Similarity of ‘fast-growing perturbations’ and an illustrative experiment with ensemble forecasting

David Noone and Ian Simmonds
School of Earth Sciences, University of Melbourne, Australia
(Manuscript received November 1996; revised October 1997)

We address an aspect of the role that ensemble forecasting may play in five-day forecasts both globally and over the Australian region. A central question that arises in this topic is how to choose the most useful perturbations from which to run the ensemble.

The philosophy adopted here is that the ‘best’ perturbations are those that are representative of the analysis errors and that project onto growing synoptic modes. Such perturbation modes are found using a method designed to ‘breed’ a perturbation that is representative of errors introduced in an analysis cycle. These fast-growing modes (FGMs) are discussed in terms of both synoptic variability and within a theoretical framework of the dynamics of initial uncertainty. The use of pattern correlation and empirical orthogonal function (EOF) analysis to compare FGMs in an ensemble is seen to give an indication of the number of regional modes sampled in the FGM ensemble. In this paper we also wish to quantify the sensitivity of the structure of the FGMs to the nature of the ‘seeding perturbations’ and to the synoptic patterns obtained during the period of their generation. Statistically significant improvements have been seen as a reduction in root mean squared forecast error of three per cent globally and four per cent in the Australian region by using an averaged, ensemble ‘forecast’ with only two FGMs as member perturbations.

Introduction

In a simple numerical convection model with three degrees of freedom, Lorenz (1963) observed evolution that was highly dependent on the initial conditions. He found that if two runs are started with only tiny initial differences, the solutions will eventually diverge and will bear no more resemblance to each other than if they had been randomly selected. Vastly more complex atmospheric models, in particular general circulation models (GCMs), display such chaotic behaviour. Behaviour of this nature is of obvious concern in meteorological prognosis, as the observed initial state for numerical weather prediction (NWP) (i.e. ‘today’s’ atmosphere around the globe) can never be determined with infinite certainty. Hence, the range of atmospheric predictability is limited.

In view of such a limitation on deterministic forecasting, it is believed that forecast skill can be improved by use of ensemble forecasting techniques to allow consideration of possible initial uncertainties (e.g. Leith 1974). A decision must be made in ensemble forecasting as to what statistical properties should be prescribed for such uncertainties. In models with a small number of equations (e.g. Lorenz 1965), ensembles can be created with each model variable being perturbed separately to represent the error variance in the initial conditions. In modern weather prediction, such an ensemble would be impractical as NWP models can have more than $10^6$ degrees of freedom. It is thought, however, that most of

Corresponding author address: David Noone, School of Earth Sciences, University of Melbourne, Parkville, Vic. 3052, Australia.
Ideal choices for ensemble perturbations are therefore those which are representative of the analysis errors and that project onto the growing synoptic modes.

The identification of the fastest growing error modes in the analysis cycle is analogous to choosing ensemble members perturbed along the longest axis of the ellipsoid of uncertainty in phase space described by Lorenz (1965). Thus, adding and subtracting the fast-growing perturbations from a best-guess state would represent the ellipsoid with a minimum number of ensemble members (i.e. two). Further, a reasonably large divergence of members immediately after the commencement of model integration could be expected. This is in fact a necessary condition to be considered in designing a method for perturbation generation. The method for breeding fastest growing modes as described by Toth and Kalnay (1993) finds such perturbations.

Briefly stated, the method involves adding a small multivariate random perturbation field to the atmospheric analysis. The forecast model being used is then integrated from both the perturbed and unperturbed (control) state until the next time for which an atmospheric state is known (12 hours in our case). The difference between these two forecasts is found and scaled to have the same (spatial) root mean square (rms) value as the initial random perturbation. This new, scaled perturbation is added to the following analysis and again a control and a perturbed forecast are made for 12 hours. The process is repeated until the rate of error growth has reached saturation. Typically, this is at around three to four days after breeding has commenced. The mode can then be considered fully developed. Any initial projection of analysis errors onto non-growing modes can be regarded as having a negligible contribution to expected forecast errors assuming the growing modes in the model match those of the atmosphere. It had been thought that the precise nature of the initial perturbation is of little consequence to the properties of the FGM. Through use of the GCM, this perturbation can be assumed dynamically balanced, suggesting that an ensemble will diverge immediately when the model forecast is started. Scaling the perturbation repetitively at each step of the breeding causes the ratio between the non-growing and growing errors to become smaller and therefore makes the perturbation strongly biased toward the fast-growing meteorologically important modes. Thus, the final bred perturbation contains, by definition, the modes that have grown fastest during the analysis cycle and are most unstable. The error growth in the forecast, both during the breeding cycle and during model integration, will therefore cause a divergence from the evolution of the true atmosphere associated with physically growing waves.

The breeding method

Let us consider the wave structure of the atmosphere to comprise well-organised meteorologically important modes and high-frequency gravity waves. The errors in any analysis are considered to be either of the well-organised fast-growing type or the non-growing type in either a high (gravity waves) or low (Rossby waves) frequency form. In this study, the fast-growing errors are of most interest. These are the modes that tend to grow because of small initial perturbations and are symptomatic of the intrinsic errors associated with the analysis of the synoptic situation. Because they are balanced synoptic solutions to the governing equations, their evolution represented in phase space will show a desirable divergence of different initial uncertainties immediately after model integration has begun.

In the analysis (and assimilation) procedure, observational data are added to an initial state. The resulting imperfect state is balanced and consequently the analysis error (which was partly caused by observational error) will project onto the synoptic degrees of freedom.
Discussion of various aspects of the breeding method has been presented in the literature. These include similarities with singular vector analysis (Palmer 1996) and the role of the breeding method in identifying the global and local Lyapunov vectors into the non-linear domain (Toth and Kalnay 1997).

**Atmospheric model**

The model used in this study (the Melbourne University GCM) is a full non-linear physically based primitive equation spectral GCM. The four main prognostic variables (vorticity, divergence, temperature and moisture) are represented in the vertical at nine discrete levels (L9) in the terrain-following ‘sigma’ coordinate. The spectral series is truncated rhomboidally at wave number 21 (R21) to give a gridded horizontal resolution equivalent to 5.6°x3.3°.

Many physical processes not represented in the primitive equations are included in the model via appropriate parametrisations. The radiation scheme follows that of Fels and Schwarzkopf (1975) and Schwarzkopf and Fels (1991). Clouds are generated interactively (Argete and Simmonds 1996) as either a convective or non-convective type and are formed in three discrete layers. The convection is considered via the moist convective adjustment scheme suggested by Manabe et al. (1965). A two-layer soil hydrology is included following the Deardorff (1977) model. Surface and boundary fluxes follow Monin-Obukhov theory as discussed in detail by Simmonds (1985). Heat fluxes over sea-ice are parametrised to allow for leads as described by Simmonds and Budd (1990). The topography used originated from continental heights of Smith et al. (1966) on a 1°x1° grid. Snow cover is prescribed by climatology from passive microwave observations from the Nimbus 7 satellite and there is forcing from the sea-surface temperatures of Reynolds (1988) at the lower boundary. Further discussion of the GCM can be found in Simmonds (1985), Simmonds et al. (1988), Simmonds and Lynch (1992) and Simmonds and Law (1995).

The initial and verifying analyses used here were obtained from the Australian Bureau of Meteorology. These were the twice-daily ‘GASP’ operational analyses produced by a model of similar architecture (Seaman et al. 1995; Bourke et al. 1995). As the models are not identical, it can be expected that the data will not completely project onto the normal modes of the Melbourne University GCM. These will show up as high-frequency gravity waves in the forecasts undertaken by the Melbourne model. A frequency filter (Asselin 1972) is used to efficiently damp the undesirable and meteorologically unimportant waves efficiently throughout the forecast range considered.

**Generating fast-growing modes**

Although choice of the initial perturbation in the breeding cycle has been in the past considered arbitrary, care must be taken. The random errors added initially should be representative of the uncertainty in the analysis. If, for example, the spherical harmonics of the initial state are perturbed randomly, a model could be forced into a regime of altered climate that is substantially different from any other state on the attracting set. In particular, perturbing all spherical harmonics will alter those that define the climate onto which the synoptic state is superimposed. In essence, this will alter the climate in the initial state and the model will react to return it to the attractor. One choice for the random errors is to use a field whose variance in the wave number spectrum reflects the amount of uncertainty expected for a given wave number. Another choice is to perturb only the transient part of the flow. Houtekamer and Derome (1994) scale the magnitude of their FGMs to be ten per cent the magnitude of the transient component of the flow. This is thought to represent the typical magnitude of the analysis error (Toth and Kalnay 1993). Further, it allows the perturbations to be closer to the attractor initially and remain close after a relatively small number of breeding steps. The expected analysis errors in the southern mid-latitudes are large compared to the northern mid-latitudes, thus scaling the transient component by ten per cent may not be appropriate for approximating the analysis errors in the southern regions. Notwithstanding this uncertainty, this method is adopted for this study. After scaling the transient component to ten per cent of its initial size, the amplitude of spherical harmonics coefficients were perturbed by normally distributed random numbers with a mean of zero and unit standard deviation. The resulting field was introduced as the perturbation at the start of the breeding cycle.

It must be remembered that a perturbation that is random will project on both growing and non-growing error modes. This implies that in the first few cycles of the breeding process, there will be an apparent reduction in total error as the non-growing modes are filtered out of the flow. By the end of the breeding period, however, the FGM will be balanced, by definition, with respect to the governing equations.

Once the FGM has been generated, a choice must be made as to how strongly the initial state should be perturbed, or specifically, what value the coefficient $\alpha$ should take if we represent states ($\psi$, at the initial time $t_0$ and dependent on all model variables $x$ ) perturbed with the bred perturbation (as a function, FGM) as:

$$\psi_{\text{perturbed}}(\tilde{x}, t_0) = \psi_{\text{control}}(x, t_0) + \alpha \cdot \text{FGM}(\tilde{x})$$

...
Houtekamer and Derome (1994) show, in studying the three-dimensional system of Lorenz (1963), that the best choice for the amplitude, $\alpha$, is dependent on the measure of forecast skill that is required to be minimised by using an ensemble technique. They state that measures of skill that are sensitive to large errors can be minimised by using a large value of $\alpha$. They argue that a small $\alpha$ will produce mostly linear evolution causing the sphere of uncertainty to evolve in a mostly symmetrical manner and hence the ensemble forecast will be very similar to the control forecast. Houtekamer and Derome show that the FGM ensemble forecast will be superior to the control forecast when the analysis errors are larger than those for which linear theory is applicable. They reason that for large analysis errors, a large amplitude perturbation is optimal. Thus, to force non-linear error growth early in the forecast period a larger amplitude perturbation would be more appropriate.

In this study, the chosen rms size of the initial perturbation in the breeding cycle is ten per cent of the average transient component of the flow on the day under consideration. As this is considered representative of the size of the errors present in the analysis cycle, $\alpha$ was taken to be unity.

A period of five days prior to forecast integration of the model was chosen over which the FGMs were bred.

**Dependence of FGMs on initial perturbation**

We firstly examine the influence of different random perturbations on the final bred mode for a number of arbitrarily chosen days. We have taken day 6 July in 1990, 1991 and 1992, analysed for 1100 UTC with the breeding commencing five days prior on 1 July at 1100 UTC. As the results from these three individual cases are very similar, the 6 July 1990 example is examined in detail to facilitate discussion. The mean sea-level pressure (MSLP) for 6 July 1990 is displayed in Fig. 1 and shows a dominant three Rossby wave pattern in the southern hemisphere. A dipolar system consisting of a high pressure cell is centred over the Australian continent with the associated cyclone south of the Bight. A storm near the Antarctic Peninsula is coupled with two anticyclones on either side of the South American continent. There is a third major storm to the south of Africa. For convenience, these three systems are referred to as 1, 2 and 3 respectively.

To examine the extent to which different choices for the random initial perturbations affect the structure of the FGM bred, a sample of 18 bred modes was generated for each of the three days under examination. The only difference in the breeding cycle of the 18 cases was the initial random perturbation. Figure 2 shows the streamfunction at the 0.5 sigma level for three of the 18 FGMs bred for the 1990 case; these three examples are sufficient to establish the points we wish to make below. It is perhaps surprising that these three patterns show rather different structures as one would expect the FGM to be related to the analysis error (Toth and Kalnay 1993). As they appear different, it must be concluded that each FGM captures slightly different aspects of the analysis error. This observation is also true for fields other than streamfunction (not shown) and the other days considered.

**Fig. 1** MSLP at 1100 UTC 6 July 1990. Contour interval is 8 hPa.
Fig. 2 Three examples of FGMs with different initial random perturbations for initialisation at 1100 UTC on 6 July 1990. Streamfunction is plotted in units of $10^6 \text{ s}^{-1}$ with a contour interval of $2.5 \times 10^5 \text{ s}^{-1}$.

(a) 

(b) 

(c)
To present one measure of the degree of similarity (or otherwise) of the 18 modes in each of the three years, we have calculated the spatial global pattern correlation (PC) between two fields \((x\) and \(y\)). The PC is defined as:

\[
PC = \sum_{i=1}^{l} \sum_{j=1}^{j} \frac{(x_{ij} - \bar{x})(y_{ij} - \bar{y})}{\sqrt{\sum_{i=1}^{l} \sum_{j=1}^{j} (x_{ij} - \bar{x})^2 (y_{ij} - \bar{y})^2}}
\]

where the \(x_{ij}\) and the \(y_{ij}\) are values of \(x\) and \(y\) at the \((i, j)\) location and the overbar refers to the spatial mean. The variables are area weighted by the cosine of the latitude. The pattern correlation allows us a simple measure of the similarity between spatial maps. However, as we shall see shortly, more complex measures of similarity to compare directly the shape of the ellipsoid of the FGM through the use of EOF analysis (e.g. Buizza 1995) are more powerful and lead to somewhat different conclusions. For the PC, the 0.5 sigma level was used to compare each of the 18 FGMs with each other in each of the three base days (i.e. \(3 \times 153 = 459\) individual correlations).

The distribution of these correlations is most conveniently represented in the form of a histogram of relative frequency. Figure 3 shows these distributions for PC calculated over spatial domains of the whole globe, the southern hemisphere and the Australian region (defined as being from \(10^\circ\) to \(50^\circ\)S and \(105^\circ\) to \(160^\circ\)E). As the initial perturbation is arbitrary, the sign of the bred mode is arbitrary, which suggests that it is the magnitude of the correlations that are of interest. Although it may have been expected that the structure of the FGMs is mostly independent of the initial perturbation, Fig. 3 shows that, according to this measure, there is very little relationship between spatial maps of the 18 different FGMs on each day. Indeed, each of the domains considered has a mean correlation close to zero. In the global case, only 30 of the 459 global PC values exceed the 95 per cent confidence interval suggesting the distribution of values in Fig. 3 assumes the form that could be expected from random sampling. Hence, despite what one might expect from the design of the method, our results suggest that the FGMs are structurally dissimilar. Note also that there is an increase in the spread of the distributions in Fig. 3 as we reduce the size of the domain. This is quantified by standard deviations of 0.172, 0.167 and 0.314 for the global, southern hemisphere and Australian region respectively. This is consistent with the classical dependence on sample size.

Another way of understanding the last result is by observing that any two randomly selected vectors in multi-dimensional space will have, on average, a correlation of zero. Similarly, if the phase space dimension is sufficiently large (as it is here) and there are many directions (or regional features, as seen in Fig. 2) that are important to the perturbation, the mean of the correlations should be small. Further, the variance of the correlations between any pair of perturbations will be small. This is what is seen in Fig. 3(a) and Fig. 3(b) where there are a large number of regional features with respect to the size of the domain considered. Conversely, if there are only a few directions in phase
Various regions of the atmosphere exhibit differing amount of daily variability. For example, tropical latitudes experience little variability compared to areas that are influenced by the baroclinic waves. Further, the largest quantitative forecast errors in deterministic NWP are associated with regions of greatest variability and analysis errors may be expected to be largest there. As we have stated above, the method for breeding FGMs is designed to propagate from the largest analysis errors that project onto the synoptic modes of the flow. Therefore, it can be expected that the largest amplitudes of the FGMs would be associated with regions of greatest variability.

As can be seen in Fig. 2, the FGMs tend to show most structure in the baroclinic latitudes. In the southern mid-latitudes (in particular the systems labelled 1, 2 and 3 earlier), we might have expected to see the regional features of the FGMs to be spatially correlated with those of the MSLP. While a greater variability in baroclinic regions is observed in the FGMs, the fact that the regional modes are not coincident with the synoptic features suggests that there is a more subtle association. Further, it is seen that the spatial variability of the FGMs is similar to that of the synoptic flow. This again suggests that the largest amplitudes of features in the FGMs are associated with well-defined synoptic systems.

Toth and Kalnay (1997) have commented that the regional modes are a combination of the local Lyapunov vectors in phase space. These are seen as a particular feature in the FGMs at some geographical location. It is thought that the amplitude or weighting of the individual vectors is determined by the exact nature of the initial random perturbation. Where regional modes bred very similar, the largest Lyapunov vector is said to be dominant. This is in agreement with the structure displayed in Fig.2.

In the region of system 2, we see a number of regional modes with large amplitude. This is also true for systems 1 and 3. The areas where there is interaction between, say, systems 1 and 2, the FGM will have a regional mode with large amplitude (Fig. 2(c)). Similarly where, for example, the two features in the dipolar structure of system 1 interact, there is significant regional mode in the FGMs. It is seen that the error modes will grow to emphasise the areas that will change significantly in the forecast (and breeding) period. For example, consider a system, say, the anticyclone of system 1 during the breeding cycle. If in the perturbed run, the system advanced downstream more rapidly than in the control run, a regional error feature will develop in front of the system rather than perhaps coincident with it. The converse would also be true if the development was slower in the perturbed run. This argument can be applied in a meridional direction as well as an east-west propagation of background synoptic flow. Such displacement errors will show up as dipolar structure in the difference field of the control and a perturbed model run (as in Fig. 2 to the north east of New Zealand). The implication of this is that we would expect (and indeed, observe in Fig. 2) that the features of the FGMs are ‘associated’ with the baroclinic storm activity but not coincident with it.

With this background, we can understand why the bred modes developed above appear to be so different. To discuss the role of the initial perturbation and the aspects of baroclinic instability with which the FGMs are associated, let us consider the role of small disturbances from which the baroclinic eddies develop. Specifically, the random disturbances introduced at the initial time in the breeding cycle will dictate, to some extent, the places where baroclinic development is encouraged or inhibited. That is, the regional position of eddy growth during the breeding period will depend strongly on the regional location of initial perturbations (or in our case, the random perturbations). Therefore, by choosing one particular initial perturbation, the location for baroclinic development or modification will be, to some extent, regionally ‘locked-in’. Thus, it is reasonable that, given a number of different initial perturbations, as in the example described above, different baroclinically unstable modes will develop as determined by the precise location of the initial perturbations. At the same time, the errors associated with this development will correspond to the dynamic eddy systems already present. This argument suggests why the 18 FGMs are not similar and that any one of them can be considered an ideal choice for perturbing the baroclinic modes. There may, however, be some places which will not be able to generate significant amplitude FGMs because of only weak background baroclinicity or, conversely, places which are supercritical that will be unstable for a whole range of perturbations.
Bred perturbations for experiments

We have seen above that the structure of the FGMs depends strongly on the nature of the arbitrary and random method by which the breeding is seeded. It seems natural, then, to ask by how much do the FGMs differ when bred over different time epochs. To answer this, a second set of FGMs was analysed to allow a number of short to medium-range forecasts to be run. In this case the FGMs were all seeded with the same random field but were bred over five days commencing on 1, 6, 11,..., 26 in July of 1990, 1991 and 1992, giving a total of 18 bred perturbation modes. Inspection of these modes (not shown) revealed that they showed considerable similarity to each other (and much more than those shown in Fig. 2). This perception was verified by the pattern correlations between all pairs of 0.5 sigma level streamfunction of the 18 modes (resulting, as before, in 153 PC values). These values are displayed in Fig. 4 for the global domain used above. In general, Fig. 4 shows higher correlations than those of Fig. 3. The PC has a mean of 0.497 (with a standard deviation of 0.098) and all values are positive. As such, the use of the PC indicates that this set of FGMs samples phase space in a different manner compared to the FGMs examined earlier. It is important to appreciate that the pattern correlation may not be the best measure for determining the degree of similarity (or otherwise) of our FGMs. A more rigorous approach is to conduct an EOF analysis on the global bred modes. We firstly conducted such an analysis on the 18 FGMs derived from random seeds but with the same synoptic situation. The variance explained by the first 17 modes is represented by the solid line in Fig. 5. The first mode is seen to explain in excess of 25 per cent of the variance. A similar analysis was performed and the FGMs bred with the same initial perturbation but different synoptic situations (as discussed above) and the variance spectrum is shown in Fig. 5.

Fig. 4 Frequency distribution of global pattern correlations (PC) using 0.5 sigma level streamfunction between each of the 18 July FGMs.

Fig. 5 Variance explained (%) by EOFs of FGM ensembles bred using different random seeding (solid line) and different synoptic conditions (dashed line).

If the FGMs were randomly related, we would expect the spectrum to be close to 'white'. However, we can see this is not so. This implies a stronger bias in the direction that the FGMs are sampling in phase space. Although the spectrum of the FGMs generated from different synoptic conditions is slightly 'whiter' than that of the FGMs described earlier, they are qualitatively similar. Such an observation leads to the conclusion that the two ensembles of FGMs have similar growth properties in phase space irrespective of the apparent similarity of their maps. This finding based on the EOF analysis leads us to examine more closely the tentative conclusions we had formed on the basis of Fig. 4.

Using a three-level quasigeostrophic model with triangular truncation T21, HouTKAMER and DEROME (1995) consider the impact of the length of the breeding cycle on the FGMs. They notice that initially, around 85 per cent of the number of FGMs in their ensemble are required to explain most of the variance in the FGM ensemble. After 40 iterations (i.e. 20 days), this fraction is reduced to about 16 per cent (five EOFs) after which the number is relatively constant. This suggests that what we observe in Fig. 4 represents a transient state of the FGM ensemble. If allowed a longer breeding period, one would expect the mean correlation to approach zero. This is in agreement with the observations of HouTKAMER and DEROME (1995). Further, we would expect that the standard deviation of the distribution would remain smaller than that of Fig. 3(a), indicating that the FGMs from the same day have more in common than if they had been chosen at random.

Toth and Kalny (1997) discuss the role of multiple breeding cycles to obtain many initial states in an ensemble. Their FGMs had a different relative composition of local modes. Thus, they obtained a group of perturbations that were quasi-orthogonal in phase space and thus span the initial uncertainty with more accuracy.
The results shown in Fig. 3 support this notion with low correlation expected if modes are to be orthogonal.

In the 18 July cases, the structure of the FGMs is reminiscent of that of baroclinic storms. In particular, the strong features in the FGMs represent the variability in both the rate of storm development and zonal propagation of baroclinic systems on the background flow. The FGMs indeed show large amplitude structures corresponding to the regions between significant MSLP cyclonic and anticyclonic systems.

Forecast experiments

Ensembles comprising three members were generated. These were the best guess ($\psi_{BG}$) analysis (control) and the best guess perturbed with unit amplitude for both positive ($\alpha = 1$, denoted as ($\psi_+$) and negative ($\alpha = -1$, $\psi_-$) FGM. With $\psi$ denoting some atmospheric state and FGM the perturbation function, these can be expressed as:

$$
\psi_{BG}(\vec{x}, t_0) = \psi_{control}(\vec{x}, t_0)
$$

$$
\psi_+(\vec{x}, t_0) = \psi_{control}(\vec{x}, t_0) + FGM(\vec{x}) \quad \text{...3}
$$

$$
\psi_-(\vec{x}, t_0) = \psi_{control}(\vec{x}, t_0) - FGM(\vec{x})
$$

The FGMs used in the 18 cases described above were added to, and subtracted from, the control analyses at the end of the five-day breeding period. Thus, the forecast integration started from 6, 11, 16, ..., 31 July in 1990, 1991 and 1992. The three were used to start evolution of the GCM and five-day forecasts were made. A fourth and fifth ‘forecast’ were constructed simply by averaging the two FGM forecasts (AV2) at each output time and by averaging the two FGM and the best guess forecasts (AV3). Specifically these can be expressed as:

$$
\psi_{AV2}(\vec{x}, t) = \frac{\psi_+(\vec{x}, t) + \psi_-(\vec{x}, t)}{2} \quad \text{...4}
$$

$$
\psi_{AV3}(\vec{x}, t) = \frac{\psi_+(\vec{x}, t) + \psi_-(\vec{x}, t) + \psi_{BG}(\vec{x}, t)}{3}
$$

These two forecasts are not NWPs in the normal deterministic sense. Nonetheless, it is the AV2 and AV3 forecasts that are of most interest in this study designed to examine the advantages of using the ensemble mean as a stochastic method of prediction. The use of the ensemble mean can be considered the first step in evaluating ensemble performance. More complicated methods (e.g. clustering) can be applied to the forecasts and higher order moments of the ensemble are thought to describe other useful properties of the ensemble (e.g. variance linked to predictability).

To highlight the main features of the FGM ensemble forecast, one of the 18 cases is discussed in detail. This case was that integrated from 6 July 1990 (as in Fig. 1) and verified five days later on 11 July 1990 (MSLP analysis shown in Fig. 6(a)). The GCM prediction is shown in Fig. 6(b). The difference between the verifying analysis and the control forecast is presented in Fig. 6(c). Figure 6(c) shows both errors resulting from the use of an imperfect model in combination with errors associated with the chaotic nature of the flow regime (as compared to a ‘perfect model’ approach that isolates errors associated with the flow regime (e.g. Houtekamer and Derome 1995). It is seen that the greatest errors are present in the baroclinic zones, particularly in the southern hemisphere. The model fails to predict the cyclone to the southeast of New Zealand. The evolution of the low to the south of South America is modelled to be slower than in reality. The Tasman Sea high and the low to the southwest of Australia are also modelled poorly. The cyclone over Asia is seen as an anticyclone in the model evolution while other weak Asian and European systems tend to be poorly forecast with respect to the central pressures. Atlantic features are modelled reasonably well although again the central pressures are not well simulated. In view of the variability expected in the equatorial region, the model shows an unsurprising small error.

The difference between the $+FGM$, $-FGM$ and AV2 forecasts and the control forecast are shown in Fig. 7 and can be used in conjunction with Fig. 6 to assess improvements over the control forecast without the influence of model errors. Figures 7(a) and 7(b) both show substantial deviation from the control over Asia. This, in effect, introduces the low seen in Fig. 6(a) to the control forecast where there was normally a high in the control forecast (Fig. 6(b)). An intense anomaly is seen over northern Asia in the $+FGM$ difference associated with this improvement. The northwest Pacific also shows a great deal of structure in the differences associated with the model picking up a region of high variability in the forecasts. In southern mid-latitudes, deviations from the control forecast are mostly on a synoptic scale. Large magnitude disturbances are seen in the region to the south of Australia and to the south of South America in the case of the $-FGM$ forecast.

In general, the $-FGM$ forecast is superior to the control for most of the forecast period and issuing only this perturbed forecast operationally would effectively improve the prognosis, however, it would not be known a priori which one of the FGM forecasts is preferred. Because of this, the average may be a better forecast to issue than the control. This is because it captures some of the error-reducing properties of one FGM forecast ($-FGM$ in this case) while still being able to be calculated and issued in a predictive manner.

The AV2 forecast is qualitatively superior to the control forecast. The two highs and the low over Asia are
Fig. 6  (a) MSLP on verification day – 11 July 1990 (b) Five-day control forecast of MSLP initialised on 6 July 1990 forecast (contour intervals are 8 hPa) (c) MSLP difference between verifying analysis and control forecast (contour interval is 1.5 hPa)
Fig. 7 Five-day forecast difference in MSLP on 11 July 1990 between (a) +FGM and control (b) -FGM and control (c) AV2 and control (contour intervals are 1.5 hPa).
now represented, although still not accurate. The regions south of Australia and South America show a small improvement in terms of differences and position of features, while the forecast still shows limited usefulness. It is expected that a quantitative improvement will be seen. To ensure that any improvement is not simply a manifestation of smoothing inherent in the averaging, it is useful to quantify the amount of spatial variability via the rms size or spatial standard deviation (SD).

$$SD = \sqrt{\frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} (x_{ij} - \bar{x})^2}$$ ...

where, again, the variable $x_{ij}$ is area weighted.

The SD for streamfunction is plotted for all forecasts defined (Fig. 8). Also plotted is the standard deviation of the corresponding GASP analyses to enable comparison of the GCM’s ability to reproduce synoptic variability. Figure 8 shows the spatial SD of AV2 and AV3 forecasts differ negligibly from the control forecast in the global domain and the order of 0.1 - 0.2 in the Australian region. Therefore there appears no loss of structure in the AV2 and AV3 forecasts compared to the two FGM perturbed runs and the control run from the 18 July forecasts. This result is similar if the 18 cases are examined individually (not shown).

The rms error (or difference as used earlier) between two fields can be defined as:

$$rms = \sqrt{\frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} (x_{ij} - y_{ij})^2}$$ ...

(as with the pattern correlation, the variable under consideration is area weighted).

For each of the five forecasts, this quantity was calculated for a variety of atmospheric variables. The calculations were done for the entire globe and over the Australian region to allow assessment of localised effects on the forecasts issued. The average rms error for the 18 cases over two spatial domains is plotted for two fields in Fig. 9 as differences from those obtained in the control forecast.

The most obvious feature of Fig. 9 is the consistently lower rms error of the AV2 and AV3 forecasts. Of these two, the AV2 forecast is seen to be better in a rms sense and will be used in the following discussion. The AV2 forecast is, in fact, better than control (MSLP and 0.5 sigma level temperature) in 17 of the cases (with the remaining one indistinguishable from control), although divergence from the control forecast increased at around 48-60 hours (Fig. 9(a)). This last finding is not surprising as this is the time when non-linear errors are expected to become very important, causing the average ensemble forecast to lose coincidence with the central or control forecast. After this two-day period, the effects on averaging the +FGM and the -FGM forecasts can be thought of as filtering out the unpredictable components of the flow to some extent.

In the 0.5 sigma level temperature fields of the 18 cases, the global rms improvement at the end of the fifth model day is around 0.12K globally and 0.18K in the Australian region (Fig. 9(b)). Although this is only a small numerical reduction, it represents about three per cent and four per cent of the total rms error respectively. For the 18 cases, this is shown to be statistically significant at over the 99 per cent level over the entire globe and at around 80 per cent in the Australian region. The improvement in the latter region is less statistically significant essentially due to the greater spatial variance associated with having less grid-points to consider in the rms calculations. The difference in rms of any of the five forecasts can be up to 2-3 K for the 0.5 sigma temperature field.

Fig. 8 Mean SD for 18 forecasts (as a function of time - in hours) of the GASP analyses (heavy broken), and BG (heavy solid), +FGM and -FGM (light solid), AV2 (long broken) and AV3 (short broken) forecasts. Shown are the values for 0.5 sigma level streamfunction (10^5 s^-1) for the 18 July forecasts made. Both global and Australian region domains shown as upper and lower (respectively) clustering of curves.
Fig. 9 Mean difference for 18 forecasts (as a function of time in hours) between rms error of the BG and the +FGM and -FGM (light solid), AV2 (long broken) and AV3 (short broken) forecasts. Shown are the errors for (a) MSLP (hPa) and (b) 0.5 sigma level temperature (°C). Both global (right vertical axis) and Australian region (left vertical axis) are shown.

An average rms of perturbed forecasts would have a larger rms than the control. For any one case, however, while the error growth is mostly linear, the rms error of the two FGM forecasts will be greater and less than the rms error of the best guess. After this time, the errors will become larger and the rms error will approach the maximum or asymptotic rms error of the forecast model. This can be seen in some instances as a convergence of the FGM rms error to the rms error curve of the control forecast in Fig. 9.

Concluding remarks

It has been accepted for over 30 years that if the initial state of a non-linear dynamical system is not known precisely, numerical model forecasts made will eventually become of little or no use (Lorenz 1963). It is not possible to reduce the initial error indefinitely simply by improving observations. This follows from Lorenz's (1969) study of non-linear flow with many scales of motion. Specifically, Lorenz recognised that uncertainty in motions of the smallest scale will contaminate predictions of larger scales and thus the atmosphere has a limited range of predictability. It is likely that through consideration of the dynamics of errors, greater use of existing forecasts can be made and forecasting methods can be developed to enhance 'model' performance.

Ensemble techniques of forecasting have been suggested in numerous studies (e.g. Lorenz 1965; Leith 1974; Molteni and Palmer 1993; Toth and Kalnay 1993). The advantage of such methods is that it is possible to gain more information about the forecast based on ensemble statistics. Distribution of the member forecasts can illustrate the likelihood that one particular evolution is favoured. Other higher order moments of the ensemble can be used to diagnose the ensemble performance. For example, the variance associated with the divergence of ensemble members at the end of the prediction can act as a measure to predictability. Here, the mean of the ensemble has been shown to produce a forecast that is better in a rms sense than the single best guess forecast.

Two methods have been suggested to generate fast-growing perturbations. Namely, singular vector analysis following the work of Lorenz (1965) (applied to weather forecasting by Lacarra and Talagrand (1988) and Molteni and Palmer (1993)) and the method for breeding fastest-growing modes of Toth and Kalnay (1993). An attractive feature of the latter approach is that the FGMs are generated in a straightforward manner with little computational expense using an existing non-linear primitive equation model. It is thought that both methods generate perturbations that have similar evolutionary properties. Specifically, perturbing along the fast-
evolving modes ensures that ensemble members will diverge quickly.

The FGM is designed to capture possible errors in the analysis (Toth and Kalnay 1993). This is achieved by allowing the growing modes in the state to develop while being forced by the raw observations when available. The observations will have some uncertainty, which will project predominantly onto non-growing modes and be promptly reduced during initialisation. It can be assumed that the bred modes contain large-scale disturbances that were present and growing in the initial state. These disturbances will grow during the analysis cycle both dynamically and with the addition of new observations. As they will then be present as synoptic modes, they will not be removed by, for example, normal-mode initialisation. Therefore, they could be present as errors in any analysis, and perturbing the best guess with the FGM will tend to bring one member closer to the actual state of the atmosphere at the initial time, thus leading to more accurate forecasting in that member and also for an average of the FGM members.

The FGMs we have bred displayed considerable spatial structure in the mid-latitudes and features were qualitatively correlated to zones of baroclinic instability. In this case, correlation refers to the fact that there are error modes associated with the baroclinic systems, although their exact position may be displaced by a few thousand kilometres depending on the specific FGM bred. It was seen that the use of the spread of pattern correlation distribution can act as an indication of the number of regional modes (phase space directions) that are represented by the FGM ensemble. However, it was seen that the use of an EOF analysis was required to appreciate the distribution of the FGMs in phase space adequately. Ensemble forecasts were made by averaging two forecasts made with the initial state perturbed with a positive and negative FGM. Using this method, it was found with a high degree of statistical significance (99 per cent) that there was over a three per cent improvement in the rms error in 0.5 sigma temperature of the control forecast. This achievement can be thought of as partially filtering out the unpredictable components of the initial flow in an a posteriori sense.

Acknowledgments

In the initial stages of this work discussion with Jorgen Frederiksen and Bill Bourke proved most useful. The interest and suggestions of Zoltan Toth and Jacques Derome are appreciated.

References

Argete, A. and Simmonds, I. 1996. Comparison of temporal cloud vari-
ability simulated by a GCM with observations from the Nimbus-7 satellite. Atmosfera, 9, 1-21.
Simmonds, I., Trigg, G. and Law, R. 1988. The Climatology of the
Melbourne University General Circulation Model, Department of Meteorology, University of Melbourne, *Pub. No. 31, 67* [NTIS PB 88 227491].


