Tellus (1999), 51A, 431–441
Printed in UK – all rights reserved

TELLUS
ISSN 0280–6495

LETTERS TO THE EDITOR

Notes on the appropriateness of "bred modes" for generating initial perturbations used in ensemble prediction

By RONALD M. ERRICO[1]* and ROLF LANGLAND[2], [1]*National Center for Atmospheric Research*†, *Boulder, Colorado, USA;* [2]*Naval Research Laboratory, Monterey, California, USA*

(Manuscript received 3 February 1998; in final form 24 September 1998)

ABSTRACT

Papers by Szunyogh and co-workers and Iyengar and co-workers, among others, claim that "bred growing vectors" (BGVs) are appropriate for generating initial perturbations for application to ensemble weather prediction. The theoretical bases for this claim are purported relationships between the bred vectors and what are called local Lyapunov vectors (LLVs). Several statements in these papers are inaccurate, however, regarding: the properties of LLVs in contrast to those of singular vectors; the relationship between LLVs and BGVs; the relationship between BGVs and analysis errors; the relationship between singular vectors and dynamic imbalances; and the adequacy of very low-resolution models to characterize singular vectors of models used for numerical weather prediction. These inaccuracies and their implications for the use of BGVs in ensemble weather prediction are discussed here.

1. Introduction

For several reasons, weather prediction should be posed as a statistical, rather than strictly deterministic, problem. Numerical forecast models, their boundary conditions, and their initial conditions typically have errors which subsequently make forecasts inaccurate. All these errors are unknown, and therefore only their statistics can be described. Although models that determine some statistics of forecast errors from statistics of model and initial condition errors can be derived (Epstein, 1969; Fleming, 1971; Ehrendorfer, 1994), nonlinearity of the forecast model and the resulting coupling of statistical moments of the errors makes this currently impractical for weather prediction models. An alternative, perhaps more computationally efficient strategy is to estimate the statistics from an ensemble of forecasts (Leith, 1974). These forecasts are generated using the same (Molteni et al., 1996; Toth and Kalnay, 1993) or different (Houtekamer et al., 1996) models starting from a set of several, differently perturbed initial conditions. Palmer (1996) and Ehrendorfer (1997) provide two recent reviews of this problem.

Since operational ensemble forecasting began at the European Centre for Medium Range Weather Forecasts (ECMWF) and the National Centers for Environmental Prediction (NCEP), there has been both public and private controversy about the appropriateness of their very different methods employed for generating the ensembles' initial perturbations (see Anderson, 1996; Houtekamer and Derome, 1995; Szunyogh et al.,

* Corresponding author: National Center for Atmospheric Research, PO Box 3000, Boulder, CO 80307–3000; e-mail: rmerrico@ucar.edu

† The National Center for Atmospheric Research is sponsored by the US National Science Foundation.

1997; Palmer et al., 1998; Ehrendorfer, 1997). Singular vectors (SVs) are used at ECMWF and "breeding of growing vectors" (BGVs, or "bred modes") are used at NCEP. These two perturbation techniques and the resulting ensembles of forecasts have extremely different characteristics.

The paper by Szunyogh et al. (1997; hereafter denoted SKT) summarizes many of the dynamical arguments that are invoked to justify the use of BGVs in ensemble prediction. In particular, they suggest a relationship between BGVs and local Lyapunov vectors (LLVs) to claim the appropriateness of BGVs. The conference preprint by Iyengar et al. (1996; hereafter denoted ITKW) also claims to justify their appropriateness based on purported relationships between BGVs and initial condition errors. There are several additional papers that make similar claims (Toth and Kalnay, 1993; Pu et al., 1997; Toth et al., 1997; Tracton and Kalnay, 1993).

Several statements in SKT and ITKW, are erroneous, inaccurate, or require clarification. These statements regard: (1) the properties of LLVs in contrast to those of SVs, (2) the relationship between LLVs and BGVs, (3) the relationship between BGVs and analysis errors, (4) the relationship between SVs and dynamic imbalances, and (5) the adequacy of very low-resolution models to characterize SVs of models used for numerical weather prediction. These issues will be described in separate sections of this note.

2. Local Lyapunov vectors and SVs

The definition and properties of LLVs presented in SKT and their references are not clear in many respects. Clearer definitions of LLVs are presented in several places (Eckhardt and Yao, 1993, and Abarbanel et al., 1993), but most of these references concentrate on evaluation of local Lyapunov exponents, particularly the single leading exponent, which is a measure of growth rate rather than the structure of a vector. In contrast, a focus on the vectors themselves, including consideration of the space of leading vectors and their relation to SVs, appears in Legras and Vautard (1996). The latter's definition will be repeated here in a slightly altered form to make the consideration of a norm more explicit (as in Ehrendorfer and Errico, 1995, but with less generality). Although similar developments appear in many other places, the repeated presentation facilitates the comparison of SVs and LLVs here.

A nonlinear model

$$\boldsymbol{x}(t_2) = \mathscr{M}[\boldsymbol{x}(t_1), t_2] \tag{2.1}$$

will be considered, where $\mathscr{M}$ describes the relationship between model state vectors $\boldsymbol{x}$ at time t_1 and subsequent time t_2. The SVs will be obtained for a corresponding linearized model

$$\boldsymbol{y}(t_2) = \boldsymbol{M}(t_1, t_2)\boldsymbol{y}(t_1), \tag{2.2}$$

where $\boldsymbol{y}$ refers to perturbations of the state vector and matrix $\boldsymbol{M}$ is the propagator specified by partial derivatives of the components of $\boldsymbol{x}(t_2)$ with respect to those of $\boldsymbol{x}(t_1)$.

In the meteorological literature (Farrell, 1988; Farrell and Ioannou, 1996; Barkmeijer et al., 1993; Buizza and Palmer, 1995; Gelaro et al., 1998; Palmer et al., 1998; Ehrendorfer and Tribbia, 1997; Ehrendorfer and Errico, 1995; and others), SVs are often denoted as solutions of the problem: determine the pair $y(t_1)$, $y(t_2)$ such that

$$\begin{aligned} &\boldsymbol{y}^{\mathrm{T}}(t_2)\boldsymbol{C}\boldsymbol{y}(t_2) \\ &\text{is maximum, given} \quad \boldsymbol{y}^{\mathrm{T}}(t_1)\boldsymbol{C}\boldsymbol{y}(t_1) = 1 \end{aligned} \tag{2.3}$$

with (2.2) as a strong constraint, where superscript T indicates a transpose and $\boldsymbol{C}$ is a symmetric, positive-definite matrix that defines an ℓ_2-norm:

$$\langle \boldsymbol{y}_2, \boldsymbol{y}_1 \rangle = \boldsymbol{y}_2^{\mathrm{T}}\boldsymbol{C}\boldsymbol{y}_1. \tag{2.4}$$

The solution of (2.2–2.3) may be determined using a Lagrange multiplier λ: maximize

$$\begin{aligned} J[\boldsymbol{y}(t_1)] = {}& \boldsymbol{y}^{\mathrm{T}}(t_1)\boldsymbol{M}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{M}\boldsymbol{y}(t_1) \\ &+ \lambda[1 - \boldsymbol{y}^{\mathrm{T}}(t_1)\boldsymbol{C}\boldsymbol{y}(t_1)]. \end{aligned} \tag{2.5}$$

The solutions satisfy

$$\boldsymbol{M}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{M}\boldsymbol{y}(t_1) = \lambda\boldsymbol{C}\boldsymbol{y}(t_1). \tag{2.6}$$

With the change of variable $\boldsymbol{z} = \boldsymbol{C}^{1/2}\boldsymbol{y}(t_1)$, (2.6) takes the form of an eigenvalue problem for a symmetric operator:

$$\boldsymbol{C}^{-1/2}\boldsymbol{M}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{M}\boldsymbol{C}^{-1/2}\boldsymbol{z} = \lambda\boldsymbol{z}. \tag{2.7}$$

The constraint on the perturbations at the initial time becomes $\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z} = 1$.

There are as many solutions to (2.7) as the size of the state vector. These may be ordered $\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots$ based on the corresponding order $\lambda_1 \geqslant \lambda_2 \geqslant \ldots$. If $\lambda_1 \neq \lambda_2$, only $\boldsymbol{y}_1 = \boldsymbol{C}^{-1/2}\boldsymbol{z}_1$ is the solution to (2.2–2.3) as stated, but the corresponding solutions for the other λ_n may be considered additional solutions if the previous ones are excluded. The $\boldsymbol{z}_n$ are right singular vectors of $\boldsymbol{C}^{1/2}\boldsymbol{M}\boldsymbol{C}^{-1/2}$, but the corresponding $\boldsymbol{y}_n$ are called the SVs of $\boldsymbol{M}$ with respect to the norm defined by the weighting $\boldsymbol{C}$ (see Ehrendorfer and Errico, 1995 for some generalizations). The λ_n are squares of the singular values of $\boldsymbol{M}$ with respect to the norm expressed by $\boldsymbol{C}$ (i.e., of $\boldsymbol{C}^{1/2}\boldsymbol{M}\boldsymbol{C}^{-1/2}$).

What SKT call an LLV is apparently what Legras and Vautard (1996) call a backward Lyapunov vector. The latter is a vector $\boldsymbol{y}_n(t_2) = \boldsymbol{M}\boldsymbol{y}_n(t_1)$ obtained from (2.6) in the limit that $t_1 \to -\infty$. This is in contrast to a forward Lyapunov vector that is a $\boldsymbol{y}_n(t_1)$ obtained from (2.6) in the limit that $t_2 \to \infty$. In this sense, a LLV is a special, limiting case of a time-evolved SV.

Several general LLV results are independent of the specific norm selected for defining the corresponding SVs. Legras and Vautard (1996) state that: (1) the leading LLV is norm independent, (2) the space spanned by the leading m LLVs is norm independent, and (3) Lyapunov exponents

$$e_n = \lim_{t_1 \to -\infty} \frac{1}{t_2 - t_1} \log \lambda_n, \tag{2.8}$$

corresponding to the LLVs are norm independent. Other important properties of the LLVs, however, are norm dependent. In particular: (1) instantaneous growth rates are norm dependent, as is the result of eq. (16) of SKT, since a measure of growth of a solution whose structural shape is not preserved in time must depend on how different components are weighted, e.g., on the norm used; (2) when orthogonalization is required, as when a distinct set of m leading LLVs are to be determined by orthogonalization, a norm must be specified; and (3) the time period required for finite-period SVs to closely approximate corresponding infinite-period LLVs can be norm dependent. For these reasons, the consideration and selection of a suitable norm should not be quickly dismissed. Another critical property of a norm to be considered is discussed in Section 5.

It should be noted that SVs describe behaviors of corresponding perturbations in the nonlinear model (2.1) only in a tangent linear sense; i.e., as long as the perturbations remain small. Otherwise, the perturbed result in (2.1) would depend nonlinearly on the amplitudes of the components of the perturbation $\boldsymbol{y}$. In applications of SVs to numerical weather prediction models (Buizza et al., 1993, 1997; Gelaro et al., 1998; Ehrendorfer and Errico 1995; and others), the time period $t_2 - t_1$ over which the SVs are computed is usually restricted to 3 days or less to ensure that, when begun from perturbations the sizes of the uncertainty of initial analyses, the perturbations do not grow so much that the linearization becomes invalid at some time during the period. A 2–3 day time period has been shown acceptable for dry models, but may be significantly shorter when linearizations of moist physics are considered (Ehrendorfer et al., 1999; Errico and Raeder, 1999).

The limitation of perturbation size also applies to any application of LLVs: If they are supposed to describe behaviors in a nonlinear model, the initial perturbation size or the time period of applicability must be limited. If initial perturbations are meaningfully chosen to be the size of initial condition (i.e., analysis) uncertainty, then it is the time period that must be restricted. This restriction, however, inherently contradicts the requirement that LLVs start in the infinite past. Although an SV determined over a finite time period may be a good approximation to an LLV, unless such a time period is not longer than 3 days, the SV cannot be associated with both an LLV and the behavior of a relevant perturbation in the nonlinear model. Neither can it then be claimed that the LLV describes the perturbation behavior in the nonlinear model. Although claims have been made in SKT that convergence to LLVs occurs for periods as short as 3 days based on evidence from simple or low-resolution models, periods longer than 5 days appear necessary in a T21 quasi-geostrophic model with realistic climatology (Reynolds and Errico, 1999).

Although SKT state that all perturbations tend toward the same LLV eventually, that is only in the tangent linear context or for infinitesimal size perturbations in the nonlinear context. In the non-infinitesimal nonlinear context, in fact, chaos prevails, and almost all perturbations (one exception being a displacement precisely along the control trajectory) eventually diverge until they saturate at a perturbation variance twice that of the local model climatology (Lorenz, 1963, 1982, 1993; Tribbia and Baumhefner, 1988; Simmons et al., 1995). In other words, two forecasts become as unalike as randomly chosen forecasts for the same day from different years (excluding the additional effects of climate drift or low-frequency climatic variability).

3. Lyapunov vectors and bred modes

The leading LLV valid at time t_2 can be generated by repeating a forecast with the nonlinear model begun far enough in the past, but initialized with a very small perturbation, and then rescaling the perturbation during the forecast as required to keep the perturbation size small. The rescaling must be done using a single scaling factor applied to all the fields, otherwise the result is no longer identical to the LLV computed for the tangent linear version of the model; i.e., in the limit of a long forecast with infinitesimal perturbation. As long as the initial perturbation has non-zero projection onto the vector that evolves into the first LLV, the dominance of that vector will ensure that it appears as the perturbed result at the end of a sufficiently long forecast. Even if there is initially no projection onto the vector that evolves into the leading LLV, in numerical practice that LLV will still likely appear as further perturbations are introduced due to small nonlinear and computational effects (Legras and Vautard, 1996). For that reason it can be difficult to obtain even approximations to many subdominant LLVs. This method of generating LLVs can be termed "breeding" although it is critically different from the BGV breeding method used at NCEP, for reasons to be made clear in the remainder of this paper.

Although the modes bred using the method just described will converge to at least the leading LLVs, the forecast time period required for convergence will depend on the spread between distinct singular values. If the spread is not so great after a short time, as suggested by the studies of Ehrendorfer and Errico (1995), Gelaro et al. (1998), and Buizza et al. (1997), then such a short time would be insufficient to claim that the result corresponds to the single, most dominant LLV, although the result should of course be expected to be dominated by a larger set of leading LLVs. The degree to which bred modes begun from different randomly-perturbed initial conditions still appear dissimilar after a forecast period indicates a lack of convergence to the leading LLV.

What makes the breeding described here most unlike BGV at NCEP, is their application of what they call a "mask" in place of the rescaling of all the fields by a single scalar. It is not clear how this "mask" is applied, but from ITKW, it appears that it is a rescaling factor that varies with location on the globe based on some estimate of analysis uncertainty. This may partially explain why the 50 kPa height field of the BGVs in the northern hemisphere usually have maximum amplitudes poleward of 60° (see Section 9). Since this is a location-dependent rescaling, it causes the BGVs to depart from LLVs that would be obtained from the forecast model when no such scaling is used.

Also it should be noted that if the perturbations produced by the BGV method are determined using differences between perturbed nonlinear forecasts and a "control forecast" produced by the periodic cycling of the unperturbed model during the NCEP data assimilation procedure, then it is not obvious that the theorems regarding LLVs for dynamical systems apply without modification (J. Anderson, personal communication). In the method just described, the control forecast has a stochastic element due to the effects of observations external to the model. Also, this "control forecast" is not temporally continuous. More properly, it should be considered as a stochastically-driven dynamical system, for which the driving "noise", determined by the observations assimilated, is neither temporally or spatially white.

4. Bred modes and analysis errors

It is sometimes claimed that bred modes or LLVs describe analysis errors. Is this claim consistent with the behavior of perturbations (i.e., differences or errors) in forecasts and assimilation systems?

An analysis is produced by blending two components: one is given by observations and the other by a prior estimate of the atmospheric state produced by a previous short-term forecast. Due to the latter, errors introduced into an earlier forecast will spatially and temporally propagate through the data assimilation system. If such errors remain unconstrained by subsequent observations, they would eventually grow sufficiently large that nonlinear effects would become important, resulting in chaos and a departure of the analysis from reality (Lorenz, 1993). That such divergence from reality is not observed is due to the constraining influence of the observations. The character of the errors in the analyses therefore depends critically on the nature of this constraining influence.

SVs of short enough time period can tell us something about possible analysis errors under certain conditions. In regions where no observations are present, the analysis error is identical to the error in the prior estimate. The leading SVs of the short-period forecast used to produce the prior estimate indicate where errors initially present in that forecast will grow the most. Only if the initial error in that forecast projects sufficiently on the leading initial SVs, will the concluding SVs also indicate where errors in the prior estimate are largest. For longer periods, however, SVs can suggest little about analysis error, since observations will not remain absent throughout the extensive region of the evolving SVs over long periods. In particular, since LLVs are simply SVs defined for such long periods, they should not be expected to describe analysis errors produced by an assimilation system, even approximately.

Some information about the nature of observations can be introduced during a breeding process by a modified rescaling, as apparently done with BGV at NCEP (ITKW). In this case, the results should depart dramatically from those of the LLVs without such rescaling if the profound effects of observations are incorporated. Also, details of the rescaling become critical and no such breeding results can be properly interpreted without knowing the exact form of this “masking” and without seeing its effect on the otherwise non-rescaled perturbations.

Estimating the structures of real analysis errors is very difficult. Even estimating statistics of the structures of analyses errors is not easy (Houtekamer and Mitchell, 1998; Reynolds and Palmer, 1998). Estimating the variances of analysis errors is much easier, but these tell us little about error structures.

It should be mentioned that the NCEP bred modes do not at all resemble differences among analyses produced at different centers. The BGV structures have extremely large horizontal scale and large variance. In the northern hemisphere, they tend to be dominant northward of the storm tracks, rather than within them (see Section 9).

ITKW claim that bred modes are representative of analysis errors, as referenced by, e.g., Pu et al. (1997) and Toth et al. (1997). This claim is partly based on the fact that the masking coefficients applied to the BGVs are determined from analysis error statistics derived from differences between a pair of corresponding NCEP analyses. For the other comparisons made in ITKW, analysis error is mistaken for other kinds of errors. For example, ITKW make comparisons of bred modes with spatially averaged innovations (actually combinations of forecast and observation error), with analysis residuals (actually combinations of observation and analysis error), and with 24-h ensemble spread intended as a proxy for 6-h forecast error. No explanation is given in ITKW as to why any of these latter should be considered indicative of analysis error.

5. SVs and quasi-geostrophy

SKT misinterpret the results of Ehrendorfer and Errico (1995; hereafter denoted EE) regarding effects of geostrophic adjustment on SV determination. EE found that with their model, geostrophic adjustment significantly contributed to growth of an energy-based norm. They recognized this as unphysical, because past studies of predictability (Daley, 1981), stability analysis (Errico, 1981, 1984), and general experience with forecast models have shown that deep gravitational modes, which are responsible for geostrophic adjustment on time scales of a few hours, simply propagate without increasing their amplitude or physically appropriate measures of energy. In fact, if their propagation is not strictly adiabatic, their amplitude tends to damp (Daley, 1981; Errico and Williamson, 1988; Gelaro and Errico, 1993). Furthermore, in the limited-area model used by EE, fast-propagating gravity waves quickly passed out of the model domain, thereby reducing local-measurements of gravity-wave magnitudes (Errico and Baumhefner, 1987).

EE carefully traced their anomalous result to two sources of “error.” One was an actual bug in their model software that affected the nonlinear vertical mode initialization scheme (following Bourke and McGregor (1983)) applied at the start of the forecast, although it did not affect a corresponding uninitialized result that showed the same growth over 2 h due to geostrophic adjustment.

The other "error" was that the energy norm used,

$$E=\sum_{i,j}\Delta A_{i,j}\left[\frac{RT_r}{p_r^2}\,p_{i,j}'^2+\sum_k\Delta\sigma_k\times\left(u_{i,j,k}'^2+v_{i,j,k}'^2+\frac{c_p}{T_r}\,T_{i,j,k}'^2\right)\right], \quad (5.1)$$

where u, v, T, c_p, $\Delta\sigma$ have their usual meaning, i, j are horizontal indices, ΔA is a grid area, k is a vertical index, p is surface pressure, and subscript r refers to a spatially invariant reference value, was inconsistent with the model's vertical finite difference scheme. When the vertically continuous forms of the adiabatic primitive equations are linearized about an isothermal, resting state, the integral (rather than summation over sigma level) form of (5.1) is conserved in time (Talagrand, 1981). More importantly, contributions to E by distinct vertical modes in such a linearized continuous model are independent. Consequently, contributions to E by rotational and gravitational modes (the latter corresponding to inertial gravity waves) are also independent (Errico, 1989). In the inconsistent model formulation, however, E expressed in terms of these modes had contributions by cross terms; i.e., terms involving products of rotational and gravitational components.

In EE, it was possible to actually *increase* E by *filtering* gravitational modes due to the contribution to E by cross terms. This filtering was accomplished by gravity waves either propagating out of the limited model domain (i.e., local geostrophic adjustment) or by the instantaneous effect of the vertical mode initialization scheme. Due to this mechanism, EE obtained much greater growth of E than reported by Gelaro et al. (1998) or Buizza et al. (1997), among others, for other models. When EE used an expression of quadratic energy consistent with their model's finite difference scheme, this new E was in fact reduced by the filtering of gravity waves. Also, when they used a norm that excluded gravitational modes, they obtained maximum growth rates very similar to those that others had reported (Ehrendorfer et al., 1999).

Subsequently, the model used by EE has been modified to be energy consistent with respect to the usual E norm so that rotational and gravitational modes do contribute independently. The consequence, reported in Ehrendorfer et al. (1997), is that the SVs in the dry version of the energy-consistent model are in fact nearly balanced. There is negligible effect on the norm due to vertical mode initialization applied to the SV; the growth of the norm with time is smooth, with no noticeable effects by any short term adjustment; and, when compared with an initially geostrophically constrained SV solution, the SV structures and growth rates are fairly similar. These results are consistent with the result obtained at ECMWF several years ago that their normal mode initialization had negligible impact on the norm (P. Courtier, personal communication). There is thus ample evidence that the fast growth of E reported for SVs in all the cited works is not due to significant dynamic imbalances in the SV fields.

This last fact therefore begs the question: Why do SKT obtain SVs for which E grows due to geostrophic adjustment? They claim that their model vertical differencing is consistent with conservation of E, but is this the same E as in (5.1)? Note that E defined by (5.1) is not the model kinetic plus potential energy (the latter is not quadratic in temperature). Neither is this E the kinetic plus available potential energy, as comparison with eq. (57) in Lorenz (1960) suggests, except when the model has been linearized about an isothermal, resting state. Note also that (5.1) applies to perturbation fields, not the full model fields, and that, in general, a model formulation that conserves an expression for energy in terms of the full model fields does not necessarily yield a conserved expression in a corresponding linearized model applicable to perturbed fields. Rather, the conservation property described by Talagrand (1981) should be explicitly checked. Furthermore, SKT say they do some kind of vertical interpolation of their fields before determining the sum over k in (5.1). This would likely render the sum no longer consistent with the model's finite difference scheme. Hence, their expression for E likely no longer has the necessary property of independent contributions by rotational and gravitational components, and may result in a similarly gravity-wave influenced norm as used by EE. This is the most likely explanation of their results.

It should also be noted that while there is evidence suggesting that 6-h forecast errors from global models are dominated by geostrophic components in the extratropics (Daley, 1991), it is less clear what the partitioning is in the analysis errors, where the more gross geostrophic errors have presumably been corrected by assimilation of observations. It is therefore not yet clear to what degree a perturbation that is is intended to be consistent with analysis error statistics should in fact be geostrophic or dynamically balanced.

6. Inadequacy of model

The SVs presented in SKT look nothing like those determined by Ehrendorfer and Errico (1995), Buizza et al. (1997), Gelaro et al. (1998), etc. This is likely because the T10 truncation makes the fields much too smooth, almost zonal, judging from the appearance of the stream function in Fig. 1 in SKT. Even the zonal structure is not very realistic judging from the cross section shown in Fig. 1b in SKT, although it is not stated whether a zonal mean or longitudinal slice is being plotted there. As a result of this smoothness, the SV problem being determined by SKT is more like a Charney (1947) baroclinic instability, normal mode problem than the usual SV problem determined for ensemble numerical weather prediction models. In particular, nearly redundant eigenvalues λ_n are likely in the T10 model, corresponding to the indeterminacy of the phases of the unstable modes in Charney's problem. Even the spectrum of unstable modes in Charney's problem may not be adequately resolved. Certainly the transfer of perturbation energy between small and large scales reported by Buizza et al. (1997) cannot be replicated in a T10 model. The generality of conclusions made at this resolution to results made at the resolutions (even T21) employed in the other studies cited here should therefore not be taken for granted.

While the works by Trevisan and Legnani (1995) and Pires et al. (1996) are claimed to support many of the conclusions in SKT, those works are based on the Lorenz (1963) model. While this model is extremely useful for demonstrating certain critically important concepts, it is dangerous to use it to infer quantitative behavior in numerical weather prediction models. The Lorenz (1963) model has only one dominant unstable direction, and knowing the character of that component alone is sufficient to describe much of the system behavior. In contrast, knowing only the leading LLV in an operational numerical weather prediction model is grossly inadequate for describing the model's unstable manifold. In particular, the way a large model ingests observations distributed inhomogeneously in space and time, and the lack of great separation between its singular values suggests that greater care should be taken in generalizing quantitative results from assimilation experiments or SV convergence tests from the Lorenz model.

7. Summary

Legras and Vautard (1996) define backward Lyapunov vectors as end-time SVs begun at an infinite time in the past. These are apparently what Szunyogh et al. (1997) call LLVs. From this definition, many properties of the LLVs are easily deduced. In particular, what results are norm independent is made clear. When orthogonalization of a set of vectors is required to achieve independence, a particular norm must be selected. Also, instantaneous growth rates are necessarily norm dependent. These properties of LLVs are not stated in Szunyogh et al. (1997).

LLVs of the model are significantly different from BGVs since a non-scalar re-scaling is periodically applied to the latter. Model LLVs do not describe analysis errors because such errors are constrained by observations, unlike model LLVs. Neither do they describe the behavior of typical forecast errors beyond a few days (at most) because the behaviors of such errors quickly become nonlinear.

No credible theoretical reason why BGVs should resemble analysis errors has been offered. The behavior of perturbations in the data assimilation modeling system, either when constrained by observations or unconstrained so that the behavior becomes nonlinear, should not be expected to be like the behaviors of BGVs. Most of the comparisons cited in Iyengar et al. (1996) are invalid because they are made with respect to estimates of observation or forecast errors, not strictly analysis errors. The support of Pires et al. (1996) using the Lorenz (1963) model is insufficient because of the model's simplicity with respect to characteristics of the ingestion of observations in operational data assimilation systems.

If there is any resemblance of BGV results to analysis errors, it must be built in by the "masking" applied. This "masking" procedure therefore becomes critical to the appropriateness of BGVs for ensemble prediction or other applications. It is imperative that the "mask" and how it is applied be thoroughly described and its alteration of the forecast component of the breeding be thoroughly documented.

The re-interpretation by Szunyogh et al. (1997) of the geostrophic adjustment problem described by Ehrendorfer and Errico (1995) is erroneous, as a great wealth of studies indicate. The large growth rates reported for SVs in the literature using dry model linearizations are not due to initial geostrophic imbalances in the SVs (the result reported by Ehrendorfer and Errico is an exception). Properly computed, with an appropriate norm, these SVs are devoid of imbalances that project onto fast gravity waves. Consequently, the dynamic imbalances obtained by SKT in their SV determinations are likely due to either an energy-inconsistent formulation of vertical finite-difference equations in their model or due to an inconsistent computation of the energy.

In conclusion, anyone using BGVs should carefully investigate how they were produced and what their structures and magnitudes are before using them in any application. They should also carefully consider any theory claiming to justify their use for any particular application. While lack of a theory explaining their appropriateness for ensemble prediction does not imply that they are not useful for that purpose in practice, it does suggest that the BGV method should be more carefully evaluated in the future.

Also, it should be noted that the SVs computed for dry tangent linear and adjoint models using the energy norm described in most of the literature are not grossly geostrophically unbalanced structures. The growth rates observed are consistent with behaviors within the model's slow manifold. In examples where this is not the case, as in SKT, the properties and appropriateness of the implemented norm with respect to the contributions and behavior of gravitational modes should be carefully investigated.

8. Acknowledgments

The authors are grateful to the comments and encouragement provided by T. Palmer, M. Ehrendorfer, J. Anderson, R. Gelaro, R. Daley, J. Tribbia, G. Branstator, C. Reynolds, and L. Fillion on drafting this manuscript. The manuscript, however, does not necessarily reflect all the opinions of those named. Helpful comments were also provided by anonomous reviewers. The figures appearing in Section 9 were provided by D. Baumhefner.

9. Appendix

An example of bred modes

As an example of BGM results, the 50 kPa geopotential heights in the northern extratropics for the 5 BGV perturbations which were used to generate the NCEP ensemble forecast starting 1 December 1995 are presented in Fig. 1. This date was randomly drawn from the period during which SKT was under review. Other dates during this period were also examined to ensure that the qualitative statements made here regarding their BGV structures, amplitudes, and mutual resemblances apply in general.

Note that the largest absolute values appear north of 60°N. The structures have a very large horizontal scale. The modes are not all the same, although modes 3 and 5 strongly resemble each

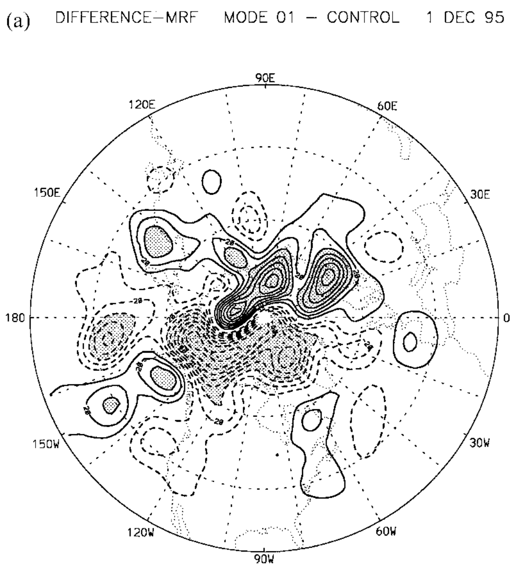

Fig. 1. Initial perturbations of the 50 kPa geopotential height field north of 25°N for the five distinct bred modes produced by the BGV method at NCEP for 1 December 1995. The contour interval is 10 m, negative-valued contours are dashed, and the 0-valued contour is omitted.

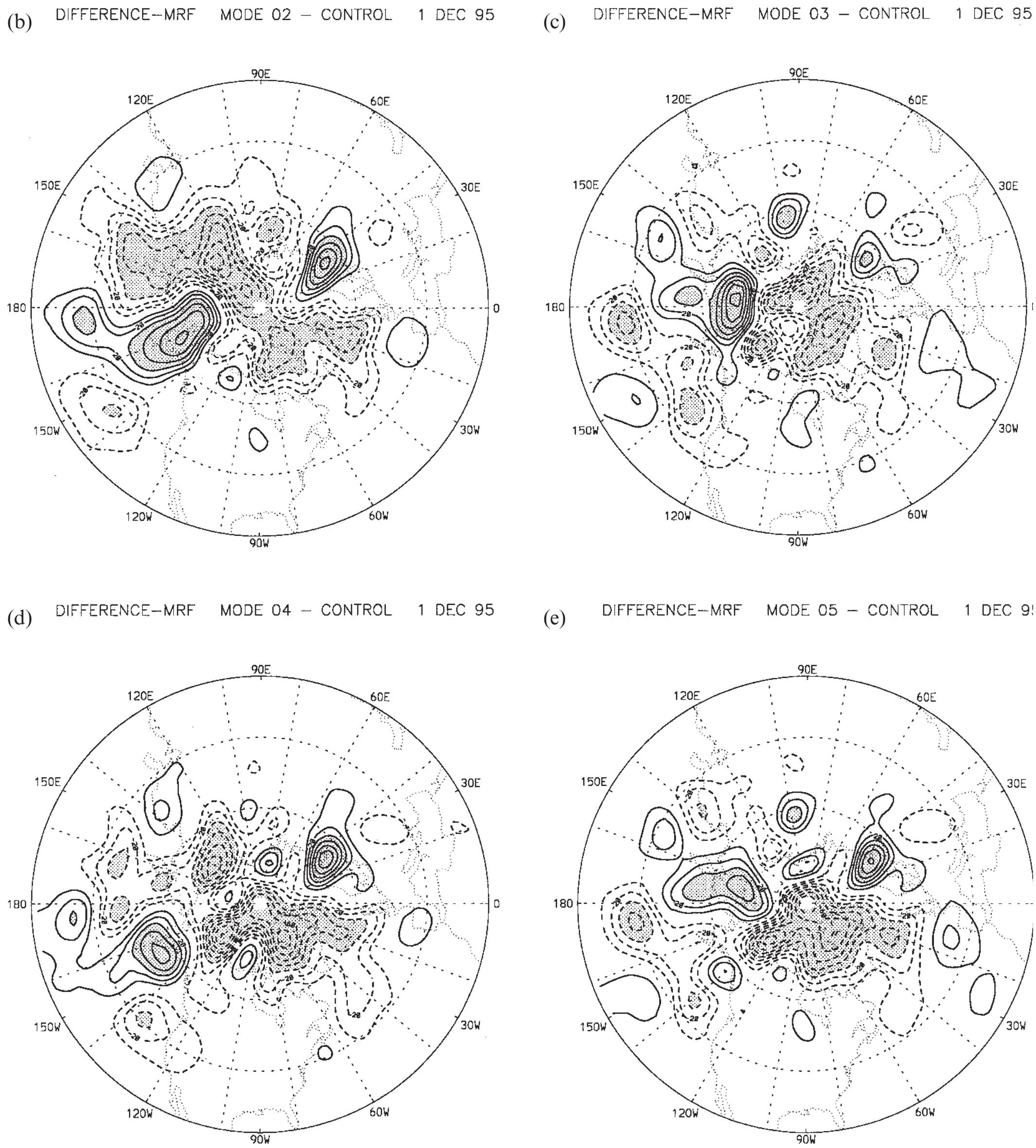

Fig. 1. (cont'd).

other as do modes 2 and 4: about one-half of the local extrema within each pair are colocated. Whether other fields for the same pairs of modes resemble each other as strongly is unknown to us from the data we have available.

The similarity and differences between modes raises two questions: why are the modes not more alike if they have nearly converged to the leading LLVs as claimed? Alternatively, for application to ensemble forecasting with very few members, why should individual perturbations as apparently similar (i.e., redundant) as some shown here be used?

REFERENCES

Abarbanel, H. D. I., Brown, R., Sidorowich, J. J. and Tsimring, L. S. 1993. The analysis of observed data in physical systems. *Rev. Modern Phys.* **65**, 1331–1392.

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

Barkmeijer, J., Houtekamer, P. and Wang, X. 1993. Validation of a skill prediction method. *Tellus* **45A**, 424–434.

Bourke, W. and McGregor, J. L. 1983. A nonlinear vertical mode initialization scheme for a limited area prediction model. *Mon. Wea. Rev.* **111**, 2285–2297.

Buizza, R., Tribbia, J., Molteni, F. and Palmer, T. N. 1993. Computation of optimal unstable structures for a numerical prediction model. *Tellus* **45A**, 388–407.

Buizza, R., Gelaro, R., Molteni, F. and Palmer, T. N. 1997. The impact of increased resolution on predictability studies with singular vectors. *Quart. J. Roy. Meteor. Soc.* **123**, 1007–1033.

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

Charney, J. G. 1947. The dynamics of long waves in a baroclinic westerly current. *J. Meteorol.* **4**, 135–162.

Daley, R. 1981. Predictability experiments with a baroclinic model. *Atmos. Ocean* **19**, 77–89.

Daley, R. 1991. *Atmospheric data analysis.* Cambridge University Press, 420 pp.

Eckhardt, B. and Yao, D. 1993. Local Lyapunov exponents in chaotic systems. *Physica* **D65**, 100–108.

Ehrendorfer, M. 1994. The Liouville equation and its potential usefulness for the prediction of forecast skill. Part I: *Theory. Mon. Wea. Rev.* **122**, 703–713.

Ehrendorfer, M. 1997. Predicting the uncertainty of numerical weather forecasts. *Meteorol. Zeitschrift* **6**, 147–183.

Ehrendorfer, M. and Errico, R. M. 1995. Mesoscale predictability and the spectrum of optimal perturbations. *J. Atmos. Sci.* **52**, 3475–3500.

Ehrendorfer, M. and Tribbia, J. J. 1997. Optimal prediction of forecast error covariances through singular vectors. *J. Atmos. Sci.* **54**, 286–313.

Ehrendorfer, M., Errico, R. M. and Raeder, K. D. 1999. Singular-vector perturbation growth in a primitive-equation model with moist physics. *J. Atmos. Sci.*, in press.

Epstein, E. S. 1969. Stochastic dynamic prediction. *Tellus* **21**, 730–759.

Errico, R. M. 1981. An analysis of interactions between geostrophic and ageostrophic modes in a simple model. *J. Atmos. Sci.* **38**, 544–553.

Errico, R. M. 1984. The statistical equilibrium solution of a primitive-equation model. *Tellus* **36A**, 42–51.

Errico, R. M. 1989. *Theory and application of nonlinear normal mode initialization.* NCAR Technical Note, NCAR/TN–344+IA, 145 pp.

Errico, R. M. and Baumhefner, D. P. 1987. Predictability experiments using a high-resolution limited-area model. *Mon. Wea. Rev.* **115**, 488–504.

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

Errico, R. M. and Williamson, D. L. 1988. The behavior of gravitational modes in numerical forecasts. *Mon. Wea. Rev.* **116**, 1737–1756.

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

Farrell, B. F. and Ioannou, P. J. 1996. Generalized stability theory. Part II: Nonautonomous operators. *J. Atmos. Sci.* **53**, 2041–2053.

Fleming, R. J. 1971. On stochastic dynamic prediction. I: The energetics of uncertainty and the question of closure. *Mon. Wea. Rev.* **99**, 851–872.

Gelaro, R., Buizza, R., Palmer, T. N. and Klinker, E. 1998. Sensitivity analysis of forecast errors and the construction of optimal perturbations using singular vectors. *J. Atmos. Sci.* **55**, 1012–1037.

Gelaro, R. and Errico, R. M. 1993. The forcing and balance of zonally symmetric modes in a global model. *Mon. Wea. Rev.* **121**, 470–481.

Houtekamer, P. L. and Derome, J. 1995. Methods for ensemble prediction. *Mon. Wea. Rev.* **122**, 2181–2196.

Houtekamer, P. L., Lefaivre, L., Derome, J., Ritchie, H. and Mitchell, H. L. 1996. A system simulation approach to ensemble prediction. *Mon. Wea. Rev.* **124**, 1225–1242.

Houtekamer, P. L. and Mitchell, H. L. 1998. Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.* **126**, 796–811.

Iyengar, G., Toth, Z., Kalnay, E. and Woolen, J. S. 1996. Are the bred vectors representative of analysis errors? Preprints, *11th Conference on Numerical weather prediction*, 19–23 August 1996, Norfolk, VA. American Meteorological Society, J64–J66.

Legras, B. and Vautard, R. 1996. A guide to Liapunov vectors. *Proceedings 1995 ECMWF Seminar on Predictability*, Vol. I, pp. 143–156.

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

Lorenz, E. N. 1960. Energy and numerical weather prediction. *Tellus* **12**, 364–373.

Lorenz, E. N. 1963. Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130–141.

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

Lorenz, E. N. 1993. The essence of chaos. University of Washington Press, 227 pp.

Molteni, F., Buizza, R., Palmer, T. N. and Petroliagis, T. 1996. The ECMWF ensemble prediction system:

Methodology and validation. *Quart. J. Roy. Meteor. Soc.* **122**, 73–119.

Palmer, T. N. 1996. Predictability of the atmosphere and oceans: from days to decades. *Proceedings 1995 ECMWF Seminar on Predictability*, Vol. I, pp. 83–141.

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

Pires, C., Vautard, R. and Talagrand, O. 1996. On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus* **48A**, 96–121.

Pu, Z.-X., Kalnay, E., Parrish, D., Wu, W. and Toth, Z. 1997. The use of bred vectors in the NCEP global 3-D variational analysis system. *Wea. Forecasting* **12**, 689–695.

Reynolds, C. A. and Errico, R. M. 1999. On the convergence of singular vectors toward Lyapunov vectors. *Mon. Wea. Rev.*, in press.

Reynolds, C. A. and Palmer, T. N. 1998. The impact of decaying singular vectors on forecast and analysis correction. *J. Atmos. Sci.* **55**, 3005–3023.

Simmons, A. J., Mureau, R. and Petroliagis, T. 1995. Error growth and estimates of predictability from the ECMWF forecasting system. *Quart. J. Roy. Meteor. Soc.* **121**, 1739–1771.

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

Talagrand, O. 1981. A study of the dynamics of four-dimensional data assimilation. *Tellus* **33**, 43–60.

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

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

Tracton, M. S. and Kalnay, E. 1993. Operational ensemble prediction at the National Meteorological Center: practical aspects. *Weather and Forecasting* **8**, 379–398.

Trevisan, A. and Legnani, R. 1995. Transient error growth and local predictability: A study in the Lorenz system. *Tellus* **47A**, 103–117.

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