Typeset using LATEX2e under UNIX

Low-dimensional representation of error covariance

Michael K. Tippett[*], Stephen E. Cohn[†],
Ricardo Todling[‡], and Dan Marchesin[§]

January 14, 2000

[*]Corresponding author: Dr. Michael K. Tippett, International Research Institute for climate prediction, PO Box 1000 / 61 Route 9W, Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY 10964-8000, USA (tippett@iri.ldgo.columbia.edu). This work was done while the author was with the Centro de Previsão de Tempo e Estudos Climáticos, Cachoeira Paulista, SP, Brazil.

[†]Data Assimilation Office, NASA/GSFC, Greenbelt, MD, USA.

[‡]General Sciences Corp./SAIC, NASA/GSFC/DAO, Greenbelt, MD, USA.

[§]Instituto de Matemática Pura e Aplicada, Rio de Janeiro, RJ, Brazil.

Abstract

Ensemble and reduced-rank approaches to prediction and assimilation rely on low-dimensional approximations of the estimation error covariances. Here stability properties of the forecast/analysis cycle for linear, time-independent systems are used to identify factors that cause the steady-state analysis error covariance to admit a low-dimensional representation. A useful measure of forecast/analysis cycle stability is the *bound matrix*, a function of the dynamics, observation operator and assimilation method. Upper and lower estimates for the steady-state analysis error covariance matrix eigenvalues are derived from the bound matrix. The estimates generalize to time-dependent systems. If much of the steady-state analysis error variance is due to a few dominant modes, the leading eigenvectors of the bound matrix approximate those of the steady-state analysis error covariance matrix.

The analytical results are illustrated in two numerical examples where the Kalman filter is carried to steady state. The first example uses the dynamics of a generalized advection equation exhibiting nonmodal transient growth. Failure to observe growing modes leads to increased steady-state analysis error variances. Leading eigenvectors of the steady-state analysis error covariance matrix are well approximated by leading eigenvectors of the bound matrix. The second example uses the dynamics of a damped baroclinic wave model. The leading eigenvectors of a lowest-order approximation of the bound matrix are shown to approximate well the leading eigenvectors of the steady-state analysis error covariance matrix.

1. Introduction

Accurate and realistic representations of forecast and analysis errors are key to the performance of atmospheric data assimilation and ensemble prediction systems. Unfortunately, forecast and analysis error statistics such as the covariance are impractical to compute directly for operational-scale problems, and approximate methods must be used. Primary sources of difficulty are the large number of degrees of freedom present in numerical forecast models and poor understanding of forecast model error, making it infeasible to store and evolve the complete error covariances.

Data assimilation systems combine observations with a forecast model first guess to produce an analysis or estimate of the state of the atmosphere. If the error statistics of the forecast and of the observations are known, then the relative weights given to the first guess and to the observations can be chosen so that the analysis error is minimized. Data assimilation systems are sensitive to the specification of the forecast error statistics, and misspecification of them increases analysis error levels. In current operational practice, the forecast error covariance is generally approximated by an analytical state-independent model depending on a number of estimated parameters (Courtier et al., 1998; Rabier et al., 1998; Dee and da Silva, 1999; Dee et al., 1999). Though the details of these error models vary, their simplicity means that realistic features, such as anisotropy and flow-dependent structure, are not included. A general way of including some of these features is through reduced-rank methods, where evolution of the error covariances or the error covariances themselves are replaced by low-dimensional representations (Evensen, 1994; Cohn and Todling, 1996; Cane et al., 1996; Verlaan and Heemink, 1997).

Ensemble prediction schemes produce an ensemble of forecasts, each starting from slightly different initial conditions. The quality of the ensemble forecast, particularly for short to medium-range forecasting, is sensitive to the structure of the perturbations used to form the ensemble of initial conditions. Ideally, the ensemble of initial conditions should reflect the errors present in the analysis. Operational ensemble forecast systems use state-dependent initial perturbations that are related to either past or future growing modes of the dynamics (Toth and Kalnay, 1993; Molteni et al., 1996). The existence of a distinction between past and future growing modes stems from the time-dependence and nonnormality of the tangent linear dynamics which is linearized about a nonlinear, time-varying trajectory. If the dynamics were time-independent and normal (a linear operator is *normal* when it has a complete set of orthogonal eigenvectors), both past and future growing modes would coincide with the eigenvectors or normal modes. Time-independent dynamics, for instance, the result of linearizing about a mean state or of statistical modeling, will typically be nonnormal and present behavior that is primarily nonmodal (Blumenthal, 1991; Whitaker and Sardeshmukh, 1998). The rationale for using past growing modes (approximate Lyapunov vectors) as initial perturbations is that they represent the analysis error due to errors in the first guess. However, there is no guarantee that past growing modes efficiently sample the space of rapidly growing analysis errors. On the other hand, future growing modes (singular vectors of the dynamics), while presenting optimal growth, intrinsically have little to do with the analysis error. If information about the analysis errors is available, the singular vectors of the dynamics can be used to identify the analysis errors that will contribute most to forecast error (Barkmeijer et al., 1998; Swanson et al., 1998). In any case, since an ensemble covariance is by construction low-rank, it is useful to identify factors that lead to the actual error covariances having good low-dimensional approximations.

In this article the analysis error covariance structure is examined in the context of linear, time-independent dynamics and observation operators. The aim is to understand under what conditions the analysis error covariance admits a low-dimensional representation. Such information would be useful in both ensemble forecast systems and data assimilation systems, though our results in this idealized context can only give qualitative information. Our approach is to identify errors which are not efficiently reduced by the forecast/analysis cycle. If the system is observable and the Kalman filter, or more generally a stable data assimilation method, is used, then all analysis errors are *eventually* reduced, that is to say, the forecast/analysis cycle is exponentially stable. A linear system and its measurements is completely observable when it is possible to reconstruct uniquely the state of the system from past measurements (Cohn and Dee, 1988). However, it may be the case that the damping of analysis errors by the forecast/analysis cycle is slow or that analysis errors may be amplified nonmodally on transient time-scales before being reduced. Only in special cases is the behavior of forecast/analysis cycle purely modal, for instance if the dynamics is normal, all variables are observed, and the observational and forecast model errors are homogeneous (Daley and Ménard, 1993).

Our motivation is complementary to that of Swanson et al. (1998), who studied the analysis error structure of four-dimensional variational assimilation applied to a nonlinear quasi-geostrophic model. There, in the perfect model setting with all state variables observed, it was found that in the limit of a long assimilation period, analysis errors projected primarily onto Lyapunov vectors. These Lyapunov vectors were found to project weakly onto the leading right singular vectors of the dynamics, although this part of the analysis error was responsible for most of the forecast error. Distinguishing features of the present work are treatment of the interaction of observing pattern and dynamics, and analytical results that include model error but do not require its complete details.

We make extensive use of the fact that the steady-state analysis error covariance matrix satisfies a discrete algebraic Lyapunov equation (DALE) in the case of time-independent dynamics, observation operators, forecast model error and observation error covariances. Investigation of the DALE permits the identification and estimation of the dominant part (leading eigenvalues and eigenvectors) of the steady-state analysis error covariance and understanding of its dependence on the dynamics and observing pattern. The transient and asymptotic behavior of the analysis errors, as measured by the *bound matrix*, provides rigorous estimates for the eigenvalues and eigenvectors of the steady-state analysis error covariance. A bound matrix giving time-dependent estimates for the time-dependent analysis error covariance can also be defined when the dynamics, observing pattern and error covariance sources vary in time.

The organization of this paper is as follows. In Section 2 the generic linear, time–independent forecast and analysis system is described. In Section 3 the bound matrix is defined and employed to investigate the steady-state analysis error covariance matrix. Sections 4 and 5 illustrate the results of Section 3 by applying the Kalman filter to two simple dynamical models: a generalized advection equation and a model for damped baroclinic waves. Finally, a summary and conclusions are given in Section 6. Some technical details and the time-dependent problem are presented in the Appendix.

2. Linear Time-Independent Forecast/Analysis Cycle

A general linear forecast/analysis cycle is

$$\mathbf{y}_k^f = \mathbf{\Psi}_k \mathbf{y}_{k-1}^a \,, \tag{1}$$

$$\mathbf{y}_k^a = \mathbf{y}_k^f + \mathbf{K}_k \left(\mathbf{y}_k^o - \mathbf{H}_k \mathbf{y}_k^f \right) \,, \tag{2}$$

where the forecast, analysis and observation vectors at time step t_k are denoted by $\mathbf{y}_k^f$, $\mathbf{y}_k^a$ and $\mathbf{y}_k^o$ respectively. The dynamics matrix $\mathbf{\Psi}_k$ advances the system from one analysis time to the next. The matrix $\mathbf{H}_k$ is the observation operator or generalized interpolation operator that maps forecast state variables to observation space. The dynamics and observation operator are taken to be linear and also time-independent for most of our results. This assumption on $\mathbf{H}_k = \mathbf{H}$ corresponds to a fixed, *in situ*, observing system. The assumption of linear dynamics (linearized about a non-linear trajectory) is not unreasonable for short times, i.e., up to one or two days in atmospheric models (Courtier and Talagrand, 1987). Stochastically forced linear models are also sometimes appropriate in an average sense (Blumenthal, 1991; Xue et al., 1994). The time-independence of the dynamics, $\mathbf{\Psi}_k = \mathbf{\Psi}$, is a stronger restriction, although recent studies have shown that a simple model linearized about the long-term mean flow and stochastically forced may be capable of reproducing aspects of observed storm-track variability (Whitaker and Sardeshmukh, 1998). In the Appendix we show how the time-independence requirements on the dynamics and observation operator can be relaxed.

The state estimate $\mathbf{y}_k^a$ and the observations $\mathbf{y}_k^o$ are real vectors of length n and p respectively, usually with $p \ll n$; $\mathbf{\Psi}_k$ and $\mathbf{H}_k$ are real $n \times n$ and $p \times n$ matrices respectively; the gain $\mathbf{K}_k$ is an $n \times p$ matrix. Typically in statistical interpolation and 3D-Var data assimilation methods the gain at time t_k is effectively given by

$$\mathbf{K}_k = \mathbf{S}_k^f \mathbf{H}_k^T (\mathbf{H}_k \mathbf{S}_k^f \mathbf{H}_k^T + \mathbf{R}_k)^{-1} , \tag{3}$$

where $\mathbf{S}_k^f$ is a model of the forecast error covariance and $\mathbf{R}_k$ is a model of the observation error covariance (Cohn, 1997; Cohn et al., 1998).

We assume the observation error to be additive and white in time with mean zero and $p \times p$ covariance matrix $\mathbf{R}_k$, and the forecast model error to be additive and white in time with mean zero and $n \times n$ covariance matrix $\mathbf{Q}_k^m$. In other words, denoting the true state of the atmosphere by $\mathbf{y}_k^t$, the observation vector $\mathbf{y}_k^o$ is related to the true state by

$$\mathbf{y}_k^o = \mathbf{H}_k \mathbf{y}_k^t + \mathbf{b}_k^o , \quad \langle \mathbf{b}_k^o \rangle = 0 , \quad \left\langle \mathbf{b}_k^o (\mathbf{b}_l^o)^T \right\rangle = \delta_{kl} \mathbf{R}_k , \tag{4}$$

and the evolution of the true state is given by

$$\mathbf{y}_k^t = \mathbf{\Psi}_k \mathbf{y}_{k-1}^t + \mathbf{b}_k^m , \quad \langle \mathbf{b}_k^m \rangle = 0 , \quad \left\langle \mathbf{b}_k^m (\mathbf{b}_l^m)^T \right\rangle = \delta_{kl} \mathbf{Q}_k^m ; \tag{5}$$

$\mathbf{b}_k^o$ and $\mathbf{b}_k^m$ are the observation and forecast model error respectively; δ_{kl} is the Kronecker delta. Additionally the forecast model and observation errors are assumed to be uncorrelated with each other, $\langle \mathbf{b}_k^m \, (\mathbf{b}_l^o)^T \rangle = 0$, and with the initial error, $\left\langle \mathbf{b}_k^m (\mathbf{y}_0^t - \mathbf{y}_o^a)^T \right\rangle = 0$ and $\left\langle \mathbf{b}_k^o (\mathbf{y}_0^t - \mathbf{y}_o^a)^T \right\rangle = 0$. We have neglected biases, which may be a particularly important part of the model error and should be estimated separately when present (Dee and da Silva, 1998).

Under these assumptions, the time evolution of the analysis error $\mathbf{y}_k \equiv \mathbf{y}_k^t - \mathbf{y}_k^a$ is given by

$$\mathbf{y}_k = \mathsf{A}_k \mathbf{y}_{k-1} + \mathbf{b}_k \,, \quad \langle \mathbf{b}_k \rangle = 0 \,, \quad \left\langle \mathbf{b}_k \mathbf{b}_l^T \right\rangle = \delta_{kl} \mathsf{Q}_k \,, \quad \left\langle \mathbf{y}_0 \mathbf{b}_l^T \right\rangle = 0 \,, \tag{6}$$

where the *system matrix* A_k and the *system error covariance matrix* Q_k are defined by

$$\mathsf{A}_k \equiv (\mathsf{I} - \mathsf{K}_k \mathsf{H}_k) \boldsymbol{\Psi}_k, \quad \mathsf{Q}_k \equiv (\mathsf{I} - \mathsf{K}_k \mathsf{H}_k) \mathsf{Q}_k^m (\mathsf{I} - \mathsf{K}_k \mathsf{H}_k)^T + \mathsf{K}_k \mathsf{R}_k \mathsf{K}_k^T \,; \tag{7}$$

I is the $n \times n$ identity matrix. The system matrix and system error covariance matrix both depend on the gain matrix K_k. The system matrix A_k propagates analysis errors from one analysis time to the next. Roughly speaking, one expects analysis errors to grow in the forecast step $(\boldsymbol{\Psi}_k)$ and to be reduced by the analysis step $(\mathsf{I} - \mathsf{K}_k \mathsf{H}_k)$. The stochastic forcing part $\mathbf{b}_k$ of the analysis error propagation (6), with error covariance Q_k, is due to both model and observation errors.

Henceforth we assume time-independence of the dynamics and the observation operator, $\boldsymbol{\Psi}_k = \boldsymbol{\Psi}$ and $\mathsf{H}_k = \mathsf{H}$, and also of the model and observation error covariances, $\mathsf{Q}_k^m = \mathsf{Q}^m$ and $\mathsf{R}_k = \mathsf{R}$. In this case, if the system (4), (5) is observable and if the (time-*dependent*) Kalman gain is used, then $\mathsf{A}_k \to \mathsf{A}_\infty \equiv \mathsf{A}$ and $\mathsf{Q}_k \to \mathsf{Q}_\infty \equiv \mathsf{Q}$, and the eigenvalues of the steady-state system matrix A all lie inside the unit circle (Kalman, 1960). Suboptimal choices of the gain matrix may also yield an asymptotic steady state (e.g., Cohn and Todling, 1996). In any case, we assume that A_k and Q_k tend to steady state, and that the steady-state system matrix $\mathsf{A} \equiv \mathsf{A}_\infty$ is *stable,* that is, all its eigenvalues lie inside the unit circle. The steady-state analysis error covariance P is then given by the large-time limit (see the Appendix)

$$\mathsf{P} \equiv \lim_{k \to \infty} \left\langle \mathbf{y}_k \mathbf{y}_k^T \right\rangle \,, \tag{8}$$

and satisfies

$$\mathsf{P} = \mathsf{A} \mathsf{P} \mathsf{A}^T + \mathsf{Q} \,. \tag{9}$$

The linear matrix equation (9) is known as the discrete algebraic Lyapunov equation (DALE; Gajić and Qureshi, 1995). Some properties of its solution will be discussed in the next section.

3. Theory

3.1. Normal system matrix

When the steady-state system matrix $\mathbf{A}$ is normal, the eigenvectors $\mathbf{z}_i$ of $\mathbf{A}$ form a complete and orthonormal basis, and $\mathbf{Q}$ can be decomposed in this basis as

$$\mathbf{Q} = \sum_{i=1}^{n}\sum_{j=1}^{n} \left(\mathbf{z}_i^* \mathbf{Q} \mathbf{z}_j\right) \mathbf{z}_i \mathbf{z}_j^* \, . \tag{10}$$

The solution of (9) is then

$$\mathbf{P} = \sum_{k=0}^{\infty} \mathbf{A}^k \mathbf{Q} \left(\mathbf{A}^T\right)^k = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\mathbf{z}_i^* \mathbf{Q} \mathbf{z}_j}{1 - \lambda_i(\mathbf{A})\overline{\lambda_j(\mathbf{A})}} \mathbf{z}_i \mathbf{z}_j^* \, ; \tag{11}$$

the notation $\lambda_i(\mathbf{A})$ is used to denote the i-th eigenvalue of the matrix $\mathbf{A}$, where the eigenvalues are ordered in decreasing magnitude; $\mathbf{z}_i$ is the corresponding eigenvector of $\mathbf{A}$ and $\mathbf{z}_i^*$ the conjugate transpose. Equation (11) shows that when $\mathbf{A}$ is normal, $\mathbf{P}$ is controlled by the proximity of the eigenvalues of $\mathbf{A}$ to the unit circle and by the relationship between the eigenvectors of $\mathbf{A}$ and $\mathbf{Q}$. For instance, the lower bound

$$\lambda_1(\mathbf{P}) \geq \mathbf{z}_1^* \mathbf{P} \mathbf{z}_1 = \frac{\mathbf{z}_1^* \mathbf{Q} \mathbf{z}_1}{1 - |\lambda_1(\mathbf{A})|^2} \tag{12}$$

shows that if the leading eigenvalue of the steady-state system matrix, $\lambda_1(\mathbf{A})$, is near the unit circle, and if the projection $\mathbf{z}_1^* \mathbf{Q} \mathbf{z}_1$ of the steady-state system error covariance $\mathbf{Q}$ onto the corresponding eigenvector $\mathbf{z}_1$ does not happen to be small, then $\mathbf{P}$ projects strongly onto the leading eigenvector $\mathbf{z}_1$ of the system matrix and the leading eigenvalue of $\mathbf{P}$ is large. In other words, the steady-state analysis error covariance $\mathbf{P}$ must have a large eigenvalue when the steady-state system matrix $\mathbf{A}$ has a weakly-damped eigenmode that is well-excited by the steady-state system error forcing. Alternatively, $\mathbf{z}_1^* \mathbf{Q} \mathbf{z}_1$ being large causes $\mathbf{z}_1^* \mathbf{P} \mathbf{z}_1$ to be large even when $|\lambda_1(\mathbf{A})| \ll 1$.

The connection between the eigendecompositions of $\mathbf{P}$ and $\mathbf{A}$ is especially simple when the eigenvectors of $\mathbf{A}$ are also eigenvectors of $\mathbf{Q}$ (equivalently when $\mathbf{A}$ and $\mathbf{Q}$ commute), in which case Eq. (11) simplifies to

$$\mathbf{P} = \sum_{i=1}^{n} \frac{\mathbf{z}_i^* \mathbf{Q} \mathbf{z}_i}{1 - |\lambda_i(\mathbf{A})|^2} \mathbf{z}_i \mathbf{z}_i^* \, ; \tag{13}$$

the eigenvalues of $\mathbf{P}$ are $(\mathbf{z}_i^* \mathbf{Q} \mathbf{z}_i)(1 - |\lambda_i(\mathbf{A})|^2)^{-1}$. In this case, when the leading eigenvalues of $\mathbf{A}$ are near the unit circle and the associated leading eigenvectors of $\mathbf{A}$ are well-excited by the system error, $\mathbf{P}$ will have large eigenvalues with eigenvectors that lie in the space spanned by the leading eigenvectors of $\mathbf{A}$.

In general, nonnormality of the dynamics and inhomogeneity of the observing network cause $\mathbf{A}$ to be nonnormal, and there is no simple connection between the eigenvectors of $\mathbf{P}$ and those of $\mathbf{A}$. However, as will be shown in the next subsection, there is still a connection between the leading eigenvectors of $\mathbf{P}$ and more general stability properties of the steady-state system matrix $\mathbf{A}$.

3.2. Nonnormal system matrix

When $\mathbf{A}$ is nonnormal its stability properties may not be well-described by its eigenvalues and eigenvectors (Trefethen et al., 1993; Farrell and Ioannou, 1996; Tippett et al., 2000). The damping of analysis errors may be much less efficient than predicted by the eigenvalues of $\mathbf{A}$, and the dominant error structures may not be described well by eigenvectors of $\mathbf{A}$. To obtain a more useful measure of the stability properties of $\mathbf{A}$, we examine the quantity $\mathbf{x}^T \mathbf{A}^N \mathbf{y}$ for any two real unit vectors $\mathbf{x}$ and $\mathbf{y}$. This quantity gives the response in the $\mathbf{x}$-direction to an analysis error in the $\mathbf{y}$-direction after N forecast/analysis cycles in the steady-state regime. The maximum of $\left|\mathbf{x}^T \mathbf{A}^N \mathbf{y}\right|$ defines the Euclidean matrix norm of $\mathbf{A}^N$,

$$||\mathbf{A}^N||_2 \equiv \max_{||\mathbf{x}||_2 = ||\mathbf{y}||_2 = 1} \left|\mathbf{x}^T \mathbf{A}^N \mathbf{y}\right| \, . \tag{14}$$

When $\mathbf{A}$ is normal and stable, $||\mathbf{A}^N||_2 = \left|\mathbf{z}_1^* \mathbf{A}^N \mathbf{z}_1\right| = |\lambda_1(\mathbf{A})|^N$, and $||\mathbf{A}^N||_2$ decays monotonically and exponentially as N increases since $|\lambda_1(\mathbf{A})| < 1$. Therefore when $\mathbf{A}$ is normal and stable, $||\mathbf{A}^N \mathbf{y}_k||_2 \leq ||\mathbf{y}_k||_2$ and the analysis error evolution equation (6) shows that near steady state, any transient analysis error growth must be due to the observational and model error forcing of the system error $\mathbf{b}_k$. In the general case when $\mathbf{A}$ is nonnormal, $\left|\mathbf{x}^T \mathbf{A}^N \mathbf{y}\right|$ attains its largest value when $\mathbf{x}$ and $\mathbf{y}$ are respectively the leading eigenvectors of $\mathbf{A}^N \left(\mathbf{A}^T\right)^N$ and $\left(\mathbf{A}^T\right)^N \mathbf{A}^N$ (equivalently, the leading left and right singular vectors of $\mathbf{A}^N$). The stability of $\mathbf{A}$ does not constrain $||\mathbf{A}^N||_2$ to be monotonically decreasing as N increases. As a consequence, $||\mathbf{A}^N \mathbf{y}_k||_2$ need not be less than $||\mathbf{y}_k||_2$; analysis errors in the steady-state regime can grow on transient time-scales even in the absence of observational and model error forcing. Asymptotically, however, $||\mathbf{A}^N||_2$ must behave like $|\lambda_1(\mathbf{A})|^N$ (Horn and Johnson, 1985):

$$\lim_{N\to\infty} \exp \frac{1}{N} \log \|\mathbf{A}^N\|_2 = |\lambda_1(\mathbf{A})| \,, \tag{15}$$

and eventually analysis errors are reduced. In this limit of large N, the leading left and right singular vectors of $\mathbf{A}^N$ are approximately the leading eigenvectors of $\mathbf{A}$ and its adjoint $\mathbf{A}^T$ respectively (Farrell, 1989).

In the steady-state regime, the maximum response in the $\mathbf{x}$-direction after N forecast/analysis cycles applied to a unit vector $\mathbf{y}$ is

$$\|\mathbf{x}^T \mathbf{A}^N\|_2 = \max_{\|\mathbf{y}\|_2=1} \left|\mathbf{x}^T \mathbf{A}^N \mathbf{y}\right| = \sqrt{\left(\mathbf{x}^T \mathbf{A}^N\right)\left(\mathbf{x}^T \mathbf{A}^N\right)^T} \,. \tag{16}$$

A measure of the maximum response in the $\mathbf{x}$-direction accumulated over all times-scales is obtained by summing the square of (16) over N:

$$\sum_{N=0}^{\infty} \|\mathbf{x}^T \mathbf{A}^N\|_2^2 = \sum_{N=0}^{\infty} \mathbf{x}^T \mathbf{A}^N \left(\mathbf{A}^T\right)^N \mathbf{x} = \mathbf{x}^T \mathbf{B} \mathbf{x} \,, \tag{17}$$

where the *bound matrix* $\mathbf{B}$ is defined by

$$\mathbf{B} = \sum_{N=0}^{\infty} \mathbf{A}^N (\mathbf{A}^T)^N \,. \tag{18}$$

Since $\mathbf{B}$ is a symmetric matrix, its eigenvalues have a "minimax" characterization (Golub and Van Loan, 1996),

$$\lambda_i(\mathbf{B}) = \max_{\dim(D)=i} \; \min_{\mathbf{x}\in D} \frac{\mathbf{x}^T \mathbf{B} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} \,. \tag{19}$$

Therefore the eigenvectors of the bound matrix $\mathbf{B}$ order the directions in state space according to the maximum response in those directions. In this sense, the eigenvectors of $\mathbf{B}$ play a role similar to that of Lyapunov vectors, identifying directions in state space associated with accumulated growth.

It is also useful to know which system error forcing produces the maximum or minimum steady-state analysis error variance since, as (7) shows, modifying the observing pattern can change the portion of the system error covariance due to model error. Applying the above arguments to the quantity $\|\mathbf{A}^N \mathbf{y}\|_2$ shows that the eigenvectors of the matrix $\mathbf{B}_T$ given by (Tippett and Marchesin, 1999*a*)

$$\mathbf{B}_T = \sum_{k=0}^{\infty} (\mathbf{A}^T)^k \mathbf{A}^k \,, \tag{20}$$

order the directions of state space according to the analysis error variance produced by system error forcing in those directions. When $\mathbf{A}$ is nonnormal, the eigenvectors of $\mathbf{B}$ and $\mathbf{B}_T$ are generally different, and therefore the directions of optimal response and forcing are different.

3.3. Estimates derived from the bound matrix

Given only limited information about the steady-state system error covariance $\mathbf{Q}$, the bound matrix can be used to estimate the steady-state analysis error covariance $\mathbf{P}$. Rigorous upper and lower matrix bounds for $\mathbf{P}$ are (Tippett and Marchesin, 1999a; see (54) of the Appendix for the corresponding bounds in the time-dependent case):

$$\lambda_n(\mathbf{Q})\mathbf{B} \leq \mathbf{P} \leq \lambda_1(\mathbf{Q})\mathbf{B}\,; \tag{21}$$

for two symmetric matrices $\mathbf{X}$ and $\mathbf{Y}$, $\mathbf{X} \leq \mathbf{Y}$ means that $\mathbf{Y} - \mathbf{X}$ is positive semidefinite. When the ratio $\lambda_1(\mathbf{Q})/\lambda_n(\mathbf{Q})$ is near unity, these upper and lower bounds tightly restrict $\mathbf{P}$. If more information about $\mathbf{Q}$ is available, tighter bounds can be obtained (see (56) of the Appendix and Tippett and Marchesin, 1999*a*). The estimates in (21) can be used to derive upper and lower bounds for the eigenvalues of $\mathbf{P}$ and therefore for the trace of $\mathbf{P}$:

$$\lambda_n(\mathbf{Q})\lambda_i(\mathbf{B}) \leq \lambda_i(\mathbf{P}) \leq \lambda_1(\mathbf{Q})\lambda_i(\mathbf{B})\,, \tag{22}$$

and

$$\lambda_n(\mathbf{Q})\,\mathrm{tr}\,\mathbf{B} \leq \mathrm{tr}\,\mathbf{P} \leq \lambda_1(\mathbf{Q})\,\mathrm{tr}\,\mathbf{B}\,, \tag{23}$$

where tr denotes the trace. Again, the upper and lower bounds are tight when the spectrum of $\mathbf{Q}$ is fairly flat.

When the bounds in (21) are tight and $\mathbf{P}$ has a well-separated set of leading eigenvalues, the leading eigenvectors of $\mathbf{P}$ and $\mathbf{B}$ span approximately the same subspaces (Theorem 7.2.4 of Golub and Van Loan, 1996). However, if $\mathbf{P}$ does not have a well-separated set of leading eigenvalues, $\mathbf{P} - \lambda_1(\mathbf{Q})\mathbf{B}$ may be small without the leading eigenvectors of $\mathbf{P}$ and $\mathbf{B}$ spanning approximately the same subspaces.

3.4. Estimates derived from approximation of the bound matrix

For systems where n is large, calculating the bound matrix $\mathbf{B}$ directly may be as impractical as calculating $\mathbf{P}$. However, if the series in (18) converges rapidly, a few terms may provide a good approximation for $\mathbf{B}$, and therefore an additional theoretical tool. Iterative Lanczos methods can be used to calculate the leading eigenvalues and eigenvectors of the partial sum (Golub and Van Loan, 1996). The partial sum $\mathbf{B}_L$ is defined by either

$$\mathbf{B}_L = \sum_{k=0}^{L} \mathbf{A}^k \left(\mathbf{A}^T\right)^k , \tag{24}$$

or recursively by $\mathbf{B}_L = \mathbf{A}\mathbf{B}_{L-1}\mathbf{A}^T + \mathbf{I}$, $\mathbf{B}_0 = \mathbf{I}$. The cost of calculating eigenvectors of $\mathbf{B}_L$ with an iterative method is about a factor L greater than that of iteratively calculating singular vectors of $\mathbf{A}$. Monte Carlo simulations of (6) are another method of calculating $\mathbf{B}$ (see the Appendix). However, it is difficult to estimate *a priori* the number of terms needed to approximate $\mathbf{B}$ to a given accuracy. The relation

$$\|\mathbf{B} - \mathbf{B}_L\|_2 = \|\mathbf{A}^{L+1}\mathbf{B}\left(\mathbf{A}^T\right)^{L+1}\|_2 \leq \|\mathbf{A}^{L+1}\|_2^2\, \|\mathbf{B}\|_2 \tag{25}$$

shows that the number of terms in (24) needed to approximate $\mathbf{B}$ to a given accuracy depends on how rapidly $\|\mathbf{A}^L\|_2$ tends to zero. On the other hand, the size of $\|\mathbf{B} - \mathbf{B}_L\|_2$ does not necessarily determine how well the eigenvectors of the truncated series $\mathbf{B}_L$ approximate those of $\mathbf{B}$, for even when $\|\mathbf{B} - \mathbf{B}_L\|_2$ is not small, $\mathbf{B}$ and $\mathbf{B}_L$ may have approximately the same eigenvectors. For a normal system matrix, $\mathbf{B}$ and $\mathbf{B}_L$ have precisely the same eigenvectors, independently of $L \geq 1$; the error patterns established after a single time step with the steady-state system matrix $\mathbf{A}$ are not subsequently altered. When $\mathbf{A}$ is nonnormal the situation is less clear, though there may be situations where the eigenvectors of $\mathbf{B}_L$ approximate those of $\mathbf{B}$ for moderate values of L. We explore this point further in the numerical experiments of Section 5.2.

The simplest approximation for the bound matrix (and one which allows some analytical treatment) is to take only the first two terms of the series in (18):

$$\mathbf{B} \approx \mathbf{B}_1 = \mathbf{I} + \mathbf{A}\mathbf{A}^T . \tag{26}$$

The eigenvalues of $\mathbf{B}_1$ are $(1 + \sigma_i^2(\mathbf{A}))$ and the eigenvectors are the left singular vectors of $\mathbf{A}$; the singular values of $\mathbf{A}$ are defined by $\sigma_i^2(\mathbf{A}) = \lambda_i(\mathbf{A}\mathbf{A}^T)$. Since $\mathbf{B}_1 \leq \mathbf{B}$, from (22) one has the lower bound

$$\lambda_i(\mathbf{P}) \geq \lambda_n(\mathbf{Q})\lambda_i(\mathbf{B}_1) = \lambda_n(\mathbf{Q})(1 + \sigma_i^2(\mathbf{A}))\,, \tag{27}$$

and summing over i gives

$$\operatorname{tr}\mathbf{P} \geq \lambda_n(\mathbf{Q})(1 + \operatorname{tr}\mathbf{A}\mathbf{A}^T)\,. \tag{28}$$

These two estimates highlight the role of nonmodal growth ($\sigma_i(\mathbf{A}) > 1$) on the size of the eigenvalues of $\mathbf{P}$. On the other hand, if $\mathbf{A}$ is normal, or nearly normal in the sense that nonmodal growth does not exceed the modal growth ($\sigma_1(\mathbf{A}) = |\lambda_1(\mathbf{A})| < 1$), bounds for the eigenvalues of $\mathbf{P}$ are (Mori et al., 1982)

$$\lambda_n(\mathbf{Q})\frac{|\lambda_n(\mathbf{A})|^2}{1-|\lambda_n(\mathbf{A})|^2} + \lambda_i(\mathbf{Q}) \leq \lambda_i(\mathbf{P}) \leq \lambda_i(\mathbf{Q}) + \lambda_1(\mathbf{Q})\frac{|\lambda_1(\mathbf{A})|^2}{1-|\lambda_1(\mathbf{A})|^2}\,. \tag{29}$$

When $|\lambda_1(\mathbf{A})| \ll 1$, as might be expected for an accurate and dense observing network for instance, (29) suggests that the eigenvalues of the steady-state analysis error covariance $\mathbf{P}$ are determined more by the steady-state system error covariance $\mathbf{Q}$ than by the steady-state system matrix $\mathbf{A}$.

4. Generalized advection model

The analytical results presented above are illustrated in numerical experiments with two different dynamical models. We have chosen dynamics that are exponentially stable to demonstrate that even in the absence of modal instability, nonmodal growth is sufficient to cause the steady-state analysis error covariance to have dominant parts. The first dynamical model is a one-dimensional advection equation with a term added to cause nonmodal growth. Simple advection equations have been a valuable theoretical tool in data assimilation (e.g., Daley, 1992; Daley and Ménard, 1993; Cohn, 1997; Mitchell and Daley, 1997). The simplicity of the dynamics allows the eigenvectors and singular vectors to be computed analytically.

4.1. Model

The generalized one-dimensional advection equation considered is

$$\mu_t + a\mu_x + c'(x)\mu = 0\,, \tag{30}$$

with initial and periodic boundary conditions

$$\mu(x, t = 0) = \mu_0(x), \quad \mu(x = 0, t) = \mu(x = 1, t)\,, \tag{31}$$

respectively. The quantity a plays the role of a constant zonal velocity and $-c'(x)$ that of a zonally varying vertical velocity. When the vertical velocity is identically zero, $c'(x) \equiv 0$, the dynamics is normal. As we shall show, the larger the vertical velocity, the more nonnormal is the dynamics.

The physical basis for the model is the following. Considering an $xz-$plane representing the zonal and vertical directions, a two-dimensional nondivergent wind field can be written in terms of a stream function $\psi(x, z)$ as $-\psi_z\hat{\mathbf{x}} + \psi_x\hat{\mathbf{z}}$. Suppose that the stream function has the form $\psi(x, z) = -az - c(x)$; a is the constant zonal velocity; the vertical velocity is $-c'(x)$ where $'$ is used to indicate the derivative with respect to x. The streamlines of such a wind field are shown in Figure 1 for $c(x)$ taken to be

$$c(x) = \frac{c_0}{16}(1 + \cos 2\pi(x - 0.5))^4\,, \tag{32}$$

with $c_0 = 0.15$ and $a = 1/9$. These non-dimensional parameters correspond to a zonal length scale of 3105 km, a vertical length scale of 1 km, a zonal velocity of 15 m/s, a maximum vertical velocity of 10 cm/s and a time-scale of 6.4 hours. Suppose that a quantity $\tilde{\mu}$ is advected by such a wind field and that initially the z-dependence of $\tilde{\mu}$ has the form $\tilde{\mu}(x, z, t = 0) = \mu(x, t = 0)\exp(-z)$. Then the equation for the time evolution of μ is (30).

The operator $\mathbf{\Psi}_\tau$ that advances the solution τ time units is given by

$$\mathbf{\Psi}_\tau\{\mu(x, t)\} = \mu(x - a\tau, t)\exp\left[(c(x - a\tau) - c(x))/a\right]. \tag{33}$$

An "exact" discrete dynamics $\mathbf{\Psi}_\tau^d$ is obtained by restricting $\mathbf{\Psi}_\tau$ to act on functions defined on a spatially regular grid with spacing Δx such that the Courant number $a\tau/\Delta x$ is an integer. If the time-step τ is small compared to the advection time-scale $1/a$, the factor $\exp\left[(c(x - a\tau) - c(x))/a\right]$ can be approximated by $\exp\left(-c'(x)\tau\right)$ and in a single time-step the solution is seen to grow in regions where the vertical velocity is positive ($-c'(x) > 0$). This behavior is consistent with the physical interpretation of $\tilde{\mu}$ having vertical density stratification that decreases with height. This growth is not evident from the eigenvalues of $\mathbf{\Psi}_\tau$. The eigenvalues of $\mathbf{\Psi}_\tau$ are $\exp\left(-2\pi a\tau li\right)$ with corresponding eigenfunctions $\exp\left[-c(x)/a + 2\pi lxi\right]$, for any integer l. Regardless of the vertical velocity $-c'(x)$, all the eigenvalues of $\boldsymbol{\Psi}_\tau$ have unit modulus. Only in the case that $c'(x) \equiv 0$ and hence $\boldsymbol{\Psi}_\tau$ is normal, does this property actually imply that solutions neither decay nor grow in time.

4.2. Singular values and vectors

The effect of applying the operator $\boldsymbol{\Psi}_\tau$ once is explained by its singular values (or more generally s-numbers; Gohberg and Kreĭn, 1969). The singular values of $\boldsymbol{\Psi}_\tau$ are the square roots of the eigenvalues of $\boldsymbol{\Psi}_\tau\boldsymbol{\Psi}_\tau^*$ (or of $\boldsymbol{\Psi}_\tau^*\boldsymbol{\Psi}_\tau$). The adjoint of equation (30) with respect to the usual inner product on the space of square integrable functions is

$$\mu_t - a\mu_x + c'(x)\mu = 0\,. \tag{34}$$

Hence, the adjoint operator $\boldsymbol{\Psi}_\tau^*$ is

$$\boldsymbol{\Psi}_\tau^*\{\mu(x,t)\} = \mu(x + a\tau, t)\exp\left[-(c(x + a\tau) - c(x))/a\right]. \tag{35}$$

The eigenvalues of $\boldsymbol{\Psi}_\tau^*$ are $\exp(2\pi a\tau li)$ with corresponding eigenfunctions $\exp[c(x)/a + 2\pi lxi]$. Using the definitions in (33) and (35),

$$\boldsymbol{\Psi}_\tau\boldsymbol{\Psi}_\tau^*\{\mu(x)\} = \mu(x)\exp\left[2(c(x - a\tau) - c(x))/a\right]. \tag{36}$$

The operator $\boldsymbol{\Psi}_\tau\boldsymbol{\Psi}_\tau^*$ is a multiplication operator and its spectrum consists of the values taken on by the multiplying factor $\exp[2(c(x - a\tau) - c(x))/a])$ for $0 \le x \le 1$ (Halmos, 1967, Problem 52).*

Although the eigenvalues of the operator $\boldsymbol{\Psi}_\tau$ are necessarily the complex conjugates of those of its adjoint $\boldsymbol{\Psi}_\tau^*$, the eigenfunctions of $\boldsymbol{\Psi}_\tau$ and $\boldsymbol{\Psi}_\tau^*$ can be very different, leading to transient growth and limiting predictability (Farrell, 1989). If $c'(x) \equiv 0$, then $\boldsymbol{\Psi}_\tau$ is normal, and the eigenfunctions of $\boldsymbol{\Psi}_\tau$ and $\boldsymbol{\Psi}_\tau^*$ are the same. More generally the cosine S of the angle between the eigenfunctions

*In general, the spectrum of an operator can be divided into three disjoint sets: the point spectrum, the residual spectrum and the continuous spectrum. For a multiplication operator $\mathcal{A}$ induced by a multiplier ϕ: (i) the point spectrum is equal to the set of complex numbers for which $\phi^{-1}(\lambda)$ has positive measure; (ii) the residual spectrum is empty; and (iii) the continuous spectrum is the essential range of ϕ minus the point spectrum (Halmos, 1967, Problem 66). An eigenvalue λ is in the point spectrum of $\mathcal{A}$ if and only if there exists a function f such that $\mathcal{A}f = \lambda f$. For an eigenvalue λ in the continuous spectrum there is no eigenfunction but there is a sequence of approximate eigenfunctions f_j such that $\|\mathcal{A}f_j - \lambda f_j\| \to 0$. Therefore, for non-constant $c(x)$ the operator $\boldsymbol{\Psi}_\tau\boldsymbol{\Psi}_\tau^*$ has a continuous spectrum and no eigenfunctions but does have approximate eigenfunctions, for example those given by the eigenvectors of the discrete approximation $\boldsymbol{\Psi}_\tau^d\boldsymbol{\Psi}_\tau^{d*}$.

of $\mathbf{\Psi}_\tau$ and $\mathbf{\Psi}_\tau^*$ is given by

$$\frac{1}{S^2} = \int_0^1 \int_0^1 \exp\left(2(c(x) - c(y))/a\right) \, dxdy \,. \tag{37}$$

When the vertical velocity $-c'(x)$ is large, the cosine S will be small, indicating that the eigenfunctions of $\mathbf{\Psi}_\tau$ and $\mathbf{\Psi}_\tau^*$ are nearly orthogonal. Similarly, the smaller the zonal velocity a is, the smaller S is. The quantity S is also a measure of the sensitivity of the eigenvalues to perturbations of the operator $\mathbf{\Psi}_\tau$ (Golub and Van Loan, 1996, Section 7.2.2).

The singular values of the spatially discrete dynamics $\mathbf{\Psi}_\tau^d$ are obtained by restricting the values of x to be grid points. That is, according to Eq. (36), the n singular values σ_m of $\mathbf{\Psi}_\tau^d$ are (not ordered by size):

$$\sigma_m = \exp\left[(c(x_m - a\tau) - c(x_m))/a\right], \quad x_m = (m-1)/n\,, \quad m = 1, \ldots, n\,. \tag{38}$$

This expression shows that if $a\tau \ll 1$, the size of the largest singular value is an (exponentially) increasing function of the maximum vertical velocity and the time-step τ. In Fig. 2, $c(x)$ and $s(x) = \exp\left[(c(x - a\tau) - c(x))/a\right]$ are plotted for $c(x)$ given by (32) and $\tau = 0.5, 1.0, 2.0$. For larger values of τ the magnitude of the largest singular value increases and its localization moves upstream.

The $n \times n$ matrices $\mathbf{\Psi}_\tau^d \mathbf{\Psi}_\tau^{d*}$ and $\mathbf{\Psi}_\tau^{d*} \mathbf{\Psi}_\tau^d$ are diagonal, a reflection of the fact that $\mathbf{\Psi}_\tau \mathbf{\Psi}_\tau^*$ and $\mathbf{\Psi}_\tau^* \mathbf{\Psi}_\tau$ are multiplication operators; the squared singular values σ_m^2 appear along the diagonal. The diagonal entries of $\mathbf{\Psi}_\tau^{d*} \mathbf{\Psi}_\tau^d$ are those of $\mathbf{\Psi}_\tau^d \mathbf{\Psi}_\tau^{d*}$ shifted by $a\tau n$ units, the Courant number, as can be seen by comparing the expression for $\mathbf{\Psi}_\tau^* \mathbf{\Psi}_\tau$ with that of $\mathbf{\Psi}_\tau \mathbf{\Psi}_\tau^*$ given in (36). Thus the left singular vector of $\mathbf{\Psi}_\tau^d$ corresponding to σ_m is an n-vector with value 1 at the m-th position and zero elsewhere, and the right singular vector corresponding to the same singular value has 1 at the $(m - a\tau n)$ position and is elsewhere zero. When there are repeated singular values, the associated singular vectors of $\mathbf{\Psi}_\tau^d$ are not uniquely defined; only the subspace they span is uniquely defined. The localized structure of the singular vectors seen here is observed, to some extent, in more physically realistic models (e.g., Buizza et al., 1997). The left and right singular vectors of $\mathbf{\Psi}_\tau^d$ are different in general but for this problem it is only their ordering that is different.

4.3. Kalman filter experiments

Three sets of experiments were performed. In all experiments, the solution of the DALE is obtained by iterating the time-dependent Kalman filter to steady state and using the resulting steady-state gain $\mathbf{K}$ to compute the system matrix. In the first set of experiments we take $c'(x) \equiv 0$ and the dynamics $\mathbf{\Psi}_\tau^d$ is normal; an inhomogeneous observing pattern is the source of system matrix nonnormality (Daley, 1992). In the second and third sets of experiments nonnormal dynamics $\mathbf{\Psi}_\tau^d$ is considered and $c(x)$ is that given in (32). The second and third sets differ in their choices of observing pattern. In the second set of experiments, the nine *trailing* left singular vectors of the dynamics are observed, leading to a system matrix with large singular values. In the third set of experiments, the nine *leading* left singular vectors of the dynamics are observed. This is equivalent to observing upstream (experiment 2) and in the middle (experiment 3) of the instability (Todling and Ghil, 1996). Observations are assimilated at unit time intervals, $\tau = 1$; the τ subscript and the d superscript on $\mathbf{\Psi}_\tau^d$ will be dropped from the notation. A grid spacing of $\Delta x = 1/36$ is used, so that $n = 36$. In all cases, the observations are made at grid points and the number of observations $p = 9$ is one fourth of the total number of grid points.

The observation error is taken to be spatially homogeneous and uncorrelated, $\mathbf{R} = 0.025\mathbf{I}_p$ for all experiments; $\mathbf{I}_p$ is the $p \times p$ identity matrix. We consider two types of model error covariances: one with $\mathbf{Q}^m = 0.025\mathbf{I}$, for which errors at different grid points are uncorrelated, and another whose entries are constructed as

$$\mathbf{Q}_{ij}^m = \delta_{ij} + \rho(d_{ij}; L)\,, \tag{39}$$

where d_{ij} is the distance between the i-th and j-th grid points and ρ is the compactly supported piecewise-rational correlation function with correlation length $L = 0.2$ (Gaspari and Cohn, 1999, Eq. 4.10). The trace of $\mathbf{Q}^m$ is normalized to $0.025n$. A 36×36 “checkerboard plot” of the elements of the $\mathbf{Q}^m$ given by Eq. (39) is shown in Fig. 3.

In the first experiment $c'(x) \equiv 0$ and the dynamics is normal. Observations are taken at the first 9 grid points. Figure 4a shows the singular values and the absolute values of the eigenvalues of the steady-state system matrix $\mathbf{A}$ for $\mathbf{Q}^m = 0.025\mathbf{I}$. Though $\mathbf{\Psi}$ is normal, the inhomogeneous observing pattern makes $\mathbf{A}$ nonnormal; the difference between the singular values and magnitudes of the eigenvalues is one indication of nonnormality. The singular values of the system matrix control how analysis errors are modified by one forecast/analysis cycle in the steady-state regime. Figure 4a shows that there is a nine-dimensional subspace where analysis errors are reduced after a single forecast/analysis cycle; the magnitude of analysis errors outside of this subspace remain unchanged. The eigenvalues of $\mathbf{A}$ are all less than one in modulus, demonstrating that asymptotically all analysis errors are reduced.

The spectra of the observation and model error covariance matrices $\mathbf{R} = 0.025\mathbf{I}_p$ and $\mathbf{Q}^m = 0.025\mathbf{I}$ are flat by construction. However, the spectrum of $\mathbf{P}$ shown in Fig. 4b varies by an order of magnitude between its largest and smallest eigenvalues (note the logarithmic scale on panel b). This feature is similar to that seen in the example of Tippett and Marchesin (1999*b*) where for the same dynamics, assimilation of a single observation produced a steady-state analysis error covariance whose largest and smallest eigenvalues differed by a factor equal to the ratio of the domain size to the advection speed; here, the ratio of the domain size to the advection speed is 9. The clustering of the eigenvalues of $\mathbf{P}$ in groups of four is the result of the spatial homogeneity of the model and observational errors and of the advection speed, 4 grid-points/unit time. Also shown in Fig. 4b are the upper and lower estimates for the eigenvalues of $\mathbf{P}$ obtained from (22). The corresponding results (not shown) for the experiment with spatially correlated model error are qualitatively similar to the spatially uncorrelated model error case though the bounds from (22) are not tight because $\lambda_n(\mathbf{Q}) \ll \lambda_1(\mathbf{Q})$ for the spatially correlated case.

The magnitude of the eigenvalues and singular values of $\mathbf{\Psi}$ for the nonnormal case are shown in Fig. 5. The singular values of $\mathbf{\Psi}$ are the values of the function plotted (dashed line) in Fig. 2, evaluated at the grid points. As was shown analytically, though the eigenvalues of the dynamics all have magnitude one, the presence of singular values greater than unity allows transient growth to occur. To examine the effect of not observing the growing modes of the dynamics, first only the trailing 9 left singular vectors of $\mathbf{\Psi}$ were observed. We take $\mathbf{H} = [\mathbf{u}_{n-p+1}, \dots, \mathbf{u}_n]^T$, where $\mathbf{u}_m$ is the m-th left singular vector of $\mathbf{\Psi}$ and $p = 9$. Figure 2 shows that this choice of $\mathbf{H}$ corresponds to observing to the left or upstream of the instability. Figure 6a shows that as in the case with normal dynamics, the analysis errors are reduced in a nine-dimensional subspace by a single forecast/analysis cycle. Since the observations were chosen to lie in the region where errors are not amplified by the dynamics, the smallest singular values of $\mathbf{A}$ shown in Fig. 6a are smaller than those in the normal case, shown in Fig. 4a. Not observing the growing modes of the dynamics leads to $\mathbf{A}$, like $\boldsymbol{\Psi}$, having 13 singular values greater than unity, indicating that there is a thirteen-dimensional subspace associated with transiently growing analysis errors. The asymptotic rate at which analysis errors are reduced, as given by the eigenvalues of $\mathbf{A}$ (crosses in Fig. 6a), is roughly the same as in the normal case (crosses in Fig. 4a). Figure 6b shows that not observing growing modes of the dynamics leads to $\mathbf{P}$ having eigenvalues that are larger than in the normal case (compare Fig. 4b). Similar results (not shown) were found for the spatially correlated model error case.

Figure 7a shows that with the same nonnormal dynamics, if the 9 *leading* left singular vectors of the dynamics are observed ($\mathbf{H} = [\mathbf{u}_1, \ldots, \mathbf{u}_p]^T$), the singular values of $\mathbf{A}$ are reduced and most of the growing modes of $\mathbf{A}$ are removed. This $\mathbf{H}$ corresponds to observing directly in the middle of the instability. The result of this change in observing pattern is a reduction of the analysis error to a level comparable to that of the normal case (compare Fig. 7b and Fig. 4b). This reduction in size of the eigenvalues of $\mathbf{P}$ is *suggested* by the inequality in (27). However, as the relation in (27) is a *lower* bound on the eigenvalues of $\mathbf{P}$, reducing the singular values of the steady-state system matrix does not necessarily reduce the asymptotic analysis error levels. If the analysis error levels were strictly controlled by the singular values of the system matrix, then the optimal observing strategy would be simply to minimize the singular values of the system matrix. We return to this point in Section 5 and in the Conclusions.

We have shown how the eigenvalues of the bound matrix $\mathbf{B}$ are related to those of the steady-state analysis error covariance $\mathbf{P}$. Now we examine how the eigenvectors of $\mathbf{B}$ project onto those of $\mathbf{P}$. For the three pairs of experiments described, the invariant subspaces (subspaces spanned by the eigenvectors) of $\mathbf{P}$ and $\mathbf{B}$ were compared in the following manner: let $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_n]$ and $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n]$ be the $n \times n$ matrices whose columns are the eigenvectors, in order of decreasing eigenvalue, of $\mathbf{P}$ and $\mathbf{B}$ respectively. Consider the matrix $\mathbf{M}$ defined to be the product $\mathbf{M} = \mathbf{Z}^T\mathbf{W}$. The entries of $\mathbf{M}$ are $\mathbf{M}_{ij} = \mathbf{z}_i^T\mathbf{w}_j$. That is, the i-th row of $\mathbf{M}$ contains the components of the i-th eigenvector of $\mathbf{B}$ written in the basis given by the eigenvectors of $\mathbf{P}$. Gelaro et al. (1998) used a similar approach, referring to $\mathbf{M}$ as a projection matrix, to compare the subspaces spanned by initial and final singular vectors. If $\mathbf{P}$ and $\mathbf{B}$ have the same eigenvectors with the same ordering, then $\mathbf{M} = \mathbf{I}$; if only the ordering is different then $\mathbf{M}$ is the identity matrix with some columns permuted. Since $\mathbf{W}$ and $\mathbf{Z}$ are orthogonal matrices, the entries of $\mathbf{M}$ have absolute value between zero and one.

The structure of $\mathbf{M}$ for the three cases considered is shown in Figs. 8 – 10. In these "checker-board" plots of the elements of $\mathbf{M}$, the darkness of the ij-th square is proportional to the absolute value of $\mathbf{M}_{ij}$. The darkest squares indicate vectors that are parallel; white squares indicate vectors that are orthogonal. The image on the left is for $\mathbf{Q}^m = 0.025\mathbf{I}$ and that on the right for the spatially correlated $\mathbf{Q}^m$. Figs. 8 – 10 (left panels) show that in all the cases when the model error covariance is proportional to the identity matrix, $\mathbf{P}$ and $\mathbf{B}$ have the same eigenvectors; only the ordering is different. Additionally in all the cases the difference in ordering is not very great. When the spatially correlated $\mathbf{Q}^m$ is used, the eigenvectors of $\mathbf{P}$ and $\mathbf{B}$ do not coincide (Figs. 8 – 10, right panels). However, some diagonal dominance is still seen in the plots of $\mathbf{M}$, indicating that the leading invariant subspace of $\mathbf{B}$ is still a good approximation of that of $\mathbf{P}$; that is, the leading eigenvectors of $\mathbf{P}$ project mostly onto the leading eigenvectors of $\mathbf{B}$.

5. Damped baroclinic wave model

The next illustration uses the dynamics of a model of damped baroclinic waves (Farrell, 1989). Similar baroclinic models have been studied extensively in the context of optimal growth and predictability (Farrell, 1985; Farrell and Ioannou, 1993).

5.1. Model

Using the same scaling parameters as in Farrell (1989), the non-dimensionalized potential vorticity equation based on the non-dimensional stream function

$$\Phi(x, y, z, t) = \psi(z, t)e^{z/2}e^{ikx}\cos(ly) \tag{40}$$

is

$$\left(\frac{\partial}{\partial t} + ikz\right)\left[\psi_{zz} - \left(\alpha^2 + \frac{1}{4}\right)\psi\right] + ik(\beta + 1)\psi = 0\,, \tag{41}$$

with boundary conditions

$$\frac{\partial}{\partial t}\left(\psi_z + \frac{\psi}{2}\right) - ik(1 - \Gamma)\psi = 0\,, \quad z = 0\,, \tag{42}$$

$$\left(\frac{\partial}{\partial t} - ikz\right)\left(\psi_z + \frac{\psi}{2}\right) - ik\psi = 0, \quad z = z_T\,, \tag{43}$$

where $\alpha = \sqrt{k^2 + l^2}$ and Γ is the non-dimensionalized Ekman damping parameter taken here to be $0.2i$, corresponding to a vertical eddy viscosity of 4.5m^2/s; $\beta = 0.533$ and a unit time interval is equivalent to 9.3 hours. The $k = l = 2$ wave with meridional wavelength 3100 km is considered. The upper boundary condition at $z_T = 4.0$ corresponds to a height of 40 km.

Discretizing equation (41) and the boundary conditions (42) and (43) using centered differences on a regular grid in the vertical coordinate z gives a system of equations of the form

$$\frac{\partial \psi_j}{\partial t} = \sum_{m=1}^{n} \mathbf{F}_{jm}\psi_m\,, \tag{44}$$

where $\psi_j(t) = \psi(z_j, t)$, z_j is the j-th level, and n is the number of levels; details of $\mathbf{F}_{jm}$ can be found in Farrell (1989). The matrix $\mathbf{\Psi}_\tau$ that advances the solution by time τ is $\mathbf{\Psi}_\tau = \exp(\tau\mathbf{F})$. Figure 11 shows the singular values and eigenvalues of $\mathbf{\Psi}_\tau$ with the number of levels $n = 36$, for $\tau = 0.6452$ corresponding to a time step of 6 hours, which is used here. The qualitative features of the singular values and eigenvalues of the baroclinic dynamics are similar to those of the generalized advection equation shown in Fig. 5, with eigenvalues that have approximately unit magnitude and several singular values greater than unity, indicating transient, nonmodal growth. The perturbation that leads to maximum initial growth was calculated in Farrell (1989) and has most of its structure concentrated near the surface. In the absence of Ekman damping, the baroclinic wave dynamics has one exponentially increasing and one exponentially decaying eigenmode in addition to the neutral eigenmodes. The vertical dissipation used here is sufficient to damp the exponentially growing mode (Farrell, 1989). However, an effect of the relatively low vertical resolution used here is that the exponential growth rate is quite small but nonzero.

5.2. Kalman filter experiments

The steady-state system matrix $\mathbf{A}$ and analysis error covariance $\mathbf{P}$ are calculated as in Section 4.3. The observation error is taken to be uncorrelated, with $\mathbf{R} = 0.1\mathbf{I}_p$; the forecast model error given in (39) is used with $L = 5$ km and the normalization $\operatorname{tr}\mathbf{Q}^m/n = 1$. To give an indication of the effect of changing the observing pattern, the squared Frobenius norm of $\mathbf{A}^k$, $\|\mathbf{A}^k\|_F^2 \equiv \operatorname{tr}\mathbf{A}^k\left(\mathbf{A}^T\right)^k$, is plotted as a function of k in Fig. 12 for three different observing patterns. This quantity can be interpreted as the expected value $\langle \mathbf{y}_k^T \mathbf{y}_k \rangle$ of the response of the unforced system $\mathbf{y}_{k+1} = \mathbf{A}\mathbf{y}_k$ to a spatially uncorrelated random initial condition $\mathbf{y}_0$ with $\left\langle \mathbf{y}_0 \mathbf{y}_0^T \right\rangle = \mathbf{I}$; that is, $\left\langle \mathbf{y}_k^T \mathbf{y}_k \right\rangle = ||\mathbf{A}^k||_F^2$, and initially $||\mathbf{A}^0||_F^2 = ||\mathbf{I}||_F^2 = n = 36$. This unforced, time-independent system, like the bound matrix, can be used as a tool to study the steady state of the time-dependent system (6). The three observing patterns are (1) observations of ψ at every fourth vertical level; (2) observations of the leading 6 right singular vectors of the dynamics; and (3) observations of the leading 6 left singular vectors of the dynamics. Both the leading left and the leading right singular vectors of the dynamics have most of their structure localized near the surface. Figure 12 shows that with observing pattern 1, $||\mathbf{A}^k||_F^2$ has a large initial transient in comparison to that of observing patterns 2 and 3. This behavior is a reminder that the asymptotic gain $\mathbf{K}$, while optimal in the asymptotic steady state, can give far from optimal performance in the unsteady regime. After the initial peak, $||\mathbf{A}^k||_F^2$ decays faster for pattern 1 than for patterns 2 and 3. The size of the steady-state analysis error variance depends on the behavior of the steady-state system matrix on all time-scales. In particular, from (18),

$$\mathrm{tr}\,\mathbf{B} = \sum_{k=0}^{\infty} ||\mathbf{A}^k||_F^2 \,, \tag{45}$$

and therefore the trace of the bound matrix is proportional to the area under the graph of $||\mathbf{A}^k||_F^2$ plotted as a function of k. Although the temporal behavior of the steady-state system matrices associated with the three observing patterns is different, the areas under the graphs in Fig. 12 are not so different, and thus the steady-state values of the total analysis error variance are comparable; for pattern 1, $\mathrm{tr}\,\mathbf{P} = 236.4$, for pattern 2, $\mathrm{tr}\,\mathbf{P} = 251.7$, and for pattern 3, $\mathrm{tr}\,\mathbf{P} = 309.2$.

The remainder of the results shown use observing pattern 1. Figure 13a shows two properties of the steady-state system matrix $\mathbf{A}$ that lead to large asymptotic analysis error levels: (i) the system matrix has eigenvalues inside but near the unit circle; and (ii) it has many singular values greater than unity. Note that the largest singular values of the system matrix are larger than those of the dynamics, another indication of the suboptimality of the asymptotic gain outside of the steady-state regime; we show later that these growing modes of the system matrix are “unlikely” in the steady state since they project mostly onto the trailing eigenvectors of the steady-state analysis error covariance. Figure 13b shows that the steady-state analysis error covariance has large eigenvalues with the bounds from (22) predicting well the shape but not the values of the analysis error covariance spectrum.

We examine the eigenvectors of the steady-state analysis error covariance in the same manner used in Section 4.3. Figure 14 shows checkerboard plots of the elements of $\mathbf{M} = \mathbf{Z}^*\mathbf{W}$ for $\mathbf{W}$ taken to be the eigenvectors of $\mathbf{P}$, and for $\mathbf{Z}$ taken to be (a) the eigenvectors of $\mathbf{B}$, (b) the eigenvectors of $\mathbf{B}_1$ (left singular vectors of $\mathbf{A}$) and (c) the right singular vectors of $\mathbf{A}$. Figure 14(a) shows that the leading eigenvectors of the bound matrix $\mathbf{B}$ are nearly the same as those of the asymptotic analysis error covariance matrix. Figure 14(b) shows that the leading invariant subspace of $\mathbf{B}_1$ (leading left singular vectors of $\mathbf{A}$) is also a good approximation for that of the asymptotic analysis error covariance matrix. The dark squares in the lower left-hand corner of Fig. 14(c) show that the *trailing* right singular vectors of the steady-state system matrix are a good approximation for the leading invariant subspace of the asymptotic analysis error covariance matrix. Additionally, the leading right singular vectors of the system matrix project strongly onto the trailing eigenvectors of $\mathbf{P}$ (see upper right-hand corner of Fig. 14(c)). The leading right and left singular vectors of the steady-state system matrix $\mathbf{A}$ are therefore approximately orthogonal since they approximate the trailing and leading eigenvectors, respectively, of $\mathbf{P}$, and since the leading and trailing eigenvectors of $\mathbf{P}$ are orthogonal by construction.

Similar relations between right and left singular vectors of *dynamics* have been noted elsewhere. For the discrete dynamics of Sec. 4, we saw that the left and right singular vector corresponding to any given singular value are orthogonal, being different columns of the identity matrix. Gelaro et al. (1998) observed near-orthogonality of leading right and left singular vectors of the dynamics in a more realistic model. An explanation for this similar behavior of the steady-state system matrix is that since the system matrix is stable, its leading left singular vectors must project mostly onto its trailing right singular vectors, which are orthogonal to its leading right singular vectors by construction, so that growth is not maintained.

We stated earlier that the eigenvectors of $\mathbf{B}$, by ordering state space according to accumulated growth, play a role analogous to that of Lyapunov vectors. That analogy is supported by the relations observed here among the eigenvectors of $\mathbf{B}$, the singular vectors of the steady-state system matrix, and the eigenvectors of the steady-state analysis error covariance matrix. Szunyogh et al. (1997) noted that generally leading right singular vectors of the dynamics are quite different from leading Lyapunov vectors, while leading left singular vectors resemble leading Lyapunov vectors.

Also, Swanson et al. (1998), observing all state variables, and using perfect-model, 4D-Var data assimilation, found that the dominant analysis errors were well described by the Lyapunov vectors.

6. Summary and Conclusions

The properties of a general linear, time-independent forecast/analysis system were investigated. The steady-state analysis error covariance matrix of such a system satisfies a discrete algebraic Lyapunov equation (DALE). Using properties of the DALE we examined why the steady-state analysis error covariance might have a dominant part and how that dominant part can be approximated. Nonnormality of the system matrix is a key feature; the system matrix is the linear operator that combines the forecast and analysis steps. If the system matrix is normal, the properties of the DALE and hence of the steady-state analysis error covariance are trivially related to the eigenvalues and eigenvectors of the system matrix. The role of observation and forecast model errors is also important since their presence prevents repeated iterations of the forecast/analysis cycle from relaxing to modal behavior. However, the view taken here has been to assume that the impact of the system matrix on the steady-state analysis error covariance is greater than that of the observation and forecast model error forcing. This choice is a practical one since the forecast model error is poorly known.

The bound matrix provides a quantitative connection between the structure of the steady-state analysis error covariance and the nonmodal stability properties of the system matrix. The eigenvalues of the bound matrix give estimates for the eigenvalues of the steady-state analysis error covariance matrix, and in some circumstances the leading eigenvectors of the bound matrix approximate those of the steady-state analysis error covariance matrix. This bound matrix estimate approach extends naturally to the time-dependent problem. The eigenvectors of the bound matrix, identifying the directions in state space associated with accumulated past growth, play a role similar to that of Lyapunov vectors in deterministic systems. Methods of computing the leading singular values and vectors of the system matrix might be adapted to compute the leading eigenvectors of an approximate bound matrix. In the simplest approximation, the leading eigenvectors of the bound matrix are approximated by the leading left singular vectors of the system matrix, and the leading eigenvalues of the bound matrix depend on the squares of the singular values of the system matrix. The singular values and vectors of the system matrix are in general different from those of the dynamics. If growing modes of the dynamics, in particular leading left singular vectors of the dynamics, are not observed, then the analysis step will not be able to reduce the associated errors in the first guess. In that case, growing modes of the system matrix are related to those of the dynamics.

Failure to observe the leading left singular vectors of the dynamics was seen to lead to a system matrix with large singular values and elevated steady-state analysis error variances in a numerical example using a generalized advection equation. This result has implications for targeted observing strategies. However, there are several caveats. The first is the simplicity of the dynamics, in particular, its time-independence. Second, in this work the focus has been on the analysis error covariance, whereas investigations of observing strategies as well as initial ensemble selection typically have used forecast skill as the measure of optimality (Gelaro et al., 1998). The choice of whether to observe the modes that grew (left singular vectors) or the modes that will grow (right singular vectors) depends on whether forecast or analysis accuracy is the optimality criterion. If analysis accuracy is desired, then observations of the leading left singular vectors of the dynamics are generally more important than observations of the leading right singular vectors of the dynamics. The reason for this conclusion is that if observations of the leading right singular vectors of the dynamics are used, the size of the forecast error that projects onto the leading left singular vectors of the dynamics depends on the observation error *amplified* by the forecast dynamics. The nonnormality of the dynamics implies that the subspaces spanned by the leading right and left singular vectors may be nearly orthogonal. In that case, with observations of only the leading right singular vectors, the analysis is unable to reduce the first guess error that projects onto the leading left singular vectors of the dynamics. On the other hand, if observations of the left singular vectors are used, the first guess error projecting onto the leading left singular vectors of the dynamics is reduced by the observations, and the size of the resulting analysis error depends on the observation error *unamplified* by the forecast dynamics. Of course, applying the same arguments to the steady-state forecast error covariance leads to the conclusion that observations of the leading right singular vectors of the dynamics reduce singular values of the forecast error propagation system matrix and consequently forecast error.

Since the dependence of the steady-state analysis error covariance on the singular values of the system matrix presented in this work is in the form of a lower bound, increasing the singular values of the system matrix guarantees that the steady-state analysis error variance will be increased, but reducing them does not guarantee reduction of the variance. That is to say, observing the leading left singular vectors of the dynamics may not be the *optimal* strategy since the steady-state analysis error covariance does not depend on a single application of the system matrix. More generally, the dependence of the bound matrix on the system matrix should be used to examine the role of the observing pattern in determining steady-state analysis error levels. It must be remembered that the observing pattern also plays a role in the system error covariance matrix, which includes terms for both the model and the observation error covariances. Failure to observe the leading eigenvectors of the model error covariance leads to large system error covariance eigenvalues and consequently to high analysis error levels. In a similar fashion, the dependence of the bound matrix on the dynamics gives a means of investigating covariance evolution schemes based on low-dimensional representation of the dynamics.

In summary we find that:

- Nonnormality of the forecast/analysis system is a key factor in determining the properties of the steady-state analysis error covariance.
- Theoretical estimates based on the bound matrix show the dependence of the steady-state analysis error covariance on the dynamics and observing pattern through the system matrix.
- In the simplest approximation, the dominant eigenvectors of the bound matrix are approximated by the *left* singular vectors of the system matrix; when growing modes of the dynamics are not observed, they are related to the *left* singular vectors of the dynamics.

This last point seems to have implications in the context of approximating error covariances via ensemble methods (e.g., Evensen, 1994; Houtekamer et al., 1996), at least for linear systems. The accuracy of an ensemble method depends on capturing growing modes of the system *and* on representing the analysis uncertainty well. The work here suggests that it may not always be possible to satisfy these criteria simultaneously. Choosing perturbations that correspond to leading right singular vectors, while capturing the growing modes, would not necessarily result in an ensemble whose sample covariance approximates well the analysis error covariance. In fact, in the model for damped baroclinic waves, the leading eigenvectors of the steady-state analysis error covariance projected strongly onto the *trailing* right singular vectors of the system matrix. Conversely, the sample forecast error covariance resulting from a small initial ensemble that approximates well the analysis uncertainty may not explain well forecast errors due to growing modes. Possible solutions are to project analysis errors on to the growing modes or to calculate the growing modes weighted by the analysis error (Barkmeijer et al., 1998; Swanson et al., 1998).

7. Acknowledgments

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) Grants 91.0029/95-4, 381737/97-7 and 30.0204/83-3, Financiadora de Estudos e Projetos (FINEP) Grant 77.97.0315.00, and the NASA EOS Interdisciplinary Project on Data Assimilation. Thanks are extended to the anonymous reviewers whose careful reading, constructive comments and insightful criticism greatly improved the readability and presentation of this work. We also thank Greg Gaspari for his attentive reading and useful comments.

Appendix

Define $\mathbf{P}_k \equiv \left\langle \mathbf{y}_k \mathbf{y}_k^T \right\rangle$ where $\mathbf{y}_k$ is defined by (6), and suppose that (6) is in steady state: $\mathbf{A}_k = \mathbf{A}$ and $\mathbf{Q}_k = \mathbf{Q}$. Then

$$\begin{aligned} \mathbf{P}_k &= \left\langle (\mathbf{A}\mathbf{y}_{k-1} + \mathbf{b}_k)(\mathbf{A}\mathbf{y}_{k-1} + \mathbf{b}_k)^T \right\rangle \\ &= \mathbf{A} \left\langle \mathbf{y}_{k-1}\mathbf{y}_{k-1}^T \right\rangle \mathbf{A}^T + \mathbf{Q} + \mathbf{A} \left\langle \mathbf{y}_{k-1}\mathbf{b}_k^T \right\rangle + \left\langle \mathbf{b}_k \mathbf{y}_{k-1}^T \right\rangle \mathbf{A}^T . \end{aligned} \tag{46}$$

But $\left\langle \mathbf{y}_{k-1}\mathbf{b}_k^T \right\rangle = 0$ since

$$\mathbf{y}_{k-1} = \mathbf{A}^{k-1}\mathbf{y}_0 + \sum_{j=1}^{k-1} \mathbf{A}^{k-j-1}\mathbf{b}_j , \tag{47}$$

$\left\langle \mathbf{y}_0 \mathbf{b}_k^T \right\rangle = 0$, and $\left\langle \mathbf{b}_j \mathbf{b}_k^T \right\rangle = 0$ for $j \neq k$. Therefore, $\mathbf{P}_k = \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^T + \mathbf{Q}$, or

$$\mathbf{P}_k = \mathbf{A}^k \mathbf{P}_0 \left(\mathbf{A}^T\right)^k + \sum_{j=0}^{k-1} \mathbf{A}^j \mathbf{Q} \left(\mathbf{A}^T\right)^j . \tag{48}$$

Since the eigenvalues of $\mathbf{A}$ lie inside the unit circle, the first term in (48) tends to zero, and the terms of the series in (48) decay geometrically and the series converges uniformly (Golub and Van Loan, 1996, Lemma 7.3.2). Therefore, the limit

$$\mathbf{P} = \lim_{k\to\infty} \mathbf{P}_k \tag{49}$$

exists and is the solution of $\mathsf{P} = \mathsf{APA}^T + \mathsf{Q}$.

The bound matrix approach can be extended to the time-dependent case as follows. Again define $\mathsf{P}_k \equiv \left\langle \mathbf{y}_k \mathbf{y}_k^T \right\rangle$ where $\mathbf{y}_k$ is defined by (6). Then using the arguments above,

$$\mathsf{P}_k = \mathsf{A}_k \mathsf{P}_{k-1} \mathsf{A}_k^T + \mathsf{Q}_k \,, \tag{50}$$

and we can define a time-dependent bound matrix B_k by the iteration

$$\mathsf{B}_k = \mathsf{A}_k \mathsf{B}_{k-1} \mathsf{A}_k^T + \mathsf{I} \,, \tag{51}$$

with $\mathsf{B}_0 = \mathsf{I}$. Subtracting (50) from (51) multiplied by a scalar γ_k gives

$$(\gamma_k \mathsf{B}_k - \mathsf{P}_k) = \mathsf{A}_k (\gamma_k \mathsf{B}_{k-1} - \mathsf{P}_{k-1}) \mathsf{A}_k^T + (\gamma_k \mathsf{I} - \mathsf{Q}_k) \,. \tag{52}$$

Define the nonnegative constants γ_k^+ and γ_k^- by

$$\gamma_k^+ = \max(\lambda_1(\mathsf{P}_0), \max_{1 \le j \le k} \lambda_1(\mathsf{Q}_j)) \,, \text{ and } \gamma_k^- = \min(\lambda_n(\mathsf{P}_0), \min_{1 \le j \le k} \lambda_n(\mathsf{Q}_j)) \,. \tag{53}$$

Since $(\gamma_k^- \mathsf{I} - \mathsf{Q}_k)$ is negative semidefinite for all k, it follows from (52) that $(\gamma_k^- \mathsf{B}_k - \mathsf{P}_k)$ is negative semidefinite for all k. Likewise it follows that $(\gamma_k^+ \mathsf{B}_k - \mathsf{P}_k)$ is positive semidefinite for all k. In other words,

$$\gamma_k^- \mathsf{B}_k \le \mathsf{P}_k \le \gamma_k^+ \mathsf{B}_k \,, \tag{54}$$

with the relation in (21) following as a special case.

If there is some information available about the system error covariance, it can be used in the definition of the bound matrix. Suppose we have a time-dependent model $\tilde{\mathsf{Q}}_k$ for the system error covariance Q_k and time-independent constants γ^- and γ^+ such that $\gamma^- \tilde{\mathsf{Q}}_k \le \mathsf{Q}_k \le \gamma^+ \tilde{\mathsf{Q}}_k$. Then defining a bound matrix $\tilde{\mathsf{B}}_k$ by

$$\tilde{\mathsf{B}}_k = \mathsf{A}_k \tilde{\mathsf{B}}_{k-1} \mathsf{A}_k^T + \tilde{\mathsf{Q}}_k \,, \tag{55}$$

it follows that

$$\gamma^- \tilde{\mathsf{B}}_k \le \mathsf{P}_k \le \gamma^+ \tilde{\mathsf{B}}_k \,. \tag{56}$$

If the system error covariance Q_k is known at least approximately, Monte Carlo methods can be used to calculate P_k by approximating the expectation with an ensemble average (Evensen, 1994).

Such an approach is similar to that of system simulation (Houtekamer et al., 1996) where an attempt is made to account for all sources of error; it is also conceptually similar to the "breeding of growing modes" (BGM) method (Toth and Kalnay, 1993). In the BGM method the factor $(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)$ appearing in the system matrix is approximated by a constant or by a "regional rescaling" factor, and there is no explicit stochastic forcing.

In the steady-state problem, the dependence of the properties of $\mathbf{B}$ on the nonnormality of the system matrix highlights the role played by nonnormality in stochastically forced systems. If the forcing term is not included, the iteration in (6) with $\mathbf{A}_k = \mathbf{A}$ is the power method without renormalization (Golub and Van Loan, 1996, Section 7.3.1). For this reason, it has been stated that for linear, stationary dynamics the BGM method is similar to the power method and that in this case, the leading BGM vector is the leading eigenvector of the dynamics (Buizza and Palmer, 1995). If the system matrix $\mathbf{A}$ is normal, this conclusion is correct even in the presence of stochastic forcing, since in that case, the leading eigenvector of $\mathbf{B}$ is identical to the leading eigenvector of $\mathbf{A}$. However, when the system matrix is nonnormal, the eigenvectors of the bound matrix $\mathbf{B}$ are in general not eigenvectors of $\mathbf{A}$. Another reason not to interpret BGM as the power method is that the power method may be excessively slow to converge for nonnormal matrices (Golub and Van Loan, 1996, Lemma 7.3.1).

The eigenvectors of the bound matrix $\mathbf{B}$ do not correspond to Lyapunov vectors of the unforced system. For time-independent dynamics the method for computing Lyapunov vectors (Legras and Vautard, 1996, Section 5) reduces to orthogonal iteration, a standard method for calculating the subspace spanned by the leading eigenvectors (Golub and Van Loan, 1996, Section 7.3.2); the Lyapunov vectors themselves are not eigenvectors but an orthogonal basis for the invariant subspaces of $\mathbf{A}$ (for an alternative Lyapunov vector definition see Trevisan and Pancotti, 1998). However, conceptually the eigenvectors of $\mathbf{B}$ do play the role of Lyapunov vectors of the system with spatially uncorrelated, homogeneous random forcing, since they order the directions of state space according to accumulated growth.

References

Barkmeijer, J., Gijzen, M. V. and Bouttier, F. 1998. Singular vectors and estimates of the analysis-error covariance matrix. *Q. J. R. Meteorol. Soc.* **124**, 1695–1713.

Blumenthal, M. B. 1991. Predictability of a coupled ocean-atmosphere model. *J. Clim.* **4**, 766–784.

Buizza, R. and Palmer, T. N. 1995. The singular vector structure of the atmospheric global circulation. *J. Atmos. Sci.* **52**, 1434–1472.

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

Cane, M. A., Kaplan, A., Miller, R., Tang, B., Hackert, E. and Busalacchi, A. J. 1996. Mapping tropical Pacific sea level: Data assimilation via a reduced state space Kalman filter. *J. Geophys. Res.* **101**, 22,599–22,617.

Cohn, S. E. 1997. An introduction to estimation theory. *J. Meteor. Soc. Japan* **75**, 257–288.

Cohn, S. E. and Dee, D. P. 1988. Observability of discretized partial differential equations. *SIAM J. Numerical Analysis* **25**, 586–617.

Cohn, S. E. and Todling, R. 1996. Approximate data assimilation schemes for stable and unstable dynamics. *J. Meteor. Soc. Japan* **74**, 63–75.

Cohn, S. E., da Silva, A. M., Guo, J., Sienkiewicz, M. and Lamich, D. 1998. Assessing the effects of data selection with the DAO physical-space statistical analysis system. *Mon. Wea. Rev* **126**, 2913–2926.

Courtier, P. and Talagrand, O. 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. II: Numerical results. *Quart. J. Roy. Meteor. Soc.* **113**, 1329–1347.

Courtier, P., Andersson, E., Heckley, W., J, P., Vasiljevic, D., Hamrud, M., Hollingsworth, A., Rabier, E. and Fisher, M. 1998. The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Q. J. R. Meteorol. Soc.* **124**, 1783–1807.

Daley, R. 1992. Forecast-error statistics for homogeneous and inhomogeneous observation networks. *Mon. Wea. Rev.* **120**, 627–643.

Daley, R. and Ménard, R. 1993. Spectral characteristics of Kalman filter systems for atmospheric data assimilation. *Mon. Wea. Rev.* **121**, 1554–1565.

Dee, D. P. and da Silva, A. M. 1998. Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc.* **124**, 269–297.

Dee, D. P. and da Silva, A. M. 1999. Maximum-likelihood estimation of forecast and observation error covariance parameters. Part I: Methodology. *Mon. Wea. Rev.* **127**, 1822–1834.

Dee, D. P., , Gaspari, G., Redder, C., Rukhovets, L. and da Silva, A. M. 1999. Maximum-likelihood estimation of forecast and observation error covariance parameters. Part II: Aplications. *Mon. Wea. Rev.* **127**, 1835–1849.

Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**, 1043–1062.

Farrell, B. F. 1985. Transient growth of damped baroclinic waves. *J. Atmos. Sci.* **42**, 2718–2727.

Farrell, B. F. 1989. Optimal excitation of baroclinic waves. *J. Atmos. Sci.* **46**, 1193–1206.

Farrell, B. F. and Ioannou, P. J. 1993. Stochastic dynamics of baroclinic waves. *J. Atmos. Sci.* **50**, 4044–4057.

Farrell, B. F. and Ioannou, P. J. 1996. Generalized stability theory. Part I: Autonomous operators. *J. Atmos. Sci.* **53**, 2025–2040.

Gajić, Z. and Qureshi, M. 1995. *Lyapunov Matrix Equation in System Stability and Control*. Academic Press, San Diego. 255 pp.

Gaspari, G. and Cohn, S. E. 1999. Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.* **125**, 723–757.

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.

Gohberg, I. C. and Kreĭn, M. G. 1969. *Introduction to the Theory of Linear Nonselfadjoint Operators*. American Mathematical Society, Providence, RI.

Golub, G. H. and Van Loan, C. F. 1996. *Matrix Computations*. Third edn, The Johns Hopkins University Press, Baltimore. 694 pp.

Halmos, P. R. 1967. *A Hilbert Space Problem Book*. Van Nostrand-Reinhold, New York. 365 pp.

Horn, R. A. and Johnson, C. R. 1985. *Matrix Analyis*. Cambridge University Press, New York. 561 pp.

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

Kalman, R. 1960. A new approach to linear filtering and prediction problems. *Trans. ASME, Ser. D, J. Basic Eng.* **82**, 35–45.

Legras, B. and Vautard, R. 1996. A guide to Lyapunov vectors. In: Proceedings of a Seminar on Predictability, Vol. 1, European Centre for Medium-Range Forecasts, Reading, England, pp. 143–156.

Mitchell, H. L. and Daley, R. 1997. Discretization error and signal/error correlation in atmospheric data assimilation (I) All scales resolved. *Tellus* **49A**, 32–53.

Molteni, F., Buizza, R., Palmer, T. N. and Petroliagis, T. 1996. The ECMWF ensemble prediction system: Methodology and validation. *Q. J. R. Meteorol. Soc.* **122**, 73–119.

Mori, T., Fukuma, N. and Kuwahara, M. 1982. Upper and lower bounds for the solution to the discrete Lyapunov matrix equation. *International Journal of Control* **36**, 889–892.

Rabier, F., McNally, A., Andersson, E., Courtier, P., Unden, P., Eyre, J., Hollingsworth, A. and Bouttier, F. 1998. The ECMWF implementation of three-dimensional variational assimilation (3D-Var). II: Structure functions. *Q. J. R. Meteorol. Soc.* **124**, 1809–1829.

Swanson, K. L., Vautard, R. and Pires, C. 1998. Four-dimensional variational assimilation and predictability in a quasi-geostrophic model. *Tellus* **50A**, 369–390.

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

Tippett, M. K. and Marchesin, D. 1999*a*. Bounds for solutions of the discrete algebraic Lyapunov equation. *IEEE Trans. Automat. Contr.* **44**, 214–218.

Tippett, M. K. and Marchesin, D. 1999*b*. Upper bounds for the solution of the discrete algebraic Lyapunov equation. *Automatica* **35**, 1485–1489.

Tippett, M. K., Cohn, S., Todling, R. and Marchesin, D. 2000. Conditioning of the stable, discrete-time Lyapunov operator. *SIAM J. Matrix Anal. Appl.* , in press.

Todling, R. and Ghil, M. 1996. Tracking atmospheric instabilities with the Kalman filter. Part I: Methodology and one-layer results. *Mon. Wea. Rev.* **122**, 183–204.

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

Trefethen, L. N., Trefethen, A. E. and Reddy, S. C. 1993. Hydrodynamic stability without eigenvalues. *Science* **261**, 578–584.

Trevisan, A. and Pancotti, F. 1998. Periodic orbits, Lyapunov vectors and singular vectors in the Lorenz system. *J. Atmos. Sci.* **55**, 390–298.

Verlaan, M. and Heemink, A. W. 1997. Tidal flow forecasting using reduced rank square filters. *Stochastic Hydrology and Hydraulics* **11**, 349–368.

Whitaker, J. S. and Sardeshmukh, P. D. 1998. A linear theory of extratropical synoptic eddy statistics. *J. Atmos. Sci.* **55**, 237–258.

Xue, Y., Cane, M. A., Zebiak, S. E. and Blumenthal, M. B. 1994. On the prediction of ENSO: A study with a low-order Markov model. *Tellus* **46A**, 512–528.

List of Figures

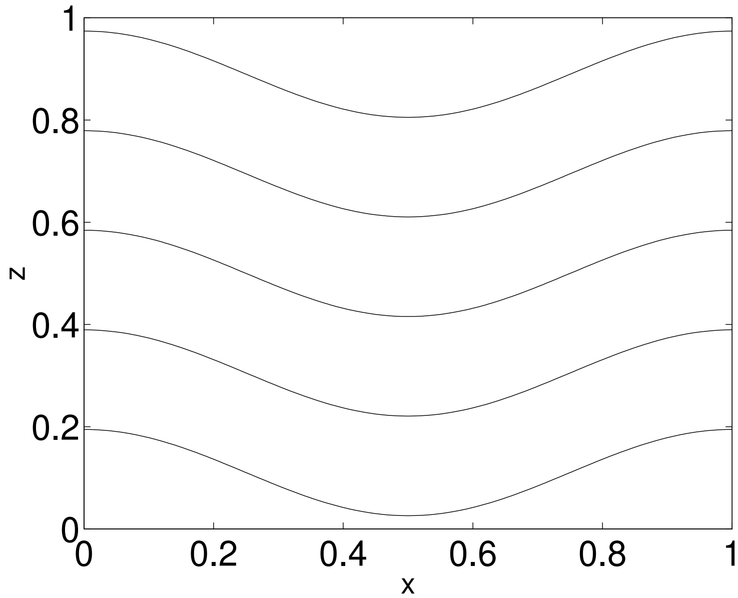

Fig. 1. The streamlines of $\psi(x, z) = -az - c(x)$ for $c(x)$ given by (32).

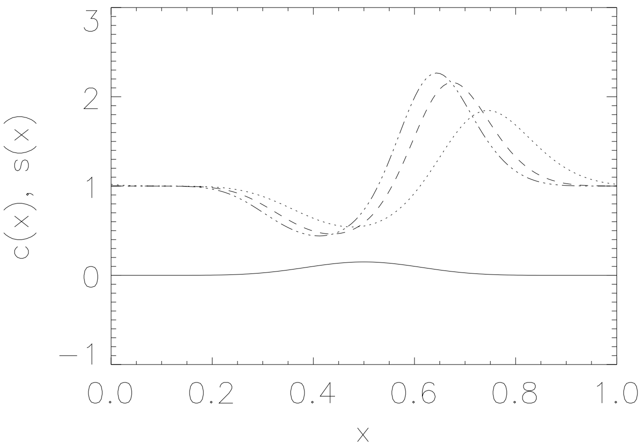

Fig. 2. Solid line $c(x)$; $s(x) = \exp[(c(x - a\tau) - c(x))/a]$ for $\tau = 0.5$ (dotted line), $\tau = 1.0$ (dashed line), and $\tau = 2.0$ (dotted-dashed line).

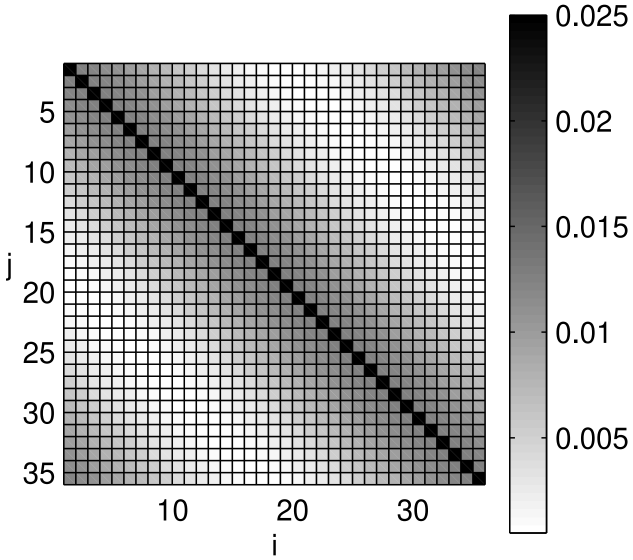

Fig. 3. Checkerboard plot of the elements of the spatially correlated $\mathbf{Q}^m$ used in the experiments and grey scale.

(a) (b)

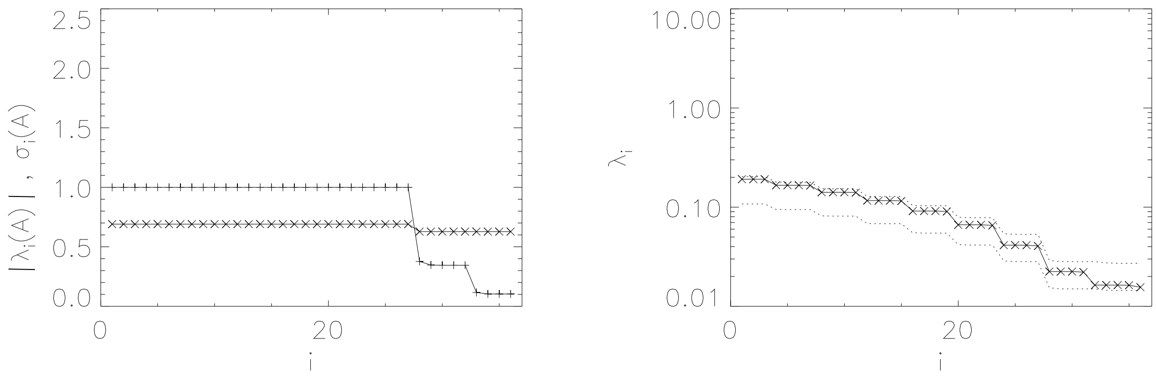

Fig. 4. Singular values and magnitude of the eigenvalues of the steady-state system matrix for normal dynamics ($c'(x) = 0$) and observations at the first 9 grid points. Panel (a) shows the singular values (plus signs) and magnitude of the eigenvalues (crosses) of **A** for $\mathbf{Q}^m = 0.025\mathbf{I}$. Panel (b) shows the eigenvalues of **P** (crosses), and upper and lower bounds from (22) (dotted lines).

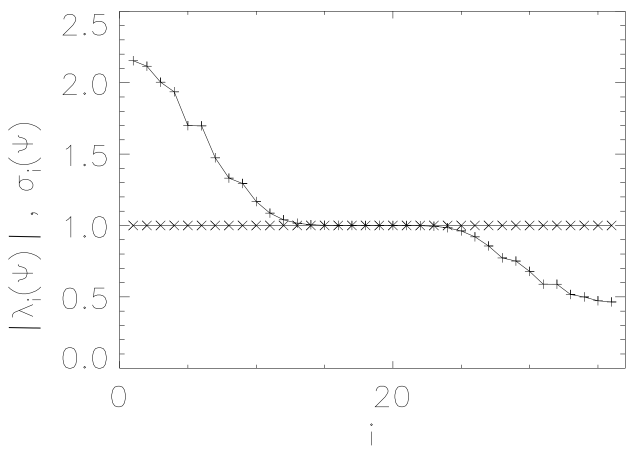

Fig. 5. Singular values (plus signs) and magnitude of the eigenvalues (crosses) of $\mathbf{\Psi}$ for the non-normal dynamics case.

(a) (b)

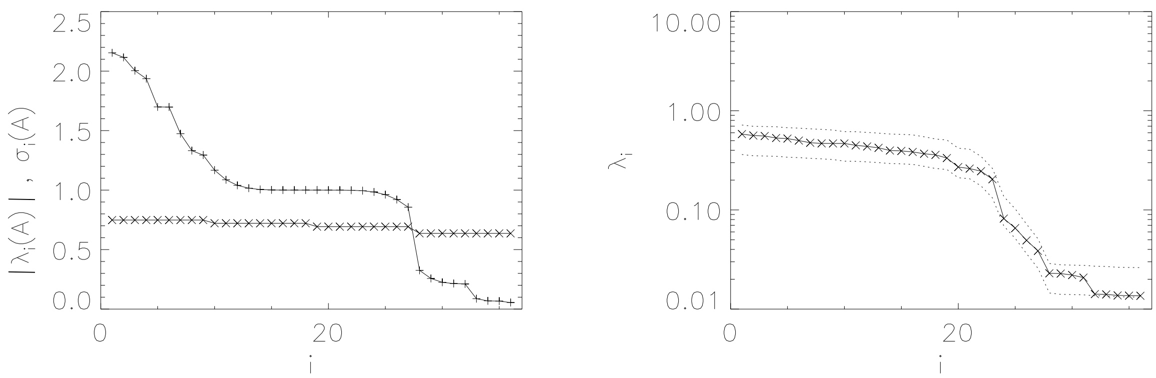

Fig. 6. As in Fig. 4, but for nonnormal dynamics and an observation operator that corresponds to observing the trailing left singular vectors of the dynamics $\mathbf{\Psi}$.

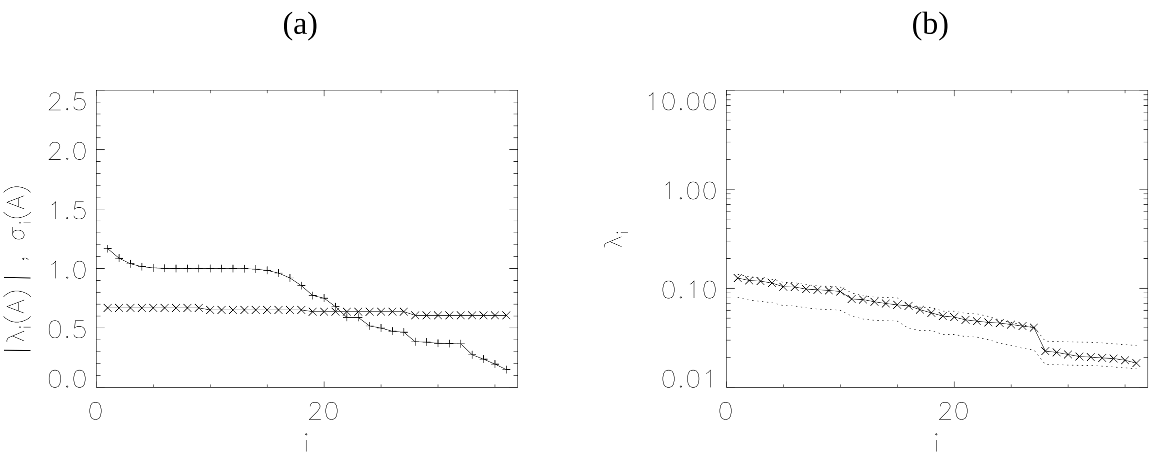

Fig. 7. As in Fig. 6, but when the leading left singular vectors are observed.

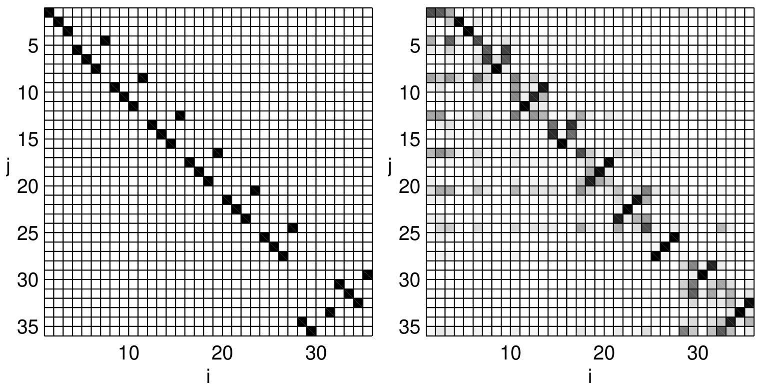

Fig. 8. Checkerboard plots representing the projection of the eigenvectors of $\mathbf{P}$ onto the eigenvectors of $\mathbf{B}$ for the normal dynamics case. The image on the left is for $\mathbf{Q}^m = 0.025\mathbf{I}$ and that on the right for the spatially correlated model error.

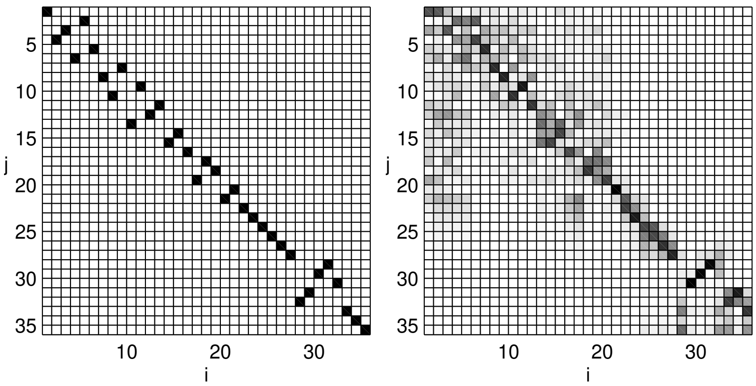

Fig. 9. As in Fig. 8, but for nonnormal dynamics and with observations of the trailing left singular vectors of the dynamics.

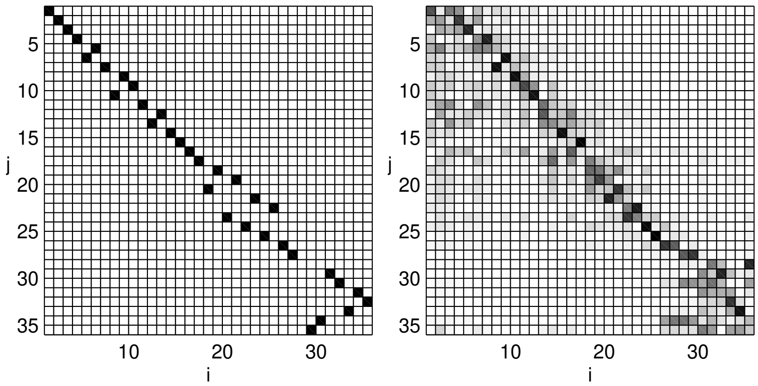

Fig. 10. As in Fig. 9, but with observations of the leading left singular vectors of the dynamics.

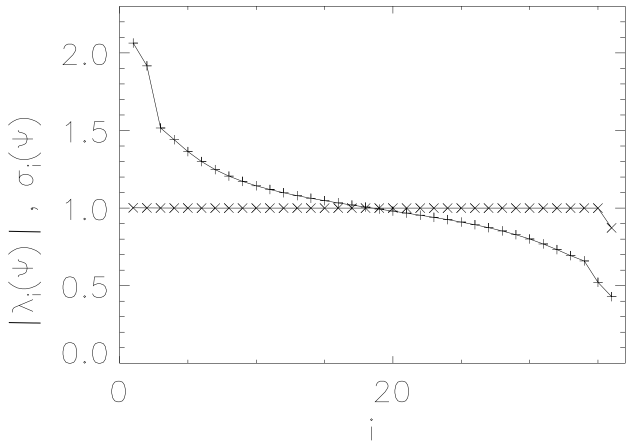

Fig. 11. Singular values (plus signs) and magnitude of the eigenvalues (crosses) of $\mathbf{\Psi}$ for the damped baroclinic wave model.

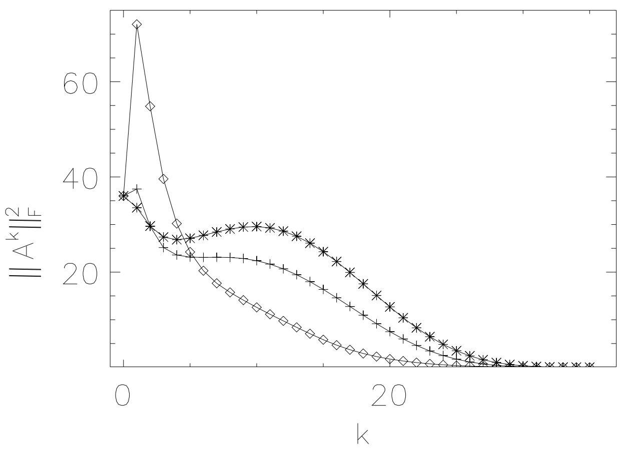

Fig. 12. The quantity $||\mathbf{A}^k||_F^2$ plotted as a function of k for the damped baroclinic wave dynamics and observing patterns pattern 1 (diamonds), pattern 2 (plus signs) and pattern 3 (asterisks) as described in the text.

(a) (b)

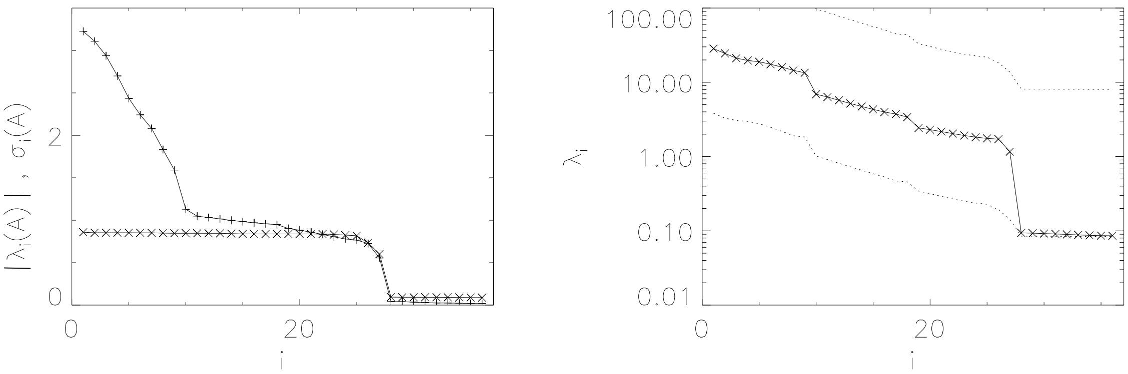

Fig. 13. Singular values and magnitude of the eigenvalues of the system matrix for dynamics of the damped baroclinic wave model. Panel (a) shows the singular values (plus signs) and magnitude of the eigenvalues (crosses) of $\mathbf{A}$. Panel (b) shows the eigenvalues of $\mathbf{P}$ (crosses), and upper and lower bounds from (22) (dotted lines).

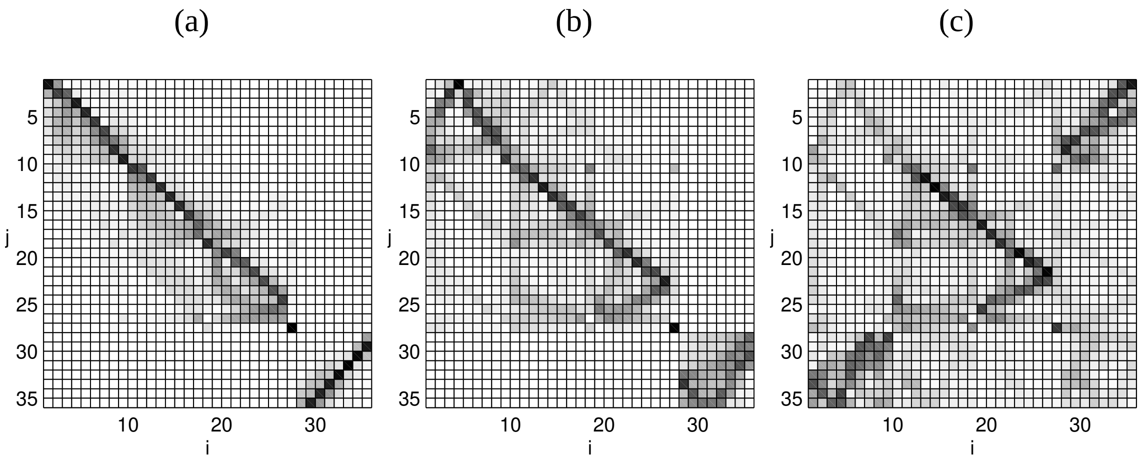

Fig. 14. Checkerboard plots of the elements of the matrix $\mathbf{M} = \mathbf{Z}^*\mathbf{W}$ for $\mathbf{Z}$ taken to be (a) the eigenvectors of $\mathbf{B}$, (b) the eigenvectors of $\mathbf{B}_1$ and (c) the right singular vectors of $\mathbf{A}$.