

# Singular-spectrum analysis: A toolkit for short, noisy chaotic signals

Robert Vautard, Pascal Yiou<sup>1</sup> and Michael Ghil<sup>2</sup>

*Laboratoire de Météorologie Dynamique du CNRS, Ecole Normale Supérieure, F-75231 Paris Cedex 05, France*

Received 18 October 1991

Revised manuscript received 18 February 1992

Accepted 2 March 1992

Singular-spectrum analysis (SSA) is developed further, based on experience with applications to geophysical time series. It is shown that SSA provides a crude but robust approximation of strange attractors by tori, in the presence of noise. The method works well for short, noisy time series.

The lagged-covariance matrix of the processes studied is the basis of SSA. We select subsets of eigenelements and associated principal components (PCs) in order to provide (i) a noise-reduction algorithm, (ii) a detrending algorithm, and (iii) an algorithm for the identification of oscillatory components. Reconstructed components (RCs) are developed to provide optimal reconstruction of a dynamic process at precise epochs, rather than averaged over the window length of the analysis.

SSA is combined with advanced spectral-analysis methods – the maximum entropy method (MEM) and the multi-taper method (MTM) – to refine the interpretation of oscillatory behavior. A combined SSA–MEM method is also used for the prediction of selected subsets of RCs.

The entire toolkit is validated against a set of four prescribed time series generated by known processes, quasi-periodic or chaotic. It is also applied to a time series of global surface air temperatures, 130 years long, which has attracted considerable attention in the context of the global warming issue and provides a severe test for noise reduction and prediction.

## 1. Introduction

### 1.1. Motivation

The analysis of observed time series is often a prerequisite for progress in modeling and forecasting the physical system which generates them. Three cases have to be distinguished. First, when the evolution equations governing the physical system are already known, and are relatively insensitive to initial data, forecasting is

based on these equations and its accuracy depends largely on the quality of initial data. Such is the case of celestial mechanics [1], on the whole.

For other systems, the knowledge of exact evolution equations is useless for long-term forecasting purposes, even when good initial data are available. This typically happens when the dynamical system has instabilities and non-linearities that give rise to deterministic chaos, as shown by Lorenz [2]. Chaos, however, does not mean that, for large time scales, the behavior is totally irregular or random. Some macroscopic regularities, such as near periodicities, may still contribute a large part of the variability of the system. This is the second possibility: we know the equations, detailed forecasting based on

<sup>1</sup>Now at Laboratoire de Modélisation du Climat et de l'Environnement/DSM, Commissariat à l'Energie Atomique, F-91191 Gif-sur-Yvette, France.

<sup>2</sup>On sabbatical leave from the Department of Atmospheric Sciences and the Institute of Geophysics and Planetary Physics, University of California, Los Angeles, USA.

them for a long time is impossible because of the sensitivity to initial data, but there might be other ways to predict the regular part of the asymptotic behavior. Indeed, the phase-space trajectories of such a chaotic system converge generically to a *strange attractor*. Weakly unstable periodic orbits, contained in the latter, can attract trajectories intermittently, and therefore lead to spells of periodic activity. The underlying periodic orbits have to be identified by data analysis, and can help extended-range prediction.

The climatic system is such a dynamical system. The deterministic predictability limit of detailed weather is not longer than a couple of weeks [3, 4 pp. 182–190, 438–441]. On longer time scales, instabilities and nonlinearities make the atmosphere unpredictable. However, there are some near-periodicities such as the El Niño–Southern Oscillation (ENSO) cycle in the atmosphere and the oceans [5], with periods of two to five years, or the 40–50 day oscillation [6] in the tropical atmosphere. The regularity of these phenomena can make them easier to predict with empirical models [7], based on time-series analysis, than with elaborate general circulation models, based on the discretization and numerical solution of atmosphere–ocean-coupled systems of partial differential equations.

The third class includes systems with unknown evolution equations. An example is given by complex biomedical systems, such as the human brain [8]. Often, only noisy measurements of one variable from an intrinsically high-dimensional system are available, in either one of the two latter cases.

The purpose of this paper is to review the capabilities of a data-analysis method, called *singular-spectrum analysis* (SSA): SSA extracts as much reliable information as possible from short and noisy time series without using prior knowledge about the underlying physics or biology of the system; based on this information, it also provides prediction models. If only measurements of one variable are available, *single-channel*

*SSA* applies. When several variables are measured, the cross-correlations between the time series can be taken into account by using *multi-channel SSA*.

SSA is essentially a linear analysis and prediction method. Its superiority over classical spectral methods, and the sense in which it can use concepts from and be useful in nonlinear dynamics, lies in the data-adaptive character of the eigenelements it is based on. Truly nonlinear information about and high predictive skill for intrinsically low-dimensional systems requires tens of thousands of data points [9], and many more for typical systems with intermediate and high phase-space dimensions. As we shall see, SSA can provide useful physical insight and modest, but unprecedented, medium-term predictive skill starting with the few hundred data points typically available for geophysical and other natural systems.

## 1.2. Background

SSA as a data-analysis method has been used for years in digital signal processing [10, 11]. It was introduced into oceanography by Colebrook [12], and into nonlinear dynamics by Broomhead and King [13] and by Fraedrich [14]. SSA is based on principal component analysis (PCA) in the vector space of delay coordinates for a time series. Classical PCA [15] is used with multi-channel time series, and gives the principal axes of a sequence of  $M$ -dimensional vectors ( $X_i$ ,  $1 \leq i \leq N$ ), by expanding it with respect to an orthonormal basis ( $E^k$ ,  $1 \leq k \leq M$ ):

$$X_{ij} = \sum_{k=1}^M a_i^k E_j^k, \quad 1 \leq j \leq M. \quad (1.1)$$

The projection coefficients  $a_i^k$  are called the *principal components* (PCs), and the basis vectors  $E^k$  the *empirical orthogonal functions* (EOFs). The vectors  $E^k$  are the eigenvectors of the cross-covariance matrix of the sequence ( $X_i$ ). For single-channel SSA, if the scalar series values are

denoted by  $(x_i, 1 \leq i \leq N)$ , the equivalent expansion is

$$x_{i+j} = \sum_{k=1}^M a_i^k E_j^k, \quad 1 \leq j \leq M. \quad (1.2)$$

The analogy is made by augmenting the single time series  $x_i$  into the multi-variate time series  $X_i = (x_{i+1}, x_{i+2}, \dots, x_{i+M})$ . Aside from this definition, there is no formal difference between the two expansions (1.1) and (1.2).  $M$  in the latter is called the *window length*, or *embedding dimension* – and is chosen by the user – in contradistinction with classical PCA, where  $M$  is the fixed dimension of the data vectors.

The vectors  $E^k$  are the eigenvectors of the *Toeplitz matrix* of  $x$ ,  $T_x$ , that contains in column  $j$  and row  $i$  the covariance of  $x$  at lag  $i - j$ . In both situations, the eigenvalue–eigenvector decomposition of the covariance matrix (with respect to space or lag) is related to singular-value decomposition (SVD [16]) of a rectangular matrix; in the case of SSA, the trajectory matrix has the  $N - M$  augmented vectors  $X_i$  as its columns. SVD is a class of algorithms of great generality in numerical linear algebra; we prefer not to confuse matters and distinguish between it and SSA, which is a methodology for time-series analysis.

For multi-channel SSA [17–19], with original  $L$ -dimensional data vectors  $X_{l,i}$ ,  $1 \leq l \leq L$ ,  $1 \leq i \leq N$ , the expansion becomes

$$X_{l,i+j} = \sum_{k=1}^{L \times M} a_i^k E_{l,j}^k, \quad 1 \leq l \leq L, 1 \leq j \leq M. \quad (1.3)$$

Here, the state vector considered at time  $i$  is  $(X_{1,i+1}, X_{1,i+2}, \dots, X_{1,i+M}, X_{2,i+1}, \dots, X_{2,i+M}, \dots, X_{L,i+1}, \dots, X_{L,i+M})$ .  $M$  is still the window length, but now the problem is of embedding dimension  $L \times M$ . The  $k$ th basis vector is the eigenvector of the block-Toeplitz matrix  $T_x$  containing the cross-covariance coefficients of the different channels  $l$  at lags 0 to  $M - 1$ .

The three formulae (1.1)–(1.3) are all applications of the general Karhunen–Loève bi-orthogonal expansion [20], and are most often used in signal processing for information compression and signal-to-noise (S/N) ratio enhancement. Usually, the eigenvalues  $\lambda_k$  of the symmetric, nonnegative covariance matrix of the problem are sorted in descending order. The orthogonality in both time (zero cross-covariance of two different PCs at lag 0) and “space” (orthogonality of the EOFs) imply in particular that  $\lambda_k$  is the variance of the  $k$ th PC. Therefore, truncating the sum in eq. (1.1) at an order  $p < M$  reduces the information to the first  $p$  principal components, instead of the  $M$  initial components. This truncation is done, in PCA, in an optimal way: the first  $p$  principal directions describe the largest fraction of the total variance that one can obtain using a projection onto  $p$  orthogonal vectors.

Vautard and Ghil [21] (VG hereafter) showed that the SSA expansion (1.2) yields other powerful tools for time-series analysis than information compression and S/N enhancement, and applied these to a set of paleoclimatic time series. In particular, the near-equality of a pair of eigenelements is associated with periodic activity in the signal. In contradistinction from classical spectral analysis, where the basis functions are prescribed sines and cosines, SSA can easily and automatically localize in time intermittent oscillation spells. The shape of these oscillations is determined adaptively from the data, which makes SSA more flexible and better suited for the analysis of nonlinear, anharmonic oscillations.

SSA also provides a qualitative decomposition of the signal into its significant and noisy parts. Indeed, in the presence of white noise, the eigenvalue spectrum levels off after a certain order, and the PCs of higher order are dominated by noise. VG showed that the order  $S$  of this break in the eigenvalue spectrum and the capacity dimension  $D$  of the underlying dynamical system are not equal to each other, even approximately, but that SSA can help verify a

capacity estimate (see eqs. (4.2a)–(4.2d) of VG).

SSA has been applied by now to over a dozen geophysical data sets, on time scales from days to millenia, of various lengths and with spatial extents going from single channel to hundreds of grid points. Rasmusson et al. [22] showed that the irregular ENSO phenomenon in the coupled ocean–atmosphere system contains a rather regular quasi-biennial signal modulated by a lower-frequency, less regular 4–5 year oscillation. Ghil and Vautard [23] applied SSA to a 135 year long global–surface-temperature time series and found evidence of interannual and interdecadal oscillations, confirmed since then by Allen et al. [24]. Ghil and Mo [25] gave a comprehensive description of intraseasonal oscillations in the tropical and extratropical atmosphere. Multi-channel SSA applied to geopotential height data in the Northern Hemisphere extratropics reveals cycles with periods of 40–50 days and 20–25 days [18, 19], and of 70 days [19]. Penland et al. [26] showed that SSA prefiltering allows the use of the maximum entropy method (MEM) with low-order autoregressive (AR) models in spectral estimation. Based on this combination of robust, low-order AR models with SSA, Keppenne and Ghil [7] predicted the Southern Oscillation index for ENSO with considerable skill at 30 month lead times.

Each one of these applications required a better understanding of SSA’s properties, of its strengths and weaknesses. The present paper relies heavily on the experience thus accumulated, and goes considerably further in its methodological development of a coherent SSA toolkit.

### 1.3. Outline of the present study

In this paper we concentrate on single-channel SSA. The studies cited in the previous subsection raise four major, and several minor problems. First, the apparent arbitrariness in the choice of window length  $M$  gives pause. After analyzing

further the eigenvalue problem central to SSA in section 2.1, we show in section 2.2 how the choice of  $M$  can influence the analysis and the results. The possibility to vary  $M$  makes in fact SSA much more adaptable to a large range of time scales than other statistical tools such as complex PCA [27], or principal oscillation patterns (POPs [28]).

Another major problem is that of robustness and statistical confidence. Analytical formulae for confidence levels on the eigenvalues [29] apply only to sets of independent realizations. We examine this problem in section 2.3, using the methodology described in section 1.4. Sections 2.4 and 2.5 are devoted to the interpretation of the time-dependent by-products of SSA. In particular, we derive in section 2.5 an algorithm capable of extracting the components of the signal corresponding to individual eigenelements, at a given epoch.

Third, the identification of noise plateaus in the eigenvalue spectrum was quite subjective in VG and most of the subsequent work in atmospheric and climatic applications. Indeed, observed time series do not, in general, provide the ideal break described above, nor is noise ever perfectly white. In section 3, objective algorithms for noise reduction are developed. The analyzed signal is decomposed explicitly as the sum of an intrinsic *dynamical component* and an external *noisy component*.

Noise reduction is a well-known signal-processing problem. Techniques like Kalman filtering, as well as more sophisticated nonlinear dynamical algorithms, like optimal shadowing [30] have shown very satisfactory results in noise reduction. This type of methods requires, however, a knowledge of the evolution equations. Approximations of the equations can be found empirically, as shown by Casdagli [31], but this in turn requires very long data sets in order to provide a useful reconstruction. Quite to the contrary, SSA is best at extracting – by essentially linear but data-adaptive methods – useful information about nonlinear systems from short,

noisy time series, in the absence of – or without using – the knowledge about governing equations.

Finally, the interpretation of the eigenelements has to be sharpened: When is SSA successful at removing trends and nonstationarities? How can SSA capture nearly periodic behavior – is the occurrence of a pair of degenerate eigenvalues always associated with a peak in the spectrum? We attempt to answer these two questions in sections 4.1 and 4.2, respectively. Maximum entropy spectral estimation [32–34] also uses the Toeplitz matrix  $T_x$  in order to determine the associated autoregression (AR) coefficients. We propose in section 4.3 an estimation of the power spectrum of a time series which is consistent with both MEM and SSA. In section 4.4, SSA is compared with multi-taper spectral estimation [35].

Linear forecasting algorithms, based on SSA expansion and AR models, are presented in section 5, following Keppenne and Ghil [7]. Concluding remarks appear in section 6.

#### 1.4. Algorithm validation

In order to validate the algorithms we develop here, four single-channel processes with simple and known properties are analyzed by SSA. For each process, 100 Monte Carlo realizations of 150 points are generated. These sets are used to provide nonparametric estimates of the statistical confidence. For each realization of each process, SSA is applied using  $M = 40$  and  $M = 20$ , in order to examine the influence of the window length. The processes  $x_n$ ,  $1 \leq n \leq 4$ , are given below by defining  $y_n$  and  $w_n$  in

$$x_n = y_n + w_n. \quad (1.4)$$

For the processes 1, 2 and 3 (P1, P2 and P3)

$$y_{n,i} = 2 \cos(\Omega_1 i + \phi_1) + \cos(\Omega_2 i + \phi_2), \quad (1.5)$$

and  $w_{n,i}$  is a Gaussian white-noise process with

variance  $\sigma^2$  and zero expectation.  $\Omega_1 = \frac{2}{20} \pi$  and  $\Omega_2 = \frac{2}{7} \pi$  are fixed frequencies, with a long common period (equal to 140), while  $\phi_1$  and  $\phi_2$  are constant, random phases depending on the realization; since the common period is of the same order of magnitude as the interval of 150 time units over which the processes P1–P3 are sampled, the behavior of P1–P3 approximates that of stochastically perturbed quasi-periodic signals, where *no* common period is present. For P1, we take  $\sigma^2 = 2.5$ , so that it has the same variance as the “signal”  $y_1$ . For P2, a larger noise variance is taken,  $\sigma^2 = 4$ . For P3, the noise variance is the same as the variance of the secondary oscillation,  $\sigma^2 = 0.5$ .

For the fourth process (P4)  $y_{4,i}$  are the consecutive values of the  $Y$ -variable in an integration of the Lorenz [2] equations, with a predictor–corrector scheme using a time step of 0.02 and sampling rate of 0.1 (one value sampled every 5 time steps). The parameter values are those of Lorenz [2]. The variance of the white-noise process  $w_4$  is taken equal to that of P2,  $\sigma^2 = 4$ . Fig. 1 shows the first realization of each process. The dashed curves represent the first 150 values of the noisy signal  $x_n$  to be analyzed, and the dotted ones the pure signal  $y_n$ . The light solid curves are the processed data using the noise-reduction algorithm described in section 4, where this figure is discussed further.

SSA is also applied to the IPCC global surface air temperature data [36]. The yearly averaged values are shown in fig. 2. The data studied by Ghil and Vautard [23] were slightly different [37], as was the method. The IPCC data are quite noisy, but show a marked trend as well as a significant year-to-year variability and inter-decadal oscillations. This data set is studied as a worst-case example: the time series is short, noisy and nonstationary. Fortunately, the data are at least regularly sampled! As we do not dispose of independent realizations, the stability of the results is tested by repeating the analysis with data from 1861–1950, 1861–1951, and so on, up to 1861–1990.

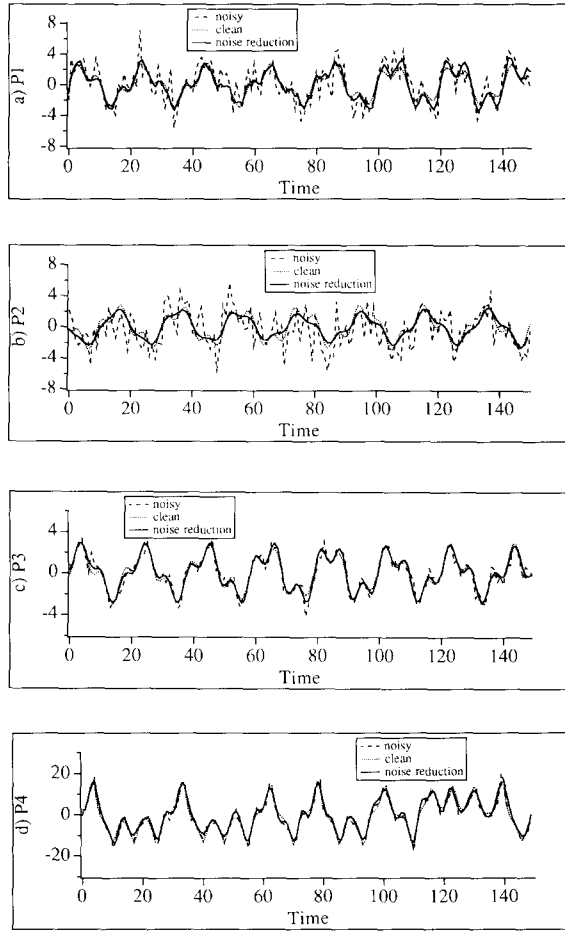


Fig. 1. One realization of the four synthetic processes under study, with the clean signal (dotted), the noisy time series (dashed), and the noisy series after application of the SSA noise-reduction filter (solid);  $M = 40$ . (a) P1, (b) P2, (c) P3 and (d) P4; note the scale difference on the ordinate between panels.

## 2. Theoretical and computational preliminaries

### 2.1. The Toeplitz matrix

The cornerstone of SSA is the Karhunen–Loève expansion theorem; this in turn is based on the lagged-covariance matrix of the process  $x$ , whose sample is the time series  $(x_i, 1 \leq i \leq N)$ , assumed – without loss of generality – to have zero expectation [38]. This matrix  $T_x$  has a Toeplitz structure, i.e., constant diagonals corresponding to equal lags:

$$T_x = \begin{pmatrix} c(0) & c(1) & \cdot & \cdot & \cdot & c(M-1) \\ c(1) & c(0) & c(1) & & & \cdot \\ \cdot & c(1) & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot & \cdot & \cdot \\ \cdot & & & \cdot & \cdot & c(1) \\ c(M-1) & \cdot & \cdot & \cdot & c(1) & c(0) \end{pmatrix}, \quad (2.1)$$

where  $c(j)$ ,  $0 \leq j \leq M-1$ , is the covariance of  $x$  at lag  $j$ .

There are different ways to estimate  $T_x$  [29, 39, 40]. Among the most frequently used are the Yule–Walker estimate

$$c(j) = \frac{1}{N} \sum_{i=1}^{N-j} x_i x_{i+j}, \quad (2.2)$$

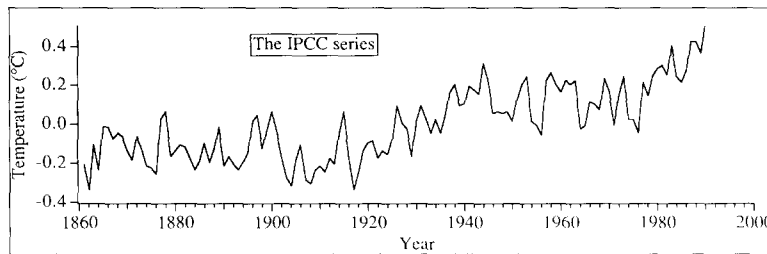


Fig. 2. The IPCC global temperature series [36] of yearly averages, as a function of time. The grand mean has been removed, i.e., the series is centered.

as well as the estimate

$$c(j) = \frac{1}{N-j} \sum_{i=1}^{N-j} x_i x_{i+j}. \quad (2.3)$$

Burg's [32] algorithm, which gives the AR coefficients associated with MEM, also estimates implicitly  $c(j)$ ; this estimate is again an average, like in (2.2, 2.3), but with larger weights toward the middle of the series than at the ends (see also refs. [26, 41]).

The first estimate, eq. (2.2), used by Box and Jenkins [39], is strongly biased when the number of data  $N$  is small; the estimate in (2.3) has larger variance but less bias [29]. When  $N$  is small, the Burg estimate can also be biased at large lags  $j$ , if there is power at periods larger than  $N$ , as shown by the following example: consider the process  $x_i = \cos(\frac{2}{150}\pi i + \phi)$ , where  $\phi$  is a random constant phase, and  $N = 100$ . Fig. 3 shows the average of the various estimates. In this case, (2.3) is slightly biased, whereas both (2.2) and the Burg estimates are strongly biased.

This problem occurs only when there are very low frequencies in the system, or trends. Once those frequencies are removed, both the Burg estimate and (2.3) are equivalent.

One could also estimate the covariance matrix of the process as in classical PCA, by averaging over lagged copies of the window [42]. This has a double disadvantage: (i) it does not conserve the

Toeplitz properties of the sample covariance (see the next subsection), and (ii) it tends to give, like the Burg estimate, larger weights to the middle of the series and thus produce large biases as well. In the absence of prior information about the signal, we prefer therefore the estimate (2.3), and use it hereafter to compute the Toeplitz matrix.

## 2.2. Eigenelements and choice of window length

The Toeplitz matrix  $T_x$  is symmetric and non-negative. Its eigenvalues  $\lambda_k$  are strictly positive, except when data are perfectly clean and come from a dynamical system with purely quasi-periodic behavior; in the latter case, all but a finite number – equal to twice the number of frequencies – are zero. They are sorted into decreasing order, and the eigenvectors  $E^k$  are normalized so that

$$\sum_{k=1}^M E_j^k E_l^k = \delta_{jl}, \quad 1 \leq j \leq M, 1 \leq l \leq M, \quad (2.4)$$

and that the spectral decomposition formula should hold:

$$\sum_{k=1}^M \lambda_k E_j^k E_l^k = T_{x,jl} = c(j-l), \quad 1 \leq j \leq M, 1 \leq l \leq M. \quad (2.5)$$

An eigenvector here is a lag sequence of length  $M$ . Since  $T_x$  is a Toeplitz matrix, eigenvectors are either symmetric or antisymmetric with respect to  $\frac{1}{2}M$ . When the sampling rate of the signal is increased, for a given window length in time units, the shape of the first eigenvectors does not change much, as shown by VG; thus, the eigenvectors have a limit as the sampling interval goes to zero. On the other hand, if the embedding dimension is constant and the sampling interval goes to zero, i.e., the window length also goes to zero, the eigenvectors tend to

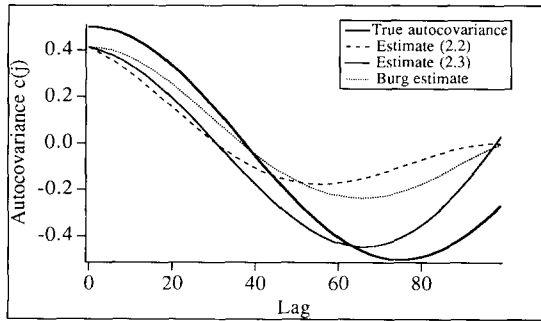


Fig. 3. Autocovariance function of the process  $x_i = \cos(\frac{2}{150}\pi i)$ : true (heavy solid curve) versus three different estimates from a 100 point long time series, as described in the text (see inset for legend).

fixed linear combinations of the  $M$  successive time derivative operators [43].

For fixed, finite window length, Fortus [44] has shown that the largest eigenvalue corresponds to the maximum of the spectrum, under certain restrictive assumptions. In the asymptotic limit  $M \rightarrow +\infty$ , with fixed sampling interval, i.e., the window length goes to infinity, the eigenvectors tend to pairs of sines and cosines and the associated eigenvalues tend to the corresponding *spectral density* values [45, 21]. For finite  $M$ , all eigenvalues fall between the maximum and the minimum of the spectral density [46].

A key problem in SSA is the proper choice of  $M$ . As we shall see forthwith, SSA does not resolve periods longer than the window length. Hence, if one wishes to reconstruct a strange attractor, whose spectrum includes periods of arbitrary length, the larger  $M$  the better, as long as statistical errors do not dominate the last values of the autocovariance function. To prevent this, one should not exceed  $M = \frac{1}{3}N$ .

In many physical and engineering applications, however, one wishes to concentrate on oscillatory phenomena in a certain band, which may be associated with the least-unstable periodic orbits embedded in a strange attractor. Such periodic orbits typically generate oscillations of strongly varying amplitude [18, 19, 23, 25]: the system's trajectory approaches and follows them for a certain time – comparable to or longer than the period in question – only to wander off into other parts of phase space. When the ratio of  $M$  to the life time of such an intermittent oscillation – the typical time interval of sustained high amplitudes – is large, the corresponding eigenvector pair suffers from the same Gibbs effect as in classical spectral analysis (see section 2.4 below). Spells of the oscillation – weak or strong – will be smoothed out. The following arguments should help understand the difficulty and make the correct choice of  $M$ .

Let us denote, by analogy with the time-continuous case treated in section 2 of VG, by  $\tilde{E}^k(f)$  the *reduced Fourier transform* of  $E^k$ , i.e.,

$$\tilde{E}^k(f) = \sum_{j=1}^M E_j^k \exp(2\pi i j f). \quad (2.6)$$

$\tilde{E}^k(f)$  is also the response function of the filter transforming  $x$  into its  $k$ th PC.  $\tilde{E}^k(f)$  is a sum of periodic functions of the frequency  $f$  with periods  $1, \frac{1}{2}, \frac{1}{3}, \dots, 1/M$ . Therefore the spectral resolution is  $1/M$ .

If  $M$  is too small, the coarse resolution may cause several neighboring peaks in the spectrum of  $x$  to coalesce. When there is an intermittent oscillation, reflected by a broad spectral peak, on the contrary, large  $M$  values (high resolution) will split the peak into several components with neighboring frequencies. Eq. (2.6) also shows that the filters are unable to isolate peaks at frequencies lower than the resolution  $1/M$ , i.e., periods larger than  $M$ . Given a peak in the power spectrum  $P_x(f)$  of  $x$  – with maximal spectral density at  $f_0$  and width  $2\delta f$ , say – these considerations suggest that SSA will isolate correctly the intermittent oscillation if

$$\frac{1}{f_0} \leq M \leq \frac{1}{2\delta f}. \quad (2.7)$$

In other words, the window length has to be chosen between the period of the oscillation and the average life time of its spells. In practice, this latter quantity cannot be estimated a priori, but SSA is typically successful at analyzing periods in the range  $(\frac{1}{3}M, M)$ .

### 2.3. Statistical stability of eigenvalues

Statistical stability is crucial in spectral analysis. SSA is based on estimates of lagged autocovariances, i.e., on second-order moments. The eigenvalues should converge therefore at least as well as Blackman–Tukey spectral estimates, heuristically speaking. Estimating the statistical confidence to be placed in the eigenvectors, however, is rather difficult. Simple confidence-interval formulae were used by Fraedrich [14], VG, and Ghil and Mo [25], who all assume



the independence of successive points of the augmented time series, either one sampling time or one window length apart. Here we use a nonparametric Monte Carlo method to check for stability, as we dispose of several realizations of the synthetic processes P1–P4.

Fig. 4 shows the average of the eigenvalues  $\lambda_k$ , calculated with the 100 randomly generated data sets, for the four synthetic processes, with  $M = 40$  (solid circles) and  $M = 20$  (open circles). The 95% error bars are calculated as  $\lambda_k \pm 1.96\sigma_k$ , where  $\sigma_k$  is the variance – estimated with the 100 realizations – of  $\lambda_k$ . Also shown are the confidence intervals estimated with the heuristic variance formula of Ghil and Mo [25],  $\sigma_k = \lambda_k \sqrt{M/2N}$ , for P1 only; the latter is quite conservative, except for the smallest  $\lambda_k$ 's.

For P1 (fig. 4a), the two spectra are quite similar, with two leading pairs standing out, and a regular weakly descending ramp after  $k = 5$ . These two pairs are associated with the periodic components of the signal, whereas the slowly decreasing part corresponds to the white-noise

dominated components. Note that the tails of the spectra are not flat. This is essentially due to the finiteness of the data.

Processes P2 and P3 (figs. 4b, 4c) behave in the same way as P1, with different noise plateaus. The second pair of P2 stands out less from the rest of the spectrum, since it is reached by the noise. We anticipate that for P2 SSA would not, based on most realizations, identify the secondary oscillation of period 7 with great confidence. For P3, the level of noise is low, and the last average eigenvalue is even negative (not shown on this logarithmic plot). Indeed, the covariance estimate (2.3) may lead to a nonpositive Toeplitz matrix, but negative eigenvalues are quite rare, and small in absolute value. For the stochastically perturbed Lorenz process (P4: fig. 4d), no clear break is apparent, although an inflection point of the average spectrum lies at  $k \cong \frac{1}{2}M$ .

The near-disappearance of the second pair into the noise level for P2 and  $M = 20$  raises another point to consider in choosing the win-

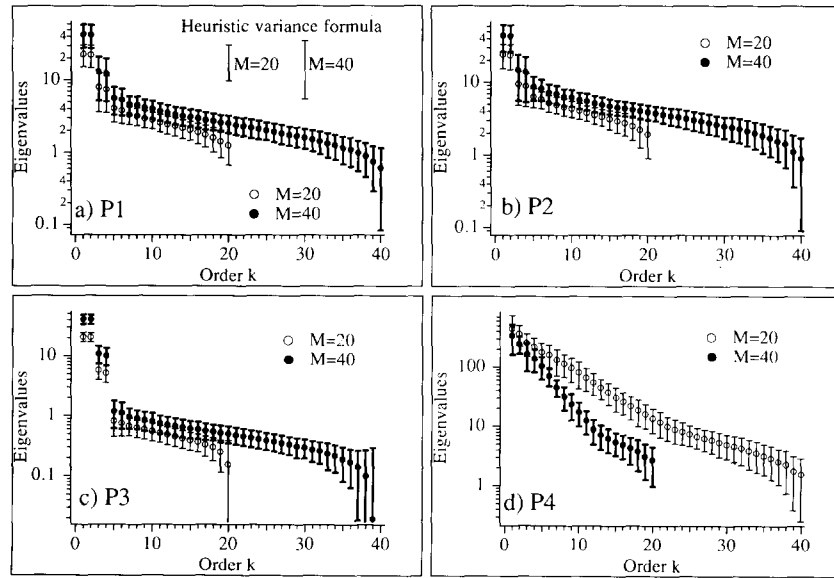


Fig. 4. SSA eigenvalue spectra, for  $M = 40$  (averages shown as solid circles), and  $M = 20$  (open circles), with the 95% confidence limits calculated using the actual variance of the 100 realizations (thick bars:  $M = 40$ ; thin bars:  $M = 20$ ). (a) P1, (b) P2, (c) P3 and (d) P4. At the top of panel (a) are represented also the error bars (thin long I's) estimated by the heuristic variance formula of Ghil and Mo [25]; note that the latter is quite conservative.

dow length: if the process is a pure sine function of variance  $v$ , the two nonvanishing eigenvalues are close to  $v' = \frac{1}{2}Mv$  (compare eqs. (2.7a)–(2.7e) of VG). On the other hand, if the process is noisy, the noise floor always lies at the value of the noise variance, for all values of  $M$  (with different slopes). Therefore, some pairs can be extracted from the noise level by increasing  $M$ . The fact that the noisy part of the eigenvalue spectrum is flatter for larger  $M$  values enhances pair detection further as  $M$  is increased.

For the IPCC data, the eigenvalue spectra (not shown) present no evidence of a noise floor, and are rather similar to the spectrum [23] of the Jones et al. [37] data. Since the length of the record is small (130 numbers), however, white-noise floors may be steep and hardly recognizable from signal. The quantitative method developed in section 3 shows that, in fact, about 20 eigenvalues are above the noise floor. The first two eigenvalues, representing the nonstationarity of the data, are about one order of magnitude above the other ones.

When SSA is applied to the IPCC data, with the ending year varying continuously, the order of the eigenvalues is not stable, whereas the shape of the eigenvectors is, i.e., it is possible to follow the eigenvectors continuously, from ending year to ending year, but the associated eigenvalues may undergo exchanges in their respective order (see also refs. [24, 42]).

#### 2.4. Principal components (PCs)

The  $k$ th PC is defined as the orthogonal projection coefficient of the original series onto the  $k$ th EOF:

$$a_i^k = \sum_{j=1}^M x_{i+j} E_j^k, \quad 0 \leq i \leq N - M. \quad (2.8)$$

PCs are therefore processes of length  $N - M + 1$ , which can be considered as weighted *moving averages* of the process  $x$ . If we denote by  $B$  the

backward shift operator, and by  $\Psi_k$  the polynomial

$$\Psi_k(\zeta) = E_M^k + E_{M-1}^k \zeta + \cdots + E_1^k \zeta^{M-1}, \quad (2.9)$$

then the  $k$ th PC can be written as

$$a_i^k = \Psi_k(B)x_{i+M}. \quad (2.10)$$

Like in classical PCA, principal components are orthogonal to each other, i.e.,  $\mathcal{E}(a^k a^l) = \lambda_k \delta_{kl}$ , where  $\mathcal{E}$  is the expectation operator. It does not mean that the PCs are independent from each other, since this relation holds for SSA only at lag zero. PCs give the representation of the augmented time series in a new coordinate system, with most information represented along the first coordinates.

The PCs can be interpreted in another way. Let us consider the portion of signal  $x$  contained between instants  $i + 1$  and  $i + M$ , and, for a given  $k$ , the function

$$J(\alpha) = \sum_{j=1}^M (x_{i+j} - \alpha E_j^k)^2. \quad (2.11)$$

It is easy to show that  $J(\alpha)$  is minimum for  $\alpha = a_i^k$ , and hence that the PCs can be obtained from the local fit, in the least-squares sense, of the  $k$ th EOF to the original series  $x$ . This property is conserved if one fits several EOFs, i.e., the PCs are the coefficients of the linear combination of any subset of EOFs that minimizes the square distance to the series  $x$  over the window considered. The implication is, again, that if an intermittent oscillation has a typical life time shorter than the window length, Gibbs effects will reduce the amplitude of the fit within the spells and produce artificial periodicity off the spells.

From the spectral point of view, EOFs correspond to data-adaptive moving-average filters. The power spectrum  $P_k(f)$  of  $a^k$ , at frequency  $f$ , is

$$P_k(f) = P_x(f) |\tilde{E}^k(f)|^2, \quad (2.12)$$

with  $\tilde{E}^k$  being given by (2.6), or

$$P_k(f) = P_x(f) |\Psi_k(\zeta)|^2, \quad (2.13)$$

for  $\zeta = \exp(2\pi if)$ . The orthogonality constraints of the problem give the identity

$$1 = \frac{1}{M} \sum_{k=1}^M |\tilde{E}^k(f)|^2 \quad (2.14)$$

for any frequency  $f$ , so that the sum of the spectra of the PCs is identical to the power spectrum of  $x$ , i.e.,

$$P_x(f) = \frac{1}{M} \sum_{k=1}^M P_k(f). \quad (2.15)$$

It is thus interesting to build *stack spectra* by piling up the contributions of the various components (see section 4.3).

### 2.5. Reconstructed components (RCs)

PCs are filtered versions of the original series. However, they do not allow a unique expansion of the signal into a sum of the different components. Indeed, in the expansion (1.2), individual terms depend on the index  $j$ , varying from 1 to  $M$ . Therefore, there are  $M$  different ways of reconstructing the components of the signal, which do not give, in general, the same results. Another problem in using eq. (1.2) for filtering, and in particular for *real-time* filtering and prediction, is that the resulting series are of length  $N - M + 1$ , and not of length  $N$  as desired. We show here how to extract, in an optimal way, series of length  $N$  – corresponding to a given set of eigenelements – that we shall call *reconstructed components* (RCs).

Let us consider a subset  $\mathcal{A}$  of eigenelements  $k$  over which the reconstruction is to be performed. By analogy with eq. (1.2), we seek a series of length  $N$ ,  $y = R_{\mathcal{A}}x$ , such that the quantity

$$H_{\mathcal{A}}(y) = \sum_{i=0}^{N-M} \sum_{j=1}^M \left( y_{i+j} - \sum_{k \in \mathcal{A}} a_{i-j}^k E_j^k \right)^2 \quad (2.16)$$

is minimized. In other words, the optimal series  $y$  is the one whose augmented version  $Y$  is the closest, in the least-squares sense, to the projection of the augmented series  $X$  onto EOFs with indices belonging to  $\mathcal{A}$ . The solution  $y = R_{\mathcal{A}}x$  to this least-squares problem is given by

$$(R_{\mathcal{A}}x)_i = \frac{1}{M} \sum_{j=1}^M \sum_{k \in \mathcal{A}} a_{i-j}^k E_j^k \quad \text{for } M \leq i \leq N - M + 1, \quad (2.17a)$$

$$(R_{\mathcal{A}}x)_i = \frac{1}{i} \sum_{j=1}^i \sum_{k \in \mathcal{A}} a_{i-j}^k E_j^k \quad \text{for } 1 \leq i \leq M - 1, \quad (2.17b)$$

$$(R_{\mathcal{A}}x)_i = \frac{1}{N - i + 1} \sum_{j=i-N+M}^M \sum_{k \in \mathcal{A}} a_{i-j}^k E_j^k \quad \text{for } N - M + 2 \leq i \leq N. \quad (2.17c)$$

When  $\mathcal{A}$  consists of a single index  $k$ , the series  $R_{\mathcal{A}}x$  is called the  $k$ th RC, and will be denoted by  $x^k$ . RCs have additive properties, i.e.,

$$R_{\mathcal{A}}x = \sum_{k \in \mathcal{A}} x^k. \quad (2.18)$$

In particular, the series  $x$  can be expanded as the sum of its RCs:

$$x = \sum_{k=1}^M x^k. \quad (2.19)$$

Note that, despite its linear aspect, the transform changing the series  $x$  into  $x^k$  is, in fact, non-linear, since the eigenvectors  $E^k$  depend non-linearly on  $x$ . A drawback of RCs is that they are correlated even at lag 0.

The main advantage of using RCs instead of PCs is the recovery of the epochs: indeed, if there are short spells of oscillations in the signal, PCs do not allow to localize them precisely.

whereas RCs do. Moreover, there is no phase shift between  $x$  and  $x^k$ , except possibly near the ends of the series: In the middle of the data set – where (2.17a) is used – the transform from  $x$  to  $R_{\mathcal{A}}x$  is a linear filter whose response function

$$\rho_{\mathcal{A}}(f) = \frac{1}{M} \sum_{k \in \mathcal{A}} |\tilde{E}^k(f)|^2 \quad (2.20)$$

is real valued. Therefore, at no frequency is there a phase shift between  $x$  and  $x^k$ . Within  $M$  points of the ends of the series, however, there might be some phase shift. This effect is small as long as  $\mathcal{A}$  is large, or suitably chosen, and, of course, totally disappears when  $\mathcal{A} = \{1, \dots, M\}$ . Based on a number of SSA experiments, we also found that there is no phase shift at the ends of the series when  $\mathcal{A}$  is made up of an oscillatory pair  $(k, k+1)$ .

Our reconstruction algorithm may be compared to the Wiener filtering method [47], which also provides optimal filters in a least-squares sense. The main differences are (i) that the Wiener filter uses harmonic functions as a basis, and (ii) that it is not fully adaptive, in the sense that one has to prescribe the shape of the desired filtered spectral density. Smooth and very reliable estimates of the power spectrum are therefore required in the Wiener method, which are impossible to obtain with short data sets.

### 3. Noise reduction

The simplest kind of noise reduction is by applying a fixed, prescribed low-pass filter to the data. This procedure is successful when the power spectrum of the process is rapidly decreasing to zero. When the spectrum is not monotonic, and has lines or peaks distributed over a wide range, the problem is more complicated. The gaps between these significant elements in the spectrum are filled by spurious noise, and to

filter out this noise requires more-complex filters. As stated at the end of the last section, it is still possible to use the Wiener method in this case, but only by assuming either accurate spectral estimates or prior hypotheses on the noise variance; the former is impossible with short data sets, the latter is arbitrary.

In this section we show that SSA is a powerful tool for signal reconstruction from noisy data. We assume that we have at our disposal a realization of a finite process  $x$  of length  $N$ , which is the sum of a dynamical process  $y$  and of some external noise process  $w$  due to measurement errors. In VG, we showed that the last PCs of such processes are dominated by noise, after the break in the eigenvalue spectra. We develop here a systematic method for determining the break, and derive a process  $R_{\mathcal{A}}x$ , with  $\mathcal{A} = \{1, \dots, S\}$ , approximating  $y$  with the knowledge of only a few  $x$ -data.

Fig. 5 shows the average of the ratio  $n(p)$  as a function of  $p$ , over the 100 realization sets of the four synthetic processes,

$$n(p) = \frac{\sum_{i=1}^N (y_i - \sum_{k=1}^p x_i^k)^2}{\sum_{i=1}^N \{y_i - x_i\}^2}; \quad (3.1)$$

$n(p)$  represents the noise-reduction ratio when  $p$  RCs are considered. For P1, P2 and P3 (figs. 5a–5c), the average optimum noise reduction is at  $p=3$  or  $p=4$ , and the reduced noise is less than 10% of the original. With  $M=20$ , the best reduction that can be achieved is not quite as good as with  $M=40$ . Even with short data sets and dominant noise, like for P2, reconstruction with  $p=3$  or 4 is very close to the underlying signal. For the process P4 (fig. 5d), the optimal reduction – to about 0.6 – occurs around  $p = \frac{1}{2}M$ . This relatively bad score is due to the fact that the Lorenz system has a monotonically decreasing power spectrum, with no particular frequency standing out, and therefore any linear noise-reduction technique unavoidably filters out a part of the signal as well.

In the real world, we do not have several

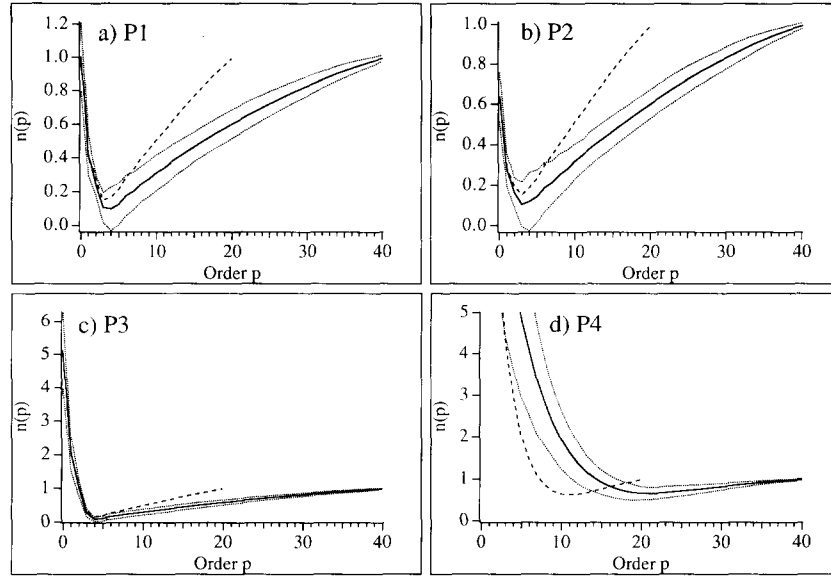


Fig. 5. Average error made when approximating the clean signal by the reconstruction filters  $R_{\mathcal{A}}$ , with  $\mathcal{A} = \{1, \dots, p\}$ , divided by the noise variance, as a function of  $p$ , for  $M=20$  (dashed) and  $M=40$  (solid). Dotted curves show the 95% confidence interval for  $M=40$ . (a) P1, (b) P2, (c) P3 and (d) P4.

realizations of the same process at our disposal, and the pure signal is not known. The optimal order, therefore, needs to be calculated in a different way. For a given order  $p$ , let us denote again by  $\mathcal{A}$  the subset  $\{1, \dots, p\}$ , and by  $\mathcal{A}'$  the subset  $\{p+1, \dots, M\}$ . The algorithm is based on the remark that if the reconstructed part of  $x$  involving the subset  $\mathcal{A}'$  of indices is really dominated by white noise, it should not be significantly different from the reconstruction, using the same EOFs, of some pure gaussian white-noise process. Let  $v$  be a normal white-noise process. The problem is to find a positive number  $\beta$ , such that  $R_{\mathcal{A}}x$  behave like  $R_{\mathcal{A}}(\beta v)$ . In this case, at least the autocovariance functions of  $R_{\mathcal{A}}x$  and  $R_{\mathcal{A}}(\beta v)$  should be nearly equal at lags 0 to  $M-1$ . That is, if SSA is reapplied to the two processes, it will provide statistically indistinguishable results.

To test this idea, we use again Monte Carlo simulation. This type of simulation, going back to the work of Ulam in the 1950s, has been used extensively of late in meteorological time-series analysis [48]; it has been applied recently in

nonlinear dynamics under the name of the *surrogate-data* method [49]. We generate 100 normal realizations of length  $N$  of the white-noise process  $v$ . The operator  $R_{\mathcal{A}}$  is applied – for a given order  $p$  – to the 100 realizations of  $v$  and to the signal  $x$ . Let us denote by  $\omega$  the realization, by  $c_{\omega}(j)$ ,  $0 \leq j \leq M-1$ , the autocorrelation of the reconstructed  $\omega$  with realization  $R_{\mathcal{A}}v$  at lag  $j$ , and by  $c_p(j)$  the autocorrelation of  $R_{\mathcal{A}}x$ . The average  $c_v(j)$  over these realizations, the variances  $s_v(j)$  of these estimates, and the 95% confidence intervals  $(c_-(j), c_+(j))$ , with  $c_{\pm}(j) = c_v(j) \pm 1.96\sqrt{s_v(j)}$ , are estimated. Then we look for  $\beta$  such that, for  $0 \leq j \leq M-1$ ,  $c_p(j)$  lies in the interval  $(\beta^2 c_-(j), \beta^2 c_+(j))$ ;  $\beta$  must therefore lie in the intersection of  $M$  intervals and must be positive. If this intersection is empty, the null hypothesis of white-noise dominated behavior of the last  $M-p$  components is rejected. Otherwise, this intersection is itself an interval  $(\gamma_p, \delta_p)$ , and the last  $M-p$  RCs may be considered as a reconstruction of mere white noise. The smallest  $p$  satisfying the above condition is called the *statistical dimension* of the data set,

and is denoted here (as in VG) by  $S$ . The interval  $(\gamma_S, \delta_S)$  gives bounds for the standard deviation  $\beta$  of the noise present in the data.

As shown by VG,  $S$  has nothing to do with the dimension of an underlying attractor:  $S$  simply gives the number of significant components in an SSA expansion; it depends almost linearly on the window  $M$  for synthetic and observed data, although VG only used a heuristic criterion for finding  $S$ . For almost all  $p > S$ , the conditions of nonrejection are fulfilled; sampling errors are responsible for the other cases. For  $p = S$  and  $\mathcal{A} = \mathcal{I} = \{1, \dots, S\}$ , we obtain

$$x = R_{\mathcal{I}}x + R_{\mathcal{J}}x. \quad (3.2)$$

$R_{\mathcal{I}}x$  is an approximation of  $y$ ,

$$y = R_{\mathcal{I}}x + \varepsilon, \quad (3.3)$$

where

$$\varepsilon = R_{\mathcal{I}}y - R_{\mathcal{I}}w. \quad (3.4)$$

The error  $\varepsilon$  made when approximating  $y$  by the

significant reconstructed part  $R_{\mathcal{I}}x$  of  $x$  is due to the difference of two quantities. The first one describes the part of signal that has been removed by the reconstruction filter; the second one is the part of the noise that has not been removed by it. The skill of the noise-reduction algorithm depends on minimizing both of these spurious quantities. The quantity plotted in fig. 5 shows, in fact, the average variance ratio of  $\varepsilon$  to  $w$  as a function of  $p$ .

The first contribution to  $\varepsilon$  dominates for small values of  $p$ , and is a monotonically decreasing function, whereas the second contribution dominates for large values of  $p$ , and is monotonically increasing. Processes for which the noise reduction is successful are those for which the variance of the first contribution in  $\varepsilon$  decreases rapidly with  $p$ , i.e., processes for which a few components explain a large part of the variance. Quasi-periodic processes and processes dominated by low frequencies fall within this category. More generally, the algorithm will be successful when a significant part of the power lies within a small fraction of the frequency range of interest.

Fig. 6 shows the distribution of  $S$  obtained

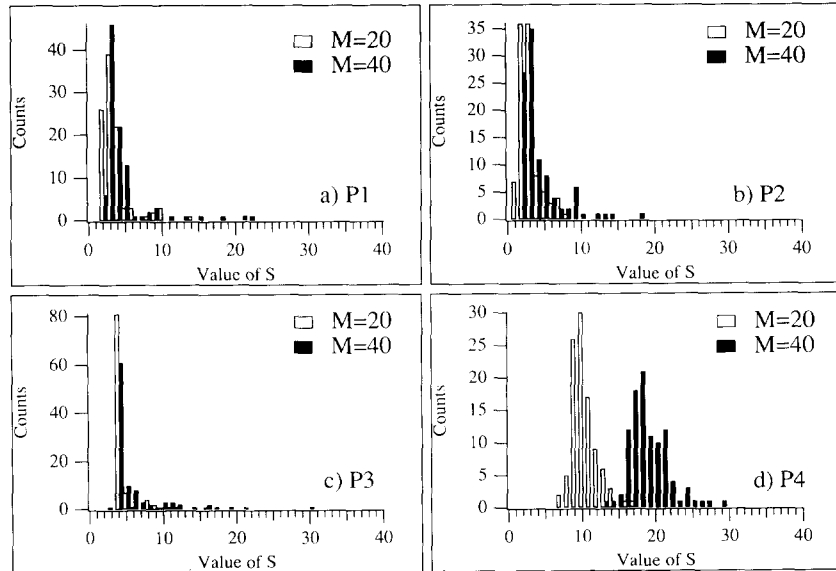


Fig. 6. Histograms of the statistical dimension estimate  $S$ , for  $M = 40$  (solid) and  $M = 20$  (open). Bars are placed to the left of the abscissa point for  $M = 20$ , and to the right for  $M = 40$ . (a) P1, (b) P2, (c) P3 and (d) P4.

with 100 realizations of the various test processes, using  $M = 40$  and  $M = 20$ . For P1, P2, and P3 (figs. 6a–6c), there is but little difference between the distribution with  $M = 20$  and  $M = 40$ , while the histogram peak is, as desired, at  $S = 3$  or  $S = 4$  (cf. fig. 5). Only for P2, the noisiest process, do many realizations actually give  $S = 2$ . For P3, most of the values are  $S = 4$ ; therefore, both oscillations are recognized as significant, even though the noise has the same variance as the oscillation of period 7. For P4 (fig. 6d), the peak lies at  $S = 10$  for  $M = 20$  and at  $S = 18$  for  $M = 40$  (i.e.,  $S \cong \frac{1}{2}M$ , like the  $p$ -value for optimal noise reduction).

For the IPCC data set, the estimates of  $S$  vary – with the ending year – from 14 to 22 (most of the values equaling 18) for  $M = 40$ , and from 7 to 12 for  $M = 20$ .  $S$  tends to increase towards the end of the time series, probably due to the increasing amount of data. The estimate of  $S$  is therefore almost proportional to  $M$ , as expected (cf. VG), and otherwise well behaved. The present values are larger than the estimate  $S = 12$  of Ghil and Vautard [23], obtained for  $M = 40$ , but coming from a different data set [37] and using a different Monte Carlo method; the earlier method [23] compared the eigenvalue spectra (fig. 1 there) instead of the autocovariance function. For the Jones et al. [37] data set the present method also gives lower  $S$ -values.

The characteristics of the noise's estimated standard deviation  $\beta$  are summarized in table 1.

The estimates  $\gamma_S$  and  $\delta_S$  are stable, since their standard deviation at given  $M$  is about one tenth of their respective average values. These estimates are also fairly insensitive to the window length, since doubling  $M$  only changes the average by a few percent at most. For P1, P2, and P3,  $\gamma_S$  and  $\delta_S$  are both nearly equal to the true noise standard deviation  $\sigma$ , all but 4 out of 12 values in the table being well within their own standard deviations. The error made in estimating the noise is not only small – about 10% for  $\frac{1}{2}(\gamma_S + \delta_S)$  versus  $\sigma$  – it is also on the safe side.

The excess of variance for the estimated noise is due to the fact that the algorithm tries to fit  $R_{\mathcal{G}}x = R_{\mathcal{G}}y + R_{\mathcal{G}}w$  to  $R_{\mathcal{G}}v$ , where  $v$  is a white-noise process. Since  $c_-$  and  $c_+$  are, in fact, bounds on the variance of  $v$ , and  $y$  and  $w$  are uncorrelated, the excess arises from the fraction of the variance of  $y$  contained in the frequency bands involved by the filter  $R_{\mathcal{G}}$ . It follows that, while in principle  $\gamma_S \leq \sigma \leq \delta_S$ , the only completely reliable inequality in practice is  $\sigma < \delta_S$ . The closeness of individual reconstructions to the process  $y$  can be checked for the four processes in fig. 1a–1d. The values of  $\gamma_S$  and  $\delta_S$  for the IPCC data set are also quite stable for the time intervals under study.

Fig. 7a shows the average and the standard deviation of the 41 reconstructions  $R_{\mathcal{G}}x$  of the significant temperature signal, based on ending dates between 1950 and 1990. It is noteworthy that the standard deviation at all times is very

Table 1

Average and standard deviation of the bounds  $\gamma_S$  and  $\delta_S$  on the standard deviation of the noise, as identified by the reduction algorithm. These values are calculated from the 100 synthetic realizations – for processes P1–P4 – and with the final year moving from 1950 to 1990 – for the IPCC data. Each column contains the average (left number) and the standard deviation (right number).

Process	$M = 40$		$M = 20$	
	$\gamma_S$	$\delta_S$	$\gamma_S$	$\delta_S$
P1	$1.67 \pm 0.17$	$1.78 \pm 0.18$	$1.66 \pm 0.21$	$1.83 \pm 0.17$
P2	$2.09 \pm 0.20$	$2.24 \pm 0.18$	$2.07 \pm 0.25$	$2.30 \pm 0.20$
P3	$0.71 \pm 0.08$	$0.77 \pm 0.07$	$0.68 \pm 0.09$	$0.79 \pm 0.07$
P4	$2.43 \pm 0.33$	$2.57 \pm 0.31$	$2.29 \pm 0.40$	$2.56 \pm 0.34$
IPCC	$0.065 \pm 0.007$	$0.070 \pm 0.004$	$0.070 \pm 0.007$	$0.074 \pm 0.003$

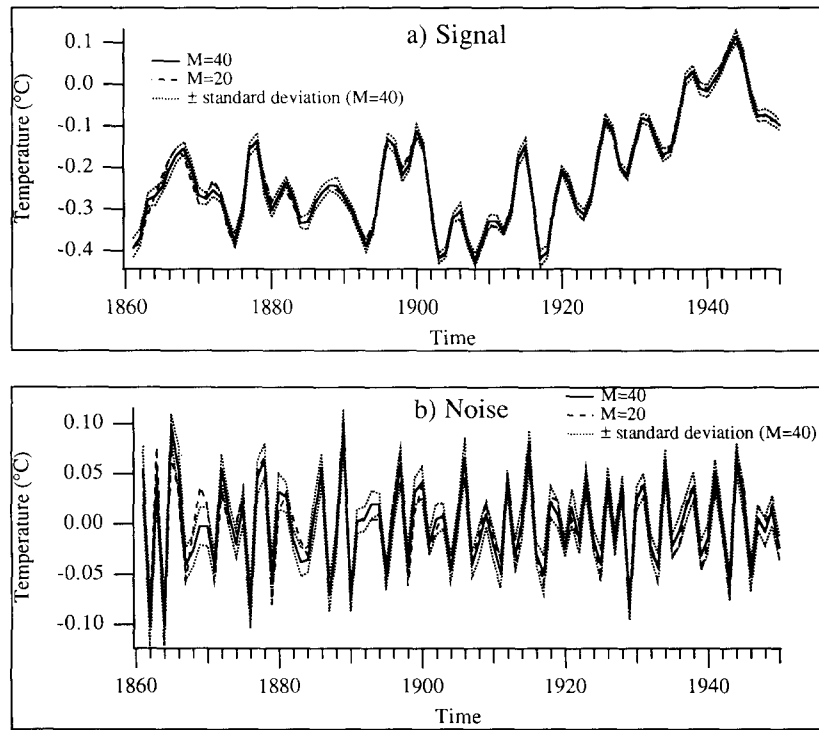


Fig. 7. (a) Noise-reduced IPCC series, calculated with the 41 ending dates running from 1950 to 1990, for  $M = 40$  (solid) and  $M = 20$  (dashed). The standard deviation of the noise-reduced data is represented by dotted lines (for  $M = 40$  only). (b) Same as (a) for the noise part only. Curves for  $M = 20$  and  $M = 40$  are almost indistinguishable, for the noise as well as the signal.

small compared with the fluctuations over time of the average. The agreement between the filtered series for  $M = 20$  and  $M = 40$  is also quite remarkable, even at the ends, in spite of the dependence of  $S$  on  $M$  and on ending date.

In fig. 7b are plotted the same quantities for the reconstructed noise part  $R_q x$  of the signal. Again, the variability from one estimate to another is considerably smaller than the noise variability itself. We conclude that, despite the fact that individual eigenelements in the tail of the expansion are not statistically stable, the global reconstruction process is stable – as the length of the data set as well as the window length  $M$  is varied.

The Monte Carlo method becomes rapidly untractable numerically as  $N$  increases, and the above algorithm can be shown to have a bias towards large values of  $S$ . Still, some experiments performed with  $N = 10\,000$  on P4 showed

that the noise-reduction factor (fig. 5) has a minimum value below 1 even when  $N \rightarrow \infty$ , and one can identify  $S$  as the order at which this minimum occurs. For large values of  $N$ , the eigenvalue spectrum itself becomes much more reliable and hence should provide a solution to this identification problem: as a noise plateau emerges more clearly with larger  $N$ , the break point should give an estimate of the best order  $p = S$  to use for noise reduction. Other methods, like the Wiener filter, can also be used when  $N$  is large.

#### 4. Interpretation of the eigenelements

##### 4.1. Trends and nonstationarities

The interpretation of SSA results relies on the assumption that the process  $x$  under study is



stationary in the weak sense, i.e., that the second-order moments are invariant under translation, although the Karhunen–Loève expansion theorem does *not* require the stationarity of the process [38]. Individual realizations of a process of length  $N$  may indeed appear nonstationary. This happens typically when periods larger than  $N$  are present in the system, even if the process is stationary, like in the example of fig. 3 in section 2.1.

When several realizations are available, it is possible to check for stationarity. For instance, let us assume that the visual trend present in the IPCC temperature data (figs. 2 and 7a) results from a natural climatic oscillation with a period of, say, 500 years; then another realization of this process might show a decreasing trend. However, when only one realization is available, it is impossible to distinguish between actual trends or nonstationarities and the presence of ultra-low frequencies. In practice, SSA still works quite well, just as if the stationarity of the process, along with the ergodicity of the realizations, were satisfied.

The distinction between trend and stationary ultra-low frequency can be crucial in a given application. For instance, if the temperature data over the last century reflect a true trend, and this trend is caused by anthropogenic increases in greenhouse trace gases, such as carbon dioxide ( $\text{CO}_2$ ), then a number of technological and socio-economic consequences follow [36]. The statistical significance of the trend in the data was established by Kuo et al. [50], while the causal role played by  $\text{CO}_2$  increases is plausible but not definitively confirmed.

From the point of view of studying the higher frequencies clearly manifest in a time series, the presence of either a trend or an ultra-low frequency is a major impediment. Various detrending methods, such as prewhitening [40], polynomial fits [51] and spline fits [52] exist. They all have some advantages and serious drawbacks. Ghil and Vautard [23] showed that SSA provides an effective and adaptive method of detrending,

with little if any undesirable aliasing. We derive here a systematic data-adaptive algorithm for removing trends or ultra-low frequencies in a given data set.

The algorithm is based on the same principle as noise reduction. If the trend is sizable, it should appear in the first few PCs. Bearing this in mind, we use the nonparametric test of Kendall for global trend identification [29]: consider a sequence of values  $(x_i, 1 \leq i \leq n)$  and count the number  $K_r$  of pairs of indices  $(i, j)$ , with  $i < j$ , such that  $x_i < x_j$ . Roughly speaking, if  $K_r$  is large, there is a positive trend in the series, and if  $K_r$  is small, there is a negative trend in the series. More precisely, the distribution of the coefficient

$$\tau = \frac{4K_r}{n(n-1)} - 1 \quad (4.1)$$

tends rapidly to a normal distribution with zero expectation and standard deviation

$$s = \sqrt{\frac{2(2n+5)}{9n(n-1)}}. \quad (4.2)$$

The hypothesis of no trend is rejected, therefore, when the measured value of  $\tau$  is outside the interval  $(-1.96s, +1.96s)$ , with a 5% chance of being wrong.

The test is applied to the successive PCs, using  $n = N - M + 1$ , and we denote by  $T_1$  the first order such that the corresponding PC has no significant trend. Detrending is then performed by reconstruction over the set  $\mathcal{T}' = \{T_1, T_1 + 1, \dots, M\}$ . Since the ends of the time series may lead to artificial trends, a second Kendall test is performed on the detrended series. If a trend is still detectable, reconstruction over the set  $\mathcal{T}' = \{T_1 + 1, \dots, M\}$  is performed and again tested, and so on. The first order  $T^*$  for which the reconstructed series has no significant trend determines the order of the detrending process, and – in obvious notation –  $R_{\mathcal{T}^*}x$  is the detrended series.

For the processes P1–P4,  $T^*$  was always found to be 1. For the IPCC data  $T^*$  was found to lie between 2 and 4, the variations being mainly due to changes in the order of the eigenvalues and to our restricting attention – for simplicity – to the leading eigenelements, as far as the trend is concerned. Fig. 8 shows the average estimate of the IPCC temperature trend  $R_{\mathcal{T}}x$ ,  $\mathcal{T} = \{1, \dots, T^* - 1\}$ , calculated with different ending dates, for  $M = 40$  and  $M = 20$ , in the same format as figs. 7a, 7b. Again, individual trend estimates are almost identical. For  $M = 20$ , a low-frequency component is mixed in with the trend, justifying the use of larger windows by Ghil and Vautard [23]. The detrending algorithm is quite stable, even at the ends of the series, for  $M = 40$ .

#### 4.2. Pairs of eigenelements

In VG, we showed that when a vigorous – albeit irregular – oscillation is present, a pair of nearly equal eigenvalues stands out of the spectrum and that the associated eigenvectors and PCs are in quadrature. Even for a pure sinusoid, the two associated eigenvalues are not exactly equal (see eqs. (2.7d), (2.7e) of VG), so in practice it can be difficult to tell statistical degeneracy from oscillatory pairing. Ghil and Mo [25] introduced therefore an *ad hoc* criterion for the significance of quadrature between the PCs of a pair, based on their lag correlation. The main difficulty in this approach is the lack of reliable statistical significance estimators of the lag corre-

lation, especially when the processes are nearly periodic, as the PCs are suspected of being.

We propose here instead two natural criteria based on the spectral properties of the eigenvectors. The first criterion is based on the remark that oscillating pairs of eigenelements  $(k, k + 1)$  must be spectrally localized around the same frequency. The squares of the reduced Fourier transforms  $|\tilde{E}^k(f)|^2$  of EOFs  $k$  and  $k + 1$  are calculated, cf. eq. (2.6), at 500 equally sampled frequencies  $f$  between 0 and 0.5 cpy, and the frequency  $f_k$  corresponding to the maximum value is estimated. Then, for orders  $k$  and  $k + 1$  to represent an oscillatory pair, the difference  $\delta f_k = |f_k - f_{k+1}|$  has to be small. For a pure red-noise process (see VG for the analytical calculation of the EOFs in this case), one has  $\delta f_k \approx 1/2M$ . Since we want to exclude pairs for this type of process, we impose the criterion  $2M \delta f_k < 0.75$ .

Although necessary, this criterion is not sufficient, since the amplitude of the peaks must also be high. In fact, if the presumed oscillatory pair completely resolves a frequency  $f^*$  between  $f_k$  and  $f_{k+1}$ , the response function  $\rho_k(f^*) + \rho_{k+1}(f^*)$  of the reconstruction filter based on components  $k$  and  $k + 1$ , given by eq. (2.20), must be close to 1 at this frequency, cf. eq. (2.14). Therefore, the maximum value of  $\rho_{\mathcal{A}}(f)$ ,  $\mathcal{A} = \{k, k + 1\}$ , is calculated and the pair is kept as an oscillatory pair only when this maximum is larger than  $\frac{2}{3}$ , i.e., only if at least two thirds of the variance of  $x$  at the peak frequency  $f^*$  is described by the pair in question.

These two criteria are applied to the realizations of the four synthetic processes. Results are summarized in table 2. For  $M = 40$ , despite the large variance of the noise in P2, the algorithm still finds a pair corresponding to the period 20 for all realizations, whereas the period 7 is identified as a pair in 34 realizations only. For P1 and P3, the algorithm is successful in general, with pair 1–2 corresponding to the period 20 and pair 3–4 to the period 7. Each realization in which the pair 3–4 (and 1–2) is not recovered corre-

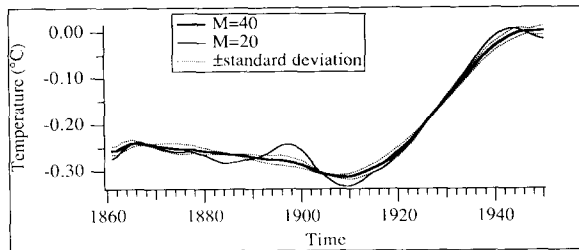


Fig. 8. Same as in the figs. 7a, 7b for the trend components.

Table 2

Oscillatory pairs for the four synthetic processes P1–P4, based on 100 independent realizations, for  $M = 20$  and  $M = 40$ . First row: average number of pairs per realization. Second row: number of realizations showing a pair for the oscillation with period 20. Third row: same as second row for the oscillation with period 7. Last row: average number of spurious pairs.

P1		P2		P3		P4	
$M = 20$	$M = 40$	$M = 20$	$M = 40$	$M = 20$	$M = 40$	$M = 20$	$M = 40$
1.49	1.68	1.14	1.60	2.14	2.46	1.78	5.51
100	100	92	100	100	100	/	/
35	47	18	34	99	100	/	/
0.13	0.20	0.04	0.24	0.16	0.44	1.78	5.51

sponds to values of  $S$  smaller than 4 (or 2, respectively). The number of spurious pairs is small for P1, P2 and P3. For  $M = 20$ , the number of spurious pairs is slightly reduced, as is the number of successful pairs, showing that our criteria are not very sensitive to a change in the window length.

P4 should not be reducible to nearly periodic components, except for the problems arising in the Lorenz system from the presence of embedded (unstable) periodic orbits of arbitrary length, on the one hand, and from finite sample length, on the other. The average number of detected pairs is in fact high, due to these problems. We shall see in section 4.4 that even the most sophisticated spectral methods of classical type find, for these data, a certain number of significant peaks, for the same reasons: they are simply there for any finite segment of a trajectory. Indeed, spells of oscillations occur when the trajectory spirals around the unstable fixed points, for an exponentially distributed length of time, and with varying mean periods. As a consequence, the frequency of the oscillations detected varies with realization, rather than being fixed, as it is for P1, P2 and P3 (see fig. 10 below).

The evolution of the pairing for the IPCC time series was calculated as a function of the ending date (not shown). After the components corresponding to the trend, an average of about three pairs are found (for  $M = 40$ ), associated with rather stable frequencies (see fig. 11 below). For the whole series (ending in 1990), we find five

oscillatory pairs with peak periods similar to those found in Ghil and Vautard [23], i.e., 26 years (pair 3–4), 15 years (pair 7–8), 10 years (pair 5–6), 5.2 years (pair 9–10) and 4.6 years (pair 16–17). Fig. 9 shows selected RCs of the IPCC series. The last two pairs just mentioned describe most of the variance with periods of 4–6 years, associated with the low-frequency component of the ENSO phenomenon (see also refs. [7, 22, 26, 53]. While the pairing itself, as a function of ending date, is not very stable for the IPCC data set (cf. also ref. [24]), both the reconstructions (fig. 9) and the dominant periods (fig. 11 below) are quite stable.

#### 4.3. SSA and MEM spectral estimates

Penland et al. [26] advocated the use of SSA for noise reduction before applying the maximum entropy method (MEM) to estimate the power spectrum of a time series. They estimated the power spectra of each of the  $S$  significant PCs with MEM, based on the formulation of Burg [32], and truncated eq. (2.15) to these PCs to obtain an approximation of the power spectrum  $P_x$  of  $x$ . The generally regular behavior of the PCs allowed them to obtain good spectral resolution with low-order AR models. The striking advantage of this SSA prefiltering is to eliminate therewith the spurious peaks inherent to high-order MEM estimates. We refine here this approach by deriving a fully consistent SSA–MEM spectral estimate.

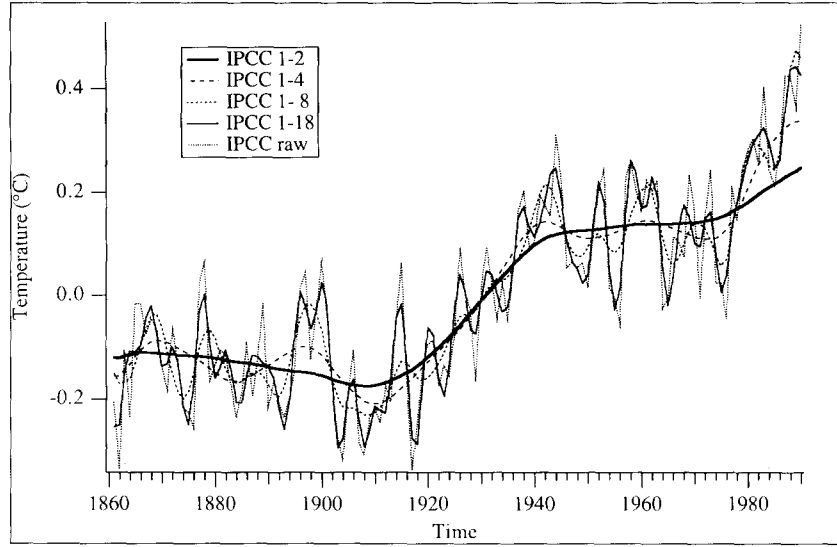


Fig. 9. Reconstructed subsets  $R_{\alpha}x$  of the IPCC series with ending date in 1990 only. The global average is removed.

We assume that the autocovariance function  $c(j)$  is known up to lag  $j = M - 1$ . Using the notation of eqs. (2.9), (2.10), an AR process  $z$  of order  $M - 1$  is fitted to the time series  $x$ , using the Yule [33]–Walker [34] method,

$$\Phi(B) z_i = v_i, \quad (4.3)$$

where  $v$  is a white-noise process of variance  $\sigma^2$  and  $\Phi(\zeta)$  a polynomial of degree  $M - 1$ , with real coefficients, of the complex variable  $\zeta$ . The PCs  $b^k$  of the fitted process satisfy

$$b_i^k = \Psi_k(B) \Phi^{-1}(B) v_{i+M}. \quad (4.4)$$

The power spectrum  $P_k(f)$  of  $a^k$  may thus be approximated by the spectrum  $P_k(f)$  of  $b^k$ ,

$$P_k(f) = \frac{|\Psi_k(\zeta)|^2}{|\Phi(\zeta)|^2} \sigma^2, \quad \zeta = \exp(2\pi if), \quad (4.5)$$

where  $\sigma^2$  is the variance of the noise process  $v$ .

In figs. 10a–10d, the histograms of the peak frequencies of the estimates (4.5) of  $P_k(f) + P_{k+1}(f)$  are displayed. Fifty bins between 0 and 0.5 are used for this calculation. For P1, P2 and P3, the spurious peaks are distributed almost

uniformly along the frequency axis. The probability of success in the identification of the oscillation with period 20 is close to 1 for all three processes, whereas it varies for the oscillation with period 7, P2 being the worst case and P3 the best, as expected. The spurious peaks, which are described by the oscillatory pairs for P4 (cf. table 2), are concentrated towards the lower frequencies, where the actual power is higher. These remarks hold both for  $M = 40$  and  $M = 20$ , although there is somewhat less success for the latter in identifying the period 7 at high noise levels (figs. 10a, 10b). It follows that, almost independently of (sufficient) window length, SSA separates oscillations from noise even when the noise variance is higher than that of the oscillations.

For the IPCC series, the results are presented in fig. 11, as a function of the ending year. The stablest oscillations are those with periods around 4–5 years, 10 years and 15 years. An interdecadal oscillation with a period of 20–30 years is also in evidence for about half the ending years. Note the similarities with the peaks estimated by Ghil and Vautard [23], using the Jones et al. [37] time series up to 1988 only.

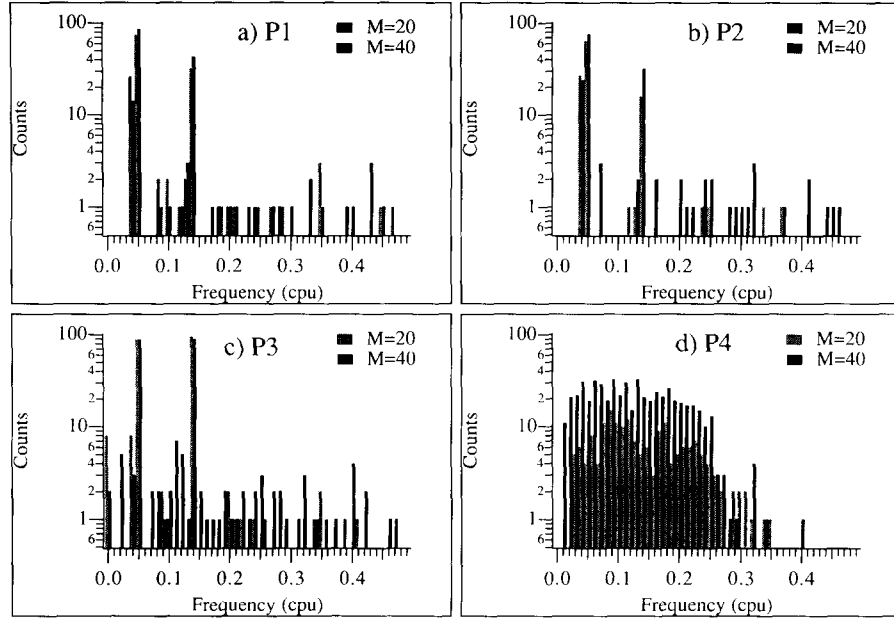


Fig. 10. Histogram of frequencies associated with the oscillatory pairs; frequencies are estimated by the maximum entropy method (MEM). Fifty bins of width 0.01 have been used to count the number of frequencies; the place of the bars for  $M = 20$  (shaded) and  $M = 40$  (solid) relative to the frequency bins is the same as in fig. 6. (a) P1, (b) P2, (c) P3 and (d) P4.

The theoretical advantage of the formulation (4.5), compared to that of Penland et al. [26], where an AR model is fitted to each individual significant PC, is that the additive property (2.15) of the spectra is conserved exactly. It justifies rigorously the display of *stack spectra*  $\mathcal{Q}_p$  of the PCs of the process  $x$ : the stack spectrum

of order  $p$  is the partial sum of (2.15) truncated at order  $p$ , i.e.,

$$\mathcal{Q}_p(f) = \frac{\sigma^2}{M|\Phi(\zeta)|^2} \sum_{k=1}^p |\psi_k(\zeta)|^2, \quad \zeta = \exp(2i\pi f). \quad (4.6)$$

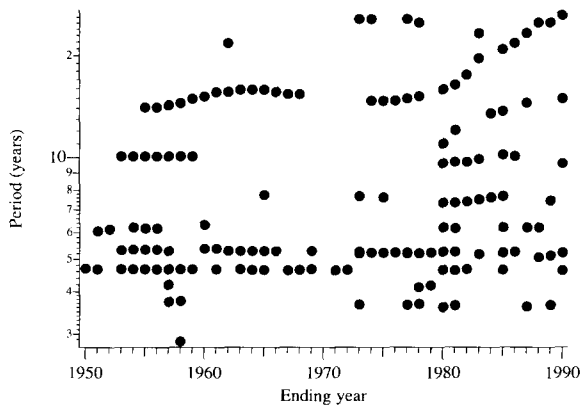


Fig. 11. Periods associated with the quasi-periodic components for the IPCC series, as a function of the moving final year.

Fig. 12 shows the stack spectra for the IPCC series ending in 1990, with  $p = 2$ ,  $p = 8$ ,  $p = S = 18$ , and  $p = M = 40$  (raw MEM). The three low-frequency oscillations described by the pairs (3, 4), (5, 6) and (7, 8) have periods of about 26, 15, and 10 years. Most of the variance associated with PCs 9-to-18 is related to ENSO oscillations, with periods between 4 and 6 years. Note that the peaks found at periods 4.2 and 3.2 years are rejected as noise, as well as most of the variance at periods below 4 years. There is remarkable agreement with the results of Ghil and Vautard [23], using somewhat different raw data [37] and MTM (their fig. 2). Despite the difference between the data sets and the method, we still

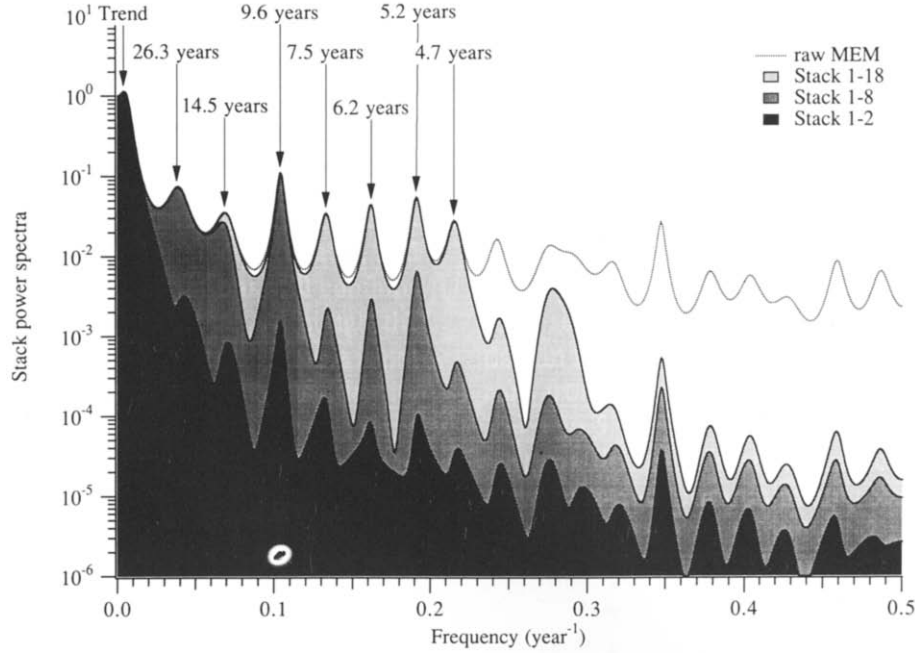


Fig. 12. Stack MEM spectra of the IPCC series (1861–1990), for various subsets of PCs, together with the raw MEM spectrum (i.e., the stack 1–40);  $M = 40$ .

obtain an interdecadal oscillation, the ENSO 4–6 year activity, and the two other periods (of 10 and 15 years).

#### 4.4. Multi-taper spectral estimates

The multi-taper method (MTM) – devised by Thomson [35], based on the work of Slepian [54] – is a nonparametric spectral analysis method, unlike MEM which assumes an AR signal. It provides, moreover, an array of statistical tests for the spectral estimates. MTM was applied to various geophysical time series [55–57] and more recently to several climatic time series, from historic temperature data [23, 50] through tree-ring data [58], to late Pleistocene data [52, 59]. We review briefly the main properties of MTM (see also appendix A of ref. [52] for succinct details) and apply it to the previously defined synthetic and IPCC data sets.

MTM aims to eliminate the spectral bias induced by finite sampling. Given a line frequency signal perturbed by white noise,  $x_i =$

$\mu \sin(2\pi f_0 t) + w_i$ ,  $1 \leq i \leq N$ , and a frequency band  $B = (f_0 - W, f_0 + W)$ , one wants to minimize the power leakage of  $f_0$  outside  $B$ , with a tapered signal  $x_i v_i$ ,  $1 \leq i \leq N$ . This goal is achieved with a subset of the  $K = [2WN]$  first Slepian sequences, otherwise called *discrete prolate spheroidal sequences* (DPSSs). These sequences  $(v^{(k)}, 1 \leq k \leq K)$  are easily computed, given  $W$  and  $N$  [54, 59].

The tapers are explicitly devised to optimize the estimate of line frequencies  $f_0$  and the corresponding amplitudes  $\mu$ . An estimate  $\nu$  of  $\mu(f_0)$ , for  $f_0$  known, is derived from a least-squares regression with respect to the coefficients of the  $K$  tapered signals, yielding

$$\nu(f_0) = \frac{\sum_{k=1}^K V_k(0) y_k(f_0)}{\sum_{k=1}^K V_k(0)^2}, \quad (4.7)$$

where  $y_k(f)$  is the Fourier transform of the  $k$ th tapered  $xv^{(k)}$  and  $V_k(f)$  is the Fourier transform of the  $k$ th taper  $v^{(k)}$ . Notice that the amplitude estimate  $\nu(f_0)$  is unbiased when the noise  $w$  is

white [56]. In practice,  $f_0$  is not known a priori, and the estimate (4.7) is computed for  $\nu(f)$  at different  $f$ , to estimate the position  $f_0^*$  of the maximum, as well as its value  $\nu(f_0^*)$ .

One can test the validity of this estimate by calculating the ratio of the explained to the unexplained variance. It turns out that the variable

$$F(f) = \frac{(K-1)|\nu(f)|^2 \sum_{k=1}^K V_k(0)^2}{\sum_{k=1}^K |y_k(f) - \nu(f)V_k(0)|^2} \quad (4.8)$$

follows a Fisher–Snedecor distribution – with 2 and  $2K-2$  degrees of freedom for the numerator and the denominator, respectively – when the noise is white [56]. Consequently, one rejects the null hypothesis  $\mu = 0$  (i.e., that the series is white) with a probability  $1 - q$  of being wrong, when the value  $F(f)$  exceeds a threshold  $F_q$  such that  $\mathcal{P}(F < F_q) = q$ .

Notice that for a single, finite realization, the maxima of  $\nu(f)$  and  $F(f)$  will not coincide in position. The position of a maximum in  $F(f)$  provides an unbiased – to first order in the detuning  $f - f_0$  – estimate of a line component when the noise is white. Hence we use the latter to estimate the position of the peaks. Moreover, the value of  $F$  does not depend – to first order – on the magnitude of  $\mu$ , allowing the detection of small-amplitude oscillations with less ambiguity than traditional spectral analysis methods. This estimate also appears to be robust in practice to the slope of the noise spectrum: the usual white-noise assumption can be replaced by colored noise, yielding a negative rather than zero slope in the frequency domain, like that of most climatic time series [60].

If the process  $x$  is composed of several lines, separated by at least  $2W$ , the procedure still applies in principle, due to the rough independence of the estimates at the different lines. Nonetheless, as we shall see, the test (4.8) detects more sharp lines than actually present in the signal, and the number of spurious peaks appears to be stable in our Monte Carlo experi-

ments, even though the lines themselves occur at random frequencies. For a pure white noise,  $F$ -values follow a Fisher distribution. In theory, therefore, if  $q$  is the probability of  $F$  being less than  $F_q$ , the expected number of  $F$ -values above  $F_q$  is  $(1 - q)N$ . It turns out in our experiments on pure white noise (not shown) that the number of peaks above  $F_q$  depends in fact on the bandwidth  $2W$  and the number of tapers  $K$ ;  $(1 - q)N$  appears therefore to be only a rough upper bound for the number of spurious peaks MTM will exhibit.

We investigate here the effects of SSA noise reduction on MTM. Contrary to MEM, there is no simple algebraic link between MTM and SSA. We have calculated, for the raw series and for the reconstructed noise-reduced series, the MTM spectral estimates, with  $WN = 6$  and  $K = 8$ . The position of the peaks significant at the 95% confidence level is displayed in histogram form in figs. 13a–13d. Only the case of noise reduction with  $M = 40$  is discussed, results being similar for  $M = 20$ .

For P1, the two lines appear for every realization, the period 20 being distributed within two histogram bins and the period 7 within one bin only (fig. 13a). No significant peaks occur in the vicinity of the two lines. The spectral power in a  $2W$  band around those frequencies is absorbed by the lines. Indeed – in this case –  $W = 0.04$ , which corresponds to the gaps on each side of the line peaks. The frequency band beyond 0.2 cpu yields a roughly continuous and flat histogram: all frequencies are spuriously detected in this noise band with equal probability. Hence MTM is an efficient estimator of the real lines in this case, and its spurious frequencies are uniformly distributed. With a 95% statistical confidence level, one expects at most seven peaks in the MTM spectrum of a white-noise process. For P1 (fig. 13a) there are four spurious peaks on average above this threshold, each peak being randomly distributed in the frequency band  $(0, 0.5)$  – the two gaps around the line frequencies excepted.

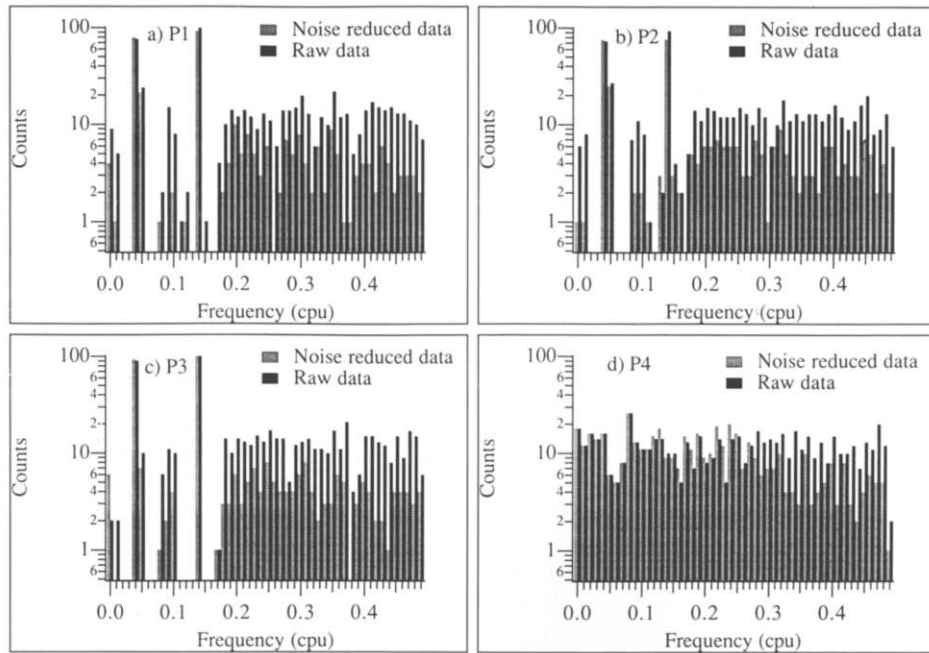


Fig. 13. Histogram of line frequencies identified by the MTM's  $F$ -test – at a 95% significance level – before noise reduction (solid bars) and after (shaded bars);  $M = 40$ . (a) P1, (b) P2, (c) P3 and (d) P4.

The noise reduction algorithm had a drastic effect on the tail of the MEM spectrum, whereas little change is observed for the occurrences of the MTM line peaks. For P1, this procedure reduces the number of spurious peaks exhibited by a factor of 2 – uniformly throughout the noisy part of the spectrum – while the power of the noise is divided by a factor of 10 (not shown). If we increase (P2), or decrease (P3) the white-noise variance, the MTM histograms of peaks are quite similar to those for P1. Therefore, MTM is relatively insensitive to the signal-to-noise ratio, in the case of a quasi-periodic process. Note that the SSA estimates (figs. 10a–10d) count much fewer spurious peaks than MTM. The number of SSA detections of the period 7 is unfortunately also affected by the noise-reduction algorithm, especially for P2 (fig. 10b), which is not the case of MTM estimations. SSA estimates are thus more sensitive to the signal-to-noise ratio, but tend to produce fewer spurious peaks; i.e., SSA is more conservative.

The application of MTM to the process P4

indicates no preferred frequency, proving the robustness of the method. Indeed, MTM considers the Lorenz system as a colored-noise process, with no greater probability of having a line in some band than in another. SSA noise reduction filters out higher frequencies, but the histogram shape for MTM line detections does not change much below 0.3  $\text{cpu}$ . SSA estimates, on the other hand, clearly detect peaks with a higher probability at lower frequencies. The difference between the two estimates is clear: SSA does not try to find lines, but rather spectral bands with a high percentage of explained variance.

The detrended IPCC series is again analyzed with a final date moving from 1950 to 1990. In fig. 14, we plot the periods of the peaks significant at the 95% confidence level, in a display format similar to fig. 11. For the detrended, but still noisy data (reconstruction of components from  $T^*$  to  $M = 40$ ; open circles in the figure), a sequence of peaks at 15 and 9 years stand out. At lower periods, a number of peaks occur between 2.5 years and 6 years. The SSA noise-



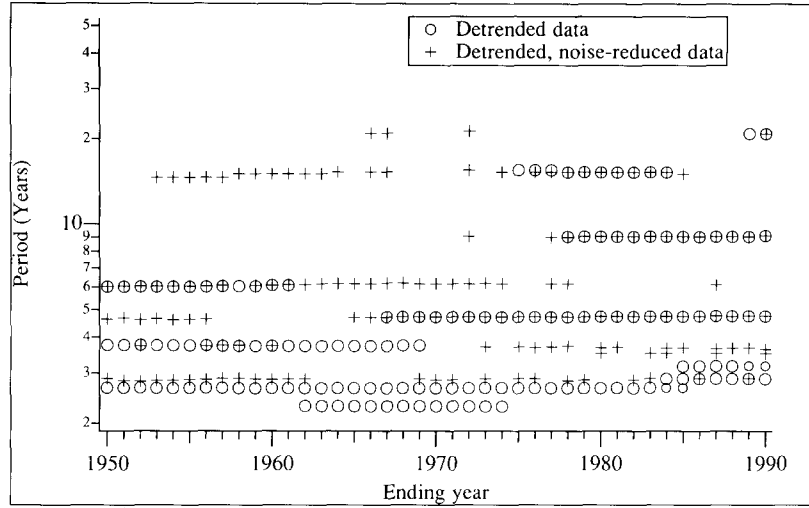


Fig. 14. Evolution of the periods identified by MTM—at the 95% significance level—calculated from the RCs  $R_{\mathcal{A}}x$ , with  $\mathcal{A} = \{T^*, \dots, 40\}$  (circles), and with  $\mathcal{A} = \{T^*, \dots, S\}$  (plus signs), as a function of the ending year.

reduction filter (RCs  $T^*$  to  $S$ ; pluses in the figure) essentially removes the two highest-frequency peaks. The peaks with periods of 21, 15, 9, 6, 4.6, 3.8, and 2.8 years all appear significant, although the last two have a much lower amplitude. This is in agreement with our MEM results (fig. 12). Note that the interdecadal oscillation with a period of 20–30 years detected by Ghil and Vautard [23] is not very stable at the 95% confidence level. Still, this interdecadal oscillation is significant at the 90% confidence level for every final date (not shown).

## 5. Application to prediction

Since the PCs are filtered versions of the signal (cf. section 3) and typically band-limited (cf. section 4), their behavior is more regular than that of the raw series, and hence more predictable. This leads to the heuristic idea of forecasting only a subset of PCs, e.g., all the significant ones or the oscillatory pairs exclusively. Thus a—hopefully large—fraction of the variance can be predicted with reasonable skill. The forecast for the entire series is the sum of the expected value of the complementary PCs, i.e., zero, and

of the forecast for the selected subset. Keppenne and Ghil [7] showed, by forecasting the Southern Oscillation index (SOI) of ENSO with considerable skill out to two-and-a-half years, that this approach can produce more accurate forecasts on an important and well-studied, but very irregular time series than any other method in current use.

We follow Keppenne and Ghil [7] and fit an AR model to each individual PC using the AR coefficient estimates of Burg [32], with different orders  $L$ . The order  $L = M - 1$ , which seems the most consistent with the SSA analysis (cf. section 4.3) was found to be too large compared with the simple behavior of the PCs: indeed, the variance of the AR coefficient estimates increases with the order [61]. Forecast errors were minimal, for all signals analyzed, when the order was quite low,  $L \leq 10$ . For small orders, forecast errors are dominated by the lack of resolution; for large orders, by the error made on the coefficients of the model. The results presented here are for  $L = \text{const.} = 10$ , and remain valid for even smaller orders.

Once forecasts of the individual PCs are produced, the reconstruction algorithm is applied, in order to compare forecasts with real data. Let us

denote by  $\mathcal{A}$  the subset of PCs used in the forecast. If the original series is of length  $N$ , the forecast for time  $N + \tau$  of the sub-signal in question is the reconstruction of the corresponding extrapolated PCs  $a_i^k$ , with  $1 \leq i \leq N - M + \tau$ . Note that this forecast for time  $N + \tau$  differs from the value at time  $N + \tau$  produced by a forecast for time  $N + \tau'$ , with  $\tau' > \tau$ . Indeed, the forecast for time  $N + \tau$  is the end of the reconstructed series, whereas the reconstruction process takes into account PC forecasts out into the future in the latter case. The validation of the forecasts is made by applying the *exact* same procedure to the actual PCs as to the extrapolated PCs.

In fig. 15, we plot – as a function of lead time  $\tau$  – the average forecast error for different sub-sets  $\mathcal{A}$ . The dotted curve represents the average, over the 100 realizations of the four synthetic processes, of the ratio

$$r_1(\tau) = \frac{[F_E(N + \tau) - y(N + \tau)]^2}{u_1}, \quad (5.1)$$

where  $F_E(N + \tau)$  is the MEM forecast at time  $N + \tau$ , using the raw series  $x$ , with  $L = 39$ ;  $y(N + \tau)$  is the extrapolated clean signal, calculated from formula (1.5) for P1, P2, and P3, and with the Lorenz equations for P4. In eq. (5.1),  $u_1$  is the measured variance of  $y$ . With this normalization, the perfect forecast has an error of 0, and the worst an error of 1, corresponding to the “climatological” forecast, i.e., using the past average as the future forecast. For P1, P2 and P3, these forecast errors are smaller for longer lead times – around  $\tau = 40$  – before increasing again. This results from the fact that the spurious oscillations generated by noise have an  $e$ -folding time shorter than the significant ones. The forecast is therefore polluted by the effect of noise on the AR coefficient estimates.

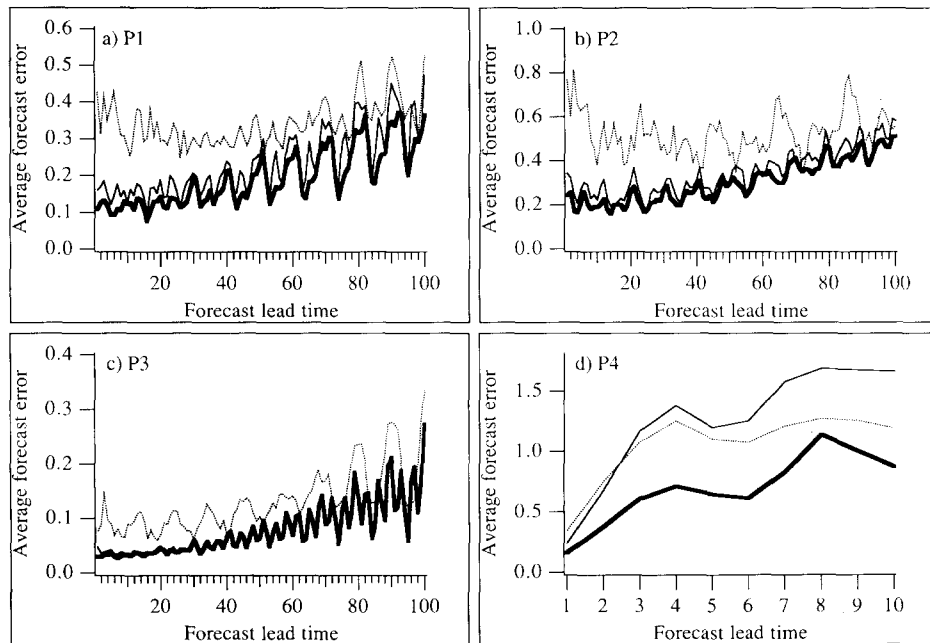


Fig. 15. Average errors of the SSA-MEM forecasts for the RCs  $R_{\mathcal{A}}x$ , normalized by the variance of the RCs of the clean signal (solid curves). Light lines stand for  $\mathcal{A} = \mathcal{S} = \{1, \dots, S\}$ , and heavy lines for  $\mathcal{A}$  being the union of the significant oscillating pairs. The dotted line represents the average error of the standard MEM forecast of the raw series, using the AR order  $L = 39$ . (a) P1, (b) P2, (c) P3 and (d) P4.

The two solid curves show, for two subsets  $\mathcal{A}$ , the average of the ratio

$$r_2(\mathcal{A}, \tau) = \frac{[F_{\mathcal{A}}(N + \tau) - R_{\mathcal{A}}y(N + \tau)]^2}{u_2} \quad (5.2)$$

where now  $u_2$  is the measured variance of  $R_{\mathcal{A}}y$ , while  $F_{\mathcal{A}}(N + \tau)$  is the forecast using the above algorithm with  $\mathcal{A} = \mathcal{S} = \{1, \dots, S\}$  ( $S$ -forecast: light solid) and  $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \dots \cup \mathcal{A}_Q$ , the quasi-periodic pairs  $\mathcal{A}_q$  defined in section 4.2 ( $Q$ -forecast: heavy solid). Thus,  $r_2(\mathcal{A}, t)$  represents the average forecast error made with respect to the corresponding reconstructed part of the (clean) signal.

Comparison of the solid curves in figs. 15a–15c with the dotted one shows that all forecasts produced by AR extrapolation of the PCs are much better, at all lags, than forecasts performed on the raw signal. For the three processes P1–P3, the error is reduced by a factor of 3 at the beginning and the PC forecasts are good for about 100 time units, i.e., five times the longest deterministic period involved. Raw MEM forecast errors grow more rapidly at longer lags than PC forecasts. Forecasts errors based on the oscillatory pairs only ( $Q$ -forecasts) are smaller than forecasts based on all the significant components ( $S$ -forecasts).

For the Lorenz equations (fig. 15d), PC-based forecasts are not better than raw MEM forecasts, when compared with the clean signal. After two time units, forecast errors are greater than 1, indicating a total loss of predictive skill. When only oscillatory PCs are considered, a predictability limit of seven time units is found, corresponding to the average period of spiral motion around the unstable fixed points. The (intermittently) oscillatory components describe, in general, more than 50% of the total variance of the finite-length Lorenz signal.

These results leave us with the hope that, for signals with intermittent oscillations, a certain fraction of the variance can be forecast by linear models. When the oscillations are sustained, the

predictability limit should be pushed even further, as shown by the results of P1, P2 and P3. The results are fairly stable to a change in window length, since for  $M = 20$  (not shown), the behavior of the forecast errors is the same, with somewhat lesser skill.

For the IPCC data, forecast skill is established in a different way. For each year  $N_y$ , starting in 1950, the AR coefficients of the PCs are calculated with SSA from the training period (1861,  $N_y$ ). Then, the AR models for the PCs are used to perform a series of forecasts starting at years  $N_y$ ,  $N_y + 1, \dots, 1989$ . The forecasts are compared to the actual reconstructions, and the error is compared with the error made by the *climatological* forecast (CF hereafter), i.e., the extrapolation of the PCs with the value 0, and by the *persistence* forecast (PF hereafter), i.e., the extrapolation of the PCs with the last known value; CF and PF are standard benchmarks for skill in numerical weather prediction and its extensions.

The average forecast error at lead time  $\tau$  is given by the average of the quantity

$$e_F(N_y, \mathcal{A}, t, \tau) = [F_{\mathcal{A}}(N_y + t + \tau) - R_{\mathcal{A}}x(N_y + t + \tau)]^2 \quad (5.3)$$

with respect to the initial years  $N_y + t$ . The CF error is the average of the quantity

$$e_C(N_y, \mathcal{A}, t, \tau) = [R_{\mathcal{A}}x(N_y + t + \tau)]^2, \quad (5.4)$$

and the PF error is the average of

$$e_P(N_y, \mathcal{A}, t, \tau) = [R_{\mathcal{A}}x(N_y + t) - R_{\mathcal{A}}x(N_y + t + \tau)]^2. \quad (5.5)$$

Note that for the IPCC time series, we only have at our disposal the raw data to compare with, not the clean signal. This is why, in eqs. (5.3)–(5.5),  $R_{\mathcal{A}}x$  stands where  $R_{\mathcal{A}}y$  does in eqs. (5.1), (5.2). The forecast errors with respect to the clean signal should obviously be lower.

In table 3 appear the statistics on the ratio of the average forecast error to both the average CF and PF errors, for lead times of  $\tau = 1$ –10 years, and  $M = 40$ . Only  $N_y$  values ranging from 1950 to 1979 were considered, since after that, average errors are calculated over less than 10 values and lose their statistical significance. If, for year  $N_y$ , lead time  $\tau$  and subset  $\mathcal{A}$ , the average of the forecast error  $e_F$  is lower than the average CF error  $e_C$  (or average persistence forecast  $e_P$ ), the model is successful for those values of the parameters. In table 3, we count the number of successes of the PC-based AR model, as well as the number of failures. Italic characters stand for cases when the number of successes is larger than the number of failures. These numbers are estimated for the subsets  $\mathcal{A} = \{T^*, \dots, S\}$ , representing the trendless significant components, as well as for the subset  $\mathcal{A}$  including only the quasi-periodic trendless components ( $S$ - and  $Q$ -forecasts, respectively).

Table 3

The left/right element in each column is the number of initial years  $N_y$ ,  $1950 \leq N_y \leq 1979$ , having an average forecast error lower/higher than the climatological forecast (CLI) or the persistence forecast (PER). The results are calculated with  $M = 40$  for the IPCC series.  $S$  stands for forecasts of the significant components, i.e., from order  $T^*$  to  $S$ .  $Q$  stands for the quasi-periodic components, i.e., for the union of pairs lying in the interval  $(T^*, S)$ . The rows indicate the lead time  $\tau$ , in years. Italic characters indicate that the SSA-MEM forecast is more successful than the comparison forecast. The sum of the elements in each cell is the number of initial years, 30 for the  $S$ -forecasts and 29 only for the  $Q$ -forecasts – since for year 1970, no oscillatory pair was detected (see also fig. 11).

$\tau$	$S/CLI$	$S/PER$	$Q/CLI$	$Q/PER$
1	30/0	30/0	29/0	29/0
2	29/1	24/6	29/0	29/0
3	23/7	19/11	29/0	28/1
4	23/7	19/11	26/3	26/3
5	24/6	14/16	26/3	25/4
6	25/5	7/23	23/6	24/5
7	23/7	8/22	19/10	26/3
8	20/10	8/22	18/11	26/3
9	19/11	7/23	15/14	21/8
10	22/8	9/21	16/13	21/8

Our  $S$ -forecasts are better than CFs at all lead times and better than PFs up to 4 years. Thus PFs are harder to beat than CFs. Careful examination shows that the CFs are, in fact, bad forecasts, because the “detrended” time series has still a small positive trend toward the end. All temperature forecasts produced were slightly below the observed. Climatological forecasts, in particular, lie below the average of the end of the time series. Persistence forecasts are better since the last known value is a better estimate for a series with a (small) trend. Our forecasts decay eventually towards the average (of the past known values), since there is no noise forcing in the AR forecasts. The fact that they remain more successful than CFs indicates, however, that the sign of the anomalies is well forecasted. The phase of the oscillations is correctly estimated, whereas the amplitude is typically underestimated.

The  $Q$ -forecasts are better than CFs and PFs for all lead times considered, indicating that pairs do really correspond to linearly predictable phenomena. The price to pay is that  $Q$ -forecasts only predict – on average – about 15% of the total variance, 35% of the trendless variance, and 50% of the trendless and noiseless variance. Fig. 16a shows the particularly good forecast initiated from the year  $N_y = 1953$ , for the significant components 4–14. The forecast matches the series up to 1968, i.e., 15 years ahead, quite well.

Fig. 16b shows the global temperature forecast to the end of the century. Global temperatures should decrease by about 0.2°C up to 1995–1996 before increasing again. For a complete forecast, the trend has to be added back in. If the increase of temperature given by the trend does not change (about 0.08° for the last five years, cf. fig. 9), we should witness a decrease of about 0.12°C for 1995–1996. The global temperature would still be high compared with the beginning of the century, but would be close to the values for the early eighties.

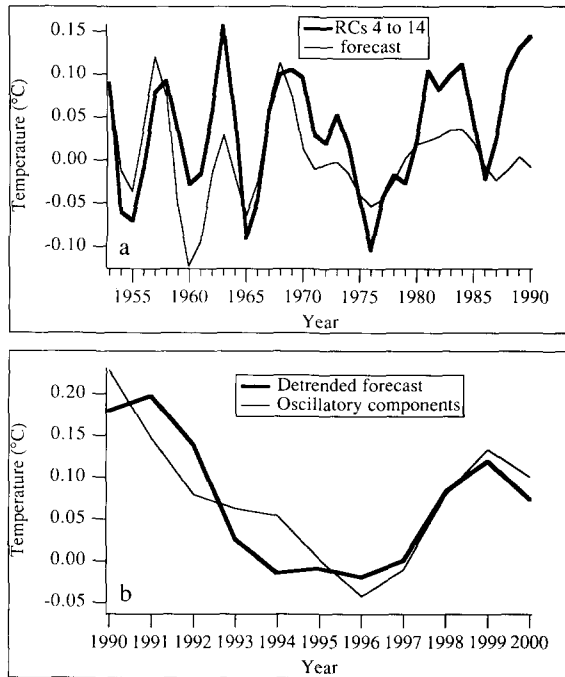


Fig. 16. Global forecasts of detrended temperature, from IPCC data. (a) Based on data prior to 1953 only, of the reconstructed series  $R_{\mathcal{A}}x$ , with  $\mathcal{A} = \{4, \dots, 14\}$ , together with the evolution of the actual quantity forecasted. (b) For the next ten years: heavy curve:  $\mathcal{A} = \{3, \dots, 18\}$ ; light curve:  $\mathcal{A} = \{3, \dots, 10, 16, 17\}$ .

## 6. Summary and discussion

We have reviewed theoretical and algorithmic properties of singular-spectrum analysis (SSA), from the signal-processing point of view. This method of analysis is particularly useful when little or no knowledge of the underlying dynamical system is available, and the time series at hand are short and noisy. An outline of the main results follows.

(1) Parameter ranges are obtained for applying SSA successfully to the analysis of short signals. A particular method of estimating of the Toeplitz matrix is shown to have little bias compared to other estimates. In order to correctly localize the spells of an oscillation of period  $T$  and typical life time  $T'$ , we suggest to choose the window length  $M$  so that  $T \leq M \leq T'$ .

(2) An algorithm for the reconstruction of isolated components, or groups of components, is presented. It allows one to expand the series  $x$  under study into a sum of reconstructed components  $x^k$ , corresponding to the  $k$ th eigenelement of the Toeplitz matrix, at any given epoch.

(3) We develop a method for the identification of noisy and of significant components. The reconstruction based on the significant components only gives good and statistically stable estimates of the clean signal, whether the noise is white (as in the synthetic examples) or not (as in the IPCC data). This noise-reduction algorithm is shown to be particularly efficient, given short data sets, for processes with a large quasi-periodic component. For purely chaotic processes perturbed by noise, the proposed noise-reduction algorithm involves substantial reduction in the signal as well.

(4) A method is derived to recognize oscillatory pairs of eigenelements. These pairs are shown to correspond to the main periodicities in the data. The periodicities do not necessarily represent sustained harmonic components, but indicate at least intermittent periodic activity in the time series under study. SSA recognizes systematically an oscillation in a short signal (we assumed the knowledge of 150 sample points only) perturbed by white noise with variance as large as twice the variance of the oscillation in question. We also show that SSA can be used for detrending purposes, even for other than linear trends.

(5) Two advanced spectral-analysis methods can be used to good advantage in combination with SSA. First, we derive a fully consistent SSA–maximum-entropy-method (MEM) for power spectrum estimates; this approach respects the additive property of SSA response filters, and allows one to stack spectral estimates of the different components, quantifying therein the spectral contribution of each eigenelement. Second, when the noise reduction and detrending algorithms are applied prior to multi-taper spectral estimation, the number of spurious

spectral peaks is significantly reduced. This eliminates a major drawback of the multi-taper method (MTM), which otherwise provides a significance test for the line components it detects and is robust to noise characteristics.

(6) The regular behavior of the principal components (PCs) makes them easier to forecast than the complete signal. We show that using autoregressive (AR) models for the significant components increases the predictability, especially in the case of quasi-periodic signals embedded in noise. For chaotic dynamical systems with no particular frequency excited, linear models do not perform very well. However, these systems may hide intermittent oscillations. Depending on the variance explained by the latter, at least a fraction of the total variance can be forecasted. For the IPCC temperature data, oscillatory components account for about 50% of the significant variance, after trend removal, and are predictable more than 10 years ahead.

We have focused in this paper on single-channel analysis. The same calculations can be carried out in multi-channel problems. Reconstruction of selected components is done essentially in the same way [19]. The noise-reduction problem can also be approached in a similar manner. The interpretation of the eigenelements is different, since in multi-channel analysis the EOFs are time sequences of “spatial” patterns [18, 62]. Therefore, when a pair of eigenelements emerges, it is usually associated with nearly linear traveling waves, as in the extended EOFs with fixed, small time window already used in meteorology [62, 63].

MEM spectral estimates can be computed using multi-channel AR models [19]. Linear forecasting of the PCs is performed in the same way as in single-channel analysis. An interesting climate modeling application of multi-channel SSA could be to use such simple linear models in order to forecast low-frequency quasi-periodic components, such as the El Niño–Southern Oscillation (ENSO), and use this forecast in combination with a general circulation model, which –

in many cases – does not reproduce such oscillations very accurately, but can provide other physical details of interest, related to shorter periodicities [64].

## Acknowledgements

It is a pleasure to thank our collaborators in the SSA enterprise: M. Dettinger, C.L. Keppenne, M. Kimoto, K.-C. Mo, M.C. Penland, G. Plaut, Y. Sezginer, and E. Toniolo. G.E. Birchfield, D. Broomhead, L. Smith, P. Switzer, H. Tong and many others asked interesting questions; we hope they will find some of the answers in here. C.K. Folland and D.E. Parker graciously provided the IPCC data set. We also benefited from constructive remarks by P. Manneville, two anonymous reviewers, and the editors of this special issue. The animated atmosphere of the IUTAM/NATO workshop at the University of Warwick was very stimulating. This work was supported at UCLA by NSF grant ATM 90-13217, at LMD by the CNRS, and by a Guggenheim Fellowship to M. Ghil. Computations were done at LMD on SUN Spark workstations.

## References

- [1] D. Brouwer and M. Clemence, *Methods of Celestial Mechanics* (Academic Press, New York/London, 1961).
- [2] E.N. Lorenz, *J. Atmos. Sci.* 20 (1963) 130.
- [3] E.N. Lorenz, *J. Atmos. Sci.* 26 (1969) 636.
- [4] M. Ghil and S. Childress, *Topics in Geophysical Fluid Dynamics: Atmospheric Dynamics, Dynamo Theory and Climate Dynamics* (Springer, Berlin, 1987).
- [5] S.G.H. Philander, *El Niño, La Niña, and the Southern Oscillation* (Academic Press, San Diego, 1990).
- [6] R.A. Madden and P.R. Julian, *J. Atmos. Sci.* 28 (1971) 702.
- [7] C.L. Keppenne and M. Ghil, Adaptive spectral analysis and prediction of the Southern Oscillation Index, in: *Proc. 15th Annual climate diagnostics workshop* (US Department of Commerce, NOAA/Climate Analysis Center, 1991), 30; see also Adaptive filtering and prediction of the Southern Oscillation Index, *J. Geophys. Res.*, sub judice (1992).
- [8] J. Eisenfeld and M. Witten, eds., *Modeling of Biomedical Systems*, *Proc. 11th Intern. Assoc. Math. Comput.*

- Simul. (IMACS) World Congress on Systems, Simulation and Scientific Computing (North-Holland, Amsterdam, 1986).
- [9] G. Mayer-Kress, ed., *Dimensions and Entropies in Chaotic Systems* (Springer, Berlin, 1986); L.A. Smith, *Phys. Lett. A* 113 (1988) 283; D. Ruelle, *Proc. R. Soc. London* 427A (1990) 241.
- [10] R. Kumaresan and D.W. Tufts, Data-adaptive principal component signal processing, in: *IEEE Proc. Conf. on Decision and Control* (Albuquerque, 1980) 949.
- [11] E.R. Pike, J.G. McWhirter, M. Bertero and C. de Mol, *IEEE Proc.* 131 (1984) 660.
- [12] J.M. Colebrook, *Oceanol. Acta* 1 (1978) 9.
- [13] D.S. Broomhead and G.P. King, *Physica D* 20 (1986) 217.
- [14] K. Fraedrich, *J. Atmos. Sci.* 43 (1986) 419.
- [15] R.W. Preisendorfer, *Principal Component Analysis in Meteorology and Oceanography*, ed. C.D. Mobley (Elsevier, Amsterdam, 1988).
- [16] G.H. Golub and C.F. Van Loan, *Matrix Computations* (Johns Hopkins Univ. Press, 1983).
- [17] D.S. Broomhead and G.P. King, On the qualitative analysis of experimental dynamical systems, in: *Non-linear Phenomena and Chaos*, ed. S. Sarkar (Adam Hilger, Bristol, 1986).
- [18] M. Kimoto, M. Ghil and K.-C. Mo, Spatial structure of the extratropical 40-day oscillation, in: *Proc. 8th Conf. on Atmospheric and Oceanic Waves and Stability* (American Meteorological Society, Boston, 1991) 17.
- [19] R. Vautard and G. Plaut, Low-frequency oscillations and weather regimes in the Northern Hemisphere, in: *Proc. ECMWF Workshop on Predictability* (European Centre for Medium Range Weather Forecasts, Reading, 1992) in press.
- [20] M. Loève, *Probability Theory*, 3rd ed. (Van Nostrand, Princeton, 1962).
- [21] R. Vautard and M. Ghil, *Physica D* 35 (1989) 395.
- [22] E.M. Rasmusson, X. Wang and C.F. Ropelewski, *J. Mar. Syst.* 1 (1990) 71.
- [23] M. Ghil and R. Vautard, *Nature* 350 (1991) 324.
- [24] M.R. Allen, P.L. Read and L.A. Smith, Bidecadal oscillations and the global temperature record, *Nature* 355 (1992) 686.
- [25] M. Ghil and K.-C. Mo, *J. Atmos. Sci.* 48 (1991a,b) 752, 780.
- [26] C. Penland, M. Ghil and K. Weickmann, *J. Geophys. Res.* 96 (1991) 22659.
- [27] J.D. Horel, *J. Clim. Appl. Meteor.* 23 (1984) 1660.
- [28] K. Hasselmann, *J. Geophys. Res.* 93 (1988) 11005; C. Penland, *Mon. Weather Rev.* 117 (1989) 2165.
- [29] M.G. Kendall and A. Stuart, *The Advanced Theory of Statistics* (Griffin, London, 1968).
- [30] J.D. Farmer and J.J. Sidorowich, *Physica D* 47 (1991) 373.
- [31] M. Casdagli, *Physica D* 35 (1989).
- [32] J.P. Burg, Maximum entropy spectral analysis, paper presented at the 37th Annual Meeting of the Soc. Explor. Geophys. (Oklahoma City, OK, 1967); also in *Modern Spectrum Analysis*, ed. D.G. Childers (IEEE Press, New York, 1978) 34.
- [33] G.U. Yule, *Phil. Trans. R. Soc. Lond. A* 226 (1927) 267.
- [34] G. Walker, *Proc. R. Soc. Lond. A* 131 (1931) 518.
- [35] D.J. Thomson, *Proc. IEEE* 70 (1982) 1055.
- [36] J.T. Houghton, G.J. Jenkins and J.T. Ephraums, eds., *Climate Change, The IPCC Scientific Assessment* (Cambridge Univ. Press, Cambridge, 1990).
- [37] P.D. Jones, T.M.L. Wigley and P.B. Wright, *Nature* 322 (1986) 430.
- [38] A. Blanc-Lapierre and B. Picinbono, *Fonctions Aléatoires* (Masson, Paris, 1981).
- [39] G.E.P. Box and G.M. Jenkins, *Time Series Analysis, Forecasting and Control* (Holden-Day, San Francisco, 1970).
- [40] G.M. Jenkins and D.G. Watts, *Spectral Analysis and its Applications* (Holden-Day, San Francisco, 1968).
- [41] T.J. Ulrych and T.N. Bishop, *Rev. Geophys. Space Phys.* 13 (1975) 183.
- [42] J.G. Elsner and A.A. Tsonis, *Nature* 353 (1991) 551.
- [43] J.F. Gibson, M. Casdagli, S.E. Eubank and J.D. Farmer, Principal component analysis and derivatives of time series, *Physica D*, submitted (1991).
- [44] M.I. Fortus, *Izv., Atmospheric and Oceanic Physics* 9 (1973) 34.
- [45] P.A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach* (Prentice-Hall, New York, 1982).
- [46] U. Grenander and G. Szegő, *Toeplitz Forms and Their Applications* (University of California Press, Berkeley, 1958).
- [47] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, *Numerical Recipes, The Art of Scientific Computing* (Cambridge Univ. Press, New York, 1989).
- [48] S. Ulam, On the Monte Carlo method, in: *Proc. 2nd Symp. on Large-Scale Digital Calculating Machinery* (1951) 207; R.M. Dole and N.D. Gordon, *Mon. Weather Rev.* 111 (1983) 1567; R.E. Livezey and W.Y. Chen, *Mon. Weather Rev.* 111 (1983) 46; R. Vautard, K.-C. Mo and M. Ghil, *J. Atmos. Sci.* 47 (1990) 1926.
- [49] J. Theiler, B. Galdrikian, A. Longtin, S. Eubank and J.D. Farmer, *Physica D* 58 (1992) 77, these Proceedings.
- [50] C. Kuo, C. Lindberg and D.J. Thomson, *Nature* 343 (1990) 709.
- [51] P. Pestiaux, I. van der Mersch, A. Berger and J.C. Duplessy, *Climatic Change* 12 (1988) 9.
- [52] P. Yiou, C. Genthon, M. Ghil, J. Jouzel, H. Le Treut, J.M. Barnola, C. Lorius and Y.N. Korotkevitch, *J. Geophys. Res.* 96 (1991) 20356.
- [53] P.D. Jones, Global temperature variations since 1861, in: *Long-Term Variability of Pelagic Fish Populations and their Environment*, eds. T. Kawasaki, S. Tanaka, Y. Toba and A. Taniguchi (Pergamon, Oxford, 1989).
- [54] D. Slepian, *Bell Syst. Tech. J.* 57 (1978) 1371.
- [55] L.J. Lanzerotti, D.J. Thomson, A. Meloni, L.V. Medford and C.G. MacLennan, *J. Geophys. Res.* 91 (1986) 7417.

- [56] C.R. Lindberg, Multiple Taper Spectral Analysis of Terrestrial Free Oscillations, Ph.D. Thesis (University of California, San Diego, Scripps Institution of Oceanography, 1986).
- [57] J. Park, C.R. Lindberg and F.L. Vernon III, *J. Geophys. Res.* 92 (1987) 14027.
- [58] D.J. Thomson, *Phil. Trans. R. Soc. Lond. A* 330 (1990) 601.
- [59] D.J. Thomson, *Proc. R. Soc. London A* 332 (1990) 539.
- [60] H. Le Treut, J. Portes, J. Jouzel and M. Ghil, *J. Geophys. Res.* 93 (1988) 9365.
- [61] H. Akaike, *Ann. Inst. Statist. Math.* 21 (1969) 243; H. Akaike, *IEEE Trans. Autom. Control* AC-19 (1974) 716.
- [62] K.C. Mo and M. Ghil, Interannual oscillations in the 700 Mb geopotential height field over the Northern Hemisphere, submitted (1992).
- [63] B.C. Weare and J.N. Nasstrom, *Mon. Wea. Rev.* 110 (1982) 481; K.-M. Lau and P.H. Chan, *Mon. Weather Rev.* 113 (1985) 1889.
- [64] J.D. Neelin, M. Latif, M.A.F. Allaart, M.A. Cane, U. Cubasch, W.L. Gates, P.R. Gent, M. Ghil, C. Gordon, N.C. Lau, C.R. Mechoso, G.A. Meehl, J.M. Oberhuber, S.G.H. Philander, P.S. Schopf, K.R. Sperber, A. Sterl, T. Tokioka, J. Tribbia and S.E. Zebiak, *Climate Dyn.* 7 (1992) 73.