5.8 PREDICTABILITY OF LINEAR STOCHASTIC DYNAMICS

Michael K. Tippett* and Ping Chang
International Research Institute for climate prediction, Palisades, NY

1. INTRODUCTION

The original work of Hasselmann (1976) on stochastic climate theory pioneered the use of linear stochastic dynamics for modeling and predicting various modes of climate variability. Since then, convincing evidence has been presented that many climate phenomena are described by linear stochastic dynamics, at least as a first-order approximation (Penland and Matrosova, 1994; Whitaker and Sardeshmukh, 1998). Linear stochastic models are also routinely used in operational climate forecasts (Penland and Matrosova, 1998; Winkler et al., 2001). Therefore, a general understanding of the predictability of linear stochastic dynamics is of theoretical interest and practical value.

A system is understood to be predictable on those time-scales where prediction errors do not exceed climate variability (Lorenz, 1969). Predictability depends on both the physical system and on the prediction system. Phenomena modeled by linear stochastic dynamics evolve according to stochastically-forced deterministic linear dynamics. In a prediction model consisting of the linear deterministic dynamics alone, prediction errors are due to initial condition errors and to the absence of stochastic forcing. Here we focus on the latter source of error and assume perfect initial conditions.

In this case, prediction errors depend on the deterministic dynamics of the prediction system and the stochastic forcing. Since it is difficult to determine the structure of stochastic forcing in reality, often theoretical studies assume that the noise forcing is unitary; a spatially white noise is a trivial example of unitary noise forcing. Under such forcing, all normal modes of the system are excited and error covariance growth is maximized in the spectral norm (Tippett and Cohn, 2001). Prediction error growth then depends on the temporal (eigenvalues) and spatial (eigenvectors) characteristics of the normal modes of the dynamics. An interesting question is how the predictability of a linear stochastic model relates to the eigenmodes of the system?

A key factor in describing and understanding prediction error growth is orthogonality of the eigenvectors of the dynamics. Dynamics with a complete set of orthogonal eigenvectors is said to be *normal*, and error growth in a such system is simply determined by the eigenvalues of the dynamics. Error growth in nonnormal systems is magnified compared to that in normal systems with the same eigenvalues (Ioannou, 1995; Farrell and Ioannou, 1996). The enhanced error growth of nonnormal systems raises the question of whether predictability is lost at a faster rate in nonnormal systems.

Since predictability depends on the relative sizes of prediction error and climatological variability, and since nonnormality increases both, it is not immediately obvious whether the effect of nonnormality on predictability is positive, negative or indeterminate. Chang et al. (2001) recently attempted to address this issue using a set of conventional predictability measures. Here we show that nonnormality in fact *increases* predictability under a set of generalized predictability measures. We define predictability using the eigenvalues of the *predictive information matrix*, a measure of relative prediction error size (Schneider and Griffies, 1999).

We begin our discussion in Section 2 with an introduction to a general linear stochastic system and associated prediction error growth; in Section 3 we analyze the predictability of the linear stochastic system under a set of general predictability measures; in Section 4 a theoretical example illustrates how nonnormality can give enhanced predictability; in Section 5 we summarize the general findings of this study and discuss their general implications.

2. LINEAR STOCHASTIC DYNAMICS

We assume that the observed phenomenon of interest is represented by a real n-dimensional state-vector $\mathbf{w}^{\text{obs}}$ and whose evolution is governed by linear stochastic dynamics. That is, $\mathbf{w}^{\text{obs}}$ satisfies

$$\frac{d\mathbf{w}^{\text{obs}}}{dt} = \mathbf{A}\mathbf{w}^{\text{obs}} + \mathbf{F}\xi\,, \quad \mathbf{w}^{\text{obs}}(t_0) = \mathbf{w}_0^{\text{obs}}\,, \qquad (1)$$

where the dynamics matrix **A** and the forcing matrix **F** are real $n \times n$ matrices; ξ is Gaussian, spatially uncorrelated, white-noise with $\langle \xi(t_1)\xi(t_2)^T \rangle = \delta(t_1 - t_2)\mathbf{I}$; the notation $()^T$ denotes transpose. A general deterministic prediction system has the form

$$\frac{d\mathbf{w}^{\text{pred}}}{dt} = \mathbf{A}^{\text{pred}}\mathbf{w}^{\text{pred}}\,, \quad \mathbf{w}^{\text{pred}}(t_0) = \mathbf{w}_0^{\text{pred}}\,. \qquad (2)$$

Differences between the observed state $\mathbf{w}^{\text{obs}}$ and the predicted state $\mathbf{w}^{\text{pred}}$ are due to (i) differences between the observed initial condition $\mathbf{w}_0^{\text{obs}}$ and the prediction initial condition $\mathbf{w}_0^{\text{pred}}$, (ii) deficiencies in the approximate dynamics $\mathbf{A}^{\text{pred}}$ and, (iii) the presence of stochastic processes in the observations. Here we assume that the deterministic prediction system is perfect with exact deterministic dynamics $\mathbf{A}^{\text{pred}} = \mathbf{A}$ and perfect initial conditions $\mathbf{w}_0^{\text{pred}} = \mathbf{w}_0^{\text{obs}}$. In this scenario, the prediction error

*Corresponding author address: IRI, Lamont-Doherty Earth Observatory of Columbia University PO Box 1000 / 61 Route 9W, Palisades, NY 10964-8000 *email*: tippett@iri.columbia.edu

$\mathbf{w} \equiv \mathbf{w}^{\text{obs}} - \mathbf{w}^{\text{pred}}$ evolves according to

$$\frac{d\mathbf{w}}{dt} = \mathbf{A}\mathbf{w} + \mathbf{F}\xi\,, \quad \mathbf{w}(t=0) = 0\,. \tag{3}$$

A more detailed discussion on the perfect initial condition scenario can be found in Chang et al. (2001). Additionally, we assume that **F** is unitary and that the dynamics matrix **A** is stable, i.e., all its eigenvalues $\lambda_i(\mathbf{A})$ have negative real part. The eigenvalues of **A** are ordered so that $0 > \operatorname{Re}\lambda_1(\mathbf{A}) \geq \operatorname{Re}\lambda_2(\mathbf{A}) \cdots \geq \operatorname{Re}\lambda_n(\mathbf{A})$. The assumption of unitary stochastic forcing is arbitrary and depends on the choice of inner product.

The prediction error covariance at lead-time τ is defined as $\mathbf{C}_\tau \equiv \langle \mathbf{w}(\tau)\mathbf{w}(\tau)^T \rangle$ and given by

$$\mathbf{C}_\tau = \int_0^\tau e^{\mathbf{A}t} e^{\mathbf{A}^T t}\, dt\,. \tag{4}$$

Suppose the dynamics matrix **A** is diagonalizable and has eigendecomposition $\mathbf{A} = \mathbf{Z}\Lambda\mathbf{Z}^{-1}$; the matrix Λ of eigenvalues is an $n \times n$ diagonal matrix with diagonal entries $\lambda_i(\mathbf{A})$; the eigenvector $\mathbf{z}_i$ is the i-th column of the $n \times n$ matrix **Z** and satisfies $\mathbf{A}\mathbf{z}_i = \lambda_i(\mathbf{A})\mathbf{z}_i$. The matrix **Y** of adjoint eigenvectors of **A** is defined by $\mathbf{Y} \equiv \left(\mathbf{Z}^{-1}\right)^\dagger$ where the notation $()^\dagger$ denotes conjugate transpose. We assume without loss of generality that the columns $\mathbf{y}_i$ of **Y** are unit vectors with $\mathbf{y}_i^\dagger \mathbf{y}_i = 1$. Now the prediction error covariance $\mathbf{C}_\tau$ can be expressed in the basis of the eigenvectors of the dynamics as $\mathbf{C}_\tau = \mathbf{Z}\tilde{\mathbf{C}}_\tau\mathbf{Z}^\dagger$. Using the relation $\mathbf{Y}^\dagger\mathbf{Z} = \mathbf{Z}^\dagger\mathbf{Y} = \mathbf{I}$, the matrix $\tilde{\mathbf{C}}_\tau$ is determined by

$$\tilde{\mathbf{C}}_\tau = \mathbf{Y}^\dagger\mathbf{C}_\tau\mathbf{Y} = \int_0^\tau e^{\Lambda t}\mathbf{Y}^\dagger\mathbf{Y}e^{\Lambda^\dagger t}\, dt = \mathbf{Y}^\dagger\mathbf{Y} \circ \mathbf{E}_\tau\,, \tag{5}$$

where the notation $\circ$ denotes Hadamard product[1], and the entries of the positive semidefinite matrix $\mathbf{E}_\tau$ are

$$[\mathbf{E}_\tau]_{ij} = \int_0^\tau e^{(\lambda_i(\mathbf{A})+\overline{\lambda}_j(\mathbf{A}))t}\, dt = \frac{e^{(\lambda_i(\mathbf{A})+\overline{\lambda}_j(\mathbf{A}))\tau} - 1}{\lambda_i(\mathbf{A}) + \overline{\lambda}_j(\mathbf{A})}\,. \tag{6}$$

The matrix $\mathbf{E}_\tau$ depends only on the lead-time τ and the eigenvalues of the dynamics. The dependence of the prediction error covariance on the eigenvalues of the dynamics and on the adjoint eigenvectors **Y** is shown by the representation $\mathbf{C}_\tau = \mathbf{Z}\left(\mathbf{Y}^\dagger\mathbf{Y} \circ \mathbf{E}_\tau\right)\mathbf{Z}^\dagger$.

This dependence is particularly clear in the special case when **A** has a complete set of orthogonal eigenvectors, i.e., when the dynamics is normal. In this case, $\mathbf{Y}^\dagger\mathbf{Y} = \mathbf{I}$ and $\mathbf{C}_\tau = \mathbf{Z}(\operatorname{diag}\mathbf{E}_\tau)\mathbf{Z}^\dagger$. The eigenvalues $\lambda_i(\mathbf{C}_\tau)$ of the prediction error covariance $\mathbf{C}_\tau$ are given by

$$\lambda_i(\mathbf{C}_\tau) = \lambda_i(\operatorname{diag}\mathbf{E}_\tau) = \frac{e^{2\operatorname{Re}\lambda_i(\mathbf{A})\tau} - 1}{2\operatorname{Re}\lambda_i(\mathbf{A})}\,, \tag{7}$$

recalling that $\operatorname{Re}\lambda_i(\mathbf{A}) < 0$. The eigenvalue $\lambda_i(\mathbf{C}_\tau)$, the variance explained by the i-th eigenvector or EOF of $\mathbf{C}_\tau$, depends only on the distance in the complex plane from $\lambda_i(\mathbf{A})$ to the imaginary axis. Consequently, useful measures of the error covariance size, such as $\operatorname{tr}\mathbf{C}_\tau$ and $\det\mathbf{C}_\tau$, are entirely determined by the eigenvalues of **A** for normal dynamics. When the dynamics matrix **A** is normal, the i-th eigenvector of $\mathbf{C}_\tau$ is $\mathbf{z}_i$, if $\mathbf{z}_i$ is real (equivalently, if $\lambda_i(\mathbf{A})$ is real); when $\lambda_i(\mathbf{A})$ is complex, $\mathbf{C}_\tau$ has a repeated eigenvalue and two nonunique eigenvectors $\operatorname{Re}\mathbf{z}_i$ and $\operatorname{Im}\mathbf{z}_i$. Remarkably, for normal dynamics the eigenvectors of $\mathbf{C}_\tau$ are independent of the lead-time τ.

For general nonnormal dynamics, the prediction error covariance is less simply described and its size not fully determined by the eigenvalues of the dynamics matrix **A**. However, there are useful estimates for the eigenvalues of $\mathbf{C}_\tau$ that depend on the eigenvalues of the dynamics matrix **A**. For instance, the variance explained by the leading eigenvector of $\mathbf{C}_\tau$ has the lower bound

$$\lambda_1(\mathbf{C}_\tau) \geq \mathbf{y}_1^\dagger\mathbf{C}_\tau\mathbf{y}_1 = \frac{e^{2\operatorname{Re}\lambda_1(\mathbf{A})\tau} - 1}{2\operatorname{Re}\lambda_1(\mathbf{A})}\,. \tag{8}$$

Physically, this means that more variance is explained by the first eigenvector of $\mathbf{C}_\tau$ for nonnormal dynamics than for normal dynamics with the same eigenvalues. The trailing eigenvalue of $\mathbf{C}_\tau$ is bounded above by

$$\lambda_n(\mathbf{C}_\tau) \leq \mathbf{y}_n^\dagger\mathbf{C}_\tau\mathbf{y}_n = \frac{e^{2\operatorname{Re}\lambda_n(\mathbf{A})\tau} - 1}{2\operatorname{Re}\lambda_n(\mathbf{A})}\,. \tag{9}$$

This means that the trailing eigenvector of $\mathbf{C}_\tau$ explains less variance for nonnormal dynamics than for equivalent normal dynamics. Together, these inequalities show that the condition number $\lambda_1(\mathbf{C}_\tau)/\lambda_n(\mathbf{C}_\tau)$ of the prediction error covariance matrix $\mathbf{C}_\tau$ is larger for nonnormal dynamics than for equivalent normal dynamics. Large condition number is a necessary condition for low-rank approximations (Tippett et al., 2000). The condition number can also be used to bound maximum growth $\|e^{t\mathbf{A}}\|^2 \leq \lambda_1(\mathbf{C}_\infty)/\lambda_n(\mathbf{C}_\infty)$ (Hewer and Kenney, 1988).

Similar lower bounds for the total variance and volume in phase-space of the error variance matrix $\mathbf{C}_\tau$ are:

$$\operatorname{tr}\mathbf{C}_\tau \geq \operatorname{tr}\mathbf{E}_\tau\,; \tag{10}$$

$$\det\mathbf{C}_\tau \geq \det(\operatorname{diag}\mathbf{E}_\tau)\,. \tag{11}$$

The right-hand-sides of (10) and (11) are the total variance and phase-space volume of the error covariance of a normal system with the same eigenvalues; (10) is due to Ioannou (1995); (11) is Lemma 2 of the Appendix. The relations in (8),(10) and (11) lead to the general conclusion that global measures of prediction error are larger at all lead-times for nonnormal dynamics than for normal dynamics with the same eigenvalues. Next we compare the predictability of normal and nonnormal systems.

3. PREDICTABILITY

Prediction utility at lead-time τ depends on the size of the prediction error covariance $\mathbf{C}_\tau$ relative to the climatological covariance $\mathbf{C}_\infty$. Here, we use measures of predictability defined by the eigenvalues of the *predictive information matrix* $\mathbf{G}_\tau \equiv \mathbf{C}_\infty^{-1}\mathbf{C}_\tau$ (Schneider and Griffies,

[1]The Hadamard product of two matrices **X** and **Y** with entries $\mathbf{X}_{ij}$ and $\mathbf{Y}_{ij}$, respectively, is the matrix whose entries are $\mathbf{X}_{ij}\mathbf{Y}_{ij}$.

1999). The predictive information matrix is a multivariate generalization of the univariate relative error variance $\sigma_\tau^2/\sigma_\infty^2$ where σ_τ^2 is the prediction error variance at lead-time τ and σ_∞^2 is the climatological variance. We consider measures of predictability that are multivariate generalizations of the univariate predictability measure $1-\sigma_\tau^2/\sigma_\infty^2$.

The eigenvalues of $\mathbf{G}_\tau$ are invariant with respect to linear transformations of the state variable $\mathbf{w}$ since if we define a new variable $\hat{\mathbf{w}} = \mathbf{L}\mathbf{w}$ and its prediction error covariance $\hat{\mathbf{C}}_\tau \equiv \langle \hat{\mathbf{w}}(\tau)\hat{\mathbf{w}}(\tau)^T\rangle$, the new predictive information matrix $\hat{\mathbf{G}}_\tau \equiv \hat{\mathbf{C}}_\infty^{-1}\hat{\mathbf{C}}_\tau$ is related to $\mathbf{G}_\tau$ by a similarity transformation

$$\hat{\mathbf{G}}_\tau = \left(\mathbf{L}\mathbf{C}_\infty\mathbf{L}^T\right)^{-1}\left(\mathbf{L}\mathbf{C}_\tau\mathbf{L}^T\right) = \mathbf{L}^{-T}\mathbf{C}_\infty^{-1}\mathbf{C}_\tau\mathbf{L}^T\,. \quad (12)$$

Therefore, predictability measures defined by the eigenvalues of the predictive information matrix $\mathbf{G}_\tau$ are, unlike the error growth, invariant with respect to linear transformations. The assumption of unitary forcing is, however, dependent on the choice of inner product. Initially, the prediction error is zero and consequently the predictive information matrix $\mathbf{G}_0$ is zero. In the limit of large lead-time τ, the prediction error is identical with the climatological variability and $\mathbf{G}_\infty = \mathbf{I}$. The eigenvalues of $\mathbf{G}_\tau$ are between zero and unity for intermediate lead-times since if λ is an eigenvalue of $\mathbf{G}_\tau$ with eigenvector $\mathbf{p}$, then

$$\lambda = \frac{\mathbf{p}^T\mathbf{C}_\tau\mathbf{p}}{\mathbf{p}^T\mathbf{C}_\infty\mathbf{p}}\,, \quad (13)$$

and $\mathbf{p}^T\mathbf{C}_\infty\mathbf{p} \geq \mathbf{p}^T\mathbf{C}_\tau\mathbf{p} > 0$. The spatial pattern $\mathbf{p}$ that minimizes the ratio in (13) is the first *predictable pattern* and the associated λ is its relative error variance (Schneider and Griffies, 1999). The eigenvalues of $\mathbf{G}_\tau$ increase with lead-time, i.e., if $\tau_1 \leq \tau_2$, then $\lambda_i(\mathbf{G}_{\tau_1}) \leq \lambda_i(\mathbf{G}_{\tau_2})$.

Predictable patterns and their relative error variances are simply related to the dynamics when the dynamics is normal. In this case, the predictive information matrix $\mathbf{G}_\tau$ has the simple form

$$\mathbf{G}_\tau = \mathbf{C}_\infty^{-1}\mathbf{C}_\tau = \mathbf{Z}\,\mathrm{diag}(\mathbf{E}_\infty)^{-1}\,\mathrm{diag}(\mathbf{E}_\tau)\mathbf{Z}^{-1}\,, \quad (14)$$

since $\mathbf{Y} = \mathbf{Z}$. Additionally, (14) is the eigendecomposition of $\mathbf{G}_\tau$ so that the eigenvalues of $\mathbf{G}_\tau$ are

$$\lambda_i(\mathbf{G}_\tau) = 1 - e^{2\,\mathrm{Re}\,\lambda_{n-i+1}(\mathbf{A})\tau} \quad (15)$$

with eigenvectors of $\mathbf{G}_\tau$ being the corresponding eigenvectors of the dynamics. The first predictable pattern of a system with normal dynamics is the leading eigenvector $\mathbf{z}_1$ of the dynamics.

In the general case, the prediction error covariance $\mathbf{C}_\tau$ can be written as

$$\mathbf{C}_\tau = \mathbf{C}_\infty - e^{\tau\mathbf{A}}\mathbf{C}_\infty e^{\tau\mathbf{A}^T}\,, \quad (16)$$

so that

$$\mathbf{G}_\tau = \mathbf{C}_\infty^{-1}\mathbf{C}_\tau = \mathbf{I} - \mathbf{C}_\infty^{-1}e^{\mathbf{A}\tau}\mathbf{C}_\infty e^{\mathbf{A}^T\tau}\,. \quad (17)$$

The eigenvalues of $\mathbf{G}_\tau$ are then

$$\begin{aligned}\lambda_i(\mathbf{G}_\tau) &= \lambda_i(\mathbf{I} - \mathbf{C}_\infty^{-1}e^{\tau\mathbf{A}}\mathbf{C}_\infty e^{\tau\mathbf{A}^T}) \\ &= \lambda_i(\mathbf{I} - \mathbf{C}_\infty^{-1/2}e^{\tau\mathbf{A}}\mathbf{C}_\infty e^{\tau\mathbf{A}^T}\mathbf{C}_\infty^{-1/2}) \\ &= \lambda_i(\mathbf{I} - \mathbf{W}_\tau\mathbf{W}_\tau^T) \\ &= 1 - \sigma_{n-i+1}^2(\mathbf{W}_\tau)\,,\end{aligned} \quad (18)$$

where we introduce the matrix $\mathbf{W}_\tau \equiv \mathbf{C}_\infty^{-1/2}e^{\mathbf{A}\tau}\mathbf{C}_\infty^{1/2}$. Consequently, the eigendecomposition of $\mathbf{G}_\tau$ is determined by the singular value decomposition of $\mathbf{W}_\tau$. The eigenvectors of $\mathbf{G}_\tau$ are obtained by applying $\mathbf{C}_\infty^{-1/2}$ to the left singular vectors of $\mathbf{W}_\tau$.

The matrix $\mathbf{W}_\tau$ is invertible for finite τ since its eigenvalues are given by $\lambda_i(\mathbf{W}_\tau) = e^{\lambda_i(\mathbf{A})\tau}$. When the dynamics is normal, $\mathbf{W}_\tau$ is normal and $\sigma_i^2(\mathbf{W}_\tau) = |\lambda_i(\mathbf{W}_\tau)|^2$. Therefore, a comparison between the singular values and eigenvalues of the matrix $\mathbf{W}_\tau$ is equivalent to comparing the eigenvalues of the predictive information matrix $\mathbf{G}_\tau$ for general dynamics and for equivalent normal dynamics. For instance, from the general property $\sigma_1^2(\mathbf{W}_\tau) \geq |\lambda_1(\mathbf{W}_\tau)|^2$, it follows that $\lambda_n(\mathbf{G}_\tau) \leq 1 - e^{2\,\mathrm{Re}\,\lambda_1(\mathbf{A})\tau}$. Physically this means that the relative error of the first predictable pattern of a nonnormal system is less than or equal to that of a normal system with the same eigenvalues.

A simple global predictability measure is the quantity $1 - 1/n\,\mathrm{tr}\,\mathbf{G}_\tau$. This measure is more useful than $\mathrm{tr}\,\mathbf{C}_\tau/\,\mathrm{tr}\,\mathbf{C}_\infty$ since it is invariant under linear change of state variable. For normal dynamics,

$$1 - \frac{1}{n}\,\mathrm{tr}(\mathbf{G}_\tau) = \frac{1}{n}\sum_{i=1}^n e^{2\,\mathrm{Re}\,\lambda_i(\mathbf{A})\tau}\,. \quad (19)$$

This measure of predictability is larger for general nonnormal dynamics since

$$\begin{aligned}1 - \frac{1}{n}\,\mathrm{tr}\,\mathbf{G}_\tau &= \frac{1}{n}\sum_{i=1}^n \sigma_i^2\,(\mathbf{W}_\tau) \geq \frac{1}{n}\sum_{i=1}^n |\lambda_i\,(\mathbf{W}_\tau)|^2 \\ &= \frac{1}{n}\sum_{i=1}^n e^{2\,\mathrm{Re}\,\lambda_i(\mathbf{A})\tau}\,,\end{aligned} \quad (20)$$

where we have used inequality (32).

Another global predictability measure is the *predictive power* α_τ defined by (Schneider and Griffies, 1999)

$$\alpha_\tau \equiv 1 - (\det\mathbf{G})^{1/2n}\,. \quad (21)$$

The predictive power for normal dynamics is, using (7),

$$\alpha_\tau = 1 - \prod_{i=1}^n \left(1 - e^{2\,\mathrm{Re}\,\lambda_i(\mathbf{A})\tau}\right)^{1/2n}\,. \quad (22)$$

The predictive power is larger for general nonnormal dynamics since Lemma 4 of the Appendix gives that

$$\det\left(\mathbf{G}_\tau\right) = \det\left(\mathbf{I} - \mathbf{W}_\tau\mathbf{W}_\tau^T\right) \leq \prod_{i=1}^n \left(1 - e^{2\,\mathrm{Re}\,\lambda_i\tau}\right)\,. \quad (23)$$

Similar arguments show that nonnormality increases predictability as measured by *relative entropy* defined in Kleeman (2001).

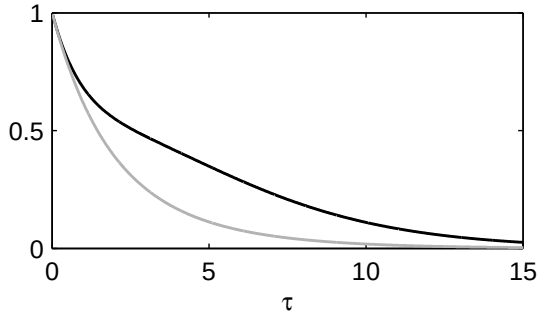

Figure 1: Predictability of the normal $\theta = 0$ (gray line) and nonnormal $\theta = \pi/2 + 0.05$ (black line) system in (24).

4. THEORETICAL EXAMPLE

Consider the dynamics matrix **A** given by

$$\mathbf{A} = \frac{\lambda}{\mathbf{z}^T\mathbf{y}}\mathbf{z}\mathbf{y}^T + \lambda\mathbf{I}\,, \quad \|\mathbf{z}\| = \|\mathbf{y}\| = 1\,, \tag{24}$$

where $\mathbf{z}^T\mathbf{y} \neq 0$, $\lambda < 0$ and $\epsilon > 0$. This dynamics has two real, negative eigenvalues λ and 2λ, and is normal when the unit vectors $\mathbf{z}$ and $\mathbf{y}$ are parallel and nonnormal otherwise. The angle θ between $\mathbf{z}$ and $\mathbf{y}$ is given by $\cos\theta = \mathbf{z}^T\mathbf{y}$. Initial transient growth is arbitrarily large as the angle θ is approaches $\pi/2$. The relation

$$e^{t\mathbf{A}} = e^{\lambda t}\left(\sec(\theta)\left(e^{\lambda t} - 1\right)\mathbf{z}\mathbf{y}^T + \mathbf{I}\right)\,, \tag{25}$$

can be used to calculate the prediction error covariance. In the special case that $n = 2$ and **A** is a 2×2 matrix, we find that

$$1 - \frac{1}{2}\operatorname{tr}\mathbf{G}_\tau \approx \frac{1}{2}e^{2\,\lambda\,\tau}\left(9 - 16\,e^{\lambda\,\tau} + 9\,e^{2\,\lambda\,\tau}\right) \tag{26}$$

where the terms neglected are $\mathcal{O}\left(\theta - \pi/2\right)^2$. In a normal system with the same eigenvalues

$$1 - \frac{1}{2}\operatorname{tr}\mathbf{G}_\tau = \frac{1}{2}e^{2\,\lambda\,\tau}\left(1 + e^{2\,\lambda\,\tau}\right)\,. \tag{27}$$

Initially the predictability of the normal and nonnormal system is the same. The predictability of the nonnormal system is larger by a factor of nine for large lead-times, independent of λ. Figure 1 shows the nonnormal and equivalent normal system predictability for $\theta = \pi/2 + 0.05$ and $\lambda = -1/6$.

5. DISCUSSION

The purpose of this study is to explore theoretically the relationship between deterministic dynamics and predictability in general linear stochastic systems, without reference to any specific system. We consider general predictability definitions that depend on the spectrum of the predictive information matrix. The main result is that normal-mode analysis provides lower bounds for predictability. Nonnormality can simultaneously produce error growth and predictability that is larger than that in normal systems with the same eigenvalues. In a simple theoretical example, a system with nonmodal growth exhibits both enhanced error growth and predictability.

Computation of the complete spectrum of the predictive information matrix is impractical when the forecast error covariance is incompletely known. Therefore it is of practical value to note that lower bounds for predictability obtained from normal mode analysis are also valid when a truncated set of trailing eigenvalues of the predictive information matrix are used.

The assumption of unitary stochastic forcing is arbitrary and introduces an undesirable dependence on the choice of inner product. Work to be reported elsewhere examines the impact of general forcing on predictability. We find that normal mode analysis also gives lower bounds for predictability for general, nonsingular stochastic forcing.

APPENDIX

Lemma 1 (Oppenheim's inequality). *Let* $\mathbf{A} = (a_{ij})$ *and* $\mathbf{B} = (b_{ij})$ *be positive semidefinite matrices. Then (Horn and Johnson, 1985)*

$$\det(\mathbf{A} \circ \mathbf{B}) \geq \left(\prod_{i=1}^{n} a_{ii}\right)\det\mathbf{B}\,. \tag{28}$$

Lemma 2. *Suppose* $\mathbf{C}_\tau = \mathbf{Z}\left(\mathbf{Y}^\dagger\mathbf{Y} \circ \mathbf{E}_\tau\right)\mathbf{Z}^\dagger$ *with* $\mathbf{Y} = \left(\mathbf{Z}^{-1}\right)^\dagger$ *and the entries of the matrix* $\mathbf{E}_\tau$ *are*

$$\left[\mathbf{E}_\tau\right]_{ij} = \int_0^\tau e^{(\lambda_i + \overline{\lambda}_j)t}\,dt = \frac{e^{(\lambda_i+\overline{\lambda}_j)\tau} - 1}{\lambda_i + \overline{\lambda}_j}\,. \tag{29}$$

Then, $\det\mathbf{C}_\tau \geq \det\operatorname{diag}(\mathbf{E}_\tau)$.

Proof. First we show that $\mathbf{E}_\tau$ is positive semi-definite because it is a Gram matrix. The functions $h_i(t) = e^{\lambda_i t}$, $0 \leq t \leq \tau$ are in the space $\mathbf{L}_2([0,\tau])$. The inner product in this space is

$$(h_i, h_j) \equiv \int_0^\tau h_i\overline{h}_j\,dt = \int_0^\tau e^{(\lambda_i+\overline{\lambda}_j)t}\,dt = \left[\mathbf{E}_\tau\right]_{ij}\,. \tag{30}$$

Now,

$$\begin{aligned}\det\mathbf{C}_\tau &= \det\left(\mathbf{Z}\left(\mathbf{Y}^\dagger\mathbf{Y} \circ \mathbf{E}_\tau\right)\mathbf{Z}^\dagger\right)\\ &= \det\left(\mathbf{Z}^\dagger\mathbf{Z}\right)\det\left(\mathbf{Y}^\dagger\mathbf{Y} \circ \mathbf{E}_\tau\right)\\ &\geq \det\left(\mathbf{Z}^\dagger\mathbf{Z}\right)\det\left(\mathbf{Y}^\dagger\mathbf{Y}\right)\det(\operatorname{diag}(\mathbf{E}_\tau))\\ &= \det(\operatorname{diag}(\mathbf{E}_\tau))\,,\end{aligned} \tag{31}$$

where we use (28) and $\det\left(\mathbf{Z}^\dagger\mathbf{Z}\right)\det\left(\mathbf{Y}^\dagger\mathbf{Y}\right) = 1$. □

Lemma 3 (An inequality of Weyl). *Suppose that* **W** *is invertible. Then, for* $s > 0$ *(Marcus and Minc, 1992, II.4.2)*

$$\sum_{i=1}^{k}\sigma_i^s(\mathbf{W}) \geq \sum_{i=1}^{k}\left|\lambda_i(\mathbf{W})\right|^s\,, \quad k = 1, 2, \ldots, n\,. \tag{32}$$

Lemma 4. *Suppose that* $\mathbf{W}$ *is an* $n \times n$ *matrix with* $\sigma_1(\mathbf{W}) < 1$ *and* $\sigma_n(\mathbf{W}) > 0$*. Then*

$$\det\left(\mathbf{I} - \mathbf{W}\mathbf{W}^T\right) \leq \prod_{i=0}^{n} \left(1 - |\lambda_i(\mathbf{W})|^2\right) . \qquad (33)$$

Proof. Since

$$\det\left(\mathbf{I} - \mathbf{W}\mathbf{W}^T\right) = \prod_{i=0}^{n} \lambda_i\left(\mathbf{I} - \mathbf{W}\mathbf{W}^T\right) , \qquad (34)$$

and the eigenvalues of $\mathbf{I} - \mathbf{W}\mathbf{W}^T$ are $1 - \sigma_i^2(\mathbf{W})$, it remains to show that

$$\prod_{i=0}^{n} \left(1 - \sigma_i^2(\mathbf{W})\right) \leq \prod_{i=0}^{n} \left(1 - |\lambda_i(\mathbf{W})|^2\right) . \qquad (35)$$

We generalize the proof of Marcus and Minc (1992, II.4.2). Since $\sigma_n(\mathbf{W}) > 0$ we can take the logarithm of both sides of (Marcus and Minc, 1992, II.4.1(9))

$$\prod_{i=1}^{k} \sigma_i(\mathbf{W}) \geq \prod_{i=1}^{k} |\lambda_i(\mathbf{W})| , \quad k = 1, 2, \ldots, n , \qquad (36)$$

with equality for $k = n$, to give

$$\sum_{i=1}^{k} \log \sigma_i(\mathbf{W}) \geq \sum_{i=1}^{k} \log |\lambda_i(\mathbf{W})| , \qquad (37)$$

with equality for $k = n$. The entries of the two n-vectors $\mathbf{u}$ and $\mathbf{v}$ defined by $\mathbf{u} \equiv \{\log|\lambda_1(\mathbf{W})|, \ldots, \log|\lambda_n(\mathbf{W})|\}$ and $\mathbf{v} \equiv \{\log \sigma_1(\mathbf{W}), \ldots, \log \sigma_n(\mathbf{W})\}$ are negative and related by $\mathbf{u} = \mathbf{S}\mathbf{v}$ where $\mathbf{S}$ is some doubly stochastic matrix (Marcus and Minc, 1992, II.3.5.4). The function

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{k} -\log(1 - e^{2x_i}) \qquad (38)$$

is a convex function on $(-\infty, 0)^n$ since $-\log(1 - e^{2x})$ is convex for $x < 0$. Consequently, $g(\mathbf{R}) \equiv f(\mathbf{R}\mathbf{v})$ is a convex function on the set of doubly stochastic matrices $\mathbf{R}$ (Marcus and Minc, 1992, II.4.1.2). In addition, $g(\mathbf{R}) \leq g(\mathbf{P})$ for some permutation matrix $\mathbf{P}$ (Marcus and Minc, 1992, II.4.1.4). Thus taking $\mathbf{R} = \mathbf{S}$ gives,

$$f(\mathbf{u}) = f(\mathbf{S}\mathbf{v}) = g(\mathbf{S}) \leq g(\mathbf{P}) = f(\mathbf{P}\mathbf{v}) . \qquad (39)$$

Substitution gives

$$-\sum_{i=1}^{k} \log\left(1 - |\lambda_i(\mathbf{W})|^2\right) \leq -\sum_{i=1}^{k} \log\left(1 - \sigma_{s(i)}^2(\mathbf{W})\right) , \qquad (40)$$

for some permutation s. Since the logarithm is an increasing function it follows that

$$\sum_{i=1}^{k} \log\left(1 - \sigma_{s(i)}^2(\mathbf{W})\right) \geq \sum_{i=1}^{k} \log\left(1 - \sigma_i^2(\mathbf{W})\right) , \qquad (41)$$

for any permutation s. Finally,

$$\sum_{i=1}^{k} \log\left(1 - \sigma_i^2(\mathbf{W})\right) \leq \sum_{i=1}^{k} \log\left(1 - |\lambda_i(\mathbf{W})|^2\right) \qquad (42)$$

and, the theorem follows. □

References

Chang, P., R. Saravanan, and T. DelSole, 2001: Normal-modes, Noice Forcing and Predictability in Linear Coupled Systems. *J. Climate*, submitted.

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

Hasselmann, K., 1976: Stochastic climate models I. Theory. *Tellus*, **28**, 473–485.

Hewer, G. and C. Kenney, 1988: The sensitivity of the stable Lyapunov equation. *SIAM J. Control and Optimization*, **26**, 321–344.

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

Ioannou, P. J., 1995: Nonnormality increases variance. *J. Atmos. Sci*, **52**, 1155–1158.

Kleeman, R., 2001: Measuring dynamical prediction utility using relative entropy. *J. Atmos. Sci.*, submitted.

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

Marcus, M. and H. Minc, 1992: *A survey of matrix theory and matrix inequalities*. Dover, New York.

Penland, C. and L. Matrosova, 1994: A Balance Condition for Stochastic Numerical Models with Application to the El Niño-Southern Oscillation. *J. Climate*, **7**, 1352–1372.

—, 1998: Prediction of tropical Atlantic sea aurface temperatures using Linear Inverse Modeling. *J. Climate*, **11**, 483–496.

Schneider, T. and S. Griffies, 1999: A Conceptual Framework for Predictability Studies. *J. Climate*, **12**, 3133–3155.

Tippett, M. K. and S. E. Cohn, 2001: Adjoints and low-rank covariance representation. *Nonlinear Processes in Geophysics*, in press.

Tippett, M. K., S. E. Cohn, R. Todling, and D. Marchesin, 2000: Conditioning of the stable, discrete-time Lyapunov operator. *SIAM J. Matrix Anal. Appl.*, **22**, 56–65.

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

Winkler, C. R., M. Newman, and P. D. Saradeshmukh, 2001: A Linear Model of Wintertime Low-Frequency Variability. Part I: Formulation and Forecast Skill. *J. Climate*, in press.