

Four-dimensional variational assimilation and predictability in a quasi-geostrophic model

By KYLE SWANSON* and ROBERT VAUTARD, *Laboratoire de Météorologie Dynamique, Ecole Normale Supérieure, 24 Rue Lhomond, 75231 Paris Cedex 05, France*, and CARLOS PIRES, *2 Faculdade de Ciências, Departamento de Física, Edifício C1, Piso 4, Campo Grande, 1700 Lisboa, Portugal*

(Manuscript received 21 July 1997; in final form 20 February 1998)

ABSTRACT

Four-dimensional variational assimilation (4DVAR) of noisy observations in a multi-layer quasi-geostrophic model is studied, in both the perfect and imperfect model settings. Within the perfect model setting, the quality of the assimilated state improves significantly when the assimilation period is extended more than one week into the past. Specifically, when observations are supplied every 6 h, the squared error in the assimilated state at the end of the assimilation time period (the present) saturates at a value two orders of magnitude smaller than the imposed observational error for an assimilation period of 10 days. Further, this reduction in error occurs not only in measures explicitly minimized by 4DVAR, but for all standard measures of error. For realistic levels of observational error, the extension of forecast lead times is large, exceeding 15 days for global forecasts when the assimilation period is 10 days. This holds even for weather regime transitions, which are shown to be predictable at lead times of 10 days. The use of long assimilation periods extends forecast lead times approximately 5 days over the case when assimilation periods are on the order of one day. The structure of the analysis error when long assimilation period 4DVAR is applied is examined. This error is primarily concentrated in the midlatitude storm tracks. The reduction in analysis error is increasingly efficient at small scales as the assimilation period is increased; consequently, for long assimilation periods the analysis error projects strongly into the subspace of the leading Lyapunov vectors. The performance of 4DVAR in an imperfect model setting is also examined, and is found to depend upon the growth rate of the model errors. For rapidly growing model errors, extension of the assimilation period beyond about 1–2 days results in a degradation in the quality of the assimilated state as well as in the forecast quality. However, for model error growth rates similar to the growth rates of the leading Lyapunov vectors of the system, improvements in the assimilated state similar to those found for the perfect model are obtained. As such, it is estimated that assimilation times of 3–5 days for current levels of model error may improve the quality of assimilated states and forecasts in an operational setting.

1. Introduction

It is well established that dynamical motions in the extra-tropical atmosphere are fundamentally

chaotic. While this fact imposes an ultimate limit upon useful deterministic weather forecasts estimated at about 2 weeks, present performance of numerical weather prediction models shows that this limit is still far from being reached. Since the quality of weather forecasts depends strongly upon the quality of the initial conditions, it is hoped that more accurate specification of those conditions will significantly improve the quality of

* Corresponding author, current address: Department of Geosciences, University of Wisconsin-Milwaukee, Milwaukee, WI 53201 USA.
E-mail: kwanson@csd.uwm.edu

short-range forecasts and extend the range of useful forecasts.

One obvious way to improve forecast initial conditions is to improve how observational data are assimilated into numerical prediction models. It is widely agreed that advanced data assimilation techniques should utilize data better than current statistical interpolation techniques (Ghil and Malanotte-Rizzoli, 1991). This includes not only extracting as much dynamically relevant information from the data as possible, but also filtering out the spurious information (noise). Recently, one particular advanced technique, four-dimensional variational assimilation of data (4DVAR), has been studied intensively. This interest was spawned by the work of Lewis and Derber (1985), Le Dimet and Talagrand (1986) (hereafter LDT86), and Talagrand and Courtier (1987) (TC87), who showed that it is feasible to perform 4DVAR with fairly realistic models. 4DVAR differs from sequential algorithms, such as variants of Kalman filtering (Kalman, 1960; Ghil and Malanotte-Rizzoli, 1991), as it seeks a nonlinear best estimate of the flow through direct explicit minimization of some scalar function measuring the misfit between the observations and the model solution over a finite assimilation time period T_a . The misfit, or cost function to be minimized is typically the sum of the squared observation-minus-model differences, weighted by the inverse covariance matrix of the observational errors.

The computational appeal of 4DVAR is that adjoint methods provide a means to systematically search for the minimum of the cost function. This is because the adjoint equations allow explicit and efficient calculation of the gradient of the cost function with respect to the state of the model at the beginning of the assimilation period (LDT86; TC87). Currently, 4DVAR is operational at the European Center for Medium Range Weather Forecasts (ECMWF) with an assimilation period of 6 h. However, both practical and theoretical research is needed before the limits and power of 4DVAR in an operational setting can be assessed.

In operational mode 4DVAR is used sequentially, and previous forecasts are treated as 'observations' weighted by their respective error covariances. This adds a background term to the cost function. From a theoretical viewpoint, there is no fundamental difference between these forecast values and the observed values. Therefore, for this

study we do not distinguish between the two sources of information. Rather, we consider 4DVAR as globally adjusting a model solution to the available, albeit erroneous, information over the entire assimilation period.

In this work, we examine a number of issues regarding the application of 4DVAR in an idealized setting. One class of these issues concerns how variational assimilation performs in the face of chaos. The performance of 4DVAR for low degree of freedom chaotic dynamical systems, such as the classic Lorenz (1963) three variable system, has been studied by a number of authors. Gauthier (1992), Stensrud and Bao (1992), Miller et al. (1994), and Pires et al. (1996) all found that the performance of 4DVAR varies significantly as a function of the assimilation time period T_a . Further, it was found in these studies that assimilation is particularly difficult at times when the system experiences regime transitions between the two lobes of the Lorenz attractor. As there are conceptual similarities between the regime transitions in the Lorenz system and weather regime transitions in the atmosphere (Vautard, 1990), such as from a blocked to an unblocked flow, this result suggests that predicting atmospheric weather regime transitions is inherently difficult. Finally, assimilation over long time periods T_a is difficult due to the appearance of multiple minima in the cost function. In the presence of multiple minima, the result of the minimization, and hence the assimilated model state estimate, can depend upon the starting point of the minimization. The purpose of this work is to examine the extent to which these results from low degree of freedom models apply to variational assimilation in a more realistic atmospheric model, namely the 3-layer quasi-geostrophic model of Marshall and Molteni (1993).

As such, the first issue we address concerns the optimal length of time over which to perform variational assimilation. Pires et al. (1996) found that increasing the assimilation time period T_a beyond the error doubling time of the system τ_{NL} significantly improved the quality of the assimilated state, and by extension, the quality of the forecasts. They proposed an algorithm, quasi-static variational assimilation (QSVA), that allows tracking of the absolute minimum of the cost function over progressively longer assimilation periods provided that the temporal density of

observations is sufficiently large and the level of observational error sufficiently small. Based upon their findings with the Lorenz model, they suggested that assimilation time periods on the order of one week should be optimal in more realistic atmospheric models. We examine whether this estimate holds in a more complicated model, and to quantify improvements in forecasts obtained when the assimilation time period is extended.

Another important issue concerns the structure of the analysis error at the forecast initiation time. Analysis error has been and continues to be a topic of extreme interest and importance in operational weather forecasting, since it guides the development of ensemble prediction strategies (Toth and Kalnay, 1993; Molteni et al., 1996). At the National Center for Environmental Prediction (NCEP), forecast ensembles are initiated by adding perturbations to the analysis that supposedly represent the analysis error, while ensembles at ECMWF are initiated by adding perturbations that yield the maximum future forecast spread. Our purpose here is not to arbitrate between these strategies, but to provide scientific material to this debate by examining the main characteristics of analysis errors within the perfect model hypothesis. Since present models are far from perfect, the results presented here are not strictly valid. However, model improvements and the operational application of 4DVAR should make these results applicable in the future.

The general result demonstrated by Pires et al. (1996) is that in a perfect model environment, the 4DVAR-assimilated state error projects essentially onto the unstable manifold of the system.* This implies that for long assimilation periods, the assimilation error should project primarily upon the amplifying local Lyapunov vectors (Eckmann and Ruelle, 1981; Legras and Vautard, 1995), since these vectors span the unstable manifold. Pires et al. (1996) explicitly confirmed that such a projection occurs for the 3-variable Lorenz (1963) system.

Interestingly, it may be argued that confining

* For reference, the unstable manifold is spanned by those error directions that decay toward the true solution trajectory in the distant past; conversely, the stable manifold is spanned by those error directions that decay toward the true solution trajectory in the distant future (Guckenheimer and Holmes, 1983).

the error in this manner may reduce the overall growth rate of the error compared to the case where no assimilation, or assimilation over a short time period is done. This is because disturbances that grow more rapidly than the local exponential divergence of trajectories on the unstable manifold, such as singular vectors (Farrell, 1990; Molteni and Palmer, 1993), are typically comprised of strongly decaying as well as amplifying error components (Palmer, 1995). The prototype of singular vector growth is an initial disturbance composed of two very non-orthogonal eigenmodes, one that strongly decays with time and a second that amplifies or only weakly decays with time. Since the two eigenmodes are very non-orthogonal, the initial disturbance formed by the addition of these eigenmodes can initially be small. However, since the first eigenmode decays rapidly, the disturbance will rapidly acquire the structure of second eigenmode as it evolves with time, yielding rapid disturbance growth. Transient growth of singular vectors in fluid systems was first noted by Orr (1907), and a review of this process in plane parallel shear flow is given by Lindzen (1988). In the context here, however, if extending the assimilation period eliminates error projections in decaying directions, singular vector growth that relies upon strongly decaying error components will not occur, leading to a consequent reduction in overall error growth rates.

The structure of the assimilated state error has important practical consequences for predictability. Provided the unstable manifold projects only weakly into the subspace of leading singular vectors, 4DVAR should drastically reduce the error of the initial conditions in those directions that grow the fastest over the first few days of the forecast. Our aim is to address this issue within the context of a reasonably realistic atmospheric model. We seek to understand the structure of the assimilation error and the growth of the forecast error under 4DVAR. This includes how the structure of the error at the forecast initiation time depends upon the length of the assimilation time interval, whether error at small scales plays a significant role in the growth of forecast error, and whether the predictability is directly related to the instability of the flow itself, i.e., to synoptic scales associated with baroclinic instability.

The final issue we address regards how well 4DVAR performs in a non-perfect model environ-

ment. 4DVAR is formally valid only in the perfect model framework, where the only source of error is in the observations themselves. However, any realistic application of 4DVAR in an operational setting must necessarily account for model errors as well. To understand to what extent 4DVAR improves forecast quality in the presence of model errors, we examine assimilation and forecast error growth using versions of the 3-layer quasi-geostrophic model with model errors in the truncation as well as improper specification of the model forcing.

The outline of the article is as follows. In Section 2, we introduce the 3-layer quasi-geostrophic intermediate model used herein, outline the application of 4DVAR to this model, and review the arguments presented by Pires et al. (1996) why long assimilation time periods are desirable to optimally reduce analysis errors. In Section 3, we study 4DVAR and predictability in the perfect model situation, examining the enhanced predictability made possible by the use of 4DVAR in both case studies as well as on average. In Section 4, the structure of the analysis error under 4DVAR is examined, focusing on the spatial structure of that error and its projection onto the leading singular vectors and Lyapunov vectors for the system. In Section 5, we extend the results above to the imperfect model case, and conclusions are drawn in Section 6.

2. Model and methodology

2.1. The dynamical model

The intermediate model we use to study 4DVAR is the Laboratoire de Météorologie Dynamique (LMD) version of the 3-layer quasi-geostrophic (QG) model introduced by Marshall and Molteni (1993; hereafter MM93). This model adequately simulates both extra-tropical synoptic-scale and low frequency variability (MM93; Liu, 1994; Michelangeli, 1996), and the simplicity of the QG framework minimizes computational requirements. The model is spectral on the sphere, with triangular truncation at total wavenumber 21 (T21). The vertical coordinate is pressure, with the QG pseudo-potential vorticity q predicted at 200, 500, and 800 mb levels (levels $l = 1, 2, \text{ and } 3$, respectively). The evolution equations are discussed at length in MM93 (their Section 3 and

Appendix A), and may be schematically written

$$\partial q / \partial t = -J(\psi, q + f) - \mathcal{D}q + S. \quad (2.1)$$

Here, q is the relative QG pseudo-potential vorticity (PV), related by an elliptic operator $q = \mathcal{A}\psi$ to the geostrophic streamfunction ψ , $J(A, B) = (a \cos \theta)^{-1}(\partial_\lambda A \partial_\theta B - \partial_\lambda B \partial_\theta A)$ is the 2-D Jacobian operator on a sphere of radius a ; f is the Coriolis parameter; and the contribution to the PV from topography in the lowest layer is implicitly assumed. In each layer, the dissipation $\mathcal{D}q$ consists of Newtonian relaxation of the baroclinic flow component to a state of rest, hyperviscosity to damp small scales, and in the bottom layer, Ekman friction. The forcing term S is empirically calculated from the wintertime (DJFM) ECMWF 1983–1993 analysis to yield a vanishing time average vorticity tendency in each layer, i.e.,

$$S = \overline{J(\psi_a, q_a + f)} + \overline{\mathcal{D}q_a}, \quad (2.2)$$

where the overline indicates time average, and subscript a indicates analyzed ECMWF fields. This forcing prevents the model from drifting away from the observed ECMWF climatology. The only substantive difference between the MM93 model and the version herein is the time difference scheme; we use a first order Adams–Bashforth–Moulton predictor/corrector scheme (Stoer and Bulirsch, 1993), while MM93 used a leap frog scheme.

The experimental situations that we consider are idealized. First, a truth trajectory is generated using the fully nonlinear model. From this trajectory, “observations” O are generated every 6 h by adding error to the truth trajectory. This error is modeled by an isotropic Gaussian white noise process with a standard deviation of (60, 40, 20) m in the QG geopotential height field $Z = f_0 g^{-1} \psi$ at (200, 500, 850) mb, respectively, where g is the acceleration due to gravity and f_0 the Coriolis parameter at 45°N latitude. These observations are provided to 4DVAR in both the perfect and imperfect model settings.

2.2. The cost function and its minimization

To assimilate observations spread over the finite time interval $(t_0, t_0 + T_a)$, we seek the initial condition $\psi_0 = \psi(t_0)$ leading to the solution $\psi(t)$ that

minimizes the misfit between this solution and the data. This misfit is measured by the cost function.

$$\mathcal{J}[\psi_0] = \frac{1}{2} \sum_{i=1}^N \langle \psi_i - O_i, S_i^{-1}(\psi_i - O_i) \rangle. \quad (2.3)$$

Here, $\psi_i = \psi(t_i)$ (where the layer subscript has been suppressed), O_i is an observation at time t_i , S_i is the covariance of the i th observation error, and $\langle \cdot, \cdot \rangle$ is the inner product chosen as a measure of the error. For the study here, we consider the inner product based on QG streamfunction squared, which is simply

$$\langle \psi_a, \psi_b \rangle = \iint (\psi_a, \psi_b) dA, \quad (2.4)$$

where the double integral is over the surface of the sphere, and summation over the three layers is implicit. For future reference, we also define the entropy inner product

$$\langle q_a, q_b \rangle = \iint (q_a, q_b) dA; \quad (2.5)$$

this alternative error measure is used to assess the extent to which error reduction by 4DVAR is norm-specific.

The minimization of the cost function is done iteratively using algorithms such as conjugate gradient or quasi-Newton methods, where the latter are generally more efficient (Gilbert and Lemaréchal, 1989). All minimizations in this paper use a slightly modified version of the quasi-Newton algorithm DFPMIN from Numerical recipes (Press et al., 1992). The iterative quasi-Newton method requires at each iteration the knowledge of the gradient of the cost function with respect to initial condition ψ_0 . As demonstrated by LDT86 and TC87, the value of the cost function and its gradient may be obtained by a forward integration of the direct nonlinear model followed by a backward integration of the adjoint model.

For the QG model (2.1), the adjoint model in the streamfunction squared norm can be written

$$\partial \delta^* q / \partial t = -\mathcal{L}(q) \delta^* q \quad (2.6)$$

where

$$\begin{aligned} \mathcal{L}(q) = & -\mathcal{A}J(q + f, \mathcal{A}^{-2} \delta^* q) \\ & + \mathcal{A}^2 J(\mathcal{A}^{-1} q, \mathcal{A}^{-2} \delta^* q) - \mathcal{A} \mathcal{D} \mathcal{A}^{-2} \delta^* q. \end{aligned} \quad (2.7)$$

Here, $\delta^* q$ is the sensitivity perturbation to the integrated backwards in time, and $\mathcal{A}^2(\cdot) = \mathcal{A}(\mathcal{A}(\cdot))$. The choice of the streamfunction squared norm (in contrast to a kinetic energy based norm, for example) simply reflects that our observational errors are white in this norm; it shall become apparent that for long assimilation periods, the choice of norm is not critical.

The formulae above are valid provided the model is not temporally discretized. In practice, to obtain an exact gradient calculation the adjoint of the numerical time difference scheme must also be taken. For the case here, the forward model employs the first order Adams–Bashforth–Moulton predictor–corrector time difference scheme, and the calculation of the adjoint to this scheme is straightforward. The adjoint code was verified using the scalar product identity (Thépaut and Courtier, 1991).

The adjoint model (2.6) involves twice the number of nonlinear terms as the nonlinear forward model (2.1), so that each adjoint integration is approximately twice as costly as the forward integration. Given that the minimization algorithm requires several iterations, each with one or more evaluations of the functional and one of its gradient, the cost of one iteration corresponds roughly to the time required to integrate the model over a period of $3T_a$.

2.3. Long assimilation periods and quasi-static variational assimilation

The successful assimilation of data for long T_a in a nonlinear, chaotic system is not trivial, as multiple minima in the cost function emerge as T_a is increased. The emergence of multiple minima is discussed at length in the context of low order chaotic systems by a number of authors (Li, 1991; Gauthier, 1992; Miller et al., 1994; Pires et al., 1996). For T_a sufficiently long in a chaotic system, the true solution trajectory will pass through regions in the system's phase space where the flow is highly sensitive to small errors. For example, in the Lorenz (1963) system, these sensitive regions occur whenever the x and y components of the trajectory are simultaneously small (Palmer, 1993). In these sensitive regions, assimilated trajectories lying near the true trajectory at the beginning of the assimilation period may be diverted into erroneous solution regimes for some

period of time, only to later return to the vicinity of the true trajectory before the end of the assimilation time interval. Any trajectory that is diverted and rejoins the true trajectory in this manner must lie within a secondary minima, as discussed in Subsection 2.2 of Pires et al. (1996).

Increasing T_a leads to ever more encounters with sensitive regions within the system's phase space. Given the exponential sensitivity of flow trajectories in chaotic systems, assimilated trajectories lying ever nearer to the true trajectory at the beginning of the assimilation period will be diverted into secondary minima as T_a is increased. This makes assimilation by descent methods effectively useless for very long T_a , as assimilated trajectories will nearly always become trapped in a secondary minima of the cost function rather than shadow the true trajectory (Miller et al., 1994).

Yet, it can be shown that for optimal 4DVAR, the assimilation period T_a should be as long as possible. As discussed in Pires et al. (1996), this is because of the manner in which 4DVAR distributes the assimilation error during the assimilation time period. For example, at the beginning of the assimilation period, 4DVAR has significant information about the future behavior of trajectories, and hence errors in phase space directions containing trajectories that depart from the true trajectory in the future are strongly constrained. However, since the assimilation has no information about the past behavior of the system, error components that decay toward the true trajectory in the future, i.e., trajectories on the stable manifold of the system, are not strongly constrained. Conversely, at the end of the assimilation period 4DVAR has significant information about the past behavior of trajectories, and hence errors in phase space directions containing trajectories that depart from the true trajectory in the past are strongly constrained. However, since the assimilation has no information about the future behavior of the system, error components that decay toward the true trajectory in the past, i.e., trajectories on the unstable manifold of the system, are not strongly constrained. At intermediate times, 4DVAR has significant information about both the past and future behavior of the true trajectory and the error tends to be small in all phase space directions.

It is of interest to note that if 4DVAR is properly cycled, as it would be in an operational setting,

then the error at the beginning of the assimilation period would be much more strongly constrained. Cycling involves the addition of a background term at the beginning of the assimilation period that contains information from previous assimilations. If this background term has appropriate error statistics (in particular, flow dependent statistics), 4DVAR will have significant information about the past. As such, the error at the beginning of the assimilation period would be much reduced. However, if the assimilation period is already quite long, essentially no new information about the unstable components of the system would be added. Thus, cycling would not significantly affect the error at the end of the assimilation period, and hence by extension, the forecast error.

To provide the best estimate of the model state at the forecast initiation time, we are confronted with two conflicting objectives: make T_a as long as possible to improve the quality of the assimilated state, but keep T_a small enough to avoid the emergence of multiple minima. Pires et al. (1996) proposed an algorithm, quasi-static variational assimilation (QSVA), that allows extension of T_a while avoiding the problems associated with multiple minima. The QSVA algorithm can be decomposed into three steps, summarized below. Without loss of generality we denote the present (forecast initiation time) as $t = 0$, implying assimilation is done over the time interval $[-T_a, 0]$. As outlined above, the minimization is performed with respect to the state $\psi_0 = \psi(t_0)$ at the beginning of the assimilation period.

(1) Initialization. Use the observation $\psi_0^{(0)} = O(T_a)$ as the starting point for the minimization of the intermediate cost function $\mathcal{J}[\delta T_a, \psi_0]$, defined over the time interval $[-T_a, -T_a + \delta T_a]$. The state at $-T_a$ corresponding to this minimizing trajectory is denoted $\psi_0^{(1)}$.

(2) Stage i to stage $i + 1$: Use $\psi_0^{(i)}$ as the starting point for the minimization of the intermediate cost function $\mathcal{J}[(i + 1)\delta T_a, \psi_0]$, defined over the time interval $[-T_a, -T_a + (i + 1)\delta T_a]$. The state at $-T_a$ corresponding to this minimizing trajectory is denoted $\psi_0^{(i+1)}$. The index i varies from 1 to $M - 1$ where $T_a = M\delta T_a$.

(3) Final stage. The starting point of the final minimizing trajectory $\psi_0^{(M)}$ is intended to be the absolute minimizing solution of $\mathcal{J}[\delta T_a, \psi_0] = \mathcal{J}[\psi_0]$ defined over the entire assimilation period $[-T_a, 0]$. The corresponding state estimate for

times $t > -T_a$ is obtained by integrating the model over the assimilation period with $\psi_0^{(M)}$ as the initial condition at time $-T_a$.

In the experiments described below, we consider the time increment $\delta T_a = 1$ day. The value $T_a = 0$ corresponds to a case where no assimilation is performed, where the assimilation error is identical to the observation error. As a computational note, about 20 minimizing steps using the quasi-Newton algorithm are used for each QSVA increment.

3. 4DVAR and predictability in the perfect model case

In this section, we estimate the reduction in analysis and forecast error by using 4DVAR in a perfect model environment. In particular, we focus on estimating the efficient past time scale τ_e beyond which observations no longer lead to improvements in the assimilation and forecast skill. Pires et al. (1996) found that for the Lorenz (1963) system, this time scale was about 3 times the doubling time of small errors τ_{NL} for the system. Extrapolating, they argued that for the real atmosphere, the efficient time scale should be about one week. We now examine the relevance of this estimate for the MM93 system.

3.1. Performance of QSVA and the use of far past observations

To understand how far QSVA allows us to extend T_a consider assimilation of a sequence of "observations" available every 6 h that are generated by adding noise to a reference solution of the nonlinear model (2.1). For reference, the error doubling time for the model calculated by integrating a random initial perturbation over a sufficiently long time interval using the tangent linear model is $\tau_{NL} \sim 3$ days.

Previous attempts to extend T_a without QSVA yielded improved assimilated states to $T_a \sim \tau_{NL}$ (Gauthier, 1992; Tanguay et al., 1995). However, extending T_a much beyond τ_{NL} was shown to lead to a rapid decay of the quality of the assimilated state, as the assimilated trajectories became trapped in secondary minima. For the MM93 model, we find that using QSVA circumvents some of the difficulties presented by the appearance of multiple minima. Fig. 1 shows that as T_a increases

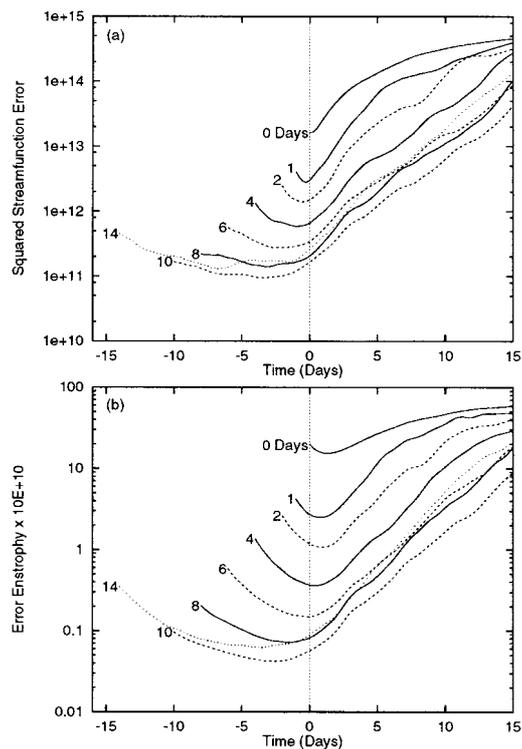
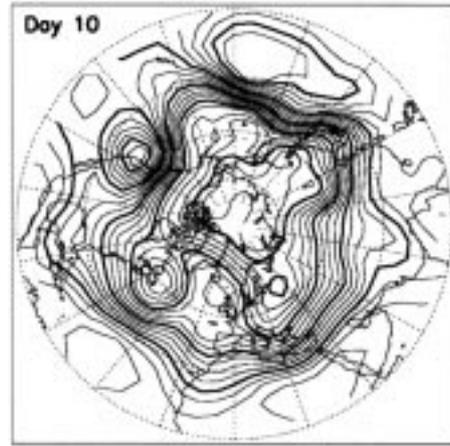
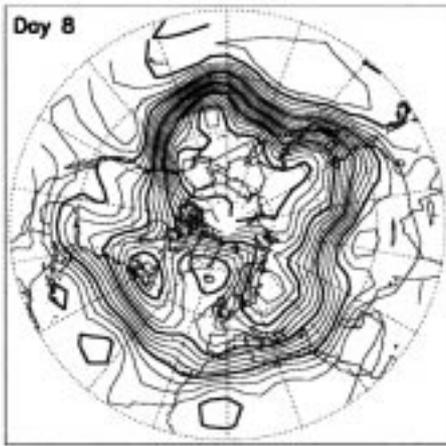
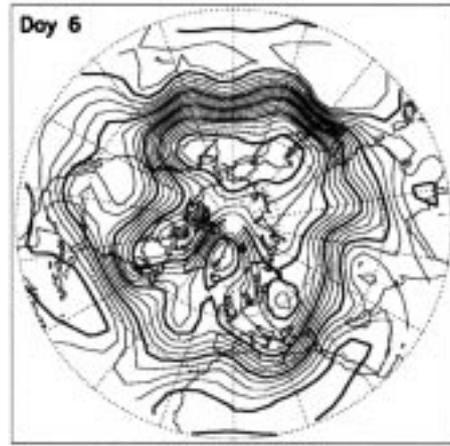
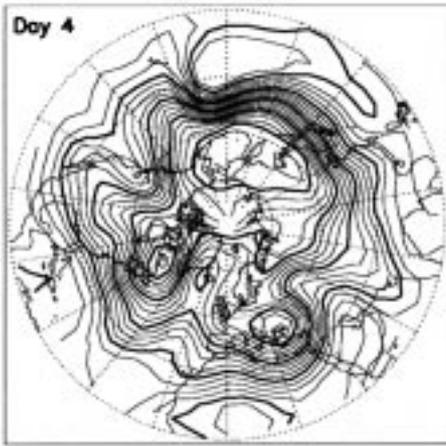
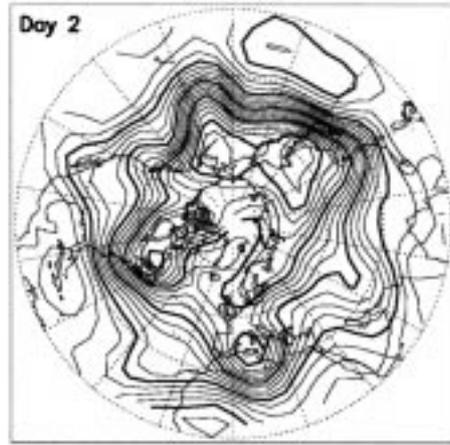
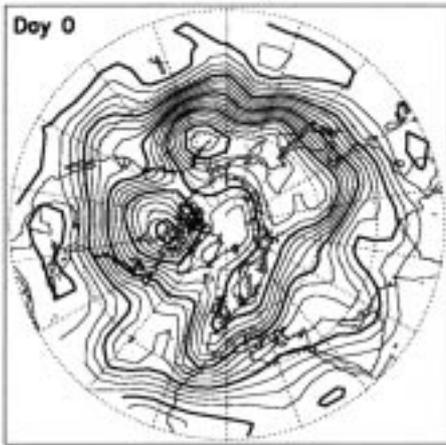


Fig. 1. Variation of the assimilation quality and forecast skill as measured by (a) the squared streamfunction error, and (b) the enstrophy error as a function of time (abscissa) and T_a (the numbers at the start of each curve). Note that the cost function minimized herein measures errors in the squared streamfunction norm. This assimilation experiment is performed for one case only.

to 10 days, the error in the assimilated state decreases substantially in both squared streamfunction and enstrophy norms. Somewhat surprisingly, the reduction in error enstrophy is larger than the reduction in error squared streamfunction, despite the fact that QSVA explicitly minimizes the latter quantity. However, the assimilated trajectory is dynamically consistent over the assimilation time period; hence it is only natural that error in all dynamically relevant quantities should be reduced, regardless of whether a given measure of error is explicitly minimized.

The QSVA scheme succeeds in maintaining the assimilated trajectory within the basin of attraction for the absolute minima of the cost function for T_a as long as 10 days. This is in spite of the existence of multiple minima in the cost function



for $T_a > 5$ days. For example, performing the assimilation directly without QSVA for $T_a = 6$ days (not shown) yields an assimilated state with error streamfunction and error enstrophy both about 1.5 times larger than the QSVA assimilated state errors shown in Fig. 1. In general, for the MM93 model we have found that multiple minima in the cost function emerge for $T_a = 4$ to 6 days, implying that QSVA is required to obtain any reduction in the assimilated state error and forecast quality that may occur for longer values of T_a .

However, it is also apparent from Fig. 1 that the error in the assimilated state cannot be reduced indefinitely. Extending T_a to 14 days increases the error in the assimilated state as well as the forecast error compared to their respective values for $T_a = 10$ days. The exponential proliferation of multiple minima associated with the chaotic nature of the system ultimately catches up with the QSVA algorithm for T_a sufficiently long. The barrier at 10 to 12 days is roughly independent of both δT_a and the observation frequency, as it still exists when the QSVA increment is changed to $\delta T_a = 1/2$ day or 2 days, as well as when the frequency of observations is doubled. This indicates that extending T_a beyond 10 days may be difficult regardless of the length of the QSVA increment δT_a . Curiously, Pires et al. (1996) were able to extend T_a to $8\tau_{NL}$ for the low order Lorenz (1963) system compared to only $3-4\tau_{NL}$ for the MM93 model here. Evidently, maintaining the assimilated state in the absolute minima of the cost function is more difficult for long T_a in high dimensional systems than in low dimensional systems like the Lorenz system. Nevertheless, assimilation to a value of $T_a = 3\tau_{NL}$ appears readily possible, which by the arguments presented in Pires et al. (1996) should be sufficient to extract the bulk of the dynamical information contained within the observations.

It is interesting that the minimum in error variance is located near the end rather than at the middle of the assimilation period. This is because there are many more stable phase space directions in the MM93 model than unstable phase space directions: Vannitsen and Nicolis (1997) found

that about 100 of the 1518 degrees of freedom in the MM93 model are unstable. This imbalance between the number of stable and unstable directions means that the stable manifold is much more difficult for 4DVAR to characterize, and as such, reduction of errors on the stable manifold is more difficult than reduction of errors in the unstable manifold within the assimilation period.

Another interesting feature apparent in Fig. 1 is the drastic improvement in predictability in the medium range (days 5–15) that is obtained by using long assimilation periods. Fig. 1 shows $t = 0$ day assimilation/analysis error levels for short T_a are equal to forecast error levels more than 5 days in the future for $T_a = 10$ days. This result suggests that if current NWP models were perfect, the implementation of a long-period 4DVAR analysis system would lead to a substantial gain in forecast skill.

3.2. Assimilation and predictability of a blocking onset

Turning to a more observationally relevant situation, we now examine the extent to which 4DVAR improves the predictability of a particular weather regime transition in a reference model solution. Regime transitions, such as from a zonal to a blocked flow, are well known to be difficult to predict operationally (Tibaldi and Molteni, 1990). Based upon the behavior of simple dynamical systems such as the Lorenz (1963) system, it has been argued that the difficulty in predicting such transitions occurs because the flow is inherently more unstable to small perturbations in the vicinity of such transitions (Palmer, 1993). This instability can be quantified in any number of ways; the growth of finite time singular vectors (Molteni and Palmer, 1993), local Lyapunov vectors (Trevisan and Legnani, 1995), or the flow sensitivity as measured by the growth of perturbations optimally configured to cause such weather regime transitions (Oortwijn and Barkmeijer, 1995) all indicate faster error growth when such transitions occur.

Consider a synoptic situation consisting of the

Fig. 2. Onset and decay of the blocking event studied in this sub-section. Panels read from left to right, top to bottom, and show 500mb QG geopotential height $Z = f_0 g^{-1} \psi$ at days 0, 2, 4, 6, 8, and 10. The contour interval is 50 m, with the 200 m contours darkened.

onset and decay of a block in a model reference solution. Several time slices of the 500 mb height field for this reference solution are shown in Fig. 2, revealing the onset and maturation of an Ω -type block in the North Atlantic between $t = 0$ and $t = 10$ days. This block then subsequently decays quickly, with the flow returning to near climatological values by $t = 15$ days. The experiments are constructed as follows: to the 10 days preceding day 0 of this solution, we generate 20 different observational realities by adding Gaussian white noise error to the reference solution. These observations are provided every 6 h and assimilation is performed over the time interval $[-T_a, 0]$ for $T_a = 0, 2,$ and 10 days for each realization. The QSVA algorithm is then applied to each series of noisy observations.

To measure the skill of regional blocking prediction, we define a blocking index in a manner similar to that of Liu (1994) by introducing the QG geopotential height anomaly pattern Z_B shown in Fig. 3. This anomaly pattern is constructed by subtracting the time mean from the 5-day running average 500 mb geopotential height field centered at day 8, and setting to zero every point of this difference field outside the Europe–North Atlantic (ENA) region (30 to 90°N latitude, 60°W to 40°E longitude), as well as those points

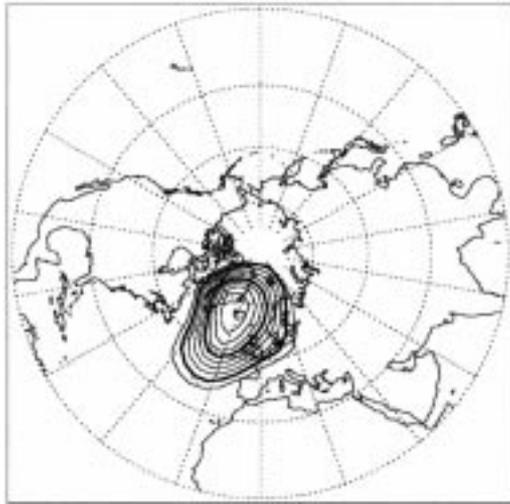


Fig. 3. Geopotential height anomaly pattern used to quantify regional predictability skill for the 20 manifestations of observational error for the experiments in this sub-section. The contour interval is 25 m.

within the ENA region whose geopotential height is smaller than the time mean. Although this construction is arbitrary, it provides a case-specific spatial disturbance pattern to compare against the time evolution of the block. Given this pattern we can define the blocking index

$$C = \frac{\langle Z, Z_B \rangle}{\langle Z_B, Z_B \rangle}, \quad (3.1)$$

where $Z = f_0 g^{-1} \psi$ is the instantaneous 500 mb QG geopotential height field. We then arbitrarily normalize C to vanish for $t = 0$ days and to have a value of unity at $t = 8$ days for the true model trajectory. The variability of C with time for each individual forecast provides a quantitative measure of the skill of regional blocking prediction.

The forecast values of C , shown in Fig. 4 for the 20 realizations of the observational error, indicate that extending T_a greatly improves regionally predictability. For $T_a = 0$ days (no assimilation), neither the blocking onset nor its decay are well simulated, as the median of C for the 20 realizations is quite small and the spread of C quickly becomes as large as the blocking anomaly itself. Extending the assimilation period to $T_a = 2$ days, the blocking onset is successfully predicted for all but two realizations of the observational error. However, the spread of the forecasts beyond $t = 8$ days indicates rather poor case-wise predictability of both the timing and the speed of the decay of the block. With an assimilation period of $T_a = 10$ days, both the onset and decay of the block are captured for all 20 realizations of the observational error. The forecast values of C only deviate significantly from the truth trajectory for forecast lead times longer than 10 days.

3.3. Average predictability

In addition to improving predictability in specific synoptic situations, extending T_a also improves the average predictive skill of the model. We now consider a 150-day model reference solution, of which noisy observations are generated every 6 h as above. A sequential version of QSVA proposed by Pires et al. (1996) is applied to this time series to quantify the average predictability for a 100 day segment of this solution, where assimilation and forecasting are carried out every $\delta t = 12$ h. The sequential version of QSVA implemented is quite simple. We start by doing QSVA for the interval $[t - T_a, t]$ for $t = 0$ days. From

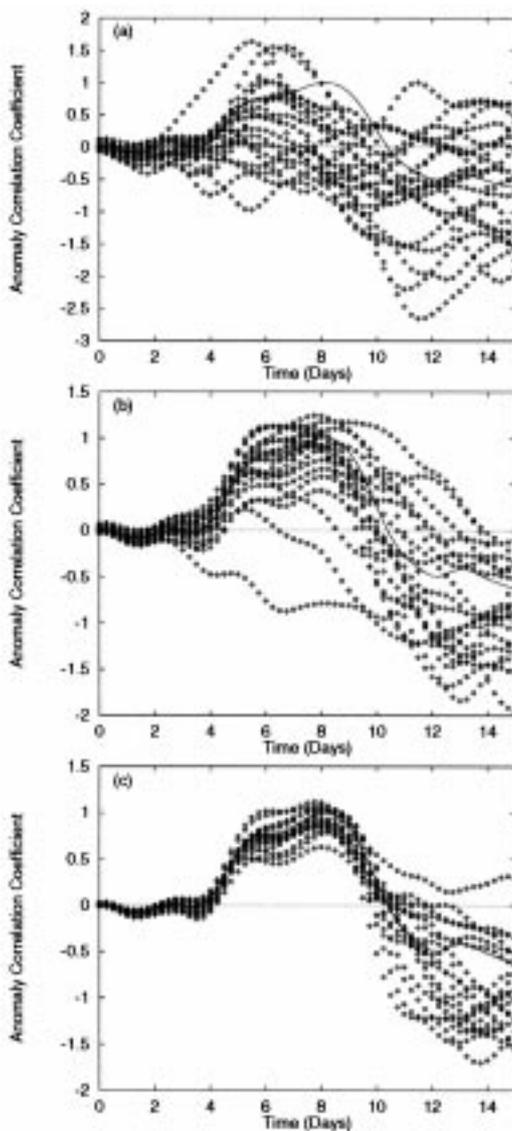


Fig. 4. Dependence upon T_a of the forecast anomaly correlation coefficient C for the 20 realizations of observational error. (a) $T_a = 0$ days, (b) 2 days, and (c) 10 days. For all cases, the solid line is the anomaly correlation coefficient for the true solution, and the impulses are the individual forecast realizations.

that point forward, the assimilated state at time $t - T_a + \delta t$ is used as the starting point for the next assimilation's minimization. No QSVA appears to be necessary after the first time step for $T_a = 10$ days, as the assimilated state does not

escape from the attracting basin of the absolute minima of the cost function. However, taking a few QSVA minimizing steps at the intermediate time steps $T_a = 6$ and 8 days speeds the overall convergence of the minimization. The assimilation and forecast statistics described below are for the 200 individual assimilations and forecasts.

Fig. 5 shows that extending T_a from 0 to 10 days reduces the median error of the assimilated state and consequently increases the forecast skill. The quantitative reduction in the magnitude of the error converges to a level about 70 times smaller than the observational error in the squared streamfunction norm explicitly minimized by QSVA, and about 100 times smaller than the observational error in the enstrophy norm for the assimilation period $T_a = 10$ days. For T_a greater than 4 days, the error growth is approximately exponential. The growth rate of the squared error measures suggests an error e -folding time of 5 days,

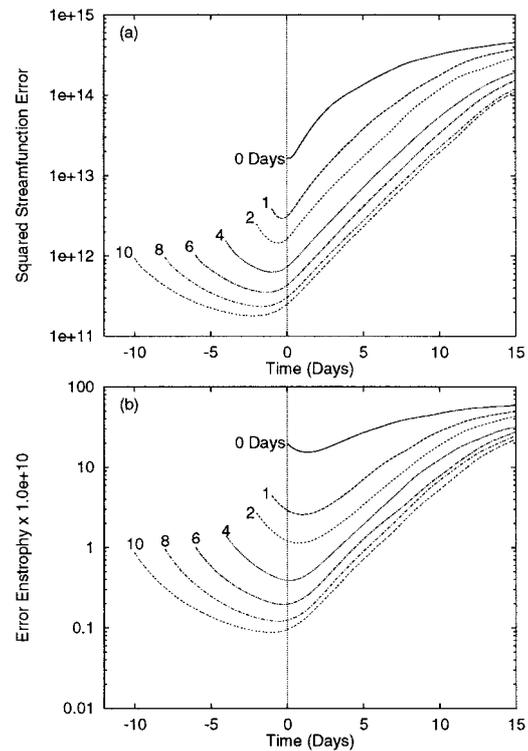


Fig. 5. Median values of the (a) streamfunction squared error, and (b) enstrophy error for the 200 forecast set as a function of forecast time and of the assimilation time T_a .

or an error doubling time of approximately 3.5 days. This doubling time is somewhat longer than the doubling time based upon the leading Lyapunov exponent for the MM93 model of 3 days, primarily due to smaller local error growth rates in the Equatorial and Southern Hemispheric regions of the model domain. The lead time for global forecast skill is quite long, exceeding 15 days for T_a larger than 6 days. Note that in the squared streamfunction norm, the growth rate of the error decreases as T_a is made larger. The extent to which extending T_a changes the error growth, as well as changes in the structure of the error as a function of T_a shall be explored further in the next Section.

The power of the 4DVAR to reduce error in this perfect model setting when used in assimilation mode, i.e., for negative times in Fig. 5, is also apparent. In assimilation mode with $T_a = 10$ days, the error squared streamfunction is reduced by two orders of magnitude at time $t = -2$ days. The reduction in error for the assimilated states is greater than that at the forecast initiation time, as the assimilated state has the benefit of information propagating not only forward but also backward in time. We emphasize this fact because the potential of variational assimilation to provide extremely accurate estimations of the atmosphere for re-analysis of data should not be overlooked.

Finally we note that Fig. 5 reveals the systematic nature of the behavior depicted by Fig. 1. Long period assimilation with $T_a = 10$ days increases the predictability time scale by 5 days over short assimilation periods of 1–2 days.

4. Structure of the analysis error under 4DVAR

4.1. Spatial structure

To begin our analysis of the error structure in the MM93 model using QSVA, we consider the spatial patterns of the error averaged over the 200 forecast experiments discussed above. Fig. 6 shows the 500 mb QG geopotential error variance for $T_a = 10$ days at the forecast initiation time (day 0) and 4 days later. At both times, the error is primarily concentrated in the mid-latitude storm tracks. However, comparison of panels (a) and (b) shows that the error grows faster in the Pacific storm track than in the Atlantic storm track. This is consistent with larger lower tropospheric Eady growth rates in the Pacific storm track for the

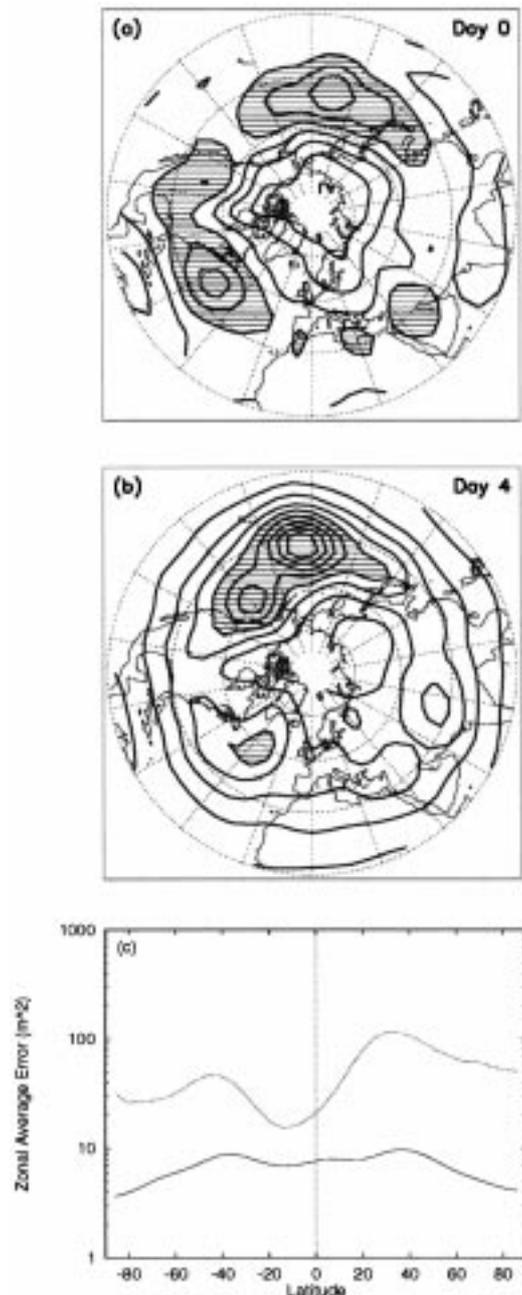


Fig. 6. (a) 500 mb QG geopotential height error variance at day 0 for $T_a = 10$ days. Contour interval is 1.25 m^2 , with values greater than 8.75 m^2 shaded. (b) Same, but day 4 error. Contour interval is 25 m^2 , with values greater than 125 m^2 shaded. (c) Zonal average error variance at day 0 (solid) and day 4 (dashed). Note the log scale for the ordinate.

MM93 model, and implicates baroclinic instability as the primary source of error growth. Fig. 6c shows the zonally averaged 500 mb geopotential height error variance at 0 and 4 days, again for $T_a = 10$ days. The zonally averaged error at day 0 does not vary strongly with latitude, but the growth of error in the tropics is weak compared to that in the mid-latitudes and particularly the Northern Hemisphere extra-tropics. The tropical error amplifies by a factor of only 3 over the 4-day time period, compared to more than a factor of 10 amplification in the error over the Northern Hemisphere extra-tropics.

It should be noted that these error structures in general depend upon the space-time distribution of the observations. Hence, it is encouraging from the operational perspective that error growth is slow in the tropics, which are poorly observed. In contrast, one expects that long period 4DVAR would greatly reduce analysis errors over the north Atlantic, due to the influence of observations made upstream over the North American continent.

4.2. Spectral structure of the error

Next, we examine the spectral structure of the error as a function of T_a . Fig. 7 shows the power spectra of the streamfunction squared variance as

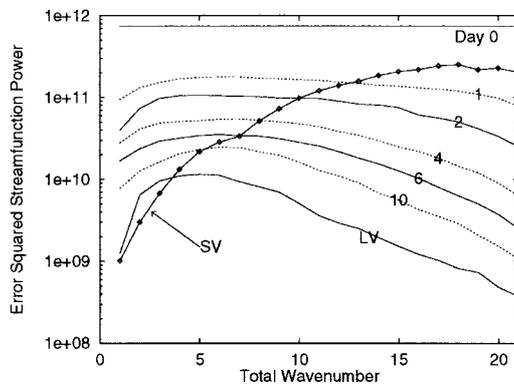


Fig. 7. Time average power spectra of day 0/ forecast initiation time error in the squared streamfunction norm for the 200 forecast experiments. Individual curves are labeled by their respective values of T_a . Also included are the spectral structure of the leading Lyapunov vector (labeled LV) and the leading 2-day future singular vector (labeled SV) averaged over the 200 forecast experiments. The amplitude for these vectors is arbitrary.

a function of total wavenumber on the sphere at the forecast initiation time for $T_a = 0, 1, 2, 4, 6,$ and 10 days. In addition to the substantial overall decrease in the magnitude of the error as T_a is increased consistent with Fig. 5, the shape of the spectra for long T_a in Fig. 7 indicates that the error is preferentially reduced at small scales. This is consistent with the larger reduction in error enstrophy compared to error squared streamfunction as T_a is increased as shown in Fig. 5. While this preferential reduction of error at small scales does not prove that error cascading from small to large spatial scales during the forecast is unimportant, QSVA does reduce error effectively at scales not heavily weighted by the norm defining the cost function.

This reduction in error is virtually independent of the information provided at small scales. Assimilation and forecasting on the same time series with observations supplied only to scales larger than T13 yields virtually identical power spectra and error reduction (not shown). This conclusion only holds at the end of the assimilation period for $T_a = 10$ days; at the beginning of the assimilation period, the error spectral structure is strongly peaked at small scales. Evidently, this information at small scales primarily lies in phase space directions that do not amplify over the assimilation period, and hence this loss of information does not adversely affect the assimilated solution at the end of the assimilation period.

The behaviour of the error in the above experiments is similar to that observed in the 4DVAR experiments of Tanguay et al. (1995). In their study of 4DVAR in a turbulent barotropic cascade, extending the assimilation period also reduced assimilation errors preferentially at small scales (their Fig. 7), although their error never acquired the spectral structure of the unstable manifold for their turbulent barotropic model due to their observation scheme*. This agreement is comforting, as instabilities in their barotropic model were also found on large scales, albeit the result of forcing rather than baroclinic instability as in the case herein. Note, however, that the application of QSVA allows the assimilation period here to be extended further than it was in study Tanguay et al. (1995) (10 versus 6 days); presumably, applying QSVA in that study would have led to even greater reduction of assimilation error.

4.3. Analysis errors, Lyapunov vectors, and singular vectors

Also shown in Fig. 7 are the average spectra of the leading Lyapunov vector (LV)** and that of the leading singular vector (SV) optimized over the first two forecast days. For $T_a = 10$ days, the power spectra of the error at forecast initiation time strongly resembles that of the leading LV, while the leading SV peaks at such larger wavenumbers (see also Molteni and Palmer, 1993). This suggests that the analysis error strongly projects onto the unstable manifold of the system error when long assimilation period 4DVAR is applied, as indicated by the theoretical result of Pires et al. (1996).

To show this explicitly, we project the 200 analysis errors at $t = 0$ days onto the 100 leading LVs and the 100 leading SVs, respectively, for varying values of T_a , and calculate the ratio between the error variance of the projection and that of the total error. Fig. 8 shows that for short assimilation periods, the variance explained by both the leading SVs and LVs is quite low. However, as T_a increases, the analysis error mostly projects onto the LVs, while no increase of the variance explained by SVs is observed. Since the unstable manifold of the system is locally spanned by the LVs associated with a positive Lyapunov exponent (about 100 in the MM93 model), we conclude that the analysis errors for $T_a = 10$ days are indeed contained within the unstable manifold. Conversely, the leading SV directions are not representative of this analysis error. Since the unstable manifold of a dynamical system is parallel to its attractor sheets (Eckmann and Ruelle, 1983; see also Legras and Vautard, 1995), we also con-

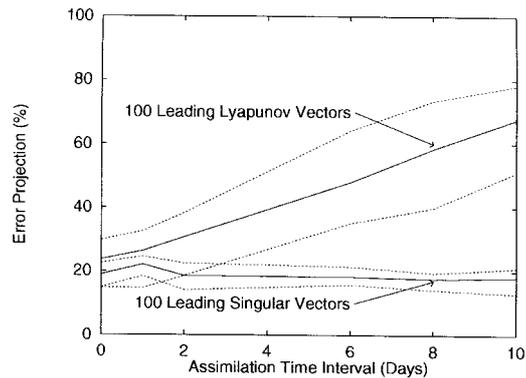


Fig. 8. Projection of the day 0 assimilation error onto the 100 leading Lyapunov vectors and the 100 leading singular vectors as a function of assimilation time T_a . The solid lines for each are the median for the 200 experiments, and the dashed lines are the 5th and 95th percent confidence intervals.

clude that analysis errors under 4DVAR are also parallel to these sheets.

4.4. Structure of the analysis error and predictability

The fact that analysis errors project onto the leading LVs does not mean that non-modal singular vector growth is totally eliminated, since LVs have a nonzero (albeit small) projection onto SVs. To show this, we project the 200 analysis errors at time $t = 0$ for assimilation period $T_a = 10$ days onto the leading 20 SVs, and integrate these projected perturbations forward in time. For comparison, the projection onto the leading 20 LVs is also performed and integrated forward in time. Fig. 9 shows the growth of these projected errors, as well as the entire forecast error for $T_a = 10$ days from Fig. 5 as a function of lead time. The small projection of the analysis error onto the leading 20 SVs at day 0 grows very rapidly with time, and the associated forecast error amplitude is the same as that of the forecast issued from the non-projected error after day 3. In contrast, forecasts issued from the analysis errors projected onto the leading 20 LVs maintain the same growth rate as the non-projected errors. The mean pattern correlation between the SV-projected forecast errors and the non-projected forecast errors is about 0.7. After day 3, it appears that the forecast error is

* Tanguay et al. (1995) provide a comprehensive discussion of the structure of the unstable manifold in a forced, turbulent barotropic system; we refer the interested reader to that study for more details.

** The Lyapunov vectors used herein are constructed by integrating arbitrary initial conditions forward for 30 days using the tangent linear model. After the first 10 days, these perturbations were reorthogonalized and normalized using singular value decomposition in the streamfunction squared norm every half day. Since these vectors are only integrated over a finite period, they are only approximate to the true Lyapunov vectors. We retain 100 of these LV's over a given 30 day interval, consistent with the approximately 100 asymptotically growing directions for the MM93 model (Vannitsen and Nicolis, 1997).

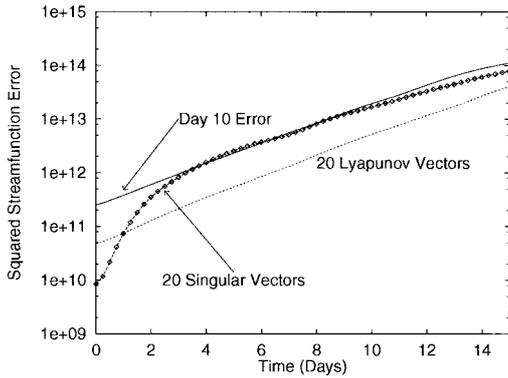


Fig. 9. Time evolution of the projection of the 200 day 0 assimilation errors for $T_a = 10$ days onto the 20 leading LVs and 20 leading SVs at day 0. Also shown is the actual $T_a = 10$ day forecast error.

dominated by the component of the initial error along the leading SVs.

In terms of spatial scales, this result indicates that in spite of the preferential reduction of error at small scales by QSVA, forecast errors undergo a significant upscale cascade of variance over the subsequent 4 days of evolution. However, the importance of this upscale cascade depends upon the length of the assimilation period T_a . To see this, we truncate the actual assimilation error at (and only at) the forecast initiation time, and compare the growth and spatial structure of this truncated error in the future to that of the true forecast error. If error information on small scales at the forecast initiation time is important, the loss of this information should result in significant, quantitative changes to the large scale structure and growth of the forecast error in the medium range. To quantify the effect of error truncation, we consider the two quantities

$$C_{pat} = \langle \psi_{act}, \psi_{trunc} \rangle / \langle \psi_{act}, \psi_{act} \rangle, \quad (4.1)$$

$$C_{grwth} = \langle \psi_{trunc}, \psi_{trunc} \rangle / \langle \psi_{act}, \psi_{act} \rangle, \quad (4.2)$$

where the subscript 'act' refers to the actual field and the subscript 'trunc' refers to fields that are spectrally truncated at $t=0$ days. The quantity C_{pat} is a pattern correlation, measuring the spatial resemblance between the truncated and actual errors at some forecast time t , while C_{grwth} measures the relative growth of the truncated and actual errors. Note that C_{grwth} is quadratic, hence

the actual difference in the error growth rate goes as the log of the square root of C_{grwth} .

Fig. 10 shows the median of C_{pat} and C_{grwth} at forecast time $t = 4$ days for the assimilation periods $T_a = 1$ and 10 days for the 200 forecast experiments. Spatial truncation of the initial error reduces both the median growth and spatial correlation of the error for both values of T_a . However, the truncation affects the $T_a = 1$ day error much more strongly. For example, removing all scales smaller than T15 cuts the median $T_a = 1$ day error growth in half over the 0- to 4-day time period, while a similar reduction in error growth for $T_a = 10$ days requires removing all error information initially on scales smaller than T11. The slopes of the C_{grwth} and C_{pat} curves provide an estimate of the sensitivity of the $t = 4$ day forecast error to information contained at

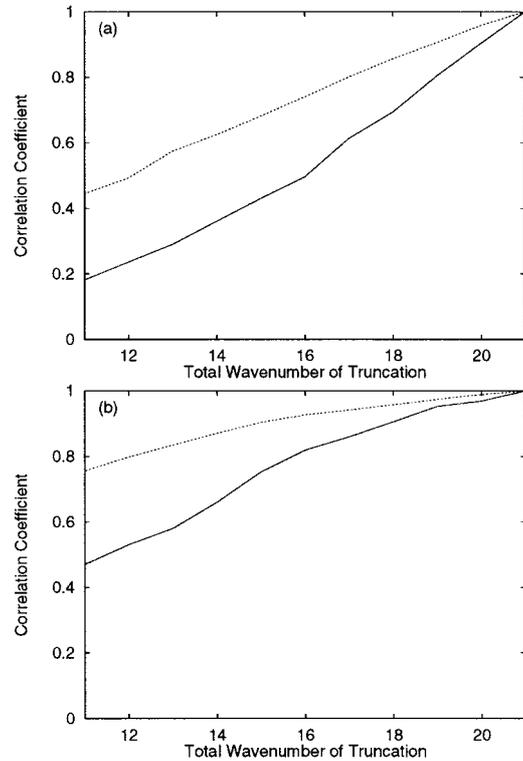


Fig. 10. C_{pat} (solid) and C_{grwth} (dashed) for the forecast error at day 4 for the assimilation periods (a) $T_a = 1$ day and (b) $T_a = 10$ days. The abscissa denotes the largest wavenumber retained when the error is truncated at the forecast initiation time.

a given scale. For $T_a = 1$ day, the forecast error is equally sensitive to errors at all scales, as both C_{grwth} and C_{pat} have relatively constant slopes over the T11 to T21 range of scales. However, for $T_a = 10$ days, the slopes of both C_{grwth} and C_{pat} are significantly greater in the T12 to T16 range than in the T17 to T21 range. Extending T_a diminishes the importance of information initially on small scales in the error to the overall growth and structure of the forecast error in the medium range.

5. Imperfect models

Operational forecasting, of course, is manifestly not done in a perfect model setting. Model errors are a fact of any operational model, and the value of 4DVAR must in part be based upon how effectively it can deal with model error. To study the effect of model error on 4DVAR more closely, we examine in detail assimilation using two distinct imperfect models, one where the source of model error is due to truncation error, and the other where the model error results from the incorrect forcing.

5.1. Truncation error

For our first set of experiments using 4DVAR in an imperfect model setting, we use a version of the MM93 model identical in all respects to that used for the perfect model experiments, except that it is truncated at total wavenumber T20 instead of T21. The experiments are as follows: a reference solution is generated using the T21 model, just as in the perfect model experiments, and "observations" are generated by adding Gaussian white noise to this solution every 6 h, as above. These observations are then truncated at total wavenumber T20, and their spectral coefficients are supplied to a T20 version of the MM93 model and its adjoint for assimilation and forecasting. The results presented below are averages over 20 distinct reference solutions.

Consider first the model error in the absence of observational noise, shown in Fig. 11. The model error grows quite quickly, reaching the level of the observational noise in about 3 days. The error saturates at a value similar to the climatological error level for the perfect model experiments of

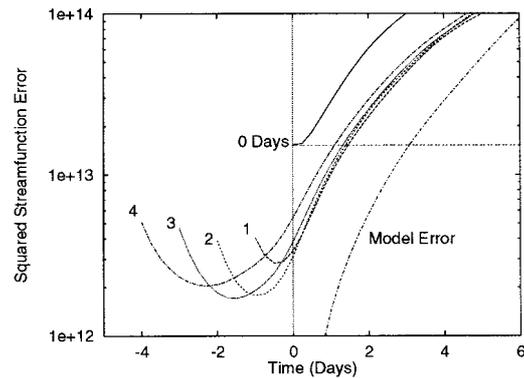


Fig. 11. Effect of extending the assimilation period T_a in the presence of model error when the predictive model is truncated to T20.

Fig. 5. The dynamical origin of this model error is the loss of eddy energy due to the harsher truncation in the meridional direction. This is consistent with the result of Held and Phillips (1993), who showed that the level of eddy energy in spectral models depends strongly on model resolution at these relatively coarse resolutions.

Adding observational noise, assimilation is performed for the time periods $T_a = 0, 1, 2, 3,$ and 4 days. The assimilation and forecast error as a function of T_a are shown in Fig. 11. As in the perfect model experiments, there is a marked reduction in error when assimilation is carried out with $T_a = 1$ day compared to the $T_a = 0$ day case with no assimilation. The magnitude of this error reduction is similar to that in the perfect model experiments. However, extending T_a to 2 days results in very little additional forecast improvement, and the forecast skill actually degrades as T_a is extended to 3 and 4 days. Note that the quality of the assimilated state also degrades as T_a is extended (compare $T_a = 3$ and 4 days), although this occurs for somewhat larger T_a than does the degradation in forecast skill. There is also a significant increase in the growth rate of the forecast error compared to the perfect model forecast error growth rate. Comparison of Figs. 11 and 5 reveals that while the error growth rate for $T_a = 0$ days is similar to the error growth rate in the perfect model in the absence of assimilation, the forecast error growth rate for $T_a = 1$ day is nearly twice as large as the $T_a = 1$ day perfect model rate. Within this context, however, model

error emerges as the reason for the rapid growth of forecast error, as the curve for the $T_a = 1$ day assimilated forecast error follows the model error curve exactly. This highlights that in the presence of model error, it is unrealistic to expect the assimilation scheme to produce an initial condition leading to a forecast error that grows more slowly than the underlying model error itself.

The truncated imperfect model results do present one additional intriguing result, namely that the optimal assimilation time is approximately half the time it takes model error to reach the level of the observational error. To test this hypothesis, we turn to an imperfect model where the magnitude of the model error can be better controlled.

5.2. Errors in forcing

Returning to the T21 version of the model used in the perfect model experiments above, we introduce a model error by adding a perturbation to the empirical forcing term S in the nonlinear equation of motion (2.1). If we let S denote the true forcing used in the perfect model experiments, an erroneous forcing is generated by scaling that forcing by some constant α ,

$$S_{\text{err}} = (1 + \alpha)S. \quad (5.1)$$

This error in empirical forcing results in a drift of the modeled climate with systematic error for $\alpha = 0.08$ is shown in Fig. 12. It has a maximum magnitude of approximately 50 m geopotential height on the 500 mb surface, similar to the 60 m geopotential height errors of typical of the current generation of forecast models (Anderson, 1993; Déqué and Royer, 1992).

In contrast to the truncated imperfect case, Fig. 13 shows that in the absence of observational noise, model error resulting from erroneous forcing amplifies at about the same rate as the $T_a = 10$ days perfect model forecast error for all values of α . As α increases, the model error reaches the level of the observational noise more quickly, ranging from 14 days for $\alpha = 0.02$ and 3 days for $\alpha = 0.32$.

Adding observational noise, assimilation is carried out on 20 distinct reference solutions for model error forcing values of $\alpha = 0.02, 0.08, 0.16,$ and 0.32 , and for a variety of values of T_a . Fig. 14a

shows the assimilation and forecast errors for these values of α for the assimilation time period T_a when the median $t = 0$ day error for the 20 forecasts is the smallest. As one would expect, smaller levels of erroneous forcing allow T_a to be extended further, ranging from as long as $T_a = 7$ days for $\alpha = 0.02$ to as short as $T_a = 2$ days for $\alpha = 0.32$. As in the truncated imperfect model experiments, however, extending T_a beyond this optimal assimilation time leads to a rapid degradation in the quality of the assimilated state. Again, a reasonable estimate of this optimal T_a is about half the time it takes the model error to reach the level of the observational error.

With regard to extending these results to an operational environment, however, it is encouraging to note that even in the presence of an erroneous forcing of $\alpha = 0.08$, generating a systematic error on the order of 50 m in the 500 mb geopotential height field, the optimal assimilation time period is $T_a = 5$ days. Fig. 14b shows the assimilation and forecast error for $T_a = 0, 1,$ and 5 days for model error $\alpha = 0.08$. Even for this level of model error, the gain in predictability time by extending the assimilation period from 1 to 5 days is about 3 days. This leaves hope for the implementation of 4DVAR with long assimilation periods even using current operational models.

The forecast skill of the forced imperfect model is clearly better than that of the truncated imperfect model (compared Figs. 14a and 11). This suggests that the growth rate of the model error determines whether the gain in forecast lead times using long assimilation periods in the perfect model setting can be achieved in an imperfect model setting. The model error growth rates for the erroneous empirical forcing are similar to the growth rates leading LVs for the MM93 model, provided α is sufficiently small ($\alpha < 0.16$). This differs from the truncated imperfect model, where the model error grows so quickly that it swamps the growth of error due to amplification of observational errors by internal dynamical processes. While in an operational setting model error growth is impossible to measure absolutely, it can be estimated by comparing different generations of forecast models. Provided that the growth rate of this model error is similar to that of the leading LVs of the system, we expect the perfect model results here to extend to the imperfect model case. Hence, reductions in model error can be expected

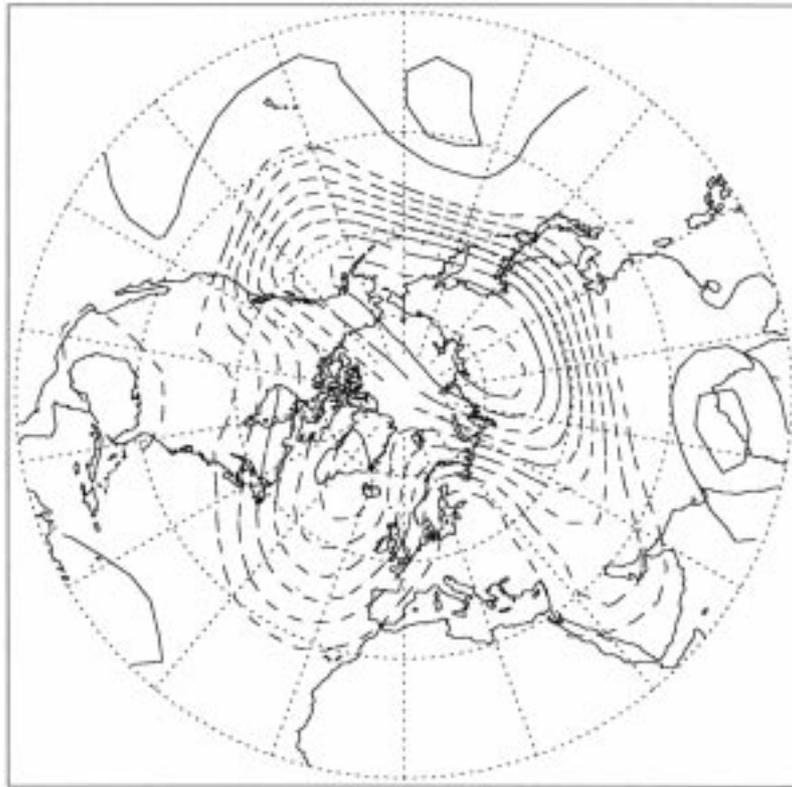


Fig. 12. Model climate drift of the 500 mb geopotential height for a forcing value of $\alpha = 0.8$. Contour interval is 5 m, with negative contours dashed and the zero contour omitted.

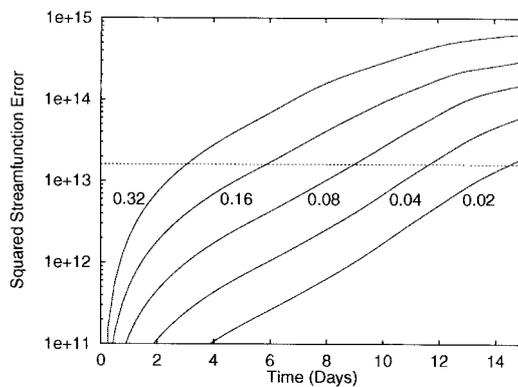


Fig. 13. Model error generated by the addition of anomalous forcing of strength α , denoted by the label closest to each curve. The dashed curve denotes the error level of the observational noise used throughout the text.

to provide a two-fold benefit to forecasting. Not only will the model better mimic the real atmosphere, but smaller model errors will also allow for longer assimilation periods and the accompanying improvements in forecast initial conditions and slower overall forecast error growth rates.

6. Summary and conclusions

We have performed an investigation into the performance of four-dimensional variational assimilation of noisy observations in a model of relatively realistic phenomenology compared to the real atmosphere, the quasi-geostrophic model designed by Marshall and Molteni (1993) (MM93). The study is made within both the perfect model setting, by assuming that the only source of forecast error lies in the initial condition,

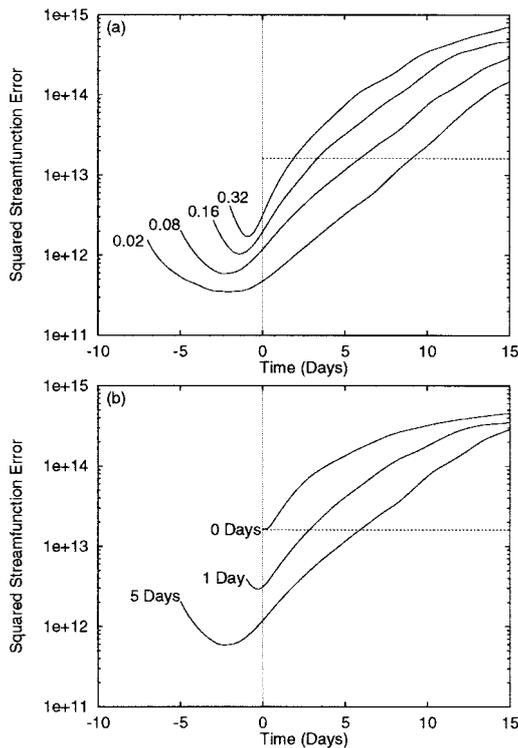


Fig. 14. (a) Assimilation and forecast error as a function of time for the error forcing values of $\alpha = 0.02, 0.08, 0.16$, and 0.32 . The assimilation time period T_a used for each value of α is the period for which the median $t = 0$ day error is the smallest. (b) Assimilation and forecast skill for the forced imperfect model with $\alpha = 0.08$ and for several values of T_a . The dashed line denotes the level of observational noise in both panels.

and the imperfect model setting, where both observational and model error are included. Within the perfect model setting, the quality of the assimilated state improves significantly when the assimilation time interval is extended significantly into the past, with the state estimation saturating at an assimilation period of 10 days. This result is consistent with the conjecture of Pires et al. (1996) that the optimal assimilation time period for the atmosphere should be about 3 times the nonlinear time scale of the model as measured by the e -folding time of small initial errors. Extension of the assimilation time interval to long times requires the use of the quasi-static variational assimilation (QSVA) algorithm designed by Pires et al. (1996) in order to keep track of the absolute minimum of the cost

function. For the MM93 model, multiple minima that degrade the quality of the assimilated state if the assimilation is attempted without the QSVA iterations emerge in the cost function for assimilation time intervals greater than about 5 days.

The reduction in error in the assimilated solutions is quite impressive, with error in the streamfunction squared norm reduced by a factor of 70 over the level of observational error for an assimilation period of 10 days (Fig. 5). The increase in forecast lead times is equally impressive, exceeding 15 days for global forecasts for realistic levels of observational error. An assimilation period of 10 days extends the forecast lead times about 5 days beyond the case where the assimilation period is 1 day. Even weather regime transitions in the extra-tropics, known to be difficult to predict in operational settings, are predictable at lead times of greater than 10 days using QSVA with an assimilation period of 10 days.

The structure of the analysis error under 4DVAR was examined. We showed that this error is most concentrated in the storm tracks and at relatively large scales (total wavenumbers 3–7). The error reduction is increasingly efficient at small scales as the assimilation period is increased. To examine the relative efficiency of Lyapunov vector (LVs) (Toth and Kalnay, 1993; Houtekamer, 1995) versus singular vectors (Lacarra and Talagrand, 1988; Molteni and Palmer, 1993; Molteni et al., 1996) in representing the analysis error for the design of ensemble prediction strategies, we compared the projection of the assimilated solutions into the subspaces spanned by the leading LVs and singular vectors, respectively. As suggested by the work of Pires et al. (1996), our results show that when long assimilation periods are used, assimilation errors are mostly concentrated on the unstable manifold of the model, or alternatively, along the leading LVs. However, the leading LVs and SVs are not strictly orthogonal, and the very small projection of analysis errors onto the leading SVs is the component of the error that will dominate the future forecast error at later times. This result hints at a hybrid strategy for generating perturbations for probabilistic weather prediction in the medium range. Our results show that in the perfect model setting, LVs rather than SVs represent analysis errors. Hence, perturbations generated for probabilistic forecasting should be restricted to

the subspace spanned by the leading LVs. However, the number of potentially useful LVs can be large. For the simple MM93 model here, it takes 100 LVs to describe 70% of the analysis error variance. It is expected that more realistic models will require hundreds of LVs in order to have a realistic representation of the analysis error. Obviously, such a large subspace cannot be sampled efficiently given current limitations on computing power. However, for an assimilation period of 10 days, we found that as few as 20 leading SV directions account for much of the forecast error after day 3. Thus, we hypothesize that the projection of the leading LVs onto a small number of SV directions should achieve the best compromise between an ensemble's ability to represent the analysis error and limitations due to sample size. Needless to say, the efficiency of the last projection stage for ensemble prediction is not guaranteed and must be tested using simple models such as the one used here.

The performance of variational assimilation in an imperfect model setting varies strongly with both the amount of growth rate of the model error. For large, rapidly growing model error, extension of assimilation periods beyond 1–2 days results in a degradation in the quality of the assimilated state as well as in the forecast quality. However, for model error growth rates similar to the growth rates of the leading LVs of the system, improvements in the assimilated state similar to those found for the perfect model appear possible. Based upon this, it is estimated that assimilation times of as large as 5 days for current values of model error might provide improved assimilation states and forecast initial conditions in an operational setting. More generally, we found that the optimal assimilation period length should be about half the time taken by the model error to reach the level of observational error when perfect initial conditions are assumed. A fundamental outstanding issue is whether assimilation methods in which model error as well as observational error are represented, such as the Extended Kalman Filter, can improve the assimilation and forecast skill shown for variational schemes. Examination of this question with a model of intermediate complexity such as the MM93 model is certainly feasible, but necessarily must be left to future study. However, it should be noted that QSVA-like algorithms have additional appeal from the

operational point of view. This is because there algorithms can be started before all observations are available to maximize the computational time available to produce an assimilated state, and observations that arrive after the "cut-off" time can be blended in as they arrive (Jarvinen et al., 1996). This is a desirable feature due to the heavy computational burden that is required by this assimilation method.

Another important outstanding question regarding the application of variational assimilation concerns the performance of 4DVAR in the presence of instabilities of unobserved scales. The MM93 model used here as baroclinic instability active on synoptic scales, but the small scales of the model are stable, as witnessed by the preferential reduction in error at small scales by the 4DVAR (Fig. 7). However, in the real atmosphere there are both balanced instabilities on subsynoptic scales, such as frontal-type instabilities, as well as meso and micro-scale moist convective instabilities. These small scale instabilities undoubtedly will affect the ability to perform data assimilation even on synoptic scales.

Above all, the results presented in this article are intended to provide hope for the development of more elaborate 4DVAR strategies. Fig. 5 exemplifies the potential of long period 4DVAR; the 5-day gain in predictability time for assimilation periods of 10 days compared to assimilation periods of 1 day roughly represents the potential gain in predictability that can be expected in the future due to proper application of 4DVAR. As such, there is good cause to believe that substantial gains in predictability are still possible, as models become ever more "perfect" and computer power increases, relative to current applications of optimal interpolation and 3DVAR.

7. Acknowledgements

We particularly wish to thank Olivier Talagrand and Philippe Courtier for their encouragement and advice throughout this work. Fruitful discussions with Tim Palmer, Anthony Holingsworth, and Roberto Buizza during a visit of RV and ECMWF in April 1997 were also very helpful, and led to a closer collaboration between LMD and ECMWF. The project was partly sponsored by the French Groupement De Recherche (GDR)

on variational methods in oceanography and meteorology. The stay of KLS at LMD was made possible by the North Atlantic Treaty Organization under a Grant awarded in 1995.

REFERENCES

- Anderson, J. L. 1993. The climatology of blocking in a numerical forecast model. *J. Climate* **6**, 1041–1056.
- Courtier, P., J.-N. Thépaut and A. Hollingsworth, 1994. A strategy for operational implementation of 4D-VAR, using an incremental approach. *Q. J. R. Met. Soc.* **120**, 1367–1387.
- Déqué, M. and J. F. Royer, 1992. The skill of extended range extra-tropical winter dynamical forecasts. *J. Climate* **5**, 1346–1356.
- Eckmann, J.-P. and D. Ruelle, 1985. Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.* **57**, 617–656.
- Farrell, B. F. 1990. Small error dynamics and the predictability of atmospheric flows. *J. Atmos. Sci.* **47**, 2409–2416.
- Gauthier, P. 1992. Chaos and quadri-dimensional data assimilation. A study based on the Lorenz model. *Tellus* **44A**, 2–17.
- Ghil, M. and P. Malanotte-Rizzoli, 1991. Data assimilation in meteorology and oceanography. *Adv. Geophys.* **33**, 141–266.
- Gilbert, J. C. and C. Lemaréchal, 1989. Some numerical experiments with variable storage quasi-Newton algorithms. *Mathematical Programming* **45**, 407–435.
- Guckenheimer, J. and P. Holmes, 1983. *Nonlinear oscillations, dynamical systems and bifurcations of vector fields*. New York: Springer-Verlag, 453 pp.
- Hartmann, D. L., R. Buizza and T. N. Palmer, 1995. The effect of spatial scale on the linear growth of disturbances. *J. Atmos. Sci.* **52**, 3885–3894.
- Held, I. M. and P. J. Phillips 1993. Sensitivity of the eddy momentum flux of meridional resolution in atmospheric GCM's. *J. Climate* **6**, 499–507.
- Houtekamer, P. L. 1995. The construction of optimal perturbations. *Mon. Wea. Rev.* **123**, 2888–2898.
- Houtekamer, P. L. and J. Derome, 1995. Methods for ensemble prediction. *Mon. Wea. Rev.* **123**, 2181–2196.
- Jarvinen, H., J.-N. Thépaut and P. Courtier, 1996. Quasi-continuous variational data assimilation. *Q. J. R. Met. Soc.* **122**, 515–534.
- Kalman, R. 1960. A new approach to linear filtering and prediction theory. *J. Basic. Engr.* **82D**, 35–45.
- Lacarra, J. and O. Talagrand, 1988. Short range evolution of small perturbations is a barotropic model. *Tellus* **40A**, 81–95.
- Le Dimet, F.-X. and O. Talagrand, 1986. Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus* **38A**, 97–110.
- Legras, B. and R. Vautard, 1995. A guide to Liapunov vectors. *ECMWF Seminar on Predictability* **1**, 143–156.
- Lewis, J. and J. C. Derber, 1985. The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus* **37A**, 309–322.
- Li, Y. 1991. A note on the uniqueness problem of variational adjustment approach to four-dimensional data assimilation. *J. Met. Soc. Japan* **69**, 581–585.
- Lindzen, R. S. 1988. Instability of plane parallel shear flow (towards a mechanistic picture of how it works). *PAGEOPH* **126**, 103–121.
- Liu, Q. 1994. On the definition and persistence of blocking. *Tellus* **46A**, 286–298.
- Lorenz, E. N. 1963. Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130–141.
- Marshall, J. and F. Molteni, 1993. Toward a dynamical understanding of the planetary scale flow regimes. *J. Atmos. Sci.* **50**, 1792–1818.
- Michelangeli, P. A. 1996. *Variabilité atmosphérique basse-fréquence observée et simulée aux latitudes tempérées*. PhD dissertation, Université de Paris VI.
- Miller, R. N., M. Ghil and F. Gauthiez, 1994. Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.* **51**, 1037–1056.
- Molteni, F. and T. N. Palmer, 1993. Predictability and finite time instability of the northern hemisphere winter circulation. *Q. J. R. Met. Soc.* **119**, 269–298.
- Molteni, F., R. Buizza, T. N. Palmer and T. Petroliagis, 1996. The ECMWF ensemble prediction system: methodology and validation. *Q. J. R. Met. Soc.* **122**, 73–120.
- Oortwijn, J. and J. Barkmeijer, 1995. Perturbations that optimally trigger weather regimes. *J. Atmos. Sci.* **52**, 3932–3944.
- Oor, W. M. F. 1907. The stability or instability of the steady motions of a perfect liquid and of a viscous liquid. *Proc. Roy. Irish Acad.* **A27**, 9–138.
- Palmer, T. N. 1993. Extended range atmospheric prediction and the Lorenz model. *Bull. Am. Met. Soc.* **74**, 49–65.
- Palmer, T. N. 1995. Predictability of the atmosphere and oceans: From days to decades. *ECMWF Seminar on predictability* **1**, 83–140.
- Pires, C., R. Vautard, and O. Talagrand, 1996. On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus* **48A**, 96–121.
- Press, W. H., B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, 1992. *Numerical Recipes in Fortran*. New York: Cambridge University Press, 702 pp.
- Stensrud, D. J. and J. W. Bao, 1992. Behaviors of variational and nudging assimilation techniques with a chaotic low-order model. *Mon. Wea. Rev.* **120**, 3016–3028.
- Stoer, J. and R. Bulirsch, 1993. *Introduction to numerical analysis*. New York: Springer-Verlag, 468 pp.
- Talagrand, O. and P. Courtier, 1987. Variational assimilation of meteorological observations with the adjoint

- vorticity equation (I). Theory. *Q. J. R. Met. Soc.* **113**, 1311–1328.
- Tanguay, M., P. Bartello and P. Gauthier, 1995. Four-dimensional data assimilation with a wide range of scales. *Tellus* **47A**, 974–997.
- Thépaut, J. N. and Courtier, 1991. Four dimensional variational data assimilation using the adjoint of a multilevel primitive equation model. *Q. J. R. Met. Soc.* **117**, 1225–1254.
- Tibaldi, S. and F. Molteni, 1990. On the operational predictability of blocking. *Tellus* **42A**, 343–365.
- Toth, Z. and E. Kalnay, 1993. Ensemble forecasting at NMC: The generation of perturbations. *Bull. Am. Met. Soc.* **74**, 2317–2330.
- Trevisan, A. and R. Legnani, 1995. Transient error growth and local predictability: A study in the Lorenz system. *Tellus* **47A**, 103–117.
- Vannitsen, S. and C. Nicolis, 1997. Lyapunov vectors and error growth patterns in a T21L3 quasigeostrophic model. *J. Atmos. Sci.* **54**, 347–361.
- Vautrad, R. 1990. Multiple weather regimes over the North Atlantic. Analysis of precursors and successors. *Mon. Wea. Rev.* **118**, 2056–1081.
- Vukicevic, T. 1991. Nonlinear and linear evolution of initial forecast errors. *Mon. Wea. Rev.* **119**, 1602–1611.