*Tellus (2005), 57A, 12–20*
*Printed in UK. All rights reserved*


**TELLUS**

# On the prediction of linear stochastic systems with a low-order model

*By* FAMING WANG* and ROBERT SCOTT, *Institute for Geophysics, University of Texas at Austin, TX 78759, USA*

(Manuscript received 24 December 2003; in final form 21 June 2004)

ABSTRACT

Three methods for approximating the high-dimensional stochastic system with a low-dimensional model are examined, and the prediction error and predictability of the reduced-order models are evaluated. It is shown that during reduction both the normal modes of deterministic dynamics and the spatial structures of stochastic forcing need to be taken into account. In addition to stability, which determines the asymptotic behavior, non-normality, which controls the error growth at short lead times, should also be preserved. An experiment with tropical Atlantic variability illustrates that the empirical orthogonal function and balanced truncation are superior to modal reduction in capturing the predictable dynamics.

## 1. Introduction

Geophysical fluids form a complex dynamical system, which usually possesses many degrees of freedom and poses a challenge for simulation and analysis. One standard strategy is to project the full system into a subspace with fewer dimensions, and to study the principal dynamical properties of the complex system with the aid of a low-order model. This is known as a model reduction problem (Obinata and Anderson 2000), which, in dynamical system theory, is to find the finite-dimensional manifold in the infinite-dimensional phase space (Holmes et al. 1996; Patil et al. 2001). A successful example is Lorenz's work on atmospheric predictability (Lorenz 1965, 1969). Using double-Fourier series as basis, Lorenz derived low-order atmospheric models and found that useful weather forecasts were limited to two weeks. Even with today's operational numerical weather prediction, this limit still holds (Kalnay et al. 1998).

While spherical harmonics are the default basis functions of global atmospheric models, they are not efficient at describing large-scale atmospheric dynamics. For example, the behavior of a barotropic T21 model with 231 degrees of freedom is well reproduced by a low-order model with only tens of degrees of freedom (Selten 1995; Kwasniok 2004). To obtain that compact description, however, an adaptive projection needs to be employed, i.e. the basis functions are derived specifically for the system at hand rather than specified beforehand. One such method is the empirical orthogonal function (EOF) truncation, in which the total variance is mostly explained by a few dominant patterns of variability. This has been used in the reduction of El Niño Southern Oscillation (ENSO) dynamics (Timmermann et al. 2001; Roulston and Neelin 2003), barotropic and baroclinic shear flows (Selten 1995, 1997), mid-latitude storm tracks (Zhang and Held 1999), and atmospheric low-frequency variability (Achatz and Branstator 1999; D'Andrea and Vautard 2001; Winkler et al. 2001). An alternative method is the principal interaction pattern (PIP) reduction proposed by Hasselmann (1988). This is designed to reveal the internal structure of the system, and offers a substantial improvement over the reduction based on EOFs (Achatz and Schmitz 1997; Kwasniok 2004; Crommelin and Majda 2004) if the low-order model is properly closed for the neglected interactions. Recently, Farrell and Ioannou (2001a,b) introduced yet another method, the balanced truncation (BT), into atmospheric science and showed its potential use in model reduction and data assimilation.

Depending on the dynamical system of interest and the characteristics of concern, all these reduction methods have their strengths and weaknesses (Antoulas and Sorensen 2001). Therefore, a quantitative understanding of the reduction error is crucial in assessing and interpreting the results of the low-order model. Farrell and Ioannou (2001a) examined the performance of various methods for reducing the dimension of linear non-normal systems, and found that accurate reduction requires BT. In this paper, we take a further look at this problem from a practical perspective, namely the low-order model is used to predict the dynamical system (e.g. Penland and Magorian 1993; Xue et al. 1994; Johnson et al. 2000) and estimate the predictability in truncated space (e.g. Blumenthal 1991; Schneider and Griffies

*Corresponding author.
e-mail: fwang@ig.utexas.edu

1999; Wang 2001). A good reduction means that the prediction error is small and the predictability estimate is accurate. This is achieved by choosing optimal basis of projection and appropriate parametrization of the neglected dynamics. While applying a closure scheme may produce a better low-order representation of the full dynamics (Achatz and Schmitz 1997; Selten 1997; Majda et al. 2003), such effort is not attempted in this study.

Within the framework of linear stochastic dynamics (see Farrell and Ioannou 1996a,b; Chang et al. 2004a,b), the objective of this work is to investigate the effect of model reduction on the prediction and predictability estimate of the full system. The paper is organized as follows. First, the problem is formulated in Section 2, where the prediction error and predictability are defined and three reduction methods are briefly reviewed. Then, the reduction error is discussed in Section 3 for system with EOFs representing physical modes, and in Section 4 for EOFs distinct from physical modes. Finally, some concluding remarks and discussion are given in Section 5.

## 2. Reduction of linear stochastic dynamics

### 2.1. Prediction and predictability of a linear stochastic system

In this paper, we assume that the dynamical system of interest is fully described by the linear time-invariant model

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathsf{A}\mathbf{x} + \boldsymbol{\eta}, \tag{1}$$

where $\mathbf{x}$ is an $n$-dimensional vector representing the state of the system, $\mathsf{A}$ is an $n \times n$ matrix describing the deterministic dynamics, and $\boldsymbol{\eta}$ is a forcing vector denoting the influence of unresolved processes and other external perturbations. It is further assumed that $\mathsf{A}$ is stable, i.e. all its eigenvalues have negative real parts, and that $\boldsymbol{\eta}$ is a Gaussian white noise, i.e. the time evolution of the forcing is totally unpredictable. Then the linear stochastic model (1) describes a stationary response $\mathbf{x}$ whose statistical characteristics can be solved analytically. In the following, only the main results about system (1) are reviewed; the interested reader is referred to the papers by Farrell and Ioannou (1996a,b) and Chang et al. (2004a,b) for detailed derivation.

Given the initial state $\mathbf{x}(t_0)$, the solution of eq. (1) is

$$\mathbf{x}(t_0+\tau) = \mathrm{e}^{\mathsf{A}\tau}\mathbf{x}(t_0) + \int_0^\tau \mathrm{e}^{\mathsf{A}(\tau-s)}\boldsymbol{\eta}(s+t_0)\,\mathrm{d}s, \tag{2}$$

where $\tau$ is the lead time. The first term on the right-hand side represents the effect of initial condition. The second term, independent of the first, represents the influence of random noise forcing. If system (1) evolves long enough to approach a statistically steady state, then the covariance matrix of $\mathbf{x}$ satisfies the Lyapunov equation

$$\mathsf{A}\mathsf{C} + \mathsf{C}\mathsf{A}^\dagger = -\Sigma, \tag{3}$$

where $\mathsf{C} = \langle \mathbf{x}\mathbf{x}^\dagger \rangle$, $\Sigma = \langle \eta\eta^\dagger \rangle$, $(\cdot)^\dagger$ denotes the complex conjugate transpose, and $\langle\cdot\rangle$ means the ensemble average. Physically, eq. (3) represents the energy balance between the fluctuation and the dissipation. Mathematically, it determines the stability of eq. (1) via the Lyapunov theorem: $\mathsf{A}$ is stable if and only if there exists a positive definite $\mathsf{C}$ for some positive definite $\Sigma$. One observation of eq. (3) is that only for a very special case, such as a normal system ($\mathsf{A}\mathsf{A}^\dagger = \mathsf{A}^\dagger\mathsf{A}$) under unitary forcing ($\Sigma = \mathsf{I}$), can the eigenanalysis of $\mathsf{C}$ be used to reveal the physical modes of a dynamical system.

The effect of stochastic forcing is also illustrated by its influence on the total variance

$$\mathrm{var}\{\mathbf{x}\} = \mathrm{tr}\{\mathsf{C}\} = \mathrm{tr}\{\langle \boldsymbol{\eta}^\dagger \mathsf{B}\boldsymbol{\eta}\rangle\}, \tag{4}$$

where the stochastic dynamical operator $\mathsf{B}$ is determined via another Lyapunov equation

$$\mathsf{A}^\dagger\mathsf{B} + \mathsf{B}\mathsf{A} = -\mathsf{I}. \tag{5}$$

The eigenvector of $\mathsf{B}$ defines the optimal structure of $\boldsymbol{\eta}$, and the eigenvalue measures the variance excited by $\boldsymbol{\eta}^\dagger\,\boldsymbol{\eta} = 1$ (Farrell and Ioannou 1996a). Because these forcing patterns are most effective in exciting stationary response, they are referred to as stochastic optimals. Generally, the stochastic optimals, EOFs (eigenvectors of covariance matrix $\mathsf{C}$) and normal modes (eigenvectors of dynamical operator $\mathsf{A}$) are all distinct.

Knowing the initial condition $\mathbf{x}(t_0)$, we can make a prediction of $\mathbf{x}(t_0+\tau)$ based on a deterministic forecasting model $P[\tau, \mathbf{x}(t_0)]$. Then the normalized prediction error is

$$\begin{aligned}\epsilon_{\mathrm{p}}(\tau) &= \frac{\mathrm{var}\{\mathbf{x}(t_0+\tau) - P[\tau,\mathbf{x}(t_0)]\}}{\mathrm{var}\{\mathbf{x}\}} \\ &= \frac{\mathrm{var}\left\{\mathrm{e}^{\mathsf{A}\tau}\mathbf{x}(t_0) - P[\tau,\mathbf{x}(t_0)]\right\}}{\mathrm{var}\{\mathbf{x}\}} + \varepsilon(\tau),\end{aligned} \tag{6}$$

where

$$\varepsilon(\tau) = 1 - \frac{\mathrm{tr}\{\mathrm{e}^{\mathsf{A}\tau}\mathsf{C}\mathrm{e}^{\mathsf{A}^\dagger\tau}\}}{\mathrm{tr}\{\mathsf{C}\}} \tag{7}$$

represents the uncertainty due to stochastic forcing, which is independent of the prediction scheme $P$. Obviously, the minimal prediction error $\epsilon_{\mathrm{p}} = \varepsilon$ is achieved by the optimal prediction $P[\tau, \mathbf{x}(t_0)] = \bar{\mathbf{x}}(t_0,\tau) = \mathrm{e}^{\mathsf{A}\tau}\mathbf{x}(t_0)$. However, unlike $\epsilon_{\mathrm{p}}$, which measures the performance of a prediction scheme, $\varepsilon$ is an inherent property of eq. (1) and measures its predictability.

To illustrate the role of coherent forcing $\boldsymbol{\eta}$ and dynamical operator $\mathsf{A}$, an upper bound and a lower bound for predictability are given here (see the appendix for the proof)

$$1 - \mathrm{e}^{\lambda_n(\mathsf{A}+\mathsf{A}^\dagger)\tau} \geq \varepsilon(\tau) \geq 1 - \mathrm{e}^{\lambda_1(\mathsf{A}+\mathsf{A}^\dagger)\tau}, \tag{8}$$

where $\lambda_i(\mathsf{X})$ denotes the $i$th eigenvalue of $\mathsf{X}$ in descending order of the real parts. For a normal system, $\lambda(\mathsf{A}+\mathsf{A}^\dagger) = \lambda(\mathsf{A}) + \lambda^*(\mathsf{A})$, inequality (8) tells us that its predictability is bounded by the skills of the most and least-damped normal modes. Also, the maximum and minimum predictabilities are achieved when

forcing takes the structure of the first and last eigenvectors of operator $\mathsf{A}$. For a non-normal system, we have $\lambda_n(\mathsf{A}) + \lambda_n{}^*(\mathsf{A}) \geq \lambda_n(\mathsf{A}+\mathsf{A}^\dagger)$ and $\lambda_1(\mathsf{A}) + \lambda_1{}^*(\mathsf{A}) \leq \lambda_1(\mathsf{A}+\mathsf{A}^\dagger)$. Therefore, eq. (8) essentially says that a non-normal system could be more predictable than its least-damped mode, or less predictable than its most damped mode. In other words, depending on the setting of forcing structure, non-normality can either increase or decrease the predictability. For unitary forcing, Tippett and Chang (2003) have shown that non-normality always increases predictability in terms of predictive information, whereby the lower bound for predictability is given by normal dynamics with the same eigenvalues.

### 2.2. *Methods for reducing the order of a linear stochastic system*

Through linear transformation $\mathbf{x} = \mathsf{T}\boldsymbol{\alpha}$, eq. (1) is equal to

$$\frac{\mathrm{d}\boldsymbol{\alpha}}{\mathrm{d}t} = \widetilde{\mathsf{A}}\boldsymbol{\alpha} + \boldsymbol{\zeta}, \tag{9}$$

where $\boldsymbol{\alpha} = \mathsf{T}^{-1}\mathbf{x}$, $\widetilde{\mathsf{A}} = \mathsf{T}^{-1}\mathsf{A}\mathsf{T}$, and $\boldsymbol{\zeta} = \mathsf{T}^{-1}\boldsymbol{\eta}$. Suppose, according to some criteria, the first $m$ $\alpha$ are far more important than the rest, then an $m$-dimensional low-order model can be constructed by neglecting the unimportant $\alpha$

$$\frac{\mathrm{d}\boldsymbol{\alpha}_{m\times 1}}{\mathrm{d}t} = \widetilde{\mathsf{A}}_{m\times m}\boldsymbol{\alpha}_{m\times 1} + \boldsymbol{\zeta}_{m\times 1}, \tag{10}$$

where $\boldsymbol{\alpha}_{m\times 1} \cong \mathsf{L}_{m\times n}\mathbf{x}_{n\times 1}$, $\widetilde{\mathsf{A}}_{m\times m} = \mathsf{L}_{m\times n}\mathsf{A}_{n\times n}\mathsf{R}_{n\times m}$, and $\zeta_{m\times 1} = \mathsf{L}_{m\times n}\eta_{n\times 1}$. While the left projector $\mathsf{L}_{m\times n}$ is the first $m$ rows of $\mathsf{T}^{-1}$, the right projector $\mathsf{R}_{n\times m}$ is the first $m$ columns of $\mathsf{T}$, and they satisfy $\mathsf{L}\mathsf{R} = \mathsf{I}_{m\times m}$. The low-order model (10) may be thought of as a projection (or truncation) of the high-order model (1) into subspace $\mathsf{R}$. Back into original space, eq. (10) describes a stochastic process $\mathsf{R}\boldsymbol{\alpha}$ which serves as an approximation of $\mathbf{x}$. In the following, eq. (1) is referred to as the full-order model (FOM), and eq. (10) as the reduced-order model (ROM).

Within the context of this study, two intimately related questions can be asked.

First, if we use the ROM to predict the full system, how good is it? To answer this, we define the normalized prediction error in physical space

$$\begin{aligned}\epsilon_{\mathrm{p}}(\tau) &= \frac{\mathrm{var}\{\mathbf{x} - \mathsf{R}\bar{\boldsymbol{\alpha}}\}}{\mathrm{var}\{\mathbf{x}\}} \\ &= \frac{\mathrm{var}\{\mathrm{e}^{\mathsf{A}\tau}\mathbf{x}(t_0) - \mathsf{R}\mathrm{e}^{\widetilde{\mathsf{A}}\tau}\mathsf{L}\mathbf{x}(\mathsf{t}_0)\}}{\mathrm{var}\{\mathbf{x}\}} + \varepsilon(\tau),\end{aligned} \tag{11}$$

where $\bar{\boldsymbol{\alpha}} = \mathrm{e}^{\widetilde{\mathsf{A}}\tau}\boldsymbol{\alpha}$ is the prediction in reduced space (10). The first term on the right-hand side is the error due to reduction, and the second term $\varepsilon$ is the predictability of the FOM (1). As we know, smaller $\epsilon_{\mathrm{p}}$ means better prediction, hence better model reduction.

Secondly, do the FOM and ROM have similar predictability? A positive answer will ensure the use of the low-order model as a substitute for studying the predictability of a high-order complex system. Noticing that error variance depends on coordinates, we define the ROM's predictability as the ratio of variance in physical space (space of FOM)

$$\varepsilon_r(\tau) = \frac{\mathrm{var}\{\mathsf{R}\boldsymbol{\alpha} - \mathsf{R}\bar{\boldsymbol{\alpha}}\}}{\mathrm{var}\{\mathsf{R}\boldsymbol{\alpha}\}} = 1 - \frac{\mathrm{tr}\{\mathsf{R}\mathrm{e}^{\widetilde{\mathsf{A}}\tau}\widetilde{\mathsf{C}}\mathrm{e}^{\widetilde{\mathsf{A}}^\dagger\tau}\mathsf{R}^\dagger\}}{\mathrm{tr}\{\mathsf{R}\widetilde{\mathsf{C}}\mathsf{R}^\dagger\}} \tag{12}$$

where $\widetilde{\mathsf{C}} = \langle\boldsymbol{\alpha}\boldsymbol{\alpha}^\dagger\rangle$ is determined via the Lyapunov equation: $\widetilde{\mathsf{A}}\widetilde{\mathsf{C}} + \widetilde{\mathsf{C}}\widetilde{\mathsf{A}}^\dagger = -\widetilde{\Sigma}$ with $\widetilde{\Sigma} = \langle\boldsymbol{\zeta}\boldsymbol{\zeta}^\dagger\rangle = \mathsf{L}\Sigma\mathsf{L}^\dagger$. A good reduction requires the ROM's predictability $\varepsilon_r(\tau)$ close to the FOM's predictability $\varepsilon(\tau)$.

Mathematically, the above two problems can be written as variations, min $\epsilon_{\mathrm{p}}(\tau)$ and min $|\varepsilon_r(\tau) - \varepsilon(\tau)|$, as they depend on the basis function $\mathsf{R}$. Unfortunately, we do not have definite answers for them unless the system at hand is very special. However, we do have an answer for the initial error $\epsilon_{\mathrm{p}}(0)$. From eq. (11), we know that minimizing $\epsilon_{\mathrm{p}}(0)$ is equivalent to minimizing $\mathrm{var}\{\mathbf{x} - \mathsf{R}\mathsf{L}\mathbf{x}\}$, which is a total least-squares problem (Huffel and Vandewalle 1991). According to the Eckart–Young–Mirsky theorem (Eckart and Young 1936; Mirsky 1960), the optimal projection at $\tau = 0$ is given by the first $m$ EOFs of $\mathbf{x}$, i.e. eigenvectors of the covariance matrix $\mathsf{C}$. In other words, EOF truncation preserves more variance of $\mathbf{x}$ than any other projections for the same degree of reduction $m$. As demonstrated in the next two sections, however, for non-zero lead times the EOF is no longer the optimal basis.

Instead of searching for the optimal basis, in this paper we just compare the performance, measured by $\epsilon_{\mathrm{p}}$ and $\varepsilon_r$, of three common model reduction methods, as follows.

*2.2.1. Modal reduction.* The use of normal modes as basis functions is called modal reduction. It separates the $n$-dimensional system into $n$ unrelated one-dimensional systems, which makes the computation much easier ($\widetilde{\mathsf{A}}$ is diagonal). It essentially is a similarity transform, and hence preserves the eigenvalues. The traditional wisdom is to choose the least-damped $m$ modes, with the belief that low modes dominate the long-term response while short-living high modes control the rapid transitions. During reduction, the forcing structure has been totally ignored, which will cause serious consequences.

*2.2.2. EOF truncation.* EOF decomposition is not unfamiliar to atmospheric scientists and oceanographers (Preisendorfer 1988). It is also known as principal component analysis, factor analysis or total-least-squares estimation in statistics and data processing; proper orthogonal decomposition in the context of turbulence (Holmes et al. 1996); and Karhunen–Loève decomposition in mathematical physics. The wide use is partially due to the easy implementation, partially rooted in the firm mathematical ground – the Eckart–Young–Mirsky theorem which guarantees the optimality of EOF. Note that the problem formulated here is different from the definition of the EOF; therefore, the ROM in EOF subspace is only suboptimal at best. Using the orthogonality of the EOF and Lyapunov theorem, however, we can prove that the reduced model is stable and $\widetilde{\mathsf{C}} = \mathrm{diag}[\lambda_i(\mathsf{C})]$,

where $\lambda_i$ are the variances of the first $m$ EOFs (Farrell and Ioannou 2001a).

*2.2.3. Balanced truncation.* Another suboptimal model reduction is the BT method proposed by Moore (1981) in the context of control theory. It has been tested extensively since and has proved superior over most other methods (Obinata and Anderson 2000). It has several useful properties, such as retaining the stability, the truncation error being upper bounded, and extendability to a non-linear system.

Even though BT is a well-known procedure in standard textbooks (Dullerud and Paganini 2000), some details are given here because of its new appearance in our field. The autonomous system (1) can be written as a trivial controlled system:

$$\begin{aligned} \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} &= \mathsf{A}\mathbf{x} + \boldsymbol{\eta} \\ \mathbf{y} &= \mathbf{x}. \end{aligned} \tag{13}$$

With respect to eq. (13), internal balancing is seeking a similarity transformation $\mathsf{T}$ that results in a diagonal covariance matrix and stochastic operator (see the appendix for a diagonalizing procedure)

$$\widetilde{\mathsf{C}} = \widetilde{\mathsf{B}} = \Phi = \mathrm{diag}\,[h_1, h_2, \cdots, h_n] \tag{14}$$

where $\widetilde{\mathsf{C}} = \mathsf{T}^{-1}\mathsf{C}\mathsf{T}^{-\dagger}$, $\widetilde{\mathsf{B}} = \mathsf{T}^{\dagger}\mathsf{B}\mathsf{T}$, and the Hankel singular value $h_1 \geq h_2 \geq \cdots \geq h_n \geq 0$ reflects the relative contribution of the corresponding mode. Using the first $k$ modes to form $\mathsf{L}$ and $\mathsf{R}$, eq. (10) would be the BT version of eq. (1). Here, it is helpful to think of BT as a combination of EOFs and stochastic optimals. Taking both dynamics and forcing into consideration, BT gives better performance in the reduction of fluid flow problem (Farrell and Ioannou 2001a).

## 3. Rduction of simple case: EOFs are normal modes

EOFs represent normal modes of eq. (1) only when the system is normal and the stochastic forcing is uncorrelated in normal-mode space (Tippett and Chang 2003). In other words, if the dynamical operator and forcing covariance have the same set of eigenvectors,

$$\mathsf{A} = \mathsf{V}\,\mathrm{diag}[a_i]\mathsf{V}^{\dagger}, \tag{15}$$

$$\Sigma = \mathsf{V}\,\mathrm{diag}[d_i]\mathsf{V}^{\dagger}, \tag{16}$$

so do the stochastic operator and covariance of response,

$$\mathsf{B} = \mathsf{V}\,\mathrm{diag}[b_i]\mathsf{V}^{\dagger}, \tag{17}$$

$$\mathsf{C} = \mathsf{V}\,\mathrm{diag}[c_i]\mathsf{V}^{\dagger}, \tag{18}$$

where the unitary matrix $\mathsf{V}$ is made up of columns of orthonormal modes such that $0 > \Re(a_1) \geq \Re(a_2) \cdots \geq \Re(a_n)$. According to the Lyapunov eqs. (3) and (5), $b$ and $c$ depend on $a$ and $d$:

$$c_i = \frac{-d_i}{a_i + a_i^*} \quad \text{and} \quad b_i = \frac{-1}{a_i + a_i^*}. \tag{19}$$

Note that $d_i$, hence $c_i$, are not necessarily in descending order.

Now, different model reduction methods are simplified to be different ways of selecting $m$ modes from set $\mathsf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]$. For example, modal reduction use the least-damped modes, i.e. $\mathbf{v}_1, \ldots, \mathbf{v}_m$: EOF reduction picks out the most energetic modes $c_i$, while BT chooses the modes with the largest Hankel singular values $h_i = \sqrt{c_i b_i}$. Generally speaking, let $[l_1, \ldots, l_m]$ be an $m$-combination of $[1, \ldots, n]$, $\mathsf{R} = [\mathbf{v}_{l_1}, \ldots, \mathbf{v}_{l_m}]$, then $\mathsf{L} = \mathsf{R}^{\dagger}$, prediction error of ROM (eq. 11)

$$\epsilon_{\mathrm{p}} = 1 - \frac{\sum_{i=1}^{m} \mathrm{e}^{(a_{l_i} + a_{l_i}^*)\tau} c_{l_i}}{\sum_{i=1}^{n} c_i}, \tag{20}$$

and the predictability of ROM (eq. 12)

$$\varepsilon_r = 1 - \frac{\sum_{i=1}^{m} \mathrm{e}^{(a_{l_i} + a_{l_i}^*)\tau} c_{l_i}}{\sum_{i=1}^{m} c_{l_i}}. \tag{21}$$

If $m = n$, then $\epsilon_{\mathrm{p}} = \varepsilon_r = \varepsilon$. When $m < n$, the minimum prediction error at lead time $\tau$ is achieved if the reduction is through the first $m$ modes with largest variance $\mathrm{e}^{(a_i + a_i^*)\tau} c_i$. Comparing $\varepsilon_r$ with $\varepsilon$, we see that the ROM can either overestimate or underestimate the predictability. The details depend on the eigenspectra of deterministic dynamics $a_i$ and stochastic forcing $d_i$.

One special case, for which we can obtain the upper and lower bounds of model reduction, is spatially-white noise, i.e. $d_1 = d_2 = \cdots = d_n = 1$. Under such forcing, $c_i = b_i$, and all the reduction methods discussed above select the first $m$ modes, and hence give the same result. The reduction via the first $m$ modes gives

$$\begin{aligned} \epsilon_{\mathrm{p}}^{\min} &= 1 - \frac{\sum_{i=1}^{m} \left[\mathrm{e}^{(a_i + a_i^*)\tau}/a_i + a_i^*\right]}{\sum_{i=1}^{n} \left(1/a_i + a_i^*\right)} \qquad \text{and} \\ \varepsilon_r^{\min} &= 1 - \frac{\sum_{i=1}^{m} \left[\mathrm{e}^{(a_i + a_i^*)\tau}/a_i + a_i^*\right]}{\sum_{i=1}^{m} \left(1/a_i + a_i^*\right)}, \end{aligned} \tag{22}$$

while the reduction via the last $m$ modes has

$$\begin{aligned} \epsilon_{\mathrm{p}}^{\max} &= 1 - \frac{\sum_{i=n-m+1}^{n} \left[\mathrm{e}^{(a_i + a_i^*)\tau}/a_i + a_i^*\right]}{\sum_{i=1}^{n} \left(1/a_i + a_i^*\right)} \qquad \text{and} \\ \varepsilon_r^{\max} &= 1 - \frac{\sum_{i=n-m+1}^{n} \left[\mathrm{e}^{(a_i + a_i^*)\tau}/a_i + a_i^*\right]}{\sum_{i=n-m+1}^{n} \left(1/a_i + a_i^*\right)}. \end{aligned} \tag{23}$$

We can easily prove for an arbitrary reduction, orthogonal or oblique,

$$\epsilon_{\mathrm{p}}^{\min} \leq \epsilon_{\mathrm{p}} \leq \epsilon_{\mathrm{p}}^{\max} \quad \text{and} \quad \varepsilon_r^{\min} \leq (\varepsilon, \varepsilon_r) \leq \varepsilon_r^{\max}, \tag{24}$$

where $\epsilon_{\mathrm{p}}$ is the prediction error of the ROM, $\varepsilon_r$ is the predictability of the ROM, and $\varepsilon$ is the predictability of the FOM. Therefore, within this case the usual model reduction (modal, EOF or BT) is the global optimal reduction, in the sense that its prediction error is minimal at every lead time $\tau$, although it always overestimates the predictability.

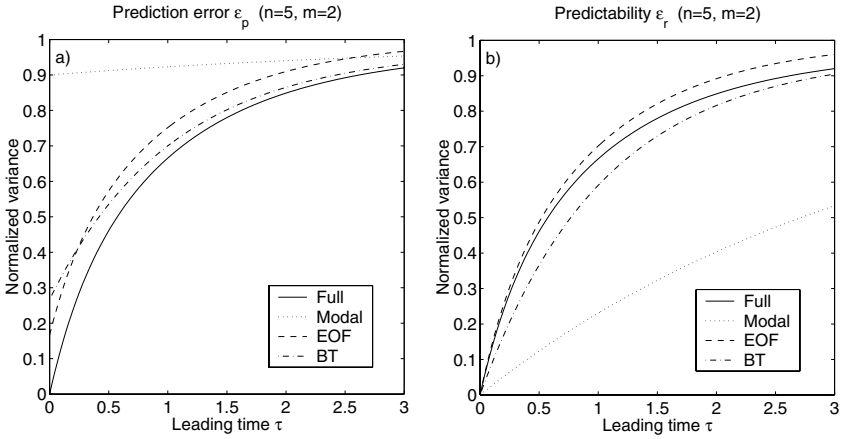


*Fig 1.* Performance of different model reductions for simple normal system: (a) normalized prediction error; (b) predictability of low-order model. The solid lines denote skill of FOM; dotted lines, modal reduction; dashed lines, EOF truncation; dot-dashed lines, BT reduction. Note that the two solid lines in (a) and (b) are identical because we define the normalized error variance of optimal prediction as the predictability.

To illustrate the basic idea, we present an example here.

Let us assume that the eigenvalues of the dynamical operator $\mathsf{A}$ and the stochastic forcing covariance $\Sigma$ are

$$[a_i] = -[0.1, 0.2, 0.5, 1, 2] \quad \text{and} \quad [d_i] = [0.1, 0.1, 5, 1, 5],$$

respectively. Then the skill of modal, EOF and BT reduction can be calculated according to eqs. (20) and (21). The results for the ROM with two dimensions are presented in Fig. 1.

Figures 1(a) and (b) show that modal reduction is the worst both in terms of prediction and assessing predictability. This is not a surprising result given the setting of $\mathsf{A}$ and $\Sigma$, i.e. the least-damped modes are not representative in this system. On the contrary, taking the forcing effect into account, EOF and BT reduction produce much better results. While the EOF truncation underestimates the predictability and BT overestimates it (Fig. 1b), their performances are comparable. However, there are some subtle differences in prediction errors (Fig. 1a). When $\tau = 0$, $\epsilon_{\mathrm{p}}$ represents the error in initial condition due to truncation. From the definition, we know that the EOF gives the smallest error. When $\tau \to \infty$, the least-damped normal mode dominates, which is well captured by modal reduction. BT somehow combines the benefits of modal and EOF reduction, yielding an overall better prediction.

## 4. Reduction of general case: EOFs are not normal modes

Generally, there are two reasons for EOFs $\neq$ normal modes: the system is non-normal and/or the stochastic forcing is correlated in normal-mode space. Such a system is too complex to be solved analytically. We mainly illustrate some ideas with the help of a simple model of tropical Atlantic variability (TAV).

Before proceeding, let us first consider a very special case where the system is represented by one mode and forcing is unitary. When $m = 1$, $\mathsf{L}$ and $\mathsf{R}$ would degrade to row vector $\mathbf{l}$ and column vector $\mathbf{r}$, and $\varepsilon_r$ is simply an exponential function of $\tau$,

$$\varepsilon_r(\tau) = 1 - \mathrm{e}^{(\tilde{a}+\tilde{a}^*)\tau} \tag{25}$$

where $\tilde{a} = \mathbf{l}\mathsf{A}\mathbf{r}$ is a scalar. The exponent $\tilde{a} + \tilde{a}^*$ satisfies the (elementary) algebraic Lyapunov equation:

$$(a + a^*)\mathbf{l}\mathsf{C}\mathbf{l}^\dagger = -\mathbf{l}\mathbf{l}^\dagger. \tag{26}$$

For modal, EOF and BT reduction, we have simple solutions for $a + a^*$,

$$2\Re[\lambda_1(\mathsf{A})], \ \frac{-1}{\lambda_1(\mathsf{C})}, \quad \text{and} \ \frac{-1}{h_1}.$$

However, the optimization property of eigenvalues tells us that $a + a^*$ is bounded,

$$\frac{-1}{\lambda_n(\mathsf{C})} \le a + a^* \le \frac{-1}{\lambda_1(\mathsf{C})}. \tag{27}$$

In other words, among all reduction methods, the one-dimensional model from EOF truncation possesses maximum predictability.

Now, let us turn to a simple but physically more relevant model – a coupled climate model of TAV (see Chang et al. 2001, 2004b, for details). The purpose of this model is to study the inter-hemispheric sea surface temperature (SST) variability with a set of reduced physical elements: a zonal averaged temperature equation with northward oceanic mean transport, oceanic damping and positive air–sea feedback. With all these simplifications, we arrive at a simple linear stochastic climate model:

$$\frac{\partial T}{\partial t} + V\frac{\partial T}{\partial y} + \lambda T - \kappa\frac{\partial^2 T}{\partial y^2} = \beta S(y) T_{\mathrm{ITCZ}} + \eta. \tag{28}$$

Here, mean current $V$, damping rate $\lambda$ and coupling pattern $S$ are spatial functions based on GCM analysis as well as observations; $T_{\mathrm{ITCZ}}$ denotes the SST over the active coupling region, ITCZ; diffusion $\kappa = 1 \times 10^4$ m$^2$ s$^{-1}$; coupling strength $\beta = 1/200$ d$^{-1}$; and $\eta$ is the spatial-white Gaussian noise. To solve eq. (30), the differential equation is first transformed into matrix form of eq. (1) by employing finite difference in $y$ with a grid resolution of 0.5° within 30°S–30°N, which gives $n = 121$.

In eq. (30), it is the coupling between atmosphere and ocean that makes the system non-normal, i.e. the normal modes are not orthogonal. The least-damped mode, a dipole, is an oscillatory

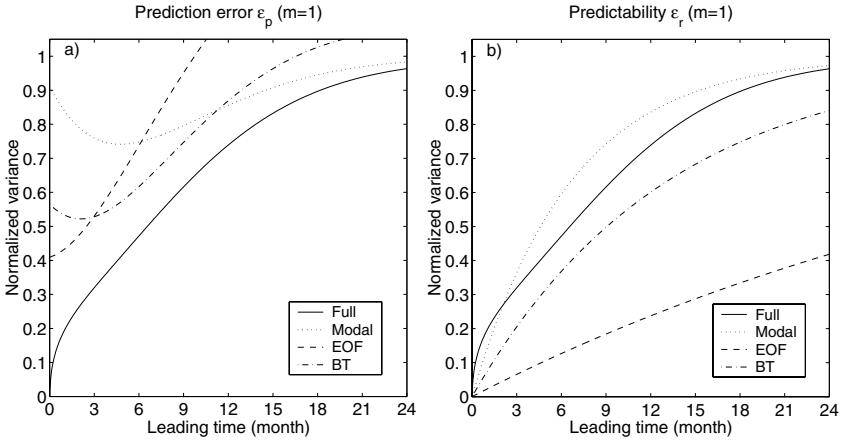


*Fig 2.* ROM of tropical Atlantic variability ($m = 1$): (a) normalized prediction error; (b) predictability of low-order model. The solid lines denote FOM; dotted lines, modal reduction; dashed lines, EOF truncation; dot-dashed lines, BT reduction.

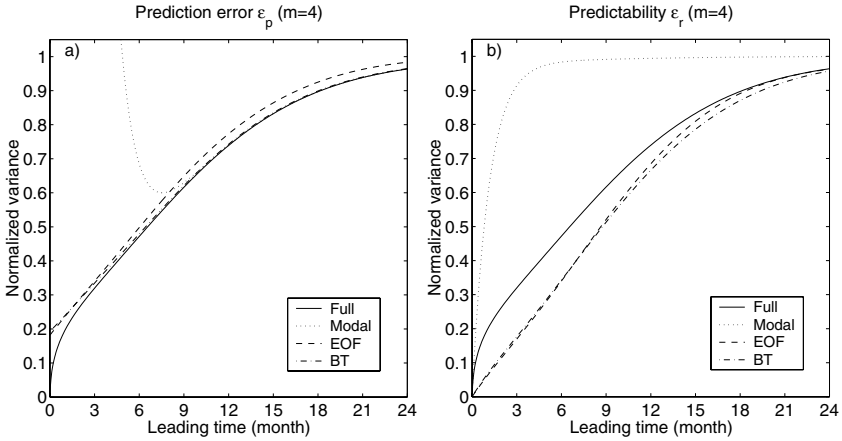


*Fig 3.* ROM of tropical Atlantic variability ($m = 4$): (a) prediction error; (b) predictability of various lower-order models. Line styles are the same as in Fig. 2.

one with period of about 6 yr, which is determined by the mean current. If we define $\epsilon \leq 0.5$ as useful skill, TAV is predictable up to two seasons (solid line in Figs. 2 and 3). However, the predictability comes mainly from coupling (non-normality) and less from oscillating modes. See Chang et al. (2004b) for an extensive discussion about predictability, including predictable component analysis, of model (30).

In this paper, we are more interested in the influence of model reduction on the evaluation of predictability. More precisely, do the ROMs agree well with the FOM in terms of predictability? To answer this question, let us consider a severely truncated model with only one mode (Fig. 2). Similar to the example in Section 3, here we see the same trend in prediction error (Fig. 2a): EOF truncation gives a better prediction at small lead time, while modal reduction does well at longer lead time; BT balances these two, and hence performs better overall. However, the large error of short-term prediction by modal reduction and long-term prediction by EOF truncation is due to the non-normal effect, not the forcing effect as shown in Fig. 1. Figure 2b shows that EOF and BT reductions of TAV overestimate its predictability while modal reduction underestimates its predictability. Also, EOF truncation produces the most predictable one-dimensional model, which is proved in inequality (29). Another interesting finding from Fig. 2 is that the full system (TAV) is more predictable than the least-damped normal mode (dipole). This is due to non-normality induced by the coupling, which is firmly established by inequality (8). Although the predictability of TAV is roughly captured by one normal mode, such agreement does not hold as more modes are included in the ROM.

Figure 3 shows the skill of a truncated model of TAV with four modes. As the number of modes increases, the truncated models using EOF and BT methods give a more accurate description of TAV, both in terms of prediction error and predictability. Nevertheless, there are still some subtle differences for long lead time: while BT produces the best prediction, EOF reduction preserves the predictability better. However, a surprising result comes from modal reduction. It does not converge to the full system at all. When four normal modes are used, not only does prediction become worse, except for longer lead time (Fig. 3a), the predictability estimate also further deviates from its real value. The poor performance of modal reduction is due to the lack of non-normality: when we project the full system on to its normal mode space, a non-normal operator $\mathsf{A}$ becomes a normal operator. In other words, modal reduction does not preserve non-normality.

However, it is the non-normality (coupling) that determines the error growth of TAV. Without this air–sea feedback, TAV is far less predictable. As we said before, non-normal growth only dominates the short-time prediction; for the long lead time only the least-damped modes survive. This explains the almost perfect prediction made by four normal modes for lead time greater than nine months.

## 5. Summary

Modal, EOF and BT, which all are popular methods for reducing the dimension of complex dynamical system, give optimal representations in very different senses. Farrell and Ioannou (2001a) have systematically compared these three methods in terms of optimal growth, power spectrum and maximum singular value, and concluded that BT gives a more accurate reduction of a non-normal system such as barotropic Couette flow. In this paper, we examined the same problem but took a different path. In particular, the performance of a reduction method is measured by the predictability and prediction error of the resulted low-order model. We not only confirmed the findings of Farrell and Ioannou (2001a), but also derived several bounds for predictability and model reduction that are useful in estimating the role of non-normality and guiding the interpretation of a low-order model. We further showed that a global optimal reduction exists for normal system under unitary forcing, and all three methods can be used to identify this optimal.

As illustrated by a simple tropical Atlantic climate model, however, the performance of modal reduction, EOF truncation and BT are considerably different for complicated system. Among them, modal reduction is not an encouraging technique. It ignores the forcing effect and destroys the non-normality, and hence the inherent properties of the full system are lost. EOF truncation has long been used in prediction and predictability analysis because of its robustness and easy implementation. However, it should be kept in mind that EOF truncation only keeps the leading modes, which are usually large-scale and long-lived. Such low-pass filtering is likely to artificially enhance the persistence and the predictability (Munk 1960). In our context, EOF truncation overestimates the system's predictability, the degree of which depends on the number of EOFs being used. BT, introduced by Farrell and Ioannou (2001a,b), is another promising method. It combines the merits of modal reduction and EOF truncation and creates an overall better representation of the full dynamical system. Its use, however, largely depends on how easy it is to find the stochastic operator $\mathsf{B}$. In climate studies, usually at least one realization is available, so we know the covariance matrix $\mathsf{C}$ by assuming ergodicity. Finding $\mathsf{B}$, on the other hand, is not a trivial task. It is equivalent to solving the covariance matrix of the adjoint system.

Some insight of the dynamical system (1) is also gained from this study. The evolution of a linear stochastic system is jointly controlled by the deterministic dynamics and the stochastic forcing. As a consequence, the EOFs may not necessarily represent the normal modes or the forcing patterns. For a normal system, however, maximal (minimal) predictability is achieved when forcing takes the structure of the least (most) damped mode, by which such a mode is excited. For a non-normal system, non-normality determines the short-term error growth. Therefore, the system could be more predictable than its first normal mode. In general, predictability of a stochastic system cannot be inferred from its normal modes, i.e. the stability analysis, alone. A systematic analysis of the interactions between normal modes and coherent forcing is needed to achieve a profound understanding of the predictable dynamics (Farrell and Ioannou 1996a,b; Neumaier and Schneider 2001; Chang et al. 2004a,b).

In the light of climate and weather prediction, the work presented here has several limitations. First, a dynamical system is usually non-linear, and the interaction between truncated components and retained components is likely strong. Reduction of such a complex system requires a careful selection of basis functions as well as a systematic strategy for modeling the effects of discarded processes on the dynamics of the resolved processes, i.e. a closure scheme (Majda et al. 2003). Most attempts at constructing closure schemes involve certain statistical and empirical tuning (Achatz and Schmitz 1997; Selten 1997; Roulston and Neelin 2003), which makes it hard to analyze without referring to a particular model. Secondly, the initial condition is not perfectly known because observations always have error. An analysis/initialization procedure, known as data assimilation, is widely used to produce initial states for numerical weather predictions. Because the dimension of atmosphere–ocean models is very high, assimilation is usually carried out in reduced space (Cane et al. 1996; Kaplan et al. 1997; Farrell and Ioannou 2001b; Buehner and Malanotte-Rizzoli 2003). And, such reduced state space data assimilation could improve forecast skill significantly.

## 6. Acknowledgments

We thank three anonymous reviewers for their valuable comments that helped to improve the presentation of this paper. This research was supported by the G. Unger Vetlesen Foundation and the Institute for Geophysics, University of Texas. Part of this work is based on FW's PhD thesis supported by grants from the National Oceanic and Atmospheric Administration (NA16GP1572) and National Science Foundation (ATM-99007625).

## 7. Appendix

Two inequalities are used in the proof of predictability bound (8). First, the trace of the product inequality (Marcus and Minc 1992): for any $n \times n$ positive semidefinite matrices $\mathsf{X}$ and $\mathsf{Y}$

$$\lambda_n(\mathsf{X})\mathrm{tr}\{\mathsf{Y}\} \leq \mathrm{tr}\{\mathsf{XY}\} \leq \lambda_1(\mathsf{X})\mathrm{tr}\{\mathsf{Y}\}. \tag{A1}$$

Secondly, Coppel's inequality (Coppel 1965; Mori et al. 1987): for any real square $n \times n$ matrix $\mathsf{X}$

$$\begin{aligned} \lambda_1\left(\mathrm{e}^{\mathsf{X}^{\mathrm{T}}t}\mathrm{e}^{\mathsf{X}t}\right) &\leq \mathrm{e}^{\lambda_1(\mathsf{X}+\mathsf{X}^{\mathrm{T}})t}, \quad t \geq 0 \\ \lambda_n\left(\mathrm{e}^{\mathsf{X}^{\mathrm{T}}t}\mathrm{e}^{\mathsf{X}t}\right) &\geq \mathrm{e}^{\lambda_n(\mathsf{X}+\mathsf{X}^{\mathrm{T}})t}, \quad t \geq 0 \end{aligned} . \tag{A2}$$

Balancing two covariance matrices, $\mathsf{B}$ and $\mathsf{C}$, can be done in three steps (Lall et al. 2002). First, construct a Cholesky factorization of $\mathsf{B}$ such that $\mathsf{B} = \mathsf{Z}^{\dagger}\mathsf{Z}$, where $\mathsf{Z}$ is an upper triangular matrix. Then, let $\mathsf{U}\Phi^2\mathsf{U}^{\dagger}$ be a singular value decomposition of $\mathsf{ZCZ}^{\dagger}$, and let $\mathsf{T} = \mathsf{Z}^{-1}\mathsf{U}\Phi^{1/2}$ and $\mathsf{T}^{-1} = \Phi^{-(1/2)}\,\mathsf{U}^{\dagger}\mathsf{Z}$. Finally, $\mathsf{T}^{\dagger}\mathsf{BT} = \Phi$ and $\mathsf{T}^{-1}\mathsf{CT}^{-\dagger} = \Phi$.

## References


Achatz, U. and Branstator, G. 1999. A two-layer model with empirical linear corrections and reduced order for studies of internal climate variability. *J. Atmos. Sci.* **56**, 3140–3160.

Achatz, U. and Schmitz, G. 1997. On the closure problem in the reduction of complex atmospheric models by PIPs and EOFs: a comparison for the case of a two-layer model with zonally symmetric forcing. *J. Atmos. Sci.* **54**, 2452–2474.

Antoulas, A. C. and Sorensen, D. C. 2001. Approximation of large-scale dynamical systems: An overview. *Int. J. Appl. Math. Comput. Sci.* **11**, 1093–1121.

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

Buehner, M. and Malanotte-Rizzoli, P. 2003. Reduced-rank Kalman filters applied to an idealized model of the wind-driven ocean circulation. *J. Geophys. Res.* **108**, doi:10.1029/2001JC000873.

Cane, M. A., Kaplan, A., Miller, R. N., Tang, B., Hackert, E. C. 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.

Chang, P., Ji, L. and Saravanan, R. 2001. A hybrid coupled model study of tropical Atlantic variability. *J. Climate* **14**, 361–390.

Chang, P., Saravanan, R., DelSole, T. and Wang, F. 2004a. Predictability of linear coupled systems. Part I: theoretical analysis. *J. Climate* **17**, 1474–1486.

Chang, P., Saravanan, R., Wang, F. and Ji, L. 2004b. Predictability of linear coupled systems. Part II: an application to a simple model of tropical Atlantic variability. *J. Climate* **17**, 1487–1503.

Coppel, W. A. 1965. *Stability and Asymptotic Behavior of Differential Equations.* D.C. Heath and Co., Boston, 166 pp.

Crommelin, D. T. and Majda, A. J. 2004. Strategies for model reduction: comparing different optimal bases. *J. Atmos. Sci.* **61**, 2206–2217.

D'Andrea, F. and Vautard, R. 2001. Extratropical low-frequency variability as a low dimensional problem. Part I: a simplified model. *Q. J. R. Meteorol. Soc.* **127**, 1357–1375.

Dullerud, G. E. and Paganini, F. 2000. *A Course in Robust Control Theory: A Convex Approach.* Springer-Verlag, Berlin, 440 pp.

Eckart, C. and Young, G. 1936. The approximation of one matrix by another of lower rank. *Psychometrica* **1**, 211–218.

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

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

Farrell, B. F. and Ioannou, P. J. 2001a. Accurate low-dimensional approximation of the linear dynamics of fluid flow. *J. Atmos. Sci.* **58**, 2771–2789.

Farrell, B. F. and Ioannou, P. J. 2001b. State estimation using a reduced-order Kalman filter. *J. Atmos. Sci.* **58**, 3666–3680.

Hasselmann, K. 1988. PIPs and POPs: the reduction of complex dynamical systems using principal interaction and oscillation patterns. *J. Geophys. Res.* **93**, 11 015–11 021.

Holmes, P., Lumley, J. L. and Berkooz, G. 1996. *Turbulence, Coherent Structures, Dynamical Systems and Symmetry.* Cambridge University Press, Cambridge, 420 pp.

Huffel, S. V. and Vandewalle, J. 1991. *The Total Least Squares Problem: Computational Aspects and Analysis.* SIAM, Philadelphia, 300 pp.

Johnson, S. D., Battisti, D. S. and Sarachik, E. S. 2000. Empirically derived Markov models and prediction of tropical Pacific sea surface temperature anomalies. *J. Climate* **13**, 3–17.

Kalnay, E., Lord, S. J. and McPherson, B. D. 1998. Maturity of operational numerical weather prediction: medium range. *Bull. Am. Meteorol. Soc.* **79**, 2753–2892.

Kaplan, A., Kushnir, Y., Cane, M. A. and Blumenthal, M. B. 1997. Reduced space optimal analysis for historical data sets: 136 years of Atlantic sea surface temperature. *J. Geophys. Res.* **102**, 27 835–27 860.

Kwasniok, F. 2004. Empirical low-order models of barotropic flow. *J. Atmos. Sci.* **61**, 235–245.

Lall, S., Marsden, J. E. and Glavaški, S. 2002. A subspace approach to balanced truncation for model reduction of non-linear control systems. *Int. J. Robust Nonlinear Control* **12**, 519–535.

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

Lorenz, E. N. 1969. The predictability of a flow which possesses many scales of motion. *Tellus* **21**, 289–307.

Majda, A. J., Timofeyev, I. and Vanden-Eijnden, E. 2003. Systematic strategies for stochastic mode reduction in climate. *J. Atmos. Sci.* **60**, 1705–1722.

Marcus, M. and Minc, H. 1992. *A Survey of Matrix Theory and Matrix Inequalities.* Dover, New York, reprint edition. 180 pp.

Mirsky, L. 1960. Symmetric gauge functions and unitarily invariant norms. *Q. J. Math.* **11**, 50–59.

Moore, B. C. 1981. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Trans. Autom. Control* **AC-26**, 17–32.

Mori, T., Fukuma, N. and Kuwahara, M. 1987. Bounds in the Lyapunov matrix differential equation. *IEEE Trans. Autom. Control* **AC-32**, 55–57.

Munk, W. H. 1960. Smoothing and persistence. *J. Meteorol.* **17**, 92–93.

Neumaier, A. and Schneider, T. 2001. Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Trans. Math. Software* **27**, 27–57.

Obinata, G. and Anderson, B. D. O. 2000. *Model Reduction for Control System Design.* Springer-Verlag, Berlin, 165 pp.

Patil, D. J., Hunt, B. R., Kalnay, E., Yorke, J. A. and Ott, E. 2001. Local low dimensionality of atmospheric dynamics. *Phys. Rev. Lett.* **86**, 5878–5881.

Penland, C. and Magorian, T. 1993. Prediction of Niño 3 sea surface temperatures using linear inverse modeling. *J. Climate* **6**, 1067–1076.

Preisendorfer, R. W. 1988. *Principal Component Analysis in Meteorology and Oceanography.* Elsevier, Amsterdam, 425 pp.

Roulston, M. S. and Neelin, J. D. 2003. Non-linear coupling between modes in a low-dimensional model of ENSO. *Atmos. Ocean* **41**, 217–231.

Schneider, T. and Griffies, S. M. 1999. A conceptual framework for predictability studies. *J. Climate* **12**, 3133–3155.

Selten, F. M. 1995. An efficient description of the dynamics of barotropic flow. *J. Atmos. Sci.* **52**, 915–936.

Selten, F. M. 1997. Baroclinic empirical orthogonal functions as basis functions in an atmospheric model. *J. Atmos. Sci.* **54**, 2099–2114.

Timmermann, A., Voss, H. U. and Pasmanter, R. 2001. Empirical dynamical system modeling of ENSO using non-linear inverse techniques. *J. Phys. Oceanogr.* **31**, 1579–1598.

Tippett, M. K. and Chang, P. 2003. Some theoretical considerations on predictability of linear stochastic dynamics. *Tellus* **55A**, 148–157.

Wang, R. 2001. Prediction of seasonal climate in a low-dimensional phase space derived from the observed SST forcing. *J. Climate* **14**, 77–97.

Winkler, C. R., Newman, M. and Sardeshmukh, R. D. 2001. A linear model of wintertime low-frequency variability. Part I: formulation and forecast skill. *J. Climate* **14**, 4474–4494.

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.

Zhang, Y. and Held, I. M. 1999. A linear stochastic model of a GCM's midlatitude storm tracks. *J. Atmos. Sci.* **56**, 3416–3435.