27 Stochastic Linear Models of Nonlinear Geosystems

Cécile Penland[1]

[1] NOAA/ESRL/Physical Sciences Division, Cecile.Penland@noaa.gov

Abstract. When nonlinearities are strong enough, they can often be treated in terms of linear stochastic differential equations. We discuss the conditions under which this can be done, and illustrate the linear approximation with several examples having varying success. In particular, we consider the low-dimensional chaotic Lorenz system, a linear process driven by the chaotic Lorenz system, and the real-world application to tropical sea surface temperatures.

1 Introduction: Random Thoughts on Nonlinearity

It has been long known that dynamic systems of different spatial and temporal scales are often coupled. That is, linear low-pass or linear high-pass filters, while often useful, are also limited in applicability. Fortunately, if the behavior of the long timescales is of interest and the detailed behavior of the short ones is less so, it is often possible to estimate the effect of the fast processes on the slower ones through stochastic parameterization. The procedure for developing the stochastic parameterization of rapidly varying processes is far from arbitrary and requires some statistical knowledge of the "uninteresting" fast dynamical component.

The very use of the relative terms "fast" and "slow" indicates a separation of time scales. The limit theorems (Khasminskii 1966; Papanicolaou and Kohler 1974; Majda et al. 1999) describing asymptotic stochastic behavior quantify time scale separation by introducing a smallness parameter ε into the equations of motion. Powers of ε are written as coefficients of individual terms in the equations and indicate the rapidity with which these terms' particular autocorrelation functions decrease with respect to each other. Although ε is a smallness parameter, it is not an importance parameter; terms multiplied by higher powers of ε are not only retained but represent the "interesting" deterministic dynamics. Said another way, we divide the dynamical evolution of a system into large effects that don't last very long and smaller effects that behave coherently. The short-lasting effects are quantitatively treated as Gaussian white noise, possibly modulated by the coherent processes, on a coarse-grained timescale short enough to resolve the slow dynamics but too long to resolve the details of the fast dynamics. The validity of treating rapidly varying components as Gaussian white noise requires that the "noise" be the combined effects of a variety of weakly interacting physical phenomena. This is the reason for coarse-graining the dynamical equations; the unit of time is large enough that the integrated effects of unresolved dynamics can be approximated as a Gaussian stochastic variable (the Central Limit Theorem).

Naturally, the most interesting cases are generally those in which a clear timescale separation is not possible. There are several ways around this problem, although the

mathematical underpinnings are not always as rigorous as when the separation is evident. For example, one might model the stochastic perturbations as red noise with an appropriate spectrum rather than as white noise (e.g., Horsthemke and Léfèver 1984; Newman et al. 1997). This procedure works well as long as one has knowledge of what that spectrum is, and if one is working on a "forward" problem that is *a priori* completely defined. The so-called "inverse problems," where one assumes only the basic form of the dynamical equations and estimates the quantitative description from data, are more difficult.

The difficulty in inverse problems lies in the non-uniqueness of the estimated results, and this difficulty is an issue whether or not there exists a clear timescale separation. Even when it is known that a system is governed by a matrix *linear* equation, with a coefficient of Gaussian white stochastic forcing varying *linearly* with the state vector, one can use sample moments to deduce the matrix coefficient of the noise only up to an arbitrary orthogonal matrix factor. Further, since this article deals with systems where the noise is temporally continuous and only *approximately* white, there is a noise-induced modification (the so-called "Ito correction" or "noise-induced drift," depending on the community) to the systematic linear feedback involving this unobservable orthogonal matrix (Wong and Zakai 1965; Khasminskii 1966). If one is interested only in the predictability of the system, or in covariance statistics of the noise, the arbitrary matrix factor is irrelevant. However, if one is interested in using the results of the inverse model for process studies investigating the detailed interaction between stochastic and deterministic dynamics, the arbitrariness can be problematic. In this study, we shall use inverse modeling to diagnose covariance statistics of a low-order deterministic chaotic process acting as additive forcing to a linear system, thus avoiding the issue of non-unique results.

The idea that low-order chaos can act as stochastic forcing is not new. Jarzynski (1995) has shown that fast deterministic chaos coupled to a slow Brownian particle drives the slow degrees of freedom toward a state of statistical equilibrium with the fast degrees. This result is general, and was achieved by analyzing the Liouville equation for the combined fast-slow system to investigate the probability distribution of the slow part. In other work, Rödenbeck et al (2001) considered a 9-dimensional model introduced by Lorenz (1995), coupled to a faster 8-dimensional system. The important restriction in these studies was the well-separated timescales of the fast and slow dynamics. And so, since the issue of whether driving chaos is of low or high dimensionality is not really important to the validity of a stochastic approximation, we are back to the issue of how far we can go with such an approximation when the timescales are not well-separated.

Why would one want to do such a thing? Why not just consider the nonlinear system as a whole? The problem, of course, is in the word "just;" most of the time we do not know the equations for the full nonlinear system. We may know some of them, but we should not expect a good description of reality by a geophysical model confined to those equations that are well understood, when the effect of unknown processes might be as large as that of the known ones. We can replace the unknown dynamics with detailed conceptual models based on data, with the added advantage of being able to analyze these "toy" models in detail. This is an excellent approach as

long as one is able to justify quantitatively the rejection of competing possibilities, but bears the danger of legitimizing unjustified assumptions, depending on how ad hoc the toy model is. A model ought not be accepted on the basis of a test that can pass mutually exclusive dynamics. For example, prediction error alone is not always a good criterion for evaluating competing physical models; for more than a thousand years before Copernicus, Kepler and Galileo appeared on the scene, Aristarchus' (4th century BCE) heliocentric model of the solar system was routinely rejected in favor of Ptolemy's (2nd century CE) earth-centered model because Ptolemy's model predicted planetary positions more accurately.

If there exists a plethora of detailed deterministic models having different dynamical properties but similar resemblances to observations (this is the case with El Niño, for example), it is often useful to step back to the most general description of that system's dynamics and use the results of the Central Limit Theorem (CLT) to devise a rigorous approximation, including an estimation of how accurate the approximation is on realistic space and time scales. That is, "rigorous" does not always mean "accurate," or vice versa; it does mean that the conditions for application of the CLT (see discussion below) are met.

In the following, we shall explore the completely rigorous, but sometimes inaccurate, approximation that a nonlinear system can be described as a *linear* process augmented with stochastic terms resulting from rapidly varying nonlinearities. It must be emphasized that we are not throwing away the nonlinearities; we retain them in an approximate form. This approximation cannot be made arbitrarily, and so we review its methodology in some detail. In particular, we discuss a test for the validity of the linear approximation that is difficult to pass. This test is subject to "Type 1 errors," often rejecting the linear hypothesis when it is true. Most statistical tests are subject to either Type 1 errors (rejecting the hypothesis when it is true) or Type 2 errors (accepting the hypothesis when it is false). While neither type of error is desirable, Type 2 errors are worse. They are also the errors likely to be made by those who look for evidence in favor of theories rather than for evidence against them.

After considering some theoretical aspects of the analysis, we use the familiar chaotic Lorenz (1963) system to show an example of a rigorous but useless application of the approximation. After that, we show an example where the approximation is somewhat problematic but is still extremely useful in diagnosing many dynamical properties of the system. The next stage of the article applies the approximation to a real geophysical system. Finally, we discuss the results presented here. By elucidating that the dependence of a dynamical description of natural phenomena in terms of linear and/or nonlinear processes is largely a matter of scale, it is hoped that the commonalities rather than the conflicts between different views of dynamical systems will be emphasized.

2 Theoretical Considerations

2.1 The central limit theorem

The first part of this discussion is taken from Sardeshmukh et al. (2001) and is repeated here mostly verbatim because of its importance to this study. (Let the reader be assured that I am plagiarizing myself.) Consider a dynamical system of equations as follows:

$$\frac{dx}{dt} = F'(x,t) + G'(x,t) \tag{1}$$

where x is a vector in the N-dimensional Euclidean space $\mathbb{R}^N$ and where $F'(x, t)$ and $G'(x, t)$ are characterized by short and long correlation times, respectively. For our purposes, an alternative description of x in terms of a dimensionless parameter ε is preferable

$$\frac{dx}{dt} = \varepsilon F(x,t) + \varepsilon^2 G(x,t) \quad , \tag{2}$$

where ε^2 can be thought of as a ratio of characteristics timescales of F' and G'. It should be noted that the parameter ε is not intended here to be a measure of the relative importance of $F(x, t)$ and $G(x, t)$ but rather of the relative rapidity with which the autocorrelation functions of these terms decay, as will be clear in what follows. The theorem of Papanicolaou and Kohler (1974; PK74 hereafter) describes the conditions under which a singular scaling of time allows Eq.(2) to converge weakly to a stochastic differential equation. This is a dynamical form of the Central Limit Theorem. The conditions require that the fast process $\varepsilon F(x, t)$ be sufficiently variable and that the probability density function (pdf) of any value of $\varepsilon F(x, t)$ becomes independent of any initial conditions as time increases indefinitely, and at a sufficiently rapid rate. Further, $F(x, t)$ is required to be sufficiently smooth with respect to the components of x, where "sufficiently rapid" and "sufficiently smooth" are made explicit in PK74. The time coordinate is now scaled

$$s = \varepsilon^2 t \tag{3}$$

and Eq.(2) becomes

$$\frac{dx}{ds} = \frac{1}{\varepsilon} F(x, s / \varepsilon^2) + G(x, s / \varepsilon^2) \tag{4}$$

The proof in PK74 that Eq.(4) converges weakly to a stochastic differential equation is very general, and often difficult to apply in a forward sense. Therefore, for illustrative purposes, we restrict the problem by putting more conditions on $F(x,s/\varepsilon^2)$, and by stating that F, G, and x are all vectors with N elements. Let $F(x,s/\varepsilon^2)$ be of the form (see Remark 8 of PK74)

$$F_i(x, s / \varepsilon^2) = \sum_{k=1}^{K} F_i^k(x, s)\eta_k(s / \varepsilon^2),$$ (5)

where $\eta_k(s/\varepsilon^2)$ is stationary, centered and bounded. The integrated lagged covariance matrix of η has elements

$$C_{km} = \int_0^{\infty} <\eta_k(t)\eta_m(t+\tau)> d\tau, \qquad k,m = 1,2,...K,$$ (6)

where angle brackets denote expectation value. With these restrictions, the theorem by PK74 states that in the limit of long times ($t \to \infty$) and small ε ($\varepsilon \to 0$), taken so that $s = \varepsilon^2 t$ remains fixed, the conditional pdf $p(x, s \mid x_0, s_0)$ for x at time s given an initial condition $x_0(s_0)$ satisfies the "backward Kolmogorov equation," so called because the operator $\mathcal{L}$ describes the dependence of $p(x, s \mid x_0, s_0)$ on the initial conditions:

$$\frac{\partial p(x, s \mid x_0, s_0)}{\partial s_0} = \mathcal{L}(p(x, s \mid x_0, s_0)),$$ (7)

where

$$\mathcal{L}(\bullet) = \sum_{ij=1}^{N} a^{ij}(x_0, s_0)\frac{\partial^2}{\partial x_{0i}\partial x_{0j}}(\bullet) + \sum_{j=1}^{N} b^j(x_0, s_0)\frac{\partial}{\partial x_{0j}}(\bullet)$$ (8)

and

$$a^{ij}(x, s) = \sum_{km=1}^{K} C_{km}F_i^k(x, s)F_j^m(x, s)$$ (9a)

$$b^j(x, s) = \sum_{km=1}^{K} C_{km} \sum_{i=1}^{N} F_i^k(x, s)\frac{\partial F_j^m(x, s)}{\partial x_i} + G_j(x, s).$$ (9b)

In this limit, if the formal adjoint $\mathcal{L}^*$ of $\mathcal{L}$ exists, the conditional pdf also satisfies a "forward Kolmogorov equation" in the scaled coordinates:

$$\frac{\partial p(x, s \mid x_0, s_0)}{\partial s} = \mathcal{L}^* (p(x, s \mid x_0, s_0))$$ (10)

Where

$$\mathcal{L}^*(\bullet) = \sum_{ij=1}^{N} \frac{\partial^2}{\partial x_i \partial x_j}(a^{ij}(x, s)\bullet) - \sum_{j=1}^{N} \frac{\partial}{\partial x_j}(b^j(x, s)\bullet)$$ (11)

and where the superscript asterisk indicates the formal operator adjoint. Eq. (10) is called a "Fokker-Planck equation" in the scientific literature, and we adopt this terminology hereafter. As stated above, the conditional pdf of x in the scaled

coordinate system obeys Eqs. (7) and (10) in a weak sense. That is, the moments of x can be approximated with an error of order ε by the moments of the solution to the stochastic differential equation,

$$dx = G(x, s)\ ds + \sum_{k,\alpha} F^k(x, s)S_{k\alpha} \bullet dW_\alpha , \qquad (12)$$

where the symmetric matrix C has been written as the product of two matrices ($C = SS^\mathrm{T}$) and has absorbed the factor of 1/2 present in most formulations of the Fokker-Planck equation. The quantity W is a vector of independent Wiener processes and the expression $\bullet dW$ denotes the fact that Eq. (12) is to be interpreted in the sense of Stratonovich. That is, the white noise is an approximation to a continuous system with small but finite decorrelation time. One implication of this result is that stochastic integrals reduce to standard Riemann integrals (Kloeden and Platen 1992). Since we are usually interested in moments of the system rather than sample paths, the stochastic integral property of Stratonovich systems is less important than the dependence of moments, especially the mean, of the limiting process on the form of the stochasticity. Unless F' is independent of x, the ensemble mean of x will be different from the deterministic solution for the differential equation with only the slow part G'.

In forward problems, the form of η usually dictates the form of S. Inverse problems are usually defined only up to an estimation of C and, therefore, infinitely many matrices S consistent with C may be constructed by multiplying any estimation of S on the right by an arbitrary orthogonal matrix.

As a final comment on this section, we note that the CLT is posed for "random" processes $F(x, t)$. It should be noted that this definition of "random" is from Kolmogorov (1933) and includes the deterministic processes. The most important condition on $F(x, t)$ is the "mixing condition," which states how rapidly the autocorrelation function of $F(x, t)$ decreases. The mixing condition in PK74 is much weaker than that of Khasminsii (1966) and is fairly easy for most physical processes to pass. Beck and Roepstorff (1987) further explored this issue further in the context of chaotic systems. Thus, we interpret the CLT to mean that there is a timescale on which the moments of a deterministic chaotic process are indistinguishable from those of a system governed by the stochastic differential equation (12). Whether or not this timescale is useful is another matter. For example, the approximation is rigorous but useless for prediction studies if the behavior of a nonlinear system is coherent enough that the timescale for the stochastic approximation is too long to resolve the differences between conditional and marginal probabilities. One valuable use of the CLT is in model evaluation. If nature can be shown to obey the CLT at some timescale, then any realistic model of nature must also obey it at that same timescale. We shall explore this point further below.

2.2 Linear inverse modeling

We now explore the possibility of a natural process obeying a rather extreme version of the CLT, the multivariate, stable linear process driven by stochastic forcing. This approximation is surprisingly useful, particularly when the deterministic feedback matrix is not orthogonal.

Linear Inverse Modeling (LIM) is a statistical method of deriving the best linear description from observations of a dynamical system. The first step of this procedure is also known as Principal Oscillation Pattern (POP) analysis (Hasselmann 1988; Von Storch et al. 1988). Unlike POP analysis, however, LIM attempts to identify the dynamical nature of the system being observed. Inherent in the procedure is a test to see whether or not the procedure is appropriate for the time series, i.e., whether the dynamical system generating the time series is, indeed, linear. We also review LIM's ability to describe the transient growth that can occur when the linear dynamical operator is not orthogonal (e.g., Farrell 1988).

Consider a multivariate linear dynamical system x driven by additive, stationary, Gaussian white noise ξ:

$$\frac{dx}{dt} = \mathbf{B}x + \xi .$$

(13)

In Eq. (13), ξ is related to a vector Brownian motion $W(t)$ as follows:

$$\xi \, dt = \mathbf{S}dW,$$

(14)

so that $\mathbf{Q} = \mathbf{S}\mathbf{S}^{\mathrm{T}}$ is a matrix measuring the covariance properties of the stochastic forcing. The analysis that follows can be adapted to several variations of Eq. (13). In particular, we may allow $\mathbf{S}$ to vary periodically in time, or to be a linear function of x. We shall discuss these adaptations to the procedure during the exposition.

Moments of the Fokker-Planck equation for the transition pdf $p(x,t+\tau \,|x,t)$ corresponding to Eq. (13) can be manipulated to yield

$$<x(t+\tau)x^{\mathrm{T}}(t)> = \exp(\mathbf{B}\tau) <x(t)x^{\mathrm{T}}(t)>,$$

(15)

where angle brackets denote ensemble averages. In practice, the contemporaneous and lagged autocovariance matrices are estimated as time averages. If $\mathbf{S}$ is a periodic function rather than a constant matrix, Eq. (15) is true as long as the matrices are averaged over an integer number of periods. The difference between this derivation and derivations for discrete analyses by, for example, Von Storch et al (1988) is that their noise is required to be contemporaneously uncorrelated with the state variable. In a continuous system Eq. (13), where the driving noise is a physical process with a nonzero but negligible correlation time, the system $x(t)$ and $\xi(t)$ are correlated as follows (García et al. 1987):

$$<x(t)\xi^{\mathrm{T}}(t)> + <\xi(t)x^{\mathrm{T}}(t)> = \mathbf{S}\mathbf{S}^{\mathrm{T}} = \mathbf{Q}.$$

(16)

Principal Oscillation Pattern (POP) analysis (Hasselmann 1988; Von Storch et al. 1988) estimates the eigenstructure of an operator $\mathbf{A}$ that replaces $\exp(\mathbf{B}\tau)$ in Eq. (16). POP analysis is the first step of LIM. However, instead of attempting to identify a

single dominant mode, LIM identifies **A** with the Green function $\mathbf{G}(\tau) \equiv \exp(\mathbf{B}\tau)$, with τ chosen to optimize the accuracy of the calculation (Penland and Sardeshmukh 1995a). The entire set of normal modes (eigenvectors of $\mathbf{G}(\tau)$ and, hence, **B**) and adjoints (eigenvectors of $\mathbf{G}^T(\tau)$ and, hence, $\mathbf{B}^T$) is then used to infer the dynamical operator **B**.

One very useful property of using the stationary continuous system is the Fluctuaion-Dissipation relation (FDR: e.g., Penland and Matrosova 1994)

$$\mathbf{B}<x(t)x^T(t)> + <x(t)x^T(t)>\mathbf{B}^T + \mathbf{Q} = \mathbf{0}. \tag{17}$$

Equation (18) states that the system's stationary statistics, represented by $<x(t)x^T(t)>$, are maintained by a balance between the stochastic forcing, represented by **Q**, and the dissipation effected by the deterministic dynamics represented by **B**. For periodically varying **S**, the right side of Eq. (18) is not zero, but is rather the derivative with respect to time of $<x(t)x^T(t)>$.

Given any initial condition $x(t)$, the most probable prediction $x'(t+\tau)$ of $x(t+\tau)$ is performed simply by multiplying the initial condition by the Green function as follows:

$$x'(t+\tau) = \mathbf{G}(\tau)x(t) \equiv \exp(\mathbf{B}\tau)\,x(t). \tag{18}$$

For a perfect model of the stochastic system (13), the prediction error

$$\varepsilon(t+\tau) \equiv x(t+\tau) - x'(t+\tau) \tag{19}$$

is not zero, but is rather a Gaussian random variable with covariance matrix σ.

$$\sigma(t+\tau) \equiv <\varepsilon\varepsilon^T(t+\tau)> = <x(t+\tau)x^T(t+\tau)> - \mathbf{G}(\tau)<x(t)x^T(t)>\mathbf{G}^T(\tau). \tag{20}$$

For constant **S**, of course, x is wide-sense stationary. One now has a test for the validity of Eq. (13). A lag τ_o is chosen and Green functions at other lags are estimated using the following relation:

$$\mathbf{G}(\tau) = [\mathbf{G}(\tau_o)]^{\tau/\tau o}. \tag{21}$$

(Eq. (21) is obtained by applying the Cayley-Hamilton theorem to the spectral decomposition of $\mathbf{G}(\tau)$). If Eq. (13) is valid, the prediction error variance does not depend on τ_o. Penland and Magorian (1993) and Penland and Sardeshmukh (1995b, PS95 hereafter) both found this "tau test" to be passed by LIM applied to tropical IndoPacific sea surface temperature anomalies.

There are, unfortunately, linear systems for which the tau test is not valid. If **B** is not a constant matrix, then the tau test will fail. The tau test will also fail if the dimensionality of the observed state vector is too small to span the linear vector space in which the dynamics reside (Penland and Ghil 1993). Sometimes, the lag τ_o happens to lie close to the half-period of a normal mode's eigenvalue, thereby precluding the diagnosis of that mode's imaginary part and causing an inaccurate estimation of $\mathbf{G}(\tau_o)$. Yet another source of Type 1 errors (rejecting the linear approximation hypothesis when it is true) is observational error in the data from which the model is estimated. Nevertheless, if the validity of the linear approximation is the question, Type 1 errors of this genre are probably safer in

diagnosing the dynamical behavior of a system than Type 2 errors (accepting the linear approximation hypothesis when it is false) would be.

If $\mathbf{S}dW$ is a linear function of x as follows:

$$\left(\mathbf{S}dW\right)_i = \sum_j S^{(1)}_{ij} x_j \bullet dW^{(1)} + \sum_\beta S^{(2)}_{i\beta} dW_\beta^{(2)}, \tag{22}$$

with $\mathbf{S}^{(1)}$ a constant matrix, the above formalism follows similarly with some exceptions. First of all, the matrix Green function $\mathbf{G}(\tau)$ becomes

$$\mathbf{G}(\tau) = \exp\{[\mathbf{B} + (\mathbf{S}^{(1)})^2/2]\tau\}. \tag{23}$$

Secondly, it may not be possible to prove that the prediction $x'(t+\tau) = \mathbf{G}(\tau)x(t)$ is the most probable prediction (if the existence of a probability current around the boundaries of the phase space cannot be shown to vanish), but $x'(t+\tau)$ is still the best prediction in the mean square sense. Next, Eq. (16) does not hold. Finally, the FDR is modified to read

$$[\mathbf{B} + (\mathbf{S}^{(1)})^2/2]{<}x(t)x^T(t){>} + {<}x(t)x^T(t){>}[\mathbf{B} + (\mathbf{S}^{(1)})^2/2]^T + ...$$

$$... + \mathbf{S}^{(1)}{<}x(t)x^T(t){>}\mathbf{S}^{T(1)} + \mathbf{S}^{(2)}\mathbf{S}^{T(2)} = \mathbf{0}. \tag{24}$$

The annoyance of these expressions is that it is generally possible to infer only the combination $[\mathbf{B} + (\mathbf{S}^{(1)})^2/2]$ from data, rather than $\mathbf{B}$ and $\mathbf{S}^{(1)}$ separately. However, the tau test for the linear approximation is still valid. Thus, LIM can distinguish linear dynamics from demonstrably nonlinear dynamics in data, but it cannot distinguish an additive from a multiplicative noise process.

For simplicity of explanation, we now return to the case of constant $\mathbf{S}$. The eigen-ectors of $\mathbf{B}$, and hence $\mathbf{G}(\tau)$, are generally not orthogonal to each other, in spite of their being known as "normal" modes. When $\mathbf{B}$ is orthogonal and $\mathbf{Q}$ is diagonal, then the normal modes are indeed orthogonal and equivalent to the Empirical Orthogonal Functions (EOFs). However, this is a highly specialized case. Whenever the normal modes are orthogonal, the amplitude of a prediction $x'(t+\tau)$ is always smaller than that of the initial condition. When they are not orthogonal, the non-normality some-imes causes temporary amplitude growth (Farrell 1988; Penland and Sardeshmukh 1995b). The condition for this temporary growth is that the initial condition project strongly onto a right singular vector of the operator $\mathbf{G}(\tau)$, i.e., an eigenvector of $\mathbf{G}^T\mathbf{G}(\tau)$, and that this singular vector correspond to an eigenvalue greater than one. In fact, the eigenvalue is the ratio of the predicted amplitude at lead time τ to the initial condition amplitude and is called the "growth factor at lead time τ."

3 Experiments Using the Chaotic Lorenz System

In this section we shall consider the deterministic equations of Lorenz (1963):

$$\frac{dx}{dt} = -a(x - y) \tag{25a}$$

$$\frac{dy}{dt} = rx - xz - y \tag{25b}$$

$$\frac{dz}{dt} = xy - bz \tag{25c}$$

where the choice of parameters $a = 10$, $b = 8/3$ and $r = 28$ give the familiar chaotic behavior. Equations (25) are integrated using a simple Euler scheme with a model timestep $\Delta = 0.001$, and sampled every 20 timesteps, so that the sampling interval Δ_s is 0.02. The first 1000 samples were discarded to ensure that the time series used in the experiments are representative of the attractor. The time series $x(t)$, $y(t)$ and $z(t)$ are shown in Fig. 27.1; the usual projections of the attractor are shown in Fig. 27.2. A time series of 500 000 samples (10 000 model time units) was generated; hereafter we call it the *L63 time series*.
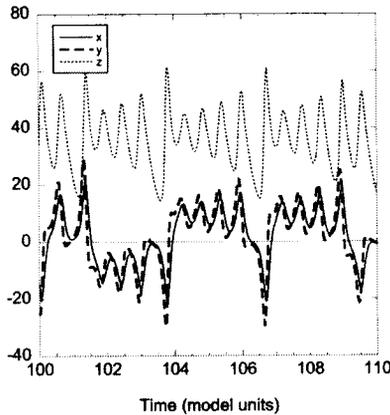

Fig. 27.1. Time series of the chaotic Lorenz (1963) system used in this study. Solid line: x. Dashed line: y. Dotted line: z.

3.1 A linear model of the chaotic Lorenz system

This does not work. It is not expected to work. The reason we tried it is to give credibility to results for systems where the linear approximation *does* work, at least, in comparison.

Let us linearize Equations (25) around $l_0 = (x_0, y_0, z_0) = (0, 0, <z>)$ so that the equation for the vector $l = (x, y, \zeta)$, where $\zeta = z - z_0$, can be written

$$\frac{dl}{dt} = Ll + \eta \tag{26a}$$

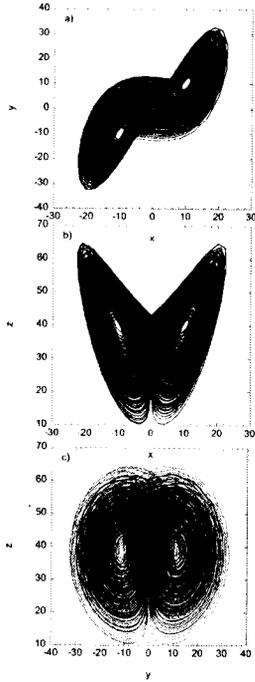**Fig. 27.2.** Maps of the chaotic Lorenz attractor generated for this study. *a) y vs. x. b) z vs. x. c) z vs. y.*

with

$$L = \begin{pmatrix} -a & a & 0 \\ r - z_0 & -1 & 0 \\ 0 & 0 & -b \end{pmatrix}$$ (26b)

and

$$\eta = \begin{pmatrix} 0 \\ -x\varsigma \\ xy - bz_0 \end{pmatrix}.$$ (26c)

To see whether Eq. (26) can be estimated as a stochastic linear process on the sampling time scale, we centered the L63 time series and subjected it to LIM, using values of $\tau_0 = 0.02, 0.04, 0.06, 0.08, 0.10, 0.12, 0.14,$ and 0.16. Traces of the

resulting estimates **B'** of **B** are shown in Fig. 27.3. Clearly, **B'** does not pass the tau test. In fact, trace(**B'**) behaves exactly as expected for a nonlinear system subjected to LIM with small values of τ_0, as shown below.

Consider a sample matrix

$$\mathbf{G}'(\tau_0) = <\mathbf{x}(t+\tau_0)\mathbf{x}^{\mathrm{T}}(t)><\mathbf{x}(t)\mathbf{x}^{\mathrm{T}}(t)>^{-1} \tag{27}$$

that is *not* well approximated by a matrix exponential. We do assume that it is smooth enough to be Taylor expanded in an absolutely convergent series for small values of τ_0:

$$\mathbf{G}'(\tau_0) = \mathbf{1} + \mathbf{A}'\tau_0 + \mathbf{A}''\tau_0^2 + ...$$

$$= \mathbf{1} + \mathbf{A}'\tau_0 + (\mathbf{A}'\tau_0)^2/2 + (\mathbf{A}''-\mathbf{A}'^2/2)\tau_0^2 + ...$$

$$= \exp(\mathbf{A}'\tau_0) + O(\tau_0^2)$$

$$\approx \exp([\mathbf{A}'+\mathbf{a}\tau_0]\tau_0). \tag{28}$$

In Eq.(28), the symbol $O(\tau_0^2)$ indicates all terms of at least second order in τ_0, and the matrix **a** is the difference between the second term in the Taylor expansion of **G'**(τ_0) and that of a matrix exponential. Thus, we identify the slope (-140.3/model unit^2) of the curve at small values of τ_0 in Fig. 27.3 as trace(**a**).
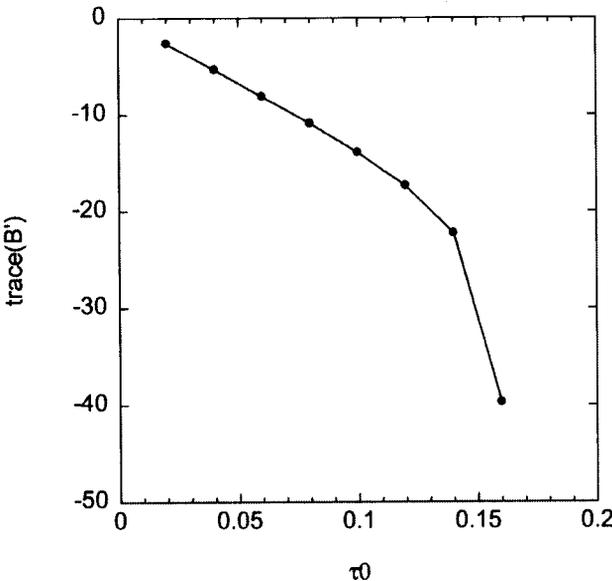

Fig. 27.3. Trace(**B'**) *vs.* τ_0 for LIM applied to L63.

It was shown in Penland and Ghil (1993) that a similar dependence on τ_0 obtains when the system is linear but the observations do not sufficiently span the dynamical space. This is an indication of the well-known mathematical equivalence between nonlinear models and linear models of larger dimensionality. If the system in fact obeys dynamics with negligible nonlinearity and the observations do not span the dynamical space, one must obviously generate the unobserved directions with a nonlinear model in order to fit the data. Conversely, one can usually fit a nonlinear model with a linear model of sufficiently high dimensionality by redefining temporal derivatives. In that case, the number of dimensions needed to fit the data can provide valuable information on the character of the nonlinearity in physical space. As humans, we are anxious to ask, "But what is it *really?*" Heartless Nature responds that it depends on how we choose (or are able) to look at it.

Returning to the Lorenz system, it may be hoped that linear predictions may be useful even though we know from Fig. 27.3 that the system is nonlinear in the observed space. Fig. 27.4 obliterates this hope. The observed error variance (solid lines), normalized to the total variance of the 3-dimensional time series, show deviations from the expected error variance (trace($\sigma(\tau)$), see Eq. 20)) for all values of τ_0. Further, the curves are strongly dependent upon the value of τ_0, with $\tau_0 = 0.02$ (open circles) estimating much *smaller* prediction error variance than $\tau_0 = 0.16$ (heavy solid line) but delivering much *larger* errors. Although Fig. 27.4 reaches the same conclusion as PS95 (their Fig. 19), the figures themselves look somewhat different, even though both studies considered the same Lorenz system, using the
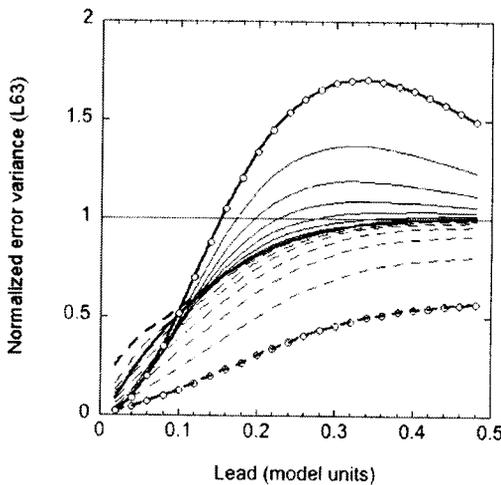

Fig. 27.4. Variance of linear prediction error, normalized to the total variance of the Lorenz system. Dashed lines: Theoretically expected error variance predicted by LIM. Solid lines: True error variance. Plain heavy solid line: $\tau_0 = 0.16$ model units. Heavy solid line with open circles: $\tau_0 = 0.02$. Light solid lines correspond to the other values of τ_0.

same values of τ_0, and for the same lead times. The resolution of this apparent paradox is that PS95, in an attempt to mirror the treatment of their sea surface temperature data set as closely as possible, used a verification time series much shorter than the one we used here. It is therefore possible that their verification time series sampled only one lobe of the attractor. If the linear approximation hypothesis were correct, of course, there would no difference. It is also noted that the PS95 study carefully distinguished between training period and verification period. Here, we have used the entire length of the time series for both training and verification since we are more interested in whether the curves vary with τ_0 rather than the actual size of the error variance.

Note that the observed error variance can be larger than the saturation value of unity, particularly for small values of τ_0. The normalized error variance of the persistence forecasts (not shown) is also larger than the saturation value; its peak value of two occurs at a lead of about 0.4 units.

Finally, one may ask if a coarse-grained version of L63 may be treated as a linear stochastic process. Starting with the sampling interval, we smoothed the L63 time series on progressively longer time scales and were unable to find any coarse-graining that gave a useful linear description of the system. The oscillations of the L63 are rapid enough that Nyquist problems occurred before the linear regime could be reached. That is, oscillations were too rapid to be resolved at the smallest τ_0 allowed by the coarse-graining, resulting in the imaginary part of an eigenvalue of $\mathbf{B'}\tau_0$ equal to $n\pi$, where n is a nonzero integer. We conclude that any possible three-dimensional linear regime of L63 would be at a scaling coarse enough that oscillations around the center of a lobe are unresolved.

3.2 A linear process coupled to the chaotic Lorenz system

A situation is now considered where the Lorenz system is considered as external forcing for a linear system. That is, we wish to investigate whether a modest coarse-graining allows the L63 to be treated as a stochastic component. Redefining x, y, and z,

$$\frac{dx}{dt} = \alpha x + l_x$$
$$\frac{dy}{dt} = \beta y + l_y \, . \tag{29}$$
$$\frac{dz}{dt} = \gamma z + l_z$$

In Eq.(29), l is defined as in Eq.(26); that is, we use the centered Lorenz system to force the components $\mathbf{u} = (x, y, z)$. The values of (α, β, γ) are (-0.02, -0.033, -0.05). Eq. (29) is integrated using a time step of 0.02 model units, i.e., the sampling of the Lorenz system shown in Fig. 27.1. The result is coarse-grained on intervals of two model units, so that each recorded component of $\mathbf{u}$ is an average over 100 samples.

Because the model has been coarse-grained over two model units, we shall call this case Model 2. Note that we rewrite Eq. (29) as

$$\frac{du}{dt} = \mathbf{B}u + l, \tag{30}$$

where, we repeat, u and l have been coarse grained over 100 time steps, and the matrix $\mathbf{B}$ is diagonal with trace($\mathbf{B}$) = -0.103.

LIM was applied to the coarse grained time series and samples $\mathbf{B}'(\tau_0)$ estimated for τ_0 = 2, 4, 6 , 8, 10, 12, 14 and 16 model units. For values of τ_0 larger than 16, Nyquist problems occurred. The trace($\mathbf{B}'$) is shown in Fig. 27.5 (solid circles), along with the trace (-0.103) expected if the L63 system truly acted as stationary white noise forcing on those timescales. Clearly, $\mathbf{B}'$ does not perfectly pass the tau test, particularly for values of τ_0 smaller than 6. Nevertheless, even for $\tau_0 \leq 6$, the slope (0.013/model unit^2) is four orders of magnitude smaller than that of the linearized regime of L63. This is less impressive than it sounds, but not much. One should actually compare that slope with the trace of the average trace($\mathbf{B}'$). For L63, that ratio is about -20; for the system Eq. (29), coarse grained at 2 model units, the ratio is closer to 0.1.

We also coarse-grained u over 200 time steps, equivalent to 4 model units. Hereafter, this case will be known as Model 4. We can see from Fig. 27.1 that 4 model units do not allow sufficient time for many passages between lobes. Nevertheless, the doubling of the averaging time does seem to improve the linear approximation in that the slope of trace($\mathbf{B}'$) with time is greatly reduced, and the values are closer to the expected value of -0.103.

From Table 27.1, we see that the reproduction of $\mathbf{B}$ by the estimates $\mathbf{B}'$ is imperfect, but does at least suggest that $\mathbf{B}$ is dominated by its diagonal elements. The high correlation of l_x and l_y (0.85) causes LIM to confuse the linear behavior of x and y
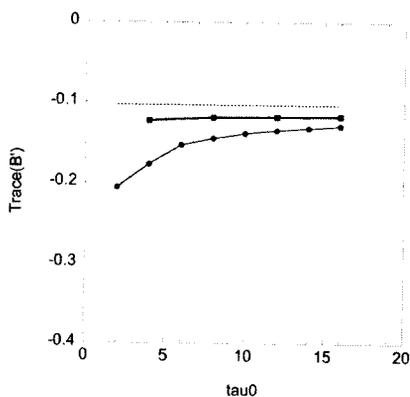

Fig. 27.5. Trace($\mathbf{B}'$) for Eq. (28). Light solid line: Model 2. Heavy solid line: Model 4. Dotted line: Trace($\mathbf{B}'$)= − 0.103/model unit.

somewhat. Nevertheless, the amount of damping introduced by **B** is approximately the right size, with the coarse graining, performed after the model integration, introducing a bit of extra damping to z.

Table 27.1

Empirical estimates of the linear operator in Eq. (28)

Model 2:

$$\mathbf{B'}_{ave} = \begin{pmatrix} -0.023 \pm 0.016 & 0.006 \pm 0.026 & 0.008 \pm 0.010 \\ -0.001 \pm 0.015 & -0.031 \pm 0.023 & 0.007 \pm 0.011 \\ 0.002 \pm 0.0004 & -0.002 \pm 0.0006 & -0.097 \pm 0.033 \end{pmatrix}$$

Model 2($\tau_0 \geq 6$):

$$\mathbf{B'}_{ave} = \begin{pmatrix} -0.014 \pm 0.002 & -0.008 \pm 0.004 & 0.001 \pm 0.004 \\ 0.008 \pm 0.002 & -0.044 \pm 0.003 & 0.0002 \pm 0.004 \\ 0.001 \pm 0.0002 & -0.002 \pm 0.0002 & -0.078 \pm 0.007 \end{pmatrix}$$

Model 4:

$$\mathbf{B'}_{ave} = \begin{pmatrix} -0.026 \pm 0.014 & 0.012 \pm 0.022 & 0.014 \pm 0.012 \\ -0.004 \pm 0.013 & -0.024 \pm 0.021 & 0.013 \pm 0.012 \\ 0.001 \pm 0.0002 & -0.002 \pm 0.0003 & -0.069 \pm 0.010 \end{pmatrix}$$

Integrated Model:

$$\mathbf{B} = \begin{pmatrix} -0.020 & 0.000 & 0.000 \\ 0.000 & -0.033 & 0.000 \\ 0.000 & 0.000 & -0.050 \end{pmatrix}$$

The numerical values for **B'** and **Q'** as estimated by the following: 1) Model 2, averaged over values of $\tau_0 = 2, 4, 6, 8, 10, 12, 14$ and 16 model units; 2) again Model 2, but averaging over only $\tau_0 = 8, 10, 12, 14$ and 16 model units; i.e., the linear regime as indicated by the flattening of trace(**B'**) shown in Fig. 27.5; and 3) Model 4, averaged over values of $\tau_0 = 4, 8, 12$ and 16 model units. Table 1 gives the averaged values of B'_{ij}; Table 27.2 gives the averaged values of Q'_{ij} along with the contemporaneous covariance matrix and correlation matrix of L63. The uncertainty estimates are not those associated with the length of the time series, but are rather indicative of the variation with τ_0.

Table 27.2

Empirical estimates of the forcing covariance matrix

Model 2:

$$\mathbf{Q'}_{ave} = \begin{pmatrix} 53.14 \pm 7.39 & 53.45 \pm 7.35 & 0.12 \pm 0.38 \\ 53.45 \pm 7.35 & 53.86 \pm 7.27 & 0.13 \pm 0.04 \\ 0.12 \pm 0.38 & 0.13 \pm 0.04 & 1.21 \pm 0.42 \end{pmatrix}$$

Model 2 $(\tau_0 \geq 6)$:

$$\mathbf{Q'}_{ave} = \begin{pmatrix} 57.07 \pm 0.86 & 57.37 \pm 0.90 & 0.12 \pm 0.05 \\ 57.37 \pm 0.90 & 57.74 \pm 0.92 & 0.14 \pm 0.05 \\ 0.12 \pm 0.05 & 0.14 \pm 0.05 & 0.97 \pm 0.09 \end{pmatrix}$$

Model 4:

$$\mathbf{Q'}_{ave} = \begin{pmatrix} 47.88 \pm 6.88 & 48.22 \pm 6.95 & 0.02 \pm 0.03 \\ 48.22 \pm 6.95 & 48.56 \pm 6.99 & 0.03 \pm 0.03 \\ 0.02 \pm 0.03 & 0.03 \pm 0.03 & 0.73 \pm 0.10 \end{pmatrix}$$

Covariance matrix of L63:

$$\langle \mathbf{ll}^T \rangle = \begin{pmatrix} 95.49 & 95.32 & -0.26 \\ 95.32 & 131.07 & -0.33 \\ -0.26 & -0.33 & 107.14 \end{pmatrix}$$

Correlation matrix of L63:

$$\mathbf{C}_{L63} = \begin{pmatrix} 1.00 & 0.852 & -0.003 \\ 0.852 & 1.00 & -0.003 \\ -0.003 & -0.003 & 1.00 \end{pmatrix}$$

The effect of the coarse graining is particularly noticeable in Table 27.2. Although the estimates $\mathbf{Q'}$ are not directly related to $\langle \mathbf{ll}^T \rangle$ (they have different units, for one thing, since the white noise approximation to L63 would be a complicated time integral over its lagged covariance structure), a surprising amount of information about L63 is obtained in this way. The high correlation between l_x and l_y is reproduced, and each is shown to be only weakly correlated with l_z. That the value of Q'_{zz} is severely reduced by the coarse graining is clear from Fig. 27.1; the variation in the

coarse-grained l_x and l_y comes primarily from jumping between the lobes of the attractor while most of the l_z variation is simply averaged out.
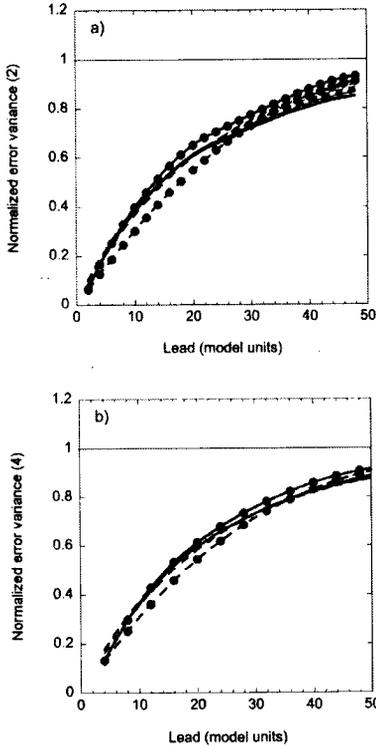

Fig. 27.6. Error variance curves, normalized to the field variance. Dashed and solid lines as in Fig. 27.4. Filled circles: smallest value of τ_0. Heavy line with no symbols: largest value of τ_0. *a)* Model 2. *b)* Model 4.

In terms of linear predictability, we see from Fig. 27.6 that the linear models of both coarse-grained systems pass the tau test reasonably well, though imperfectly. One notes in each case that the smallest value of τ_0 (filled circles) promises the best, and delivers the worst, prediction skill. Nevertheless, the spread of prediction error with τ_0 compared with that in Fig. 27.4 is very small, providing quantitative evidence for the validity of linearly approximating these coarse-grained systems.
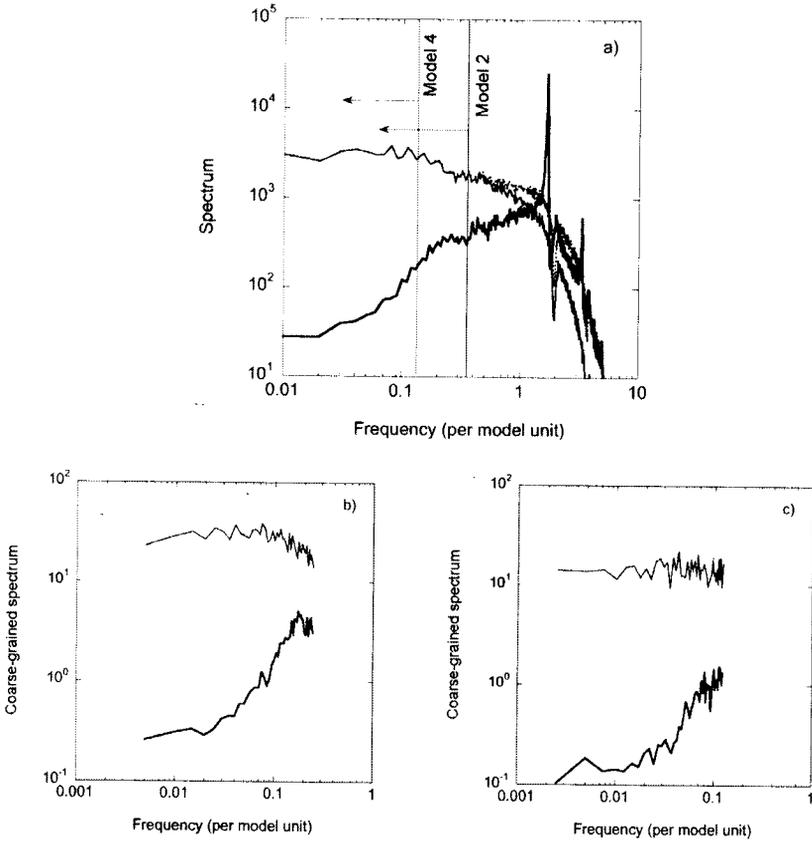
Fig. 27.7. Fast Fourier Transform spectra for l_x (light solid line), l_y (dotted line) and l_z (heavy solid line) components of the chaotic Lorenz system. *a*) Original system. *b*) Model 2. *c*) Model 4. At low frequencies, the spectra for l_x and l_y cannot be distinguished.

As a final diagnosis of the validity of treating the coarse-grained Lorenz model as white noise in Eq. (30), we present spectra of l_x, l_y and l_z in Fig. 27.7 corresponding to the original Lorenz system sampled at 0.02 model units (Fig. 27.7a), Model 2 (Fig. 27.7b) and Model 4 (Fig. 27.7c). In Fig. 27.7a, we have denoted the frequencies resolved by Models 2 and 4 for comparison with spectra in Figs. 27.7b and 27.7c. It is clear that the coarse-graining does more than change the resolution; it also affects both the magnitude and the shape of the spectra. While it is no surprise that the sharp peaks in the spectrum of l_z are eliminated, one might remark that the amplitude of the broadband part of that spectrum is also reduced by orders of magnitude. The same is true of l_x and l_y. This property underscores the importance of $1/\varepsilon$ in the scaled Eq. (4);

if the rapidly varying contribution to the underlying dynamical system is not large enough, coarse-graining will simply average out its effects. Further, for the white noise approximation to be valid in the coarse-grained system, the high-frequency band of the original spectrum must have a broad enough nature that at least some of that variance can be aliased into the low frequencies by the coarse graining, thus flattening out the spectrum. This is the case for l_x and l_y, whose spectra in the coarse-grained system are virtually identical. However, the component l_z of the original system has much of its variance in nearly sinusoidal behavior that the coarse-graining simply averages out.

4 Experiments Using Tropical Sea Surface Temperatures

In our examples using linear stochastic approximations so far, we have greatly profited from the blessings of knowing in advance the architecture of the underlying dynamical system. When we applied LIM to coarse-grained models, we had a "truth" with which to compare our answers. When we considered the spectra of the coarse-grained forcing, we had a time series of that forcing to analyze. This is not the case when the dynamics of natural phenomena are diagnosed from observed time series, which is why so many researchers like to start with a model "having many features of the observed system" and assume that it describes reality.

One purpose of this manuscript is to present one method for quantifying the "many features" statement. The accuracy with which observations can be described by a linear system, even when the underlying dynamics are chaotic, is a physical characteristic of the phenomenon. Comparison between LIM applied to data and LIM applied to a model can therefore aid in judging the fidelity of the model to reality. We believe this procedure could be a valuable tool in evaluating the enormous number of models claiming to represent the dynamics of tropical sea surface temperatures related to "El Niño," described below.

Originally known as a phenomenon occurring off the western coast of tropical South America, the term "El Niño" now generally indicates an episodic warming of sea surface temperatures (SSTs) over most of the tropical Pacific. With the availability of longer and better data sets, it is gradually becoming obvious that the "El Niño signal" is a global phenomenon, with clear signatures in data such as south tropical Atlantic SSTs (Mo and Häkkinen 2001; Penland and Matrosova 2006). Multiscale interactions are crucial to the evolution of El Niño in every theory proposed to explain it.

Unfortunately for diagnostics, salient features of El Niño can be found in models based on every possible genre of attractor, whether they be strange attractors (e.g., Jin et al. 1994; Tziperman et al. 1994), limit cycles or quasi-limit cycles (e.g., Battisti and Hirst 1989; Jin and Neelin 1993a,b; Neelin and Jin 1993; Blanke et al. 1997; Sun et al. 2004; Sun 2007) or fixed points (e.g., Penland and Magorian 1993; PS95; Moore and Kleeman 1996; Kleeman and Moore 1997; Thompson and Battisti 2000, 2001). Simple conceptual models of El Niño are therefore unable to give dynamical clarification of what the El Niño attractor looks like.

Linear theories of El Niño invoke the nonnormality of the deterministic linear operator (e.g., Farrell 1988; PS95) to explain the growth of SST anomalies during the development of an El Niño or, in its opposite phase, La Niña event.

As we have seen, there may be a timescale on which the paradigm of a chaotic attractor and that of a fixed point are mutually consistent. This statement should be taken in the spirit of the statement that the classical dynamics, the ideal gas law for example, is consistent with quantum mechanics. There is much evidence that on seasonal timescales El Niño dynamics reside in the linear regime and that the chaotic nonlinear dynamical component may be treated as stochastic forcing, with variance depending on the annual cycle. For modeling evidence, we refer the reader to works by R. Kleeman, and A. Moore, and the later (post-2000) modeling studies of D. Battisti. Some of the observational evidence is summarized here.

As stated in Section 2, linear theories of El Niño invoke the nonnormality of the deterministic linear operator to explain the growth of SST anomalies during the development of an El Niño or, in its opposite phase, La Niña event. One dominant optimal initial structure for growth is found when LIM is applied to tropical SSTs, and this structure evolves into a mature El Niño event about 6-9 months after its appearance (Fig. 27.8). The optimal structure does not significantly change with values of τ_0 (Penland 1996). PS95 and Penland (1996) confined their studies to the tropical Indo-Pacific, but inclusion of the entire tropical belt (Penland and Matrosova 1998, 2006) only makes their conclusions more robust. In fact, using the constituent normal modes to isolate the El Niño signal (Penland and Matrosova 2006) underscored the global nature of El Niño by revealing that its signal in the north tropical Atlantic ocean is very similar to the El Niño signal in the central Indian ocean (contemporaneous correlation +0.8), and that a strong precursor to SST
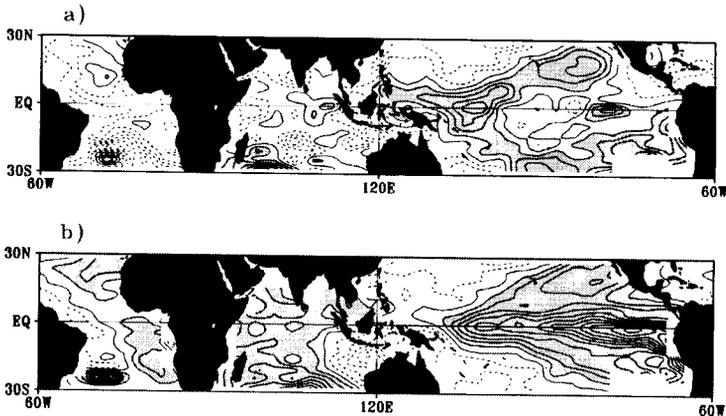

Fig. 27.8. Optimal structure of tropical SSTA and b) its evolution in eight months. The contour interval in the domain 60W-120E is 0.015. The contour interval in the domain 120E-60W is 0.03. Areas enclosed by the first positive contour are shaded. Dotted lines represent negative contours.

anomalies in the traditional Niño 3.4 region (6°N-6°S, 170°W-120°W) is found in the tropical south Atlantic ocean (lagged correlation, S. Atlantic leading, -0.6). This precursor occurs because the projection onto the optimal structure is so strong in the south tropical Atlantic, so using SSTs in that region to predict El Niño is no better than using the projection of the optimal structure as a whole. These global teleconnections indicate that any credible theory of El Niño cannot be based on a single time series such as the Niño 3.4 SST anomaly or the Southern Oscillation Index alone; such a theory must also be able to predict the global El Niño signal.

3.1 The data set

We consider COADS monthly SST data (Woodruff et al. 1993) in the entire tropical strip between 30°N and 30°S. The data were consolidated onto a 4°×10° grid and subjected to a three-month running mean. The 1950-2000 climatology was then removed from the SST data and the anomalies were projected onto 20 Empirical Orthogonal Functions (EOFs: Hotelling 1933) containing about two-thirds of the variance. Obviously, massaging the data as just described is tantamount to filtering the data. However, for the purposes of isolating the El Niño signal in this study, we shall refer to this data set as the "unfiltered data."

LIM was applied to the unfiltered data using $\tau_0 = 4$ months. The usual optimal structure was found to project onto three complex normal mode pairs associated with periods of about 2 years, 5 years, and 20 years. The dominance of these three modal pairs in the nonnormally evolving El Niño signal was found to be robust for those values of τ_0 uncompromised by Nyquist problems. The global El Niño signal was extracted from the multivariate data set by projecting the data onto the adjoint patterns of the three modal pairs. As shown by Penland and Sardeshmukh (1995a), the uncertainty in these timescales is large. However, Penland and Matrosova (2006) showed that the data filtered in this way accounted for the relevant Fourier spectral frequencies, as well as most of the low-frequency variability of the data set. The red noise background remained in the residual data.

To show that isolating the El Niño signal this way is consistent with the traditional Niño 3.4 SST anomaly index, we show both the filtered and unfiltered versions of that index in Fig. 27.9. The two indices are highly correlated (correlation
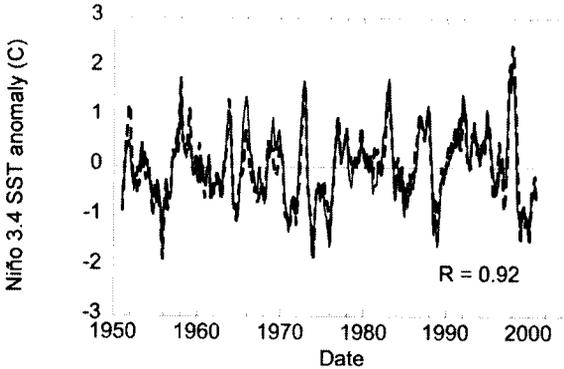
Fig. 27.9. Time series of Niño 3.4 SST anomaly. Solid line: unfiltered data. Dashed line: El Niño pass data.

+0.92), and mostly indistinguishable from each other. If one accepts the National Oceanic and Atmospheric Administration's designation of the unfiltered Niño 3.4 SST anomaly index as the official El Niño indicator, this check verifies that the filter, which is a pattern-based method, captures most of the El Niño variability. Although the modal and adjoint patterns were derived using an assumption of linear dynamics, there is no reason to believe that a map-by-map projection onto these time-independent patterns would eliminate any important nonlinear effects in the temporal evolution of El Niño, particularly since there are no large differences between the filtered and unfiltered versions of this index. We now proceed with the six-dimensional El Niño-pass filtered data set.

3.2 Results for SSTs

Our first order of business in treating the filtered data is to project them onto their EOFs, of which there are six. A detailed description of these EOFs and their time-dependent coefficients is found in Penland and Matrosova (2006). LIM was applied to this system and estimates $\mathbf{B}'$ of the linear operator were obtained using values of $\tau_0 = 3, 4, 5, 6, 7, 8, 9$ and 10 months. Nyquist problems appeared for larger values of τ_0. Interestingly, Nyquist problems did not appear for τ_0 near 6 months, in agreement with Penland's (1996) suggestion that the well-known dependence of El Niño on the annual cycle is mainly due to the yearly varying variance of the fast processes acting as stochastic forcing. As long as statistics are accumulated over an integral number of years, a periodically varying $\mathbf{S}$ in Eq. (14) can easily be handled by LIM; the only modification occurs in interpreting the matices $\langle x(t)x^T(t)\rangle$ and $\mathbf{Q}$ in Eq. (17) as annual averages.

A plot of trace($\mathbf{B}'$) *vs.* τ_0 is shown in Fig. 27.10. The dependence on τ_0 is linear for values of τ_0 less than about 6 months, after which the curve flattens out. The slope before the break is about 0.04 (months)$^{-2}$. Interestingly, this result implies a role for nonlinearities somewhat slower than that found by PS95 (their Fig. 12).
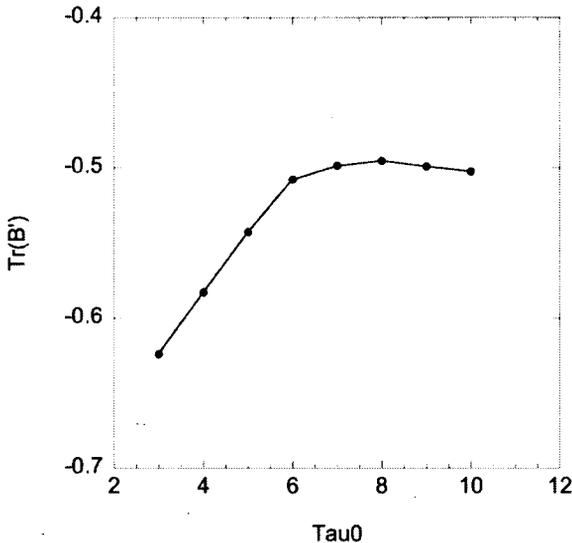
Fig. 27.10. Trace(**B'**) *vs.* τ_0 for tropical SST anomalies.

Of course, their 15-dimensional system contained dynamical process other than El Niño, and their graph was contaminated with Nyquist effects. However, the two studies agree in that the linear approximation is fairly good, particularly at timescales larger than 6 months.

A similar story is told by the error variance of the linear predictions (Fig. 27.11). Unlike the linear system forced by L63, the worst predictions are not at the smallest value of τ_0, but rather at the largest (10 months). This may occur because of nonlinearities, or it may be due to proximity to the half period of the quasibienniel oscillation. In any case, it is encouraging to see that the error variance for $\tau_0 = 4$ months is not significantly larger than that for $\tau_0 = 6$ months, in spite of Fig. 27.10, since forecasts currently provided to the general public from the website http://www.cdc.noaa.gov were developed with $\tau_0 = 4$ months.

Nevertherless, the variation with τ_0 in Fig. 27.11 is not as tight as that found by PS95. It appears that application of the nonnormal filter, by eliminating much of the red noise background associated with other linear processes in tropical SST, may actually magnify the presence of nonlinearity in the El Niño signal. This does not mean that nonlinear dynamics dominate its evolution on the seasonal scale. On the contrary, comparing the results of LIM applied to SST with those of the L63 results verifies the usefulness of the linear approximation for tropical SST anomaly dynamics.

The estimates **B'** for the 6-dimensional system spanned by the leading EOFs of the El Niño-pass SST data are provided in Table 27.3. Estimates **Q'** are provided in Table 27.4. It is encouraging that **Q'** is positive definite, a situation not at all obvious with the linear system driven by the Lorenz model.
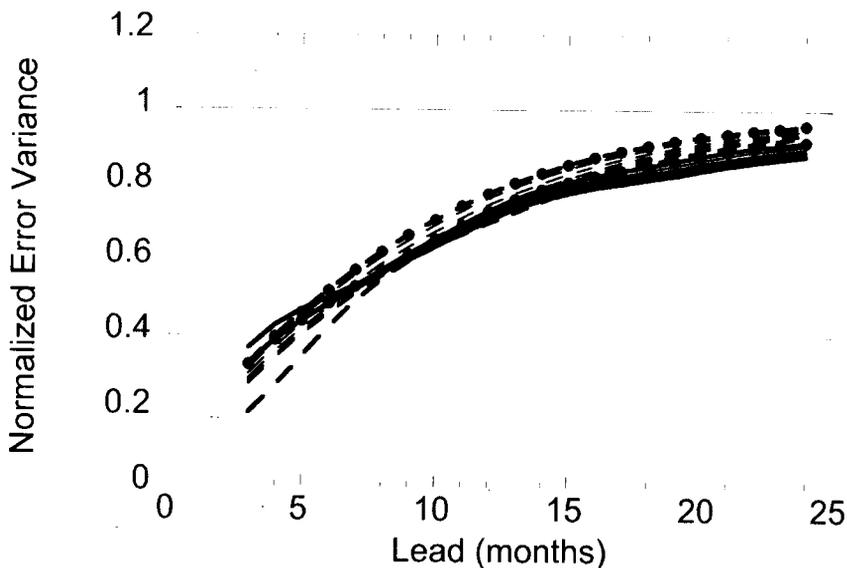
Fig. 27.11. Error variance normalized to SST anomaly field variance. Conventions as in Fig. 27.6.

Table 27.3

Empirical estimates of the linear operator from tropical SST

All values of τ_0:

B' =

$$
\begin{pmatrix}
-0.03 \pm 0.001 & 0.05 \pm 0.002 & -0.02 \pm 0.001 & 0.02 \pm 0.003 & 0.02 \pm 0.001 & -0.03 \pm 0.001 \\
-0.11 \pm 0.001 & -0.05 \pm 0.01 & 0.03 \pm 0.004 & -0.03 \pm 0.002 & 0.01 \pm 0.005 & -0.01 \pm 0.002 \\
0.11 \pm 0.002 & 0.07 \pm 0.004 & -0.09 \pm 0.01 & -0.13 \pm 0.01 & -0.08 \pm 0.003 & 0.02 \pm 0.005 \\
-0.21 \pm 0.005 & 0.04 \pm 0.01 & 0.17 \pm 0.004 & -0.15 \pm 0.02 & 0.13 \pm 0.009 & -0.06 \pm 0.002 \\
-0.13 \pm 0.004 & -0.04 \pm 0.007 & 0.05 \pm 0.001 & -0.20 \pm 0.01 & -0.15 \pm 0.005 & -0.09 \pm 0.005 \\
0.18 \pm 0.006 & 0.045 \pm 0.002 & -0.05 \pm 0.001 & 0.06 \pm 0.01 & 0.08 \pm 0.002 & -0.08 \pm 0.01
\end{pmatrix}
$$

Values of $\tau_0 \geq 6$ months:

B' =

$$
\begin{pmatrix}
-0.03 \pm 0.001 & 0.05 \pm 0.001 & -0.02 \pm 0.01 & 0.03 \pm 0.01 & 0.01 \pm 0.002 & -0.03 \pm 0.003 \\
-0.11 \pm 0.002 & -0.04 \pm 0.004 & 0.03 \pm 0.002 & -0.03 \pm 0.004 & 0.01 \pm 0.002 & -0.01 \pm 0.003 \\
0.078 \pm 0.05 & -0.07 \pm 0.01 & -0.06 \pm 0.03 & -0.16 \pm 0.03 & -0.08 \pm 0.01 & 0.01 \pm 0.01 \\
-0.27 \pm 0.03 & 0.04 \pm 0.02 & 0.17 \pm 0.01 & -0.16 \pm 0.05 & 0.10 \pm 0.03 & -0.06 \pm 0.001 \\
-0.18 \pm 0.09 & -0.03 \pm 0.003 & 0.07 \pm 0.04 & -0.26 \pm 0.08 & -0.14 \pm 0.01 & -0.11 \pm 0.02 \\
0.18 \pm 0.006 & 0.19 \pm 0.01 & 0.04 \pm 0.01 & -0.05 \pm 0.01 & 0.06 \pm 0.01 & -0.07 \pm 0.01
\end{pmatrix}
$$

Table 27.4

Empirical estimates of the stochastic forcing from tropical SST

All values of τ_0:

$Q' =$

$$
\begin{pmatrix}
1.52 \pm 0.50 & -0.28 \pm 0.31 & -0.19 \pm 0.21 & 0.15 \pm 0.81 & 0.09 \pm 0.27 & 0.14 \pm 0.17 \\
-0.28 \pm 0.31 & 1.07 \pm 0.15 & 0.22 \pm 0.08 & -0.04 \pm 0.27 & -0.01 \pm 0.14 & -0.13 \pm 0.21 \\
-0.19 \pm 0.21 & 0.22 \pm 0.08 & 1.13 \pm 0.42 & 0.11 \pm 0.26 & 0.34 \pm 0.08 & 0.02 \pm 0.03 \\
0.15 \pm 0.81 & -0.04 \pm 0.27 & 0.11 \pm 0.26 & 1.52 \pm 0.29 & 0.19 \pm 0.21 & 0.10 \pm 0.10 \\
0.09 \pm 0.27 & -0.01 \pm 0.14 & 0.34 \pm 0.08 & 0.19 \pm 0.21 & 1.43 \pm 0.07 & 0.08 \pm 0.22 \\
0.14 \pm 0.17 & -0.13 \pm 0.21 & 0.02 \pm 0.03 & 0.10 \pm 0.10 & 0.08 \pm 0.22 & 0.52 \pm 0.20
\end{pmatrix}
$$

Values of $\tau_0 \geq 6$ months:

$Q' =$

$$
\begin{pmatrix}
1.55 \pm 0.50 & -0.16 \pm 0.04 & 0.05 \pm 0.54 & 0.44 \pm 0.08 & 0.47 \pm 0.45 & 0.19 \pm 0.13 \\
-0.16 \pm 0.04 & 0.90 \pm 0.08 & 0.19 \pm 0.13 & 0.11 \pm 0.16 & 0.04 \pm 0.04 & -0.03 \pm 0.01 \\
0.05 \pm 0.55 & 0.19 \pm 0.06 & 0.83 \pm 0.43 & 0.24 \pm 0.23 & 0.24 \pm 0.23 & 0.11 \pm 0.12 \\
0.43 \pm 0.08 & 0.11 \pm 0.16 & 0.24 \pm 0.23 & 1.79 \pm 0.61 & 0.68 \pm 0.52 & 0.10 \pm 0.04 \\
0.47 \pm 0.45 & 0.04 \pm 0.04 & 0.24 \pm 0.23 & 0.68 \pm 0.52 & 1.40 \pm 0.08 & 0.28 \pm 0.13 \\
0.19 \pm 0.13 & -0.03 \pm 0.01 & 0.11 \pm 0.12 & 0.10 \pm 0.04 & 0.28 \pm 0.13 & 0.53 \pm 0.06
\end{pmatrix}
$$

4 Conclusions

It is true, generally speaking, that the universe is governed by nonlinear dynamics. It is also true, generally speaking, that the universe is governed by quantum mechanics. In fact, given the ubiquity of electromagnetic radiation, vacuum fluctuations, and interstellar media comprising atomic and subatomic particles, not to mention the relatively small amount of space taken up by bright matter, there is a case to be made that the subset of dynamical systems for which the classical approximation obtains is a very small set indeed. Nevertheless, for us who are confined to a terrestrial environment, that small set is extremely important. Further, there is a grey area, particularly in molecular dynamics, where hybrid models, basically using classical dynamics but incorporating quantum mechanical effects with relevant approximations, have proved themselves as valuable research tools. The relationship between classical and quantum mechanics as an allegory for that between linear and nonlinear dynamics is not exact, but it is useful to remind us that there is a very large grey area where the dynamical system may be treated as basically linear, although nonlinearities may not be ignored.

The importance of weather and climate predictions to the well-being of terrestrial inhabitants demands that meteorologists and climatologists provide timely forecasts

of the highest possible quality. Linearly based forecasts may be timely and easy to make, but no credible scientist would advocate using them for these reasons alone. However, they *are* easy to make and verify, and are sometimes even the best forecasts available. For example, LIM forecasts of river flows in Colombia, a system this author would never have the courage to consider linear, have shown themselves to be a skillful component of estimating electricity needs in that nation (Germán Poveda 2002, personal communication). It therefore behooves us to examine quantitatively the extent to which the linear approximation is valid. Recall that this approximation does not ignore nonlinearities; rather, it is valid when they vary rapidly enough to be treated as a stochastic component, possibly modulated by a linear term.

We have reviewed a procedure for finding the optimal linear model of a multivariate process, along with quantitative criteria for assessing its validity (the tau test). We have also briefly reviewed some of the limit theorems (e.g., Khasminsii 1966; Papanicolaou and Kohler 1974) that are relevant to the part of the approximation that treats the rapidly varying system as a stochastic term. There are later articles extending the limit theorems, notably by Kifer (2001). Further, the decomposition of a chaotic attractor into Markov partitions, and the resulting probabilistic description of those systems, is discussed by Dorfman (1999) and references therein. Our approach here is not theoretical rigor as much as it is to give a flavor of when the linear approximation may be useful. This is done in the context of three systems: the low-order chaotic Lorenz model (L63), a coarse-grained linear system for which L63 was treated as external forcing, and observed seasonal sea surface temperature anomalies.

It is not surprising that a linear approximation of L63 yields poor predictions of that system, or that tests for validity of the linear approximation fail spectacularly. In fact, this failure may be compared with the failure of LIM applied to another low dimensional system by Tsonis et al. (2006). The usefulness of these exercises lies in their providing benchmarks for how highly nonlinear, low-order chaotic systems perform on such tests.

Our second experimental set-up consisted of a linear system with decay times long compared with the timescales of L63 (Fig. 27.1), which maintained the linear system by acting as an external forcing. Although the linear decay times were long, the coarse-graining and sampling interval of 2 model units (Model 2) was purposely chosen to resolve some of L63's strong nonlinearity in order to strain the linear approximation. We compared Model 2 with a model coarse-grained and sampled at 4 model units (Model 4). Again, our expectations were confirmed when Model 4 passed the tau test better than Model 2 did. On the other hand, Model 2 gave us more information about the system: it told us the validity of the linear system was probably best for coarse-grainings greater than 6 model units, it told us the relative size, compared to linear dynamics, of the nonlinearities for coarse-grainings between 2 and 6 model units, and it gave just as good insight as Model 4 into the relative variance and correlation of the three L63 components.

It is encouraging that many characteristics of L63 could be inferred from a linear fluctuation-dissipation relation (for the purists: "of the second kind", Cugliandolo, Dean and Kurchan 1997) derived from Models 2 and 4. That it is able to diagnose

high correlations, such as that between l_x and l_y, suggests that a similar analysis of numerical model errors might reveal unsuspected relations between different sources of error.

Analysis of tropical sea surface temperatures basically confirmed the results of PS95, who found a linear approximation of their dynamics to be good and useful, though imperfect. A dynamically based filter (Penland and Matrosova 2006) used to isolate the El Niño signal seemed to unmask a small nonlinear component to this signal, even though the filter itself was based on an assumption of linear dynamics. Of course, Fourier filtering is useful even when it falsely assumes that the system is perfectly periodic with a period equal to the length of the time series; similarly, the assumption of linear dynamics in the development of the nonnormal filter is used to estimate spatial patterns and cannot force these patterns to be associated with a linear signal if they are not so in the unfiltered data.

Our results, as well as all of our previously published results (e.g., PS95, Penland and Magorian 1993), are also consistent with Kondrashov, Kravtsov, Robertson and Ghil (2005), who show that accounting for quadratic nonlinearities has an advantage in predicting extreme warm events as manifested by Niño 3 (5°S–5°N, 150°–90°W) SST anomalies, but otherwise is mostly indistinguishable from the linear model. Prediction of extremes is an important issue, and faster-than-seasonally varying nonlinearities (Kondrashov et al. 2005 used monthly rather than seasonal data) undoubtedly play an important role in maintaining extreme warm events. However, neither our studies nor theirs show any indication that nonlinear dynamics dominate the evolution of tropical SST anomalies except during the development of these warm extremes.

The importance of nonlinearities as a source of stochastic forcing must not be understimated. It has been stated by several of my colleagues that the term "noise" is unfortunate in that it implies processes to be filtered out if possible. In fact, these processes are crucial to the maintenance of those systems for which the linear approximation is valid, and the linear fluctuation-dissipation relation should be used as a guide for further study rather than an end in itself. After all, the linear approximation itself is best used as a diagnostic tool to investigate the strength of nonlinearities, thereby narrowing the possible physical mechanisms responsible for important weather and climate phenomena.

The time is ripe for research into the Central Limit Theorem properties of specific nonlinear processes. Are there timescales on which synchronized chaotic systems (e.g., Duane and Tribbia 2001; 2007) act as stochastic forcing and, if so, what are those timescales? We have already stated that LIM applied to dynamical models should reproduce the results of LIM applied to data. Conversely, it is time to apply nonlinear data analysis (e.g., Tsonis and Elsner 1996; Tsonis et al. 2006; Tsonis 2007) to realistic linear stochastic models and compare those results with those obtained from applying the analysis to data. If they are different, is there a timescale on which the results converge? Whether or not there is such a timescale, a careful analysis of the differences is likely to yield valuable information about the nature of the physical system.

Acknowledgements

References

Archimedes, ed. Heiberg, vol. II, p. 244 (Arenarius I. 4-7) *The Works of Archimedes*, ed. Heath, pp. 221-222.

Aristarchus of Samos (310BCE-230BCE) Personal Communications to Archimedes. See Archimedes.

Battisti, D., and A. Hirst, (1989) Interannual Variability in a Tropical Atmosphere–Ocean Model: Influence of the Basic State, Ocean Geometry and Nonlinearity. J. Atmos. Sci. 46, 1687-1712.

Beck, C., and Roepstorff, G. (1987) From dynamical systems to the Langevin equation. Physica, 145A, 1-14.

Blanke, B., Neelin, J.D., Gutzler, D. (1997) Estimating the Effect of Stochastic Wind Stress Forcing on ENSO Irregularity. J. Climate 10, 1473-1486.

Cugliandolo, L. F., Dean, D.S., and Kurchan, J. (1997) Fluctuation-dissipation theorems and entropy production in relaxational systems. Physical Review Letters 79, 2168-2171.

Dorfman, J.R. (1999) Introduction to Chaos in Nonequilibrium Statistical Mechanics. Cambridge University Press, Cambridge.

Duane, G.S., and Tribbia, J.J. (2001) Synchronized Chaos in Geophysical Fluid Dynamics. Phys. Rev. Lett., 86, 4298-4301.

Duane, G.S., and Tribbia, J.J. (2007) Deterministic and stochastic parameter adaptation in the "Jungian" synchronization-based approach to data assimilation. *20 Years of Nonlinear Dynamics in Geosciences*, Springer, this volume.

Farrell, B., (1988) Optimal excitation of neutral Rossby waves. J. Atmos. Sci. 45, 163-172.

Garcia, A.L., Malek Mansour, M., Lie G., and Clementi, E. (1987) Numerical Integration of the Fluctuating Hydrodynamic Equations. J. Stat. Phys. 47, 209-228.

Hasselmann, K. (1988:) PIPs and POPs — The reduction of complex A general formalism for the reduction of dynamical systems in terms of Principal Interaction Patterns and Principal Oscillation Patterns. J. Geophys. Res. 93, 11015-11020.

Horsthemke W., and Lefever, R. (1984) *Noise-Induced Transitions: Theory and Applications in Physics, Chemistry and Biology*. Springer-Verlag, 318 pp.

Hotelling, H. (1933) Analysis of a complex of statistical variables into principal components. J. Educ. Psych. 24, 417-520.

Jarzynski, C. (1995) Thermalization of a Brownian particle via coupling to low-dimensional chaos. Physical Review Letters 74, 2937-2940.

Jin, F.-F., and Neelin, J.D. (1993a) Modes of Interannual Tropical Ocean–Atmosphere Interaction—a Unified View. Part I: Numerical Results. J. Atmos. Sci. 50, 3477-3503.

Jin, F.-F., and Neelin, J.D. (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.

Jin, F.-F., Neelin, J.D., Ghil, M. (1994) El Niño on the Devil's Staircase: annual subharmonic steps to chaos. Science 264, 70-72.

Khasminskii, R.Z. (1966) A limit theorem for solutions of differential equations with random right-hand sides. Theory Prob. Applications 11, 390-406.

Kifer, Y. (2001) Averaging and climate models. *Stochastic Climate Models*, ed. P. Imkeller and J.-S. von Storch, Progress in Probability, 49, Birkhaueser, Basel.

Kleeman, R., and Moore, A.M. (1997) A Theory for the Limitation of ENSO Predictability Due to Stochastic Atmospheric Transients. J. Atmos. Sci. 54, 753-767.

Kolmogorov, A. (1933) *Grundbegriffe der Warscheinlischkeitrechnung*. Springer, Berlin.

Kloeden P. E., and Platen, E. (1992) Numerical Solution of Stochastic Differential Equations. Springer-Verlag, 632 pp.

Lorenz, E. (1963) Deterministic non-periodic flow. J. Atmos. Sci. 20, 130-141.

Lorenz, E. (1995) Predictability: a problem partly solved. *Seminar Proceedings "Predictability,"* European Center for Medium-Range Weather Forecasts, 1-18.

Majda A. J., Timofeyev, I., and Vanden Eijnden, E. (1999) Models for stochastic climate prediction. Proc. Natl. Acad. Sci. USA 96, 14687-14691.

Mo, K. C., and Häkkinen,S. (2001) Interannual Variability in the Tropical Atlantic and Linkages to the Pacific. J. Climate 14, 2740-2762.

Moore, A. M., and Kleeman, R. (1996) The dynamics of error growth and predictability in a coupled model of ENSO. Quart. J. Roy. Meteor. Soc. 122, 1405-1446.

Newman, M., Sardeshmukh, P. D., and Penland, C. (1997) Stochastic forcing of the winter time extratropical flow. J. Atmos. Sci. 54, 435-455.

Neelin, J.D., and Jin, F.-F. (1993) Modes of Interannual Tropical Ocean–Atmosphere Interaction—a Unified View. Part II: Analytical results in the weak coupling limit. J. Atmos. Sci. 50, 3504-3522.

Papanicolaou, G.C., and Kohler, W. (1974) Asymptotic theory of mixing stochastic ordinary differential equations. Comm. Pure Appl. Math. 27, 641-668.

Penland, C., and Magorian, T. (1993) Prediction of Niño 3 Sea-surface Temperature Anomalies using Linear Inverse Modeling. J. Climate 6, 1067-1075.

Penland, C., and Ghil, M. (1993) Forecasting Northern Hemisphere 700mb Geopotential Height Anomalies using Empirical Normal Modes. Monthly Weather Review 121, 2355-2372.

Penland, C., And Matrosova, L. (1994) A balance condition for stochastic numerical models with application to the El Niño - Southern Oscillation. J. Climate 7, 1352-1372.

Penland, C., and Sardeshmukh, P. D. (1995a) Error and sensitivity analysis of geophysical systems. J. Climate 8, 1988-1998.

Penland, C., and Sardeshmukh,P. D. (1995b) The optimal growth of tropical sea surface temperature anomalies. J. Climate 8, 1999-2024.

Penland, C. (1996) A stochastic model of Indo-Pacific sea surface temperature anomalies. Physica D 98, 534-558.

Penland, C., and L. Matrosova, L. (1998) Prediction of tropical Atlantic sea surface temperatures using Linear Inverse Modeling J. Climate, 11, 483-496.

Penland, C., and Matrosova, L. (2006) Studies of El Nino and interdecadal variability in tropical sea surface temperatures using a nonnormal filter. J. Climate 19, 5796-5815.

Ptolemy (148) *Almagest*, Canopus.

Rödenbeck, C., Beck, C., and Kantz, H, (2001) Dynamical systems with timescale separation. *Stochastic Climate Models*, ed. P. Imkeller and J.-S. von Storch, Progress in Probability, 49, Birkhaueser, Basel.

Sardeshmukh, P. D., Penland, C., and Newman, M. (2001) Rossby waves in a fluctuating medium. *Stochastic Climate Models*, ed. P. Imkeller and J.-S. von Storch, Progress in Probability, 49, Birkhaueser, Basel.

Sun, D.-Z., Zhang, T., and Shin, S.-I., (2004) The effect of subtropical cooling on the amplitude of ENSO: A numerical study. J. Climate 17, 3786-3798.

Sun, D.-Z. (2007) The role of ENSO in regulating its background state. *20 Years of Nonlinear Dynamics in Geosciences*, Springer, this volume.

Thompson, C.J., and Battisti, D.S., (2000) A linear stochastic model of ENSO. Part I: Model Development. J. Climate 13, 2818-2832.

Thompson, C.J., and Battisti, D.S., (2001) A linear stochastic model of ENSO. Part II: Analysis. J. Climate 14, 445-466.

Tsonis, A.A., and Elsner, J.B. (1996) Mapping the channels of communication between the tropics and higher latitudes in the atmosphere. Physica D 92, 237-244.

Tsonis, A.A., Swanson, K.L., and Roebber, P.J. (2006) What do networks have to do with climate? Bulletin of the American Meteorological Society 87, 585-595.

Tsonis, A.A. (2007) What do networks have to do with climate? *20 Years of Nonlinear Dynamics in Geosciences*, Springer, this volume.

Tziperman, E., Stone, L., Cane, M.A., Jarosh, H. (1994) El Niño chaos: Overlapping residences between the seasonal cycle and the Pacific ocean-atmosphere. Science 264, 72-74.

Von Storch, H., Bruns, T.,Fischer-Bruns, I., and Hasselmann, K. (1988) Principal Oscillation Pattern analysis of the 30–60 day oscillation in a GCM equatorial troposphere. J. Geophys. Res. 93, 11021-11036.

Wong, E., and Zakai, M. (1965) On the convergence of ordinary integrals to stochastic integrals. Annals of Mathematical Statistics 36, 1560-1564.

Woodruff, S., Lubker S., Wolter K., Worley S., and J. Elms, 1993: Comprehensive Ocean-Atmosphere Data Set (COADS) Release 1a: 1980–92. *Earth Sys. Monitor* 4, 1–8.