

## SINGULAR SPECTRUM ANALYSIS IN NONLINEAR DYNAMICS, WITH APPLICATIONS TO PALEOCLIMATIC TIME SERIES

R. VAUTARD\* and M. GHIL

*Climate Dynamics Laboratory, Department of Atmospheric Sciences and Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA 90024-1565, USA*

Received 18 June 1988

Revised manuscript received 13 December 1988

Communicated by J.E. Marsden

We distinguish between two dimensions of a dynamical system given by experimental time series. *Statistical* dimension gives a theoretical upper bound for the minimal number of degrees of freedom required to describe the attractor up to the accuracy of the data, taking into account sampling and noise problems. The *dynamical* dimension is the intrinsic dimension of the attractor and does not depend on the quality of the data.

Singular Spectrum Analysis (SSA) provides estimates of the statistical dimension. SSA also describes the main physical phenomena reflected by the data. It gives adaptive spectral filters associated with the dominant oscillations of the system, and clarifies the noise characteristics of the data.

We apply SSA to four paleoclimatic records. The principal climatic oscillations, and the regime changes in their amplitude are detected. About 10 degrees of freedom are statistically significant in the data. Large noise and insufficient sample length do not allow reliable estimates of the dynamical dimension.

### 1. Introduction and motivation

#### 1.1. *Deterministic chaos, stochastic processes, and dimension*

The classical view, going back to Boltzmann, Gibbs and Einstein, that irregular behavior in nature results of necessity from the interaction of a large number of degrees of freedom (*d-o-f*) has been shaken recently by mathematicians [1], meteorologists [2] and physicists [3]. According to the new dispensation, as few as three *d-o-f* suffice, by interacting nonlinearly, to create “deterministic chaos”. Chaos in time series, and its phase-space manifestation, “strange attractors”, are immensely attractive (pun intended) because they hold the promise of much deeper understanding of natural irregularity than heretofore possible through stochastic modeling.

At the same time, nobody doubts that many *d-o-f* are still present in nature, and affect any given phenomenon. The right question therefore is not whether a given time series is of purely deterministic or purely stochastic origin, i.e. whether very few or very many *d-o-f* have interacted to produce it. It is rather, how much and what part of the observed variability is due to a few *d-o-f* and what part is due to the infinite rest. The former part can presumably be modeled deterministically, and analyzed therewith rather completely; the latter part has still to be relinquished to the obscure kingdom of means and variances.

The outlook described in the previous paragraph corresponds to the way science always had, and still has to proceed: by isolating those parts of the system under study of greatest interest, and those processes most important in producing the phenomenon to be explained. The only difference is that dynamical systems theory seemed to hold out the Grail of instant, automatic determination

\*Present address: Laboratoire de Météorologie Dynamique, ENS/CNRS, 75231 Paris Cedex 05, France.

of something like the number of *d-o-f*, more precisely the phase-space dimension(s) of the system's variability [4–6]. The fly in the Grail's ointment is that the various algorithms proposed to determine this dimension of the “deterministic” part work well on some simple examples, but have trouble coping with the inescapable “stochastic” part of any natural time series.

In fact, such algorithms work quite well when applied to a few mathematical examples of strange attractors. More and more serious difficulties are encountered as one proceeds to apply them first to numerical solutions of differential equations, ordinary and partial, next to controlled laboratory experiments, and finally to actual observations of natural systems [6]. The obvious reason for the increase in difficulty is that the time series to which the algorithms have to be applied become shorter and noisier in this progression.

The case of large-scale natural systems is thus the most difficult, but in a sense also the most interesting one. A noteworthy attempt to investigate the question of dimension in the definitely interesting and certainly difficult case of paleoclimatic records is that of Nicolis and Nicolis [7]. They applied the Grassberger–Procaccia [8] algorithm for determining the correlation dimension to one of the better-documented marine-sediment cores recording proxy data of Quaternary climate changes, i.e., the waxing and waning of continental ice sheets during the last two million years. For the time series obtained from this core by Shackleton and Opdyke [9], Nicolis and Nicolis reported a correlation dimension of approximately 3.1. Grassberger [10] analyzed the same record, and showed that no finite upper bound for the dimension is apparent from the data.

Both studies obtained the record from the same source [11], but in slightly different versions. As we shall see, paleoclimatic raw data are uniformly spaced with respect to depth in the core, rather than uniformly sampled in time. Pestiaux [11] thus used different interpolation methods for the purposes of spectral analysis: one interpolated series had about 500 data points [7], the other had 230

[10]. But both were based on the same 184 raw data points [9, 12]. The time series is therefore much too short for a reliable estimate of dimension [13] and, as discussed in the body of this article, also too noisy.

What can be determined from such short and noisy records, that is interesting for nonlinear dynamics in general, and for paleoclimatology in particular? The answer, as we shall see, is provided by *Singular Spectrum Analysis* (SSA). SSA is an ingenious application of the Karhunen–Loève expansion for random processes [14] and is widely used in signal processing [15]. It provides qualitative and quantitative information about the deterministic and stochastic parts of system behavior recorded in a time series even when the series is short and noisy.

Broomhead and King [16] (BK hereafter) recently proposed the application of SSA to the problems of dynamical systems theory. It was applied, independently, by Fraedrich [17] to another paleoclimatic marine record. In this article, we extend their treatment of SSA and refine certain aspects of its application, such as the influence of window size, sampling interval, and length of the sample on the results of SSA. We then apply the extended and refined SSA approach to a number of paleoclimatic time series, including the marine cores V28-238 and V28-239, which triggered the dimension debate mentioned above. In a companion paper [18], the implications of the present results for paleoclimatic modeling are explored in further detail.

## 1.2. Measurement noise and statistical dimension

The concept of number of *d-o-f* or of *dimensionality* considered so far may be rendered precise in different ways, from either a statistical or a dynamical point of view [4, 5]. We settle here on two distinct definitions of dimensionality that complement each other in facilitating the statistical and dynamical understanding of any measured system.

Following BK, we consider a discrete version of the time series  $Y(t)$ , consisting of measurements  $Y_k = Y(k\tau_s)$  sampled regularly in time, for  $k = 1, \dots, N$ , where  $N$  is the sample length, and  $\tau_s$  is the sampling interval. The measured values include certain contributions that are not part of the underlying dynamical system itself, such as white or colored noise for direct measurements of a natural system or round-off and truncation errors for numerical solutions of a model system.

The time series  $Y(t)$  is therefore the sum of an *intrinsic part*  $X(t)$  and an *extrinsic part*  $W(t)$ . The intrinsic part  $X(t)$  refers to the actual physical or biological system we wish to study, and may itself contain deterministic, as well as stochastic, components. The extrinsic part  $W(t)$  refers to measurement or discretization effects and is purely stochastic.

In the language of signal processing,  $X$  is the signal and  $W$  is the noise. This additional noise  $W$  could hide a part of the relevant dynamics. In other words, the right question to ask before designing models is how many variables are required to fit the data up to their measurement accuracy.

Further limits to model fitting are imposed by the sample length  $N$  and the sampling interval  $\tau_s$ . That is, the data can only be modeled in a certain range of time scales: one person's noise is another person's signal. Thus, for instance, the irregularities of weather motivated the seminal paper on deterministic chaos by Lorenz [2]. But fast weather phenomena can appear only as stochastic noise in either a model or a record of slow, paleoclimatic evolution [19, 20]. Accordingly, the number of variables required to fit the data is not an invariant of the underlying dynamical system, but is also a function of data characteristics.

The simplest way to tackle both the noise and the finite-sampling problems is to compute how many mutually-orthogonal directions of a (reconstructed) attractor are significant. We define this number, in a precise sense given in section 2, to be the *statistical dimension*  $S$  of the data set, provided the range of the time scales is determined. It is likely that certain nonlinear changes of variables

could reduce this number; hence the statistical dimension provides an upper bound for the minimum number of *d.o.f* of the measured system. We emphasize again that this dimension is not a characteristic of the system itself, as long as the data are noisy and finitely sampled.

In fact, it is shown in section 2 that the statistical dimension of a chaotic system increases as the noise level decreases. This can be explained in the following terms: when data are perfect, a nonlinear dynamical system, even of small order, requires an infinite number of variables for its approximation by a *linear* model. Statistical dimension is therefore a more useful measure of dimensionality for the noisy data encountered in practice.

BK did not use the term statistical dimension for the number they computed, and they did not emphasize the two problems above. But they did develop the basic tools to calculate the statistical dimension, which we recall briefly here. The sampled time series  $Y$  is embedded in an  $M$ -dimensional space by taking as state vectors the consecutive sequences  $Z_n = (Y_n, Y_{n+1}, \dots, Y_{n+M-1})$ , for  $n = 1$  to  $N - M + 1$ . The reconstruction of the attractor is then guaranteed by Whitney's [21] embedding lemma and its application to single time series [4, 22–24], under certain hypotheses on  $M$ ,  $\tau_s$  and the smoothness of  $Y(t)$  (see BK). The time scales of the dynamics which can be reconstructed are bounded from below by  $\tau_s$  and from above by the *window length*  $\tau_w = M\tau_s$ .

A simple way of thinking about the use of (sub)sequences from a single time series for attractor embedding purposes is to recall the (approximate) equivalence of a single higher-order ordinary differential equation (ODE) and of a system of first-order ODEs. While the actual (discrete) differentiation of  $Y(t)$  is not recommended, since it amplifies noise, this is the underlying intuitive concept.

The dynamics of the cloud of points covering the embedded attractor can be described statistically in a linear way by its principal axes. Typi-

cally, many or most directions in the extended phase space, or embedding space, are dominated by the noisy part of the signal; hence there is no loss of significant information by neglecting the variance associated with these directions. Therefore, to within the accuracy of the data, the cloud of points is contained within a subspace of the embedding space.

The principal directions of the cloud are determined by the eigenvalue problem

$$\kappa Ce = \lambda e, \quad (1.1)$$

where  $\kappa$  is a scaling factor chosen for convergence purposes (see section 2.2),  $C$  is the (symmetric) covariance matrix of the augmented time series  $Z$ , with each diagonal being a constant and equal to a lag-correlation coefficient of the original time series  $Y$ ,  $\lambda$  is an eigenvalue of  $C$  and  $e$  the associated eigenvector.

This type of analysis is usually called Principal Component (PC) Analysis and is widely used in the study of multivariate time series [25, 26], including the study of short-term variability of climatic fields [27]. Its counterpart in time series analysis is described in [14] and has been used for optimization of the signal-to-noise ratio [28]. The square roots of the eigenvalues are called the *singular values* and their set is called the *singular spectrum*. The successive singular values can be arranged in monotonically-decreasing order, and the noise level then appears in the singular spectrum as the flat “floor” at its tail, associated with the nonsignificant directions. The statistical dimension  $S$  is the number of singular values above the noise floor.

### 1.3. Deterministic chaos and dynamical dimension

We have seen that statistical dimension, as defined above, is related to the quality of the data, as well as to the system itself. It does permit, however, a minimal representation of the attractor without prior knowledge of the dynamical system.

The usual concept of fractal dimension, which we shall call by contraposition the *dynamical dimension*  $D$ , characterizes the dynamical system itself, when it can be measured reliably.

Actually there exist numerous definitions of dimension of this type; some equalities and inequalities among them are known, others are not [4, 5]. We shall make use here of only one operational definition of dynamical dimension, called the *correlation dimension*, which is easily computed [8]. Its determination relies on the fact that the correlation histogram  $K(r)$ , which gives the number of distances between pairs of points on the (finitely-sampled) attractor smaller than  $r$ , decreases as  $r^D$  for  $r$  going to zero.  $D$  therefore measures the average of the local dimensions of the foliation of the attractor.

If  $D = \infty$  the intrinsic dynamics contains, in all likelihood, a stochastic component.  $D < \infty$ , on the other hand, indicates a purely deterministic process. In the latter case, noninteger  $D$  suggests a chaotic process, possessed of a “strange” invariant set, with possibly self-similar, or *fractal* structure. For a forced-dissipative system [2], like the climatic system [29], this “strange” set would have to be an attractor.

The purpose of this article is twofold. First, we wish to explore fully the potential of SSA in studying the dynamics recorded in the data. Second, we want to verify whether a connection between the statistical dimension  $S$  and the dynamical dimension  $D$ , as claimed in [16, 17], exists.

On the positive side, we show that SSA provides an enlightening quantitative analysis of the underlying dynamics. In particular, it extracts the important components of the variability of the system, as well as its spectral properties, even when stationarity fails for the data set. On the negative side, it is shown that there is no simple connection between  $S$  and  $D$ :  $S$  depends on the quality and quantity of the data, while  $D$  should be independent of the sampling properties of the data, when they are plentiful and accurate enough. More specifically, available paleoclimatic records for the Quaternary do not permit a reliable deter-

mination of whether dynamical dimension is finite and low, or infinite. SSA, on the other hand, proves successful in describing the statistically-significant number of *d-o-f* contained in the data, as well as the dynamics of paleoclimatic oscillations [20, 30].

Section 2 is devoted to a theoretical study of the properties of SSA. In particular, the link between SSA and classical spectral analysis is investigated. We show that the statistical dimension depends crucially on the quality of the data, as well as on the sampling time scale. In section 3, SSA is applied to four marine-sediment cores. The number of *d-o-f* for each core, as well as the dynamics of the climatic oscillations, as revealed by SSA, are discussed. In section 4, estimating the dynamical dimension of the climatic attractor is attempted, in the light of SSA results. We show that much longer and cleaner data sets are necessary to estimate this quantity with any reliability. Conclusions follow in section 5, and two appendices contain some technical details.

## 2. Singular spectrum analysis

This section is devoted to a theoretical survey of the properties of SSA. Some of these properties were discussed by BK, but additional theoretical results are required in order to extract all the quantitative information this method can provide. The simplest treatment is obtained with continuous time series (section 2.1). The application to discrete, sampled data follows in section 2.2.

### 2.1. Continuous formulation

We deal here with the general case of a zero-mean, continuous, infinite and ergodic time series  $X(t)$ . The eigenvalue problem (1.1) defined in the introduction generalizes to the search of the eigenfunctions and eigenvalues of the linear, integral operator  $A$  acting on functions  $\rho(t)$  belonging to

$L^2(-\tau, \tau)$  according to

$$A\rho(t) := \frac{1}{2\tau} \int_{-\tau}^{\tau} C_X(t-s)\rho(s) ds; \quad (2.1a)$$

we use the symbol  $:=$  for defining identities, and  $C_X(t)$  is the autocorrelation function of the series  $X$  at lag  $t$ . The operator has been centered with respect to the discrete problem (1.1), so that  $\tau = \tau_w/2$ . Eigenvalues  $\lambda$  and eigenfunctions  $\rho(t)$  are solutions of

$$A\rho = \lambda\rho, \quad (2.1b)$$

where  $\lambda$  is the square of the associated singular value.

$H := L^2(-\tau, \tau)$  plays the role of the embedding space and is a Hilbert space with respect to the usual inner product

$$(f, g) = \frac{1}{2\tau} \int_{-\tau}^{\tau} f(t)g(t) dt. \quad (2.2)$$

We assume that  $C_X(t)$  does not vanish almost everywhere (a.e.) and that it belongs to  $L^2(-2\tau, 2\tau)$ . Since  $C_X(t)$  is symmetric with respect to the origin  $t=0$ ,  $A$  is a symmetric non-negative operator, i.e.  $(f, Ag) = (Af, g)$  for every  $f$  and  $g$  in  $H$ , and  $(Af, f) \geq 0$ . Under the above assumptions, Hilbert-Schmidt theory (e.g. [31]) ensures the existence of a countable set (finite or infinite) of nonnegative eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k \geq \dots \geq 0$ , and the fact that the corresponding eigenfunctions  $\rho_k$  form an orthonormal basis of  $H$ . The eigenfunctions of  $A$  are called empirical orthogonal functions (EOFs).

Each segment of length  $2\tau$  centered at time  $t$  of the series  $X$  is, by assumption, a function in  $H$  that may be expanded with respect to the EOF basis. More precisely, one may write for each  $t$  and  $s$  such that  $|s| \leq \tau$ , provided  $\lambda_k > 0$  for all  $k$ ,

$$X(t+s) = \sum_{k=1}^{\infty} X_k(t)\rho_k(s). \quad (2.3)$$

The projection coefficient  $X_k(t)$  onto the  $k$ th EOF,

$$X_k(t) = \frac{1}{2\tau} \int_{-\tau}^{\tau} X(t+s) \rho_k(s) ds, \quad (2.4)$$

constitutes a zero-mean time series which is called the  $k$ th principal component (PC).  $X_k(t)$  has variance  $\lambda_k$ , and it represents a filtered version of  $X$ , the filter being the moving average weighted by  $\rho_k$ . The PCs are uncorrelated with each other at zero lag, but their cross-covariances are not vanishing for lags different from zero.

Every averaging process involving a weight function which vanishes off the interval  $(-\tau, \tau)$  may be expanded as an infinite linear combination of the  $\rho_k$ 's. Therefore, the filter that retains most information is  $\rho_1(t)$ . According to information theory, the minimal number of *d-o-f*,  $n$ , to represent a given fraction of the total variance is given by the  $n$  first EOFs whose eigenvalues add up to the given fraction. In our case, the eigenfunctions form an orthonormal basis of filters that possesses useful spectral properties, as will be seen shortly.

From Mercer's theorem [31] applied to the Hilbert-Schmidt operator  $A$ , it follows that the autocorrelation function can be expanded as

$$C_X(t-s) = \sum_{k=1}^{\infty} \lambda_k \rho_k(t) \rho_k(s), \quad (2.5)$$

for almost every  $t$  and  $s$  in  $(-\tau, \tau)$ . Taking  $t=s$  and integrating (2.5) from  $-\tau$  to  $\tau$  gives

$$\sum_{k=1}^{\infty} \lambda_k = C_X(0), \quad (2.6)$$

where  $\sigma_X^2 := C_X(0)$  is the total variance of  $X(t)$ . Eq. (2.6) holds as long as (2.5) remains valid on the line  $s=t$ , a.e. with respect to measure on this line. If this condition is not satisfied, the sum of the eigenvalues can take on arbitrary values.

The latter situation cannot occur in a finite representation of the problem, since Lebesgue measure is replaced by a discrete one. Yet we are interested in time series given by a deterministi-

cally-chaotic signal perturbed by finite-variance "white" noise  $W(t)$  as defined in appendix A. Assuming that the deterministic part  $X(t)$  has a continuous autocorrelation function, the noise adds only a discontinuity at the origin, where  $C_{X+W}(0) = C_X(0) + C_W(0)$ . In this case, (2.5) no longer holds for the series  $X+W$  on the diagonal  $t=s$ , but it is easy to see that the eigenvalues and eigenvectors of  $A$  are not affected by the noise. The sum of the eigenvalues in this case is the variance of the deterministic part  $X$  only.

Eq. (2.5) also shows that, in general, the number of strictly positive eigenvalues is infinite, since the set of autocorrelation functions satisfying (2.5) with a finite sum in the right-hand side (RHS) is not generic. Consequently, the eigenvalues of  $A$  will generally tend to zero, and the rank of  $A$  is not finite. The most important consequence for our applications is that, in the continuous case, there exists no noise floor, as long as the noise is as defined in appendix A.

In the discrete case (see section 2.2 and appendix B), the rank of the correlation matrix  $C$  of (1.1) will not be bounded as the embedding dimension increases, not even when the time series  $X(t)$  represents a trajectory of a finite-order dynamical system. A flat tail of the singular spectrum in this case is typically the consequence of extrinsic noise being present, with a variance in excess of a given fraction of that of the intrinsic process. This indicates the considerable theoretical discrepancy between statistical dimension  $S$  and dynamical dimension  $D$ .

The symmetry of the problem implies that the eigenfunctions are either even (symmetric) or odd (antisymmetric). Their respective roles can be understood from the following example.

The autocorrelation function of a *purely-periodic* oscillation of angular frequency  $\omega$  is

$$C_X(t) = \sigma_X^2 \cos(\omega t). \quad (2.7a)$$

In this particular case, only two eigenvalues are nonzero, independently of  $\tau$ . The eigenfunctions

are

$$\rho^S(t) = \alpha^S \cos(\omega t) \quad (\text{even}), \quad (2.7b)$$

$$\rho^A(t) = \alpha^A \sin(\omega t) \quad (\text{odd}), \quad (2.7c)$$

where  $\alpha^S$  and  $\alpha^A$  are suitable constants. The corresponding eigenvalues are

$$\lambda^S = \frac{\sigma_X^2}{2} \left( 1 + \frac{\sin(2\omega\tau)}{2\omega\tau} \right), \quad (2.7d)$$

$$\lambda^A = \frac{\sigma_X^2}{2} \left( 1 - \frac{\sin(2\omega\tau)}{2\omega\tau} \right). \quad (2.7e)$$

Depending on the values of  $\tau$ , the order of  $\lambda^S$  and  $\lambda^A$  varies periodically. When  $\tau$  is a multiple of  $\pi/2\omega$ , the eigenvalues merge; this case is known as a degeneracy of the problem in terms of PC analysis. That is, every combination of  $\rho^S$  and  $\rho^A$  is also an eigenfunction. In other words, any moving average of  $X$  weighted by such a combination will have the variance  $\sigma_X^2/2$ . Hence the phase of the weight does not affect the variance of the filtered signal for these special, resonant values of the window length. Moreover, when  $\tau$  goes to infinity, the two eigenvalues also merge. This property can be extended to the case of a multiply-periodic signal, and will be of particular interest in section 3 for detecting oscillations of physical significance in the records, without using the more laborious tools of spectral analysis.

To establish a closer connection between SSA and classical spectral analysis, the eigenfunctions  $\rho_k$ , defined at first in  $H$ , have to be extended over the whole real axis, using (2.1). This can be easily done as long as  $C_X$  is integrable over each interval of length  $2\tau$ . Indeed, if  $\lambda_k$  does not vanish, the knowledge of  $\rho_k$  within  $(-\tau, \tau)$  only allows its extension, using (2.1). If we now assume that  $C_X$  is in  $L^2(\mathbb{R})$ , each eigenfunction is also square-integrable over  $\mathbb{R}$ . This assumption facilitates considerations of convergence, but is not necessary if (2.1) is considered in the sense of distributions.

The Fourier transform of (2.1b) gives

$$\lambda_k \hat{\rho}_k(f) = \frac{1}{2\tau} P_X(f) \tilde{\rho}_k(f), \quad (2.8a)$$

where  $\hat{\rho}_k$  denotes the Fourier transform of  $\rho_k$ ,

$$\hat{\rho}_k(f) = \int_{-\infty}^{+\infty} \rho_k(t) e^{2i\pi f t} dt. \quad (2.8b)$$

By the Khinchin–Wiener theorem, the Fourier transform  $P_X(f)$  of  $C_X(t)$  is the spectral density of  $X(t)$ , commonly called its power spectrum;

$$\tilde{\rho}_k(f) = \frac{1}{2\tau} \int_{-\tau}^{\tau} \rho_k(t) e^{2i\pi f t} dt \quad (2.8c)$$

is the *reduced* Fourier transform of  $\rho_k$ . The operator in (2.8c) is a frequency-band average of the total Fourier transform, since it acts as a convolution in the frequency domain with the function  $a(f) := \sin(2\pi f \tau)/2\pi f \tau$ . The equivalent bandwidth of this average is  $1/\tau$ . The ratio between  $\hat{\rho}_k$  and its band average is thus proportional to the power spectrum of the series. In particular, when the power spectrum contains no frequency gap, both  $\hat{\rho}_k$  and  $\tilde{\rho}_k$  must vanish at the same frequencies.

From (2.8) it also follows that  $\hat{\rho}_k$  vanishes whenever  $P_X$  does. If for instance  $P_X$  is a band-limited spectrum around a frequency  $f_0$  with a bandwidth of  $\Delta f \ll f_0$ , and vanishes elsewhere, the  $\hat{\rho}_k$ 's are limited to the same band as well, whatever  $\tau$  may be. The shape of the  $\hat{\rho}_k$ 's may, however, vary widely depending on  $\tau$ . When the frequency-averaging band is small,  $\tau \gg 1/f_0$ , the averaging effects do not propagate very far from the frequency band of interest:  $\tilde{\rho}_k(f)$  decays rapidly away from the neighborhood of the latter. All eigenfunctions  $\rho_k$  then oscillate on  $(-\tau, \tau)$  with a number of nodes in the range  $2(f_0 \pm \Delta f)\tau$ .

For small  $\tau$ ,  $\tau \ll 1/f_0$ , the orthogonality constraints on the eigenfunctions  $\rho_k$  will necessarily distort their approximately-sinusoidal shape near the origin. Still, it is observed that, independently of  $\tau$ , SSA is capable of separating  $\rho_k$ 's into distinct groups, each corresponding to a given peak in the power spectrum, even in the fairly general situation where these peaks rise above a background level of continuous spectrum.

The ability of SSA to separate contributions from different frequency domains in this case is justified by numerical evidence (section 3) and may be explained as follows: Let us assume that  $P_X$  has two distinct peaks.  $X$  is then the sum of two independent processes involving one frequency band each. As  $\tau$  goes to infinity, the averaging effects on  $\hat{\rho}_k$  leading to  $\tilde{\rho}_k$  become very small. Therefore each eigenfunction for each separate process approximates well a solution to the total eigenproblem (2.8). Hence the eigenfunctions separate into two distinct and nearly orthogonal sets.

In practice, effective splitting occurs even for small windows and more complex spectra. This separation property follows apparently from the orthogonality of the eigenfunctions, i.e.

$$(\rho_i, \rho_j) \equiv \frac{1}{2\tau} \int_{-\tau}^{\tau} \rho_i(t) \rho_j(t) dt = \delta_{ij}. \quad (2.9)$$

The orthogonality (2.9) must be due either to the two eigenfunctions being out-of-phase, cf. (2.7b, c), or to their involving different frequency domains. The latter situation obtains for Sturm–Liouville problems and seems to extend to a large class of stochastic eigenvalue problems.

So much for pairs of singular values and spectral peaks. Geophysical records in general and paleoclimatological ones in particular exhibit typically, besides more or less broad peaks, a large *red-noise* component [29].

The autocorrelation function for a red-noise process is

$$C_X(t) = \sigma_X^2 e^{-a|t|} \quad (2.10a)$$

and the spectral density is

$$P_X(f) = \frac{2a\sigma_X^2}{a^2 + 4\pi^2 f^2}. \quad (2.10b)$$

The spectrum decreases monotonically and no particular frequency band is excited. It can be shown, after some algebraic manipulations, that

the symmetric eigenfunctions take the form

$$\rho_k^S(t) = A_k^S \cos(2\pi f_k^S t) \quad \text{for } |t| \leq \tau, \quad (2.11a)$$

$$\rho_k^S(t) = B_k^S e^{-a|t|} \quad \text{for } |t| > \tau, \quad (2.11b)$$

where  $A_k^S$  and  $B_k^S$  are suitable constants so that  $\rho_k^S$  is continuous, and  $f_k^S$  are the (ordered) solutions of

$$2\pi f \tan(2\pi f \tau) = a, \quad (2.11c)$$

which has a solution in each interval  $((k-1)/2\tau, (k-1/2)/2\tau)$ . The associated eigenvalues are

$$\lambda_k^S = \frac{a\sigma_X^2}{\tau[a^2 + 4\pi^2(f_k^S)^2]}, \quad (2.11d)$$

thus decreasing as  $k$  increases. The antisymmetric eigenfunctions are given by

$$\rho_k^A(t) = A_k^A \sin(2\pi f_k^A t) \quad \text{for } |t| \leq \tau, \quad (2.12a)$$

$$\rho_k^A(t) = \pm B_k^A e^{-a|t|} \quad \text{for } |t| > \tau. \quad (2.12b)$$

The eigenvalues  $\lambda_k^A$  are still given by (2.11d), but the frequencies  $f_k^S$  are now replaced by the solutions  $f_k^A$  of

$$2\pi f \tan(2\pi f \tau) = -a. \quad (2.12c)$$

Consequently, the spectrum of the operator  $A$  is ordered as

$$\lambda_1^S \geq \lambda_1^A \geq \lambda_2^S \geq \lambda_2^A \geq \dots \geq \lambda_k^S \geq \lambda_k^A \geq \dots \quad (2.13)$$

The reduced Fourier transform of an even eigenfunction is

$$\tilde{\rho}_k^S(f) = C_k^S \left( \frac{\sin(2\pi\tau(f + f_k^S))}{2\pi\tau(f + f_k^S)} + \frac{\sin(2\pi\tau(f - f_k^S))}{2\pi\tau(f - f_k^S)} \right). \quad (2.14)$$



The formula for the odd ones is obtained from (2.14) through multiplication by the imaginary unit  $i$  and replacement of  $f_k^S$  by  $f_k^A$ .

The maximum of  $\tilde{\rho}_k^S$  occurs at  $\pm f_k^S$  and the bandwidth is roughly  $1/\tau$ . There is but little overlap in the frequency domain between eigenfunctions of consecutive order  $k$ . The frequencies increase with the order as expected: the rate of decrease of the singular spectrum is in fact the same as for the power spectrum.

Exact correspondence between the two types of spectra is not to be expected for more complex processes and finite  $\tau$ , but turns out to be observed in most cases. Devijver and Kittler [32] show that exact correspondence between power and singular spectra is actually achieved when  $\tau$  goes to infinity. In this limit, the singular eigenset consists of pairs of even and odd eigenfunctions corresponding to one eigenvalue. The singular spectrum is no longer countable, and each eigenvalue corresponds to a value of the power spectrum (multiplied by a scaling factor), with the eigenfunctions being pure oscillations of the associated frequency, cf. (2.7).

The variation of the singular spectrum as a function of  $\tau$  follows also from (2.11). When  $\tau$  goes to zero, the first frequency  $f_1$  behaves like  $(1/2\pi)\sqrt{a/\tau}$ , hence  $\lambda_1^S \approx \sigma_X^2/(1+a\tau)$ . In this limit, the first eigenvalue absorbs all the variance. The spectrum decreases as  $\lambda_k^S \approx a\tau\sigma_X^2/\pi^2(k-1)^2$ , for large  $k$ .

When  $\tau$  is large, on the other hand, each individual eigenvalue tends to be proportional to  $1/\tau$  and the spectrum therefore is flatter, with an enhanced tail, compared to the first limit. The behavior of the spectrum for increasing  $\tau$  is pictured in fig. 1, using  $a = \sigma_X^2 = 1$ , for  $\tau = 1, 10$ , and  $100$ . These observations will prove to be extremely useful for the choice of embedding dimension (see section 2.2).

The spectral properties of the principal components (PCs) also play an important role. Indeed, PC analysis can detect the characteristic oscillations of the input signal, as pointed out above. The power spectrum of the  $k$ th PC, computed

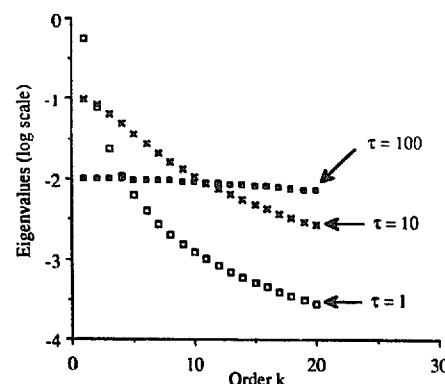


Fig. 1. The first twenty singular values squared of a red-noise process, for window length  $\tau = 1, 10$  and  $100$ .

from (2.4) is

$$P_{X_k}(f) = |\tilde{\rho}_k(f)|^2 P_X(f). \quad (2.15)$$

According to signal-processing terminology,  $\tilde{\rho}_k$  is the transfer function of the  $k$ th filter. SSA thus concentrates the transfer functions within regions where sharp peaks in the power spectrum occur, facilitating the detection of such oscillations. This data-adaptive choice of filters provides a significantly more flexible tool than standard spectral analysis. Moreover, the evolution in time of the PCs yields quantitative information about the behavior of each oscillation band in time. In particular, for short records such as those we are going to analyze, the usual stationarity hypothesis of spectral analysis is questionable and SSA helps detect discontinuities or breaks in the amplitude of oscillations (see section 3).

Moreover, the reconstruction of the time series  $X$  from the PCs is possible, cf. (2.3). In fact, only knowledge of the even eigenfunctions is required, taking  $t = 0$  in (2.3). The sum in this equation, however, is not built with elementary independent processes since the  $X_k$ 's are correlated for lags different from 0. A time series which has the same spectral characteristics as  $X$  can nevertheless be built up using a countable set of independent time series, owing to another nice property of SSA.

Let us denote by  $\Delta$  the equivalent of the  $\delta$  function for  $H = L^2(-\tau, \tau)$  instead of  $L^2(\mathbb{R})$ .

Consider the expansion of the distribution  $\Delta(x - y)$  with respect to the orthonormal filter basis

$$\Delta(x - y) = \frac{1}{2\tau} \sum_{k=1}^{\infty} \rho_k(x) \rho_k(y), \quad (2.16)$$

valid for every  $x$  and  $y$  inside  $(-\tau, \tau)$ . Taking now the reduced Fourier transform of (2.16) in both  $x$  and  $y$  yields

$$1 = \sum_{k=1}^{\infty} |\tilde{\rho}_k(f)|^2, \quad (2.17)$$

for any frequency  $f$ . It follows that the power spectrum of  $X$  is the sum of the power spectra of the  $X_k$ 's. Therefore, it is possible to consider  $X$  as a realization of a random process  $X^{(R)}(t, \tilde{\omega})$  that may be obtained as the infinite sum of the elementary processes  $X_k^{(R)}(t, \tilde{\omega})$  having the same spectrum as  $X_k$  but uncorrelated with each other at all lags. Any realization of  $X^{(R)}$  will also have the same singular spectrum as  $X$ .

In terms of information compression, SSA therefore provides a number of quantitative results, in particular spectral characteristics of the dominant phenomena to be represented, before proceeding to their actual modeling. SSA also differs in character from classical PC analysis applied to multivariate time series [27]. The former gives truly dynamical information, whilst the latter provides mostly a geometrical description.

## 2.2. Sampling and noise

Here we consider a sample of length  $N$  of successive values of the continuous time series  $X(t)$ , separated by a regular sampling interval  $\tau_s$ . Finite-variance white noise  $W(t)$  of variance  $\sigma_W^2$ , as defined in appendix A, has been added due to measurement or round-off errors. It follows that the autocorrelation function of the resulting process  $Y(t)$  has a discontinuity at the origin, in the limit  $\tau_s \rightarrow 0$ . The covariance matrix is estimated using this sample and is subject to random errors

due to the finiteness of the sample. It was found, however, that this finite-sampling error is small when compared to errors due to noise (see the end of this subsection). We assume therefore throughout this section that  $N$  is large enough to give good estimates of the covariance matrix  $C$  of the process  $Y$ . The details of the discrete formalism, paralleling eqs. (2.1)–(2.6) and (2.8), are given in appendix B.

We want the discrete estimates in embedding dimension  $M$  to converge to the “true” values of the spectrum, hence the scaling factor of the problem (1.1) is  $\kappa = 1/M$ . Indeed, it is expected that the estimates will be good as long as the approximation

$$\begin{aligned} \frac{1}{M} \sum_{n=1}^M C_X((p-n)\tau_s) \rho(n\tau_s) \\ \approx \frac{1}{\tau_w} \int_0^{\tau_w} C_X(p\tau_s - t) \rho(t) dt \end{aligned} \quad (2.18)$$

remains valid, for every  $p$ .

According to our assumptions, the covariance matrix for the sampled series is now  $C = C^{(T)} + C^{(W)}$ , where  $C^{(T)}$  is the covariance for the “true” or “intrinsic” sampled signal, and  $C^{(W)}$  is the diagonal noise matrix with elements  $\sigma_W^2$ . The estimated eigenvalues of the problem (1.1) are

$$\lambda_k = \lambda_k^{(T)} + \frac{1}{M} \sigma_W^2, \quad (2.19)$$

where  $\lambda_k^{(T)}$  now denotes an estimate for the true sampled signal, converging to an eigenvalue of the continuous signal as  $M$  goes to infinity keeping  $\tau_w$  fixed. Since  $\lambda_k^{(T)}$  does not depend on the noise, the estimates  $\{\lambda_k\}$ ,  $k = 1, \dots, n, \dots$ , turn out to be biased by the noisy component. According to (2.19), the noise floor occurs at  $\sigma_W^2/M$ , and depends therefore on the sampling time  $\tau_s$  for a fixed window length  $\tau_w$ .

It follows that the number of significant eigenvalues also depends on the sampling time, contradicting the claim of BK. Reducing the sampling time  $\tau_s$  for fixed  $\tau_w$  not only results in adding

eigenvalues to the noise floor, but also to the significant part of the spectrum, as long as the noise remains white for high frequencies. This problem is also encountered for power spectrum estimates, the level of noise decreasing with decreasing sampling time as well. In both analyses, the reduction of the sampling time increases the information about noise, as well as about the signal. Furthermore, the bias due to sampling is of order  $1/M$ , whereas the bias on covariance estimates due to finiteness of the sample is of order  $1/N$ , thus showing the minor importance of the latter when  $M \ll N$ .

### 2.3. Embedding dimension

The main problem in determining a suitable embedding dimension is that two independent parameters are involved:  $\tau_s$  and  $\tau_w$ . In practice, as pointed out in appendix A, the data should not be sampled too finely, since there is a time scale under which experimental noise is no longer uncorrelated. Once the data are suitably sampled, it is possible to take multiples of the sampling time as the basic lag for covariance matrix calculations.

We consider in the following  $\tau_s$  as fixed, and given by the lag between successive available measurements. The problem then reduces to the choice of  $\tau_w$  or equivalently  $M$ . The sampling time  $\tau_s$ , on the one hand, and the total length of the time series  $N\tau_s$ , on the other, impose obvious limits on the bandwidth of the signal one is able to analyze. The choice of  $\tau_w = M\tau_s$ , with  $1 \leq M \leq N$ , corresponds to a compromise between the amount of significant information one hopes to retain – the larger the  $M$  the better, and the statistical confidence one needs to achieve – the smaller the  $M$  the better. No optimal  $M$  exists, and one has to evaluate the stable features of the eigenset by varying  $M$  over a reasonable range.

To understand more clearly the effects of changes in window length, let us consider again the simple analytical example of red noise, eqs. (2.10). We assume that a white noise of variance  $\sigma_w^2$  is added to this signal. The significant eigen-

values are those satisfying  $\lambda_k^{S,A} > \sigma_w^2/M$ . It can be shown that their number, namely the statistical dimension, is

$$S = \frac{2M}{\pi} \left\{ a\tau_s \left( \frac{\sigma_x^2}{\sigma_w^2} - a\tau_s \right) \right\}^{1/2}. \quad (2.20)$$

When the signal-to-noise ratio is large enough,  $\sigma_x^2/\sigma_w^2 \geq a\tau_s$ , then  $S \geq 0$ . When this condition is not satisfied, the method is unable to distinguish any significant correlation between successive data and therefore the singular spectrum is roughly constant, i.e., the noise floor fills the spectrum. On the other hand, the noise floor is never reached when  $S > M > 0$ . This typically happens when  $\sigma_w^2$  is very small or  $\tau_s$  is not small enough. When  $M > S > 0$ , the singular spectrum does split into noise and signal components, and  $S$  may be evaluated from SSA.

Eq. (2.20) shows clearly the dependence of  $S$  on  $M$ ;  $S$  in this case is a linear function of  $M$  when sampling time is fixed. The simple example of red noise shows that  $S$  can change dramatically when the embedding dimension  $M$  is varied. This occurs in fact for the paleoclimatic time series analyzed in the next two sections.

Thus far we have assumed that the sample is taken from a series of measurements at regular time intervals, and that the noise is white. Yet it may happen that different levels of noise occur in different frequency ranges. Such is the case for instance when the data undergo successive filtering processes involving different spectral bands, or when measurements are not performed regularly in time.

The latter case applies to paleoclimatic records, since the measurements are taken uniformly with sediment depth in the core. Because the sampling uniformity in time is required for both spectral analysis and SSA, while sedimentation rates have varied greatly throughout the Quaternary, the data are interpolated. The most severe effect of this processing is that the noisy component of the signal is also interpolated, violating the hypothesis of whiteness.

On different time scales, the noise amplitude in the records shows a large amount of intermittency. A spectral consequence is the presence of several noise bands. Again, SSA is able to detect such inhomogeneities, and the analysis of the PCs will provide considerable improvement with respect to pure spectral analysis. Indeed, when the noise bands are relatively narrow, it is impossible to decide from spectral analysis whether they correspond to significant power peaks or not, while the visual inspection of the corresponding PCs will give a definite answer. When regular oscillations are observed we can conclude that the band is significant, whereas when the PCs display strong intermittency the band must result from noise interpolation.

More generally, both SSA (eqs. (2.3)–(2.5)) and spectral analysis *assume* stationarity. The latter approach deals with nonstationarity in practice via *evolutive spectral analysis* [11], which requires the separate investigation of many overlapping short time series. SSA, on the other hand, provides the required information through a straightforward evaluation of the PCs, based on the same eigenset for the entire time series. The price to pay is that the peaks of the singular spectrum are not as sharp as in some advanced spectral methods (maximum entropy [11] or multi-taper [56]).

### 3. The singular spectra of paleoclimatic records

We investigate here four records obtained from deep-sea cores, RC11-120, V19-240, V28-238 and V28-239. The convention used to identify marine cores is a letter acronym to designate the ship which retrieved them, followed by two digits to designate the cruise number, followed by a hyphen and the three-digit station number. The variable to be analyzed is the ratio  $\delta^{18}\text{O}$  of the rare  $^{18}\text{O}$  to the common  $^{16}\text{O}$  isotope, as found in the calcareous shells of foraminiferal species deposited over time in the bottom sediments.

To first order,  $\delta^{18}\text{O}$  is a proxy indicator of global ice volume [9, 29, 33]. Isotopically-lighter

oxygen, evaporated from the tropical and midlatitude oceans, gets trapped in continental ice sheets, instead of returning as rain or river run-off to the oceans. This makes the foraminiferal shells isotopically heavier during episodes of increased glaciation. Local temperature effects [34] and lack of isotopic equilibrium between the foraminifera and the surrounding water also affect the shells'  $\delta^{18}\text{O}$  at deposition. Post-depositional effects, such as bioturbation, act in turn to complicate the record [11] and increase the measurement noise.

Data are sampled at regular depth intervals, which correspond to unequal time intervals. To obtain uniform sampling in time, various techniques to date the sediments are used, followed by interpolation of the data to the new time scale [11]. Total geometric length of the four cores is limited by the coring technique used to obtain them [35].

The first two cores, RC11-120 [36] and V19-240 [37] (C1 and C2 hereafter), are characterized by a relatively high sedimentation rate, and cover therefore a rather small fraction of the Quaternary, but allow for the investigation of relatively high frequencies. Since there are a number of events during the time interval covered by these two cores which can be dated with reasonable accuracy, the conversion from depth scale to time scale is rather successful.

The other two cores, V28-238 (C3: [9]) and V28-239 (C4: [33]), are among the cores covering the longest time interval, being characterized by low sedimentation rates, and are thus not adequate for the analysis of high frequencies. The four records are sampled at intervals which achieve a reasonable compromise between the opposite dangers of oversampling, i.e. of encountering correlated noise, and of information loss:  $\tau_s = 1$  ka for C1 and C2, and  $\tau_s = 5$  ka for C3 and C4, where 1 ka = 1000 years. These cores are described in detail and analyzed spectrally by various methods in [11]. The sedimentation rate undergoes dramatic changes for C1 and C3, and is more homogeneous for C2 and C4. The data are displayed in fig. 2, after subtracting the mean of

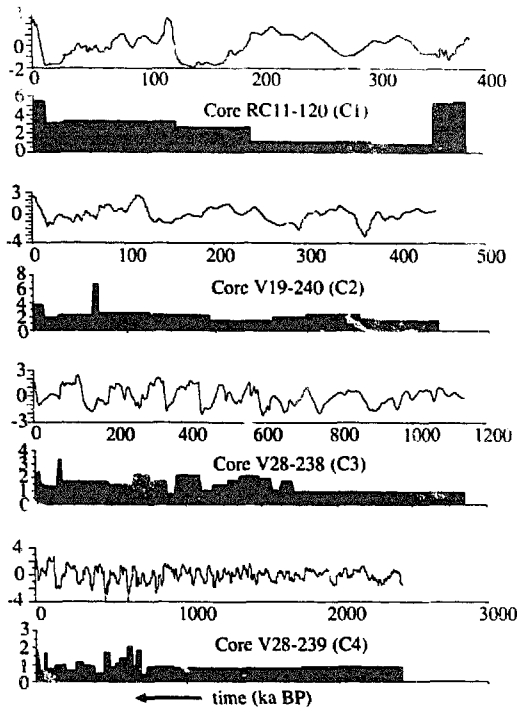


Fig. 2. Four paleoclimatic time series, showing  $\delta^{18}\text{O}$ , a proxy indicator of global ice volume, as a function of time; geological convention is to show time increasing into the past, in ka BP (thousands of years before present). For each time series, the mean has been subtracted and the values normalized by the standard deviation. Estimated sedimentation rates are given as solid bars underneath each core, in cm/ka (after Pestiaux [11]).

each record and normalizing it. Also shown in fig. 2 is the estimated sedimentation rate as a function of time, according to [11].

We do not expect the singular spectra of the four cores to have the same shape, since they involve different time scales, i.e. different frequency bands. The four cores come from different locations in the world ocean, and are thus influenced by different local temperatures [34] and other environmental conditions. Even if the measured indicator  $\delta^{18}\text{O}$  were a function of global ice volume only, the ambient water  $\delta^{18}\text{O}$  undergoes nonlinear biological transformations different in each case. Hence the  $\delta^{18}\text{O}$  of each species, different from core to core, does not provide the same “image” of the paleoclimatic attractor, if any. We thus anticipate that the statistical dimension, and other properties, will differ in each case. Therefore each record will be examined separately.

### 3.1. C1: RC11-120

This record has been chosen because it illustrates very well the impact of compression and stretching in the time scale. As shown in fig. 2, the sedimentation rate varies by a factor of 5, and two long homogeneous periods with significantly different rates are detected. The singular spectra for  $M = 10, 20$  and  $40$  are displayed in fig. 3; the corresponding window lengths are  $\tau_w = 10$  ka,  $20$  ka and  $40$  ka. Also shown in fig. 3 is the singular spectrum obtained when taking an embedding dimension of  $M = 10$  and doubling  $\tau_s$ .

The values in the latter case match closely the first ten singular values of the spectrum in the case  $M = 20$ ,  $\tau_s = 1$  ka, indicating that one is close to convergence towards the continuous problem, with  $\tau_s \rightarrow 0$  at  $\tau_w = 2\tau$  fixed, in this spectral range. The bias induced by the noise is different in the two cases, cf. eq. (2.19). For  $M = 10$ ,  $\tau_s = 2$  ka, the estimates of the singular spectrum are indeed slightly larger than for  $M = 20$ ,  $\tau_s = 1$  ka. This upward shift appears in the figure only for the last few, smallest singular values, because of the logarithmic scale.

As anticipated, the spectra do not show the simple structure proposed by BK, which obtains only when the noise is truly white. Several plateaus occur in each spectrum. The significance of this

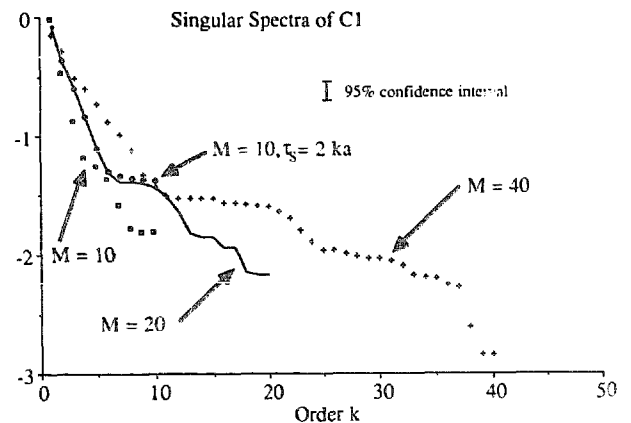


Fig. 3. Singular values of time series C1, for different values of the embedding dimension  $M$ ; also shown is the 95% confidence interval obtained for  $M = 20$ .

behavior may be questionable as the number of data, 388 points for this core, is rather limited. To check the statistical significance of the plateaus, we also show the 95% confidence interval that would be obtained for a normal process with the same variance. The error interval associated with each eigenvalue  $\lambda$ , which estimates a variance, is given by  $\delta\lambda = 2\lambda(2/N)^{1/2}$ , where  $N$  is the sample length. This error estimate does not take into account the fact that the data are correlated, nor the numerical error in diagonalizing the matrix.

The tight error bars show that the fluctuations in the spectrum for  $M=20$  do not arise from statistical errors made in estimating the covariance matrix. The sharp downward jump observed for  $M=40$  in the last two singular values is due to computational difficulties in the diagonalization process, which do not occur for  $M \leq 30$ . The three spectra show a similar shape, with an exponential decrease up to  $k=6$  (for  $M=20$ ), followed by a plateau sloping down to a second plateau, and a third plateau which stands out clearly for  $M=20$ . The floors occur at about the same levels, only slightly higher for smaller values of  $M$ . Comparison between the EOFs obtained for each value of  $M$  and  $\tau_s$  (not shown) demonstrates that the plateaus, which are shifted to higher  $k$ -values as  $M$  increases, correspond to the same dynamic phenomena.

Fig. 4 shows the principal components #1, #2, #3, #4, #5, #8 and #14 for  $M=20$ . The local excitation of the first four PCs near 120 ka BP (before present) is generated by the sharp transition from the maximum glacial stage 6, peaking at roughly 130 ka BP, to the interglacial stage 5e, peaking near 110 ka BP [34, 38]. A similar local excitation is visible in all the PCs at the top of the core. It is hard to tell whether this is due to the similarly sharp transition from the last glacial maximum (LGM), peaking at 18 ka BP (isotopic stage 2: [9, 29, 34]), to the warmer Holocene (stage 1), in which we presently live, or to mechanical coring effects. The last few thousand years of paleoclimatic history are more often missing than not from deep-sea cores due to the latter effects.

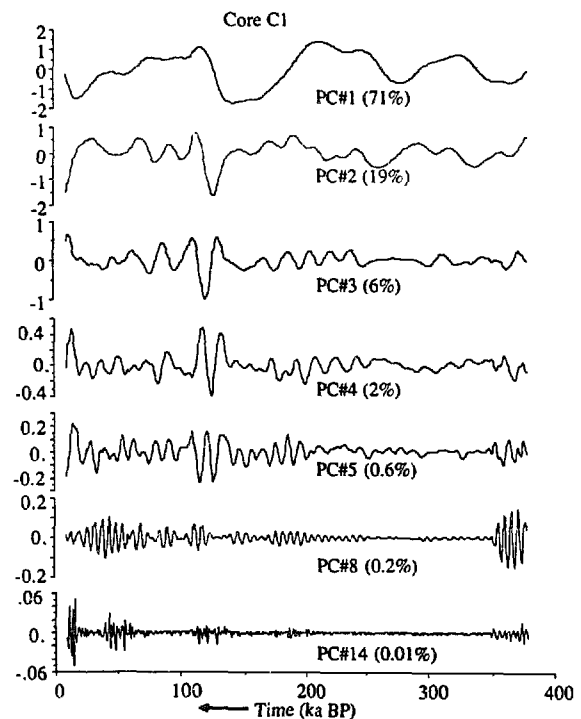


Fig. 4. Time evolution of the principal components (PCs) #1, #2, #3, #4, #5, #8, #14, for C1;  $M=20$ .

Aside from the excitation due to the stage 6-to-stage 5e transition at 120 ka BP, the first four PCs display a rather homogeneous behavior, while the fifth one suggests changes in the regime of the oscillations at 200 ka BP and 350 ka BP. As a matter of fact, both regime changes occur at the same time as the changes in sedimentation rate.

PC#8, as well as the other PCs of the first plateau (not shown) indicate that the time interval between 200 and 350 ka BP, associated with a uniformly low sedimentation rate, contains only small amplitude variations. The variance of this group of PCs is due to the variations during the latest 200 ka, and to large-amplitude oscillations at the beginning of the record, when the sedimentation rate was also high.

The interaction between noise and local stretching is now clear, and the first plateau is associated obviously with a noise floor. The same interpretation holds for the second and third plateaus (PC#14), characterized by an enhancement of the amplitude of the noise over the latest 10 ka and the first 40 ka, corresponding to rapid-sedimenta-

tion intervals. The plateaus are also found in the power spectrum of this series [11] and their explanation is now clear.

The local stretching of the time scale thus turns the true, uniform noise floor into a number of distinct plateaus associated with different frequency ranges. The inhomogeneity exhibited by PC#5 across 200 ka BP indicates that the PCs start to be contaminated by noise at this order. It is reasonable therefore to set the statistical dimension  $S$  equal to 5 in this case, while  $S = 3$  for  $M = 10$  and  $S = 10$  for  $M = 40$ .  $S$  seems to behave quite linearly with  $M$ , as in the theoretical case of red noise (eq. (2.20)). The statistical dimension  $S$  as a function of  $M$  is given in table I, for this and the other three cores.

Looking in greater detail at the significant PCs only, one can easily see that PC#1 corresponds to the dominant periodicity of Quaternary glaciations, which is nominally 100 ka. In fact the period here, and also from spectral analysis [11], is

closer to 130 ka, and the PC displays a somewhat irregular behavior. The other significant PCs are associated with other types of oscillations, with decreasing periods as the order increases. Yet they all exhibit an intermittent behavior.

PC#3 displays an oscillatory behavior between 250 ka BP and 50 ka BP, with a period in the range 20–25 ka, known to be associated with variations of the precessional parameter of the Earth's rotation around the Sun [36, 39], as shown also by spectral analysis. The change in behavior around 250 ka BP could be real, since sedimentation rate is constant across this epoch, but the frequencies may still be affected by a bad matching of the time scales.

On the other hand, the obliquity-related period [39] near 40 ka, though present in the power spectrum [36], does not appear clearly here. Some traces of this periodicity are apparent in PC#1 and PC#2. PC#2 shows large variations in the period of the oscillations. The main periodicities present in the PCs of this core and the other three are shown in table II.

PC#4 and #5 are associated with periods in the range 15–20 ka and around 10 ka respectively. Spectral peaks to the right of the orbital frequencies associated with obliquity and precession are considered to be high frequencies in Quaternary studies. Their existence has been predicted by the nonlinear deterministic climate model of Ghil and Le Treut [40]. This model is based on coupling the

Table I

Statistical dimension  $S$  as a function of embedding dimension  $M$  for the four Quaternary records under study.

$M$	$S$			
	C1	C2	C3	C4
10	3	3	4	5
20	5	4	8	9
40	10	7	11	18

Table II

Oscillatory behavior revealed by the principal components (PCs) of SSA in four paleoclimatic time series. The periods in the table heading correspond to predictions of the descriptive astronomical hypothesis (A: [36, 39]) and of three nonlinear, deterministic models (B: [49], G: [30, 40, 42], and S: [47]).

Period	Low frequencies			Orbital frequencies (A, B, G, S)		High frequencies
	109 ka	100 ka	50–70 ka	41 ka	19–23 ka	14.7 ka, 13 ka, 10.4 ka
Prediction	G	S	B	Obliquity	Precession	G
C1	130 ka (PC #1, irregular)			40 ka (PC #2, variable)	20–25 ka (PC #3)	15–20 ka (PC #4) 10 ka (PC #5)
C2	100 ka (PC #1)			30–40 ka (PC #2)		10–15 ka (PC #4)
C3	103 ka (PC #1, #2)			40–50 ka (PC #3)	20–30 ka (PC #4)	
C4	100 ka	65 ka (PC #3, PC #1, #2)	also 300–500 ka	40–45 ka (#9–10)		

energy balance of the atmosphere–ocean system with the mass balance of continental ice sheets, and was analyzed in considerable detail by Ghil and Tavantzis [41], Le Treut and Ghil [42], Ghil [30], and Le Treut et al. [43]. It also exhibits chaotic behavior, with the 100 ka peak, the orbital peaks, and the high-frequency peaks superimposed on a broad-band continuous spectrum. Four of the predicted high-frequency peaks, near 15 ka and 10 ka, have been confirmed by the spectral analysis of three high sedimentation-rate marine cores by Pestiaux and colleagues [11, 44]. SSA of these three “short” cores is left for the companion paper [18].

### 3.2. C2: V19-240

C2 has been chosen because it is the most homogeneous of the available intermediate-length cores. A narrow bump in the sedimentation rate occurring around 65 ka BP (see fig. 2) should not modify greatly the eigenvalue problem. Fig. 5 shows singular spectra for C2 as a function of embedding dimension  $M$ .

The spectra all indicate three distinct regions: the first part ( $1 \leq k \leq 4$  for  $M = 20$ ), with a sharp slope, dominated by the significant part of the signal, an intermediate region with a smaller slope ( $5 \leq k \leq 11$ ), and a final noise floor. The intermediate range turns out (see below) to be associated with an amplitude enhancement during the inter-

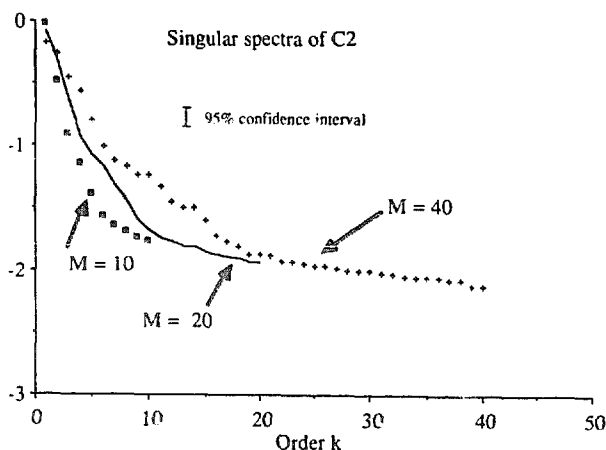


Fig. 5. Singular spectra for C2; compare fig. 3 for details.

vals of slightly higher sedimentation rate, i.e. 0–20 ka BP, 60–200 ka BP, and 300–350 ka BP, whereas the final floor exhibits some amplification during the latest 100 ka.

No intermediate plateau occurs here, indicating that the mixing between noise and signal is smooth. This is due, as expected, to the more homogeneous structure of the data. For  $k \geq 5$ , however, the effects of noise are significant. We determine therefore that  $S = 4$  for  $M = 20$ , as compared to  $S = 5$  for C1 and the same  $M$ . The approximately linear dependence of  $S$  on  $M$  holds here too (see table I).

The first four PCs are represented in fig. 6. The oscillations are now more regular. The 100 ka cycle is clearly present in the evolution of the first PC, with a small trend. PC#2 shows oscillations with periods in the range 30–40 ka. This result is in good agreement with the power spectrum [11] which displays a wide frequency band within this range. The 10–15 ka period of the oscillation in PC#4, however, is not clearly recovered by spectral analysis, due to the low ratio between its amplitude and that of the noise. Yet SSA shows this oscillation, predicted by Ghil [30], and Le Treut and Ghil [42], to be significant, as confirmed by the spectral analysis of high-resolution cores [44]. The higher resolving power of SSA for

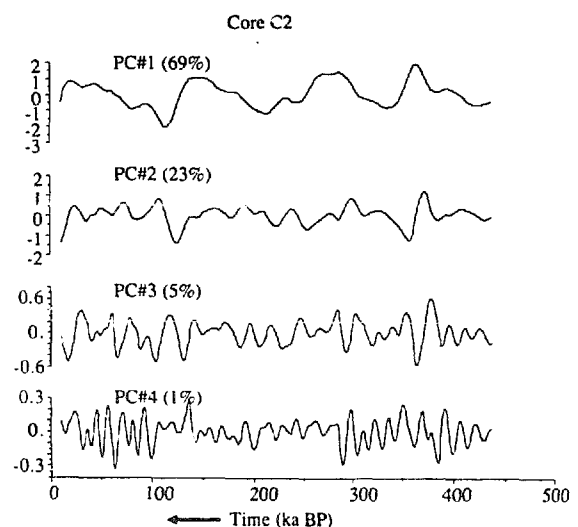


Fig. 6. Representative PCs for C2;  $M = 20$ .



the same record is a result of the optimal, adaptive nature of the filters  $\{\rho_k\}$  it uses.

### 3.3. C3: V28-238

This core has been the basis of numerous Quaternary studies [9, 29], and it was recently the object of the dimension controversy [7, 10, 12] cited in section 1.1. The application of SSA to this record shows clearly the impact of inhomogeneous time-scale stretching and, as pointed out in section 1.2, indicates the limitations of dynamical dimension estimates from such cores.

This core and the next one are sampled every 5 ka. Hence the length of the windows used are considerably larger than for C1 and C2, running from 50 ka ( $M = 10$ ) to 200 ka ( $M = 40$ ). These time scales are comparable to or larger than the periods of the main oscillations of the climatic system. We may therefore expect some singular values to be associated in pairs, as explained in section 2.1, in connection with eqs. (2.7a–e). On the other hand, as shown in fig. 1, large windows create a serious problem in properly defining the noise floor, since the entire spectrum is much flatter, and may exhibit a smooth transition to the noise floor.

The singular spectra for C3 are shown in fig. 7. Their decrease is far more uniform than for C1 or C2, and they do not seem to exhibit a noise floor at large  $k$ . Still, there is a noticeable feature com-

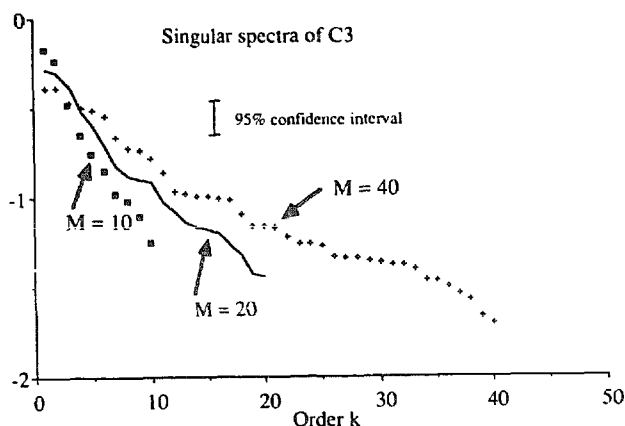


Fig. 7. Singular spectra for C3.

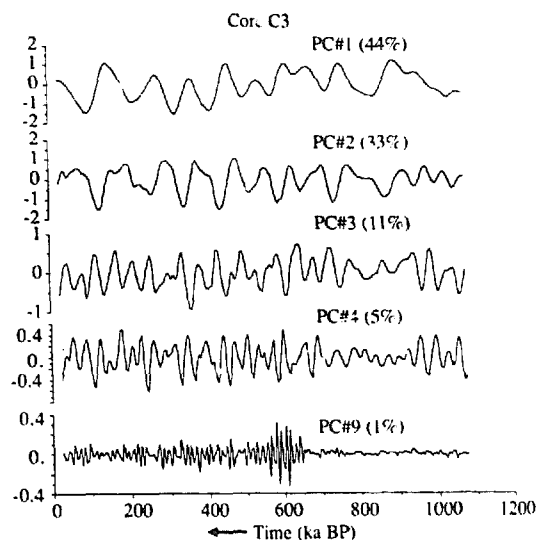


Fig. 8. Representative PCs for C3;  $M = 10$ .

mon to all spectra, which appears at  $k = 7$  for  $M = 20$ , in the form of a small step in the spectrum. This step, though, does not extend further, and it is not possible to treat the rest of the spectrum as a white-noise floor, a fact confirmed also by the statistical significance test.

For  $M = 10$ , PC#1, 2, 3, 4 and 9 are shown in fig. 8. From  $k = 5$  to 10, the PCs possess the same features as PC#9, in different frequency ranges. A clear jump of the variance is observed in these PCs near 650 ka BP, where large changes in the sedimentation rate occur (cf. fig. 2). The two artificial regimes of higher and lower variance only appear for  $k \geq 5$ ; hence  $S = 4$  for  $M = 10$ ,  $S = 7$  for  $M = 20$  and  $S = 12$  for  $M = 40$ . In the course of data interpolation to a uniform time scale, both noise and signal are seriously altered for larger  $k$ . This causes severe difficulties for dynamical dimension estimates, as shown in section 4.

Again, the first PC is associated with the 100 ka cycle. Pestiaux [11] found the peak to lie at 103 ka using a maximum entropy method, also in close agreement with the prediction of 109 ka by Le Treut and Ghil [42]. The main period involved in the second PC is comparable, but higher-frequency dynamics is superimposed on it, especially during the latest 300 ka. The phase shift

between PC#1 and #2 is almost constant during the last 600 ka, indicating that they are coupled indeed, with PC#2 leading #1. This phase relationship between the two leading PCs also holds for  $M = 20$  and 40 (not shown).

The very different behavior of this oscillatory couple in the complementary time interval could be a hint of bad estimates of the sedimentation rate, owing to the lack of absolute reference dates during the middle Pleistocene. As a matter of fact, a clear decrease of the period for the second PC is observed in the interval 1100–900 ka BP. Alternatively, this change of period may indicate that the series is not stationary in statistical terms, and an actual regime change cannot be excluded from consideration [29, 45–48].

PC#3 shows a more regular oscillation with a period in the range 40–50 ka, and PC#4 is associated with a 20–30 ka cycle, both in agreement with the power spectrum [11]. The latter two oscillations appear more strikingly using larger windows, e.g.  $M = 40$ , but they will be discussed in greater detail in connection with C4.

### 3.4. C4: V28–239

This record covers 2420 ka and extends therewith over the entire Quaternary and into the preceding Pliocene. The temporal and spectral characteristics of this record have been examined with considerable care. In particular, Pisias and Moore [45] and Pestiaux [11] have shown that climatic oscillations change in both frequency and amplitude at about 900 ka and 1450 ka BP. Again, SSA provides a detailed and convincing picture of these regime changes.

In C4, the sedimentation rate remains almost constant during the earliest 1500 ka. Its variations afterwards are more abrupt and hence affect higher frequencies than for C3, where the changes in sedimentation rate are more gradual. Here only the interval between 400 ka BP and 700 ka BP contains a significant low-frequency enhancement of the rate. Roughly speaking, C4 is more homogeneous at low frequencies than C3, where there is

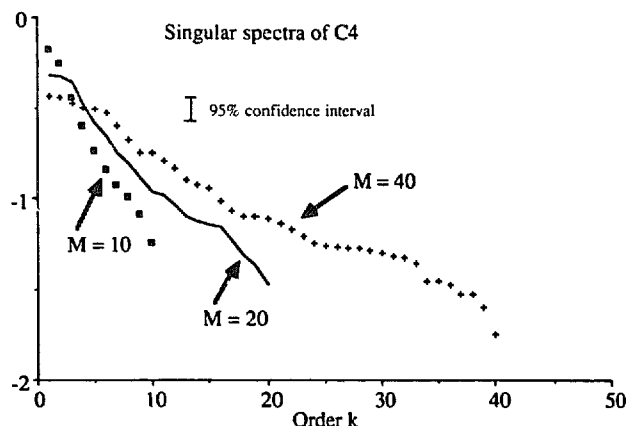


Fig. 9. Singular spectra for C4.

a contrast in mean sedimentation rates between the two halves of the core.

The singular spectra (fig. 9) still exhibit a break at their middle. A visual inspection of the PCs corresponding to the spectral tail (not shown) indicates a small local increase of the amplitude near 500 ka BP. The latter inhomogeneity occurs over a wide range of frequencies, reflecting a noisy component of the signal. This irregularity shows up from  $k = 6$  on for  $M = 10$ ,  $k = 9$  for  $M = 20$ , and  $k = 18$  for  $M = 40$ . More information is available therewith from this core than from C3, since the statistical dimension is slightly higher (cf. table I).

Fig. 10 shows PC#1, 3, 5, 6, 9 and 12 for  $M = 40$ . For  $1 \leq k \leq 6$ , each PC exhibits a different behavior. The exception to this are the couples (1, 2) and (4, 5), each PC in a pair being in almost perfect quadrature with the other. The six leading PCs represent phenomena of essentially the same variance, since there are no significant differences between their singular values (cf. fig. 9).

From  $k = 6$  to  $k = 9$ , the singular values decrease monotonically, indicating a “reddish” behavior in the corresponding frequency range (periods of 40–70 ka). The PCs in the couple (9, 10) correspond to an oscillation with a period of 40–45 ka. This orbitally-forced oscillation has much smaller variance than the rest of the singular spectrum. The spectrum decreases further up to  $k = 18$ , in a red-noise manner.

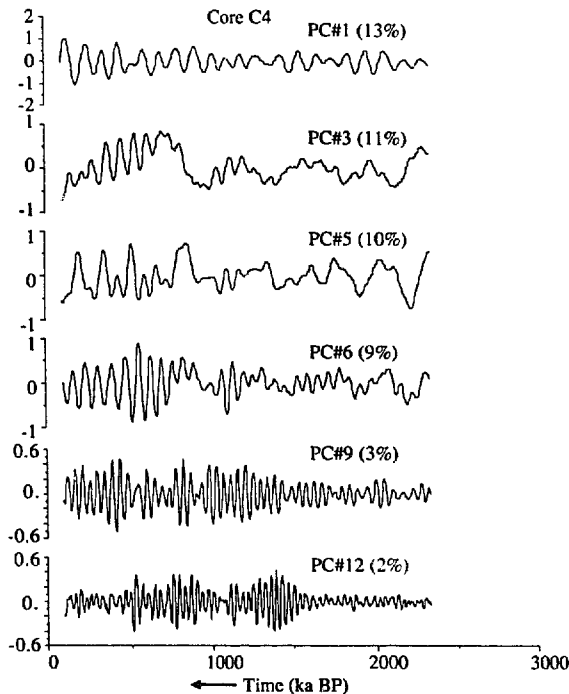


Fig. 10. Representative PCs for C4;  $M = 40$ .

Fig. 10 clarifies the climatic regime changes. The 100 ka oscillation represented in PC#1 undergoes a marked amplitude increase from the earlier to the later half of the record, and shorter oscillations with periods around 70 ka, clearly present in PC#3, 4, 5, 6, exhibit several jumps in amplitude.

PC#3 is the most illustrative of this feature. During the latest 700 ka, the late Pleistocene, the 70 ka oscillation has a regular behavior whereas it appears intermittently before. This oscillation was detected in several long paleoclimatic records by spectral analysis [11], and lies close to the free, self-sustained oscillation of the nonlinear ice-sheet model of Birchfield and Grumbine [49].

PC#3 also hints at a lower-frequency variability. It actually decomposes into the 70 ka oscillation and another one with a period in the range 300–500 ka, the latter being possibly linked with the 400 ka period of the eccentricity parameter [29, 39]. A clear break in the amplitude of these two oscillations is observed near 700 ka BP, being more pronounced for the longer cycle. Such addi-

tional changes do not appear in the 100 ka oscillation, whose variance only undergoes a transition at 1000 ka BP, approximately.

PC#5 shows the same kind of intermittency, with a larger low-frequency component before 700 ka BP, and more rapid oscillations afterwards. In PC#6, the low-frequency part seems eliminated, but still the 700 ka BP break is present. On the contrary, PC#9 does not exhibit this change, only a pronounced amplitude decrease before 1400–1500 ka BP. This modification is also present up to PC#17 (not shown). It affects periods running from 20 ka to 40 ka, and is particularly noticeable in PC#12.

SSA thus provides an excellent descriptive tool for climate regime changes. The changes occur at different dates for different frequency ranges. The present results are in good agreement with, and extend those of Pisias and Moore [45], and of the evolutive spectral analysis of Pestiaux [11]. But they are obtained at the much lower computational price of an eigenvalue problem, and appear to be more robust than the often-criticized [50] maximum-entropy method. In contradistinction from C3, neither the regime changes nor the relatively short interval of increased noise in C4 affect negatively dynamical dimension estimates, to which we proceed forthwith.

#### 4. Dynamical dimension estimates

##### 4.1. SSA and dynamical dimension

At the center of the quest for a finite and fractional dimension lies the concept of a strange attractor. In practice, such an attractor, if present, is only measured to within observational noise, and might be perturbed by intrinsic noise. Thus one wishes to draw inferences from a time series about a phase-space object which has three levels of complexity: (i) a macroscopic structure, (ii) a self-similar structure, and (iii) superimposed stochastic noise. The macroscopic structure is as-

sociated with the large-scale behavior of the smooth manifolds which make up the attractor, and is subject to distortion through linear changes of coordinates. The self-similar structure is associated with the directions of Cantor-set type behavior which are linearly independent of the smooth part [1–6]. The portion of this fine structure which is hidden by the noise depends on the latter's variance.

To reduce the variance of the noise, it might appear desirable to prefilter the time series before embedding it into an augmented-vector phase-space. Grassberger [10] showed, for instance, that the severely-filtered SpecMap time series [51] had much slower growth of correlation dimension with embedding dimension than either V28-238 or V28-239. But even the SpecMap record did not show signs of saturation of this growth for reasonable embedding dimensions,  $M \leq 12$ .

For the fine, self-similar structure to be apparent between the large scale and the noise scale, the two scales have to be separated by at least one order of magnitude. The attractor may even contain several self-similarity ranges [52]. This introduces the supplementary condition that the data set must be large enough and the noise small enough for the sample to reach the self-similarity scale (e.g., Casswell and Yorke [6], pp. 123–136). As shown by Smith [13], the latter scale is reached in the case of a fractal attractor only when the correlation histogram displays log-periodic oscillations superimposed on a straight line.

Smith's criterion leads to the stringent estimate that  $N > 42^\nu$ , where  $\nu = [D]$  is the integer part of  $D$ , in order to estimate  $D$  with an accuracy of 5% [13]. One may suspect that such extreme accuracy is not really necessary for a reasonable inference of the leading-order deterministic dynamics. Indeed, models will not reproduce it, owing to the classical problems associated with discretization and the parametrization of subgrid-scale processes. But an exponential behavior of the minimal number of points necessary for an accurate estimate of dimension is rendered plausible by just recalling that a hypercube in dimension  $M$  has  $2^M$

vertices. From the macroscopic viewpoint, the number of points required for a consistent covering of the space must thus increase at least exponentially with the embedding dimension. This simple fact can easily lead to a spurious saturation of dimension estimates when increasing  $M$ .

There are two ways in which SSA can help verify a dynamical dimension estimate. The first is simply that it is necessary, for  $D^*$  to be a valid estimate of the true dynamical dimension  $D$ , that

$$D^* \leq S(M) \leq M \quad (4.2a)$$

for

$$2D^* + 1 \leq M \leq M^*, \quad (4.2b)$$

where  $M^*$  is the value of the embedding dimension at which  $D(M)$  reaches its (apparent) saturation value  $D^*$ . For very low values of  $D^*$  and  $M$ , such as those of [7], (4.2b) would have to be replaced by

$$M_1 \leq M \leq M_2, \quad (4.2b')$$

where

$$M_1 = \min \{2D^* + 1, M^*\}, \quad (4.2c)$$

$$M_2 = \max \{2D^* + 1, M^*\}. \quad (4.2d)$$

In either case,  $D^* > S$  clearly implies one has included the noisy part of the signal into the – supposedly – intrinsic part, and that saturation of  $D(M)$  with  $M$  is due to extraneous causes, such as too small an  $N$  for the attempted  $M$  [13, 17].

The second check on  $D^*$  provided by SSA is based on the observation that any estimate of the dynamical dimension should be independent of the metric used. That is, the attractor must have invariant topological characteristics under linear transformations of the coordinates. This condition is especially important since models deal in gen-

eral with scalings different from those used for data analysis.

Due to the finite sampling, topological properties can only be approximated; but dimension estimates should stay within reasonable bounds when the cloud of points in extended phase space is stretched or shrunk. The difficulty is highlighted by the following simple example: if the variance in two coordinate directions is much larger than in the others, one may expect to find a dimension close to 2 when the attractor is undersampled, while a higher dimension is expected for the same sample if presented as a nearly isotropic cloud.

This consideration suggests using always a normalized version of the raw-data cloud, by simply dividing each PC by its standard deviation. That part of the correlation histogram for which the slope is modified by normalization belongs definitely to the macroscopic scale range. Dimension estimates within this range are meaningless. We emphasize, moreover, that the two conditions derived above, eqs. (4.2a–d) and invariance under normalization, are both only *necessary* conditions; even when they are satisfied, no rigorous conclusion on the validity of an estimate for  $D$  is possible.

Fig. 11(a) illustrates the effects of normalization on a very clean data set, arising from the numerical integration of the Lorenz [2] equations by a predictor-corrector scheme, but with a number of points comparable to that of the marine-sediment cores,  $N = 300$ . The correlation histogram (see section 4.2 for precise definition) is built up for the first coordinate  $x$  using a sampling time  $\tau_s = 10\Delta t = 0.2$ , where  $\Delta t = 0.02$  is the integration time step, and an embedding dimension  $M = 10$ . For these parameter values no noise floor is reached, due to the window length  $\tau_w = M\tau_s = 2.0$  being considerably larger than the mean period of rotation on each of the attractor's two symmetric lobes. Fig. 11(b) shows the correlation histograms obtained when perturbing the data set of fig. 11(a) with a white noise of relative variance comparable to that of the cores C3 and C4, i.e., explaining 5% of the total variance in the data.

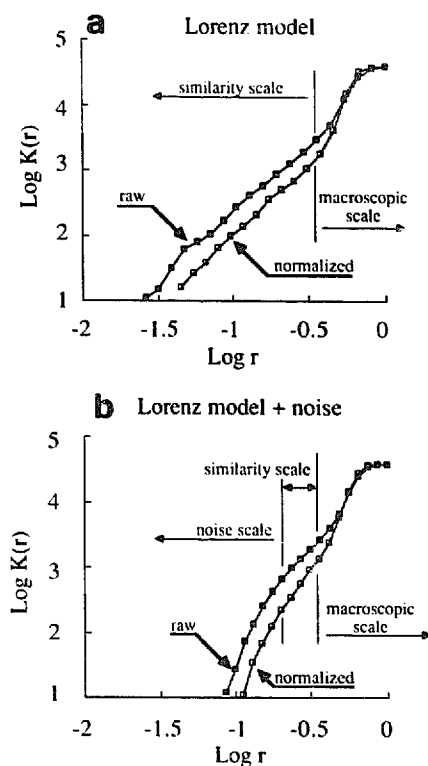


Fig. 11. Raw and normalized correlation histograms for 300 consecutive points sampling the variable  $x$  of the Lorenz [2] equations integrated numerically with a time step of 0.02 and a sampling time  $\tau_s = 0.2$ . Embedding dimension is  $M = 10$ . (a) Non-perturbed data; (b) data perturbed by white noise (see text). The macroscopic, self-similar, and noise portion of the histograms are indicated when present.

The correlation histogram displays two distinct linear regions for the non-perturbed data (fig. 11(a)) in both the raw and normalized cases. The slope corresponding to the region  $\text{log } r < -0.5$  of small point-to-point distances does not change through normalization, and equals approximately the well-known value of 2.06. The second slope, corresponding to the macroscopic view of the reconstructed attractor,  $\text{log } r \geq -0.5$ , is larger,  $D \approx 4$ , and is further increased by normalization.

For the perturbed case (fig. 11(b)), the first part of the correlation histogram ( $\text{log } r < -0.5$ ) bends down, allowing one to define the noise region,  $\text{log } r < -0.7$ . In this range, the slope increases as  $r$  decreases. Even though the noise explains but a very small fraction of the variance (5%), the self-similarity region is considerably reduced. An un-

dersampled or noisy time series from the Lorenz model might suggest therefore the existence of an attractor of dynamical dimension much larger than the correct one. This indicates clearly the possibility of misleading conclusions about dimensionality when only large distances are available for consideration [53]. It is very likely that, for more complex systems, with the same amount of data and signal-to-noise ratio, the self-similarity region totally disappears.

This example also shows that a bad estimate of the dimension in the large- $r$  domain will not provide a lower bound on the dimension either. Hence, if the paleoclimatic records C3 and C4 are under-sampled in the precise sense just described, neither the low dimension estimate found by Nicolis and Nicolis [7] from C3, nor the lower bound on the dimension claimed by Grassberger [10] from C3 and C4 can be rigorously justified.

In our case, the time series for C3 has 230 points (compare [10]) and C4 has 485 points. This number does not allow higher embedding dimensions than 15, according to tests using uncorrelated Gaussian white noise with 300 points. Dimension estimates for this test case are recapitulated in table III, showing saturation near  $M = 15$  (see also [17]). Beyond this limit, the number of points is not large enough to allow sufficiently many pairs to be as close as the algorithm requires.

#### 4.2. Dynamical dimension from cores C3 and C4

We apply the Grassberger–Procaccia algorithm [8] to the cores C3 and C4. The embedding dimension

Table III  
Dynamical dimension  $D$  as a function of embedding dimension  $M$  for a Gaussian white-noise sample of 300 points.

$M$	$D$
5	4.2
10	8.7
15	11.4
20	12.8
25	13.1

is taken to be  $M = 10$  in both cases according to the previous discussion. The algorithm is modified following the suggestions of Theiler [54] to avoid redundancy between data, so that the correlation histogram is given by

$$K(r) = \frac{2}{N(N-d)} \times \sum_{j>i+d} H\left(r - \sqrt{\sum_{k=0}^{M-1} (X_{i+k} - X_{j+k})^2}\right), \quad (4.3)$$

where  $H$  is the Heaviside function,  $H(x) = 0$  when  $x < 0$  and  $H(x) = 1$  when  $x \geq 0$ , and  $d$  is a delay for independence between pairs of points. We choose  $d = 10$ , but the results are insensitive to a choice of  $d$  between 5 and 15. Notice that it is best to take  $\tau_s$  as small as possible (cf. section 2.3), and use only pairs of uncorrelated points  $d\tau_s$  apart in (4.3). This is distinct from, and much better

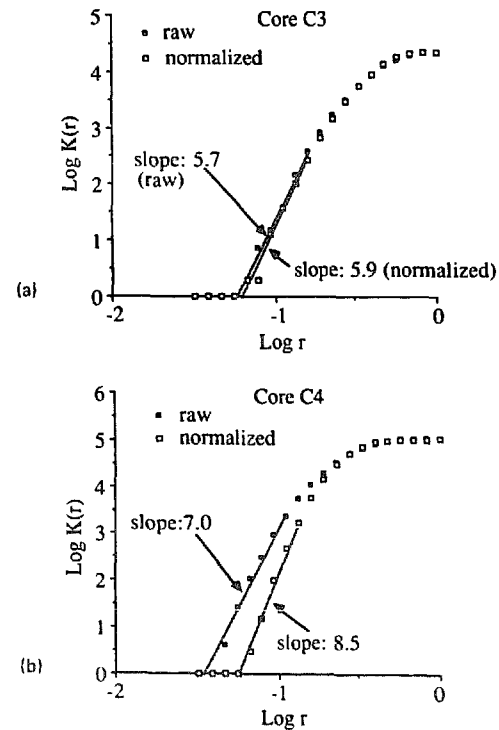


Fig. 12. Raw and normalized correlation histograms for  $M = 10$ , and the straight-line fits giving the slopes. (a) Core C3; (b) core C4. The logarithm of the actual number of pairs is shown, without the normalizing factor in eq. (4.3); the same holds for figs. 11(a, b).

Table IV

Contributions to correlation histogram (4.1) of pairs of points from different parts of the time series. Columns 2 through 7 refer to C3; see text for details of columns 7 and 8. The mutual distance intervals defining the boxes are 0.167 long. See text for definition of "Normal scores".

Box	Upper bound of the box	Number of pairs	Percentage in Category			% Weight of PC #1-4 (C3)	% Weight of PC #1-4 (C4)
			1	2	3		
1	-1.167	2	100	0	0	59	38
2	-1.000	15	73	0	27	57	47
3	-0.833	131	84	0	16	52	44
4	-0.667	833	67	3	30	61	47
5	-0.500	3551	57	5	38	65	49
6	-0.333	8176	13	18	68	46	46
7	-0.167	8327	43	45	55	39	43
8	0.000	1330	0	85	15	29	16
Normal scores (%)	-	-	16	30	54	40	40

than, using  $d\tau_s$  instead of  $\tau_s$  as the basic sampling time [10, 12].

The correlation histograms for the two cores are displayed in figs. 12(a, b). The similarity with the macroscopic portion of  $K(r)$  for the Lorenz model (figs. 11(a, b)) is striking and, as we shall see, not at all fortuitous. The distances  $r$  have been scaled by the maximal distance encountered, for curve comparison purposes, inducing only a shift of the first coordinate in the logarithmic plot. The dimension estimates are computed by fitting the curve over the six first points (as shown by the lines in fig. 12(a)). Since no significant change is observed through normalization, one could conclude that a purely-deterministic attractor of dimension  $D \approx 5.8$  is revealed by the time series. This is about twice as large as the estimate  $D^* = 3.1$  given for the same core by [7] and comparable with  $D(10) \approx 4.8$  of [10].

Table I shows that  $S(10) = 4$  for C3. Since  $S$  decreases nearly linearly with  $M$  for this core, in agreement with eq. (2.20), it is clear that  $S(M^*) \leq 3$ , where  $M^* = 4$  is the value at which  $D(M)$  appears to saturate in [7]. Thus conditions (4.2) are violated for the estimate  $D^* = 3.1$  of [7].

For  $M = 10$  here, we also get  $D(M) \approx 5.8 > 4 = S(M)$ . Why does the normalization criterion

fail to alert us to the lack of self-similarity in the data, while the comparison (4.2) does alert us to it? This is due to the subtle effect of the inhomogeneous time scale. Let us consider the normalized case. SSA has shown that the first four PCs are nearly stationary, whereas the last six show a clear break in the variance near 650 ka BP. After normalization, each PC gives the same average contribution to the variance. Hence, the pairs detected for small  $r$  correspond mostly to the earliest part of the record and are largely given by their projection onto the first four PCs. On the other hand, pairs at large distances  $r$  correspond mostly to the nonstationary, noisy components.

This idea is tested using 8 logarithmic boxes and looking at the characteristics of the pairs counted in each box. All pairs are classified into three categories:

- 1) pairs of dates occurring before 650 ka BP;
- 2) pairs of dates occurring after 650 ka BP;
- 3) mixed pairs.

For each box, the percentage of pairs in each category is computed, and we compute the "normal average" scores obtained by dividing the total number of possible pairs in each case by the total number of pairs, independently of their distance. The results are given in table IV. Also shown in

table IV is the average weight of the first four PCs in the computation of the distances. That is, we compute, for each box, the average over the pairs  $(X(t_i), X(t_j))$  with distance  $r$  within the box of the quantity  $p(r, t_i, t_j)$ :

$$p(r, t_i, t_j) = \frac{\sum_{k=1}^4 (X_k(t_i) - X_k(t_j))^2}{\sum_{k=1}^{10} (X_k(t_i) - X_k(t_j))^2}. \quad (4.4)$$

The “normal” values of  $p$  that would result from an isotropic cloud of points are 40%. It is clear from table IV that the latest part of the record accounts for large  $r$  and inhomogeneous components, whereas the earliest part accounts for small  $r$  and the first four PCs. In a schematic way, the normalized attractor is composed of two orthogonal pieces of relatively small dimension, and therefore one finds a small average dynamical dimension. The nonstationarity of the data set is so large that even raw estimates are affected by this bias, and hence no reliable estimates of dynamical dimension can be drawn from the core V28-238.

For C4, the data are more homogeneous and do not cluster into two such subsets with different characteristics. Now the contribution (4.4) of the first four PCs (see table IV) lies between 38% and 49% except for the last box (largest  $r$ ); this box is dominated by the contribution from the last PC, due to the local inhomogeneity of the sedimentation rate occurring near 500 ka BP.

Fig. 12(b) shows the correlation histograms. The dimension estimates in this case are greatly affected by normalization, changing from  $D \approx 7.0$  for the raw data to  $D \approx 8.5$ . The value for the raw data is also in reasonable agreement with Grassberger's  $D(9) \approx 6$ , and  $D$  still growing with  $M$  [10]. Furthermore,  $S(10) = 6$  for this core, so that (4.2) is also violated by  $D(10) \approx 7.0$  for the raw data. According to the previous discussion, we conclude that the self-similarity scale, if present in the dynamics, is not reached in the data: too few points are available to estimate the dynamical dimension, and the measurement noise is too large.

It follows that no reliable estimates  $D^*$  of dynamical dimension  $D$  for a paleoclimatic attractor can be obtained from the previous two cores. Lower bounds for  $D$  cannot be found either, since the extrinsic noise amplitude is too large or the sample too short. Longer data sets, covering both short and long time scales, are necessary to settle the issue in a satisfactory manner.

## 5. Concluding remarks

Detecting deterministic chaos in time series perturbed by measurement noise is a difficult enterprise. It is very tempting to reduce the problem to the computation of a single, nonlinear functional of the data, the *dynamical dimension*  $D$ . If  $D$  is finite, the underlying dynamics is deterministic. If  $D$  is also fractional, the dynamics is chaotic, while a finite integer value could indicate a multiply-periodic phenomenon, associated with a (two- or higher-dimensional) torus in phase space.

However tempting, the determination of  $D$  finite and fractional has proven rather illusive [6, 7, 10, 12]. In this article, we tried to formulate the problem in terms which are more complex, as well as more realistic. The calculation and assessment of a single functional has to be replaced by the evaluation and analysis of the linear eigenproblem (2.1) for the lagged autocorrelation operator of the time series under study.

This eigenproblem was introduced in the present context by Broomhead and King [16] and by Fraedrich [17]. We called it Singular Spectrum Analysis (SSA) and investigated its properties first for simple phenomena, such as a pure oscillation, a red-noise process, or the Lorenz [2] system. Next we applied it to four paleoclimatic records from the Quaternary. It turns out that SSA is a powerful descriptive tool for nonlinear dynamics in general and climate dynamics in particular.

To appreciate the power of SSA, one has to realize first that the information necessary for the understanding of a physical phenomenon is not



limited to an estimate of the dimension of the attractor. Now as before, modeling the phenomenon and comparing model results with the data is the high road to understanding. The role of dimensional estimates is simply to give an idea of the number  $n$  of distinct variables and of interactions between them necessary to model the phenomenon. This number of “degrees of freedom” ( $d-o-f$ ) is also a function of the accuracy of the data against which model results will be eventually compared.

First of all, SSA provides a significant estimate of this number of  $d-o-f$ , which we called *statistical dimension*  $S$ . For a dynamical system which contains more  $d-o-f$   $n$  than data points  $N$ ,  $S = N$ . Otherwise,  $S < N$ , indicating that the dynamics can be modeled, up to the accuracy of the available measurements, with fewer  $d-o-f$ ,  $n \leq S$ . Whitney’s [21] embedding lemma and the reconstruction theorems of Mañé and of Takens [24] guarantee at best, under suitable smoothness conditions not verified for strange attractors, that  $n \leq 2D + 1$  (roughly speaking). From SSA we get  $n \leq S$ .

This result follows directly from the study of the *scalar* part of the eigenproblem, the singular spectrum  $\{\lambda_k\}$  itself. The *vector* part of the eigenproblem yields further enlightening information. This information comes in two guises: (i) The eigenvectors  $\{\rho_k(s)\}$  are adaptive time-domain moving-average filters determined directly by the data; they are called empirical orthogonal functions (EOFs) and their Fourier transforms reduced to the window length  $\tau_w$  of the SSA are band-pass filters of equivalent bandwidth  $2/\tau_w$ . (ii) The expansion coefficients  $\{X_k(t)\}$  of the lagged data  $X(t+s)$  with respect to the eigenvectors  $\{\rho_k(s)\}$ ; these coefficients are called principal components (PCs) and they permit an evaluation of the stationarity in the data.

The properties of the filters and PCs provide a close connection with classical spectral analysis. Thus a pure oscillation above a red-noise background will express itself in a pair of PCs associated with the same singular value  $\lambda_k$  and in phase

quadrature with each other, even when other oscillations are present. The optimal, data-adaptive nature of the filters  $\{\rho_k\}$  permits the detection of low-variance oscillations, not apparent from spectral analysis, with high statistical confidence.

The dependence of all eigenelements on the parameters of the SSA, sampling interval  $\tau_s$ , embedding dimension  $M$  and window length  $\tau_w = M\tau_s$  has been clarified. Thus larger  $\tau_w$  leads to flatter spectra, so that the desired distinction between a strong negative slope of the singular spectrum’s dynamically-significant part and the measurement-error “noise floor” becomes more difficult.

This apparatus, when applied to the four time series of proxy data on global ice volume, shows its flexibility and incisiveness. First of all, it allows an assessment of the reliability of dynamical dimension estimates. The result here is rather negative, since problems of uneven sedimentation rates in core V28-238 (C3) and of insufficient sample length for core V28-239 (C4) permit neither a low estimate of  $D$  [7], nor a lower bound for  $D$  [10].

SSA provides two criteria for testing a dynamical dimension estimate  $D^*$ . First, for  $D^*$  to be significant, it is necessary that  $D^* < S$  (eqs. (4.2)). Second, by projecting the data onto the EOFs, the correlation histogram (4.3) of the raw data must yield the same  $D^*$  as that resulting from the data with variance-normalized components. SSA also permits the diagnosis of the reasons for failing these two statistical significance tests, such as severe nonstationarity or insufficient sample length.

On the other hand, SSA provides us with a data-dependent, but reliable estimate of statistical dimension  $S$ . This estimate follows directly from the singular spectrum in simple cases, such as the noise-perturbed Lorenz system [16], when a negative slope of  $\lambda_k$  as a function of  $k$  is followed by a flat noise floor, with  $\lambda_k$  constant as a function of  $k$ . In more complex cases typical of measured time series, examination of the spectrum itself has to be accompanied by that of the PCs.

The results of this assessment for our cores were given in table I. We recall that cores C1 and C2 cover time scales roughly between 10 ka and 200

ka, while C3 and C4 cover about 20 ka to 500 ka. The theoretical estimate (2.20) indicates that  $S$  should depend linearly on  $M$ . This dependence is confirmed rather well by the table.

The results for  $M = 20$  are most representative of the actual dynamic information in the time series, achieving a well-balanced compromise between as large an embedding dimension as possible and good statistical significance. Thus it can be said that, to within the accuracy of the data, five-to-ten  $d-o-f$  should be sufficient to represent the information contained in these cores. In fact, hundreds of marine, continental and ice cores exist in various repositories [35]. Considerable geographic detail is known therefore at different epochs in the past, such as stage 2 (LGM, 18 ka BP), as well as for the present. To represent such detail, models of great complexity, so-called general circulation models (GCMs) with  $10^5$  to  $10^6$  variables distributed over a regular grid in space, are needed. But these models cannot be used to simulate climatic phenomena on time scales longer than a few years.

Moreover, the number of cores which provide time-continuous, relatively-accurate coverage for  $10^2$  to  $10^3$  ka is rather limited. The four cores selected here are quite representative of this small subset. It is the dynamic information on *global* climate which they contain that can be represented by 5–10  $d-o-f$ .

A number of simple, deterministic models exist, which attempt to reproduce explicitly this information [20, 29, 30]. Their variables include typically global mean temperature, global ice volume, and global mean sea level. The interactions between these variables, accounted for by one model or another, include energy balance of the atmosphere–ocean system, mass balance and viscoplastic flow of the Quaternary ice sheets, isostatic rebound of the Earth's lithosphere as the ice sheets melt, changes in atmospheric composition and in ocean circulation. In addition, external forcing due to variations in insolation caused by secular changes in the Earth's orbital parameters is taken into account. The number of physically-plaus-

ible mechanisms and of associated  $d-o-f$  is comparable therewith to the estimate provided by SSA.

But this estimate is not sufficient to discriminate between the models. The additional information provided by the PCs and summarized in table II is appreciably more useful. SSA, as we saw, is related to ordinary spectral analysis, but is considerably more robust to the nonstationarity pervasive in climatic records. From table II it is clear that none of the current nonlinear climate models reproduces *all* the information in the data. Only model [49] captures the 70 ka oscillation present in C4 as PC#3. All three models, as well as the descriptive astronomical hypothesis, exhibit the orbital peaks near 40 ka and 20 ka (PCs#2, 3, 4, or 9 here). Two of the models also capture the dominant broad peak of the Quaternary glaciations near 100 ka, manifest here in PC#1 of all four cores. And model [42] alone captures the high-frequency peaks between 10 ka and 15 ka, manifest in PCs#4 and #5 of the shorter cores C1 and C2.

Thus SSA confirms the need for simple nonlinear models with a few  $d-o-f$  in order to understand the dynamic information contained in existing records of paleoclimatic change. It also highlights novel aspects of this information, which can help the modeling enterprise, as well as the verification of model results.

### Acknowledgements

M. Kimoto brought to our attention the work of Broomhead and King. P. Pestiaux provided the time series, shared with us his experience in analyzing them, and helped formulate the problems about SSA which we tried to solve in this article. D. Boggs and C. de Mol helped with the singular-value decomposition algorithm. K. Fraedrich, R.N. Hoffman, G. Mayer-Kress, D. Müller, A.P. Mullhaupt (cf. also [55]), L.A. Smith and I. van der Mersch pointed out some of the pitfalls of

dimension computations. P. Yiou carried out an independent check of our SSA computations for the four cores, and caught a few minor mistakes in the text. C. Nicolis and two anonymous referees made perceptive comments, which improved the final manuscript. To all of them, but especially to Pierre Pestiaux, our deep gratitude. Computations were carried out on an Apple Macintosh II and an IBM 4361. This work was supported by NSF grant ATM86-15424 and by NASA grant NAG-5713.

## Appendix A

### Assumptions about data noise

We can expect the noise in direct measurements to be white, i.e. to have a flat power spectrum. Theoretical white noise, however, is not a physically-realizable process since its variance is infinite. It is usual, therefore, to assume that the noise is white up to a cut-off frequency, after which the power spectrum eventually decreases

If the sampling frequency is chosen close to this cut-off, the confidence in the data is necessarily poor. It can mean, for instance, that we are close to the time scale of the measurement process itself. One has to assume therefore that the sampling interval  $\tau_s$  is much larger than this cut-off time scale at which data are spuriously correlated.

With no real loss of rigor, we assume therewith that the plateau in the noise spectrum extends to infinity, but the noise has a finite variance, leading to a vanishing power spectrum with a nonzero integral. In other words, we assume that the correlation function of the noise is zero everywhere but at the origin, where it equals the variance.

In the case of computational round-off, the errors are also uncorrelated, down to the computer's inner-clock period, and have a finite variance. Our assumption therefore seems quite reasonable as long as successive values of a solution are separated by much longer intervals.

## Appendix B

### Discrete formulation of SSA

We consider the finite, discrete, time series of length  $N$ :

$$X_i := X(i\tau_s), \quad i = 1, \dots, N. \quad (\text{B.1})$$

The average is removed and the series is normalized by its standard deviation. The autocorrelation function is estimated in the following way:

$$C_{-j} = C_j := C_x(j\tau_s) = \frac{1}{N-j} \sum_{i=1}^{N-j} X_i X_{i+j},$$

$$j = 0, \dots, N-1. \quad (\text{B.2})$$

We use the subscript  $i$  for discretized time  $t$ , and  $j$  for discretized lag  $s$ . Then the discrete form of (2.1a, b) is

$$\lambda_k \rho_{k,j} = \frac{1}{M} \sum_{j'=1}^M C_{j-j'} \rho_{k,j'},$$

$$k = 1, \dots, M, \quad j = 1, \dots, M, \quad (\text{B.3})$$

where  $\rho_{k,j}$  denotes the  $j$ th coordinate of the  $k$ th eigenvector (EOF). The scalar product used is

$$(f, g)_M = \frac{1}{M} \sum_{j=1}^M f_j g_j. \quad (\text{B.4})$$

Formulas (2.3) and (2.4) have their discrete equivalent as well: the expansion with respect to the augmented time series onto the EOF basis is

$$X_{i+j} = \sum_{k=1}^M X_{k,i} \rho_{k,j},$$

$$i = 1, \dots, N-M, \quad j = 1, \dots, M. \quad (\text{B.5})$$

It is convenient to center the PCs in time as we did for the continuous formulation; hence the epoch associated with  $X_{k,i}$  is  $(i + (M-1)/2)\tau_s$ . For consistency with eq. (B.1), one would have to make indices  $j$  and  $j'$  in (B.3)–(B.7) run from

$-(M-1)/2$  to  $+(M-1)/2$ , and  $i$  in (B.5), (B.6) from  $(M-1)/2$  to  $N-(M-1)/2$ ; then the form of (B.5) remains unchanged and the epochs would correspond exactly with the multiplication of  $i$  by  $\tau_s$ .

The exact expression of the  $k$ th PC is

$$X_{k,i} = \frac{1}{M} \sum_{j=1}^M X_{i+j} \rho_{k,j}. \quad (\text{B.6})$$

Eq. (2.5) reads, in the discrete case,

$$C_{i-j} = \sum_{k=1}^M \lambda_k \rho_{k,i} \rho_{k,j}, \quad (\text{B.7})$$

and eq. (2.6) is valid for the sum of the  $M$  eigenvalues.

When considering the spectral properties of SSA, we have to assume that the series  $X$  is periodic with a period of  $N\tau_s$ . We denote the Fourier transform of any such series  $Y_i$ ,  $i = 1, \dots, N$ , by

$$\hat{Y}_p = \sum_{m=1}^N Y_m e^{2i\pi m p/N}, \quad p = 0, \dots, N-1, \quad (\text{B.8})$$

where the index  $p$  now denotes the frequency  $p/N\tau_s$ . Then, neglecting side effects in the calculation of the autocorrelation function, the power spectrum  $PY$  of the series  $Y$ , defined as

$$(PY)_p = \frac{1}{N} |\hat{Y}_p|^2, \quad (\text{B.9})$$

is the Fourier transform of the autocorrelation function of the series  $Y$ . Eq. (2.8a) is transformed into

$$\lambda_k \hat{\rho}_{k,p} = (PX)_p \tilde{\rho}_{k,p}, \quad (\text{B.10})$$

provided that we define the reduced Fourier transform  $\tilde{\rho}_k$  of  $\rho_k$  by the analog of (2.3c),

$$\tilde{\rho}_{k,p} = \frac{1}{M} \sum_{m=1}^M \rho_{k,m} e^{2i\pi m p/N}. \quad (\text{B.11})$$

According to the above definitions, it can be shown that eqs. (2.15)–(2.17) remain valid, with obvious changes of notation. Hence the fundamental result that the sum of the power spectra of the principal components  $X_k$  is the power spectrum of the original time series  $X$  holds. So far, we estimated the power spectrum by a periodogram, which is not a statistically-robust estimate. However, equivalent formulas can be worked out for other estimates, such as the classical single-window methods of Blackman–Tukey, or the multi-taper method of Thompson [56].

## References

- [1] S. Smale, Differentiable dynamical systems, *Bull. Am. Math. Soc.* 73 (1967) 747–817.
- [2] E.N. Lorenz, Deterministic nonperiodic flow, *J. Atmos. Sci.* 20 (1963) 130–141.
- [3] D. Ruelle and F. Takens, On the nature of turbulence, *Commun. Math. Phys.* 20 (1971) 167–192.
- [4] J.-P. Eckmann and D. Ruelle, Ergodic theory of chaos and strange attractors, *Rev. Mod. Phys.* 54 (1985) 617–656.
- [5] J.D. Farmer, E. Ott and J.A. Yorke, The dimension of chaotic attractors, *Physica D* 7 (1983) 153–180.
- [6] G. Mayer-Kress, *Dimensions and Entropies in Chaotic Systems* (Springer, Berlin, 1986).
- [7] C. Nicolis and G. Nicolis, Is there a climatic attractor? *Nature* 311 (1984) 529–532.
- [8] P. Grassberger and I. Procaccia, Characterization of strange attractors, *Phys. Rev. Lett.* 50 (1983) 346–349; Measuring the strangeness of strange attractors, *Physica D* 9 (1983) 189–208.
- [9] N.J. Shackleton and N.D. Opdyke, Oxygen isotope and paleomagnetic stratigraphy of the equatorial Pacific core V28-238: oxygen isotope temperatures and ice volumes on a  $10^4$  years and  $10^5$  years scale, *Quat. Res. New York* 3 (1973) 39–55.
- [10] P. Grassberger, Do climatic attractors exist? *Nature* 323 (1986) 609–612.
- [11] P. Pestiaux, Approche spectrale en modélisation paléoclimatique, Ph.D. thesis, Université Catholique de Louvain (1984). A. Berger and P. Pestiaux, An introduction to the PACDATA bank, *Techn. Rep. No. 28*, Institute of Astronomy and Geophysics, UCL (1982).
- [12] C. Nicolis and G. Nicolis, Evidence for climatic attractors, and P. Grassberger, Grassberger replies, *Nature* 326 (1986) 523–524.
- [13] L.A. Smith, Lacunarity and Chaos in Nature, Ph.D. thesis, Columbia University (1987); Intrinsic limits on dimension calculations, *Phys. Lett. A*, in press.
- [14] K. Fukunaga, *Introduction to Statistical Pattern Recognition* (Academic Press, New York, 1970).

- [15] E.R. Pike, J.G. McWhirter, M. Bertero and C. de Mol, Generalized information theory for inverse problems in signal processing, *IEEE Proc.* 131 (1984) 660–667.
- [16] D.S. Broomhead and G.P. King, Extracting qualitative dynamics from experimental data, *Physica D* 20 (1986) 217–236.
- [17] K. Fraedrich, Estimating the dimensions of weather and climate attractors, *J. Atmos. Sci.* 43 (1986) 419–432.
- [18] M. Ghil, P. Pestiaux, R. Vautard and P. Yiou, Nonlinear variability of the climatic system, from singular and power spectra of Quaternary records, in preparation (1989).
- [19] K. Hasselmann, Stochastic climate models: Part I, Theory, *Tellus* 28 (1976) 473–485.
- [20] B. Saltzman, Paleoclimatic modeling, in: *Paleoclimatic Analysis and Modeling*, A.D. Hecht, ed. (Wiley, New York 1985), pp. 341–396.
- [21] H. Whitney, Differentiable manifolds, *Ann. Math.* 37 (1936) 645–680.
- [22] N.H. Packard, J.P. Crutchfield, D. Farmer and R.S. Shaw, Geometry from a time series, *Phys. Rev. Lett.* 45 (1980) 712–716.
- [23] J.C. Roux, A. Rossi, S. Bachelart and C. Vidal, Representation of a strange attractor from an experimental study of chemical turbulence, *Phys. Lett. A* 77 (1980) 391–393.
- [24] R. Mañé, On the dimension of the compact invariant sets of certain nonlinear maps, pp. 230–242, and F. Takens, Detecting strange attractors in turbulence, pp. 366–381, in: *Dynamical Systems and Turbulence*, D.A. Rand and L.-S. Young, eds., *Lecture Notes in Mathematics* 898 (Springer, Berlin, 1981).
- [25] J.-O. Kim and N.H. Nie, Factor analysis, in: *A Statistical Package for the Social Sciences*, D.H. Bent and C.H. Hull, eds. (McGraw-Hill, New York, 1970), pp. 208–244.
- [26] D.F. Morrison, *Multivariate Statistical Methods* (McGraw-Hill, New York, 1976).
- [27] M. Ghil, Dynamics, statistics and predictability of planetary flow regimes, in: *Irreversible Phenomena and Dynamical Systems Analysis in Geosciences*, C. Nicolis and G. Nicolis, eds. (Reidel, Dordrecht/Boston, 1987), pp. 241–283. K. Mo and M. Ghil, Cluster analysis of multiple planetary flow regimes, *J. Geophys. Res.* D 93 (1988) 10927–10952.
- [28] R. Kumaresan and D.W. Tufts, Data-adaptive principal component signal processing, *IEEE Proc. Conf. Decision and Control*, Albuquerque (1980) pp. 949–954.
- [29] M. Ghil and S. Childress, *Topics in Geophysical Fluid Dynamics: Atmospheric Dynamics, Dynamo Theory and Climate Dynamics* (Springer, New York, 1987).
- [30] M. Ghil, Theoretical climate dynamics: an introduction, in: *Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics*, M. Ghil, R. Benzi and G. Parisi, ed. (North-Holland, Amsterdam, 1985), pp. 347–402.
- [31] F. Riesz and B. Sz. Nagy, *Functional Analysis* (F. Ungar, New York, 1955).
- [32] P.A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach* (Prentice-Hall, New York, 1982).
- [33] N.J. Shackleton and N.D. Opdyke, Oxygen isotope and paleomagnetic stratigraphy of Pacific core V28-239. Late Pliocene to latest Pleistocene, *Geol. Soc. Am. Memoir* 145 (1976) 449–463.
- [34] C. Emiliani, Pleistocene temperatures, *J. Geol.* 63 (1955) 538–578.
- [35] T.J. Crowley, The geologic record of climatic change, *Rev. Geophys. Space Phys.* 21 (1983) 828–877.
- [36] J.D. Hays, J. Imbrie and N.J. Shackleton, Variations in the earth's orbit: pacemaker of the ice ages, *Science* 194 (1976) 1121–1132.
- [37] N.J. Shackleton, The oxygen isotope stratigraphic record of the late Pleistocene, *Philos. Trans. R. Soc. Lond. B* 280 (1977) 169–182.
- [38] J. Jouzel, C. Lorius, J.R. Petit, C. Genthon, N.I. Barkov, V.M. Kotlyakov and V.M. Petrov, Vostok ice core: a continuous isotopic temperature record over the last climatic cycle (160,000 years), *Nature* 329 (1987) 403–408.
- [39] A. Berger, Long-term variations of daily insolation and Quaternary climate changes, *J. Atmos. Sci.* 35 (1978) 2362–2367.
- [40] M. Ghil and H. Le Treut, A climate model with cryodynamics and geodynamics, *J. Geophys. Res.* 86 (1981) 5262–5270.
- [41] M. Ghil and J. Tavantzis, Global Hopf bifurcation in a simple climate model, *SIAM J. Appl. Math.* 43 (1983) 1019–1041.
- [42] H. Le Treut and M. Ghil, Orbital forcing, climatic interactions and glaciation cycles, *J. Geophys. Res.* 88 (1983) 5167–5190.
- [43] H. Le Treut, J. Portes, J. Jouzel and M. Ghil, Isotopic modeling of climatic oscillations: implications for a comparative study of marine and ice-core records, *J. Geophys. Res.* D 93 (1988) 9365–9383.
- [44] P. Pestiaux, I. van der Mersch, A. Berger and J.C. Duplessy, Paleoclimatic variability at frequencies ranging from 1 cycle per 10000 years to 1 cycle per 1000 years: Evidence for nonlinear behaviour of the climate system, *Climatic Change* 12 (1988) 9–37.
- [45] N.G. Pisias and T.C. Moore, The evolution of Pleistocene climate, *Earth Planet. Sci. Lett.* 52 (1981) 450–458.
- [46] M. Ghil, Climate sensitivity, energy balance models, and oscillatory climate models, *J. Geophys. Res.* 89 (1984) 1280–1284.
- [47] B. Saltzman and A. Sutera, The mid-Quaternary climatic transition as the free response of a three-variable dynamical model, *J. Atmos. Sci.* 44 (1987) 236–241.
- [48] K.A. Maasch, Statistical detection of the mid-Pleistocene transition, *Climate Dyn.* 2 (1988) 133–143.
- [49] G.E. Birchfield and R.W. Grumbine, “Slow” physics of large continental ice sheets and underlying bedrock and its relation to the Pleistocene ice ages, *J. Geophys. Res.* 90 (1985) 11294–11302.
- [50] S.M. Kay and S.L. Marple, Spectrum analysis – A modern perspective, *Proc. IEEE* 69 (1981) 1380–1419.
- [51] J. Imbrie, J.D. Hays, D.G. Martinson, A. McIntyre, A.C. Mix, J.J. Morley, N.G. Pisias, W.L. Prell and N.J. Shackleton, The orbital theory of Pleistocene climate: support from a revised chronology of the marine  $\delta^{18}\text{O}$

- record, in: *Milankovitch and Climate, Understanding the Response to Astronomical Forcing*, A. Berger et al., eds. (Reidel, Dordrecht/Boston, 1984), pp. 269–305.
- [52] U. Frisch and G. Parisi, On the singularity structure of fully developed turbulence, in: *Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics*, M. Ghil, R. Benzi and G. Parisi, eds. (North-Holland, Amsterdam, 1985), p. 347.
- [53] E.N. Lorenz, Atmospheric predictability as revealed by naturally occurring analogues, *J. Atmos. Sci.* 26 (1969) 636–646.
- [54] J. Theiler, Spurious dimensions from correlation algorithms applied to limited time-series data, *Phys. Rev.* 343 (1986) 2427–2432.
- [55] J.-P. Eckmann and D. Ruelle, Addendum: Ergodic theory of chaos and strange attractors, *Rev. Mod. Phys.* 57 (1985) 1115.
- [56] D.J. Thomson, Spectrum estimation and harmonic analysis, *Proc. IEEE* 70 (1982) 1055–1096.