Climate Dynamics (1996) 12:775–784

Climate Dynamics

Distinguishing modulated oscillations from coloured noise in multivariate datasets

M. R. Allen[1]*, **A. W. Robertson**[2]

[1] NOAA Postdoctoral Program in Climate and Global Charge, Center for Meteorology and Physical Oceanography, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2] Department of Atmospheric Sciences and Institute of Geophysics and Planetary Physics, University of California, Los Angeles, CA 90095-1565, USA

Received: 13 September 1995 / Accepted: 4 April 1996

Abstract. Extended empirical orthogonal functions (EEOFs), alternatively known as multi-channel singular systems (or singular spectrum) analysis (MSSA), provide a natural method of extracting oscillatory modes of variability from multivariate data. The eigenfunctions of some simple non-oscillatory noise processes are, however, also solutions to the wave equation, so the occurrence of stable, wave-like patterns in EEOF/MSSA is not sufficient grounds for concluding that data exhibits oscillations. We present a generalisation of the "Monte Carlo SSA" algorithm which allows an objective test for the presence of oscillations at low signal-to-noise ratios in multivariate data. The test is similar to those used in standard regression, examining directions in state-space to determine whether they contain more variance than would be expected if the noise null-hypothesis were valid. We demonstrate the application of the test to the analysis of interannual variability in tropical Pacific sea-surface temperatures.

1 Extended EOFs and reduced principal components

Standard principal component analysis, or PCA (Lorenz 1956; Jolliffe 1986), involves sliding a wide, flat window (length and width $1 \times L$) over each field in a dataset **d** which consists of N fields each containing L datapoints ($d_{ts:\, t=1,N;\, s=1,L}$) and identifying spatial patterns (commonly known as EOFs) which account for a high proportion of the variance in the N views of the dataset thus obtained. Equivalently, PCA can be described as sliding a narrow, long ($N \times 1$) window across the input channels and identifying high-variance temporal patterns (PCs) in the corresponding L views.

Correspondence to: M. R. Allen

* *Permanent address:* Space Science Department, Rutherford Appleton Laboratory, Chilton, Didcot, OX11 0QX, UK and Department of Physics, University of Oxford, UK (*e-mail:* m.allen1@physics.oxford.ac.uk)

Since the signals of interest to climatologists often persist in time, a logical generalisation of standard PCA is to extend the spatial window to include elements of **d** at times $t+1$, $t+2$, up to some user-specified maximum lag, $t+M-1$ (for consistency with previous authors, we use M to refer to the number of lags included in the lag-window and L for the number of channels). We thus replace our $1 \times L$ window with an $M \times L$ window and search for spatio-temporal patterns which maximize variance in the $N' = N - M + 1$ overlapping views of **d** thus obtained. This is the extended EOF algorithm, early applications of which include Barnett and Hasselmann (1979) and Weare and Nastrom (1982). EEOF analysis is equivalent to performing conventional PCA on an augmented dataset, **D** ($D_{ij:\, i=1,N';\, j=1,LM}$), which consists of M lagged copies of **d** arranged thus:

$$D_{ij} = d_{i+l-1,s}. \tag{1}$$

The index i indicates window position (specifically, the time coordinate of the earliest points in the window) while j indicates position within the lag-window:

$$j = l + M(s-1), \tag{2}$$

where $l-1 = 0, M-1$ corresponds to time-lag from the beginning of the window and $s = 1, L$ corresponds to spatial position or input channel number. The elements of **D** correspond exactly to those in a three-dimensional array with indices i, l, s arranged in the conventional manner, showing how lag plays the role of an additional spatial dimension in EEOF analysis.

The singular value decomposition of **D**,

$$\mathbf{D} = \eta \mathbf{P}_D \Lambda_D{}^{1/2} \mathbf{E}_D^T, \tag{3}$$

where Λ_D is diagonal and η is a convenient normalisation equal to the larger of N' and ML, yields a set of orthonormal vectors of rank ML (the columns of $\mathbf{E}_D$) and a second set rank N' (the columns of $\mathbf{P}_D$). Various names have been used in the literature to refer to these vectors, including extended EOFs and PCs, temporal EOF/PCs and spatio-temporal EOF/PCs. Since standard PCA is a special case of EEOF analysis with $M = 1$,

we will use EOFs and PCs to refer respectively to all spatio-temporal patterns and timeseries of coefficients obtained in this way, making clear as necessary whether or not $M>1$.

In introducing Eq. (3) above, we followed the usual approach to EEOFs, stressing the idea of augmenting the spatial window to search for high-variance spatio-temporal patterns in N' views through an $M\times L$ window. An equally valid description of the algorithm is reducing the length of the PCs found in standard PCA from N to $N'=N-M+1$, and searching for high-variance temporal patterns in ML views through a window with dimensions $N'\times 1$. Thus we can think of the algorithm as identifying dominant spatio-temporal modes of variability (the EOFs) each of which involve a number of channels, or we can think of it as identifying purely temporal patterns (the PCs) which explain a high proportion of the variance in lagged copies of the individual channels without explicitly making use of cross-channel covariances. The fact that this complementary window is always one-dimensional proves useful in formulating a significance test for EEOFs.

Broomhead and King (1986) noted that the basic EEOF algorithm could be used to examine the qualitative dynamics of chaotic systems in noise-contaminated experimental data. Drawing on the signal-processing literature, they refer to the algorithm as singular systems analysis (SSA), generally qualified as multi-channel (MSSA) when $L>1$. Vautard and Ghil (1989) and Vautard et al. (1992) further developed the single-channel version of SSA, while MSSA has recently been implemented in the analysis of laboratory (Read 1993), meteorological (Keppenne and Ghil 1993; Plaut and Vautard 1994) and oceanographic (Jiang et al. 1995; Unal and Ghil 1995) data, as well as in the interpretation of results from climate models (Robertson et al. 1995a; Robertson 1996; Zorita and Frankignoul 1996).

An important property of SSA, first noted by Vautard and Ghil (1989), is that it may be used directly to identify modulated oscillations in the presence of noise. In the single channel case, any view of an oscillation through a window of width M can be described completely in terms of only two vectors, a sine and cosine with periods equal to the oscillation, provided the period is less than M and time scales of amplitude- and phase-modulation are much greater than M. Thus, if the variance of a series is dominated by such an oscillation, SSA will generate a pair of sinusoidal EOFs $\pi/2$ out of phase with each other. Likewise, in the multi-channel context, a stationary or propagating oscillatory mode can be represented by EEOF/MSSA as two spatio-temporal patterns which are sinusoidal in time and $\pi/2$ out of phase (Plaut and Vautard 1994).

Unfortunately, such sinusoidal EOF-pairs are also generated by finite realisations of a number of non-oscillatory processes (such as first-order autoregressive – AR(1) – noise) so their occurrence does not unambiguously indicate an oscillation. This problem led to the development of the formal Monte Carlo significance test for single channel SSA documented in Allen (1992), Allen and Smith (1994) and Allen and Smith (1996). Similar problems also occur in the multi-channel context, discussed in the following section, again prompting the need for a formal test which is the objective of this study.

2 Stable patterns due to noise

Why should we be concerned that an ostensibly stable and apparently oscillatory pattern (or pair of patterns) which we find in our data might in fact be attributable to noise? The eigenbases $\mathbf{P}_D$ and $\mathbf{E}_D$ in Eq. (3) are the eigenvectors of the lag-covariance matrices $\mathbf{C}_D^{(P)}\equiv\frac{1}{ML}\,\mathbf{D}\mathbf{D}^T$ and $\mathbf{C}_D^{(E)}\equiv\frac{1}{N'}\,\mathbf{D}^T\mathbf{D}$ respectively, with the k^{th} diagonal element of $\Lambda_D=\mathbf{P}_D^T\mathbf{C}_D^{(P)}\mathbf{P}_D$ $(=\mathbf{E}_D^T\mathbf{C}_D^{(E)}\mathbf{E}_D)$ being proportional to the power in the k^{th} PC (EOF). Except in the near-singular case of $N'=ML$, only the smaller of these matrices is full-rank (even this smaller matrix may be rank-deficient if the data is completely noise-free, but this situation need not concern us here). The eigenbasis of the larger matrix is necessarily underdetermined, with large numbers of its eigenvalues identically zero. To appreciate how physical-looking patterns which are stable between process realisations can be obtained from pure noise, it is helpful to break the algorithm down into three steps: we estimate the smaller of the two covariance matrices; diagonalise it to obtain a complete eigenbasis (PCs or EOFs); and then obtain the incomplete basis (EOFs or PCs) by projecting $\mathbf{D}$ onto the complete basis.

2.1 Stable PCs

The most commonly encountered situation when dealing with short time series, $N'<ML$, is also conceptually the simplest. The elements of $\mathbf{C}_D^{(P)}$ are given by

$$C_{Dij}^{(P)}=\frac{1}{L}\sum_{s=1}^{L}\left[\frac{1}{M}\sum_{l=1}^{M}d_{i+l-1,s}\,d_{j+l-1,s}\right]. \tag{4}$$

The term in square brackets is an estimate of the autocovariance of channel s lagged by $i-j$ in time. Thus the ij^{th} element of $\mathbf{C}_D^{(P)}$ is simply the average of the ij^{th} elements of L individual $N'\times N'$ single-channel lag-covariance matrices each of rank M (assuming $M<N'$). Recall the point made already concerning the one-dimensional complemetary window in EEOF/MSSA: no cross-channel covariance information is used in the estimation of $\mathbf{C}_D^{(P)}$.

Suppose $\mathbf{d}$ is generated by a set of independent AR(1) processes (red noise) in which temporal autocorrelation decays exponentially with increasing lag in each of the channels and the lag-1 autocorrelation in channel s is given by γ_s:

$$u_{ts}=\gamma_s u_{t-1,s}+\alpha_s z_{ts}, \tag{5}$$

where $c_{0s}=\alpha_s^2/(1-\gamma_s^2)$ is the variance in channel s and

z_{ts} is a unit-variance gaussian white noise. In this example,

$$(C_{Dij}^{(P)}) = \frac{1}{L} \sum_{s=1}^{L} c_{0s}\, \gamma_s^{|i-j|}, \tag{6}$$

where is the expectation operator. Since no cross-channel covariance information is used in the estimation of $\mathbf{C}_D^{(P)}$, the fact that the individual channels are mutually independent in Eq. (5) has no impact on the RHS of Eq. (6).

The eigenvectors of this matrix are sinusoidal so, despite the fact that this is a non-oscillatory input process, it will generate sinusoidal PCs. In the many-channel, large-window limit ($N' \ll ML$), with c_{0s} and γ_s the same in all channels, each PC will be associated with a single frequency, with the lowest frequencies corresponding to the largest eigenvalues, and the frequencies of adjacent PCs separated by $1/2N'$ (Vautard and Ghil 1989).

Orthogonality constraints coupled with the presence of a strong signal such as a trend or annual cycle can, however, force PCs which are not associated with any signal to form sine-cosine pairs (Allen 1992) and thus be mistakenly identified as an oscillation. Sampling fluctuations also cause PCs to pair up, for the following reason. We obtain the PCs by moving the one-dimensional complementary window a distance M over each channel. Chance fluctuations in a finite sample will always mean that there will be more power (but not significantly more) than the process average in some sinusoidal patterns for any position of this complementary window. The sliding window ensures that we are also likely to find such anomalous power in those same patterns phase-shifted by $\pi/2$ (Allen and Smith 1996). Thus, we should expect to obtain pairs of sinusoidal PCs $\pi/2$ out of phase with each other through chance fluctuations in a non-oscillatory red noise process. The presence of such a PC-pair is not sufficient grounds to conclude that the data exhibits an oscillation. Because red noise projects more variance onto the lowest frequencies, low-frequency PC-pairs which are entirely due to noise will appear high in the eigenvalue rank-order.

2.2 Stable EOFs

By definition, EOF-k is the projection of the augmented dataset $\mathbf{D}$ onto PC-k. Thus if PC-k and PC-$k+1$ are a pair of sinusoids in quadrature, EOF-k and EOF-$k+1$ will also be sinusoidal in time $\pi/2$ out of phase. Sinusoidal PCs in a sliding window act as narrow-band filters, so the temporal structure of EOFs obtained by the EEOF/MSSA algorithm, and any reconstruction of the data based on a small number of PCs, will necessarily look like an oscillation. The probability of generating sinusoidal EOFs and PCs is enhanced if we use the algorithm described in Plaut and Vautard (1994) to impose Toeplitz structure on the lag-covariance matrices (we do not do this in any of the examples shown here). The reason for this is that the eigenvectors of the Toeplitz matrix are necessarily either symmetric or anti-symmetric and converge either towards pure sinusoids or towards discrete Legendre polynomials (depending on the properties of the generating process) in the long series/short window limit (Gibson et al. 1992).

We therefore expect the temporal structure of EOFs to be stable between process realisations even if they are due to noise. But what if their spatial structure is also stable? This, unfortunately, is equally inconclusive. Suppose, for example, the c_{0s} and γ_s are different in the different channels in Eq. (5). Different channels may dominate variance on different time scales (think of a superposition of exponentials with different intercepts and decay times). In this case, the expected channel-dependence (spatial structure) of each EOF will simply be a unit vector centred on the channel which contributes most variance at the time scale picked out by the corresponding PC. This channel-dependence will be stable between realisations of the generating process.

A common check on the robustness of results from EEOF/MSSA is to split the dataset in two and compare results obtained from the two halves independently. Quite apart from the undesirable subjectivity of a test based on pattern-similarities, sinusoidal EOFs and PCs may be perfectly stable even if they are generated by pure noise.

Similar arguments apply in the case $ML < N'$, although cross-channel covariances now play a larger role. The elements of $\mathbf{C}_D^{(E)}$ are given by

$$C_{Dij}^{(E)}) = \frac{1}{N'} \sum_{i'=1}^{N'} d_{i'+l_1-1,s_1} d_{i'+l_2-1,s_2} \tag{7}$$

with $i(l_1,s_1)$ and $j(l_2,s_2)$ defined as in Eq. (2) above. Thus $(\mathbf{C}_{Dij}^{(E)})$ is the covariance between channels s_1 and s_2 lagged by $l_1 - l_2$ in time. If the channels are mutually independent at all lags, then (apart from possible rotations due to degeneracies) the channel-dependence of the EOFs will again be the unit vectors, while their time-dependence [for an AR(1) process in the long-series limit $ML \ll N'$] will be sinusoids with associated frequencies separated by $1/2M$. Sampling fluctuations can, however, cause EOFs to pair up at frequencies not associated with any signal. If the channels all have equal variance but are not independent, with cross-channel correlations declining exponentially with separation, then the EOFs will appear as solutions to the two-dimensional wave equation in an $M \times L$ domain (Thacker 1995). Again, we have stable wave-like patterns with high-ranked eigenvalues being generated by a non-oscillatory process.

Another common method of evaluating the robustness of an oscillation is to estimate its damping time, which may be thought of as the e-folding time of amplitude- and phase-modulation. If this damping time is greater than or comparable to the period then the oscillation is regarded as genuine. Extreme caution must be used in interpreting the damping times of signals reconstructed using the EEOF/MSSA algorithm. A slid-

ing-window-based reconstruction is, by construction, an order-M moving-average (MA) process, so it will necessarily have an e-folding time at least comparable to M regardless of the true e-folding time of the underlying process (Box and Jenkins 1976).

Plaut and Vautard (1994) stress the difficulty of testing the statistical significance of individual EOFs and PCs in MSSA, and recommend that the technique should always be used in conjunction with conventional Fourier techniques. The preceding discussion supports the conclusion that informal significance tests are likely to be misleading. Our aim in this study is to demonstrate that it is possible to formulate an objective significance test within the framework of EEOF/MSSA. While we do not, of course, recommend complete reliance on any single statistical test, this should enhance the utility of the EEOF/MSSA algorithm in signal detection applications.

We have found that the sinusoidal appearance of a (pair of) pattern(s) and/or its stability between realisations are not sufficient grounds to reject the hypothesis that it is entirely attributable to non-oscillatory stochastic processes. The standard practice of "truncating the eigenspectrum" (retaining only the EOF/PCs corresponding to the largest eigenvalues) is also invalid as a method of separating oscillatory signals from non-oscillatory noise, since low-frequency EOFs and PCs which are entirely due to noise may have high-ranked eigenvalues. Instead, we base our procedure, previously outlined in Robertson et al. (1995b), on this simple question: does this pattern contain more variance than we would expect if the data were generated by noise?

3 An objective hypothesis-testing procedure

3.1 Specifying the null-hypothesis

The test we propose can be applied to a wide range of null-hypotheses. Following the surrogate data approach to the analysis of non-linear systems (Smith 1992; Theiler et al. 1992), our main requirement is that it should not be possible to reject the null-hypothesis a priori or by the application of a trivial statistical test. The white noise null-hypothesis, for example, is generally unsuitable for geophysical problems, since few geophysical systems are capable of generating white noise output (Hasselmann 1976). Moreover, the hypothesis of no temporal or spatial correlation can generally be rejected by simple inspection of most geophysical datasets (Livezey and Chen 1982). Thus rejection of the white noise hypothesis through the application of a sophisticated statistical test provides us with no new information.

As a demonstration, we focus on a particular null-hypothesis which is appropriate when the L input channels are variance-weighted PCs obtained from standard PCA. Pre-filtering with standard PCA is a common procedure in EEOF/MSSA (Plaut and Vautard 1994) which becomes computationally essential when large datasets are involved. At present, we have no formal check that the signals of interest are completely described by the L highest-variance PCs obtained in standard PCA so we simply vary the number of PCs retained to ensure that results are insensitive to this parameter.

The L input channels are pairwise uncorrelated at lag-0, since the PCs of standard PCA are mutually orthogonal. If, as in many applications, the data being analysed consist of anomalies about a climatology consisting of statistical means derived from the input data, the input channels will also be centred (i.e. sample means will be zero by construction). Our null-hypothesis, therefore, is that the data have been generated by L independent AR(1) processes, being the model (5) with the α_s and γ_s chosen such that the u_{ts}, after centring the individual channels, have the same variance and time-lag-1 autocorrelation as the centred input channels, d_{ts}. It is important to take into account the effect of centring if we are dealing with anomaly data, since failure to do so increases the probability of incorrect detection of spurious low-frequency oscillations in pure noise (Allen and Smith 1996).

We focus on the null-hypothesis of L independent AR(1) processes because, following the philosophy of surrogate data testing, it reproduces basic linear statistics of the dataset while being incapable of supporting the behaviour which we are interested in detecting (oscillations). We wish to minimise the probability of rejecting the null-hypothesis for the wrong reason, but we also want to maximise the chance of rejection if the data does in fact display oscillations. For this reason, the null-hypothesis of an L-channel multivariate AR(1) process is inappropriate. A multivariate AR(1) process can itself support oscillations (this is, indeed, the principal oscillation pattern model: von Storch et al. 1995) so it is inappropriate as a null-hypothesis in a test for oscillatory behaviour.

The key advantage of using PCs from standard PCA as input channels d_{ts} is that it provides a natural way of taking into account the fact that temporal autocorrelation tends to be spatial-scale-dependent in geophysical timeseries: large-scale patterns tend to have long decorrelation times, even for non-oscillatory processes. Because of this fact, it is generally easy (and uninteresting) to reject the null-hypothesis that the d_{ts} are indistinguishable from a set of AR(1) processes with a uniform decorrelation time.

3.2 Formulating the test (I): using the PCs and EOFs of the data

Having formulated a model of the null-hypothesis we generate a large ensemble of surrogate data segments, $\mathbf{u}$, each with the same length and number of channels as $\mathbf{d}$. If $N' < ML$ ($N' > ML$) we compute, for each surrogate segment, the lag-covariance matrix $\mathbf{C}_R^{(P)}$ ($\mathbf{C}_R^{(E)}$) and the matrix of projections $\Lambda_R \equiv \mathbf{P}_D^T \mathbf{C}_R^{(P)} \mathbf{P}_D$ ($\mathbf{E}_D^T \mathbf{C}_R^{(E)} \mathbf{E}_D$). The k^{th} diagonal element of Λ_R indicates the power in the state-space-direction defined by the k^{th} PC (EOF) in this surrogate segment. Note that

$(\mathrm{tr}(\Lambda_R)) = \mathrm{tr}(\Lambda_D)$. This is the reason complete eigenbases must be used throughout, since if we used the incomplete basis, all data variance would be concentrated in the highest ML (N') PCs (EOFs), whereas surrogate variance would also be distributed over other eigen-directions.

We store the diagonal elements of the Λ_R and compute their distribution statistics by summing over the surrogate ensemble. If the k^{th} diagonal element of Λ_D is larger than 97.5% of the k^{th} diagonal elements of the Λ_R, we would conclude, at the 97.5% confidence level, that there is more power in $\mathbf{d}$ in this state-space-direction (PC or EOF) than we would expect on this null-hypothesis.

Following standard practice in spectral analysis (see, for example, Thomson 1990), we quote confidence levels appropriate to the local tests of the individual PCs or EOFs, but it should be born in mind that we expect an average of 2.5% of the diagonal elements of the Λ_R themselves to lie above their 97.5^{th} percentiles, so simply observing a small number of such excursions would not be sufficient grounds for rejection of the null-hypothesis without more information (von Storch 1982; Livezey and Chen 1982). These issues are discussed in detail in Allen and Smith (1996). There is a second reason why we should be cautious in interpreting results from any test based on PCs or EOFs derived from the data which is addressed in the following subsection.

Allen and Smith (1996) remark that the Monte Carlo procedure can be replaced with a simple χ^2 test in the single-channel case provided we are dealing with a normally distributed null-hypothesis and near-sinusoidal PCs or EOFs. A similar approach could be taken here, exploiting the fact that the diagonal elements of the Λ_R will be distributed as the sum of L χ^2-distributed variables, each with approximately $3N/N'$ degrees of freedom, weighted by the individual channel variances. If all the channels are of equal variance, this is simply χ^2-distributed with $3LN/N'$ degrees of freedom, but with unequal channel variances, the distribution is more complicated. As a rough guide, we can assume the Λ_R are approximately χ^2-distributed with the number of degrees of freedom lying between $3N/N'$ (the case when one channel dominates) and $3LN/N'$ (the case when all channels contain equal variance). This may be useful as an initial data-screening procedure, but the Monte Carlo approach gives the simplest and most robust significance estimates for use in reporting results.

3.3 Formulating the test (II): using the PCs and EOFs of the null-hypothesis

Allen and Smith (1996) argue that when we are uncertain whether or not the oscillation under investigation actually exists the data-adaptive eigenbasis provided by $\mathbf{P}_D$ or $\mathbf{E}_D$ is not a reliable analysis tool. By construction, it is tailored to compress the maximum possible amount of variance in this particular dataset into the highest-ranked PCs/EOFs. So we should expect to find anomalously high variance in these PCs/EOFs even if the data consists of a segment of pure noise. The extent of this artificial variance compression is difficult to predict except for very simple processes such as white noise. A more rigorous approach is to use the PCs/EOFs appropriate to the null-hypothesis to decide whether or not the null-hypothesis can be rejected, and only introduce the PCs/EOFs of the data when we are confident we have a signal which is worth characterising.

We obtain this null-hypothesis basis by computing the expected lag-covariance matrix of the surrogates, $\mathbf{C}_N^{(P)}$ or $\mathbf{C}_N^{(E)}$, defined as the matrix which we would obtain by averaging the $\mathbf{C}_R^{(P)}$ or $\mathbf{C}_R^{(E)}$ over a large surrogate ensemble. Often, this matrix may be known analytically, see Eq. (6). Unlike the $\mathbf{C}_D$, the $\mathbf{C}_N$ will generally both be of full rank. Nonetheless, we should still use the smaller of the two, to simplify interpretation of results.

We diagonalise $\mathbf{C}_N^{(P)}$ ($\mathbf{C}_N^{(E)}$) to obtain a set of null-hypothesis PCs (EOFs), $\mathbf{P}_N$ ($\mathbf{E}_N$) and proceed exactly as before, projecting both data and surrogates onto these vectors: $\Lambda'_D \equiv \mathbf{P}_N^T \mathbf{C}_D^{(P)} \mathbf{P}_N$ ($\mathbf{E}_N^T \mathbf{C}_D^{(E)} \mathbf{E}_N$), and similarly for Λ'_R. For the null-hypothesis of L independent AR(1) processes, and for most null-hypotheses in which autocorrelation decays monotonically with time, the null-hypothesis PCs and the temporal structure of the null-hypothesis EOFs will be close to sinusoidal, so it is straightforward to associate a time scale with an individual PC or EOF. If the k^{th} diagonal element of Λ'_D is anomalously high relative to the ensemble-derived distribution of the corresponding elements of the Λ'_R, this indicates anomalous power in the data-set on this timescale against this null-hypothesis, which would be evidence for an oscillation.

Using the PCs or EOFs of the null-hypothesis provides us with a more conservative test for the presence of a modulated oscillation which is not subject to the problem of artificial variance compression. It does not provide an optimal basis for the characterisation and reconstruction of the null-hypothesis-violating behaviour. We recommend, therefore, using the null-hypothesis basis to establish whether or not EEOF/MSSA indicates any evidence of an oscillation, and then examining the data basis for a pair of PCs or EOFs which characterise that oscillation. We can associate vectors between eigenbases either by computing cross-products (as in Allen and Smith 1996) or, as in the examples presented here, by finding dominant associated frequencies from the reduced Fourier transform (Vautard et al. 1992).

4 An example: tropical Pacific SST

We present, as an example, an analysis of sea-surface temperatures (SSTs) for the period 1951 to 1992 in the equatorial Pacific region 20°N to 20°S and 150°E to 80°W. The dataset used is the GISST data from the UK Meteorological Office (Parker et al. 1994). Since

we are interested in interseasonal and interannual variability, we use 4°×4°×3-month means as our input data, expressed as anomalies about the 1951–80 seasonal climatology with the statistical mean at each point removed prior to the analysis.

We initially perform a conventional PCA of the GISST data and retain 10 conventional PCs which account for 77% of the variance. These PCs, weighted by their singular values (so that variances reflect variance in the original data), provide the L input channels for the EEOF/MSSA algorithm. Increasing L to 20 has only a negligible effect on results. We have 42 years of data, so $N=168$. Our input channels are centred, pairwise uncorrelated at lag-0, and have different variances, c_{0s}, and lag-1 auto-correlations, γ_s, so we test the hypothesis of L independent AR(1) processes, individually centred, with the same c_{0s} and γ_s as the input channels.

Following Jiang et al. (1995) we initially use a window-width of ~5 y ($M=21$, so $N'=148$ – more on the choice of window below). Figure 1 shows the result, projecting both data and surrogates onto the data PCs, $\mathbf{P}_D$. Note that $N'<ML$ in this example, so it is the PCs which provide our complete basis. Diamonds show the diagonal elements of Λ_D, plotted against the dominant frequency of the corresponding PC. Vertical bars show the 2.5$^{\text{th}}$ and 97.5$^{\text{th}}$ percentiles of the distributions of the corresponding elements of Λ_R. The first thing to notice about this figure is that very few of the data eigenvalues lie above their corresponding surrogate data bars, despite the fact that the surrogates have the same expected variance as the data and we have used the data-adaptive basis, $\mathbf{P}_D$, which we expect to compress data variance into high-ranked PCs. Two data eigenvalues, with associated frequencies 0.023 and 0.017 months^{-1} (periods 43–58 months) are clearly visible in the upper left-hand corner of the plot well separated from their corresponding surrogate data bars; one more, with a lower associated frequency, is also visible in that region (consistent with spectral broadening of the low-frequency component of El Niño: Penland and Sardeshmukh 1995); and a further pair, with associated frequencies both around 0.043 months^{-1} (period 23 months) is also indicated as significant.

Out of a possible 148 PCs, the surrogate data test against red noise has picked out two pairs associated with the quasi-biennial (QB) and quasi-quadrennial (QQ) components of El Niño. If, following Allen and Smith (1996), we were to incorporate these signals into the null-hypothesis and readjust the noise parameters to exclude variance associated with these oscillatory modes, more signals might emerge, but our main point here is to stress that the test has pickled out a relatively small number of components as containing improbably high variance. If we were to use eigenvalue rank-order as a criterion of significance, we would have to retain the highest 14 PCs in order to include the QB component. The surrogate data test indicates that 9 of these contain no more variance than we would expect if the data consists of a set of independent AR(1) processes. Testing against red noise, therefore, promises a much more economical description of the data, in terms of two oscillatory PC-pairs and AR noise, rather than 14 or more signal PCs plus noise.

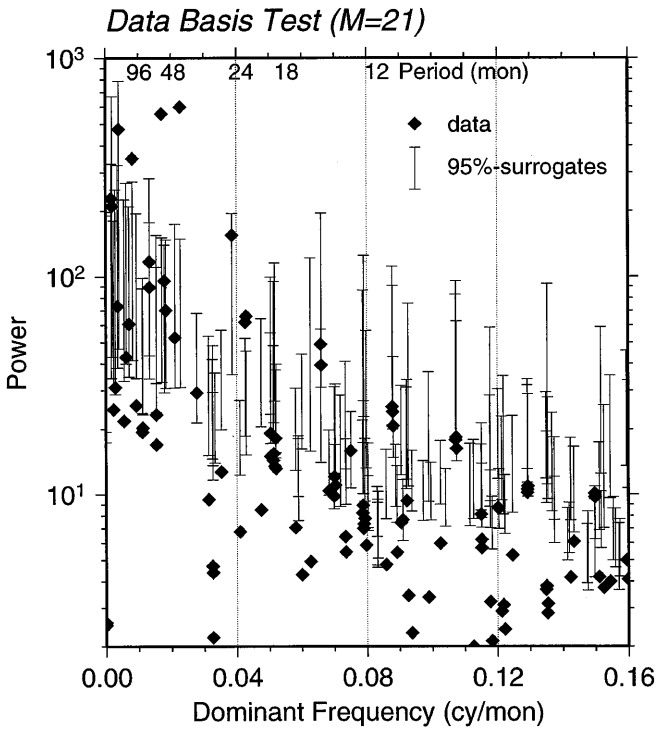

Fig. 1. Test of GISST tropical Pacific SST (3-month means with the seasonal cycle removed) for the period 1951–1992, using $L=10$ PCs from a conventional PCA (76.8% of the total variance) as the input channels. The spatial domain is 20°N–20°S, 150°E–80°W, with the 1°-resolution GISST data averaged into 4°×4° boxes. The data-adaptive basis $\mathbf{P}_D$ is used, with a ~5-year window: $M=21$, $N'=148$. *Diamonds* show the data eigenvalues, Λ_D, plotted against the dominant frequency associated with the corresponding PC, computed via a reduced Fourier transform. The vertical surrogate data bars span the 2.5$^{\text{th}}$ to 97.5$^{\text{th}}$ percentiles of the corresponding diagonal elements of Λ_R computed from 1000 realisations of a noise model consisting of L independent AR(1) processes with the same γ_s and c_{0s} as the input data channels. Two PCs are clearly significant in the *upper left corner,* with frequencies around 0.02 (50 months), and a *second pair* are above the 97.5$^{\text{th}}$ percentiles at 0.034 (23 months). One other low-frequency PC is also indicated

A striking feature of Fig. 1 is the large number of low-ranked data eigenvalues which lie below the 2.5$^{\text{th}}$ percentiles of their corresponding surrogate data bars. The reason is the artificial variance compression effect noted above. In this example, we have $N'=148$ and $ML=210$, so $\mathbf{C}_D^{(P)}$ is relatively close to singular. Variance compression is therefore very pronounced, and starves low-ranked PCs of power. We must, therefore, confirm results using the null-hypothesis basis, for which artificial variance compression is not a problem.

Results are shown in Fig. 2: again, the QQ and QB frequencies are picked out as containing improbably high variance, although the QB mode appears at a slightly lower frequency (25–30 months) and they are no longer separated by data eigenvalues which are

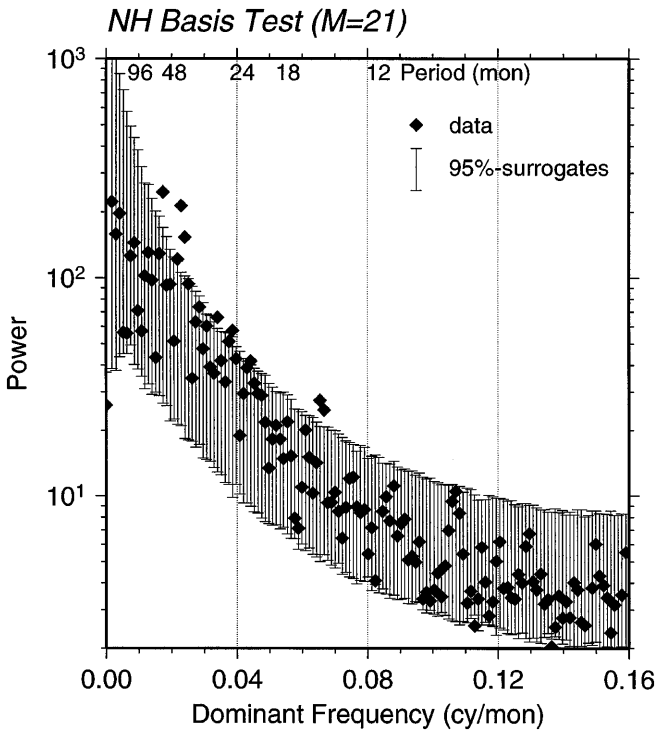

Fig. 2. As Fig. 1 but using the null-hypthesis basis, $\mathbf{P}_N$, derived from the expected lag-covariance matrix of the surrogates, $\mathbf{C}_N^{(P)}$

clearly not significant. We should expect leakage of power between frequencies to be quite pronounced, since we know that El Niño signals are amplitude- and phase-modulated. But the use of this basis allows us to confirm that the anomalous power which we observe on QQ and QB time scales with the data-adaptive basis cannot be attributed solely to artificial variance compression. We also observe anomalous power at 15-month periods, which was not picked out as significant by the data-adaptive basis. Although we generally expect the test based on the null-hypothesis basis to be the more conservative, there may be cases in which a signal is obscured in the data-adaptive basis due to degeneracy in the eigen-decomposition and the use of an alternative basis may bring it out.

A total of 7 excursions above the 97.5$^{\text{th}}$ percentiles are observed in Fig. 2. If these excursions are assumed to be mutually independent (which they are not), the probability of 7 or more excursions, given by the binomial distribution, is approximately 8%. Allen and Smith (1996) propose a two-pass Monte Carlo approach to computing the probability of a given number of excursions which should be used if precise significance levels are a priority. The binomial approximation is, however, sufficient to demonstrate that, without prior knowledge of which frequencies are of interest, the overall confidence level at which we can reject the red noise null-hypothesis is significantly lower than 97.5%.

It would also still be premature to conclude that the QQ and QB components represent distinct modes of El Niño, since the frequencies associated with them in the data-adaptive basis are separated by almost exactly $1/M$. When we are dealing with two oscillatory signals whose spatial components (channel-dependence) are not mutually orthogonal, $1/M$ is the limiting frequency resolution of the data-adaptive basis in EEOF/MSSA, recall that $\mathbf{C}_D^{(P)}$ is the average of L single-channel lag-covariance matrices each of rank M. This is a common problem in EEOF/MSSA and SSA-based analyses of El Niño, see, for example, Rasmusson et al. (1990), Keppenne and Ghil (1992), Jiang et al. (1995). At five-year window is typically used, and a 4–5-year and a $\sim$2-year component identified. These have generally been thought of as distinct modes but, since they are invariably separated by something close to the frequency resolution of the algorithm, they are equally consistent with a broad-band signal spanning the 2–5-year range.

The observation that two components may have been prematurely identified as distinct modes in these previous studies does not mean that they are not, in fact, distinguishable given adequate spectral resolution. Indeed, Allen and Smith (1996) use a 15-year window on an extended Southern Oscillation Index, and find some evidence of significant power at 2-year and 4–5-year periods, separated by frequencies which do not contain more power than we would expect on a red noise null-hypothesis. They urge caution in the interpretation of this result, since no account has been taken of the phase-locking between El Niño and the annual cycle.

A full analysis taking phase-locking into account would take us beyond the scope of this study, so we simply demonstrate a second application of the surrogate data test, now with $N'=60$ 3-month intervals (15 years), thus $M=109$ ($\sim$27 years). With both N' and M now close to half the length of these series, we approach the upper limit of spectral resolution available in EEOF/MSSA when spatial patterns are not mutually orthogonal. Results are shown in Fig. 3 for the data-adaptive basis. Note first how the artificial variance-compression effect is still present but much reduced. With $N' \ll ML=1090$ the matrix $\mathbf{C}_D^{(P)}$ is no longer close to singular.

Anomalous power against the red-noise hypothesis is still observed in the QQ and QB frequencies, and we also observe data eigenvalues near the 97.5$^{\text{th}}$ percentiles at 15-month and 9-month periods. Although these signals should not be taken as significant in themselves, they are intriguing, since Robertson et al. (1995b) and Robertson et al. (1995a) report power at these frequencies in a coupled model simulation of El Niño. Note that the QQ mode is no longer characterised by a PC-pair: one member of the pair has evidently been scrambled in the eigen-decomposition through degeneracy with some other low-frequency component.

Figure 4 shows the analysis with $N'=60$ using the eigenbasis of the null-hypothesis. Again, the QQ and QB frequencies are indicated as significant, as are a number of PCs at intermediate frequencies. The 15-month and 9-month periods are also indicated as potentially of interest, but not clearly significant. A num-

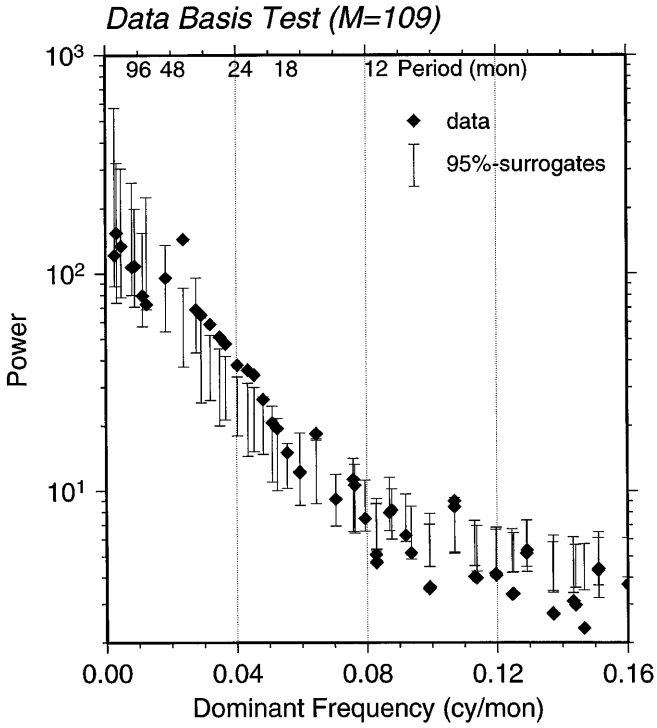

Fig. 3. As Fig. 1 but using the data-adaptive basis with $N' = 60$ (15 years)

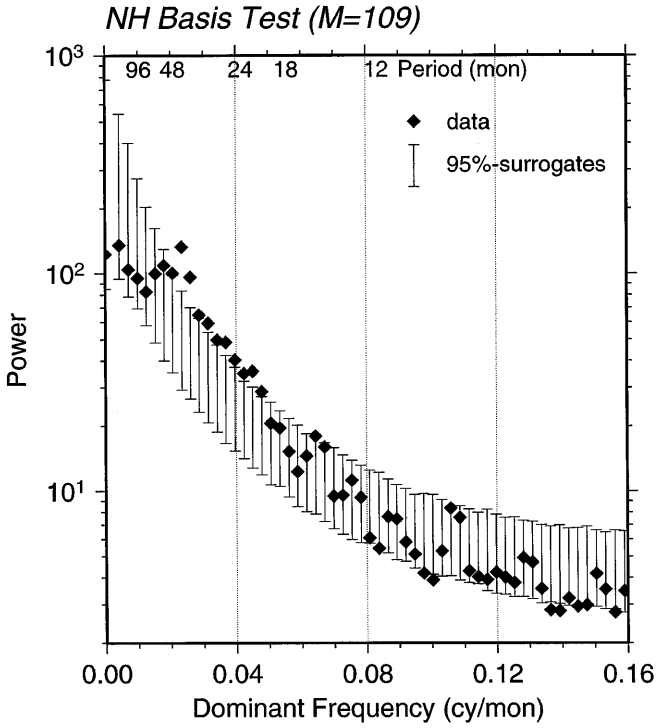

Fig. 4. As Fig. 3, using the null-hypothesis basis

ber of low-ranked elements of Λ_D' fall below the 2.5th percentiles. This is to be expected, since we observe anomalously high variance in Λ_D' in the QQ to QB frequency range, and the total expected variance of the surrogates is equal to the total variance in the data by construction.

We observe 6 excursions above the 97.5th percentiles in Fig. 4, which the binomial approximation indicates as $<0.5\%$ probable, as compared to 8% for 7 excursions with $N' = 148$. This is an example of how reducing the spectral resolution of the EEOF/MSSA algorithm, although it typically reduces the confidence level at which individual PCs are indicated as significant, may increase the overall confidence level at which we can reject a null-hypothesis without any prior information.

5 Discussion

A number of conclusions can be drawn from this simple demonstration.

5.1 Eigenvalue rank-order and "informal" tests are unreliable

The PCs corresponding to the QB mode of El Niño were ranked 13th and 14th in the eigenvalue rank-order in Fig. 1. Simply picking out the PCs with the largest associated eigenvalues is not valid as a method of separating signals from red noise. The main problem with the EEOF/MSSA algorithm, as previously implemented, is not that it may fail to detect very weak signals, but that it may appear to detect far too much. Stable, physical-looking sinusoidal PCs and EOFs are generated by pure noise and perfectly genuine signals may occur relatively low in the eigenvalue rank-order. The surrogate data test applied to the GISST data allowed us to pick out the two frequency ranges in which the null-hypothesis was clearly violated, and left us with the plausible conclusion that the rest of the variability in this dataset could be described as red noise.

5.2 Red noise is hard to beat

We chose, as our example, a dataset which we knew to display the strongest known mode of tropospheric interannual climate variability: El Niño. We did reject the red noise null-hypothesis, but not by a large margin. Perhaps the most remarkable aspect of Figs. 2 and 4 is not the excursions we see above the surrogate data bars, but the fact that, over the full Nyquist interval and over 1.5 orders of magnitude in variance, the fit to the red noise spectrum is extremely good. Even with the effects of artificial variance compression, the power in the QQ-mode PCs in the data-adaptive basis (Fig. 1) is only $\sim 50\%$ higher than we would expect on a pure red noise null-hypothesis: 2/3^{rds} of the variance in the data at these frequencies can be attributed to red noise.

We could, of course, have obtained a null-hypothesis which we would have rejected at a much higher confidence level if we had been less careful to ensure that

the expected variance and lag-1 autocorrelation of the surrogates were the same as those of the data. But rejection of a null-hypothesis with incorrectly specified parameters, at whatever confidence level, is a meaningless result. Given that even El Niño is a relatively close call against a correctly-specified red noise hypothesis, algorithms which indicate climate oscillations at much higher confidence levels should be investigated carefully to ensure that these levels are not just a trivial consequence of poor specification of the noise.

5.3 Rejecting the hypothesis of L independent AR(1) processes does not necessarily imply an oscillation

The algorithm outlined here allows us to test the hypothesis that the data input channels are consistent with L independent AR(1) processes. While rejection of this hypothesis is a necessary condition for us to conclude that EEOF/MSSA indicates an oscillation, it is not, in general, sufficient. For example, two or more of the L input channels may be significantly correlated at some lag other than zero for reasons other than an oscillation. If $N' < ML$ this will not directly affect Λ_D or the expected values of the Λ_R, since cross-channel covariance information is not used in $\mathbf{C}_D^{(P)}$ or $\mathbf{C}_R^{(P)}$. It would affect the second-order moments of the Λ_R, since such correlations would reduce the number of degrees of freedom of the system, but except in pathological cases such as one channel being a lagged copy of another (which should sorted our prior to the analysis), the impact on results should be minimal.

If $ML < N'$, then non-oscillatory cross-channel correlations at non-zero lags would affect Λ_D and the expected values of the Λ_R, since cross-channel covariances are used in computing $\mathbf{C}_D^{(E)}$ and $\mathbf{C}_R^{(E)}$. In this situation, rejecting the L-independent-AR(1)-process hypothesis is less conclusive with respect to the alternative hypothesis that the data exhibits an oscillation, although it remains a necessary condition. The test remains informative, however, since if information propagates consistently from one channel to another at some constant lag in time, it is worth looking for a physical explanation whether or not an oscillation is involved. If evaluating the strength of the evidence for an oscillation is of paramount importance, it will always be possible, through a different choice of window, to reformulate the problem such that $N' < ML$.

Summary

We have shown how the sliding window algorithm in EEOF/MSSA can generate oscillatory-looking patterns from pure noise, motivating the need for an objective hypothesis test in the detection and reconstruction of oscillatory signals. We have presented a simple test and demonstrated its application to a relatively uncontroversial geophysical problem. Rejection of the red noise null-hypothesis using this test should be considered a necessary condition for EEOF/MSSA to have detected an oscillation, although in certain situations non-oscillatory processes might also lead to rejection. The test described should provide a valuable filtering tool to ensure that investigative energy is not dissipated on developing oscillatory explanations for patterns which turn out to be attributable to red noise.

Acknowledgements. The immediate motivation for this study was an interesting discussion with Claude Frankignoul, Hans von Storch and Eduardo Zorita. We hope it helps. Comments from Lenny Smith, Pascal Yiou and an anonymous reviewer were also useful. This work was supported by a UCAR Visiting Scientist position with the NOAA Postdoctoral Program in Climate and Global Change and by the University of California's INCOR program and NOAA grant NA46GP0244 (AWR). We would also like to thank Richard Lindzen and Michael Ghil for support and encouragement.

References

Allen MR (1992) Interactions between the atmosphere and oceans on time scales of weeks to years. Ph.D. Thesis, University of Oxford

Allen MR, Smith LA (1994) Investigating the origins and significance of low-frequency modes of climate variability. Geophys Res Lett 21:883–886

Allen MR, Smith LA (1996) Monte Carlo SSA: detecting irregular oscillations in the presence of coloured noise. J Clim (in press)

Barnett TP, Hasselmann K (1979) Techniques of linear prediction, with application to oceanic and atmospheric fields in the Tropical Pacific. Rev Geophys Space Phys 17:949–968

Box GEP, Jenkins GM (1976) Time series analysis, forecasting and control. Holden Day, San Francisco

Broomhead DS, King G (1986) Extracting qualitative dynamics from experimental data. Physica D 20:217–236

Gibson JF, Farmer JD, Casdagli M, Eubank S (1992) An analytical approach to practical state space reconstruction. Physica D 57:1–30

Hasselmann K (1976) Stochastic climate models. Part I: theory. Tellus 28:473–485

Jiang N, Neelin JD, Ghil M (1995) Quasi-quadrennial and quasi-biennial variability in equatorial Pacific sea surface temperatures and winds. Clim Dyn 12:101–112

Jolliffe IT (1986) Principal component analysis. Springer, Berlin Heidelberg New York

Keppenne CL, Ghil M (1992) Adaptive spectral analysis and prediction of the Southern Oscillation Index. J Geophys Res 97:20449–20454

Keppenne CL, Ghil M (1993) Adaptive filtering and prediction of noisy multivariate signals: an application to subannual variability in atmospheric angular momentum. Int J Bifurcat Chaos 3:625–634

Livezey RE, Chen WY (1982) Statistical field significance and its determination by Monte Carlo techniques. Mon Weather Rev 111:46–57

Lorenz EN (1956) Empirical orthogonal functions and statistical weather prediction. Tech. Rep. 1. Statistical forecasting project, Department of Meteorology, Massachusetts Institute of Technology

Parker DE, Folland CK, Bevan A, Ward MN, Jackson M, Maskell K (1994) Marine surface data for analysis of climatic fluctuations on interannual to century time scales. In: Martinson DG, Bryan K, Ghil M, Hall MM, Karl TR, Sarachik ES, Soroostian S, Talley LF (eds), Natural climate variability on decade to century time scales. National Academy Press, Washington

Penland C, Sardeshmukh P (1995) Error and sensitivity analysis of geophysical eigensystems. J Clim 8:1988–1998

Plaut G, Vautard R (1994) Spells of low-frequency oscillations and weather regimes in the Northern Hemisphere. J Atmos Sci 51:211–220

Rasmusson EM, Wang X, Ropelewski CF (1990) The biennial component of ENSO variability. J Mar Syst 1:71–96

Read PL (1993) Phase portrait reconstruction using multivariate singular systems analysis. Physica D 69:353–365

Robertson AW (1996) Interdecadal variability over the North Pacific in a coupled ocean-atmosphere model. Clim Dyn (in press)

Robertson AW, Ma C-C, Mechoso CR, Ghil M (1995a) Simulation of the Tropical-Pacific climate with a coupled ocean-atmosphere general circulation model. Part II: interannual variability. J Clim 8:1199–1216

Robertson AW, Allen MR, Ghil M, Smith LA (1995b) Tests for distinguishing multivariate oscillatory behaviour from AR(1) noise. In: Proc. 19th Clim Diagn Workshop. CAC/NOAA, US Department of Commerce

Smith LA (1992) Identification and prediction of low-dimensional dynamics. Physica D 58:50–76

Thacker WC (1995) Statistical modeling for numerical modelers. Tech. Rep. GKSS 95/E/4. GKSS-Forschungszentrum Geesthacht GmbH

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 DJ (1990) Time series analysis of Holocene climate data. Philos Trans R Soc London A 330:601–616

Unal YS, Ghil M (1995) Interannual and interdecadal oscillation patterns in sea level. Clim Dyn 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

von Storch H (1982) A remark on Chervin-Schneider's algorithm to test significance. J Atmos Sci 39:187–189

von Storch H, Bürger G, Schnur R, von Storch J-S (1995) Principal oscillation patterns: a review. J Clim 8:377–400

Weare BC, Nastrom JN (1982) Examples of extended empirical orthogonal function analyses. Mon Weather Rev 110:481–485

Zorita E, Frankignoul C (1996) Modes of North Atlantic decadal variability in the ECHAM/LSG coupled ocean-atmosphere general circulation model, J Clim (in press)