Random Forcing and Forecasting Using Principal Oscillation Pattern Analysis

CECILE PENLAND

Department of Atmospheric Sciences, University of California, Los Angeles, California

(Manuscript received 28 October 1988, in final form 1 May 1989)

ABSTRACT

The effects of random forcing and deterministic feedback are combined in a measured multivariate time series. It is shown here how the characteristics of the driving noise can be found after the deterministic effects have been identified by the principal oscillation pattern (POP) analysis. In addition, the POP analysis is extended to enable the prediction of the most probable meteorological pattern at some future time when the present pattern is known, and the conditional probability of finding the process at any location within a range of values given the value of the process at another location at an earlier time. Estimates of how well these predictions can be trusted are also given. The basic assumption of POP analysis is that the system can be optimally modeled by a linear Markov process.

1. Introduction

A major effort in meteorological study during the past few decades has been how to extract important patterns and dynamical variables directly from measured data. The data systems to be considered in this study are multivariate time series, each component of which represents a particular measurement location. A related but separate question, which ought not to be confused with the first one, is how to describe complex dynamical systems with many degrees of freedom by simpler representative processes that reproduce the important physical effects of the original system using a greatly reduced number of degrees of freedom. These processes should be determined as far as possible directly from the data.

Philosophically consistent with the second goal is the principal component (PC) analysis, introduced in its present form by Hotelling (1933) and accepted in a meteorological context after application by the Statistical Forecasting Project at the Massachusetts Institute of Technology (Lorenz 1956). A concise review of classical PC analysis is given by Kutzbach (1967). The principal components, called amplitude time series by Rasmussen et al. (1981), are statistically stationary quantities and their associated empirical orthogonal functions (EOFs) provide an objectively determined basis where error due to system truncation is clear.

The fact that the EOFs are orthogonal is an advantage or a disadvantage, depending upon the purpose of the analysis. The statistical independence of errors in parameters fitted to a model is nearly always assumed in estimating confidence intervals; this assumption is more likely to be valid if the parameters themselves describe the behavior of statistically uncorrelated components. On the other hand, EOFs are by construction unable to detect traveling waves in geophysical phenomena. It should be pointed out that this static nature of classical EOFs gives them a wide range of applicability; PC analysis orders the variance contained in the corresponding EOFs for any stable, stationary Gaussian-distributed stochastic process. Very little dynamical information can be obtained empirically from a system without some assumptions as to its nature, and classical PC analysis does not pretend to try.

It has been proposed that propagating structures could be observed by determining the eigenvectors of lagged covariance matrices at varying lags. Weare and Nasstrom (1982) introduced extended empirical orthogonal functions, which combine the spatial variability as described by the classical EOFs with the propagation information obtained from the lagged covariance matrices. Subsequent work by Broomhead and King (1986) and by Vautard and Ghil (1989) also examine the eigenvalues and principal components of lagged covariance matrices.

Among the extensions to classical PC analysis proposed to detect propagating structures, the complex principal component (CPC) analysis in the time domain (Rasmussen et al. 1981; Barnett 1983, 1985; Horel 1984, and references within) is perhaps the closest in philosophy to classical PC analysis. The measured time series is assumed to be the real part of a causal, complex system, the imaginary part of which is found using Hilbert transforms (Thomas 1969). The complex EOFs are the eigenfunctions of the Hermitian correlation matrix for the complex system. Horel (1984) gives a concise review of the statistical properties, in-

Corresponding author address: Dr. Cecile Penland, Dept. of Atmospheric Sciences, University of California, 405 Hilgard Avenue, Los Angeles, CA 90024.

cluding an estimate of the sampling errors [compare with North et al. (1982) for a discussion of sampling errors in the case of real EOFs].

Although the complex EOFs provide an orthogonal basis within which any complex system, linear or not, can be described, physical interpretation of the phase information obtained from a complex EOF analysis is difficult if the dynamics are not linear. Therefore, the next logical step is to make the assumption of linear Markovian dynamics, and see just how far one could go with the theory. In particular, the effects of (linearized) deterministic feedback could be separated from those of the random driving forces. The eigenfunctions of the deterministic feedback matrix, called principal oscillation patterns, or POPs, (Hasselmann 1988) give the spatial properties of the feedback. The corresponding eigenvalues indicate the time scales on which these feedback processes occur. An example of the approach has been provided by von Storch et al. (1988) where a POP analysis of the T21-GCM high-pass filtered velocity potential output yielded a POP with most of the properties of the observed 30–60 day tropical wave.

It has not been recognized that the POP analysis can be used to find the spatial properties of the driving random noise or that these properties are important. Previous POP analyses concentrate on deterministic effects and seem to take the view of traditional signal processing that the noise is simply an unwanted error term. Here, the philosophy is more consistent with Hasselmann's (1976) consideration of "fast" and "slow" variables; the system is driven by the fast variables which are treated as white noise. We believe the spatial properties of the driving energy source to be important. Rather than consider discrete Markov processes driven by white noise, as has been the previous approach, we consider continuous processes which have been discretely sampled. The distinction is irrelevant to the results obtained previously but is crucial to determining the properties of the driving forces.

The primary purpose of the POP analysis as presented here is the separation of deterministic effects, which occur on time scales at least as slow as the sampling rate, from the quickly varying random forcing. Consistent with previous work, the spatial properties of the deterministic feedback processes and the time scales upon which they occur can be found. In contrast with any of the inverse modeling methods discussed above, the spatial properties of the stationary random forcing are sought. We also investigate the possibility of using the POP analysis as a forecasting tool. Under the assumptions of linearity and Markovianity stated above, it is shown that, given a pattern at time t, the most probable pattern at later time $t + \tau$ can be predicted. In addition, expressions are provided for the conditional probability of finding the value of the process at any particular location given that we know the value of the process at another location at an earlier time. It is not expected that the procedure would outperform general circulation models in short-term forecasting. However, it is possible that the procedure would be valuable for longer-range forecasting if nonlinear effects propagate slowly compared with the sampling rate. This is a very important "if." Nevertheless, the analysis can provide an estimate of when these nonlinear effects become important.

There is some modification of the terminology used in previous POP analyses; the "feedback" matrix as used by von Storch et al. (1988) is a function of the sampling interval and is called here the Green function, consistent with Riskin (1984). We reserve the term "feedback matrix" to describe the constant matrix in the continuous stochastic propagation equation. The POPs are eigenfunctions of both matrices.

In section 2, we review some properties of Markov processes, emphasizing the probability theory, which are interesting to the meteorologist. In section 3, we concentrate on the theory of POPs and their relation to Markov processes. Section 4 presents a method for estimating how well pattern predictions can be trusted. This is followed in section 5 by a recipe for calculating 1) the spatial properties of the feedback processes, 2) the time scales on which these feedback processes occur, 3) the noise covariance matrix, which shows the regions where a large amount of short-term activity occurs, 4) the matrix of conditional probabilities; i.e., the probability that a large value of the process at position j and time t will be followed by a large (or small) value at position i and later time $t + \tau$, 5) the most probable pattern at time $t + \tau$ given the pattern at earlier time t, and 6) reliability limits on the predictions. Two numerical examples of the POP analysis are provided in section 6. The first example is an artificially generated time series governed by linear, Markovian dynamics. Of particular interest is how the results depend on the number of samples in the series. The second numerical example investigates the POP analysis of a system that violates nearly all of the necessary assumptions: the highly nonlinear, purely deterministic Lorenz system. Finally, a discussion of the analysis and its relation to complex EOFs is presented in section 7.

2. Some properties of Markov processes

Consider a continuous N-dimensional Markov process $\mathbf{x}(t)$:

$$\frac{d\mathbf{x}}{dt} = \mathbf{B}\mathbf{x} + \boldsymbol{\xi}. \tag{1}$$

Here, x_i is the value of the process at position i and time t, $\mathbf{B}$ is the constant $N \times N$ feedback matrix, and ξ_i is Gaussian white noise at position i and time t. Note that the process as shown here is centered; we assume that its mean value has been removed. The values of x_i shall therefore be referred to as "anomalies." The noise has constant covariance matrix $\mathbf{Q}$

$$\langle \boldsymbol{\xi}\boldsymbol{\xi}^{\mathrm{T}} \rangle = \mathbf{Q}. \tag{2}$$

We form the covariance matrix of the process $\mathbf{x}$:

$$\langle \mathbf{x}\mathbf{x}^{\mathrm{T}} \rangle = \mathbf{\Lambda}. \tag{3}$$

In most cases of practical interest, it is desirable to reduct the dimensionality of the system by projecting the system onto a chosen number of EOFs. If this has been done, then $x_i(t)$ represents the time-varying coefficient of the ith EOF rather than the value of the process at the ith geographical location. Also, $\mathbf{\Lambda}$ is diagonal in EOF space. The formalism here is valid in both geographical space and EOF space.

The Langevin equation (1) is a stochastic differential equation. The transition probability $p(\mathbf{x}, t+\tau|\mathbf{x}', t)$ that the pattern $\mathbf{x}'$ at time t will be followed by the pattern $\mathbf{x}$ at later time $t+\tau$ obeys a Fokker–Planck equation (Arnold 1974):

$$\frac{dp(\mathbf{x}, t+\tau|\mathbf{x}', t)}{d\tau} = \sum_{i,j} \left\{ -B_{ij} \frac{\partial}{\partial x_i} (x_j p(\mathbf{x}, t + \tau|\mathbf{x}', t)) + \frac{1}{2} Q_{ij} \frac{\partial^2}{\partial x_i \partial x_j} p(\mathbf{x}, t+\tau|\mathbf{x}', t) \right\} \tag{4}$$

with initial condition

$$p(\mathbf{x}, t|\mathbf{x}', t) = \delta(\mathbf{x} - \mathbf{x}'). \tag{5}$$

Multiplying Eq. (4) by $x_m x_n$, integrating by parts, and using the fact that $\langle x_m x_n \rangle$ is stationary, we obtain

$$\sum_i B_{im} \langle x_m x_n \rangle + \sum_i B_{in} \langle x_m x_n \rangle + Q_{mn} = 0, \tag{6a}$$

or, in matrix form,

$$\mathbf{B}\mathbf{\Lambda} + \mathbf{\Lambda}\mathbf{B}^{\mathrm{T}} + \mathbf{Q} = \mathbf{0}. \tag{6b}$$

Similarly, we multiply Eq. (4) by $x_m(t+\tau)x_n(t)$ and again integrate by parts. Because the equation is expressed with the value of $\mathbf{x}(t)$ given (and equal to $\mathbf{x}'$), the derivatives operate only on the value of $\mathbf{x}$ at time $t+\tau$. Solving the resulting differential equation for $\langle x_m t+\tau) x_n(t) \rangle$, we find, in matrix form,

$$\langle \mathbf{x}(t+\tau)\mathbf{x}^{\mathrm{T}}(t) \rangle = \exp(\mathbf{B}\tau)\mathbf{\Lambda}. \tag{7}$$

We note in passing that comparison of Eq. (6) with results from the Langevin equation (1) yields an interesting theoretical identity:

$$\langle \mathbf{x}\boldsymbol{\xi}^{\mathrm{T}} \rangle + \langle \boldsymbol{\xi}\mathbf{x}^{\mathrm{T}} \rangle = \mathbf{Q}. \tag{8}$$

For notational convenience we identify the Green function

$$\mathbf{G}(\tau) = \exp(\mathbf{B}\tau) = \langle \mathbf{x}(t+\tau)\mathbf{x}^{\mathrm{T}}(t) \rangle \mathbf{\Lambda}^{-1}. \tag{9}$$

If the system is stable, then $\mathbf{G}(\tau)$ must tend to zero at long lags τ. The Fokker–Planck equation (4) has a Gaussian solution (Riskin 1984)

$$p(\mathbf{x}, t+\tau|\mathbf{x}', t) = (2\pi)^{-N/2} (\det \sigma(\tau))^{-1/2} \times \exp\left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{G}(\tau)\mathbf{x}')^{\mathrm{T}} \sigma^{-1} (\mathbf{x} - \mathbf{G}(\tau)\mathbf{x}') \right\}, \tag{10}$$

where

$$\sigma(\tau) = \mathbf{\Lambda} - \mathbf{G}(\tau)\mathbf{\Lambda}\mathbf{G}^{\mathrm{T}}(\tau). \tag{11}$$

The multivariate Gaussian is maximized when $\mathbf{x}(t+\tau)$ is equal to $\mathbf{G}(\tau)\mathbf{x}'(t)$; therefore, the most probable pattern at time $t+\tau$ given pattern $\mathbf{x}'$ at t is simply $\mathbf{G}(\tau)\mathbf{x}'(t)$.

The physical interpretation of $\sigma(\tau)$ is obtained by considering the discretized process. By integrating Eq. (1) from t to $t+\tau$, we find

$$\mathbf{x}(t+\tau) = e^{\mathbf{B}\tau}\mathbf{x}(t) + e^{\mathbf{B}(t+\tau)} \int_t^{t+\tau} e^{-\mathbf{B}t'} \boldsymbol{\xi}(t') dt', \tag{12}$$

or, defining the colored noise $\boldsymbol{\zeta}(t, \tau)$,

$$\mathbf{x}(t+\tau) = \mathbf{G}(\tau)\mathbf{x}(t) + \boldsymbol{\zeta}(t, \tau). \tag{13}$$

This is the form of the Markov process which has been used in past POP analyses (e.g., von Storch et al. 1988), with $\boldsymbol{\zeta}(t, \tau)$ being approximated as white. Here, however, we do not consider an intrinsically discrete process with $\boldsymbol{\zeta}(t, \tau)$ being an instantaneous blip at time $t+\tau$, but rather a continuous process which has been discretely sampled.

From Eqs. (6) and (12), we find the covariance matrix of $\boldsymbol{\zeta}$ as a function of τ:

$$\langle \boldsymbol{\zeta}\boldsymbol{\zeta}^{\mathrm{T}} \rangle = \mathbf{\Lambda} - \mathbf{G}(\tau)\mathbf{\Lambda}\mathbf{G}^{\mathrm{T}}(\tau) = \sigma(\tau). \tag{14}$$

At $\tau = 0$, this matrix is null, as required by Eqs. (5) and (12). At long time lags τ, the covariance matrix of the colored noise tends to the covariance of the process $\mathbf{x}$ itself.

The transition probability [Eq. (10)] can be manipulated to derive several other useful distributions. Most useful would be the probability $p(x_m, t+\tau|x_n', t)$ that a large anomaly at position n and time t will be followed by a large (or small) anomaly at position m and later time $t+\tau$. In order to obtain this, we need the stationary probability $W(\mathbf{x})$ of any pattern $\mathbf{x}$, the stationary probability $W(x_m)$ of any mth component of $\mathbf{x}$, the joint probability $p(\mathbf{x}, t+\tau, \mathbf{x}', t)$ of having both pattern $\mathbf{x}'$ at time t and $\mathbf{x}$ at later time $t+\tau$, and the joint probability $p(x_m, t+\tau, x_n', t)$ that the nth component of $\mathbf{x}'$ has value x_n at time t and the mth component of $\mathbf{x}$ has value x_m at later time $t+\tau$.

As noted above, the stability of the process demands that $\mathbf{G}(\tau)$ tend to zero at long time lags τ. The transition probability [Eq. (10)] then tends to the stationary distribution $W(\mathbf{x})$ of the process $\mathbf{x}$,

$$W(\mathbf{x}) = (2\pi)^{-N/2} (\det \mathbf{\Lambda})^{-1/2} \exp\left\{ -\frac{1}{2} \mathbf{x}^{\mathrm{T}} \mathbf{\Lambda}^{-1} \mathbf{x} \right\}. \tag{15}$$

That is, $\mathbf{x}$ is normally distributed with mean zero and covarance matrix $\mathbf{\Lambda}$:

$$W(\mathbf{x}) = \mathcal{N}(\mathbf{0}, \mathbf{\Lambda}). \tag{16}$$

The stationary distribution of the mth component x_m

of x is also Gaussian (Wilks 1962, p. 168, Theorem 7.4.4),

$$W(x_m) = \mathcal{N}(0, \Lambda_{mm}). \quad (17)$$

The joint probability that the pattern at time t is $\mathbf{x}'$ and that the pattern at time $t + \tau$ is $\mathbf{x}$ is simply the product of Eqs. (10) and (18)

$$p(\mathbf{x}, t + \tau, \mathbf{x}', t) = p(\mathbf{x}, t + \tau | \mathbf{x}', t) W(\mathbf{x}'). \quad (18)$$

This is again a Gaussian distribution for the $2N$-dimensional random process $[\mathbf{x}^T, \mathbf{x}'^T]^T$

$$p(\mathbf{x}, t + \tau, \mathbf{x}', t) = \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}), \quad (19)$$

where, from Eqs. (10), (15) and (18),

$$\boldsymbol{\Gamma}^{-1} = \begin{bmatrix} \sigma^{-1} & -\sigma^{-1}\mathbf{G} \\ -\mathbf{G}^T\sigma^{-1} & (\mathbf{G}^T\sigma^{-1}\mathbf{G} + \boldsymbol{\Lambda}^{-1}) \end{bmatrix}. \quad (20)$$

Inverting this matrix,

$$\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{\Lambda} & \mathbf{G}\boldsymbol{\Lambda} \\ \boldsymbol{\Lambda}\mathbf{G}^T & \boldsymbol{\Lambda} \end{bmatrix}. \quad (21)$$

The upper left and lower right blocks of this matrix represent the equal-time covariance matrices of $\mathbf{x}(t + \tau)$ and $\mathbf{x}'(t)$, respectively. We remember that the covariance matrix is stationary. The upper right and lower left blocks represent the covariances between the patterns at time t and those at time $t + \tau$.

We are now in a position to determine the joint probability $p(x_m, t + \tau, x'_n, t)$ that the nth component of $\mathbf{x}'$ has value x_n at time t and the mth component of $\mathbf{x}$ has value x_m at later time $t + \tau$ (Wilks 1962, 168–169, Theorem 7.4.5),

$$p(x_m, t + \tau, x'_n, t) = \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Gamma_{mm} & \Gamma_{m,N+n} \\ \Gamma_{N+n,m} & \Gamma_{N+n,N+n} \end{bmatrix}\right). \quad (22)$$

The conditional probability $p(x_m, t + \tau | x'_n, t)$ that a value x'_n at position n and time t will be followed by the value x_m at position i and later time $t + \tau$ now follows:

$$p(x_m, t + \tau | x'_n, t) = p(x_m, t + \tau, x'_n, t)/W(x'_n). \quad (23)$$

The goal now is to obtain these probabilities directly from the time series for any lag τ. This can be done using the POP analysis.

3. Principal oscillation pattern (POP) analysis: background

Let the αth column of $\mathbf{u}$ be the αth eigenvector (or POP) $\mathbf{u}_\alpha$ of $\mathbf{B}$, corresponding to eigenvalue β_α, then

$$\mathbf{Bu} = \mathbf{u}\boldsymbol{\beta}, \quad (24)$$

where $\boldsymbol{\beta}$ is the diagonal eigenvalue matrix. The transpose of $\mathbf{B}$ has the same eigenvalues as $\mathbf{B}$, but different eigenvectors $\mathbf{v}_\alpha$,

$$\mathbf{B}^T\mathbf{v} = \mathbf{v}\boldsymbol{\beta}. \quad (25)$$

The matrices $\mathbf{u}$ and $\mathbf{v}$ form a biorthogonal set,

$$\mathbf{u}\mathbf{v}^T = \mathbf{u}^T\mathbf{v} = \mathbf{I}, \quad (26)$$

where $\mathbf{I}$ is the identity matrix. Note that an infinity of normalizations satisfies Eq. (26) and recall that $\mathbf{B}$ and $\mathbf{G}(\tau)$ share the same eigenfunctions. The spectral decomposition of $\mathbf{B}$ and $\mathbf{G}(\tau)$ is now immediate:

$$\mathbf{B} = \mathbf{u}\boldsymbol{\beta}\mathbf{v}^T \quad (27)$$

$$\mathbf{G}(\tau) = \mathbf{u}e^{\mathbf{B}\tau}\mathbf{v}^T. \quad (28)$$

In practice, we find the eigenvalues and eigenvectors of $\mathbf{G}(\tau)$ and $\mathbf{G}^T(\tau)$ directly from the data [Eq. (9)] before generating $\mathbf{B}$ using the spectral decomposition. The covariance matrix of the driving noise $\mathbf{Q}$ (not σ!) can then be calculated using $\mathbf{B}$ and $\boldsymbol{\Lambda}$. The αth eigenvalue of $\mathbf{G}(\tau)$ is $\exp(\beta_\alpha\tau)$, where β_α is the αth eigenvalue of $\mathbf{B}$. The quantities β_α are generally complex, so that a period of oscillation as well as a decay rate is obtained from the analysis. It is therefore possible to determine if the feedback mechanisms occur in traveling waves, standing waves or simple exponentially decaying structures. Note that since $\mathbf{u}$ and $\mathbf{v}^T$ are orthogonal, the POPs with the longest e-folding times become the most probable structures after several time steps.

4. Skill

No forecasting method is complete without some estimate of its accuracy and the length of time over which the forecasts can be believed. Because the dynamical system is stochastic, the theory predicts that perfect forecasts at nonzero lag are highly improbable. As a measure of the expected forecast error, we introduce the relative discrepancy $\delta(\tau)$. This is the ratio of the mean square discrepancy between the observation $\mathbf{x}(t + \tau)$ at time $t + \tau$ and its prediction $\mathbf{G}(\tau)\mathbf{x}(t)$ to the trace of the covariance matrix. That is,

$$\delta(\tau) = \langle (\mathbf{x}(t + \tau) - \mathbf{G}(\tau)\mathbf{x}(t))^T(\mathbf{x}(t + \tau) - \mathbf{G}(\tau)\mathbf{x}(t)) \rangle / \mathrm{tr}(\boldsymbol{\Lambda}). \quad (29)$$

This quantity is generally nonzero for finite lag τ, even for processes well described by a first-order Markov process. For a system whose true dynamics are governed by such a system, the theoretical discrepancy is

$$\delta(\tau)_{\mathrm{th}} = \mathrm{tr}(\sigma(\tau))/\mathrm{tr}(\boldsymbol{\Lambda}), \quad (30)$$

a quantity which is zero at $\tau = 0$ and goes to one as τ approaches infinity.

5. The POP analysis: recipe

1) The theory has been presented for centered (i.e., zero-mean) multivariate time series. It is possible to

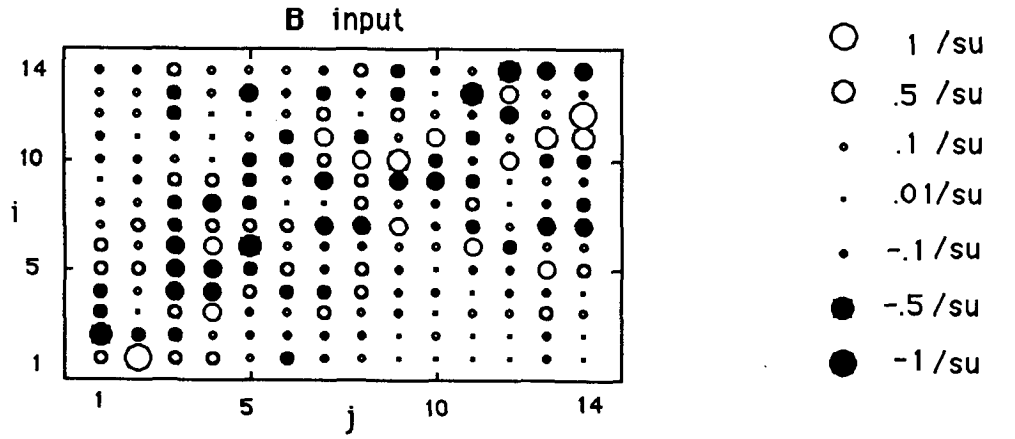


FIG. 1. Deterministic feedback matrix **B** as input. Area of symbol is proportional to matrix element. Filled symbols are negative; open symbols are positive.

derive most of the theory for noncentered processes, but it is easiest to center the process, work with the anomalies, and then add in the pattern means at the end if necessary.

2) Form the covariance matrix of the process [Eq. (3)]. It is now possible to calculate the stationary probability distribution of $\mathbf{x}$ [Eq. (15)] and of any component of $\mathbf{x}$ [Eq. (17)].

3) Project the centered data onto a chosen number of EOFs. This is a practical necessity for two reasons: i) the usually huge number of degrees of freedom can be thereby reduced; ii) the covariance matrix $\mathbf{\Lambda}$ and its inverse appear often in the calculations. Program run time can be drastically reduced by performing the calculations entirely in EOF space and then transforming them into geographical space for interpretation. After projection onto EOFs, we may calculate the stationary probability distribution of $\mathbf{x}$ [Eq. (15)] and of any component of $\mathbf{x}$ [Eq. (17)], where $x_m(t)$ is now the time-varying coefficient of the mth EOF.

4) Calculate the sample estimate of the one-lag correlation matrix and the inverse of $\mathbf{\Lambda}$ to find $\mathbf{G}(\tau)$ [Eq. (9)]. Find the eigenvalues $\exp(\beta_\alpha \tau)$ and eigenvectors (POPs) $\mathbf{u}_\alpha$ of $\mathbf{G}(\tau)$. Then find the eigenvectors $\mathbf{v}_\alpha$ of $\mathbf{G}^{\mathrm{T}}(\tau)$. It is now possible to generate the most probable anomaly pattern $\mathbf{G}(\tau)\mathbf{x}(t)$ at time $t+\tau$ given the pattern $\mathbf{x}(t)$ at earlier time t. *Programmer's note*: because the POPs with longest e-folding times eventually become the most probable structures, it is logical to sort the eigenvalues by decreasing e-folding time. Complex conjugate eigenvalues have the same e-folding time. Therefore, careless sorting can result in a complex $\mathbf{u}_\alpha$ being paired not with its corresponding $\mathbf{v}_\alpha$, to which it is parallel, but rather with the complex conjugate of

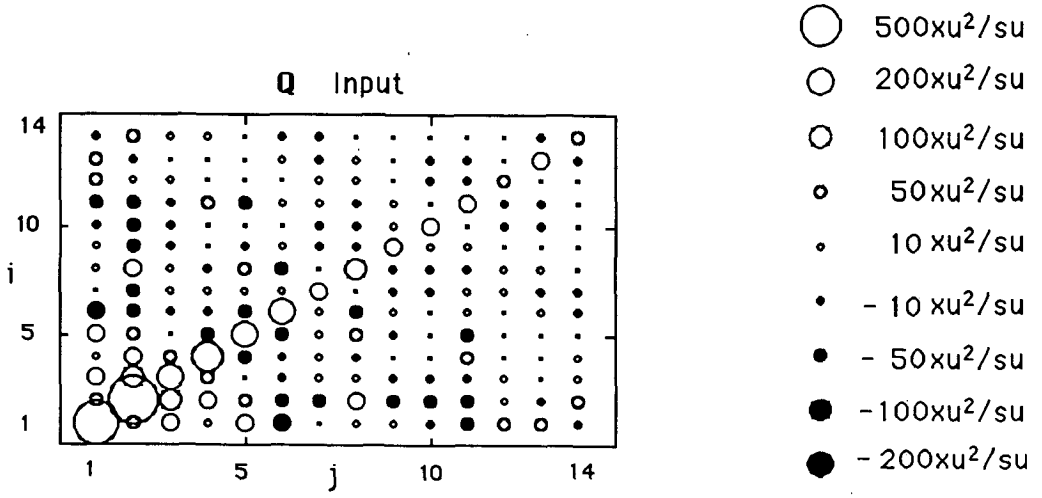


FIG. 2. Noise covariance matrix **Q** as input. Area of symbol is proportional to matrix element. Filled symbols are negative; open symbols are positive.

TABLE 1. POP eigenvalues and time scales for true Markov process.

	Input B	2000 points	1000 points	500 points	300 points
			POP 1		
β (1/su)	$-.031 \pm .833i$	$-.022 \pm .828i$	$-.022 \pm .818i$	$-.018 \pm .815i$	$-.015 \pm .810i$
e-fold (su)	31.82	46.29	45.91	56.48	67.67
period (su)	7.54	7.59	7.68	7.71	7.75
			POP 2		
β (1/su)	$-.052 \pm .342i$	$-.060 \pm .343i$	$-.068 \pm .339i$	$-.095 \pm .337i$	$-.078 \pm .342i$
e-fold (su)	19.27	16.65	14.79	10.48	12.72
period (su)	18.36	18.30	18.56	18.60	18.37
			POP 3		
β (1/su)	$-.063 \pm .563i$	$-.071 \pm .562i$	$-.083 \pm .582i$	$-.084 \pm .568i$	$-.050 \pm .565i$
e-fold (su)	15.96	14.04	12.01	11.87	19.74
period (su)	11.15	11.18	10.80	11.07	11.12
			POP 4		
β (1/su)	$-.075 \pm .185i$	$-.099 \pm .192i$	$-.088 \pm .199i$	$-.066 \pm .200i$	$-.096 \pm .209i$
e-fold (su)	13.27	10.07	11.40	15.09	10.42
period (su)	33.92	32.65	31.63	31.47	30.03
			POP 5		
β (1/su)	$-.092 \pm .523i$	$-.090 \pm .525i$	$-.065 \pm .496i$	$-.074 \pm .502i$	$-.108 \pm .494i$
e-fold (su)	10.90	11.15	15.33	12.42	9.23
period (su)	12.00	11.97	12.66	12.51	12.71
			POP 6		
β (1/su)	$-.099 \pm .758i$	$-.090 \pm .754i$	$-.080 \pm .753i$	$-.091 \pm .754i$	$-.097 \pm .748i$
e-fold (su)	10.13	11.06	12.54	10.96	10.29
period (su)	8.29	8.34	8.34	8.33	8.40
			POP 7		
β (1/su)	$-.417 \pm .651i$	$-.429 \pm .691i$	$-.416 \pm .738i$	$-.429 \pm .772i$	$-.386 \pm .786i$
e-fold (su)	2.40	2.33	2.40	2.33	2.59
period (su)	9.65	9.09	8.40	8.14	8.00

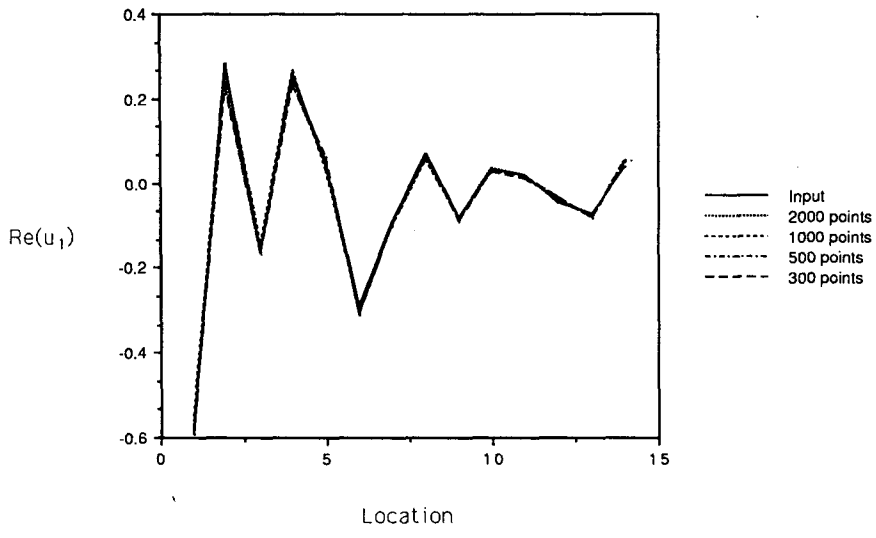
$\mathbf{v}_\alpha$, to which $\mathbf{u}_\alpha$ is orthogonal. This error manifests itself by a "program crash" in the normalization routine (see step 5) or by unrealistically large entries in the normalized eigenfunction elements.

5) Normalize $\mathbf{u}_\alpha$ and $\mathbf{v}_\alpha$ according to Eq. (26). *Programmer's note*: this normalization allows purely imaginary $\mathbf{u}_\alpha$ and $\mathbf{v}_\alpha$ for a real β_α. Although this is acceptable analytically, it can be a practical annoyance. It is worth the few extra lines of computer code to ensure that $\mathbf{u}_\alpha$ and $\mathbf{v}_\alpha$ are real for real β_α. The spectral decomposition of $\mathbf{G}$ can now be used to calculate $\mathbf{G}$ for any lag τ and the most probable pattern $\mathbf{x}(t+\tau) = \mathbf{G}(\tau)\mathbf{x}'(t)$ predicted from the current pattern $\mathbf{x}'(t)$.

6) Find the eigenvalues β_α of $\mathbf{B}$ from the eigenvalues $\exp(\beta_\alpha\tau)$ of $\mathbf{G}(\tau)$. The real part of β_α, which must be negative for a stable system, gives the decay rate of our ability to predict the αth POP; the complex part of β_α gives the αth POP's oscillation frequency.

7) Generate $\sigma(\tau)$. This can be done using either Eq. (11) or by taking the covariance of ζ directly from the data [Eqs. (13)–(14)]. Using Eq. (11) takes fewer operations. Actually, $\sigma(\tau)$ could have been calculated after step 4; however, by waiting until after step 5 it is possible to use the spectral decomposition of $\mathbf{G}$ to calculate $\mathbf{G}$ and σ as a function of varying lag τ. If desired, it is now possible to calculate the transition probability $p(\mathbf{x}, t+\tau|\mathbf{x}', t)$ that the pattern $\mathbf{x}'$ at time t will be followed by the pattern $\mathbf{x}$ at later time $t+\tau$ [Eq. (10)]. Also, the joint probability $p(\mathbf{x}, t+\tau, \mathbf{x}', t)$ that the pattern at time t is $\mathbf{x}'$ and that the pattern at time $t+\tau$ is $\mathbf{x}$ can be calculated at this point using Eq. (18). Since both these probabilities describe all the components of $\mathbf{x}$ simultaneously, their values at any particular pattern $\mathbf{x}$ are very small.

8) Generate the matrix $\mathbf{\Gamma}$ using Eq. (21). At last, the conditional probability $p(x_m, t+\tau|x'_n, t)$ that a value x'_n at position n and time t will be followed by the value x_m at position m and later time $t+\tau$ may



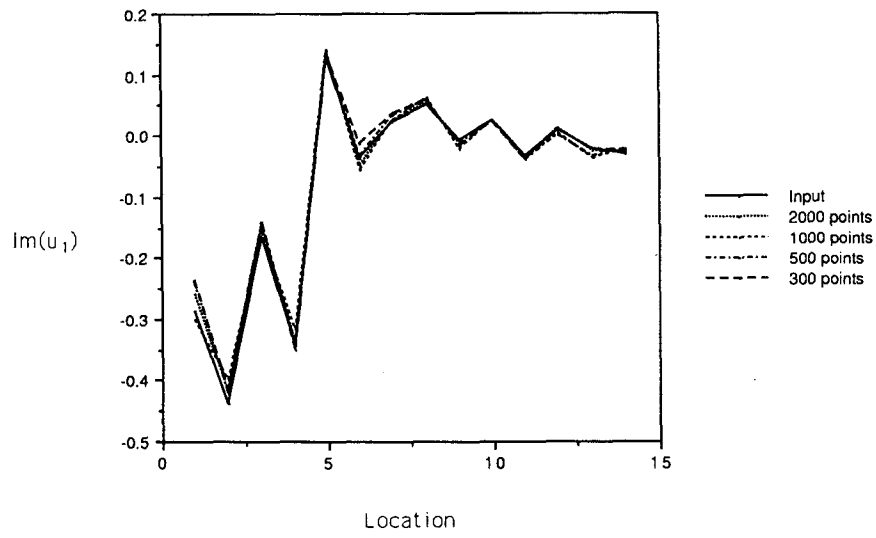


FIG. 3. Real (Fig. 3a) and imaginary (Fig. 3b) parts of $\mathbf{u}_1$ for the four series lengths considered compared with eigenfunction $\mathbf{u}_1$ of input series.

be calculated from Eqs. (17), (22) and (23). This step may actually be performed without ever performing an eigenfunction analysis of **G**, but that becomes very cumbersome if several values of τ are required.

9) Generate **B** using its spectral decomposition [Eq. (27)]. The driving noise covariance matrix **Q** can now be obtained from Eq. (6).

10) Calculate the relative discrepancy $\delta(\tau)$ [Eq. (29)] and compare with its theoretical value [Eq. (30)] for increasing lag τ.

6. Numerical examples

The typical meteorological dataset is highly multivariate and frustratingly short. The two numerical examples presented here reflect these conditions. The first example is a true Markov process that has been numerically generated to exhibit the strengths and weaknesses of the POP analysis. The second example, the deterministically chaotic Lorenz system, will show that the POP analysis can sometimes be valuable even when most of the assumptions behind it are violated.

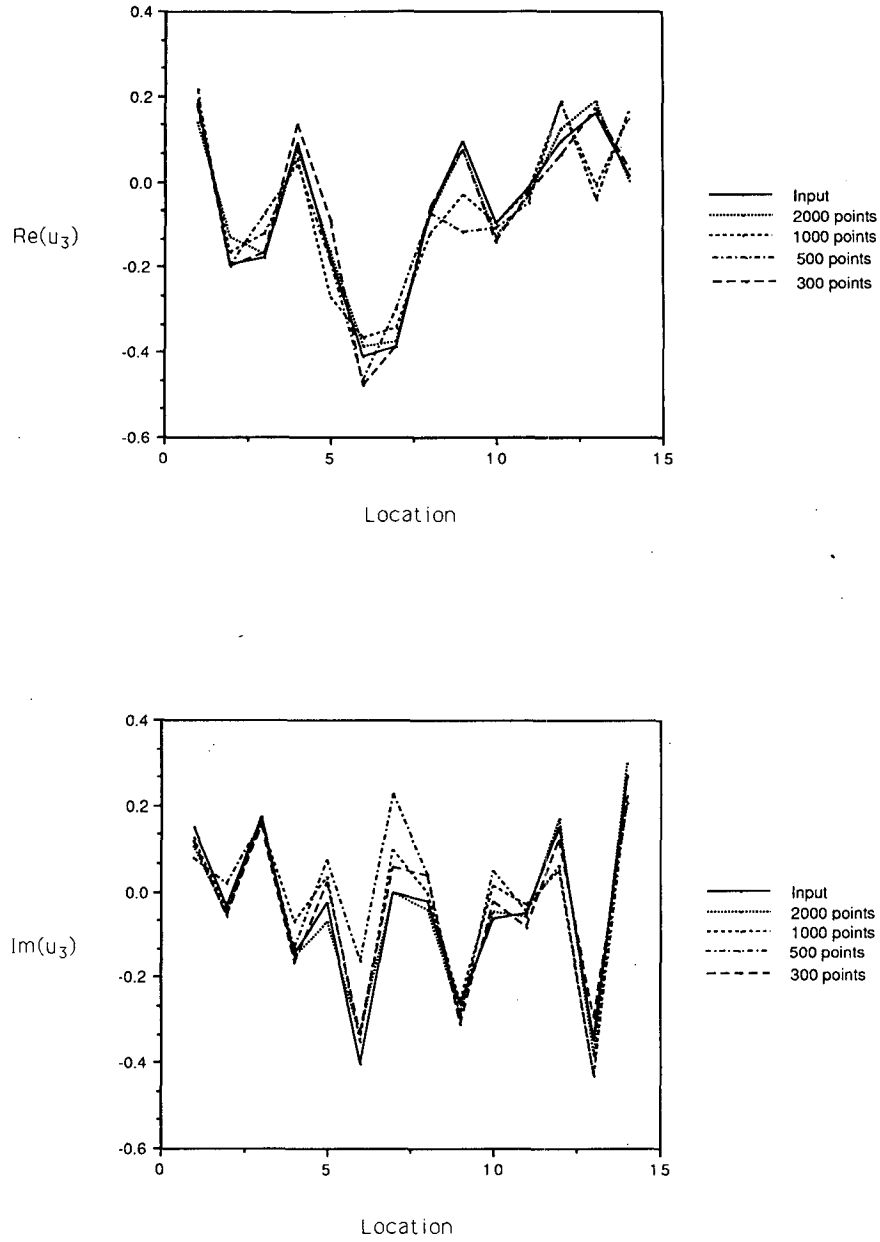


FIG. 4. As in Fig. 3a, b, but for POP u_3.

a. A true Markov process

Consider a 14-dimensional process whose "true" dynamics are governed by Eqs. (1)–(3). The matrices **B** (Fig. 1) and **Q** (Fig. 2) are selected from equations describing randomly driven, coupled, damped harmonic oscillators. There is no reason to believe that the driving noise is geographically uncorrelated and it is not uncorrelated here, although **Q** is indeed diagonally dominant at low indices. The low-indexed locations are randomly driven most strongly; the series variation in the high-indexed locations is due mainly to transfer of amplitude out of the highly excited regions by means of deterministic feedback into the less strongly excited regions. This system does not represent an optimal situation; three problems that can be expected to cause difficulty for the POP analysis are built into the matrix **B**: two POP eigenvalues are close together, one eigenvalue has an *e*-folding time much greater than the oscillation period, and one eigenvalue has an oscillation period much greater than the *e*-folding time. The chosen unit of time is our sampling interval, defined to equal 1 "su," and the units of **x** are defined as "xu." We integrate Eq. (1) using 60 time

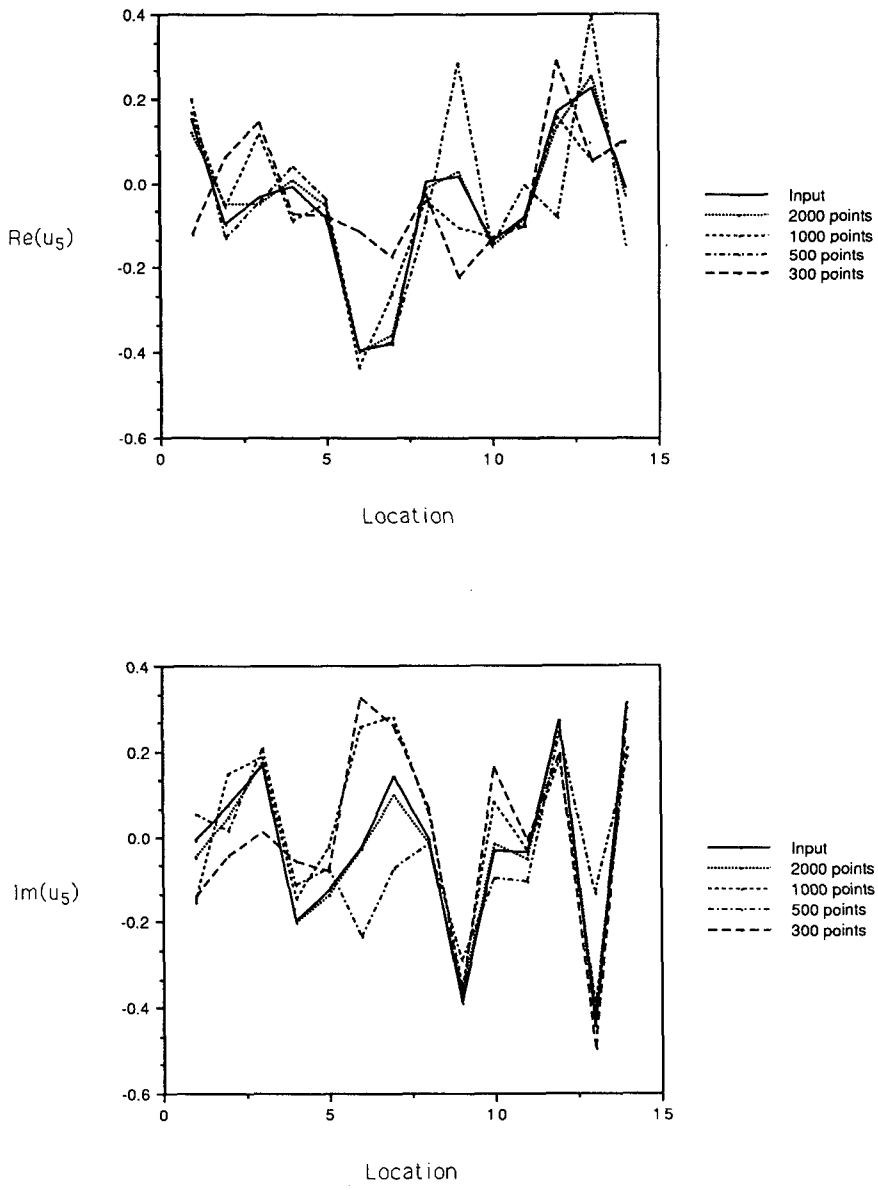


FIG. 5. As in Fig. 3a, b, but for POP $\mathbf{u}_5$.

steps per sampling interval ($dt = 1/60$ su). The number of time steps between samplings has no physical meaning other than that it is large enough to estimate the integral over a process that is nowhere differentiable and everywhere continuous (except at isolated points) by a discrete sum. The numerical generation of continuous Gaussian white noise demands some care.

Eq. (1) is rewritten as

$$d\mathbf{x} = \mathbf{B}\mathbf{x}dt + d\mathbf{Y}, \tag{31}$$

where $\mathbf{Y}$ is proportional to an N-dimensional Wiener process and has the property (Arnold 1974)

$$\langle d\mathbf{Y} d\mathbf{Y}^{\mathrm{T}} \rangle = \mathbf{Q}dt. \tag{32}$$

Let q_α be the αth eigenvalue of $\mathbf{Q}$ corresponding to eigenfunction ψ_α; that is,

$$\mathbf{Q}\boldsymbol{\psi} = \boldsymbol{\psi}\mathbf{q}. \tag{33}$$

An Euler method (Rümelin 1982) was used to integrate (31) with the algorithm

$$x_i(t + dt) = x_i(t) + \sum_{j=1}^{N} B_{ij}x_j dt + \sum_{\alpha=1}^{N} \psi_{i\alpha}\sqrt{q_\alpha dt}\,\mathcal{R}. \tag{34}$$

The zero-mean Gaussian deviate $\mathcal{R}$ has unit variance and was newly chosen at each time step. The Gaussian random deviate generator (Press et al. 1986) was ex-

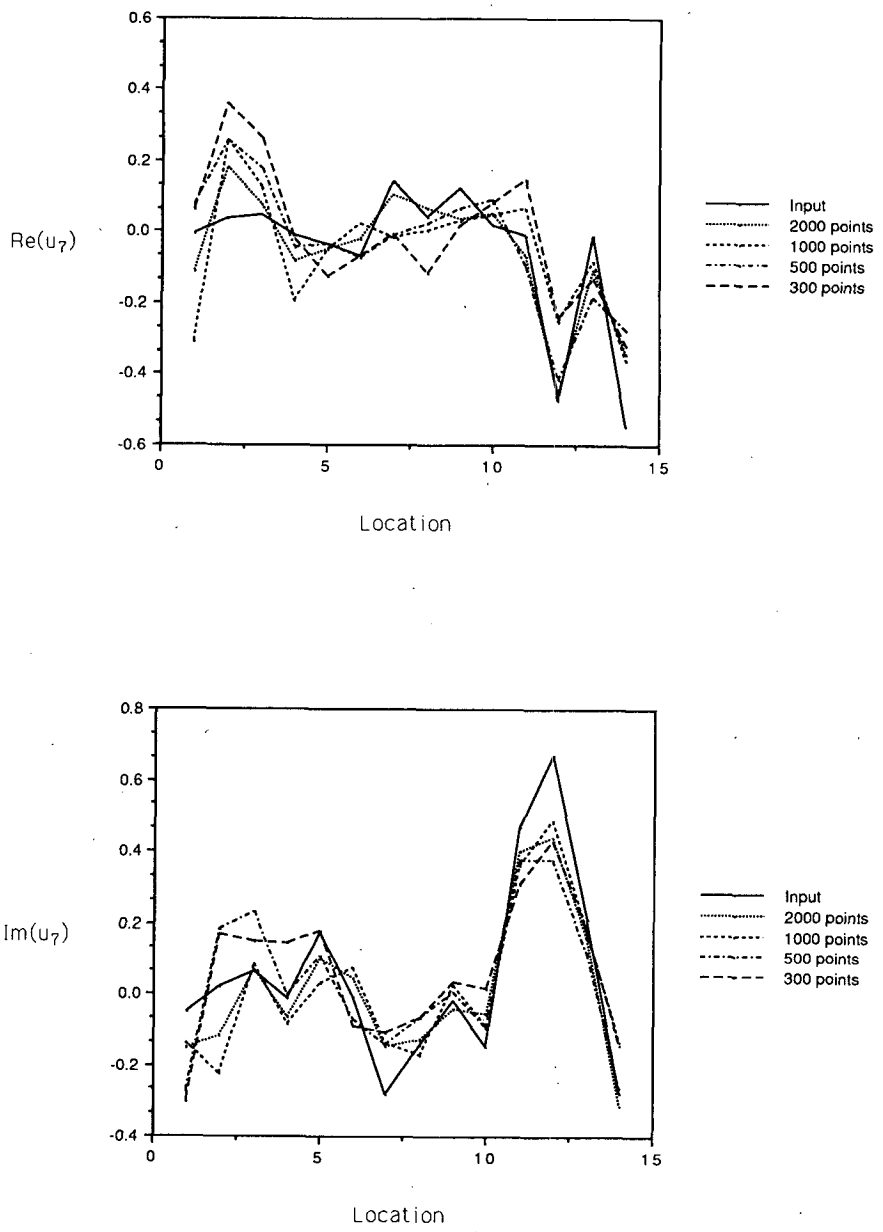


FIG. 6. Same as Fig. 3a, b, but for POP $\mathbf{u}_7$.

ercised for 1000 time steps. Initial conditions $x_i = 0$ [the most probable value of x_i; see Eq. (17)] were then chosen and the system (34) was allowed to propagate for 1000 time steps before the first sample was taken. In all, 2000 samples of the 14-dimensional multivariate series were recorded. The POP analysis using the first 300 samples, the first 500 samples, the first 1000 samples, and all 2000 samples was performed using the recipe provided above; the results are compared with the parameters used to generate the series. All 14 EOFs were retained so that discrepancies may be explained solely by the length of the time series and not by truncation in EOF space.

The generating feedback matrix **B** has seven pairs of complex conjugate eigenvalues with negative real parts, a situation that was faithfully reproduced in each of the four POP analyses. The eigenvalues of the generating **B**, their associated time scales, and corresponding values from the POP analyses are presented in Table 1. The POP analysis determines the period of oscillation very well in all cases. It is less successful in determining the e-folding time when the decay is slow; i.e., when the real part of β_α is small enough that a small absolute error in β_α results in a large error in the e-folding time. This is the case in POP 1 and, to a lesser extent, POP 2. In contrast, the oscillation period of POP 4 was well-

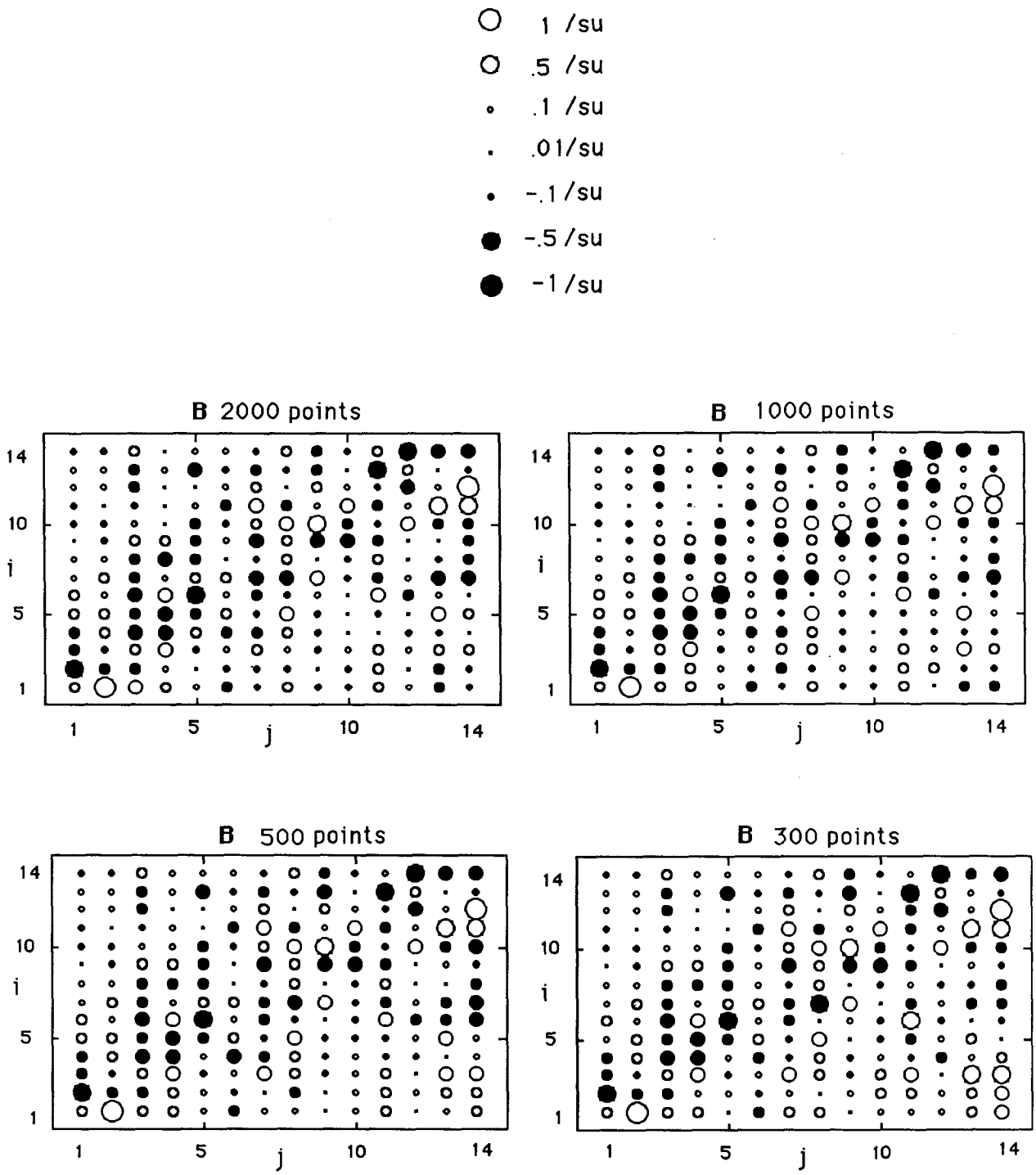


FIG. 7. Deterministic feedback matrix **B** as determined using 2000 points (Fig. 7a), 1000 points (Fig. 7b), 500 points (Fig. 7c) and 300 points (Fig. 7d). Area of symbol is proportional to matrix element. Filled symbols are negative; open symbols are positive.

determined even though it is approximately as long as the *e*-folding time of POP 1. The analysis was generally successful in determining the short to intermediate *e*-folding times.

The real and imaginary parts of $\mathbf{u}_1$, $\mathbf{u}_3$, $\mathbf{u}_5$ and $\mathbf{u}_7$ as determined by the four POP analyses are compared with the eigenfunctions of the generating matrix **B** in Figs. 3a, b–6a, b. The worst agreement in the entire

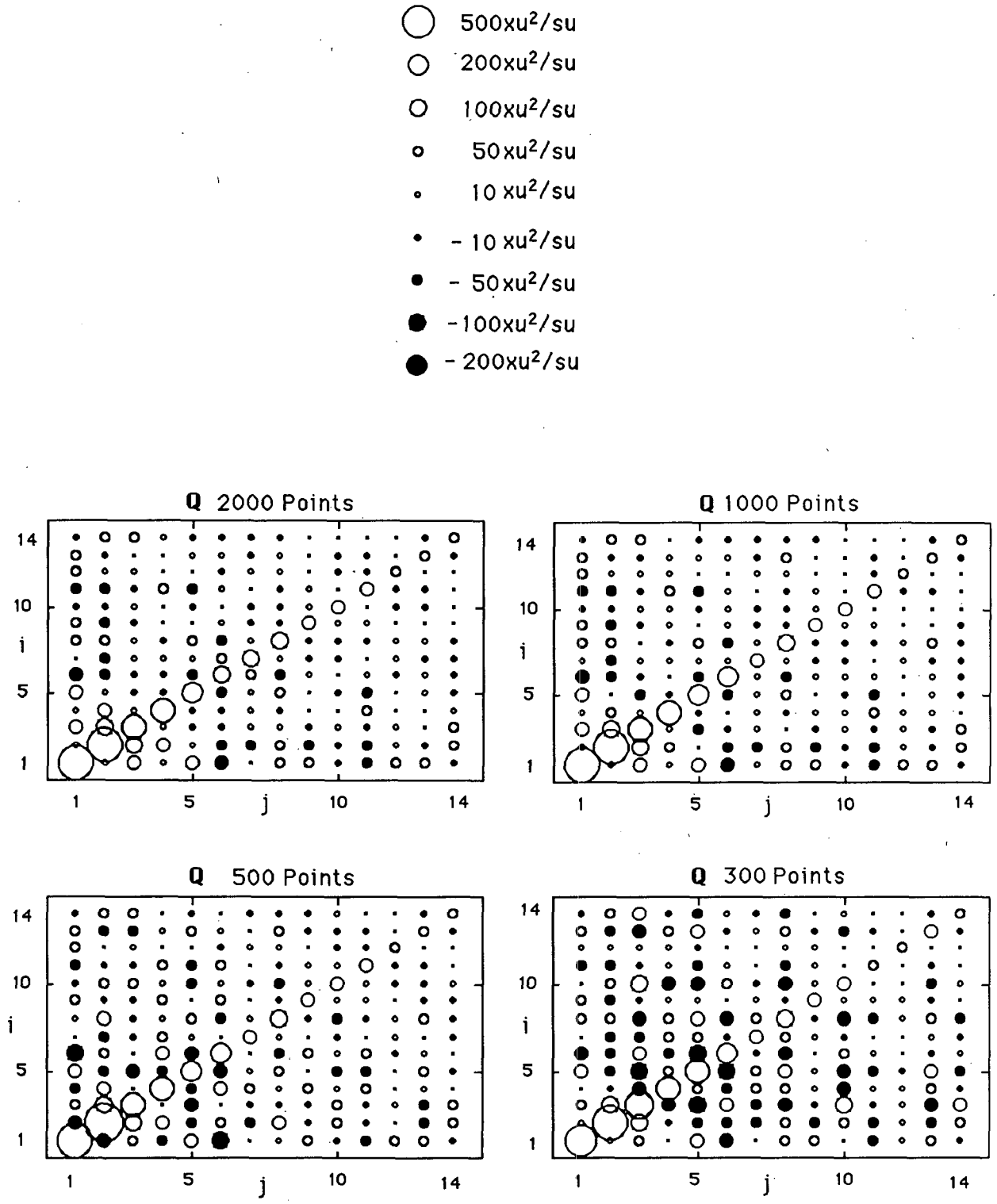


FIG. 8. Noise covariance matrix **Q** as determined using 2000 points (Fig. 8a), 1000 points (Fig. 8b), 500 points (Fig. 8c) and 300 points (Fig. 8d). Area of symbol is proportional to matrix element. Filled symbols are negative; open symbols are positive.

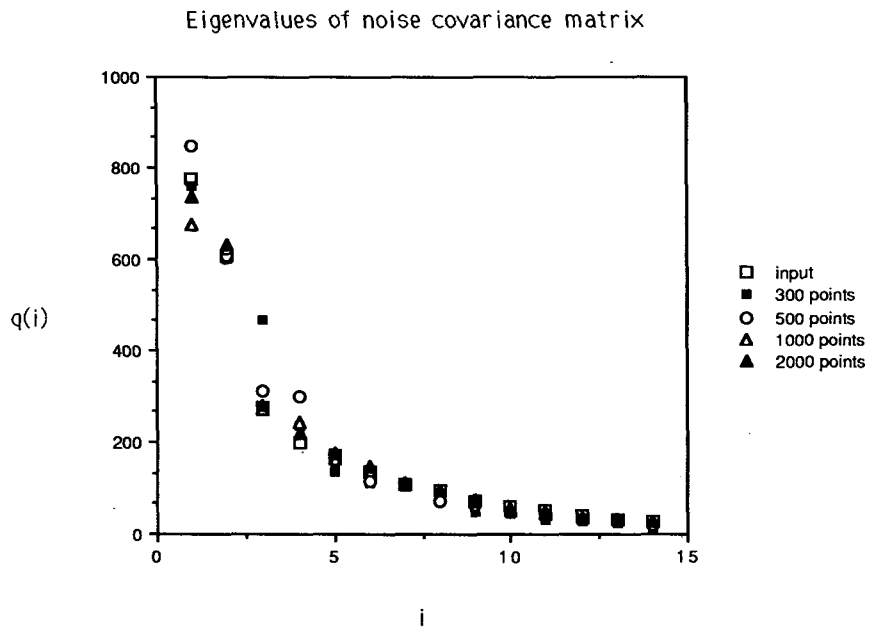


FIG. 9. Noise eigenvalues.

set is exhibited by POP 5, whose eigenvalue is very close to that of POP 6. The best agreement is found in POP 1. Typical agreement is shown in POPs 3 and 7. Eigenfunctions obtained using the 2000 point series always agreed very well with the eigenfunctions of the generating **B** and the other analyses could be trusted to reproduce the general features of each POP.

A graphical representation of the feedback matrix **B** as reconstituted by the spectral decomposition obtained from the POP analyses is compared with the generating feedback matrix in Fig. 7a–d and should be compared with the generating feedback matrix (Fig. 1). The agreement increases with the number of samples in the times series, of course, but the major features of the generating matrix are reproduced even in the 300 point case.

The matrix **Q** as determined from Eq. (6) and using the **B** found from the POP analysis is compared with the covariance matrix of the driving noise in Fig. 8a–d and is to be compared with the covariance matrix of the driving noise (Fig. 2). The general features of the random forcing are well reproduced by the analysis.

The eigenvalues of **Q** as determined from the POP analyses are compared with those of the generating noise matrix in Fig. 9; the first four eigenfunctions are shown in Fig. 10a–d. Only the 300-point series shows any trouble in reproducing reality and even it is generally successful in finding regions of high excitation. The third eigenfunction of **Q** obtained from this shortest series is similar to the true second eigenfunction, and the fourth eigenfunction of **Q** obtained from the 300-point series looks very much like it should be the third one.

The conditional probabilities calculated using the parameters obtained from the POP analysis are now compared with those resulting from histograms. Two lags, $\tau = 1$ su and $\tau = 5$ su were considered. Each (ith) location was assigned a set of ten bins centered at $x_i = 0$ and having width $\Delta x_i = 0.3\sqrt{\Lambda_{ii}}$, except bins 1 and 10, which extend to negative and positive infinity, respectively. The number of occurrences of values x_i lying within each bin at each location was counted in the 2000-point series. Next, the number of occurrences of values x_i lying within each mth bin following at lag τ the value of x_j' lying in the nth bin was counted for all i, j, m and n. The conditional probability of finding a value of x_i in the mth bin given that x_j' lies in the nth bin τ amount of time earlier was then calculated from these histograms. This method of calculating conditional probabilities is clumsy, time-consuming, and inaccurate at the tails of the distribution due to the paucity of samples there.

More satisfying are the conditional probabilities calculated from parameters resulting from the POP analysis. The conditional probability of finding a value of $x_i \pm \frac{1}{2}\Delta x_i$, given x_j', τ amount of time earlier is estimated as $p(x_i, t + \tau \mid x_j', t)\Delta x_i$. Figures 11a, b show contour plots of this quantity for $i = j = 1$, $\tau = 1$ su, as calculated from parameters obtained from POP analysis of the 2000-point series and the 300-point series. The smooth surfaces have been obtained using negative exponential interpolation (Wilkinson 1988). These contours are compared with results from the histograms described above (Fig. 11c). The structure near $x_1' = -300$ in Fig. 11c is not real but is rather due to the paucity of samples in this region. As expected,

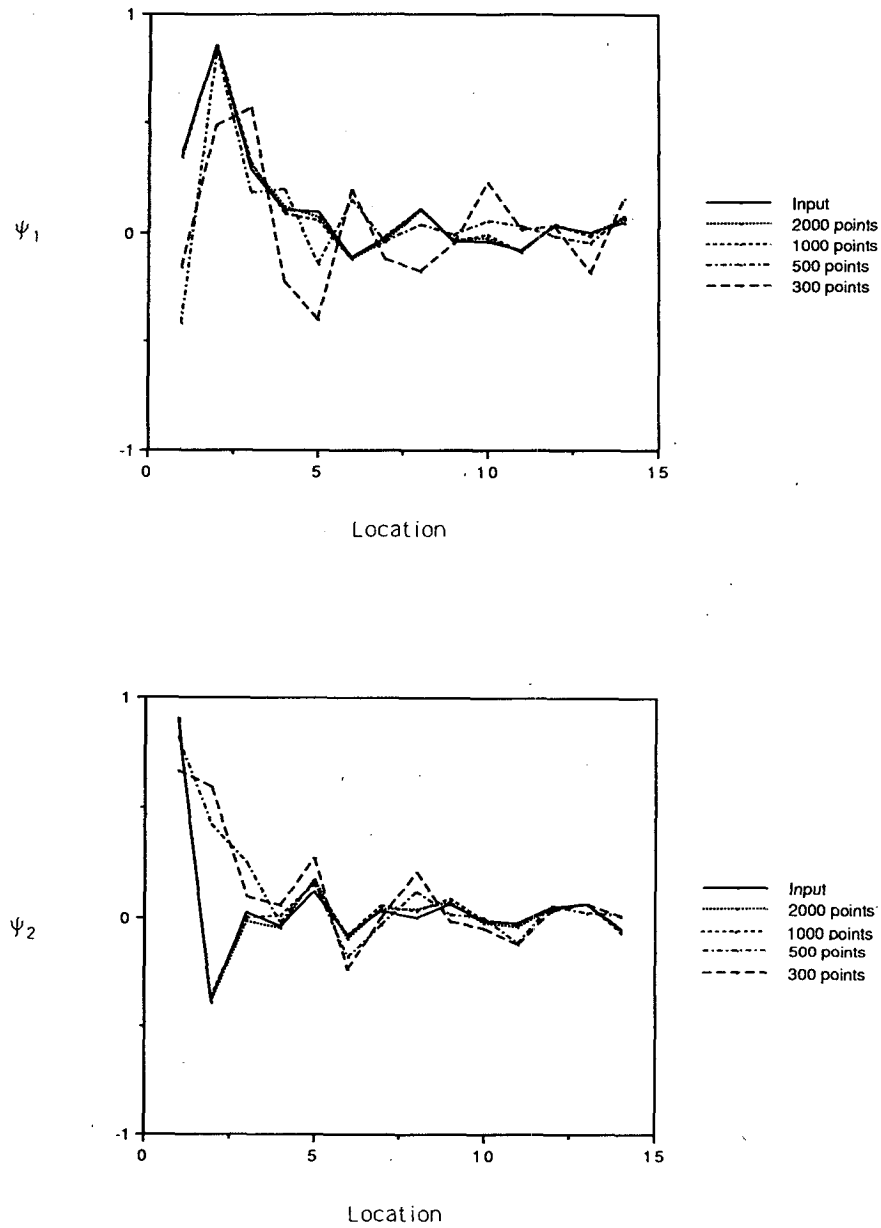


FIG. 10. Eigenfunctions of noise covariance matrix: eigenfunction 1 (Fig. 10a), eigenfunction 2 (Fig. 10b), eigenfunction 3 (Fig. 10c) and eigenfunction 4 (Fig. 10d).

the value of x_1 at $t + 1$ su is highly (positively) correlated with its value at time t. The value of x_1 is negatively correlated with its value 5 su earlier, as shown in Fig. 12. Note that the slope is smaller, the peak is lower and the ridge is broader, indicating the smaller certainty in our predictions at longer lag. Independence would be indicated by horizontal contours. The probabilities shown in Figs. 12–14 were calculated using the 2000-point series.

Independence need not increase with lag. Consider, for example, the dependence of the value of x at location 4 upon x_1'. In Fig. 13 we see a negative correlation at $\tau = 1$ su; the peak is rather low and the ridge is broad. Remember, please, that the time scales in this numerical example are all greater than 2 su and many are greater than 10 su. At a lag of $\tau = 5$ su, sufficient time has elapsed to allow interaction between the various EOFs. We see (Fig. 14), therefore, that at $\tau = 5$ su the value of x at location 4 is about as dependent upon the value of x_1' at time 5 su earlier as x_1 itself was at shorter lag $\tau = 1$ su. The slope is steep, the peak is high, and the ridge is narrow.

The estimates of conditional probability using parameters obtained from the POP analysis do not suffer

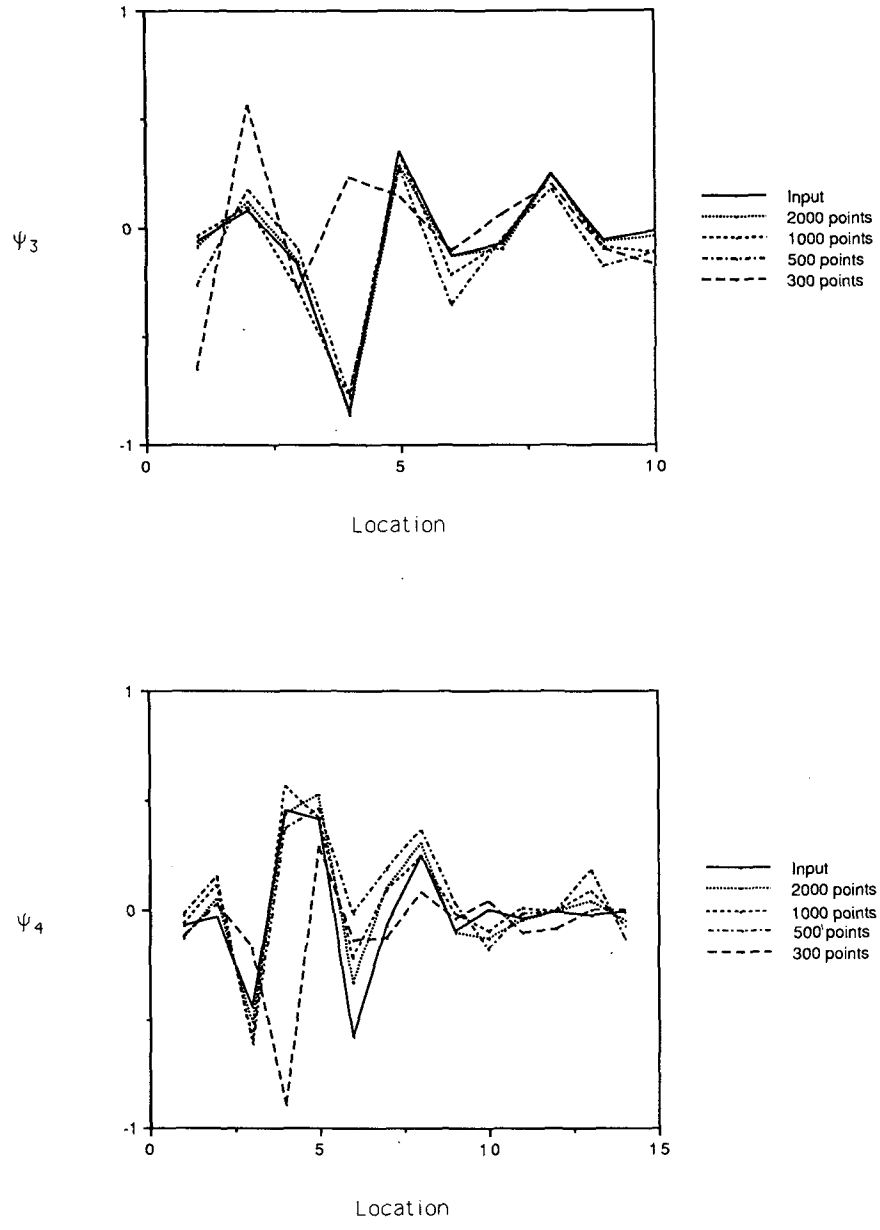


FIG. 10. (*Continued*)

from the error due to a rarity of data in the low-probability regions of phase space. For this reason, they are generally more trustworthy than those calculated using histograms. Even the 300-point series gives accurate estimates of conditional probabilities.

Finally, the efficacy of the procedure in predicting probable patterns is investigated. Here, the parameters obtained from POP analysis of the first thousand points of the time series were used to predict patterns using the second thousand. The relative discrepancy $\delta(\tau)$ was calculated from the data using Eq. (29) and compared with its theoretical value [Eq. (30)]. The excellent agreement shown in Fig. 15 is no surprise since the process considered truly is a linear Markov process. With real data, the two curves would diverge as nonlinear effects became important.

b. *The Lorenz system*

Because the atmosphere is in reality governed by nonlinear (though possibly linearizable) dynamics, it is necessary to consider what violence to the predictive ability of the POP analysis the existence of strong nonlinearities will do. For this reason, we consider the Lorenz system:

$$\frac{dx}{dt} = \sigma(y - x) \tag{34a}$$

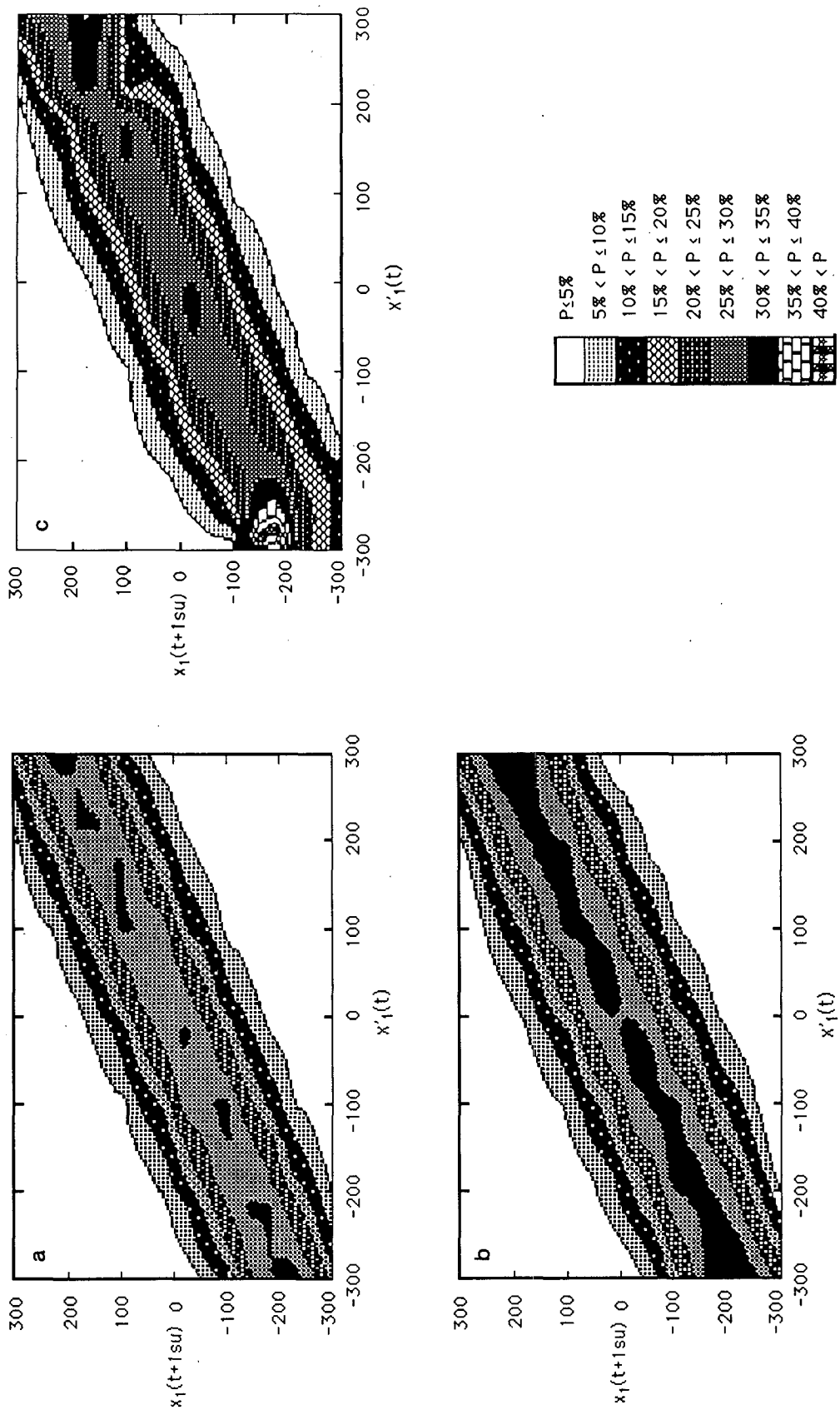


FIG. 11. Conditional probability of finding anomaly at location 1 between values $x_1 \pm \Delta x_1$ ($\Delta x_1 = 0.3\sqrt{\Lambda_{11}}$) at time $\tau = 1$ su later than given anomaly x'_1. Probabilities using 2000-point series (Fig. 11a) are compared with results from the 300-point series (Fig. 11b) and with those obtained from histograms (Fig. 11c).

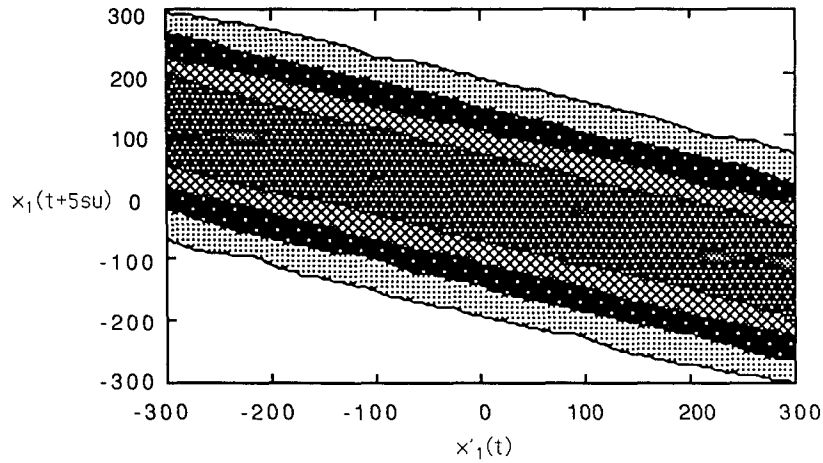


FIG. 12. Conditional probability of finding anomaly at location 1 between values $x_1 \pm \Delta x_1$ at time $\tau = 5$ su later than given anomaly x'_1.

$$\frac{dy}{dt} = rx - y - xz \tag{34b}$$

$$\frac{dz}{dt} = xy - bz \tag{34c}$$

where the parameters $\sigma = 10$, $r = 28$ and $b = 8/3$ ensure a chaotic system. The system was started with initial conditions $(x, y, z) = (1, 1, 1)$ and then integrated with a time step of 0.001 dimensionless time units (tu). The variables x, y and z are presumed to be unitless. After the system had been exercised for 1000 time steps, data was recorded and centered. Two samplings are considered. In the first case, data is recorded at every time step for a total of 8000 points. That is, the total length of the time series is 8 tu. The second sampling consists of 8000 points taken every 100 time steps (total length of time series = 800 tu).

The system in the first case is confined to one lobe of the attractor (Fig. 16). The centering of the data has no other effect than to shift the center of the spiral to position $(x, y, z) = (0, 0, 0)$. The three dimensions of the system require either three real POPs, or one real POP and one complex conjugate pair. The latter result obtains (Table 2) and it is noted that the imaginary part of the POP eigenvalues is almost two orders of magnitude larger than the real part. That is, the system estimates the path in phase space between adjacent paths as an essentially circular decaying spiral. Although the true case is an expanding spiral, a calculation of the relative discrepancy (Fig. 17) shows that the results of the POP analysis yield a good approxi-

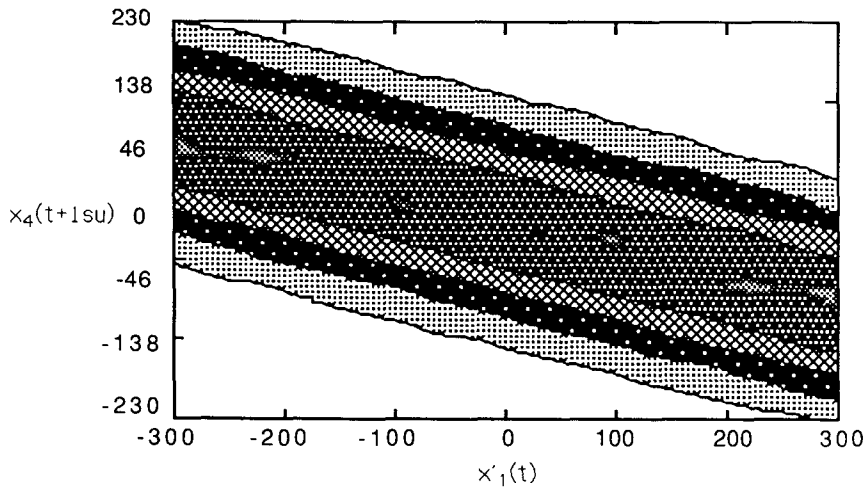


FIG. 13. Conditional probability of finding anomaly at location 4 between values $x_4 \pm \Delta x_4$ ($\Delta x_4 = 0.3\sqrt{\Lambda_{44}}$) at time $\tau = 1$ su later than given anomaly x'_1.

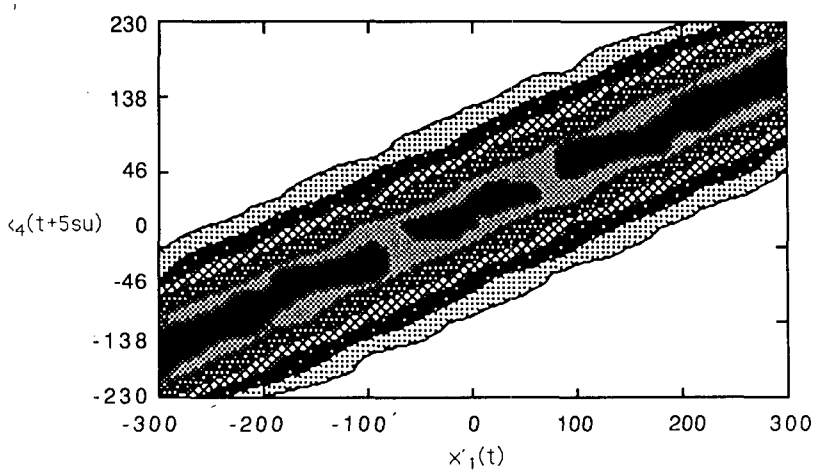


FIG. 14. Conditional probability of finding anomaly at location 4 between values $x_4 \pm \Delta x_4$ at τ time 5 su later than given anomaly x'_1.

mation for propagation prediction for the very small sampling intervals considered. That is, the nonlinearities do not have enough time to significantly couple to the field between samplings.

Calculation of the "noise" matrix **Q** [Eq. (6)], into which the POP analysis places the effects of nonlinearities, yields a result which has negative eigenvalues (Table 3). That is, the nonlinearities cannot be represented as a random driving noise. It cannot be expected that significant nonlinearities will always result in negative eigenvalues of **Q**, but negative eigenvalues of **Q** are a pretty good indication of nonlinearities.

Turning to the second sampling, we find that both lobes of the Lorenz attractor are well-represented in the trajectory of the system through phase space. A scatter plot of the first 2664 of the 8000 (centered) points is shown in Fig. 18. The analysis applied to this system yields POPs with significant real part (Table 2). The skill of prediction is not good (Fig. 17). In addition, the calculated "noise" matrix **Q** possesses a negative eigenvalue (Table 3).

7. Discussion

The POP analysis has been extended to obtain properties of the random forcing and to enable calculation of conditional probabilities. As an example, the POP analysis was applied to a numerically generated Markov process. Time series as short as 300 sampling points reproduced the parameters used in generating the series. Estimations of oscillation frequencies seemed to be generally better than those of decay rates, but only decay rates much slower than the corresponding oscillation frequencies needed to be seriously questioned. The spatial properties of the noise covariance matrix **Q** and the feedback matrix **B** were reproduced and used to calculate conditional probabilities that agreed with conditional probabilities obtained from histograms. However, the calculated probability matrices

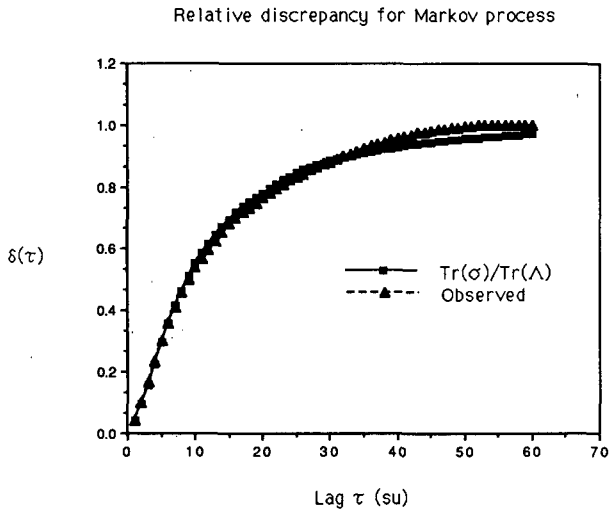


FIG. 15. Theoretical relative discrepancy (solid line) for the linear Markov process compared with observation (dotted line).

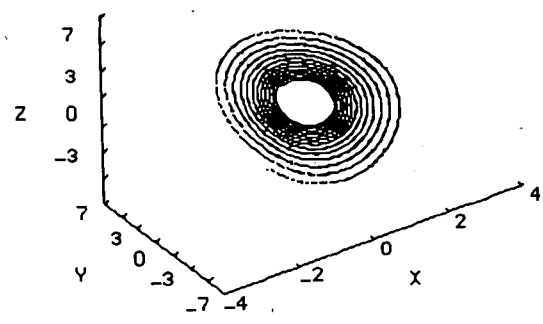


FIG. 16. Scatter diagram of Lorenz time series confined to one lobe of the attractor. Sampling = 0.001 tu.

TABLE 2. POP eigenvalues β (1/tu) obtained from analysis of Lorenz attractor.

Sampling = .001 tu	Sampling = .1 tu
$-.113 \pm 9.983i$	$-2.600 \pm 5.235i$
$-.238$	-4.609

did not suffer from the spurious structure to which probabilities obtained from histograms are subject. In addition, the spectral decomposition of **B** allows conditional probabilities at varying lags to be calculated without having to repeat the analysis.

The Green function $\mathbf{G}(\tau)$ (which is actually what von Storch et al. 1988, call the feedback matrix) is important in predicting the most probable patterns in the future; given a pattern $\mathbf{x}'(t)$ at some time t, the most probable pattern $\mathbf{x}(t+\tau)$ at a later time $t+\tau$ is $\mathbf{G}(\tau)\mathbf{x}'(t)$. As we try to predict farther and farther into the future, the most probable pattern tends to zero, and our predictive ability decays with it.

Just how fast this predictive ability decays depends upon the eigenvalues of the feedback matrix **B**; the POPs with the longest e-folding times become the most probable structures at long τ. The amplitude of these structures, however, depends upon the size of the spectral components corresponding to the slowly decaying POPs when the prediction is made. The stochastic nature of the dynamics makes perfect prediction highly improbable and a theoretical estimate of the mean square discrepancy $\delta(\tau)$ betwen the forecast and the observation has been provided. Although the agreement between the theoretical discrepancy and the observed discrepancy is very good in the example above, it should be remembered that the time series generated here are, after all, linear Markov processes. In most cases found in nature, this is not so, although the linear Markov process may be the optimal model. The discrepancy can therefore be expected to increase more quickly with lag than the theoretical value as nonlinear effects take over.

To consider the effect of nonlinearities on prediction, the POP analysis was applied to the prototype chaotic system of Lorenz, where the dynamics are nonlinear and deterministic rather than linear and stochastic. Nevertheless, the relative discrepancy between prediction and observation is very good as long as the sampling interval is small enough that $\mathbf{x}(t+\tau)$ is well described by dynamics linearized around $\mathbf{x}(t)$. This implies another restriction on the use of the POP analysis to describe a nonlinear system; the behavior of the sampled time series must be regular enough that the system linearized around the state at time t not be materially different from the system linearized around the state at any other time t'.

The connection between the statistical model used here and physical reality is provided by Hasselmann's (1976) consideration of "slow" variables, which constitute the system, and "fast" variables, which constitute

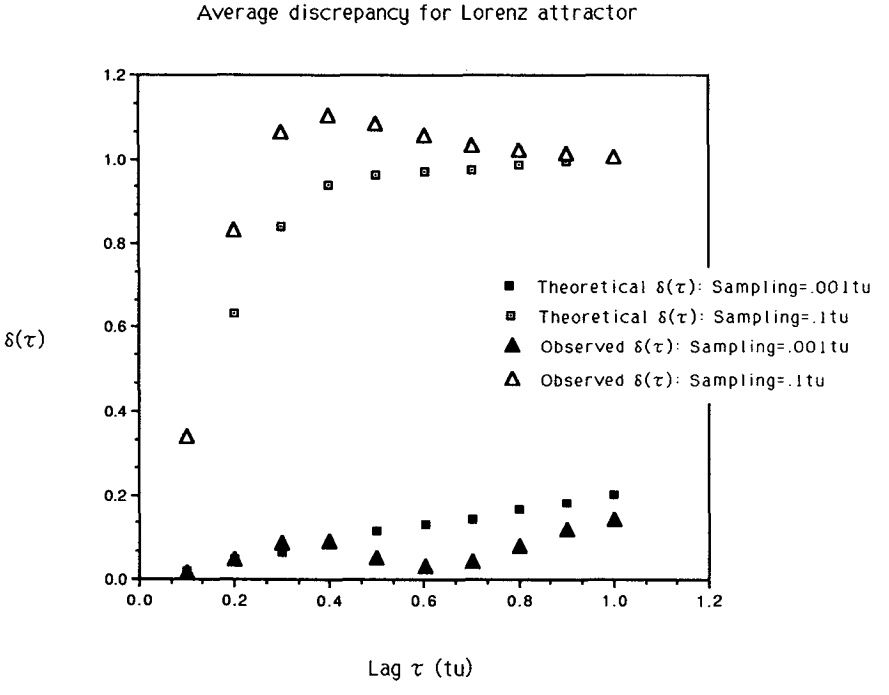


FIG. 17. Comparison of relative discrepancies for predictions of POP analysis applied to Lorenz system. Open symbols: sampling = 0.1 tu. Filled symbols: sampling = 0.001 tu.

the driving noise. That is, the system is approximated by a continuous Markov process that has been discretely sampled. Therefore, the noise term in the discretized propagation equation is not white, as has been the assumption in previous POP analyses (von Storch et al. 1988), but is rather white noise integrated over a sampling interval. How similar these two noises are depends upon the efficiency of the feedback. The results of von Storch et al. (1988) do not depend upon this distinction since that study was concerned exclusively with deterministic effects, which were completely decoupled from the noise. Here, the direction of emphasis is different; the noise is not an unwanted error but rather represents the effect of the fast variables which drive the system. The POP analysis is an effective method of obtaining the spatial properties of both the deterministic feedback and the driving energy source, the effects of which have been combined in the multivariate times series.

Another major difference between this work and previous POP analyses is that we do not concentrate on the POPs themselves as patterns, but rather on their role in the propagation of pattern changes in any appropriate space—geographical, EOF or whatever. For this reason, the POP and EOF analyses need not compete with each other; EOFs describe spatial patterns and POPs describe how they propagate. Indeed, the large number of degrees of freedom and the necessary inversion of the anomaly covariance matrix make operating in EOF space a practical necessity. The numerical calculations are performed entirely in EOF space and then transformed into geographical space for interpretation.

The main contrast between POP analysis and complex EOF analysis (Barnett 1983) lies in the consideration of energy transfer between patterns. The average energy in a meteorological pattern depends on equal-time averages of quadratic functions of quantities such as velocity components and streamfunctions. These averages are generally assumed to be stationary. When the complex EOF analysis is applied to such quantities, a basis is found where the patterns do not, *on average*, interfere or exchange energy, although the complex principal components may indeed be correlated at nonzero lag. A POP analysis, on the other hand, provides a spectral decomposition of the linear feedback matrix and particularly describes the linear deterministic processes by which energy is transferred from one geographical point to another.

TABLE 3. Eigenvalues (1/tu) of calculated matrix **Q** obtained from analysis of Lorenz attractor.

Sampling = .001 tu	Sampling = .1 tu
$-.125 \times 10^4$	$-.191 \times 10^3$
$-.643 \times 10^{-1}$	$.179 \times 10^4$
$.267 \times 10^4$	$.673 \times 10^4$

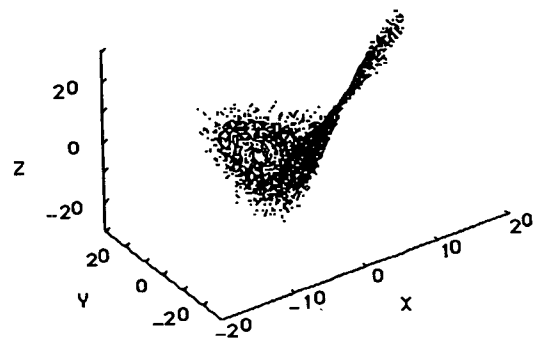


FIG. 18. Scatter diagram of Lorenz time series representing entire attractor. Sampling = 0.1 tu.

Although most climatic processes appear to be non-Markovian and are certainly nonlinear, the length of a measured time series is often too short to justify any model more complicated than a linear Markov process. When rainfall data is modeled by Markov chains, the order of which is determined by maximum-likelihood techniques, it is well known that the order so determined tends to increase with the length of the measured time series (e.g., Eidsvik 1980) and that the length of the time series needed to unambiguously identify a higher-order process increases with the number of states in the multivariate process. For the short (10^3 samples, say), highly multivariate time series often found in meteorology, we do not expect more complicated models to significantly improve the predictions made by a linear Markov process unless there is independent knowledge of the physical processes involved.

Nevertheless, we suggest two tests for the validity of the model. For a truly linear Markov process, the calculated matrices **B** and **Q** should be independent of lag. The analysis should be performed at different lags and the results compared with each other. The second test involves comparison of the relative discrepancy between predictions and observations using the data from which the parameters were obtained [Eq. (29)] with that expected for a linear Markov process [Eq. (30)]. Good agreement at unit lag is expected, since that is what is used to get the parameters in the first place. The agreement will break down as nonlinearities manifest themselves.

Acknowledgments. The author would like to thank H. von Storch, K. Hasselmann and I. Fisher-Bruns for useful discussions. Michael Ghil and a referee provided valuable comments on the manuscript. Much of this work was performed while C. Penland was employed at the Max-Planck-Institut für Meteorologie in Hamburg, FRG, and she gratefully acknowledges this support. Completion of the work was supported by NSF Grant ATM88-08797.

REFERENCES

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

Barnett, T. P., 1983: Interaction of the monsoon and trade wind system at interannual time scales. Part 1: The equatorial zone. *Mon. Wea. Rev.*, **111**, 756–773.

——, 1985: Three-dimensional structure of low-frequency pressure variations in the tropical atmosphere. *J. Atmos. Sci.*, **42**, 2798–2803.

Broomhead, D. S., and G. P. King, 1986: Extracting qualitative dynamics from experimental data. *Physica*, **20D**, 217–236.

Eidsvik, K. J., 1980: Identification of models for some time series of atmospheric origin with Akaike's information criterion. *J. Appl. Meteor.*, **19**, 357–369.

Hasselmann, K., 1976: Stochastic climate models. Part 1: Theory. *Tellus*, **28**, 473–485.

——, 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**, 11 105–11 021.

Horel, J. D., 1984: Complex principal component analysis: Theory and examples. *J. Climate Appl. Meteor.*, **23**, 1660–1673.

Hotelling, H. 1933: Analysis of a complex of statistical variables into principal components. *J. Educ. Psych.*, **24**, 417–520.

Kutzbach, J. E., 1967: Empirical eigenvectors of sea-level pressure surface temperature and precipitation complexes over North America. *J. Appl. Meteor.*, **6**, 791–802.

Lorenz, E. N., 1956: Empirical orthogonal functions and statistical weather prediction. Sci. Rep. No. 1, Statistical Forecasting Project, Dept. of Meteorology, M.I.T., Cambridge, MA, Contract AF19(604)-1566, 49 pp.

North, G. R., T. L. Bell, R. F. Cahalan and F. J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.*, **110**, 699–706.

Press, W., B. Flannery, S. Teukolsky and W. Vetterling, 1986: *Numerical Recipes, The Art of Scientific Computing.* Cambridge University Press, 818 pp.

Rasmussen, E. M., P. A. Arkin and W.-Y. Chen, 1981: Biennial variations in surface temperature over the United States as revealed by singular decomposition. *Mon. Wea. Rev.*, **109**, 587–598.

Riskin, H., 1984: *The Fokker–Planck Equation.* Springer-Verlag, 454 pp.

Rümelin, W., 1982: Numerical treatment of stochastic differential equations. *SIAM J. Numer. Anal.*, **19**, 604–613.

Thomas, J. B., 1969: *An Introduction to Statistical Communications Theory.* Wiley & Sons, 663 pp.

Vautard, R., and M. Ghil, 1989: Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. *Physica D* (in press).

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**, 11 020–11 036.

Weare, B. C., and J. S. Nasstrom, 1982: Examples of extended empirical orthogonal function analysis. *Mon. Wea. Rev.*, **110**, 481–487.

Wilkinson, L., 1988: *SYGRAPH.* SYSTAT, Inc. 922 pp.

Wilks, S., 1962: *Mathematical Statistics.* Wiley & Sons, 644 pp.