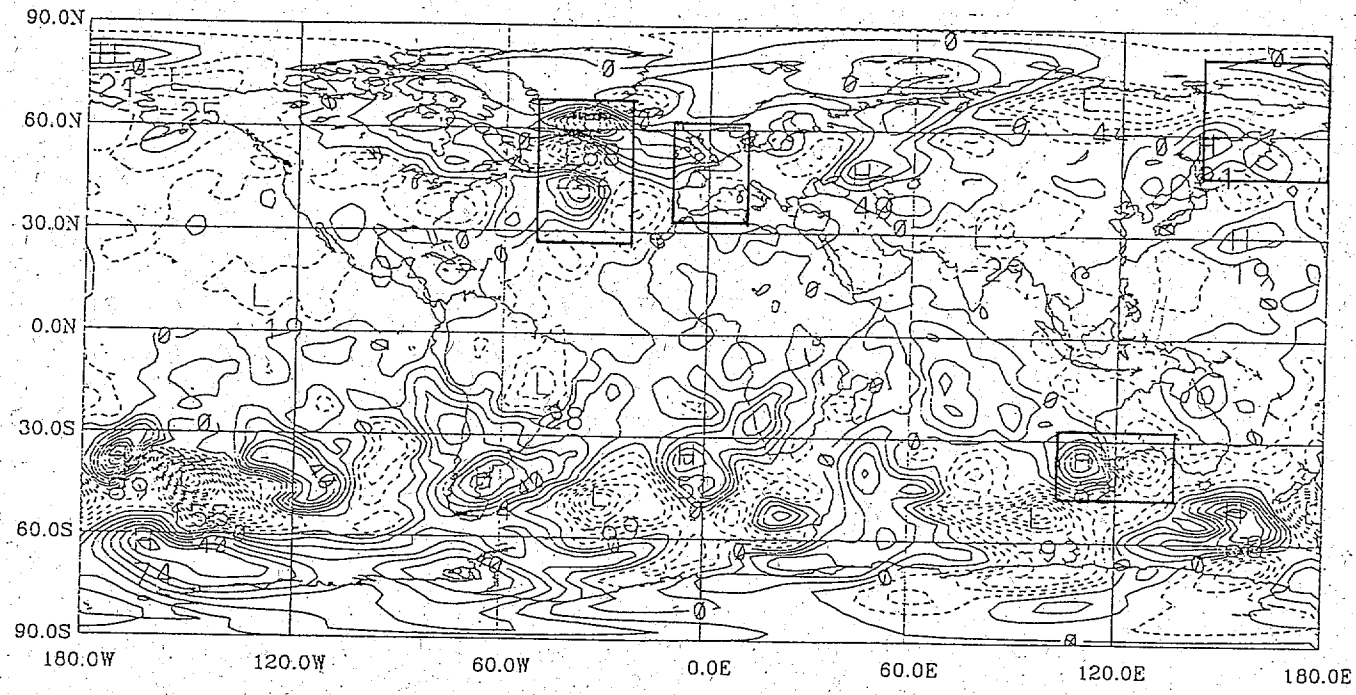


Fig. 46

CONTOUR FROM -0.15000E+08 TO 0.80000E+07 CONTOUR INTERVAL OF 0.10000E+07 PT(3,3)= -0.52458E+06 LABELS SCALED BY 0.10000E-04

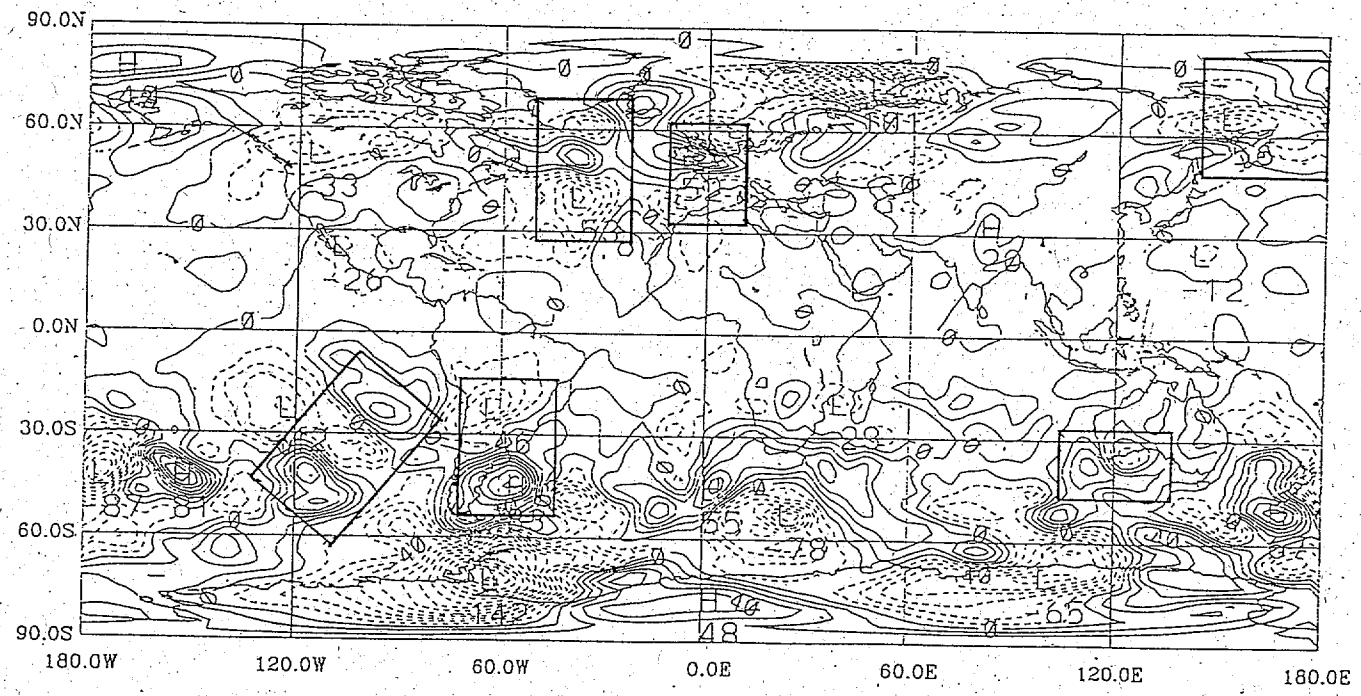
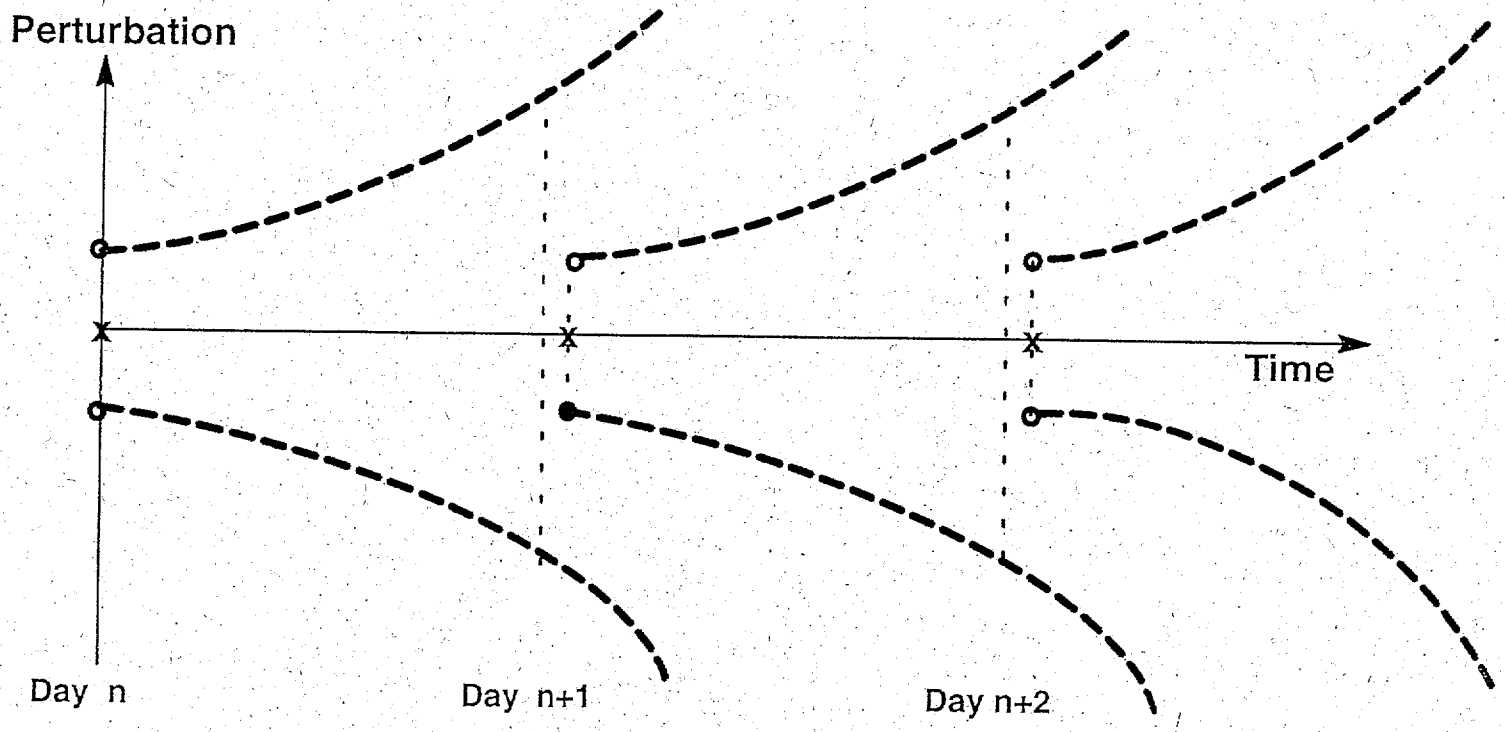


Fig. 4C

CONTOUR FROM -0.14000E+08 TO 0.80000E+07 CONTOUR INTERVAL OF 0.10000E+07 PT(3,3)= 0.33442E+07 LABELS SCALED BY 0.10000E-04

SELF–BREEDING OF TWIN FORECASTS



x Analysis

o Perturbed Initial Condition

Perturbed forecast

Bred Growing Mode

Fig. 5

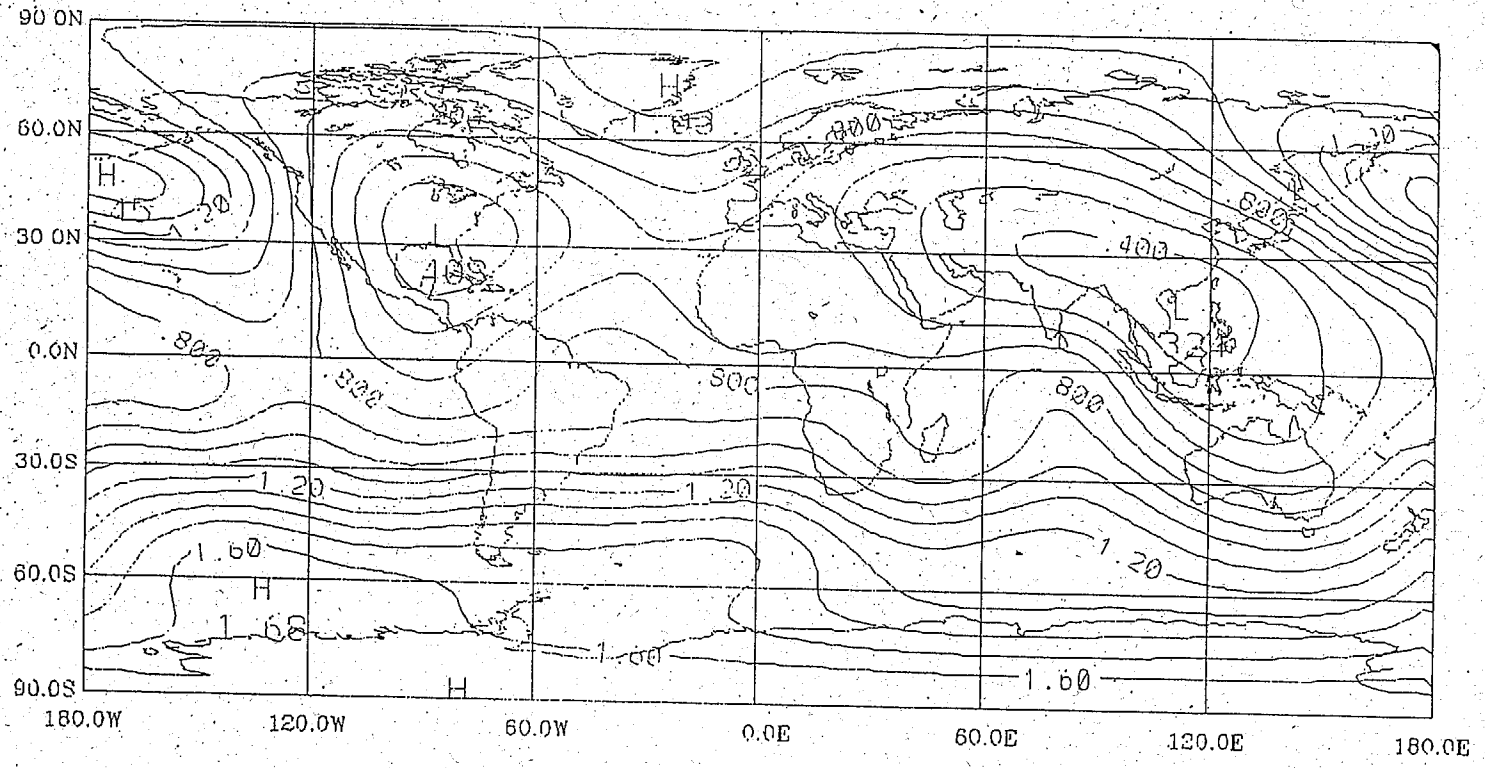


Fig. 6

CONTOUR FROM 0.00000E+00 TO 1.0000 CONTOUR INTERVAL OF 0.10000 PT(3,3)= 1.0588

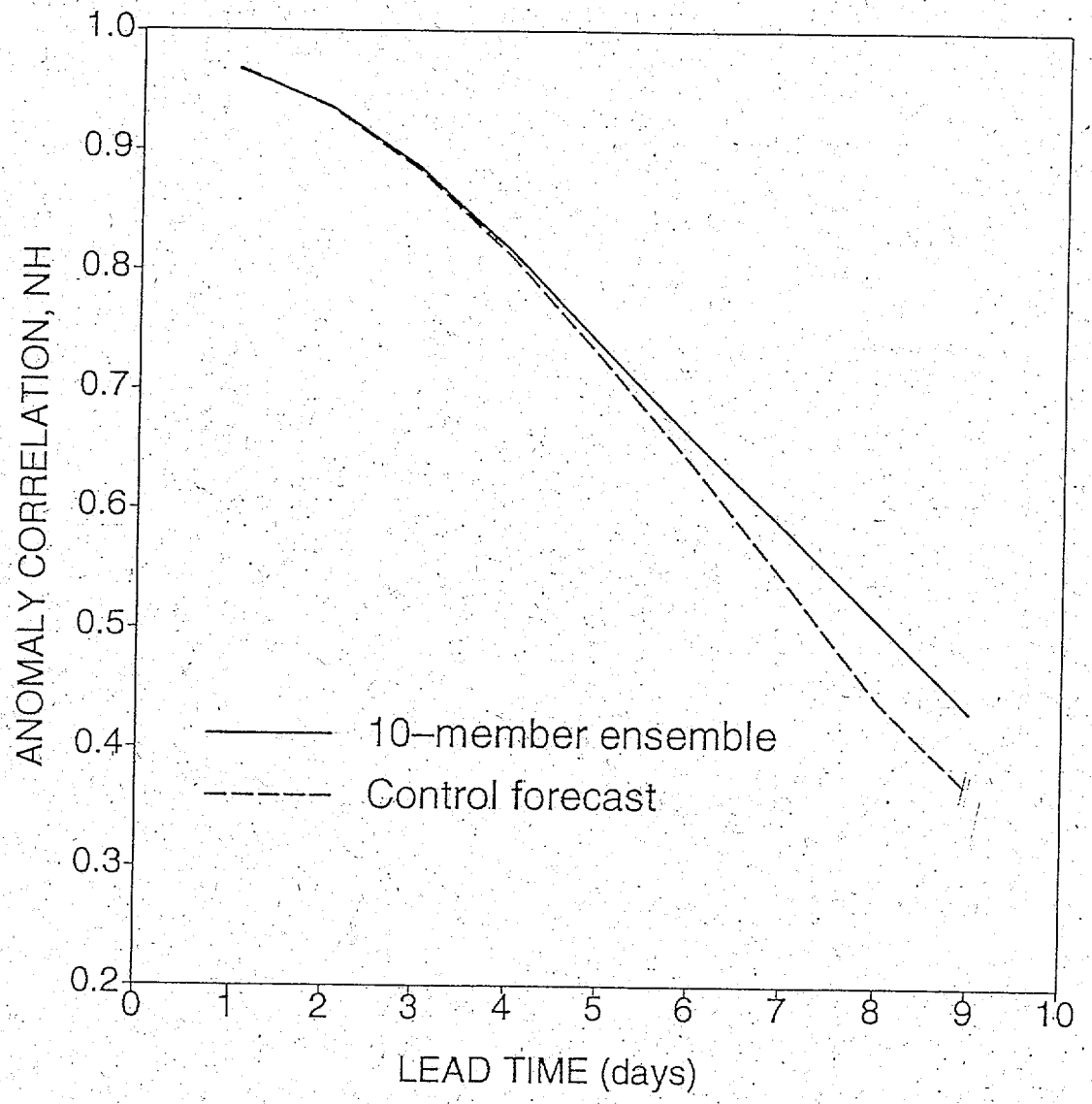


Fig. 7.a

Fig 7: Forecast skill (pattern anomaly correlation) of a 10–member ensemble mean (solid curve) as a function of lead time for the (a) Northern and (b) Southern extratropics and for the (c) tropics, for 1992 May 6 – June 14. The score for a single control forecast is also shown (dotted curve).

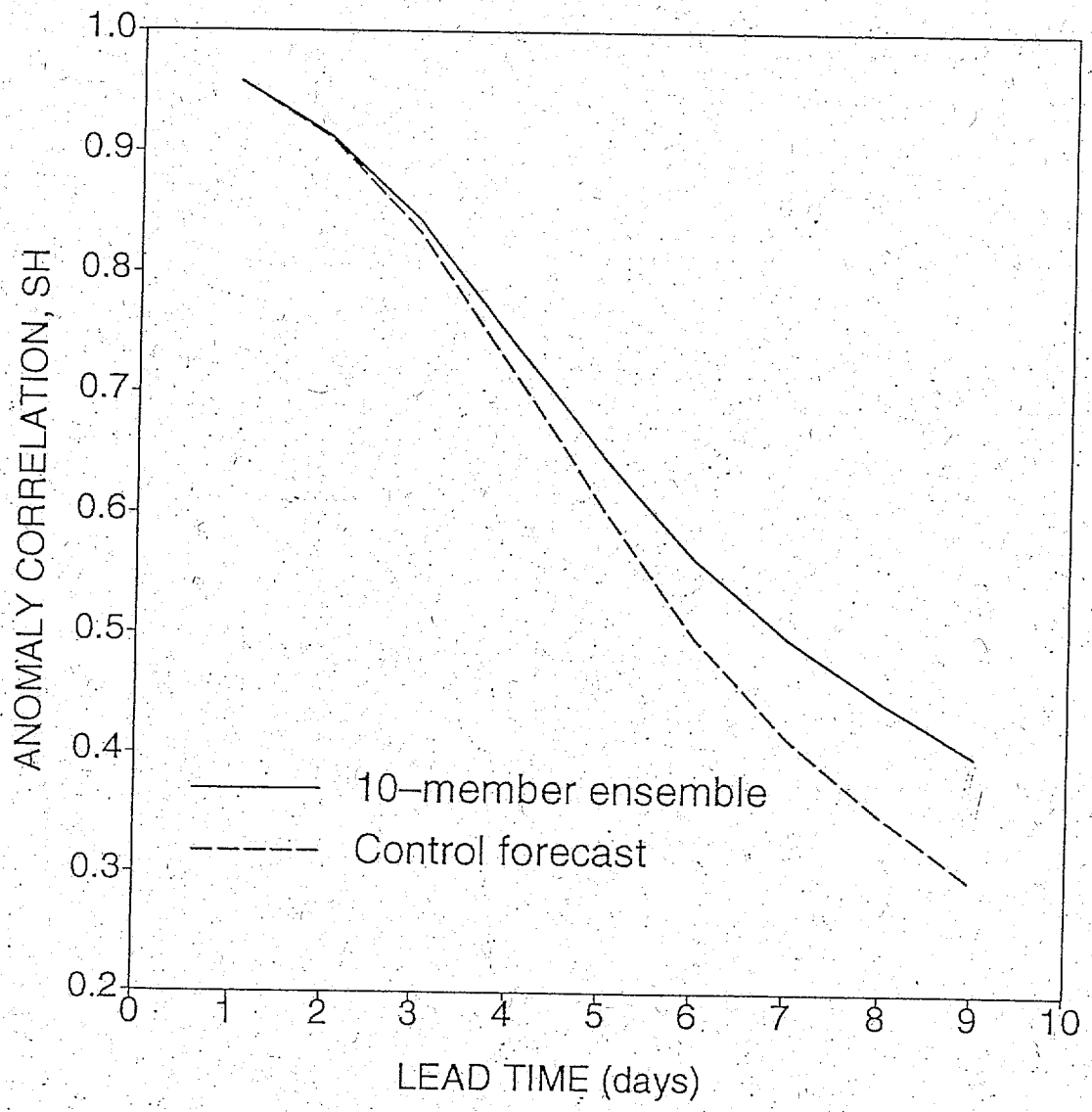

1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
ANOMALY CORRELATION, SH
10–member ensemble
Control forecast
0
1
2
3
4
5
6
7
8
9
10
LEAD TIME (days)

Fig. 7.b

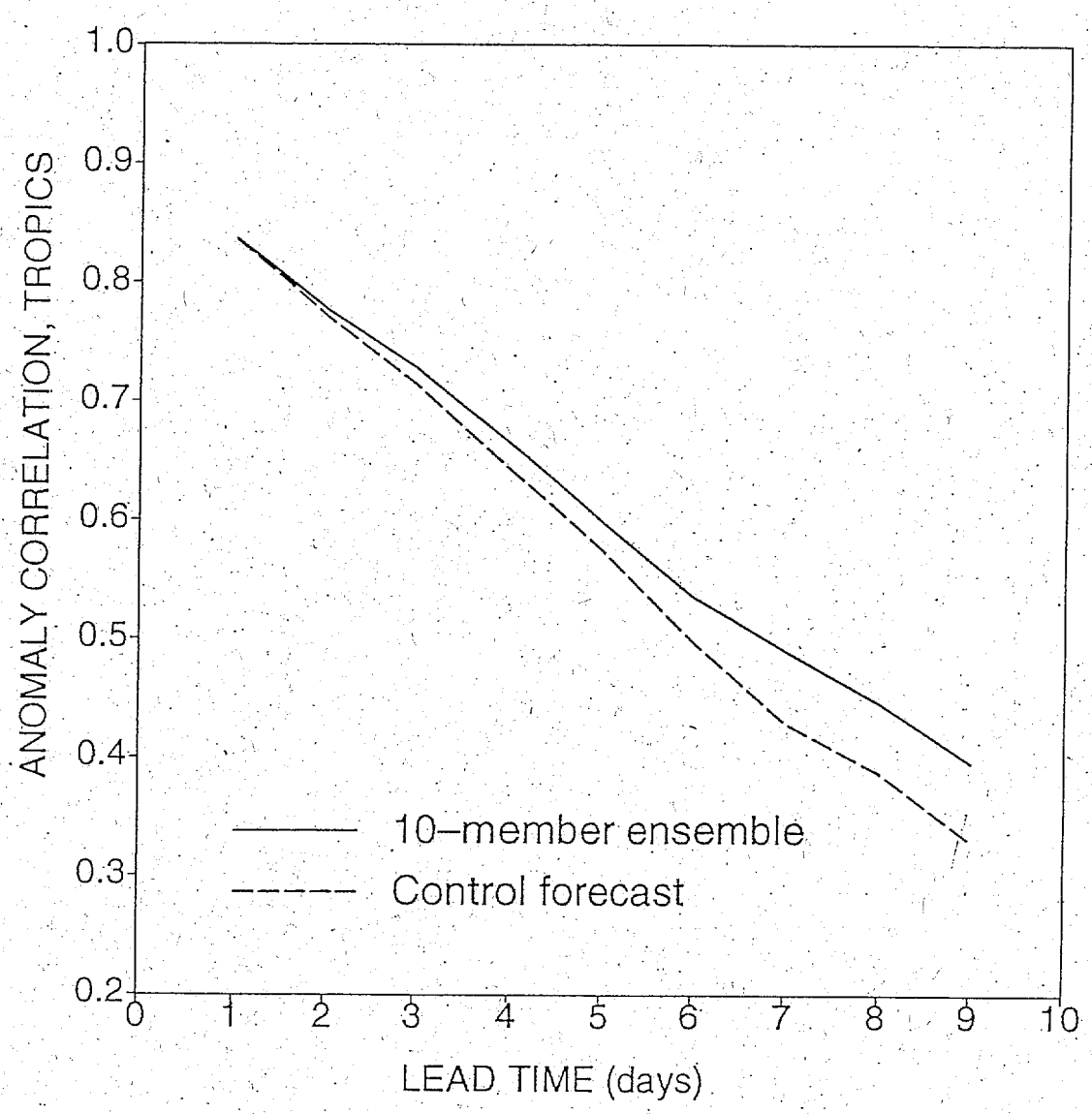

ANOMALY CORRELATION, TROPICS
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0
1
2
3
4
5
6
7
8
9
10
LEAD TIME (days)
10–member ensemble
Control forecast

Fig. 7c

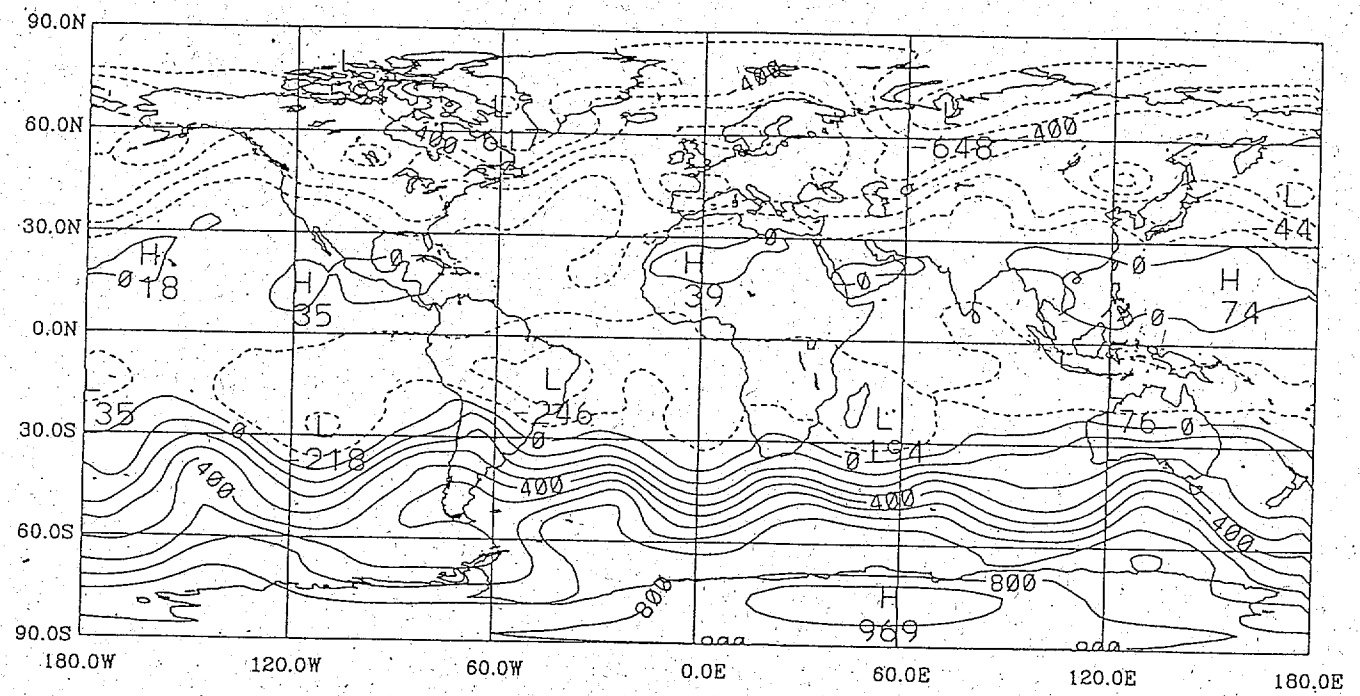


Fig. 8a

CONTOUR FROM -0.60000E+08 TO 0.90000E+08 CONTOUR INTERVAL OF 0.10000E+08 PT(3,3)= 0.77932E+08 LABELS SCALED BY 0.10000E-04

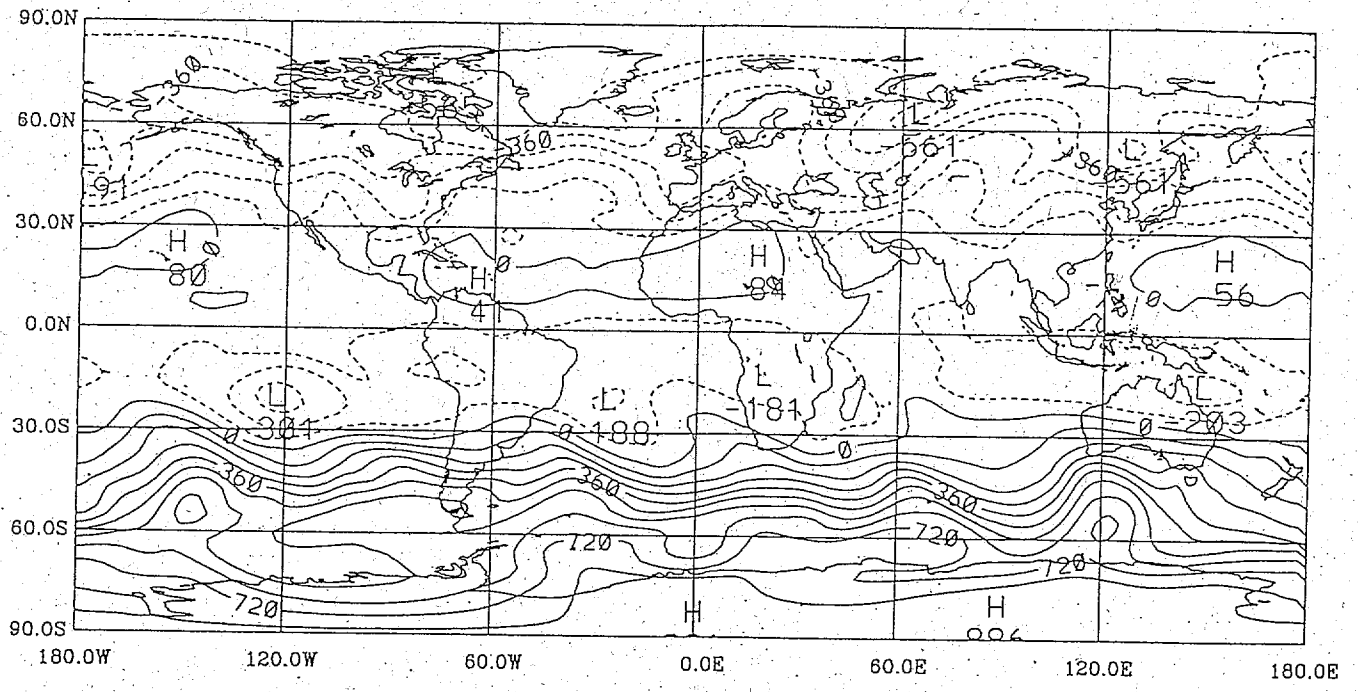


Fig. 86

CONTOUR FROM -0.54000E+08 TO 0.81000E+08 CONTOUR INTERVAL OF 0.90000E+07 PT(3,3)= 0.81993E+08 LABELS SCALED BY 0.10000E-04

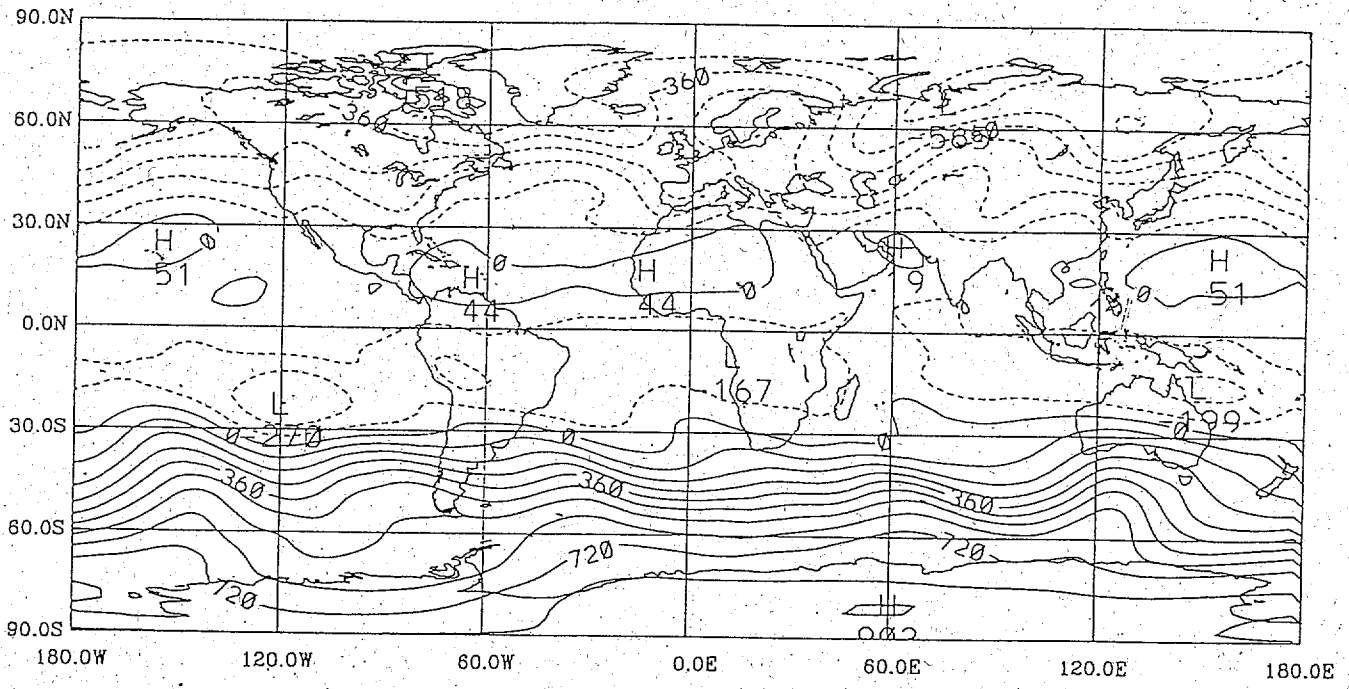


Fig. 8c

CONTOUR FROM -0.54000E+08 TO 0.90000E+08 CONTOUR INTERVAL OF 0.90000E+07 PT(3,3)= 0.79248E+08 LABELS SCALED BY 0.10000E-04

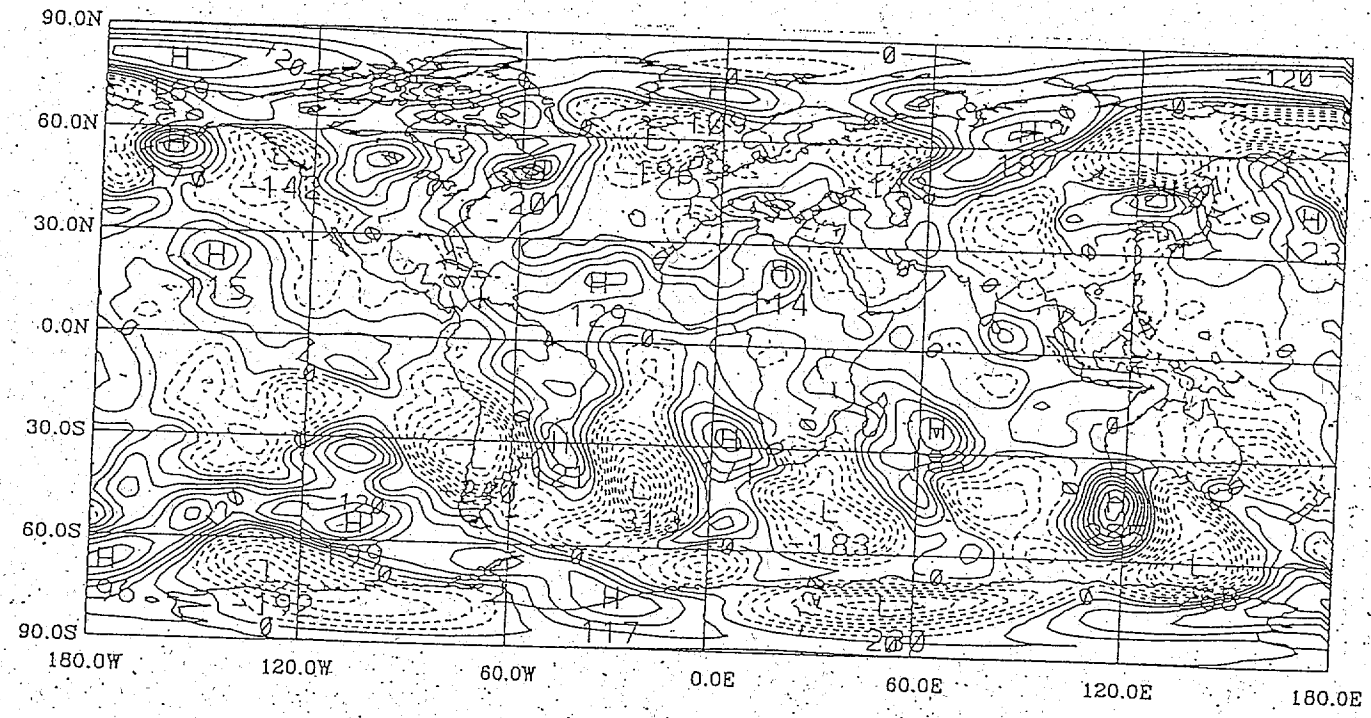


Fig. 8d

CONTOUR FROM -0.30000E+08 TO 0.24000E+08 CONTOUR INTERVAL OF 0.30000E+07 PT(3,3)= 0.40617E+07 LABELS SCALED BY 0.10000E-04

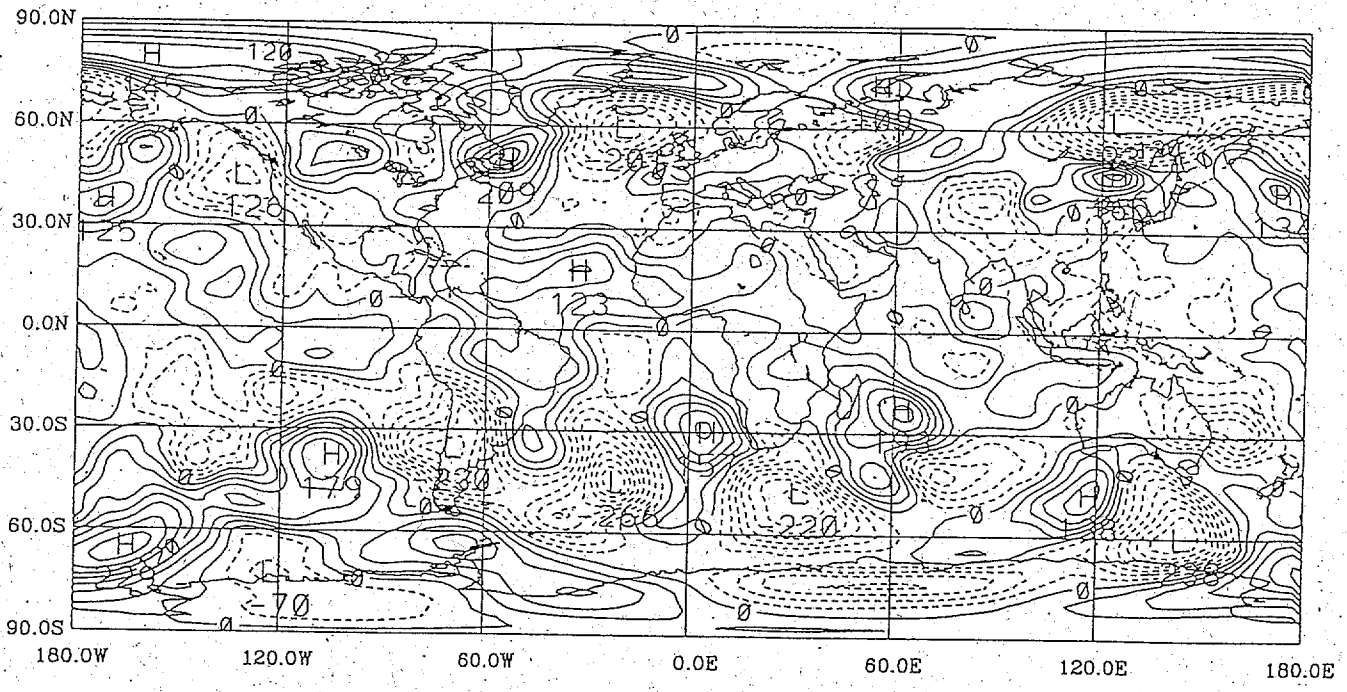


Fig. 8e

CONTOUR FROM -0.24000E+08 TO 0.21000E+08 CONTOUR INTERVAL OF 0.30000E+07 PT(3,3)= 0.13159E+07 LABELS SCALED BY 0.10000E-04

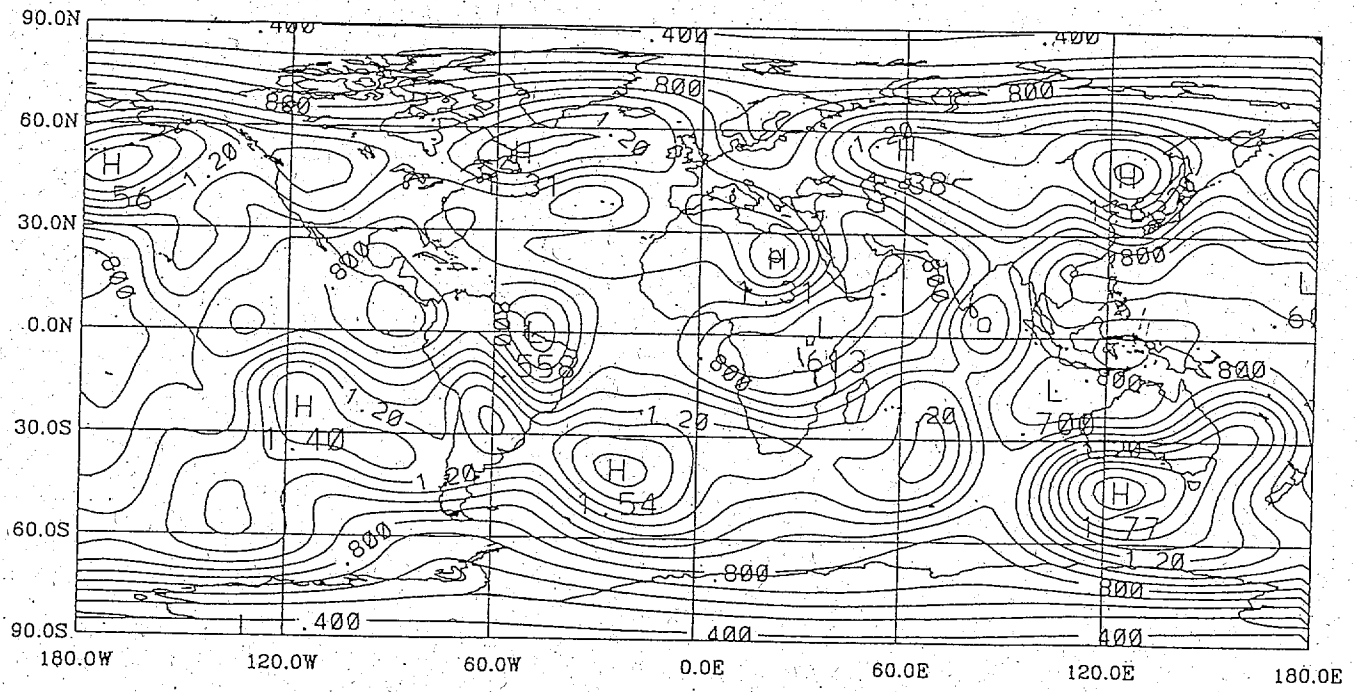


Fig. 8f

CONTOUR FROM 0.00000E+00 TO 1.7000 CONTOUR INTERVAL OF 0.10000 PT(3,3)= 0.40648

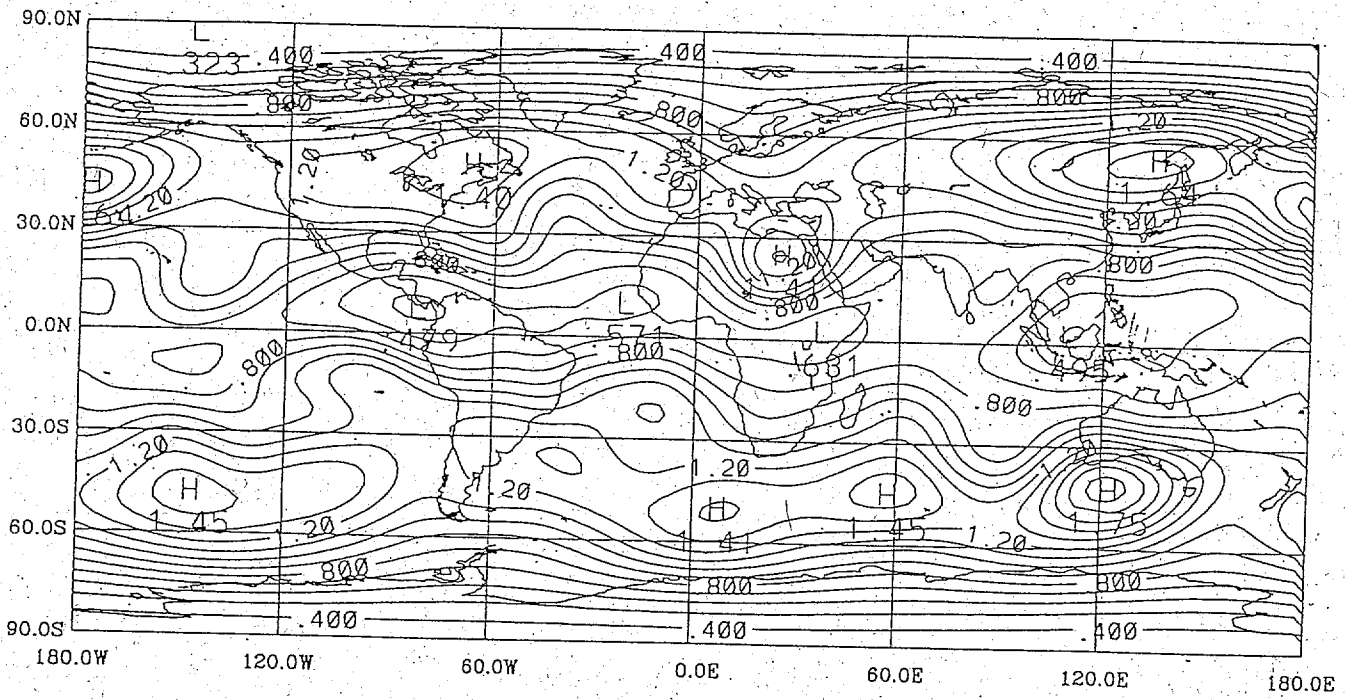


Fig. 89

CONTOUR FROM 0.00000E+00 TO 1.7000 CONTOUR INTERVAL OF 0.10000 PT(3,3)= 0.37854

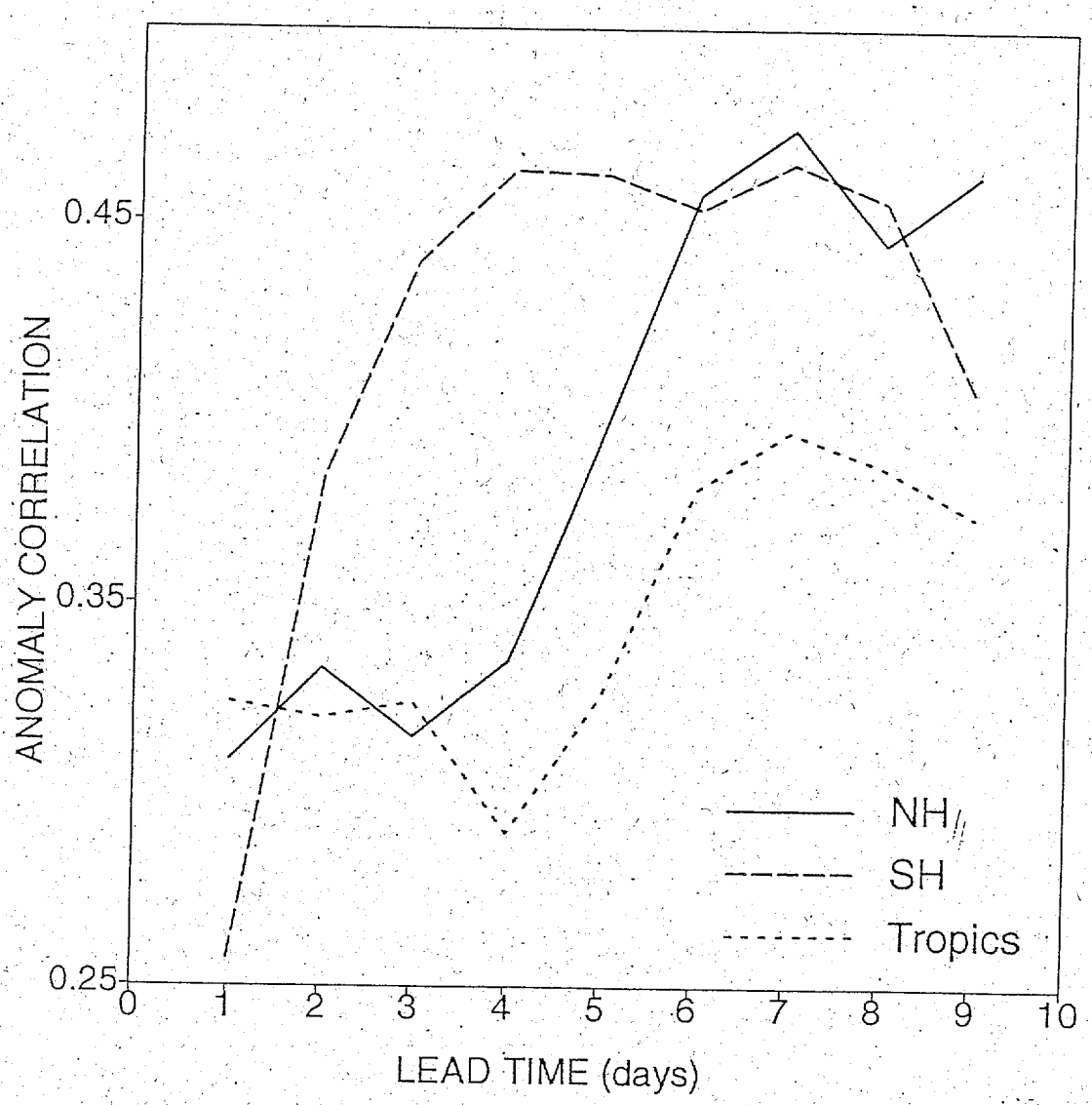


Fig. 9

Fig. 9: Pattern correlation between predicted and actual error in the forecasts, averaged for the period 6 May–14 June 1992. For further details, see text.

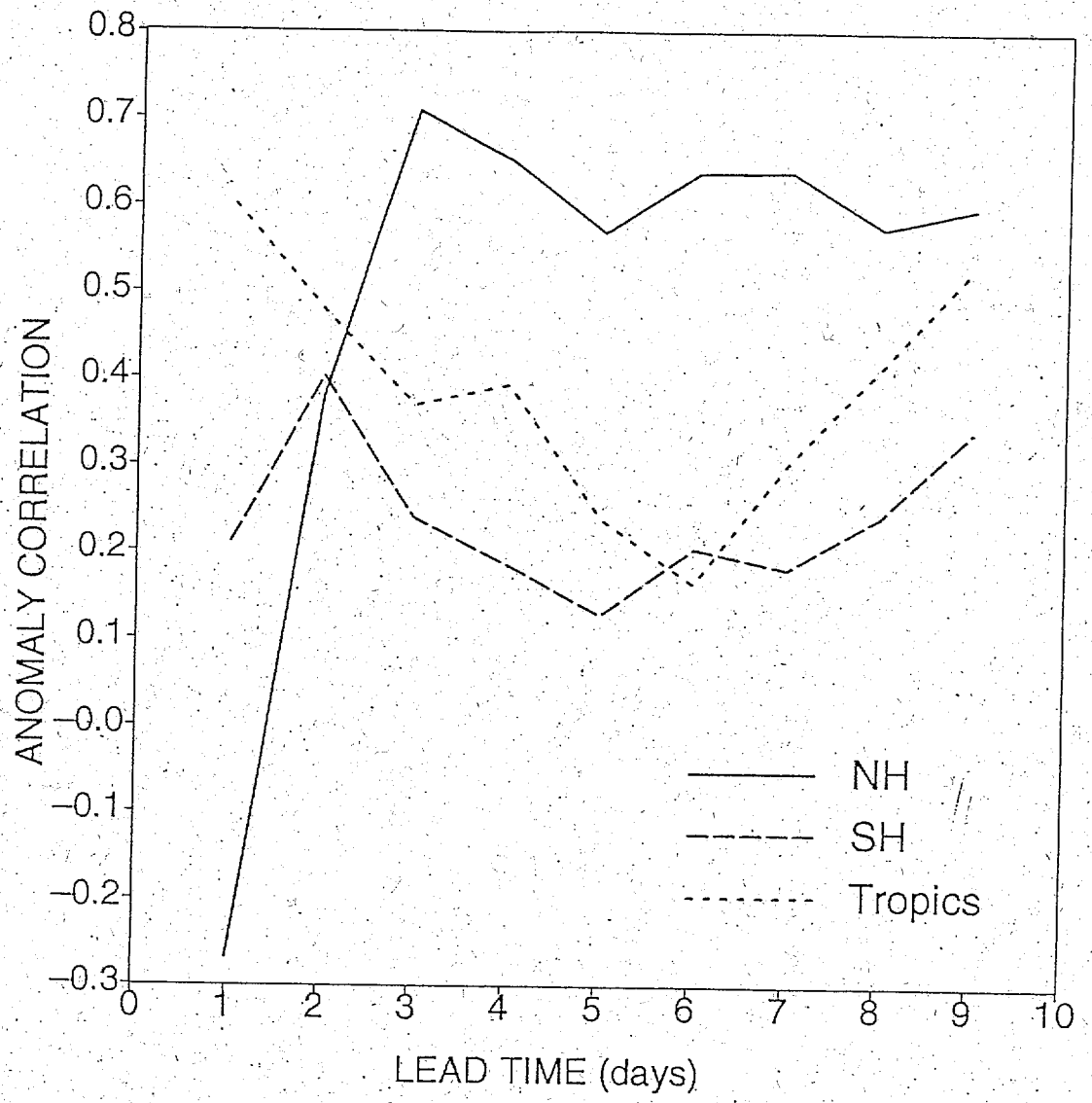


Fig. 10

Fig. 10: Time correlation of predicted and actual forecast errors, for 6 May–14 June 1992. The correlation values at the 0.1 and 0.001 statistical significance levels are 0.264 and 0.501, respectively. For further details, see text.

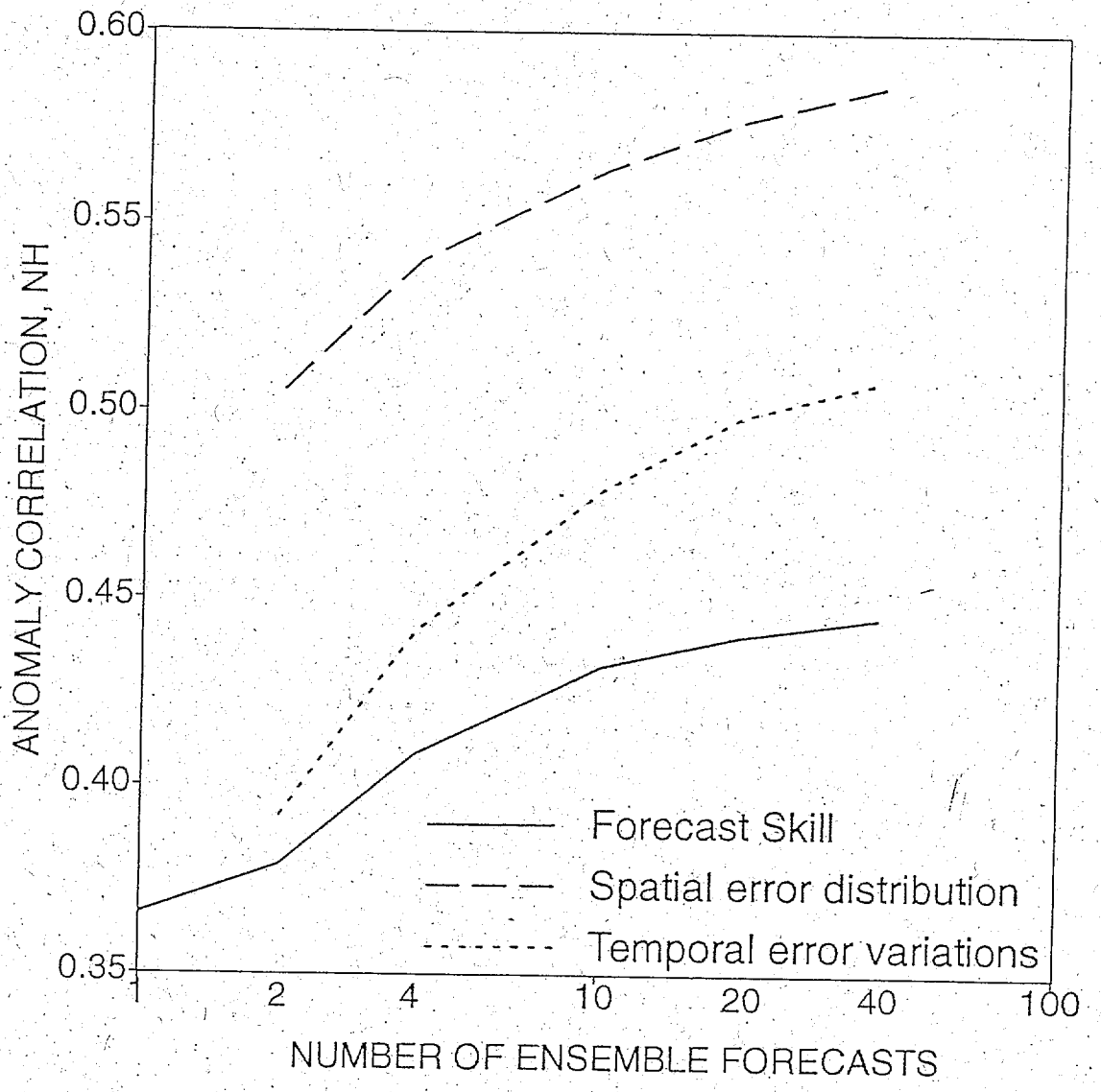


Fig. 11.a

Fig. 11: Forecast skill (solid line), and evaluation of the prediction of the spatial distribution (dashed lines) and temporal variations (dotted line) in the forecast skill for 1992 May 6 – June 14, as a function of ensemble size, for the NH (a), for the SH (b) and for the tropics (c). For further details, see text.

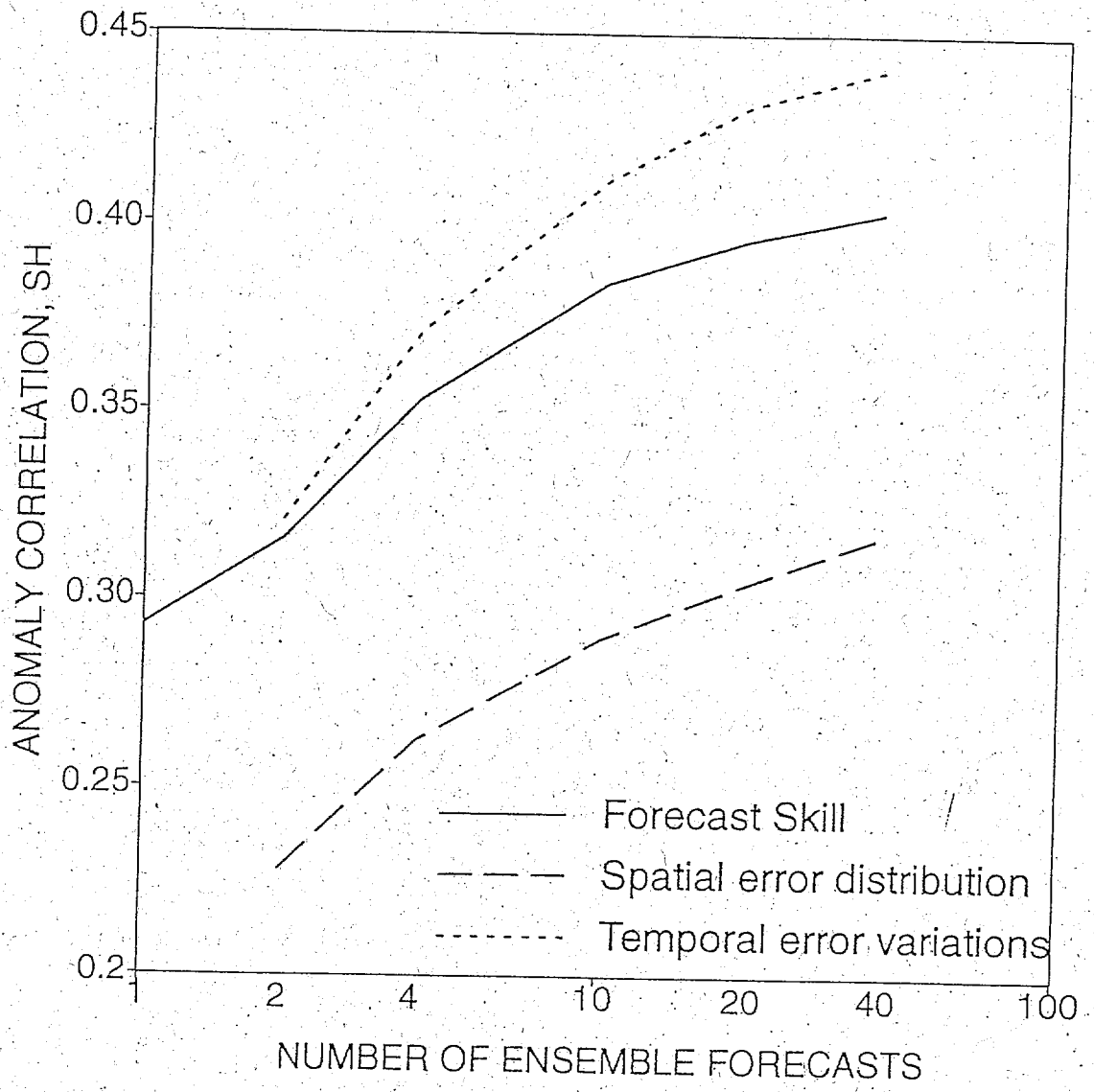

0.45
0.40
0.35
0.30
0.25
0.2
ANOMALY CORRELATION, SH
Forecast Skill
Spatial error distribution
Temporal error variations
2
4
10
20
40
100
NUMBER OF ENSEMBLE FORECASTS

Fig. 11.b

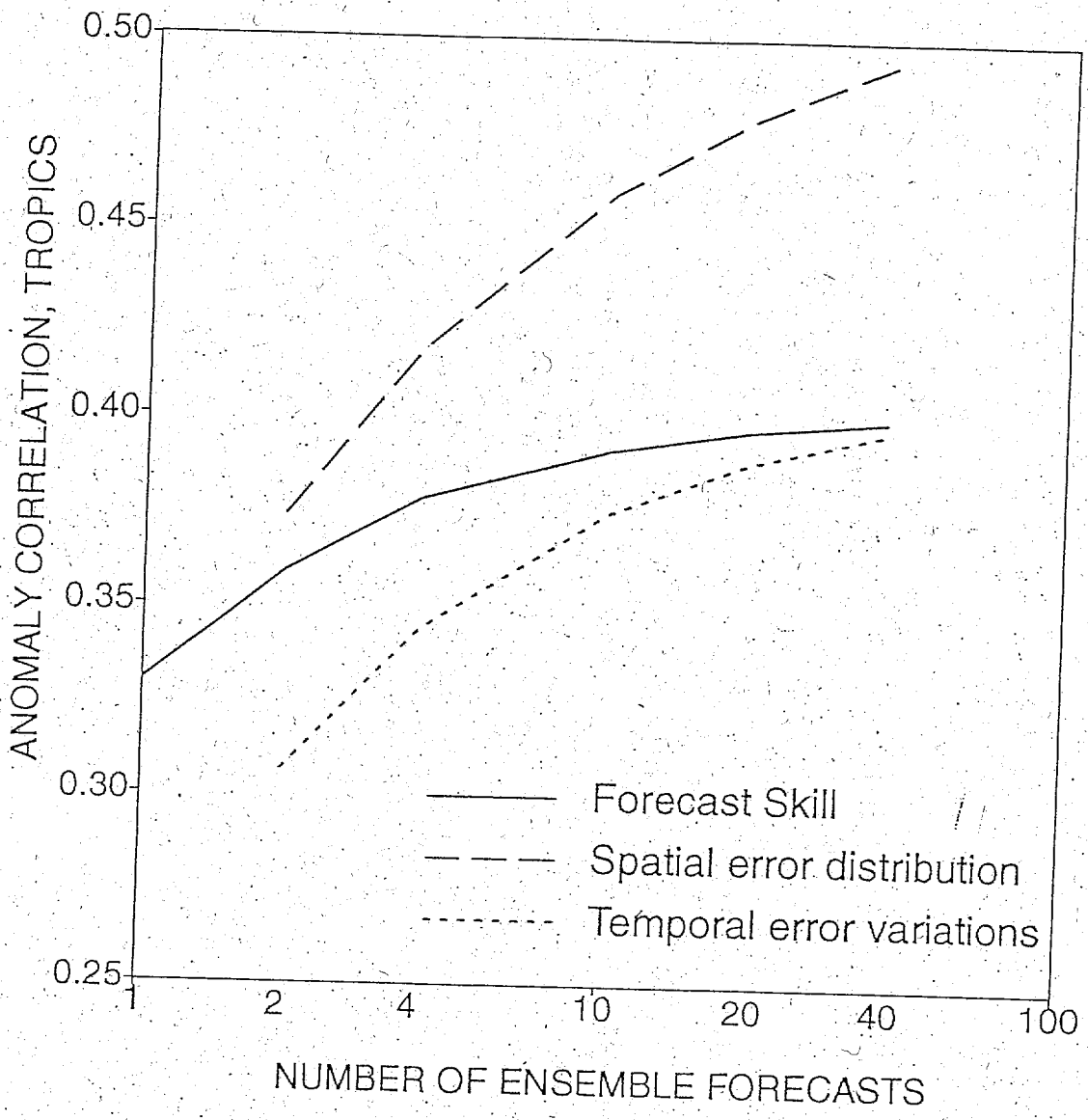
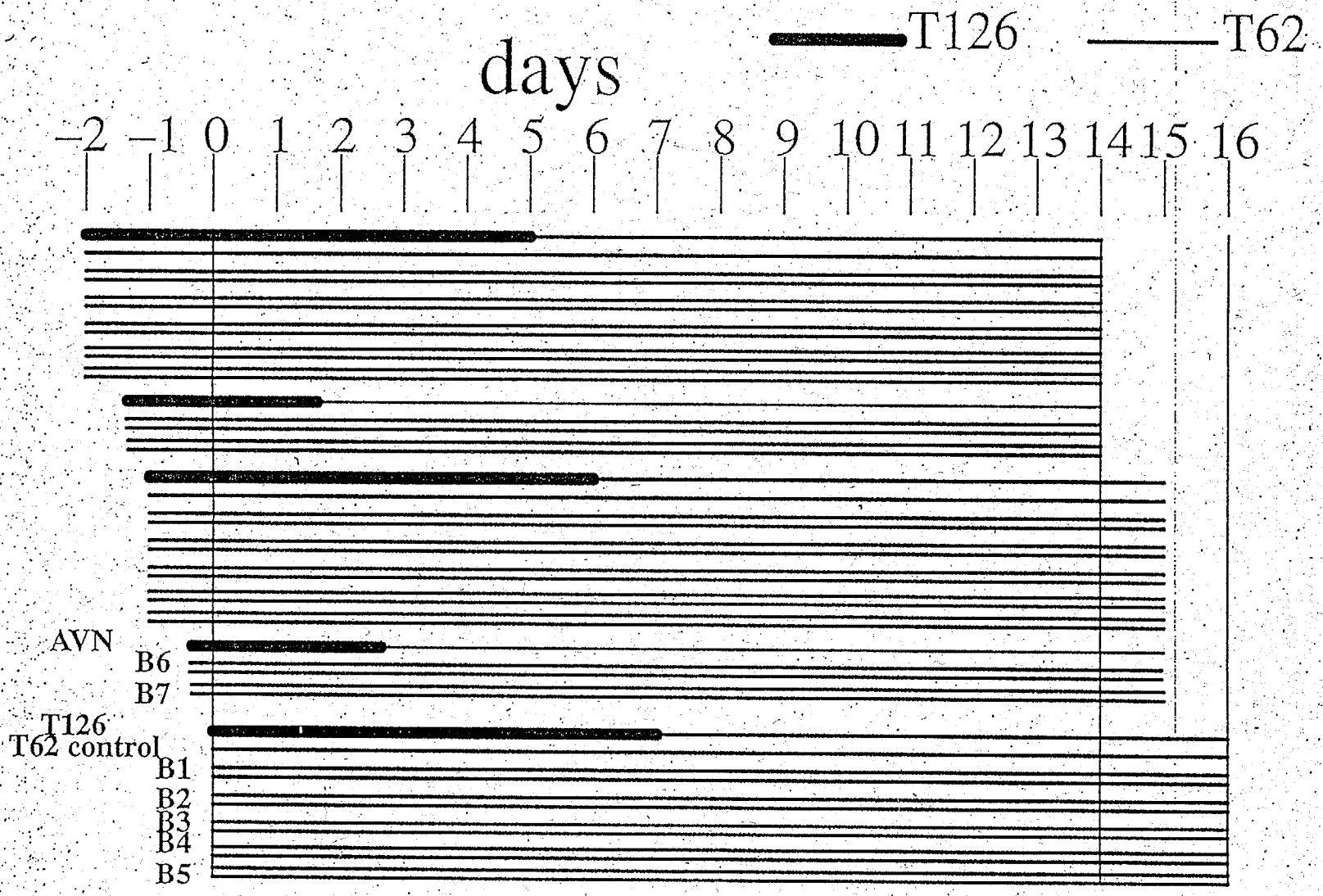

0.50
0.45
0.40
0.35
0.30
0.25
ANOMALY CORRELATION, TROPICS
1
2
4
10
20
40
100
NUMBER OF ENSEMBLE FORECASTS
Forecast Skill
Spatial error distribution
Temporal error variations

Fig. 11.c



1994 Ensemble Configuration

Fig. 12