The Optimal Growth of Tropical Sea Surface Temperature Anomalies

CÉCILE PENLAND AND PRASHANT D. SARDESHMUKH

Cooperative Institute for Research in the Environmental Sciences, University of Colorado, Boulder, Colorado

ABSTRACT

It is argued from SST observations for the period 1950–90 that the tropical Indo-Pacific ocean–atmosphere system may be described as a stable linear dynamical system driven by spatially coherent Gaussian white noise. Evidence is presented that the predictable component of SST anomaly growth is associated with the constructive interference of several damped normal modes after an optimal initial structure is set up by the white noise forcing. In particular, El Niño–Southern Oscillation (ENSO) growth is associated with an interplay of at least three damped normal modes, with periods longer than two years and decay times of 4 to 8 months, rather than the manifestation of a single unstable mode whose growth is arrested by nonlinearities. Interestingly, the relevant modes are not the three least damped modes of the system. Rather, mode selection, and the establishment of the optimal initial structure from which growth occurs, are controlled by the stochastic forcing. Experiments conducted with an empirical stochastic-dynamical model show that stochastic forcing not only adds energy to the system, but also plays a role in setting up the optimal structure.

It is shown that growth from modal interference can occur for as long as 18 months, which within the confines of this model defines a predictability limit for growth events. Growth associated with the stochastic forcing is also possible, but is unpredictable. The timescale on which the predictable and unpredictable components of SST growth become comparable to each other gives a more conservative predictability limit of 15 months.

The above scenario is based on empirical evidence obtained from SST anomalies alone. From the results of several tests based on statistical properties of linear and nonlinear dynamical systems, one may conclude that much of the ENSO cycle in nature is dominated by stable, forced dynamics.

1. Introduction

Interannual variations in tropical Pacific sea surface temperatures (SSTs) play a key role in the interannual variability of the global climate system. Observing, understanding, and predicting such SST changes and their worldwide implications are the subject of intense current research. Considerable progress has already been made on all three fronts (Philander 1990; Glantz et al. 1991). This paper is a contribution in the area of inverse modeling, broadly defined as the extraction of dynamical properties of a system from its observed statistics. Inverse modeling is to be contrasted with direct or forward modeling in which one derives the dynamical properties from first principles. Although dealing exclusively with observations, inverse modeling is also to be distinguished from phenomenological observational studies. The latter are useful in documenting characteristics of a system that one would ideally wish to reproduce in models; however, these may or may not reveal anything about the underlying dynamics themselves.

The inverse modeling of the tropical atmosphere–ocean system suggests that on interannual timescales it may be viewed to a surprisingly good approximation as a linear system driven by white noise. We construct several tests for this premise and find that the system generally passes them, except during the warmest phase of El Niño–Southern Oscillation (ENSO) events. Other supporting evidence is contained in two previous papers (Penland and Magorian 1993; Penland and Matrosova 1994) and will be summoned here where appropriate. Briefly, Penland and Magorian show that the linear normal modes (perturbation eigenfunctions) of the system are determined reliably enough from the observed simultaneous and time-lag covariance statistics that SST forecasts, based on their free evolution, are useful for up to nine months in the eastern Pacific. A notable result from their study is that the forecast skill is much higher when considering the totality of modes than any one particular mode alone. Given these modes, Penland and Matrosova consider the problem of what the forcing must be in order to generate the observed variance and covariance statistics. They first show that a forcing that is white in both time and space is inconsistent with a fluctuation–dissipation relationship, and then, assuming that the forcing is white in time, determine its properties in space from that same relationship.

Corresponding author address: Dr. Cécile Penland, CIRES, University of Colorado, Campus Box 449, Boulder, CO 80309-0449.

The statement that a system is linear and driven by Gaussian white noise has several implications, which will be discussed at various stages in this article. One implication is that its normal modes must all have exponentially decaying amplitudes. This means that without forcing every initial perturbation must eventually decay. Indeed, the fluctuation–dissipation relationship mentioned above arises from the need to balance this general decaying tendency with forcing in a statistical sense; the system cannot have stationary statistics otherwise. The question immediately arises as to whether growth always occurs due to forcing in such a system. If it does, the prospects for predictability must be bleak, since a white noise forcing has a zero decorrelation timescale. If on the other hand it turns out that the system is predictable, with predictable growth events, then there must be an alternative mechanism for growth. That mechanism is the constructive interference of the modes, which can give not only local but also global growth over a finite time interval if the modes are nonorthogonal. Because more than one mode is involved, such growth is fundamentally nonmodal.

Evidence is presented here that not only is such a global growth, defined as an increase of a domain-integrated quadratic measure of SST anomalies, possible in the tropical Pacific and Indian Oceans, but that it does in fact often occur. Indeed, it is only a slight exaggeration to suggest that this is the rule rather than the exception. Most warm and cold ENSO episodes are associated with such a global growth. The effect is mostly associated with the interference of several damped modes, which, interestingly, are not the least damped modes of the system.

A case can be made that not only is the SST anomaly growth nonmodal, but that it is close to being "optimal." This is done in two steps. First, an optimal perturbation eigenanalysis is performed to determine the initial SST anomaly pattern associated with the maximum possible nonmodal growth over a given time interval, say seven months. It is then shown that whenever an SST anomaly pattern projects substantially onto this theoretical optimal structure, growth almost invariably occurs over the next seven months.

Optimal structures have been emphasized recently in studies of *error* variance growth in both atmospheric models (e.g., Lacarra and Talagrand 1988; Molteni and Palmer 1993) and ENSO models (e.g., Blumenthal 1991; Xue et al. 1994). Farrell (1988, 1989, 1990), Farrell and Ioannou (1993), and Borges and Hartmann (1992) have also considered *actual* variance growth in the atmosphere. The present study represents perhaps the first attempt at establishing the relevance of such structures in *actual* ENSO growth.

This mechanism of growth is very different from that sought by many authors in models of ENSO of varying complexity (e.g., Philander et al. 1984; Cane

and Zebiak 1985; Schopf and Suarez 1988, 1990; Hirst 1988; Battisti and Hirst 1989; Cane et al. 1990; Münnich et al. 1991; Wakata and Sarachik 1991; Neelin 1991; Jin and Neelin 1993a,b; Neelin and Jin 1993). The reigning paradigm in these model studies is that SST anomaly growth is associated with growing modes in a nonlinear dynamical system whose linearized dynamics are unstable. Much attention has been given in these studies to model parameters, such as the strength of the air–sea coupling, which lead to instability. The view propounded in this paper is that the linearized system need not be unstable at all for basinwide warming or cooling to occur.

If a dynamical system is nonlinear, then its linearized form, such as studied in the papers cited above, may certainly admit unstable modes whose growth is limited by nonlinearities. However, even when one finds a most unstable mode resembling observed structures by this means, there remain the questions of how nonlinear saturation is attained, whether this leaves intact the mode selection based upon linear growth rates, why the observed frequency spectrum is broadly red and not strongly peaked at the modal frequency, what are the relative roles of nonlinearity and forcing, and what causes decay. On the other hand, if the equations governing SST anomalies are dominated by linear dynamics, that is, if the nonlinear terms are always small or parameterizable as linear terms, then they can only support decaying modes. The problem now is not explaining decay, but growth. As discussed above, this can occur through either modal interference or forcing. The forcing plays an important role in mode selection and in determining the frequency spectrum. This represents a complication, but no more so than in the nonlinear regime of the unstable scenario described above and without the added complexities of nonlinear dynamics. The simplest model of SST anomalies for much of the ENSO cycle is therefore a *damped linear model* that makes the simplest assumptions about the forcing. The original motivation for this paper was in fact provided by a desire to determine the extent to which this alternate paradigm could be sustained.

It need hardly be emphasized that the approximation of the undoubtedly complex coupled tropical ocean–atmosphere system as a linear stochastically driven system is only an approximation, but in our view it is a good and a useful one for benchmarking and diagnostic purposes. It should be noted that our statements are based upon empirical normal modes (ENMs) derived from the SST observations alone. It is assumed that on the long timescales of ENSO evolution the SST equation is the only equation with explicit time dependence in the coupled system. This amounts to assuming that on these timescales the forcing of anomalous SST by anomalous surface winds, ocean currents, thermocline depths, and surface radiative, sensible, and latent heat fluxes can collectively be represented as lin-

ear terms in SST, plus a remainder which can be approximated as white noise. The normal modes obtained in this paper are thus closest in character to the so-called SST modes in the fast ocean-wave limit discussed by Neelin (1991). Insofar as our inverse modeling approach is successful, our results are consistent with the relevance of such SST modes in the real physical system. However, the satisfaction of the fast-wave limit is sufficient but not necessary for the validity of our dynamical interpretation.

The characterization of a system as "deterministic" or "stochastic" is a matter of which timescales are being emphasized. Our stochastic forcing may be deterministic, nonlinear, even chaotic at some timescale shorter than the interannual scale considered here. The assumption made is that internal nonlinearities vary quickly enough that their effect on the slow, deterministic dynamics on interannual timescales may be described by a stochastic differential equation (Wong and Zakai 1965; Papanicolaou and Kohler 1974; Arnold 1974; Hasselmann 1976; Gardiner 1985; Kloeden and Platen 1992). Note that this description does not preclude the existence of additional, *external* stochastic forcing.

The inverse modeling technique used here is based upon the principal oscillation pattern (POP) technique (Hasselmann 1988; von Storch et al. 1988), which has been used by many researchers (e.g., Blumenthal 1991; Xu 1993; Xue et al. 1994). As discussed in Penland (1989), the use of the POP technique implies certain assumptions about a dynamical system, and examining those assumptions allows further extensive diagnosis of the system. An especially appealing feature of the inverse modeling formalism presented here is that it incorporates tests for determining when the linear approximation is *not* valid. Even when this happens, however, one can still hope to learn something useful about a system.

The possibility that the coupled ocean–atmospheric dynamics may be dominated by stable modes driven by stochastic forcing has been considered previously. Such a situation obtained in the model of McWilliams and Gent (1978) when realistic parameters were chosen. Schopf and Suarez (1988) argued that if the McWilliams–Gent scenario was a proper paradigm of ENSO the spectrum would show frequencies of the "selected" modes. This argument assumes that the oscillating behavior is dominated by one or just a few chosen modes and is not only inconsistent with Penland and Magorian (1993), who found that many modes are necessary to make accurate predictions, but also ignores the possibility of nonmodal growth. Some studies (e.g., Lau 1985; Zebiak 1989; Kleeman and Power 1994) have considered the effect of stochasticity in models dominated by deterministic nonlinear ENSO dynamics. Latif (1987) found the random component of the wind stress to be important to the ocean dynam-

ics of the Hamburg Ocean General Circulation Model. Finally, Neelin (1990) discusses the difficulty of determining whether the system is dominated by a deterministic limit cycle and is perturbed by noise, or is stable and is sustained by noise. The particular contribution made here is the evidence obtained from data that stochastic forcing plays an important role in ENSO dynamics.

The paper is organized as follows. Section 2 recapitulates the basic procedures to be followed in approximating any dynamical system as a stable, linear, stochastically driven system, and describes some tests for establishing whether or not the approximation is valid. The ENMs of the linear system are presented in section 3. The maximum potential of SST growth over finite time intervals from modal interference is quantified in section 4 by means of a maximum amplification (MA) curve, determined through an optimal perturbation eigenanalysis. The eigenanalysis also determines the optimal initial SST patterns associated with such growth, the "right singular vectors" of the linear evolution operator. In principle, different initial patterns exist for optimizing growth over different time intervals. However, in our problem practically the same pattern is obtained for a wide range of time intervals, allowing us to call it *the* optimal structure. Evidence that growth as indicated by the MA curve actually occurs in the physical system is presented in section 5. The three ENMs whose interference is most important in this growth are also discussed there. Given the inverse-modeling results, one is now in a position to consider tests for linearity in section 6. In section 7, an experiment with Penland and Matrosova's empirical dynamical model of ENSO is described to show that the stochastic forcing plays an important role in exciting the optimal structure. Some annual cycle effects are presented in section 8. Section 9 considers the implications of our observational results for the theory and modeling of ENSO variability, and concluding remarks are made in section 10.

2. Inverse modeling preliminaries

Any dynamical system may be represented in the form

$$dX/dt = \mathcal{B}X + n(X) + f, \qquad (1)$$

where X is the state vector, $\mathcal{B}$ is the linear system matrix (also known as the deterministic feedback matrix), and $n(X)$ and f are nonlinear and external forcing terms. The components of X represent the values of all system variables in some linear vector space, not necessarily of finite dimension. For example, in the coupled global atmosphere–ocean system, X could represent the deviation of all relevant atmospheric and oceanic variables from their climatological values at all grid points. Understanding the evolution of any subset of these,

such as tropical SSTs, first involves determining what the relevant variables are. In essence one looks for a dynamical subsystem. If one does not have preconceptions of the system, only observations, then the simplest viable trial subsystem for tropical SSTs is a linear system driven by Gaussian white noise. Thus, one rewrites (1) as

$$dx/dt = \mathbf{B}x + \underline{\xi}, \qquad (2)$$

where x now contains only tropical SSTs, $\mathbf{B}$ is the appropriate submatrix of $\mathcal{B}$, and the nonlinear and forcing terms have been replaced by a white noise forcing $\underline{\xi}$. This trial model states that the evolution of anomalous tropical SSTs is governed by a multivariate linear Markov process. As discussed previously, this is not to say that influences such as anomalous surface heat fluxes and advection by anomalous ocean currents forced by anomalous winds are unimportant, but rather that they may be represented as linear terms in SST plus Gaussian white noise. This may seem like an oversimplification, but its validity may be checked in several ways as described below.

If (2) is an accurate representation of reality, then given a state vector $x(t)$ at time t, the most probable vector $\hat{x}(t + \tau_0)$ at time $t + \tau_0$ is (e.g., Penland 1989)

$$\hat{x}(t + \tau_0) = \exp(\mathbf{B}\tau_0)x(t). \qquad (3)$$

As discussed further by Penland (1989), the statistics of such a system must be Gaussian. The matrix $\mathbf{B}$ can therefore be determined from the observed time series $x(t)$ by an error variance minimization procedure as

$$\mathbf{B} = \tau_0^{-1} \ln\{\mathbf{C}(\tau_0)\mathbf{C}(0)^{-1}\}, \qquad (4)$$

where $\mathbf{C}(\tau_0)$ and $\mathbf{C}(0)$ are the covariance matrices at lag τ_0 and lag 0,

$$\mathbf{C}(\tau_0) = \langle x(t + \tau_0)x^\mathsf{T}(t) \rangle$$

$$\mathbf{C}(0) = \langle x(t)x^\mathsf{T}(t) \rangle. \qquad (5)$$

Angle brackets here denote an ensemble average, estimated as a time average for variables with stationary statistics. Note that empirical orthogonal functions (EOFs) are the eigenfunctions of $\mathbf{C}(0)$. For a system with stationary statistics it can be shown (appendix A) that the eigenvalues of $\mathbf{B}$ obtained in this manner have negative real parts.

Having obtained $\mathbf{B}$, forecasts of x can now made for an arbitrary lead time τ as

$$\hat{x}(t + \tau) = \exp(\mathbf{B}\tau)x(t) = \mathbf{G}(\tau)x(t), \qquad (6)$$

where $\mathbf{G}(t) \equiv \exp(\mathbf{B}\tau)$. This is the essence of the prediction technique used by Penland and Magorian (1993). Because of the presence of the white noise forcing $\underline{\xi}$ in (2), the forecasts in (6) differ from the

actual $x(t + \tau)$, by $e(\tau)$. Penland (1989) shows that the expected value of the global error $\delta(\tau) = \langle e(\tau)^\mathsf{T}e(\tau) \rangle$ is the trace of the error covariance matrix,

$$\langle e(\tau)e(\tau)^\mathsf{T} \rangle = \mathbf{C}(0) - \mathbf{G}(\tau)\mathbf{C}(0)\mathbf{G}^\mathsf{T}(\tau). \qquad (7)$$

The forecast error is stationary in the "wide sense" in that on average it depends only on the lead time τ and not on the time t the forecast is made. It is convenient to normalize δ by the trace of $\mathbf{C}(0)$ so that it increases from $\delta = 0$ at $\tau = 0$ to $\delta = 1$ at $t = \infty$.

Given $\mathbf{B}$ and $\mathbf{C}(0)$, the properties of the forcing $\underline{\xi}$ can be determined from the fluctuation–dissipation relation (Gardiner 1985; Penland 1989; Farrell and Ioannou 1993; Penland and Matrosova 1994)

$$\mathbf{B}\mathbf{C}(0) + \mathbf{C}(0)\mathbf{B}^\mathsf{T} + \mathbf{Q} = 0, \qquad (8)$$

where $\mathbf{Q} = \langle \underline{\xi}\underline{\xi}^\mathsf{T} \rangle dt$ is the covariance matrix of the white noise forcing multiplied by dt. A rigorous derivation of (8) involves considering the multivariate Fokker-Planck equation. Nevertheless, the existence of some such relation can readily be anticipated from (7) in the limit of small $t \to \Delta t$ in which the error covariance matrix is identified with $\mathbf{Q}dt \sim \mathbf{Q}\Delta t$. It is important to recognize that $\mathbf{Q}$ obtained from (8) is in general symmetric but not diagonal. Thus, the forcing must have spatial coherence; that is, it cannot be white in both time and space. The dominant patterns of forcing are then determined as the eigenfunctions of $\mathbf{Q}$ with the largest eigenvalues. These are also referred to as noise EOFs. A full discussion is given in Penland and Matrosova (1994).

The above procedure can of course be applied to observations $x(t)$ of any dynamical system, not just a system of the form (2). Thus, one may choose any lag τ_0, determine $\mathbf{B}$ via (4) and $\mathbf{Q}$ via (8), and use (6) to make forecasts. However, if the system is of the form (2), one will also find that, in general,

- its statistics are Gaussian, (9a)
- $\mathbf{B}$ is independent of the lag τ_0 that is chosen to determine it, (9b)
- $\mathbf{Q}$ determined from (8) is positive definite, (9c)
- the forecasts based on (6) are good, and (9d)
- the forecast errors grow as predicted by (7), assuming no error in the initial conditions. (9e)

These are demanding tests and some allowance should be made for the errors in estimating the covariance matrices in (5) from the available observations. Also a failure to meet these tests should not immediately be taken to imply that the dynamical equations for the phenomena in question must be nonlinear. It could be that one has not specified the correct subset x of $\mathcal{X}$ in (2). In the tropical SST problem this might indicate the need to include anomalous ocean currents and thermocline depths explicitly in x instead of implicitly

as, for example, in the fast ocean-wave limit. Steps (4)–(8) could then be repeated with the extended x to see if the requirements (9) are satisfied better. One might also consider an explicit treatment of surface wind stresses, although this is probably less crucial given the relative success of several "slave-atmosphere" type coupled atmosphere–ocean models (e.g., Neelin 1991; see also Penland and Matrosova 1994). Finally, there is the possibility that the system is linear but the external forcing cannot reasonably be approximated as white noise. These facts should be kept in mind when examining the observational evidence for the validity of (2) and (9) for the tropical SST evolution problem in the following sections.

3. Empirical normal modes

a. Data

Our data source is the Comprehensive Ocean–Atmosphere Data Set (COADS) (Slutz et al. 1985). SSTs for the period January 1950–December 1990 in the tropical Indo-Pacific basin (30°N–30°S, 30°E–70°W) were consolidated onto a 4° × 10° latitude–longitude grid. Anomalies were obtained by subjecting the data to a three-month running mean, and then removing the grand mean and the annual cycle at each grid point. An EOF analysis of these smoothed anomalies was performed. All subsequent calculations were performed in the space spanned by the first 15 EOFs and transformed into geographical space for interpretation. The time-dependent coefficients of the first 15 EOFs define the 15-component state vector $x(t)$ of this paper. In the following, it will be useful to keep in mind that the typical magnitudes of the components of $x(t)$ are proportional to the square root of the fraction of the SST variance explained by the corresponding EOFs.

The first 15 EOFs explain 67% of the total SST variance, and more than 90% of the variance in the area known as Niño-3, indicated by the outlined box (6°N–6°S, 150°–90°W) in Fig. 1a. The decision to truncate the EOF series at 15 was guided partly by the desire to focus on large-scale variability and partly by the need to eliminate small-scale errors in the data. The calculations were repeated with 10 and 20 EOFs to confirm that none of our main conclusions depended upon this choice. Henceforth, all references to SST in this paper will be to the 15 EOF truncated dataset unless specified otherwise.

Figure 1a shows the geographical distribution of the SST variance associated with the 15 EOFs. For future reference, the first two EOFs are shown in Figs. 1b and 1c. They explain 31% and 7.3% of the total variance of the anomalous 3-month running means, respectively. The time coefficients of these patterns (the principal components) are the components $x_1(t)$ and $x_2(t)$ of the 15-component state vector $x(t)$. Figure 1 high-
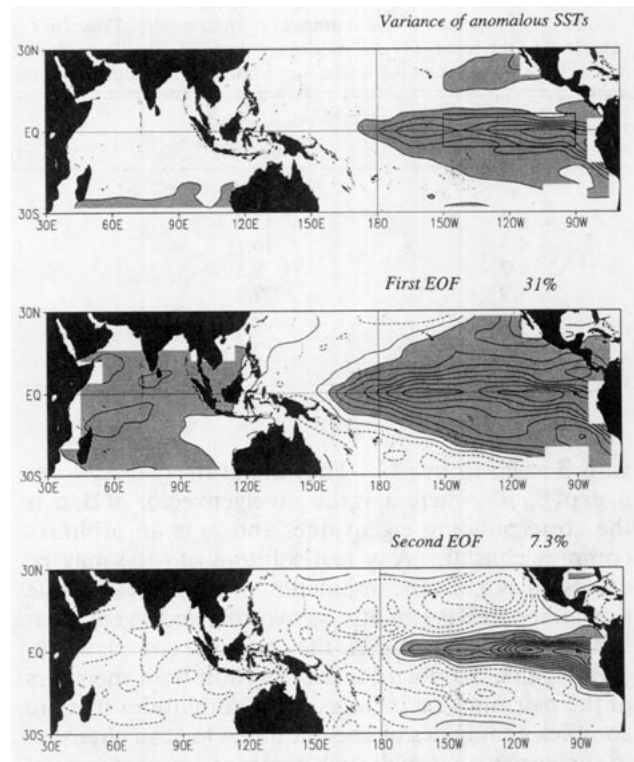

FIG. 1. SST variance associated with the first 15 EOFs of 3-month running mean SST anomalies from the annual cycle. The contour interval is 0.2°C². The outlined box in the eastern Pacific denotes the Niño-3 region. The first two EOFs are shown in the bottom two panels. In these plots positive values are indicated by solid and negative by dashed contours. The contour interval is 0.025 in arbitrary units.

lights the familiar fact that SST variability in the Tropics is dominated by large-scale fluctuations in the equatorial central and eastern Pacific Ocean associated with the ENSO phenomenon. The panels are consistent with the appearance of warm waters first off the coast of Peru, and their subsequent intensification and westward expansion along the equator into the central Pacific, especially when the time series of $x_1(t)$ and $x_2(t)$ (not shown) are taken into account. It will be interesting later in this section to compare EOFs 1 and 2 with the phase-quadrature components of one of our ENMs, and assess to what extent that single ENM captures the essence of ENSO.

b. Calculation of the empirical normal modes

ENMs are the solutions of (2) without the forcing term ξ, that is, of

$$dx/dt = \mathbf{B}x, \qquad (10)$$

TABLE 1. Properties of the empirical normal modes. Properties of the solutions of (10) are given below assuming that they have been expressed in the form (11). All modes are damped. The "growth index" for the mode in the last column is the quantity (13). If the growth index is less than 1 then the mode is not only damped but overdamped; that is, its energy (12) decays monotonically.

Mode	Decay time (months)	Period (months)	$b^T b$	Growth index
1	23.8	∞	—	—
2/3	17.4	365	3.1	0.54
4/5	10.1	159	4.6	0.69
6/7	8.5	46	2.5	0.66 ←
8/9	7.9	25	1.9	0.71 ←
10/11	7.3	85	6.7	0.84
12/13	4.4	18	2.7	0.85
14/15	3.7	72	5.4	0.72 ←

with **B** as given by (4).[1] They are of the general form $u_j \exp(\beta_j t) c_j$, where u_j is the jth eigenvector of **B**, β_j is the corresponding eigenvalue, and c_j is an arbitrary complex constant. Any real solution of (10) may be expressed as a linear combination of these ENMs. Note that although **B** is real, its eigenvectors and eigenvalues can be complex; however, they come in complex conjugate pairs. To obtain a real solution both members of the pair must be included with amplitudes that are complex conjugates of one another. One can choose c_j in such a way that each such combination is expressed as

$$x_j(t) = \{ a_j \cos\omega_j t + b_j \sin\omega_j t \} \exp(\sigma_j t), \quad (11)$$

where σ_j and ω_j are the real and imaginary parts of β_j, the vectors a_j and b_j are real and satisfy the conditions $a_j \cdot b_j = 0$, $a_j \cdot a_j = 1$, and $b_j \cdot b_j \geq 1$. The inner product $a \cdot b$ of two real vectors is defined in general as $a \cdot b = a^T D b$, where **D** is a constant positive-definite real symmetric matrix. The L_2 norm given by $D = I$ will be implied here unless stated otherwise. We will refer to the combination (11) of a complex-conjugate mode pair as a single oscillatory mode with phase-quadrature components a_j and b_j, respectively. Note that one may also represent purely exponentially damped ENMs in the form (11) by setting $\omega_j \equiv 0$ and $b_j \equiv 0$.

For a system with stationary statistics it can be shown (appendix A) that the σ_j's in (11) are negative. Table 1 shows the periods $2\pi/\omega$ and decay times $-1/\sigma$ of the ENMs of our 15-component dynamical system, displayed in order of decreasing decay time (compare with Fig. 2 of Ghil and Vautard 1991). There is one

purely exponential mode (which happens to be the least damped mode in our system) and seven oscillatory modes. The modes with periods of 46, 25, and 72 months are highlighted in Table 1 for future reference. These timescales are reasonably close to those usually associated with ENSO. However, the modal periods are not reliably determined from the relatively short data record. We show here the results for $\tau_0 = 7$ months; repeating the analysis at other lags yields periods for the "46-month" mode, as identified through pattern correlations, between 37 and 59 months, roughly consistent with a perturbation theory estimate (Penland and Ghil 1993; Penland and Sardeshmukh 1995) for this period lying between 28 and 54 months. The "true" period of the "25-month" mode is expected to be between 19 and 30 months, while the uncertainty in the "72-month" mode is so large that the mode is
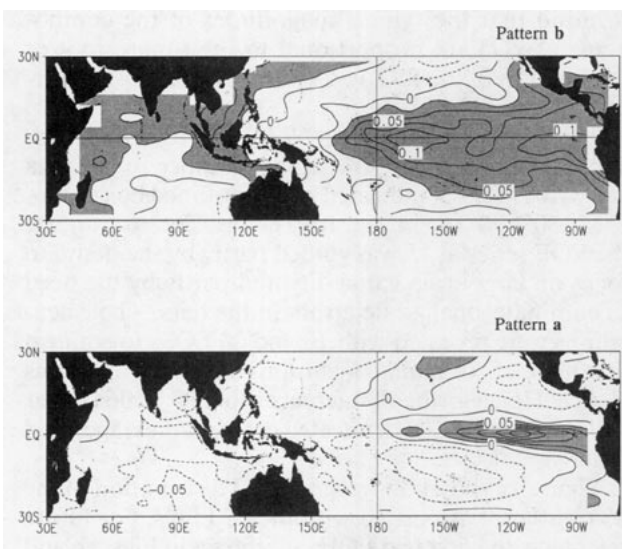

FIG. 2. The phase-quadrature components a and b of mode 6/7 with a period of 46 months and decay time of 8.5 months. The mode evolves as $a \rightarrow b \rightarrow -a \rightarrow -b \rightarrow a$ over 46 months while undergoing decay. The ratio of the spatial rms of the patterns, $(b^T b/a^T a)^{1/2}$, is $(2.5)^{1/2} \sim 1.6$. The contour interval is arbitrary, but is the same in the two panels. Negative values are indicated by dashed contours.

---

[1] These are also referred to as POPs in the literature. The term empirical normal mode is more appropriate in cases where the dynamical system can be shown to be of the form (2) through satisfaction of the requirements (9). Also, not all solutions of (10) are necessarily oscillatory, and as shown later, the ordering of POPs in terms of their damping rates does not always reflect their true importance in the system. It is also not always the case that any one oscillatory mode dominates the system's statistics. The adjective "principal" can therefore be ambiguous.

statistically indistinguishable from an exponentially decaying structure. As tempting as it is, therefore, to assign a harmonic relationship to these modes, it is dangerous to do so.

Mode 6/7, whose estimated period of 46 months is closest to the 4-yr quasi periodicity of ENSO, deserves a closer look. Figure 2 shows its phase quadrature components **a** and **b** in geographical space. The mode evolves from **a** to **b** to −**a** to −**b** to **a** over 46 months. This sequence is in broad agreement with the composite ENSO event described by Rasmusson and Carpenter (1982). If mode 6/7 were the only dynamical mode of importance in the system, the first two SST EOFs would be identical to **b** and **a** and explain $(2.5/3.5) \times 100 = 71\%$ and $(1/3.5) \times 100 = 29\%$ of the SST variance, respectively. The EOFs in Fig. 1 do indeed show some striking similarities to Fig. 2. However, the similarity is superficial. First of all, the two EOFs explain much smaller fractions of the total variance. One can accommodate this by imagining that the system consists of mode 6/7 plus noise. However, even in this scenario, Fig. 2 predicts the ratio of the variance explained by the two EOFs to be $(\mathbf{b}^T\mathbf{b}/\mathbf{a}^T\mathbf{a}) = 2.5$, whereas the actual ratio is $31/7.3 \sim 4.5$. Second, and more importantly, the apparently realistic growth of SST anomalies suggested by the evolution from **a** to **b** is seen to be illusory when the damping time of 8.5 months for the mode is taken into account. In the quarter cycle (11.5 months) that it takes the mode to evolve from **a** to **b**, its amplitude diminishes by a factor $e^{(11.5/8.5)} \sim 3.5$, so no growth is in fact seen.

A freely evolving mode 6/7, despite its apparently realistic period and spatial structures, thus fails to explain the most basic feature of ENSO, that of SST growth. In fact, it can be shown that no single one of our empirically derived modes can do this. In this context it is necessary to be more precise about the term "growth." We define it as an increase of the domain integral of squared SST anomalies; in our framework, of the "energy" $E(t) = \mathbf{x}(t) \cdot \mathbf{x}(t)$. The light curve in Fig. 3 shows the behavior of observed $E(t)$ over 1950–90. For reference, the time series of Niño-3 SSTs is also shown (dark curve) with some prominent ENSO events highlighted. Bearing in mind that $E(t)$ is a positive quantity, its correspondence with the Niño-3 series is high. Thus, ENSO events, both warm and cold, dominate increases of $E(t)$.

Now, because all the modes in Table 1 have $\sigma_j < 0$, $\mathbf{x} = 0$ is a stable fixed point of the system (2), and any initial perturbation that is allowed to evolve freely decays eventually, both locally and globally. Thus, without forcing, every component of $\mathbf{x}(t)$ as well as $E(t)$ tends to zero for large t. However, the decrease need not be monotonic, even for a single mode. One can show from (11) that $E(t)$, for an oscillatory mode, evolves as
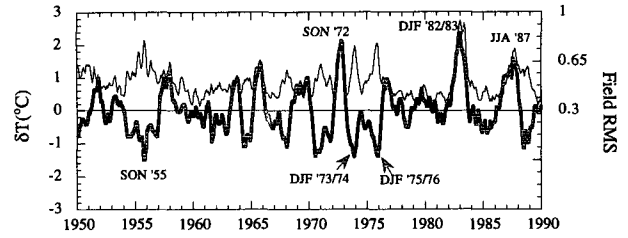

FIG. 3. Thick curve: Time series of 3-month running mean Niño-3 SST anomalies over the period 1950–90. Thin curve: Time series of the spatial rms $E^{1/2}$ of SST anomalies over the same period. The domain over which the spatial rms is calculated is the entire tropical Indo-Pacific basin. The minimum $E^{1/2}$ is 0.33 (°C).

$$E_j(t) = \{\mathbf{a}_j \cdot \mathbf{a}_j \cos^2\omega_j t + \mathbf{b}_j \cdot \mathbf{b}_j \sin^2\omega_j t\} \exp(2\sigma_j t), \tag{12}$$

that is, in the form of a damped oscillation. For small σ_j, the energy oscillates between $\mathbf{a}_j \cdot \mathbf{a}_j$ and $\mathbf{b}_j \cdot \mathbf{b}_j$ while undergoing a slow overall decay; thus, the energy increases between times $t \approx n\pi/\omega_j$ and $t \approx (n + 0.5)\pi/\omega_j$ for $n = 0, 1, 2, 3, \cdots$. There is, therefore, still the possibility of attributing temporary increases of $E(t)$ to the underdamped energy oscillations of a single mode. A necessary condition for such an energy oscillation to occur is that a growth parameter μ_j defined as

$$\mu_j = \{1 + (\omega_j/\sigma_j)^2\}^{1/2}\left[\frac{(\mathbf{b}_j \cdot \mathbf{b}_j - 1)}{(\mathbf{b}_j \cdot \mathbf{b}_j + 1)}\right] \tag{13}$$

be greater than 1. If this condition is not met, the energy oscillation is overdamped and $E_j(t)$ decreases monotonically. This is the case with all of the ENMs in Table 1.

No single unforced mode can therefore explain the growth of SST anomalies in the Indo-Pacific basin. Growth can, however, be explained by the constructive interference of several modes, even though the modes themselves are individually overdamped. This is demonstrated in the following section.

4. The maximum amplification curve

Given overdamped normal modes for each of which E_j decreases monotonically, one may ask if E also has to decrease monotonically for any linear combination of them. The answer is yes if the modes are orthogonal, since then E is just the sum of the individual E_j. However, in most systems of interest such as the one studied here, the normal modes are not orthogonal, so the possibility exists for temporary energy growth from modal interference. The growth or decay over any time interval τ may be written
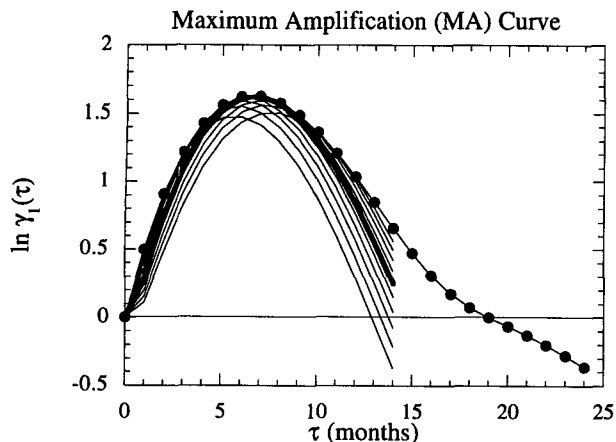
Maximum Amplification (MA) Curve

FIG. 4. (a) Thick curve with filled circles: The maximum amplification curve. Thick curve: Energy amplification of a forecast (6) begun with the initial condition $\phi_1(7)$ optimized to give maximum amplification at 7 months. Light curves: Energy amplification of forecasts begun with initial conditions optimized to give maximum growth at other τs.

$$\mu(\tau) = \frac{E(\tau)}{E(0)} = \frac{\{x(\tau) \cdot x(\tau)\}}{\{x(0) \cdot x(0)\}}$$

$$= \frac{\{G(\tau)x(0) \cdot G(\tau)x(0)\}}{\{x(0) \cdot x(0)\}}$$

$$= \frac{\{x(0) \cdot G^T G(\tau)x(0)\}}{\{x(0) \cdot x(0)\}}, \qquad (14)$$

where $G(\tau) \equiv \exp(B\tau)$ with B as given by (4) and G^T is the transpose of $G(\tau)$. The matrix $G^T G$ is real, symmetric, and positive-definite; therefore, its eigenvectors ϕ_j form a complete orthonormal basis, its eigenvalues γ_j are real and positive and can be arranged in decreasing order. It is evident that $\mu(\tau)$ will be greater than 1 for any $x(0)$ lying in the subspace of ϕ_js with eigenvalues $\gamma_j > 1$. The maximum growth $\{\mu(\tau)\}_{max} = \gamma_1$ will be obtained for $x(0)$ parallel to ϕ_1. This eigenanalysis of $G^T G$ can be repeated for a range of τs, and γ_1 plotted against τ. We call this plot the "Maximum Amplification" (MA) curve of the system, since it quantifies the maximum growth possible over an interval τ in the absence of forcing. Note that growth associated with individual underdamped oscillatory modes is also included in this description. If the maximum of the MA curve is less than 1 (or its logarithm less than 0), then no growth is possible without forcing, and all growth events are unpredictable.

The MA curve for our system is shown as the thick curve with filled circles in Fig. 4. What is actually plotted is the logarithm of $\gamma_1(\tau)$. The curve indicates that no growth can be sustained by modal interference for more than 18 months in this system. However, growth of as much as a factor of $e^{1.7} \sim 5$ is possible over seven months. Associated with each τ in Fig. 4 is an optimal

initial condition $x(0) = \phi_1(\tau)$, the first right singular vector of $G(\tau)$. It is optimal in the sense that of all possible initial conditions of unit amplitude, it evolves into the longest state vector at time $t = \tau$. [It turns out that in this system all other optimals have $\gamma_j(\tau) \ll \gamma_1(\tau)$.] The plain thick curve in Fig. 4 shows the evolution of μ when $x(0) = \phi_1(7)$ is specified as the initial condition in (10). It remains below the MA curve and becomes tangent to it at $t = \tau = 7$ months, as expected. The behavior of μ for several other initial conditions that are optimal for other τs is given by the light curves. These show that although, in general, an initial condition that is optimal for one τ need not be optimal for another τ, in our system it is almost optimal. In fact, we find that the ϕ_1s for τ ranging from 3 to 15 months are spatially correlated with $\phi_1(7)$ at levels of more than 0.9. We shall henceforth refer to $\phi_1(7)$ as the optimal initial structure for growth by modal interference in this system.

As stated above, the MA curve indicates that growth from modal interference cannot be sustained for more than 18 months. This defines a predictability limit of 18 months for growth events, such as a non-ENSO developing into either a warm or a cold ENSO event. In view of the fact, however, that the dynamical system is (2), not (10), this predictability limit is somewhat optimistic. The system loses predictability due to the white noise forcing according to (7). That curve may also be recast as the growth curve of error energy $\mu_{error}(\tau) = 1 + \delta(\tau)$ and compared with the MA curve in Fig. 5 on the same log scale. The time τ_p at which the two curves cross gives a more conservative predictability limit than previously, of 15 months. In effect, this is the forecast lead time at which growth from the unpredictable stochastic forcing is expected to become
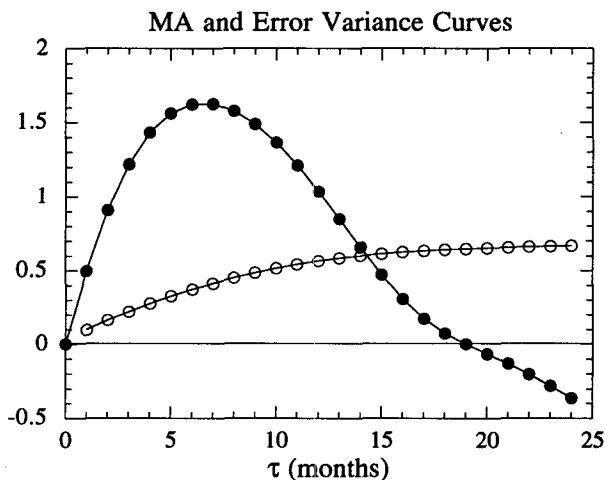
MA and Error Variance Curves

FIG. 5. Filled circles: The maximum amplification curve. Open circles: Logarithm of average error energy growth $1 + \delta(\tau)$ associated with the stochastic forcing.

comparable to that from the predictable modal interference. For ENSO growth to be predictable, it is therefore not only necessary that $\ln\{\gamma_I(\tau)\}$ be greater than 0 over some range of τ, but that it be greater than $\ln\{\mu_{error}(\tau)\}$ over that same range.

Finally, it should be recognized that the MA curve by itself only reveals the *possibility* of growth from modal interference in any system. Whether and how often such growth actually occurs depends upon the likelihood of the state vector projecting on to the subspace of growing optimals at any given time. This issue is examined in the next section.

5. Optimal SST growth

Figure 6a shows the optimal initial SST anomaly structure $\phi_I(7)$ for growth by modal interference in the tropical atmosphere–ocean system. When this is specified as the initial condition in (10), the SST anomaly pattern evolves in seven months into the structure shown in Fig. 6b, strongly reminiscent of the mature phase of an ENSO event. Both Figs. 6a and 6b have been contoured with the same contour interval.

Figure 7 demonstrates that not only is SST growth as indicated by Fig. 6 possible, but that it does in fact often occur. Shown is a scatterplot of the correlation between the SST anomaly pattern with Fig. 6b versus the correlation of the SST anomaly pattern seven months earlier with the optimal structure, Fig. 6a. Evidently, a large-enough projection (correlation greater than 0.3) of an SST pattern on the optimal structure is usually followed by a mature ENSO pattern seven
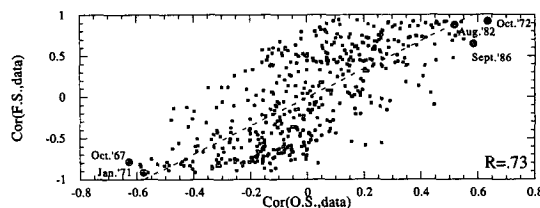

FIG. 7. Scatter diagram of the spatial correlation between Fig. 6b and the SST anomaly field vs the spatial correlation of the SST anomaly pattern seven months earlier with the optimal initial structure for growth (Fig. 6a).

months later. Data points corresponding to the most mature stage of some prominent warm and cold ENSO episodes are highlighted as filled black circles on the plot. Consistent with linear theory, a positive projection of SST on the optimal structure leads to a warm event, a negative projection to a cold event.

A straight line fitted to the scatter plot should obviously have positive slope, and it does. Although the scatter around the fit is large (the line explains about 53% of the variance), this scatter cannot be unambiguously ascribed to failure of the model. Even if there were no stochastic forcing to introduce error into the 7-month forecasts, one would expect to see some scatter in Fig. 7, unless the SST anomaly pattern *always* projected solely onto the optimal structure in Fig. 6a, which it cannot. Besides, a linear stochastic numerical model (discussed below) yields similar scatter. Neither should the fact that ENSOs often follow a projection as modest as 0.3 be taken as evidence against the importance of the optimal structure. We expect the statistically obtained pattern to be a combination of the regions that are sensitive to the excitation of ENSO growth; we do not expect the entire pattern to exhibit itself before every extreme event. To understand this, consider the stable, linear dynamics of a stretched string, the optimal structures of which are equivalent to the normal modes themselves. One need not force the string with a pattern exactly equal to the optimal structure corresponding to, say, the first even harmonic in order to excite it; one need only force the string at any position away from a modal node. Of course, since more than one mode and more than one optimal structure are allowed by that system, other modes will be excited as well. The point is that although the whole pattern of an optimal structure gives those regions that are sensitive to forcing, the projection onto the entire pattern need not be large for its corresponding mode (or modes) to be excited.

Repeating the analysis with 10 EOFs yielded the same optimal structure, with maximum growth occurring over 8 months. This was done in order to verify the robustness of our results. The time series obtained by cutting from the record into two equal, nonoverlapping pieces of 240 months each are not long enough
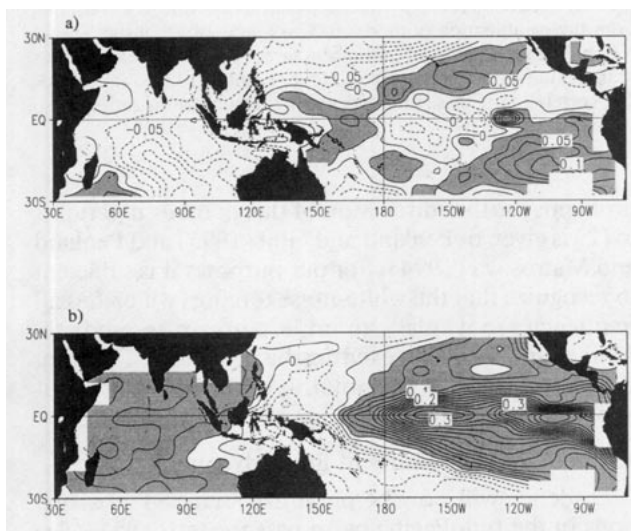

FIG. 6. (a) Top: Optimal initial SST perturbation $\phi_1(7)$ optimized to give maximum amplification of SST anomalies at seven months. When this is specified as the initial condition in (6) [or (10)], the SST anomaly pattern evolves into the pattern shown in the bottom panel (b). The contour interval is arbitrary (0.025) but the same in both panels. Negative values are indicated by dashed contours.

to obtain accurate parameters for a 15-dimensional system. However, acceptable results could be obtained for a 10-dimensional system; that is, the modes were recognizable as those found in the longer record. The prediction errors were larger, sometimes by a factor of 2 at short lead times. Nevertheless, the first short record (February 1950–January 1970) yielded an optimal structure that correlated with that obtained from the 420-month record at 88%. Analogous treatment of the second short record (February 1970–January 1990) produced a pattern that correlated with the one in the longer record at 92%.

The optimal growth shown in Figs. 6 and 7 is associated mostly with the constructive interference of modes 6/7, 8/9, and 14/15 in Table 1. The individual and collective contribution of these modes to Niño-3 SSTs is given in Fig. 8. The individual series by themselves explain only 16%, 8.4%, and 1% of the variance of the Niño-3 time series respectively, but their sum explains 69%. The fact that $16 + 8.4 + 1 \neq 69$ again emphasizes the importance of the spatial nonorthogonality of the normal modes in optimal SST growth.

In addition to growth, modal interference can also shed light on the east–west propagation characteristics of SST anomalies, not well understood at present. The composite event of Rasmusson and Carpenter (1982) emphasizes westward propagation, whereas events in the last decade (including 1982/83, the strongest event of the century) have tended to display eastward propagation (Rasmusson and Wallace 1983; Weickmann and Kiladis 1994). Figure 9 shows longitude–time sections of equatorial SST during 1981–85. The sum of the three modes is shown in panel (a) of Fig. 9, and modes 6/7, 8/9, and 14/15 in panels (b), (c), and (d), respectively. The individual modes show little if any east–west propagation, but their sum shows substantial eastward propagation. Figure 10 depicts the 1971–75 case in an identical format. Again the individual modes are mostly standing oscillations; this time their sum shows strong westward propagation. One can thus obtain eastward or westward propagation depending upon the relative amplitudes of the three interfering normal modes.

It is interesting that the modes that emerge as the most important in our analysis of SST growth and propagation are not the three least damped modes in Table 1. Indeed mode 14/15 is the *most* damped. Something other than growth rates is clearly important in mode selection, and the only other mechanism in our system is stochastic forcing. The influence of the forcing can be seen in the three modal time series in Fig. 8, which would all show a monotonic decrease without it. Because of the forcing, the Fourier spectra of the modal time series are not delta functions centered at the modal frequencies in Table 1. As for any forced oscillator, the spectrum of the forced response depends on both the natural and the forcing frequencies in the
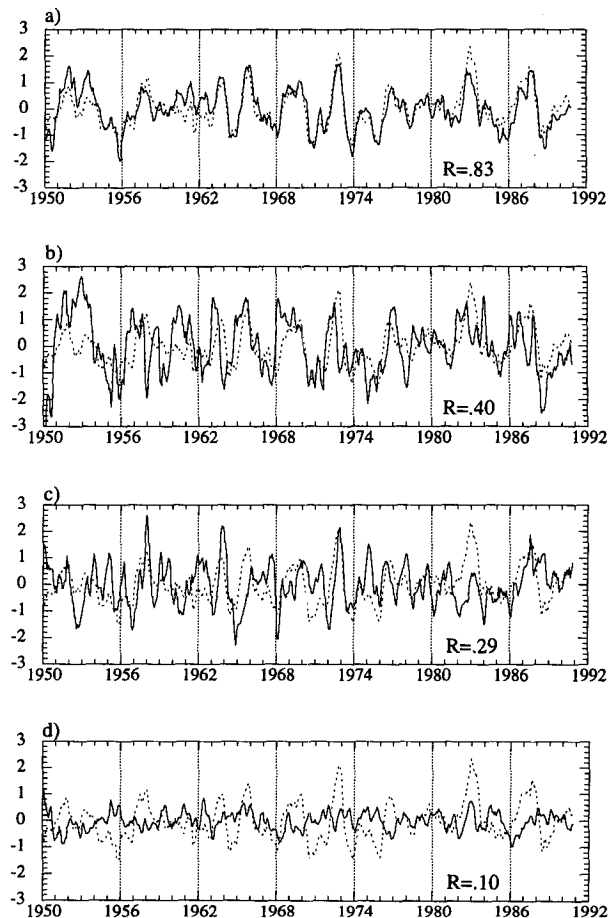

FIG. 8. The dotted curve in all four panels is the time series of the observed 3-month running mean Niño-3 SST anomalies (°C) for 1950–90. The solid curve (a) in the top panel is the time series resulting from the combination of modes 6/7, 8/9, and 14/15 in the Niño-3 region. The solid curves in the bottom three panels are the time series of these individual modes (b) 6/7, (c) 8/9, and (d) 14/15 in Niño-3. The correlation of these time series with the dotted time series is also indicated.

problem. Further discussion of this point as it pertains to (2) is given in Penland and Ghil (1993) and Penland and Matrosova (1994). For our purposes it is sufficient to recognize that the white-noise forcing (which has all frequencies in it) plays an important part in mode selection, and thence in exciting the optimal SST pattern shown in Fig. 6a. A quantitative estimate of this influence is given in section 7.

6. Tests for the validity of linear dynamics

Evidence will now be presented that SST observations in the IndoPacific basin pass the tests (9a)–(9e) within the limitations of the data.

a. Are the SST statistics Gaussian?

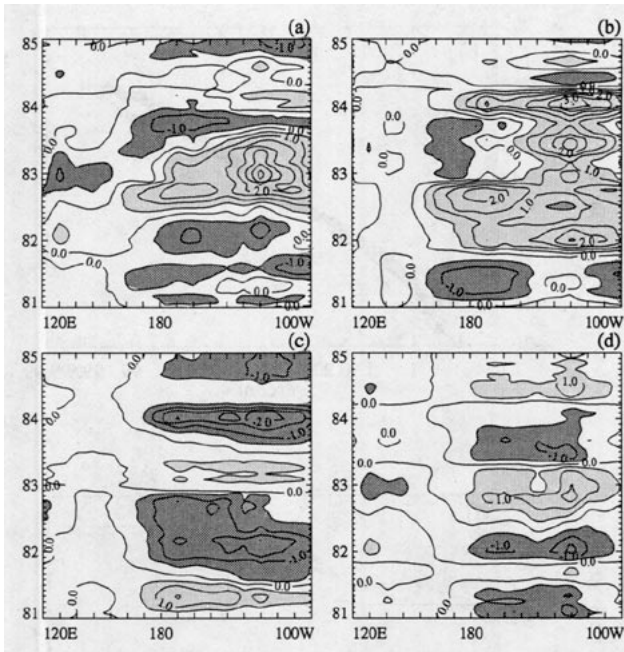Figure 11a (heavy solid line) shows the cumulative distribution function (cdf) of Niño-3 SSTs on a

FIG. 9. Hovmöller plot of SST anomalies averaged over the equatorial belt 2°N–2°S during the 1982–83 warm event obtained from (a) the combination of modes 6/7, 8/9, and 14/15, (b) mode 6/7 alone, (c) mode 8/9 alone, and (d) mode 14/15 alone. The contour interval is 0.5°C; positive anomalies are indicated by light and negative by dark shading.

stretched Gaussian coordinate. The dashed straight line shows the cdf for a normally distributed variable with the same variance. If the distribution of Niño-3 SSTs were truly Gaussian, the points would fall on this straight line. Niño-3 was chosen because it is a region of large SST variability, that is, with a large range of SST anomalies, so presumably the likelihood of the system deviating from small-amplitude linear dynamics would be large there. Further, it is well known that our predictions break down during the warmest phase of the ENSO cycle, and since this phase is strongly represented in the Niño-3 region, we hope that the anomalies there might present a sort of worst-case scenario. A total of 490 monthly values from the 41-yr record were used in the construction of this plot. It is evident that all but 5% of the coldest and 5% of the warmest months fall on the straight line. It is interesting that there is no discernible S-shaped kink in Fig. 11a (see Figs. 11c and 11d below), and therefore little evidence of the bimodality of Niño-3 SSTs sometimes claimed in the literature.

It is also worth mentioning that although Fig. 11a suggests the SST anomalies deviating equally from Gaussian behavior during both extreme warm and cold events, the stretched coordinate system is based on the variance in the entire time series, and inclusion of points deviating from the true Gaussian distribution causes the others to deviate from the empirically de-

termined line. Further, the forecasts based on (6) are substantially better during extreme cold events (Penland and Magorian 1993; Penland and Matrosova 1994, their Fig. 2). In other words, (2) is not as bad an approximation during extreme cold events as might be inferred from Fig. 11a. In fact, although the data shown in Fig. 11a show deviations from the line at both extremes, we believe that it would be more accurate to ignore the warmest 10%–16% and fit the line to the colder data.

Modes 6/7, 8/9, and 14/15 were shown above to be important to the development of the ENSO event, and their contribution to the Niño-3 anomaly appears more Gaussian, particularly during cold events, than did the full Niño-3 anomaly (note again that the straight line in Fig. 11a was fit to the full anomaly). We quantify this further below. Figure 11b shows the cdfs for the real and imaginary parts of normal mode 6; plots for modes 8 and 14 (not shown) are similar to Fig. 11b. The first two maps, corresponding to January–March 1950 and February–April 1950, although represented in Fig. 11b (the two largest negative points on each curve), did not contain sufficient data to project accurately onto the EOF field in the optimal sense defined by Reynolds and Livezey (1995; see also Xu 1993). We repeated all of our calculations without these two maps and found no discernible differences in the optimal structure or its growth rate; neither was there any difference in the modal makeup of the field, except that the period of the modal pair 14/15 was reduced to 55 months. For this modal pair, the decay time is so small as to make its much longer period almost
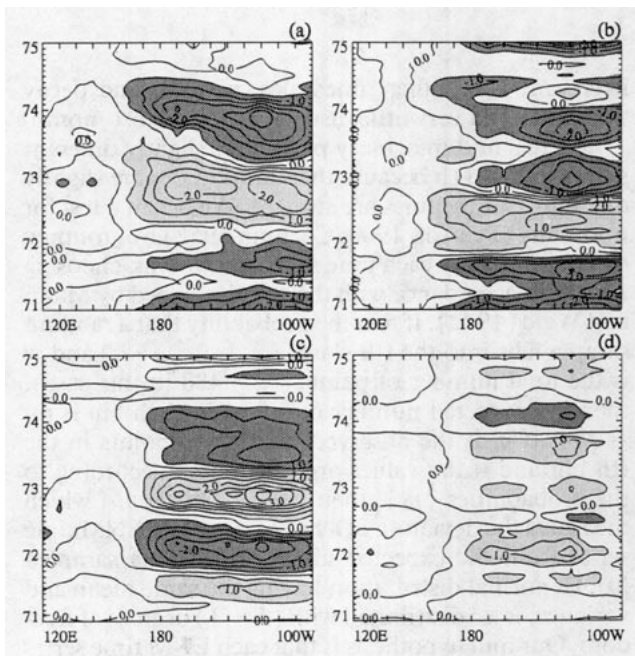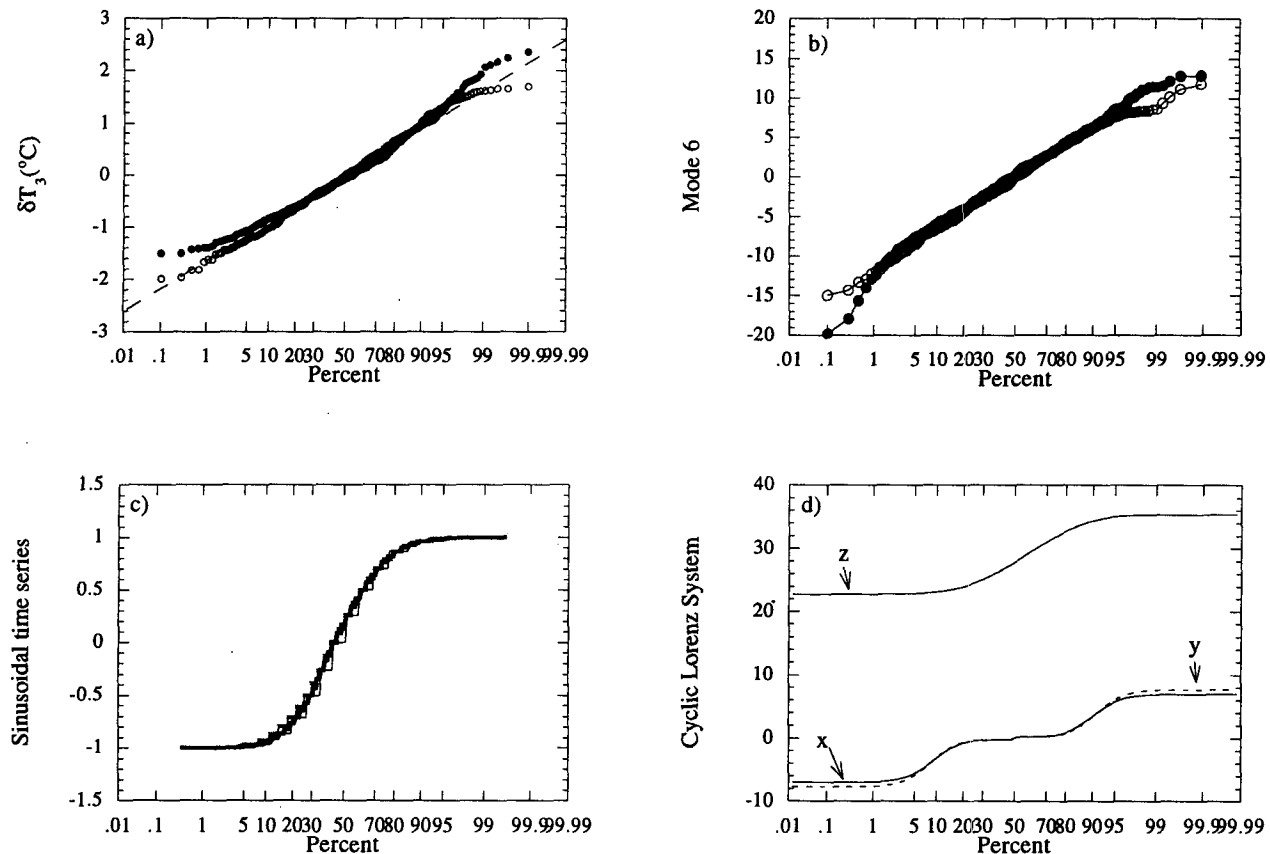

FIG. 10. As in Fig. 9 but for the 1972–73 event.

FIG. 11. (a) Cumulative probability distribution (cdf) of 3-month running mean Niño-3 SST anomalies (°C) over the period 1950–90 (filled circles). The dashed straight line is the cdf for a normally distributed variable with the same variance as the Niño-3 SSTs. Also shown (open circles) is the contribution to the Niño-3 anomaly from modes 6/7, 8/9, and 14/15. Analogous cdfs are shown for (b) mode 6 (filled circles: real part; open circles: imaginary part), (c) circular limit cycles with periods of 2, 4, and 6 yr, and (d) the x, y, and z components of the Lorenz attractor in the limit cycle regime.

irrelevant. The other timescales, periods, and decay times changed very little, usually by less than 1 month.

The real and imaginary parts have slightly different slopes in Fig. 11b because their amplitudes are slightly different, but the lines are straight. We began a test for normality (Pearson 1900; Cochran 1952) by grouping the amplitudes of each time series into K bins, choosing $K = 45$ in accordance with the rule developed by Mann and Wald (1942). If p_i is the probability that a random sample falls into the ith class ($i = 1, 2, \cdots K$) and N is the total number of points ($N = 488$ in this case), then the expected number of points in each bin is $m_i = N p_i$. If n_i is the observed number of points in the ith bin and if the values are distributed according to the probabilities $\{p_i\}$, then $\sum (n_i - m_i)^2 / m_i$, which measures the deviation of a histogram of an ENM time series from the expected histogram of data sampled from a normal distribution having the same mean and variance, is x^2 distributed with ($K - 1$) degrees of freedom. Our null hypothesis is that each ENM time series is normally distributed around its sample mean (close

to zero) with its sample variance. Accordingly, we estimate p_i as $p(x_i)\Delta x$, where $p(x_i)$ is the normal distribution evaluated at x_i, the center of the ith bin, and Δx is the bin width. The limits of the distribution were chosen as plus and minus the smallest integer larger than the maximum amplitude of the time series. Values of x^2 for the real and imaginary time series coefficients of modes 6, 8, and 14 are shown in Table 2, along with the probability that a histogram with 45 bins obtained from random samplings of a normal distribution would result in a value of x^2 larger than that obtained from the ENM coefficient time series. For example, 76% of an ensemble of histograms generated from samplings of a normal distribution having the same mean and variance as the time series coefficients of $\mathrm{Re}(u_6)$ would be expected to have values of x^2 greater than 37.03.

The ideal value of x^2 is zero, but the expected value is 44 and its expected variance is 88. The sample mean of our six time series is 46.56, which is well within one standard deviation of the expected mean. The calculated variance around the sample mean is 71.02, and

TABLE 2. Test for normality of ENM time series. Values of χ^2 measuring deviations of histograms of ENM coefficient time series from expected histograms of normal distributions having the same mean and variance (see text for details). The probability that a histogram with 45 bins obtained from random samplings of a normal distribution would result in a value of χ^2 larger than that obtained from the ENM coefficient time series is given as $1 - P(\chi^2)$, where $P(\chi^2)$ is the cumulative distribution. For example, 76% of an ensemble of histograms with 45 bins generated from samplings of a normal distribution having the same mean and variance as the time series coefficients of $Re(u_6)$ would be expected to have values of χ^2 greater than 37.03. The ideal value of χ^2 is zero, but the expected value is 44.

	χ^2	$1 - P(\chi^2)$	
$Re(u_6)$	37.03	76%	
$Im(u_6)$	56.32	10%	
$Re(u_8)$	46.98	35%	
$Im(u_8)$	37.88	73%	
$Re(u_{14})$	45.05	43%	
$Im(u_{14})$	56.11	10%	
Sample mean	46.56	Expected mean	44.00
Variance around sample mean	71.02	Expected variance	88.00
Variance around expected mean	78.89		

the calculated variance around the *expected* mean is 78.89, both values being within the expected variance of 88. Similar numbers (not shown) are obtained from analyzing the output of a numerical model that is linear by construction. Note that although the hypothesis of normality of any single ENM coefficient time series need not be rejected at the 95% level, only 5% of all appropriate normal distributions occurring in nature would pass this test anyway, and the probability of all six passing that test is orders of magnitude smaller.

Considering the contribution to the Niño-3 anomaly from these three modal pairs (Fig. 11a, open circles), we found a value of $\chi^2 = 48$, a value that would be superceded by 32% of all appropriate Gaussian processes. In contrast, the Niño-3 anomaly from all of the normal modes (Fig. 11a, filled circles) has a value of $\chi^2 = 73$, which would be superceded by fewer than 1% of Gaussians having the same mean and variance. The departure from normality is due to the strong positive skew in the histogram (not shown), consistent with our contention that the model breaks down during warm events. These χ^2 results therefore further support the hypothesis that much of the relevant dynamics are dominated by Gaussian dynamics except during warm events. At the very least, the hypothesis cannot be ruled out at any meaningful level.

We end this section by comparing Figs. 11a and 11b with cdfs corresponding to systems with circular (i.e., sinusoidal time series; Fig. 11c) and highly noncircular limit cycles (Fig. 11d). Three sinusoidal time series with periods of two, four, and six years were generated and sampled at monthly intervals. As seen in Fig. 11c,

their S-shaped cdfs lie on top of one another in this projection. Fig. 11d shows the cdfs for the x, y, and z components of the Lorenz system in the regime where its attractor is a stable limit cycle ($\sigma = 10$, $r = 30$, $b = 0.25$, but otherwise as discussed in appendix B) (see also Sparrow 1982). The kinks in these distributions, not evident in Figs. 11a and 11b, are clear.

b. Is **B** independent of the lag τ_0 that is chosen to determine it?

This is a hard test, and failure does not imply nonlinearity (appendix B). Thirteen estimates of **B** (which in EOF space is a 15×15 matrix) were generated from (4) using data for the period February 1950–January 1985 and $\tau_0 = 3, 4, 5, \cdots, 15$ months, respectively.
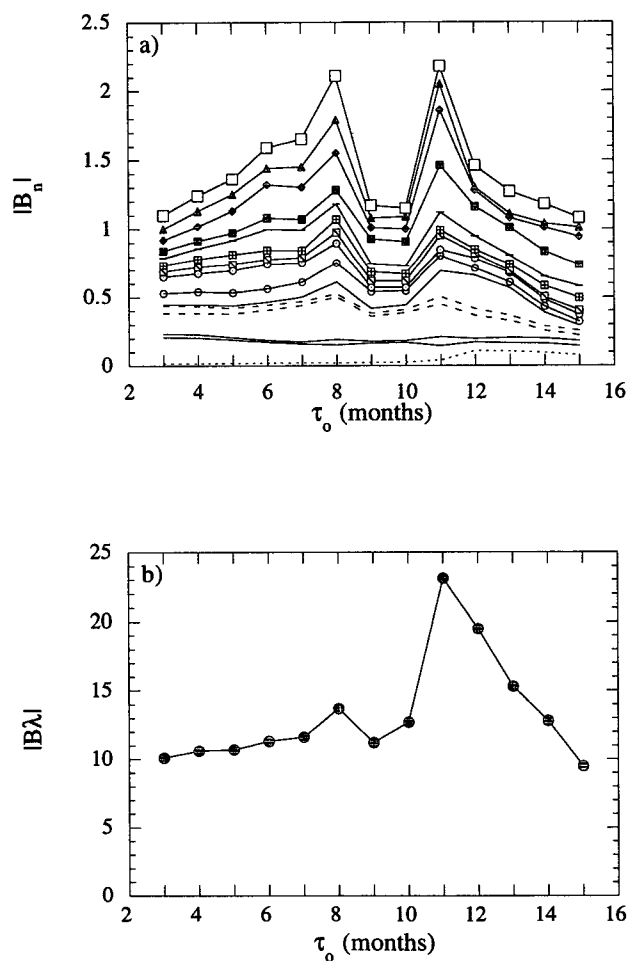

FIG. 12. (a) Top curve: The Euclidean norms of **B** obtained using different lags τ_0 (in months) in Eq. (4). The 14 curves below it are the norms of the 14 submatrices **B**$_n$ of decreasing rank, each with first element B_{11}. Thus, the bottom curve shows **B**$_I = B_{11}$ as a function of τ_0. (b) The magnitude of **B** times a typical constant vector λ as a function of τ_0. See text for further explanation.

The top curve in Fig. 12a shows the Euclidean norm of these matrices as a function of τ_0. Clearly the curve is not flat. What matters in (2), however, is not so much **B** itself as **Bx**, and since the higher-indexed components of **x** have typically smaller magnitudes (in EOF space), the 15th column of **B** is less important than say the 4th column and should be given less weight in the comparisons. One way to do this is to compare the submatrices $\mathbf{B}_n$ of **B** of rank $n < 15$, starting with the upper left corner: $\mathbf{B}_1 = [B_{11}]$, which acts on x_1; $\mathbf{B}_2 = [B_{11}, B_{12}; B_{21}, B_{22}]$, which acts on x_1 and x_2; and so on. The 15 curves in Fig. 12a show the Euclidean norms of these submatrices as a function of τ_0. Note that $\mathbf{B}_{15} = \mathbf{B}$. It is evident that starting with the $n = 1$ curve at the bottom, the curves remain approximately flat until about $n = 8$, although some slope in the curve is seen starting with $n = 4$. This is a consequence of a restriction on the inverse modeling procedure, discussed in appendix B, analogous to restricting Fourier analysis to frequencies lower than the Nyquist frequency. POP analysis and related procedures involving lagged covariances can give erroneous results if performed at lags greater than half the smallest oscillation period in the system, and large uncertainties can occur when the lag is near this half-period. From Table 1 it is seen that one modal pair has a period of 18 months, and the analysis at $\tau_0 = 9$ months yields an eigenvalue corresponding to this mode with $\mathrm{Im}(\beta_\alpha \tau_0)$ equal to π. This particular modal pair does not have much variance, and so its effect on the product **Bx** is small. However, modal pair 8/9, with a period of 25 months, is very important to the system, so the large uncertainties obtained using $\tau_0 = 11$ months cannot be ignored.

Figure 12b depicts this in another way. A "typical" state vector $\underline{\lambda}$ is defined as one with components proportional to the square root of the SST variance explained by the corresponding EOF; the figure shows the length of $\mathbf{B}\underline{\lambda}$ as a function of τ_0. The curve is nearly flat up to $\tau_0 = 10$. However, as in Fig. 12a, $\tau_0 = 11$ is an outlier, and the curve for larger values of τ_0 is not flat.

c. Is **Q** determined from (8) positive definite?

The 13 leading eigenvalues of **Q** are positive; two eigenvalues are negative. Two negative eigenvalues are also found when the analysis is repeated with 20 SST EOFs instead of 15. However, using 10 EOFs, all eigenvalues are positive. Further analysis of this case yields an optimal structure very similar to the one found with 15 EOFs and also dominated by three ENSO-type modes with periods of about 30, 40, and 60–80 months (see also Penland and Magorian 1993, their Table 1). This result can be interpreted as implying either that the higher-indexed EOFs are not determined reliably from the COADS data, or that the trial model (2) is only approximately valid, with non-

linearities represented primarily in the high-index, low-variance EOFs. Both interpretations are likely. We note in passing that Penland and Matrosova (1994) neglect the noise EOFs associated with the two negative eigenvalues of **Q**, and report no discernible adverse effect on their results.

d. Are the forecasts based on (6) good?

For this test it is important that the training and verification periods be separate. As already stated, **B** was determined (using $\tau_0 = 7$ months) from the first 420 months of data in the complete 490-month record. Forecasts were then made using the remaining data (February 1985–November 1990).
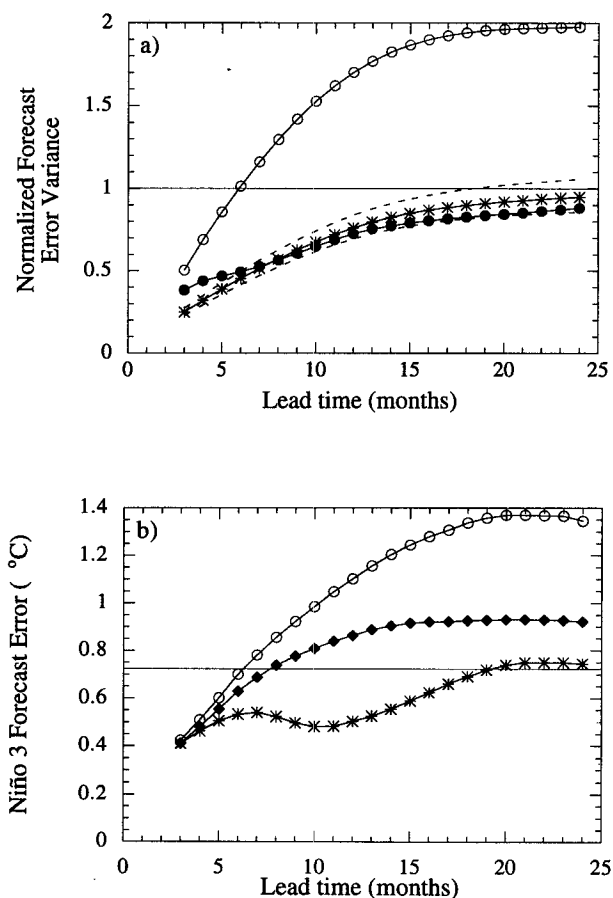

FIG. 13. (a) Filled circles: Normalized domain-integrated error variance of SST forecasts based on Eq. (6) as a function of forecast lead time. Asterisks: Error variance predicted by Eq. (7). Open circles: Error variance of persistence forecasts. Horizontal line: Error variance of climatology forecasts. Dashed lines: 95% confidence interval appropriate to the theoretical error curve. See text for further explanation. (b) Rms prediction error of the Niño-3 anomaly as a function of lead time. Asterisks: Rms error of linear model forecasts. Filled diamonds: Rms error of forecasts based on a univariate autoregressive process. Open circles: Rms error of persistence forecasts. Horizontal line: Rms error of climatology forecasts.

The filled circles in Fig. 13a show the global forecast error variance δ as a function of forecast lead time, normalized with respect to the domain-integrated SST variance. As the data have been subjected to a 3-month running mean, only lead times of 3 months and larger are considered. The curve represents an ensemble average of 47 to 68 forecasts, depending on lead time, and is to be compared with the theoretically expected error curve (asterisks). Also shown is the 95% confidence interval (dashed lines) around the expected error curve for independent predictions made by the linear process for the number of forecasts made at a given lead time (e.g., for 47 forecasts at a lead time of 24 months, 48 forecasts at a lead time of 23 months, etc.). Of course, the 3-month running mean means that adjacent predictions are *not* independent, and the number of degrees of freedom is therefore reduced by a factor of 3. Taking this into account increases the width of the 95% confidence interval by about 60%. We have chosen the more stringent confidence interval to provide a more conservative test of fitness. Because the eigenvalues of $\mathbf{B}$ have negative real parts, the eigenvalues of $\mathbf{G}(\tau)$ are smaller than unity and approach zero at large τ. The forecast at long lead times is thus a zero anomaly forecast, that is, climatology. Thus, δ increases from 0 to 1. The curve with open circles shows the average error of persistence forecasts. This increases from 0 to 2, because at long lead times the error approaches the expected difference between two arbitrary state vectors instead of the difference between an arbitrary state vector and climatology. The solid horizontal line shows the expected error of climatology forecasts, which remains constant at 1 because it is the expected difference between an arbitrary state vector and climatology.

The forecasts based on (6) are clearly much better than persistence forecasts. They are also much better than climatology for lead times of less than about 15 months. The forecasts for Niño-3 SSTAs (Fig. 13b) display essentially the same level of skill and may be compared with forecasts performed using a univariate AR model (dashed line in Fig. 13b). In fact, the forecasts made with the inverse model (2) are of comparable quality to those made with the best "forward" dynamical models at present (see, for example, the National Oceanic and Atmospheric Administration's *Experimental Long-Lead Forecast Bulletin*).

e. Do the forecast errors grow as predicted by (7)?

In Fig. 13a, the curve with asterisks is the theoretically predicted error growth curve. It closely approximates the actual error growth curve and, except for lead times of five months or less, the differences are statistically insignificant. (The difference between the theoretical and observed error curves in a system dominated by nonlinearities is much larger; see appendix

B.) Nevertheless, a tendency for the actual error to be larger than that predicted at short lead times and smaller at long lead times is evident. The larger, statistically significant error at short lead times is most probably due to the inability of the system to make accurate predictions when initialized during the warm phases of ENSO. The smaller error at long lead times is harder to explain. It is possible that it reflects a slight overestimate of the negative eigenvalues of $\mathbf{B}$ obtained by specifying $\tau_0 = 7$ months in (4). Note that the warm phases of ENSO, in which the linear system has greatest difficulty, are included in the training statistics. An improperly fitted linear model can be expected to give degraded forecasts sooner than the true decay rates would indicate. On the other hand, Penland and Magorian (1993), whose system was truncated to 10 EOFs, found that eliminating warm events from the training statistics resulted in slightly degraded predictions of *cold* events. Therefore, either some of the stable, linear processes continue to operate during the warm phases, or eliminating 14% of the training period resulted in less accurate parameters, or both.

In sum, the requirements (9a)–(9e) are reasonably well met by the model (2) for slowly evolving (3-month running mean) tropical SST anomalies. This enables a meaningful discussion of the normal modes of the real physical system as the free solutions of (2).

7. The role of the forcing in exciting the optimal structure

The system matrix $\mathbf{B}$ estimated from (4) and the forcing matrix $\mathbf{Q}$ estimated from (8) can be used to generate a time-dependent empirical dynamical model of tropical SSTs of the form (2). As discussed by Penland and Matrosova (1994), the appropriate method of integrating (2) when using a finite time step Δt is a two step process:

$$\mathbf{y}(t + \Delta t) = (\mathbf{I} + \mathbf{B}\Delta t)\mathbf{y}(t) + (\Delta t)^{1/2}\mathbf{S}\mathbf{r}(t) \quad (15a)$$

$$\mathbf{x}(t + \Delta t/2) = \{\mathbf{y}(t + \Delta t) + \mathbf{y}(t)\}/2 \quad (15b)$$

with $\mathbf{y}(0) = \mathbf{0}$ (see also Kloeden and Platen 1992; Kloeden et al. 1994). Here $\mathbf{r}$ is a vector whose components are independent Gaussian deviates, each with unit variance, that is generated once at each time step. The term $\mathbf{S}$ is a constant matrix whose columns are the eigenvectors $\{\psi_\alpha\}$ of $\mathbf{Q}$ multiplied by the square root of the corresponding eigenvalues $\{q_\alpha\}$. Thus, $\mathbf{Q} = \mathbf{\Psi}\mathbf{q}\mathbf{\Psi}^T = \{\mathbf{\Psi}\mathbf{q}^{1/2}\}\{\mathbf{\Psi}\mathbf{q}^{1/2}\}^T = \mathbf{S}\mathbf{S}^T$.

The model was exercised for 10 yr and output collected for the subsequent 2000 yr. The time step was 3 h. The covariance matrices $\mathbf{C}_{model}(\tau_0)$ and $\mathbf{C}_{model}(0)$ were estimated from model output and confirmed to reproduce the observed $\mathbf{C}(\tau_0)$ and $\mathbf{C}_{model}(0)$ used to determine the $\mathbf{B}$ and $\mathbf{S}$ matrices specified in (15). In some modeling exercises, one might wish to specify $\mathbf{B}$ and $\mathbf{S}$

and test for dynamical consistency when an observed $\mathbf{C}(0)$ is not available. In this case, $\mathbf{C}_{model}(0)$ should be compared with the covariance matrix consistent with the fluctuation–dissipation relation (appendix C). The numerical algorithm (15) is carefully designed to ensure consistency with (8), provided a small enough time step is used. We find a time step of 3 hours to be necessary. It should be noted that this is much smaller than the time step required for computational stability using a forward time step in this problem. By construction, then, our model is guaranteed to reproduce the observed $\mathbf{C}(0)$ and $\mathbf{C}(\tau_0 = 7)$. However, it is not guaranteed to reproduce coherent phenomena, such as a realistically evolving ENSO, as well. The latter is nevertheless true as shown by Penland and Matrosova (1994) and further attests to the validity of (2). We obtain essentially the same results using 15 SST EOFs, as opposed to their 20 EOFs. As in their analysis, we use the time series δT_3 of model Niño-3 SSTs to characterize the model output and its sensitivity to modifications in the white noise forcing. Warm events in the model are identified as upward excursions of δT_3 across a 1-standard deviation threshold with the proviso that δT_3 dip below $-0.4°C$ between events. A composite event derived from them (not shown, but see Penland and Matrosova 1994) has many features in common with a composite event derived from observations. Furthermore, a composite map 7 months before peak warming correlates with the optimal structure in Fig. 6a at 85%. Figure 14 shows a scatter plot analogous to Fig. 7 for the first 2000 months of the model run. Clearly, the large scatter in Fig. 7 is not inconsistent with a linear, stochastically driven system. Note that the nearly identical fit (slope = 1.73 vs 1.75) and scatter (correlation of data with fit = 0.73 vs 0.71) in Figs. 7 and 14 are not guaranteed by our generation of the model from the observed $\mathbf{C}(\tau_0)$ and $\mathbf{C}(0)$.

Agreement of statistics by itself does not guarantee realistic realizations, and so we present three 40-yr segments of the Niño-3 sea surface temperature anomaly (SSTA) from the model run along with the true Niño-3 record (Figs. 15a–d). Some readers will quickly identify which plot is the real one: first because of the relative location of the warm events, and second because we did not subject our model output to a three-month running mean since the empirically derived parameters are supposed to account for that. However, the verisimilitude of the model output is good.

As stated previously, our principal aim in this section was to determine the influence of the forcing on warm ENSO events through its excitation of the optimal initial structure for SST growth. The noise EOFs with the largest projection on $\phi_1(7)$ are $\psi_3(0.68)$ and $\psi_5(0.30)$ (not shown). These explain 12.3% and 6.9% of the total noise variance, respectively. A modified white-noise forcing was generated by setting the third and
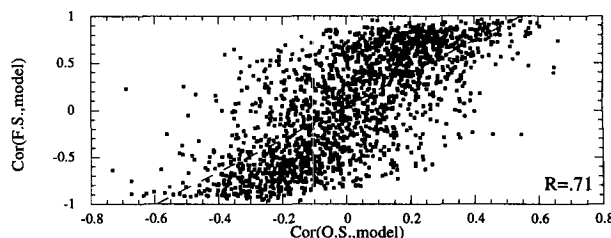

FIG. 14. Scatter diagram of the spatial correlation between Fig. 6b and the SST anomaly field as generated by the linear model vs the spatial correlation of the model's SST anomaly pattern seven months earlier with the optimal initial structure for growth (Fig. 6a). The slope and scatter are nearly identical to that of Fig. 7.

fifth columns of $\mathbf{S}$ to zero in (15) and enhancing the other columns by a constant factor so as to retain the same total forcing variance. The model was then rerun for 2000 years with the modified forcing.

Figure 16 shows the amplitudes of the real and imaginary parts of modes 6, 8, and 14 for both model runs; it is seen that the effect of the modified forcing is to reduce the basinwide warming as described by these modes. These effects may seem somewhat modest. This is partly by design, since we chose only to reduce (by about 20%) but not eliminate the noise projection on the optimal structure in the modified run. The idea was to modify the forcing just enough to alter the statistics of the modeled ENSO events without radically affecting the character of the events themselves.

8. The annual cycle

The accuracy of our prediction method is far more susceptible to the phase of the ENSO cycle than it is to the phase of the annual cycle. In particular, accurate predictions through the spring season are common (Penland and Magorian 1993; Penland and Matrosova 1994). To further investigate the annual cycle, we divided our verification period (February 1985–November 1990) into seasons and plotted the forecast error variance as a function of lead time for initial conditions taken from each season (not shown). The error during the verification period is dominated by the 1986–87 warm event, and a large variation of the error with season at short lead times is observed, primarily due to the seasons in which the 1986–87 warm event was strong (specific cases are shown in Penland and Matrosova 1994). On the other hand, the 5-yr verification period is too short to draw any unambiguous conclusion about seasonality. Verification on the training set obviously cannot give a good idea of model accuracy; however, the model parameters were calculated regardless of season, so one might get a hint of the seasonal dependence of model accuracy by calculating error curves stratified by season for the entire dataset.
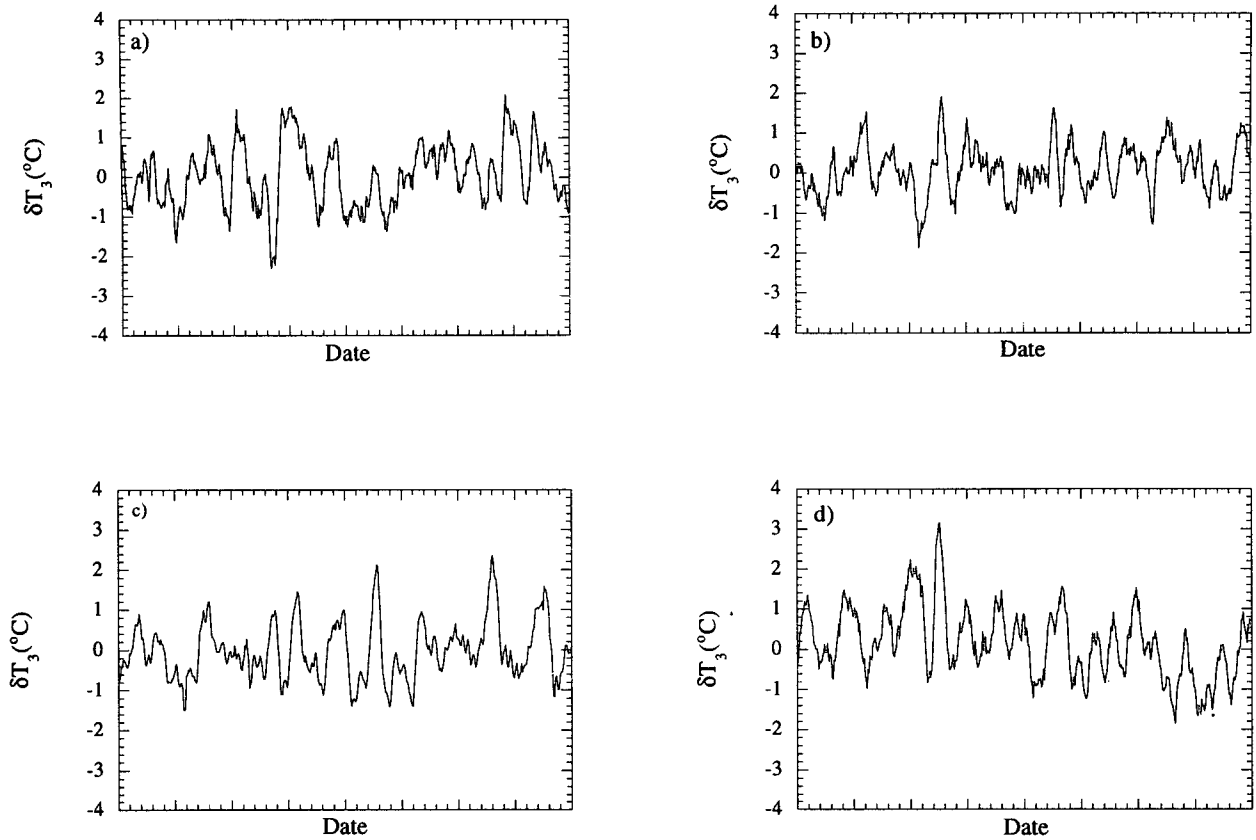
FIG. 15. (a)–(d) Three 40-yr segments of the Niño-3 SST anomaly calculated from output generated by the linear model. Also shown is the measured record. Which is which?

This was done, and the error variance was normalized by the stationary variance, that is, the amount of variance contained in *all* the seasons during the verification period (Fig. 17a). The seasonality of the error curves is weak at short lead times; again, examination of individual cases shows that the phase of the ENSO cycle dominates the error. However, Fig. 17a does show some influence of the annual cycle at large lead times ($\tau > 1$ yr).

It turns out that the annual cycle seen in Fig. 17a at large lead times is due to the stationary normalization. When the normalizing variance appropriate for the season of the verifications is specified, the annual cycle disappears (Fig. 17b). Consistent with linear dynamics, this implies that while the variance of the field does indeed contain an annual cycle, this measure of the accuracy of the predictions does not.

A stronger presence of the annual cycle is felt in the frequency of occurrence of the optimal structure. Figure 18 shows the number of times the size of the projection of the optimal structure onto the SSTA's in any particular month exceeds 0.3 in our dataset. For clarity, two cycles are exhibited. Although every month is rep-

resented, it is clear that the optimal structure occurs primarily in the boreal spring.

These results do not contradict our basic premise of a linear stable system driven by stochastic forcing; neither do they contradict the validity of the τ-test. However, a modification is in order. If one assumes that Eq. (2) holds, but where $\langle \underline{\xi}(t)\underline{\xi}(t')^{\mathrm{T}} \rangle = \mathbf{Q}(t)\delta(t - t')$ and $\mathbf{Q}(t)$ is periodic, then the method of obtaining $\mathbf{B}$ is unchanged as long as the estimated statistics are accumulated over an integral number of periods. This is why we chose our training period to be exactly 35 years. The fluctuation–dissipation theorem is generalized to

$$d\langle \mathbf{x}\mathbf{x}^{\mathrm{T}} \rangle / dt = \mathbf{B}\langle \mathbf{x}\mathbf{x}^{\mathrm{T}} \rangle + \langle \mathbf{x}\mathbf{x}^{\mathrm{T}} \rangle \mathbf{B}^{\mathrm{T}} + \mathbf{Q}(t), \quad (16)$$

so that Eq. (8) holds as long as $\mathbf{C}(0)$ and $\mathbf{Q}$ are interpreted as averages over the period. That is, the eigenvectors of $\mathbf{Q}$ as calculated from Eq. (8) have survived the annual averaging process. In particular, the prediction method [Eq. (6)] is unchanged. It can be shown that the probability density for the system $\mathbf{x}(t + \tau)$ conditioned on an initial condition $\mathbf{x}'(t)$ is

$$p(\mathbf{x}, t + \tau | \mathbf{x}', t) = (2\pi)^{-N/2}[\mathrm{Det}\Sigma(t, \tau)]^{-1/2}$$

$$\times \exp\{-\tfrac{1}{2}(\mathbf{x} - \mathbf{G}(\tau)\mathbf{x}')^{\mathrm{T}}\Sigma(t, \tau)^{-1}(\mathbf{x} - \mathbf{G}(\tau)\mathbf{x}')\}, \quad (17)$$
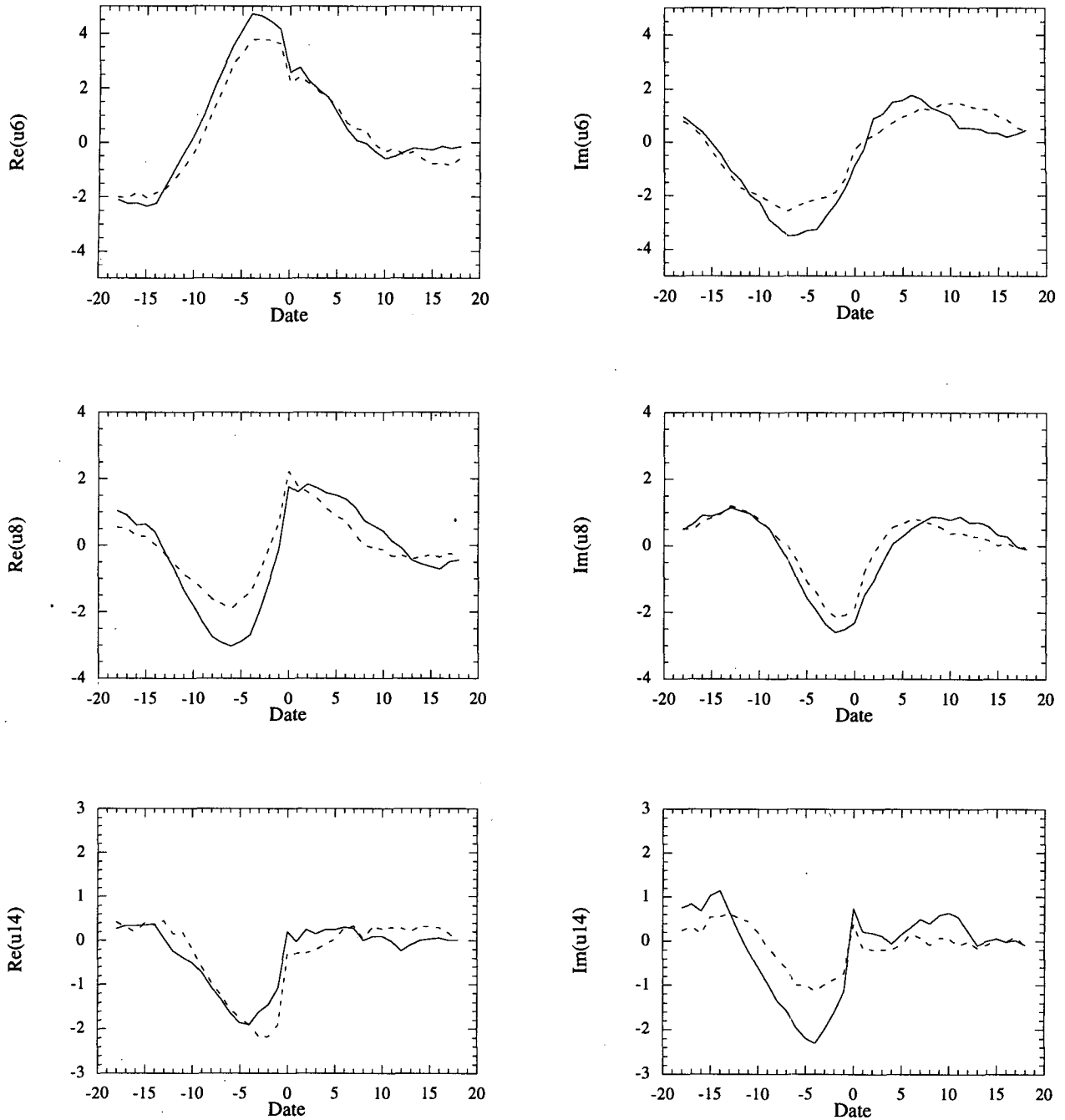
FIG. 16. Amplitude of normal modes during the composite warm event. Solid line: Full model. Dotted line: Modified forcing.

where

$$\Sigma(t, \tau) = \left\langle e(t, \tau)e(t, \tau)^{\mathrm{T}} \right\rangle = \left\langle x(t + \tau)x(t + \tau)^{\mathrm{T}} \right\rangle$$

$$- G(\tau)\langle x(t)x(t)^{\mathrm{T}}\rangle G(\tau)^{\mathrm{T}}. \quad (18)$$

The probability density [Eq. (17)] is of the same form as that for the stationary case and is maximized

at $x(t + \tau) = G(\tau)x'(t)$, with $G(\tau)$ independent of t. That is, the τ-test still holds. Note that Eq. (18) allows periodicity in the predictability, which could allow for the small annual cycle seen in Fig. 17. A more detailed study of the role of the annual cycle in the propagation of SST anomalies is a topic of current research.
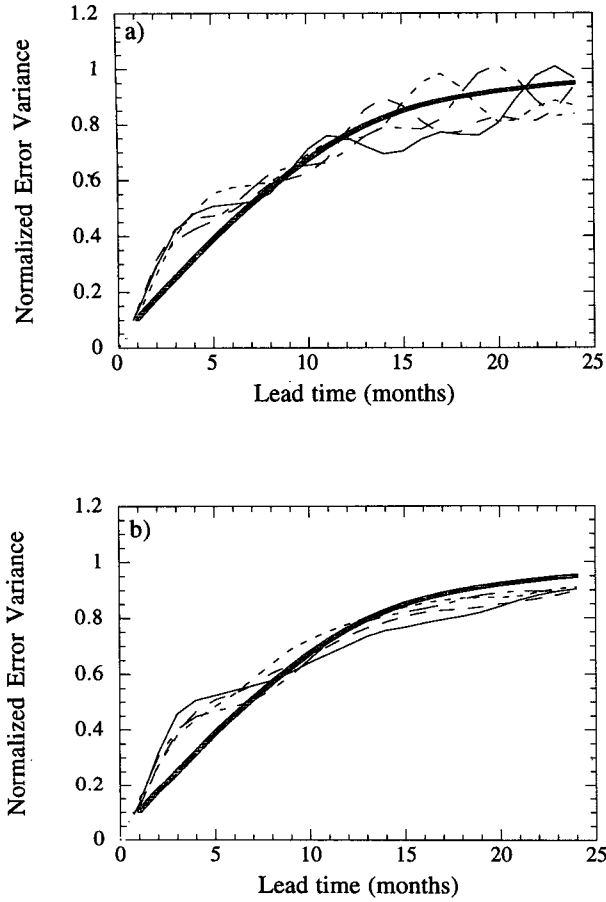
FIG. 17. Error variance stratified by season of the initial condition. Light solid line: December–January–February initial conditions. Dot-slash line: March–April–May initial conditions. Short dashes: June–July–August initial conditions. Long dashes: September–October–November initial conditions. Heavy solid line: Theoretical curve for the stationary process. (a) Errors calculated from the entire record normalized to the stationary variance. (b) Errors calculated from the entire record normalized to the seasonal variance.

9. Discussion

The inverse modeling analysis of this paper can be put in the context of the forward dynamical models by considering the evolution equation for SSTs in the form

$$T_t = -uT_x - vT_y - wT_z + F. \tag{19}$$

Here T is the temperature of the surface ocean layer, whose changes are determined by advection by the ocean current velocities u, v, and w in the x, y, and z directions and by F. This last term represents the combined effect of surface sensible and latent heat fluxes, shortwave and longwave radiation, as well as sub-grid-scale diffusive processes. In many simple models F is parameterized in terms of T. Further, the vertical derivative is simplified as

$$T_z = \{ T - T_{\text{sub}} \}/H = \frac{\{ T - (T_s + \Gamma h) \}}{H}, \tag{20}$$

where T_{sub} is a characteristic subsurface temperature at depth H, which is communicated to the surface by the upwelling w. The term T_{sub} is itself parameterized in some models (e.g., Neelin 1991) in terms of thermocline displacements h as $T_s + \Gamma h$, where T_s is a characteristic thermocline temperature and Γ a constant of order 0.1 K m^{-1}. Writing all variables as climatology plus an anomaly $a = \bar{a} + a'$, the equation for the SST anomalies may be written

$$T_t = -\bar{u}T_x - \bar{v}T_y - \bar{w}T/H + F(T) \tag{21a}$$

$$- u\bar{T}_x - v\bar{T}_y - w\bar{T}_z + \bar{w}\Gamma h/H \tag{21b}$$

$$+ \text{ nonlinear terms}, \tag{21c}$$

where the primes on the anomalies have been dropped for convenience. The terms on the right-hand side of (21a) are already of the form $B_l(T)$, where B_l is a linear nonlocal spatial operator. The anomalous currents and thermocline depth in (21b) evolve in response to the anomalous surface winds V, which are themselves a response to T. One thus needs an ocean model to obtain (u, v, w, h) from V and an atmospheric model to obtain V from T. If on the timescales of ENSO evolution the winds adjust quickly to SSTs and the ocean currents and thermocline depth adjust quickly to winds, then one can represent (21b) as $A_l(\text{V}) = A_l(A_2(T)) = B_2(T)$. Defining $B(T) \equiv B_l(T) + B_2(T)$, (21) becomes

$$T_t = B(T) + \text{ nonlinear terms}$$

$$= B(T) + \xi, \tag{22}$$

where the nonlinear terms have been approximated as a white noise forcing ξ. The nonlinear terms here
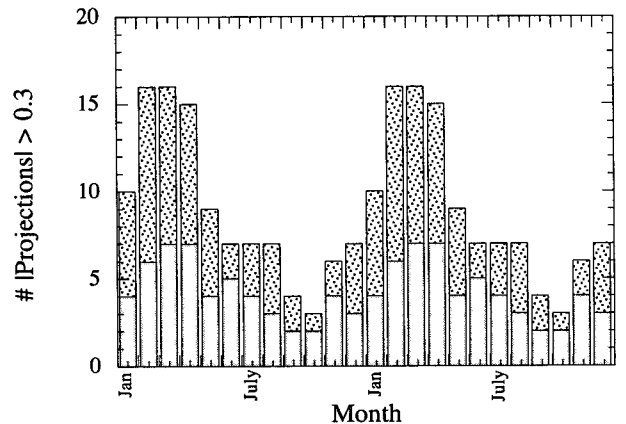

FIG. 18. Number of times the size of the projection by the optimal structure onto the SST anomaly field exceeds 0.3 in any month. For clarity, two cycles are shown. Coarse shading: positive projection. Fine shading: negative projection.

should be interpreted as the sum of the nonlinear terms in (21c) plus the error made in parameterizing (21b) as a linear nonlocal spatial operator $B_2(T)$. Equation (22) is of the form (2).

One can thus derive (2) from (21) through a series of well-defined approximations. It should be stressed that this derivation of (2) is not unique; in fact, it is possible to derive (2) assuming that SSTs adjust quickly to changes in ocean dynamics (the so-called fast SST limit) rather than vice versa. The important question is whether (22) is valid. The observations as analyzed in this paper suggest that it is. The five conditions (9a)–(9e) for the validity of (22) are all reasonably well satisfied. The occurrence of optimal growth as predicted by the eigenstructure of a matrix derived from **B** provides further evidence in support of (22).

Although we have not given here the explicit form of **B** in terms of climatological variables, we have demonstrated its relevance in the tropical climate system and determined some of its important properties from the observations. Any forward dynamical model of the system should have its equivalent **B** whose eigenstructure should match that of the observed **B**. It is not difficult to see how **B** could be misrepresented. Its accurate representation depends upon accurate mean ocean currents and temperature gradients, an accurate ocean anomaly model A_1, and an atmospheric anomaly model A_2. Note that A_2 also affects the formulation of $F(T)$ in (21a), rendering it nonlocal. This is because the anomalous latent heat flux, which makes a large contribution to $F(T)$, depends not only on anomalous ocean temperatures T but also on anomalous winds **V** $= A_2(T)$.

As discussed in several studies (e.g., Neelin 1991; Barnett et al. 1991; Lau et al. 1992), the precise balance of terms in (21) is delicate, but can nevertheless strongly affect the dynamics of SST evolution. Neelin (1991) argues for the general importance of unstable SST modes [the free solutions of (22)] in the system over that of coupled Kelvin and Rossby modes on the grounds that although their propagation is sensitive to which terms dominate in (21), their instability itself is not. Our conclusion differs from his in one fundamental respect. This is that the SST modal solutions are stable. This then necessitates the consideration of modal interference, optimal structures, and forcing.

Some further insight can be gained by going back to the formalism of equations (1) and (2). Again, we consider the "fast-wave limit," where oceanic variables adjust quickly to SSTs. Again, this limit is chosen for exposition purposes and is not crucial for the validity of our dynamical interpretation. Let x_T denote SSTs [the same as x in (2)], x_a denote atmospheric variables, and x_O denote oceanic variables other than SST. Then one may represent the anomaly equations in the coupled system as

$$\frac{dx_T}{dt} = L_{TA}x_a + L_{TO}x_O + L_{TT}x_T + \xi_T$$

$$\text{SST equation} \quad (23a)$$

$$\frac{dx_a}{dt} = L_{AA}x_a + L_{AO}x_O + L_{AT}x_T + \xi_a$$

$$\text{Atmospheric model} \quad (23b)$$

$$\frac{dx_O}{dt} = L_{OA}x_a + L_{OO}x_O + L_{OT}x_T + \xi_O$$

$$\text{Ocean model.} \quad (23c)$$

If the atmosphere experiences the effect of ocean dynamics only through the SSTs, then L_{AO} is zero. Assuming that on the long timescales of SST evolution,

$$\frac{dx_a}{dt} = 0 \quad \text{winds adjust quickly to SST, and} \quad (24a)$$

$$\frac{dx_O}{dt} = 0 \quad \text{currents and thermocline depth}$$

$$\text{adjust quickly to winds,} \quad (24b)$$

enables one to eliminate the ocean variables and write

$$\frac{dx_T}{dt} = L_{ta}x_a + L_{tt}x_T + \xi'_T \quad (25a)$$

$$\frac{dx_a}{dt} = L_{AA}x_a + L_{AT}x_T + \xi_a = 0, \quad (25b)$$

where

$$L_{tt} = L_{TT} - L_{TO}L_{OO}^{-1}L_{OT},$$

$$L_{ta} = L_{TA} - L_{TO}L_{OO}^{-1}L_{OA},$$

$$\xi'_T = \xi_T - L_{TO}L_{OO}^{-1}\xi_O.$$

Equations (25) embody the Bjerknes hypothesis in its simplest form, that is, that winds affect SSTs, and SSTs affect winds. Invoking (24a) now enables the atmospheric variables to be eliminated as well, giving

$$dx_T/dt = Bx_T + \xi, \quad (26)$$

where $B = \{L_{tt} - L_{ta}L_{AA}^{-1}L_{AT}\}$ and $\xi = \xi_T - L_{TO}L_{OO}^{-1}\xi_O - L_{ta}L_{AA}^{-1}\xi_a$. Thus, one recovers equation (2).

Now, the facts that (i) atmospheric models forced with prescribed SSTs give reasonably realistic winds, and (ii) ocean models forced with prescribed winds give reasonably realistic SSTs suggest that the operators L_{AA}, L_{tt}, L_{ta}, and L_{AT} in (25) are reasonably well known. This also suggests that the stochastic forcing term is generally of less importance than $L_{ta}x_a$ in the SST Eq. (25a). The coupled system, however, has much longer timescales than either of the uncoupled systems, if the limits (24a) and (24b) are relevant. This indicates a substantial cancellation between the two component

parts of **B** in (26), and a strong sensitivity of forward coupled models to model parameters.

At present, most forward coupled systems do not include ξ. But they do include nonlinearities of the type represented in (21c). It is easy to imagine how this can be crucially important. If the model's **B** has negative eigenvalues (as our verifying tests suggest), then without ξ the system tends to die, that is, to go permanently to a steady state. This encourages a modeler to tune his model, effectively, we suspect, until **B** has positive eigenvalues. Now the system is unstable, with unstable growth arrested by nonlinearities, so ENSO-like vacillations can be sustained even in a long run without external forcing. By ignoring the stochastic forcing, one can thus end up with the paradigm of a chaotic nonlinear coupled system (e.g., Jin et al. 1994) as opposed to a stochastically driven stable linear system suggested by the observations.

10. Concluding remarks

It has been argued that the tropical climate system can be approximated as a stochastically driven damped linear system on interannual timescales, with the evolution of SST governed by a stable linear operator **B** and a spatially coherent white noise forcing ξ. The properties of both have been determined from the observations, and several tests have been applied to see whether the assumptions inherent in the inverse modeling actually hold. The two most important properties of **B** are that it is not self-adjoint and that its eigenvalues have negative real parts. Thus, the normal modes of the system are evanescent and, without forcing, every SST perturbation eventually decays. Perturbation growth is nevertheless possible in the short term from the modes interfering constructively with one another. An optimal initial structure giving the maximum possible growth by this mechanism has been determined theoretically, and its relevance in the real physical system has been demonstrated. The unpredictable forcing plays a secondary role while optimal growth is in progress. In fact this is why growth is predictable at all, up to 15 months ahead. The forcing is nevertheless important in setting up the optimal structure and in determining the recurrence time of ENSO events. It is also crucial for keeping the system alive in the long run.

The optimal structure and the forcing patterns (the noise EOFs) that are highly correlated with it have strong loadings in the south tropical Indian Ocean and the South China Sea. Using cross-spectral analysis, K. M. Weickmann (1994, personal communication) has found SST anomalies in these areas to act has skillful predictors of SSTA in the eastern equatorial Pacific at a lead time of about two seasons. The proximity of these locations to the subtropical jets must invite speculation that extratropical atmospheric processes play

some role in triggering ENSO events (see also Barnett et al. 1989).

The geographical coherence of the stochastic forcing is important, and so it is not surprising that Zebiak (1989) found little effect when a stochastic forcing, which did not account for the geographical coherence demanded by the fluctuation–dissipation relation, was added to a model already tuned to give interannual variability. Note also that the geographical properties of a temporally white stochastic forcing cannot be reliably obtained by high-pass filtering the data. Even if the patterns obtained in this manner are provided with random coefficients so that their time series are white, one must still contend with the dispersion relation. Since high frequencies are generally associated with high wavenumbers, the forcing obtained in this way is not likely to project significantly onto the adjoints of the low-frequency modes, and is therefore unlikely to excite those modes significantly. A model driven by this forcing, therefore, will need nonlinearities to generate low-frequency variability, not necessarily as a mirror of the real system, but by construction of the model.

While a case can be made in support of linearity, that is, of our model (2), it is important not to overstate it. Conditions (9a)–(9e) are satisfied, but only approximately. The scatter in Fig. 7 is not small. However, this scatter is also produced in a numerical simulation of the stochastic linear system (15), so the scatter by itself cannot be used as evidence against the model.

It should be noted that the optimal structure shown in Fig. 6a is in fact not unfamiliar. Some of its features, such as warm SSTAs in the southeastern equatorial Pacific and a secondary maximum of warm SSTAs on the equator at about 120°W, are recognizable in the precursors to the mature phase of composite ENSO events in several observational and model studies (e.g., Rasmusson and Carpenter 1982; Lau et al. 1992). Barnett's (1991) complex EOF (CEOF) analysis of the quasibiennial band shows negative SST anomalies in the Indian Ocean and the South China Sea and over Indonesia, as well as positive anomalies stretching from the equatorial western Pacific toward Baja California (although these latter are displaced northward in comparison to those in our optimal structure) six months before maximum warm SSTAs in the eastern Pacific. Similar analysis of a low-frequency band shows (in the opposite phase) a cool eastern Pacific about six months before maximum cold Pacific SSTAs. However, the Indian Ocean and the eastern Pacific in the low-frequency CEOF are too cool to be consistent with our structure in its negative phase. Still, since his study and ours both used COADS data, we believe our analysis explains his precursors as manifestations of the optimal initial structure for SST growth. An optimal structure found in the output of the Lamont

coupled model (Xue et al. 1994) agrees with the observed structure in that it has a cool–warm dipole in the eastern equatorial Pacific. However, their pattern has very little else in common with the observed one; in particular, the narrow extent of their warm eastern Pacific precursor is unrealistic, and there is no hint of the band of warm water stretching from the equatorial western Pacific (160°–170°E) up to Baja California. Xu and von Storch (1990) obtained an SST "conjugate correlation pattern" associated with ENSO based on its relationship to a POP found in tropical sea level pressure. Their SST precursor pattern had some of the characteristics of our optimal structure, except that the southeastern tropical Pacific SSTAs were of the same (rather than of opposite) sign as the Indian Ocean in their pattern. We speculate that this may be explained by their filtering out all variance on timescales shorter than 15 months. Again, we believe our results explain their precursor pattern as being related to the optimal anomaly pattern for nonmodal growth of SSTs.

The annual cycle, which has not been treated explicitly in constructing the inverse model (2), is nevertheless present implicitly. To the extent that the anomaly amplitudes tend to be largest during a particular season, the covariance matrices in (5) emphasize that season in all quantities derived from them. Thus, the optimal structure shown in Fig. 6a tends to occur more often during the boreal spring than other seasons, consistent with the fact that the mature phase of ENSO occurs more often during the (following) winter. The obvious annual cycle in the occurrence of the optimal structure (Fig. 18) does not at all compromise our dynamical interpretations; we find only a weak annual cycle in the predictability (Fig. 17b). This may be contrasted with the necessity to include the annual cycle in the predictors when POP analysis is applied to some dynamical models (e.g., Blumenthal 1991; Balmaseda et al. 1994; Xue et al. 1994). However, more detailed work on the role of the annual cycle is necessary and is currently underway.

Our study indicates that prediction skill is more dependent on the phase of the ENSO cycle than it is on the phase of the annual cycle, but this phenomenon is not unique to our stable linear system. In particular, Davey et al. (1993) have found that the seasonal dependence of forecast skill in a simple nonlinear delayed-oscillator system is indirect; the seasonal variation of skill in that model depends on the phase locking of ENSO to the annual cycle.

The inverse modeling analysis of this paper should be viewed as nothing more or less than a dynamically meaningful summary of the observations. One could think of fitting a more complicated model than (2), say (23), to the observations. However, it has been shown that (2) is already a reasonable approximation. This has enabled an estimation of the normal modes

of the system. The most important result from this analysis is the observational evidence that one cannot explain ENSO growth in terms of any single unforced normal mode.

We end by offering the following scenario of ENSO evolution, a scenario that is as yet unproven but fits the observations considered here: ENSO involves at least three stable normal modes, and SST anomaly growth occurs from their constructive interference after an optimal nudge from the stochastic forcing. When SSTAs become warm enough, instabilities do occur and dominate the ENSO cycle until nonlinear saturation is attained. The oceanic basin then cools until the system becomes stable again, and stable linear dynamics dominate the subsequent evolution, including any cold event, until stochastic forcing sets up the next warm episode. Note that this scenario implies that the relevant dynamical system hovers fairly near the critical point for stability, and that the regimes of stable, nonmodal growth or unstable, modal growth may be affected by a one or two degree temperature change in climatology.

APPENDIX A

Empirically Derived Eigenvalues of the Linear Operator

By "stationary statistics" we mean that the covariance matrices in (5) do not depend upon the definition of the time origin. To show that the eigenvalues of **B** derived from (4) have negative real parts, it is sufficient to show that the eigenvalues of

$$\mathbf{G}(\tau) \equiv \exp(\mathbf{B}\tau)$$
$$= \langle \mathbf{x}(t+\tau)\mathbf{x}^T(t) \rangle \langle \mathbf{x}(t)\mathbf{x}^T(t) \rangle^{-1} \quad (A1)$$

are less than unity, that is, $|g_\alpha| < 1$.

Let $\mathbf{e}(\tau) = \mathbf{x}(t+\tau) - \mathbf{G}\mathbf{x}(t)$. Then

$$\langle \mathbf{e}(\tau)\mathbf{e}(\tau)^T \rangle$$
$$= \langle \{\mathbf{x}(t+\tau) - \mathbf{G}\mathbf{x}(t)\} \{\mathbf{x}(t+\tau) - \mathbf{G}\mathbf{x}(t)\}^T \rangle$$
$$= \langle \mathbf{x}(t+\tau)\mathbf{x}^T(t+\tau) \rangle - \mathbf{G}\langle \mathbf{x}(t)\mathbf{x}^T(t+\tau) \rangle$$
$$\quad - \langle \mathbf{x}(t+\tau)\mathbf{x}^T(t) \rangle \mathbf{G}^T + \mathbf{G}\langle \mathbf{x}(t)\mathbf{x}^T(t) \rangle \mathbf{G}^T$$
$$= \langle \mathbf{x}(t)\mathbf{x}^T(t) \rangle - \mathbf{G}\langle \mathbf{x}(t)\mathbf{x}^T(t) \rangle \mathbf{G}^T,$$

where use has been made of $\langle x(t + \tau)x^T(t)\rangle = G(\tau)\langle x(t)x^T(t)\rangle$ and $\langle x(t + \tau)x(t + \tau)^T\rangle = \langle x(t)x(\tau)^T\rangle$.

Consider now the eigenstructure of G and G^T:

$$G(\tau)u = ug(\tau), \quad G^T(\tau)v = vg(\tau),$$

$$\text{and} \quad uv^T = u^Tv = I.$$

In mode space,

$$v^T\langle ee^T\rangle v*$$

$$= v^T\langle xx^T\rangle v* - v^TGuv^T\langle xx^T\rangle v*u^\dagger G^Tv*,$$

where the superscript $\dagger$ denotes a Hermitian conjugate. Thus, in mode space,

$$\langle \epsilon\epsilon^\dagger\rangle = \langle dd^\dagger\rangle - g\langle dd^\dagger\rangle g^\dagger.$$

Now consider the αth element along the diagonal:

$$\langle \epsilon_\alpha\epsilon_\alpha^\dagger\rangle = \langle d_\alpha d_\alpha^\dagger\rangle - g_\alpha\langle d_\alpha d_\alpha^\dagger\rangle g_\alpha^\dagger$$

$$= (1 - g_\alpha g_\alpha^\dagger)\langle d_\alpha d_\alpha^\dagger\rangle,$$

which gives $\langle |\epsilon_\alpha|^2\rangle = (1 - |g_\alpha|^2)\langle |d_\alpha|^2\rangle$. But this is only true if $|g_\alpha|$ is less than 1.

APPENDIX B

The Tau Test

The "τ-test" is a check that the logarithm of the eigenvalues of the Green function

$$G = \langle x(t + \tau_0)x^T(t)\rangle\langle x(t)x^T(t)\rangle^{-1}$$

vary linearly with lead time τ_0. Equivalently, the matrix $B = \tau_0^{-1}\ln G$ should be independent of τ_0. This test is prone to a "Type I error" (rejecting the hypothesis of linearity when it is true). Among other reasons, a system can be expected to fail the τ-test if 1) the system consists of fewer time series than variables are needed to describe it, 2) lead times longer than half the shortest period of oscillation are chosen, or 3) nonlinear dynamics dominate. As an example of the first case, consider the two-component linear system,

$$dx_1/dt = x_2, \quad dx_2/dt = -\omega^2 x_1, \quad \text{(B1)}$$

but where only the time series $x_1(t) = A\sin(\omega t + \phi)$ is provided for the construction of the inverse model. Calculating the sample Green function at a lead time of τ_0 gives

$$\langle x(t + \tau_0)x(t)\rangle/\langle x^2(t)\rangle = \cos(\omega\tau_0), \quad \text{(B2)}$$

which obviously fails the τ-test. However, the τ-test is satisfied if another sinusoidal time series with the same frequency as the first but somewhat out of phase is provided, and the method is able to identify the oscillation frequency since the eigenvalues of the sample Green function are $\exp(\pm i\omega\tau_0)$, regardless of the phase (except phases of $n\pi$). The uncertainty in the identi-

fication depends, of course, upon the oscillation frequency and the length of the time series. One would expect, for example, a small but spurious decay rate that would decrease with the length of the sample, unless one were lucky enough to have a time series exactly equal to an integral number of oscillation periods.

The property that excessive EOF truncation results in τ_0-dependent system matrices B (an error of $O(\tau_0)$: Penland and Ghil 1993) is a further example of case 1. In fact, a τ_0-dependent B can be expected any time the dimensionality of the measured system is insufficient to account for its dynamics, as the case of the sinusoid discussed above exemplifies.

Lead times should be chosen less than $1/2$ the shortest period of oscillation, and the length of the time series necessary to identify accurately the oscillation period and its associated normal mode increases as τ_0 approaches the half period. When the lead time is close enough to the half period that it corresponds to a period near the spectral peak of the damped mode, the mode acquires an imaginary part $\text{Im}(\beta_\alpha\tau_0)$, approximately equal to π. At larger lead times τ_0, harmonics of the oscillation period rather than the fundamental are yielded by the analysis. In this way, POP analysis is restricted in much the same way that Fourier analysis is, with the half-period of the shortest modal oscillation analogous to the Nyquist period.

To see this effect, consider the submatrices B_n of B as described in the text. The 15 curves in Fig. 12a show the Euclidean norms of these submatrices as a function of τ_0. Two peaks are evident: one at $\tau_0 = 8$ months and the other at $\tau_0 = 11$ months, near the half periods of the 18- and 24-month modes. Similar peaked behavior is observed in a sample time series of equal length obtained from a stable, linear model driven by stochastic forcing (discussed in section 7) and, in fact, this peak can be as far as three sample intervals smaller than the half period if the corresponding decay time is very small. Fortunately, the τ-test applied to the Green function, or, equivalently, the prediction error curve, is not so dependent on sampling.

As an example of the third case, we consider the τ-test applied to the three-component Lorenz system:

$$dx/dt = \sigma(y - x) \quad \text{(B2a)}$$

$$dy/dt = rx - y - xz \quad \text{(B2b)}$$

$$dz/dt = xy - bz. \quad \text{(B2c)}$$

As in Penland (1989), the parameters $\sigma = 10$, $r = 28$, and $b = 8/3$ were chosen to ensure a chaotic system. The method of integration is explained in Penland (1989); the time step was taken to be 0.001 dimensionless time units (tu) and the system was sampled every 20 time steps for a total of 4000 samples. The training set, consisting of half the time series, samples both lobes of the attractor. In fact, Penland (1989)
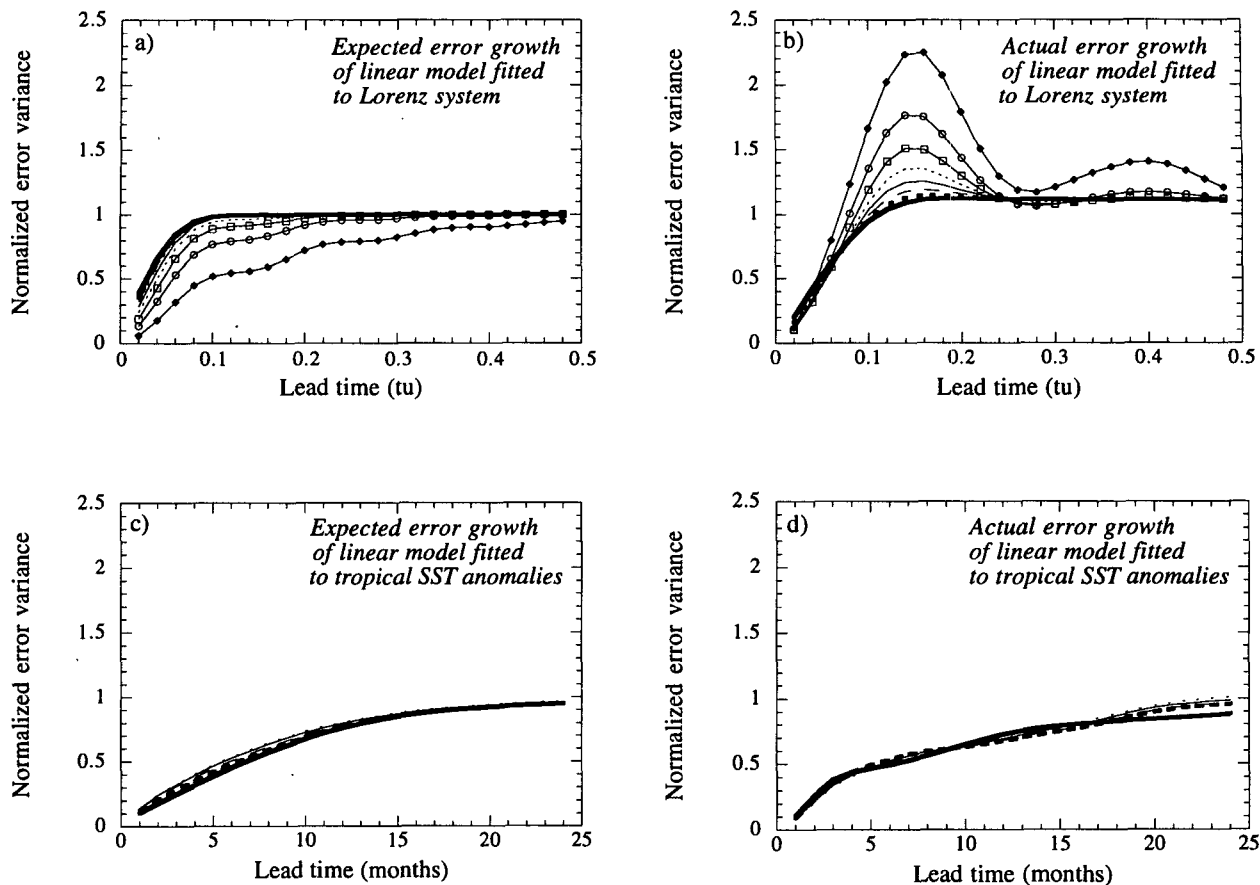
FIG. B1. (a) Theoretical and (b) calculated error variance for the Lorenz attractor as a function of lead time for τ_0 = 0.02tu (filled diamonds), 0.04tu (open circles), 0.06tu (open squares), 0.08tu (light dotted line), 0.10tu (light solid line), 0.12tu (dashed line), 0.14tu (heavy dotted line), and 0.16tu (heavy solid line). (c) Theoretical and (d) calculated error variance for COADS SST anomalies as a function of lead time for τ_0 = 3 months (light dotted line), 4 months (light solid line), 5 months (dashed line), 6 months (heavy dotted line), and 7 months (heavy solid line).

found that the chaotic Lorenz system was better at masquerading itself as a stable linear system if only one finely sampled lobe was considered, and perhaps the tangent linear analysis considered, for example, by Lacarra and Talagrand (1988) would be appropriate there. However, as we hope that our 35-yr sampled SSTA system is representative of the entire ENSO attractor, be it fixed point, limit cycle, or strange, the appropriate comparison is with a sampling of the Lorenz system that experiences both lobes of the attractor.

Applying the inverse-modeling technique to the Lorenz model output yielded an eigenvalue with an imaginary part equal to π at a sampling of τ_0 = 0.2tu. Accordingly, the analysis was performed for lags in the interval (0.0tu, 0.2tu) at τ_0 = 0.02tu, 0.04tu, 0.06tu, 0.08tu, 0.10tu, 0.12tu, 0.14tu, and 0.16tu. The theoretically expected error curves for the various τ_0s is shown in Fig. B1a, and the corresponding

error curves obtained from trying to predict the behavior of the second half of the series is shown in Fig. B1b. The variation with τ_0 is severe. For comparison, we show on the same scale analogous error curves taken from the SSTA system (Figs. B1c,d) corresponding to lags in the interval (0 months, 9 months) at τ_0 = 3, 4, 5, 6, and 7 months. Although there is some variation with τ_0, both bundles of curves are much tighter than those corresponding to the Lorenz system and, although the correspondence between the theoretically expected error curves and those obtained from predictions performed during the verification period is not exact (probably due to our failure to predict during the warm phase of ENSO), the correspondence is much better than that exhibited by the chaotic system. Penland (1989) and Penland and Ghil (1993) provide a flavor of what this error curve looks like for other linear and nonlinear systems.

APPENDIX C

Derivation of Covariance Matrix for a Linear Stochastic Process

In many applications, it would be very useful to be able to generate the statistics of a linear stochastically forced system, for which the dynamical equation is known, without actually having to run a numerical model. This is, in fact, possible. Consider a linear system

$$\frac{dx_i}{dt} = \sum_{k=1}^{M} B_{ik}x_k + \xi_i, \tag{C1}$$

where ξ is a stationary Gaussian white-noise vector with covariance matrix $\langle \xi\xi^T \rangle dt = \mathbf{Q}$ and where the matrix $\mathbf{B}$ has eigenvectors (normal modes) $\{\mathbf{u}_\alpha\}$ and adjoints $\{\mathbf{v}_\alpha\}$ corresponding to eigenvalues $\{\beta_\alpha\}$. Both $\mathbf{Q}$ and $\mathbf{B}$ are known, as is, therefore, the eigenstructure of each. As in the text, we shall denote the equal-time covariance matrix $\langle \mathbf{xx}^T \rangle$ as $\mathbf{C}_0$. The fluctuation–dissipation relation for this system is

$$\mathbf{BC}_0 + \mathbf{C}_0\mathbf{B}^T + \mathbf{Q} = 0. \tag{C2}$$

The problem is to determine $\mathbf{C}_0$ given $\mathbf{B}$ and $\mathbf{Q}$. This can be done by transforming Eq. (C2) into normal mode space. Let $\mathbf{u}$ be the matrix the αth column of which is the αth normal mode $\mathbf{u}_\alpha$, and let $\mathbf{v}$ be the corresponding adjoint matrix. The diagonal eigenvalue matrix of $\mathbf{B}$ is called β. Operating on the left of Eq. (C2) by $\mathbf{v}^T$ and on the right by $\mathbf{v}^*$, where (*) denotes complex conjugate, we find that

$$\beta\mathbf{D}_0 + \mathbf{D}_0\beta^* + \mathbf{M} = 0, \tag{C3}$$

where $\mathbf{D}_0 = \mathbf{v}^T\mathbf{C}_0\mathbf{v}^*$ and $\mathbf{M} = \mathbf{v}^T\mathbf{Q}\mathbf{v}^*$. Physically, $\mathbf{D}_0$ is the equal-time covariance matrix of the time series coefficients of the normal modes, and $\mathbf{M}$ is the covariance matrix of the white noise driving the normal modes. In component form, since β is diagonal,

$$\beta_\alpha D_{\alpha\beta} + D_{\alpha\beta}\beta_\beta^* + M_{\alpha\beta} = 0,$$

or

$$D_{\alpha\beta} = -M_{\alpha\beta}/(\beta_\alpha + \beta_\beta^*). \tag{C4}$$

Once the elements of $\mathbf{D}_0$ are obtained, the covariance matrix is easily calculated as

$$\mathbf{C}_0 = \mathbf{u}\mathbf{D}_0\mathbf{u}^\dagger, \tag{C5}$$

where (†) denotes complex conjugate-transpose.

REFERENCES

Arnold, L., 1974: *Stochastic Differential Equations: Theory and Applications.* Wiley and Sons, 228 pp.

Balmaseda, M., D. Anderson, and M. Davey, 1994: ENSO prediction using a dynamical ocean model coupled to statistical atmospheres. *Tellus,* **46A,** 497–511.

Barnett, T. P., 1991: The interaction of multiple time scales in the tropical climate system. *J. Climate,* **4,** 269–285.

——, L. Dümenil, U. Schlese, E. Roeckner, and M. Latif, 1989: The effect of Eurasian snow cover on regional and global climate variations. *J. Atmos. Sci.,* **46,** 661–685.

——, M. Latif, E. Kirk, and E. Roeckner, 1991: On ENSO physics. *J. Climate,* **4,** 487–515.

Battisti, D. S., and A. C. Hirst, 1989: Interannual variability in a tropical ocean model: Influence of the basic state, ocean geometry, and nonlinearity. *J. Atmos. Sci.,* **46,** 1687–1712.

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

Borges, M., and D. Hartmann, 1992: Barotropic instability and optimal perturbations of observed nonzonal flows. *J. Atmos. Sci.,* **49,** 335–354.

Cane, M., and S. Zebiak, 1985: A theory for El Niño and the Southern Oscillation. *Science,* **228,** 1085–1087.

——, M. Münnich, and S. E. Zebiak, 1990: A study of self-excited oscillations of the tropical ocean–atmosphere system. Part I: Linear analysis. *J. Atmos. Sci.,* **47,** 1562–1577.

Cochran, W. G., 1952: The χ^2 test of goodness of fit. *Ann. Math. Stat.,* **23,** 315–345.

Davey, M. K., D. Anderson, and S. Lawrence, 1993: A simulation of seasonality in ENSO forecast skill. Climate Res. Tech. Note 38, Hadley Centre, Meteorological Office.

Farrell, B., 1988: Optimal excitation of neutral Rossby waves. *J. Atmos. Sci.,* **45,** 163–172.

——, 1989: Optimal excitation of baroclinic waves. *J. Atmos. Sci.,* **46,** 1193–1206.

——, 1990: Small error dynamics and the predictability of atmospheric flows. *J. Atmos. Sci.,* **47,** 2409–2416.

——, and P. J. Ioannou, 1993: Stochastic dynamics of baroclinic waves. *J. Atmos. Sci.,* **50,** 4044–4057.

Gardiner, C. W., 1985: *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences.* Springer Verlag, 442 pp.

Ghil, M., and R. Vautard, 1991: Interdecadal oscillations and the warming trend in global temperature time series. *Nature,* **350,** 324–327.

Glantz, M. H., R. W. Katz, and N. Nicholls, Eds., 1991: *Teleconnections linking worldwide climate anomalies.* Cambridge University Press, 535 pp.

Hasselmann, K., 1976: Stochastic climate models. Part I. Theory. *Tellus,* **28,** 474–485.

——, 1988: PIPs and POPs—a general formalism for the reduction of dynamical systems in terms of Principal Interaction Patterns and Principal Oscillation Patterns. *J. Geophys. Res.,* **93,** 11 015–11 020.

Hirst, A. C., 1988: Slow instabilities in tropical ocean basin–global atmosphere models. *J. Atmos. Sci.,* **45,** 830–852.

Jin, F.-F., and J. D. Neelin, 1993a: Modes of interannual tropical ocean–atmosphere interaction—a unified view. Part I: Numerical results. *J. Atmos. Sci.,* **50,** 3477–3503.

——, and ——, 1993b: Modes of interannual tropical ocean–atmosphere interaction—a unified view. Part III: Analytical results in fully coupled cases. *J. Atmos. Sci.,* **50,** 3523–3540.

——, ——, and M. Ghil, 1994: El Niño on the Devil's Staircase: Annual subharmonic steps to chaos. *Science,* **264,** 70–72.

Kleeman, R., and S. Power, 1994: Limits to predictability in a coupled ocean-atmosphere model due to atmospheric noise. *Tellus,* **46A,** 529–540.

Kloeden, P., and E. Platen, 1992: *Numerical Solution of Stochastic Differential Equations.* Springer Verlag, 632 pp.

——, ——, and H. Schurz, 1994: *Numerical Solution of SDE Through Computer Experiments.* Springer Verlag, 292 pp.

Lacarra, J.-F., and O. Talagrand, 1988: Short-range evolution of small perturbations in a barotropic model. *Tellus,* **40A,** 81–95.

Latif, M., 1987: Tropical ocean circulation experiments. *J. Phys. Oceanogr.,* **17,** 246–263.

Lau, K.-M., 1985: Elements of a stochastic-dynamical theory of the long-term variability of the El Niño–Southern Oscillation. *J. Atmos. Sci.*, **42**, 1552–1558.

Lau, N.-C., S. G. H. Philander, and M. J. Nath, 1992: Simulation of ENSO phenomena with a low-resolution coupled general circulation model of the global ocean and atmosphere. *J. Climate*, **5**, 284–307.

Mann, H. B., and A. Wald, 1942: On the choice of the number of class intervals in the application of the chi square test. *Ann. Math. Stat.*, **13**, 306–317.

McWilliams, J., and P. Gent, 1978: A coupled air and sea model for the tropical Pacific. *J. Atmos. Sci.*, **35**, 962–989.

Molteni, F., and T. Palmer, 1993: Predictability and finite-time instability of the northern winter circulation. *Quart. J. Roy. Meteor. Soc.*, **119**, 269–298.

Münnich, M., M. A. Cane, and S. E. Zebiak, 1991: A study of self-excited oscillations of the tropical ocean–atmosphere system. Part II: Nonlinear cases. *J. Atmos. Sci.*, **48**, 1238–1248.

Neelin, J. D., 1990: A hybrid coupled general circulation model for El Niño studies. *J. Atmos. Sci.*, **47**, 674–693.

——, 1991: The slow sea surface temperature mode and the fast wave limit: Analytical theory for tropical interannual oscillations and experiments in a hybrid coupled model. *J. Atmos. Sci.*, **48**, 584–606.

——, and F.-F. Jin, 1993: Modes of interannual tropical ocean–atmosphere interactions—unified view. Part II: Analytical results in the weak coupling limit. *J. Atmos. Sci.*, **50**, 3504–3522.

Papanicolaou, G., and W. Kohler, 1974: Asymptotic theory of mixing stochastic ordinary differential equations. *Commun. Pure Appl. Math.*, **27**, 641–668.

Pearson, K., 1900: On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philos. Mag. Series 5*, **50**, 157–172.

Penland, C., 1989: Random forcing and forecasting using principal oscillation pattern analysis. *Mon. Wea. Rev.*, **117**, 2165–2185.

——, and M. Ghil, 1993: Forecasting Northern Hemisphere 700-mb geopotential heights using Principal Oscillation Patterns. *Mon. Wea. Rev.*, **121**, 2355–2372.

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

——, and L. Matrosova, 1994: A balance condition for stochastic numerical models with application to the El Niño–Southern Oscillation. *J. Climate*, **7**, 1352–1372.

——, and P. Sardeshmukh, 1995: Error and sensitivity analysis of geophysical eigensystems. *J. Climate*, **8**, 1988–1998.

Philander, S. G. H., 1990: *El Niño, La Niña, and the Southern Oscillation.* Academic Press, 293 pp.

——, T. Yamagata, and R. C. Pacanowski, 1984: Unstable air–sea interactions in the tropics. *J. Atmos. Sci.*, **41**, 604–613.

Rasmusson, E. M., and T. H. Carpenter, 1982: Variations in tropical sea surface temperature and surface wind fields associated with the Southern Oscillation/El Niño. *Mon. Wea. Rev.*, **110**, 354–384.

——, and J. M. Wallace, 1983: Meteorological spects of the El Niño/Southern Oscillation. *Science*, **222**, 1195–1202.

Reynolds, R., and R. Livezey, 1995: Improvement of tropical Pacific SSTs using EOFs. *Proc. 19th Climate Diagnostics Workshop*, Camp Springs, MD, U.S. Dept. of Commerce,

Schopf, P. S., and M. J. Suarez, 1988: Vacillations in a coupled ocean–atmosphere model. *J. Atmos. Sci.*, **45**, 549–566.

——, and ——, 1990: Ocean wave dynamics and the time scale of ENSO. *J. Phys. Oceanogr.*, **20**, 629–645.

Slutz, R. J., S. J. Lubker, J. D. Hiscox, S. D. Woodruff, R. L. Jenne, D. H. Joseph, P. M. Steurer, and J. D. Elms, 1985: Comprehensive Ocean-Atmosphere Data Set. Release 1. NOAA Environmental Research Laboratory, Boulder, Colorado. 268 pp.

Sparrow, C., 1982: *The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors.* Springer Verlag, 269 pp.

von Storch, H., T. Bruns, I. Fischer-Bruns, and K. Hasselmann, 1988: Principal Oscillation Pattern analysis of the 30-60 day oscillation in a GCM equatorial troposphere. *J. Geophys. Res.*, **93**, 11 021–11 036.

Wakata, Y., and E. S. Sarachik, 1991: Unstable coupled atmosphere-ocean basin modes in the presence of a spatially varying basic state. *J. Atmos. Sci.*, **48**, 2060–2077.

Weickmann, K. M., and G. N. Kiladis, 1994: Slow and fast sea surface temperature modes and the ENSO cycle during 1979–93. *J. Climate*, submitted.

Wong, E., and M. Zakai, 1965: On the convergence of ordinary integrals to stochastic integrals. *Ann. Math. Stat.*, **36**, 1560.

Xu, J.-S., 1993: The joint modes of the coupled atmosphere-ocean system observed from 1967 to 1986. *J. Climate*, **6**, 816–838.

——, and H. von Storch, 1990: Predicting the state of the Southern Oscillation using principal oscillation pattern analysis. *J. Climate*, **3**, 1316–1329.

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

Zebiak, S. E., 1989: On the 30–60 day oscillation and the prediction of El Niño. *J. Climate*, **2**, 1381–1387.

——, and M. A. Cane, 1987: A model El Niño–Southern Oscillation. *Mon. Wea. Rev.*, **115**, 2262–2278.