ELSEVIER

13 October 1997

PHYSICS LETTERS A

Physics Letters A 234 (1997) 419–428

Optimal filtering in singular spectrum analysis

Myles R. Allen [a,b], Leonard A. Smith [c]

[a] *Space Science Department, Rutherford Appleton Laboratory, Chilton, Didcot, OX11 0QX, UK*

[b] *Department of Physics, University of Oxford, Oxford, UK*

[c] *Mathematical Institute, University of Oxford, Oxford, UK*

Received 17 December 1996; revised manuscript received 23 June 1997; accepted for publication 1 July 1997
Communicated by A.P. Fordy

Abstract

Singular spectrum analysis (SSA) provides a robust method of separating an arbitrary signal from "white" (independent, identically distributed) noise. In the presence of "coloured" noise, or any autocorrelated process, high-variance components of the noise can confuse the singular value decomposition, thereby obscuring genuine signals which are, in principle, detectable. A generalization of SSA is presented which yields both an optimal filter to discriminate against an arbitrary coloured noise and an objective method of quantifying uncertainty in signal reconstruction. The algorithm is applied to a simple synthetic signal-separation problem and used to resolve a degeneracy in the SSA of interannual and interdecadal variability of the Earth's global mean temperature. © 1997 Elsevier Science B.V.

1. Singular spectrum analysis

Singular Spectrum Analysis (SSA) is widely used to extract qualitative dynamics from noise-contaminated data [1–4]. The technique may be visualized as sliding a window of width M down a series of $\mathbf{d}$ of length N and determining the orthogonal patterns which best capture the variance in the views of the series thus obtained. These "empirical orthogonal functions" (EOFs) are eigenvectors of the $M \times M$ lag-covariance matrix $\mathbf{C}_D$,

$$(C_D)_{ij} = \frac{1}{k_{max}} D^T_{ik} D_{kj} = \frac{1}{k_{max}} \sum_{k=1}^{k_{max}} d_{i+k-1} d_{j+k-1}, \quad (1)$$

where [1] $k_{max} = N - M + 1$ and $\mathbf{D}$ is the "augmented" dataset, $D_{kj} = d_{j+k-1}$. For a stationary process, $\mathcal{E}((C_D)_{ij})$ is the process covariance at lag $|i-j|$.

If $\mathbf{d}$ consists of the sum of two ergodic, linearly independent processes (called, for convenience, "signal" and "noise") with sample lag-covariance matrices $\mathbf{C}_S$ and $\mathbf{C}_R$ then,

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

If the noise is "white" ($\mathcal{E}(\mathbf{C}_R \equiv \mathbf{C}_N = \sigma^2 \mathbf{I})$ and $\mathcal{E}(\mathbf{C}_S)$ is non-diagonal, then the eigenvectors of $\mathcal{E}(\mathbf{C}_D)$ are the eigenvectors of $\mathcal{E}(\mathbf{C}_S)$ with each eigenvalue increased by σ^2. The high-ranked EOFs of $\mathbf{C}_D$ (those with the largest eigenvalues) provide a consistent estimate of the high-ranked EOFs of $\mathbf{C}_S$: they converge to these "signal EOFs" as $N \to \infty$.

[1] We use the summation convention of Ref. [1], for reasons explained in Refs. [5,6].

PII S0375-9601(97)00559-8

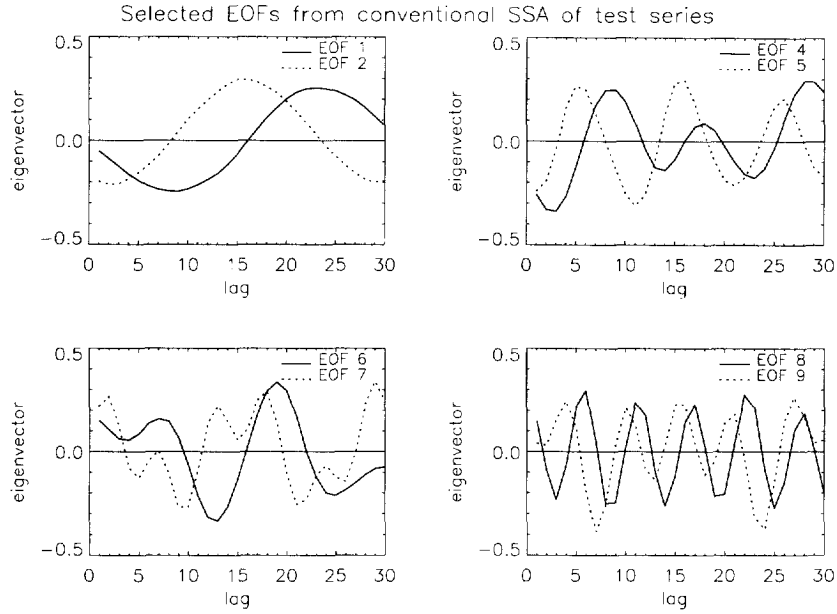

Fig. 1. Selected EOFs with high-ranked eigenvalues from conventional SSA of test series. EOFs 1 and 2 form a relatively pure sine-cosine pair, but are not associated with any signal, indicating the danger of using the properties of EOFs to identify oscillations. EOFs 8 and 9 contain most of the signal variance, but EOF 7 is close to degenerate with both of them, corrupting EOF 9.

A projection onto the κ highest-ranked data EOFs provides an optimal linear filter for this noise: the signal-to-noise variance ratio is maximized in the filtered augmented dataset

$$\mathbf{D}^{(\kappa)} \equiv \mathbf{D}\mathbf{E}_{\mathrm{D}}\mathbf{L}^{(\kappa)}\mathbf{E}_{\mathrm{D}}^{\mathrm{T}}, \tag{3}$$

where $\mathbf{E}_{\mathrm{D}}$ contains the data EOFs arranged column-wise in order of decreasing eigenvalue and $\mathbf{L}^{(\kappa)}$ is diagonal with $L_{ij}^{(\kappa)} = 1$ if $i \leqslant \kappa$, and $L_{ii}^{(\kappa)} = 0$ otherwise.

If the noise is anything other than white (if $\mathbf{C}_{\mathrm{N}} \neq \sigma^2\mathbf{I}$) then none of these optimality properties hold. High-ranked EOFs of $\mathbf{C}_{\mathrm{D}}$ need have nothing to do with the EOFs of $\mathbf{C}_{\mathrm{S}}$. In particular, sine-cosine EOF pairs, which for reasons given in Ref. [3], are frequently used to identify modulated oscillations, may occur in the absence of any genuine oscillatory signal [6,7]. A variety of signal selection criteria [4,8] have been proposed for SSA based on the expected properties of $\mathbf{C}_{\mathrm{S}}$ and $\mathbf{C}_{\mathrm{R}}$ in the pure-signal and pure-noise limits. Because the eigenbasis of the sum of two matrices does not, in general, share eigenvectors with either of the two constituent matrices, these criteria cannot be justified; the appearance of an EOF pair "corresponding" to an oscillation may depend as much on luck and a judicious choice of the window width as it depends on the signal.

Consider a "signal" which consists of stochastically-triggered sinusoidal bursts, with unit initial amplitude, period 5.5 units and decay-time 30 units, contaminated with unit variance first-order autoregressive – AR(1) – noise: the test series shown in Fig. 1 of Ref. [6]. Conventional SSA of this series with $M = 30$ gives only one sinusoidal EOF with a period equal to the oscillation (number 8 in Fig. 5 of Ref. [6]). The other EOF required for an optimal filter for this signal is scrambled with lower-frequency components of the noise: EOFs 7 and 9 (shown in Fig. 1) are almost equal, so the decomposition is undetermined to rotations within the subspace which they define. Both contain power at the signal frequency, so neither provides a consistent estimate of the dominant signal EOFs. How can this degeneracy be lifted?

2. Generalizing SSA to deal with coloured noise

Since the eigenvectors of $\mathbf{C}_{\mathrm{D}}$ are consistent estimators of the eigenvectors of $\mathbf{C}_{\mathrm{S}}$ if and only if the noise is white, we introduce a "pre-whitening" transforma-

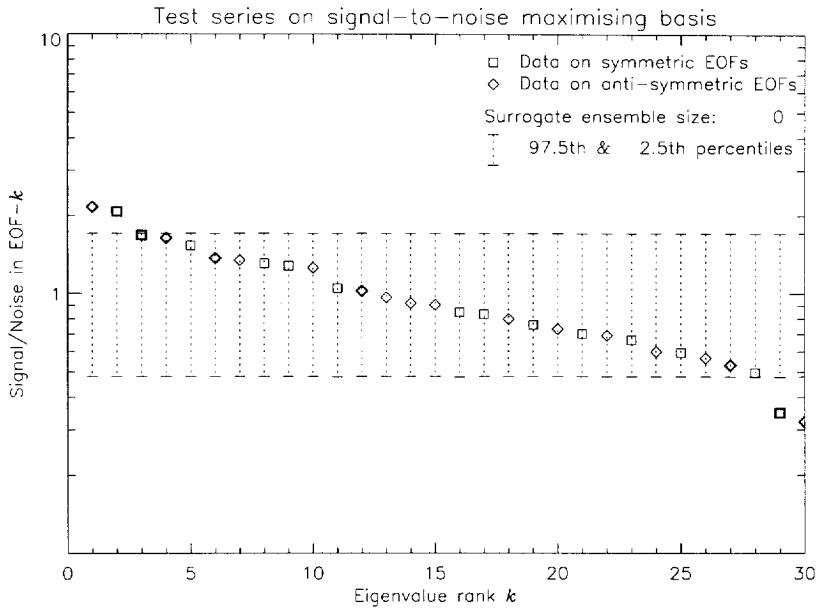

Fig. 2. Eigenspectrum of $\mathbf{C}'_{\mathrm{D}}$. Squares (diamonds) show signal-to-noise ratio in the test series on signal-to-noise maximizing EOFs, $\tilde{\mathbf{E}}_{\mathrm{D}}$, which are predominantly symmetric (anti-symmetric) about the mid-point of the window. Bold symbols indicate EOFs whose maximum correlation with a pure sinusoid is > 0.85. Vertical bars show the expected distribution of the noise variance in these eigendirections: a surrogate ensemble size of 0 indicates that the χ^2 parametric distribution has been used (see text).

tion, as in generalized regression or canonical (co-) variance analysis [9], and recently applied to spatial EOFs by Thacker [10,11]. The expected data variance to noise variance ratio (loosely termed the "signal-to-noise ratio") in the state-space direction defined by the vector $\boldsymbol{e}$ is

$$\rho \equiv \frac{\boldsymbol{e}^{\mathrm{T}}\mathbf{C}_{\mathrm{D}}\boldsymbol{e}}{\boldsymbol{e}^{\mathrm{T}}\mathbf{C}_{\mathrm{N}}\boldsymbol{e}}. \tag{4}$$

Given the noise covariance, $\mathbf{C}_{\mathrm{N}}$, is positive-definite, with eigenvalues forming the diagonal elements of $\mathbf{\Lambda}_{\mathrm{N}}$ and eigenvectors $\mathbf{E}_{\mathrm{N}}$, we define a coordinate transformation

$$\boldsymbol{e}' \equiv \mathbf{\Lambda}_{\mathrm{N}}^{1/2}\mathbf{E}_{\mathrm{N}}^{\mathrm{T}}\boldsymbol{e}, \quad \boldsymbol{e} \equiv \mathbf{E}_{\mathrm{N}}\mathbf{\Lambda}_{\mathrm{N}}^{-1/2}\boldsymbol{e}'. \tag{5}$$

In these transformed coordinates, the noise has equal variance in all directions, so

$$\rho = \frac{\boldsymbol{e}'^{\mathrm{T}}\mathbf{C}'_{\mathrm{D}}\boldsymbol{e}'}{\boldsymbol{e}'^{\mathrm{T}}\boldsymbol{e}'}, \tag{6}$$

where $\mathbf{C}'_{\mathrm{D}}$ and $\mathbf{C}'_{\mathrm{S}}$ are the transformed covariance matrices, defined thus,

$$\mathbf{C}' \equiv \mathbf{\Lambda}_{\mathrm{N}}^{1/2}\mathbf{E}_{\mathrm{N}}^{\mathrm{T}}\mathbf{C}\mathbf{E}_{\mathrm{N}}\mathbf{\Lambda}_{\mathrm{N}}^{-1/2}. \tag{7}$$

The vector $\boldsymbol{e}'$ which maximizes ρ in Eq. (6) is simply the eigenvector of $\mathbf{C}'_{\mathrm{D}}$ with the largest eigenvalue. By virtue of the coordinate transformation, $\mathcal{E}(\mathbf{C}'_{\mathrm{D}}) = \mathcal{E}(\mathbf{C}'_{\mathrm{S}}) + \mathbf{I}$, so the eigenvectors of $\mathbf{C}'_{\mathrm{D}}$ are consistent estimators of the eigenvectors of $\mathbf{C}'_{\mathrm{S}}$, in the sense defined above. Equating the $\boldsymbol{e}'$ with the eigenvectors of $\mathbf{C}'_{\mathrm{D}}$ (the columns of $\mathbf{E}'_{\mathrm{D}}$, arranged in order of decreasing eigenvalue) thus provides an optimal and consistent set of signal-to-noise maximizing vectors, with signal-to-noise ratios given by the eigenvalues

$$\mathbf{\Lambda}'_{\mathrm{D}} = \mathbf{E}'^{\mathrm{T}}_{\mathrm{D}}\mathbf{C}'_{\mathrm{D}}\mathbf{E}'_{\mathrm{D}}. \tag{8}$$

Fig. 2 shows the eigenspectrum of $\mathbf{C}'_{\mathrm{D}}$ for the test series. The first two eigenvalues contain more variance than expected in pure AR(1) noise, lying above the 99.5th percentiles of the distribution of power in these state-space directions expected from an AR(1) process [6]. Note that the noise has the same expected variance in all directions in these transformed coordinates, and we have neglected the selection effect (compression of variance into high-ranked EOFs) – see Ref. [6] for details.

To facilitate the interpretation, we transform back to our original coordinates,

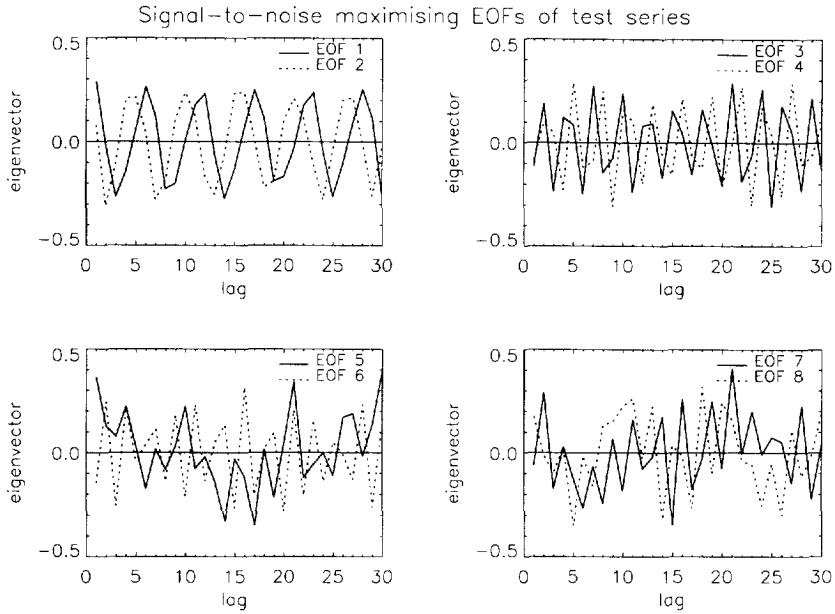

Fig. 3. Signal-to-noise maximizing EOFs of test series. EOFs 1 and 2 contain the 5.5-unit period signal and EOFs 3 and 4 now contain the first harmonic, which was almost completely obscured in conventional SSA.

$$\tilde{\mathbf{E}}_D \equiv \mathbf{E}_N \mathbf{\Lambda}_N^{-1/2} \mathbf{E}'_D. \tag{9}$$

Fig. 3 shows $\tilde{\mathbf{E}}_D$ for the test series, each renormalised for clarity. Note how the two highest-ranked EOFs are now reasonably "clean" sinusoids: their maximum correlations with a pure sinusoid are 0.96 and 0.93, while the corresponding values for the closest pair in Fig. 1 (EOFs 8 and 9) are 0.98 and 0.56.

The columns of $\tilde{\mathbf{E}}_D$ (the desired signal-to-noise maximizing patterns) are orthonormal in the metric defined by $\mathbf{C}_N$, rather than orthonormal in the conventional sense [10,12], that is $\tilde{\mathbf{E}}_D \tilde{\mathbf{E}}_D^T = \mathbf{C}_N^{-1}$ and $\tilde{\mathbf{E}}_D^T \mathbf{C}_N \tilde{\mathbf{E}}_D = \mathbf{I}$. We refer to these vectors as "signal-to-noise maximizing EOFs", although "noise" here represents any process we wish to discriminate against. The diagonal elements in the projection of a covariance matrix onto these S/N maximizing EOFs, $\mathbf{\Lambda}'_D = \tilde{\mathbf{E}}_D^T \mathbf{C}_D \tilde{\mathbf{E}}_D$, represent signal-to-noise variance ratios. These are proportional to the absolute variance if and only if the noise is white. Thus standard EOFs are a special case of S/N maximizing EOFs when $\mathbf{C}_N = \sigma^2 \mathbf{I}$.

The crucial property of S/N maximizing EOFs is that their expected orientation is independent of the noise variance, provided $\mathbf{C}_N$ correctly reflects noise autocorrelation. They thus provide a consistent estimate of the patterns which we would obtain by analysing the same signal in the absence of any noise, and an optimal linear filter for the reconstruction of that signal in the presence of noise with this correlation structure.

To confirm that S/N maximizing EOFs consistently provide an improved estimate of the true signal EOFs, we generate an ensemble of realisations of the test process by retaining the first 12 000 200-point segments in which at least one oscillatory burst occurs [2]. We then apply both conventional and S/N maximizing SSA to each segment and correlate the resulting EOFs with the two dominant EOFs of the underlying noise-free process. In each case, we select the two EOFs which are best correlated with these signal EOFs and record their rank in the eigendecomposition.

Fig. 4, left panel, shows that the EOFs from conventional SSA which correlate most closely with the pure-signal EOFs typically occur between 5 and 10 in the eigenvalue rank-order, and that their average correlation with the signal EOFs ranges between 0.6 and

[2] In a stochastic process, there is a chance that no burst will occur in a 200-point segment. Since the influence of bursts from previous segments is small, we would not expect any algorithm to distinguish most burst-free segments from pure noise.

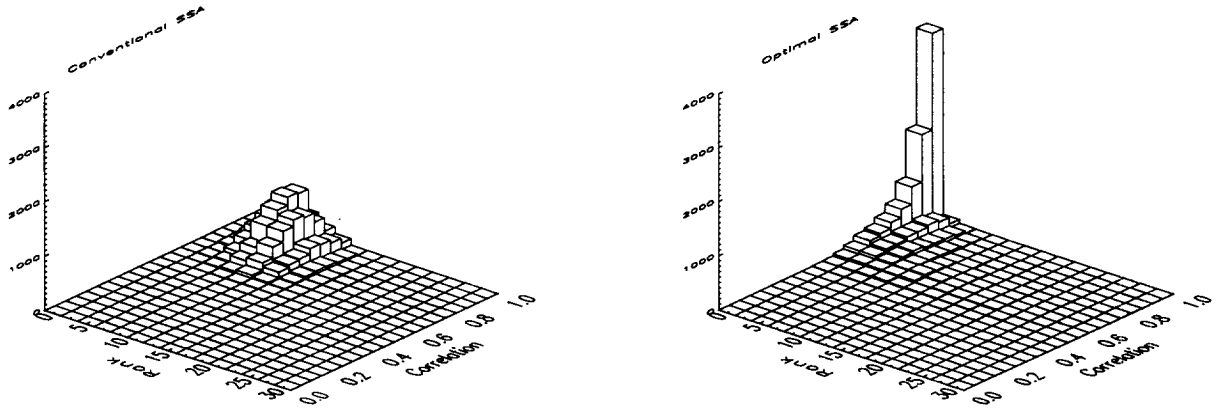

Fig. 4. Histogram showing how S/N maximizing SSA (right panel) outperforms conventional SSA (left panel) in identifying patterns which correspond to the true "signal" EOFs. In each member of an ensemble of 12 000 segments containing damped sinusoidal oscillations, the two EOFs which correlated best with the pure-signal EOFs were selected. Bars show the distribution of average rank (x-axis) and average correlation with the pure-signal EOFs (y-axis) of these EOF pairs. A perfect algorithm would show a single non-zero peak in the far corner of the plane, corresponding to a mean rank of 1–2 and a mean correlation of 0.95–1.0. S/N maximizing SSA clearly comes much closer to this ideal result.

1.0, with the most probable correlation being ~ 0.8. The fact that EOFs occur relatively low in the rank-order means that some additional test will be required to identify them, and even if they are correctly identified, a relatively low correlation with the signal EOFs means that they will not provide an effective filter for separating signal from noise. In contrast, Fig. 4, right panel, shows that the S/N maximizing EOFs which correlate best with the pure-signal EOFs are almost invariably ranked at the top of the eigendecomposition and in the majority of cases the mean correlation with the signal EOFs is close to 1.0. With this revised approach to SSA, the chance of obtaining an effective filter for separating signal from noise in the highest-ranked EOFs is significantly enhanced.

Fig. 4 shows that the uncertainty in the orientation of the two highest-ranked S/N maximizing EOFs is low (they almost always point in approximately the right direction), although the uncertainty in the noise variance in individual state-space directions (indicated by the vertical bars in Fig. 2) is relatively high. A number of authors (e.g. Ref. [13]) have interpreted similar "error bars" as indicating the uncertainty in the eigendecomposition itself. This is incorrect: the correct treatment of uncertainty in the orientation of the individual EOFs must be based on the fact that $\mathbf{C}_D$ conforms to a Wishart distribution, details of which we will consider elsewhere.

When the properties of the noise are unknown, the "correct" specification of $\mathbf{C}_N$ remains a problem. Several observations are relevant here. First, unlike conventional EOFs, the expected orientation of S/N maximizing EOFs is independent of the total noise variance which can therefore be grossly in error without affecting the procedure. Second, for high signal-to-noise patterns, errors in the noise autocorrelation structure do not have a significant effect on the orientation of the highest-ranked S/N maximizing EOFs, due to the nonlinearity of the eigendecomposition procedure: in this example, we can increase and decrease the characteristic decorrelation time of the noise by a factor of two without affecting the shape of highest-ranked S/N maximizing EOFs. Significance estimates are, of course, affected, so standard SSA supplemented by a significance test is a poor substitute for S/N maximizing SSA if the noise properties are in doubt. Third, the specification of $\mathbf{C}_N$ may be an integral part of the problem to be solved. For example, if we are looking for patterns in an observational dataset which are inconsistent with "natural climate variability" as simulated by a climate model, then $\mathbf{C}_N$ may be computed directly from a control run of the model following the procedure of optimal fingerprinting [14]. Finally, it may be possible to frame the detection problem in terms of identifying and characterising a *change* in the dynamics of the system [15,16]. This is a well-

established approach to fault detection and diagnosis in mechanical systems – see Ref. [17], and references therein. In this case, $\mathbf{C}_N$ is estimated from the historical "normal" system behaviour.

Since the eigendecomposition in Eq. (8) ranks eigenvectors by signal-to-noise rather than by variance, problems of degeneracy between EOFs corresponding to the "signal" and high-variance, low-frequency components of the noise do not arise. Different signals with equal signal-to-noise ratios may, however, be degenerate. While such degeneracies are unimportant in constructing an optimal filter, they complicate signal separation. If the noise is red, however, then components which have the same signal-to-noise ratio are likely to have different variances, resolving any degeneracies. We construct the filtered transformed covariance matrix

$$\mathbf{C}_D'^{(\kappa)} \equiv \mathbf{E}_D' \mathbf{\Lambda}_D' \mathbf{L}^{(\kappa)} \mathbf{E}_D'^T, \tag{10}$$

with $\mathbf{L}^{(\kappa)}$ defined as above to extract the κ highest signal-to-noise components; back-transform to obtain a rank-κ covariance matrix, which we re-diagonalize,

$$\mathbf{C}_D^{(\kappa)} = \mathbf{E}_N \mathbf{\Lambda}_N^{1/2} \mathbf{C}_D'^{(\kappa)} \mathbf{\Lambda}_N^{1/2} \mathbf{E}_n^T \tag{11}$$

$$= \mathbf{E}_D^{(\kappa)} \mathbf{\Lambda}_D^{(\kappa)} \mathbf{E}_D^{(\kappa)T}. \tag{12}$$

This represents a variance-maximizing rotation, since the κ elements of $\mathbf{E}_D^{(\kappa)}$ with non-zero eigenvalues span the subspace defined by the κ highest-ranked S/N maximizing EOFs, with the variance in each given by the corresponding eigenvalue (diagonal element of $\mathbf{\Lambda}_D^{(\kappa)}$). These rotated S/N maximizing EOFs are orthonormal in the conventional Euclidean metric. This two-stage procedure is reminiscent of rotated principal component analysis (RPCA [18]), except that filtering is performed on the basis of signal-to-noise, rather than on the basis of variance. RPCA will break down if too many EOFs need to be retained in order to include important, but low-variance, components [19]. The procedure outlined here identifies components directly and has the additional advantage of consistency: the expected orientation of the $\mathbf{E}_D^{(\kappa)}$ is independent of the noise variance. This is not the case in RPCA.

3. Quantifying uncertainty in signal-reconstruction

The basic assumption underlying SSA is that each row of the augmented dataset $\mathbf{D}$ (each "view through the window") can be well described by the linear sum of a small number ($\kappa < M$) of patterns, denoted by the columns of $\mathbf{E}^{(\kappa)}$, each modulated by a time series of coefficients ("principal components", or PCs), denoted by the columns of $\mathbf{P}^{(\kappa)}$,

$$\mathbf{D} = \mathbf{P}^{(\kappa)} \mathbf{E}^{(\kappa)T} + \text{noise}. \tag{13}$$

Given the noise covariance, $\mathbf{C}_N$, the best linear unbiased (BLUE) estimators for the elements of $\mathbf{P}^{(\kappa)}$ are given by the generalized linear regression formulae [9],

$$\tilde{\mathbf{P}}^{(\kappa)} \equiv \mathbf{D} \mathbf{C}_N^{-1} \mathbf{E}^{(\kappa)} (\mathbf{E}^{(\kappa)T} \mathbf{C}_N^{-1} \mathbf{E}^{(\kappa)})^{-1}. \tag{14}$$

The time series of coefficients obtained in this way are called generalized PCs [10].

The estimators $\tilde{\mathbf{P}}^{(\kappa)}$ are *blue* irrespective of the normalisation or orthogonality of the $\mathbf{E}^{(\kappa)}$, allowing a complete separation between the procedure for deriving the $\mathbf{E}^{(\kappa)}$ and the estimation of the elements of $\mathbf{P}^{(\kappa)}$. Eq. (14) may be used for a signal-reconstruction based on standard EOFs or the S/N maximizing EOFs described here, the only condition being that no two EOFs are collinear. If S/N maximizing EOFs are used, Eq. (13) yields generalized PCs identical to those proposed by Ref. [10]. We believe there is a gain in simplicity in separating the EOF-estimation step and the PC-reconstruction step, particularly given prior information to constrain the EOFs, but in the absence of such information the two approaches are essentially equivalent.

In the long series limit ($N \gg M$), the variance (uncertainty) in the PCs is given by

$$\mathcal{E}((\tilde{\mathbf{P}}^{(\kappa)} - \mathbf{P}^{(\kappa)})_{ik_1} (\tilde{\mathbf{P}}^{(\kappa)} - \mathbf{P}^{(\kappa)})_{ik_2}) = (\mathbf{E}^{(\kappa)T} \mathbf{C}_N^{-1} \mathbf{E}^{(\kappa)})^{-1}_{k_1 k_2}. \tag{15}$$

With a short series, Eq. (15) will be biased due to sampling uncertainty in $\mathbf{E}^{(\kappa)}$. This bias can be reduced by using independent data to estimate $\mathbf{E}^{(\kappa)}$ and $\tilde{\mathbf{P}}^{(\kappa)}$, but even a moderately biased error-estimate remains useful, since the band-pass filtering effect of SSA can give deceptively regular and physical-looking PCs even in

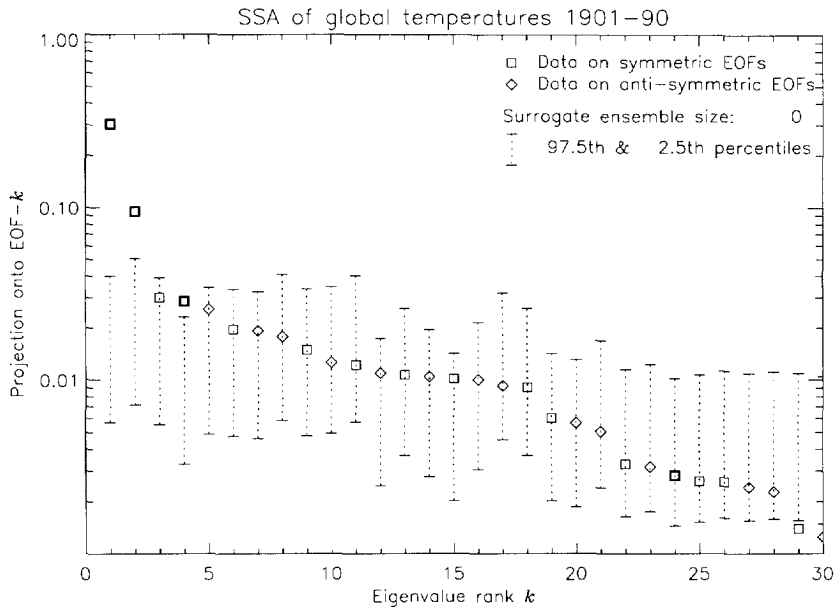

Fig. 5. Conventional SSA of 1901–90 global temperature record, with noise distributions computed for an AR(1) process after EOFs 1 and 2 (corresponding to the trend) have been eliminated from the noise-parameter-estimation. Note how EOFs 3–5 and 6–8 form two, almost completely degenerate, triplets. The noise distribution bars indicate EOF 4 as significant (with associated period 5 years), but the other member of the pair is scrambled with a low-frequency component of the noise.

the absence of any signal. Since the PCs are a form of weighted moving average of the original data series, erroneous excursions will, by construction, persist over a time scale defined by the window width M. An objective measure of uncertainty in PCs is therefore essential, particularly if they are to be used for empirical prediction [4,20].

Because the columns of $\mathbf{E}^{(\kappa)}$ need not be mutually orthogonal, Eqs. (14) and (15) also provide a natural method of estimating PCs and associated uncertainties in the presence of data gaps. A further application would be to employ the S/N maximizing EOFs to form a natural set of bases of increasing dimension, each consisting of the κ highest signal-to-noise components, from which to construct nonlinear inverses to linear filters, as in Ref. [21], and in other nonlinear noise-reduction algorithms.

4. Resolving degeneracies in the analysis of climate time series

In the exchange which originally focused attention on the naïve application of SSA to signal detection in climate research, it was observed [22] that the EOF pair advanced as evidence for an interdecadal oscillation in global temperatures [23] was unstable to minor changes in the length of the series analysed (or, indeed, to the summation convention used in $\mathbf{C}_D$). The problem was one of degeneracy [24], the power spectral density at interdecadal periods is close to that in the low-frequency ($\sim$ 5-year) component of the El Niño/Southern Oscillation (ENSO) signal. Thus the eigendecomposition is underdetermined to a rotation, making results extremely unstable. Such degeneracies have nothing to do with the existence of deterministic low-frequency components in the generating process. Statistical tests show that the interdecadal component of global temperatures is not distinguishable from AR(1) noise, while the El Niño/Southern Oscillation (ENSO) signal is only significant at around the 90% level [25].

The analysis of the 1901–90 global temperature record [26,27] also provides an interesting example of how standard SSA can fail to distinguish a genuine, albeit weak, signal (ENSO) from high-variance, low-frequency noise (the interdecadal component): this shortcoming is corrected through the application of the

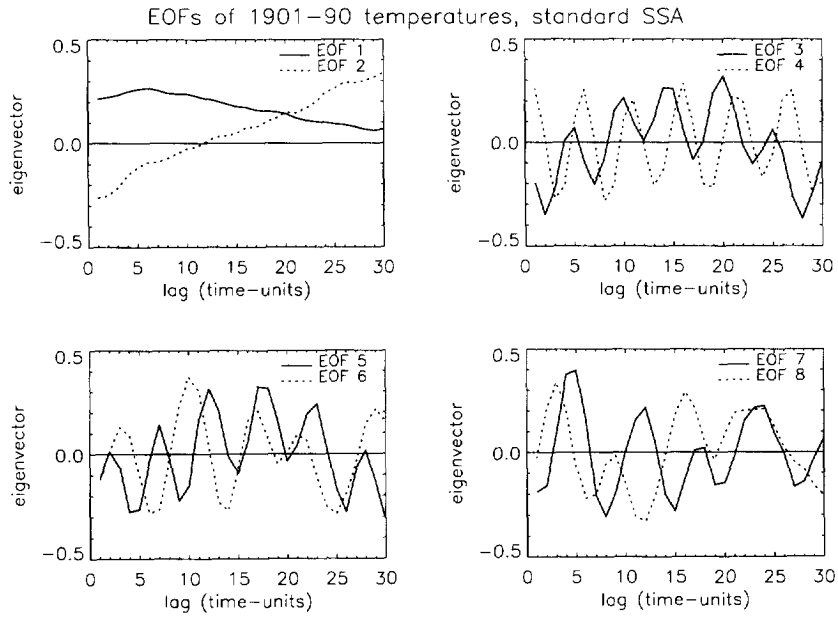

Fig. 6. EOFs 1–8 from conventional SSA of 1901–90 global temperature record. EOFs 3–8 all contain power at 5-year and (much less obviously) interdecadal periods, but because they are scrambled due to degeneracy, only EOF 4 is anything like a pure sinusoid.

technique presented in this paper. Fig. 5 shows conventional SSA of 1901–90 global temperatures [28]. At 5-year periods, only one eigenvalue (number four in the rank-order) appears above the 97.5th percentile of the χ^2 AR(1) noise distributions (unlike Fig. 2, these distributions are not identical, since the expected red-noise variance in a particular EOF depends on the dominant frequencies associated with that EOF). The other member of the pair is scrambled with the interdecadal component of the noise, so no sine-cosine EOF pair is observed at either frequency. The eight highest-ranked EOFs are shown in Fig. 6.

The noise in Fig. 5 was estimated by eliminating the trend (EOFs 1 and 2 from conventional SSA, which are significant against red noise) and fitting the AR parameters to the lag-0 and lag-1 covariances of the remainder [6]. We now use this noise model to compute S/N maximizing EOFs as described above. Because we are obliged to estimate the noise model from the data, the possibility remains that it may be incorrectly specified. We cannot, however, reject the hypothesis of a trend-plus-AR(1)-noise for this data, so the procedure is internally consistent. In contrast, we can reject the hypothesis that the residuals are white noise at a very high confidence level. The use of standard SSA, which assumes white noise, in such an analysis is therefore incoherent.

Ratios of data variance to expected noise variance on the S/N maximizing basis $\tilde{\mathbf{E}}_D$ (eigenvalues Λ'_D of $\mathbf{C}'_D$) are shown in Fig. 7. Five ratios appear above the 97.5th percentiles of the noise distribution, but we know that some (if not all) of these excursions may be due to artificial compression of variance into high-ranked EOFs. A more conservative test, based on the EOFs of the null-hypothesis [6], still indicates that S/N maximising EOFs 1–4 are significant at the 97.5% level.

The first four S/N maximizing EOFs are plotted in the upper two panels of Fig. 8. They are clearly dominated by the secular trend and the 5-year ENSO signal, but we also note that there is some cross-contamination between them since, ranked on the basis of signal-to-noise rather than variance, these two components are close to degenerate with each other. EOF-2 in particular contains power at 5-year time scales. To separate out the two components, we apply a variance-maximizing rotation as given by Eqs. (10)–(12) with $\kappa = 4$. The rotated EOFs are shown in the lower two panels of Fig. 8: a much clearer signal-separation has been achieved. The problem of degeneracy between

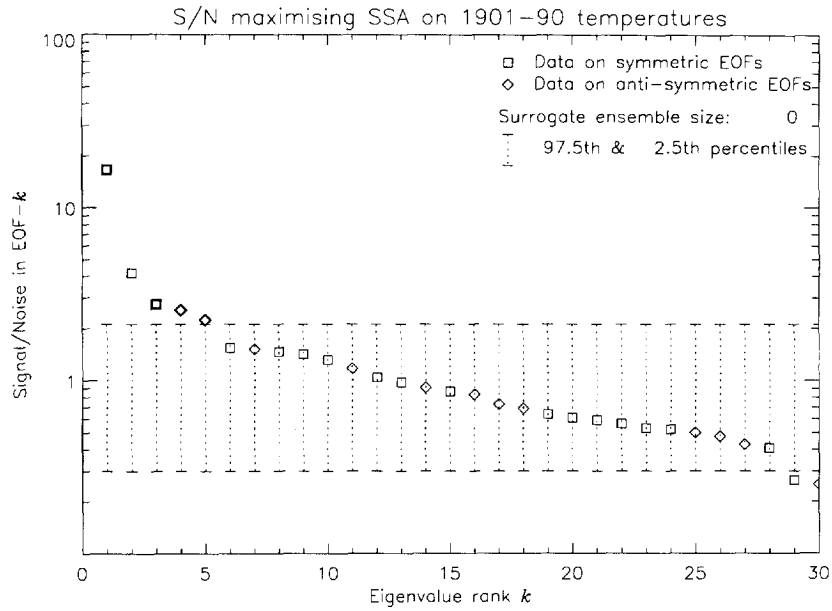

Fig. 7. Eigenspectrum of C'_D for the global temperature series, plotted against signal-to-noise in the corresponding S/N-maximizing EOF. Vertical bars show the 2.5th and 97.5th percentiles of the distribution of power expected from the noise component of a trend-plus-AR(1)-noise null-hypothesis.

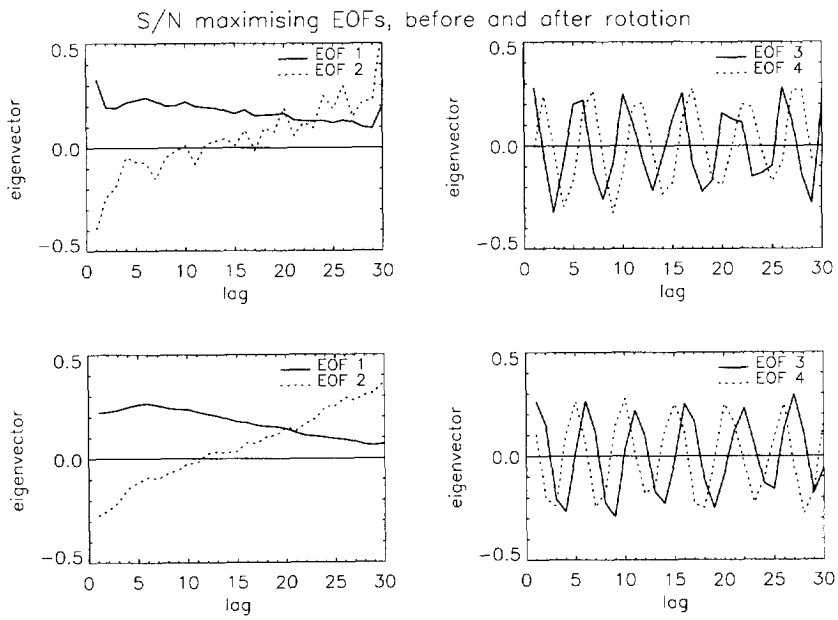

Fig. 8. Highest-ranked 4 S/N maximizing EOFs, before (upper panels) and after (lower) a variance-maximizing rotation separates the 5-year ENSO signal from the trend.

the 5-year component of ENSO and the interdecadal component of the noise is completely resolved. S/N maximising SSA has no trouble distinguishing between them because the signal-to-noise ratio is very different at these two frequencies.

Since the interdecadal component is indistinguishable from noise, empirical prediction of interdecadal temperature variations [4,20] cannot be justified. Even if we had an a priori reason to believe in a deterministic interdecadal oscillation, application of Eq. (15) reveals that the phase of this component cannot be determined from the data at any point in the 136-year global temperature series (or over any but the earliest years of the 335-year Central England Temperature series [29]). It is therefore unrealistic to attempt SSA-based decadal prediction of these data, since such forecasts depend entirely on the estimated phase near the current end of the series.

5. Summary

We have presented a simple generalisation of SSA, allowing a self-consistent treatment of problems involving autocorrelated noise by determining patterns (EOFs) which maximize the signal-to-noise ratio rather than maximizing variance. Our approach resolves ambiguities which arise due to degeneracies between high-frequency signals and low-frequency components of the noise. It may, however, introduce degeneracies between different signals with similar signal-to-noise ratios. These are resolved in a two-stage procedure, filtering first by signal-to-noise, then by variance. Noting that an SSA based reconstruction from a limited number of EOFs can appear deceptively "physical" even in the absence of any signal, we have also proposed a method of quantifying uncertainty in such reconstructions. This is essential if SSA is to be used for empirical prediction.

Acknowledgement

We would like to thank C. Thacker for helpful and stimulating exchanges and for providing prepublication material. M.R.A. was supported as a Visiting Scientist on the NOAA Postdoctoral Program in Climate and Global Change at MIT, and by the UK NERC Ocean Dynamics SLA at RAL; L.A.S. by a Senior Research Fellowship from Pembroke College, Oxford, and Guest Professorship in Innovationskolleg Formale Modelle Kognitiver Komplexität of the University of Potsdam.

References

[1] D.S. Broomhead and G. King, Physica D 20 (1986) 217.
[2] K. Fraedrich, J. Atmos. Sci. 43 (1986) 419.
[3] R. Vautard and M. Ghil, Physica D 35 (1989) 395.
[4] R. Vautard, P. Yiou and M. Ghil, Physica D 58 (1992) 95.
[5] M.R. Allen, Interactions between the atmosphere and oceans on time-scales of weeks to years, Ph.D. thesis, University of Oxford, 1992.
[6] M.R. Allen and L.A. Smith, J. Climate 9 (1996) 3373.
[7] M.R. Allen and A.W. Robertson, Climate Dynamics 12 (1996) 775.
[8] M.D. Dettinger, M. Ghil, C.M. Strong, W. Weibel and P. Yiou, EOS, Trans. Am. Geophys. Union 76 (1995) 12.
[9] K. Mardia, J. Kent and J. Bibbiy, Multivariate analysis (Academic Press, New York, 1979).
[10] W.C. Thacker, Tellus 48 (1996) 584.
[11] W.C. Thacker and R. Lewandowicz, J. Climate 9 (1996) 1942.
[12] D.B. Stephenson, Tellus, to appear (1997).
[13] M. Ghil and K.-C. Mo, J. Atmos. Sci. 48 (1991) 752.
[14] K. Hasselmann, J. Climate 6 (1993) 1957.
[15] L.A. Smith, K. Godfrey, P. Fox and K. Warwick, in Control 91, Vol. 1 of IEEE Conf. Publication 332, p. 1062, March (1991).
[16] W.G. Früh and M.R. Allen (1997) in preparation.
[17] Q. Zhang, M. Brasseville and A. Beneviste, Automatica 30 (1994) 95.
[18] M.B. Richman, J. Climatology 6 (1986) 293.
[19] E.A. O'Lenic and R.E. Livesey, Mon. Wea. Rev. 116 (1988) 1682.
[20] G. Plaut, M. Ghil and R. Vautard, Science 268 (1995) 710.
[21] D.S. Broomhead, J.P. Huke and M.A.S. Potts, Physica D 89 (1996) 439.
[22] J.B. Elsner and A.A. Tsonis, Nature (London) 353 (1991) 551.
[23] M. Ghil and R. Vautard, Nature (London) 350 (1991) 324.
[24] M.R. Allen, P.L. Read and L.A. Smith, Nature (London) 355 (1992) 686.
[25] M.R. Allen and L.A. Smith, Geophys. Res. Lett. 21 (1994) 883.
[26] A.A. Tsonis and J.B. Elsner, Nature (London) 356 (1992) 751.
[27] M.R. Allen, P.L. Read and L.A. Smith, Nature (London) 359 (1992) 679.
[28] C.K. Folland et al., Observed climate variability and change, in J.T. Houghton et al. eds., Climate change 1992, chapter C (Cambridge University Press, Cambridge, 1992).
[29] D.E. Parker, T.P. Legg and C.K. Folland, Technical Report CTRN11, Hadley Centre, Meteorological Office, London Road, Bracknell, RG12 2SY, UK (1991).