State Estimation Using a Reduced-Order Kalman Filter

BRIAN F. FARRELL

Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts

PETROS J. IOANNOU

Department of Physics, National and Capodistrian University of Athens, Athens, Greece

(Manuscript received 5 September 2000, in final form 22 February 2001)

ABSTRACT

Minimizing forecast error requires accurately specifying the initial state from which the forecast is made by optimally using available observing resources to obtain the most accurate possible analysis. The Kalman filter accomplishes this for a wide class of linear systems, and experience shows that the extended Kalman filter also performs well in nonlinear systems. Unfortunately, the Kalman filter and the extended Kalman filter require computation of the time-dependent error covariance matrix, which presents a daunting computational burden. However, the dynamically relevant dimension of the forecast error system is generally far smaller than the full state dimension of the forecast model, which suggests the use of reduced-order error models to obtain near-optimal state estimators. A method is described and illustrated for implementing a Kalman filter on a reduced-order approximation of the forecast error system. This reduced-order system is obtained by balanced truncation of the Hankel operator representation of the full error system and is used to construct a reduced-order Kalman filter for the purpose of state identification in a time-dependent quasigeostrophic storm track model. The accuracy of the state identification by the reduced-order Kalman filter is assessed by comparison to the true state, to the state estimate obtained by the full Kalman filter, and to the state estimate obtained by direct insertion.

1. Introduction

An important component of forecast error is error in the analysis of the initial state from which the forecast is made. Analysis error can be reduced by taking more observations, by taking more accurate observations, by taking observations at locations chosen to better constrain the forecast, and by extracting more information from the observations that are available. The last of these, obtaining the maximum amount of information from observations, is attractive because it makes existing observations more valuable and because, at least for linear systems, there is a solution to the problem of extracting the maximum information from a given set of observations: under appropriate assumptions the problem of extracting the maximum amount of information from a set of observations of a linear system in order to minimize the uncertainty in the state estimate is solved by the Kalman filter (KF; Kalman 1960; Ghil and Malanotte-Rizzoli 1991; Wunch 1996; Ide et al. 1997; Lermusiaux and Robinson 1999). Moreover, application of the Kalman filter to the local tangent error equations of a nonlinear system provides a first-order approximation to the optimal data assimilation method, which is valid in the limit of sufficiently small errors. This nonlinear extension of the KF is referred to as the extended Kalman filter (EKF; Ghil et al. 1981; Miller et al. 1994; Ide and Ghil 1997; Ghil 1997).

Unfortunately, the Kalman filter and the extended Kalman filter require statistical description of the forecast error in the form of the error covariance and obtaining the required error covariance involves integrating a system with dimension equal to the square of the dimension of the forecast system. Direct integration of a system of such high dimension is not feasible. Attempts to circumvent this difficulty (see review of Ghil 1997) have involved various approximations to the error covariance (Bishop et al. 2001; Tippett et al. 2000) and approximate integration methods (Evensen 1994; Dee 1995; Fukumori and Malanotte-Rizzoli 1995; Cohn and Todling 1996; Verlaan and Heemink 1997; Houtekamer and Mitchell 1998).

While the formal dimension of the forecast error system obtained by linearizing the forecast model about a base trajectory is the same as that of the forecast system itself, there are reasons to believe that the effective dimension is far lower. The trajectory of the system state in a high-dimensional dynamical system typically lies

Corresponding author address: Dr. Brian F. Farrell, Division of Engineering and Applied Sciences, Harvard University, Oxford St., Mail Area H0162, Cambridge, MA 02138.
E-mail: farrell@deas.harvard.edu

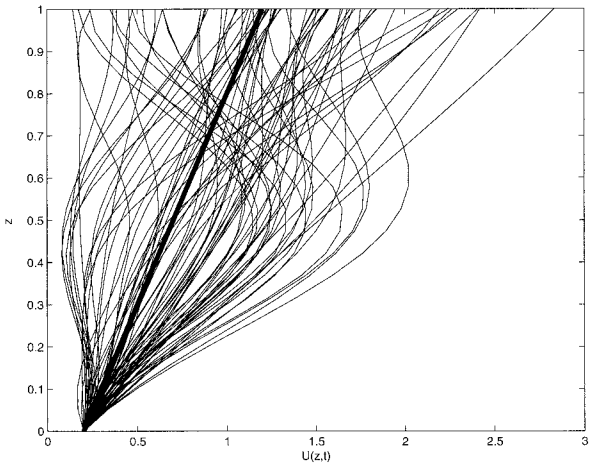

FIG. 1. Realizations of the time-dependent velocity $U(z, t) = 0.2 + z + u(z, t)$, where $u(z, t)$ is given by (31). The random functions $f_i(t)$ are chosen with zero mean and standard deviation 0.5. The bold line is the mean wind, $U(z) = 0.2 + z$.

on a small dimensional subspace of the entire phase space. In chaotic systems all initial conditions approach this attractor, which can be embedded in a space of dimension at most $2d + 1$, where d is the attractor dimension. An estimate of the attractor dimension can be made from the the number of positive Lyapunov exponents [the Kaplan–Yorke dimension; Kaplan and Yorke (1979)] but in any case the attractor dimension is bounded above by the number of Lyapunov exponents associated with positive volume growth along the system trajectory in phase space (Illyashenko 1983). While this is useful conceptually for bounding the dimension of the embedding space, identifying the subspace itself is more difficult in the case of nonlinear and time-dependent systems. However, in the case of stochastically forced linear normal systems the analogous subspace to which the solution is primarily confined can be easily found by eigenanalysis of the covariance matrix of the system forced white in space and time. The resulting empirical orthogonal function (EOF) spectrum typically falls off rapidly in physical models. The eigenvectors may be identified with the modes of the normal operator and the corresponding eigenvalues are the variance accounted for by the modes (North 1984; Farrell and Ioannou 1996, henceforth FI96). The fact that a restricted number of EOFs account for nearly all of the variance in normal systems shows that the effective dynamical dimension of these systems is small compared with the dimension of their phase space. This notion of quantifying the effective dimension of normal linear systems can be extended to bound the effective dimension of nonnormal systems (Farrell and Ioannou 2001, henceforth FI01).

In the case of the tangent linear forecast error system the spectrum of optimal perturbations of the error propagator over the forecast interval typically comprises a few hundred growing structures (Buizza and Palmer 1995) and Lyapunov spectra for error growth have shown similar numbers of positive exponents (Palmer et al. 1998), which suggests from the above considerations that the effective dimension of the error system is $O(10^3)$.

The problem of reducing the order of a linear dynamical system can be cast mathematically as that of finding a finite-dimensional representation of the dynamical system so that the Eckart–Schmidt–Mirsky (ESM) theorem (Stewart and Sun 1990) can be applied to obtain an approximate truncated system with quantifiable error. The ESM theorem states that the optimal k-order truncation of an n-dimensional matrix in the Euclidean or Frobenius norm is the matrix formed by truncating the singular value decomposition of the matrix to its first k singular vectors and singular values. A method for exploiting the ESM theorem to obtain a reduced-order approximation to a dynamical system was developed in the context of controlling lumped parameter engineering systems and is called balanced truncation (Moore 1981; Glover 1984; Zhou and Doyle 1998). Balanced truncation was applied to the set of ordinary differential equations approximating the partial differential equations governing perturbation growth in time-independent atmospheric flows by FI01.

In this work a reduced-order Kalman filter based on balanced truncation is applied to a time-dependent Lyapunov unstable quasigeostrophic model of a forecast tangent linear error system. We first briefly review the method of balanced truncation of time-independent dynamics and then construct a reduced-order Kalman filter for a linear time-dependent storm track error model and study its accuracy in a series of experiments.

2. Reducing model order by balanced truncation

The error dynamics are assumed to be governed by the linear system:

$$\frac{d\psi}{dt} = \mathbf{A}\psi, \tag{1}$$

where ψ is the error state vector and $\mathbf{A}$ is the matrix dynamical operator, which may be time dependent, but will for the time being be assumed to be time independent. Because of the high dimension of the error system (1) in forecast applications we are interested in exploring the accuracy of reduced-order approximations to the system.

Before proceeding with the description of the method of balanced order reduction we must first choose the norm that will be used to measure the accuracy of the approximation. The accuracy is measured by the norm of the Euclidean length of the errors incurred in a chosen variable. This norm is the square root of the Euclidean inner product in this variable. If another norm is selected to measure the accuracy of the approximation, then the

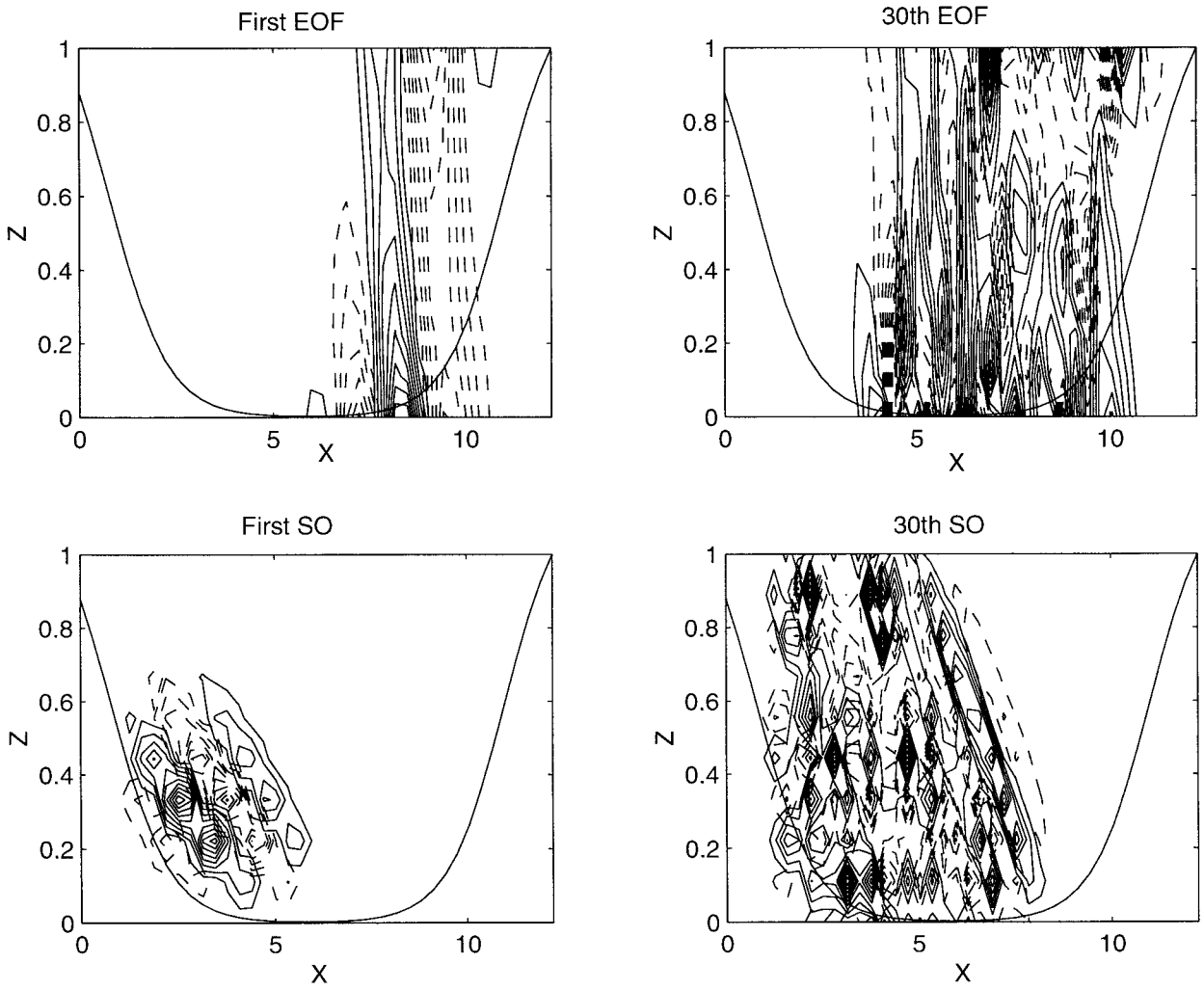

FIG. 2. (top) The streamfunction of the 1st and the 30th EOFs along $y = 0$. The 1st EOF accounts for 23% of the variance, the 30th EOF accounts for 0.35% of the variance. (bottom) The structure of the streamfunction of the 1st and 30th stochastic optimals. The 1st SO is responsible for producing 19.7% of the variance; the 30th SO is responsible for producing 0.48% of the variance. The solid curve indicates the structure of the Rayleigh damping coefficient $r(x)$. The spatially localized damping region simulates the enhanced dissipation and disruption that occurs when perturbations encounter the continental landmasses. The damping region suppresses the convective or modal instabilities that would otherwise occur if the flow were of infinite extent or had periodic conditions, respectively.

most direct method of accounting for this choice is to transform the variable used to represent the state of the system so that the Euclidean inner product in the transformed variable corresponds to the new norm. The reduced-order approximate system resulting from balanced truncation will in general depend on the norm. As discussed in FI01, optimal order reduction for stable normal systems is immediate: it is the Galerkin method based on projection of the dynamics onto the least damped modes. Difficulties in the reduction process arise in cases for which the system is nonnormal in the variable corresponding to the chosen norm. Then the Galerkin method based on projection on the least damped modes is suboptimal and the reduction must proceed by including in the retained subspace the distinct subspaces of the preferred excitations and preferred responses of the system. Throughout this paper we have chosen streamfunction as the error variable, the rms of which is to be minimized in the construction of the model order reduction. However, we find that the results do not change qualitatively if the energy norm is chosen instead.

The preferred structures of response of the nonnormal storm track system are revealed by stochastically forcing the system with spatially and temporally uncorrelated unitary forcing and calculating the eigenfunctions of the resulting mean covariance matrix $\mathbf{P} = \langle \psi\psi^{\dagger} \rangle$ (the brackets denote an ensemble average, and † the Hermitian transpose of a vector or a matrix). The covariance matrix under such forcing is given by

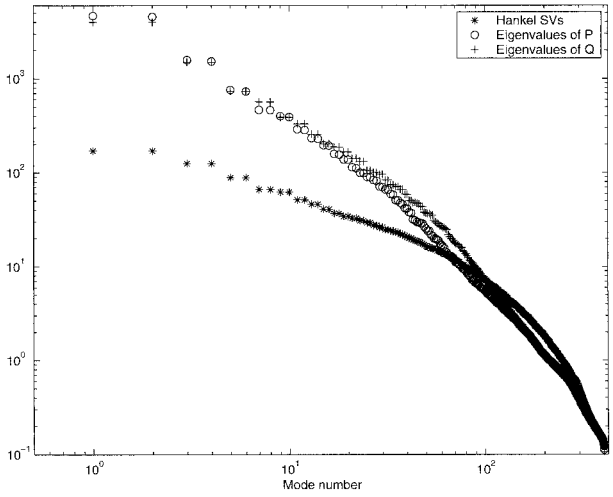

FIG. 3. The Hankel singular values (stars) compared to the eigenvalues of the covariance matrix **P** (circles), and the eigenvalues of the stochastic optimal matrix **Q** (crosses). The Hankel singular values are the square roots of the eigenvalues of the product **PQ**. Note that the variance accounted for by the EOFs (the eigenvalues of **P**), and the variance generated by the SOs (the eigenvalues of **Q**) fall much more rapidly with mode number than do the Hankel singular values.

$$\mathbf{P} = \int_0^\infty e^{\mathbf{A}t} e^{\mathbf{A}^\dagger} \, dt, \tag{2}$$

and this integral is readily calculated by solving the Lyapunov equation (FI96),

$$\mathbf{AP} + \mathbf{PA}^\dagger = -\mathbf{I}, \tag{3}$$

which **P** satisfies, as can be easily verified. The Hermitian and positive definite matrix **P** characterizes the response of the system and its orthogonal eigenvectors, ordered in decreasing magnitude of their eigenvalue, are the EOFs of the system under spatially and temporally uncorrelated forcing.

The preferred structures of excitation of the system are determined from the stochastic optimal matrix,

$$\mathbf{Q} = \int_0^\infty e^{\mathbf{A}^\dagger t} e^{\mathbf{A}t} \, dt, \tag{4}$$

the orthogonal eigenvectors of which, in decreasing magnitude of their eigenvalue, order the forcing structures according to their effectiveness in producing the statistically maintained variance (for a deterministic interpretation of **Q** see FI01). The eigenvectors of **Q** are called the stochastic optimals (SOs) and because of the nonnormality of the system are distinct from the EOFs. The stochastic optimal matrix **Q** satisfies the back Lyapunov equation:

$$\mathbf{A}^\dagger\mathbf{Q} + \mathbf{QA} = -\mathbf{I}. \tag{5}$$

Lyapunov equations (3) and (5) have unique positive definite solutions **P** and **Q** if **A** is stable. The covariance matrix **P** and stochastic optimal matrix **Q** need to be determined in order to proceed with order reduction by balanced truncation.

A successful order reduction must accurately approximate the dynamics of the system, which can be expressed as the mapping of all past (square integrable) forcings to all future responses of the system. This linear mapping of inputs to outputs is called the Hankel operator. Application of the ESM theorem to the Hankel operator provides the optimal low-order truncation of the dynamics. Remarkably, because of the separation between past forcings and future responses in the Hankel operator representation of the dynamics this operator has finite rank equal to the order of the system; its singular values, denoted by h, are the square root of the eigenvalues of the product of the covariance and stochastic optimal matrix, **PQ**. The balanced truncation transforms the internal coordinates of the system so that the transformed covariance matrix **P** and stochastic optimal matrix **Q** become identical and diagonal (while preserving the inner product of the physical variables). The dynamical system is then truncated in these transformed balanced coordinates. The balanced truncation retains a leading subset of EOFs and SOs of the dynamical system and preserves the norm. Balanced truncation preserves the stability of the full system and provides an approximation with known error bounds that is found in practice to be nearly optimal (Moore 1981; Glover 1984; FI01). The procedure used to implement the balanced truncation is now briefly reviewed.

Consider a general k-order truncation of the N-dimensional system (1):

$$\frac{d\tilde{\psi}_k}{dt} = \mathbf{A}_k\tilde{\psi}_k, \tag{6}$$

where $\mathbf{A}_k$ is the reduced $k \times k$ dynamical matrix, with $k < N$, and $\tilde{\psi}_k$ the associated reduced-order k-dimensional state vector, which is related to the full state vector by the transformation $\tilde{\psi} = \mathbf{X}\tilde{\psi}_k$. Similarly, the reduced state vector $\tilde{\psi}_k$ is related to the full state vector by $\tilde{\psi}_k = \mathbf{Y}^\dagger\tilde{\psi}$ (the dagger denotes the Hermitian transpose of a matrix), which implies that $\mathbf{Y}^\dagger\mathbf{X} = \mathbf{I}_k$, where $\mathbf{I}_k$ is the k-order identity matrix. Matrices **Y** and **X** determine the transformation from the full system to the reduced system. The matrix $\mathbf{A}_k$, governing the dynamics in (6), is

$$\mathbf{A}_k = \mathbf{Y}^\dagger\mathbf{AX}. \tag{7}$$

Further details on the construction of the biorthogonal matrices **X** and **Y** are given in the appendix.

A measure of the accuracy of the truncation is the maximum difference that can occur between the full system response, $\psi(t)$, and the reduced-order system response, $\tilde{\psi}(t)$. This measure is the H_∞ norm of the error system:

$$\|\mathbf{A} - \mathbf{A}_k\|_\infty = \sup_\omega \|\mathbf{R}(\omega) - \tilde{\mathbf{R}}(\omega)\|_2, \tag{8}$$

in which the resolvent of the full system, $\mathbf{R}(\omega)$, is defined as $\mathbf{R}(\omega) = (i\omega\mathbf{I} - \mathbf{A})^{-1}$ and the resolvent of the

full-order projection of the reduced system is $\tilde{\mathbf{R}}(\omega) = \mathbf{X}(i\omega\mathbf{I}_k - \mathbf{A}_k)^{-1}$. It is to be recalled that the L_2 norm of a matrix, denoted as $\|\cdot\|_2$, is equal to its largest singular value.

Assuming the Hankel singular values have been ordered as decreasing in magnitude, it can be shown that the error in the H_∞ norm (8) of the approximation of the full system by any k-order system $\mathbf{A}_k$ satisfies the inequality

$$h_{k+1} \leq \|\mathbf{A} - \mathbf{A}_k\|_\infty \leq 2 \sum_{i=k+1}^{N} h_i, \tag{9}$$

where h_{k+1} is the first neglected Hankel singular value (Zhou and Doyle 1998). Although h_{k+1} is only a lower bound on the error, we have found in examples that this lower bound is nearly attained.

3. Constructing a reduced-order Kalman filter

a. The full-order Kalman filter

Consider the perturbation field to be observed only at specific locations and times. The perturbation dynamics are linear and fully described by the sure time-dependent operator $\mathbf{A}(t)$, which in the case of a forecast model is the tangent linear operator.

The perturbation field, ψ, in the true forecast model is thus assumed to be governed by the deterministic equation:

$$\frac{d\psi}{dt} = \mathbf{A}(t)\psi, \tag{10}$$

while the error forecast model is assumed to advance the perturbation field from time t_i to time t_{i+1} according to the equation

$$\psi(t_{i+1}) = \mathcal{M}(t_i)\,\psi(t_i) + \eta(t_i), \tag{11}$$

where η is a white noise process, describing the model error, with zero mean and covariance[1] $\mathcal{Q}$. The propagator $\mathcal{M}(t_i)$ is defined as

$$\mathcal{M}(t_i) \equiv \lim_{\tau \to 0} \prod_{j=1}^{m} e^{\mathbf{A}(t_i + j\tau)\tau}, \tag{12}$$

obtained by m advances of the system by the infinitesimal propagators $e^{\mathbf{A}(t_i+j\tau)\tau}$, where m and τ satisfy the relation $t_{i+1} - t_i = m\tau$, and the matrix $\mathbf{A}(t)$ is assumed to govern the evolution. We select the times t_i to be the times at which observations of the perturbation streamfunction are made.

A vector of observations y_i^o (of length N_{ob}) is made at time t_i and at prescribed locations. These observations are assumed to be linear combinations of the true perturbation field and a measurement uncertainty

$$y_i^o = \mathcal{H}\psi(t_i) + e_i, \tag{13}$$

where $\mathcal{H}$ is the observation matrix (of size $N_{\text{ob}} \times N$) and e_i is the error in the observations at time t_i, which will be assumed to be white in time with known spatial covariance, $\mathcal{R}$. These observations are entrained serially with the forecast to produce the analysis

$$\psi^a(t_{i+1}) = \psi^f(t_{i+1}) + \mathcal{K}(y_{i+1}^o - \mathcal{H}\psi^f(t_{i+1})), \tag{14}$$

in which $\mathcal{K}$ is referred to as the gain matrix (of size $N \times N_{\text{ob}}$, $\psi_a(t_{i+1})$ is the analyzed state at time t_{i+1}, and $\psi^f(t_{i+1})$ the forecast state at time t_{i+1}. The forecast is obtained by advancing the previous analyzed state $\psi^a(t_i)$ using the system propagator (12); that is,

$$\psi^f(t_{i+1}) = \mathcal{M}(t_i)\psi^a(t_i). \tag{15}$$

The gain $\mathcal{K}$ interpolates between the observations, y_i^o, and the forecast, $\psi^f(t_i)$. A plausible choice, if the observations are good, is to trust the observations over the forecast and simply correct the analysis to the observations. This is called direct substitution (Daley 1991). While direct substitution creates imbalance in primitive equation models, it provides a useful example in the balanced quasigeostrophic model used here. Alternatively, a variational problem can be formulated to find the gain that minimizes the expected error between the analysis, ψ^a, and the true state, ψ, which determines an optimal $\mathcal{K}$ called the Kalman gain (Kalman 1960; Ghil and Malanotte-Rizzoli 1991; Wunch 1996). The optimality of $\mathcal{K}$ is predicated on knowledge of the statistical structure of the forecast error and of the observation error. The forecast and analysis error structure are characterized, respectively, by the error covariance matrices $\mathcal{P}^f(t_i) = \langle[\psi(t_i) - \psi^f(t_i)][\psi(t_i) - \psi^f(t_i)]^{\mathrm{T}}\rangle$ and $\mathcal{P}^a(t_i) = \langle[\psi(t_i) - \psi^a(t_i)][\psi(t_i) - \psi^a(t_i)]^{\mathrm{T}}\rangle$ ($\langle\cdot\rangle$ denotes an ensemble average). The Kalman gain at time t_{i+1} is given by

$$\mathcal{K}(t_{i+1}) = \mathcal{P}^f(t_{i+1})\mathcal{H}^{\mathrm{T}}[\mathcal{H}\,\mathcal{P}^f(t_{i+1})\mathcal{H}^{\mathrm{T}} + \mathcal{R}]^{-1}. \tag{16}$$

If the observation error covariance, $\mathcal{R}$, predicts a large error in a structure compared to the forecast error in that structure, $\mathcal{H}\mathcal{P}^f(t_{i+1})\mathcal{H}^{\mathrm{T}}$, the analysis tends to adopt the forecast for that structure, ignoring the observations; and if the observation error is expected to be small compared to the forecast error in a structure, the analysis tends to adopt the observations while ignoring the forecast.

The extent to which the forecast is adopted depends on the analysis error covariance matrix, which is first advanced in time according to the dynamics given by Eq. (11) to produce the forecast error covariance,

$$\mathcal{P}^f(t_{i+1}) = \mathcal{M}(t_i)\mathcal{P}^a(t_i)\mathcal{M}^{\mathrm{T}}(t_i) + \mathcal{Q}, \tag{17}$$

and then corrected by the observation statistics to produce the analysis error covariance:

$$\mathcal{P}^a(t_{i+1}) = \mathcal{P}^f(t_{i+1}) - \mathcal{K}(t_{i+1})\mathcal{H}\mathcal{P}^f(t_{i+1}). \tag{18}$$

[1] Note that the $\mathcal{Q}$ that denotes the model error covariance hereafter is distinct from the $\mathbf{Q}$ that we used in section 2, which denotes the stochastic optimal matrix.

b. The reduced-order Kalman filter

One problem with implementing the Kalman filter for state identification for use in forecast is that the matrix manipulations involved are prohibitively expensive computationally. This is principally because the error covariance matrix corresponding to an atmospheric state with $N = O(10^7)$ degrees of freedom has dimension $N^2 = O(10^{14})$.

In the previous section we showed how to reduce the order of an autonomous linear system by obtaining an accurate balanced truncation of the time-independent operator $\mathbf{A}$. Consider now a time-dependent operator of the form $\mathbf{A}(t) = \mathbf{A}_0 + \mathbf{A}_1(t)$, and assume that the mean operator $\mathbf{A}_0$ is dominant. In previous work we showed that perturbation growth in time-dependent nonnormal systems occurs primarily in the nonnormal subspace of the time mean operator (Farrell and Ioannou 1999), a result verified by Reynolds and Errico (1999) and Gelaro et al. (2000, manuscript submitted to *J. Atmos. Sci.*, hereafter GRE). In the following we take advantage of this result to obtain an approximate reduced-order model of a time-dependent system by reducing the order of the time-dependent operator using the balancing transformation derived for the mean operator. This procedure is found to produce an accurate model order reduction for example nonautonomous stationary systems at far less computational cost than incurred by balancing at each time as in van Dooren (2000). In the following the error covariance matrix and the Kalman gain are obtained using this reduced model and transformed back to the full space for use in updating the state estimate.

Let the dimension of the full system perturbation vector, ψ, be N, and of the reduced system, $\tilde{\psi}_k$, be k with $k < N$. The variables in the reduced system, $\tilde{\psi}_k$, are related to the variables in the full system, $\tilde{\psi}$, by the transformation $\tilde{\psi}_k = \mathbf{Y}^\dagger\tilde{\psi}$. The evolution equation in the $\tilde{\psi}_k$ coordinates is

$$\frac{d\tilde{\psi}_k}{dt} = \mathbf{A}_k(t)\tilde{\psi}_k, \tag{19}$$

where $\mathbf{A}_k(t) = \mathbf{Y}^\dagger\mathbf{A}(t)\mathbf{X}$, as in (7). In this approximation the biorthogonal bases $\mathbf{Y}$ and $\mathbf{X}$ remain the balancing transformation of the mean operator $\mathbf{A}_0$ instead of being recalculated at each step.

In the reduced-order system variables the observation matrix is $\mathcal{H}^k = \mathcal{H}\mathbf{X}$, so that the error covariance matrix in the reduced-order system, $\mathcal{P}^k(t_i) = \mathbf{Y}\mathcal{P}(t_i)\mathbf{Y}^\dagger$, is evolved according to the reduced-order dynamics (10):

$$\mathcal{P}^{kf}(t_{i+1}) = \mathcal{M}^k(t_i)\mathcal{P}^{ka}(t_i)\mathcal{M}^{k\mathrm{T}}(t_i) + \mathcal{Q}^k, \tag{20}$$

where $\mathcal{Q}^k$ is the model error covariance projected on the reduced-order space, that is, $\mathcal{Q}^k = \mathbf{Y}\mathcal{Q}\mathbf{Y}^\dagger$, and $\mathcal{M}^k$ is the reduced-order propagator. The reduced-order covariance matrix is corrected by the observation statistics:

$$\mathcal{P}^{ka}(t_{i+1}) = \mathcal{P}^{kf}(t_{i+1}) - \mathcal{K}^k(t_{i+1})\mathcal{H}\mathcal{P}^{kf}(t_{i+1}), \tag{21}$$

in which the reduced-order Kalman gain is

$$\mathcal{K}^k(t_{i+1}) = \mathcal{P}^{kf}(t_{i+1})\mathcal{H}^{k\mathrm{T}}[\mathcal{H}^k\mathcal{P}^{kf}(t_{i+1})\mathcal{H}^{k\mathrm{T}} + \mathcal{R}^k]^{-1}, \tag{22}$$

where $\mathcal{R}^k = \mathbf{Y}\mathcal{R}\mathbf{Y}^\dagger$.

The Kalman gain that will be used in the full system in order to entrain the observations as in (14) is obtained from the reduced-order system gain:

$$\mathcal{K}^r = \mathbf{X}\mathcal{K}^k, \tag{23}$$

with the superscript indicating that the Kalman gain is obtained from the reduced-order system.

4. Constructing a reduced-order Kalman filter for a model storm track error system

a. Formulating the storm track error model

Consider an idealized model of the midlatitude storm track consisting of a Boussinesq atmosphere with constant stratification and constant shear in thermal wind balance on a β-plane channel with periodic boundary conditions in the zonal, x, direction; solid walls located at two latitudes in the meridional, y, direction; and a solid lid at height $z = H$, simulating the tropopause. The observed zonal localization of a midlatitude storm track is simulated in the model by terminating the channel with a linear damping modeling the storm track exit region. The stability properties of such a storm track model are discussed in FI96.

Zonal and meridional lengths are nondimensionalized by $L = 1200$ km, vertical scales by $H = fL/N = 10$ km, velocity by $U_0 = 50$ m s^{-1}, and time by $T = L/U_0$, so that a time unit is approximately 6.7 h. The Brunt–Väisälä frequency is $N = 10^{-2}$ s^{-1}, and the Coriolis parameter is $f = 10^{-4}$ s^{-1}. The corresponding nondimensional value of the planetary vorticity gradient is $\beta = 0.46$.

The nondimensional linearized equation that governs evolution of streamfunction perturbations is

$$\frac{\partial\nabla^2\psi}{\partial t} = -U(z)\nabla^2 D\psi - \left[\beta - \frac{d^2U(z)}{dz^2}\right]D\psi - r(x)\nabla^2\psi, \tag{24}$$

in which the perturbation is assumed to be in the form $\psi(x, z, t)\, e^{ily}$, where l is the meridional wavenumber; $\nabla^2\psi$ is the perturbation potential vorticity, with $\nabla^2 \equiv \partial^2/\partial x^2 + \partial^2/\partial z^2 - l^2$; and $D \equiv \partial/\partial x$. The perturbation potential vorticity damping rate $r(x)$ is taken to vary smoothly in the zonal direction with form

$$r(x) = \frac{\mu}{2}\left[2 - \tanh\left(\frac{x - \pi/4}{\delta}\right) + \tanh\left(\frac{x - 7\pi/2}{\delta}\right)\right], \tag{25}$$

in which parameters controlling the maximum damping rate and the width of the damping region have been chosen to be $\mu = 5$ and $\delta = 1.5$, respectively. The mean

velocity profile is $U(z) = 0.2 + z$. The zonal extent of the reentrant channel is $0 < x < 4\pi$; latitudinal walls are located at $y = 0$ and $y = 1$; and the ground and tropopause boundaries are located at $z = 0$ and $z = 1$, respectively. In the following we consider perturbations with $l = 1$.

Conservation of potential temperature at the ground and tropopause provides the boundary conditions:

$$\frac{\partial^2\psi}{\partial t\partial z} = -U(0)D\frac{\partial\psi}{\partial z} + U'(0)D\psi - r(x)\frac{\partial\psi}{\partial z} - \Gamma_g(D^2 - l^2)\psi \quad \text{at } z = 0, \tag{26}$$

$$\frac{\partial^2\phi}{\partial t\partial z} = -U(1)D\frac{\partial\psi}{\partial z} + U'(1)D\psi - r(x)\frac{\partial\psi}{\partial z} \quad \text{at } z = 1, \tag{27}$$

where $U'(0)$ and $U'(1)$ denote the velocity shear at $z = 0$ and $z = 1$, respectively. The coefficient of Ekman damping $\Gamma_g \equiv (N/U_0)(\nu/2f)^{1/2}$ is given the value $\Gamma_g = 0.0632$ corresponding to a vertical eddy momentum diffusion coefficient $\nu = 20\ \mathrm{m^2\ s^{-1}}$ in the boundary layer.

The waves evolve with nearly zero damping in the middle third of the channel (a length of $2\pi L \approx 7500$ km), which models the core of the storm track. Because in this model absolute instabilities do not exist with everywhere westerly flow, the storm track is asymptotically stable for all meridional wavenumbers because all perturbations are eventually absorbed on entering the highly dissipative sponge (FI96).

Two scenarios are investigated. In the first a transiently growing disturbance excited near the western boundary of the storm track is modeled using the reduced-order system, the purpose being to illustrate the accuracy of the reduced-order model approximation of the autonomous dynamics. In the second, time dependence is added to produce a Lyapunov unstable model of a tangent linear forecast error system, the time mean operator remaining stable, with the purpose of evaluating the accuracy of the Kalman filter obtained by the reduced-order model in an unstable time-dependent system. Such an unstable time-dependent system provides an even more stringent test of the state estimator than does the time-independent stable and unstable model error systems studied by Todling and Ghil (1994), Ghil and Todling (1996), and Cohn and Todling (1996).

The perturbation dynamics of the time mean storm track are governed by

$$\frac{d\psi}{dt} = \mathbf{A}_0\psi, \tag{28}$$

where

$$\mathbf{A}_0 = (\nabla^2)^{-1}[-(0.2 + z)D\nabla^2 - \beta D - r(x)\nabla^2], \tag{29}$$

in which the Helmholtz operator, ∇^2, has been made invertible by incorporating the boundary conditions.[2]

[2] For waves with a constant meridional wavenumber l, the operator ∇^2 is invertible even for homogeneous boundary conditions.

In the time-dependent storm track the perturbation dynamics described by the time-dependent operator $\mathbf{A}(t)$, which is the sum of the operator, $\mathbf{A}_0$, arising from linearizing about the time mean zonal flow, is given by (29), and the time-dependent deviation operator,

$$\mathbf{A}_1(t) = (\nabla^2)^{-1}\left[-u(z, t)D\nabla^2 + \frac{d^2u(z, t)}{dz^2}D\right], \tag{30}$$

obtained from the linearization of the simple storm track model about zonal velocity fluctuations, $u(z, t)$. We consider a fluctuating zonal wind of form

$$u(z, t) = \frac{f_1(t)}{2}\left[1 - \cos\left(\frac{\pi z}{2}\right)\right] + \frac{f_2(t)}{2}[1 - \cos(2\pi z)] + f_3(t)\sin\left(\frac{\pi z}{2}\right) + f_4(t)z^2, \tag{31}$$

where $f_i(t)$ is a red noise process with mean zero, standard deviation 0.5, and decorrelation time 1.5 days. This red noise process is generated by

$$df_i = -a\ f_i dt + \sigma\ dW, \tag{32}$$

with W a Wiener process. The standard deviation of f_i is given by $\sigma(2a)^{-1/2}$ and the decorrelation time is $1/a$. The desired values are achieved for $a = 1/5$, and $\sigma = (1/10)^{1/2}$. The profiles were chosen to give variance increasing with height, and also to lead to almost surely westerly winds. Typical realizations of the resulting mean flow are shown in Fig. 1. It should be emphasized that the time-dependent operator $\mathbf{A}(t)$ is not to be regarded as an uncertain operator. The operator is certain and the time dependence of the operator is assumed to be deterministic. We use the stochastic system (32) only for convenience to generate a sample realization from a class of statistically similar realizations.

The dynamical operator is approximated spectrally in the zonal direction and with finite differences in the vertical. With 40 zonal harmonics and 10 levels in the vertical the resulting dynamical system has $N = 400$ degrees of freedom.

b. Applying balanced truncation to the mean storm track error model

In order to obtain a balanced truncation of the storm track model governed by mean operator $\mathbf{A}_0$ given in (29) we first obtain the covariance matrix, $\mathbf{P}$, and the stochastic optimal matrix, $\mathbf{Q}$, by solving Lyapunov equations (3) and (5), respectively. The eigenfunction of $\mathbf{P}$ associated with the largest eigenvalue is the first EOF of the perturbation field, and the eigenfunction of $\mathbf{Q}$ associated with the largest eigenvalue is the first SO of the perturbation field. The structure of the first EOF, which accounts for 23% of the streamfunction perturbation variance, is concentrated in the exit region of the storm track as can be seen in Fig. 2 (top-left panel). By

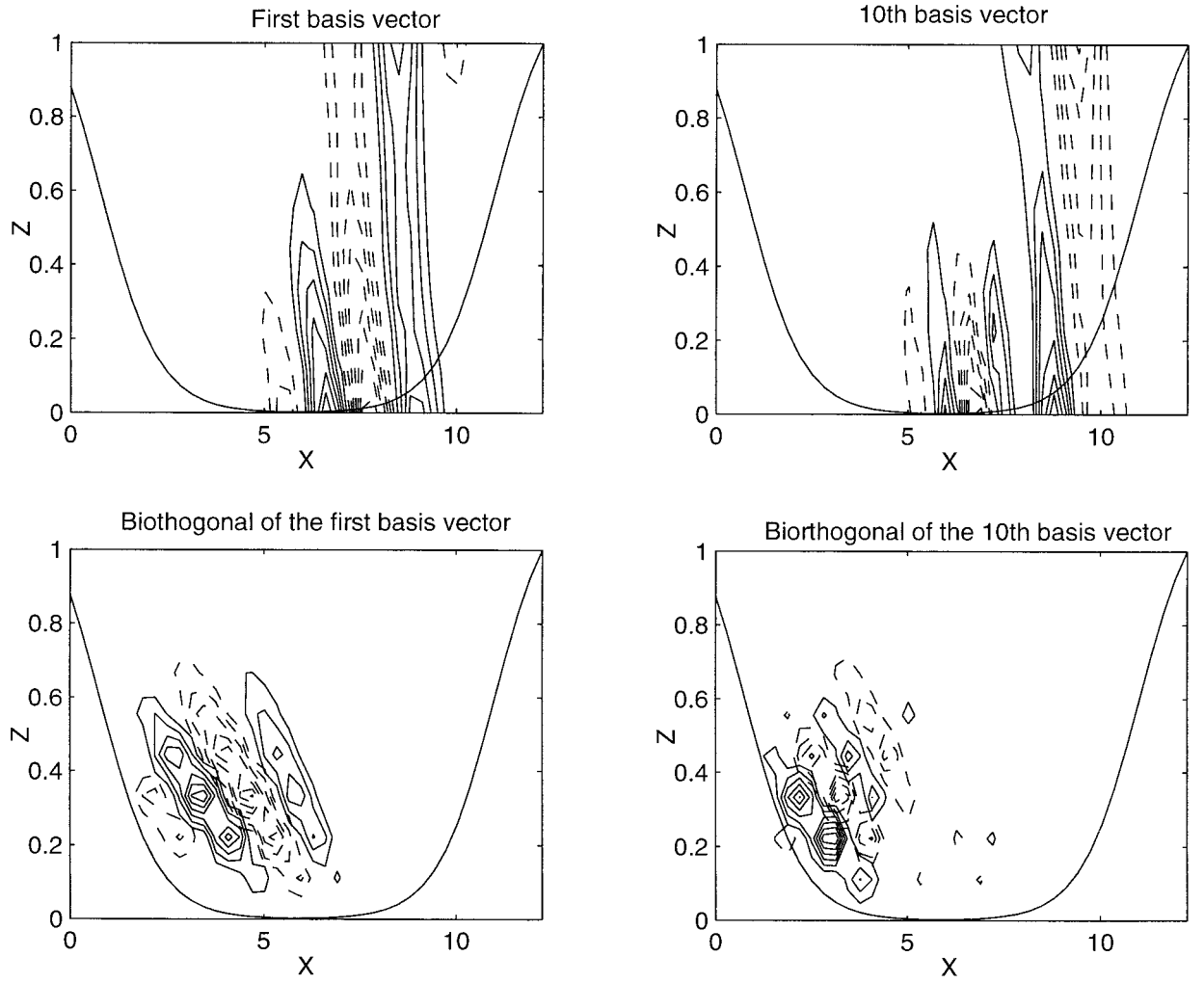

FIG. 4. (top left) The streamfunction along $y = 0$ of the first retained basis vector of the balanced truncation of the system, given by the first column of **X**. (top right) The streamfunction of the 10th basis vector of the expansion of the balanced truncation, given by the 10th column of **X**. (bottom left) The streamfunction of the biorthogonal of the first basis vector. It is given by the first column of **Y**. (top right) the streamfunction of the 10th basis vector. It is given by the 10th column of **Y**.

contrast, the first SO, which is responsible for generating 19.7% of the streamfunction perturbation variance, is concentrated at the entrance region of the storm track and is nearly orthogonal to the first EOF as can be also seen in Fig. 2 (bottom-left panel). This near orthogonality between the EOF structures and SO structures remains even at order 30. Balanced truncation accomplishes an accurate representation of the dynamics by retaining both the structure of the dominant EOFs and of the SOs. It is clear from Fig. 2 that truncations based on projections on the leading EOFs will be very suboptimal as the leading EOFs span well only the exit region of the storm track, leaving the dynamically important entry region of the storm track, where perturbations start growing, virtually without support in the span of the retained basis.

Although the error in the frequency response of a balanced truncation [cf. Eq. (9)] is bounded above by twice the sum of the neglected Hankel singular values and below by the first neglected Hankel singular value, experience shows balanced truncation of fluid systems results in errors close to the lower bound. The Hankel singular values and the eigenvalues of **P** and the **Q** for the storm track model are shown in Fig. 3. Note that the decrease with mode number of the eigenvalues of **P** and of the eigenvalues of **Q** is more rapid than that of the Hankel singular values. But this more rapid decrease with mode number of the eigenvalues of **P** and **Q** does not indicate the order required for an accurate approximation; this is instead determined by the first neglected Hankel singular value, which falls more slowly with mode number.

It is often assumed that a system can be well approximated by the Galerkin method based on projection onto a subspace of its leading EOFs; with the effectiveness of the truncation being judged from the

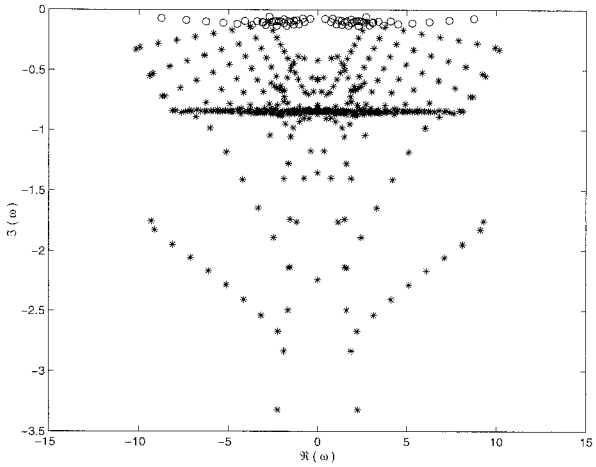

FIG. 5. The frequency [$\Re(\omega)$] and decay rate [$\Im(\omega)$] of the eigenvalues of the original system (stars), and of the order-60 balanced operator $\mathbf{A}_{0\,60}$ (circles).

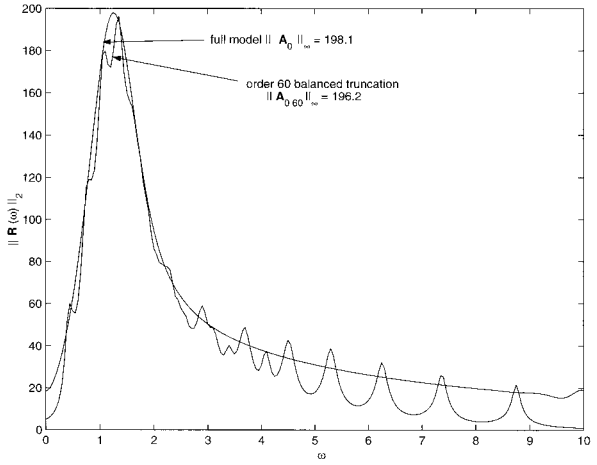

FIG. 6. The maximum singular value of the resolvent $\mathbf{R}(\omega) = (i\omega\mathbf{I} - \mathbf{A}_0)^{-1}$ of the full system as a function of frequency, ω. The frequency is in dimensional units of $1/d$. The maximum of the curve is the H_∞ norm of $\mathbf{A}_0$ which is found here to be 198.1. Also shown is the maximum singular value of the resolvent $\mathbf{R}_{60}(\omega) = (i\omega\mathbf{I} - \mathbf{A}_{0\,60})^{-1}$ of $\mathbf{A}_{060}$, which is the operator obtained from an order-60 balanced truncation of $\mathbf{A}_0$. The maximum of this curves is the H_∞ norm of $\mathbf{A}_{0\,60}$, which is found to be 196.2.

magnitude of the eigenvalues of the neglected EOFs. While this is valid for normal systems, we see here that for nonnormal systems the decrease with mode number of the eigenvalues of the covariance matrix is misleading and generally optimistic as an estimate of the order required for an accurate approximation.

A subset of the columns of $\mathbf{X}$ is retained in the balanced truncation. This nonorthogonal basis and its biorthogonal, the columns of $\mathbf{Y}$, are constructed so as to capture the structures supporting the dynamics most efficiently, simultaneously accounting for the preferred responses (EOFs) and the preferred excitations (SOs) of the dynamics. The 1st and the 10th structures retained in the dynamics (the 1st and the 10th columns of $\mathbf{X}$) and their biorthogonal structures (the 1st and 10th columns of $\mathbf{Y}$) are shown in Fig. 4.

The storm track model and its reduced-order approximate have very different eigenvalue spectra (Fig. 5). The eigenvalue spectrum of the reduced-order approximate is such that the frequency response of the approximate system is as close as possible to that of the original system, which is shown in Fig. 6. This results both from a decrease in the stability of the reduced system compared to that of the full system and from the nonnormality of the reduced system.

The accuracy of the approximation is measured by the H_∞ norm of the error dynamical system $\|\mathbf{A}_0 - \mathbf{A}_{0\,60}\|_\infty$, which, as discussed in the previous section, lies between the lower bound given by the first neglected Hankel singular value, $h_{61} = 13.8$, and the upper bound: $2\sum_{i=61}^{400} h_i = 1.8 \times 10^3$. The largest singular value of the error system resolvent as a function of frequency is shown in Fig. 7, where it can be seen that $\|\mathbf{A}_0 - \mathbf{A}_{0\,60}\|_\infty = 28.5$, which shows that the balanced truncation error in this example is only approximately twice its lower bound. The error is nearly white for the span of frequencies that correspond to the frequencies of the system eigenmodes (5). For comparison, the error incurred in the order-60 Galerkin method based on projection of the dynamics onto the first 60 EOFs and the error incurred in an order-60 Galerkin projection onto the first 60 least damped modes, are also shown in Fig. 7. It can be seen that the EOF projection performs appreciably worse than the bal-

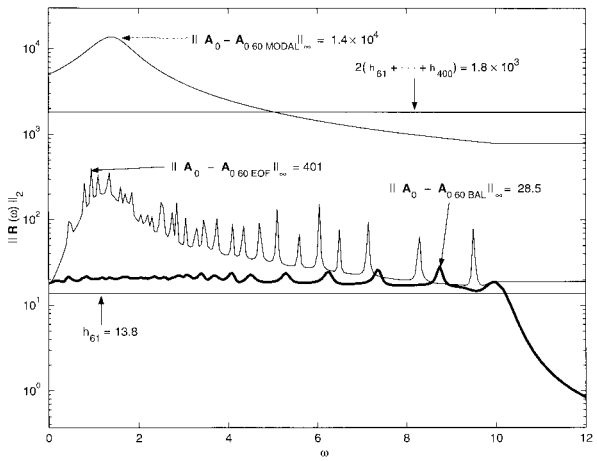

FIG. 7. The maximum singular value of the error system $\mathbf{A}_0 - \mathbf{A}_{0\,60}$ as a function of frequency, ω. The frequency is in dimensional units of $1/d$. The system $\mathbf{A}_{0\,60}$ is an order-60 approximation obtained from $\mathbf{A}$ by balanced truncation. The maximum of this curve is the H_∞ error of the order-60 balanced truncation, which is seen here to be 28.5. The theoretical minimum error of an order-60 truncation, which equals the first neglected Hankel singular value $h_{61} = 13.8$, and the theoretical upper bound on the error, which equals the sum of the neglected singular values $\sum_{i=61}^{400} h_i = 1800$, are also shown. The balanced truncation is seen to be nearly optimal. Similarly, H_∞ errors are shown for order-60 modal and the Galerkin method based on a 60 leading EOF projection.

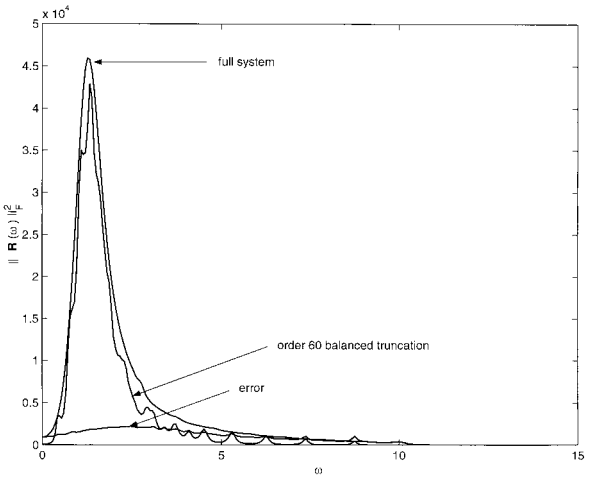

FIG. 8. The square of the Frobenius norm of the resolvent of the full system $\mathbf{A}_0$, of the order-60 system obtained by balanced truncation $\mathbf{A}_{60}$, and of the error system $\mathbf{A}_0 - \mathbf{A}_{0\,60}$, as a function of frequency, ω. The frequency is in units of $1/d$. The square of the Frobenius norm of the resolvent at each frequency is the sum of the squares of the singular values. The areas under these curves give the variance maintained under spatially and temporally white forcing of unit spectral density in each degree of freedom.

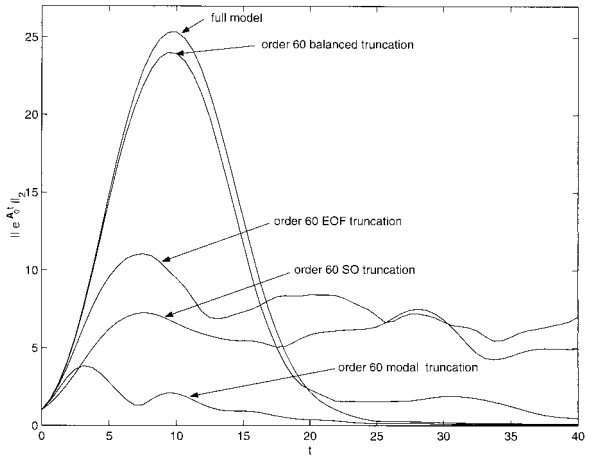

FIG. 9. Optimal growth, $\|e^{\mathbf{A}_0 t}\|_2$, as a function of time for the storm track model. Shown is the optimal growth for the full system with 400 degrees of freedom, the optimal growth produced by an order-60 approximate system obtained by balanced truncation of the full system, the optimal growth attained by the order 60-approximate system obtained by the Galerkin method based on projection on the first 60 EOF's, the first 60 SOs, and the first 60 least damped modes.

anced truncation, while the modal truncation at this order is useless.

The streamfunction variance maintained by spatially and temporally white forcing of unit spectral magnitude is given by

$$\text{trace}(\mathbf{P}) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)\, d\omega, \tag{33}$$

where the streamfunction spectral density $F(\omega)$,

$$F(\omega) = \text{trace}(\mathbf{R}^{\dagger}(\omega)\, \mathbf{R}(\omega)), \tag{34}$$

is the square of the Frobenius norm[3] of the resolvent, $\mathbf{R}(\omega)$ (FI96). A plot of the Frobenius norm of the spectral density $F(\omega)$ is shown in Fig. 8 for both the full system and the order-60 balance truncation. Note that while the response streamfunction variance is well approximated by the reduced system over the frequency range for which the variance is large (0.5 d^{-1}–9 d^{-1}), the error variance exceeds the response variance itself at very small frequencies and at very high frequencies.

The optimal growth[4] as a function of optimizing time attained by the full system and by the following are all shown in Fig. 9: the order-60 balanced truncation, the order-60 system obtained by the Galerkin method based on projection on the first 60 EOFs, the order-60 system obtained by the Galerkin method based on projection on the first 60 SOs, and the order-60 system obtained by Galerkin projection on the first 60 least damped modes. Note that the balanced truncation performs very well, reproducing the optimal growth nearly perfectly up to $t = 5$, corresponding to about 2 days. By comparison the EOF and SO truncations perform appreciably worse and the modal truncation gives even poorer results.

The structure of the initial perturbation that leads to greatest square streamfunction growth at $t = 10$ in the full system, together with the resulting structure, is shown in Fig. 10; for comparison these structures as obtained by the truncated system are also shown. The structures are well captured by the order-60 reduced system.

c. *Constructing a reduced-order Kalman filter for a time-dependent storm track error model*

We have shown how to obtain an accurate balanced truncation of a time-independent operator $\mathbf{A}$. In previous work we showed that perturbation growth in time-dependent nonnormal systems occurs primarily in the nonnormal subspace of the time mean operator (Farrell and Ioannou 1999). In the following we take advantage of this result to obtain an approximate reduced-order model of a time-dependent system by reducing the order of the time-dependent operator using the balancing transformation derived for the mean operator. This reduced system is used to advance the error covariance matrix that determines the Kalman gain.

Although the time mean system, $\mathbf{A}_0$, is asymptotically stable, inclusion of physically realistic time dependence of the zonal wind results in a nonautonomous system

[3] The Frobenius norm of a matrix is also equal to the rms of the singular values of the matrix.

[4] The optimal growth at time, t, is defined as the maximum perturbation growth that can occur over time t. For an autonomous system, governed by $\mathbf{A}_0$, the optimal growth at t is given by the largest singular value of $e^{\mathbf{A}_0 t}$ or by $\|e^{\mathbf{A}_0 t}\|_2$.

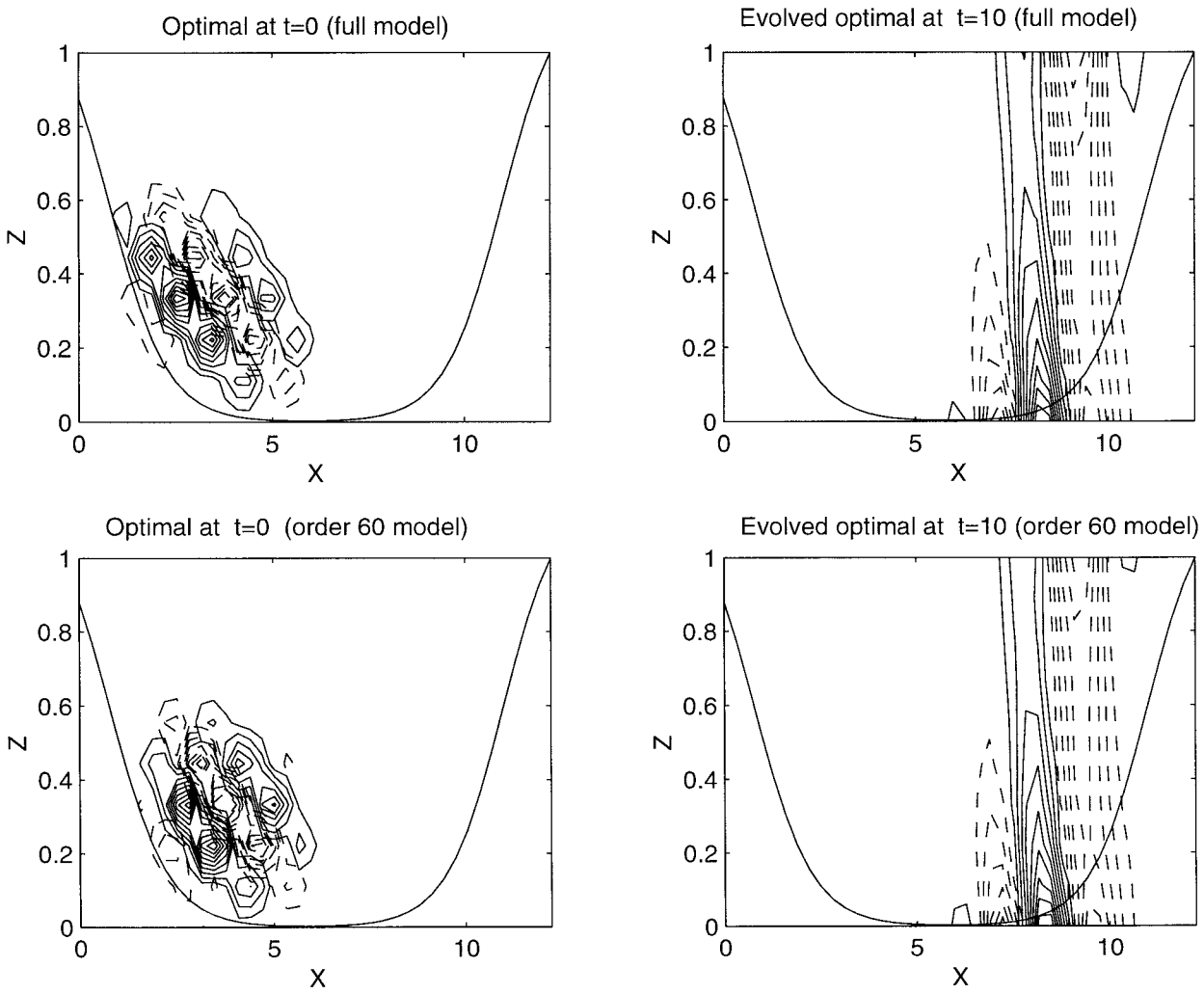

FIG. 10. (left) The streamfunction of the optimal perturbation leading to the greatest square streamfunction growth at $t = 10$, and (right) the evolved optimal streamfunction, at the optimizing time $t = 10$. (top) Streamfunction for the full system; (bottom) the corresponding streamfunction for the order-60 balanced truncation.

governed by the Lyapunov unstable composite operator, $\mathbf{A}_0 + \mathbf{A}_1(t)$, which models the tangent linear forecast system (Farrell and Ioannou 1999; GRE). We attempt to recover the state of this Lyapunov unstable system in order to test the accuracy of the reduced-order Kalman filter. Because the system $\mathbf{A}_0 + \mathbf{A}_1(t)$ is sure the system (11) has Gaussian probability density function under the Gaussian stochastic forcing, η, and the Kalman filter is provably optimal.[5]

We seek to perform state estimation on the time-dependent storm track model introduced in the previous section with damping parameter $\mu = 1.5$ in (25). This system is asymptotically unstable with Lyapunov exponent $\lambda \approx 0.075$.

[5] If $\mathbf{A}(t)$ were uncertain, then the probability density function of the streamfunction would not necessarily remain Gaussian and the Kalman filter could be suboptimal.

We assume for the example that the model error in the true system has error covariance $\mathcal{Q} = q^2\mathbf{I}_N$, where $\mathbf{I}_N$ is the $N \times N$ identity, and $q = 0.075$. Four observations of the true system are taken at longitudes $x = 4.7$ and $x = 6.9$ and at heights $z = 0.1$ and $z = 0.78$ at time intervals $\tau = 0.1$. The observations are noisy with variance $r = 2 \times 10^{-3}$, corresponding to a dimensional rms velocity error of 5 m s^{-1}. The observational errors are uncorrelated and the observation error covariance is therefore $\mathcal{R} = r\mathbf{I}_4$, where $\mathbf{I}_4$ is the 4×4 identity.

A Kalman gain is computed using (16)–(18) in order to estimate the true state using (14) and (15). The initial estimate of the true state is the zero state. The initial estimate of the covariance matrix, $\mathcal{P}(0)$, in (17) is not important if the model error is full rank. We assume that initially we do not possess any knowledge of the structure of the error covariance and take

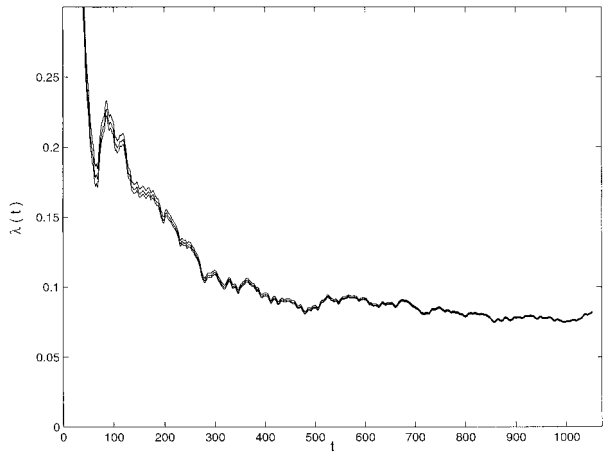

FIG. 11. The growth rate of a random initial perturbation as a function of time. For large times the growth rate of any perturbation approaches the first Lyapunov exponent, which for the time-dependent flow considered is approximately $\lambda = 0.075$. Also shown is the growth rate as identified with a full Kalman filter and with the order-40 reduced Kalman filter.

$$\mathcal{P}(0) = \alpha \mathbf{C}^{\dagger}\mathbf{C}, \tag{35}$$

where $\mathbf{C}$ is a random matrix with normalized columns so that the variance of $\mathbf{C}$ is N, where N is the dimension of the full system. We consider the case $\alpha = 0.1$.

The reduced-order Kalman filter is obtained using an order-40 balanced truncation[6] of the time-dependent equations with the balancing transformation derived for the time-independent mean operator, $\mathbf{A}_0$. The initial estimate of the true state is the zero state and the initial error covariance in the reduced space, $\mathcal{P}^k(0)$, is the initial error covariance assumed for the full-order Kalman filter (35) projected onto the reduced space, that is, $\mathcal{P}^k(0) = \mathbf{Y}\mathcal{P}(0)\mathbf{Y}^{\dagger}$.

First we compare the growth rate of a random initial perturbation in the true system with the growth rate obtained in the analyzed system using a full Kalman filter, and using an order-40 reduced-order Kalman filter.

[6] In the previous section we used as an example of the method of balanced truncation a model storm track (with a time-independent mean zonal flow) retaining 60 degrees of freedom. However, we found that the Kalman filter can successfully identify the state of the system with balanced truncation at order 40.

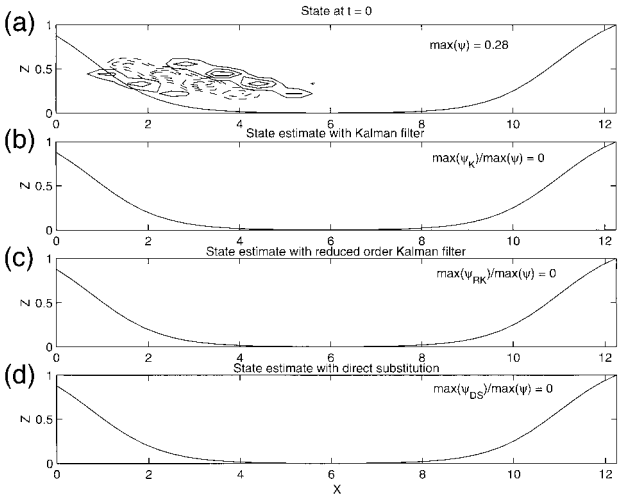

FIG. 12. (first panel) The state at $t = 0$. (second, third, and fourth panels) The initial estimate before observations are entrained (defined to be the zero state). Observations are made at longitudes $x = 4.7$ and $x = 6.9$, and at heights $z = 0.1$ and at $z = 0.78$, every 0.1 units of time. The solid curve indicates the structure of the Rayleigh damping coefficient $r(x)$.

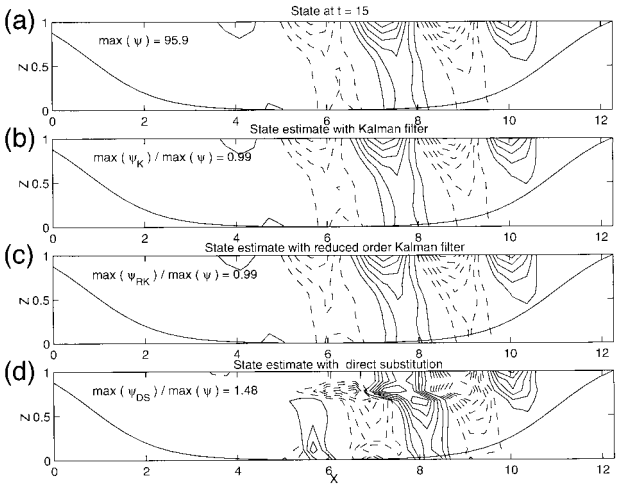

FIG. 13. (a) Streamfunction of the true state at $t = 15$. (b) Streamfunction of the analyzed state obtained using a full Kalman filter. (c) Streamfunction of the analyzed state obtained using a Kalman filter calculated from a balanced truncation of the full system to 40 degrees of freedom. (d) Streamfunction obtained by direct substitution.

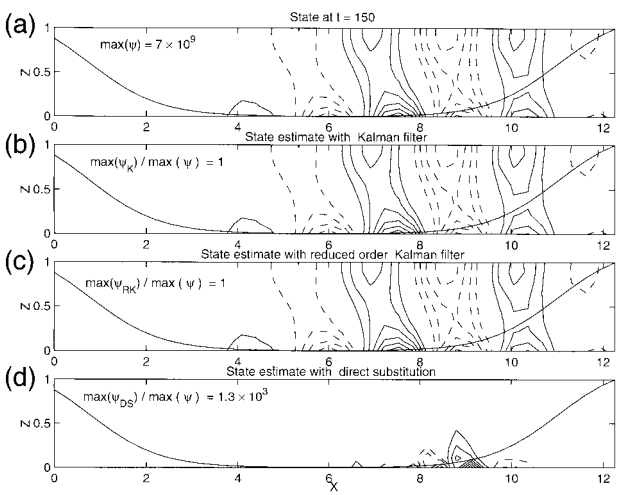

FIG. 14. (a) Streamfunction of the true state at $t = 150$. The state is increasing with Lyapunov exponent $\lambda \approx 0.075$. (b) Streamfunction of the analyzed state obtained using a full Kalman filter. (c) Streamfunction of the analyzed state obtained using a Kalman filter calculated from a balanced truncation of the full system to 40 degrees of freedom. (d) Streamfunction obtained by direct substitution.

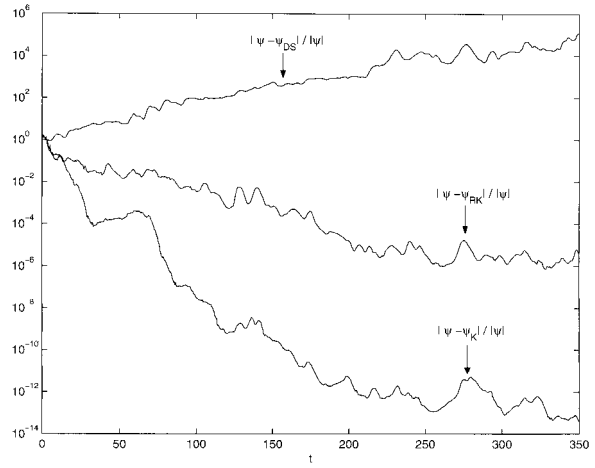

FIG. 15. The relative error of the Kalman filter estimate, ψ_K, the order-40 reduced-order Kalman filter estimate, ψ_{RK}, and the direct substitution estimate, ψ_{DS}, as a function of time. The relative error is defined as the Euclidean distance between the true state and its estimate normalized by the Euclidean length of the true state.

In Fig. 11 we plot the three growth rates, $\lambda(t)$, obtained by three systems, all initialized with the same random initial perturbation ψ_0. The growth rate at time t is defined as

$$\lambda(t) = \frac{1}{t}\ln\left(\frac{|\psi(t)|}{\psi_0}\right), \tag{36}$$

where $\psi(t)$ is the streamfunction at time t. The growth rate curves plotted in Fig. 11 are indistinguishable and all three systems converge to the same Lyapunov exponent, λ. Convergence to the Lyapunov exponent of the true system by the analyzed system using a full-order Kalman filter and by the analyzed system using an order-40 reduced-order Kalman filter demonstrates that the reduced-order Kalman filter can reproduce the growth of perturbations in the full-order time-dependent system.

We investigate now the convergence in structure. The true system starts at $t = 0$ with an optimal perturbation, ψ_{opt}, that gives optimal growth in streamfunction at $t_{\text{opt}} = 10$ when the system is advanced to t_{opt} by the time mean operator $\mathbf{A}_0$; that is, the initial state of the error system is the first right singular vector of $e^{\mathbf{A}_0 t_{\text{opt}}}$. For comparison we also include an estimation of the error state based on direct substitution of the true state at the observational sites.

In the four panels of Fig. 12 are shown the true state (top panel), and the three initial state estimates. By $t = 5$ both the Kalman filter based on integrating the full error covariance matrix and the order-40 reduced-order Kalman filter give a good estimate of the true state, while direct substitution is unable to give a good estimate. By $t = 15$ the state has been almost perfectly estimated by both the Kalman filter and the reduced Kalman filter, as can be seen in Fig. 13. The good estimation is preserved asymptotically as can be seen in Fig. 14. The estimate by direct substitution on the other hand remains poor.

The estimation error, defined as the rms difference between the estimated state and the true state, normalized by the magnitude of the true state is shown as a function of time in Fig. 15. Both the Kalman filters are seen to produce analyses that are very accurate, while direct substitution performs poorly.

5. Discussion and conclusions

Optimal utilization of observing resources requires that the structure of the time-dependent error be taken into account in identifying the state. The error covariance matrix contains the required information but the high dimension of the forecast system precludes directly obtaining it. In this work we described a method for obtaining an approximate error covariance and an approximate state identification using a Kalman filter based on balanced truncation of the tangent linear forecast error system.

As a first example balanced truncation was used to reduce an autonomous forecast error model of order 400 to order 60. The truncated model provides an accurate approximation to the dynamics as measured by the maximum error in the system frequency response. Both the optimal perturbation evolution and the dominant EOFs and SOs were also found to be well approximated. A Kalman filter for the full system obtained using a Kalman gain obtained from the reduced system with 40 degrees of freedom was found to be as effective in observing the state.

However, the error dynamics of an atmospheric forecast is controlled by the tangent linear system where the linearization has been performed about the time-dependent forecast trajectory. Because the forecast error system is time dependent it is important to extend the system approximation method to time-dependent systems. This was done by performing a second experiment on a time-dependent system in which the balanced truncation derived from the mean operator was applied to reduce the order of the full time-dependent operator to order 40 over the forecast period.

Comparison of the performance of a full Kalman filter and approximate filters obtained by balanced truncation on the order-400 storm track model reveals that truncation at order 40 is sufficient to provide accurate flow-dependent covariances for the purpose of approximating the Kalman gain.

Acknowledgments. We thank Athanasios Antoulas and Paul van Dooren for valuable exchanges on reduced-order systems and Michael Ghil for his reviewing comments. This work was supported by NSF Grants ATM-9623539 and by ONR N00014-99-1-0018.

APPENDIX

Calculation of the Balancing Transformation

In balanced truncation the transformation of coordinates implied by the biorthogonal bases $\mathbf{X}$ and $\mathbf{Y}$ is such as to simultaneously diagonalize matrices $\mathbf{P}$ and $\mathbf{Q}$, which have as their eigenfunction the EOFs and the SOs, respectively. We follow the method of constructing the biorthogonal basis $\mathbf{X}$ and $\mathbf{Y}$ given by van Dooren (2000). Alternative methods of balancing and further details can be found in the seminal papers of Moore (1981) and Glover (1984), in the textbook by Zhou and Doyle (1998), or in FI01.

In the transformed coordinates the covariance matrix becomes $\mathbf{P}_k = \mathbf{Y}^\dagger\mathbf{P}\mathbf{Y}$ and the stochastic optimal matrix becomes $\mathbf{Q}_k = \mathbf{X}^\dagger\mathbf{Q}\mathbf{X}$. The balancing transformations $\mathbf{Y}$ and $\mathbf{X}$ are found by decomposing the positive definite covariance matrix, $\mathbf{P}$, and stochastic optimal matrix, $\mathbf{Q}$, as follows:

$$\mathbf{P} = \mathbf{S}^\dagger\mathbf{S}, \qquad \mathbf{Q} = \mathbf{R}^\dagger\mathbf{R}, \tag{A1}$$

which requires a Cholesky factorization of $\mathbf{P}$ and $\mathbf{Q}$ (Golub and van Loan 1996; van Dooren 2000). The singular value decomposition of the product $\mathbf{S}\mathbf{R}^\dagger$ is then obtained:

$$\mathbf{S}\mathbf{R}^\dagger = \mathbf{U}\Sigma\mathbf{V}^\dagger, \tag{A2}$$

in which the diagonal elements of Σ are the Hankel singular values, h, that is, the square roots of the eigenvalues of $\mathbf{P}\mathbf{Q}$.

A k-order truncation is formed by retaining the first k columns of $\mathbf{U}$, denoted by $\mathbf{U}_k$; the first k columns of $\mathbf{V}$, denoted by $\mathbf{V}_k$; and the diagonal matrix consisting of the top k Hankel singular values, denoted by Σ_k and defining the projection matrices as

$$\mathbf{Y} = \mathbf{R}^\dagger\mathbf{V}_k\Sigma_k^{-1/2} \quad \text{and} \quad \mathbf{X} = \mathbf{S}^\dagger\mathbf{U}_k\Sigma_k^{-1/2}, \tag{A3}$$

whose columns form a biorthogonal basis, that is, $\mathbf{Y}^\dagger\mathbf{X} = \mathbf{I}_k$, where $\mathbf{I}_k$ is the identity matrix of order k. These transformation matrices determine the k-order balancing transformation from which the reduced k-order system (6) is derived.

REFERENCES

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.,* **129,** 420–436.

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

Cohn, S. E., and R. Todling, 1996: Approximate Kalman filters for stable and unstable dynamics. *J. Meteor. Soc. Japan,* **74,** 63–75.

Daley, R. A., 1991: *Atmospheric Data Analysis.* Cambridge University Press, 457 pp.

Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.,* **123,** 1128–1145.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99** (C5), 10 143–10 162.

Farrell, B. F., and P. J. Ioannou, 1996: Generalized stability. Part I: Autonomous operators. *J. Atmos. Sci.,* **53,** 2025–2041.

——, and ——, 1999: Perturbation growth and structure in time-dependent flows. *J. Atmos. Sci.,* **56,** 3622–3639.

——, and ——, 2001: Accurate low-dimensional approximation of the linear dynamics of fluid flow. *J. Atmos. Sci.,* **58,** 2771–2789.

Fukumori, I., and P. Malanotte-Rizzoli, 1995: An approximate Kalman filter for ocean data assimilation: An example with an idealized Gulf Stream model. *J. Geophys. Res.,* **100C,** 6777–6793.

Gelaro, R., C. Reynolds, and R. M. Errico, 2000: Transient and asymptotic perturbation growth in a simple model. *J. Atmos. Sci.,* submitted.

Ghil, M., 1997: Advances in sequential estimation for atmospheric and oceanic flows. *J. Meteor. Soc. Japan,* **75,** 289–304.

——, and P. Malanotte-Rizzoli, 1991: Data assimilation in meteorology and oceanography. *Advances in Geophysics,* Vol. 33, Academic Press, 141–266.

——, and R. Todling, 1996: Tracking atmospheric instabilities with the Kalman filter. Part II: Two layer results. *Mon. Wea. Rev.,* **124,** 2340–2352.

——, S. Cohn, J. Tavantzis, K. Bube, and E. Isaacson, 1981: Applications of estimation theory to numerical weather prediction. *Dynamic Meteorology: Data Assimilation Methods,* L. Bengtsson, M. Ghil, and E. Kallen, Eds., Springer-Verlag, 139–224.

Glover, K., 1984: An optimal Hankel-norm approximation of linear multivariable systems and their Ł$^\infty$-error bounds. *Int. J. Control,* **39,** 1115–1193.

Golub, G. H., and C. F. van Loan, 1996: *Matrix Computations,* The Johns Hopkins University Press, 694 pp.

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

Ide, K., and M. Ghil, 1997: Extended Kalman filtering for vortex systems. Part I: Methodology and point vortices. *Dyn. Atmos. Oceans,* **27,** 301–332.

——, P. Courtier, M. Ghil, and A. C. Lorenc, 1997: Unified notation for data assimilation: Operational, sequential, and variational. *J. Meteor. Soc. Japan,* **75,** 181–189.

Illyashenko, Yu. S., 1983: On the dimension of attractors of k-contracting systems in infinite-dimensional space. *Vestn. Mosk. Univ. Ser. 1: Mat. Mekh.,* **3,** 52–58.

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

Kaplan, J. L., and J. A. Yorke, 1979: Preturbulence: A regime observed in a fluid flow model of Lorenz. *Commun. Math. Phys.,* **67,** 93–108.

Lermusiaux, P. F. J., and A. R. Robinson, 1999: Data assimilation via error statistical estimation. Part I: Theory and schemes. *Mon. Wea. Rev.,* **127,** 1385–1407.

Miller, R. N., M. Ghil, and F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.,* **51,** 1037–1056.

Moore, B. C., 1981: Principal component analysis in linear systems: controllability, observability, and model reduction. *IEEE Trans. Autom. Control,* **AC-26,** 17–31.

North, G., 1984: Empirical orthogonal functions and normal modes. *J. Atmos. Sci.,* **41,** 879–887.

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

Reynolds, C. A., and R. M. Errico, 1999: Convergence of singular vectors toward Lyapunov vectors. *Mon. Wea. Rev.,* **127,** 2309–2323.

Stewart, G. W., and J.-G. Sun, 1990: *Matrix Perturbation Theory.* Academic Press, 365 pp.

Tippett, M. K., S. E. Cohn, R. Todling, and D. Marchesin, 2000: Low-dimensional representation of error covariance. *Tellus,* **52,** 533–553.

Todling, R., and M. Ghil, 1994: Tracking atmospheric instabilities with the Kalman filter. Part I: Methodology and one-layer results. *Mon. Wea. Rev.,* **122,** 183–204.
van Dooren, P., 2000: Gramian based model reduction of large-scale dynamical systems. *Numerical Analysis 1999,* Chapman and Hall/CRC Press, 231–247.
Verlaan, M., and A. W. Heemink, 1997: Tidal flow forecasting using reduced rank square root filters. *Stochastic Hydrol. Hydraul.,* **11,** 349–368.
Wunsch, C., 1996: *The Ocean Circulation Inverse Problem.* Cambridge University Press, 442 pp.
Zhou, K., and J. C. Doyle, 1998: *Essentials of Robust Control.* Prentice-Hall, 411 pp.