TECHNICAL MEMORANDUM

328

On the duration of the linear regime: Is 24 hours a long time in weather forecasting?

I. Gilmour, L.A. Smith and R. Buizza

Research Department

January 2001

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen

Abstract

Day to day variations in the growth of uncertainty in the current state of the atmosphere have led to operational ensemble weather predictions in which an ensemble of different initial conditions, each perturbed from the best estimate of the current state yet still consistent with the observations, is forecast. Contrasting competing methods for the selection of ensemble members is a subject of active research; the assumption that the ensemble members represent sufficiently small perturbations so as to evolve within the 'linear regime' is implicit to several of these methods. This regime, in which the model dynamics are well represented by a linear approximation, is commonly held to extend to 2 or 3 days for operational forecasts. It is shown that this is rarely the case. A new measure is introduced which quantifies the duration of the linear regime by monitoring the evolution of 'twin' pairs of ensemble members. Both European and American ensemble prediction systems are examined; in the cases considered for each system (87 and 25 respectively), the duration of the linear regime is often less than a day and never extends to 2 days. The internal consistency of operational ensemble formation schemes is discussed in light of these results. By decreasing the optimization time, a modified singular vector based formation scheme is shown to improve consistency while maintaining traditional skill and spread scores in the 7 cases considered. The relevance of the linear regime to issues regarding data assimilation and adaptive observations are also noted.

1 Introduction

Uncertainty in the initial condition combined with model error renders prediction of a chaotic system nontrivial. Indeed, the day to day variation in growth of uncertainties of the initial state of the atmosphere has led many operational weather forecasting centers to adopt ensemble forecasting (*i.e.* the use of ensembles of initial conditions evolved under a model). Investigations which aim either to further understanding of uncertainty growth in the system, or to improve predictions, often are based on the assumption that error growth is linear. The extent to which this approximation reflects the true dynamics defines the 'linear regime'; as discussed in the next section, the duration of the linear regime will depend on the size and orientation of the perturbation as well as the particular initial condition. In this paper we exploit the symmetry of linear dynamics to construct a new measure, Θ, which can be used to bound the duration of the linear regime.

Ensemble prediction systems (EPS) of the European Centre for Medium-range Weather Forecasts (ECMWF) and of the National Center for Environmental Prediction (NCEP) will be examined. Operational perturbations are commonly held to evolve approximately linearly for at least two or three days (Palmer *et al.* 1994); our results indicate that this is *not* the case, either for ECMWF singular vector (SV) based ensemble members or for NCEP breeding vector (BV) based ensemble members. This result is important inasmuch as the operational ensemble members will not reflect the properties of the theoretical SV (or BV) which motivated their selection in the first place. While we will focus on the implications for ensemble formation, these results also have significance for data assimilation schemes, adaptive observation strategies and model sensitivity analyses. Given the fundamental role the linear regime plays in these contexts, the fact that its duration is significantly less than what is commonly assumed is of some importance.

Ensemble formation schemes aim to select a few dozen ensemble members which provide an operationally viable sample of the dynamics in a $\sim 10^7$ dimensional model state space. Debates over both what constitutes a good sample and how such a sample is best obtained keep this an active, often acrimonious, field of research (Anderson 1996; ECMWF 1999; Hamill *et al.* 2000; Houtekamer and Derome 1995; Palmer *et al.* 1998; Szunyogh *et al.* 1997). It is not the purpose of this paper to enter into this discussion directly; rather, we consider the theoretical framework behind the operational ensemble formation schemes of both ECMWF and NCEP and contrast their assumptions regarding the linear regime with measurements of its duration. This paper provides a test of internal consistency: Our aim is to determine whether, or not, the dynamics of the operational ensembles realized in practice are consistent with the linearity assumptions made in their definition. (For the SV the assumption is explicit in their definition; for the BV it is implicit.) Note that operational forecasts extend over a time span such that the ensemble member trajectories will typically be nonlinear at some point during the forecast; indeed, herein lies much of the value of an ensemble forecast. We investigate linearity of evolution of perturbations over much shorter time-scales, relevant to the definition of the subspace sampled by the ensemble.

In the following section we define the linear regime in order to introduce methods which quantify its duration before motivating the study with an illustration from laboratory geophysical fluid dynamics and describing the implementation in operational models. A new quantitative measure of the validity of the linear approximation is defined which can be calculated whenever twin perturbations are included as members of the ensemble (which is a common procedure; see *e.g.* Molteni *et al.* (1996); Toth and Kalnay (1997)). This approach provides an *upper* bound on the duration of the linear regime for some prescribed degree of accuracy. The duration of the linear regime is quantified for operational ensembles in section 3, where it is shown that the error in assuming the linear approximation to hold at 48 hours is typically 70% of the magnitude of the evolved perturbation. Implications of this lack of internal consistency and possible remedies are considered; a modified SV ensemble is shown to be more internally consistent, while maintaining traditional skill and spread scores, for the 7 cases considered. The relationship of the new results in this paper to previous work is considered in section 4, and is followed by a general discussion of how these results may be extended and applied in section 5. Section 6 provides a summary.

2 Definition of the linear regime & determination of its extent

The extent of the linear regime for a given perturbation depends upon the relative importance of the nonlinear terms of the flow for that perturbation (as shown schematically in figure 1). For any given initial perturbation, the duration over which the so-called 'tangent linear model' (TLM) will provide a useful approximation will vary with the initial condition, the initial magnitude of the perturbation about this initial condition, the orientation of the perturbation and, of course, the particular model used. We return to issues of modeling the linear regime later in this section, after first describing the dynamics under a perfect TLM.

Suppose we have a set of analysis values, $\mathbf{A}(t)$ (best guesses of the initial conditions computed using assimilation techniques (Talagrand and Courtier 1987)), corresponding to the true system values $\mathbf{x}(t)$. Then the initial condition $\mathbf{A}(0)$ may be considered to be displaced by a perturbation $\boldsymbol{\delta}(0)$ from the system value; the trajectory of this perturbed initial condition ($\mathbf{A}(0) = \mathbf{x}(0) + \boldsymbol{\delta}(0)$) can then be described by a Taylor expansion about the trajectory of $\mathbf{x}$:

$$\mathbf{A}(t) = \mathbf{x}(t) + \mathcal{M}(\mathbf{x}, t)\boldsymbol{\delta}(0) + \boldsymbol{\delta}^T(0)\mathcal{H}(\mathbf{x}, t)\boldsymbol{\delta}(0) + \dots \tag{1}$$

where $\mathcal{M}$ is the linear propagator, $\mathcal{H}$ the Hessian, and so on. For sufficiently small $|\boldsymbol{\delta}(0)|$ and sufficiently short t, the linear approximation

$$\mathbf{A}(t) \approx \mathbf{x}(t) + \mathcal{M}(\mathbf{x}, t)\boldsymbol{\delta}(0) \tag{2}$$

holds to high accuracy in any smooth dynamical system (Eubank and Farmer 1990). Thus the linear propagator $\mathcal{M}(\mathbf{x}, \tau)$ is often said to describe the *linear evolution* of a perturbation about the trajectory of $\mathbf{x}$ over the time interval $[0, \tau]$ (Lorenz 1965); an initial perturbation $\boldsymbol{\delta}(0)$ about $\mathbf{x}(0)$,

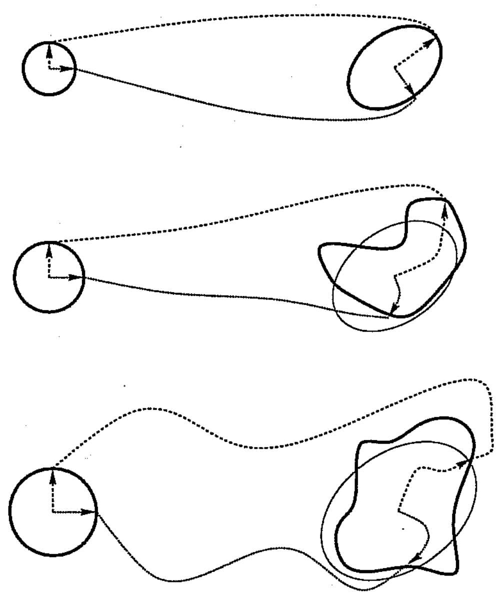

Figure 1: A schematic illustration of the extent of the linear regime. Perturbations of varying magnitude are made from an observation at initial time (dotted and dashed arrows on left hand side) and evolved forwards (indicated by lines from left to right) to some time t (dotted and dashed arrows on right hand side). For small perturbations (upper panel) the linear evolution (represented by bold circle on left to ellipse on right) is an excellent approximation to the full nonlinear dynamics (nonlinearly evolved circle is represented by bold outline on right). As the perturbation magnitude increases (middle panel) the linear approximation is a less accurate description until it has little relation to the full nonlinear dynamics (lower panel).

evolves linearly to $\boldsymbol{\delta}(\tau) = \mathcal{M}(\mathbf{x}, \tau)\boldsymbol{\delta}(0)$. Thus the error in making the linear approximation is simply $\mathbf{A}(\tau) - (\mathbf{x}(\tau) + \boldsymbol{\delta}(\tau))$. Obviously if $|\boldsymbol{\delta}(0)|$ is too large or $[0, \tau]$ too long, the approximation is expected to break down. In practice, a particular time τ is fixed and a set of perturbations defined based on $\mathcal{M}(\mathbf{x}, \tau)$[1]; the linearity of evolution of these perturbations may be quantified as a function of time, t. One aim of this paper is to provide a computationally convenient method for quantifying the quality of the linear approximation in cases where the computational cost of direct simultaneous integration of linear and nonlinear trajectories is prohibitive, and/or in cases where the TLM is not exact.

If a TLM is used which is not an exact linearization of the full nonlinear model, then there are various potential sources of error quite independent from the linearity, or otherwise, of evolution. An obvious concern in numerical weather models arises when the TLM has been computed with a lower spatial resolution than the nonlinear model, but the commonplace exclusion of processes which yield non-differentiable terms from the model is also a concern (Buizza and Montani 1999; Errico *et al.* 1993). Errico *et al.* (1993) also discuss error which arises from the numerical computation. These additional sources of error may curtail the duration over which the TLM is an acceptable approximation.

There are two approaches to determining the duration of the linear regime: One is to contrast the evolution of a perturbation under the full nonlinear model with its evolution under the TLM in order to quantify the error in the TLM as a function of time. An alternative approach, used in this paper, is to quantify the relative impact of the nonlinear terms by monitoring the evolution of twin perturbations under the full nonlinear model. 'Twin' perturbations are initially equal in magnitude and opposite in orientation about an analysis; as long as the model dynamics are approximately linear the twin perturbations will remain roughly equal and opposite with respect to the evolved analysis.

[1] At ECMWF the time τ is called the 'optimization time' since the related perturbations have maximum linear growth over the 'optimization time interval' $[0, \tau]$.

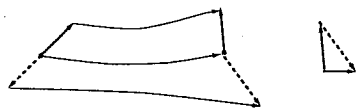

Figure 2: Defining Θ: Equal and opposite perturbations at $t = 0$, $\delta^{\pm}(0)$, evolve so as to be no longer symmetric at time t: the error in assuming linear dynamics, $||\delta^{+}(t) + \delta^{-}(t)||$, is scaled by the average magnitude of the evolved perturbations to give the test statistic Θ.

Comparing the nonlinear evolution of twin perturbations determines when nonlinear processes may not be neglected, thereby providing an upper bound for the relevance of the TLM (and for the duration of the linear regime) for the perturbations considered. Such a comparison does not verify that the properties of a particular TLM match those of the linearized nonlinear model since ensemble members may evolve linearly even if the particular TLM used is inaccurate. We do not, therefore, test the accuracy of a particular TLM, but rather we extract an *upper* bound on the duration of its usefulness as an approximation.

We wish to define a measure to quantify the error made in assuming linear evolution which is sensitive to variations in both magnitude and orientation of a perturbation. Given that we wish to apply this measure to operational ensembles, the diagnostic to be employed should be calculable from data produced routinely. Consider the control trajectory initiated from the analysis, $\mathbf{A}(0)$, as the fiducial trajectory, and denote the deviations of a positive (negative) perturbation from the evolved control at time t by $\delta^{+}(t)$ ($\delta^{-}(t)$). If growth is exactly linear then $\delta^{+}(t) + \delta^{-}(t) = \mathbf{0}$; this sum is sensitive to both the relative magnitudes and orientations of the evolved perturbations, as illustrated in figure 2. Scaling by the average magnitude of the evolved perturbations defines a suitable statistic, the *relative nonlinearity* of evolution Θ, given by

$$\Theta(\hat{\delta}, ||\delta||, t) = \frac{||\delta^{+}(t) + \delta^{-}(t)||}{0.5\{||\delta^{+}(t)|| + ||\delta^{-}(t)||\}} \tag{3}$$

where $\hat{\delta}$ is the unit vector and $||\cdot||$ is one of several possible metrics (typically based on the 500 h Pa geopotential height) defined by the inner product, $(\cdot,\cdot)$. Θ will, of course, vary with the initial condition (*i.e.* the analysis value $\mathbf{A}(0)$). Note that $\Theta = 0$ implies that the dynamics of the perturbation may be linear[2], while $\Theta = 0.5$ implies that the error made in assuming linear evolution will be at least 50% of the average magnitude of the evolved perturbations.

How might such information be used? Figure 3 shows how the fraction of initial conditions for which the linear approximation is deemed acceptable varies as a function of time. These insights are drawn from a radial basis function (RBF) model of the thermally driven rotating fluid annulus (Hide 1958; Read 1992; Smith 1992). This fluid annulus is a classic geophysical fluid dynamics experiment: a laboratory analogy to the Earth's mid-latitude large scale circulation (Früh and Read 1997; Hide and Mason 1975). From this figure, differences in the duration of the linear regime for different magnitudes of perturbations can be clearly seen. The two distinct branches of behavior reflect initial perturbations of differing magnitude: in the upper cluster the perturbation δ has an initial magnitude 10 times that of those in the lower branch; several values of optimization times τ are included. Suppose we were interested in the relevance of the linear regime at $t = 5$. For the smaller perturbations the regime is seen to be a valid approximation for more than 80% of the initial conditions; for the larger perturbations the approximation is valid for less than 10% of cases. Hence if our ensemble formation scheme, for example, relies on validity of the linear approximation at $t = 5$, we must achieve an initial uncertainty in the initial condition corresponding to that of the upper branch. If we cannot, then

[2]It is crucial to remember that $\Theta = 0$ is only a necessary condition for linear evolution; one can contrive examples (*e.g.* pure cubic nonlinearities) where $\Theta = 0$ for some perturbations and yet the linear approximation is wildly inaccurate.

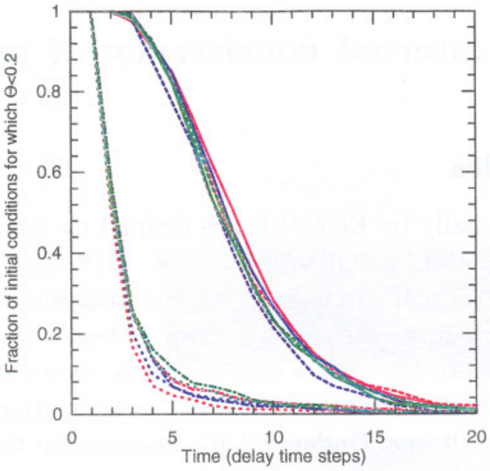

Figure 3: Illustration to show how the fraction of initial conditions for which linear approximation is deemed acceptable varies as a function of time for a RBF model of the annulus. Twin SV perturbations of magnitude ϵ_{SV} and optimized over the time interval $[0, t_{opt}]$ are applied to distinct initial conditions and iterated forward under the model. At each time step, the fraction of initial conditions for which the assumption of linear perturbation growth is in error by less than 20% of ϵ_{SV} (equivalently $\Theta < 0.2$, see equation 3) is calculated. Results shown are for positive perturbations, $\epsilon_{SV} = 0.01$ (solid line), $\epsilon_{SV} = 0.1$ (dot-dashed line) and negative perturbations, $\epsilon_{SV} = 0.01$ (dashed line), $\epsilon_{SV} = 0.1$ (dotted line); the proximity of the results makes it difficult to distinguish between members of paired perturbations. At each magnitude there are 3 sets of results, obtained by varying the optimization time, t_{opt}, from 2 (red) to 4 (blue) or 8 (green) time steps. (Figure 6.7 from Gilmour (1998).)

there is no reason to believe that the operational ensemble will share the properties of the envisaged ensemble. If we rely upon the linear approximation at $t = 15$, neither of the magnitudes considered in figure 3 are small enough. We return to these results in section 3.1.1, but they are introduced here in order to illustrate the utility of knowing the likely duration of the linear regime. Further discussion of the experiment, model, and linear analysis may be found in Gilmour (1998).

The breakdown of the linear regime in an operational model has been explored by Buizza (1995). Following Houtekamer and Derome (1994), Buizza utilizes twin perturbations to define the time after which "*nonlinearity becomes dominant*" in the system as that at which the (anti-)correlation between the evolved twin pair perturbations crosses the value of 0.7. In this case the *anti-correlation*, ℓ, is given by

$$\ell = -\frac{(\delta^+(t), \delta^-(t))}{||\delta^+(t)||\ ||\delta^-(t)||} \quad . \tag{4}$$

Note that the correlation measure reflects *only* the orientation of the evolved perturbations and is blind to their relative magnitudes; if one perturbation grows while the other shrinks, with no change in orientation, the correlation will remain as unity, indifferent to this nonlinear evolution.

The value $\ell = 0.7$ corresponds to the perturbations deviating by $\sim 45°$ from anti-parallel. The corresponding error in assuming linear evolution is at least 75% of the mean magnitude of the evolved perturbations and thus $\Theta > 0.75$; there is no inconsistency here since each of the statistics poses a necessary condition for linear evolution, yet neither provides a sufficient condition. We note here in passing that although the Θ statistic is computed in the model state space, regions of large Θ can be isolated in physical space and compared to synoptic structures of the day, as is done in section 5 below.

3 Evaluation of the internal consistency of operational NWP ensembles

3.1 ECMWF SV ensembles

Ensembles employed operationally by ECMWF are defined by perturbations which are restricted to a subspace spanned by the leading singular vectors of $\mathcal{M}(\mathbf{x}, \tau_{\text{opt}})$ where the optimization time, τ_{opt}, is fixed. By constructing SV perturbations about the analysis, and using an initial magnitude comparable to the analysis uncertainty (the average error between the analysis and the true system value), SV ensembles aim to capture the 'worse case scenario' (Mureau *et al.* 1993). Implicit in this approach is the assumption that the evolution over the optimization time interval $[0, \tau_{\text{opt}}]$ is well described by the TLM and the additional (independent) assumption that these evolved perturbations at time τ_{opt} are likely to sample worst case scenarios over the remainder of the forecast period. In other words, the SV ensemble is designed to approximate the singular subspace in which growth over the optimization time interval is largest; this need only occur if the finite amplitude operational perturbations grow approximately linearly during the optimization time interval.

3.1.1 Construction of the singular vectors

As of May 1998, three models are run operationally at ECMWF (Buizza *et al.* 1998): (1) T42L31, with which the SV are determined and the ensemble formed as detailed below, (2) T_l159L31, used to evolve the ensembles (the analysis is also evolved at this resolution to give a low resolution 'control') and (3) T213L31, under which the analysis is also evolved to give a high resolution control forecast. The TLM for the T42L31 model, developed by the Centre, includes the tangent version of the adiabatic part of the model, linearized horizontal diffusion and simple vertical diffusion and surface drag; among the processes not included are radiation, convection and gravity wave drag (Buizza and Palmer 1995; Park and Droegemeier 1997). Since this TLM is computed at a lower spatial resolution than the nonlinear model and excludes some processes from the linearization, there are sources of error in the TLM which are independent from linearity considerations; the estimate of the duration of the linear regime given by Θ will be an upper bound. The 25 twin pair SV perturbations which form the ECMWF SV ensemble are calculated as follows (further details may be found in Buizza *et al.* (1998) and Molteni *et al.* (1996)):

- Each day at 1200 UTC, the TLM is started from the T42L31 resolution value of the analysis; all the singular vectors, optimized over 48 hours, are calculated by the Lanczos algorithm (Strang 1986).

- Two sets of 25 SV, $\mathbf{v}_j$; $j = 1, 25$, are then selected, one over the Northern Hemisphere and another over the Southern Hemisphere (both extra Tropics, *i.e.* 30°N/S-90°N/S). The first 4 SV for each set are always selected. Each subsequent SV is selected if at least half of its total energy is outside the localized regions of the SV already chosen.

- Both sets of SV are then independently rotated such that the resulting perturbations $\mathbf{p}_j$ have the same globally average energy as the singular vectors but smaller local maxima and more uniform spatial distribution. Note that rotation of singular vectors implies that none of the perturbations need to be in any 'optimal' direction (Vannitsem and Nicolis 1997).

- The SV in each set are then rescaled by choosing constants α_{jn} and R_n to satisfy two considerations: Initial perturbations should have, locally, an initial amplitude similar to analysis error estimates; initial perturbations are larger over oceans, where data is sparse, than over more densely sampled land (Buizza *et al.* 1998; Molteni *et al.* 1996). Secondly, the ensemble standard

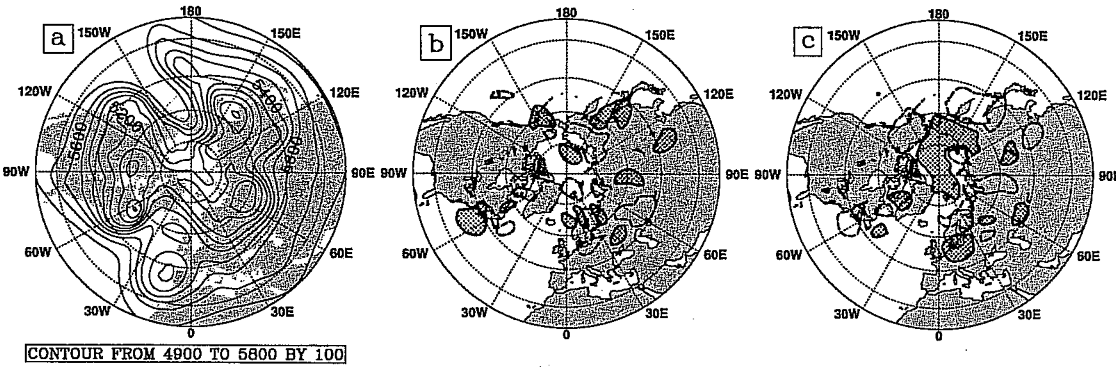

Figure 4: b) and c) 500 h Pa anomaly fields for a twin pair of evolved SV perturbations (negative anomalies less than -30 m outlined, positive anomalies greater than +30 m stippled) along with a) the corresponding evolved control. The perturbations were initiated at 1200 UTC on 19.xii.96 and evolved forwards 48 hours to give a forecast for 1200 UTC on 21.xii.96; shown here is their deviation from the evolved control in the 500 h Pa field for the area over which measures are calculated, namely 22.5°N-90°N.

deviation should be comparable to the estimated error of the ensemble mean (Molteni *et al.* 1996). These two aspects define, respectively, an upper and a lower bound on the perturbation initial amplitude. Practically, a scaling factor R_n is chosen by experimentation (here $R_n = 0.6$) and then the constants α_{jn} are chosen such that $\mathbf{p}_j = \sum_{k=1}^{25} \alpha_{jk}\mathbf{v}_k$ and $||\mathbf{p}_j||\sqrt{R_n} \leq ||\mathbf{a}_e||$ (where $\mathbf{a}_e$ represents the approximated analysis error).

- The corresponding members of the NH and SH sets of SV are then added together, and each resulting perturbation both added to and subtracted from the T42L31 analysis value.

Each ensemble member is then interpolated from T42L31 to T_l159L31 resolution and evolved at this resolution out to 10 days with data output at 12 hour intervals. An example of the 500 h Pa anomaly fields of an evolved pair of twin SV members is shown in figure 4, along with the corresponding evolved (low resolution) control.

3.1.2 Calculation of linearity measures

The relative nonlinearity (Θ) and anti-correlation (ℓ) results are calculated using a norm based on the 500 h Pa geopotential heights (with an 'equal area' weighting) over the Northern Hemisphere, excluding the Tropics, (22.5°N-90°N) for the 25 twin pairs for each of 87 different cases, giving over 2000 twin SV pairs in total. The 500 h Pa geopotential height norm is chosen both because it is the field traditionally used to represent the dynamical state of the atmosphere[3] (Stroe and Royer 1993), in which we are interested, and because it is considered to be the norm in which evolution will be most linear (Toth 1997, private communication); linearity measures calculated using other norms are discussed in section 5 The first 25 cases are SV initiated at 1200 UTC on each day between 11 December 1996 and 4 January 1997 inclusive; results for the 31 days in each of January 1998 and August 1998 were also computed.

At the optimization time of 48 hours, the error made in assuming linear evolution ranges from 42% to 108% of the average magnitude of the evolved perturbations, with an average of 70%. Results for the 25 cases in December 1996 and January 1997 are shown in figure 5; variations in results between the 3 different periods are minimal.

[3]Flow divergence is considered to be minimal at this level, while vertical motion is maximal; the geostrophic approximation will be most valid for this field and hence the accuracy of the barotropic approximations of early models was considered optimal for this field (Baumhefner 2000, private communication).

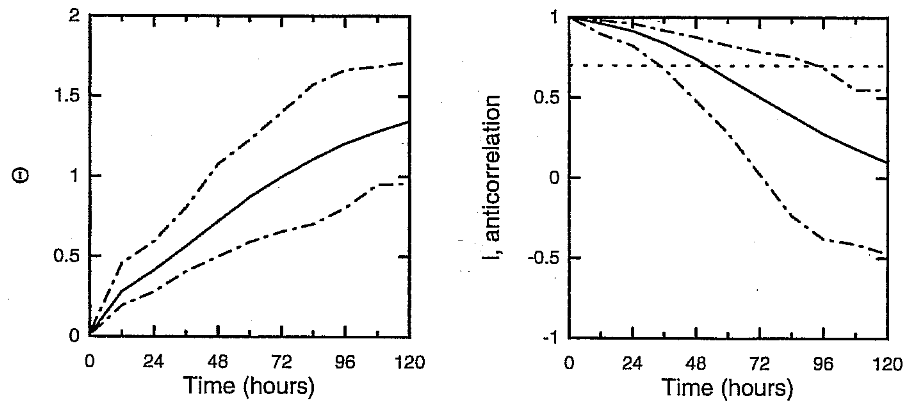

Figure 5: Linearity results for ECMWF operational twin SV perturbations ($\tau_{\rm opt} = 48$ hours), calculated using 500 h Pa geopotential height data over the Northern Hemisphere excluding the Tropics and taken over 25 days. The panels show the mean (solid line) and extent (dot-dashed lines) of error made in assuming linearity as measured by Θ (left) and the (anti-)correlation between twin pairs (right).

The mean difference in magnitude between the positive and negative perturbations is found to be less than 5% of their summed magnitudes; the nonlinearity is almost all due to the perturbations evolving in different directions, consistent with the average (anti-)correlation of 0.75 at 48 hours. A relative nonlinearity of 0.70 at optimization time, and a correlation of 0.75 suggest that, for SV optimized over 48 hours, *SV ensembles are not internally consistent.* Inasmuch as the (nonlinear) trajectories of the ensemble members *must have* diverged far from the (linearly) evolved images of the singular vectors used to define them, there are no grounds for assuming that the ensemble members share the desired properties of the singular subspace which they were designed to approximate. In particular, one should expect there to be other directions in which perturbations will have grown more over the interval $[0, \tau_{\rm opt}]$.

The relevance of the linear propagator to the nonlinear trajectories may be restored either by reducing the magnitude of the perturbation, by shortening the optimization time interval or both. It is desirable that the perturbation magnitude reflect the analysis uncertainty (Molteni *et al.* 1996), and, since the optimization time is less constrained by other considerations, the effect of shortening the interval $[0, \tau_{\rm opt}]$ is investigated in the next subsection. Note that a particular (imperfect) TLM may be improved by reducing the sources of error discussed in section 2 (Mahfouf 1999).

3.1.3 Construction & evaluation of ECMWF SV optimized over 24h

If the linearity results are independent of optimization time in the case of the ECMWF model (as they are for the annulus as described in section 2), then the relative nonlinearity for 24h SV perturbations at the optimization time of 24 hours will be the same as that for the 48h SV perturbations at 24 hours, *i.e.* on average less than 0.4. Such values would certainly suggest the assumptions of linear growth of the perturbations during the optimization time interval to be reasonable, and consequently, the dynamics of the 24h SV ensembles to be closer to those envisaged.

For comparison of perturbations optimized over the different times, two sets of SV ensembles were constructed using the ECMWF operational models, for 7 cases from 1–7 January 1998: the first set were perturbations optimized over 48 hours (48h SV) while the second set were perturbations optimized over the shorter time interval of 24 hours (24h SV).

As described in section 33.11 above, the constant of proportionality, $\sqrt{R_n}$, used in defining the SV is determined in order that the ensemble standard deviation should be comparable to the estimated error of the ensemble mean. Comparison of the amplitudes of 24h SV with those of 48h SV (at times from 48 to 84 hours) in the total energy norm for 1 case (1 January 1998) suggests that for the 24h SV

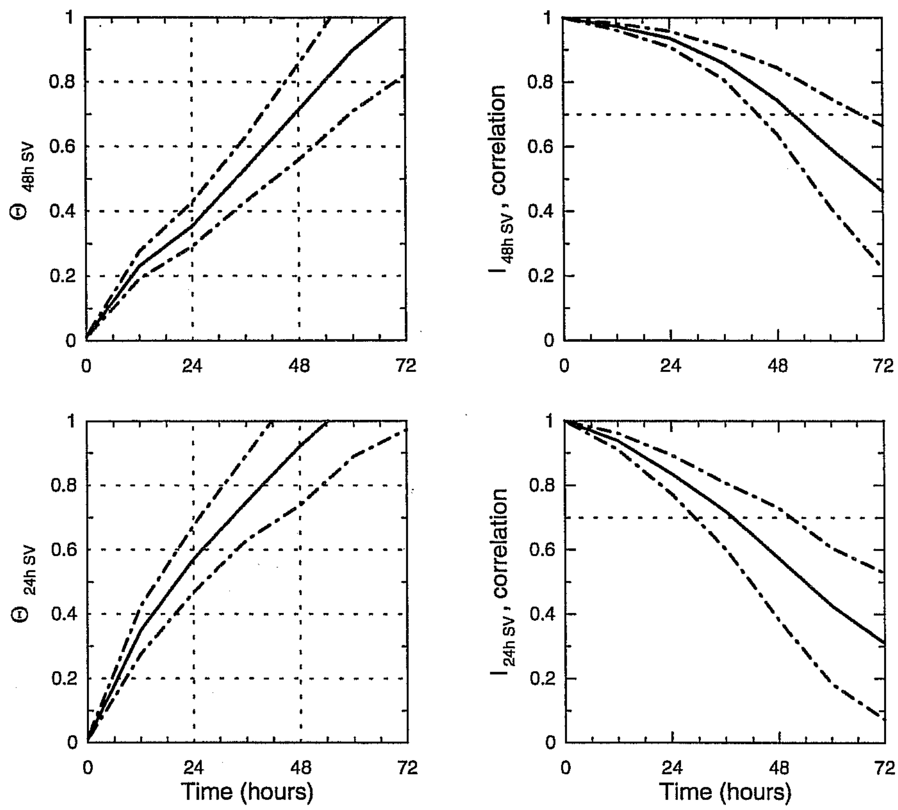

Figure 6: Linearity results for ECMWF 48h SV (upper panel) and 24h SV (lower panel) perturbations, calculated using 500 h Pa geopotential height data over the Northern Hemisphere taken over the 7 days 1.i.98-7.i.98. The panels show the mean (solid line) and extent (dot-dashed lines) of the relative nonlinearity as measured by Θ (left) and the (anti-)correlation (right) for a total of 175 twin pairs.

the constant of proportionality should be increased from $\sqrt{R_n} = \sqrt{0.6}$ to $\sqrt{R_n} = 1.0$; a 23% increase in the initial magnitude of 24h SV perturbations. For results presented here the 24h SV perturbations were constructed using this increased value of $\sqrt{R_n}$ (while the 48h SV are calculated using the original $\sqrt{R_n} = \sqrt{0.6}$); we revert to the 500 h Pa geopotential height norm.

Results for the 48h SV linearity tests from the week of 1–7 January 1998 are similar to those not only for the 25 cases in December 1996 and January 1997 (cf figure 5 and the upper panel of figure 6) but also for the months of January and August 1998; the relative nonlinearity of the 48h SV at the optimization time of 48 hours is ~ 0.7, as is the correlation. From figure 3 it can be seen that changing the optimization time of SV perturbations in the annulus has little effect on the linearity results. In contrast, changing the optimization time of SV perturbations in the ECMWF numerical weather model has a much larger effect on the linearity results (see figure 6). The average error in assuming linear evolution at the optimization time of 24 hours is $\sim 55\%$ of the average evolved perturbation magnitude, with an average correlation of 0.84. These higher values show that, on average, the 24h SV perturbations are better described by the linear approximation at their optimization time (of 24 hours) than the 48h SV perturbations are at 48 hours. The 24h SV results are not, however, sufficient to claim that reducing the optimization time to 24 hours renders SV ensembles internally consistent.

Comparing the evolution of SV perturbations to their optimization time, 24h SV evolve more linearly than 48h SV, thereby making 24h SV more internally consistent. The evolution of the 24h SV is, however, less linear over the 0-48 hour period than that of the 48h SV and it is of interest to ask whether this is due to more rapid amplification or more nonlinear evolution of the 24h SV compared to that of the 48h SV. For the single case of 1 January 1998 the amplification factors, given

	48h SV	24h SV
$t_0 = 0, \quad t_1 = 24$	4.7	4.8
$t_0 = 24, \quad t_1 = 48$	2.6	2.5
$t_0 = 0, \quad t_1 = 48$	12.5	12.2

Table 1: Amplification factors for 48h SV and 24h SV over the periods 0–24 hours, 24–48 hours and 0–48 hours, calculated using the 500 h Pa geopotential height data.

by $\frac{\|\delta^{\pm}(t=t_1)\|}{\|\delta^{\pm}(t=t_0)\|}$ over a period t_0–t_1, were calculated for the first (non-rotated) 48h SV and 24h SV. Over the 0–24 hour, 24–48 hour and 0–48 hour periods the amplification rates for 48h SV and 24h SV were found to be similar, as shown in table 1. This suggests that the less linear evolution of the 24h SV over the 0-48 hour period is not due to the 24h SV growing more rapidly than the 48h SV[4].

Although it is not the purpose of this paper to enter into a discussion on the performance measures used to quantify the success of SV ensembles, it is of interest to note that, according to anomaly correlation coefficient (ACC) and root mean square (RMS) skill and spread scores (as defined in Buizza (1995)), the 48h and 24h SV ensembles score almost identically over the 7 cases considered, as shown in figure 7. The 24h SV ensembles are more internally consistent, computationally less expensive and as successful as their 48h counterparts.

3.2 Evaluation of NCEP BV ensembles

Breeding vectors (BV) were developed with the aim of capturing "the possible growing error fields in the analysis" (Toth and Kalnay 1997), using only past observations and model dynamics (Toth and Kalnay 1993; Toth and Kalnay 1997; Toth *et al.* 1999). In theory they can be shown to converge to the (local) orientation of finite-sample Lyapunov vectors in certain limits (Toth and Kalnay 1993; Toth *et al.* 1999; Vannitsem and Nicolis 1997; Ziehmann *et al.* 1999; Ziehmann *et al.* 2000). One of these limits implies that the linear approximation holds.

For a given magnitude of perturbation and particular breeding cycle time τ_c, the practical connection between the BV ensembles and the validity of the linearity approximation arises from the operational implementation of BV ensembles[5]; we consider the implementation of NCEP (Toth and Kalnay 1993). Initially a perturbation is both added to and subtracted from the analysis and the two perturbed values are integrated forward a time τ_c. There is then a choice of how to define the new BV orientation (see figure 8): either as the difference between either the positive or the negative perturbation and the evolved analysis (method A), or as the difference between the two integrated perturbations (method B). Whichever method is used, the new BV is then added to and subtracted from the next analysis value and the process repeated. The methods are only equivalent if evolution

[4]Note that if the linear approximation was exact and the TLM perfect, then the leading 48h SV would always have an amplification factor at 24 hours which was less than or equal to that of the leading 24h SV.

[5]Note that the operational implementation of the BV ensemble does *not* involve the use of the TLM, and hence the additional error sources related to the TLM, as discussed in section 2, are not of concern here.

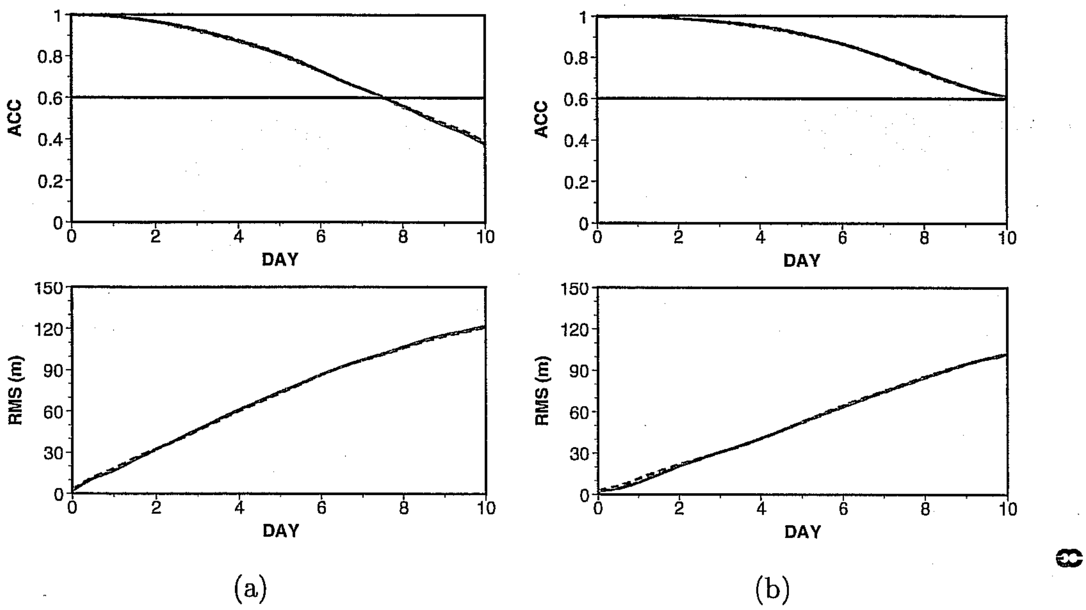

Figure 7: Spread and skill scores for the 48h SV (solid line) and 24h SV (dashed line) ECMWF ensembles, taken over 7 days, 1.i.98-7.i.98. The skill scores shown are the mean anomaly correlation coefficient (ACC) between the (evolved) ensemble members and the corresponding analysis (upper left panel) and the mean root mean square (RMS) distance between the ensemble members and the analysis (lower left panel). The spread scores shown are the mean ACC between the ensembles members and the control (upper right panel) and the mean RMS error, again between the ensemble members and the control (lower right panel).

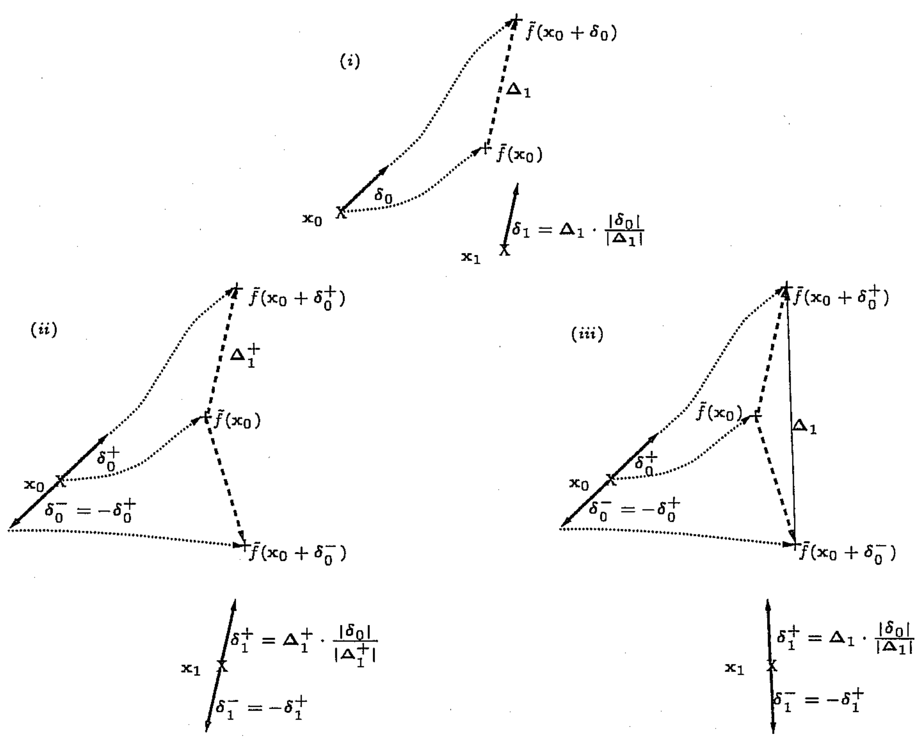

Figure 8: Schematics of breeding. (*i*) The ideas of breeding vectors: A perturbation δ_0 (solid line at $t=0$) is added to the analysis $\mathbf{x}_0$ at initial time and both are integrated forward under the model (dotted lines). The evolved perturbation $\boldsymbol{\Delta}_1$ (dashed line) is then rescaled to the magnitude $|\delta_0|$ of the initial perturbation to give the new BV $\boldsymbol{\delta}_1$ (solid line at $t=1$) which is used as the perturbation from the new analysis $\mathbf{x}_1$. If the dynamics are not linear there is a choice of how to generate BV twins. (*ii*) Method A: Chose one of the evolved twins (here the positive) over the other, re-scale and introduce its symmetric image. (*iii*) Method B: Alternatively the difference between the evolved twins may be rescaled and its symmetric image introduced.

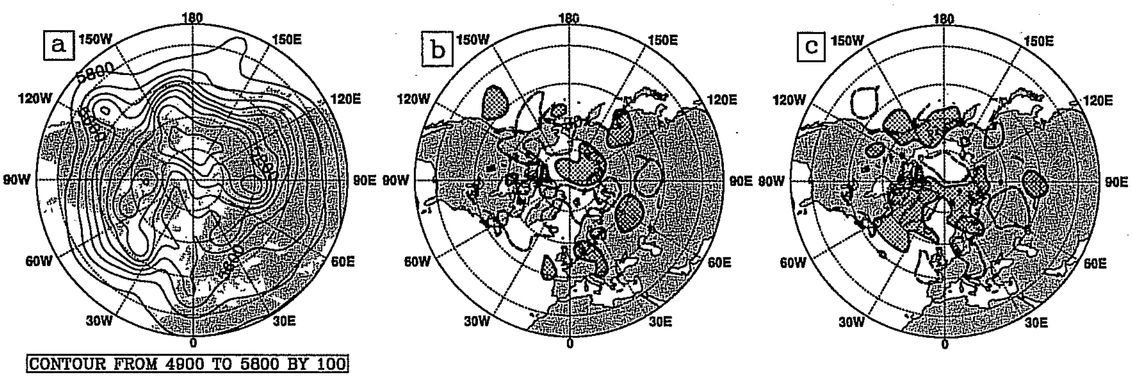

Figure 9: b) and c) 500 h Pa anomaly fields for a twin pair of evolved BV perturbations (negative anomalies less than -30 m outlined, positive anomalies greater than +30 m stippled) along with a) the corresponding evolved control. The perturbations were initiated at 0000 UTC on 25.xi.97 and evolved forwards 48 hours to give a forecast for 0000 UTC on 23.xi.97; shown here is their deviation from the evolved control in the 500 h Pa field for the area over which measures are calculated, namely 22.5°N-90°N.

is linear over the time interval $[0, \tau_c]$ for the magnitude ϵ_{BV}. Method B is used at NCEP (Toth *et al.* 1997).

In the case of the RBF model of the annulus, the similarity between method A and method B BV is found to be inversely related to the initial perturbation magnitude (*i.e.* less similar for larger magnitude), as expected. While the temperatures observed in the annulus can change by more than 1°C in a time step, initial perturbations need to be on the order of 10^{-3} °C for the method A and method B perturbations to coincide. Further, method A BV perturbations are found to better capture the analysis error orientation (see Gilmour (1998) for details).

During the period considered in this study (2 October - 26 November 1997), NCEP BV ensembles consisted of 5 twin pair perturbations, initiated at 0000 UTC and bred using a cycle time of 24 hours[6] and scaled using the kinetic energy norm to a magnitude equal to the estimated seasonal analysis error[7]; they were both constructed and evolved at T62L28 resolution (Toth *et al.* 1997). To ease comparison with the results for the ECMWF SV ensembles, the results for NCEP BV ensembles are presented using the 500 h Pa geopotential height norm over the Northern Hemisphere excluding the Tropics (22.5°N-90°N). Results are taken over 25 different cases, giving 125 twin BV pairs in total; the BV considered were initiated at 0000 UTC on each of 2, 4-6, 8-17, 20-22, 24-27 October and 21, 23, 25, 26 November, all[8] in 1997. An example of the 500 h Pa anomaly fields of an evolved pair of twin BV members is shown in figure 9, along with the corresponding evolved control.

The BV ensemble formation scheme assumes that the evolution of the operational perturbations is approximately linear for the breeding cycle time $\tau_c = 24$ hours. In fact, at τ_c the average error in assuming the linear approximation is about 50%. (The relative nonlinearity has an average value of 0.47, and the correlation an average of 0.89, see figure 10.) As in the case of the SV ensemble members, the nonlinearity which is present is mainly due to the perturbations evolving in different directions rather than having differing magnitudes. While the relative nonlinearity of BV perturbations at 24 hours is less than that of either 48h SV or 24h SV at that time, the error is however substantial. It would be of interest to see how BV ensembles constructed using method A perform in the operational

[6]The cycle time was previously 6 hours (Toth and Kalnay 1993), changed to 24 hours for computing considerations in 1994 (Toth and Kalnay 1997) but will return to 6 hours in 2001 (Toth 2000, private communication) to better coincide with the analysis cycle; a reduced cycle time should result in reduced nonlinearities.

[7]A regional rescaling is also applied; for details see Toth and Kalnay (1997).

[8]Discontinuities (missing days) in BV cases studied reflect difficulties in real-time data acquisition.

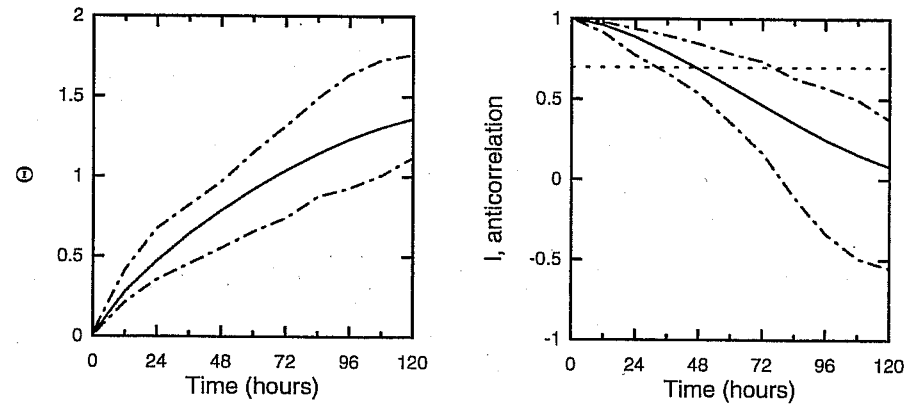

Figure 10: Linearity results for NCEP operational twin BV perturbations, calculated using 500 h Pa geopotential height data over the Northern Hemisphere excluding the Tropics and taken over 25 days. The panels are as in figure 5.

model; in contrast to method B, which becomes ambiguous as the linear approximation fails, method A makes no assumption about linear evolution. Of course, the correspondence to Lyapunov vectors requires the linear approximation to hold, in which case method A and method B coincide. In practice one may argue that the bred vectors need not converge to the Lyapunov vectors, but to do so removes the most common justification of the use of BV ensembles, specifically that they reflect the directions of sustained error growth (see the discussion in Smith (1995) and references therein).

4 Related work

Computational limitations have fueled interest in the TLMs of numerical weather models by reason of their potential application in three domains: sensitivity analysis of the model to varying parameters, the assimilation of observations to yield analysis values, and the selection of state dependent (or targeted) observations. As noted in section 3 there are two approaches to determining the duration of the linear regime: comparing the evolution of perturbations under the nonlinear model with that under a TLM, and comparing different perturbations evolved under the nonlinear model. While either of these approaches provide an estimate of the duration of the linear regime, most previous investigations use random perturbations and simplified models; few results could be found for SV or BV perturbations in operational NWP models.

Comparison of evolution under the TLM with that under the nonlinear model is more commonplace, and is usually motivated by issues concerning data assimilation; classification of what is a good approximation is obviously subjective being both user dependent and application dependent and thus it varies between investigations. Lacarra and Talagrand (1988) consider perturbations with amplitude comparable to that of forecast errors found in data assimilation in a barotropic (f-plane shallow water) model. They conclude that the TLM is a good approximation for 'ranges of up to about 48 hours', emphasizing that baroclinic instabilities have a stronger influence than barotropic instabilities. Vukićević (1991) considers perturbations given by initial data errors in a primitive equation limited-area model (which includes both baroclinic and barotropic components); the TLM is found to give a good approximation for 1-1.5 days. A low resolution (T21L19) primitive equation model is used by Rabier and Courtier (1992), with perturbations whose initial magnitude is 'far from being negligible'. Those authors show that the evolution of the eddy part is essentially linear for a range of 1-2 days, while the total evolution of zonal and eddy parts is less linear. Errico *et al.* (1993), using the same mesoscale model as Vukićević (1991), studies evolution to 72 hours of random perturbations with magnitudes comparable to analysis errors in the absence of moist physical processes (in which case more linear behavior is expected (Errico and Raeder 1999)). For both a summer case and a winter case the correlations between linear and nonlinear evolution remain high out to 72 hours; the authors note that the boundary conditions imposed artificially constrain perturbation growth. For a moist convective cloud model Park and Droegemeier (1997) find the results to be very sensitive to perturbation magnitude, and also to the frequency with which the basic state, used in calculating the TLM, was updated. Errico and Raeder (1999) compare the evolution of SV perturbations under various TLM with that under the nonlinear primitive equation moist physics model (similar to that used by (Vukićević 1991)) for a summer and a winter case. TLMs which both include and exclude linearizations of the moist processes are used and their accuracy as an approximation to the nonlinear evolution compared out to 48 hours. A variety of initial perturbation magnitudes are also contrasted, showing that agreement between linear and nonlinear evolution is larger for initially smaller perturbations, as expected. Buizza and Montani (1999) compare the nonlinear evolution of the SV perturbation, under the ECMWF NWP model, with that of the pseudo-inverse initial perturbation calculated using the TLM. (Both perturbations are evolved at the same resolution, T63L19.) Results indicate that the differences between the two integrations are mainly due to linear processes not included in the TLM version of the model, rather than to nonlinear effects.

The investigations which compare the evolution of various perturbations under the nonlinear model are usually motivated by ensemble forecasting issues. Houtekamer and Derome (1994) use a quasi-non-divergent global spectral model of low resolution (T21) to examine whether the mean of evolved twin pair BV perturbations, with magnitudes comparable to analysis errors, yields a higher quality forecast than the control (run at the same resolution). Noting that the mean of the bred perturbations and the evolved control can only differ if the evolution of the bred perturbations is nonlinear, we can deduce from figure 3 of that paper that the duration of the linear regime is at most 2 days. Buizza (1995) directly investigates the time after which "nonlinear processes can not be neglected" using an ECMWF primitive-equation model (with T63 resolution); the twin SV perturbations are optimized for 36 hours at a lower resolution (T21) and have a variety of initial amplitudes. Buizza quantifies the contribution of nonlinear processes both by considering the correlation (ℓ) of the perturbations, and by algebraically manipulating the information given by the nonlinearly evolved twin perturbations to describe it as a truncated Taylor expansion, enabling comparison of the amplitude of nonlinear terms with those of the linear terms without defining a TLM. Both methods suggest that nonlinear processes become "important" (for the correlation method this is taken to be when $\ell < 0.7$) after 2-2.5 days when the initial amplitude is comparable to analysis error estimates.

Previous investigations suggest that the duration of the linear regime varies from 24 hours to at least 3 days. This variation is unsurprising since the magnitude and orientation of perturbations differ, the model considered is not unique, the definition of validity varies between studies, and the statistics used tend to be necessary, but not sufficient, to establish the validity of the linear regime. Nevertheless, the new results presented in the current paper suggest this not to be the case for present day operational models: the large values of Θ observed at 12 and 24 hours indicate that the inferred range of the linear regime of operational models has been significantly overestimated. Just how widespread this overestimation is can be easily determined by examining the evolution of Θ in other systems.

5 Discussion and Generalizations

Previous sections presented spatially (hemispheric) averaged results focusing on the 500 h Pa geopotential height field. In this section, the results of alternative fields are summarized. In addition, patterns of the breakdown of linearity in physical space are examined in a particular case. As mentioned in section 1 the duration of the linear regime directly impacts other issues; we conclude this section by discussing a few of these.

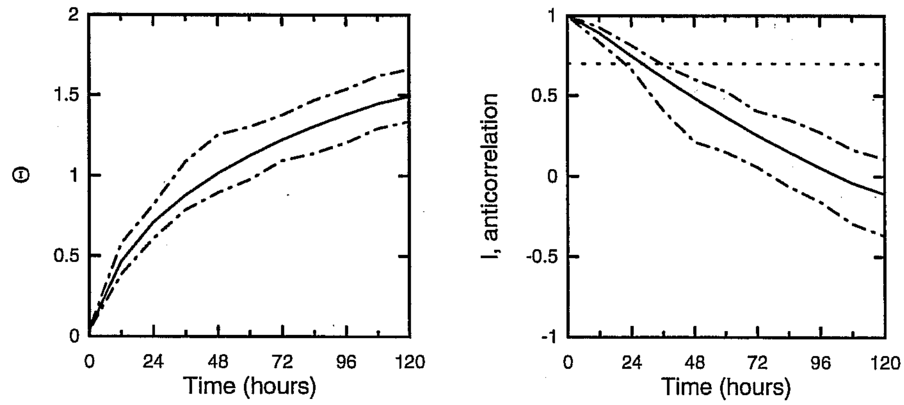

Figure 11: Linearity results for NCEP operational twin BV perturbations, calculated using the horizontal wind velocity at 500 h Pa data over the Northern Hemisphere excluding the Tropics and taken over 25 days. The panels are as in figure 5.

The relative nonlinearity results for NWP model/ensemble configurations presented above are calculated using a specified norm over a given region and averaging over a certain number of perturbations and cases. Similar calculations were carried out for the NCEP BV configuration using as the norm both the horizontal wind velocity at 500 h Pa (uv500) field (see figure11) and the temperature at 2m (t2m) and using the total energy as the norm for ECMWF SV configuration (results not shown). Given that the t2m field is considered to lie within the boundary layer, initial perturbations are adjusted so as to be consistent with the surface layer and hence are not always symmetric at initial time; nevertheless linearity results using both the t2m norm and the z500 norm saturate to similar Θ values at similar times. Of the norms considered, the z500 norm was the 'most linear' and the total energy norm the least linear. The impact of using different geographical regions was less noticeable: of the BV results for the Northern Hemisphere without Tropics, the Northern Hemisphere, North America and the Tropics the only significant variation was in the results over the Tropics, where the spread of Θ values at any given time is larger (results not shown).

These spatially-averaged results over large geographical regions do not, however, give any insight as to whether the nonlinearity is homogeneous over all grid points, or whether it is locally centered. While insight can be gained from examining plots of fields for each twin pair at each time step, such as that in figure 4, a plot summarizing the behavior of all twin perturbations may be derived using the relative nonlinearity measure. The relative nonlinearity will take a value of 2 whenever both members of a twin pair evolve to be either greater than or less than the evolved control if a scalar measure such as one based on geopotential height is used; an example of this maybe be seen in figure 4 at 80°N, 180°E where both members of the twin pair have evolved to be greater than the evolved control. By computing the fraction of twin pairs for which the relative nonlinearity is 2 at each grid point we may see the distribution of strong nonlinear evolution. In order to discount small perturbations which may affect the results, one can consider only the twin pairs in which both perturbations are over a certain threshold magnitude.

Figure 12 shows the results of such a computation for the ECMWF 48 hour forecasts initiated at 1200 UTC 19 December 1996 and discounting twin pairs in which either of the perturbations are less than 10 m; there are definite local centers of strong nonlinear evolution. Figures like this one also enable the locations of regions of strong nonlinearity to be compared to synoptic structures, seen in the corresponding control forecast of the lower panel; for the case shown there is some correlation between the regions of strong nonlinearity and the synoptic structures. Further work on identifying such correlations is being pursued.

The localized nature of the nonlinearity suggests that the results will be affected by both the

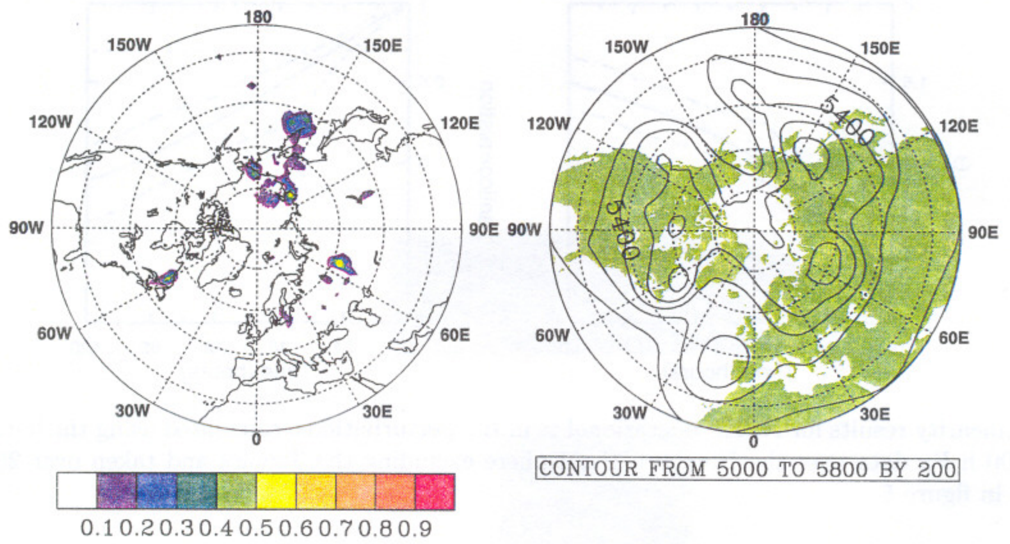

Figure 12: ECMWF 48 hour forecasts initiated at 1200 UTC on 19.xii.96 for 22.5°N-90°N. The fraction of pairs of SV twin perturbations over 10 m with maximum local nonlinearity (*i.e.* those for which $\Theta = 2$ at that grid point) is denoted by the shading (as shown by the color bar) (left panel). Note that regions of strong nonlinearity reflect the synoptic structures ('lows') seen on the low resolution control forecast (right panel); some degree of correlation is evident.

resolution at which the perturbations are evolved and the resolution at which the relative nonlinearity is calculated. Results from the ECMWF SV configuration for the *single case* of 1 January 1998 suggest that evolution of SV perturbations at higher resolution (T_l159) enhances nonlinearities which are present at lower (T42) resolution.

The duration of the linear regime for perturbations of a realistic magnitude should be tested whenever singular vectors are employed. In the selection of state dependent (or targeted) observations (Hansen 1998; Joly *et al.* 1999; Lorenz and Emanuel 1998; Snyder 1996), the aim is to select the additional observations which, when included in the forecast, are most likely to reduce the forecast error; one must consider not only the spatial locations at which the current analysis is most uncertain, but also the likely amplification factor by which this uncertainty will grow. TLMs and adjoint models have been used to target observations using singular vectors (Buizza and Montani 1999; Gelaro *et al.* 1999; Langland *et al.* 1999; Montani *et al.* 1999) and the results are encouraging (see Gelaro *et al.* (1999) and other papers in that volume); objective targeting based on linear approximations has been shown to reduce forecast error by up to $\sim 15\%$ on average over the the area of interest (the 'verification region'); see *e.g.* Buizza and Montani (1999)[9]. Singular vectors are a prime candidate for identifying which uncertainties are likely to be amplified the most *as long as* the magnitude of all uncertainties are sufficiently small so as to evolve approximately linearly (and the model sufficiently accurate). There are numerous factors which undoubtedly limit error reduction: an inexact TLM, the limited number of SVs, details of the assimilation scheme. The results presented here suggest that the poor validity of the linear approximation may be a further factor and calculations of the distribution of relative nonlinearity have been shown to be useful in the targeting context: Hansen and Smith (2000) suggests that the unproductive nature of the Lorenz and Emanuel (1998) adaptive observation strategy is unsurprising given that the error in the linear approximation "is well over 100%", as measured by the relative nonlinearity, Θ.

[9]Other techniques for targeting exist; see *e.g.* Szunyogh *et al.* (2000) which reports a 10-20% error reduction using an "Ensemble Transform" technique.

Calculation of the Θ measure is straight-forward provided that the ensemble consists of twinned perturbations. The fundamental idea of using ensemble trajectories to quantify the inaccuracy in the linear approximation may be generalized to cases where the ensemble members are not twinned by fitting an empirical linear map to the evolution of ensemble members in any suitable (empirical) basis; the inaccuracy of the linear approximation could be estimated via the forecast error of this empirical map with respect to the observed (nonlinear) evolution. Note that such empirical maps (either linear, or better still nonlinear) could be used to significantly increase the ensemble size over the duration for which the maps are a useful approximation, at least for the subspace sampled by the initial perturbations.

6 Summary

In this paper we have introduced a new measure to evaluate the duration of the linear regime and illustrated that the commonly accepted range of values, 2 to 3 days, is a significant overestimate of the relevance of the linear regime for the operational perturbations employed by NCEP as well as those of ECMWF. The implications this holds for operational NWP ensemble construction have been explored; modifying the SV ensemble construction by shortening the optimization time has been shown to yield trajectories more consistent with the linear framework which originally justified focusing attention upon them. Additionally, the modified SV ensemble was shown to give similar skill scores.

The assumption that "small" operational perturbations evolve linearly is commonplace in many aspects of forecasting. These include ensemble construction, data assimilation, and the selection of adaptive observations. Our results provide a simple approach for verifying that the operational perturbations are indeed small in this sense. We hope this simple test will find widespread employment wherever properties of linear growth are invoked, given that the duration of the linear regime appears to have been often overestimated.

Acknowledgments

This paper is based on part of the doctoral thesis of the first author (Gilmour 1998); these results could not have been obtained without the assistance of Zoltan Toth and Tim Palmer in the discussion of the operational ensembles and in obtaining the data. We thank ECMWF for use of computing facilities and provision of data. We have benefited from discussions with Dave Broomhead, Myles Allen, Christine Ziehmann, Joe Tribbia, Martin Ehrendorfer, Jan Barkmeijer, Jim Hansen and Tom Hamill. The reviews of Zoltan Toth and two anonymous referees provided useful comments and suggestions. This work was supported by an EPSRC research grant GR/K77617 and the ONR Predictability DRI under grant N00014-99-1-0056.

References

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

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

———, and A. Montani, 1999: Targeting observations using singular vectors. *J. Atmos. Sci.*, **56**, 2965–2985.

———, and T.N. Palmer, 1995: The singular-vector structure of the atmospheric general circulation. *J. Atmos. Sci.*, **52**, 1434–1456.

———, T. Petroliagis, T.N. Palmer, J. Barkmeijer, M. Hamrud, A. Hollingsworth, A. Simmons, and N. Wedi, 1998: Impact of model resolution and ensemble size on the performance of an Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **124**, 1935–1960.

ECMWF, 1999: *Proceedings of a workshop held at ECMWF on Predictability, 20–22 October, 1997.* ECMWF.

Errico, R.M., and K.D. Raeder, 1999: An examination of the accuracy of the linearization of a mesoscale model with moist physics. *Quart. J. Roy. Meteor. Soc.*, **125**, 169–195.

———, T. Vukićević, and K. Raeder, 1993: Examination of the accuracy of a tangent linear model. *Tellus A*, **45**, 462–477.

Eubank, S., and J.D. Farmer, 1990: An introduction to chaos and randomness. *Lectures in Complex Systems*, E. Jen, Ed., Vol. Lecture II, *SFI Studies in the Sciences of Complexity.* Addison-Wesley.

Früh, W.-G., and P.L. Read, 1997: Wave interactions and the transition to chaos of baroclinic waves in a thermally driven rotating annulus. *Phil. Trans. Roy. Soc. London A*, **355**, 101–153.

Gelaro, R., R.H. Langland, G.D. Rohali, and T.E. Rosmond, 1999: An assessment of the singular-vector approach to targeted observing using the FASTEX dataset. *Quart. J. Roy. Meteor. Soc.*, **125**, 3299–3327.

Gilmour, I., 1998: *Nonlinear model evaluation: ι-shadowing, probabilistic prediction and weather forecasting.* PhD thesis, University of Oxford, U.K. 184 pp.

Hamill, T.M., D.P. Baumhefner, C. Snyder, S.L. Mullen, and Z. Toth, 2000: Ensemble forecasting in the short to medium range: report from a workshop. *Bull. Amer. Meteor. Soc.* Submitted.

Hansen, J. A., 1998: *Adaptive Observations in Spatially-extended, Nonlinear Dynamical Systems.* PhD thesis, University of Oxford, U.K. 207 pp.

Hansen, J.A., and L.A. Smith, 2000: The role of operational constraints in selecting supplementary observations. *J. Atmos. Sci.* In press.

Hide, R., 1958: An experimental study of thermal convection in a rotating fluid. *Phil. Trans. Roy. Soc. London A*, **250**, 441–478.

———, and P.J. Mason, 1975: Sloping convection in a rotating fluid. *Adv. Phys.*, **24**, 47–100.

Houtekamer, P.L., and J. Derome, 1994: Prediction experiments with two-member ensembles. *Mon. Wea. Rev.*, **122**, 2179–2191.

——, and ——, 1995: Methods for ensemble prediction. *Mon. Wea. Rev.*, **123**, 2181–2196.

Joly, A., K.A. Browning, P. Bessemoulin, J.P. Cammas, G. Caniaux, J.P. Chalon, S.A. Clough, R. Dirks, K.A. Emanuel, L. Eymard, R. Gall, T.D. Hewson, P.H. Hildebrand, D. Jorgensen, F. Lalaurette, R.H. Langland, Y. Lemaitre, P. Mascart, J.A. Moore, P.O.G. Persson, F. Roux, M.A. Shapiro, C. Snyder, Z. Toth, and R.M. Wakimoto, 1999: Overview of the field phase of the Fronts and Atlantic Storm-Track EXperiment (FASTEX) project. *Quart. J. Roy. Meteor. Soc.*, **125**, 3131–3163.

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

Langland, R.H., R. Gelaro, G.D. Rohali, and M.A. Shapiro, 1999: Targeted observations in FASTEX: adjoint based targeting procedure and data impact experiments in IOP17 and IOP18. *Quart. J. Roy. Meteor. Soc.*, **125**, 3241–3270.

Lorenz, E., and K. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.*, **55**, 399–414.

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

Mahfouf, J., 1999: Influence of physical processes on the tangent-linear approximation. *Tellus A*, **51**, 147–166.

Molteni, F., R. Buizza, T.N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–120.

Montani, A., A.J. Thorpe, R. Buizza, and P. Unden, 1999: Forcast skill of the ECMWF model using targeted observations during FASTEX. *Quart. J. Roy. Meteor. Soc.*, **125**, 3219–3240.

Mureau, R., F. Molteni, and T.N. Palmer, 1993: Ensemble prediction using dynamically conditioned perturbations. *Quart. J. Roy. Meteor. Soc.*, **119**, 299–323.

Palmer, T.N., R. Gelaro, J. Barkmeijer, and R. Buizza, 1998: Singular vectors, metrics, and adaptive observations. *J. Atmos. Sci.*, **55**, 633–653.

——, R. Buizza, F. Molteni, Y.Q. Chen, and S. Corti, 1994: Singular vectors and the predictability of weather and climate. *Phil. Trans. Roy. Soc. London A*, **348**, 459–475.

Park, S.K., and K.K. Droegemeier, 1997: Validity of the tangent linear approximation in a moist convective cloud model. *Mon. Wea. Rev.*, **125**, 3320–3340.

Rabier, F., and P. Courtier, 1992: Four-dimensional assimilation in the presence of baroclinic instability. *Quart. J. Roy. Meteor. Soc.*, **118**, 649–672.

Read, P.L., 1992: Applications of singular systems analysis to baroclinic chaos. *Physica D*, **58**, 455–468.

Smith, L. A., 1992: Identification and prediction of low dimensional dynamics. *Physica D*, **58**, 50–76.

Smith, L.A., 1995: Accountability and error in ensemble prediction of baroclinc flows. *Predictability(1995)*, Vol. 1, 351–368, ECMWF. Shinfield Park, Reading, Berks., U.K. Proceedings of a seminar held at ECMWF, 4–8 September, 1995.

Snyder, C., 1996: Summary of an informal workshop on adaptive observations and FASTEX. *Bull. Amer. Meteor. Soc.*, **77**, 953–961.

Strang, G., 1986: *Introduction to applied mathematics.* Wellesley-Cambridge Press, 184 pp.

Stroe, R., and J.F. Royer, 1993: Comparison of different error growth formulas and predicability estimation in numerical extended-range forecasts. *Ann. Geophys.*, **11**, 296–316.

Szunyogh, I., E. Kalnay, and Z. Toth, 1997: A comparison of Lyapunov and optimal vectors in a low-resolution GCM. *Tellus A*, **49**, 200–227.

———, Z. Toth, R.E. Morss, S.J. Majumdar, B.J. Etherton, and C.H. Bishop, 2000: The effect of targeted dropsonde observations during the 1999 Winter Storm Reconnaissance Program. *Mon. Wea. Rev.*, **128**, 3520–3537.

Talagrand, O., and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation - part I: theory. *Quart. J. Roy. Meteor. Soc.*, **113**, 1311–1328.

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

———, and ———, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319.

———, I. Szunyogh, E. Kalnay, and G. Iyengar, 1999: Coments on: "Notes on the appropriateness of 'bred modes' for generating initial perturbations". *Tellus A*, **51**, 442–449.

———, E. Kalnay, S.M. Tracton, R. Wobus, and J. Irwin, 1997: A synoptic evaluation of the NCEP ensemble. *Wea. Forecasting*, **12**, 140–153.

Vannitsem, S., and C. Nicolis, 1997: Lyapunov vectors and error growth patterns in a T21L3 quasi-geostrophic model. *J. Atmos. Sci.*, **54**, 347–361.

Vukićević, T., 1991: Nonlinear and linear evolution of initial forecast errors. *Mon. Wea. Rev.*, **119**, 1602–1611.

Ziehmann, C., L.A. Smith, and J. Kurths, 1999: The bootstrap and Lyapunov exponents in deterministic chaos. *Physica D*, **126**, 49–59.

———, ———, and ———, 2000: Localized Lyapunov exponents and the prediction of predictability. *Phys. Letts. A*. In press.