# Monte Carlo SSA: detecting irregular oscillations in the presence of coloured noise

Myles R. Allen[1]
NOAA Postdoctoral Program in Climate and Global Change, Center for Meteorology and Physical Oceanography, Massachusetts Insititute of Technology, Cambridge, Massachusetts *and*
Space Science Department, Rutherford Appleton Laboratory, Chilton, Oxfordshire, United Kingdom

Leonard A. Smith
Mathematical Institute, University of Oxford, U.K.

*Journal of Climate*, **9**, 3373–3404, December 1996.

## Abstract

Singular Systems (or Singular Spectrum) Analysis, SSA, was originally proposed for noise reduction in the analysis of experimental data and is now becoming widely used to identify intermittent or modulated oscillations in geophysical and climatic time-series. Progress has been hindered by a lack of effective statistical tests to discriminate between potential oscillations and anything but the simplest form of noise, *i.e.*, "white" (independent, identically distributed) noise, in which power is independent of frequency. We show how the basic formalism of SSA provides a natural test for modulated oscillations against an arbitrary "coloured noise" null-hypothesis. This test, Monte Carlo SSA, is illustrated using synthetic data in three situations: (i) where we have prior knowledge of the power-spectral characteristics of the noise, a situation expected in some laboratory and engineering applications, or when the "noise" against which we are testing the data consists of the output of an independently-specified model, such as a climate model; (ii) where we are testing a simple hypothetical noise model, *viz.* that the data consists only of white or coloured noise; and (iii) where we are testing a composite hypothetical noise model, assuming some deterministic components have already been found in the data, such as a trend or annual cycle, and we wish to establish whether the remainder may be attributed to noise. We examine two historical temperature records and show that the strength of the evidence provided by SSA for interannual and interdecadal climate oscillations in such data has been considerably over-estimated. In contrast, multiple inter- and sub-annual oscillatory components are identified in an extended Southern Oscillation Index at a high significance level. We explore a number of variations on the Monte Carlo SSA algorithm, and note that it is readily applicable to multivariate series, covering standard Empirical Orthogonal Functions (EOFs) and Multi-channel SSA.

[1] *Corresponding author address:*
Dr. Myles R. Allen
Atmospheric, Oceanic and Planetary Physics, Clarendon Laboratory
Parks Road, Oxford OX1 3PU, United Kingdom
*fax:* 44 1865 272923 *email:* m.allen1@physics.oxford.ac.uk)

## 1 Background

Searching for evidence of predictability in observational time-series provides a starting point for many geophysical investigations since predictability indicates some degree of determinism in underlying system dynamics. In its simplest guise, this search consists in looking for trends and periodic oscillations. Predictability may be very limited in any nonlinear system, whether it is stochastically forced or purely deterministic; any trend may vary over time, as may the phase and amplitude of physical oscillations. Nonetheless, even limited predictability may be highly informative, making analysis techniques which allow the detection of intermittent trends or oscillations extremely valuable.

Singular Systems Analysis, SSA, is based on the idea of sliding a window down a time-series and looking for patterns which account for a high proportion of the variance in the views of the series thus obtained. SSA is closely related to the standard meteorological technique of Empirical Orthogonal Function (EOF) analysis (Lorenz, 1956; Kutzbach, 1967; Jolliffe, 1986). Those familiar with EOF analysis may think of single-channel SSA is as follows: if $M$ is the number of data-points in the window at any given time, the (overlapping) views of the scalar series form an $M$-variate vector series for which we obtain a complete set of $M$ orthonormal EOFs in the usual manner.

SSA was introduced into the study of dynamical systems by Broomhead & King, 1986a; Broomhead & King, 1986b, (BK) and Broomhead & Jones, 1989, as a method of visualising qualitative dynamics from noisy experimental data. Fraedrich, 1986, and Fraedrich & Ziehmann-Schlumbohm, 1994, observed that the algorithm could be used to estimate the number of degrees of freedom necessary to model the dynamics of an attractor (and thus an indication of the attractor dimension) and applied it to the analysis of atmospheric and palaeoclimatic time-series. Vautard & Ghil, 1989, (VG) also applied SSA to palaeoclimatic data, modifying the technique to exploit the assumption of stationarity and thereby increasing its noise-reduction power. VG emphasised the direct physical interpretation of the individual EOFs obtained with SSA, introducing the idea of searching for pairs of sinusoidal EOFs in quadrature which were taken to indicate a physical oscillation[1]. This latter application of SSA has since received considerable attention, particularly in the analysis of climate records including atmospheric angular momentum data (Penland *et al.*, 1991), the historical global temperature record (Ghil & Vautard, 1991; Elsner & Tsonis, 1991; Vautard *et al.*, 1992; Schlesinger & Ramankutty, 1994; Allen & Smith, 1994), and the Southern Oscillation Index (SOI) (Rasmusson *et al.*, 1990; Keppenne & Ghil, 1992; Elsner & Tsonis, 1994a).

The generalisation of SSA to multi-channel data, originally noted by BK, has been applied both to laboratory data (Read, 1993) and meteorological records (Keppenne & Ghil, 1993; Plaut & Vautard, 1994; Robertson *et al.*, 1995). The multi-channel SSA algorithm is mathematically identical to Extended EOF analysis (Weare & Nastrom, 1982; Graham *et al.*, 1987; Preisendorfer, 1988; Latif & Graham, 1992) or a special case of Combined Principle Component Analysis (Kutzbach, 1967; Bretherton *et al.*, 1992; Berkooz *et al.*, 1993). Many recent applications of multi-channel SSA have also emphasised identifying pairs of EOFs which characterise oscillations. Allen & Robertson, 1996, apply the basic test outlined here to the multi-channel problem; for simplicity, we restrict our discussion to the single-channel case.

A thorough exposition of single-channel SSA, emphasising its application to signal-detection problems in general and the detection of oscillations in particular, is given in Vautard *et al.*, 1992. Following the work of Allen, 1992, this paper will focus on those aspects of SSA which are relevant to the problem of distinguishing signals from noise.

## 2 Motivation

Discriminating between "signal" and "noise" is a crucial aspect of applied time series analysis. While the meaning of these two terms varies with context, there will always be a non-zero probability of incorrectly identifying noise as a deterministic trend or oscillation, given limited data. The acceptable probability of such a "false positive" or "type-1 error" must be specified, being the "nominal level" of any statistical test. If a test is misapplied, or prior assumptions concerning the properties of the noise are incorrect, then the true probability of a type-1 error, the "true level" of the test, may differ substantially from its nominal level, invalidating statements of statistical significance. This can lead, for example, to implausibly large numbers of spectral peaks being "detected" in limited data at very high nominal levels of significance. For a test to be accurate, its true level must be as close as possible to its nominal level. It

[1] SSA, like other linear analysis techniques, can only identify unstable periodic orbits in chaotic systems in those special cases where the the unstable orbit is observed in such a manner that it closely resembles one or a small number of sinusoidal oscillations.

is also important to ensure that the null-hypothesis is appropriate to the problem at hand, and not determined solely by the analysis technique employed. These sometimes conflicting aims are reflected in two slightly different approaches to Monte Carlo hypothesis testing in the analysis of nonlinear systems, jointly known as "the method of surrogate data": Theiler *et al.*, 1992, emphasise that the null-hypothesis should be well-understood while Smith, 1992, stresses that it must be physically interesting. Clearly, the ultimate goal is to attain both. This paper demonstrates how this goal can be achieved for the case of employing SSA to detect oscillations.

To illustrate the importance of an appropriate null-hypothesis in the context of geophysics, we consider the example of temperature anomalies, $T'(t)$, in a generic dissipative system with a finite heat capacity (Hasselmann, 1976; Allen *et al.*, 1994). If $Q'(t)\delta t$ is the anomalous heating or cooling due to external sources from time $t$ to $t+\delta t$, then

$$T'(t+\delta t) = T'(t)e^{-\frac{\delta t}{\tau}} + \frac{Q'(t)\delta t}{c}, \qquad (1)$$

where $\tau$ is the temperature relaxation time, and $c$ is the heat capacity. If $\tau$ and $c$ are both non-zero, $T'(t)$ will be positively autocorrelated in time ($\mathcal{E}(T'(t)T'(t+\delta t)) > 0$, where $\mathcal{E}$ is the expectation operator) unless $Q'(t)$ is strongly anti-correlated at lag $\delta t$, which is physically implausible for most natural heat sources. Thus equation (1) is inconsistent with the hypothesis that $T'(t)$ is an independent, identically distributed (i.i.d.) random variable (*i.e.,* white noise), the null-hypothesis considered in early significance tests proposed for the detection of oscillations with SSA.

If we know, on physical grounds, that a system could not appear to be white noise, then rejection of the white noise null-hypothesis provides us with no new information (Smith, 1992). In particular, it does not provide evidence for physical oscillations. For example, suppose $Q'(t)$ is pure white noise (unlikely in most geophysical systems but at least not physically impossible provided $\delta t$ is sufficiently large). The discrete form of equation (1), scaled such that $\delta t = 1$, then becomes the AR(1) model, which is widely used in time-series analysis (see Mardia *et al.*, 1979, and references therein):

$$u_t - u_0 = \gamma(u_{t-1} - u_0) + \alpha z_t, \qquad (2)$$

where $u_0$ is the process mean, $\gamma$ and $\alpha$ are process parameters, and $z_t$ is a Gaussian, unit-variance white noise. The output of an AR(1) process, or "AR(1) noise", supports no oscillations, although its power spectrum is biased towards low-frequencies; hence the common name "red noise". Its autocorrelation function decays exponentially, with an e-folding time of $\tau = \frac{-1}{\ln(\gamma)}$. A large class of geophysical processes produce output indistinguishable from noise of this type (Ghil & Childress, 1987; Zweirs & von Storch, 1995), so unlike white noise we can seldom reject the AR(1) noise null-hypothesis *a priori*. To be useful in geophysical applications, an analysis technique must consider the AR(1) null-hypothesis; if it is limited to the white noise null-hypothesis, it may falsely indicate large numbers of oscillations which are not significant if we allow the noise to be red (introducing one additional parameter into the noise model).

Monte Carlo SSA, introduced in section 4, tests for the presence of modulated oscillations against an arbitrary null-hypothesis. We describe in detail the application to testing against AR(1) noise, but the generalisation to higher-order AR and moving average (MA) processes is straightforward. While this allows us to avoid making null-hypotheses too simple, we should also guard against making them too complicated, since we clearly want to detect the presence of a genuine oscillation in the data. An AR process of order 2 or higher, for example, can itself support oscillations, which complicates the issue of what we learn by rejecting or failing to reject such a null-hypothesis. Since AR(1) noise itself has no preferred frequencies, it is a suitable null-hypothesis to test for oscillations.

The class of null-hypotheses is not restricted to linear stochastic processes. In climate research, for example, we could use the output of a climate model as our source of "noise", in which case Monte Carlo SSA ceases to be simply a method of detecting oscillations, but a method of detecting modes of variability in the observations which are inconsistent with the behaviour of that model. Given the current state of climate model development, however, it will generally be necessary to model at least some of the variability in a dataset as a stochastic residual term. If we are dealing with temperature data, the AR(1) process is the simplest appropriate model for these residuals.

Having decided on an AR(1) noise model, we then have to determine the process parameters $\gamma$ and $\alpha$. Ideally, we want to select those parameters which maximise the likelihood that we will fail to reject the null-hypothesis: rejecting a particular AR(1) noise null-hypothesis is uninteresting if we would have failed to reject another AR(1) noise null-hypothesis with a different choice of parameters. Parameter-specification becomes complicated when we have already identified a deterministic signal in the data (such as an annual cy-

cle) and are testing whether the residual is due to noise. These issues are discussed in sections 4.3 and 4.4.

In section 5 we discuss some fundamental problems with the use of a data-adaptive algorithm such as SSA for signal detection, particularly when we are dealing with short series. In section 5.2, we propose a revised approach to SSA which has the fundamental advantage for signal-detection applications that it brings the probability of a type-1 error closer to the nominal level of the test than is possible with standard SSA.

Examples of geophysical applications are given in section 6: we consider the detection of interannual and interdecadal oscillations in the historical record of global mean temperature (Folland *et al.*, 1992) and the Central England Temperature series (Manley, 1974; Parker *et al.*, 1991; Plaut *et al.*, 1995), and the detection of inter- and sub-annual oscillations in an extended Southern Oscillation Index (Jones, 1989).

# 3 Signals and noise in SSA

Like many techniques based on Singular Value Decomposition, SSA involves constructing a complete, orthonormal set of $M$ vectors, or EOFs, onto which we project a dataset. These EOFs define a coordinate system in an $M$-dimensional state-space, and the projections of consecutive segments of the data onto the EOFs (known as "EOF-coefficients" or "Principal Components") represent a distribution of points expressed in these coordinates. If the data are noisy, all $M$ dimensions will be required to describe this distribution completely, regardless of the coordinate system. The interesting aspects of its behaviour *may*, however, be confined to motion on a relatively low-dimensional subspace of this $M$-dimensional space. For example, the current state of a modulated oscillation may be described by projections onto only two vectors, a sine and a cosine with period identical to the oscillation, provided the time scales of amplitude- and/or phase-modulation are much longer than the window width.

For signal detection and reconstruction we wish to identify those EOFs (state-space directions) which are dominated by "signal", and eliminate those which are dominated by "noise". In the analysis of non-linear systems, it is important not to confuse the dimensionality of the linear subspace identified by standard (or "global" – Broomhead *et al.*, 1987) SSA, and the dimension of the underlying attractor, if any such attractor exists: a point made by BK and VG, and recently re-emphasised by Paluš & Dvořák, 1992. This article focusses on the direct application of SSA to signal-detection, and does not relate to any form of attractor dimension calculation.

## 3.1 The SSA algorithm

We present the SSA algorithm in detail to clarify the implications of various assumptions which lie behind it. For simplicity, we will confine our discussion to single-channel (univariate) SSA, although our remarks concerning signals and noise apply equally to the multi-channel case. We begin with a series $\mathbf{d}$ of length $N$, $d_{i:\ i=1,N}$, generated by a process which we assume to have zero mean. If we do not know the mean of the generating process, we can "center" the data series by removing its statistical mean, although this introduces some parameter-estimation complications, discussed in section 4.3. Throughout this article, the "process" generating $\mathbf{d}$ includes both the system under observation and the measurement procedure used to observe it; the series $\mathbf{d}$ represents all the available information. If we have independent knowledge of, for example, dynamical constraints and/or the properties of the noise, this information can and should be exploited in the analysis, but for brevity, we address these issues elsewhere Allen & Smith, 1996

We slide a window of width $M$ down the series to obtain an $N' \times M$ "trajectory matrix", $\mathbf{D}$, where the $i^{\text{th}}$ row of $\mathbf{D}$ corresponds to the $i^{\text{th}}$ "view" of $\mathbf{d}$ through the window. The total number of such views, $N'$, depends on how the ends of the series are treated. Two methods are widely used. *(i)* In the original algorithm of BK, the window stops as soon as it reaches the beginning or end of the data series, giving $N' = N - M + 1$ and $D_{ij} = d_{i+j-1}$. *(ii)* In the algorithm proposed by VG, we may think of the window "sliding off the ends" (Allen, 1992) giving $N' = N + M - 1$, and some elements missing in $\mathbf{D}$. These require special treatment, discussed below. The rows of $\mathbf{D}$ define $N'$ points on the system's trajectory in an $M$-dimensional state-space.

We define an $M \times M$ lag-covariance matrix:

$$\mathbf{C}_D \equiv \eta \mathbf{D}^T \mathbf{D}. \tag{3}$$

The normalisation constant $\eta$ is chosen such that, if the process is stationary, $\mathcal{E}(C_{ij}) \simeq c_{|i-j|} \equiv \mathcal{E}(d_i d_j)$, $c_l$ being the series' lag-$l$ auto-covariance. In the BK algorithm, there are no elements missing in $\mathbf{D}$, so calculation of $\mathbf{C}_D$ is straightforward, and $\eta = \frac{1}{N-M+1}$. In the VG algorithm, we calculate $\mathbf{D}^T\mathbf{D}$ by summing over all terms in which elements are defined in both $\mathbf{D}^T$ and $\mathbf{D}$, and divide by the number of terms in the sum, giving $C_{ij} = \eta \sum_{k=1}^{N-|i-j|} d_k d_{k+|i-j|}$, where $\eta = \frac{1}{N-|i-j|}$. Thus

the VG algorithm effectively assumes that the contribution to $\mathbf{C}_D$ from data beyond the series end-points may be estimated by the contributions to $\mathbf{C}_D$ from data within the series, an assumption which is only justified if the process which generated $\mathbf{d}$ is effectively stationary on the timescales spanned by the observations (to understand "effective stationarity", consider this example: non-stationarity on very long geological timescales would be irrelevant to the analysis of a 150-year temperature record; non-stationarity due to century-timescale changes in radiative forcing could, however, affect results).

Unless we specify otherwise, when we refer to a series' lag-covariance matrix such as $\mathbf{C}_D$, we mean the matrix which we calculate explicitly from that $N$-point series, using either the BK or VG algorithm. Our general remarks apply to both algorithms, although VG will be used in the examples presented here. We will also refer to the "expected lag-covariance matrix", $\mathcal{E}(\mathbf{C}_D)$, whose elements are the expected values of the elements of $\mathbf{C}_D$, given the algorithm (BK or VG) used to compute it. Both the BK and VG algorithms are biased, with the degree of bias depending on the length of the series available and the process under investigation, so that, in general, neither version of $\mathcal{E}(\mathbf{C}_D)$ is equal to the "process lag-covariance matrix" (the matrix obtained from an infinitely long realisation of the generating process).

VG arrive at their algorithm by a very different route. They argue that setting $C_{ij}$ equal to the most efficient and least biased estimator of the series covariance at lag $|i-j|$ gives the best possible estimate of the process lag-covariance matrix. We present this sliding-window-based interpretation to make the role of the stationarity assumption transparent since there has been some debate as to whether or not the VG algorithm is applicable to non-stationary processes (Allen, 1992; Dettinger *et al.*, 1995). Vautard *et al.*, 1992, remark that stationarity is still required in their original justification: lag-covariances may diverge for any infinitely long non-stationary series such that the process lag-covariance matrix is undefined. Even if the process is stationary, problems may still arise if $\mathbf{d}$ is too short to reflect that stationarity and the estimated lag-covariances bear no resemblance to their asymptotic values.

Applying the VG algorithm to observations of a non-stationary process can introduce significant biases in signal identification, reconstruction and prediction (Allen, 1992). For example, a reconstruction of a linear trend based on the VG algorithm will be biased towards the segment mean near the series endpoints. If the trend is positive, therefore, the estimated value of the trend at the end of the series will be systematically lower than its true value. The estimated value of the residual "stationary" component will be biased in the opposite sense.[2] Consequently, any short-term prediction of this stationary component will be subject to a significant negative bias.

Provided such biases are understood, the VG algorithm does provide significant additional noise reduction when applied to short series, making it often preferable for geophysical applications. Bias/variance trade-offs are common in statistical analysis: the VG estimate of $\mathbf{C}_D$ is subject to more bias but less variance than the BK estimate. By construction, it generates a Toeplitz lag-covariance matrix, whose eigenvectors are constrained to be either symmetric or anti-symmetric about the mid-point of the window. This proves useful for associating frequencies with EOFs.

Whichever method is used to compute $\mathbf{C}_D$, the next step in SSA is to diagonalise it and rank the eigenvalues in decreasing order:

$$\mathbf{\Lambda}_D \equiv \mathbf{E}_D^T \mathbf{C}_D \mathbf{E}_D, \tag{4}$$

where $\mathbf{\Lambda}_D$ is diagonal, the $k^{\text{th}}$diagonal element being the $k^{\text{th}}$ largest eigenvalue and the $k^{\text{th}}$ column of $\mathbf{E}_D$ being the corresponding eigenvector or EOF. Following standard practice, when we refer to a "high-ranked" EOF, we mean one whose corresponding eigenvalue lies early in the rank-order (*i.e.,* larger than most). Interpreting the results of this step is where the assumptions concerning noise in SSA play a crucial role.

First, consider the properties of $\mathbf{C}_D$ in a "pure deterministic signal" case where the system's trajectory is completely confined to an attractor embedded in an $m$-dimensional linear subspace of our $M$-dimensional state-space. Here $\mathbf{C}_D$ is identical to the lag-covariance matrix of the signal, $\mathbf{C}_S$; it will have $m$ non-zero eigenvalues, being the first $m$ diagonal elements of $\mathbf{\Lambda}_D$. These are the $m$ non-zero moments of inertia of the attractor, and their corresponding eigenvectors, the columns of $\mathbf{E}_D$, are the attractor's principal axes of inertia. Thus, in the pure-signal case, the eigenvectors of $\mathbf{C}_D$ with non-zero eigenvalues have a clear physical meaning: they define the linear subspace in which the attractor lies. As $N$ increases, these EOFs will converge to the true attractor principal axes, or "process" EOFs.

Second, consider a series contaminated with white noise, $\boldsymbol{\epsilon}$, where $\mathcal{E}(\epsilon_i \epsilon_j) = \sigma^2 \delta_{ij}$. The expected lag-

[2]It is for this reason that Vautard *et al.*, 1992, consistently observe a small positive trend towards the final points of their "detrended" global temperature series.

covariance matrix of the noise is $\mathcal{E}(\mathbf{C}_R) = \sigma^2\mathbf{I}$, where $\mathbf{I}$ is the rank-$M$ identity matrix. Since signal and noise are linearly independent, the expected lag-covariance matrix of the data series is:

$$\mathcal{E}(\mathbf{C}_D) = \mathcal{E}(\mathbf{C}_S) + \mathcal{E}(\mathbf{C}_R) = \mathcal{E}(\mathbf{C}_S) + \sigma^2\mathbf{I}. \quad (5)$$

Adding $\sigma^2\mathbf{I}$ to $\mathcal{E}(\mathbf{C}_S)$ simply increases all the eigenvalues of $\mathcal{E}(\mathbf{C}_D)$ by $\sigma^2$ without altering the eigenvectors. Thus, if the series consists only of signal and white noise, the $m$ highest-ranked EOFs of $\mathbf{C}_D$ still have a clear physical meaning. In the long-series limit, they converge onto the process EOFs of the signal, as before. If $\mathbf{C}_S$ has $m$ non-zero eigenvalues, and $m < M$, these will appear as $m$ eigenvalues of $\mathbf{C}_D$ lying above a flat "noise floor". The standard practice of "truncating the eigenspectrum", retaining only the $p$ highest-ranked eigenvalues and EOFs, is an effective method of separating signal from noise in this situation. If $p \geq m$, all of the signal variance will project onto these $p$ EOFs, and only a fraction, $p/M$, of the noise variance, since white noise projects equal variance onto all EOFs. Thus the signal-to-noise (S/N) ratio in these EOFs will have been enhanced by a factor of $M/p$. Estimation problems arise with short series, but even then, if we take the average of $\mathbf{C}_D$ for a large number of short series, it will converge to $\mathcal{E}(\mathbf{C}_S) + \sigma^2\mathbf{I}$. So even with a short series, the highest-ranked EOFs of $\mathbf{C}_D$ are estimates of physically meaningful entities.

Third, consider a series contaminated with coloured noise: *i.e.,* any noise process for which $\mathcal{E}(\mathbf{C}_R)$ is not a scalar multiple of $\mathbf{I}$. Signal and noise are still linearly independent, so $\mathcal{E}(\mathbf{C}_D) = \mathcal{E}(\mathbf{C}_S) + \mathcal{E}(\mathbf{C}_R)$. Even in the long-series limit, however, we can no longer expect the high-ranked eigenvectors of $\mathbf{C}_D$ to approximate to the eigenvectors of $\mathbf{C}_S$ except under very improbable circumstances (such as when the signal and noise have identical autocorrelation functions). If any component of the noise is not white, the eigenvectors of $\mathbf{C}_D$ will depend on $\mathbf{C}_S$, $\mathbf{C}_R$ and the signal-to-noise ratio. We stress this point because a number of signal-detection algorithms rely on truncation of the eigenspectrum to separate signal from noise and make direct use of the shape of the individual EOFs of $\mathbf{C}_D$ to identify oscillations. This is incorrect if any component of the noise may be red.

The generalisation of these points to multivariate observations is straightforward. For example, in conventional EOF analysis of a dataset consisting of $L$ "spatial" channels each of length $N$, with $L > N$, we effectively obtain the principal components (PCs) by diagonalising an estimated temporal lag-covariance matrix, $\mathbf{C}_D$, exactly as in SSA. The difference is only that $\mathbf{C}_D$ is now of rank $N$, and is estimated by summing over the spatial channels, rather than summing over "views" through a sliding window. The correspondence is even closer in the case of extended-EOFs, or multi-channel SSA. The PCs obtained from multi-channel SSA with a window width of $M$, with $L \times M > N - M + 1$, are the eigenvectors of a covariance matrix, $\mathbf{C}_D$, of rank $N - M + 1$. $\mathbf{C}_D$ is exactly equivalent to the covariance matrix which we would obtain by performing single-channel SSA, with a window-width of $N-M+1$, on each spatial channel individually and averaging the results (Allen & Robertson, 1996). If $L \gg N$, the highest-ranked PCs of a pure noise process which is AR(1) in time, like the EOFs of single-channel SSA, will have a well-defined sinusoidal appearance that will not vary appreciably between realisations (Bretherton *et al.*, 1992). In both conventional EOF analysis and multi-channel SSA, therefore, temporal autocorrelation in the noise can render eigenvalue rank-order meaningless as a significance criterion, and lead to principal components that are entirely due to noise masquerading as high-variance, low-frequency oscillations. Improved methods of discriminating between signals and noise in multi-channel problems are urgently required. Allen & Robertson, 1996, represents an initial step in this direction, and further work is in progress.

### 3.2 Problems with the interpretation of SSA

There is a growing body of literature making use of SSA which, for reasons given below, takes the occurrence of a pair of sinusoidal EOFs with high-ranked eigenvalues as *prima facie* evidence of a physically meaningful oscillation. The preceding section shows that this is unjustified. First, position in the eigenvalue rank-order is only effective in separating signals from pure white noise, which is seldom encountered in geophysics. If the noise is red, rank-order is unreliable. Second, for anything other than white noise contamination, EOF-shapes depend as much on the properties of the noise as they depend on the properties of the signal, so searching for sine-cosine EOF pairs may also be misleading.

The assumption that significance decreases with position in the eigenvalue rank-order is widely used not only in SSA but also in conventional EOF analysis and numerous other techniques based on matrix-decomposition. It is based on the assumption that variance (or power) quantitatively reflects physical significance, which is simply not true for systems contaminated with coloured noise, nor for nonlinear systems in general. This point

has practical consequences for signal detection: Vautard *et al.*, 1992, propose a method of estimating the "statistical dimension" of a dataset, $\upsilon_L$. They compute a filtered reconstruction from only the lowest-ranked $M-p$ EOFs and assess whether the autocorrelation function of this reconstruction is consistent with that of a white noise process that is filtered similarly (the interested reader may wish to consider the related arguments of Theiler & Eubank, 1993; Smith, 1994). The smallest value of $p$ for which the white noise hypothesis cannot be rejected is identified with $\upsilon_L$, and EOFs 1 to $\upsilon_L$ are taken to indicate signal. Applied to a pure AR(1) process, the Vautard *et al.*, 1992, test gives $\upsilon_L \propto M$, with the constant of proportionality dependent on the lag-1 autocorrelation of the noise: *i.e.,* some of the noise is consistently indicated as signal.

Direct generalisation of the $\upsilon_L$ algorithm to test against red noise proved difficult (*Vautard*, pers. com.) primarily because, having established that a projection of the data onto EOFs $p+1$ to $M$ is inconsistent with the null-hypothesis, the test does not indicate which EOFs contribute most to the inconsistency. As demonstrated below, we must routinely pick out "signal" EOFs whose eigenvalues are ranked below those of other EOFs which are attributable to red noise. Even when the noise is white, eigenvalue rank-order is misleading when variance has been artificially suppressed at certain frequencies through, for example, the removal of an annual cycle. In this case, the lowest-ranked EOFs will be inconsistent with the null-hypothesis, as they will contain anomalously low variance. Penland & Sardeshmukh, 1995, make a similar point concerning eigenvalue rank-order noting that the highest-ranked complex eigenvalues obtained from Principal Oscillation Pattern (or Empirical Normal Mode) analysis of equatorial Pacific sea-surface temperatures (SSTs) are not the most statistically stable.

Various tests have been proposed (e.g. North *et al.*, 1982; Ghil & Mo, 1991) to assess the stability of an eigenspectrum to sampling uncertainty. These should not be confused with tests for identifying signals in noise. Given sufficient data, the eigenspectrum of a pure red noise process will be arbitrarily stable, so every eigenvalue will appear "significantly" different from its neighbours.

Interpreting the shapes of individual EOFs in SSA is also problematic as a method of signal-identification. VG, for example, note that a pure sinusoidal oscillation will give, in the long-series limit, a lag-covariance matrix whose rows and columns consist of lagged sinusoids with the same period of the oscillation (recall that the autocorrelation function of a sinusoid is itself sinusoidal). This will have two non-zero eigenvalues, whose associated EOFs make up a "sine-cosine pair". If we add white noise to the oscillation, this EOF-pair appears (in the long-series limit) as two degenerate – nearly equal – eigenvalues above an otherwise flat noise floor. VG also note that an infinite realisation of pure AR(1) noise will also give sinusoidal EOFs but they appear alternately symmetric and anti-symmetric with frequencies separated by $1/2M$. This led to the proposal of "pair selection criteria" to identify oscillatory EOF-pairs against an AR(1) noise null-hypothesis. A pair of EOFs were taken to indicate an oscillation when their associated frequencies were separated by less than $0.75/2M$ and together they explained more than $2/3^{rds}$ of the variance in the series at some intervening frequency (Vautard *et al.*, 1992).

In general, however, the eigenbasis of the sum of two covariance matrices does not contain eigenvectors which are the same as, or even similar to, the eigenvectors of *either* of the two constituent matrices (recall the axes-of-inertia analogy). Even in the long-series limit, signals cannot be identified using EOF selection criteria based on the expected properties of the pure-signal and pure-noise EOFs. For example, Allen, 1992, shows that the presence of a trend in the historical record of global mean temperatures effectively forces the appearance of an EOF-pair indicating an interdecadal oscillation, passing these pair-selection criteria. The eigenbasis of $\mathcal{E}(\mathbf{C}_D)$ for a process consisting of a trend plus a segment of AR(1) noise contains a pair of EOFs (nos. 3 and 4, immediately following the pair corresponding to the trend) which, on the above criteria, indicate an oscillation with period $\sim 2M/3$. For a 40-year window, this corresponds to 26–27 years.

EOFs will only fail to form sine-cosine pairs in an infinitely long segment of pure red noise. Any departure from this ideal situation can cause spurious pairs to appear: through only a short series being available, through the presence of a trend or through the intentional suppression of an annual cycle.

A complementary problem with interpreting EOF shapes is that perfectly genuine signals may fail to show up as "clean" oscillatory pairs if they happen to be degenerate with other components of the series (*i.e.,* have nearly equal eigenvalues). For example Allen, 1992, notes that if we analyse the last 90 years of the historical global temperature record using the BK algorithm, we find that a low-frequency component of the El Niño / Southern Oscillation (ENSO) signal is degenerate with the interdecadal component (which turns out to be in-

distinguishable from red noise) and another noise component. The SSA eigenspectrum contains two degenerate triplets of nearly equal eigenvalues. Thus this eigendecomposition is under-determined to rotations within the subspaces defined by these EOF-triplets. Any linear superposition of these EOFs is an equally valid decomposition. Rotations which do not affect the eigenspectrum can cause the interdecadal and low-frequency-ENSO components to appear and disappear arbitrarily, a situation which led Tsonis & Elsner, 1992, to conclude that results from applying the BK algorithm to this series were inconsistent with those obtained from applying the VG algorithm (Allen *et al.*, 1992b; Allen *et al.*, 1992a).

Regardless of whether or not sinusoidal EOF-pairs are attributable to noise, they are extremely effective narrow-band-pass filters. If a component of noise has been mistakenly identified as an oscillatory EOF-pair, including that pair in an SSA-based "noise reducing" filter applied to the data series will bias any subsequent analysis. Likewise, a broadband signal will appear as a set of EOF-pairs with associated frequencies separated by $1/M$. After SSA-based filtering, these will appear as discrete spectral features in any "stack spectrum" (Penland *et al.*, 1991).

In conclusion, we stress that if pair-selection criteria are to be used, alone or in conjunction with other tests, it is essential that an an end-to-end check is performed to evaluate the true significance level of the composite algorithm. Evaluating this "probability of a type-1 error" requires nothing more than a Monte Carlo experiment, but setting up such experiments can become complicated when multiple tests are involved.

Monte Carlo SSA is an attempt to reduce the signal-detection component of SSA to its bare essentials. It allows a statistical test that is as simple as possible while still satisfying the two requirements mentioned above: (i) the true level of the test should be easy to quantify and close to its nominal level and (ii) the test should not impose a physically inappropriate null-hypothesis on the analysis.

# 4 Monte Carlo Singular Systems Analysis

In this section, we describe the Monte Carlo SSA algorithm and demonstrate its application to a simple test series. The basic idea was originally proposed in Broomhead & King, 1986a, but, to our knowledge, it was not implemented prior to the work of Allen, 1992. The complexity of the test procedure depends on how much prior knowledge we have of the properties of the noise. We begin by assuming that all the noise parameters are known *a priori*. While unrealistic in geophysical problems, this may occur in laboratory or engineering applications. We will use the VG algorithm throughout for the computation of covariance matrices since this is the most widely used in climate research.

## 4.1 Case 1: All noise parameters known *a priori*

Since one of the key advantages of SSA over conventional Fourier methods is its ability to detect amplitude- and phase-modulated oscillations, we use a randomly-triggered oscillatory burst as a test case. The dotted line in figure 1 represents the "signal", generated as follows: at each time-step, there is a 0.5% chance of initiating a damped sinusoidal oscillation with a period of 5.5 units, randomly chosen initial phase and unit initial amplitude, which then decays with an e-folding time of 30 units. Two such bursts occur in this 200-point realisation giving a total signal variance of $0.077(\text{units})^2$. We then add a zero-mean segment of AR(1) noise, generated using equation (2), with unit variance and a lag-1 autocorrelation of 0.72 (corresponding to an e-folding time $\tau$ of 3 units) to give the solid line. This solid line provides our test series in which we now attempt to detect the modulated oscillation.

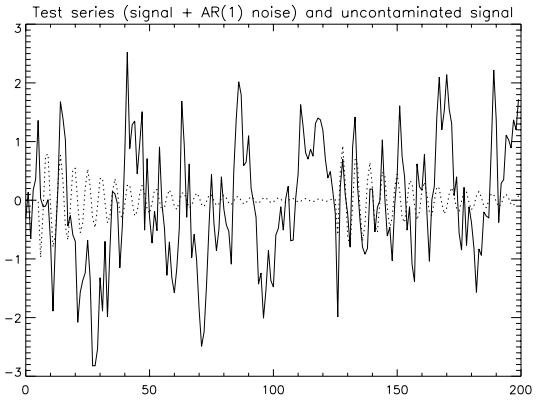


**Figure 1.** Test series, consisting of randomly-generated damped sinusoidal bursts, with unit initial amplitude, random initial phase, period of 5.5 units and e-folding time of 30 units, contaminated with red noise. Dotted line shows the uncontaminated signal, solid line shows the series obtained after adding AR(1) noise with zero mean, unit variance and autocorrelation decay time of 3 units.

The standard periodogram, applying a simple triangular (Bartlett, 1950) lag window (figure 2, upper panel), contains a peak corresponding to a period of

~5.5 units, but several other frequencies are also indicated as significant against a null-hypothesis of AR(1) noise and the variance of the spectral estimate is high. A more stable, but also more biased, spectral estimate can be obtained from the Fourier transform of the series autocovariance function (Blackman & Tukey, 1959) estimated out to $M-1$ lags, where $M \ll N$ – see figure 2 and the appendix. SSA provides a method of retaining the stability of the spectral estimate based on the autocovariance function using a data-adaptive basis in place of the rank-$M$ Fourier basis to reduce bias.

An alternative approach to reducing spectral variance while minimising bias is the multi-taper method (MTM) of spectral analysis (Thomson, 1982; Park *et al.*, 1987; Yiou *et al.*, 1991; Mann & Lees, 1995). MTM, applied to this test series, indicates a significant peak at the correct frequency, but also indicates a number of other peaks as significant (Yiou, *pers. com.*), consistent with the results of Vautard *et al.*, 1992.

If we apply standard SSA with $M = 40$ and plot the the eigenvalues of $\mathbf{C}_D$ in the conventional "rank-order" (*i.e.,* in order of decreasing size), the result is completely uninformative (the squares and diamonds in figure 3 – we use the two symbols to differentiate between EOFs which are symmetric and anti-symmetric about the mid-point of the window). Breaks appear in the eigenspectrum after EOFs 2, 4, 7 and 11, none of which have anything to do with the signal. A search for symmetric/anti-symmetric "oscillatory pairs" using the pair-selection criteria of Vautard *et al.*, 1992, suggests that EOFs 8 and 9 indicate a 5.5-unit oscillation, but also identifies EOFs 1 and 2 (period 30 units), 3 and 4 (trend), 5 and 6 (period 10 units) and a large number of other pairs further down the spectrum.

The shape of EOFs 8 and 9 is not in itself unusual, nor is the absolute amount of variance which they account for in the data series. What is remarkable about them is the variance which they account for given their shape (or, equivalently, given the direction they point to in state-space). For each EOF, a vertical bar in figure 3 indicates the power we expect to find in the state-space-direction defined by this EOF when analysing a segment of pure AR(1) noise. These "surrogate data bars" are obtained as follows:

Assuming we know the parameters in the AR(1) noise null hypothesis *a priori*, we generate an ensemble of 200-point "surrogate data" realisations (Smith, 1992; Theiler *et al.*, 1992), using the AR(1) model of equation (2) and the same parameters used to generate the noise in our test series. We compute a lag-covariance matrix, $\mathbf{C}_R$, for each surrogate realisation,

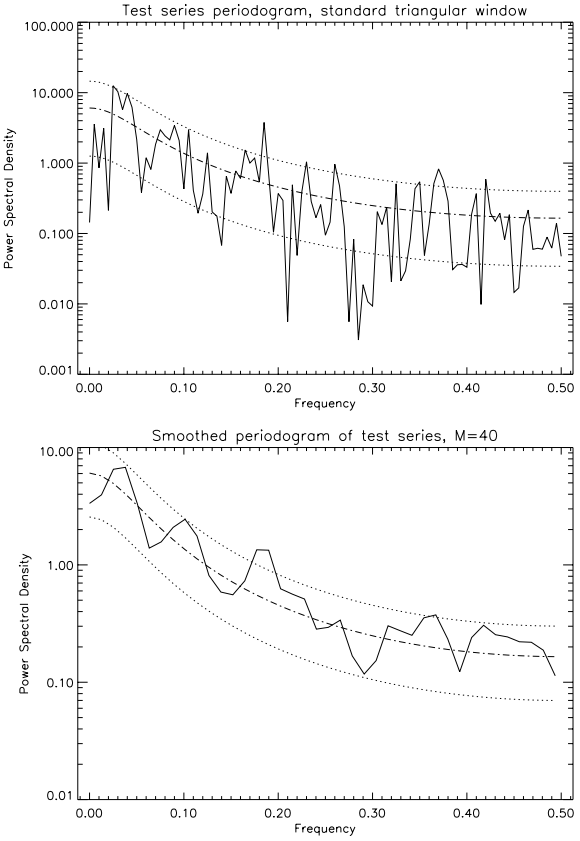


**Figure 2.** Power spectrum of test series (modulated oscillations plus AR(1) noise) using a triangular lag window applied to the full 200-point series (upper panel) and to the series autocovariance function evaluated from $-(M-1)$ to $M-1$ with $M = 40$ (lower panel). Dotted lines indicate the $2.5^{\text{th}}$ and $97.5^{\text{th}}$ percentiles for the spectral estimate of an AR(1) process with parameters equal to the noise in the test series – see Appendix.

using whichever (VG or BK) algorithm was used to compute $\mathbf{C}_D$. Since the process mean is known in this example, neither data nor surrogates are centered before computing $\mathbf{C}_D$ or $\mathbf{C}_R$. We project each surrogate realisation onto the EOFs of the data, defining the projection onto EOF-$k$ as the $k^{\text{th}}$ diagonal element of $\mathbf{\Lambda}_R$ in

$$\mathbf{\Lambda}_R \equiv \mathbf{E}_D^T \mathbf{C}_R \mathbf{E}_D. \tag{6}$$

The extrema of the vertical bars in figure 3 indicate the $2.5^{\text{th}}$ and $97.5^{\text{th}}$ percentiles of the diagonal elements of $\mathbf{\Lambda}_R$ corresponding to the EOFs whose eigenvalues they over-lie: 95% of the surrogate realisations lie within those limits.

We present the Monte Carlo approach to estimating the distribution of the diagonal elements of $\mathbf{\Lambda}_R$ since it

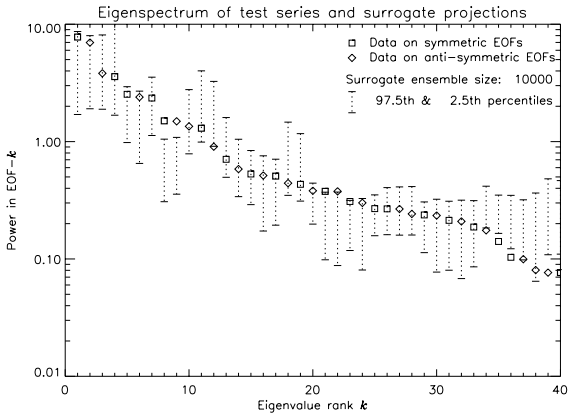


**Figure 3.** Eigenvalues of $\mathbf{C}_D$, rank $M = 40$, from the test series plotted in the conventional "rank order" (squares and diamonds). Breaks occur after EOFs 2, 4, 7 and 11, none of which correspond to a physical signal. Vertical bars show the variance we should expect in the directions defined by these EOFs in a segment of AR(1) noise. EOFs 8 and 9 of the data series contain more variance than expected on this null-hypothesis.

is applicable to a wide range of null-hypotheses. In the case of normally-distributed processes, such as AR(1) noise, and EOFs which are approximately sinusoidal, it is possible to compute these distributions analytically, avoiding the Monte Carlo step. Details are given in the appendix.

The test indicates EOFs 8 and 9 contain more power in the data series than we would expect on this null-hypothesis. The $8^{\text{th}}$ and $9^{\text{th}}$ elements of the $\mathbf{\Lambda}_R$ are greater than the corresponding elements of $\mathbf{\Lambda}_D$ in fewer than 0.1% of members of the surrogate ensemble, indicating that they are individually significant at the 99.9% level. EOFs 1 to 7, in contrast, contain only the variance which we would expect them to contain given their shape: even though they each contain more variance than EOFs 8 and 9, they all lie well within the surrogate data bars. EOFs 21 and 22 are also picked out at a slightly lower significance level (98%). These correspond to the first harmonic of our 5.5-unit oscillation which we are not surprised to find contains anomalous power given the rapid amplitude-modulation.

If the BK algorithm is used, equation (6) is equivalent to sliding the window down each surrogate realisation and summing the squared projections onto each EOF. This makes it clear how the shape of each EOF has been used in the test without our needing to describe it explicitly. If the EOF is dominated by small-scale structure (high associated frequencies), we expect the squared projections to be smaller than if it has only large-scale structure (low associated frequencies): red noise projects more variance onto larger scales. If we were testing against white noise, we would expect all the surrogate data bars to be at the same level because white noise projects almost exactly the same amount of variance onto each EOF.

An alternative way of displaying the information in figure 3 is to plot the data eigenvalues and surrogate data bars against the dominant frequency associated with their corresponding EOFs, as shown in figure 4. Since EOFs obtained with SSA are not pure sinusoids, identifying a single frequency with an EOF is problematic. For display purposes, we associate a frequency with an EOF simply by maximising the squared correlation with a sinusoid. This is essentially equivalent to the Reduced Fourier Transform of Vautard *et al.*, 1992, but it avoids the bias due to the variation of the Fejer kernel across a finite window (MacDonald, 1989; Allen, 1992). We plot in bold the eigenvalues corresponding to EOFs in which this maximum squared correlation is greater then 0.85: these are relatively "clean" sinusoids. Such a cutoff is inevitably arbitrary but is useful for the display of data. We stress that the surrogate data test itself does not require this association of EOFs with any single frequency.

A clean sine-cosine "oscillatory" EOF-pair would appear as a bold square and diamond almost superimposed on each other. Power falls off with increasing frequency in figure 4, as expected from red noise, but EOFs 8 and 9 form a pair, centered on 0.18 cycles per time-unit, which stands out from its neighbours in associated-frequency-space, indicating a period of 5.5 units. A simple visual inspection of the squares and diamonds in figure 4 is not, however, sufficient to identify physical oscillations. For example, the two highest and two lowest ranked EOFs also appear to form pairs centered on 0.03 and 0.29 cycles per time-unit: a characteristic of the data-adaptive basis of conventional SSA is that both high-ranked and low-ranked EOFs tend to "pair up", even in a segment of pure noise.

A number of authors have cited the stability of an EOF-pair to varying the window width $M$ as evidence for the significance of the corresponding oscillation. Such stability, however, does not distinguish a physically significant signal. Any finite segment of red noise will contain more power than the process average at some frequencies due simply to statistical fluctuations. EOF-pairs will tend to occur at these frequencies over a range of window widths. Figure 5 shows the effect of varying the window width: our test series eigenvalues and surrogate data bars are plotted against dominant

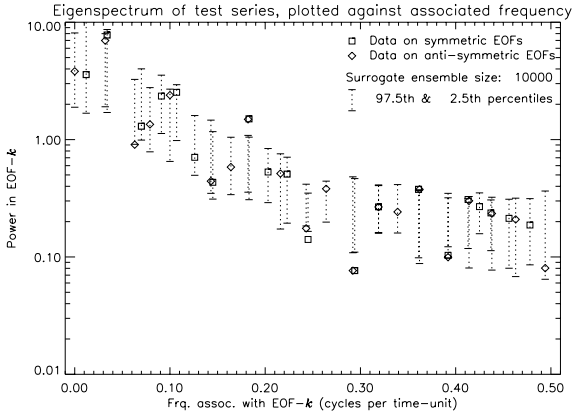


**Figure 4.** Eigenvalues of $\mathbf{C}_D$ from the test series, plotted against the dominant frequency associated with their corresponding EOFs. Clean sine-cosine pairs (bold squares and diamonds almost on top of each other) occur at frequencies of 0.18 and 0.36 $(\text{units})^{-1}$ (periods 5.5 and 2.8 units). The surrogate data test indicates these are significant at the 99.9% and 98% levels respectively. Note how the two highest-variance EOFs also appear to form a pair, indicating a frequency of 0.03 $(\text{units})^{-1}$. These would pass any selection criteria based on EOF shape but the surrogate data test shows that they do not contain more variance than we would expect from a segment of AR(1) noise.

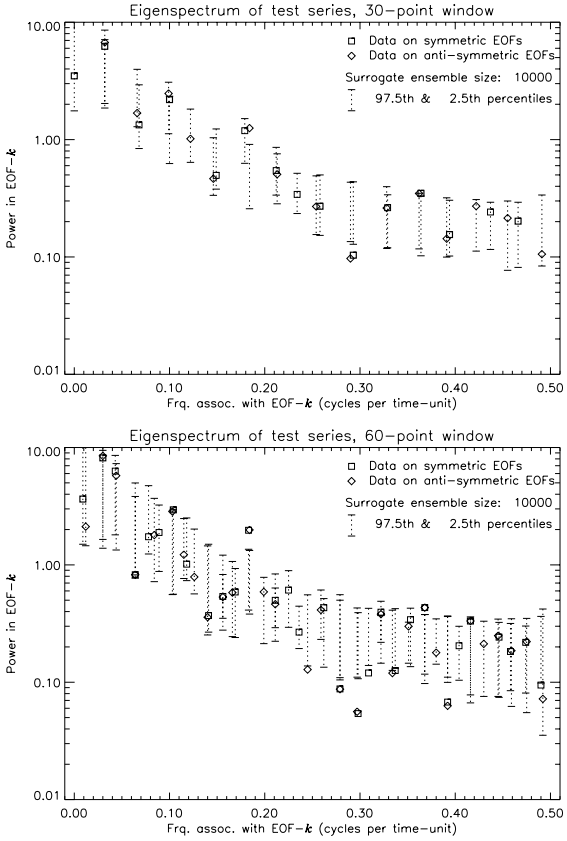


**Figure 5.** As previous figure but with $M = 30$ (upper) and $M = 60$ (lower). The 5.5-unit oscillation is still identified as significant, although with $M = 30$, one member or the pair lies inside the surrogate data bar due to a degeneracy problem (see text). Note that EOFs 1 and 2, without the surrogate data test, would still indicate a period-30 oscillation. This spurious "oscillation" is robust to varying the window width, illustrating the need for quantitative significance tests.

associated frequency for $M = 30$ and $M = 60$. Note that a pair suggesting an oscillation with period 30 units appears as EOFs 1 and 2 in all cases but is consistently rejected (correctly) by the surrogate data test.

Some sensitivity of results to window width is inevitable due to the constraint that EOFs must be mutually orthogonal, but if a pair only appears for certain values of $M$, this gives us reason to doubt its significance (although it might also be a consequence of two, perfectly genuine, signals being degenerate with each other). Unfortunately, the converse is not true. Only in the limit of an infinite segment of red noise will the frequencies associated with EOFs scale exactly with $1/M$.

The surrogate data test indicates anomalously high variance at 5.5-unit periods for both $M = 30$ and $M = 60$. With $M = 30$, however, only the anti-symmetric EOF (the diamond) is indicated as clearly significant. The reason is that the symmetric EOF, number 9, is degenerate with a lower-frequency component of the noise (EOF 7). The SSA decomposition is thus under-determined, and the result is that neither EOF 7 nor EOF 9 is sinusoidal, each containing a mixture of power at 5.5 and 15 unit periods. In this case, SSA has failed to isolate the symmetric component of the 5.5-unit oscillation. A revised algorithm, described in Allen & Smith, 1996, performs better.

### 4.2 Interpretation of Monte Carlo SSA spectra

Merely observing a pair of data eigenvalues lying above the 97.5$^{\text{th}}$ percentiles of the corresponding surrogate distributions is generally not enough to conclude that we have detected an oscillation at that frequency at the 97.5% significance level. Even if we are analysing a segment of pure noise, the average number of excursions above the 97.5$^{\text{th}}$ percentile will be $0.025M$ by construction. The correct interpretation of figure 3 therefore depends on our prior knowledge and expectations.

If we know beforehand that it is EOF-$k$ that we are interested in, then the position of the $k^{\text{th}}$ eigenvalue of

$\mathbf{C}_D$ relative to the corresponding surrogate data distribution translates straightforwardly into the significance level of the test (additional complications relating to the choice of EOF basis are discussed in section 5 below). Often, however, we use the results of such spectra to decide which EOFs to focus on. If we simply look for *any* excursions above the 97.5$^{\text{th}}$ percentiles of the surrogate distributions with a window width of 40, then we are, in fact, performing 40 "mini-tests". The probability of at least one excursion above the 97.5$^{\text{th}}$ percentiles of the surrogate distributions is clearly greater than 2.5%. This is a standard problem in power spectral analysis (MacDonald, 1989; Thomson, 1990).

Since the 40 "mini-tests" are not mutually independent, the probability of a given number of excursions does not, in general, conform to an analytically-calculable distribution. Livezey & Chen, 1982, address a similar problem in evaluating the statistical significance of relationships between continuous fields where independence of data at different locations and times cannot be assumed. We adopt a similar two-pass Monte Carlo approach to estimate probabilities directly. By storing the diagonal elements of $\mathbf{\Lambda}_R$ for each surrogate realisation and making a second pass through the ensemble after computing its distribution statistics, we estimate the probability of a given number of excursions above a pre-determined percentile directly from the relative frequency of such an event occurring in a member of the ensemble.

The probability of there being at least 2 (4) excursions above the 99.5$^{\text{th}}$ (97.5$^{\text{th}}$) percentiles, as observed in figure 3, in any given member of the surrogate ensemble is 5.4% (7.1%). So, if we did not specify beforehand that EOFs 8 and 9 were of interest, then the true confidence level at which we reject this AR(1) noise null-hypothesis is ∼94%. This is a lower limit, since the data eigenvalues 8 and 9 in fact lie in the 99.9$^{\text{th}}$ percentiles, and we have only quantified the probability of excursions above the 99.5$^{\text{th}}$ percentile.

Plotting against dominant associated frequency (figure 4) might incline us to reject the AR(1) null hypothesis for reasons over and above the simple occurrence of 2 (4) excursions above the 99.5$^{\text{th}}$ (97.5$^{\text{th}}$) percentiles: the fact that the excursions appear in two pairs, one of which corresponds to double the frequency of the other, clearly suggests a modulated oscillation. These qualitative observations might form the basis of a more stringent statistical test, but we should proceed cautiously in this direction, since there are clear dangers in tailoring a test too specifically towards an expected result (or worse still, towards a result which has already been obtained).

Increasing the window width $M$ increases the spectral resolution of SSA and also increases the potential signal-to-noise (S/N) enhancement, but it also increases the number of individual excursions above a given confidence level which we should expect to occur purely by chance, simply by increasing the effective number of "mini-tests" performed: the average number of excursions scales linearly with $M$. If, therefore, only a small number of EOFs are indicated as individually significant, it is essential that that the two-pass test is performed to quantify the probability of that number of excursions occurring by chance. The advantage of the procedure presented here (which is equally applicable to the periodograms discussed by MacDonald, 1989) is that it is applicable to null-hypotheses whose statistics are highly non-Gaussian, such as those generated by chaotic systems.

Even with a relatively short window ($M = 40$), the probability of at least 2 excursions above the 99.5$^{\text{th}}$ percentile occurring by chance is not negligible. A visual inspection of the cases in which this occurs indicates that there is a high probability of such an excursion "looking like" an oscillatory pair. The reason is simple: if a series contains above-average power in one sinusoidal EOF, then, in the long-series limit, it necessarily contains above-average power in the same sinusoid, phase-shifted by $\pi/2$ (*i.e.,* in the other member of the "pair"). In this example, the probability of 2 or more chance excursions above the 97.5$^{\text{th}}$ percentile is 28%. So if EOFs 8 and 9 were to lie in the 97.5$^{\text{th}}$ (as opposed to 99.5$^{\text{th}}$) percentile, the true level at which we could reject the AR(1) noise null-hypothesis would be only 72%. For most practical applications, the difference between rejection of the null-hypothesis at 97.5% and at 99.5% is uninteresting. But the difference between the 94.6% level and the 72% level would almost certainly incline us to view a result very differently.

Quantifying the probability of excursions above these high percentiles requires large Monte Carlo ensembles: we have used 10,000 in the examples in this section. This will only be necessary if we are dealing with a result on the margins of acceptable significance. For Gaussian distributed null-hypotheses, a rough indication of global significance may be obtained by assuming excursions conform to the binomial distribution (see appendix). If we are dealing with a more complicated null-hypothesis, such as a chaotic process, then it is preferable to pursue the two-pass Monte Carlo procedure to quantify the significance level explicitly.

### 4.3 Case 2: Simple null-hypothesis with some unknown noise parameters

The test outlined above evaluates the hypothesis that a series has been generated by a particular noise process whose parameters are known. Frequently, however, we wish to test a vaguer null-hypothesis, such as "this series was generated by an AR(1) process" where the process parameters (in this case: mean, variance and lag-1 autocorrelation) are unknown. To reject the entire class of AR(1) processes on the basis of a single test, we must identify that particular AR(1) process, or that set of parameters $u_0$, $\alpha$ and $\gamma$ in equation (2), which maximises the likelihood of our failing to reject the null hypothesis.[3]

We can deal with the fact we do not know $u_0$ simply by centering the data series (removing the statistical mean, $\bar{d}$). In order to ensure that we are treating surrogates and data identically, however, we must then center each individual surrogate, thereby complicating the estimation of $\alpha$ and $\gamma$.

The lag-$l$ covariances of an AR(1) process are given by

$$c_l = \frac{\alpha^2 \gamma^l}{1-\gamma^2}, \tag{7}$$

so we have only to estimate the covariance at two different lags to obtain estimates of $\alpha$ and $\gamma$. The most efficient estimators are obtained from the lag-0 and lag-1 covariances Stuart & Ord, 1991. A natural estimator for the lag-$l$ covariance is (Vautard *et al.*, 1992)

$$\hat{c}_l \equiv \frac{1}{N-l} \sum_{i=1}^{N-l} (d_i - \bar{d})(d_{i+l} - \bar{d}). \tag{8}$$

Although less biased than the Yule-Walker estimate (Yule, 1927; Walker, 1954), $\hat{c}_l$ is still subject to some bias because the mean of this particular segment, $\bar{d}$, is not identical to the process mean. If we generate surrogates using parameters $\gamma$ and $\alpha$ chosen to yield lag-0 and lag-1 covariances $\hat{c}_0$ and $\hat{c}_1$ in equation (7), and center the surrogates, then the expected variance of the surrogates will be less than the variance of the data series by a factor

$$\mu^2 \equiv \frac{1}{N} + \frac{1}{N^2} \sum_{k=1}^{N-1} 2(N-k)\gamma^k, \tag{9}$$

where $\mu^2 = \mathcal{E}(\bar{d}^2)/c_0$, the expected square on the mean of a finite segment of AR(1) noise, normalised by the process variance (Allen, 1992).[4] This bias is unimportant for the estimation of $\mathbf{C}_D$ in SSA, since it has only a minor impact on the eigenbasis $\mathbf{E}_D$, but it can obviously be very important in setting the parameters of the null-hypothesis for significance tests. If we use equation (8) to estimate parameters naively, then the expected variance of the surrogates will be less than that of the data series. This incorrectly exaggerates the apparent significance of any features observed. In short, it means that a segment of AR(1) noise would tend to appear improbable when tested against the AR(1) noise null-hypothesis!

We can correct for this bias by using the estimator

$$\tilde{c}_l \equiv \hat{c}_l + \tilde{c}_0 \mu^2. \tag{10}$$

When $\gamma = 0$ (the white noise hypothesis), $\tilde{c}_0$ reduces to the familiar unbiased estimator of the variance, $\frac{N}{N-1}\hat{c}_0$. When $\gamma \neq 0$, equation (10) is still only implicit in the individual $\tilde{c}_l$, but dividing $\tilde{c}_1$ by $\tilde{c}_0$ gives an explicit estimator for $\gamma$, *viz.* $\tilde{\gamma}$, being the solution of:

$$\frac{\hat{c}_1}{\hat{c}_0} = \frac{\tilde{\gamma} - \mu^2(\tilde{\gamma})}{1 - \mu^2(\tilde{\gamma})}. \tag{11}$$

The gradient of the RHS of equation (11) is always positive in $\tilde{\gamma}$, and generally close to unity, so an efficient solution procedure is provided by Newton-Raphson iteration, simplified by assuming unit gradient throughout, and starting from $\hat{\gamma} = \frac{\hat{c}_1}{\hat{c}_0}$. We find this algorithm generally converges to an acceptable accuracy (estimating $\tilde{\gamma}$ to within $\sim 10^{-5}$ of its asymptotic value) in 2–5 iterations. Once $\tilde{\gamma}$ has been found, $\tilde{c}_0$ is given by $\tilde{c}_0 = \frac{\hat{c}_0}{1-\mu^2(\tilde{\gamma})}$, and $\tilde{\alpha}$ obtained from equation (7).

Using $\tilde{\gamma}$ and $\tilde{c}_0$ allows us to avoid the bias inherent in $\hat{\gamma}$ and $\hat{c}_0$, at the cost of introducing estimators which

[3] Much of this discussion of parameter-estimation is closely related to the Generalised Linear Regression problem: see Mardia *et al.*, 1979, for a helpful introduction. In statistical jargon, we are simply obtaining the Best Linear Unbiased Estimators of AR process parameters in the particular context of SSA.

[4] The RHS of equation (9) may be summed explicitly, which is useful for examining its asymptotic properties (we are grateful to D. Broomhead for this suggestion):

$$\mu^2(\gamma) = -\frac{1}{N} + \frac{2}{N^2}\left[\frac{N-\gamma^N}{1-\gamma} - \frac{\gamma(1-\gamma^{N-1})}{(1-\gamma)^2}\right].$$

Thus:

$$\lim_{\gamma \to 1}\left[\frac{\gamma - \mu^2(\gamma)}{1-\mu^2(\gamma)}\right] = \frac{N^2 - 3N - 1}{N^2 - 1}.$$

If the LHS of equation (11) is greater than $(N^2-3N-1)/(N^2-1)$, then we cannot place an upper bound on the autocorrelation of the noise as long as the process mean is unknown. A common reason for this occurring is that we are dealing with data which contains a trend or random-walk component, which must be accounted for if possible (e.g. using the procedure outlined in section 4.4) before the surrogate data test is performed.

are non-linear functions of the second-order moments of the data. Fortunately, provided the autocorrelation time-scale $\tau$ is at least an order of magnitude smaller than the length of the series (this is the case in all the series examined here, and results should be treated with caution in any situation where this condition does not hold), this non-linearity is weak, and $\tilde{\gamma}$ and $\tilde{c}_0$ are efficient and well-behaved. The small-sample properties $\tilde{\gamma}$ and $\tilde{c}_0$ are documented in detail in section 3.2.2 of Allen, 1992.

Applying these estimators to our test series gives $\tilde{\gamma}$ within 1% of the value used in the AR(1) noise component of the generating process, but an estimate $\tilde{c}_0$ which is 14% larger than the process variance of the generating noise, since the test series also contains variance due to the signal. Applying Monte Carlo SSA with these estimated AR(1) parameters (results not shown, but they are visually very similar to figure 4), EOFs 8 and 9 remain significant at greater than the 99.5% level, but because the noise in the null-hypothesis now contains more variance than the noise in the generating process, EOFs 21 and 22 no longer appear significant even at the 95% level, and 5 low-ranked eigenvalues lie below the $0.5^{\text{th}}$ percentiles of their corresponding surrogate data distributions.

The results presented in this section allow us to reject the pure noise null-hypothesis, on the grounds that the data contains evidence of a 5.5-unit-period oscillation. The next question is, does the data contain evidence of anything else? No other EOFs are indicated as significant at even the 95% level, but this is inconclusive, since having rejected this null-hypothesis, we know that the noise variance is too high. To ensure that we have not missed anything, the testing procedure must continue until we arrive at a null-hypothesis that we cannot reject.[5] The next step is to test the composite null-hypothesis that the data consists of the oscillation indicated by EOFs 8 and 9 plus AR(1) noise.

[5]Mann & Lees, 1995, in their red-noise test for MTM, try to ensure that secondary spectral features are not rejected through misspecification of the noise parameters by using "robust" estimators which are not sensitive to the presence of narrow-band spectral features. Their test has the advantage that it is single-step procedure, but we prefer the approach described here since it is applicable to broadband signals, occupying a significant portion of the Nyquist interval, which are often encountered in geophysics. Moreover, by continuing to test until we arrive at a null-hypothesis that we cannot reject, we can check that an AR(1) process was an appropriate noise model to assume in the first place.

### 4.4 Case 3: Composite null-hypothesis with some unknown noise parameters

Given that we have identified some components of the data series as "signal" (either through a statistical test, or from *a priori* knowledge), and wish to test whether the remainder is attributable to AR(1) noise with undetermined parameters, we have several options. Conceptually, the simplest procedure is the "signal-reconstruction" approach (Allen, 1992; Allen & Smith, 1994; Dettinger *et al.*, 1995), which works as follows. The signal component of the data time-series is reconstructed explicitly using the algorithm given in Vautard *et al.*, 1992. This involves computing the filtered trajectory matrix,

$$\mathbf{D}' \equiv \mathbf{D}\mathbf{E}_D(\mathbf{I}-\mathbf{K})\mathbf{E}_D^T \equiv \mathbf{D}\mathbf{S}, \tag{12}$$

where $\mathbf{K}$ is an $M \times M$ diagonal matrix in which $K_{kk} = 0$ if EOF-$k$ has been identified as corresponding to a signal, and $K_{kk} = 1$ otherwise. In the case of our test series, $\mathbf{K}$ would simply be the unit matrix with the $8^{\text{th}}$ and $9^{\text{th}}$ diagonal elements set to zero (assuming we have reserved judgement on EOFs 21 and 22). $\mathbf{S}$, the "signal projection matrix", is idempotent ($\mathbf{SS} = \mathbf{S}$), and commutes with $\mathbf{C}_D$, so $\mathbf{S}\mathbf{C}_D\mathbf{S} = \mathbf{S}\mathbf{S}\mathbf{C}_D = \mathbf{S}\mathbf{C}_D$. An explicit filtered reconstruction, $\mathbf{d}'$, of the original scalar series $\mathbf{d}$ is then given by averaging along the diagonals of $\mathbf{D}'$ (Ghil & Vautard, 1991), such that

$$d'_i = \frac{1}{\zeta} \sum_{j=\max(1,i+M-N)}^{\min(i,M)} D'_{i-j+1,j} \quad , \tag{13}$$

$\zeta$ being the number of terms in the sum (the smallest of 1, $M$ and $N-i+1$).

The next step in the signal-reconstruction approach is to compute the process parameters required to generate AR(1) noise such that, when added to $\mathbf{d}'$, the composite signal-plus-noise has the same expected variance and lag-1 autocorrelation as the original data series. Noise covariances cannot be estimated directly from the residual $\mathbf{d}-\mathbf{d}'$, because $\mathbf{d}'$ and $\mathbf{d}-\mathbf{d}'$ are not orthogonal (Allen, 1992). Noise segments are then generated using these parameters and added to $\mathbf{d}'$ to give "composite surrogates", from which the $\mathbf{C}_R$ are computed and tested as in section 4.1 above.

A major disadvantage of the signal-reconstruction approach is that it does not yield unbiased estimates of the noise parameters, even in the long-series limit. When AR noise is added to to the reconstructed signal, the noise variance is distributed over all frequencies, including those associated with the signal. Thus for

large $N$ the surrogates will contain, on average, more variance than the data at the signal frequencies and, because we have set their total expected variance to be equal to that of the data, less variance at all other frequencies. A second source of bias arises with short series, since SSA-based reconstructions tend to be "overfitted" near the series end-points (Vautard *et al.*, 1992). Thus the AR parameters obtained with this procedure do not maximise the likelihood of our failing to reject the signal-plus-AR(1)-noise null-hypothesis.

Exactly the same problem arises if we use SSA to filter out the component of the data which we consider to be signal, and then fit the AR parameters to the filtered data. To ensure that, in distinguishing between data and surrogates, we are not simply "detecting" the filter response function, we must filter the surrogates in exactly the same way as we filter the data (this point is also important if we are applying a surrogate data test to detrended data – Schlesinger & Ramankutty, 1994; Elsner, 1995). Filtering will reduce the expected variance of the surrogates, making it less than that of the filtered data.

Precise optimisation of null-hypothesis parameters is important if we wish to report a positive result: *i.e.,* if we wish to stress the fact that the null-hypothesis has been rejected. Negative results are more robust. If we fail to reject the null-hypothesis even with a sub-optimal choice of parameters, then we would be even more likely to fail with the correct parameters. For example, the key result of Allen, 1992; Allen & Smith, 1994, is a negative one, *viz.* that SSA provides no evidence for inter-decadal oscillations in the historical global temperature record, so it is not affected by their use of a sub-optimal parameter-estimation procedure.

A second disadvantage of the signal-reconstruction approach is that SSA-based reconstructions are poorly behaved when we are dealing with irregularly-sampled and/or heteroskedastic data. Although we only deal here with regularly-sampled data with equal weight given to all data-points, an ability to generalise the technique to these other cases is clearly desirable.

Our aim is to identify null-hypothesis parameters such that the noise, after filtering to suppress variance in the directions defined by the EOFs which we have identified with signal, has the same expected variance and lag-1 autocorrelation as the data, filtered similarly. To avoid explicit reconstructions, we work entirely with lag-covariance matrices. An estimate of the lag-covariance matrix of the filtered signal is given by

$$\eta \mathbf{D'}^T \mathbf{D'} = \eta \mathbf{S} \mathbf{D}^T \mathbf{D} \mathbf{S} \simeq \mathbf{S} \mathbf{C}_D \quad (14)$$

where $\eta$ is an algorithm-dependent normalisation constant defined as in equation (3) above. If the BK algorithm is used, equation (15) is satisfied exactly. If the VG algorithm is used, equation (15) holds only in the long-series limit. Since, however, all we require are efficient and unbiased estimators of the noise parameters, we can use $\mathbf{S}\mathbf{C}_D$ as an estimate of the filtered lag-covariance matrix in this case as well. End-effects will mean that the parameters we obtain will be slightly different from those which we obtain by the signal-reconstruction approach, but since those estimates were also subject to bias, we have found no reason to favour one over the other; in addition, we wish to avoid explicit reconstructions to allow a modified algorithm for incomplete or heteroskedastic data.

We introduce a generalised trace operator $\mathrm{tr}_j$ which, when applied to a $M \times M$ symmetric matrix, is defined thus:

$$\mathrm{tr}_j (\mathbf{C}) \equiv \frac{1}{M-j} \sum_{k=1}^{M-j} C_{k,k+j} \quad , \quad (16)$$

and a "noise projection matrix", $\mathbf{Q} = \mathbf{E}_D \mathbf{K} \mathbf{E}_D^T$, with properties similar to $\mathbf{S}$. To avoid repetitive use of the expectation operator, we also introduce a matrix $\mathbf{C}_N$, being the expected lag-covariance matrix of a noise realisation, $\mathbf{C}_N = \mathcal{E}(\mathbf{C}_R)$ (recall that, since both VG and BK algorithms are biased, $\mathbf{C}_N$ is not, in general, equal to the process lag-covariance matrix of the noise). In the signal-reconstruction approach, the constraint that the noise added to the reconstructed signal should have the same expected variance as the original data is equivalent, apart from end-effects, to requiring that $\mathbf{C}_N$ satisfies $\mathrm{tr}_0 (\mathbf{C}_N) = \mathrm{tr}_0 (\mathbf{C}_D - \mathbf{S}\mathbf{C}_D) = \mathrm{tr}_0 (\mathbf{Q}\mathbf{C}_D)$. Thus the bias in the signal-reconstruction-based algorithm arises from applying the noise filter $\mathbf{Q}$ to $\mathbf{C}_D$ and not to $\mathbf{C}_N$. Unbiased estimates of noise parameters can be obtained by applying the filter to both matrices. Thus if we have two free parameters in our noise model, we estimate them by finding those values for which

$$\mathrm{tr}_0 (\mathbf{Q}\mathbf{C}_N\mathbf{Q}) = \mathrm{tr}_0 (\mathbf{Q}\mathbf{C}_D) \quad (17)$$

$$\mathrm{tr}_1 (\mathbf{Q}\mathbf{C}_N\mathbf{Q}) = \mathrm{tr}_1 (\mathbf{Q}\mathbf{C}_D) \quad (18)$$

(note that $\mathbf{Q}$ does not commute with $\mathbf{C}_N$, so we cannot simplify $\mathbf{Q}\mathbf{C}_N\mathbf{Q}$ as we can simplify $\mathbf{Q}\mathbf{C}_D\mathbf{Q}$).

If we are testing an AR(1) noise null-hypothesis, with unknown mean, variance $c_0$ and lag-1 autocorrelation $\gamma$, then $\mathbf{C}_N = c_0 \mathbf{W'}$, where $\mathbf{W'}_{ij} = \gamma^{|i-j|} - \mu^2(\gamma)$. This takes into account the effect of centering discussed in

the previous section. We estimate $\gamma$ by solving

$$\frac{\mathrm{tr}_1\left(\mathbf{QW'Q}\right)}{\mathrm{tr}_0\left(\mathbf{QW'Q}\right)} = \frac{\mathrm{tr}_1\left(\mathbf{QC}_D\right)}{\mathrm{tr}_0\left(\mathbf{QC}_D\right)} \tag{19}$$

using our modified Newton-Raphson scheme, and $c_0$ from

$$c_0 = \frac{\mathrm{tr}_0\left(\mathbf{QC}_D\right)}{\mathrm{tr}_0\left(\mathbf{QW'Q}\right)}. \tag{20}$$

Against a pure noise null-hypothesis, $\mathbf{Q}$ becomes the unit matrix, and this reduces to the algorithm given in the previous section.

We generate surrogates using these parameters, center them, and compute $\mathbf{C}_R$ and $\mathbf{\Lambda}_R$ as in section 4.1. Note that

$$\mathcal{E}\left(\mathrm{tr}_0\left(\mathbf{K\Lambda}_R\right)\right) = \mathrm{tr}_0\left(\mathbf{K\Lambda}_D\right), \tag{21}$$

which provides a useful end-to-end check.

Applying this algorithm to our test series, treating EOFs 8 and 9 as signal, gives an estimated noise variance of 1.097, and lag-1 autocorrelation of 0.72. Thus $\gamma$ is accurately estimated, but the estimated noise variance remains higher than the actual variance of the noise in the test series (1.04 in this realisation). Not all the signal has been extracted in EOFs 8 and 9, which is inevitable, since this is a rapidly modulated oscillation. We do not distinguish between signal and noise in the directions defined by the EOFs which we associate with signal (although the signal-to-noise (S/N) ratio will be considerably enhanced in these directions over the S/N ratio for the full series). We might be able to use additional information to improve signal-extraction, and therefore improve the estimation of noise parameters, if we knew we were looking specifically for damped sinusoidal bursts. The aim here, however, is to demonstrate a generally-applicable algorithm for weak modulated oscillatory signals.

With these revised parameters, EOFs 21 and 22 are individually significant at the 96% level. Out of 38 possible EOFs (EOFs 8 and 9 having been eliminated from the statistics), there is a 27% chance of 2 or more excursions above the $97.5^{\mathrm{th}}$ percentiles, so, if we did not have any other reason to be interested in EOFs 21 and 22, it would be questionable to conclude that they represent a significant signal.

This concludes our description of the basic method of hypothesis-testing in Monte Carlo SSA. Before leaving this discussion of our test series, however, we must discuss one remaining unquantifiable bias in the algorithm described so far arising from the fundamental data-adaptive properties of SSA itself.

# 5 Problems with data-adaptive EOFs

The guiding principle of surrogate data testing is that we must treat data and surrogates in exactly the same way. If we fail to do so, we may appear to distinguish the data from the surrogates simply as a result of having treated them differently. So far, we have failed to adhere to this principle in one important respect. The standard SSA algorithm determines EOFs by maximising the variance accounted for in the data series by the smallest possible number of patterns. Thus if we perform SSA on a segment of pure noise, a high (low) ranked EOF is likely to account for an improbably high (low) proportion of the variance in the particular noise-segment from which it was derived, relative to the variance it accounts for in an arbitrary series generated by the same noise process. Notice how the highest ranked data eigenvalues in figure 3 are near the top of their corresponding surrogate data bars, while the lowest ranked eigenvalues are near the bottom. The extent of this artificial "variance-compression" is difficult to quantify, since it depends not only on the length of the series, but also on the method used to compute the lag-covariance matrix, and on the noise characteristics (it is worst for short series, for the BK matrix and for white noise). It has the highly undesirable effect of making signal-detection algorithms based on the standard approach to SSA inherently under-conservative, *i.e.,* making the true probability of a type-1 error higher than the nominal level of the test.

## 5.1 A test based on eigenspectrum shape

One response to this problem is a significance test proposed by Elsner & Tsonis, 1994b; Elsner, 1995, and also implemented by Dettinger *et al.*, 1995, which we will refer to as the Florida-Milwaukee-UCLA (FMU) test. This is identical to the signal-reconstruction approach of Allen, 1992, except that, instead of projecting each surrogate lag-covariance matrix onto the data EOFs as in equation (6), the FMU test obtains a new EOF basis, $\mathbf{E}_R$, for each surrogate realisation by diagonalising the individual $\mathbf{C}_R$ thus:

$$\mathbf{\Lambda}_R^{\mathrm{FMU}} \equiv \mathbf{E}_R^T \mathbf{C}_R \mathbf{E}_R. \tag{22}$$

The elements of $\mathbf{\Lambda}_R^{\mathrm{FMU}}$ are arranged in the conventional rank-order. Thus the artificial variance-compression effect noted above will be present in both the data and the surrogate eigenspectra, consistent with the overall philosophy of surrogate data testing.

The successful application of this test, however, only allows us to say that the $k^{\mathrm{th}}$ eigenvalue of $\mathbf{C}_D$ is un-

usually large given its position in the eigenvalue rank-order. We can say nothing about the structure of "the" $k^{\text{th}}$ EOF, or any associated frequency, as there is no unique $k^{\text{th}}$ EOF. The comparison of individual elements of $\mathbf{\Lambda}_D$ with the distribution of corresponding elements of $\mathbf{\Lambda}_R^{\text{FMU}}$ is no longer meaningful since the $k^{\text{th}}$ EOF of the data series may be associated with a completely different set of frequencies (*i.e.,* have a completely different shape, or point in a completely different direction in state-space) relative to the ensemble average of the $k^{\text{th}}$ EOFs of the surrogates. This is particularly likely if the $k^{\text{th}}$ EOF of the data corresponds to a genuine oscillation. The FMU test compares the overall shape of the ranked eigenspectrum of the data with the overall shape of the ranked eigenspectra of the surrogates without taking into account the shapes of the corresponding EOFs. As such, it is not a reliable method of discriminating between oscillations and red noise, since the presence of an oscillation may not affect the overall shape of the ranked eigenspectrum.

The vertical bars in figure 6 show the distributions of the $\mathbf{\Lambda}_R^{\text{FMU}}$ obtained with the FMU eigenspectrum-shape test using exactly the same surrogate ensemble used to calculate the vertical bars in figures 3 and 4. None of the data eigenvalues are indicated as significant, because there is nothing unusual about the shape of the ranked data eigenspectrum relative to the shape of the surrogates' ranked eigenspectra.

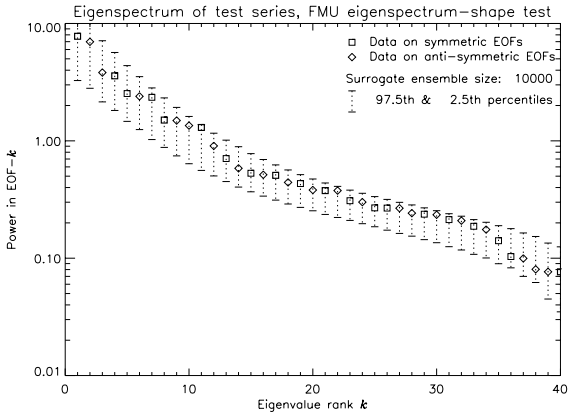


**Figure 6.** Application of the Florida/Milwaukee/UCLA eigenspectrum-shape test to our test series, testing against a pure AR(1) noise null-hypothesis. Vertical bars show the distribution of eigenvalues of $\mathbf{C}_R$, individually ranked in order of decreasing size. This test does not make use of the shape of the data EOFs, and so is not effective against red noise: no excursions above the $97.5^{\text{th}}$ percentiles occur, and EOFs 8 & 9, which contain most power at 5.5-unit periods, are not picked out as unusual.

To be effective, tests against red noise must make use of information concerning both the power in and shape of EOFs. Comparing ranked eigenspectra discards all EOF-shape information. It may be possible to reintroduce this information through some form of bootstrapping (Elsner, 1995), but the problem of determining which EOFs to boot-strap to initiate this approach remains. The FMU test requires that the eigenvalue rank-order distinguish signals from noise; in conventional SSA, rank-order is misleading. While revised approaches to SSA may exist (e.g. Allen & Smith, 1996) in which the rank-order is meaningful (and thus for which the FMU test should work), we present a simpler approach to dealing with artificial variance-compression which uses the eigenbasis of the null-hypothesis, as follows.

### 5.2 Using the EOFs of the null-hypothesis

Standard SSA determines EOFs by assuming, in effect, that the data series is noise free. It continues to extract variance-maximising patterns all the way down to the lowest-ranked EOF. This is appropriate for a data-compression tool, but clearly dangerous in a signal-detection algorithm. If we are using SSA to reconstruct a signal which we already know to be present in the data, or to "clean up" a time-series contaminated with a relatively low level of noise, as in the original work of BK, then the standard algorithm has a clear theoretical justification. Recent applications, on the other hand, aim to use SSA to identify unexpected features in a time-series. In this situation, the priority must be to quantify objectively the probability that these features may have occurred by chance. This makes the use of fully data-adaptive EOFs more difficult to justify, since effects like artificial variance-compression increase the probability of type-1 errors by an unquantifiable amount.

In this section, we introduce a different approach to SSA, which allows us to retain its data-adaptive properties for the extraction of signals which have already been detected while avoiding the problems inherent in an algorithm which implicitly assumes the existence of a signal before any signal has been identified.

We frame this revised approach on the assumption that the null-hypothesis is true until we establish otherwise. Thus, if the null-hypothesis is that a series has been generated by AR(1) noise, we represent the data using the EOFs which we *expect* from a segment of that type of noise, rather than the EOFs derived from the data itself. In support of this approach, recall that our theoretical justification for diagonalising $\mathbf{C}_D$

was that we thereby obtain an estimate of the EOFs of the process which generated the data series. If we initially assume that the null-hypothesis is true, then we know what these EOFs are without needing the data at all (things are more complicated with composite null-hypotheses when a signal has been already been identified).

We consider, first, the simple case in section 4.1 where we know the mean, variance and lag-1 autocorrelation of the noise *a priori*. The expected lag-covariance matrix of a segment of such a noise process is given by $\mathbf{C}_N = c_0\mathbf{W}$, where $W_{ij} = \gamma^{|i-j|}$ (no correction is required for the effect of centering, since the mean is known). We compute the expected EOFs of this noise by diagonalising $\mathbf{C}_N$, thus:

$$\mathbf{\Lambda}_N \equiv \mathbf{E}_N^T \mathbf{C}_N \mathbf{E}_N, \tag{23}$$

and project both data and surrogates onto these noise EOFs thus:

$$\mathbf{\Lambda}'_D \equiv \mathbf{E}_N^T \mathbf{C}_D \mathbf{E}_N \tag{24}$$

$$\mathbf{\Lambda}'_R \equiv \mathbf{E}_N^T \mathbf{C}_R \mathbf{E}_N \tag{25}$$

where the primes indicate that the null-hypothesis basis has been used.

Squares and diamonds in figure 7 show $\mathbf{\Lambda}'_D$ for our test series while the vertical bars show the $2.5^{\text{th}}$ and $97.5^{\text{th}}$ percentiles of the distributions of the $\mathbf{\Lambda}'_R$. This figure may be compared directly to figure 4, which shows the same data and surrogate ensemble on the data-adaptive EOFs. Note how the EOFs of $\mathbf{C}_N$ are regularly spaced, separated by almost exactly $1/2M$. Two elements of $\mathbf{\Lambda}'_D$, corresponding to EOFs 15 and 16 of the null-hypothesis, lie above the $99.5^{\text{th}}$ percentiles of their corresponding surrogate data bars, corresponding to EOFs associated with periods of 5.3 and 5.6 units (99.7th and $99.8^{\text{th}}$ percentiles respectively). The probability of 2 or more such excursions occuring in a member of the surrogate ensemble is 5.6%, giving a robust lower limit on the level of the test of 94.4%, without the complications associated with variance compression. Similar results are obtained in the case where we do not assume prior knowledge of the noise parameters, but confidence levels are slightly lower, because the noise variance is overestimated (see section 4.3).

Both the data projections and percentiles of the surrogate ensemble in figure 7 are very similar to the Blackman-Tukey power spectral estimate in figure 2. To the extent that the basis of pure AR(1) noise resembles the Fourier basis, there is little to choose between them. More complicated null-hypotheses, however, involve different bases, requiring the full method presented below.

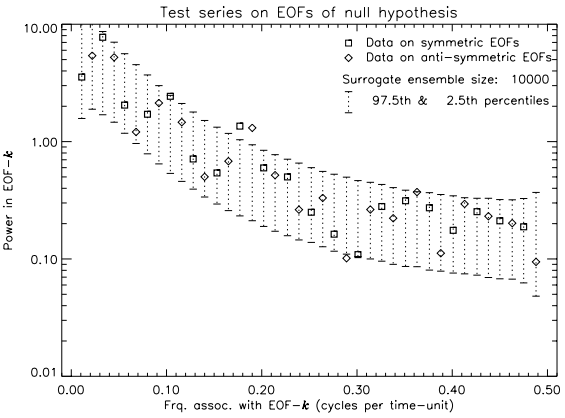


**Figure 7.** Test series against the AR(1) noise null-hypothesis, projecting both data and surrogates onto the EOFs of the expected lag-covariance matrix of the surrogates, $\mathbf{C}_N$. EOFs are regularly spaced, separated by $\sim 1/2M$, and alternately symmetric and anti-symmetric. Significant power is found at frequencies corresponding to periods of 5.3–5.6 units.

Figure 7 tells us that there is anomalous power in the data at periods between 5.3 and 5.6 units, but it does not provide us with an optimal algorithm for the extraction and reconstruction of that signal. A visual comparison of figure 7 with figure 4 is sufficient to indicate that the eigenvectors corresponding to data EOFs 8 and 9 in figure 3 contain the 5.5-unit-period signal in question, but to allow straightforward automation, we compute cross-correlations $\mathbf{E}_D^T \mathbf{E}_N$, and find the EOFs of the data which are maximally correlated with EOFs 15 and 16 of the null-hypothesis.

We now require an EOF basis corresponding to the composite null-hypothesis that the data consists of whatever signal is contained in EOFs 8 and 9, plus AR(1) noise. In this last example, we assume all noise parameters including the mean are unknown, and estimate them using the data EOFs and the matrix $\mathbf{Q}$ as in section 4.4. Using the data EOFs to represent the noise component in $\mathbf{Q}$ does not introduce any variance-compression effects, since all we require is an orthonormal basis which contains EOFs 8 and 9 of the data. We define the expected lag-covariance matrix of the composite null-hypothesis through:

$$\mathbf{C}_N = c_0 \mathbf{Q}\mathbf{W}'\mathbf{Q} + \mathbf{S}\mathbf{C}_D, \tag{26}$$

where $\mathbf{W}'_{ij} = \gamma^{|i-j|} - \mu^2(\gamma)$ as above.

The eigenvectors of $\mathbf{C}_N$ will contain EOFs 8 and 9 of the data, unless these are degenerate with some components of the noise, in which case they may be scrambled in the diagonalisation. We can sidestep de-

generacy problems through the numerical trick of temporarily setting $c_0$ in equation (26) to 90% of the smallest "signal" eigenvalue of $\mathbf{C}_D$ before diagonalising $\mathbf{C}_N$. The precise factor used does not affect the decomposition, but it should not be too small or $\mathbf{C}_N$ may become ill-conditioned.

The projections $\mathbf{\Lambda}'_D$ and $\mathbf{\Lambda}'_R$ onto the EOFs of the composite null-hypothesis are shown in figure 8. Notice how the noise EOFs tend to "pair up" in the vicinity of the signal, owing to the orthogonality constraint, and the lack of artificial variance-compression. One remaining element of $\mathbf{\Lambda}'_D$ is significant at the 95% level, number 30, with associated frequency 0.36(units)$^{-1}$. That alone is not enough to reject this null-hypothesis, but the associated frequency suggests the first harmonic, so we might wish to investigate it further.

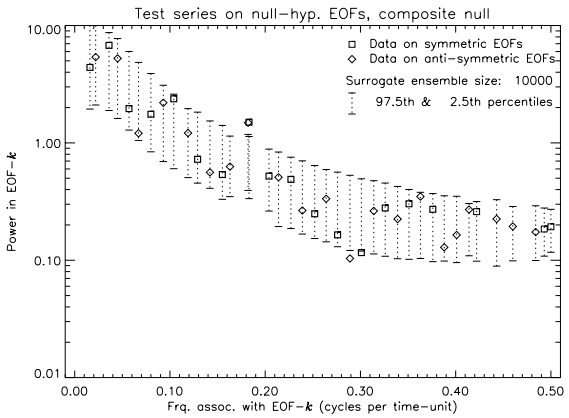


**Figure 8.** Test series against the composite null-hypothesis of AR(1) noise plus a 5.5-unit oscillation, using the EOFs of the expected lag-covariance matrix of the surrogates. A symmetric/anti-symmetric pair is used from the EOFs of $\mathbf{C}_D$ and the remaining EOFs correspond to AR(1) noise subject to the constraint that they must be orthogonal to this pair. All noise parameters have now been estimated from the data.

Using the null-hypothesis basis, we should no longer expect (much less, require) significant EOFs to appear in symmetric/anti-symmetric pairs. Neither of the EOFs adjacent to EOF 30 in figure 8 is even close to the 95% significance level. EOF 30 of $\mathbf{C}_N$ is antisymmetric and is closely aligned (cross product of 0.94) with EOF 22 of the data. An inspection of associated frequencies indicates that EOF 21 of the data is the other member of that pair, and so a signal-reconstruction should include both.

We emphasise that once a signal has been detected we then use the same EOFs as are used in standard SSA to study and reconstruct it. For genuine signals nothing is lost. There is little risk of our "missing" a signal which does not happen to align exactly with one of the EOFs of the null-hypothesis since the frequencies associated with the EOFs of pure AR(1) noise are separated by $1/2M$. Whatever the signal frequency, it must lie within $1/4M$ of either a symmetric or an anti-symmetric noise EOF, which is close enough for a high proportion of the variance which would have been associated with an optimal EOF obtained from standard SSA to project onto that noise EOF. This is particularly true if the signal variance is spread over a range of frequencies, as expected in geophysical data.

The conclusive argument in favour of using of the eigenbasis of the null-hypothesis is that we are able to quantify the probability of a false-positive result more precisely than if we use the data-adaptive basis since we are not subject to the unknown effects of artificial variance-compression. For signal detection applications, this is more important than maximising the probability of making a weak but genuine signal appear significant. Detection consists in distinguishing signals from noise. An algorithm which makes a genuine signal appear significant, at the cost of interpreting a large (and unknown) number of noise components as significant as well, cannot be said to have detected anything.

## 6 Evidence for climate oscillations

We now apply Monte Carlo SSA to a much-studied problem: the detection of low-frequency climate oscillations. We begin with the problem that first prompted this work on significance tests for SSA, *viz.* evaluating the evidence for interannual and interdecadal oscillations in global temperatures.

### 6.1 The historical global temperature record

The dataset we will focus on is the 136-year record of global, annual-mean, combined land and sea near-surface temperatures, as compiled by the Intergovernmental Panel on Climate Change (the "IPCC series") (Folland *et al.*, 1990; Folland *et al.*, 1992). SSA was first applied to the IPCC series by Ghil & Vautard, 1991, who also considered other historical records derived from closely related datasets (Jones *et al.*, 1986a; Jones *et al.*, 1986b). Ghil & Vautard, 1991, reported an interdecadal oscillation with a period slightly longer than that previously reported by Newell *et al.*, 1989, in global night-time marine air temperatures (NMAT) using a more conventional Fourier analysis. Both Newell *et al.*, 1989, and Ghil & Vautard, 1991, note that the existence of such an oscillation would have significant implica-

tions for long-range climate prediction and the detection of anthropogenic climate change. The stability of Ghil & Vautard, 1991's result was subsequently examined by Elsner & Tsonis, 1991, and Allen *et al.*, 1992b, in a correspondence which clearly indicated the need for a formal significance test for SSA against autocorrelated (red) noise.

Vautard *et al.*, 1992, applied a number of tests to this series, all of which considered the white noise null-hypothesis exclusively.[6] As demonstrated above, testing against white noise is necessarily inconclusive when dealing with temperature data since both empirical and physical arguments suggest that the noise will be red. Results from tests against white noise may, therefore, be misleading. In a 136-point segment of AR(1) noise with the same lag-1 autocorrelation as the IPCC series, the probability of an EOF-pair appearing as an interdecadal oscillation, passing all the pair-selection criteria and significance tests of Vautard *et al.*, 1992, is approximately 50% Allen, 1992.[7]

Newell *et al.*, 1989, also confine their significance analysis to white noise on the grounds that the autocorrelation function of the NMAT series does not decay exponentially with increasing lag. While we cannot comment explicitly on Newell *et al.*, 1989's results, we note that an exact exponential decay of autocorrelation is not expected for finite segments of AR(1) noise. In particular, either a trend or an interannual oscillation would force the autocorrelation function to depart considerably from an exponential even in the case that variability on all other time scales (including interdecadal) is attributable to AR(1) noise.

Figure 9 shows the IPCC series tested against a null-hypothesis of pure AR(1) noise. For continuity with previous work, we show results using the data-adaptive basis, $\mathbf{E}_D$, of conventional SSA. Noise parameters are fitted to the data series following the procedure described in section 4.3 above. Only one excursion above the 97.5$^{\text{th}}$ percentile of the surrogate distributions occurs; EOF 1 lies in the 98$^{\text{th}}$ percentile. The probability of one or more excursions out of a possible 40 above the 97.5$^{\text{th}}$ percentile is 45% (evaluated by a second pass through the surrogate ensemble – see section 4.2), so the information that one such excursion occurs is not enough for us to reject the null-hypothesis that the IPCC series is a segment of AR(1) noise. However, we have prior reason to focus attention on EOFs 1 and 2, since they represent all variability on >40-year time scales, unlike all other EOFs which represent variability only in a particular spectral interval. These EOFs, therefore, are of particular interest for reasons other than the fact that their eigenvalues are singled out as improbably high by the statistical test. We can, therefore, say that the variability on >40-year time scales (*i.e.,* the non-linear trend) in the IPCC series is inconsistent (at 97.5% confidence) with the hypothesis that the series was generated by an AR(1) process.

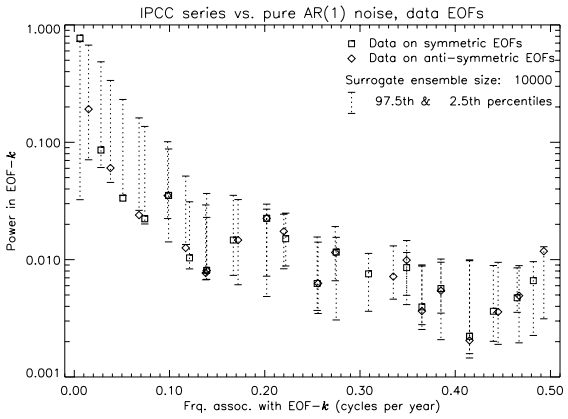


**Figure 9.** Application of Monte Carlo SSA to the IPCC record of global-mean near-surface temperatures 1858–1993, testing the null-hypothesis of pure AR(1) noise. One excursion occurs above the 97.5$^{\text{th}}$ percentiles, which in itself is not enough to reject the AR(1) noise hypothesis. If, however, we take into account the prior information that we *expect* any departure to occur on >40-year time scales (*i.e.,* in EOF 1) due to the presence of a non-linear trend, then the fact that this excursion indeed occurs in EOF 1 makes it much more significant (see text). We therefore conclude that the non-linear trend in the IPCC series is inconsistent with the pure AR(1) noise null-hypothesis at or near the 97.5% level.

If it seems counterintuitive that we can reject the specific null-hypothesis (that the non-linear trend in the IPCC series is consistent with AR(1) noise) at a much higher level than we can reject the general null-hypothesis (that the IPCC series is a segment of AR(1) noise), recall the discussion in section 4.2. The more specific a null-hypothesis, the easier it is to reject.

We now test the IPCC series against the hypothesis that it consists of the observed variability on >40-year time scales plus AR(1) noise. We treat EOFs 1 and 2

[6] A test against AR(1) noise has only recently been developed for MTM spectral analysis (Mann & Lees, 1995); consistent with the results reported here, it shows no significant evidence for low-frequency oscillations in the IPCC series.

[7] The only test not addressed explicitly in Allen, 1992, is the method of computing the statistical dimension, $\upsilon_L$, used in Ghil & Vautard, 1991; Vautard *et al.*, 1992; Plaut *et al.*, 1995. Applied to the IPCC series, this test gives $\upsilon_L \simeq 0.45M$ (Vautard *et al.*, 1992), the result we would expect for a pure AR(1) process.

as signal EOFs, following the procedure for a composite null-hypothesis given in section 4.4. EOF 2 is included even though it is not indicated as significant in figure 9. Having concluded that variability on >40-year time scales is inconsistent with the null-hypothesis, we treat all EOFs corresponding to such variability as signal, whether or not the test indicates them as individually significant, to maximise the chance of our detecting significant variability on some other time scale.

The time scale of decay of autocorrelation in the noise, after eliminating >40-year time scale variability, is 1.4 years (using the signal-reconstruction approach, Allen & Smith, 1994, found a time scale of 1.6–1.7 years, depending on whether or not ENSO variability was eliminated: the small discrepancy is due to our use here of the unbiased parameter-estimation procedure). Results are shown in figure 10. No further excursions outside the 2.5$^{\text{th}}$ and 97.5$^{\text{th}}$ percentile limits occur, even though there is a 43% chance of at least one occurring purely by chance. Moreover, we have used the data-adaptive basis that maximises the chance of high-ranked EOFs appearing significant through the artificial variance-compression effect. Excursions near the 95th percentiles occur in EOFs 9 and 10 (dominant associated period of 5 years) and EOF 17 (2 years). Note how eigenvalue rank-order fails completely as an indicator of statistical significance: EOFs 9 and 10 (associated frequency of 0.2 (years)$^{-1}$), which are unambiguously associated with the low-frequency component of ENSO (Ghil & Vautard, 1991; Allen, 1992; Allen & Smith, 1994), are indicated by the Monte Carlo SSA test as more significant than the 6 EOFs which precede them in the rank order.

To determine whether our failure to detect interdecadal oscillations in figure 10 is a consequence of inadequate signal-to-noise enhancement with $M = 40$, we increase the window width to 60 and repeat the test. With $M$ now almost half the length of the series, the artificial variance-compression effect will be very pronounced such that high-ranked EOFs will contain improbably high variance even if the data consist of a sample of pure noise. Even so, the data-adaptive basis still fails to indicate interdecadal oscillations. We do not recommend the use of the data-adaptive basis in this situation, because of the variance-compression problem. Figure 11 shows the data projected onto the basis of the null-hypothesis (section 5.2). No excursions above the 97.5$^{\text{th}}$ percentiles occur: with $M = 60$ there is a ∼68% chance of at least one such excursion occurring purely by chance.

It is not inherently surprising that the IPCC series

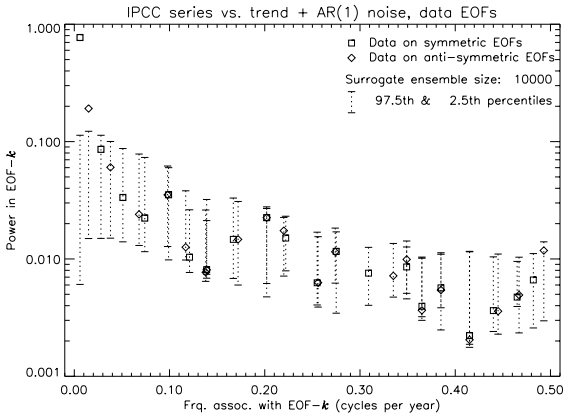


**Figure 10.** Testing the IPCC series against a null-hypothesis of the observed >40-year time-scale variability plus AR(1) noise. No interdecadal or interannual oscillations are indicated at the 97.5% confidence level, even with a >40% chance of at least one such excursion occurring by chance.

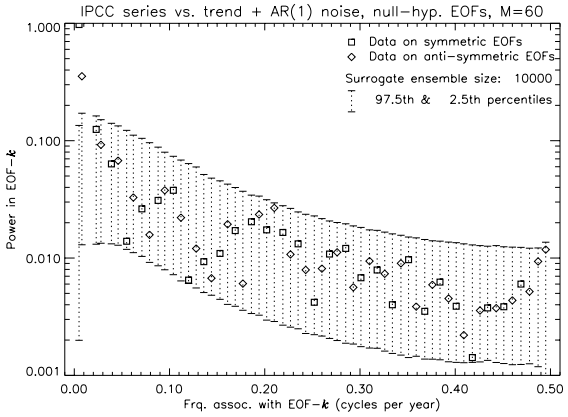


**Figure 11.** As previous figure but with a window width of 60 to increase the potential signal-to-noise enhancement, and using the EOFs of the null-hypothesis.

can be represented by such a simple model, *viz.* a nonlinear trend added to a two-parameter noise process. It consists, after all, of only 136 noisy data-points. We do not claim that all <40-year-time scale climate variability can be represented by an AR(1) process with an autocorrelation decay time of 1.4 years: simply that all variability indicated by SSA in the IPCC series is consistent with this model. Applying SSA to the IPCC series does not provide any significant evidence for interdecadal oscillations in global temperatures, and even on interannual time scales, the evidence for oscillatory behaviour is weak.

We stress that if we were to introduce more infor-

mation, in the form of other data-sets or physical models, we might well find that there are components of the IPCC series which are attributable to oscillatory phenomena. Indeed, EOFs 9 and 10 are found to correspond to ENSO, and EOFs 5 and 6 may be associated with a decadal sea-surface temperature oscillation in the equatorial Atlantic, reported by Allen & Smith, 1994, and (independently) by Mehta & Delworth, 1995, who also find a similar phenomenon in a coupled general circulation model, providing further evidence that it represents a genuine signal, and by Mann & Park, 1994, using a Multi-Taper-based analysis.

There may well be an oscillation, somewhere in the climate system, with a characteristic period of 20–30-years, and our failure to detect it in the IPCC series might simply be due to this oscillation having a dipole structure such that its impact on the global mean temperature is weak. Indeed, Latif & Barnett, 1994, find model-based and observational evidence for interdecadal-time-scale variability in North Pacific SSTs whose overall pattern (a North-West to Central Pacific dipole) closely resembles the pattern associated with the interdecadal component of the IPCC series reported in Allen & Smith, 1994. While suggestive, this does not necessarily indicate that the phenomenon reported by Latif & Barnett, 1994, is the origin of interdecadal variability in global temperatures since there is a high chance that any signal on this time scale would show a consistent phase-relationship with the interdecadal component of the IPCC series, given that the series is only long enough to span ∼3 cycles.

### 6.2 Consequences: over-confident prediction

Failure to detect an interdecadal oscillation in the IPCC series has practical consequences. Vautard *et al.*, 1992, attempt predictions of global temperature to the year 2000 using an SSA-based empirical model derived from the IPCC series. Such forecasts will clearly be influenced by the assumptions that the interdecadal oscillation exists and reached a maximum in the late 1980s. These assumptions are simply not justified by the evidence provided by these scalar series.

In light of these results, it may seem surprising that Vautard *et al.*, 1992, report skill to remarkably long lead-times in predicting the SSA-filtered IPCC series. While acknowledging that this area requires further investigation, we note that part of this skill is attributable to the fact that they are validating their forecasts not against raw data but against SSA-based filtered reconstructions of the IPCC series, as given by equation (13), (see their figure 16a). Predictions of filtered reconstructions on lead-times less than the window-width contain a strong element of artificial skill, for the following reason. Suppose we are attempting to forecast the 1993 value of the "interdecadal component" of the IPCC series (the filtered reconstruction using only EOFs 3 and 4 and a 40-year window), using only data prior to 1983. The quantity which we are attempting to forecast actually consists of the projection of the data from 1954 to 1993 onto EOFs 3 and 4, each multiplied by the final element of the corresponding EOF and added together. Thus, although no data after 1983 is used in the forecast, so there is no "look-ahead", three-quarters of the quantity being forecast actually consists of data prior to 1983, so some skill is inevitable.[8] Even though the forecasts appear to be "out-of-sample", a large fraction of the variance is, effectively, hindcast; simple tests on stochastic systems reveal how misleading this can be.

### 6.3 The extended Southern Oscillation Index

The dataset we now analyse is the 3-month averaged difference in pressure between Tahiti and Darwin, beginning in March 1866 and ending in February 1993. The data, kindly provided by P. D. Jones, are identical to those used in Jones, 1989, to compute a normalised SOI. Our analysis differs from that of Jones, 1989, in that we leave in the annual cycle; this allows us to demonstrate the algorithm described in section 4.4, by testing for subsidiary peaks in data containing a single dominant signal.

Since the spectral resolution of single-channel SSA is limited by $1/M$, we use $M = 60$ 3-month intervals, corresponding to a window-width of 15 years, to obtain sufficient resolution to distinguish the "quasi-biennial" (QB) and "quasi-quadrennial" (QQ) components of the ENSO signal (Rasmusson *et al.*, 1990; Jin *et al.*, 1994; Jiang *et al.*, 1995). Previous applications of SSA to the SOI (e.g. Keppenne & Ghil, 1992) have typically used windows of 50–60 months, applied to a shorter SOI series with the annual cycle removed. Using a 50–60-month window typically gives two EOF-pairs with associated frequencies of ∼26 and ∼52 months (Dettinger *et al.*, 1995). In the past, these have been interpreted as representing distinct spectral features but, because they are separated by almost exactly $1/M$, they are equally consistent with the interannual variability in the SOI being a broadband signal which is artificially split up by SSA into discrete frequencies. Allen, 1992,

[8] The origin of artificial skill in predictions of filtered reconstructions on lead-times less than the window width emerged through conversations with Chris Strong; Tom Mullin and coworkers have noted a similar effect.

applied a 120-month window to the standard 1935-85 SOI with the annual cycle removed and found the six leading EOFs formed three pairs with associated periods of 30, 41 and 62 months: 3 oscillatory pairs, again separated by almost exactly $1/M$.[9]

Figure 12 shows the extended SOI tested against a null-hypothesis of pure AR(1) noise. We use the basis corresponding to the null-hypothesis as described in section 5.2, to avoid artificial variance-compression. The spectrum is clearly dominated by the annual cycle, and since the noise here also includes variance due to the annual cycle, no other excursions occur above the 97.5$^{\text{th}}$ percentiles of the surrogate distributions.

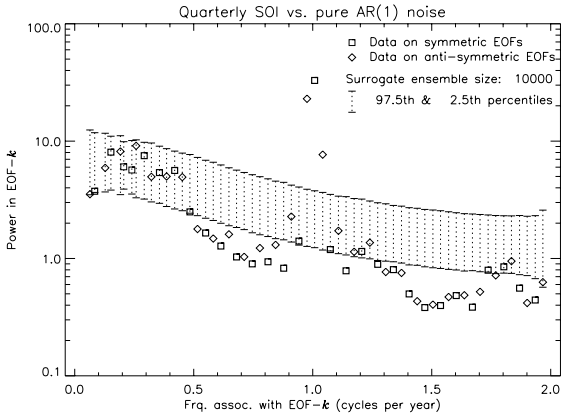


**Figure 12.** Testing the quarterly Southern Oscillation Index against pure AR(1) noise, using the EOFs of the null-hypothesis. We confidently reject this null hypothesis; the data is clearly dominated by the annual cycle, which we remove via the methods of section 4.2 before repeating the procedure to obtain figure 13.

Figure 13 shows the same data tested against a null-hypothesis of AR(1) noise added to the annual cycle using the eigenbasis of the composite covariance matrix defined in equation (26). With the AR parameters adjusted using the procedure in section 4.4, seven excursions above the 97.5$^{\text{th}}$ percentiles occur, out of a possible 58 (the EOFs corresponding to the annual cycle having been eliminated from the statistics). The probability of this many excursions is less than 0.8%, indicating that we can reject the hypothesis that the SOI consists of an annual cycle plus AR(1) noise at >99% confidence without any additional information. While not surprising, this demonstrates that the use of the null-hypothesis basis does not make Monte Carlo SSA unduly conservative.

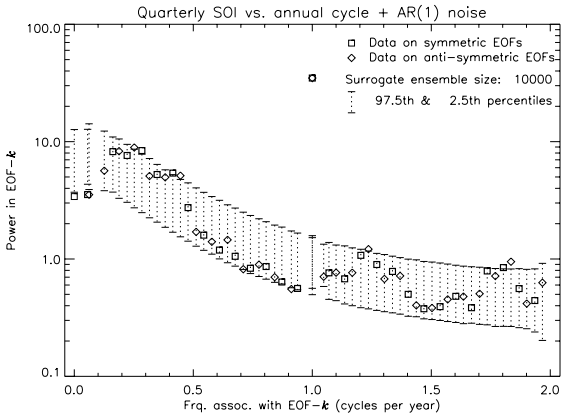


**Figure 13.** Testing the quarterly Southern Oscillation Index against the observed annual cycle plus AR(1) noise. With adjusted noise parameters, anomalously high variance is now indicated in EOFs corresponding to periods of 45, 28, 9.8 and 6.6 months.

The significant EOFs in figure 13 are located in four distinct spectral regions, two interannual and two subannual. Inspection of cross-correlations identifies them unambiguously with EOF-pairs 3 & 4, 9 & 10, 23 & 24 and 25 & 26 of the data covariance matrix. The periods associated with these pairs are 45, 28, 9.8 and 6.6 months respectively. If we include these into the null-hypothesis and repeat the test (figure 14) we observe excursions above the 97.5$^{\text{th}}$ percentiles in EOFs corresponding to frequencies below that of the QQ mode, consistent with the period of the QQ mode being much less well defined than that of the QB mode (Penland & Sardeshmukh, 1995).

This simple analysis of the extended SOI provides some support to the notion that the QQ and QB interannual components of ENSO are distinct spectral features: they both appear significant in figure 13, and are separated in the frequency domain by EOFs whose variance is well within that expected from the noise. Note, however, that QB mode only appears more significant than variability on 3-year timescales because we have been testing against AR(1) noise. Both spectral components contain similar variance, so tested against the white (or "locally white") null hypothesis, they would appear equally significant. Furthermore, the

[9] Another example of the tendency of SSA to split a broadband signal into discrete frequencies is found in Vautard *et al.*, 1992's analysis of the IPCC series. Using a 40-year window, they report five spectral peaks in the 5–10-year range, with associated periods of 9.6, 7.5, 6.2, 5.2 and 4.7 years. The separation of these peaks is $0.027 \pm 0.004$ cycles/year, which corresponds closely to $1/M = 0.025$ cycles/year. Note that a trend also introduces a bias towards peaks separated by $1/p$ in Maximum Entropy Method (MEM) spectral estimation, where $p$ is the number of poles in the MEM estimate.

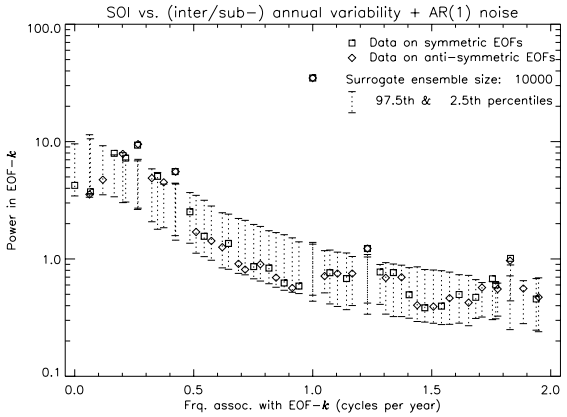


**Figure 14.** Testing the quarterly Southern Oscillation Index against the observed annual cycle, 45, 28, 9.8 and 6.6-month oscillations plus AR(1) noise. Further excursions above the 97.5$^{\text{th}}$ percentiles occur mainly in EOFs corresponding to frequencies below that of the 45-month "QQ" mode, indicating spectral broadening towards lower frequencies.

well-known phase-locking between ENSO and the annual cycle has not been taken into account. This could easily cause multiple frequencies to appear, separated by integer multiples of the annual cycle, in an underlying broadband phenomenon.

The two subannual components of the SOI, which would certainly have been missed had we relied on eigenvalue rank-order as a criterion of significance, also merit further investigation; it is interesting to note that Robertson *et al.*, 1995, also detect a 9–10-month signal in SSTs generated by a coupled model. Further investigation would have to take phase-locking into account to distinguish the effects of a non-sinusoidal annual cycle from those of a 9-10-month oscillation.

### 6.4 The Central England Temperature Series

The Central England Temperature record (CET series) is the longest set of instrumental temperature observations available: a monthly series from 1659 to the present has been obtained by combining the diaries of several observers (Manley, 1974) using indirect information and interpolation to fill any gaps (Parker *et al.*, 1991). Plaut *et al.*, 1995, have applied SSA to this series and report detecting evidence of oscillations with periods of 25, 14, 7.7 and 5.2 years against the null-hypothesis of white noise. Here, we address the question of whether SSA can distinguish between the CET series and a segment of AR(1) noise.

Figure 15 shows the CET series tested against a null-hypothesis of >40-year variability (contained in EOFs 1 & 2) plus AR(1) noise. For consistency with Plaut *et al.*, 1995, we use 12-month means of the CET series from March 1659 to February 1993 (335 years), $M = 40$, and the data-adaptive EOFs of standard SSA (hence the significance of high-ranked EOFs is enhanced through artificial variance-compression).

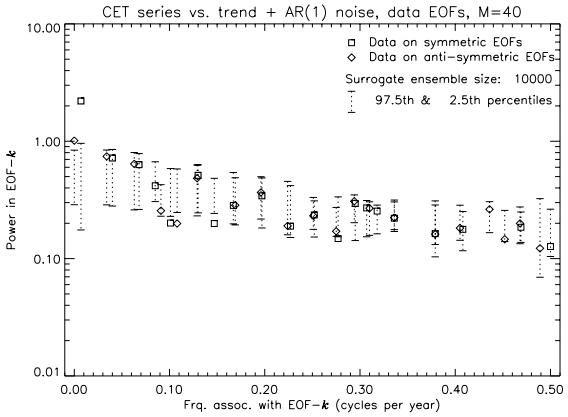


**Figure 15.** Testing the Central England Timeseries of annual temperatures from March 1659 to February 1993 against a null-hypothesis of >40-year variability and AR(1) noise. With this window width, SSA does not indicate interannual or interdecadal oscillations at the 97.5% confidence level.

EOF-pairs (adjacent, similar-variance squares and diamonds) are observed at frequencies of 0.04, 0.07, 0.13 and 0.20 cycles/year. These correspond to the four components identified by Plaut *et al.*, 1995, but none of them is significant even at the 95% level. No excursions occur above the 97.5$^{\text{th}}$ percentiles, and four occur below the 2.5$^{\text{th}}$ percentiles, but repeating the test using the null-hypothesis basis confirms that this is a consequence of artificial variance-compression "starving" the lowest-ranked EOFs of power (see figure 16). We conclude, therefore, that SSA with $M = 40$ does not provide evidence that the CET series is distinguishable from >40-year variability plus AR(1) noise.

We can increase the possible signal-to-noise enhancement in SSA by increasing the window width, at the cost of increasing the number of EOFs and thereby reducing the statistical significance of individual excursions outside the surrogate data bars. Figure 17 shows the CET series tested with a 100-year window. We now use the null-hypothesis basis, since with this longer window, the effects of artificial variance-compression are quite pronounced. One EOF with an associated period of 22 years now appears in the 98$^{\text{th}}$ percentile of the corresponding surrogate distribution, and two more ex-

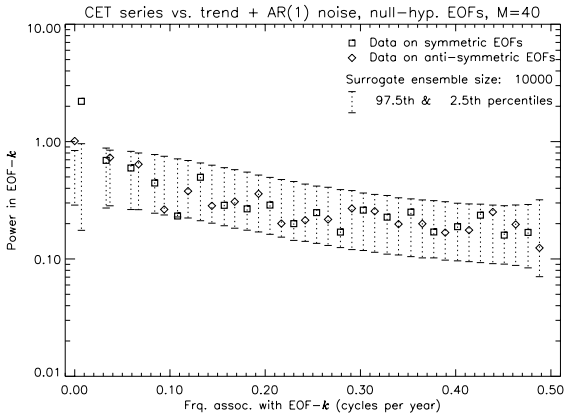


**Figure 16.** As previous figure, but using the EOFs of the null-hypothesis. Note how EOFs 3 & 4 and 5 & 6 are forced, by the constraint that they must be orthogonal to EOFs 1 & 2, to "pair up". No excursions occur below the $2.5^{\text{th}}$ percentiles, since there is no artificial variance-compression in this case.

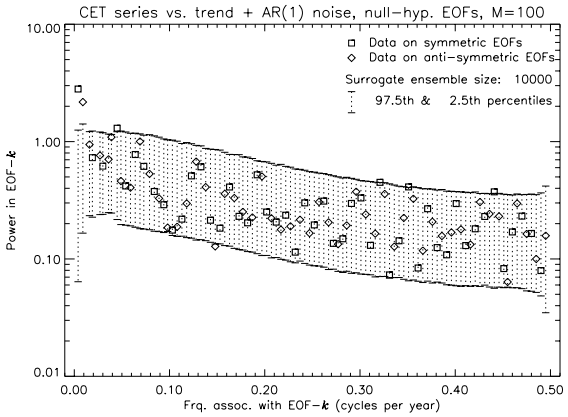


**Figure 17.** Testing the CET series against the null-hypothesis of >100-year variability and AR(1) noise using a 100-year window. The null-hypothesis basis must be used to avoid the effects of artificial variance-compression. Three EOFs appear significant at the 97.5% level, one indicating power at 22-year periods, but with $M = 100$ the probability of three or more excursions above the $97.5^{\text{th}}$ percentiles is >40%, so the trend plus AR(1) noise null hypothesis cannot be rejected on the basis of this data alone. No oscillations are indicated at 60–80 year periods.

cursions above the $97.5^{\text{th}}$ percentiles occur at higher frequencies. This might seem encouraging, but the probability of three or more excursions occurring above the $97.5^{\text{th}}$ percentiles is 42%, so on the basis of this data alone, we should still hesitate to conclude that the CET series is distinguishable from a trend-plus-AR(1)-noise. Variability on 14, 8 and 5-year time scales is consistent with this null-hypothesis.

The use of a longer window also allows us to examine evidence for oscillations on 60–80-year time scales. A 65–70-year global temperature oscillation, apparently originating in the North Atlantic, was recently reported by Schlesinger & Ramankutty, 1994, also through the application of SSA to the IPCC series, although Elsner & Tsonis, 1994b, observed that the results in Schlesinger & Ramankutty, 1994, were consistent with the hypothesis that the IPCC series consisted of a segment of AR(1) noise. More recently, Schlesinger & Ramankutty, 1995, used a version of Monte Carlo SSA, and the FMU eigenspectrum-shape test described above, to support their earlier claims. Since, however, they only tested the significance of the two highest-ranked EOFs without inspecting the remainder of the spectrum, their results remain inconclusive (it will always be possible to fit the variance in an EOF-pair using a 2-parameter noise model, so failure to do so – *i.e.,* rejection of the null-hypothesis on the basis of EOFs 1 & 2 alone – may simply indicate inadequate noise-model-specification).

A second reason why Schlesinger & Ramankutty, 1995's results require further investigation is that they were attempting to use SSA to examine possible oscillations with periods longer than the window width. They predicted the form of the trend in the IPCC series using an energy balance model, and tested whether the variance contained in EOFs 1 and 2 was consistent with the hypothesis of AR(1) noise plus a trend of this form. This approach cannot distinguish between a trend which is inconsistent with that predicted by the energy balance model and an oscillation with a period longer than the window-width: both would introduce unexplained variance into EOFs 1 and 2. It is therefore of interest to establish whether evidence for a 65–70-year oscillation can be found in a longer time-series.

Figure 17 indicates that the variance on 60–80-year time scales in the CET series is consistent with the hypothesis of long-term variability (>100-year time-scale) plus AR(1) noise. Since we would expect any oscillation in North Atlantic temperatures to have an impact on temperatures in Central England, this analysis does not support the suggestion that the unexplained non-linearity in the IPCC series is due to a 65–70-year oscillation originating in the North Atlantic. Analysis of other datasets may clarify this issue.

The conclusion of our analysis of the CET series is thus a conditional. If we use a relatively large window to maximise possible signal-to-noise enhancement, we do find there is more power on 22-year time scales than

we would expect (at the 97.5% confidence level) given a null-hypothesis of >100-year variability plus AR(1) noise. If we had prior reason to expect anomalous behaviour on this time scale in the CET series, then SSA would provide evidence in support of that expectation. In the absence of prior expectations, however, these results are consistent (at this confidence level) with the hypothesis of no significant interannual or interdecadal oscillations in the CET series.

## 7 Summary

SSA represents an extremely powerful analysis technique. Its applicability to non-stationary processes and phase- and amplitude-modulated oscillations makes it an ideal tool for the analysis of climate data. Yet the same properties which allow SSA to extract weak signals under unfavourable conditions also produce suggestively physical-looking patterns from pure noise. SSA must be used in conjunction with an adequate hypothesis-testing procedure.

Building on the foundations laid by Vautard & Ghil, 1989, this paper has illustrated these pitfalls and presented a methodology to avoid them, implementing suggestions made in Broomhead & King, 1986a. In section 3 we showed that, contrary to widespread current practice, the occurrence of a pair of sinusoidal EOFs $\pi/4$ out of phase with each other does not provide *prima facie* evidence for a physical oscillation. Indeed, EOFs will only fail to form such "oscillatory pairs" under very special circumstances (e.g. an infinite series of pure AR(1) noise). We also demonstrated that eigenvalue rank-order is not a reliable indicator of statistical or physical significance, except in those cases (rare in geophysics) when the stochastic component consists solely of white noise. The standard practice of "truncating the eigenspectrum", discarding all EOFs except those with the largest eigenvalues, is simply incorrect when employed to discriminate between signal and autocorrelated noise.

Stability of a result to varying the window width is not a sufficient condition for a physically significant signal; spurious EOF-pairs can be stable over a range of window widths. Likewise, confirming results using data from different sources cannot provide a substitute for a formal hypothesis-test: much of the "noise" in geophysical data consists genuine but unpredictable physical processes rather than observational errors, and therefore will be strongly correlated between contemporaneous datasets. A random (and insignificant) fluctuation in global temperatures, for example, would appear in many local series; analysis results from different series would be far from independent, making the overall significance level difficult to compute.

In section 4 we presented a method of distinguishing signals from arbitrary noise processes via SSA, based on the notion of "surrogate data". A Monte Carlo ensemble of surrogate series is generated using the null-hypothesis as a model, and a test is applied to establish whether it is possible to distinguish the data series from a member of the ensemble. In geometric terms, the test consists in asking, for each EOF, "does the data contain significantly more (or significantly less) variance in the direction in state-space defined by this EOF than we would expect if the null-hypothesis is true?" While we only consider the AR(1) noise null-hypothesis, the procedure is equally applicable to others: for example, identifying modes of variability in a dataset that are inconsistent with the behaviour of a climate model, treating the model as the source of "noise".

As with any analysis technique, Monte Carlo SSA is more complicated when the null-hypothesis we wish to test is not specified *a priori*. In sections 4.3 and 4.4 we address the problem of testing whether the data might arise from an unspecified AR(1) process or some deterministic signal plus an unspecified AR(1) process (*i.e.,* cases when the noise parameters – variance and lag-1 autocorrelation – are unknown). The approach we propose is a method of fitting AR(1) parameters to the data such that the process we test is, on some measure, that which is most likely to cause us to fail to reject the null-hypothesis. In this way, if we reject that process, we have reason to believe that all other AR(1) processes would also be rejected at the same or higher confidence level. The algorithm proposed makes it unnecessary to preprocess data to remove a trend or annual cycle before the analysis. Such preprocessing can lead to spurious results because the implicit response function of the preprocessing algorithm can masquerade as a physical signal.

The basic principle of surrogate data testing is that both data and surrogates must be treated in exactly the same way. Conventional SSA, however, selects the EOF basis which compresses the maximum possible variance in the data series into the highest-ranked EOFs, implicitly assuming that none of the data is noise. In section 5.2 we introduce a variant on SSA which is based on the assumption that all of the data is noise *except* that which we have established as signal. This, together with the Monte Carlo SSA test, gives us a signal detection algorithm in which, if the user-specified significance level is 97.5%, there really is a 2.5% chance of incorrect rejection of a valid null-hypothesis. To the best of our

knowledge, this property is not shared by any other signal-detection algorithm involving SSA.

In section 6 we demonstrate the application of these techniques to three well-known climatic time-series. We show that the 136-year IPCC series of global annual-mean near-surface temperatures is consistent with a non-linear trend with added AR(1) noise (*i.e.,* this null-hypothesis cannot be rejected at the 97.5%, or even the 50%, confidence level). The IPCC series does not, in itself, indicate either interannual or interdecadal oscillations. In contrast, the 126-year series of quarterly mean sea-level-pressure differences between Tahiti and Darwin shows a prominent annual cycle together with significant additional peaks at 4-year, 2-year, 9–10-month and 6–7-month time scales. These latter sub-annual peaks, which are clearly significant against an AR(1) noise null-hypothesis, would be missed by any analysis using variance or eigenvalue rank-order as a significance criterion. An analysis of the 335-year Central England Temperature series shows no interannual or interdecadal oscillations if a window width of 40 is used. Some evidence of an oscillation with a 22-year period emerges with a 100-year window, but this evidence is inconclusive since the probability of the observed number of excursions above the $97.5^{\text{th}}$ percentiles occurring purely by chance in all the EOFs tested is >40%. No evidence is found in support of a North Atlantic oscillation with a period of 65–70 years.

These negative results from our analysis of the two temperature records do not indicate any flaw in the observational records themselves; the correct conclusion to be drawn is simply that is difficult to establish that the low-frequency temperature variability of the climate system should be characterised in terms of oscillations, modulated or otherwise, solely on the basis of short scalar time-series. In short, the AR(1) model proves very hard to beat. In as much as AR(1) noise is exactly what we expect from a randomly-forced dissipative system with a finite heat capacity, this is not surprising.

SSA, as introduced by Vautard & Ghil, 1989, and extended by others, provides a much-needed tool in climate research. Monte Carlo SSA, as presented in this paper, hones this tool, allowing it to be applied to a significantly wider range of tasks. We have presented the technique in sufficient detail to make it generally available, but in conclusion, we stress that the basic idea is very simple, and equally applicable to conventional EOFs and related analysis techniques. We find the eigenvectors of an estimated covariance matrix, and use a Monte Carlo procedure to establish which, if any, of these eigenvectors account for more power in the data series than we would expect if the null-hypothesis is valid.

Many practitioners have come to distrust the results of formal statistical tests because of implausible significance claims that often result either from misspecification of the null-hypothesis or from inappropriate use of a test. Yet when an ostensibly innocent step, such as taking EOFs, can generate oscillatory patterns from pure noise, we cannot afford to rely on purely subjective criteria to identify signals of interest. Applications of time-series analysis techniques in geophysics have tended to focus on extracting weak signals which would otherwise be invisible in the noise. An equally important application is to tell us which of many all-too-visible patterns really indicate deterministic and predictable behaviour.

## Appendix: Parameterizing the distribution of surrogate projections

The Monte Carlo step to obtain the surrogate distributions represents the main computational burden of the algorithm described above. While necessary if we are dealing with complex null-hypotheses, such as "noise" which has been generated by a chaotic system, it may be eliminated if (i) the noise distribution is Gaussian (as is the case for AR(1) noise); (ii) the expected noise covariance matrix, $\mathbf{C}_N = \mathcal{E}(\mathbf{C}_R)$, is known analytically and (iii) the EOFs of interest are approximately sinusoidal. In this situation, the approximate distribution of the diagonal elements of $\mathbf{\Lambda}_R = \mathbf{E}^T\mathbf{C}_R\mathbf{E}$, where the columns of $\mathbf{E}$ are the data EOFs, the surrogate EOFs or some other orthonormal basis, can be calculated analytically. Dropping the $_R$ subscripts for clarity, we have

$$\Lambda_{k_1k_1} + \Lambda_{k_2k_2} = \sum_{k\in[k_1,k_2]} \sum_{i=1}^{M}\sum_{j=1}^{M} E^T_{ki} C_{ij} E_{jk}. \qquad (27)$$

If EOF-$k_1$ and EOF-$k_2$ form a pair of sinusoids in quadrature with angular frequency $\omega$, then with the VG summation convention for $\mathbf{C}$ (and with the BK convention, neglecting end effects),

$$\begin{aligned}
\Lambda_{k_1k_1} + \Lambda_{k_2k_2} &\simeq \sum_{i=1}^{M}\sum_{j=1}^{M} e^{-\imath\omega_k i} c_{(i-j)} e^{\imath\omega_k j} \\
&= \sum_{i=1}^{M}\sum_{j=1}^{M} c_{(i-j)} \cos\omega_k(i-j) \\
&= \sum_{\ell=1-M}^{M-1} (M-|\ell|) c_\ell \cos\omega_k\ell, \qquad (28)
\end{aligned}$$

$c_\ell$ being the estimated series covariance at lag $\ell$. The equality (28) is exact if and only if $\omega M = 2n\pi$ where $n$ is integer, $0 < n < M$, in which case $\Lambda_{k_1 k_1} = \Lambda_{k_2 k_2}$. At intervening frequencies, normalisation constraints mean that EOF-$k_1$ and EOF-$k_2$ cannot be a pair of equal-amplitude sinusoids, but this turns out to be relatively unimportant for estimating distributions.

Equation (28) is simply $M$ times the standard spectral estimate for a periodogram smoothed with a triangular (Bartlett, 1950) lag window. Its asymptotic distribution is given by

$$\Lambda_{kk} \sim \mathcal{E}(\Lambda_{kk}) \frac{\chi^2(\nu)}{\nu}, \tag{29}$$

where

$$\mathcal{E}(\Lambda_{kk}) = (\mathbf{E}^T \mathbf{C}_N \mathbf{E})_{kk}. \tag{30}$$

and the equivalent degrees of freedom, $\nu \simeq 3N/M$ (see, for example, Priestley, 1981, section 6.2). Since we use the normalised EOFs to compute $\mathcal{E}(\Lambda_{kk})$ in equation (30), the approximation in equation (28) plays a role only in the estimate of $\nu$, which is not exact in any case if $\omega M/2\pi$ is non-integer.

Surrogate projections onto non-sinusoidal EOFs are still chi-squared distributed, with a somewhat larger $\nu$. Equation (29) thus gives "surrogate data bars" which are approximately correct for sinusoidal EOFs and conservative for non-sinusoidal EOFs: when the EOFs of interest are sinusoidal, it provides a simple and computationally efficient alternative to the Monte Carlo procedure. Applied to the sample series used in this paper, testing against an AR(1) null-hypothesis, the chi-squared test gives results which are very similar to those estimated from a 10,000-member surrogate ensemble using the basis of the null-hypothesis. Differences are more apparent using the basis derived from the data, but overall conclusions are unchanged since the chi-squared approximation is reasonably accurate for the sinusoidal EOFs which we are primarily interested in.

The Monte Carlo procedure also estimates the probability of $n$ excursions above the $m^{\text{th}}$ percentile. For the AR(1) null-hypothesis, this is well approximated by the binomial distribution which we would expect if the excursions are independent, so it too can be parameterized. Again, the Monte Carlo procedure provides an essential "fall-back" if significance claims are in any doubt.

The number of degrees of freedom, $\nu$, in equation (29), is independent of the noise autocorrelation. This may appear counterintuitive (see, for example, Unal & Ghil, 1995) but it is consistent with the Monte Carlo results and with the spectral analysis literature. The explanation lies in the fact that the noise has been generated by a linear stochastic process and thus a lagged coordinate system exists in which the noise would be i.i.d. Introducing autocorrelation (transforming back from these lagged coordinates) scales the noise distributions but does not change their degrees of freedom. These issues are discussed in more detail in Allen & Smith, 1996.

An exact treatment of sampling uncertainty in SSA is complicated by the fact that $\mathbf{C}_D = \mathbf{D}^T\mathbf{D}$ does not conform to a standard Wishart distribution (Mardia *et al.*, 1979) (the "sliding window" algorithm implies that the rows of $\mathbf{D}$ are not independent). Equation (29) is reminiscent of the error formula for the eigenvalues of the data lag-covariance matrix proposed (in the context of spatial EOF analysis) by North *et al.*, 1982, and adapted for SSA as equation (3.1a,b) of Ghil & Mo, 1991. Those formulae, however, indicate the sampling uncertainty of the eigenvalues of $\mathbf{C}_D$, which may or may not reflect null-hypothesis-violating power in any particular EOF. In contrast, equation (29) describes the distribution of variance which we should expect in a given eigendirection on a specific null-hypothesis.

**Acknowledgments**

We would like to thank Phil Jones and David Parker for data; David Broomhead, Michael Ghil, Greg King and Robert Vautard for providing the primary motivation for this work; Alan Darbyshire, Mike Dettinger, Jim Elsner, Mike Mann, Cecile Penland, Guy Plaut, Peter Read, Chris Strong, Tasos Tsonis and Bill Weibel for helpful suggestions; and Chris Forest, Dennis Boccippio, James Theiler and the reviewers for useful comments on earlier drafts. Particular thanks are due to Pascal Yiou and Eric Breitenberger for thoughtful and meticulous independent implementations of the method and to Paul Billant and Andrew Robertson for motivating the work described in the appendix. MRA was supported by an Atlas Research Fellowship from the Rutherford Appleton Laboratory and Wolfson College, Oxford, and a NOAA Postdoctoral Global Change Fellowship hosted by Richard Lindzen at the Massachusetts Institute of Technology; LAS by a Senior Research Fellowship from Pembroke College, Oxford. FORTRAN source code and MATLAB routines (by Eric Breitenberger) for Monte Carlo SSA are available from the authors and via the SSA Toolkit project in UCLA.

## References


Allen, M. R. 1992. *Interactions between the atmosphere and oceans on time-scales of weeks to years.* Ph.D. thesis, University of Oxford. 202 pages.

Allen, M. R., & Robertson, A. W. 1996. Distinguishing modulated oscillations from coloured noise in multivariate datasets. *Climate Dynamics.* to appear.

Allen, M. R., & Smith, L. A. 1994. Investigating the origins and significance of low-frequency modes of climate variability. *Geophys. Res. Lett*, **21**, 883–886.

Allen, M. R., & Smith, L. A. 1996. Optimal filtering in Singular Systems Analysis. in preparation.

Allen, M. R., Read, P. L., & Smith, L. A. 1992a. Temperature oscillations. *Nature*, **359**, 679.

Allen, M. R., Read, P. L., & Smith, L. A. 1992b. Temperature timeseries? *Nature*, **355**, 686.

Allen, M. R., Mutlow, C. T., Blumberg, G. M. C., Christy, J. R., McNider, R. T., & Llewellyn-Jones, D. T. 1994. Global change detection. *Nature*, **370**, 24–25.

Bartlett, M. S. 1950. Periodogram analysis and continuous spectra. *Biometrika*, **37**, 1–16.

Berkooz, G., Holmes, P., & Lumley, J. L. 1993. The Proper Orthogonal Decomposition in the Analysis of Turbulent Flows. *Annu. Rev. Fluid Mech.*, **25**, 539–75.

Blackman, R. B., & Tukey, J. W. 1959. *The measurement of power spectra.* New York: Wiley.

Bretherton, C. S., Smith, C., & Wallace, J. M. 1992. An intercomparison of methods for finding coupled patterns in climate data. *J. Climate*, **5**, 541–560.

Broomhead, D. S., & Jones, R. 1989. Time-series analysis. *Phil. Trans. R. Soc. Lond.*, **423**, 103–121.

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

Broomhead, D. S., & King, G. 1986b. On the qualitative analysis of experimental dynamical systems. *Pages 113–144 of:* Sarkar, S. (ed), *Nonlinear Phenomena and Chaos.* Bristol: Adam Hilger.

Broomhead, D. S., Jones, R., & King, G. P. 1987. Topological dimension and local coordinates from time series data. *J. Phys. A*, **20**, L563–L569.

Dettinger, M. D., Ghil, M., Strong, C. M., Weibel, W., & Yiou, P. 1995. Software for Singular Spectrum Analysis of Noisy Timeseries. *EOS, Transactions of the American Geophysical Union*, **76(2)**, 12.

Elsner, J., & Tsonis, A. 1994a. Nonlinear Forecasting of the Evolution of ENSO. *Annales Geophysicae*, **12**(Supp. II, part II), C 528.

Elsner, J. B. 1995. Significance tests for SSA. *In: Proc. 19th Climate Diagnostics Workshop.* CAC/NOAA, U.S. Dept. of Commerce.

Elsner, J. B., & Tsonis, A. A. 1991. Do bidecadal oscillations exist in the global temperature record? *Nature*, **353**, 551–553.

Elsner, J. B., & Tsonis, A. A. 1994b. Low-frequency oscillation. *Nature*, **372**, 507–508.

Folland, C. K., Karl, T. R., Nicholls, N., Nyenzi, B. S., Parker, D. E., & Vinnikov, K. Ya. 1990. Observed climate variations and change. *Chap. 7, pages 195–238 of:* Houghton, J. T., Callander, B. A., & Varney, S. K. (eds), *Climate Change, The IPCC Scientific Assessment.* Cambridge Univ. Press.

Folland, C. K., Karl, T. R., & Vinnikov, K. Ya. 1992. Observed climate variability and change. *Chap. C, pages 135–170 of:* Houghton, J. T., Jenkins, G. J., & Ephraums, J. J. (eds), *Climate Change 1992, Supplement to the IPCC Scientific Assessment.* Cambridge Univ. Press.

Fraedrich, K. 1986. Estimating the dimension of weather and climate attractors. *J. Atmos. Sci.*, **43**, 419–432.

Fraedrich, K., & Ziehmann-Schlumbohm, C. 1994. Predictability experiments with persistence forecasts in a red-noise atmosphere. *Quart. J. Roy. Met. Soc.*, **120**, 387–428.

Ghil, M., & Childress, S. 1987. *Topics in Geophysical Fluid Dynamics: Atmospheric Dynamics, Dynamo Theory and Climate Dynamics.* New York: Springer.

Ghil, M., & Mo, K.-C. 1991. Intraseasonal oscillations in the global atmosphere – Part I: Northern Hemisphere and tropics. *J. Atmos. Sci.*, **48**, 752–779.

Ghil, M., & Vautard, R. 1991. Interdecadal Oscillations and the Warming Trend in Global Temperature Time Series. *Nature*, **350**, 324–327.

Graham, N. E., Michaelsen, J., & Barnett, T. P. 1987. An investigation of the El Niño – Southern Oscillation cycle with statistical models. 1. Predictor field characteristics. *J. Geophys. Res.*, **92**, 14251–14270.

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

Jiang, N., Neelin, J. D., & Ghil, M. 1995. Quasi-quadrennial and quasi-biennial variability in equatorial Pacific sea surface temperatures and winds. *Climate Dynamics*, **12**, 101–112.

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

Jolliffe, I. T. 1986. *Principal Component Analysis.* Springer.

Jones, P. D. 1989. The influence of ENSO on global temperatures. *Climate Monitor*, **17**, 80–89.

Jones, P. D., Raper, S. C. B., Bradley, R. S., Diaz, H. F., Kelly, P. M., & Wigley, T. M. L. 1986a. Northern Hemisphere surface air temperature variations, 1851-1984. *J. Clim. Appl. Met.*, **25**, 161–179.

Jones, P. D., Raper, S. C. B., Bradley, R. S., Diaz, H. F., Kelly, P. M., & Wigley, T. M. L. 1986b. Southern Hemisphere surface air temperature variations, 1851-1984. *J. Clim. Appl. Met.*, **25**, 1213–1230.

Keppenne, C. L., & Ghil, M. 1992. Adaptive spectral analysis and prediction of the Southern Oscillation Index. *J. Geophys. Res.*, **97**, 20,449–20,454.

Keppenne, C. L., & Ghil, M. 1993. Adaptive filtering and prediction of noisy multivariate signals: an application to

subannual variability in atmospheric angular momentum. *Intl. J. Bifurcation & Chaos*, **3**, 625–634.

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

Latif, M., & Barnett, T. P. 1994. Causes of decadal climate variability over the North Pacific/North American sector. *Science*, **266**, 634–637.

Latif, M., & Graham, N. E. 1992. How much predictive skill is contained in the thermal structure of an OGCM? *J. Phys. Oceanography*, **22**, 951–962.

Livezey, R. E., & Chen, W. Y. 1982. Statistical field significance and its determination by Monte Carlo techniques. *Mon. Wea. Rev.*, **111**, 46–57.

Lorenz, E. N. 1956. *Empirical orthogonal functions and statistical weather prediction.* Tech. rept. 1. Statistical forecasting project, Department of Meteorology, Massachusetts Institute of Technology.

MacDonald, G. J. 1989. Spectral Analysis of Time Series Generated by Nonlinear Processes. *Revs. of Geophysics*, **27**, 449–469.

Manley, G. 1974. Central England Temperatures: monthly means 1659 to 1973. *Quart. J. Roy. Met. Soc.*, **100**, 389–405.

Mann, M. E., & Lees, J. M. 1995. Robust estimation of background noise and signal detection in climatic time series. *Climatic Change*. submitted.

Mann, M. E., & Park, J. 1994. Global-scale modes of surface temperature variability on interannual to century timescales. *J. Geophys. Res.*, **99**, 25819–25833.

Mardia, K. V., Kent, J. T., & Bibby, J. M. 1979. *Multivariate Analysis*. Academic Press.

Mehta, V. M., & Delworth, T. 1995. Decadal variability of the tropical Atlantic ocean surface temperature in shipboard measurements and in a global ocean-atmosphere model. *J. Climate*, **8**, 172–190.

Newell, N. E., Newell, R. E., Hsiung, J., & Wu, Z. 1989. Global marine temperature variation and the solar magnetic cycle. *Geophys. Res. Lett*, **16**, 311–314.

North, G. R., Bell, T. L., Cahalan, R. F., & Moeng, F. J. 1982. Sampling errors in the estimation of Empirical Orthogonal Functions. *Mon. Wea. Rev.*, **110**, 699–706.

Paluš, M., & Dvořák, I. 1992. Singular value decomposition in attractor reconstruction: pitfalls and precautions. *Physica D*, **55**, 221–234.

Park, J., Lindberg, C. R., & Vernon, F. L. 1987. Multitaper spectral analysis of high-frequency seismograms. *J. Geophys. Res.*, **92**, 12675–12684.

Parker, D. E., Legg, T. P., & Folland, C. K. 1991. *A New Daily Central England Temperature Series 1772–1991.* Tech. rept. CTRN11. Hadley Centre, Meteorological Office, London Road, Bracknell, RG12 2SY, U.K. updated data kindly provided by D. E. Parker.

Penland, C., & Sardeshmukh, P. 1995. Error and sensitivity analysis of geophysical eigensystems. *J. Climate*, **8**, 1988–1998.

Penland, M. C., Ghil, M., & Weickmann, K. 1991. Adaptive filtering and maximum entropy spectra with application to changes in atmospheric angular momentum. *J. Geophys. Res.*, **91**, 22659–22671.

Plaut, G., & Vautard, R. 1994. Spells of low-frequency oscillations and weather regimes in the northern hemisphere. *J. Atmos. Sci.*, **51**, 211–220.

Plaut, G., Ghil, M., & Vautard, R. 1995. Interannual and interdecadal variability from a long temperature time series. *Science*, **268**, 710–713.

Preisendorfer, R. W. 1988. *Principal component analysis in meteorology and oceanography.* New York: Elsevier.

Priestley, M. B. 1981. *Spectral analysis and time series.* Vol. 1. Academic Press.

Rasmusson, E. M., Wang, X., & Ropelewski, C. F. 1990. The biennial component of ENSO variability. *J. Marine Syst.*, **1**, 71–96.

Read, P. L. 1993. Phase portrait reconstruction using multivariate singular systems analysis. *Physica D*, **69**, 353–365.

Robertson, A. W., Ma, C.-C., Mechoso, C. R., & Ghil, M. 1995. Simulation of the Tropical-Pacific climate with a coupled ocean-atmosphere general circulation model. Part II: Interannual Variability. *J. Climate*, **8**.

Schlesinger, M. E., & Ramankutty, N. 1994. An oscillation in the global climate system of period 65–70 years. *Nature*, **367**, 723–726.

Schlesinger, M. E., & Ramankutty, N. 1995. Is the recently reported 65–70 year surface-temperature oscillation the result of climatic noise? *J. Geophys. Res.*, **100**(D7), 13767–13774.

Smith, L.A. 1992. Identification and Prediction of Low-Dimensional Dynamics. *Physica D*, **58**, 50–76.

Smith, L.A. 1994. Local Optimal Prediction: Exploiting strangeness and the variation of sensitivity to initial condition. *Phil. Trans. R. Soc. Lond.*, **A 348**(1688), 371–381.

Stuart, A., & Ord, J. K. 1991. *Kendall's Advanced Theory of Statistics.* 5 edn. Vol. 2. London: Edward Arnold.

Theiler, J., & Eubank, S. 1993. Don't bleach chaotic data. *Chaos*, **4**(December), 1–12.

Theiler, J., Eubank, S., Longtin, A., Galdrikan, B., & Farmer, J. 1992. Testing for nonlinearity in time series: the method of surrogate data. *Physica D*, **58**, 77–94.

Thomson, D. J. 1982. Spectrum estimation and harmonic analysis. *Proc. IEEE*, **70**, 1055–1096.

Thomson, D. J. 1990. Time series analysis of Holocene climate data. *Phil. Trans. R. Soc. Lond.*, **A 330**, 601–616.

Tsonis, A. A., & Elsner, J. B. 1992. Oscillating global temperatures. *Nature*, **356**, 751.

Unal, Y. S., & Ghil, M. 1995. Interannual and interdecadal oscillation patterns in sea level. *Climate Dynamics*, **11**, 255–278.

Vautard, R., & Ghil, M. 1989. Singular Spectrum Analysis in nonlinear dynamics with applications to paleoclimatic time series. *Physica D*, **35**, 395–424.

Vautard, R., Yiou, P., & Ghil, M. 1992. Singular Spectrum Analysis: A toolkit for short, noisy and chaotic series. *Physica D*, **58**, 95–126.

Walker, A. M. 1954. The asymptotic distribution of serial correlation coefficients for autoregressive processes with dependent residuals. *Proc. Camb. Philos. Soc.*, **50**, 60–64.

Weare, B. C., & Nastrom, J. N. 1982. Examples of extended empirical orthogonal function analyses. *Mon. Wea. Rev.*, **110**, 481–485.

Yiou, P., Genthon, C., Ghil, M., Jouzel, J., Treut, H. Le, Barnola, J. M., Lorious, C., & Korotkevitch, Y. N. 1991. High-frequency paleovariability in climate and $CO_2$ levels from Vostok ice core records. *J. Geophys. Res.*, **96**, 20365–20378.

Yule, G. U. 1927. On a method of investigating periodicities in disturbed series. *Phil. Trans. R. Soc. Lond.*, **A. 226**, 267–298.

Zweirs, F. W., & von Storch, H. 1995. Taking serial correlation into account in tests of the mean. *J. Climate*, **8**, 336–351.

This preprint was prepared with the AGU LaTeX macros v3.0. File mcssajcl formatted 1997 January 21.